การแก้ไขปัญหาการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้นที่เกิดปัญหาพหุสัมพันธ์ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก

นายกิตติ วารีรัตน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
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.

CORRECTION OF ESTIMATION FOR PARAMETERS OF MULTIPLE LINEAR REGRESSION MODEL WITH MULTICOLLINEARITY USING PRINCIPAL COMPONENT ANALYSIS

Mr.Kitti Wareerat

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	การแก้ไขปัญหาการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้นที่เกิดปัญหาพหุสัมพันธ์ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก
โดย	นายกิตติ วารีรัตน์
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

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

คณะกรรมการสอบวิทยานิพนธ์

…… ประธานกรรมการ
(รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา)

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

…… กรรมการ
(อาจารย์ ดร.อนันตณัฐ กันต์ธัญญรัตน์)

…… กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.มนต์ทิพย์ เทียนสุวรรณ)

กิตติ วารีรัตน์: การแก้ไขปัญหาการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้นที่เกิดปัญหาพหุสัมพันธ์ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก. (CORRECTION OF ESTIMATION FOR PARAMETERS OF MULTIPLE LINEAR REGRESSION MODEL WITH MULTICOLLINEARITY USING PRINCIPAL COMPONENT ANALYSIS) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร.สุพล ดุรงค์วัฒนา, 116 หน้า.

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาการแก้ไขปัญหาการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้น เมื่อเกิดปัญหาพหุสัมพันธ์รุนแรงด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ได้ทำการศึกษาในขอบเขตของจำนวนตัวแปรอิสระ 2 และ 3 ตัว ที่มีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 2 และความแปรปรวนของความคลาดเคลื่อนเท่ากับ 2 โดยที่ค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 และมีขนาดตัวอย่างเท่ากับ 50, 100, 200 ซึ่งระดับความสัมพันธ์ของตัวแปรอิสระ 2 ตัว คือ 0.90, 0.91, 0.92,..., 0.99 และสำหรับตัวแปรอิสระ 3 ตัว จะพิจารณาค่าระดับความสัมพันธ์สูงสุดจากค่าระดับความสัมพันธ์ภายใต้ขอบเขตที่ทำการศึกษา

ผลการศึกษากรณีตัวแปรอิสระ 2 ตัวพบว่า ระดับความสัมพันธ์ตั้งแต่ 0.95 – 0.99 จะเกิดปัญหาพหุสัมพันธ์รุนแรง ของทุกๆขนาดตัวอย่างที่ทำการศึกษา ดังนั้น ระดับความสัมพันธ์ 0.95 คือจุดเปลี่ยนที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง และกรณีตัวแปรอิสระ 3 ตัว พบว่าระดับความสัมพันธ์สูงสุดเท่ากับ 0.93 , 0.95, 0.97, 0.99 จะเกิดปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100 และระดับความสัมพันธ์สูงสุดเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99 จะเกิดปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 200 ดังนั้น ระดับความสัมพันธ์ 0.93 คือจุดเปลี่ยนที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100 และ ระดับความสัมพันธ์ 0.95 คือจุดเปลี่ยนที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 200 ในส่วนของการแก้ไขปัญหาพหุสัมพันธ์ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก พบว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงมีค่ามากกว่าหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง แต่ค่าเฉลี่ยความเอนเอียงก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงมีค่าน้อยกว่าหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง

ภาควิชา สถิติ ลายมือชื่อนิสิต.....

สาขาวิชา สถิติ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ปีการศึกษา 2555

5381761926: MAJOR STATISTICS

KEYWORDS: MULTICOLLINEARITY/ PRINCIPAL COMPONENT ANALYSIS

KITTI WAREERAT: CORRECTION OF ESTIMATION FOR PARAMETERS OF MULTIPLE LINEAR REGRESSION MODEL WITH MULTICOLLINEARITY USING PRINCIPAL COMPONENT ANALYSIS. ADVISOR: ASSOC. PROF. SUPOL DURONGWATANA, Ph.D., 116 pp.

The objective of this research is to correct the estimation for parameters of multiple linear regression model with serious multicollinearity using Principal Component Analysis. We study Normal distribution of 2 and 3 independent variables with mean of 0, and variance of 2. The variance of error is 2, while coefficients of the regression parameters equal to 2. The samples size is 50, 100, and 200. For 2 independent variables, the correlation value is defined to be 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, and 0.99. In addition, for 3 independent variables, the maximum correlation among pairwise correlation is need as the criterion for the study.

From the study, it is found that, in 2-independent-variable case, the problems of multicollinearity occur when the correlation value is between 0.95 and 0.99 for any samples size. So the cut point of serious multicollinearity is the correlation value at 0.95. In, 3 –independent-variable case, the problems occur when the maximum correlation value is 0.93, 0.95, 0.97, 0.99 for samples size of 50 and 100, and the maximum correlation value at 0.95, 0.97, 0.99 for samples size of 200. Then the cut point of serious multicollinearity for samples size of 50, 100 is the correlation value at 0.93 and for sample size of 200 is the correlation value at 0.95. The part of correction the estimation for parameters with serious multicollinearity using Principal Component Analysis.Thus, Principal Component Analysis method can be used to solve the problem of multicollinearity. Before the problem solving, the Average Standard Error (ASE) is greater than the after one. But the average bias before solve the problem of multicollineatity is less than the after one.

Department: Statistics Student's Signature……

Field of Study: Statistics Advisor's Signature……

Academic Year: 2012

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดีเนื่องจากผู้วิจัยได้รับความกรุณาและความช่วยเหลือจาก รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา อาจารย์ที่ปรึกษา ได้กรุณาสละเวลาให้ความช่วยเหลือ คำปรึกษา ข้อชี้แนะ ข้อคิดเห็น ตลอดจนติดตามความคืบหน้าในการทำวิทยานิพนธ์ทุกขั้นตอน รวมทั้งตรวจสอบและแก้ไขข้อบกพร่องต่างๆ ที่เป็นประโยชน์ในการทำวิทยานิพนธ์ ผู้เขียนขอขอบพระคุณและสำนึกในบุญคุณยิ่ง

ขอขอบพระคุณท่านคณะกรรมการสอบวิทยานิพนธ์ ซึ่งประกอบไปด้วย รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา รองศาสตราจารย์ ดร.มนต์ทิพย์ เทียนสุวรรณ และอาจารย์ ดร.อนันตณัฐ กันต์ธัญญรัตน์ ที่ช่วยให้คำแนะนำ และแนวคิดที่เป็นประโยชน์ ตลอดจนช่วยตรวจสอบ แก้ไขให้วิทยานิพนธ์นี้สมบูรณ์มากยิ่งขึ้น

ขอขอบพระคุณ คณาจารย์ทุกท่านของภาคสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ได้ถ่ายทอดความรู้ทางสถิติให้กับผู้วิจัย

ขอขอบคุณเจ้าหน้าที่ของภาคสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่อำนวยความสะดวกในการทำวิจัยเป็นอย่างดี ขอขอบคุณเพื่อนๆ นิสิตภาคสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อันเป็นที่รักของผู้วิจัยทุกคน ที่ให้กำลังใจและความช่วยเหลือมาโดยตลอด

สุดท้ายนี้ ขอขอบพระคุณ และมอบความสำเร็จทั้งหมดจากการทำวิทยานิพนธ์เล่มนี้แด่คุณพ่อ คุณแม่ และญาติพี่น้องทุกคน ที่ให้ความห่วงใย ช่วยเป็นกำลังใจ ส่งเสริม สนับสนุนทุกๆ ด้าน จนทำให้การศึกษาครั้งนี้ประสบความสำเร็จไปได้ด้วยดี

สารบัญ

หน้า

บทคัดย่อภาษาไทย…… ง

บทคัดย่อภาษาอังกฤษ…….. จ

กิตติกรรมประกาศ…… ฉ

สารบัญ…….. ช

สารบัญตาราง…… ฌ

สารบัญรูปภาพ…….. ญ

บทที่ 1 บทนำ……. 1

1.1 ความเป็นมาและความสำคัญของปัญหา…….. 1

1.2 วัตถุประสงค์ของการศึกษา…….. 3

1.3 ขอบเขตของการศึกษา……... 3

1.4 คำจำกัดความที่ใช้ในการศึกษา……. 4

1.5 วิธีดำเนินการศึกษา……..... 5

1.6 ประโยชน์ที่คาดว่าจะได้รับ……. 8

บทที่ 2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง……..... 9

2.1 การประมาณค่าสัมประสิทธิ์ความถดถอยพหุเชิงเส้นด้วยวิธีกำลังสองน้อยสุด..... 9

2.2 ความสัมพันธ์ร่วมระหว่างตัวแปรอิสระ……..…… 10

2.3 ปัญหาพหุสัมพันธ์ (Multicollinearity) ของตัวแปรอิสระ..... 11

2.4 การแก้ไขปัญหาพหุสัมพันธ์ของตัวแปรอิสระที่สัมพันธ์กัน……..... 11

2.5 การวิเคราะห์ตัวประกอบหลัก……..... 12

2.6 การแจกแจงของตัวแปรอิสระ..... 16

2.7 การจำลองเวกเตอร์สุ่มแบบปกติ..... 17

2.8 ความเอนเอียง..... 18

2.9 สถิติที่ใช้ในการตัดสินใจ..... 18

2.9.1 มาตรวัด Variance Inflation Factor ……..... 18

หน้า

บทที่ 3 วิธีดำเนินการศึกษา..... 19
3.1 ขอบเขตการศึกษา..... 19
3.2 วิธีดำเนินการศึกษา..... 20

บทที่ 4 ผลการวิเคราะห์ข้อมูล..... 25
4.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัว..... 25
4.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัว..... 39

บทที่ 5 สรุปผลการศึกษาและข้อเสนอแนะ..... 68
5.1 สรุปผลการศึกษา..... 69
5.1.1 กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัว..... 70
5.1.2 กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัว..... 71
5.2 ข้อเสนอแนะ..... 73
5.3 แนวทางในการศึกษาต่อ..... 73

รายการอ้างอิง..... 74
ภาคผนวก..... 75
ประวัติผู้เขียนวิทยานิพนธ์..... 116

สารบัญตาราง

ตารางที่ หน้า

2.2.1 ตารางแสดงระดับความสัมพันธ์…….. 11
กรณีจำนวนตัวแปรอิสระ 2 ตัว

4.1.1 แสดงค่าเฉลี่ย VIF จำแนกตามขนาดตัวอย่าง และระดับวามสัมพันธ์…….... 25

4.1.2 แสดงค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จำแนกตามขนาด
ตัวอย่างและระดับความสัมพันธ์……..... 29

