

การเปรียบเทียบการประมาณค่าพารามิเตอร์ระหว่างวิธีกำลังสองน้อยสุดเชิงส่วน
และวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์
สำหรับตัวแบบความถดถอยพหุเชิงเส้นที่เกิดพหุสัมพันธ์

นางสาวเด่นนภา จุลเพชร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2555
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the Graduate School.

A COMPARISON OF PARAMETERS ESTIMATION AMONG PARTIAL LEAST SQUARES
AND ORDINARY LEAST SQUARES METHODS
USING GRAM-SCHMIDT'S DATA TRANSFORMATION FOR MULTIPLE LINEAR
REGRESSION MODEL WITH MULTICOLLINEARITY

Miss Dennapa Joollapet

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2012

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

การเปรียบเทียบการประมาณค่าพารามิเตอร์ระหว่างวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์ สำหรับตัวแบบความถดถอยพหุเชิงเส้นที่เกิดพหุสัมพันธ์

โดย

นางสาวเด่นนภา จุลเพชร

สาขาวิชา

สถิติ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้
วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบริหารธุรกิจ

.....คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.กัลยา วาณิชย์บัญชา)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

..... กรรมการ
(อาจารย์ ดร. อนุภาพ สมบูรณ์สวัสดิ์)

..... กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.มนต์ทิพย์ เทียนสุวรรณ)

เด่นนภา จุลเพชร: การเปรียบเทียบการประมาณค่าพารามิเตอร์ระหว่างวิธีกำลังสองน้อยสุดเชิงส่วน และวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์ สำหรับตัวแบบความถดถอยพหุเชิงเส้นที่เกิดพหุสัมพันธ์ (A COMPARISON OF PARAMETERS ESTIMATION AMONG PARTIAL LEAST SQUARES AND ORDINARY LEAST SQUARES METHODS USING GRAM-SCHMIDT'S DATA TRANSFORMATION FOR MULTIPLE LINEAR REGRESSION MODEL WITH MULTICOLLINEARITY) อ. ที่ปริภาวิทยานิพนธ์หลัก: รศ. ดร.สุพล ดุรงค์วัฒนา, 110 หน้า

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประมาณค่าพารามิเตอร์ที่เหมาะสม โดยการเปรียบเทียบจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ของตัวประมาณ สำหรับตัวแบบความถดถอยพหุเชิงเส้นของวิธีกำลังสองน้อยสุดเชิงส่วน (PLS) และวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์ (OLS_G) เมื่อเกิดพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ ซึ่งทำการศึกษาภายใต้ตัวแปรอิสระมีการแจกแจงแบบปกติค่าเฉลี่ยเท่ากับ 0 ค่าส่วนเบี่ยงเบนมาตรฐานของตัวแปรอิสระเท่ากับ 2 และค่าความแปรปรวนของความคลาดเคลื่อนเท่ากับ 10 โดยมีจำนวนตัวแปรอิสระ (p) เท่ากับ 2 และ 3 ตัวแปร มีขนาดตัวอย่าง 50, 100 และ 200 มีค่าสัมประสิทธิ์ความถดถอยเบื้องต้น $\beta_0 = \beta_1 = \dots = \beta_p = 1$ และระดับความสัมพันธ์ระหว่างตัวแปรอิสระแบ่งเป็น 2 กรณี คือ กรณีที่มีตัวแปรอิสระ 2 ตัวแปร มีระดับความสัมพันธ์ (ρ) เป็น 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99 ส่วนกรณีที่มีตัวแปรอิสระ 3 ตัวแปร มีระดับความสัมพันธ์ (ρ) เป็น 0.91, 0.93, 0.95, 0.97 และ 0.99 ซึ่งในการศึกษาจะทำซ้ำจำนวน 1000 รอบ

ผลการศึกษาพบว่ากรณี p เท่ากับ 2 ตัวแปร จะเกิดปัญหาพหุสัมพันธ์รุนแรงที่ ρ ตั้งแต่ 0.95 ขึ้นไป และกรณี p เท่ากับ 3 ตัวแปร ที่ n เท่ากับ 50 และ 100 จะเกิดปัญหาพหุสัมพันธ์รุนแรงที่ ρ ตั้งแต่ 0.93 ขึ้นไป ส่วนที่ n เท่ากับ 200 จะเกิดปัญหาพหุสัมพันธ์รุนแรงที่ ρ ตั้งแต่ 0.95 ขึ้นไป และทั้งสองกรณีมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงสูง ซึ่งส่งผลกระทบต่อค่าประมาณค่าสัมประสิทธิ์ความถดถอย ดังนั้นกรณี p เท่ากับ 2 ตัวแปร ที่ n เท่ากับ 50, 100 และ 200 ตามลำดับ และที่ ρ เท่ากับ 0.9-0.94, 0.9-0.97 และ 0.9-0.98 ตามลำดับ วิธี OLS_G มีประสิทธิภาพดีกว่าวิธี PLS แต่ที่ ρ เท่ากับ 0.95-0.99, 0.98-0.99 และ 0.99 ตามลำดับ วิธี PLS มีประสิทธิภาพดีกว่าวิธี OLS_G และกรณี p เท่ากับ 3 ตัวแปร วิธี PLS มีประสิทธิภาพดีกว่าวิธี OLS_G ทุกกรณีที่ทำการศึกษา โดยปัจจัยที่มีผลต่อค่า MSE ของทั้งสองกรณีนั้นพบว่าเมื่อ ρ เพิ่มขึ้น ค่า MSE ของวิธี PLS ไม่มีแนวโน้มที่แน่นอน ยกเว้นที่ p เท่ากับ 3 และ n เท่ากับ 200 ค่า MSE จะมีค่าลดลงเล็กน้อย ในขณะที่วิธี OLS_G จะมีค่าเพิ่มขึ้น และถ้า n เพิ่มขึ้น ค่า MSE ของทั้งสองวิธีจะมีค่าลดลง

ภาควิชา.....สถิติ..... ลายมือชื่อนิสิต.....
 สาขาวิชา.....สถิติ..... ลายมือชื่อ อ.ที่ปริภาวิทยานิพนธ์หลัก.....
 ปีการศึกษา.....2555.....

5381795226: MAJOR STATISTICS

KEYWORDS: Partial Least Squares Method/ Ordinary Least Squares Method/ Gram-Schmidt Method/ Multicollinearity

DENNAPA JOOLLAPET: A COMPARISON OF PARAMETERS ESTIMATION AMONG PARTIAL LEAST SQUARES AND ORDINARY LEAST SQUARES METHODS USING GRAM-SCHMIDT'S DATA TRANSFORMATION FOR MULTIPLE LINEAR REGRESSION MODEL WITH MULTICOLLINEARITY. ADVISOR: ASSOC.PROF. SUPOL DURONGWATANA, Ph.D., 110 pp.

The objective of this research is to estimate optimal parameters by comparison from the mean square error (MSE) of the estimate for multiple linear regression models among Partial Least Square (PLS) and Ordinary Least Square Methods by using Gram-Schmidt's data transformation (OLS_G) to solve multicollinearity violence problem of the independent variables. This study focuses on Normal distribution of the independent variables with $\mu=0$, $\sigma_x=2$ and $\sigma^2=10$. We will study on 2 and 3 independent variables (p) under the following condition; the sample size 50, 100 and 200, the initial regression coefficient is $\beta_0 = \beta_1 = \dots = \beta_p = 1$. For case 2 independent variables, the multicollinearities (p) are defined to be 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, and for the other case, 3 independent variables, p's are defined to be 0.91, 0.93, 0.95, 0.97 and 0.99. We will generate each case with 1,000 simulation runs.

In conclusion, is founded that problems of multicollinearity violence occurs when p is more than 0.95 for case 2 independent variables, and these problems depend on number of sample size for case 3 independent variables. In particular, for the latter case the problems occurs when p is more than 0.93 for sample size 50 and 100, and more than 0.95 for sample size 200. In both cases, the problems of multicollinearity violence will be occurred in high chance. Therefore, in case of p=2 at the sample size 50,100 and 200 in order and p is 0.9-0.94, 0.9-0.97 and 0.9-0.98 in order the OLS_G method is more effective than the PLS method. But, p is 0.95-0.99, 0.98-0.99 and 0.99 in order the PLS method is more effective than the OLS_G method. And in case of p=3 the PLS method is more effective than the OLS_G method. The factors affecting the MSE when p increases the MSE of the PLS method no definite trend exception p=3 and the sample size 200 the MSE decreased slightly while the OLS_G method increases. And the sample size increases the MSE of 2 cases decreases.

Department:.....Statistics..... Student's Signature.....
Field of Study:.....Statistics..... Advisor's Signature.....
Academic Year:.....2012.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลือจากอาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.สุพล คุรงค์วัฒนา ซึ่งท่านได้ให้คำปรึกษาและแนะนำตลอดจนแก้ไขข้อบกพร่องต่างๆ เป็นอย่างดีมาโดยตลอด ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้ ทั้งผู้วิจัยขอกราบขอบพระคุณท่านประธานในการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา ท่านคณะกรรมการ อาจารย์ ดร.อนุภาพ สมบูรณ์สวัสดิ์ และท่านกรรมการภายนอก รองศาสตราจารย์ ดร. มนต์ทิพย์ เทียนสุวรรณ เป็นอย่างสูงที่ท่านอาจารย์ทั้งสามท่านได้กรุณา สละเวลามาสอบและให้คำแนะนำที่ดีและมีประโยชน์ในการปรับปรุงงานของผู้วิจัยต่อไป ตลอดจน คณาจารย์ในภาควิชาสถิติทุกท่านที่ได้ประสิทธิประสาทวิชาความรู้ให้แก่ผู้วิจัย

สุดท้ายนี้ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ ที่ส่งเสริมและสนับสนุนด้านการศึกษาของผู้วิจัย และขอขอบพระคุณเพื่อนๆ พี่ๆ น้องๆ ทุกคนที่ให้ความช่วยเหลือและเป็นกำลังใจแก่ผู้วิจัยเสมอมา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฌ
สารบัญรูปภาพ.....	ฎ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการศึกษา.....	2
1.3 ขอบเขตของการศึกษา.....	2
1.4 ข้อตกลงเบื้องต้น.....	5
1.5 คำจำกัดความที่ใช้ในการศึกษา.....	5
1.6 วิธีดำเนินการศึกษา.....	6
1.7 ประโยชน์ที่คาดว่าจะได้รับ.....	7
บทที่ 2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง.....	9
2.1 การวิเคราะห์ความถดถอย.....	9
2.2 ตัวแบบและการวิเคราะห์ความถดถอยพหุเชิงเส้น.....	9
2.3 การแจกแจงของตัวแปรอิสระ.....	11
2.4 การทำ centering ข้อมูล.....	13
2.5 ตัวแบบวิธีกำลังสองน้อยสุดเชิงส่วน.....	14
2.6 ตัวแบบวิธีกำลังสองน้อยสุด ด้วยวิธีการแปลงข้อมูลแบบกรามชมิคต์.....	18
2.7 การประมาณค่าสัมประสิทธิ์ความถดถอย.....	20
2.8 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง.....	21

	หน้า
บทที่ 3 วิธีดำเนินการศึกษา.....	22
3.1 ขอบเขตการศึกษา.....	22
3.2 วิธีดำเนินการศึกษา.....	23
บทที่ 4 ผลการวิเคราะห์ข้อมูล.....	27
4.1 ผลการศึกษาจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง.	29
4.2 ผลการศึกษาเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรง (Multicollinearity).....	41
4.3 ผลการศึกษาการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ ตัวประมาณ.....	46
บทที่ 5 สรุปผลการวิจัย และข้อเสนอแนะ.....	63
5.1 สรุปผลการศึกษา.....	64
5.2 ข้อเสนอแนะ.....	67
รายการอ้างอิง.....	68
บรรณานุกรม.....	69
ภาคผนวก.....	70
รายละเอียดการเขียนโปรแกรม.....	71
ประวัติผู้เขียนวิทยานิพนธ์.....	110

สารบัญตาราง

ตารางที่	หน้า
4.1.1	แสดงค่าเฉลี่ยของ VIF กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร 29
4.1.2	แสดงค่าเฉลี่ยของ VIF กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร..... 32
4.1.2	(ต่อ) แสดงค่าเฉลี่ยของ VIF กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร..... 33
4.1.3	แสดงค่าเฉลี่ยของ VIF_j ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร..... 34
4.1.4	แสดงค่าเฉลี่ยของ VIF ที่ระดับความสัมพันธ์สูงสุด แยกตาม VIF_j ของแต่ละ ขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร..... 38
4.2.1	แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณี ที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร..... 41
4.2.2	แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณี ที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร..... 43
4.2.2	(ต่อ) แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร..... 44
4.2.3	แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ ที่ ระดับความสัมพันธ์สูงสุด กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร..... 44
4.3.1	แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ แยกตามวิธีการ วิเคราะห์ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร..... 46
4.3.2	แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ แยกตามขนาด ตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร..... 50
4.3.3	แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ กรณีที่มีจำนวน ตัวแปรอิสระ 3 ตัวแปร..... 54
4.3.3	(ต่อ) แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ กรณีที่มี จำนวนตัวแปรอิสระ 3 ตัวแปร..... 55

ตารางที่	หน้า
4.3.4 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ระดับ ความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	55
4.3.5 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ระดับ ความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	58

สารบัญภาพ

รูปที่	หน้า
4.1.1 การเปรียบเทียบค่าเฉลี่ยของ VIF แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	30
4.1.2 การเปรียบเทียบค่าเฉลี่ยของ VIF แยกตามระดับความสัมพันธ์ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	30
4.1.3 การเปรียบเทียบค่าเฉลี่ยของ VIF1 ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	34
4.1.4 การเปรียบเทียบค่าเฉลี่ยของ VIF1 ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	35
4.1.5 การเปรียบเทียบค่าเฉลี่ยของ VIF2 ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	35
4.1.6 การเปรียบเทียบค่าเฉลี่ยของ VIF2 ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	36
4.1.7 การเปรียบเทียบค่าเฉลี่ยของ VIF3 ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	36
4.1.8 การเปรียบเทียบค่าเฉลี่ยของ VIF3 ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	37
4.1.9 การเปรียบเทียบค่าเฉลี่ยของ VIF ที่ระดับความสัมพันธ์สูงสุด แยกตาม VIF_j ที่ $n=50$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	39
4.1.10 การเปรียบเทียบค่าเฉลี่ยของ VIF ที่ระดับความสัมพันธ์สูงสุด แยกตาม VIF_j ที่ $n=100$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	39
4.1.11 การเปรียบเทียบค่าเฉลี่ยของ VIF ที่ระดับความสัมพันธ์สูงสุด แยกตาม VIF_j ที่ $n=200$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	40
4.2.1 การเปรียบเทียบเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	41

รูปที่	หน้า
4.2.2 การเปรียบเทียบเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	45
4.3.1 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามวิธีการวิเคราะห์ ที่ $n=50$ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	47
4.3.2 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามวิธีการวิเคราะห์ ที่ $n=100$ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	47
4.3.3 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามวิธีการวิเคราะห์ ที่ $n=200$ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	48
4.3.4 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามขนาดตัวอย่างของวิธี PLS กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	50
4.3.5 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามระดับความสัมพันธ์ของวิธี PLS กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	51
4.3.6 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามขนาดตัวอย่างของวิธี OLS_G กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	51
4.3.7 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามระดับความสัมพันธ์ของวิธี OLS_G กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร.....	52
4.3.8 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ ที่ $n=50$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	56
4.3.9 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ ที่ $n=100$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	56
4.3.10 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ ที่ $n=200$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	57

รูปที่	หน้า
4.3.11 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่างของวิธี PLS กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	59
4.3.12 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ของวิธี PLS กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	59
4.3.13 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่างของวิธี OLS_G กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	60
4.3.14 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ของวิธี OLS_G กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร.....	60

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เนื่องจากในปัจจุบันนี้มีงานวิจัยทางด้านต่างๆ อาทิเช่น สังคมศาสตร์ วิทยาศาสตร์ วิศวกรรมศาสตร์ การแพทย์และเภสัชศาสตร์ ได้นำการวิเคราะห์ความถดถอย (Regression Analysis) มาใช้เพื่อศึกษาความสัมพันธ์ของตัวแปรที่สนใจและเพื่อประมาณค่าสัมประสิทธิ์ความถดถอยในการสร้างสมการพยากรณ์อย่างแพร่หลาย แต่หลักการของการวิเคราะห์ความถดถอยมีข้อกำหนดเบื้องต้นหลายประการ อาทิเช่น ความคลาดเคลื่อนมีความเป็นเชิงเส้น ความคลาดเคลื่อนมีความเป็นอิสระจากกัน ความคลาดเคลื่อนมีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับศูนย์ และค่าความแปรปรวนคงที่ เป็นต้น และในการสร้างสมการความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระเพื่อสร้างตัวแบบในการพยากรณ์ค่า y นั้นตัวแปรอิสระที่สนใจจะต้องเป็นอิสระซึ่งกันและกัน แต่ในความเป็นจริงตัวแปรอิสระมักเกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) ซึ่งทำให้ผลลัพธ์ของการวิเคราะห์ความถดถอยไม่มีความแม่นยำ กล่าวคือ ความแปรปรวนของสัมประสิทธิ์ความถดถอยมีค่าสูงเกินไป เครื่องหมายของสัมประสิทธิ์ความถดถอยมีทิศทางตรงข้ามกันและค่าที่ได้อาจมากหรือน้อยกว่าความเป็นจริง และทำให้ค่าสัมประสิทธิ์ความถดถอยเปลี่ยนแปลงไปเมื่อทำการเพิ่มตัวแปรอิสระ เป็นต้น ดังนั้นจึงจำเป็นต้องทำการแก้ไขปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระก่อนทำการวิเคราะห์ความถดถอย ในการศึกษาครั้งนี้จึงสนใจที่จะเปรียบเทียบประสิทธิภาพของตัวประมาณหลังแก้ไขปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระของวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชmidt โดยวิธีกำลังสองน้อยสุดเชิงส่วนเป็นเทคนิคที่พัฒนามาเพื่อใช้แก้ไข ปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ ส่วนวิธีกำลังสองน้อยสุดในการศึกษาครั้งนี้จะใช้หลักการของพีชคณิตเชิงเส้นด้วยกระบวนการแปลงข้อมูลแบบกรามชmidt เข้ามาช่วยแก้ไขปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ จากนั้นทำการประมาณค่าพารามิเตอร์ความถดถอยด้วยวิธีกำลังสองน้อยสุด ซึ่งค่าประมาณพารามิเตอร์ความถดถอยที่ได้จะเป็นค่าประมาณที่เหมาะสม ทางผู้ศึกษาจึงเห็นว่าในการแก้ปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระในการวิเคราะห์ความถดถอยหลายตัวแปรเป็นสิ่งสำคัญในการประมาณค่าพารามิเตอร์ความถดถอยที่เหมาะสมในแต่ละกรณี

ของจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับความสัมพันธ์ของตัวแปรอิสระ และการแจกแจงของตัวแปรอิสระหลายตัวแปร

ในงานวิจัยของ Garthwaite (1994)¹ ได้เสนอการประมาณค่าสัมประสิทธิ์ความถดถอยวิธีกำลังสองน้อยสุดเชิงส่วนที่มีตัวแปรตามตัวแปรเดียวและตัวแปรอิสระหลายตัวแปร ซึ่งจะใช้เป็นแนวทางในการศึกษาครั้งนี้

1.2 วัตถุประสงค์ของการศึกษา

1. เพื่อประมาณค่าพารามิเตอร์ที่เหมาะสม สำหรับตัวแบบความถดถอยพหุเชิงเส้นของวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์เมื่อเกิดพหุสัมพันธ์

2. เพื่อเปรียบเทียบการประมาณค่าพารามิเตอร์ สำหรับตัวแบบความถดถอยพหุเชิงเส้นระหว่างวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์เมื่อเกิดพหุสัมพันธ์

1.3 ขอบเขตของการศึกษา

การศึกษานี้ผู้วิจัยได้ทำการศึกษาภายใต้ขอบเขต ดังนี้

1. กำหนดจำนวนตัวแปรอิสระที่ทำการศึกษา คือ 2 และ 3 ตัวแปร
2. กำหนดขนาดตัวอย่างที่ทำการศึกษา คือ 50, 100 และ 200
3. กำหนดค่าสัมประสิทธิ์ความถดถอยเบื้องต้น โดยให้ค่า $\beta_j = 1$ สำหรับ $j = 0, 1, \dots, p$
4. กำหนดระดับความสัมพันธ์ระหว่างตัวแปรอิสระให้มีความสัมพันธ์แบบเชิงเส้นเท่านั้น
 - 4.1 โดยถ้าจำนวนตัวแปรอิสระเป็น 2 ตัวแปร ผู้วิจัยจะทำการศึกษาภายใต้ 10 ระดับ คือ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99
 - 4.2 โดยถ้าจำนวนตัวแปรอิสระเป็น 3 ตัวแปร ผู้วิจัยจะทำการศึกษาภายใต้ 5 ระดับ คือ 0.91, 0.93, 0.95, 0.97 และ 0.99

¹ Paul H. Garthwaite. "An Interpretation of Partial Least Squares.", Journal of the American Statistical Association, Vol. 89, No. 425 (Mar., 1994): 122-127.

5. กำหนดค่าความแปรปรวนของความคลาดเคลื่อน σ^2 มีค่าเป็น 10 และค่าความแปรปรวนของตัวแปรอิสระ X_1 , X_2 และ X_3 มีค่าเป็น $\sigma_1^2 = 4$, $\sigma_2^2 = 4$ และ $\sigma_3^2 = 4$ ตามลำดับ

สามารถหาค่าความแปรปรวนของตัวแปรตาม Y ได้ดังนี้

ตัวอย่าง กรณีที่จำนวนตัวแปรอิสระ 2 ตัวแปร กำหนดให้ค่า $\beta_0 = \beta_1 = \beta_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 4$ และ $\sigma^2 = 10$

$$\text{จากสมการถดถอย} \quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

$$Y_i = 1 + (1)X_{1i} + (1)X_{2i} + \varepsilon_i$$

มีความแปรปรวนของ Y_i เป็น

$$\begin{aligned} \text{Var}(Y_i) &= \text{Var}(X_{1i}) + \text{Var}(X_{2i}) + \text{Var}(\varepsilon_i) + 2\text{Cov}(X_{1i}, X_{2i}) + \\ &\quad 2\text{Cov}(X_{1i}, \varepsilon_i) + 2\text{Cov}(X_{2i}, \varepsilon_i) \end{aligned}$$

เนื่องจาก $\text{Var}(X_{1i}) = 4$, $\text{Var}(X_{2i}) = 4$, $\text{Var}(\varepsilon_i) = 10$, $\text{Cov}(X_{1i}, \varepsilon_i) = 0$

และ $\text{Cov}(X_{2i}, \varepsilon_i) = 0$

$$\begin{aligned} \text{เพราะฉะนั้น} \quad \text{Var}(Y_i) &= \sigma_1^2 + \sigma_2^2 + \sigma^2 + 2\sigma_1\sigma_2\rho_{12} \\ &= 4 + 4 + 10 + 2(2)(2)\rho_{12} \\ &= 18 + 8\rho_{12} \end{aligned}$$

ตัวอย่าง กรณีที่จำนวนตัวแปรอิสระ 3 ตัวแปร กำหนดให้ค่า $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 4$ และ $\sigma^2 = 10$

$$\text{จากสมการถดถอย} \quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

$$Y_i = 1 + (1)X_{1i} + (1)X_{2i} + (1)X_{3i} + \varepsilon_i$$

มีความแปรปรวนของ Y_i เป็น

$$\begin{aligned} \text{Var}(Y_i) &= \text{Var}(X_{1i}) + \text{Var}(X_{2i}) + \text{Var}(X_{3i}) + \text{Var}(\varepsilon_i) + \\ &\quad 2\text{Cov}(X_{1i}, X_{2i}) + 2\text{Cov}(X_{1i}, X_{3i}) + 2\text{Cov}(X_{1i}, \varepsilon_i) + \\ &\quad 2\text{Cov}(X_{2i}, X_{3i}) + 2\text{Cov}(X_{2i}, \varepsilon_i) + 2\text{Cov}(X_{3i}, \varepsilon_i) \end{aligned}$$

เนื่องจาก $\text{Var}(X_{1i}) = 4$, $\text{Var}(X_{2i}) = 4$, $\text{Var}(X_{3i}) = 4$, $\text{Var}(\varepsilon_i) = 10$, $\text{Cov}(X_{1i}, \varepsilon_i) = 0$, $\text{Cov}(X_{2i}, \varepsilon_i) = 0$ และ $\text{Cov}(X_{3i}, \varepsilon_i) = 0$

$$\begin{aligned} \text{เพราะฉะนั้น} \quad \text{Var}(Y_i) &= \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \sigma^2 + 2\sigma_1\sigma_2\rho_{12} + 2\sigma_1\sigma_3\rho_{13} \\ &\quad + 2\sigma_2\sigma_3\rho_{23} \\ &= 4 + 4 + 4 + 10 + 2(2)(2)\rho_{12} + 2(2)(2)\rho_{13} \end{aligned}$$

$$\begin{aligned}
 &+2(2)(2)\rho_{23} \\
 &= 22 + 8(\rho_{12} + \rho_{13} + \rho_{23})
 \end{aligned}$$

6. กำหนดจำนวนตัวแปรอิสระที่มีพหุสัมพันธ์และระดับความสัมพันธ์ของตัวแปรเป็น 2 กรณีดังนี้

6.1 กรณีที่จำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร ผู้วิจัยได้ทำการศึกษาเมื่อตัวแปรอิสระ 2 ตัวแปร ได้แก่ X_1 และ X_2 มีพหุสัมพันธ์กัน

โดยให้ ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_2

ซึ่งศึกษาภายใต้ระดับความสัมพันธ์ คือ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99

6.2 กรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร ผู้วิจัยได้ทำการศึกษาเมื่อตัวแปรอิสระ 3 ตัวแปร ได้แก่ X_1 , X_2 และ X_3 มีพหุสัมพันธ์กัน

ให้ $(\rho_{12}, \rho_{13}, \rho_{23})$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดย

ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_2

ρ_{13} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_3

ρ_{23} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_2 และ X_3

ซึ่งศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

(0.91, 0.91, 0.91)	(0.91, 0.91, 0.93)	(0.91, 0.91, 0.95)
(0.91, 0.91, 0.97)	(0.91, 0.91, 0.99)	(0.91, 0.93, 0.93)
(0.91, 0.93, 0.95)	(0.91, 0.93, 0.97)	(0.91, 0.93, 0.99)
(0.91, 0.95, 0.95)	(0.91, 0.95, 0.97)	(0.91, 0.95, 0.99)
(0.91, 0.97, 0.97)	(0.93, 0.93, 0.93)	(0.93, 0.93, 0.95)
(0.93, 0.93, 0.97)	(0.93, 0.93, 0.99)	(0.93, 0.95, 0.95)
(0.93, 0.95, 0.97)	(0.93, 0.95, 0.99)	(0.93, 0.97, 0.97)
(0.93, 0.97, 0.99)	(0.95, 0.95, 0.95)	(0.95, 0.95, 0.97)
(0.95, 0.95, 0.99)	(0.95, 0.97, 0.97)	(0.95, 0.97, 0.99)
(0.97, 0.97, 0.97)	(0.97, 0.97, 0.99)	(0.97, 0.99, 0.99)
(0.99, 0.99, 0.99)		

1.4 ข้อตกลงเบื้องต้น

การศึกษาครั้งนี้มีข้อตกลงเบื้องต้นสำหรับการดำเนินงานวิจัย ดังนี้

1. ศึกษาตัวแบบความถดถอยพหุเชิงเส้น โดยมีรูปแบบ ดังนี้

$$\tilde{\mathbf{Y}} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}$$

$$\tilde{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}, \tilde{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \tilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

โดยที่ $\tilde{\mathbf{Y}}$ คือ เวกเตอร์ของตัวแปรตามที่มีขนาด $n \times 1$

\mathbf{X} คือ เมทริกซ์ของตัวแปรอิสระที่มีขนาด $n \times (p + 1)$

$\tilde{\boldsymbol{\beta}}$ คือ เวกเตอร์ของพารามิเตอร์ที่มีขนาด $(p + 1) \times 1$

$\tilde{\boldsymbol{\varepsilon}}$ คือ เวกเตอร์ของความคลาดเคลื่อนที่มีขนาด $n \times 1$

n คือ ขนาดตัวอย่าง

p คือ จำนวนตัวแปรอิสระ

และ $E(\tilde{\boldsymbol{\varepsilon}}|\tilde{X} = \tilde{x}) = \tilde{0}$, $Cov(\tilde{\boldsymbol{\varepsilon}}|\tilde{X} = \tilde{x}) = \sigma^2\tilde{I}_n$, $\tilde{\boldsymbol{\varepsilon}} \sim N(\tilde{0}, \sigma^2\tilde{I}_n)$

2. ตัวแปรอิสระ คือ $\mathbf{X} = (X_1, X_2, \dots, X_p)$ มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal distribution) เขียนเป็นสัญลักษณ์ ดังนี้ $X_j \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ โดยที่ $\boldsymbol{\mu} = E(\tilde{X})$ และ $\boldsymbol{\Sigma} = Cov(\tilde{X})$
3. ตัวแปรอิสระ X และตัวแปรตาม y เป็นตัวแปรเชิงปริมาณ

1.5 คำจำกัดความที่ใช้ในการศึกษา

1. วิธีกำลังสองน้อยสุดเชิงส่วน (Partial Least Squares Method) คือ วิธีที่ใช้ประมาณค่าพารามิเตอร์ความถดถอยของการวิเคราะห์หลายตัวแปร ที่ได้ทำการศึกษาจากความสัมพันธ์ระหว่างตัวแปรตามตัวแปรเดียวและตัวแปรอิสระหลายตัวแปร ซึ่งจะนำไปใช้เพื่อการพยากรณ์

2. วิธีกำลังสองน้อยสุด (Ordinary Least Squares Method) คือ วิธีที่ใช้ประมาณค่าพารามิเตอร์ความถดถอยของการวิเคราะห์หลายตัวแปร ที่ได้ทำการศึกษาจากความสัมพันธ์ระหว่างตัวแปรตามตัวแปรเดียวและตัวแปรอิสระหลายตัวแปร ซึ่งจะนำไปใช้เพื่อการพยากรณ์

3. วิธีกรามชมิคต์ (Gram-Schmidt Method) คือ หลักการของพีชคณิตเชิงเส้น (Linear Algebra) ซึ่งเป็นวิธีที่ทำให้เวกเตอร์ของตัวแปรอิสระใหม่ตั้งฉากกัน

4. พหุสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) คือ ปัญหาที่เกิดจากตัวแปรอิสระมีความสัมพันธ์กันซึ่งกันและกัน

1.6 วิธีดำเนินการศึกษา

1. กำหนดค่าเริ่มต้นในการจำลองข้อมูล ดังนี้
 - 1.1 กำหนดจำนวนตัวแปรอิสระ p ตัว และจำนวนตัวแปรตาม 1 ตัว
 - 1.2 กำหนดขนาดตัวอย่าง n ตัว
 - 1.3 กำหนดค่าสัมประสิทธิ์ความถดถอยเบื้องต้น (β_j) สำหรับ $j = 0, 1, 2, \dots, p$
 - 1.4 กำหนดระดับความสัมพันธ์ระหว่างตัวแปรอิสระเป็นคู่ ρ_{ij} สำหรับ $i = 1, 2, \dots, n$ และ $j = 1, 2, \dots, p$
 - 1.5 กำหนดค่าความแปรปรวนของความคลาดเคลื่อน (σ^2) และค่าความแปรปรวนของตัวแปรอิสระ X_1, X_2 และ X_3
2. จำลองข้อมูลตัวแปรอิสระของการแจกแจงปกติหลายตัวแปร โดยตัวแปรอิสระสร้างจากค่าเฉลี่ย ค่าความแปรปรวนและระดับความสัมพันธ์ที่กำหนด ด้วยวิธีของโซเลสกี ดังนี้

$$X_1 = Z_1$$

$$X_2 = \rho_{12}Z_1 + \sqrt{1 - \rho_{12}^2}Z_2$$

$$X_3 = \rho_{12}Z_1 + \frac{(\rho_{23} - \rho_{12}\rho_{13})}{\sqrt{1 - \rho_{12}^2}}Z_2 + \sqrt{1 - \rho_{13}^2 - \frac{(\rho_{23} - \rho_{12}\rho_{13})^2}{1 - \rho_{12}^2}}Z_3$$