4.1.3 แสดงค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย
จำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์……..... 31

4.1.4 แสดงค่าเฉลี่ยของค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย
จำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์……..... 35
กรณีจำนวนตัวแปรอิสระ 3 ตัว

4.1.5 แสดงค่าเฉลี่ย VIF 1, VIF2, VIF3 จำแนกตามขนาดตัวอย่าง
และระดับความสัมพันธ์……. 40

4.1.6 แสดงค่าเฉลี่ย VIF 1, VIF2, VIF3 จำแนกตามขนาดตัวอย่าง
และระดับความสัมพันธ์สูงสุด……..... 41

4.1.7 แสดงค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จำแนกตามขนาด
ตัวอย่างและระดับความสัมพันธ์……..... 55

4.1.8 แสดงค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์ จำแนกตามขนาดตัวอย่าง
และระดับความสัมพันธ์สูงสุด……. 55

4.1.9 แสดงค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์
ความถดถอยจำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์…….. 58

4.1.10 แสดงค่าเฉลี่ยของค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์
ความถดถอยจำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์สูงสุด……..... 59

4.1.11 แสดงค่าเฉลี่ยของค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย
จำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์……... 63

4.1.12 แสดงค่าเฉลี่ยของค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย
จำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์สูงสุด…….. 64

สารบัญภาพ

ภาพที่ หน้า

กรณีจำนวนตัวแปรอิสระ 2 ตัว

4.1.1 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF ในแต่ละระดับความสัมพันธ์ เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200..... 26

4.1.2 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย (VIF) ของแต่ละขนาดตัวอย่าง เมื่อระดับความสัมพันธ์มีค่าตั้งแต่ 0.90 – 0.99..... 27

4.1.3 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF) ในแต่ละระดับความสัมพันธ์ เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200..... 27

4.1.4 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF) ในแต่ละขนาดตัวอย่าง เมื่อระดับความสัมพันธ์มีค่าตั้งแต่ 0.90 – 0.99..... 28

4.1.5 กราฟแสดงการเปรียบเทียบค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์ เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200..... 29

4.1.6 กราฟแสดงการเปรียบเทียบค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ในแต่ขนาดตัวอย่าง เมื่อระดับความสัมพันธ์มีค่าตั้งแต่ 0.90 – 0.99..... 30

4.1.7 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย ก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200..... 32

4.1.8 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย หลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่างเท่ากับ 50,100,200..... 32

4.1.9 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่างเท่ากับ 50..... 33

4.1.10 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่าง เท่ากับ 100..... 33

ภาพที่ หน้า

4.1.11 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของ
สัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง
เมื่อขนาดตัวอย่าง เท่ากับ 200..... 34
4.1.12 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความ
ถดถอย ก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่าง
เท่ากับ 50, 100, 200..... 36
4.1.13 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความ
ถดถอย หลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่าง
เท่ากับ 50, 100, 200..... 36
4.1.14 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความ
ถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่าง
เท่ากับ 50..... 37
4.1.15 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความ
ถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่าง
เท่ากับ 100..... 37
4.1.16 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความ
ถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่าง
เท่ากับ 200..... 38
กรณีจำนวนตัวแปรอิสระ 3 ตัว
4.1.17 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF 1 ในแต่ละระดับความสัมพันธ์
สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200..... 42
4.1.18 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF 2 ในแต่ละระดับความสัมพันธ์
สูงสุดเมื่อขนาดตัวอย่าง เท่ากับ 50, 100, 200..... 42
4.1.19 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF 3 ในแต่ละระดับความสัมพันธ์
สูงสุดเมื่อขนาดตัวอย่าง เท่ากับ 50, 100, 200..... 43
4.1.20 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1) ในแต่ละระดับ
ความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่าง เท่ากับ 50, 100, 200..... 43

ภาพที่ หน้า

4.1.21 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF2) ในแต่ละระดับ
ความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่าง เท่ากับ 50, 100, 200..... 44
4.1.22 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF3) ในแต่ละระดับ
ความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่าง เท่ากับ 50, 100, 200..... 44
4.1.23 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย (VIF1) ในแต่ละขนาดตัวอย่างเมื่อระดับ
ความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99..... 45
4.1.24 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย (VIF2) ในแต่ละขนาดตัวอย่างเมื่อระดับ
ความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99..... 45
4.1.25 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย (VIF3)ในแต่ละขนาดตัวอย่างเมื่อระดับ
ความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99..... 46
4.1.26 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1) ในแต่ละขนาดตัวอย่างเมื่อ
ระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99..... 46
4.1.27 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF2) ในแต่ละขนาดตัวอย่างเมื่อ
ระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99..... 47
4.1.28 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF3) ในแต่ละขนาดตัวอย่างเมื่อ
ระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99..... 47
4.1.29 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละระดับ
ความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50..... 48
4.1.30 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละระดับ
ความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100..... 48
4.1.31 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละระดับ
ความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200..... 49
4.1.32 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3)
ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50..... 49
4.1.33 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3)
ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100..... 50

ภาพที่ หน้า

4.1.34 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200..... 50
4.1.35 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50..... 51
4.1.36 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละค่าระดับความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 100..... 51
4.1.37 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200..... 52
4.1.38 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50..... 52
4.1.39 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100..... 53
4.1.40 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200..... 53
4.1.41 กราฟแสดงการเปรียบเทียบค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200.... 56
4.1.42 กราฟแสดงการเปรียบเทียบค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ในแต่ขนาดตัวอย่าง เมื่อระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91-0.99..... 56
4.1.43 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 50,100, 200..... 59
4.1.44 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200..... 60
4.1.45 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 50..... 60

ภาพที่ หน้า

4.1.46 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของ
สัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง
ในแต่ละระดับความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 100…….. 61
4.1.47 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของ
สัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง
ในแต่ละระดับความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 200…….. 61
4.1.48 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยของค่าเฉลี่ยความเอนเอียงของ
สัมประสิทธิ์ความถดถอยก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละ
ระดับความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่าง เท่ากับ 50, 100, 200……... 64
4.1.49 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์
ความถดถอยหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับ
ความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200…… 65
4.1.50 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์
ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับ
ความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 50…… 65
4.1.51 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์
ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับ
ความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 100…… 66
4.1.52 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์
ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับ
ความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างเท่ากับ 200…… 66

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการวิเคราะห์ทางสถิติที่เกี่ยวข้องกับการสร้างตัวแบบทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระจำนวน 1 ตัวหรือมากกว่า 1 ตัวก็ได้ (วิชิต หล่อจีระชุณห์กุลและจิราวัลย์ จิตรถเวช, 2548: 203) ซึ่งการวิเคราะห์ความถดถอยพหุเชิงเส้น เป็นการวิเคราะห์ความถดถอยอีกรูปแบบหนึ่งที่ได้รับความนิยม โดยมีการนำไปใช้ประโยชน์ในหลายสาขา เช่น ทางด้านวิทยาศาสตร์ สังคมศาสตร์ และประกันภัย เป็นต้น ซึ่งจุดประสงค์หลักของการวิเคราะห์ความถดถอย คือเพื่อศึกษาความสัมพันธ์ของตัวแปรที่ต้องการศึกษาและสร้างสมการถดถอยที่เหมาะสม โดยมีข้อตกลงเบื้องต้นในการวิเคราะห์ความถดถอยพหุเชิงเส้น คือ ความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติที่ค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนคงที่ โดยที่ความคลาดเคลื่อนสุ่มมีความเป็นอิสระกัน และตัวแปรอิสระต้องมีความเป็นอิสระกัน

สำหรับการวิเคราะห์ความถดถอยจะมีวิธีการประมาณค่าพารามิเตอร์ของความถดถอย 2 วิธี คือ วิธีกำลังสองน้อยสุด (Ordinary Least Squares Method: OLS) และวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Estimation Method: MLE) แต่ในงานวิจัยนี้จะใช้วิธีกำลังสองน้อยสุด ในการประมาณค่าพารามิเตอร์ โดยคุณสมบัติของตัวประมาณพารามิเตอร์ คือ มีการแจกแจงแบบปกติเชิงพหุที่มีค่าเฉลี่ยเท่ากับ $\widetilde{\beta}$ และเมทริกซ์ความแปรปรวนร่วมเท่ากับ $\sigma^2(X^TX)^{-1}$ โดยที่ตัวประมาณพารามิเตอร์ที่ได้จะเป็นตัวประมาณที่ไม่เอนเอียง

แต่เนื่องจากข้อมูลที่พบในปัจจุบันมีความสัมพันธ์กัน จึงเป็นปัญหาในส่วนของการวิเคราะห์และการแปลผล เมื่อนำข้อมูลมาทำการวิเคราะห์ความถดถอยพหุเชิงเส้นแล้วพบว่าค่าความคลาดเคลื่อนมาตรฐานของตัวประมาณมีค่าสูงเกินจริง หรือค่าสัมประสิทธิ์ความถดถอยมีเครื่องหมายสลับกัน ถ้าตัวแปรอิสระมีความสัมพันธ์กันสูงกับตัวแปรอิสระตัวอื่นๆ ซึ่งอาจส่งผลต่อการทดสอบสมมติฐานของค่าสัมประสิทธิ์ความถดถอยเชิงส่วนที่ไม่มีนัยสำคัญ และทำให้การสรุปผลผิดพลาดได้

ดังนั้นเราจึงต้องหาวิธีในการแก้ไขปัญหาของตัวแปรอิสระที่มีความสัมพันธ์กัน โดยการเปลี่ยนจากตัวแปรอิสระที่สัมพันธ์กันเป็นตัวแปรอิสระที่ไม่สัมพันธ์กัน ซึ่งผู้วิจัยได้ใช้วิธีการวิเคราะห์ตัวประกอบหลักในการแก้ไขปัญหาของตัวแปรอิสระที่สัมพันธ์กัน โดย (Gunst and Mason, 1977: 616-628) มีแนวคิดว่า การประมาณค่าสัมประสิทธิ์ความถดถอยพหุเชิงเส้นด้วยวิธีกำลังสองน้อยสุดจะเป็นตัวประมาณที่ดีเมื่อความแปรปรวนของความคลาดเคลื่อนมีค่าน้อย ถึงแม้ว่าตัวแปรอิสระจะเกิดปัญหาพหุสัมพันธ์ จึงได้เสนอวิธีการสร้างตัวแปรอิสระตัวใหม่ให้อยู่ในรูปของผลรวมเชิงเส้นของตัวแปรอิสระตัวเดิม และเรียกวิธีนี้ว่า การวิเคราะห์ตัวประกอบหลัก มาใช้ในการแก้ไขปัญหากรณีที่ตัวแปรอิสระเกิดปัญหาพหุสัมพันธ์ โดยใช้วิธีกำลังสองน้อยสุดในการประมาณค่าสัมประสิทธิ์ความถดถอย ซึ่งวิธีการวิเคราะห์ตัวประกอบหลักจะช่วยขจัดปัญหาของตัวแปรอิสระที่เกิดปัญหาพหุสัมพันธ์กัน ทำให้การประมาณค่าสัมประสิทธิ์ความถดถอยมีความแม่นยำมากยิ่งขึ้น ถึงแม้ว่าตัวประมาณดังกล่าวจะเป็นตัวประมาณที่เอนเอียงก็ตาม

ในงานวิจัยของ (Lafi and Kaneene, 1992: 261-275) ได้อธิบายถึงปัญหาในการวิเคราะห์ข้อมูลเมื่อเกิดปัญหาพหุสัมพันธ์ของตัวแปรอิสระ สาเหตุและผลกระทบของการเกิดปัญหาพหุสัมพันธ์ โดยมีกระบวนการตรวจสอบเมื่อเกิดปัญหาพหุสัมพันธ์ของตัวแปรอิสระ และการลดผลกระทบของการเกิดปัญหาพหุสัมพันธ์ โดยใช้วิธีการวิเคราะห์ตัวประกอบหลัก ในการแก้ไขปัญหาพหุสัมพันธ์ และยกตัวอย่างประกอบ

ในงานวิจัยของ (กรรณิกาณ์ หิรัญกสิ, 2540) ได้ศึกษาการเปรียบเทียบการประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุดแบ่งส่วน วิธีการถดถอยองค์ประกอบ และวิธีกำลังสองน้อยที่สุดในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยใช้เกณฑ์ในการเปรียบเทียบทั้ง 3 วิธีด้วย ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการพยากรณ์ โดยมีปัจจัยที่เกี่ยวข้อง คือ จำนวนตัวแปรอิสระที่มีความสัมพันธ์กัน ระดับความสัมพันธ์ของตัวแปรอิสระ และความแปรปรวนของความคลาดเคลื่อน

ในงานวิจัยของ (เปรมวดี ชูไสว, 2548) ได้ศึกษาการเปรียบเทียบวิธีการแก้ไขพหุสัมพันธ์ในการถดถอยเชิงเส้นพหุคุณ โดยทำการเปรียบเทียบ 3 วิธี คือ วิธีกำลังสองน้อยสุด วิธีริดจ์-รีเกรสชั่นและวิธีการถดถอยองค์ประกอบหลัก แล้วเปรียบเทียบด้วยค่าส่วนเบี่ยงเบนของค่าคลาดเคลื่อน โดยมีปัจจัยที่มีผลต่อความคลาดเคลื่อน คือ ขนาดตัวอย่าง ระดับความสัมพันธ์ของตัวแปรอิสระ และ ส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อน

ในการวิจัยครั้งนี้ผู้วิจัยต้องการศึกษาการแก้ไขปัญหาการประมาณค่าพารามิเตอร์ความถดถอยพหุเชิงเส้น เมื่อเกิดปัญหาพหุสัมพันธ์รุนแรงด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ซึ่งมี

ปัจจัยของจำนวนตัวแปรอิสระ ขนาดตัวอย่าง ระดับความสัมพันธ์ของตัวแปรอิสระเข้ามาเกี่ยวข้องในการพิจารณาค่าความคลาดเคลื่อนมาตรฐาน และค่าความเอนเอียงของสัมประสิทธิ์ความถดถอย

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาการแก้ไขปัญหาการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้น เมื่อเกิดปัญหาพหุสัมพันธ์รุนแรงด้วยวิธีการวิเคราะห์ตัวประกอบหลัก

1.3 ขอบเขตของการศึกษา

การวิจัยในครั้งนี้กระทำภายใต้ขอบเขตดังนี้

1.กำหนดตัวแปรอิสระ(x) คือ 2, 3 ตัวแปร เป็นตัวแปรเชิงปริมาณ ที่มีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 2 และตัวแปรอิสระมีความสัมพันธ์ดังนี้

1.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว (p=2)

ρ_{12} คือค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และ ตัวที่ 2

โดยทำการศึกษาระดับความสัมพันธ์ดังนี้

1.1.1 ระดับความสัมพันธ์ 0.90

1.1.2 ระดับความสัมพันธ์ 0.91

1.1.3 ระดับความสัมพันธ์ 0.92

1.1.4 ระดับความสัมพันธ์ 0.93

1.1.5 ระดับความสัมพันธ์ 0.94

1.1.6 ระดับความสัมพันธ์ 0.95

1.1.7 ระดับความสัมพันธ์ 0.96

1.1.8 ระดับความสัมพันธ์ 0.97

1.1.9 ระดับความสัมพันธ์ 0.98

1.1.10 ระดับความสัมพันธ์ 0.99

1.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว (p=3)

ρ_{12} คือค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และ ตัวที่ 2

ρ_{13} คือค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และ ตัวที่ 3

ρ_{23} คือค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 2 และ ตัวที่ 3

โดยทำการศึกษาระดับความสัมพันธ์ดังนี้

1.2.1 ระดับความสัมพันธ์ (0.91, 0.93, 0.95)

1.2.2 ระดับความสัมพันธ์ (0.91, 0.93, 0.97)

1.2.3 ระดับความสัมพันธ์ (0.91, 0.93, 0.99)

1.2.4 ระดับความสัมพันธ์ (0.91, 0.95, 0.97)

1.2.5 ระดับความสัมพันธ์ (0.91, 0.95, 0.99)

1.2.6 ระดับความสัมพันธ์ (0.93, 0.95, 0.97)

1.2.7 ระดับความสัมพันธ์ (0.93, 0.95, 0.99)

1.2.8 ระดับความสัมพันธ์ (0.93, 0.97, 0.99)

1.2.9 ระดับความสัมพันธ์ (0.95, 0.97, 0.99)

1.2.10 ระดับความสัมพันธ์ (0.91, 0.91, 0.91)

1.2.11 ระดับความสัมพันธ์ (0.93, 0.93, 0.93)

1.2.12 ระดับความสัมพันธ์ (0.95, 0.95, 0.95)

1.2.13 ระดับความสัมพันธ์ (0.97, 0.97, 0.97)

1.2.14 ระดับความสัมพันธ์ (0.99, 0.99, 0.99)

2. ศึกษาภายใต้ขนาดตัวอย่างดังนี้

2.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว จะศึกษาขนาดตัวอย่าง 50, 100, 200

2.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว จะศึกษาขนาดตัวอย่าง 50, 100, 200

3. กำหนดค่าสัมประสิทธิ์ความถดถอย $(\beta_0, \beta_1, \beta_2, ..., \beta_p)$ ได้ดังนี้

3.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว มีค่าสัมประสิทธิ์ความถดถอยเริ่มต้น คือ

$\beta_0 = \beta_1 = \beta_2 = 2$

3.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว มีค่าสัมประสิทธิ์ความถดถอยเริ่มต้น คือ

$\beta_0 = \beta_1 = \beta_2 = \beta_3 = 2$

4. จำลองข้อมูลความคลาดเคลื่อนที่มีการแจกแจงปกติ โดยมีค่าเฉลี่ยเท่ากับ 0 ค่าความแปรปรวนเท่ากับ 2

5. สร้างตัวแปรตาม y จากตัวแบบความถดถอยพหุเชิงเส้น $\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon}$

1.4 คำจำกัดความที่ใช้ในการศึกษา

1. การวิเคราะห์ตัวประกอบหลัก (Principal Component Analysis) คือ เทคนิคในการสร้างตัวประกอบหลัก (ตัวแปรตัวใหม่) ให้อยู่ในรูปของผลบวกเชิงเส้นของตัวแปรเดิม โดยที่ตัวประกอบหลักเป็นอิสระกัน

2. ความสัมพันธ์ร่วมเชิงพหุ (Multicollinearity) คือการที่ตัวแปรอิสระตั้งแต่2ตัวขึ้นไปเกิดความสัมพันธ์กัน

1.5 วิธีดำเนินการศึกษา

1. กำหนดค่าเริ่มต้นในการจำลองข้อมูล ดังนี้

1.1 กำหนดตัวแปรอิสระ p ตัว

1.2 กำหนดขนาดตัวอย่าง n ตัว

1.3 กำหนดค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $(\widetilde{\beta})$

1.4 กำหนดระดับความสัมพันธ์ระหว่างตัวแปรอิสระ (ρ)

2. จำลองข้อมูลตัวแปรอิสระที่มีการแจกแจงปกติหลายตัวแปร โดยมีค่าเฉลี่ย ค่าความแปรปรวนและระดับความสัมพันธ์ของตัวแปรอิสระที่กำหนด

3. จำลองข้อมูลความคลาดเคลื่อนที่มีการแจกแจงปกติ โดยมีค่าเฉลี่ยและค่าความแปรปรวนตามที่กำหนด

4. สร้างตัวแปรตาม y จากตัวแบบความถดถอยพหุเชิงเส้น $\widetilde{y} = X\widetilde{\beta} + \widetilde{\varepsilon}$

5. เกณฑ์ในการตัดสินใจว่าเกิดปัญหา Collinearity ที่รุนแรงหรือไม่ โดยใช้ค่า VIF

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

โดยที่ R_j^2 คือค่าสัมประสิทธิ์การตัดสินใจเชิงพหุของตัวแบบความถดถอย ตัวแปรอิสระตัวที่ j กับตัวแปรอิสระตัวอื่น

ถ้าค่า $VIF_j > 10; j = 1,2,...,p$ จะได้ว่าตัวแปรอิสระตัวที่ j มีความสัมพันธ์ร่วมกันกับตัวแปรอิสระตัวที่เหลือ