โดยที่ Z_1, Z_2, Z_3 เป็นตัวแปรสุ่มแบบปกติมาตรฐานที่เป็นอิสระกัน

3. สร้างเวกเตอร์ของข้อมูลตัวแปรตาม Y จากตัวแบบความถดถอยพหุเชิงเส้น
4. คำนวณค่า VIF ซึ่งพิจารณาจากสัมประสิทธิ์สหสัมพันธ์พหุของตัวแปรอิสระ โดยเกณฑ์ในการใช้ตัดสินใจว่าเกิดปัญหา Collinearity สามารถคำนวณตามมาตรวัดทางสถิติได้ ดังนี้

$$VIF_j = \frac{1}{1 - R_j^2} \quad ; \quad j = 1, 2, \dots, p$$

โดยที่ R_j^2 เป็นสัมประสิทธิ์สหสัมพันธ์พหุของตัวแปรอิสระ P_j กับตัวแปรอิสระอื่นๆ

ถ้าค่า $VIF_j \leq 10 ; j = 1, 2, \dots, p$ แสดงว่าไม่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

ถ้าค่า $VIF_j > 10 ; j = 1, 2, \dots, p$ แสดงว่าเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

5. นำเวกเตอร์ของตัวแปรตาม Y และนำเมทริกซ์ของตัวแปรอิสระ X ที่ได้มาทำ centering เพื่อให้ค่า Intercept หายไปเหลือเพียงค่า X อย่างเดียว
6. สร้างตัวแปรอิสระใหม่ให้ตั้งฉากกันด้วยการแปลงข้อมูลแบบกรามชมิตต์ และใช้วิธีกำลังสองน้อยสุดเชิงส่วนในการแก้ไขปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ
7. ประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด จากสูตรดังนี้

$$\tilde{b} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \tilde{y}$$

8. หาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการทำซ้ำจำนวน 1,000 รอบ ทั้ง 2 วิธี จากสูตรดังนี้

$$AMSE(PLS) = \frac{1}{N} \sum_{i=1}^N MSE(PLS)$$

และ

$$AMSE(OLS_G) = \frac{1}{N} \sum_{i=1}^N MSE(OLS_G)$$

โดยที่ MSE คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ

N คือ จำนวนรอบของการทำซ้ำ

เมื่อ $MSE = \frac{1}{p} \sum_{j=0}^p (Var(b_j) + (Bias(b_j))^2)$

9. เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของวิธีการประมาณ จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการทำซ้ำของทั้ง 2 วิธี
10. สรุปผลการศึกษา

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ค่าประมาณพารามิเตอร์ที่เหมาะสม สำหรับตัวแบบความถดถอยพหุเชิงเส้นของวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์ เมื่อเกิดพหุสัมพันธ์

2. เป็นแนวทางให้ผู้วิจัยเลือกวิธีการประมาณค่าพารามิเตอร์ที่เหมาะสม สำหรับตัวแบบ ความถดถอยพหุเชิงเส้นเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระน้อยสุดเชิงส่วนและตัวแปรอิสระ น้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิคต์

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

2.1 การวิเคราะห์ความถดถอย

การวิเคราะห์ความถดถอย (Regression Analysis) เป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม เพื่อศึกษาพฤติกรรมความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามว่ามีความสัมพันธ์กันหรือไม่ และถ้ามีความสัมพันธ์เกิดขึ้นจะมีความสัมพันธ์กันอย่างไร ทั้งเรื่องของระดับความสัมพันธ์และทิศทางของความสัมพันธ์ที่เกิดขึ้น และเพื่อสร้างตัวแบบความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม สำหรับการพยากรณ์หรือการประมาณค่าตัวแปรตาม ซึ่งองค์ประกอบสำคัญของการวิเคราะห์ความถดถอย คือ มีตัวแบบความถดถอยที่แสดงรูปแบบความสัมพันธ์ระหว่างตัวแปรอิสระ ตัวแปรตาม และความคลาดเคลื่อนสุ่มที่ไม่สามารถสังเกตได้ โดยความคลาดเคลื่อนสุ่มจะมีคุณสมบัติการเป็นอิสระจากกัน (Independence) การมีความเป็นเชิงเส้น (Linearity) การมีความแปรปรวนคงที่ (Homoscedasticity) และการมีการแจกแจงแบบปกติ (Normality) คุณสมบัติดังกล่าว จะทำให้สามารถนำข้อมูลค่าสังเกตที่เก็บรวบรวมได้ของตัวแปรอิสระกับตัวแปรตามมาทำการวิเคราะห์และอนุมานเกี่ยวกับพารามิเตอร์ได้อย่างถูกต้อง (Valid) ซึ่งถ้ามีตัวแปรอิสระเพียงตัวแปรเดียวจะเรียกว่า ตัวแบบความถดถอยอย่างง่ายเชิงเส้น (Simple Linear Regression Model) และถ้ามีตัวแปรอิสระมากกว่า 1 ตัวแปรจะเรียกว่า ตัวแบบความถดถอยพหุเชิงเส้น (Multiple Linear Regression Model) สำหรับในการศึกษาครั้งนี้ทางผู้วิจัยก็ได้ศึกษาตัวแบบความถดถอยพหุเชิงเส้น

2.2 ตัวแบบและการวิเคราะห์ความถดถอยพหุเชิงเส้น

การวิเคราะห์ความถดถอยพหุเชิงเส้น (Multiple Linear Regression Analysis) เป็นการศึกษาความถดถอยระหว่างตัวแปรอิสระมากกว่า 1 ตัวแปร กับตัวแปรตามจำนวน 1 ตัวแปร ซึ่งสำหรับในการศึกษาครั้งนี้จะพิจารณาข้อมูลขนาดตัวอย่าง n ตัว ในการสร้างความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ X_1, X_2, \dots, X_p กับตัวแปรตาม Y โดยที่สมการตัวแบบของความสัมพันธ์จะเขียนได้ ดังนี้

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i \quad ; i = 1, 2, \dots, n$$

หรือเขียนในรูปของเมทริกซ์ได้ ดังนี้

$$\tilde{Y} = X\tilde{\beta} + \tilde{\epsilon}$$

$$\text{เมื่อ } \tilde{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{p1} \\ 1 & X_{12} & X_{22} & \cdots & X_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{pn} \end{bmatrix}, \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \tilde{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- โดยที่
- \tilde{Y} คือ เวกเตอร์ของตัวแปรตามที่มีขนาด $n \times 1$
 - X คือ เมทริกซ์ของตัวแปรอิสระที่มีขนาด $n \times (p + 1)$
 - $\tilde{\beta}$ คือ เวกเตอร์ของพารามิเตอร์ที่มีขนาด $(p + 1) \times 1$
 - $\tilde{\epsilon}$ คือ เวกเตอร์ของความคลาดเคลื่อนที่มีขนาด $n \times 1$
 - n คือ ขนาดตัวอย่าง
 - p คือ จำนวนตัวแปรอิสระ

โดยข้อกำหนดของตัวแบบความถดถอยพหุเชิงเส้น มีดังนี้

1. ค่าความคลาดเคลื่อนสุ่ม ϵ_i มีการแจกแจงปกติ มี $E(\tilde{\epsilon}|\tilde{X} = \tilde{x}) = \tilde{0}$ และมี $Cov(\tilde{\epsilon}|\tilde{X} = \tilde{x}) = \sigma^2 \tilde{I}_n$ นั่นก็คือ $\tilde{\epsilon} \sim N(\tilde{0}, \sigma^2 \tilde{I}_n)$
2. ค่าความคลาดเคลื่อนสุ่ม ϵ_i ทุกตัวเป็นอิสระจากกัน นั่นก็คือ $E(\epsilon_i, \epsilon_j) = 0$ เมื่อ $i \neq j$ หรือกล่าวได้ว่า $Cov(\epsilon_i, \epsilon_j) = 0$
3. ตัวแปรอิสระ X ทุกตัวเป็นอิสระกัน
4. ตัวแปรตาม Y เป็นตัวแปรสุ่ม และแต่ละค่าของตัวแปรตาม Y เป็นอิสระกัน
5. เวกเตอร์สุ่ม Y มีการแจกแจงปกติพหุ มี $E(\tilde{Y}|X) = X\tilde{\beta}$ และ $Cov(\tilde{Y}|X) = \sigma^2 I_n$ โดยที่ I_n เป็นเมทริกซ์เอกลักษณ์ที่มีขนาด $n \times n$

ซึ่งจากข้อกำหนดของตัวแบบความถดถอยพหุเชิงเส้น ในการสร้างสมการความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระเพื่อสร้างตัวแบบในการพยากรณ์ค่าตัวแปรตาม y นั้นตัวแปรอิสระที่สนใจจะต้องเป็นอิสระซึ่งกันและกัน แต่ในความเป็นจริงตัวแปรอิสระมักจะเกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) ซึ่งทำให้ผลลัพธ์ของการวิเคราะห์ความถดถอยไม่มีความแม่นยำ กล่าวคือ ความแปรปรวนของสัมประสิทธิ์ความถดถอยมีค่าสูงเกินไป เครื่องหมายของสัมประสิทธิ์ความถดถอยมีทิศทางตรงข้ามกัน และค่าที่ได้อาจมากหรือน้อยกว่าความเป็นจริง และทำให้ค่าสัมประสิทธิ์ความถดถอยเปลี่ยนแปลงไปเมื่อทำการเพิ่มตัวแปรอิสระ เป็นต้น ดังนั้น

จึงจำเป็นต้องทำการแก้ไขปัญหาคู่สัมพันธ์ระหว่างตัวแปรอิสระก่อนทำการวิเคราะห์ความถดถอย ซึ่งสำหรับการศึกษาครั้งนี้ จะแก้ไขปัญหาคู่สัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์

2.3 การแจกแจงของตัวแปรอิสระ

การศึกษาค้นคว้านี้ได้ทำการศึกษาคู่สัมพันธ์ระหว่างตัวแปรอิสระที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution)¹ โดยการวิเคราะห์สถิติแบบใช้พารามิเตอร์ต้องมีข้อสมมติเกี่ยวกับการแจกแจงของข้อมูลว่าต้องมาจากการแจกแจงแบบปกติ ซึ่งถ้าข้อมูลมีหลายตัวแปร จะมีการนำการแจกแจงแบบปกติหลายตัวแปรมาพิจารณา โดยให้เวกเตอร์ของตัวแปรอิสระ คือ $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ มีการแจกแจงแบบปกติที่มีเวกเตอร์ของค่าเฉลี่ยเป็น $\boldsymbol{\mu}$ และเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) เป็น Σ โดยที่ฟังก์ชันความน่าจะเป็นของเวกเตอร์ตัวแปรอิสระเป็นดังนี้

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^p |\Sigma|^{1/2}} e^{-1/2(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

โดยที่ p เป็นจำนวนตัวแปรอิสระ

ซึ่งในการศึกษาค้นคว้านี้ได้ทำการศึกษาคู่สัมพันธ์ระหว่างตัวแปรอิสระที่มีการแจกแจงแบบปกติหลายตัวแปรที่มีพารามิเตอร์ค่าเฉลี่ย คือ $\boldsymbol{\mu} = (0, 0, \dots, 0)'$ และมีเมทริกซ์ความแปรปรวนร่วมตามระดับความสัมพันธ์ของตัวแปรอิสระที่กำหนดไว้ คือ

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{1p} & \cdots & \sigma_1\sigma_p\rho_{1p} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \cdots & \sigma_2\sigma_p\rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_p\sigma_1\rho_{p1} & \sigma_p\sigma_2\rho_{p2} & \cdots & \sigma_p^2 \end{bmatrix}_{p \times p}$$

โดยเขียนเป็นสัญลักษณ์ ดังนี้ $X_j \sim N_p(0, \sigma_j^2)$

¹ กัลยา วาณิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 4. กรุงเทพมหานคร: สำนักพิมพ์ธรรมสาร, 2552.

การจำลองเวกเตอร์สุ่มแบบปกติ²

การจำลองเวกเตอร์สุ่มแบบปกติที่ตัวแปรอิสระมีความสัมพันธ์กัน สามารถทำได้โดยการสร้างตัวแปรสุ่มแบบปกติ p ตัวที่เป็นอิสระต่อกัน Z_1, Z_2, \dots, Z_p โดยที่ $Z_p \sim N(0,1)$ ซึ่งสามารถแปลงเชิงเส้น ได้ดังนี้

$$X_i = \sum_{j=1}^p c_{ij}Z_j + \mu_i \quad ; i = 1, 2, \dots, p$$

และสามารถเขียนความสัมพันธ์อยู่ในรูปของเมทริกซ์ได้ ดังนี้

$$X = \tilde{C}Z + \mu$$

โดยที่ c_{ij} เป็นสมาชิกในแถวที่ i หลักที่ j ของเมทริกซ์ \tilde{C}

จะได้ว่า X_i ที่ได้จะมีการแจกแจงแบบปกติ เพราะว่าการแปลงเชิงเส้นของตัวแปรสุ่มแบบปกติที่เป็นอิสระต่อกันจะได้ตัวแปรสุ่มตัวใหม่ที่เป็นตัวแปรสุ่มแบบปกติ มีค่าเฉลี่ยเท่ากับ μ_i และค่า c_{ij} ที่เหมาะสมจะทำให้ความแปรปรวนร่วมสอดคล้องกับ $\tilde{\Sigma}$ จะได้ว่า

$$(X - \mu)(X - \mu)^T = \tilde{C}ZZ^T\tilde{C}^T$$

และเนื่องจาก $E[Z_i^2] = 1$ และ $E[Z_iZ_j] = 0$ เมื่อ $i \neq j$ เพราะว่า Z_i ทุกตัวเป็นอิสระกัน จะได้ว่า

$$E[(X - \mu)(X - \mu)^T] = \tilde{C}\tilde{C}^T = \tilde{\Sigma}$$

โดยที่ $E[\tilde{M}]$ คือ ค่าเฉลี่ยของสมาชิกแต่ละตัวในเมทริกซ์ \tilde{M} และเมทริกซ์ \tilde{I} เป็นเมทริกซ์เอกลักษณ์ซึ่งมีสมาชิกของเส้นทแยงมุมเป็น 1 และสมาชิกนอกเส้นทแยงมุมเป็น 0 เนื่องจาก

$$E[(X - \mu)(X - \mu)^T] = \tilde{\Sigma}$$

ดังนั้น สามารถหาค่า c_{ij} ที่เหมาะสมหรือเมทริกซ์ \tilde{C} ได้ โดยที่

$$\tilde{C}\tilde{C}^T = \tilde{\Sigma}$$

ทำให้ได้เวกเตอร์ $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ ตามที่ต้องการจากการแปลงเชิงเส้น และจากทฤษฎีการแยกตัวประกอบของโคเลสกี (Cholesky's decomposition) ถ้าเมทริกซ์ \tilde{M} เป็นเมทริกซ์สมมาตรที่เป็นบวกแน่นอน (Positive-definite symmetric matrix) จะสามารถแยก \tilde{M} เป็น

² เสกสรร เกียรติสุไพบุญ. การจำลอง. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2555.

เมทริกซ์สามเหลี่ยมล่างที่มีสมาชิกเหนือเส้นทแยงมุมเป็นศูนย์ และ \tilde{L} มีสมาชิกของเมทริกซ์เป็นเลขจำนวนจริง ซึ่งจะได้ว่า

$$\tilde{M} = \tilde{L}\tilde{L}^T$$

ทำให้การแยกตัวประกอบของโคเลสกี มีสมาชิกบนเส้นทแยงมุม L_{ii} ของ \tilde{L} เป็นดังนี้

$$L_{ii} = (m_{ii} - \sum_{j=1}^{i-1} L_{ij}^2)^{1/2}$$

และสมาชิกใต้เส้นทแยงมุม L_{ji} ของ \tilde{L} เป็นดังนี้

$$L_{ji} = \frac{1}{L_{ii}} (m_{ij} - \sum_{k=1}^{i-1} L_{ik}L_{jk}) \quad ; j = i + 1, i + 2, \dots, n$$

โดยที่ m_{ij} เป็นสมาชิกในแถวที่ i หลักที่ j ของเมทริกซ์ \tilde{M}

ดังนั้น การแยกตัวประกอบของโคเลสกี เมทริกซ์ความแปรปรวนร่วมต้องเป็นบวกแน่นอน (Positive definite) แต่หากเมทริกซ์ความแปรปรวนร่วมไม่เป็นบวกแน่นอน เมื่อทำการแยกตัวประกอบด้วยวิธีโคเลสกี จะมีสมาชิกบางตัวบนเส้นทแยงมุมเป็น 0 ซึ่งทำให้ไม่สามารถหาค่าสมาชิกใต้เส้นทแยงมุมอื่นได้

ตัวอย่าง การจำลองตัวแปรสุ่ม X_1, X_2 และ X_3 ด้วยวิธีการแยกตัวประกอบของโคเลสกี เป็นดังนี้

$$X_1 = Z_1$$

$$X_2 = \rho_{12}Z_1 + \sqrt{1 - \rho_{12}^2}Z_2$$

$$X_3 = \rho_{12}Z_1 + \frac{(\rho_{23} - \rho_{12}\rho_{13})}{\sqrt{1 - \rho_{12}^2}}Z_2 + \sqrt{1 - \rho_{13}^2 - \frac{(\rho_{23} - \rho_{12}\rho_{13})^2}{1 - \rho_{12}^2}}Z_3$$

โดยที่ Z_1, Z_2, Z_3 เป็นตัวแปรสุ่มแบบปกติมาตรฐานที่เป็นอิสระกัน

2.4 การทำ Centering ข้อมูล

จากตัวแบบความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ X_1, X_2, \dots, X_p กับตัวแปรตาม Y กำหนดให้เวกเตอร์ของตัวแปรตาม \tilde{y} เป็นตัวแปรที่ได้มาจาก \tilde{Y} และเมทริกซ์ของข้อมูลตัวแปรอิสระ \mathbf{x} เป็นตัวแปรที่ได้มาจาก \mathbf{X} จากนั้นนำมาลบด้วยค่าเฉลี่ยของตัวแปรตาม Y และค่าเฉลี่ย

ของตัวแปรอิสระ X_j เรียกว่า การทำ “Centering” ของข้อมูลเพื่อทำให้ค่า Intercept หายไปเหลือเพียงค่า X อย่างเดียว ซึ่งค่าเฉลี่ยสามารถเขียนแทนด้วย

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad \text{และ} \quad \bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ji}$$

โดยที่สมการถดถอยของค่าเฉลี่ยจะเขียนได้ ดังนี้

$$\bar{Y} = \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \dots + \beta_p \bar{X}_p + \bar{\varepsilon}$$

ซึ่งสมการถดถอยของความสัมพันธ์ใหม่จะเขียนได้ ดังนี้

$$Y_i - \bar{Y} = \beta_1 (X_{1i} - \bar{X}_1) + \beta_2 (X_{2i} - \bar{X}_2) + \dots + \beta_p (X_{pi} - \bar{X}_p) + (\varepsilon_i - \bar{\varepsilon})$$

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \dots + \beta_p X_{pi}^* + \varepsilon_i^*$$

หรือเขียนในรูปของเมทริกซ์ได้ ดังนี้

$$\tilde{Y}^* = X^* \tilde{\beta}^* + \tilde{\varepsilon}^*$$

$$\text{เมื่อ} \quad \tilde{Y}^* = \begin{bmatrix} Y_1^* \\ Y_2^* \\ \vdots \\ Y_n^* \end{bmatrix}, \quad X^* = \begin{bmatrix} X_{11}^* & X_{21}^* & \dots & X_{p1}^* \\ X_{12}^* & X_{22}^* & \dots & X_{p2}^* \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n}^* & X_{2n}^* & \dots & X_{pn}^* \end{bmatrix}, \quad \tilde{\beta}^* = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \tilde{\varepsilon}^* = \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_n^* \end{bmatrix}$$

โดยเราจะพิจารณาตัวอย่างของเวกเตอร์ตัวแปรตาม \tilde{y}^* และเมทริกซ์ของข้อมูลตัวแปรอิสระ x^*

$$\tilde{y}^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{bmatrix}, \quad x^* = \begin{bmatrix} x_{11}^* & x_{21}^* & \dots & x_{p1}^* \\ x_{12}^* & x_{22}^* & \dots & x_{p2}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n}^* & x_{2n}^* & \dots & x_{pn}^* \end{bmatrix}$$

ซึ่งข้อมูลเวกเตอร์ตัวแปรตาม \tilde{y}^* และข้อมูลเมทริกซ์ตัวแปรอิสระ x^* ที่ได้จะเป็นข้อมูลที่ได้ทำการ Centering แล้ว

2.5 ตัวแบบวิธีกำลังสองน้อยสุดเชิงส่วน

วิธีกำลังสองน้อยสุดเชิงส่วนเป็นวิธีใหม่ที่ได้พัฒนาจาก (Wold, 1980) ที่ใช้ในการสร้างสมการความถดถอยเพื่อการพยากรณ์ที่ได้รับความสนใจในปัจจุบัน ซึ่งปรากฏในเอกสารหลายฉบับ เช่น Stone and Brooks (1990) Helland (1990, 1993), และ Garthwaite (1994) เป็นต้น

โดยวิธีกำลังสองน้อยสุดเชิงส่วน เป็นวิธีที่สร้างตัวแปรอิสระตัวใหม่จากตัวแปรอิสระเดิมและตัวแปรตาม ซึ่งจะพิจารณาการสร้างตัวแปรอิสระใหม่อย่างเป็นลำดับ และตัวแปรอิสระใหม่จะสร้างมาจากผลบวกเชิงเส้นของตัวแปรอิสระเดิม โดยไม่ได้นำค่าสัมพัทธ์ระหว่างตัวแปรอิสระเดิมมาพิจารณา และเป็นวิธีที่สามารถแก้ไขปัญหาค่าพารามิเตอร์เมื่อเกิดพหุสัมพัทธ์ระหว่างตัวแปรอิสระ แต่ก็ยังเป็นตัวประมาณที่มีความเอนเอียงอยู่ ซึ่งข้อดีของวิธีกำลังสองน้อยสุดเชิงส่วน คือ ทำให้สัมพัทธ์ของตัวแปรอิสระใหม่แต่ละคู่มีค่าเป็นศูนย์ เนื่องจากสร้างตัวแปรอิสระใหม่จากส่วนเหลือของตัวแปรเดิม โดยสามารถเขียนตัวแบบความถดถอยของวิธีกำลังสองน้อยสุดเชิงส่วนได้ ดังนี้

$$\hat{Y} = b_0 + b_1 P_{1i} + b_2 P_{2i} + \dots + b_p P_{pi}$$

โดยที่ P_j เป็นตัวแปรอิสระตัวใหม่ที่เป็นผลบวกเชิงเส้นของ X_j และค่าสัมพัทธ์ตัวอย่างสำหรับแต่ละคู่ของตัวแปรอิสระใหม่มีค่าเท่ากับศูนย์

วิธีกำลังสองน้อยสุดเชิงส่วนจะพิจารณาการสร้างตัวแปรอิสระใหม่อย่างเป็นลำดับ ซึ่งตัวแปรอิสระใหม่ P_1 จะเป็นตัวแปรที่ใช้ในการพยากรณ์ค่า $\tilde{Y}_{(1)}^*$ และเป็นตัวแปรที่สร้างจากผลบวกเชิงเส้นของ $\tilde{X}_{(1)j}^*$ โดยไม่นำค่าสัมพัทธ์ตัวอย่างระหว่างตัวแปรอิสระ $X_{(1)j}^*$ มาพิจารณาในการสร้างตัวแปรใหม่ จะได้ตัวแปรอิสระใหม่ตัวแรก P_1 โดยสมการถดถอยประมาณค่า $\tilde{Y}_{(1)}^*$ ด้วยตัวแปรถดถอยตัวแรก $\tilde{X}_{(1)1}^*$ ตัวแปรถดถอยตัวที่สอง $\tilde{X}_{(1)2}^*$ และตัวแปรถดถอยต่อไปเรื่อยๆ จะถึงตัวที่ p และค่าเฉลี่ยตัวอย่างของตัวแปรอิสระใหม่แต่ละตัวมีค่าเป็นศูนย์ ซึ่งมีขั้นตอนการสร้างตัวแปรอิสระใหม่ ดังนี้

ขั้นที่ 1 กำหนดให้เวกเตอร์ $\tilde{y}_{(1)}^*$ เป็นเวกเตอร์ตัวอย่างของ $Y_{(1)}^*$ ได้จากเวกเตอร์ Y และเวกเตอร์ $\tilde{x}_{(1)j}^*$ เป็นเวกเตอร์ตัวอย่างของ $X_{(1)j}^*$ ได้จากเวกเตอร์ X_j สำหรับ $j = 1, 2, \dots, p$ ซึ่งเกิดจากการทำ Centering

โดยที่
$$\tilde{y}_{(1)}^* = \tilde{y} - \bar{y} \cdot \mathbf{1}$$

$$\tilde{y}_{(1)}^* = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{bmatrix} = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{bmatrix}$$

$$\tilde{x}_{(1)j}^* = \tilde{x}_j - \bar{x}_j \cdot \mathbf{1}$$

$$\tilde{\mathbf{x}}_{(1)j}^* = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_2 & \dots & x_{p1} - \bar{x}_p \\ x_{12} - \bar{x}_1 & x_{22} - \bar{x}_2 & \dots & x_{p2} - \bar{x}_p \\ \vdots & \vdots & \dots & \vdots \\ x_{1n} - \bar{x}_1 & x_{2n} - \bar{x}_2 & \dots & x_{pn} - \bar{x}_p \end{bmatrix} = \begin{bmatrix} x_{11}^* & x_{21}^* & \dots & x_{p1}^* \\ x_{12}^* & x_{22}^* & \dots & x_{p2}^* \\ \vdots & \vdots & \dots & \vdots \\ x_{1n}^* & x_{2n}^* & \dots & x_{pn}^* \end{bmatrix}$$

ขั้นที่ 2 ประเมินค่า $\hat{\mathbf{y}}_{(1)j}^*$ จากสมการความถดถอยด้วยวิธีกำลังสองน้อยที่สุด ได้จาก

$$\hat{\mathbf{y}}_{(1)j}^* = b_{(1)j} \tilde{\mathbf{x}}_{(1)j}^*$$

โดยที่ $b_{(1)j} = (\tilde{\mathbf{x}}_{(1)j}^* \tilde{\mathbf{x}}_{(1)j}^*)^{-1} \tilde{\mathbf{x}}_{(1)j}^* \mathbf{y}_{(1)}^*$ คือ เวกเตอร์ค่าความสัมพันธ์ของ เวกเตอร์ของตัวแปรอิสระ \mathbf{x}_1^* และเวกเตอร์ของตัวแปรตาม \mathbf{y}^*

เราจะได้สมการความถดถอยที่ประมาณค่า $\hat{\mathbf{y}}_{(1)j}^*$ ทั้งหมด p สมการ ดังนี้

$$\hat{\mathbf{y}}_{(1)1}^* = b_{(1)1} \tilde{\mathbf{x}}_{(1)1}^*$$

$$\hat{\mathbf{y}}_{(1)2}^* = b_{(1)2} \tilde{\mathbf{x}}_{(1)2}^*$$

⋮

$$\hat{\mathbf{y}}_{(1)p}^* = b_{(1)p} \tilde{\mathbf{x}}_{(1)p}^*$$

ขั้นที่ 3 คำนวณค่าตัวแปรอิสระใหม่ โดยการรวมสมการประมาณค่า $\hat{\mathbf{y}}_{(1)j}^*$ โดยไม่คำนึงถึงสหสัมพันธ์ระหว่างตัวแปรอิสระ $\mathbf{x}_{(1)j}^*$ สามารถกำหนดค่าตัวแปรอิสระใหม่ได้ ดังนี้

$$\mathbf{p}_1 = \sum_{j=1}^p w_{(1)j} b_{(1)j} \tilde{\mathbf{x}}_{(1)j}^*$$

$$\mathbf{p}_1 = \sum_{j=1}^p w_{(1)j} \hat{\mathbf{y}}_{(1)j}^*$$

$$\mathbf{p}_1 = w_{(1)1} \hat{\mathbf{y}}_{(1)1}^* + w_{(1)2} \hat{\mathbf{y}}_{(1)2}^* + \dots + w_{(1)p} \hat{\mathbf{y}}_{(1)p}^*$$

$$\text{โดยที่ } w_{(1)j} = \tilde{\mathbf{x}}_{(1)j}^* \tilde{\mathbf{x}}_{(1)j}^*$$

ขั้นตอนในการสร้างตัวแปรอิสระใหม่ $\mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_p$ จะเป็นการทำซ้ำขั้นที่ 2-3 และตัวแปรอิสระใหม่แต่ละตัวจะพิจารณาจากส่วนที่เหลือของการถดถอยก่อนหน้านี้ โดยคำนึงถึงความผันแปรส่วนที่เหลือของตัวแปรตาม \mathbf{y} ที่สัมพันธ์กับข้อมูลที่เหลือของตัวแปรอิสระ \mathbf{x}

กำหนดให้ \mathbf{p}_i โดย $i \geq 1$ เป็นค่าตัวอย่างของตัวแปรอิสระ \mathbf{P}_i และเป็นตัวแปรอิสระตัวก่อนหน้าที่สร้าง ในการสร้างตัวแปรอิสระตัวถัดไป \mathbf{p}_{i+1} จะพิจารณาจากเวกเตอร์ตัวแปร $\tilde{\mathbf{y}}_{(i+1)}^*$

ซึ่งเป็นส่วนที่เหลือของการถดถอยก่อนหน้าและ $\tilde{\mathbf{x}}_{(i+1)j}^*$ ซึ่งเป็นส่วนที่เหลือของการถดถอยก่อนหน้าเช่นกัน สำหรับ $j = 1, 2, \dots, p$ เราสามารถหา $\tilde{\mathbf{y}}_{(i+1)}^*$ ได้ดังนี้

$$\tilde{\mathbf{y}}_{(i+1)}^* = \tilde{\mathbf{y}}_{(i)}^* - \{(\mathbf{p}'_i \mathbf{p}_i)^{-1} \mathbf{p}'_i \tilde{\mathbf{y}}_{(i)}^*\} \mathbf{p}_i$$

และสามารถหา $\tilde{\mathbf{x}}_{(i+1)j}^*$ ได้ดังนี้

$$\tilde{\mathbf{x}}_{(i+1)j}^* = \tilde{\mathbf{x}}_{(i)j}^* - \{(\mathbf{p}'_i \mathbf{p}_i)^{-1} \mathbf{p}'_i \tilde{\mathbf{x}}_{(i)j}^*\} \mathbf{p}_i$$

ทำการประมาณค่า $\hat{\mathbf{y}}_{(i+1)j}^*$ จากสมการความถดถอยด้วยวิธีกำลังสองน้อยสุด ดังนี้

$$\hat{\mathbf{y}}_{(i+1)j}^* = b_{(i+1)j} \tilde{\mathbf{x}}_{(i+1)j}^*$$

$$\text{โดยที่ } b_{(i+1)j} = (\tilde{\mathbf{x}}_{(i+1)j}^* \tilde{\mathbf{x}}_{(i+1)j}^*)^{-1} \tilde{\mathbf{x}}_{(i+1)j}^* \tilde{\mathbf{y}}_{(i+1)}^*$$

ซึ่งเราสามารถสร้างรูปแบบของผลรวมเชิงเส้นของสมการประมาณค่า $\hat{\mathbf{y}}_{(i+1)j}^*$ ทำให้ได้ตัวแปรอิสระใหม่ตัวถัดไป ดังนี้

$$\mathbf{p}_{i+1} = \sum_{j=1}^p w_{(i+1)j} b_{(i+1)j} \tilde{\mathbf{x}}_{(i+1)j}^*$$

$$\mathbf{p}_{i+1} = \sum_{j=1}^p w_{(i+1)j} \hat{\mathbf{y}}_{(i+1)j}^*$$

$$\mathbf{p}_{i+1} = w_{(i+1)1} \hat{\mathbf{y}}_{(i+1)1}^* + w_{(i+1)2} \hat{\mathbf{y}}_{(i+1)2}^* + \dots + w_{(i+1)p} \hat{\mathbf{y}}_{(i+1)p}^*$$

$$\text{โดยที่ } w_{(i+1)j} = \tilde{\mathbf{x}}_{(i+1)j}^* \tilde{\mathbf{x}}_{(i+1)j}^*$$

เราจะทำซ้ำวิธีการเดิมทั้งหมดเพื่อหาตัวแปรอิสระใหม่ จนกระทั่งเราได้จำนวนตัวแปรอิสระทุกตัว $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_p$ ซึ่งตัวแปรอิสระใหม่ที่ได้จะมีความสัมพันธ์กับตัวแปรตาม y และตัวแบบความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนตัวอย่างสามารถเขียนได้ ดังนี้

$$\hat{\mathbf{y}} = b_0 + b_1 \mathbf{p}_{1i} + b_2 \mathbf{p}_{2i} + \dots + b_p \mathbf{p}_{pi}$$

หรือเขียนในรูปของเมทริกซ์ได้ ดังนี้

$$\hat{\mathbf{y}} = \mathbf{P}\tilde{\mathbf{y}}$$

$$\text{เมื่อ } \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 1 & p_{11} & p_{21} & \cdots & p_{p1} \\ 1 & p_{12} & p_{22} & \cdots & p_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & p_{1n} & p_{2n} & \cdots & p_{pn} \end{bmatrix}, \tilde{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

จากนั้นทำการประมาณค่าสัมประสิทธิ์ความถดถอยในตัวแบบความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนด้วยวิธีกำลังสองน้อยสุด

2.6 ตัวแบบวิธีกำลังสองน้อยสุด ด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์

วิธีกำลังสองน้อยสุดเป็นวิธีการประมาณค่าพารามิเตอร์ ซึ่งเป็นวิธีที่ต้องการทำให้ค่าผลบวกกำลังสองของความคลาดเคลื่อน SSE มีค่าต่ำสุด สำหรับการศึกษาคั้งนี้จะศึกษาภายใต้การเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์ในการสร้างเวกเตอร์ของตัวแปรอิสระใหม่ให้ตั้งฉากกัน โดยการหมุนแกนให้ตัวแปรอิสระใหม่ที่ได้มีความเป็นอิสระจากกัน จากนั้นประมาณค่าด้วยวิธีกำลังสองน้อยสุด ซึ่งมีตัวอย่างจากงานวิจัยของ เปรมวดี ชูไสว (2548)³ ได้ศึกษาการเปรียบเทียบวิธีการแก้ไขพหุสัมพันธ์ในการถดถอยเชิงเส้นพหุคูณ โดยใช้วิธีการแปลงข้อมูลแบบกรามซมิตต์ในการแปลงข้อมูลเมทริกซ์ตัวแปรอิสระ X ให้แต่ละเวกเตอร์ตั้งฉากกัน เพื่อนำมาใช้ในการประมาณค่าพารามิเตอร์ด้วยกำลังสองน้อยสุด ซึ่งเป็นวิธีที่ช่วยแก้ไขปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ สำหรับตัวแบบความถดถอยของวิธีกำลังสองน้อยสุดสามารถเขียนได้ ดังนี้

$$\hat{Y} = \beta_0 + \beta_1 P_{1i} + \beta_2 P_{2i} + \cdots + \beta_p P_{pi}$$

โดยที่ $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_p$ เป็นตัวแปรที่ได้มาจาก $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_p$ ซึ่งเป็นตัวแปรอิสระใหม่ที่ตั้งฉากกัน และค่าสัมพัทธ์ตัวอย่างสำหรับแต่ละคู่ของตัวแปรอิสระใหม่มีค่าเท่ากับศูนย์ ส่งผลให้การวิเคราะห์ความถดถอยมีความแม่นยำมากขึ้น กล่าวคือ ความแปรปรวนของสัมประสิทธิ์ความถดถอยมีค่าลดลง แต่ค่าประมาณสัมประสิทธิ์ความถดถอยที่ได้ก็ยังคงมีความเอนเอียงเกิดขึ้นสามารถสรุปขั้นตอนในการคำนวณ (Algorithm) โดยใช้หลักการของพีชคณิตเชิงเส้น (Linear Algebra) ด้วยกระบวนการแปลงข้อมูลแบบกรามซมิตต์ ซึ่งเป็นการสร้างตัวแปรใหม่ให้ตั้งฉากกันโดยการสร้างมูลฐานเชิงตั้งฉากปกติ (Orthonormal basis) ได้ดังนี้

³ เปรมวดี ชูไสว. การเปรียบเทียบวิธีการแก้ไขพหุสัมพันธ์ในการถดถอยเชิงเส้นพหุคูณ, หลักสูตรปริญญาโท สาขาวิชาสถิติ, จุฬาลงกรณ์มหาวิทยาลัย, 2548.