ถ้าค่า $VIF_j \le 10; j = 1,2,...,p$ จะได้ว่าตัวแปรอิสระตัวที่ j ไม่มีความสัมพันธ์ร่วมกันกับตัวแปรอิสระตัวที่เหลือ (สุพล ดุรงค์วัฒนา, 2549: 334)

6. คำนวณหาค่าสัมประสิทธิ์ความถดถอยของตัวแปรอิสระก่อนทำการแก้ไขปัญหาพหุสัมพันธ์โดยวิธีกำลังสองน้อยสุด หาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานและค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

6.1 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด

$$\widetilde{b} = (X^T X)^{-1} X^T \widetilde{y}$$

6.2 การหาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย

จากเมทริกซ์ความแปรปรวนร่วม $\hat{C}ov(\widetilde{b}) = \hat{\sigma}^2 (X^T X)^{-1}$

$$\hat{\sigma}^2 = MSE = \frac{\sum_{i=1}^{n} (y - \hat{y})^2}{n - p - 1}$$

สามารถหาความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยได้ดังนี้

$$\hat{s}e(b_j) = \sqrt{\hat{\mathrm{var}}(b_j)}$$

ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจำนวน N รอบ

$$SE(b_j) = \frac{\sum_{i=1}^{N} \hat{s}e(b_j)}{N}$$

ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย

$$ASE = \frac{\sum_{j=0}^{p} SE(b_j)}{p+1}$$

6.3 การหาค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

ค่าความเอนเอียงของสัมประสิทธิ์ความถดถอยคือ

$$bias(b_j) = \left| \frac{\sum_{i=1}^{N} (b_{ji})}{N} - \beta_j \right|$$

ค่าเฉลี่ยความเอนเอียงจำนวน N รอบ

$$Bias(b_j) = \frac{\sum_{i=1}^{N} bias(b_j)}{N}$$

ค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

$$ABias = \frac{\sum_{j=0}^{p} Bias(b_j)}{p+1}$$

โดยที่ N คือ จำนวนรอบในการทดลองซ้ำ

b_j คือ สัมประสิทธิ์ความถอยถอย j=0,1,2,..,p

7. การแก้ไขปัญหาพหุสัมพันธ์ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก โดยนำตัวแปรอิสระ คือ เมทริกซ์ X มาสร้างเมทริกซ์สหสัมพันธ์ (R) แล้วหาค่า Eigenvalue จากสมการ $|R - I\lambda| = 0$ และนำค่า λ แทนลงในสมการข้างต้น เพื่อหาค่า Eigenvector ซึ่ง Eigenvector แต่ละชุดจะเป็นสัมประสิทธิ์ของตัวแปรอิสระเดิม

8. นำค่า Eigenvector คูณกับตัวแปรอิสระเดิมที่ทำให้อยู่ในรูปมาตรฐาน จะได้ตัวประกอบหลัก (PC_j) ของแต่ละชุด ซึ่ง (PC_j) ก็คือตัวแปรอิสระตัวใหม่

9. คำนวณหาค่าสัมประสิทธิ์ความถดถอยของตัวแปรอิสระตัวใหม่โดยวิธีกำลังสองน้อยสุด หาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานและค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

9.1 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด

$$\tilde{\gamma} = (Z^T Z)^{-1} Z^T \tilde{y}$$

9.2 การหาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย

จากเมทริกซ์ความแปรปรวนร่วม $\hat{\text{var}}(\tilde{\gamma}) = \hat{\sigma}^2 (Z^T Z)^{-1}$

สามารถหาความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยได้ดังนี้

$$\hat{se}(\gamma_j) = \sqrt{\hat{\text{var}}(\gamma_j)}$$

ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจำนวน N รอบ

$$SE(\gamma_j) = \frac{\sum_{i=1}^{N} \hat{se}(\gamma_j)}{N}$$

ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย

$$ASE = \frac{\sum_{j=0}^{p} SE(\gamma_j)}{p+1}$$

9.3 การหาค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

ค่าความเอนเอียงของสัมประสิทธิ์ความถดถอยคือ

$$bias(\gamma_j) = \left| \frac{\sum_{i=1}^{N} (\gamma_{ji})}{N} - \beta_j \right|$$

ค่าเฉลี่ยความเอนเอียงจำนวน N รอบ

$$Bias(\gamma_j) = \frac{\sum_{i=1}^{N} bias(\gamma_j)}{N}$$

ค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

$$ABias = \frac{\sum_{j=0}^{p} Bias(\gamma_j)}{p+1}$$

โดยที่ N คือ จำนวนรอบในการทดลองซ้ำ

b_j คือ สัมประสิทธิ์ความถอยถอย j=0,1,2,..,p

10. พิจารณาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานและค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนและหลังแก้ไขปัญหาพหุสัมพันธ์รุนแรง

11. สรุปผลการศึกษา

1.6 ประโยชน์ที่คาดว่าจะได้รับ

ได้ค่าประมาณพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้นที่เหมาะสม ของแต่ละกรณีที่ทำการศึกษา เมื่อเกิดปัญหาพหุสัมพันธ์รุนแรง (Multicollinearity) ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก (Principal Component Analysis)

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

การศึกษาครั้งนี้ได้ศึกษาเกี่ยวกับการแก้ไขปัญหาพหุสัมพันธ์ของตัวแปรอิสระ เมื่อตัวแปรอิสระมีความสัมพันธ์กันสูงมาก จึงทำให้เกิดปัญหาพหุสัมพันธ์ที่รุนแรง ซึ่งอาจจะขัดแย้งกับข้อกำหนดเบื้องต้นของการวิเคราะห์ความถดถอยที่ว่า ตัวแปรอิสระต้องไม่มีความสัมพันธ์กัน ดังนั้นผู้วิจัยจึงได้ทำการศึกษาการแก้ไขปัญหาพหุสัมพันธ์ ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ในการสร้างตัวแปรอิสระตัวใหม่ที่ไม่มีความสัมพันธ์กัน

2.1 การประมาณค่าสัมประสิทธิ์ความถดถอยพหุเชิงเส้นด้วยวิธีกำลังสองน้อยสุด

จากรูปแบบทั่วไปของความสัมพันธ์ของการถดถอยพหุเชิงเส้นคือ

$$\widetilde{y} = X\widetilde{\beta} + \widetilde{\varepsilon}$$

สามารถเขียนให้อยู่ในรูปเมทริกซ์ได้ดังนี้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

โดยที่ $\widetilde{Y}$ เป็นเวกเตอร์ของตัวแปรตามที่มีขนาด $n \times 1$

X เป็นเมทริกซ์ของตัวแปรอิสระที่มีขนาด $n \times (p+1)$

$\widetilde{\beta}$ เป็นเวกเตอร์ของสัมประสิทธิ์ของความถดถอยพหุเชิงเส้น มีขนาด $(p+1) \times 1$

$\widetilde{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนที่มีขนาด $n \times 1$ ซึ่ง $\widetilde{\varepsilon} \sim \mathrm{NID}(\widetilde{0}, \sigma^2 I_n)$

เมื่อ n คือขนาดตัวอย่าง , p คือจำนวนตัวแปรอิสระ

โดยมีข้อตกลงเบื้องต้นในการวิเคราะห์ความถดถอยพหุเชิงเส้น คือ ความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติที่ค่าเฉลี่ยเท่ากับศูนย์ และความแปรปรวนคงที่ โดยที่ความคลาดเคลื่อนสุ่มมีความเป็นอิสระกัน และตัวแปรอิสระต้องมีความเป็นอิสระกัน

การประมาณค่าพารามิเตอร์ $\widetilde{\beta}$ ด้วยวิธีกำลังสองน้อยสุด ที่ทำให้ค่า $\widetilde{\varepsilon}^T\widetilde{\varepsilon}$ มีค่าน้อยที่สุด โดยใช้เทคนิคทางคณิตศาสตร์คือ การ Differential $\widetilde{\varepsilon}^T\widetilde{\varepsilon}$ เทียบกับ $\widetilde{\beta}$ เท่ากับ 0 ได้ดังนี้

จากสมการ $\widetilde{y} = X\widetilde{\beta} + \widetilde{\varepsilon}$

สามารถเขียนให้อยู่ในรูปของ $\tilde{\varepsilon}$ จะได้ว่า $\tilde{\varepsilon} = \tilde{y} - X\tilde{\beta}$

เนื่องจากเราต้องการ $\tilde{\varepsilon}^T\tilde{\varepsilon}$ มีค่าน้อยที่สุด จะได้ว่า

$$\tilde{\varepsilon}^T\tilde{\varepsilon} = (\tilde{y} - X\tilde{\beta})^T(\tilde{y} - X\tilde{\beta})$$
$$= \tilde{y}^T\tilde{y} - 2\tilde{\beta}^T X^T\tilde{y} + 2\tilde{\beta}^T X^T X$$

การ Differential $\tilde{\varepsilon}^T\tilde{\varepsilon}$ เทียบกับ $\tilde{\beta}$ เท่ากับ 0

$$\frac{\partial(\tilde{\varepsilon}^T\tilde{\varepsilon})}{\partial\tilde{\beta}} = -2X^T\tilde{y} + 2X^T X\tilde{\beta}$$
$$-2X^T\tilde{y} + 2X^T X\tilde{\beta} = \tilde{0}$$
$$(X^T X)\tilde{\beta} = X^T\tilde{y}$$
$$\tilde{\beta} = (X^T X)^{-1}X^T\tilde{y}$$

จะได้ว่า ค่าประมาณพารามิเตอร์ความถดถอยพหุเชิงเส้น คือ $\tilde{b} = (X^T X)^{-1}X^T\tilde{y}$ โดยที่ตัวประมาณ $\tilde{\beta}$ เป็นตัวประมาณไม่เอนเอียงและมีค่าเฉลี่ยความคลาดเคลื่อนต่ำสุด

2.2 ความสัมพันธ์ร่วมระหว่างตัวแปรอิสระ

สามารถแบ่งความสัมพันธ์ร่วมของตัวแปรอิสระโดยพิจารณาจากหลักเกณฑ์ 2 ข้อ ดังนี้

2.2.1 พิจารณาจากจำนวนตัวแปรอิสระที่สัมพันธ์กัน โดยแบ่งออกเป็น 2 ส่วน คือ จำนวนตัวแปรอิสระ 2 ตัว และ จำนวนตัวแปรอิสระมากกว่า 2 ตัว

2.2.2 พิจารณาจากระดับความสัมพันธ์ของตัวแปรอิสระ โดยแบ่งออกเป็น 2 ส่วน คือ ตัวแปรอิสระที่มีความสัมพันธ์กันแบบสมบูรณ์ และ ตัวแปรอิสระที่มีความสัมพันธ์กันแบบไม่สมบูรณ์ ซึ่งสามารถแบบออกเป็น 4 กรณี คือ

กรณีที่ 1 ตัวแปรอิสระ 2 ตัว มีความสัมพันธ์กันแบบสมบูรณ์

คือ มีตัวแปรอิสระ X_1, X_2 เราสามารถเขียนตัวแปรตัวหนึ่งให้อยู่ในรูปของฟังก์ชันของอีกตัวแปรหนึ่ง เช่น $X_2=3X_1$ เป็นต้น

กรณีที่ 2 ตัวแปรอิสระ 3 ตัว (มากกว่า 2 ตัว) มีความสัมพันธ์กันแบบสมบูรณ์

คือ มีตัวแปรอิสระ X_1, X_2, X_3 เราสามารถเขียนตัวแปรตัวหนึ่งเขียนให้อยู่ในรูปผลรวมเชิงเส้นของตัวแปรที่เหลือ เช่น $X_3=X_2+X_1$ เป็นต้น

กรณีที่ 3 ตัวแปรอิสระ 2 ตัว มีความสัมพันธ์กันแบบไม่สมบูรณ์

กรณีที่ 4 ตัวแปรอิสระ 3 ตัว (มากกว่า 2 ตัว) มีความสัมพันธ์กันแบบไม่สมบูรณ์

(สุพล ดุรงค์วัฒนา, 2549: 316-324)

โดยมีตารางแสดงระดับความสัมพันธ์ดังนี้

ค่าสัมประสิทธิ์สหสัมพันธ์	ระดับความสัมพันธ์
0.80 - 1.00	สูงหรือสูงมาก
0.60 - 0.79	ค่อนข้างสูง
0.40 – 0.59	ปานกลาง
0.20 – 0.39	ค่อนข้างต่ำ
0.00 – 0.19	ต่ำ

ตารางที่ 2.2.1ตารางแสดงระดับความสัมพันธ์

ซึ่งค่าสัมประสิทธิ์สหสัมพันธ์จะมีค่าตั้งแต่ -1 ถึง +1 โดยที่ค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเป็นบวก หมายถึงตัวแปรทั้ง 2 มีความสัมพันธ์เป็นไปในทิศทางเดียวกัน เมื่อสร้างกราฟแสดงความสัมพันธ์ของตัวแปรทั้ง 2 พบว่าเส้นกราฟจะทำมุมแหลมกับแกนนอน ค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเป็นลบ หมายถึงตัวแปรทั้ง 2 มีความสัมพันธ์เป็นไปในทิศทางตรงกันข้าม เมื่อสร้างกราฟแสดงความสัมพันธ์ของตัวแปรทั้ง 2 พบว่าเส้นกราฟจะทำมุมป้านกับแกนนอน และค่าสัมประสิทธิ์สหสัมพันธ์ที่มีค่าเป็นศูนย์ หมายถึงตัวแปรทั้ง 2 ไม่มีความสัมพันธ์กันเลย เมื่อสร้างกราฟแสดงความสัมพันธ์ของตัวแปรทั้ง 2 พบว่าเส้นกราฟจะขนานกับแกนนอน

2.3 ปัญหาพหุสัมพันธ์ (Multicollinearity) ของตัวแปรอิสระ

ปัญหาพหุสัมพันธ์เกิดจากการที่ตัวแปรอิสระมีความสัมพันธ์กันในระดับสูงมาก เมื่อนำมาวิเคราะห์ทางสถิติ พบว่าค่าความคลาดเคลื่อนมาตรฐานของตัวประมาณพารามิเตอร์มีค่าสูงขึ้น รวมไปถึงการทดสอบสมมติฐานของค่าสัมประสิทธิ์ความถดถอยเชิงส่วนที่ไม่มีนัยสำคัญ จึงทำให้เกิดปัญหาในการวิเคราะห์ อธิบายผล และการพยากรณ์จากสมการถดถอยผิดพลาดได้

2.4 การแก้ไขปัญหาพหุสัมพันธ์ของตัวแปรอิสระที่สัมพันธ์กัน

จากการที่ตัวแปรอิสระมีความสัมพันธ์กันจนเกิดปัญหาพหุสัมพันธ์ เราจะมีวิธีในการแก้ไขปัญหาได้ดังนี้

1. ไม่ต้องแก้ไขอะไรทั้งสิ้น ทั้งนี้เป็นเพราะความสัมพันธ์ร่วมที่เกิดขึ้นนั้นไม่สูงนัก หรืออาจสนใจเพียงเพื่อให้ได้สมการถดถอยสำหรับการประมาณค่าหรือพยากรณ์โดยคำจึงถึงเพียงว่าค่า R^2 มีค่าสูงพอที่จะนำไปใช้ประมาณค่าพยากรณ์ได้

2. สามารถแก้ไขด้วยการเก็บข้อมูลเพิ่มขึ้น (ถ้าเป็นไปได้) โดยพยายามหาข้อมูลตัวแปรอิสระใหม่ที่ไม่สัมพันธ์กับตัวแปรอิสระตัวอื่นๆ หรือถ้าเป็นไปไม่ได้ในกรณีดังกล่าวอาจหาข้อมูลโดยเพิ่มขนาดตัวอย่างก็ได้ ทั้งนี้ จะช่วยให้ค่าความแปรปรวนของค่าประมาณของสัมประสิทธิ์การถดถอยมีค่าลดลง

3. ถ้าปัญหาความสัมพันธ์ร่วมที่เกิดขึ้นนั้นเกิดจากการใช้ตัวแบบความถดถอยโพลิโนเมียลแล้ว แก้ไขด้วยการแปลงข้อมูลด้วยการทำให้เป็นศูนย์กลาง (Centering Method)

4. ตัดตัวแปรอิสระที่ก่อให้เกิดปัญหาออกไป ถ้าตัวแปรอิสระตัวแปรนั้นไม่ได้เป็นตัวแปรอิสระร่วมกับตัวแปรอิสระอื่นๆที่ก่อให้เกิดปัญหาดังกล่าว

5. การสร้างตัวแปรอิสระตัวใหม่ซึ่งเป็นฟังก์ชันของตัวแปรอิสระที่ก่อให้เกิดปัญหา (Respecification)

6. แก้ไขด้วยวิธีการหาองค์ประกอบหลัก (Principal Components)

7. แก้ไขด้วยการประมาณและวิเคราะห์ด้วยวิธีการวิเคราะห์การถดถอยแบบริดจ์ (Ridge Regression Analysis)

8. การใช้หลักการของการคัดเลือกตัวแปรอิสระที่เหมาะสมให้อยู่ในตัวแบบความถดถอย (Independent Variable Selection) (สุพล ดุรงค์วัฒนา, 2549: 341)

2.5 **การวิเคราะห์ตัวประกอบหลัก** (Principal Component Analysis)

การวิเคราะห์ตัวประกอบหลักเป็นการสร้างตัวแปรใหม่ที่ เรียกว่า ตัวประกอบหลัก (Principal Component) ให้อยู่ในรูปผลรวมเชิงเส้นของตัวแปรอิสระเดิม โดยที่ตัวประกอบหลัก (ตัวแปรอิสระตัวใหม่) แต่ละตัวเป็นอิสระกัน

จากตัวแบบความถดถอยเชิงพหุ จะได้ว่า

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_p X_{pi} + \varepsilon_i ; i = 1,2,..., p$$

หรือเขียนให้อยู่ในรูปทั่วไป

$$\tilde{y} = \beta_0 \tilde{1} + X\tilde{\beta} + \tilde{\varepsilon}$$

หรือเขียนให้อยู่ในรูปเมทริกซ์ได้ดังนี้

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ x_{12} & x_{22} & \cdots & x_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

พิจารณาเฉพาะตัวแปรอิสระ

$$X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{p1} \\ x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix}_{n \times p}$$

และมีเมทริกซ์ความแปรปรวนร่วมของตัวอย่าง

$$S = \begin{bmatrix} s_{11} & s_{21} & \cdots & s_{p1} \\ s_{12} & s_{22} & \cdots & s_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ s_{1p} & s_{2p} & \cdots & s_{pp} \end{bmatrix}_{p \times p}$$

โดยที่ $s_{jj} = s_j^2$ คือค่าความแปรปรวนตัวอย่างของตัวแปรที่ j ; j=1,2,…,p

$$s_{jj} = s_j^2 = \frac{\sum_{i=1}^{n} (x_{ji} - \bar{x}_j)}{n-1}$$

$s_{jj'}$ คือค่าความแปรปรวนร่วมตัวอย่างของตัวแปรที่ j และ j' ; j=1,2,…,p

$$s_{jj'} = \frac{\sum_{i=1}^{n} (x_{ji} - \bar{x}_j)(x_{j'i} - \bar{x}_{j'})}{n-1}$$

การปรับข้อมูลให้อยู่ในรูปมาตรฐาน

จากสมการถดถอยเชิงพหุ $\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon}$

กำหนดให้ $X_{ji}^* = \frac{x_{ji} - \bar{x}_j}{s_j}$, $\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ji}}{n}$, $s_j = \sqrt{\frac{\sum_{i=1}^{n} (x_{ji} - \bar{x}_j)^2}{n-1}}$; j = 1,2,…,p

จะได้ $y_i = \beta_0 + \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + ... + \beta_p^* X_{pi}^* + \varepsilon_i$