สำหรับ p มิติของผลคูณภายในปริภูมิเวกเตอร์ V เริ่มต้นด้วยการกำหนดค่าพื้นฐาน $\{x_1^*, x_2^*, \dots, x_p^*\}$ จากเมทริกซ์ x^* และวิธีการสร้างเป็นมูลฐานเชิงตั้งฉาก $\{p_1, p_2, \dots, p_p\}$ แสดงด้วยขั้นตอน ดังนี้

ขั้นที่ 1 ให้
$$p_1 = x_1^*$$

ขั้นที่ 2 เป็นการสร้างเวกเตอร์ p_2 โดยที่ $\langle p_2, p_1 \rangle = 0$ ให้

$$p_2 = x_2^* - \text{proj}_{p_1}(x_2^*)$$

เพราะฉะนั้นมี $\{p_1, p_2\}$ เป็นเซตเชิงตั้งฉากปกติ

ขั้นที่ 3 เป็นการสร้างเวกเตอร์ p_3 โดยที่ $\langle p_3, p_1 \rangle = 0$, $\langle p_3, p_2 \rangle = 0$ ให้

$$p_3 = x_3^* - \text{proj}_{p_1}(x_3^*) - \text{proj}_{p_2}(x_3^*)$$

เพราะฉะนั้นมี $\{p_1, p_2, p_3\}$ เป็นเซตเชิงตั้งฉากปกติ โดยวิธีอุปนัยทางคณิตศาสตร์ ทำให้ $\{p_1, p_2, \dots, p_p\}$ เป็นเซตเชิงตั้งฉากปกติ

ขั้นที่ p เป็นการสร้างเวกเตอร์ p_p ใน x^* ซึ่ง p_1, p_2, \dots, p_p เป็นเซตเชิงตั้งฉากปกติ ให้

$$p_p = x_p^* - \sum_{j=1}^{p-1} \text{proj}_{p_j}(x_p^*)$$

โดยกำหนดตัวดำเนินการฉาย (Projection operator) เป็นดังนี้

$$\text{proj}_p(x^*) = \frac{\langle x^*, p \rangle}{\langle p, p \rangle} p$$

และเวกเตอร์ e_i เป็นเวกเตอร์เชิงตั้งฉากปกติ คือ $e_j = \frac{p_j}{\|p_j\|}$

ดังนั้น การสร้างเวกเตอร์ของตัวแปรอิสระใหม่ให้ตั้งฉากกันด้วยวิธีการแปลงข้อมูลแบบกรามชmidt แล้ว จะทำให้ได้เมทริกซ์ของข้อมูลตัวแปรอิสระใหม่ที่เป็นอิสระกัน และสามารถเขียนมูลฐานเชิงตั้งฉาก $\{p_1, p_2, \dots, p_{p-1}, p_p\}$ ในรูปของเมทริกซ์ได้ ดังนี้

$$P^* = \begin{bmatrix} p_{11} & p_{21} & \dots & p_{p1} \\ p_{12} & p_{22} & \dots & p_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1n} & p_{2n} & \dots & p_{pn} \end{bmatrix}$$

ซึ่งจะเห็นได้ว่าข้อดีของกระบวนการแปลงข้อมูลแบบกรามซมิตต์ คือ ทำให้สหสัมพันธ์ ตัวอย่างสำหรับแต่ละคู่ของตัวแปรอิสระใหม่ที่ได้มีค่าเท่ากับศูนย์ เป็นผลให้ตัวแปรอิสระใหม่ที่ได้ ตั้งฉากกัน โดยตัวแปรอิสระใหม่ที่ได้จะมีความสัมพันธ์กับตัวแปรตาม y และสามารถเขียนตัวแบบ ความถดถอยด้วยวิธีกำลังสองน้อยสุดตัวอย่างได้ ดังนี้

$$\hat{y} = b_0 + b_1 p_{1i} + b_2 p_{2i} + \cdots + b_p p_{pi}$$

หรือเขียนในรูปของเมทริกซ์ได้ ดังนี้

$$\hat{\mathbf{y}} = \mathbf{P}\tilde{\mathbf{b}}$$

$$\text{เมื่อ } \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 1 & p_{11} & p_{21} & \cdots & p_{p1} \\ 1 & p_{12} & p_{22} & \cdots & p_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & p_{1n} & p_{2n} & \cdots & p_{pn} \end{bmatrix}, \tilde{\mathbf{b}} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}$$

จากนั้นทำการประมาณค่าสัมประสิทธิ์ความถดถอยในตัวแบบความถดถอยด้วยวิธีกำลังสองน้อยสุด

2.7 การประมาณค่าสัมประสิทธิ์ความถดถอย

สามารถประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยได้ ดังนี้

$$\tilde{\mathbf{b}} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \hat{\mathbf{y}}$$

นำค่า $\tilde{\mathbf{b}} = (b_0, b_1, b_2, \dots, b_p)'$ แทนในตัวแบบ $\hat{y} = b_0 + b_1 p_{1i} + b_2 p_{2i} + \cdots + b_p p_{pi}$ เพื่อทำการพยากรณ์ค่า \hat{y} จากนั้นทำการหาค่าความแปรปรวน (σ^2) โดยจะประมาณค่าความแปรปรวนด้วย $\hat{\sigma}^2$ ซึ่งสามารถหาได้ ดังนี้

$$\hat{\sigma}^2 = s^2 = MSE = \frac{\sum_{i=1}^n e_i^2}{n-p-1} = \frac{1}{n-p-1} (\tilde{\mathbf{e}}' \tilde{\mathbf{e}})$$

จะได้ตัวประมาณพารามิเตอร์แบบที่มีคุณสมบัติที่ไม่เอนเอียง นั่นคือ $E(MSE) = \sigma^2$ และทำการหาค่าความแปรปรวนของ $\tilde{\mathbf{b}}$ โดยจะหาได้ ดังนี้

$$\hat{Cov}(\tilde{\mathbf{b}}) = \hat{\sigma}^2 (\mathbf{x}^T \mathbf{x})^{-1} = MSE (\mathbf{x}^T \mathbf{x})^{-1} = \begin{bmatrix} \hat{V}(b_0) & \hat{C}(b_0, b_1) & \cdots & \hat{C}(b_0, b_p) \\ \hat{C}(b_1, b_0) & \hat{V}(b_1) & \cdots & \hat{C}(b_1, b_p) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}(b_p, b_0) & \hat{C}(b_p, b_1) & \cdots & \hat{V}(b_p) \end{bmatrix}$$

และค่าความเอนเอียงของตัวประมาณ b_j โดยที่ความเอนเอียงของตัวประมาณ คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ $MSE(\tilde{b})$ จะไม่เท่ากับค่าความแปรปรวนของตัวประมาณ $Var(\tilde{b})$ สามารถคำนวณได้ ดังนี้

$$Bias(b_j) = \beta_j - b_j \quad ; j = 0, 1, 2, \dots, p$$

2.8 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) เกิดจากผลบวกของความแปรปรวนของตัวประมาณกับกำลังสองของความเอนเอียง ซึ่งเป็นเกณฑ์ที่ใช้ในการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์ความถดถอยว่าตัวประมาณเป็นตัวประมาณที่ดี สามารถแสดงความสัมพันธ์ระหว่างค่าความแปรปรวนของตัวประมาณ และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณได้ ดังนี้

$$MSE(\tilde{b}) = Var(\tilde{b}) + (Bias(\tilde{b}))^2$$

หรือเขียนในรูปเมทริกซ์ได้ ดังนี้

$$\begin{bmatrix} MSE(b_0) \\ MSE(b_1) \\ \vdots \\ MSE(b_p) \end{bmatrix} = \begin{bmatrix} V(b_0) & C(b_0, b_1) & \dots & C(b_0, b_p) \\ C(b_1, b_0) & V(b_1) & \dots & C(b_1, b_p) \\ \vdots & \vdots & \ddots & \vdots \\ C(b_p, b_0) & C(b_p, b_1) & \dots & V(b_p) \end{bmatrix} + \begin{bmatrix} (bias(b_0))^2 & 0 & \dots & 0 \\ 0 & (bias(b_1))^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (bias(b_p))^2 \end{bmatrix}$$

ในการศึกษาครั้งนี้ได้นำค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ มาเป็นเกณฑ์ในการวัดประสิทธิภาพของวิธีการประมาณทั้งสองวิธี

บทที่ 3

วิธีดำเนินการศึกษา

การศึกษานี้มีเป้าหมายเพื่อหาวิธีการประมาณค่าพารามิเตอร์ที่เหมาะสม ระหว่างวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชmidt ในกรณีต่างๆ ซึ่งทำการจำลองข้อมูลโดยใช้โปรแกรม R และศึกษาภายใต้เงื่อนไขต่างๆ ดังนี้

3.1 ขอบเขตการศึกษา

1. กำหนดจำนวนตัวแปรอิสระที่ทำการศึกษา คือ 2 และ 3 ตัวแปร
2. กำหนดขนาดตัวอย่างที่ทำการศึกษา คือ 50, 100 และ 200
3. กำหนดค่าสัมประสิทธิ์ความถดถอยเบื้องต้น โดยให้ค่า $\beta_j = 1$ สำหรับ $j = 0, 1, \dots, p$
4. กำหนดระดับความสัมพันธ์ระหว่างตัวแปรอิสระให้มีความสัมพันธ์แบบเชิงเส้นเท่านั้น
 - 4.1 โดยถ้าจำนวนตัวแปรอิสระเป็น 2 ตัวแปร ผู้วิจัยจะทำการศึกษาภายใต้ 10 ระดับ คือ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99
 - 4.2 โดยถ้าจำนวนตัวแปรอิสระเป็น 3 ตัวแปร ผู้วิจัยจะทำการศึกษาภายใต้ 5 ระดับ คือ 0.91, 0.93, 0.95, 0.97 และ 0.99
5. กำหนดค่าความแปรปรวนของความคลาดเคลื่อน σ^2 มีค่าเป็น 10 และค่าความแปรปรวนของตัวแปรอิสระ X_1 , X_2 และ X_3 มีค่าเป็น $\sigma_1^2 = 4$, $\sigma_2^2 = 4$ และ $\sigma_3^2 = 4$ ตามลำดับ
6. กำหนดจำนวนตัวแปรอิสระที่มีพหุสัมพันธ์และระดับความสัมพันธ์ของตัวแปรเป็น 2 กรณีดังนี้
 - 6.1 กรณีที่จำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร ผู้วิจัยได้ทำการศึกษาเมื่อตัวแปรอิสระ 2 ตัวแปร ได้แก่ X_1 และ X_2 มีพหุสัมพันธ์กัน โดยให้ ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_2 ซึ่งศึกษาภายใต้ระดับความสัมพันธ์ คือ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99

6.2 กรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร ผู้วิจัยได้ทำการศึกษาเมื่อตัวแปรอิสระ 3 ตัวแปร ได้แก่ X_1 , X_2 และ X_3 มีพหุสัมพันธ์กัน

ให้ $(\rho_{12}, \rho_{13}, \rho_{23})$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดย

ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_2

ρ_{13} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_3

ρ_{23} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_2 และ X_3

ซึ่งศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

(0.91, 0.91, 0.91)	(0.91, 0.91, 0.93)	(0.91, 0.91, 0.95)
(0.91, 0.91, 0.97)	(0.91, 0.91, 0.99)	(0.91, 0.93, 0.93)
(0.91, 0.93, 0.95)	(0.91, 0.93, 0.97)	(0.91, 0.93, 0.99)
(0.91, 0.95, 0.95)	(0.91, 0.95, 0.97)	(0.91, 0.95, 0.99)
(0.91, 0.97, 0.97)	(0.93, 0.93, 0.93)	(0.93, 0.93, 0.95)
(0.93, 0.93, 0.97)	(0.93, 0.93, 0.99)	(0.93, 0.95, 0.95)
(0.93, 0.95, 0.97)	(0.93, 0.95, 0.99)	(0.93, 0.97, 0.97)
(0.93, 0.97, 0.99)	(0.95, 0.95, 0.95)	(0.95, 0.95, 0.97)
(0.95, 0.95, 0.99)	(0.95, 0.97, 0.97)	(0.95, 0.97, 0.99)
(0.97, 0.97, 0.97)	(0.97, 0.97, 0.99)	(0.97, 0.99, 0.99)
(0.99, 0.99, 0.99)		

3.2 วิธีดำเนินการศึกษา

1. กำหนดค่าเริ่มต้นในการจำลองข้อมูล ดังนี้
 - 1.1 กำหนดจำนวนตัวแปรอิสระ p ตัว และจำนวนตัวแปรตาม 1 ตัว
 - 1.2 กำหนดขนาดตัวอย่าง n ตัว
 - 1.3 กำหนดค่าสัมประสิทธิ์ความถดถอยเบื้องต้น (β_j) สำหรับ $j = 0, 1, 2, \dots, p$
 - 1.4 กำหนดระดับความสัมพันธ์ระหว่างตัวแปรอิสระเป็นคู่ ρ_{ij} สำหรับ $i = 1, 2, \dots, n$ และ $j = 1, 2, \dots, p$

- 1.5 กำหนดค่าความแปรปรวนของความคลาดเคลื่อน (σ^2) และค่าความแปรปรวนของตัวแปรอิสระ X_1 , X_2 และ X_3
2. จำลองข้อมูลตัวแปรอิสระของการแจกแจงปกติหลายตัวแปร โดยตัวแปรอิสระสร้างจากค่าเฉลี่ย ค่าความแปรปรวนและระดับความสัมพันธ์ที่กำหนด ด้วยวิธีของโซเลสกี้ ดังนี้

$$X_1 = Z_1$$

$$X_2 = \rho_{12}Z_1 + \sqrt{1 - \rho_{12}^2}Z_2$$

$$X_3 = \rho_{12}Z_1 + \frac{(\rho_{23} - \rho_{12}\rho_{13})}{\sqrt{1 - \rho_{12}^2}}Z_2 + \sqrt{1 - \rho_{13}^2 - \frac{(\rho_{23} - \rho_{12}\rho_{13})^2}{1 - \rho_{12}^2}}Z_3$$

โดยที่ Z_1, Z_2, Z_3 เป็นตัวแปรสุ่มแบบปกติมาตรฐานที่เป็นอิสระกัน

3. สร้างเวกเตอร์ของข้อมูลตัวแปรตาม Y จากตัวแบบความถดถอยพหุเชิงเส้น
4. คำนวณค่า VIF ซึ่งพิจารณาจากสัมประสิทธิ์สหสัมพันธ์พหุของตัวแปรอิสระ โดยเกณฑ์ในการใช้ตัดสินใจว่าเกิดปัญหา Collinearity สามารถคำนวณตามมาตรวัดทางสถิติได้ ดังนี้

$$VIF_j = \frac{1}{1 - R_j^2} \quad ; \quad j = 1, 2, \dots, p$$

โดยที่ R_j^2 เป็นสัมประสิทธิ์สหสัมพันธ์พหุของตัวแปรอิสระ P_j กับตัวแปรอิสระอื่นๆ

ถ้าค่า $VIF_j \leq 10$; $j = 1, 2, \dots, p$ แสดงว่าไม่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

ถ้าค่า $VIF_j > 10$; $j = 1, 2, \dots, p$ แสดงว่าเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

5. นำเวกเตอร์ของตัวแปรตาม Y และนำเมทริกซ์ของตัวแปรอิสระ \mathbf{X} ที่ได้มาทำ centering เพื่อให้ค่า Intercept หายไปเหลือเพียงค่า X อย่างเดียว
6. สร้างตัวแปรอิสระใหม่ให้ตั้งฉากกันด้วยการแปลงข้อมูลแบบกรามชมิตต์ และใช้วิธีกำลังสองน้อยสุดเชิงส่วนในการแก้ไขปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ
7. ประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด จากสูตรดังนี้

$$\tilde{\mathbf{b}} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \tilde{\mathbf{y}}$$

8. หาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองการทำซ้ำจำนวน 1,000 รอบ ทั้ง 2 วิธี จากสูตรดังนี้

$$AMSE(PLS) = \frac{1}{N} \sum_{i=1}^N MSE(PLS)$$

และ

$$AMSE(OLS_G) = \frac{1}{N} \sum_{i=1}^N MSE(OLS_G)$$

โดยที่ MSE คือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

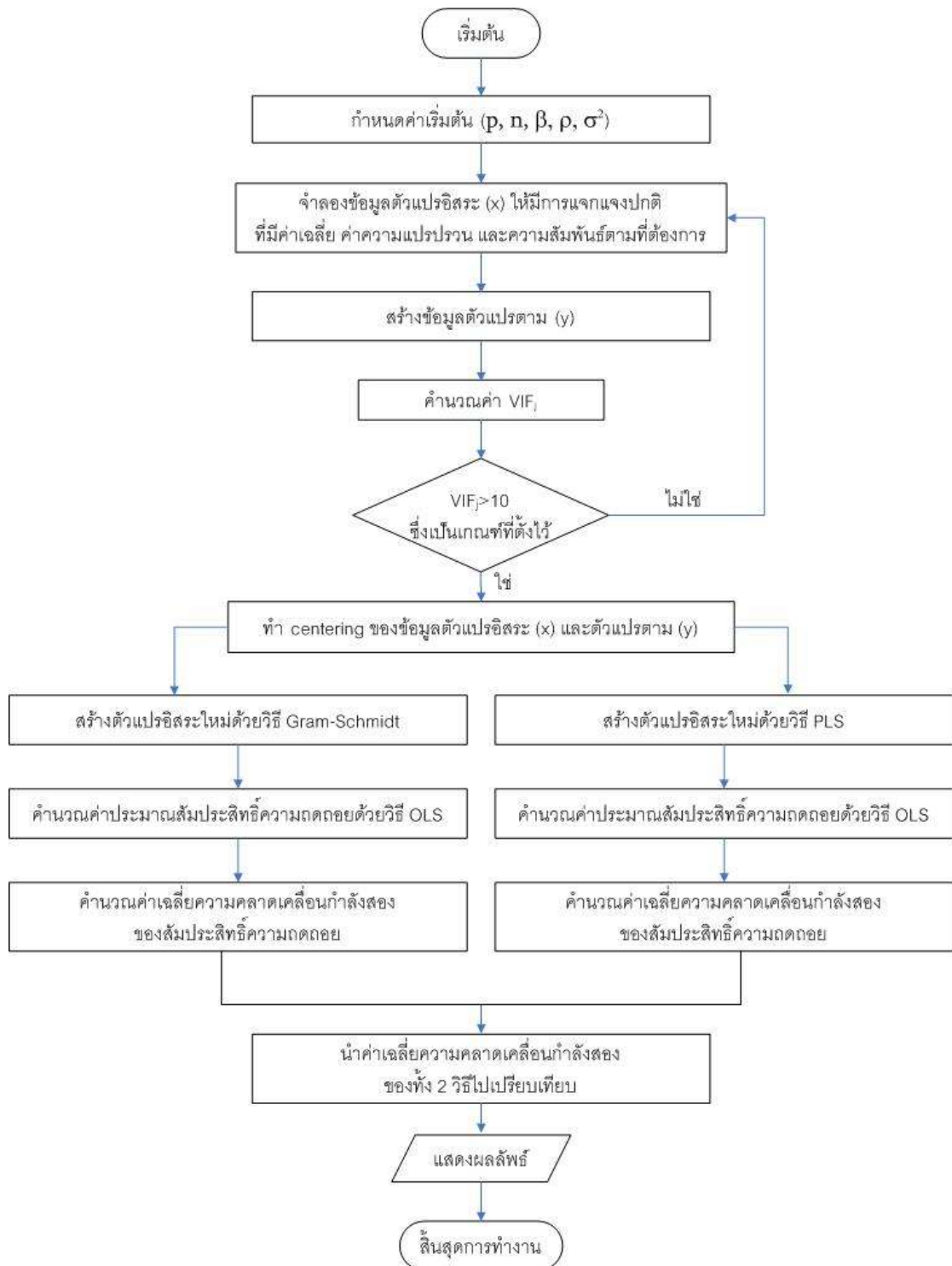
N คือ จำนวนรอบของการทำซ้ำ

เมื่อ
$$MSE = \frac{1}{p} \sum_{j=0}^p (Var(b_j) + (Bias(b_j))^2)$$

9. เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของวิธีการประมาณ จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของทั้ง 2 วิธี

10. สรุปผลการศึกษา

แผนผังการเขียนโปรแกรมที่ทำการศึกษา



บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในงานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประมาณค่าพารามิเตอร์ที่เหมาะสม โดยเปรียบเทียบ จากค่าความคลาดเคลื่อนมาตรฐานและค่าความเอนเอียงของตัวประมาณ ในแต่ละกรณีที่ทำการศึกษาจะขึ้นอยู่กับระดับความสัมพันธ์ระหว่างตัวแปรอิสระ จำนวนตัวแปรอิสระ และขนาดตัวอย่าง ซึ่งทำการศึกษาภายใต้การวิเคราะห์ความถดถอยพหุเชิงเส้น (Multiple Linear Regression) โดยทำการจำลองข้อมูลต่างๆ สำหรับการศึกษาค้นคว้าครั้งนี้ด้วยโปรแกรม R

การวิเคราะห์ผลการศึกษานี้แบ่งออกเป็น 3 ส่วน ดังนี้

1. ผลการศึกษาจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง
2. ผลการศึกษาเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรง (Multicollinearity)
3. ผลการศึกษาการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ โดยผลการศึกษาแต่ละส่วนจะทำการศึกษาภายใต้กรณีต่างๆ ดังนี้

ส่วนที่ 1 ผลการศึกษาจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง

1.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร มีทั้งหมด 30 กรณี

ตารางที่	จำนวนตัวแปรอิสระ	ระดับความสัมพันธ์	ขนาดตัวอย่าง
4.1.1	2	0.9, 0.91, 0.92, 0.93,	n=50
		0.94, 0.95, 0.96,	n=100
		0.97, 0.98, 0.99	n=200

1.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร มีทั้งหมด 93 กรณี

ตารางที่	จำนวนตัวแปรอิสระ	ระดับความสัมพันธ์	ขนาดตัวอย่าง
4.1.2	3	0.91, 0.93, 0.95, 0.97, 0.99	n=50
4.1.3			n=100
4.1.4			n=200

ส่วนที่ 2 ผลการศึกษาเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรง (Multicollinearity)

2.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร มีทั้งหมด 30 กรณี

ตารางที่	จำนวนตัวแปรอิสระ	ระดับความสัมพันธ์	ขนาดตัวอย่าง
4.2.1	2	0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99	n=50
			n=100
			n=200

2.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร มีทั้งหมด 93 กรณี

ตารางที่	จำนวนตัวแปรอิสระ	ระดับความสัมพันธ์	ขนาดตัวอย่าง
4.2.2	3	0.91, 0.93, 0.95, 0.97, 0.99	n=50
4.2.3			n=100
			n=200

ส่วนที่ 3 ผลการศึกษาการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ

3.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร มีทั้งหมด 60 กรณี

ตารางที่	จำนวนตัวแปรอิสระ	ระดับความสัมพันธ์	ขนาดตัวอย่าง
4.3.1	2	0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99	n=50
4.3.2			n=100
			n=200

3.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร มีทั้งหมด 186 กรณี

ตารางที่	จำนวนตัวแปรอิสระ	ระดับความสัมพันธ์	ขนาดตัวอย่าง
4.3.3	3	0.91, 0.93, 0.95, 0.97, 0.99	n=50
4.3.4			n=100
4.3.5			n=200

ดังนั้น กรณีที่ทำการศึกษาทั้งหมดในแต่ละส่วนสามารถอธิบายผลได้ดังต่อไปนี้

4.1 ผลการศึกษาจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง

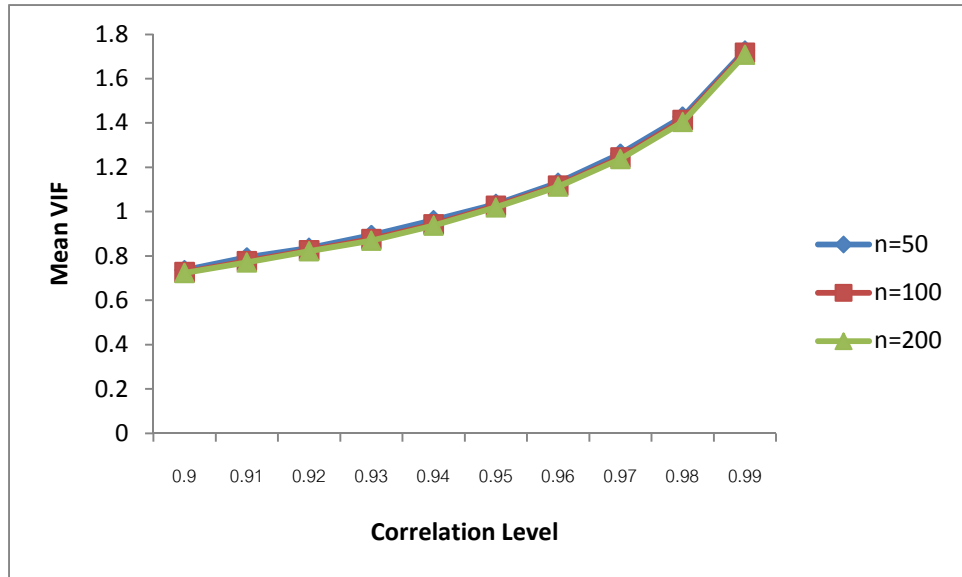
4.1.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร

จากการจำลองข้อมูลและการวิเคราะห์ผล สำหรับข้อมูลของตัวแปรอิสระที่มีพหุสัมพันธ์รุนแรงตามระดับความสัมพันธ์ต่างๆ ที่กำหนด เมื่อพิจารณา ค่าเฉลี่ยของ VIF ของแต่ละกรณีที่ทำการศึกษา จะมีผลการวิเคราะห์ดังตารางที่ 4.1.1

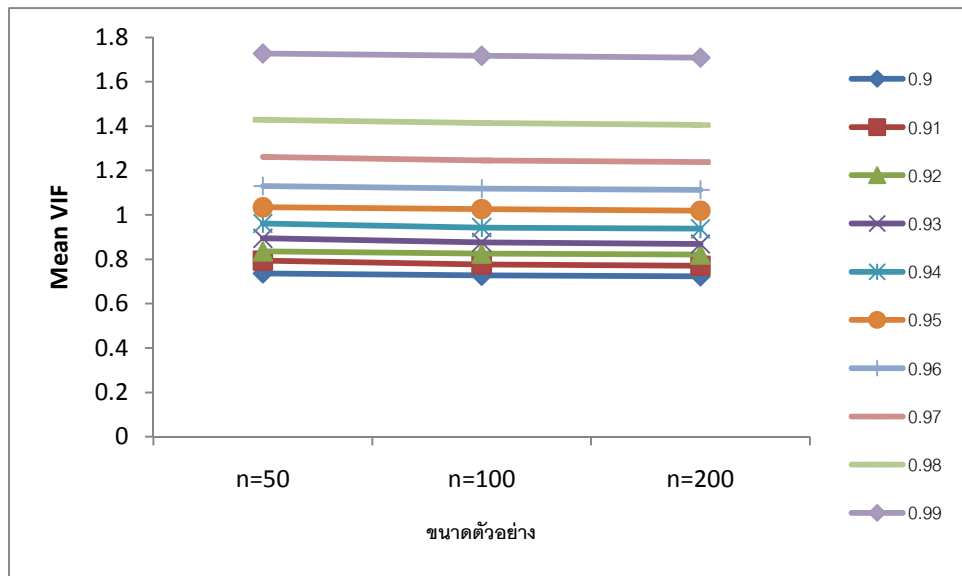
ตารางที่ 4.1.1 แสดงค่าเฉลี่ยของ VIF กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร

ระดับ ความสัมพันธ์	n=50		n=100		n=200	
	VIF1	VIF2	VIF1	VIF2	VIF1	VIF2
0.9	5.4435	5.4435	5.3304	5.3304	5.2857	5.2857
0.91	6.2150	6.2150	5.9846	5.9846	5.9004	5.9004
0.92	6.8531	6.8531	6.6972	6.6972	6.6301	6.6301
0.93	7.8389	7.8389	7.5194	7.5194	7.3991	7.3991
0.94	9.1486	9.1486	8.7682	8.7682	8.6605	8.6605
0.95	10.8193	10.8193	10.6085	10.6085	10.4399	10.4399
0.96	13.4987	13.4987	13.1183	13.1183	12.9780	12.9780
0.97	18.2553	18.2553	17.5762	17.5762	17.2889	17.2889
0.98	26.8108	26.8108	25.9457	25.9457	25.3985	25.3985
0.99	53.3150	53.3150	52.0815	52.0815	51.0690	51.0690

รูปที่ 4.1.1 การเปรียบเทียบค่าเฉลี่ยของ VIF แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



รูปที่ 4.1.2 การเปรียบเทียบค่าเฉลี่ยของ VIF แยกตามระดับความสัมพันธ์ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