หรือ $\tilde{y} = \beta_0 \tilde{1} + X^* \tilde{\beta} + \tilde{\varepsilon}$

$$X^* = \begin{bmatrix} \frac{x_{11}-\bar{x}_1}{s_1} & \frac{x_{21}-\bar{x}_2}{s_2} & \cdots & \frac{x_{p1}-\bar{x}_p}{s_p} \\ \frac{x_{12}-\bar{x}_1}{s_1} & \frac{x_{22}-\bar{x}_2}{s_2} & \cdots & \frac{x_{p2}-\bar{x}_p}{s_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{1n}-\bar{x}_1}{s_1} & \frac{x_{2n}-\bar{x}_2}{s_2} & \cdots & \frac{x_{pn}-\bar{x}_p}{s_p} \end{bmatrix}$$

$$X^{*T}X^* = \begin{bmatrix} \frac{x_{11}-\bar{x}_1}{s_1} & \frac{x_{12}-\bar{x}_1}{s_1} & \cdots & \frac{x_{1n}-\bar{x}_1}{s_1} \\ \frac{x_{21}-\bar{x}_2}{s_2} & \frac{x_{22}-\bar{x}_2}{s_2} & \cdots & \frac{x_{2n}-\bar{x}_2}{s_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{p1}-\bar{x}_p}{s_p} & \frac{x_{p2}-\bar{x}_p}{s_p} & \cdots & \frac{x_{pn}-\bar{x}_p}{s_p} \end{bmatrix} \begin{bmatrix} \frac{x_{11}-\bar{x}_1}{s_1} & \frac{x_{21}-\bar{x}_2}{s_2} & \cdots & \frac{x_{p1}-\bar{x}_p}{s_p} \\ \frac{x_{12}-\bar{x}_1}{s_1} & \frac{x_{22}-\bar{x}_2}{s_2} & \cdots & \frac{x_{p2}-\bar{x}_p}{s_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{1n}-\bar{x}_1}{s_1} & \frac{x_{2n}-\bar{x}_2}{s_2} & \cdots & \frac{x_{pn}-\bar{x}_p}{s_p} \end{bmatrix}$$

$$X^{*T}X^* = \begin{bmatrix} \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}{s_1^2} & \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(X_{2i}-\bar{x}_2)}{s_1 s_2} & \cdots & \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{pi}-\bar{x}_p)}{s_1 s_p} \\ \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1)}{s_2 s_1} & \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}{s_2^2} & \cdots & \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)(x_{pi}-\bar{x}_p)}{s_2 s_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sum_{i=1}^{n}(x_{pi}-\bar{x}_p)(x_{1i}-\bar{x}_1)}{s_p s_1} & \frac{\sum_{i=1}^{n}(x_{pi}-\bar{x}_p)(x_{2i}-\bar{x}_2)}{s_p s_2} & \cdots & \frac{\sum_{i=1}^{n}(x_{pi}-\bar{x}_p)^2}{s_p^2} \end{bmatrix}$$

เมทริกซ์สหสัมพันธ์ (Correlation Matrices)

กรณีที่มีตัวแปร 2 ตัว สามารถหาค่าสหสัมพันธ์ตัวอย่างของตัวแปรที่ j และ j'

$$r_{jj'} = \frac{s_{jj'}}{s_j s_{j'}} = \frac{\sum_{i=1}^{n}(x_{ji}-\bar{x}_j)(x_{j'i}-\bar{x}_{j'})}{s_j s_{j'}}$$

เมทริกซ์สหสัมพันธ์ตัวอย่าง กำหนดให้ R เป็นเมทริกซ์สหสัมพันธ์ของตัวแปร p ตัวแปร

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix}$$

จะได้ว่า $X^{*T}X^* = R$

<u>หลักการของการวิเคราะห์ตัวประกอบหลัก</u>

ตัวประกอบหลักตัวที่ 1 (PC_1)

ตัวประกอบหลักตัวที่ 1 จะต้องดึงความแปรปรวนจากตัวแปรทั้งหมด p ตัว มาไว้ในตัวประกอบหลักตัวที่ 1 มากที่สุด และอิสระกับตัวประกอบหลักตัวอื่นๆ

ตัวประกอบหลักตัวที่ 2 (PC_2)

ตัวประกอบหลักตัวที่ 2 จะต้องดึงความแปรปรวนที่เหลือจากการสร้างตัวประกอบหลักตัวที่ 1 ให้มากที่สุด และเป็นอิสระกับตัวประกอบหลักตัวอื่นๆ จนถึง

ตัวประกอบหลักตัวที่ p (PC_p)

ตัวประกอบหลักตัวที่ p จะต้องดึงความแปรปรวนที่เหลือจากการสร้างตัวประกอบหลักตัวที่ 1 ถึง ตัวที่ p-1 ให้มากที่สุด และเป็นอิสระกับตัวแปรใหม่ตัวอื่นๆ

<u>โดยหลักการสร้างตัวประกอบหลัก</u>

ซึ่งมีวิธีการสร้างดังนี้

<u>หาค่า Eigenvalue และ Eigenvector</u>

<u>นิยาม</u> โดยเปลี่ยนจากเมทริกซ์จัตุรัส S เป็นเมทริกซ์สหสัมพันธ์ R ค่าสเกลาร์ λ และเวกเตอร์ w ที่มีค่าไม่เป็นศูนย์

การหา Eigenvalue $(\lambda_1, \lambda_2, ..., \lambda_p)$ หาจาก $\det(R-\lambda I)=0$ จากนั้นนำค่า λ แทนกลับลงในสมการ $(R-\lambda I)\tilde{w}=0$ แล้วแก้สมการหาค่า $\tilde{w}$ (กัลยา วานิชย์บัญชา, 2552: 524-526)

โดยที่สัมประสิทธิ์ PC_j คือ $\tilde{w}_j$ มีคุณสมบัติดังนี้ $\tilde{w}_j^{\,T} = (w_{1j} \quad w_{2j} \quad \cdots \quad w_{pj})$

$\tilde{w}_j^{\,T}\tilde{w}_j = 1$ และ $\tilde{w}_j^{\,T}\tilde{w}_{j'} = 0; j \neq j'$=1, 2,..., p

จะได้ว่า W เป็นเมทริกซ์เมทริกซ์เชิงตั้งฉาก (Orthonormal Matrix) $W^TW = WW^T = I_{p\times p}$

เราสามารถเขียนตัวประกอบหลัก PC_j ให้อยู่ในรูปของผลรวมเชิงเส้นของเวกเตอร์สุ่ม $x^{*T} = (X_1^*, X_2^*, ..., X_p^*)$ ได้ดังนี้

$$PC_1 = w_{11}X_1^* + w_{21}X_2^* + ... + w_{p1}X_p^*$$
$$PC_2 = w_{12}X_1^* + w_{22}X_2^* + ... + w_{p2}X_p^*$$
$$\vdots$$
$$PC_p = w_{1p}X_1^* + w_{2p}X_2^* + ... + w_{pp}X_p^*$$

โดยสัมประสิทธิ์ของ เวกเตอร์ PC_j คือ $\tilde{w}_j^{\,T} = (w_{1j} \quad w_{2j} \quad \cdots \quad w_{pj})$; j=1,2,..., p เป็น eigenvector ที่สอดคล้องกับ eigenvalue และ เวกเตอร์ PCต้องเป็นอิสระกัน

<u>การแปลงสมการถดถอยเชิงพหุของตัวแปรอิสระตัวใหม่</u>

จากสมการถดถอยเชิงพหุ $\tilde{y} = \beta_0\tilde{1} + X\tilde{\beta} + \tilde{\varepsilon}$

เมื่อทราบว่า W เป็นเมทริกซ์เมทริกซ์เชิงตั้งฉาก $WW^T = I_{p\times p}$

จะได้ว่า $\tilde{y} = \beta_0 \tilde{1} + X^* I\tilde{\beta} + \tilde{\varepsilon}$

$$\tilde{y} = \beta_0 \tilde{1} + X^* WW^T \tilde{\beta} + \tilde{\varepsilon}$$

กำหนดให้ $Z = X^*W$ และ $\tilde{\gamma} = W^T\tilde{\beta}$

ได้สมการ $\tilde{y} = \beta_0 \tilde{1} + Z\tilde{\gamma} + \tilde{\varepsilon}$

โดยที่ Z คือ เมทริกซ์ตัวประกอบหลัก (Principal Component Matrix)

$\tilde{\gamma}$ คือ สัมประสิทธิ์ความถดถอยตัวใหม่ (New set of regression coefficients)

จาก $Z = X^*W$

$$Z^TZ = (X^*W)^T(X^*W)$$

$$Z^TZ = (W^TX^{*T})(X^*W)$$

$$Z^TZ = W^TRW$$

$$Z^TZ = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

เมื่อ $\tilde{z}_j^T\tilde{z}_j = \tilde{w}_j^TR\tilde{w}_j = \lambda_j$ และ $\tilde{z}_j^T\tilde{z}_{j'} = \tilde{w}_j^TR\tilde{w}_{j'} = 0$; $j \neq j' = 1,2,..., p$

$Z_{ji} = w_{j1}X_{1i}^* + w_{j2}X_{2i}^* + ... + w_{jp}X_{pi}^*$; j=1, 2,…,p , i=1,2,…,n

เมื่อ $\sum_{i=1}^{n} Z_{ji} = 0$, $\sum_{i=1}^{n} Z_{ji}^2 = 0$ และ $\sum_{j=1}^{p}\lambda_j = 0$, $\sum_{j=1}^{p}(\sum_{i=1}^{n} Z_{ji}^2) = p$

จากนั้นจะใช้วิธีกำลังสองน้อยสุดในการประมาณค่าสัมประสิทธิ์ความถดถอยเชิงพหุของตัวแปรอิสระตัวใหม่ ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก

$$\tilde{\gamma} = (Z^TZ)^{-1}Z^T\tilde{y}$$

ความแปรปรวนของตัวแปรอิสระตัวใหม่

$$\text{var}(\tilde{\gamma}) = \sigma^2(Z^TZ)^{-1}$$

(S.Q.Lafi and J.B.Kaneene, 1992: 261-275)

2.6 การแจกแจงของตัวแปรอิสระ

การศึกษาวิจัยนี้ได้กำหนดการแจกแจงของตัวแปรอิสระคือ การแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) ถ้าเวกเตอร์ของตัวแปร $x^T = (X_1, X_2,..., X_p)$ มี

การแจกแจงแบบปกติที่มีเวกเตอร์ของค่าเฉลี่ยเป็น μ และเมทริกซ์ความแปรปรวนร่วม (Covariance matrix) คือ $\sum$ โดยที่ฟังก์ชันความน่าจะเป็นดังนี้

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

โดยที่ p คือจำนวนตัวแปรหรือเขียนย่อๆว่า $x \sim \mathrm{N_p}(\tilde{\mu}, \sum)$

ในการศึกษาครั้งนี้ได้ศึกษาภายใต้เงื่อนไขค่าเฉลี่ย $\tilde{\mu} = (0,0,...,0)$ และเมทริกซ์ความแปรปรวนร่วมตามระดับความสัมพันธ์ของตัวแปรอิสระ (กัลยา วานิชย์บัญชา, 2552: 15)

2.7 การจำลองเวกเตอร์สุ่มแบบปกติ

กำหนดให้ $X_1, X_2, ..., X_n$ เป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติจำนวน n ตัว c_{ij} คือ ค่าคงที่แถวที่ i หลักที่ j และมีเวกเตอร์ค่าเฉลี่ย $\tilde{\mu}$ และเมทริกซ์ความแปรปรวนร่วม $\sum$ เราสามารถจำลองข้อมูลโดยการสร้างตัวแปรสุ่มร่วมแบบปกติ n ตัว ที่เป็นอิสระกัน โดยใช้สัญลักษณ์ $Z_1, Z_2, ..., Z_n$ แล้วเขียนให้อยู่ในรูปผลรวมเชิงเส้นได้ดังนี้

$$\begin{aligned} X_1 &= c_{11}Z_1 + c_{12}Z_2 + c_{13}Z_3 + ... + c_{1n}Z_n \\ X_2 &= c_{21}Z_1 + c_{22}Z_2 + c_{23}Z_3 + ... + c_{2n}Z_n \\ &\vdots \\ X_n &= c_{n1}Z_1 + c_{n2}Z_2 + c_{n3}Z_n + ... + c_{nn}Z_n \end{aligned}$$

และสามารถเขียนอยู่ในรูปของเมทริกซ์ได้ ดังนี้

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

โดยที่ c_{ij} เป็นสมาชิกในแถวที่ i หลักที่ j ของเมทริกซ์ C

ให้ $\sum$ เป็นเมทริกซ์ความแปรปรวนของ X จะได้ว่า $\sum = E[(X-\tilde{\mu})(X-\tilde{\mu})^T]$

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

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

ดังนั้น $$\sum = CC^T$$

<u>ทฤษฎีบท</u> (Cholesky's decomposition) ถ้าเมทริกซ์ M เป็นเมทริกซ์สมมาตรที่เป็นบวกแน่นอน (Positive-definite symmetric matrix) แล้ว M สามารถแยกเป็นเมทริกซ์สามเหลี่ยมล่างที่องค์ประกอบเป็นเลขจำนวนจริงซึ่ง $M = LL^T$

จะได้องค์ประกอบบนเส้นทแยงมุม L_{11} ของเมทริกซ์ L คือ

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

และองค์ประกอบใต้เส้นทแยงมุม L_{ji} ของเมทริกซ์ L คือ

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

โดยที่ m_{ij} เป็นสมาชิกในแถวที่ i หลักที่ j ของเมทริกซ์ M

(เสกสรร เกียรติสุไพบูลย์, 2555: 52-55)

2.8 ความเอนเอียง

คือ ค่าที่ใช้วัดค่าเฉลี่ยของตัวสถิตได้ห่างจากฟังก์ชันพารามิเตอร์ที่สนใจมากหรือน้อยเพียงใด และสามารถบอกทิศทางได้ด้วยว่าตัวสถิติที่ใช้ให้ค่าสูงหรือต่ำกว่าค่าพารามิเตอร์ สามารถหาค่าความเอนเอียงได้จาก $b(\theta;T) = E_\theta[T(\widetilde{X})] - q(\theta)$ โดยที่ $b(\theta;T)$ คือค่าความเอนเอียง, $E_\theta[T(\widetilde{X})]$ คือค่าเฉลี่ยของตัวสถิติ และ $q(\theta)$ คือค่าพารามิเตอร์

(ธีระพร วีระถาวร, 2536: 51)

2.9 สถิติที่ใช้ในการตัดสินใจ

1. มาตรวัด Variance Inflation Factor (VIF)

ในการศึกษานี้ใช้มาตรวัด Variance Inflation Factor (VIF) ในการตัดสินใจว่าเกิดปัญหา Collinearity ที่รุนแรงมากหรือไม่ ได้ดังนี้

มาตรวัดนี้เป็นส่วนกลับของมาตรวัด Tolerance (T_j; j=1,2,…,p) จะใช้สัญลักษณ์ด้วย VIF_j ดังนั้น มาตรวัด VIF นี้สามารถคำนวณได้ดังนี้

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

โดยที่ R_j^2 คือค่าสัมประสิทธิ์การตัดสินใจเชิงพหุของตัวแบบความถดถอย ตัวแปรอิสระตัวที่ j กับตัวแปรอิสระตัวอื่น

ถ้าค่า $VIF_j > 10; j = 1,2,..., p$ จะได้ว่าตัวแปรอิสระตัวที่ j มีความสัมพันธ์ร่วมกันกับตัวแปรอิสระตัวที่เหลือ

ถ้าค่า $VIF_j \le 10; j = 1,2,..., p$ จะได้ว่าตัวแปรอิสระตัวที่ j ไม่มีความสัมพันธ์ร่วมกันกับตัวแปรอิสระตัวที่เหลือ

บทที่ 3

วิธีการดำเนินการศึกษา

การศึกษาครั้งนี้มีวัตถุประสงค์ในการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ซึ่งจำลองข้อมูลโดยใช้โปรแกรม R มีขั้นตอนและวิธีการดำเนินการดังนี้

3.1 ขอบเขตการศึกษา

1. กำหนดตัวแปรอิสระ(x) คือ 2, 3 ตัวแปร เป็นตัวแปรเชิงปริมาณ ที่มีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 2 และตัวแปรอิสระมีความสัมพันธ์ดังนี้

1.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว (p=2)

ρ_{12} คือค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และ ตัวที่ 2

โดยทำการศึกษาระดับความสัมพันธ์ดังนี้

1.1.1 ระดับความสัมพันธ์ 0.90

1.1.2 ระดับความสัมพันธ์ 0.91

1.1.3 ระดับความสัมพันธ์ 0.92

1.1.4 ระดับความสัมพันธ์ 0.93

1.1.5 ระดับความสัมพันธ์ 0.94

1.1.6 ระดับความสัมพันธ์ 0.95

1.1.7 ระดับความสัมพันธ์ 0.96

1.1.8 ระดับความสัมพันธ์ 0.97

1.1.9 ระดับความสัมพันธ์ 0.98

1.1.10 ระดับความสัมพันธ์ 0.99

1.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว (p=3)

ρ_{12} คือค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และ ตัวที่ 2

ρ_{13} คือค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และ ตัวที่ 3

ρ_{23} คือค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 2 และ ตัวที่ 3

โดยทำการศึกษาระดับความสัมพันธ์ดังนี้

1.2.1 ระดับความสัมพันธ์ (0.91, 0.93, 0.95)

1.2.2 ระดับความสัมพันธ์ (0.91, 0.93, 0.97)

1.2.3 ระดับความสัมพันธ์ (0.91, 0.93, 0.99)

1.2.4 ระดับความสัมพันธ์ (0.91, 0.95, 0.97)

1.2.5 ระดับความสัมพันธ์ (0.91, 0.95, 0.99)

1.2.6 ระดับความสัมพันธ์ (0.93, 0.95, 0.97)

1.2.7 ระดับความสัมพันธ์ (0.93, 0.95, 0.99)

1.2.8 ระดับความสัมพันธ์ (0.93, 0.97, 0.99)

1.2.9 ระดับความสัมพันธ์ (0.95, 0.97, 0.99)

1.2.10 ระดับความสัมพันธ์ (0.91, 0.91, 0.91)

1.2.11 ระดับความสัมพันธ์ (0.93, 0.93, 0.93)

1.2.12 ระดับความสัมพันธ์ (0.95, 0.95, 0.95)

1.2.13 ระดับความสัมพันธ์ (0.97, 0.97, 0.97)

1.2.14 ระดับความสัมพันธ์ (0.99, 0.99, 0.99)

2. ศึกษาภายใต้ขนาดตัวอย่างดังนี้

2.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว จะศึกษาขนาดตัวอย่าง 50, 100, 200

2.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว จะศึกษาขนาดตัวอย่าง 50, 100, 200

3. กำหนดค่าสัมประสิทธิ์ความถดถอย $(\beta_0, \beta_1, \beta_2, ..., \beta_p)$ ได้ดังนี้

3.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว มีค่าสัมประสิทธิ์ความถดถอยเริ่มต้น คือ

$\beta_0 = \beta_1 = \beta_2 = 2$

3.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว มีค่าสัมประสิทธิ์ความถดถอยเริ่มต้น คือ

$\beta_0 = \beta_1 = \beta_2 = \beta_3 = 2$

4. จำลองข้อมูลความคลาดเคลื่อนที่มีการแจกแจงปกติ โดยมีค่าเฉลี่ยเท่ากับ 0 ค่าความแปรปรวนเท่ากับ 2

5. สร้างตัวแปรตาม y จากตัวแบบความถดถอยพหุเชิงเส้น $\tilde{y} = X\tilde{\beta} + \tilde{\varepsilon}$

3.2 วิธีดำเนินการศึกษา

1. กำหนดค่าเริ่มต้นในการจำลองข้อมูล ดังนี้

1.1 กำหนดตัวแปรอิสระ p ตัว

1.2 กำหนดขนาดตัวอย่าง n ตัว

1.3 กำหนดค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $(\widetilde{\beta})$

1.4 กำหนดระดับความสัมพันธ์ระหว่างตัวแปรอิสระ (ρ)

2. จำลองข้อมูลตัวแปรอิสระที่มีการแจกแจงปกติหลายตัวแปร โดยตัวแปรอิสระสร้างจากค่าเฉลี่ย ค่าความแปรปรวนและระดับความสัมพันธ์ของตัวแปรอิสระที่กำหนด

3. จำลองข้อมูลความคลาดเคลื่อนที่มีการแจกแจงปกติ โดยมีค่าเฉลี่ยและค่าความแปรปรวนตามที่กำหนด

4. สร้างตัวแปรตาม y จากตัวแบบความถดถอยพหุเชิงเส้น $\widetilde{y} = X\widetilde{\beta} + \widetilde{\varepsilon}$

5. เกณฑ์ในการตัดสินใจว่าเกิดปัญหา Collinearityที่รุนแรงหรือไม่ โดยใช้ค่า VIF

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

โดยที่ R_j^2 คือค่าสัมประสิทธิ์การตัดสินใจเชิงพหุของตัวแบบความถดถอย ตัวแปรอิสระตัวที่ j กับตัวแปรอิสระตัวอื่น

ถ้าค่า $VIF_j > 10; j = 1,2,...,p$ จะได้ว่าตัวแปรอิสระตัวที่ j มีความสัมพันธ์ร่วมกันกับตัวแปรอิสระตัวที่เหลือ