ผลการวิเคราะห์พบว่า เมื่อพิจารณาค่าเฉลี่ยของ VIF ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร ที่ระดับความสัมพันธ์เท่ากับ 0.95 จะเป็นจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาของความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระที่ทำการศึกษา ดังนั้น สรุปได้ว่าจะเกิดปัญหาของความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระที่ระดับความสัมพันธ์ 0.95, 0.96, 0.97, 0.98 และ 0.99 สำหรับในการศึกษาครั้งนี้

เมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF แยกตามขนาดตัวอย่าง พบว่าที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยของ VIF สูงที่สุด รองลงมาคือขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยของ VIF ต่ำที่สุด นั่นคือ ค่าเฉลี่ยของ VIF จะแปรผกผันกับขนาดตัวอย่างของทุกระดับความสัมพันธ์ที่ทำการศึกษา ซึ่งแสดงได้ดังรูปที่ 4.1.1

เมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF แยกตามระดับความสัมพันธ์ พบว่าที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยของ VIF สูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.98, 0.97, 0.96, 0.95, 0.94, 0.93, 0.92 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.91 จะมีค่าเฉลี่ยของ VIF ต่ำที่สุด นั่นคือ ค่าเฉลี่ยของ VIF จะแปรผันตรงกับระดับความสัมพันธ์ของทุกขนาดตัวอย่างที่ทำการศึกษา ซึ่งแสดงได้ดังรูปที่ 4.1.2

4.1.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร

จากการจำลองข้อมูลและการวิเคราะห์ผล สำหรับข้อมูลของตัวแปรอิสระที่มีพหุสัมพันธ์รุนแรงตามระดับความสัมพันธ์ต่างๆ ที่กำหนด เมื่อพิจารณาค่าเฉลี่ยของ VIF ของแต่ละกรณีที่ทำการศึกษา จะมีผลการวิเคราะห์ดังตารางที่ 4.1.2 และเพื่อให้ง่ายในการวิเคราะห์ผลทางผู้วิจัย จึงได้พิจารณาจากระดับความสัมพันธ์สูงสุดของแต่ละกรณี โดยพิจารณาจากค่าเฉลี่ยของ VIF ของแต่ละระดับความสัมพันธ์สูงสุด ดังตารางที่ 4.1.3-4.1.4

ตารางที่ 4.1.2 แสดงค่าเฉลี่ยของ VIF กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์	n=50			n=100			n=200		
	VIF1	VIF2	VIF3	VIF1	VIF2	VIF3	VIF1	VIF2	VIF3
(0.91,0.93,0.95)	8.3880	11.7032	14.9176	8.0679	11.3328	14.3929	7.9089	11.0800	14.0696
(0.91,0.93,0.97)	8.0025	18.3750	23.3862	7.6353	17.6239	22.3705	7.5686	17.3759	22.0708
(0.91,0.93,0.99)	8.5044	57.2131	72.9752	8.0870	54.5853	69.4700	7.9066	53.5819	68.0938
(0.91,0.95,0.97)	11.3066	18.6931	33.0023	10.8384	17.8482	31.4943	10.6104	17.5924	30.9996
(0.91,0.95,0.99)	21.5886	106.0637	187.1937	20.6261	101.3948	178.7522	20.0862	98.1358	173.1508
(0.93,0.95,0.97)	11.3254	18.5643	25.9849	10.8913	18.0152	25.1149	10.5521	17.5279	24.2901
(0.93,0.95,0.99)	11.8295	58.7841	81.2650	11.2578	55.6894	77.0217	11.0822	54.5410	75.5372
(0.93,0.97,0.99)	83.8579	251.1671	572.4963	80.6829	239.2981	547.3856	78.9390	234.4539	536.0342
(0.95,0.97,0.99)	20.2542	59.9477	98.6633	19.4750	57.5229	94.7832	19.0290	56.3961	92.9475
(0.91,0.91,0.91)	8.1309	8.0165	8.0976	7.7648	7.7216	7.7159	7.6526	7.5922	7.6182
(0.93,0.93,0.93)	10.4780	10.5499	10.4022	9.9995	10.1375	10.0333	9.8061	9.8815	9.8351
(0.95,0.95,0.95)	14.3660	14.3369	14.6414	13.9369	14.0132	14.0370	13.6613	13.6780	13.7775
(0.97,0.97,0.97)	24.3685	24.4395	24.2844	23.4630	23.4084	23.3130	22.8906	22.8673	22.7850
(0.99,0.99,0.99)	72.5221	72.5544	73.0217	69.5520	69.8710	69.8880	67.7654	68.0085	67.9696
(0.91,0.91,0.93)	7.5900	9.7498	9.7943	7.3128	9.3677	9.3276	7.1592	9.1694	9.1689

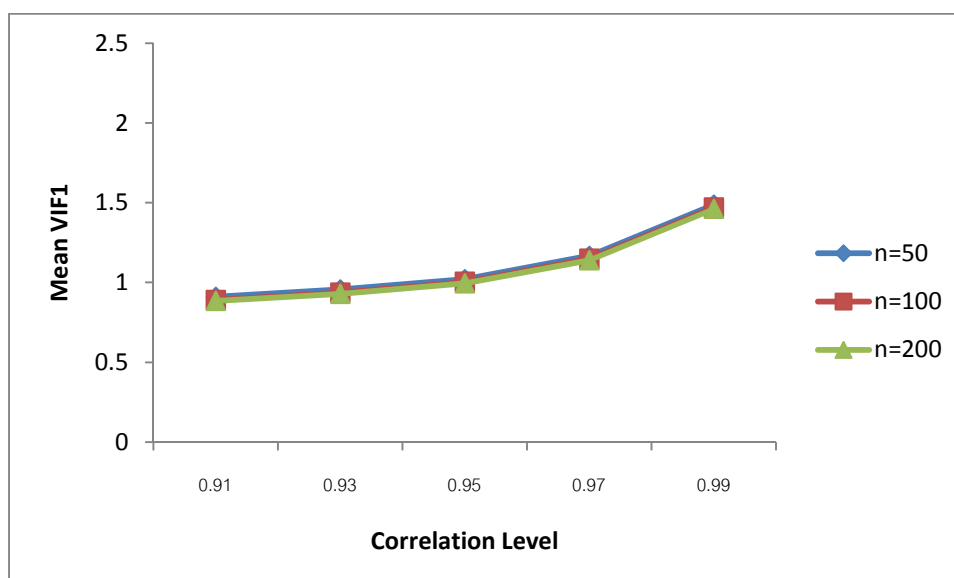
ตารางที่ 4.1.2 (ต่อ) แสดงค่าเฉลี่ยของ VIF กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์	n=50			n=100			n=200		
	VIF1	VIF2	VIF3	VIF1	VIF2	VIF3	VIF1	VIF2	VIF3
(0.91,0.91,0.95)	7.1170	12.8066	12.7360	6.9032	12.2489	12.2460	6.7790	12.0050	11.9797
(0.91,0.91,0.97)	6.7860	19.5842	19.4825	6.5157	18.8557	18.8263	6.3849	18.6155	18.5643
(0.91,0.91,0.99)	6.4302	54.9902	54.9093	6.2013	53.3590	53.0818	6.0628	52.5463	52.3706
(0.91,0.93,0.93)	9.0383	9.1284	11.5119	8.6298	8.6429	10.9993	8.5024	8.5082	10.8237
(0.91,0.95,0.95)	11.1932	11.1009	19.6984	10.7768	10.7065	19.0003	10.5172	10.5072	18.5133
(0.91,0.97,0.97)	25.1653	25.3639	73.4908	24.1590	24.3495	70.5524	23.6997	23.7277	68.9931
(0.93,0.93,0.95)	9.5897	13.2721	13.3275	9.1704	12.6535	12.6717	8.9797	12.4034	12.4276
(0.93,0.93,0.97)	8.8999	20.3749	20.3369	8.4895	19.5498	19.4278	8.3050	19.0681	19.0376
(0.93,0.93,0.99)	8.3954	57.1559	57.1747	7.9680	54.3177	54.4978	7.8416	53.2934	53.3975
(0.93,0.95,0.95)	12.1266	12.0526	16.7803	11.6023	11.4968	16.0701	11.3302	11.3644	15.7731
(0.93,0.97,0.97)	19.0011	19.3335	43.9023	18.2555	18.3912	41.9650	17.8985	17.9045	40.9768
(0.95,0.95,0.97)	12.8917	21.4079	21.3646	12.4153	20.5953	20.5578	12.1782	20.2004	20.1602
(0.95,0.95,0.99)	11.6676	57.1237	57.0875	11.2661	54.8604	54.9303	11.0833	54.1534	54.1997
(0.95,0.97,0.97)	18.9271	18.9665	31.3213	18.0232	17.9709	29.6798	17.5562	17.4523	28.9144
(0.97,0.97,0.99)	19.8712	59.1932	59.0858	19.1567	56.7905	56.6622	18.7954	55.6407	55.5455
(0.97,0.99,0.99)	73.2437	73.5952	217.7078	70.2933	70.6193	209.4184	69.2721	69.1645	205.4554

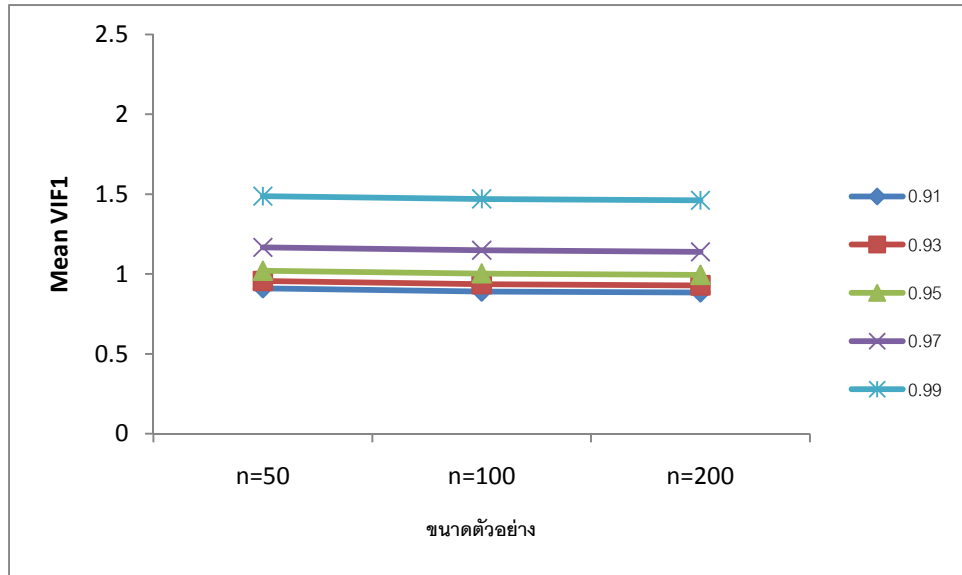
ตารางที่ 4.1.3 แสดงค่าเฉลี่ยของ VIF_j ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์สูงสุด		0.91	0.93	0.95	0.97	0.99
VIF1	n=50	8.1309	9.0354	10.4634	14.6674	30.7423
	n=100	7.7648	8.6474	10.0762	14.0686	29.5060
	n=200	7.6526	8.4892	9.8627	13.7644	28.8967
VIF2	n=50	8.0165	9.8093	12.5454	20.5103	82.5262
	n=100	7.7216	9.3827	12.0753	19.6608	78.9371
	n=200	7.5922	9.1863	11.8397	19.2332	77.2650
VIF3	n=50	8.0976	10.5695	15.3502	31.6556	139.2346
	n=100	7.7159	10.1201	14.7363	30.3302	133.2628
	n=200	7.6182	9.9426	14.4235	29.6792	130.4274

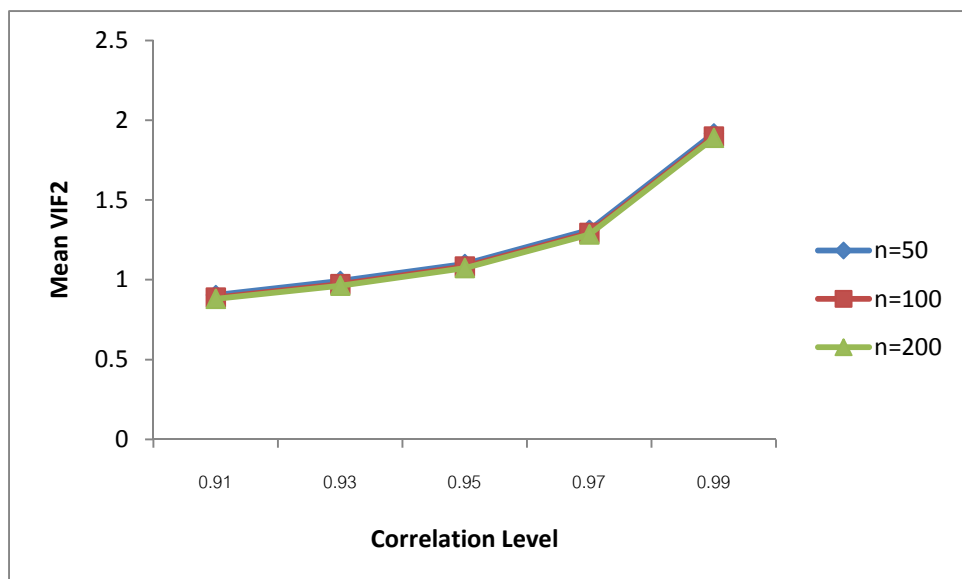
รูปที่ 4.1.3 การเปรียบเทียบค่าเฉลี่ยของ VIF_1 ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



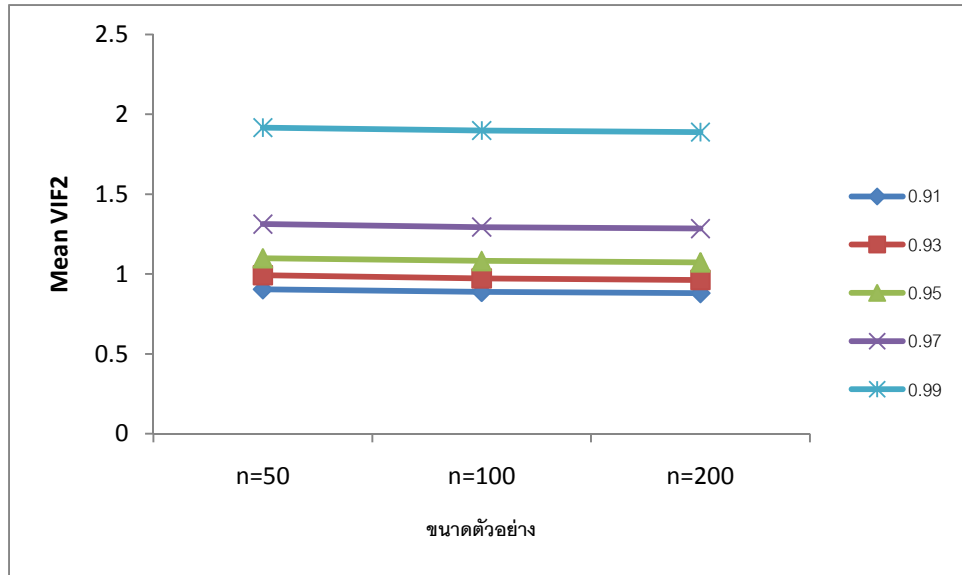
รูปที่ 4.1.4 การเปรียบเทียบค่าเฉลี่ยของ VIF1 ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



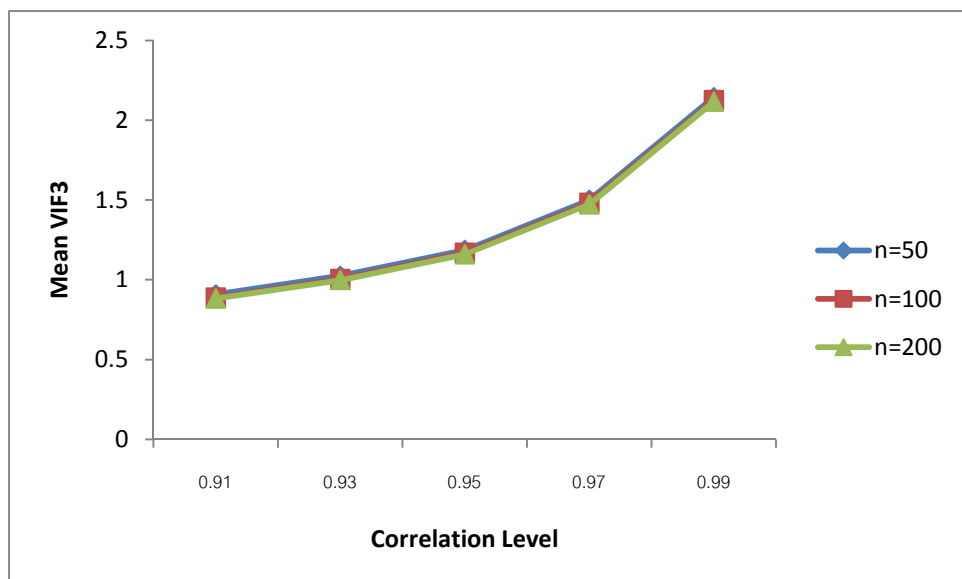
รูปที่ 4.1.5 การเปรียบเทียบค่าเฉลี่ยของ VIF2 ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



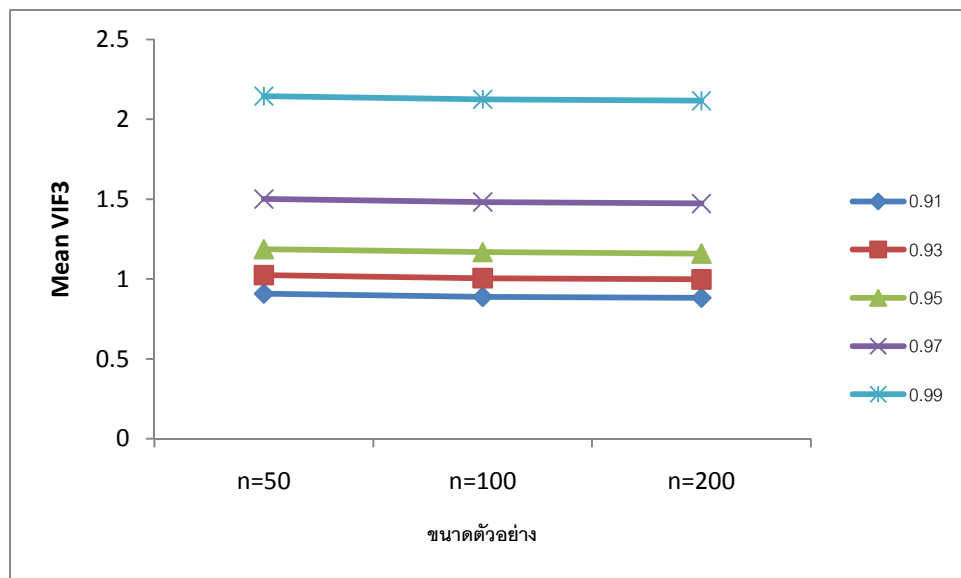
รูปที่ 4.1.6 การเปรียบเทียบค่าเฉลี่ยของ VIF2 ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



รูปที่ 4.1.7 การเปรียบเทียบค่าเฉลี่ยของ VIF3 ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



รูปที่ 4.1.8 การเปรียบเทียบค่าเฉลี่ยของ VIF3 ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



ผลการวิเคราะห์พบว่า เมื่อพิจารณาค่าเฉลี่ยของ VIF ณ ระดับความสัมพันธ์ต่างๆ สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร ที่ระดับความสัมพันธ์เท่ากับ 0.93 จะเป็นจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาของความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระของขนาดตัวอย่างเท่ากับ 50 และ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาความสัมพันธ์รุนแรงอยู่ที่ระดับความสัมพันธ์เท่ากับ 0.95 ดังนั้น สรุปได้ว่าที่ขนาดตัวอย่างเท่ากับ 50 และ 100 จะเกิดปัญหาของความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระที่ระดับความสัมพันธ์ 0.93, 0.95, 0.97 และ 0.99 และที่ขนาดตัวอย่างเท่ากับ 200 จะเกิดปัญหาของความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระที่ระดับความสัมพันธ์ 0.95, 0.97 และ 0.99 สำหรับในการศึกษาครั้งนี้

เมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF1 แยกตามขนาดตัวอย่าง พบว่าที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยของ VIF1 สูงที่สุด รองลงมาคือขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยของ VIF1 ต่ำที่สุด นั่นคือ ค่าเฉลี่ยของ VIF1 จะแปรผกผันกับขนาดตัวอย่างของทุกระดับความสัมพันธ์ ซึ่งแสดงได้ดังรูปที่ 4.1.3 และเมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF1 แยกตามระดับความสัมพันธ์ พบว่าที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยของ VIF1 สูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.97, 0.95, 0.93 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.91 จะมีค่าเฉลี่ยของ VIF1 ต่ำที่สุด นั่นคือ ค่าเฉลี่ยของ VIF1 จะแปรผันตรงกับระดับความสัมพันธ์ของทุกขนาดตัวอย่าง ซึ่งแสดงได้ดังรูปที่ 4.1.4

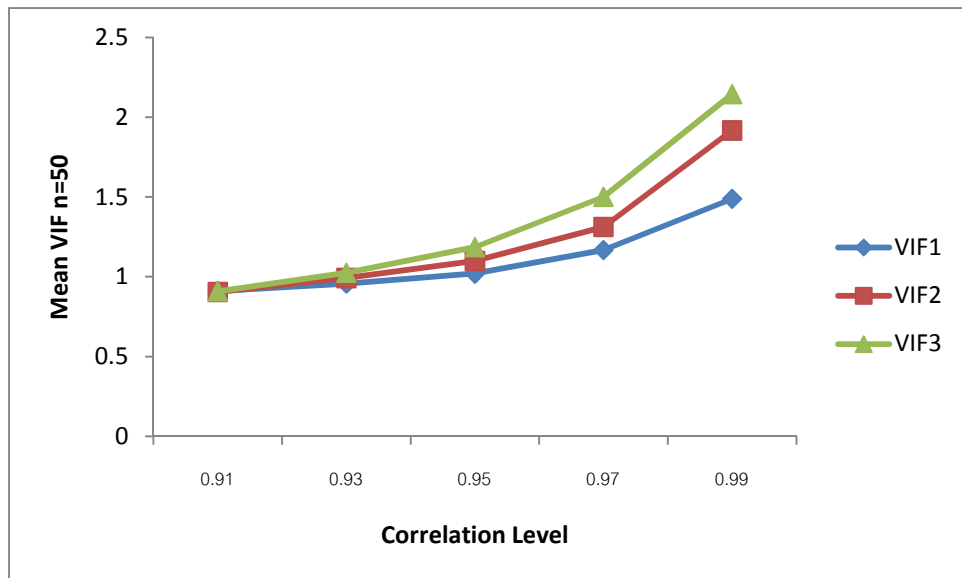
เมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF2 แยกตามขนาดตัวอย่าง พบว่าที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยของ VIF2 สูงที่สุด รองลงมาคือขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยของ VIF2 ต่ำที่สุด นั่นคือ ค่าเฉลี่ยของ VIF2 จะแปรผกผันกับขนาดตัวอย่างของทุกระดับความสัมพันธ์ ซึ่งแสดงได้ดังรูปที่ 4.1.5 และเมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF2 แยกตามระดับความสัมพันธ์ พบว่าที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยของ VIF2 สูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.97, 0.95, 0.93 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.91 จะมีค่าเฉลี่ยของ VIF2 ต่ำที่สุด นั่นคือ ค่าเฉลี่ยของ VIF2 จะแปรผันตรงกับระดับความสัมพันธ์ของทุกขนาดตัวอย่าง ซึ่งแสดงได้ดังรูปที่ 4.1.6

เมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF3 แยกตามขนาดตัวอย่าง พบว่าที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยของ VIF3 สูงที่สุด รองลงมาคือขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยของ VIF3 ต่ำที่สุด นั่นคือ ค่าเฉลี่ยของ VIF3 จะแปรผกผันกับขนาดตัวอย่างของทุกระดับความสัมพันธ์ ซึ่งแสดงได้ดังรูปที่ 4.1.7 และเมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF3 แยกตามระดับความสัมพันธ์ พบว่าที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยของ VIF3 สูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.97, 0.95, 0.93 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.91 จะมีค่าเฉลี่ยของ VIF3 ต่ำที่สุด นั่นคือ ค่าเฉลี่ยของ VIF3 จะแปรผันตรงกับระดับความสัมพันธ์ของทุกขนาดตัวอย่าง ซึ่งแสดงได้ดังรูปที่ 4.1.8

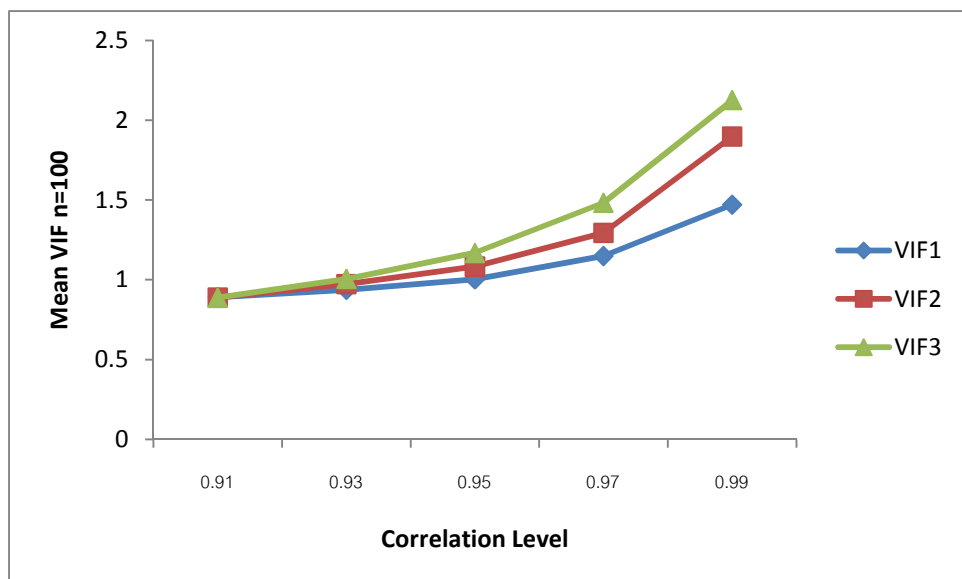
ตารางที่ 4.1.4 แสดงค่าเฉลี่ยของ VIF ที่ระดับความสัมพันธ์สูงสุด แยกตาม VIF_j ของแต่ละขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์สูงสุด		0.91	0.93	0.95	0.97	0.99
n=50	VIF1	8.1309	9.0354	10.4634	14.6674	30.7423
	VIF2	8.0165	9.8093	12.5454	20.5103	82.5262
	VIF3	8.0976	10.5695	15.3502	31.6556	139.2346
n=100	VIF1	7.7648	8.6474	10.0762	14.0686	29.5060
	VIF2	7.7216	9.3827	12.0753	19.6608	78.9371
	VIF3	7.7159	10.1201	14.7363	30.3302	133.2628
n=200	VIF1	7.6526	8.4892	9.8627	13.7644	28.8967
	VIF2	7.5922	9.1863	11.8397	19.2332	77.2650
	VIF3	7.6182	9.9426	14.4235	29.6792	130.4274

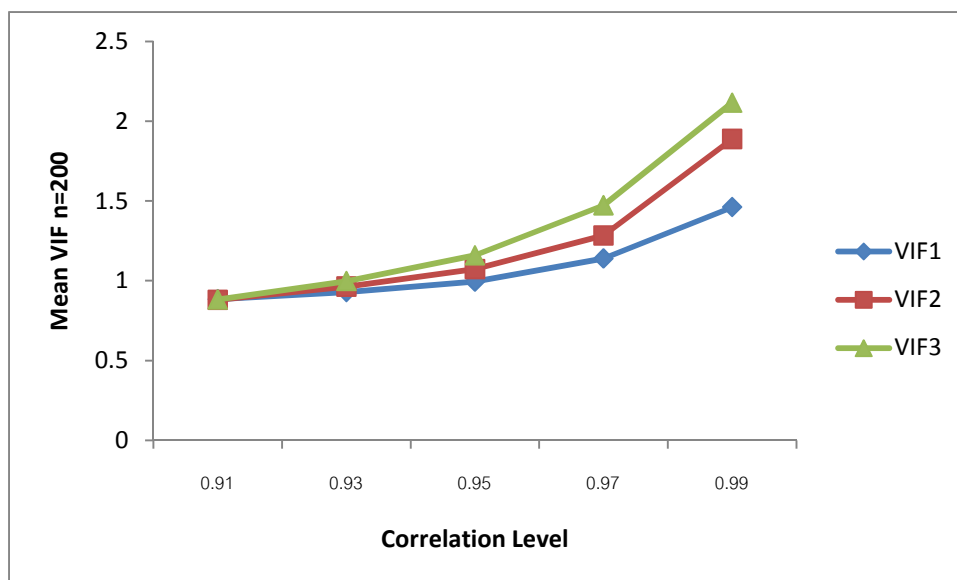
รูปที่ 4.1.9 การเปรียบเทียบค่าเฉลี่ยของ VIF ที่ระดับความสัมพันธ์สูงสุด แยกตาม VIF_j ที่ n=50 กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



รูปที่ 4.1.10 การเปรียบเทียบค่าเฉลี่ยของ VIF ที่ระดับความสัมพันธ์สูงสุด แยกตาม VIF_j ที่ n=100 กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



รูปที่ 4.1.11 การเปรียบเทียบค่าเฉลี่ยของ VIF ที่ระดับความสัมพันธ์สูงสุด แยกตาม VIF_j ที่ $n=200$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



เมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF แยกตามค่าเฉลี่ยของ VIF_j ของขนาดตัวอย่างเท่ากับ 50 พบว่าค่าเฉลี่ยของ VIF1, VIF2 และ VIF3 จะแปรผันตรงกับระดับความสัมพันธ์ และสังเกตได้ว่าที่ระดับความสัมพันธ์ 0.91 ค่าเฉลี่ยของ VIF แต่ละค่าจะไม่แตกต่างกันมากนัก ส่วนที่ระดับความสัมพันธ์ 0.93, 0.95, 0.97 และ 0.99 ค่าเฉลี่ยของ VIF3 จะมีค่ามากที่สุด รองลงมาคือค่าเฉลี่ยของ VIF2 และค่าเฉลี่ยของ VIF1 จะมีค่าน้อยที่สุด ซึ่งแสดงได้ดังรูปที่ 4.1.9

เมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF แยกตามค่าเฉลี่ยของ VIF_j ของขนาดตัวอย่างเท่ากับ 100 พบว่าค่าเฉลี่ยของ VIF1, VIF2 และ VIF3 จะแปรผันตรงกับระดับความสัมพันธ์ และสังเกตได้ว่าที่ระดับความสัมพันธ์ 0.91 ค่าเฉลี่ยของ VIF แต่ละค่าจะไม่แตกต่างกันมากนัก ส่วนที่ระดับความสัมพันธ์ 0.93, 0.95, 0.97 และ 0.99 ค่าเฉลี่ยของ VIF3 จะมีค่ามากที่สุด รองลงมาคือค่าเฉลี่ยของ VIF2 และค่าเฉลี่ยของ VIF1 จะมีค่าน้อยที่สุด ซึ่งแสดงได้ดังรูปที่ 4.1.10

เมื่อพิจารณาแนวโน้มค่าเฉลี่ยของ VIF แยกตามค่าเฉลี่ยของ VIF_j ของขนาดตัวอย่างเท่ากับ 200 พบว่าค่าเฉลี่ยของ VIF1, VIF2 และ VIF3 จะแปรผันตรงกับระดับความสัมพันธ์ และสังเกตได้ว่าที่ระดับความสัมพันธ์ 0.91 ค่าเฉลี่ยของ VIF แต่ละค่าจะไม่แตกต่างกันมากนัก ส่วนที่ระดับความสัมพันธ์ 0.93, 0.95, 0.97 และ 0.99 ค่าเฉลี่ยของ VIF3 จะมีค่ามากที่สุด รองลงมาคือค่าเฉลี่ยของ VIF2 และค่าเฉลี่ยของ VIF1 จะมีค่าน้อยที่สุด ซึ่งแสดงได้ดังรูปที่ 4.1.11

4.2 ผลการศึกษาเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรง (Multicollinearity)

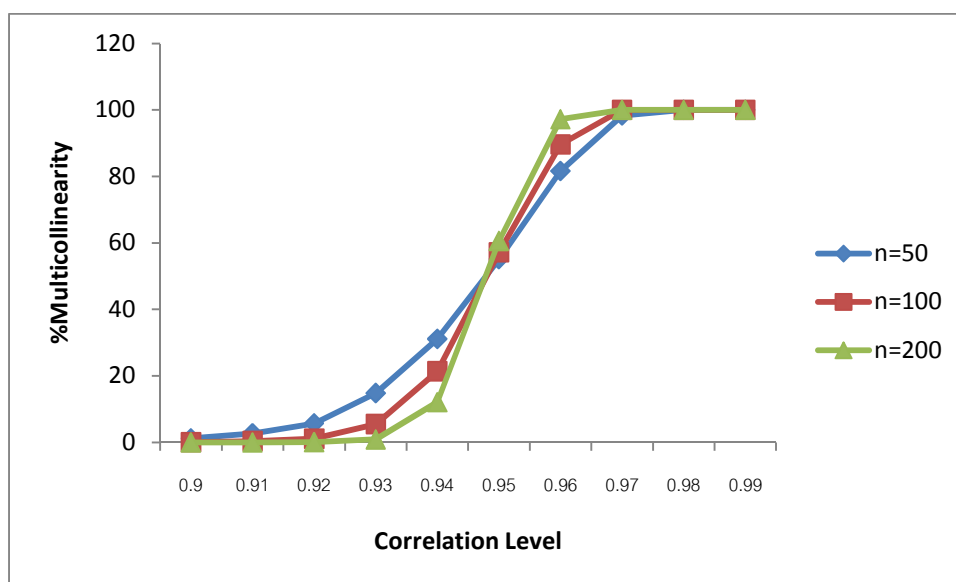
4.2.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร

เมื่อพิจารณาค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงของแต่ละกรณีที่ทำการศึกษา จะมีผลการวิเคราะห์ดังตารางที่ 4.2.1

ตารางที่ 4.2.1 แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร

ระดับความสัมพันธ์		0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
n=50	% Multi	1.2	2.7	5.7	14.8	31.1	55.1	81.6	98.3	100	100
n=100	% Multi	0.1	0.4	1.1	5.5	21.4	57.2	89.6	100	100	100
n=200	% Multi	0	0	0.1	0.9	12.1	60.6	97.2	100	100	100

รูปที่ 4.2.1 การเปรียบเทียบเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



ผลการวิเคราะห์ เมื่อพิจารณาค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ ณ ระดับความสัมพันธ์ต่างๆ พบว่าที่ระดับความสัมพันธ์ 0.9, 0.91, 0.92, 0.93 และ 0.94 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 0%-32% ส่วนที่ระดับความสัมพันธ์ 0.95, 0.96, 0.97, 0.98 และ 0.99 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 55%-100% ซึ่งจะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงสูงทุกขนาดตัวอย่างที่ทำการศึกษา

เมื่อพิจารณาแนวโน้มของค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ พบว่าที่ระดับความสัมพันธ์ 0.9, 0.91, 0.92, 0.93 และ 0.94 ค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงจะแปรผันตรงกับระดับความสัมพันธ์ของตัวแปรอิสระ แต่ที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงต่ำที่สุด รองลงมาคือขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเปอร์เซ็นต์การเกิดพหุสัมพันธ์รุนแรงสูงที่สุด ส่วนที่ระดับความสัมพันธ์ 0.95, 0.96, 0.97, 0.98 และ 0.99 ค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงจะแปรผันตรงกับระดับความสัมพันธ์ของตัวแปรอิสระ แต่ที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงสูงที่สุด รองลงมาคือขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเปอร์เซ็นต์การเกิดพหุสัมพันธ์รุนแรงต่ำที่สุด ซึ่งแสดงได้ดังรูปที่ 4.2.1

4.2.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร

เมื่อพิจารณาค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงของแต่ละกรณีที่ทำการศึกษา จะมีผลการวิเคราะห์ดังตารางที่ 4.2.2 และเพื่อให้ง่ายในการวิเคราะห์ผลทางผู้วิจัยจึงได้พิจารณาจากค่าระดับความสัมพันธ์สูงสุดของแต่ละกรณี โดยพิจารณาจากค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงที่ต่ำที่สุดของแต่ละระดับความสัมพันธ์สูงสุด ดังตารางที่ 4.2.3

ตารางที่ 4.2.2 แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์	n=50	n=100	n=200
	% Multi	% Multi	% Multi
(0.91,0.93,0.95)	91.8	96.1	99.1
(0.91,0.93,0.97)	99.8	100	100
(0.91,0.93,0.99)	100	100	100
(0.91,0.95,0.97)	100	100	100
(0.91,0.95,0.99)	100	100	100
(0.93,0.95,0.97)	100	100	100
(0.93,0.95,0.99)	100	100	100
(0.93,0.97,0.99)	100	100	100
(0.95,0.97,0.99)	100	100	100
(0.91,0.91,0.91)	33.3	15.4	3.9
(0.93,0.93,0.93)	71.1	67.5	63.9
(0.95,0.95,0.95)	96.9	99.1	99.9
(0.97,0.97,0.97)	100	100	100
(0.99,0.99,0.99)	100	100	100
(0.91,0.91,0.93)	52.2	43.4	34.7
(0.91,0.91,0.95)	85.1	88.8	94.9
(0.91,0.91,0.97)	99.2	100	100
(0.91,0.91,0.99)	100	100	100
(0.91,0.93,0.93)	70.9	68.6	69.8
(0.91,0.95,0.95)	99.2	99.9	100
(0.91,0.97,0.97)	100	100	100
(0.93,0.93,0.95)	90	93	97.2
(0.93,0.93,0.97)	99.8	99.9	100
(0.93,0.93,0.99)	100	100	100

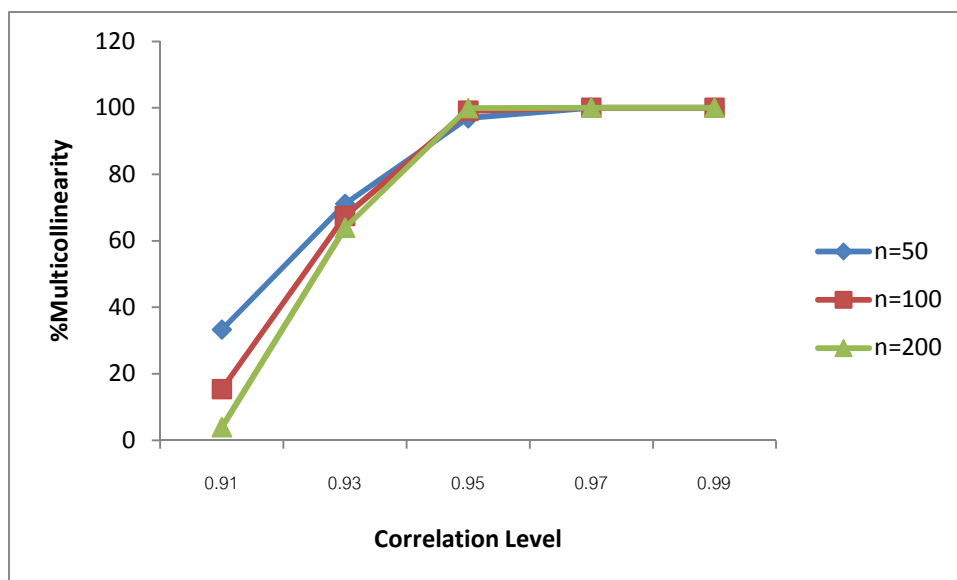
ตารางที่ 4.2.2 (ต่อ) แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์	n=50	n=100	n=200
	% Multi	% Multi	% Multi
(0.93,0.95,0.95)	95.7	99.2	100
(0.93,0.97,0.97)	100	100	100
(0.95,0.95,0.97)	99.5	100	100
(0.95,0.95,0.99)	100	100	100
(0.95,0.97,0.97)	100	100	100
(0.97,0.97,0.99)	100	100	100
(0.97,0.99,0.99)	100	100	100

ตารางที่ 4.2.3 แสดงเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ ที่ระดับความสัมพันธ์สูงสุด กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับความสัมพันธ์สูงสุด		0.91	0.93	0.95	0.97	0.99
n=50	%Multi	33.3	71.1	96.9	100	100
n=100	%Multi	15.4	67.5	99.1	100	100
n=200	%Multi	3.9	63.9	99.9	100	100

รูปที่ 4.2.2 การเปรียบเทียบเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



ผลการวิเคราะห์ เมื่อพิจารณาค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ ณ ระดับความสัมพันธ์ต่างๆ พบว่าที่ระดับความสัมพันธ์ 0.91 ขนาดตัวอย่างเท่ากับ 50 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงอยู่ประมาณ 33% ขนาดตัวอย่างเท่ากับ 100 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงอยู่ประมาณ 15% และขนาดตัวอย่างเท่ากับ 200 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงอยู่ประมาณ 4% ส่วนที่ระดับความสัมพันธ์ 0.93, 0.95, 0.97 และ 0.99 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 60%-100% ซึ่งจะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงสูงทุกขนาดตัวอย่างที่ทำการศึกษา

เมื่อพิจารณาแนวโน้มของค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระ พบว่าที่ระดับความสัมพันธ์ 0.91 และ 0.93 ค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงจะแปรผันตรงกับระดับความสัมพันธ์ของตัวแปรอิสระ แต่ที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงต่ำที่สุด รองลงมาคือขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเปอร์เซ็นต์การเกิดพหุสัมพันธ์รุนแรงสูงที่สุด ส่วนที่ระดับความสัมพันธ์ 0.95, 0.97 และ 0.99 ค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงจะแปรผันตรงกับระดับความสัมพันธ์ของตัวแปรอิสระ แต่ที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเปอร์เซ็นต์การเกิด

ปัญหาพหุสัมพันธ์รุนแรงสูงสุด รองลงมาคือขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเปอร์เซ็นต์การเกิดพหุสัมพันธ์รุนแรงต่ำที่สุด ซึ่งแสดงได้ดังรูปที่ 4.2.2

4.3 ผลการศึกษาการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ

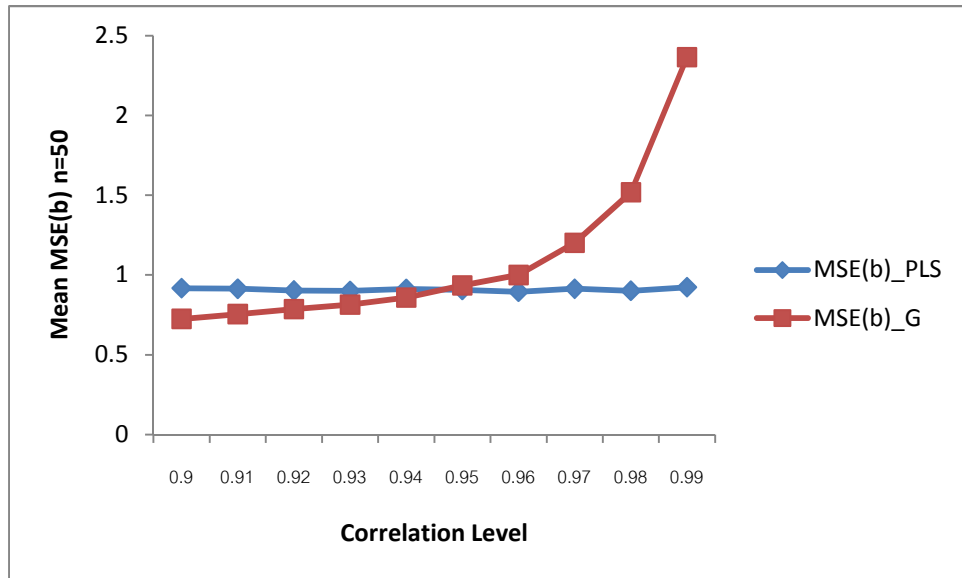
4.3.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณแต่ละกรณีที่ทำการศึกษา จะมีผลการวิเคราะห์ดังตารางที่ 4.3.1-4.3.2

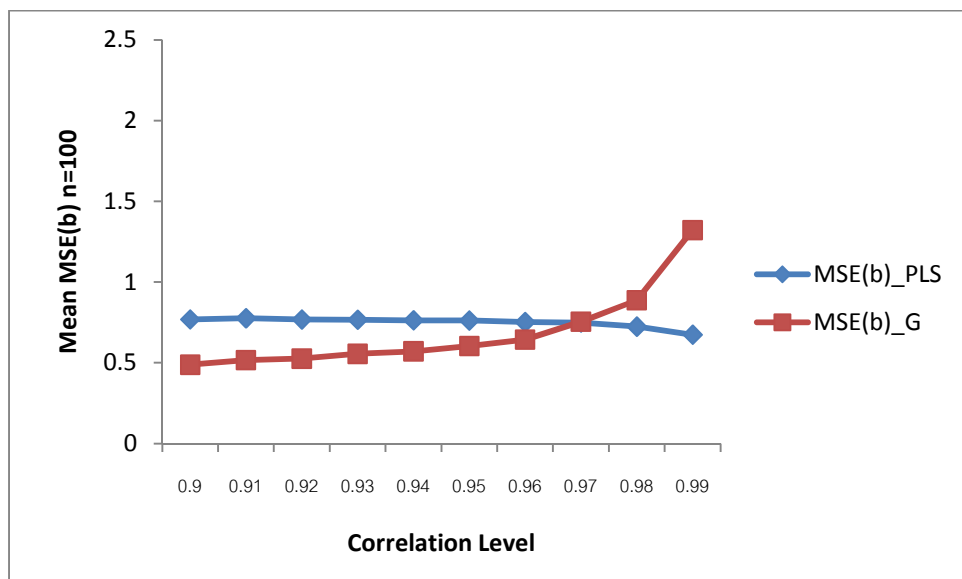
ตารางที่ 4.3.1 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ แยกตามวิธีการวิเคราะห์ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร

ระดับ ความสัมพันธ์	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
0.9	0.9172	0.7245	0.7690	0.4883	0.7179	0.3766
0.91	0.9134	0.7548	0.7764	0.5166	0.7167	0.3933
0.92	0.9023	0.7865	0.7683	0.5260	0.7127	0.4025
0.93	0.8995	0.8135	0.7670	0.5556	0.7143	0.4193
0.94	0.9120	0.8584	0.7634	0.5713	0.7098	0.4312
0.95	0.9061	0.9341	0.7618	0.6040	0.7107	0.4489
0.96	0.8952	0.9997	0.7523	0.6443	0.7056	0.4768
0.97	0.9138	1.2008	0.7489	0.7551	0.7007	0.5297
0.98	0.9004	1.5175	0.7240	0.8879	0.6880	0.6024
0.99	0.9230	2.3646	0.6743	1.3217	0.6532	0.8214

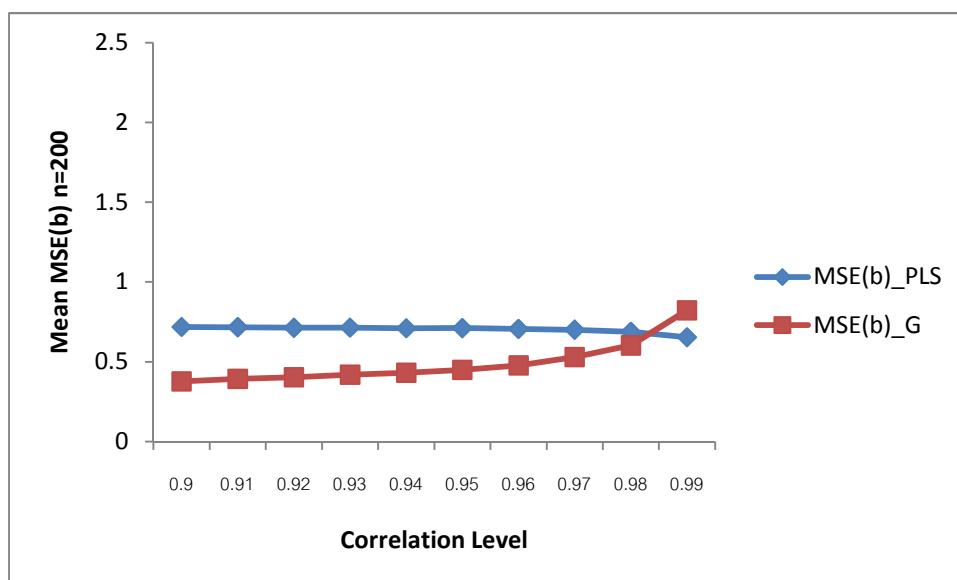
รูปที่ 4.3.1 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามวิธีการวิเคราะห์ ที่ $n=50$ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



รูปที่ 4.3.2 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามวิธีการวิเคราะห์ ที่ $n=100$ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



รูปที่ 4.3.3 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามวิธีการวิเคราะห์ ที่ $n=200$ กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ณ ระดับความสัมพันธ์ต่างๆ แยกตามวิธีการวิเคราะห์ ที่ทุกขนาดตัวอย่าง สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร เป็นดังนี้

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ขนาดตัวอย่างเท่ากับ 50 ของทั้งสองวิธี พบว่าที่ระดับความสัมพันธ์ 0.9, 0.92, 0.93 และ 0.94 วิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมากกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์ ส่วนที่ระดับความสัมพันธ์ 0.95, 0.96, 0.97, 0.98 และ 0.99 วิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์ และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยสุดเชิงส่วนไม่มีแนวโน้มที่แน่นอนเมื่อระดับความสัมพันธ์สูงขึ้น เนื่องจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองยังมีการแกว่งขึ้นแกว่งลง ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์สูงขึ้น ซึ่งแสดงได้ดังรูปที่ 4.3.1

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ขนาดตัวอย่างเท่ากับ 100 ของทั้งสองวิธี พบว่าที่ระดับความสัมพันธ์ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95 และ 0.96 วิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมากกว่าวิธีกำลังสองน้อย

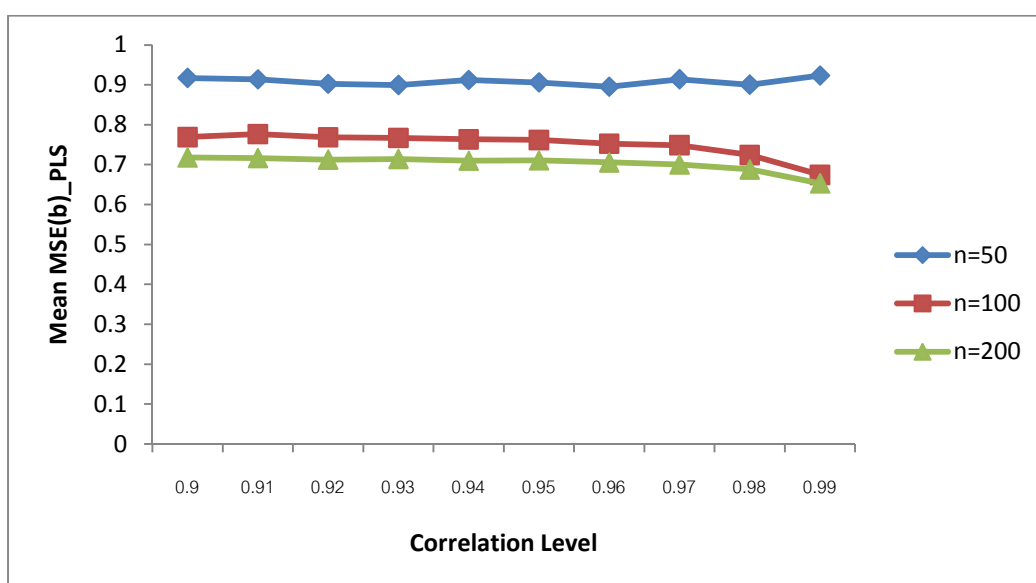
สุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์ ส่วนที่ระดับความสัมพันธ์ 0.97, 0.98 และ 0.99 วิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์ และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยสุดเชิงส่วนไม่มีแนวโน้มที่แน่นอน เมื่อระดับความสัมพันธ์สูงขึ้น เนื่องจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองยังมีการแกว่งขึ้นแกว่งลง ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์สูงขึ้น ซึ่งแสดงได้ดังรูปที่ 4.3.2

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ขนาดตัวอย่างเท่ากับ 200 ของทั้งสองวิธี พบว่าที่ระดับความสัมพันธ์ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.97 และ 0.98 วิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมากกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์ ส่วนที่ระดับความสัมพันธ์ 0.99 วิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์ และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยสุดเชิงส่วนไม่มีแนวโน้มที่แน่นอน เมื่อระดับความสัมพันธ์สูงขึ้น เนื่องจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองยังมีการแกว่งขึ้นแกว่งลง ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์สูงขึ้น ซึ่งแสดงได้ดังรูปที่ 4.3.3

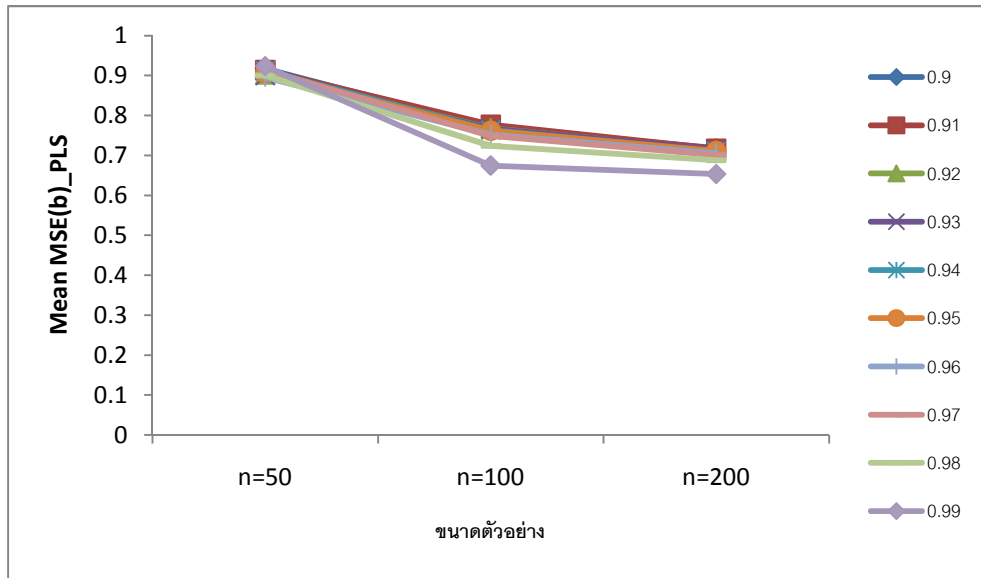
ตารางที่ 4.3.2 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร

ระดับ ความสัมพันธ์	PLS			OLS_G		
	n=50	n=100	n=200	n=50	n=100	n=200
0.9	0.9172	0.7690	0.7179	0.7245	0.4883	0.3766
0.91	0.9134	0.7764	0.7167	0.7548	0.5166	0.3933
0.92	0.9023	0.7683	0.7127	0.7865	0.5260	0.4025
0.93	0.8995	0.7670	0.7143	0.8135	0.5556	0.4193
0.94	0.9120	0.7634	0.7098	0.8584	0.5713	0.4312
0.95	0.9061	0.7618	0.7107	0.9341	0.6040	0.4489
0.96	0.8952	0.7523	0.7056	0.9997	0.6443	0.4768
0.97	0.9138	0.7489	0.7007	1.2008	0.7551	0.5297
0.98	0.9004	0.7240	0.6880	1.5175	0.8879	0.6024
0.99	0.9230	0.6743	0.6532	2.3646	1.3217	0.8214

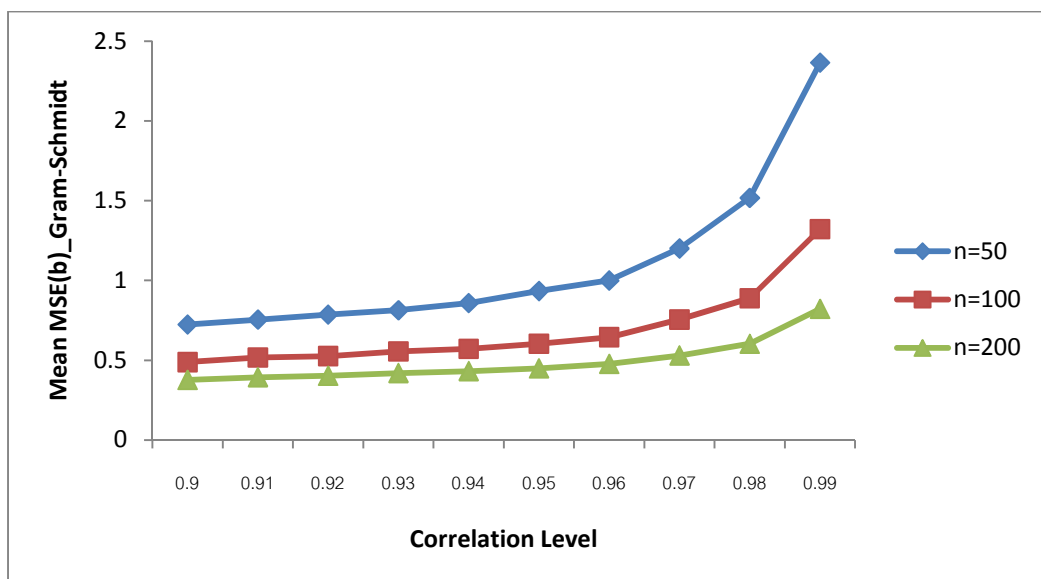
รูปที่ 4.3.4 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามขนาดตัวอย่างของวิธี PLS กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



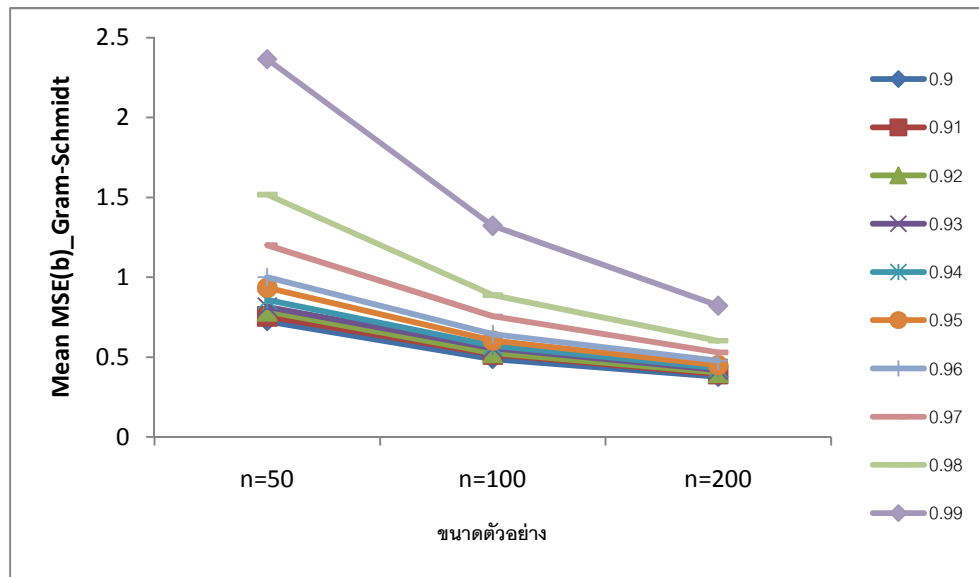
รูปที่ 4.3.5 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามระดับความสัมพันธ์ของวิธี PLS กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



รูปที่ 4.3.6 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามขนาดตัวอย่างของวิธี OLS_G กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



รูปที่ 4.3.7 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง แยกตามระดับความสัมพันธ์ของวิธี OLS_G กรณีที่มีจำนวนตัวแปรอิสระ 2 ตัวแปร



ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ณ ระดับความสัมพันธ์ต่างๆ แยกตามขนาดตัวอย่าง ของทั้งสองวิธี สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร เป็นดังนี้

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ทำการวิเคราะห์ด้วยวิธีกำลังสองน้อยสุดเชิงส่วน พบว่าที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือที่ขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด และเมื่อพิจารณาแนวโน้มพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของทุกขนาดตัวอย่างไม่มีแนวโน้มที่แน่นอนเมื่อระดับความสัมพันธ์เพิ่มขึ้น ซึ่งแสดงได้ดังรูปที่ 4.3.4

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ทำการวิเคราะห์ด้วยวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์ พบว่าที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือที่ขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด และเมื่อพิจารณาแนวโน้มพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มเพิ่มขึ้นตามระดับความสัมพันธ์ของทุกขนาดตัวอย่างที่ทำการศึกษา ซึ่งแสดงได้ดังรูปที่ 4.3.6

ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ณ ระดับความสัมพันธ์ต่างๆ แยกตามระดับความสัมพันธ์ ของทั้งสองวิธี สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 2 ตัวแปร เป็นดังนี้

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ทำการวิเคราะห์ด้วยวิธีกำลังสองน้อยสุดเชิงส่วน ตามระดับความสัมพันธ์ พบว่าที่ขนาดตัวอย่างเท่ากับ 50 ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.9, 0.97, 0.91, 0.94, 0.95, 0.92, 0.98, 0.93 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.96 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด ที่ขนาดตัวอย่างเท่ากับ 100 ระดับความสัมพันธ์เท่ากับ 0.91 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือระดับความสัมพันธ์ 0.9, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด และที่ขนาดตัวอย่างเท่ากับ 200 ระดับความสัมพันธ์เท่ากับ 0.9 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.91, 0.93, 0.92, 0.95, 0.94, 0.96, 0.97, 0.98 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด และเมื่อพิจารณาแนวโน้มพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นของทุกระดับความสัมพันธ์ที่ทำการศึกษา ซึ่งแสดงได้ดังรูปที่ 4.3.5

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ทำการวิเคราะห์ด้วยวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซิมิตต์ ตามระดับความสัมพันธ์ พบว่าที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือที่ระดับความสัมพันธ์เท่ากับ 0.98, 0.97, 0.96, 0.95, 0.94, 0.93, 0.92, 0.91 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.9 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด ทุกขนาดตัวอย่างที่ทำการศึกษา และเมื่อพิจารณาแนวโน้มพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจะมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นของทุกระดับความสัมพันธ์ที่ทำการศึกษา ซึ่งแสดงได้ดังรูปที่ 4.3.7

4.3.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ แต่ละกรณีที่ทำการศึกษา จะมีผลการวิเคราะห์ดังตารางที่ 4.3.3 และเพื่อให้ง่ายในการวิเคราะห์ผลทางผู้วิจัย

จึงได้พิจารณาจากค่าระดับความสัมพันธ์สูงสุดของแต่ละกรณี โดยพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ แต่ละระดับความสัมพันธ์สูงสุด ดังตารางที่ 4.3.4-4.3.5

ตารางที่ 4.3.3 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับ ความสัมพันธ์	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
(0.91,0.93,0.95)	0.9905	1.7719	0.8544	1.3565	0.7867	1.1427
(0.91,0.93,0.97)	0.9469	1.9864	0.8304	1.4640	0.7763	1.2122
(0.91,0.93,0.99)	0.8456	3.2668	0.7615	2.1216	0.7380	1.5712
(0.91,0.95,0.97)	0.9412	2.1887	0.8251	1.5658	0.7760	1.2660
(0.91,0.95,0.99)	0.8837	5.9408	0.6553	3.3974	0.6679	2.2203
(0.93,0.95,0.97)	0.9450	2.0938	0.8232	1.5261	0.7742	1.2456
(0.93,0.95,0.99)	0.8356	3.4402	0.7446	2.2141	0.7301	1.6237
(0.93,0.97,0.99)	3.6349	15.7172	0.9014	7.9429	0.5635	4.4271
(0.95,0.97,0.99)	0.8196	4.0917	0.7330	2.5538	0.7244	1.7649
(0.91,0.91,0.91)	0.9907	1.4912	0.8496	1.1767	0.7904	1.0340
(0.93,0.93,0.93)	1.0272	1.6851	0.8604	1.3002	0.7915	1.1088
(0.95,0.95,0.95)	0.9960	1.8991	0.8369	1.4256	0.7807	1.1877
(0.97,0.97,0.97)	0.9097	2.3223	0.8026	1.6549	0.7676	1.3373
(0.99,0.99,0.99)	0.7076	4.4300	0.6635	2.7312	0.6868	1.8763
(0.91,0.91,0.93)	1.0072	1.6106	0.8493	1.2553	0.7874	1.1427
(0.91,0.91,0.95)	1.0029	1.7255	0.8572	1.3338	0.7934	1.1498
(0.91,0.91,0.97)	0.9672	1.9151	0.8329	1.4436	0.7828	1.2294
(0.91,0.91,0.99)	0.8751	2.8873	0.7840	1.9533	0.7513	1.5014
(0.91,0.93,0.93)	1.0021	1.6552	0.8604	1.2722	0.7902	1.0916
(0.91,0.95,0.95)	0.9642	1.8586	0.8426	1.3954	0.7825	1.1531
(0.91,0.97,0.97)	0.8677	3.2396	0.7679	2.0527	0.7463	1.4849

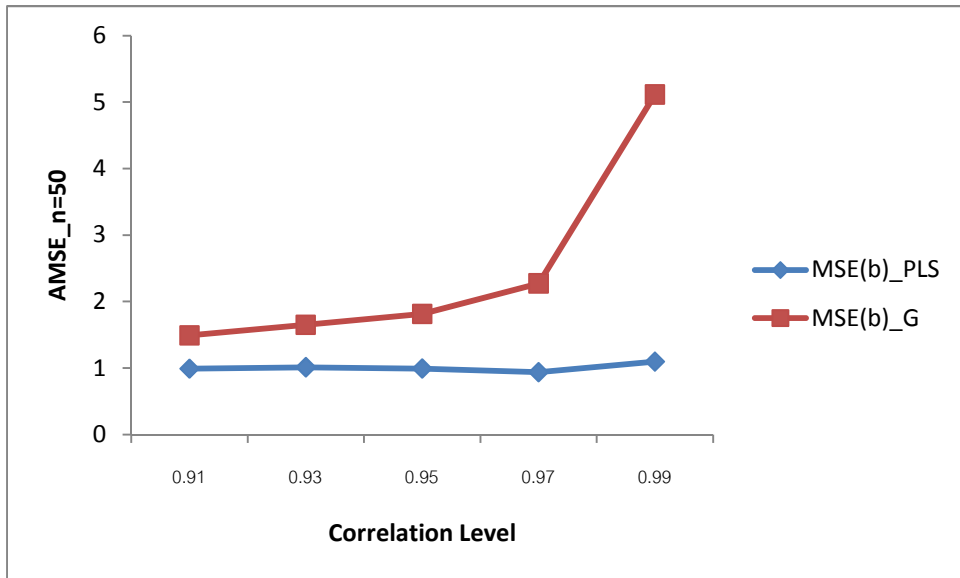
ตารางที่ 4.3.3 (ต่อ) แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับ ความสัมพันธ์	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
(0.93,0.93,0.95)	0.9887	1.7736	0.8440	1.3592	0.7826	1.1528
(0.93,0.93,0.97)	0.9718	2.0376	0.8521	1.5343	0.7836	1.2664
(0.93,0.93,0.99)	0.8711	2.9786	0.7723	1.9987	0.7446	1.5247
(0.93,0.95,0.95)	1.0097	1.8545	0.8470	1.3982	0.7856	1.1696
(0.93,0.97,0.97)	0.9207	2.4600	0.8092	1.7103	0.7661	1.3393
(0.95,0.95,0.97)	0.9692	2.1493	0.8295	1.5628	0.7789	1.2920
(0.95,0.95,0.99)	0.8860	3.0498	0.7753	2.0881	0.7452	1.5870
(0.95,0.97,0.97)	0.9331	2.3408	0.8079	1.6448	0.7697	1.3137
(0.97,0.97,0.99)	0.8495	3.2697	0.7460	2.1814	0.7377	1.6442
(0.97,0.99,0.99)	0.8504	7.1801	0.6261	4.0264	0.6420	2.4371

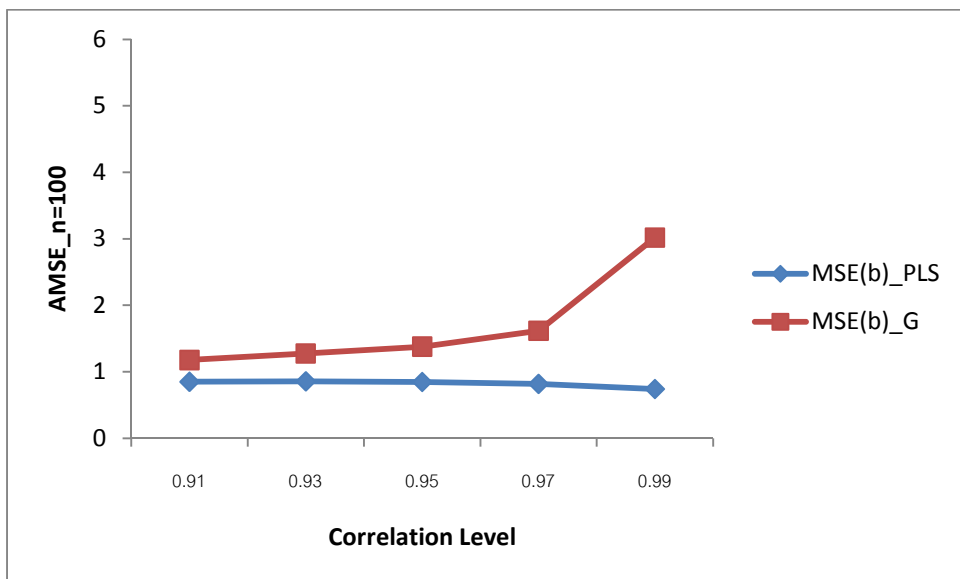
ตารางที่ 4.3.4 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ระดับความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับ ความสัมพันธ์	n=50		n=100		n=200	
	PLS	OLS_G	PLS	OLS_G	PLS	OLS_G
0.91	0.9907	1.4912	0.8496	1.1767	0.7904	1.0340
0.93	1.0122	1.6503	0.8567	1.2759	0.7897	1.1143
0.95	0.9920	1.8139	0.8470	1.3781	0.7852	1.1593
0.97	0.9373	2.2734	0.8181	1.6159	0.7722	1.2987
0.99	1.0963	5.1139	0.7421	3.0190	0.7029	2.0162

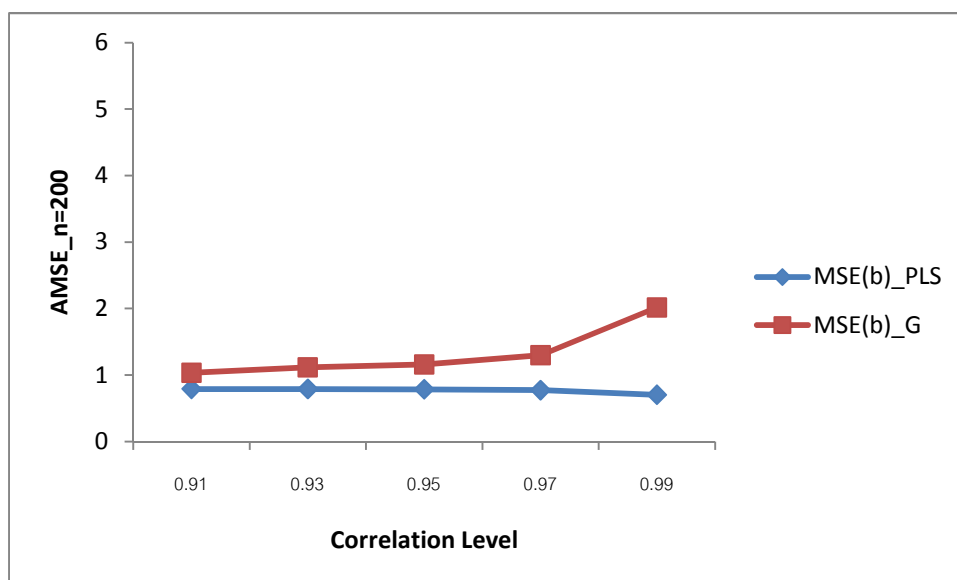
รูปที่ 4.3.8 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ ที่ $n=50$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



รูปที่ 4.3.9 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ ที่ $n=100$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



รูปที่ 4.3.10 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ ที่ $n=200$ กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ระดับความสัมพันธ์สูงสุด แยกตามวิธีการวิเคราะห์ ที่ทุกขนาดตัวอย่าง สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร เป็นดังนี้

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ขนาดตัวอย่างเท่ากับ 50 ของทั้งสองวิธี พบว่าวิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์ ในทุกระดับความสัมพันธ์ และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยสุดเชิงส่วนไม่มีแนวโน้มที่แน่นอนเมื่อระดับความสัมพันธ์สูงขึ้น เนื่องจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองยังมีการแกว่งขึ้นแกว่งลง ส่วนวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์สูงขึ้น ซึ่งแสดงได้ดังรูปที่ 4.3.8

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ขนาดตัวอย่างเท่ากับ 100 ของทั้งสองวิธี พบว่าวิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิตต์ ในทุกระดับความสัมพันธ์ และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยสุดเชิงส่วนไม่มีแนวโน้มที่แน่นอนเมื่อระดับ

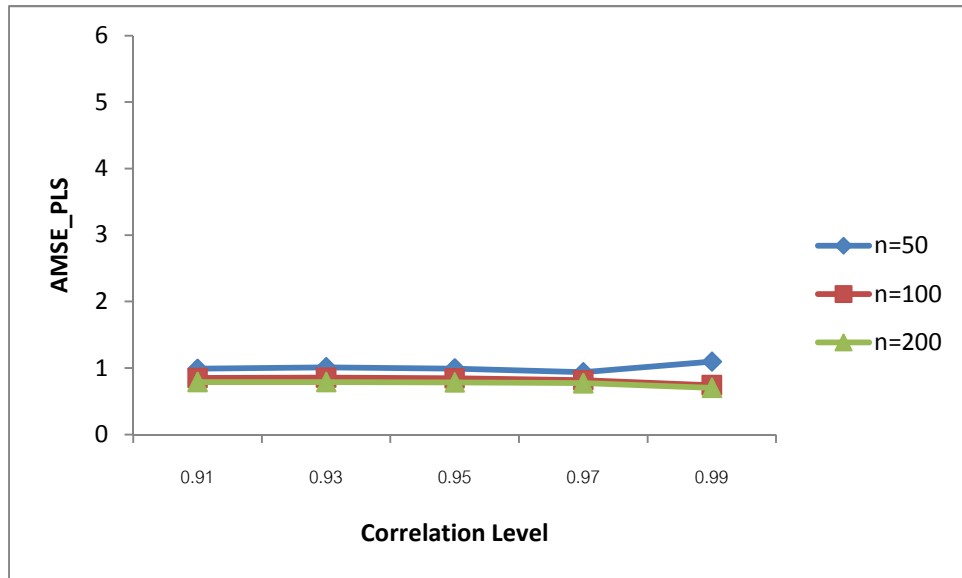
ความสัมพันธ์สูงขึ้น เนื่องจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองยังมีการแกว่งขึ้นแกว่งลง แต่วิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์สูงขึ้น ซึ่งแสดงได้ดังรูปที่ 4.3.9

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ขนาดตัวอย่างเท่ากับ 200 ของทั้งสองวิธี พบว่าวิธีกำลังสองน้อยสุดเชิงส่วนมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยกว่าวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์ ในทุกระดับความสัมพันธ์ และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของวิธีกำลังสองน้อยสุดเชิงส่วนมีแนวโน้มลดลงเล็กน้อยเมื่อระดับความสัมพันธ์สูงขึ้น แต่วิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์มีแนวโน้มเพิ่มขึ้นเมื่อระดับความสัมพันธ์สูงขึ้น ซึ่งแสดงได้ดังรูปที่ 4.3.10

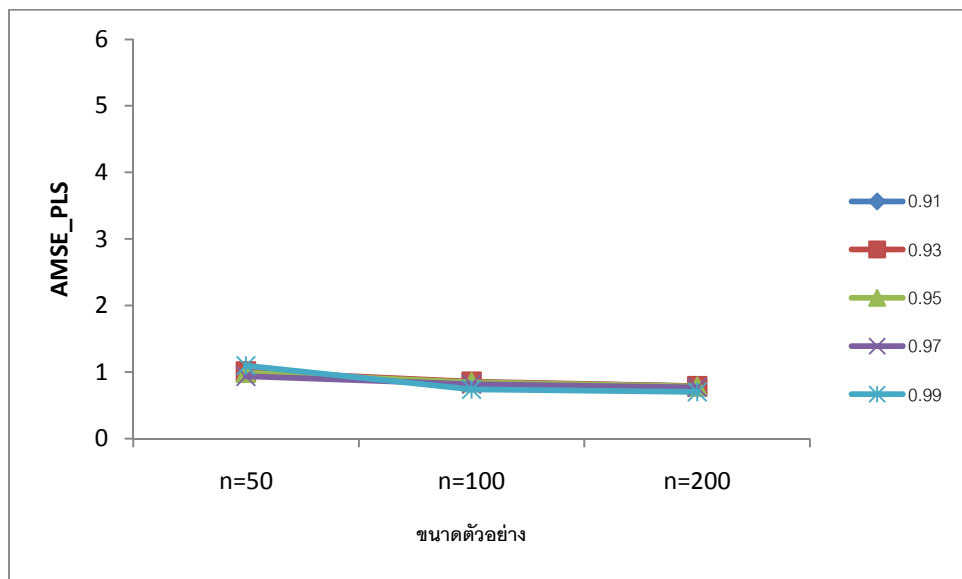
ตารางที่ 4.3.5 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่าง กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร

ระดับ ความสัมพันธ์	PLS			OLS_G		
	n=50	n=100	n=200	n=50	n=100	n=200
0.91	0.9907	0.8496	0.7904	1.4912	1.1767	1.0340
0.93	1.0122	0.8567	0.7897	1.6503	1.2759	1.1143
0.95	0.9920	0.8470	0.7852	1.8139	1.3781	1.1593
0.97	0.9373	0.8181	0.7722	2.2734	1.6159	1.2987
0.99	1.0963	0.7421	0.7029	5.1139	3.0190	2.0162

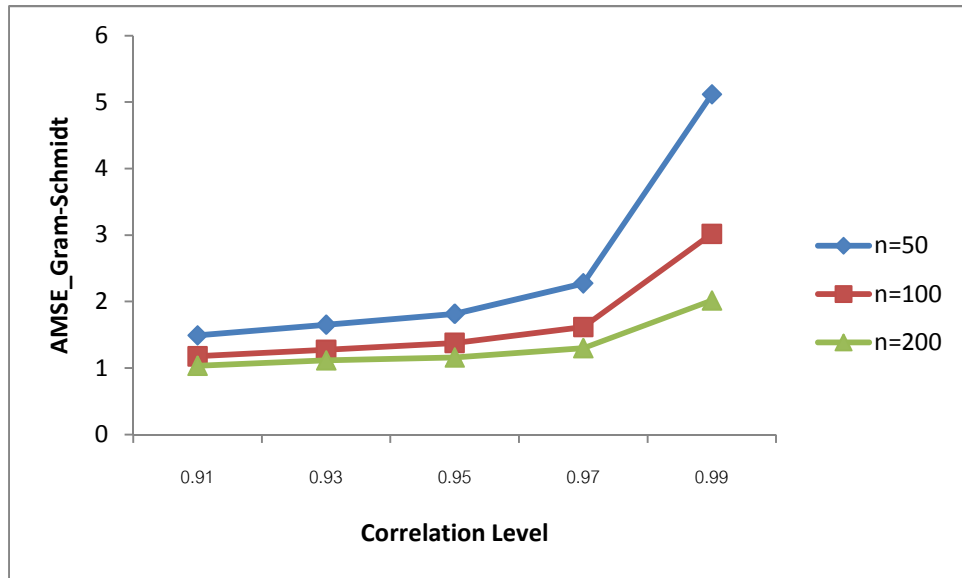
รูปที่ 4.3.11 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่างของวิธี PLS กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



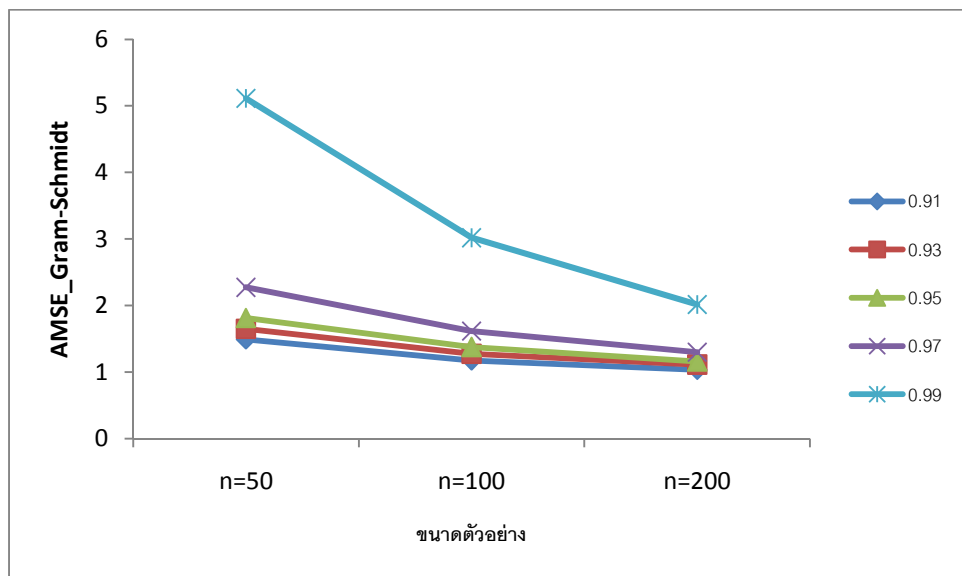
รูปที่ 4.3.12 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ของวิธี PLS กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



รูปที่ 4.3.13 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามขนาดตัวอย่างของวิธี OLS_G กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



รูปที่ 4.3.14 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์สูงสุด แยกตามระดับความสัมพันธ์ของวิธี OLS_G กรณีที่มีจำนวนตัวแปรอิสระ 3 ตัวแปร



ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ณ ระดับความสัมพันธ์ต่างๆ แยกตามขนาดตัวอย่าง ของทั้งสองวิธี สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร เป็นดังนี้

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ทำการวิเคราะห์ด้วยวิธีกำลังสองน้อยสุดเชิงส่วน พบว่าที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือที่ขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด และเมื่อพิจารณาแนวโน้มพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง พบว่าที่ขนาดตัวอย่างเท่ากับ 50 และ 100 จะไม่มีแนวโน้มที่แน่นอนเมื่อระดับความสัมพันธ์เพิ่มขึ้น แต่ที่ขนาดตัวอย่างเท่ากับ 200 จะมีแนวโน้มลดลงเล็กน้อยเมื่อระดับความสัมพันธ์เพิ่มขึ้น ซึ่งแสดงได้ดังรูปที่ 4.3.11

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ทำการวิเคราะห์ด้วยวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิคิต์ พบว่าที่ขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือที่ขนาดตัวอย่างเท่ากับ 100 และที่ขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด และเมื่อพิจารณาแนวโน้มพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มเพิ่มขึ้นตามระดับความสัมพันธ์เพิ่มของทุกขนาดตัวอย่างที่ทำการศึกษา ซึ่งแสดงได้ดังรูปที่ 4.3.13

ผลการวิเคราะห์เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ณ ระดับความสัมพันธ์ต่างๆ แยกตามระดับความสัมพันธ์ ของทั้งสองวิธี สำหรับข้อมูลของตัวแปรอิสระเท่ากับ 3 ตัวแปร เป็นดังนี้

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ทำการวิเคราะห์ด้วยวิธีกำลังสองน้อยสุดเชิงส่วน ตามระดับความสัมพันธ์ พบว่าที่ขนาดตัวอย่างเท่ากับ 50 ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.93, 0.95, 0.91 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.97 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด ที่ขนาดตัวอย่างเท่ากับ 100 ระดับความสัมพันธ์เท่ากับ 0.93 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือระดับความสัมพันธ์ 0.91, 0.95, 0.97 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด และที่ขนาดตัวอย่างเท่ากับ 200 ระดับความสัมพันธ์เท่ากับ 0.91 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.93, 0.95, 0.97 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยความคลาดเคลื่อน

กำลังสองต่ำที่สุด และเมื่อพิจารณาแนวโน้มของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง พบว่ามีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นของทุกระดับความสัมพันธ์ที่ทำการศึกษา ซึ่งแสดงได้ดังรูปที่ 4.3.12

เมื่อพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ ที่ทำการวิเคราะห์ด้วยวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามชมิสต์ ตามระดับความสัมพันธ์ พบว่าที่ระดับความสัมพันธ์เท่ากับ 0.99 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงที่สุด รองลงมาคือระดับความสัมพันธ์เท่ากับ 0.97, 0.95, 0.93 ตามลำดับ และที่ระดับความสัมพันธ์เท่ากับ 0.91 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำที่สุด ทุกขนาดตัวอย่างที่ทำการศึกษา และเมื่อพิจารณาแนวโน้มพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองจะมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้นของทุกระดับความสัมพันธ์ที่ทำการศึกษา ซึ่งแสดงได้ดังรูปที่ 4.3.14

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การศึกษานี้มีวัตถุประสงค์เพื่อหาวิธีการประมาณค่าพารามิเตอร์ที่เหมาะสม ระหว่างวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยวิธีการแปลงข้อมูลแบบกรามซมิตต์ ซึ่งทำการศึกษาและดำเนินการภายใต้ขอบเขตต่างๆ ดังนี้

1. กำหนดจำนวนตัวแปรอิสระที่ทำการศึกษา คือ 2 และ 3 ตัวแปร
2. กำหนดขนาดตัวอย่างที่ทำการศึกษา คือ 50, 100 และ 200
3. กำหนดค่าสัมประสิทธิ์ความถดถอยเบื้องต้น โดยให้ค่า $\beta_j = 1$ สำหรับ $j = 0, 1, \dots, p$
4. กำหนดระดับความสัมพันธ์ระหว่างตัวแปรอิสระให้มีความสัมพันธ์แบบเชิงเส้นเท่านั้น
 - 4.1 โดยถ้าจำนวนตัวแปรอิสระเป็น 2 ตัวแปร ผู้วิจัยจะทำการศึกษาภายใต้ 10 ระดับ คือ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99
 - 4.2 โดยถ้าจำนวนตัวแปรอิสระเป็น 3 ตัวแปร ผู้วิจัยจะทำการศึกษาภายใต้ 5 ระดับ คือ 0.91, 0.93, 0.95, 0.97 และ 0.99
5. กำหนดค่าความแปรปรวนของความคลาดเคลื่อน σ^2 มีค่าเป็น 10 และค่าความแปรปรวนของตัวแปรอิสระ X_1 , X_2 และ X_3 มีค่าเป็น $\sigma_1^2 = 4$, $\sigma_2^2 = 4$ และ $\sigma_3^2 = 4$ ตามลำดับ
6. กำหนดจำนวนตัวแปรอิสระที่มีพหุสัมพันธ์และระดับความสัมพันธ์ของตัวแปรเป็น 2 กรณีดังนี้
 - 6.1 กรณีที่จำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร ผู้วิจัยได้ทำการศึกษาเมื่อตัวแปรอิสระ 2 ตัวแปร ได้แก่ X_1 และ X_2 มีพหุสัมพันธ์กัน โดยให้ ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_2 ซึ่งศึกษาภายใต้ระดับความสัมพันธ์ คือ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98 และ 0.99
 - 6.2 กรณีที่จำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร ผู้วิจัยได้ทำการศึกษาเมื่อตัวแปรอิสระ 3 ตัวแปร ได้แก่ X_1 , X_2 และ X_3 มีพหุสัมพันธ์กัน ให้ $(\rho_{12}, \rho_{13}, \rho_{23})$ เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ โดย ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_2

ρ_{13} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_1 และ X_3

ρ_{23} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระ X_2 และ X_3

ซึ่งศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

(0.91, 0.91, 0.91)	(0.91, 0.91, 0.93)	(0.91, 0.91, 0.95)
(0.91, 0.91, 0.97)	(0.91, 0.91, 0.99)	(0.91, 0.93, 0.93)
(0.91, 0.93, 0.95)	(0.91, 0.93, 0.97)	(0.91, 0.93, 0.99)
(0.91, 0.95, 0.95)	(0.91, 0.95, 0.97)	(0.91, 0.95, 0.99)
(0.91, 0.97, 0.97)	(0.93, 0.93, 0.93)	(0.93, 0.93, 0.95)
(0.93, 0.93, 0.97)	(0.93, 0.93, 0.99)	(0.93, 0.95, 0.95)
(0.93, 0.95, 0.97)	(0.93, 0.95, 0.99)	(0.93, 0.97, 0.97)
(0.93, 0.97, 0.99)	(0.95, 0.95, 0.95)	(0.95, 0.95, 0.97)
(0.95, 0.95, 0.99)	(0.95, 0.97, 0.97)	(0.95, 0.97, 0.99)
(0.97, 0.97, 0.97)	(0.97, 0.97, 0.99)	(0.97, 0.99, 0.99)
(0.99, 0.99, 0.99)		

การวิเคราะห์ผลของการศึกษาทำการจำลองข้อมูลโดยใช้โปรแกรม R และสามารถสรุปผลการศึกษาในกรณีต่างๆ ได้ดังนี้

5.1 สรุปผลการศึกษา

1.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร

ผลการวิเคราะห์กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัวแปร พบว่าที่ระดับความสัมพันธ์เท่ากับ 0.95 จะเป็นจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาของความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระของทุกขนาดตัวอย่างที่ทำการศึกษา โดยค่าเฉลี่ยของ VIF จะแปรผันตรงกับระดับความสัมพันธ์และแปรผกผันกับขนาดตัวอย่าง และพบว่าที่ระดับความสัมพันธ์ 0.9, 0.91, 0.92, 0.93 และ 0.94 จะมีเปอร์เซ็นต์การเกิดปัญหาความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 0%-32% โดยเปอร์เซ็นต์การเกิดปัญหาความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระจะแปรผันตรงกับระดับความสัมพันธ์และแปรผกผันกับขนาดตัวอย่าง ส่วนที่ระดับความสัมพันธ์ 0.95, 0.96, 0.97, 0.98 และ 0.99 จะมีเปอร์เซ็นต์การเกิดปัญหาความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 55%-100% โดยเปอร์เซ็นต์การเกิดปัญหาความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระจะแปรผันตรงกับ

ระดับความสัมพันธ์และขนาดตัวอย่าง ซึ่งจะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงสูงทุกขนาดตัวอย่างที่ทำการศึกษา

ผลการวิเคราะห์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ พบว่าที่ขนาดตัวอย่างเท่ากับ 50 ที่ระดับความสัมพันธ์เท่ากับ 0.9, 0.91, 0.92, 0.93 และ 0.94 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์มีประสิทธิภาพดีกว่าการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วน แต่ที่ระดับความสัมพันธ์เท่ากับ 0.95, 0.96, 0.97, 0.98 และ 0.99 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนมีประสิทธิภาพดีกว่าการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์ ที่ขนาดตัวอย่างเท่ากับ 100 ที่ระดับความสัมพันธ์เท่ากับ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, และ 0.96 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์มีประสิทธิภาพดีกว่าการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วน แต่ที่ระดับความสัมพันธ์เท่ากับ 0.97, 0.98 และ 0.99 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนมีประสิทธิภาพดีกว่าการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์ และที่ขนาดตัวอย่างเท่ากับ 200 ที่ระดับความสัมพันธ์เท่ากับ 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97 และ 0.98 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์มีประสิทธิภาพดีกว่าการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วน แต่ที่ระดับความสัมพันธ์เท่ากับ 0.99 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนมีประสิทธิภาพดีกว่าการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์ เนื่องจากวิธีกำลังสองน้อยสุดเชิงส่วนเป็นวิธีที่พัฒนามาเพื่อใช้แก้ไขปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณมีค่าลดลง โดยปัจจัยที่มีผลต่อค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง นั้นพบว่าเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น วิธีกำลังสองน้อยสุดเชิงส่วนจะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองไม่มีแนวโน้มที่แน่นอน แต่วิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองแนวโน้มเพิ่มขึ้น ถ้าขนาดตัวอย่างเพิ่มขึ้น ทั้งวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มลดลง สำหรับทุกกรณีที่ทำการศึกษาในครั้งนี้

1.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร

ผลการวิเคราะห์กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัวแปร พบว่าที่ระดับความสัมพันธ์เท่ากับ 0.93 จะเป็นจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาของความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระของขนาดตัวอย่างเท่ากับ 50 และ 100 และขนาดตัวอย่างเท่ากับ 200 จะมีจุดเปลี่ยน (Cut Point) ที่ทำให้เกิดปัญหาของระดับความสัมพันธ์รุนแรงระหว่างตัวแปรอิสระที่ระดับความสัมพันธ์เท่ากับ 0.95 โดยค่าเฉลี่ยของ VIF จะแปรผันตรงกับระดับความสัมพันธ์และแปรผกผันกับขนาดตัวอย่าง และพบว่าที่ระดับความสัมพันธ์ 0.91 ขนาดตัวอย่างเท่ากับ 50 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 33% ขนาดตัวอย่างเท่ากับ 100 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 15% และขนาดตัวอย่างเท่ากับ 200 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 4% ส่วนที่ระดับความสัมพันธ์ 0.93, 0.95, 0.97 และ 0.99 จะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระอยู่ประมาณ 60%-100% ซึ่งจะมีเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงสูงทุกขนาดตัวอย่างที่ทำการศึกษา โดยที่ระดับความสัมพันธ์ 0.91 และ 0.93 เปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระจะแปรผันตรงกับระดับความสัมพันธ์และแปรผกผันกับขนาดตัวอย่าง ส่วนที่ระดับความสัมพันธ์ 0.95, 0.97 และ 0.99 เปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงระหว่างตัวแปรอิสระจะแปรผันตรงกับระดับความสัมพันธ์และขนาดตัวอย่าง

ผลการวิเคราะห์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณ พบว่าการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดเชิงส่วนมีประสิทธิภาพดีกว่าการประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์ ทุกกรณีที่ทำการศึกษา เนื่องจากวิธีกำลังสองน้อยสุดเชิงส่วนเป็นวิธีที่พัฒนามาเพื่อใช้แก้ไขปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของตัวประมาณมีค่าลดลง โดยปัจจัยที่มีผลต่อค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง นั้นพบว่า เมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น วิธีกำลังสองน้อยสุดเชิงส่วนจะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองไม่มีแนวโน้มที่แน่นอน ยกเว้นกรณีขนาดตัวอย่างเท่ากับ 200 จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองแนวโน้มลดลงเล็กน้อย แต่วิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกรามชมิตต์จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองแนวโน้มเพิ่มขึ้น ถ้าขนาดตัวอย่างเพิ่มขึ้น ทั้งวิธีกำลังสองน้อยสุดเชิงส่วนและวิธีกำลังสองน้อยสุดด้วยการแปลงข้อมูลแบบกราม

ซิมิตต์จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองแนวโน้มลดลง สำหรับทุกกรณีที่ทำการศึกษาในครั้งนี้

5.3 ข้อเสนอแนะ

เพื่อเป็นแนวทางในการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยของตัวแบบความถดถอยพหุเชิงเส้นที่เกิดพหุสัมพันธ์ที่เหมาะสม สำหรับผู้ที่สนใจศึกษาเพิ่มเติมในครั้งต่อไป

1. ในการศึกษาครั้งนี้มีจำนวนตัวแปรอิสระ 2 และ 3 ตัวแปร สำหรับในการศึกษาครั้งต่อไปควรศึกษาเพิ่มเติมในกรณีที่มีจำนวนตัวแปรอิสระมากขึ้น เพื่อให้การประมาณค่าสัมประสิทธิ์ความถดถอยเมื่อเกิดปัญหาพหุเชิงเส้นมีประสิทธิภาพมากขึ้น

2. ในการศึกษาครั้งนี้มีจำนวนตัวแปรตามเพียงตัวเดียว สำหรับในการศึกษาครั้งต่อไปควรศึกษาเพิ่มเติมในกรณีที่มีจำนวนตัวแปรตามมากกว่าหนึ่งตัว

3. ในการศึกษาครั้งนี้ทางผู้วิจัยได้ทำการเปรียบเทียบวิธีการประมาณค่า 2 วิธี ดังนั้นในการศึกษาครั้งต่อไปควรเพิ่มวิธีการประมาณค่า เพื่อให้เกิดการตัดสินใจเลือกวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยได้อย่างเหมาะสมสำหรับข้อมูลที่ทำการศึกษา

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร, พิมพ์ครั้งที่ 4. กรุงเทพมหานคร: 2552.

เปรมวดี ชูใสว. การเปรียบเทียบวิธีการแก้ไขพหุสัมพันธ์ในการถดถอยเชิงเส้นพหุคูณ. วิทยานิพนธ์

ปริญญาามหาบัณฑิต, สาขาวิชาสถิติ ภาคสถิติ คณะพาณิชยศาสตร์และการบัญชี

จุฬาลงกรณ์มหาวิทยาลัย, 2548.

เสกสรร เกียรติสุไพบูรณ์. การจำลอง, กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย,

2555.

ภาษาอังกฤษ

Paul H. Garthwaite. An Interpretation of Partial Least Squares. Journal of the American

Statistical Association, Vol. 89, No. 425 (Mar., 1994): 122-127.

บรรณานุกรม

ภาษาไทย

- กรรณิกาณ หิรัญกสิ. การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอย
พหุโดยวิธีกำลังสองน้อยที่สุดแบ่งส่วน วิธีการถดถอยองค์ประกอบ และวิธีกำลังสองน้อย
ที่สุดในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ. วิทยานิพนธ์ปริญญาามหาบัณฑิต,
 สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี
 จุฬาลงกรณ์มหาวิทยาลัย, 2540.
- ดำรง ทิพย์โยธา และ เพ็ญพรรณ ยังคง. พีชคณิตเชิงเส้น. กรุงเทพมหานคร: สำนักพิมพ์
 จุฬาลงกรณ์มหาวิทยาลัย, 2539.
- ปราณี คำแก้ว. การวิเคราะห์โครงสร้างยา โดยวิธีการถดถอยกำลังสองน้อยที่สุดบางส่วน.
 วิทยานิพนธ์ปริญญาามหาบัณฑิต, สาขาวิชาสถิติประยุกต์ ภาควิชาสถิติ
 คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่, 2552.
- สุพล ดุรงค์วัฒนา. ตัวแบบและการวิเคราะห์ความถดถอย สำหรับการวิจัยขั้นสูง,
 กรุงเทพมหานคร, 2549.

ภาษาอังกฤษ

- Dr. L. M. Bhar. Multicollinearity: Causes, Effects and Remedies. M. Sc. (Agricultural
Statistics) [Ebook]. (2005-2006).
- Inge S. Helland. Partial Least Squares Regression and Statistical Models. Scandinavian
Journal of Statistics, Vol. 17, No. 2 (1990), pp. 97-114.
- Ramesha K et al. Gram-Schmidt Orthogonalization Based Face Recognition using DWT,
International Journal of Engineering Science and Technoloe (IJEST), Vol. 3
 No. 1 Jan 2011, pp. 498.

ภาคผนวก

รายละเอียดของโปรแกรม

```
library(MASS)
library(psych)
library(MVA)
library(fmsb)
library(perturb)
library("corpcor")
function_vif <-function(x,k=10)ifelse(any(x>k),1,0)
function_det <-function(z,l=0)ifelse(all(z>l),1,0)
N=1000
B0=1
B1=1
B2=1
B3=1
V1=4
V2=4
V3=4
V=10
p1=2
p2=3
n1=50
n2=50
n3=100
VIFp2_TOTAL <-c()
MSE2_p2 <-c()
MSE2N_p2 <-c()
VIFp3_TOTAL <-c()
MSE3_p3 <-c()
```

```

MSE3N_p3 <-c()

#-----p=2-----#

for(i in 1:N){
  vifp2_total <-c()
  MSE2 <-c()
  MSE2N <-c()

  corr12 <-c(0.9,0.91,0.92,0.93,0.94,0.95,0.96,0.97,0.98,0.99)
  for(t in 1:length(corr12)){
    Cor12 <-corr12[t]
    #n=50#
    z1 <-rnorm(n1)
    z2 <-rnorm(n1)
    z3 <-rnorm(n1)
    #n=100#
    z4 <-rnorm(n2)
    z5 <-rnorm(n2)
    z6 <-rnorm(n2)
    z14 <-c(z1,z4)
    z25 <-c(z2,z5)
    z36 <-c(z3,z6)
    #n=200#
    z7 <-rnorm(n3)
    z8 <-rnorm(n3)
    z9 <-rnorm(n3)
    z147 <-c(z1,z4,z7)
    z258 <-c(z2,z5,z8)
    z369 <-c(z3,z6,z9)

    #Generate Error#
    ERROR1 <-rnorm(50,0,sqrt(V))
  }
}

```

```

ERROR1_m <-matrix(c(ERROR1),50,1)
ERROR2 <-rnorm(50,0,sqrt(V))
ERROR2_C <-c(ERROR1,ERROR2)
ERROR2_m <-matrix(c(ERROR2_C),100,1)
ERROR3 <-rnorm(100,0,sqrt(V))
ERROR3_C <-c(ERROR1,ERROR2,ERROR3)
ERROR3_m <-matrix(c(ERROR3_C),200,1)

#n=50#
q1 <-sqrt(V1)*z1
q2 <-sqrt(V2)*((Cor12)*z1+sqrt(1-(Cor12)^2)*z2)

#n=100#
r1 <-sqrt(V1)*z14
r2 <-sqrt(V2)*((Cor12)*z14+sqrt(1-(Cor12)^2)*z25)

#n=200#
f1 <-sqrt(V1)*z147
f2 <-sqrt(V2)*((Cor12)*z147+sqrt(1-(Cor12)^2)*z258)

#-----Check Multicollinearity:VIF p=2-----#

#n=50#
dat1_p2 <-data.frame(q1,q2)
lm1_p2 <-lm(q1~q2,data=dat1_p2)
lm2_p2 <-lm(q2~q1,data=dat1_p2)
Rsq1_p2 <-summary(lm1_p2)$r.squared
Rsq2_p2 <-summary(lm2_p2)$r.squared
vif1_p2 <-1/(1-Rsq1_p2)
vif2_p2 <-1/(1-Rsq2_p2)

#n=100#
dat2_p2 <-data.frame(r1,r2)
lm3_p2 <-lm(r1~r2,data=dat2_p2)
lm4_p2 <-lm(r2~r1,data=dat2_p2)

```

```

Rsq3_p2 <-summary(lm3_p2)$r.squared
Rsq4_p2 <-summary(lm4_p2)$r.squared
vif3_p2 <-1/(1-Rsq3_p2)
vif4_p2 <-1/(1-Rsq4_p2)
#n=200#
dat3_p2 <-data.frame(f1,f2)
lm5_p2 <-lm(f1~f2,data=dat3_p2)
lm6_p2 <-lm(f2~f1,data=dat3_p2)
Rsq5_p2 <-summary(lm5_p2)$r.squared
Rsq6_p2 <-summary(lm6_p2)$r.squared
vif5_p2 <-1/(1-Rsq5_p2)
vif6_p2 <-1/(1-Rsq6_p2)

#-----#

B <-matrix(c(B0,B1,B2),3,1)
x1 <-data.frame(1,q1,q2)
x1_m <-as.matrix(x1)
y1 <-(x1_m%*%B)+ERROR1_m
x2 <-data.frame(1,r1,r2)
x2_m <-as.matrix(x2)
y2 <-(x2_m%*%B)+ERROR2_m
x3 <-data.frame(1,f1,f2)
x3_m <-as.matrix(x3)
y3 <-(x3_m%*%B)+ERROR3_m

#-----#

Var_Cov1 <- var(dat1_p2)
R <- cov2cor(Var_Cov1)
pos.Var_Cov1 <-is.positive.definite(Var_Cov1)
if(pos.Var_Cov1=="TRUE"){k<-dat1_p2}
if(pos.Var_Cov1=="FALSE"){break}

```

```

a1<-k[,1]
a2<-k[,2]

#-----PLS-P=2-n=50-----#

#centering#
y1bar <-mean(y1)
a1bar <-mean(a1)
a2bar <-mean(a2)
y11 <-as.matrix(y1-y1bar)
a11 <-as.matrix(a1-a1bar)
a12 <-as.matrix(a2-a2bar)

#compute p1#
u11 <-as.numeric((ginv(t(a11)%*%a11))%*(t(a11)%*%y11))
u12 <-as.numeric((ginv(t(a12)%*%a12))%*(t(a12)%*%y11))
y1hat11 <-u11*a11
y1hat12 <-u12*a12
w1 <-as.numeric(t(a11)%*%a11)
w2 <-as.numeric(t(a12)%*%a12)
p11_1 <-w1*y1hat11
p12_1 <-w2*y1hat12
p1.PLS1 <-p11_1+p12_1
by1 <-
as.numeric((ginv(t(p1.PLS1)%*%p1.PLS1))%*(t(p1.PLS1)%*%y11))
ba1 <-
as.numeric((ginv(t(p1.PLS1)%*%p1.PLS1))%*(t(p1.PLS1)%*%a11))
ba2 <-
as.numeric((ginv(t(p1.PLS1)%*%p1.PLS1))%*(t(p1.PLS1)%*%a12))
y21 <-(y11-(by1*p1.PLS1))
a21 <-(a11-(ba1*p1.PLS1))
a22 <-(a12-(ba2*p1.PLS1))

```



```

#compute p2#
u21 <-as.numeric((ginv(t(a21)%*%a21))%*(t(a21)%*%y21))
u22 <-as.numeric((ginv(t(a22)%*%a22))%*(t(a22)%*%y21))
y1hat21 <-u21*a21
y1hat22 <-u22*a22
w3 <-as.numeric(t(a21)%*%a21)
w4 <-as.numeric(t(a22)%*%a22)
p21_1 <-w3*y1hat21
p22_1 <-w4*y1hat22
p2.PLS1 <-p21_1+p22_1
P.PLS1 <-data.frame(1,p1.PLS1,p2.PLS1)
P.PLS1_m <-as.matrix(P.PLS1)
cr12_PLS1 <-cor(p1.PLS1,p2.PLS1)

#Parameters estimation#
lmdat1 <-lm(y1~p1.PLS1+p2.PLS1)
coef1_PLS1 <-coef(lmdat1)
b50_0 <-coef1_PLS1[1]
b50_1 <-coef1_PLS1[2]
b50_2 <-coef1_PLS1[3]
y1hat.PLS1 <-b50_0+b50_1*p1.PLS1+b50_2*p2.PLS1
E.PLS1 <-as.matrix(y1-y1hat.PLS1)
mse.PLS1 <-as.numeric((t(E.PLS1)%*%E.PLS1)/(n1))
inv1 <-ginv(t(P.PLS1_m)%*%P.PLS1_m)
var.PLS1 <-mse.PLS1*inv1
var50_0 <-var.PLS1[1,1]
var50_1 <-var.PLS1[2,2]
var50_2 <-var.PLS1[3,3]
bias50_0 <-abs(B0-b50_0)
bias50_1 <-abs(B1-b50_1)

```

```

bias50_2 <-abs(B2-b50_2)
Mse50_0 <-var50_0+(bias50_0)^2
Mse50_1 <-var50_1+(bias50_1)^2
Mse50_2 <-var50_2+(bias50_2)^2
Mse50 <-(Mse50_0+Mse50_1+Mse50_2)/3

#-----#
Var_Cov2 <- var(dat2_p2)
R <- cov2cor(Var_Cov2)
pos.Var_Cov2 <-is.positive.definite(Var_Cov2)
if(pos.Var_Cov2=="TRUE"){h<-dat2_p2}
if(pos.Var_Cov2=="FALSE"){break}
b1<-h[,1]
b2<-h[,2]

#-----PLS-P=2-n=100-----#
#centering#
y2bar <-mean(y2)
b1bar <-mean(b1)
b2bar <-mean(b2)
y12 <-as.matrix(y2-y2bar)
b11 <-as.matrix(b1-b1bar)
b12 <-as.matrix(b2-b2bar)

#compute p1#
v11 <-as.numeric((ginv(t(b11)%*%b11))%*%(t(b11)%*%y12))
v12 <-as.numeric((ginv(t(b12)%*%b12))%*%(t(b12)%*%y12))
y2hat11 <-v11*b11
y2hat12 <-v12*b12
w5 <-as.numeric(t(b11)%*%b11)
w6 <-as.numeric(t(b12)%*%b12)
p11_2 <-w5*y2hat11

```

```

p12_2 <-w6*y2hat12
p1.PLS2 <-p11_2+p12_2
by2 <-
as.numeric((ginv(t(p1.PLS2)%*%p1.PLS2))%*(t(p1.PLS2)%*%y12))
bb1 <-
as.numeric((ginv(t(p1.PLS2)%*%p1.PLS2))%*(t(p1.PLS2)%*%b11))
bb2 <-
as.numeric((ginv(t(p1.PLS2)%*%p1.PLS2))%*(t(p1.PLS2)%*%b12))
y22 <-(y12-(by2*p1.PLS2))
b21 <-(b11-(bb1*p1.PLS2))
b22 <-(b12-(bb2*p1.PLS2))
#compute p2#
v21 <-as.numeric((ginv(t(b21)%*%b21))%*(t(b21)%*%y22))
v22 <-as.numeric((ginv(t(b22)%*%b22))%*(t(b22)%*%y22))
y2hat21 <-v21*b21
y2hat22 <-v22*b22
w7 <-as.numeric(t(b21)%*%b21)
w8 <-as.numeric(t(b22)%*%b22)
p21_2 <-w7*y2hat21
p22_2 <-w8*y2hat22
p2.PLS2 <-p21_2+p22_2
P.PLS2 <-data.frame(1,p1.PLS2,p2.PLS2)
P.PLS2_m <-as.matrix(P.PLS2)
cr12_PLS2 <-cor(p1.PLS2,p2.PLS2)
#Parameters estimation#
lmdat2 <-lm(y2~p1.PLS2+p2.PLS2)
coef2_PLS2 <-coef(lmdat2)
b100_0 <-coef2_PLS2[1]
b100_1 <-coef2_PLS2[2]

```

```

b100_2 <-coef2_PLS2[3]
y2hat.PLS2 <-b100_0+b100_1*p1.PLS2+b100_2*p2.PLS2
E.PLS2 <-as.matrix(y2-y2hat.PLS2)
mse.PLS2 <-as.numeric((t(E.PLS2)%*%E.PLS2)/(n1+n2))
inv2 <-ginv(t(P.PLS2_m)%*%P.PLS2_m)
var.PLS2 <-mse.PLS2*inv2
var100_0 <-var.PLS2[1,1]
var100_1 <-var.PLS2[2,2]
var100_2 <-var.PLS2[3,3]
bias100_0 <-abs(B0-b100_0)
bias100_1 <-abs(B1-b100_1)
bias100_2 <-abs(B2-b100_2)
Mse100_0 <-var100_0+(bias100_0)^2
Mse100_1 <-var100_1+(bias100_1)^2
Mse100_2 <-var100_2+(bias100_2)^2
Mse100 <-(Mse100_0+Mse100_1+Mse100_2)/3

#-----#

Var_Cov3 <- var(dat3_p2)
R <- cov2cor(Var_Cov3)
pos.Var_Cov3 <-is.positive.definite(Var_Cov3)
if(pos.Var_Cov3=="TRUE"){f<-dat3_p2}
if(pos.Var_Cov3=="FALSE"){break}
c1<-f[,1]
c2<-f[,2]

#-----PLS-P=2-n=200-----#

#centering#
y3bar <-mean(y3)
c1bar <-mean(c1)
c2bar <-mean(c2)

```

```

y13 <-as.matrix(y3-y3bar)
c11 <-as.matrix(c1-c1bar)
c12 <-as.matrix(c2-c2bar)

#compute p1#
s11 <-as.numeric((ginv(t(c11)%*%c11))%*%(t(c11)%*%y13))
s12 <-as.numeric((ginv(t(c12)%*%c12))%*%(t(c12)%*%y13))
y3hat11 <-s11*c11
y3hat12 <-s12*c12
w9 <-as.numeric(t(c11)%*%c11)
w10 <-as.numeric(t(c12)%*%c12)
p11_3 <-w9*y3hat11
p12_3 <-w10*y3hat12
p1.PLS3 <-p11_3+p12_3
by3 <-
as.numeric((ginv(t(p1.PLS3)%*%p1.PLS3))%*%(t(p1.PLS3)%*%y13))
bc1 <-
as.numeric((ginv(t(p1.PLS3)%*%p1.PLS3))%*%(t(p1.PLS3)%*%c11))
bc2 <-
as.numeric((ginv(t(p1.PLS3)%*%p1.PLS3))%*%(t(p1.PLS3)%*%c12))
y23 <-(y13-(by3*p1.PLS3))
c21 <-(c11-(bc1*p1.PLS3))
c22 <-(c12-(bc2*p1.PLS3))

#compute p2#
s21 <-as.numeric((ginv(t(c21)%*%c21))%*%(t(c21)%*%y23))
s22 <-as.numeric((ginv(t(c22)%*%c22))%*%(t(c22)%*%y23))
y3hat21 <-s21*c21
y3hat22 <-s22*c22
w11 <-as.numeric(t(c21)%*%c21)
w12 <-as.numeric(t(c22)%*%c22)

```

```

p21_3 <-w11*y3hat21
p22_3 <-w12*y3hat22
p2.PLS3 <-p21_3+p22_3
P.PLS3 <-data.frame(1,p1.PLS3,p2.PLS3)
P.PLS3_m <-as.matrix(P.PLS3)
cr12_PLS3 <-cor(p1.PLS3,p2.PLS3)

#Parameters estimation#
lmdat3 <-lm(y3~p1.PLS3+p2.PLS3)
coef3_PLS3 <-coef(lmdat3)
b200_0 <-coef3_PLS3[1]
b200_1 <-coef3_PLS3[2]
b200_2 <-coef3_PLS3[3]
y3hat.PLS3 <-b200_0+b200_1*p1.PLS3+b200_2*p2.PLS3
E.PLS3 <-as.matrix(y3-y3hat.PLS3)
mse.PLS3 <-as.numeric((t(E.PLS3)%*%E.PLS3)/(n1+n2+n3))
inv3 <-ginv(t(P.PLS3_m)%*%P.PLS3_m)
var.PLS3 <-mse.PLS3*inv3
var200_0 <-var.PLS3[1,1]
var200_1 <-var.PLS3[2,2]
var200_2 <-var.PLS3[3,3]
bias200_0 <-abs(B0-b200_0)
bias200_1 <-abs(B1-b200_1)
bias200_2 <-abs(B2-b200_2)
Mse200_0 <-var200_0+(bias200_0)^2
Mse200_1 <-var200_1+(bias200_1)^2
Mse200_2 <-var200_2+(bias200_2)^2
Mse200 <-(Mse200_0+Mse200_1+Mse200_2)/3

```

```
#-----Gram-Schmidt-P=2-n=50-----#
```

```
#centering#
```

```
yS1 <-as.matrix(y1-mean(y1))
```

```
a1S <-as.matrix(a1-mean(a1))
```

```
a2S <-as.matrix(a2-mean(a2))
```

```
#Gram-Schmidt's data transformation Orthogonal#
```

```
p1.G1 <-a1S
```

```
p2.G1 <-a2S-(sum(a2S*p1.G1)/sum(p1.G1*p1.G1))*p1.G1
```

```
p.G1 <-data.frame(1,p1.G1,p2.G1)
```

```
P.G1_m <-as.matrix(p.G1)
```

```
cr12_G1 <-cor(p1.G1,p2.G1)
```

```
#Parameters estimation#
```

```
lmdatG1 <-lm(y1~p1.G1+p2.G1)
```

```
coef1_G1 <-coef(lmdatG1)
```

```
bG50_0 <-coef1_G1[1]
```

```
bG50_1 <-coef1_G1[2]
```

```
bG50_2 <-coef1_G1[3]
```

```
y1hat.G1 <-bG50_0+bG50_1*p1.G1+bG50_2*p2.G1
```

```
E.G1 <-as.matrix(y1-y1hat.G1)
```

```
mse.G1 <-as.numeric((t(E.G1)%*%E.G1)/(n1))
```

```
var.G1 <-mse.G1*(ginv(t(P.G1_m)%*%P.G1_m))
```

```
varG50_0 <-var.G1[1,1]
```

```
varG50_1 <-var.G1[2,2]
```

```
varG50_2 <-var.G1[3,3]
```

```
biasG50_0 <-abs(B0-bG50_0)
```

```
biasG50_1 <-abs(B1-bG50_1)
```

```
biasG50_2 <-abs(B2-bG50_2)
```

```
MseG50_0 <-varG50_0+(biasG50_0)^2
```

```
MseG50_1 <-varG50_1+(biasG50_1)^2
```

```

MseG50_2 <-varG50_2+(biasG50_2)^2
MseG50 <-(MseG50_0+MseG50_1+MseG50_2)/3
#-----Gram-Schmidt-P=2-n=100-----#
#centering#
yS2 <-as.matrix(y2-mean(y2))
b1S <-as.matrix(b1-mean(b1))
b2S <-as.matrix(b2-mean(b2))
#Gram-Schmidt's data transformation Orthogonal#
p1.G2 <-b1S
p2.G2 <-b2S-(sum(b2S*p1.G2)/sum(p1.G2*p1.G2))*p1.G2
p.G2 <-data.frame(1,p1.G2,p2.G2)
P.G2_m <-as.matrix(p.G2)
cr12_G2 <-cor(p1.G2,p2.G2)
#Parameters estimation#
lmdatG2 <-lm(y2~p1.G2+p2.G2)
coef2_G2 <-coef(lmdatG2)
bG100_0 <-coef2_G2[1]
bG100_1 <-coef2_G2[2]
bG100_2 <-coef2_G2[3]
y2hat.G2 <-bG100_0+bG100_1*p1.G2+bG100_2*p2.G2
E.G2 <-as.matrix(y2-y2hat.G2)
mse.G2 <-as.numeric((t(E.G2)%*%E.G2)/(n1+n2))
var.G2 <-mse.G2*(ginv(t(P.G2_m)%*%P.G2_m))
varG100_0 <-var.G2[1,1]
varG100_1 <-var.G2[2,2]
varG100_2 <-var.G2[3,3]
biasG100_0 <-abs(B0-bG100_0)
biasG100_1 <-abs(B1-bG100_1)
biasG100_2 <-abs(B2-bG100_2)

```



```

MseG100_0 <-varG100_0+(biasG100_0)^2
MseG100_1 <-varG100_1+(biasG100_1)^2
MseG100_2 <-varG100_2+(biasG100_2)^2
MseG100 <-(MseG100_0+MseG100_1+MseG100_2)/3

#-----Gram-Schmidt-P=2-n=200-----#

#centering#
y3S <-as.matrix(y3-mean(y3))
c1S <-as.matrix(c1-mean(c1))
c2S <-as.matrix(c2-mean(c2))

#Gram-Schmidt's data transformation Orthogonal#
p1.G3 <-c1S
p2.G3 <-c2S-(sum(c2S*p1.G3)/sum(p1.G3*p1.G3))*p1.G3

p.G3 <-data.frame(1,p1.G3,p2.G3)
P.G3_m <-as.matrix(p.G3)
cr12_G3 <-cor(p1.G3,p2.G3)

#Parameters estimation#
lmdatG3 <-lm(y3~p1.G3+p2.G3)
coef3_G3 <-coef(lmdatG3)
bG200_0 <-coef3_G3[1]
bG200_1 <-coef3_G3[2]
bG200_2 <-coef3_G3[3]
y3hat.G3 <-bG200_0+bG200_1*p1.G3+bG200_2*p2.G3
E.G3 <-as.matrix(y3-y3hat.G3)
mse.G3 <-as.numeric((t(E.G3)%*%E.G3)/(n1+n2+n3))
var.G3 <-mse.G3*(ginv(t(P.G3_m)%*%P.G3_m))
varG200_0 <-var.G3[1,1]
varG200_1 <-var.G3[2,2]
varG200_2 <-var.G3[3,3]

```

```

biasG200_0 <-abs(B0-bG200_0)
biasG200_1 <-abs(B1-bG200_1)
biasG200_2 <-abs(B2-bG200_2)
MseG200_0 <-varG200_0+(biasG200_0)^2
MseG200_1 <-varG200_1+(biasG200_1)^2
MseG200_2 <-varG200_2+(biasG200_2)^2
MseG200 <-(MseG200_0+MseG200_1+MseG200_2)/3
vifp2_total <-c(vifp2_total,vif1_p2,vif2_p2,vif3_p2,vif4_p2,vif5_p2,vif6_p2)
MSE2 <-c(MSE2,mse.PLS1,mse.G1,mse.PLS2,mse.G2,mse.PLS3,mse.G3)
MSE2N <-c(MSE2N,Mse50,MseG50,Mse100,MseG100,Mse200,MseG200)
}
VIFp2_TOTAL <-c(VIFp2_TOTAL,vifp2_total)
MSE2_p2 <-c(MSE2_p2,MSE2)
MSE2N_p2 <-c(MSE2N_p2,MSE2N)
corr <-c(0.91,0.93,0.95,0.97,0.99)
m <-t(combn(corr,3))
o <-
matrix(c(0.91,0.93,0.95,0.97,0.99,0.91,0.93,0.95,0.97,0.99,0.91,0.93,0.95,0.97,0.99),5,3)
mo <-rbind(m,o)
#-----P=3-----#
vifp3_total <-c()
MSE3 <-c()
MSE3N <-c()
mo[1:5,]-> mo1
mo[7:15,]-> mo2
n <-
matrix(c(0.91,0.91,0.91,0.91,0.91,0.91,0.91,0.91,0.91,0.91,0.91,0.93,0.95,0.97,0.93,0.95,
0.97,0.99,0.93,0.95,0.97),7,3)

```

```

q <-
matrix(c(0.93,0.93,0.93,0.93,0.93,0.95,0.95,0.95,0.97,0.97,0.93,0.93,0.93,0.95,0.97,0.95,
0.95,0.97,0.97,0.99,0.95,0.97,0.99,0.95,0.97,0.97,0.99,0.97,0.99,0.99),10,3)

mo_use <-rbind(mo1,mo2,n,q)
dim(mo_use)
dim(mo_use)[1]
for(l in 1:dim(mo_use)[1]){
  Cor12 <-mo_use[l,1]
  Cor13 <-mo_use[l,2]
  Cor23 <-mo_use[l,3]

#n=50#
  z1 <-rnorm(n1)
  z2 <-rnorm(n1)
  z3 <-rnorm(n1)

#n=100#
  z4 <-rnorm(n2)
  z5 <-rnorm(n2)
  z6 <-rnorm(n2)
  z14 <-c(z1,z4)
  z25 <-c(z2,z5)
  z36 <-c(z3,z6)

#n=200#
  z7 <-rnorm(n3)
  z8 <-rnorm(n3)
  z9 <-rnorm(n3)
  z147 <-c(z1,z4,z7)
  z258 <-c(z2,z5,z8)
  z369 <-c(z3,z6,z9)

```

```

#Generate Error#

ERROR1 <-rnorm(50,0,sqrt(V))
ERROR1_m <-matrix(c(ERROR1),50,1)
ERROR2 <-rnorm(50,0,sqrt(V))
ERROR2_C <-c(ERROR1,ERROR2)
ERROR2_m <-matrix(c(ERROR2_C),100,1)
ERROR3 <-rnorm(100,0,sqrt(V))
ERROR3_C <-c(ERROR1,ERROR2,ERROR3)
ERROR3_m <-matrix(c(ERROR3_C),200,1)

B <-matrix(c(B0,B1,B2,B3),4,1)

#n=50#

q1 <-sqrt(V1)*z1
q2 <-sqrt(V2)*((Cor12)*z1+sqrt(1-(Cor12)^2)*z2)
q3 <-sqrt(V3)*((Cor13)*z1+((Cor23-(Cor12*Cor13))/sqrt(1-
(Cor12)^2))*z2+(sqrt(1-(Cor13)^2-((Cor23-(Cor12*Cor13))^2/(1-(Cor12)^2))))*z3)
x4 <-data.frame(1,q1,q2,q3)
x4_m <-as.matrix(x4)
y4 <-(x4_m%*%B)+ERROR1_m

#n=100#

r1 <-sqrt(V1)*z14
r2 <-sqrt(V2)*((Cor12)*z14+sqrt(1-(Cor12)^2)*z25)
r3 <-sqrt(V3)*((Cor13)*z14+((Cor23-(Cor12*Cor13))/sqrt(1-
(Cor12)^2))*z25+(sqrt(1-(Cor13)^2-((Cor23-(Cor12*Cor13))^2/(1-(Cor12)^2))))*z36)
x5 <-data.frame(1,r1,r2,r3)
x5_m <-as.matrix(x5)
y5 <-(x5_m%*%B)+ERROR2_m

#n=200#

f1 <-sqrt(V1)*z147
f2 <-sqrt(V2)*((Cor12)*z147+sqrt(1-(Cor12)^2)*z258)

```

```

f3 <-sqrt(V3)*((Cor13)*z147+((Cor23-(Cor12*Cor13))/sqrt(1-
(Cor12)^2))*z258+(sqrt(1-(Cor13)^2-((Cor23-(Cor12*Cor13))^2/(1-(Cor12)^2))))*z369)
x6 <-data.frame(1,f1,f2,f3)
x6_m <-as.matrix(x6)
y6 <-(x6_m%*%B)+ERROR3_m
#-----Check Multicollinearity:VIF p=3-----#
#n=50#
dat1_p3 <-data.frame(q1,q2,q3)
lm1_p3 <-lm(q1~q2+q3,data=dat1_p3)
lm2_p3 <-lm(q2~q1+q3,data=dat1_p3)
lm3_p3 <-lm(q3~q1+q2,data=dat1_p3)
Rsq1_p3 <-summary(lm1_p3)$r.squared
Rsq2_p3 <-summary(lm2_p3)$r.squared
Rsq3_p3 <-summary(lm3_p3)$r.squared
vif1_p3 <-1/(1-Rsq1_p3)
vif2_p3 <-1/(1-Rsq2_p3)
vif3_p3 <-1/(1-Rsq3_p3)
#n=100#
dat2_p3 <-data.frame(r1,r2,r3)
lm4_p3 <-lm(r1~r2+r3,data=dat2_p3)
lm5_p3 <-lm(r2~r1+r3,data=dat2_p3)
lm6_p3 <-lm(r3~r1+r2,data=dat2_p3)
Rsq4_p3 <-summary(lm4_p3)$r.squared
Rsq5_p3 <-summary(lm5_p3)$r.squared
Rsq6_p3 <-summary(lm6_p3)$r.squared
vif4_p3 <-1/(1-Rsq4_p3)
vif5_p3 <-1/(1-Rsq5_p3)
vif6_p3 <-1/(1-Rsq6_p3)

```

```

#n=200#
  dat3_p3 <-data.frame(f1,f2,f3)
  lm7_p3 <-lm(f1~f2+f3,data=dat3_p3)
  lm8_p3 <-lm(f2~f1+f3,data=dat3_p3)
  lm9_p3 <-lm(f3~f1+f2,data=dat3_p3)
  Rsq7_p3 <-summary(lm7_p3)$r.squared
  Rsq8_p3 <-summary(lm8_p3)$r.squared
  Rsq9_p3 <-summary(lm9_p3)$r.squared
  vif7_p3 <-1/(1-Rsq7_p3)
  vif8_p3 <-1/(1-Rsq8_p3)
  vif9_p3 <-1/(1-Rsq9_p3)

#-----#
  Var_Cov1 <- var(dat1_p3)
  R <- cov2cor(Var_Cov1)
  pos.Var_Cov1 <-is.positive.definite(Var_Cov1)
  if(pos.Var_Cov1=="TRUE"){k<-dat1_p3}
  if(pos.Var_Cov1=="FALSE"){break}
  a1<-k[,1]
  a2<-k[,2]
  a3<-k[,3]

#-----PLS-p=3-n=50-----#
  #centering#
  y4bar <-mean(y4)
  a1bar <-mean(a1)
  a2bar <-mean(a2)
  a3bar <-mean(a3)
  y14 <-as.matrix(y4-y4bar)
  a31 <-as.matrix(a1-a1bar)
  a32 <-as.matrix(a2-a2bar)

```

```

a33 <-as.matrix(a3-a3bar)

#compute p1#
u31 <-as.numeric((ginv(t(a31)%*%a31))%*(t(a31)%*%y14))
u32 <-as.numeric((ginv(t(a32)%*%a32))%*(t(a32)%*%y14))
u33 <-as.numeric((ginv(t(a33)%*%a33))%*(t(a33)%*%y14))
y4hat31 <-u31*a31
y4hat32 <-u32*a32
y4hat33 <-u33*a33
w1 <-as.numeric(t(a31)%*%a31)
w2 <-as.numeric(t(a32)%*%a32)
w3 <-as.numeric(t(a33)%*%a33)
p11_4 <-w1*y4hat31
p12_4 <-w2*y4hat32
p13_4 <-w3*y4hat33
p1.PLS4 <-p11_4+p12_4+p13_4
by4 <-
as.numeric((ginv(t(p1.PLS4)%*%p1.PLS4))%*(t(p1.PLS4)%*%y14))
ba11 <-
as.numeric((ginv(t(p1.PLS4)%*%p1.PLS4))%*(t(p1.PLS4)%*%a31))
ba22 <-
as.numeric((ginv(t(p1.PLS4)%*%p1.PLS4))%*(t(p1.PLS4)%*%a32))
ba33 <-
as.numeric((ginv(t(p1.PLS4)%*%p1.PLS4))%*(t(p1.PLS4)%*%a33))
y24 <-(y14-(by4*p1.PLS4))
a41 <--(a31-(ba11*p1.PLS4))
a42 <--(a32-(ba22*p1.PLS4))
a43 <--(a33-(ba33*p1.PLS4))

#compute p2#
u41 <-as.numeric((ginv(t(a41)%*%a41))%*(t(a41)%*%y24))

```

```

u42 <-as.numeric((ginv(t(a42)%*%a42))%*(t(a42)%*%y24))
u43 <-as.numeric((ginv(t(a43)%*%a43))%*(t(a43)%*%y24))
y4hat41 <-u41*a41
y4hat42 <-u42*a42
y4hat43 <-u43*a43
w4 <-as.numeric(t(a41)%*%a41)
w5 <-as.numeric(t(a42)%*%a42)
w6 <-as.numeric(t(a43)%*%a43)
p21_4 <-w4*y4hat41
p22_4 <-w5*y4hat42
p23_4 <-w6*y4hat43
p2.PLS4 <-p21_4+p22_4+p23_4
by5 <-
as.numeric((ginv(t(p2.PLS4)%*%p2.PLS4))%*(t(p2.PLS4)%*%y24))
ba111 <-
as.numeric((ginv(t(p2.PLS4)%*%p2.PLS4))%*(t(p2.PLS4)%*%a41))
ba222 <-
as.numeric((ginv(t(p2.PLS4)%*%p2.PLS4))%*(t(p2.PLS4)%*%a42))
ba333 <-
as.numeric((ginv(t(p2.PLS4)%*%p2.PLS4))%*(t(p2.PLS4)%*%a43))
y34 <-(y24-(by5*p2.PLS4))
a51 <-(a41-(ba111*p2.PLS4))
a52 <-(a42-(ba222*p2.PLS4))
a53 <-(a43-(ba333*p2.PLS4))

#compute p3#
u51 <-as.numeric((ginv(t(a51)%*%a51))%*(t(a51)%*%y34))
u52 <-as.numeric((ginv(t(a52)%*%a52))%*(t(a52)%*%y34))
u53 <-as.numeric((ginv(t(a53)%*%a53))%*(t(a53)%*%y34))
y4hat51 <-u51*a51

```



```

y4hat52 <-u52*a52
y4hat53 <-u53*a53
w7 <-as.numeric(t(a51)%*%a51)
w8 <-as.numeric(t(a52)%*%a52)
w9 <-as.numeric(t(a53)%*%a53)
p31_4 <-w7*y4hat51
p32_4 <-w8*y4hat52
p33_4 <-w9*y4hat53
p3.PLS4 <-p31_4+p32_4+p33_4
P.PLS4 <-data.frame(1,p1.PLS4,p2.PLS4,p3.PLS4)
P.PLS4_m <-as.matrix(P.PLS4)
cr12_PLS4 <-cor(p1.PLS4,p2.PLS4)
cr13_PLS4 <-cor(p1.PLS4,p3.PLS4)
cr23_PLS4 <-cor(p2.PLS4,p3.PLS4)

#Parameters estimation#
lmdat4 <-lm(y4~p1.PLS4+p2.PLS4+p3.PLS4)
coef4_PLS4 <-coef(lmdat4)
b50_00 <-coef4_PLS4[1]
b50_11 <-coef4_PLS4[2]
b50_22 <-coef4_PLS4[3]
b50_33 <-coef4_PLS4[4]
y4hat.PLS4 <-
b50_00+b50_11*p1.PLS4+b50_22*p2.PLS4+b50_33*p3.PLS4
E.PLS4 <-as.matrix(y4-y4hat.PLS4)
mse.PLS4 <-as.numeric((t(E.PLS4)%*%E.PLS4)/(n1))
inv4 <-ginv(t(P.PLS4_m)%*%P.PLS4_m)
var.PLS4 <-mse.PLS4*inv4
var50_00 <-var.PLS4[1,1]
var50_11 <-var.PLS4[2,2]

```

```

var50_22 <-var.PLS4[3,3]
var50_33 <-var.PLS4[4,4]
bias50_00 <-abs(B0-b50_00)
bias50_11 <-abs(B1-b50_11)
bias50_22 <-abs(B2-b50_22)
bias50_33 <-abs(B3-b50_33)
Mse50_00 <-var50_00+(bias50_00)^2
Mse50_11 <-var50_11+(bias50_11)^2
Mse50_22 <-var50_22+(bias50_22)^2
Mse50_33 <-var50_33+(bias50_33)^2
Mse50n <-(Mse50_00+Mse50_11+Mse50_22+Mse50_33)/4
#-----#

Var_Cov2 <- var(dat2_p3)
R <- cov2cor(Var_Cov2)
pos.Var_Cov2 <-is.positive.definite(Var_Cov2)
if(pos.Var_Cov2=="TRUE"){h<-dat2_p3}
if(pos.Var_Cov2=="FALSE"){break}
b1<-h[,1]
b2<-h[,2]
b3<-h[,3]
#-----PLS-p=3-n=100-----#

#centering#
y5bar <-mean(y5)
b1bar <-mean(b1)
b2bar <-mean(b2)
b3bar <-mean(b3)
y15 <-as.matrix(y5-y5bar)
b31 <-as.matrix(b1-b1bar)
b32 <-as.matrix(b2-b2bar)

```

```

b33 <-as.matrix(b3-b3bar)

#compute p1#
v31 <-as.numeric((ginv(t(b31)%*%b31))%*(t(b31)%*%y15))
v32 <-as.numeric((ginv(t(b32)%*%b32))%*(t(b32)%*%y15))
v33 <-as.numeric((ginv(t(b33)%*%b33))%*(t(b33)%*%y15))
y5hat31 <-v31*b31
y5hat32 <-v32*b32
y5hat33 <-v33*b33
w10 <-as.numeric(t(b31)%*%b31)
w11 <-as.numeric(t(b32)%*%b32)
w12 <-as.numeric(t(b33)%*%b33)
p11_5 <-w10*y5hat31
p12_5 <-w11*y5hat32
p13_5 <-w12*y5hat33
p1.PLS5 <-p11_5+p12_5+p13_5
by6 <-
as.numeric((ginv(t(p1.PLS5)%*%p1.PLS5))%*(t(p1.PLS5)%*%y15))
bb11 <-
as.numeric((ginv(t(p1.PLS5)%*%p1.PLS5))%*(t(p1.PLS5)%*%b31))
bb22 <-
as.numeric((ginv(t(p1.PLS5)%*%p1.PLS5))%*(t(p1.PLS5)%*%b32))
bb33 <-
as.numeric((ginv(t(p1.PLS5)%*%p1.PLS5))%*(t(p1.PLS5)%*%b33))
y25 <-(y15-(by6*p1.PLS5))
b41 <-(b31-(bb11*p1.PLS5))
b42 <-(b32-(bb22*p1.PLS5))
b43 <-(b33-(bb33*p1.PLS5))

#compute p2#
v41 <-as.numeric((ginv(t(b41)%*%b41))%*(t(b41)%*%y25))

```

```

v42 <-as.numeric((ginv(t(b42)%*%b42))%*(t(b42)%*%y25))
v43 <-as.numeric((ginv(t(b43)%*%b43))%*(t(b43)%*%y25))
y5hat41 <-v41*b41
y5hat42 <-v42*b42
y5hat43 <-v43*b43
w13 <-as.numeric(t(b41)%*%b41)
w14 <-as.numeric(t(b42)%*%b42)
w15 <-as.numeric(t(b43)%*%b43)
p21_5 <-w13*y5hat41
p22_5 <-w14*y5hat42
p23_5 <-w15*y5hat43
p2.PLS5 <-p21_5+p22_5+p23_5
by7 <-
as.numeric((ginv(t(p2.PLS5)%*%p2.PLS5))%*(t(p2.PLS5)%*%y25))
bb111 <-
as.numeric((ginv(t(p2.PLS5)%*%p2.PLS5))%*(t(p2.PLS5)%*%b41))
bb222 <-
as.numeric((ginv(t(p2.PLS5)%*%p2.PLS5))%*(t(p2.PLS5)%*%b42))
bb333 <-
as.numeric((ginv(t(p2.PLS5)%*%p2.PLS5))%*(t(p2.PLS5)%*%b43))
y35 <-(y25-(by7*p2.PLS5))
b51 <-(b41-(bb111*p2.PLS5))
b52 <-(b42-(bb222*p2.PLS5))
b53 <-(b43-(bb333*p2.PLS5))

#compute p3#
v51 <-as.numeric((ginv(t(b51)%*%b51))%*(t(b51)%*%y35))
v52 <-as.numeric((ginv(t(b52)%*%b52))%*(t(b52)%*%y35))
v53 <-as.numeric((ginv(t(b53)%*%b53))%*(t(b53)%*%y35))
y5hat51 <-v51*b51

```

```

y5hat52 <-v52*b52
y5hat53 <-v53*b53
w16 <-as.numeric(t(b51)%*%b51)
w17 <-as.numeric(t(b52)%*%b52)
w18 <-as.numeric(t(b53)%*%b53)
p31_5 <-w16*y5hat51
p32_5 <-w17*y5hat52
p33_5 <-w18*y5hat53
p3.PLS5 <-p31_5+p32_5+p33_5
P.PLS5 <-data.frame(1,p1.PLS5,p2.PLS5,p3.PLS5)
P.PLS5_m <-as.matrix(P.PLS5)
cr12_PLS5 <-cor(p1.PLS5,p2.PLS5)
cr13_PLS5 <-cor(p1.PLS5,p3.PLS5)
cr23_PLS5 <-cor(p2.PLS5,p3.PLS5)

#Parameters estimation#
lmdat5 <-lm(y5~p1.PLS5+p2.PLS5+p3.PLS5)
coef5_PLS5 <-coef(lmdat5)
b100_00 <-coef5_PLS5[1]
b100_11 <-coef5_PLS5[2]
b100_22 <-coef5_PLS5[3]
b100_33 <-coef5_PLS5[4]
y5hat.PLS5 <-
b100_00+b100_11*p1.PLS5+b100_22*p2.PLS5+b100_33*p3.PLS5
E.PLS5 <-as.matrix(y5-y5hat.PLS5)
mse.PLS5 <-as.numeric((t(E.PLS5)%*%E.PLS5)/(n1+n2))
inv5 <-ginv(t(P.PLS5_m)%*%P.PLS5_m)
var.PLS5 <-mse.PLS5*inv5
var100_00 <-var.PLS5[1,1]
var100_11 <-var.PLS5[2,2]

```

```

var100_22 <-var.PLS5[3,3]
var100_33 <-var.PLS5[4,4]
bias100_00 <-abs(B0-b100_00)
bias100_11 <-abs(B1-b100_11)
bias100_22 <-abs(B2-b100_22)
bias100_33 <-abs(B3-b100_33)
Mse100_00 <-var100_00+(bias100_00)^2
Mse100_11 <-var100_11+(bias100_11)^2
Mse100_22 <-var100_22+(bias100_22)^2
Mse100_33 <-var100_33+(bias100_33)^2
Mse100n <-(Mse100_00+Mse100_11+Mse100_22+Mse100_33)/4
#-----#
Var_Cov3 <- var(dat3_p3)
R <- cov2cor(Var_Cov3)
pos.Var_Cov3 <-is.positive.definite(Var_Cov3)
if(pos.Var_Cov3=="TRUE"){f<-dat3_p3}
if(pos.Var_Cov3=="FALSE"){break}
c1<-f[,1]
c2<-f[,2]
c3<-f[,3]
#-----PLS-p=3-n=200-----#
#centering#
y6bar <-mean(y6)
c1bar <-mean(c1)
c2bar <-mean(c2)
c3bar <-mean(c3)
y16 <-as.matrix(y6-y6bar)
c31 <-as.matrix(c1-c1bar)
c32 <-as.matrix(c2-c2bar)

```

```

c33 <-as.matrix(c3-c3bar)

#compute p1#
s31 <-as.numeric((ginv(t(c31)%*%c31))%*(t(c31)%*%y16))
s32 <-as.numeric((ginv(t(c32)%*%c32))%*(t(c32)%*%y16))
s33 <-as.numeric((ginv(t(c33)%*%c33))%*(t(c33)%*%y16))
y6hat31 <-s31*c31
y6hat32 <-s32*c32
y6hat33 <-s33*c33
w19 <-as.numeric(t(c31)%*%c31)
w20 <-as.numeric(t(c32)%*%c32)
w21 <-as.numeric(t(c33)%*%c33)
p11_6 <-w19*y6hat31
p12_6 <-w20*y6hat32
p13_6 <-w21*y6hat33
p1.PLS6 <-p11_6+p12_6+p13_6
by8 <-
as.numeric((ginv(t(p1.PLS6)%*%p1.PLS6))%*(t(p1.PLS6)%*%y16))
bc11 <-
as.numeric((ginv(t(p1.PLS6)%*%p1.PLS6))%*(t(p1.PLS6)%*%c31))
bc22 <-
as.numeric((ginv(t(p1.PLS6)%*%p1.PLS6))%*(t(p1.PLS6)%*%c32))
bc33 <-
as.numeric((ginv(t(p1.PLS6)%*%p1.PLS6))%*(t(p1.PLS6)%*%c33))
y26<-(y16-(by8*p1.PLS6))
c41<-(c31-(bc11*p1.PLS6))
c42<-(c32-(bc22*p1.PLS6))
c43<-(c33-(bc33*p1.PLS6))

#compute p2#
s41 <-as.numeric((ginv(t(c41)%*%c41))%*(t(c41)%*%y26))

```

```

s42 <-as.numeric((ginv(t(c42)%*%c42))%*(t(c42)%*%y26))
s43 <-as.numeric((ginv(t(c43)%*%c43))%*(t(c43)%*%y26))
y6hat41 <-s41*c41
y6hat42 <-s42*c42
y6hat43 <-s43*c43
w22 <-as.numeric(t(c41)%*%c41)
w23 <-as.numeric(t(c42)%*%c42)
w24 <-as.numeric(t(c43)%*%c43)
p21_6 <-w22*y6hat41
p22_6 <-w23*y6hat42
p23_6 <-w24*y6hat43
p2.PLS6 <-p21_6+p22_6+p23_6
by9 <-
as.numeric((ginv(t(p2.PLS6)%*%p2.PLS6))%*(t(p2.PLS6)%*%y26))
bc111 <-
as.numeric((ginv(t(p2.PLS6)%*%p2.PLS6))%*(t(p2.PLS6)%*%c41))
bc222 <-
as.numeric((ginv(t(p2.PLS6)%*%p2.PLS6))%*(t(p2.PLS6)%*%c42))
bc333 <-
as.numeric((ginv(t(p2.PLS6)%*%p2.PLS6))%*(t(p2.PLS6)%*%c43))
y36 <-(y26-(by9*p2.PLS6))
c51 <-(c41-(bc111*p2.PLS6))
c52 <-(c42-(bc222*p2.PLS6))
c53 <-(c43-(bc333*p2.PLS6))

#compute p3#
s51 <-as.numeric((ginv(t(c51)%*%c51))%*(t(c51)%*%y36))
s52 <-as.numeric((ginv(t(c52)%*%c52))%*(t(c52)%*%y36))
s53 <-as.numeric((ginv(t(c53)%*%c53))%*(t(c53)%*%y36))
y6hat51 <-s51*c51

```



```

y6hat52 <-s52*c52
y6hat53 <-s53*c53
w25 <-as.numeric(t(c51)%*%c51)
w26 <-as.numeric(t(c52)%*%c52)
w27 <-as.numeric(t(c53)%*%c53)
p31_6 <-w25*y6hat51
p32_6 <-w26*y6hat52
p33_6 <-w27*y6hat53
p3.PLS6 <-p31_6+p32_6+p33_6
P.PLS6 <-data.frame(1,p1.PLS6,p2.PLS6,p3.PLS6)
P.PLS6_m <-as.matrix(P.PLS6)
cr12_PLS6 <-cor(p1.PLS6,p2.PLS6)
cr13_PLS6 <-cor(p1.PLS6,p3.PLS6)
cr23_PLS6 <-cor(p2.PLS6,p3.PLS6)

#Parameters estimation#
lmdat6 <-lm(y6~p1.PLS6+p2.PLS6+p3.PLS6)
coef6_PLS6 <-coef(lmdat6)
b200_00 <-coef6_PLS6[1]
b200_11 <-coef6_PLS6[2]
b200_22 <-coef6_PLS6[3]
b200_33 <-coef6_PLS6[4]
y6hat.PLS6 <-
b200_00+b200_11*p1.PLS6+b200_22*p2.PLS6+b200_33*p3.PLS6
E.PLS6 <-as.matrix(y6-y6hat.PLS6)
mse.PLS6 <-as.numeric((t(E.PLS6)%*%E.PLS6)/(n1+n2+n3))
inv6 <-ginv(t(P.PLS6_m)%*%P.PLS6_m)
var.PLS6 <-mse.PLS6*inv6
var200_00 <-var.PLS6[1,1]
var200_11 <-var.PLS6[2,2]

```

```

var200_22 <-var.PLS6[3,3]
var200_33 <-var.PLS6[4,4]
bias200_00 <-abs(B0-b200_00)
bias200_11 <-abs(B1-b200_11)
bias200_22 <-abs(B2-b200_22)
bias200_33 <-abs(B3-b200_33)
Mse200_00 <-var200_00+(bias200_00)^2
Mse200_11 <-var200_11+(bias200_11)^2
Mse200_22 <-var200_22+(bias200_22)^2
Mse200_33 <-var200_33+(bias200_33)^2
Mse200n<-(Mse200_00+Mse200_11+Mse200_22+Mse200_33)/4
#-----Gram-Schmidt-p=3-n=50-----#
#centering#
yS4 <-as.matrix(y4-mean(y4))
a1S <-as.matrix(a1-mean(a1))
a2S <-as.matrix(a2-mean(a2))
a3S <-as.matrix(a3-mean(a3))
#Gram-Schmidt's data transformation Orthogonal#
p1.G4 <-a1S
p2.G4 <-a2S-(sum(a2S*p1.G4)/sum(p1.G4*p1.G4))*p1.G4
p3.G4 <-a3S-(sum(a3S*p1.G4)/sum(p1.G4*p1.G4))*p1.G4-
(sum(a3S*p2.G4)/sum(p2.G4*p2.G4))*p2.G4
p.G4 <-data.frame(1,p1.G4,p2.G4,p3.G4)
P.G4_m <-as.matrix(p.G4)
cr12_G4 <-cor(p1.G4,p2.G4)
cr13_G4 <-cor(p1.G4,p3.G4)
cr23_G4 <-cor(p2.G4,p3.G4)
#Parameters estimation#
lmdatG4 <-lm(y4~p1.G4+p2.G4+p3.G4)

```

```

coef4_G4 <-coef(lmdatG4)
bG50_00 <-coef4_G4[1]
bG50_11 <-coef4_G4[2]
bG50_22 <-coef4_G4[3]
bG50_33 <-coef4_G4[4]
y4hat.G4 <-
bG50_00+bG50_11*p1.G4+bG50_22*p2.G4+bG50_33*p3.G4
E.G4 <-as.matrix(y4-y4hat.G4)
mse.G4 <-as.numeric((t(E.G4)%*%E.G4)/(n1))
var.G4 <-mse.G4*(ginv(t(P.G4_m)%*%P.G4_m))
varG50_00 <-var.G4[1,1]
varG50_11 <-var.G4[2,2]
varG50_22 <-var.G4[3,3]
varG50_33 <-var.G4[4,4]
biasG50_00 <-abs(B0-bG50_00)
biasG50_11 <-abs(B1-bG50_11)
biasG50_22 <-abs(B2-bG50_22)
biasG50_33 <-abs(B3-bG50_33)
MseG50_00 <-varG50_00+(biasG50_00)^2
MseG50_11 <-varG50_11+(biasG50_11)^2
MseG50_22 <-varG50_22+(biasG50_22)^2
MseG50_33 <-varG50_33+(biasG50_33)^2
MseG50n <-(MseG50_00+MseG50_11+MseG50_22+MseG50_33)/4
#-----Gram-Schmidt-p=3-n=100-----#
#centering#
yS5 <-as.matrix(y5-mean(y5))
b1S <-as.matrix(b1-mean(b1))
b2S <-as.matrix(b2-mean(b2))
b3S <-as.matrix(b3-mean(b3))

```

```

#Gram-Schmidt's data transformation Orthogonal#
p1.G5 <-b1S
p2.G5 <-b2S-(sum(b2S*p1.G5)/sum(p1.G5*p1.G5))*p1.G5
p3.G5 <-b3S-(sum(b3S*p1.G5)/sum(p1.G5*p1.G5))*p1.G5-
(sum(b3S*p2.G5)/sum(p2.G5*p2.G5))*p2.G5
p.G5 <-data.frame(1,p1.G5,p2.G5,p3.G5)
P.G5_m <-as.matrix(p.G5)
cr12_G5 <-cor(p1.G5,p2.G5)
cr13_G5 <-cor(p1.G5,p3.G5)
cr23_G5 <-cor(p2.G5,p3.G5)

#Parameters estimation#
lmdatG5 <-lm(y5~p1.G5+p2.G5+p3.G5)
coef5_G5 <-coef(lmdatG5)
bG100_00 <-coef5_G5[1]
bG100_11 <-coef5_G5[2]
bG100_22 <-coef5_G5[3]
bG100_33 <-coef5_G5[4]
y5hat.G5 <-
bG100_00+bG100_11*p1.G5+bG100_22*p2.G5+bG100_33*p3.G5
E.G5 <-as.matrix(y5-y5hat.G5)
mse.G5 <-as.numeric((t(E.G5)%*%E.G5)/(n1+n2))
var.G5 <-mse.G5*(ginv(t(P.G5_m)%*%P.G5_m))
varG100_00 <-var.G5[1,1]
varG100_11 <-var.G5[2,2]
varG100_22 <-var.G5[3,3]
varG100_33 <-var.G5[4,4]
biasG100_00 <-abs(B0-bG100_00)
biasG100_11 <-abs(B1-bG100_11)
biasG100_22 <-abs(B2-bG100_22)

```

```

biasG100_33 <-abs(B3-bG100_33)
MseG100_00 <-varG100_00+(biasG100_00)^2
MseG100_11 <-varG100_11+(biasG100_11)^2
MseG100_22 <-varG100_22+(biasG100_22)^2
MseG100_33 <-varG100_33+(biasG100_33)^2
MseG100n <-
(MseG100_00+MseG100_11+MseG100_22+MseG100_33)/4
#-----Gramidtm-Sch-p=3-n=200-----#
#centering#
yS6 <-as.matrix(y6-mean(y6))
c1S <-as.matrix(c1-mean(c1))
c2S <-as.matrix(c2-mean(c2))
c3S <-as.matrix(c3-mean(c3))
#Gram-Schmidt's data transformation Orthogonal#
p1.G6 <-c1S
p2.G6 <-c2S-(sum(c2S*p1.G6)/sum(p1.G6*p1.G6))*p1.G6
p3.G6 <-c3S-(sum(c3S*p1.G6)/sum(p1.G6*p1.G6))*p1.G6-
(sum(c3S*p2.G6)/sum(p2.G6*p2.G6))*p2.G6
p.G6 <-data.frame(1,p1.G6,p2.G6,p3.G6)
P.G6_m <-as.matrix(p.G6)
cr12_G6 <-cor(p1.G6,p2.G6)
cr13_G6 <-cor(p1.G6,p3.G6)
cr23_G6 <-cor(p2.G6,p3.G6)
#Parameters estimation#
lmdatG6 <-lm(y6~p1.G6+p2.G6+p3.G6)
coef6_G6 <-coef(lmdatG6)
bG200_00 <-coef6_G6[1]
bG200_11 <-coef6_G6[2]
bG200_22 <-coef6_G6[3]

```

```

bG200_33 <-coef6_G6[4]
y6hat.G6 <-
bG200_00+bG200_11*p1.G6+bG200_22*p2.G6+bG200_33*p3.G6
E.G6 <-as.matrix(y6-y6hat.G6)
mse.G6 <-as.numeric((t(E.G6)%*%E.G6)/(n1+n2+n3))
var.G6 <-mse.G6*(ginv(t(P.G6_m)%*%P.G6_m))
varG200_00 <-var.G6[1,1]
varG200_11 <-var.G6[2,2]
varG200_22 <-var.G6[3,3]
varG200_33 <-var.G6[4,4]
biasG200_00 <-abs(B0-bG200_00)
biasG200_11 <-abs(B1-bG200_11)
biasG200_22 <-abs(B2-bG200_22)
biasG200_33 <-abs(B3-bG200_33)
MseG200_00 <-varG200_00+(biasG200_00)^2
MseG200_11 <-varG200_11+(biasG200_11)^2
MseG200_22 <-varG200_22+(biasG200_22)^2
MseG200_33 <-varG200_33+(biasG200_33)^2
MseG200n <-
(MseG200_00+MseG200_11+MseG200_22+MseG200_33)/4
vifp3_total <-
c(vifp3_total,vif1_p3,vif2_p3,vif3_p3,vif4_p3,vif5_p3,vif6_p3,vif7_p3,vif8_p3,vif9_p3)
MSE3 <-c(MSE3,mse.PLS4,mse.G4,mse.PLS5,mse.G5,mse.PLS6,mse.G6)
MSE3N <-
c(MSE3N,Mse50n,MseG50n,Mse100n,MseG100n,Mse200n,MseG200n)
}
VIFp3_TOTAL <-c(VIFp3_TOTAL,vifp3_total)
MSE3_p3 <-c(MSE3_p3,MSE3)
MSE3N_p3 <-c(MSE3N_p3,MSE3N)

```

```

pie(c(i,1000-i),radius=1,clockwise=T)
}

#-----#

mvif2 <-matrix(VIFp2_TOTAL,6,10000)
mvif3 <-matrix(VIFp3_TOTAL,9,31000)
mmse2 <-matrix(MSE2_p2,6,10000)
mmse3 <-matrix(MSE3_p3,6,31000)
mmse2n <-matrix(MSE2N_p2,6,10000)
mmse3n <-matrix(MSE3N_p3,6,31000)

#-----#
ตัวอย่างการดึงค่าเพื่อคำนวณค่าเฉลี่ยของ VIF และ %การเกิดปัญหาพหุสัมพันธ์
#-----Mean-And-Persent-Multicollinearity-p=2-----#
มีทั้งหมด 10 Sequence แต่ขอยกตัวอย่างมา 1 Sequence ดังนี้
#Mean vif and %Multi sequence1#
m1<-seq(1,10000,10)
mm1<-mvif2[,m1]
vi1_n50<-mean(mm1[1,])
vi1_n100<-mean(mm1[3,])
vi1_n200<-mean(mm1[5,])
Vi1<-cbind(vi1_n50,vi1_n100,vi1_n200)
e1_1<-mm1[1,]
e1_2<-mm1[2,]
e1_3<-mm1[3,]
e1_4<-mm1[4,]
e1_5<-mm1[5,]
e1_6<-mm1[6,]
vf09_n50<-data.frame(e1_1,e1_2)
vf09_n100<-data.frame(e1_3,e1_4)
vf09_n200<-data.frame(e1_5,e1_6)

```

```
VF09_n50<-as.matrix(vf09_n50)
VF09_n100<-as.matrix(vf09_n100)
VF09_n200<-as.matrix(vf09_n200)
vf1_cal09<-apply(VF09_n50,1,function_vif)
mu1.vif1<-(mean(vf1_cal09>0))*100
vf2_cal09<-apply(VF09_n100,1,function_vif)
mu2.vif1<-(mean(vf2_cal09>0))*100
vf3_cal09<-apply(VF09_n200,1,function_vif)
mu3.vif1<-(mean(vf3_cal09>0))*100
```

จะได้ $Vi1 - Vi10$ ซึ่งเป็นค่าเฉลี่ย VIF และ $mu1.vif1 - mu1.vif10$, $mu2.vif1 - mu2.vif10$, $mu3.vif1 - mu3.vif10$ ซึ่งเป็นค่าเปอร์เซ็นต์การเกิดปัญหาพหุสัมพันธ์รุนแรงของทุก Sequence สำหรับตัวแปรอิสระเท่ากับ 2 ตัวแปร

```
#-----Mean-And-Persent-Multicollinearity-p=3-----#
```

มีทั้งหมด 31 Sequence แต่ขอยกตัวอย่างมา 1 Sequence ดังนี้

```
#Mean vif and %Multi sequence1#
```

```
v1<-seq(1,31000,31)
vv1<-mvif3[,v1]
vi11_n50<-mean(vv1[1,])
vi12_n50<-mean(vv1[2,])
vi13_n50<-mean(vv1[3,])
vi11_n100<-mean(vv1[4,])
vi12_n100<-mean(vv1[5,])
vi13_n100<-mean(vv1[6,])
vi11_n200<-mean(vv1[7,])
vi12_n200<-mean(vv1[8,])
vi13_n200<-mean(vv1[9,])
Vi1p3<-
```

```
cbind(vi11_n50,vi12_n50,vi13_n50,vi11_n100,vi12_n100,vi13_n100,vi11_n200,vi12_n200,vi13_n200)
```



```

e1_11<-vv1[1,]
e1_22<-vv1[2,]
e1_33<-vv1[3,]
e1_44<-vv1[4,]
e1_55<-vv1[5,]
e1_66<-vv1[6,]
e1_77<-vv1[7,]
e1_88<-vv1[8,]
e1_99<-vv1[9,]
vf135_n50<-data.frame(e1_11,e1_22,e1_33)
vf135_n100<-data.frame(e1_44,e1_55,e1_66)
vf135_n200<-data.frame(e1_77,e1_88,e1_99)
VF135_n50<-as.matrix(vf135_n50)
VF135_n100<-as.matrix(vf135_n100)
VF135_n200<-as.matrix(vf135_n200)
vf1_cal135<-apply(VF135_n50,1,function_vif)
mu1.vif1<-((mean(vf1_cal135>0))*100)
vf2_cal135<-apply(VF135_n100,1,function_vif)
mu2.vif1<-((mean(vf2_cal135>0))*100)
vf3_cal135<-apply(VF135_n200,1,function_vif)
mu3.vif1<-((mean(vf3_cal135>0))*100)
จะได้  $V_i p_3 - V_i p_3$  ซึ่งเป็นค่าเฉลี่ย VIF และ  $\mu_1.vif1 - \mu_1.vif31$ ,  $\mu_2.vif1 - \mu_2.vif31$ ,  $\mu_3.vif1 - \mu_3.vif31$  ซึ่งเป็นค่าเปอร์เซ็นต์การเกิดปัญหาหุ้มพันธันฐานแรงของทุก Sequence สำหรับตัวแปรอิสระเท่ากับ 3 ตัวแปร
#-----#
ตัวอย่างการดึงค่าเพื่อคำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง
#-----Mean-MSE-p=2-----#
มีทั้งหมด 10 Sequence แต่ขอยกตัวอย่างมา 1 Sequence ดังนี้
#Mean mse seq=1#

```

```

s1 <-seq(1,10000,10)
ss1 <-mmse2n[,s1]
ms1_n50 <-mean(ss1[1,])
ms1_n50g <-mean(ss1[2,])
ms1_n100 <-mean(ss1[3,])
ms1_n100g <-mean(ss1[4,])
ms1_n200 <-mean(ss1[5,])
ms1_n200g <-mean(ss1[6,])
ms1 <-cbind(ms1_n50,ms1_n100,ms1_n200)
ms1g <-cbind(ms1_n50g,ms1_n100g,ms1_n200g)
จะได้ ms1 - ms10 และ ms1g - ms10g ซึ่งเป็นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของ
ตัวประมาณ ของทุก Sequence สำหรับตัวแปรอิสระเท่ากับ 2 ตัวแปร
#-----Mean-MSE-p=3-----#
มีทั้งหมด 31 Sequence แต่ขอยกตัวอย่างมา 1 Sequence ดังนี้
#Mean mse seq=1#
q1 <-seq(1,31000,31)
qq1 <-mmse3n[,q1]
ms11_n50 <-mean(qq1[1,])
ms12_n50 <-mean(qq1[2,])
ms11_n100 <-mean(qq1[3,])
ms12_n100 <-mean(qq1[4,])
ms11_n200 <-mean(qq1[5,])
ms12_n200 <-mean(qq1[6,])
ms1p3 <-cbind(ms11_n50,ms11_n100,ms11_n200)
ms1p3g <-cbind(ms12_n50,ms12_n100,ms12_n200)
จะได้ ms1p3 - ms31p3 และ ms1p3g - ms31p3g ซึ่งเป็นค่าเฉลี่ยของความ
คลาดเคลื่อนมาตรฐานของตัวประมาณ ของทุก Sequence สำหรับจำนวนตัวแปรอิสระเท่ากับ 3
ตัวแปร

```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวเด่นนภา จุลเพชร เกิดวันที่ 9 มกราคม พ.ศ. 2529 สำเร็จการศึกษาปริญญา
เศรษฐศาสตรบัณฑิต (ศศ.บ.) สาขาเศรษฐศาสตร์ธุรกิจ สำนักวิชาการจัดการ มหาวิทยาลัยวลัย
ลักษณ์ ในปีการศึกษา 2551 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต (สถ.ม.)
สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปี
การศึกษา 2553