ถ้าค่า $VIF_j \le 10; j = 1,2,...,p$ จะได้ว่าตัวแปรอิสระตัวที่ j ไม่มีความสัมพันธ์ร่วมกันกับตัวแปรอิสระตัวที่เหลือ

6. คำนวณหาค่าสัมประสิทธิ์ความถดถอยของตัวแปรอิสระก่อนทำการแก้ไขปัญหาพหุสัมพันธ์โดยวิธีกำลังสองน้อยสุด หาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานและค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

6.1 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด

$$\widetilde{b} = (X^T X)^{-1} X^T \widetilde{y}$$

6.2 การหาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย

จากเมทริกซ์ความแปรปรวนร่วม $\hat{C}ov(\widetilde{b}) = \hat{\sigma}^2 (X^T X)^{-1}$

$$\hat{\sigma}^2 = MSE = \frac{\sum_{i=1}^{n} (y - \hat{y})^2}{n - p - 1}$$

สามารถหาความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยได้ดังนี้

$$\hat{s}e(b_j) = \sqrt{\hat{\mathrm{var}}(b_j)}$$

ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจำนวน N รอบ

$$SE(b_j) = \frac{\sum_{i=1}^{N} \hat{s}e(b_j)}{N}$$

ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย

$$ASE = \frac{\sum_{j=0}^{p} SE(b_j)}{p+1}$$

6.3 การหาค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

ค่าความเอนเอียงของสัมประสิทธิ์ความถดถอยคือ

$$bias(b_j) = \left| \frac{\sum_{i=1}^{N} (b_{ji})}{N} - \beta_j \right|$$

ค่าเฉลี่ยความเอนเอียงจำนวน N รอบ

$$Bias(b_j) = \frac{\sum_{i=1}^{N} bias(b_j)}{N}$$

ค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

$$ABias = \frac{\sum_{j=0}^{p} Bias(b_j)}{p+1}$$

โดยที่ N คือ จำนวนรอบในการทดลองซ้ำ

b_j คือ สัมประสิทธิ์ความถอยถอย j=0,1,2,..,p

7. การแก้ไขปัญหาพหุสัมพันธ์ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก โดยนำตัวแปรอิสระ คือ เมทริกซ์ X มาสร้างเมทริกซ์สหสัมพันธ์ (R) แล้วหาค่า Eigenvalue จากสมการ $|R - I\lambda| = 0$ และนำค่า λ แทนลงในสมการข้างต้น เพื่อหาค่า Eigenvector ซึ่ง Eigenvector แต่ละชุดจะเป็นสัมประสิทธิ์ของตัวแปรอิสระเดิม

8. นำค่า Eigenvector คูณกับตัวแปรอิสระเดิมที่ทำให้อยู่ในรูปมาตรฐาน จะได้ตัวประกอบหลัก (PC_j) ของแต่ละชุด ซึ่ง (PC_j) ก็คือตัวแปรอิสระตัวใหม่

9. คำนวณหาค่าสัมประสิทธิ์ความถดถอยของตัวแปรอิสระตัวใหม่โดยวิธีกำลังสองน้อยสุด หาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานและค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

9.1 การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีกำลังสองน้อยสุด

$$\tilde{\gamma} = (Z^T Z)^{-1} Z^T \tilde{y}$$

9.2 การหาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย จากเมทริกซ์ความแปรปรวนร่วม $\hat{\text{var}}(\tilde{\gamma}) = \hat{\sigma}^2 (Z^T Z)^{-1}$

สามารถหาความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยได้ดังนี้

$$\hat{s}e(\gamma_j) = \sqrt{\hat{\mathrm{var}}(\gamma_j)}$$

ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจำนวน N รอบ

$$SE(\gamma_j) = \frac{\sum_{i=1}^{N} \hat{s}e(\gamma_j)}{N}$$

ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย

$$ASE = \frac{\sum_{j=0}^{p} SE(\gamma_j)}{p+1}$$

9.3 การหาค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

ค่าความเอนเอียงของสัมประสิทธิ์ความถดถอยคือ

$$bias(\gamma_j) = \left| \frac{\sum_{i=1}^{N} (\gamma_{ji})}{N} - \beta_j \right|$$

ค่าเฉลี่ยความเอนเอียงจำนวน N รอบ

$$Bias(\gamma_j) = \frac{\sum_{i=1}^{N} bias(\gamma_j)}{N}$$

ค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

$$ABias = \frac{\sum_{j=0}^{p} Bias(\gamma_j)}{p+1}$$

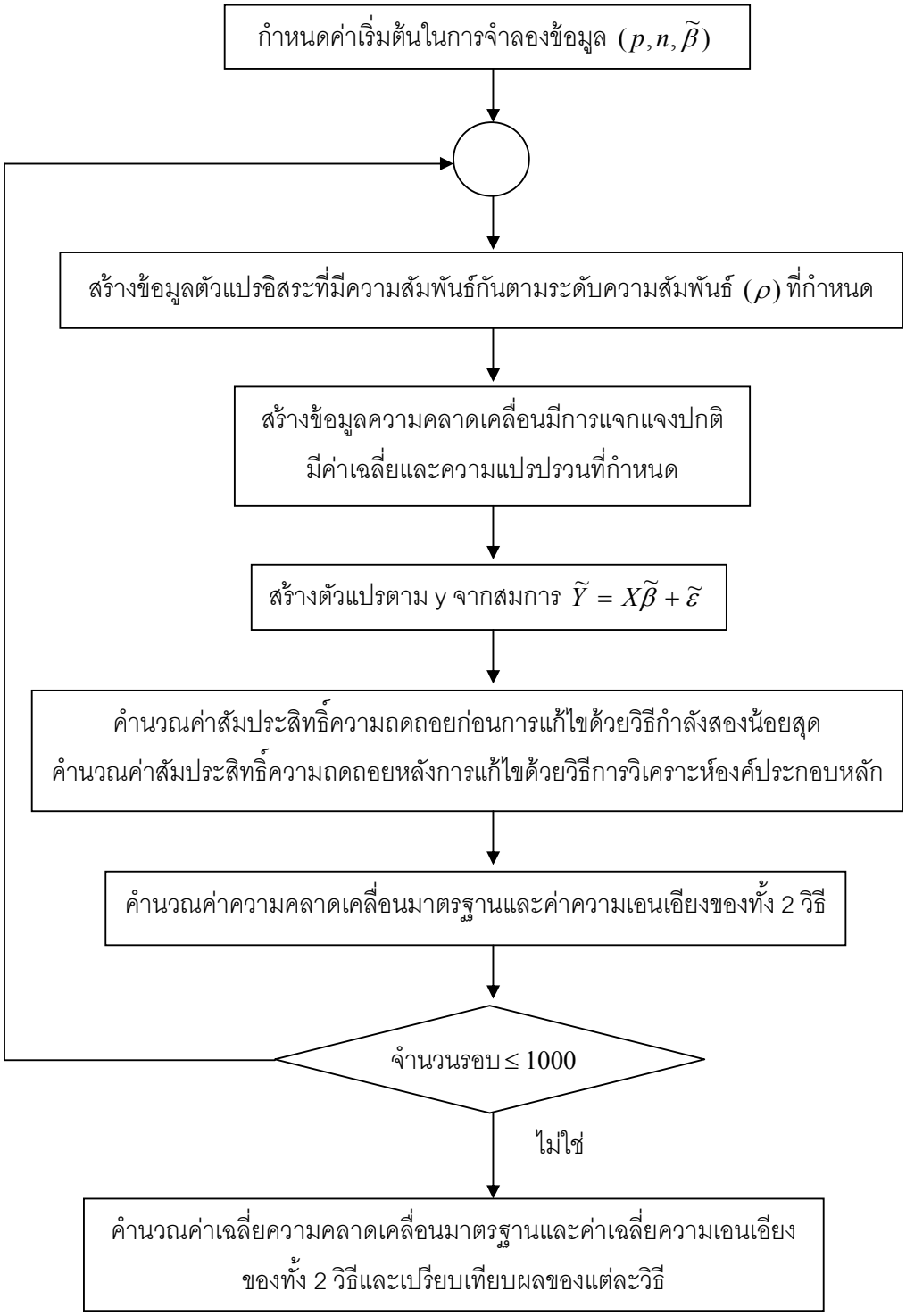
โดยที่ N คือ จำนวนรอบในการทดลองซ้ำ

b_j คือ สัมประสิทธิ์ความถอยถอย j=0,1,2,..,p

10. พิจารณาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานและค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนและหลังแก้ไขปัญหาพหุสัมพันธ์รุนแรง

11. สรุปผลการศึกษา

แผนผังการเขียนโปรแกรม



บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้มีวัตถุประสงค์เพื่อหาค่าประมาณพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้นที่เหมาะสม เมื่อเกิดปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก โดยขึ้นอยู่กับปัจจัยที่มีผลต่อการเกิดปัญหาพหุสัมพันธ์รุนแรง คือ จำนวนตัวแปรอิสระ ขนาดตัวอย่าง และค่าระดับความสัมพันธ์ของตัวแปรอิสระ โดยทำการจำลองข้อมูลด้วยโปรแกรม R

สามารถวิเคราะห์ผลการศึกษาได้ดังนี้

4.1 ผลการศึกษาในกรณีจำนวนตัวแปรอิสระ 2 ตัว

จากการจำลองข้อมูล ในการศึกษากรณีจำนวนตัวแปรอิสระ 2 ตัว จะแสดงผลการศึกษาในแต่ละหัวข้อ ได้ดังนี้

1. แสดงค่าเฉลี่ย VIF จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์ เพื่อหาค่าระดับความสัมพันธ์ที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง

2. แสดงค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

3. แสดงค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

4. แสดงค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

ตารางที่ 4.1.1 แสดงค่าเฉลี่ย VIF จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

ρ	n=50		n=100		n=200	
	VIF1	VIF2	VIF1	VIF2	VIF1	VIF2
0.9	5.575955	5.575955	5.403379	5.403379	5.338576	5.338576
0.91	6.168992	6.168992	5.974103	5.974103	5.901098	5.901098
0.92	6.910739	6.910739	6.687858	6.687858	6.60456	6.60456
0.93	7.864936	7.864936	7.605931	7.605931	7.509348	7.509348

ตารางที่ 4.1.1(ต่อ) แสดงค่าเฉลี่ย VIF จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

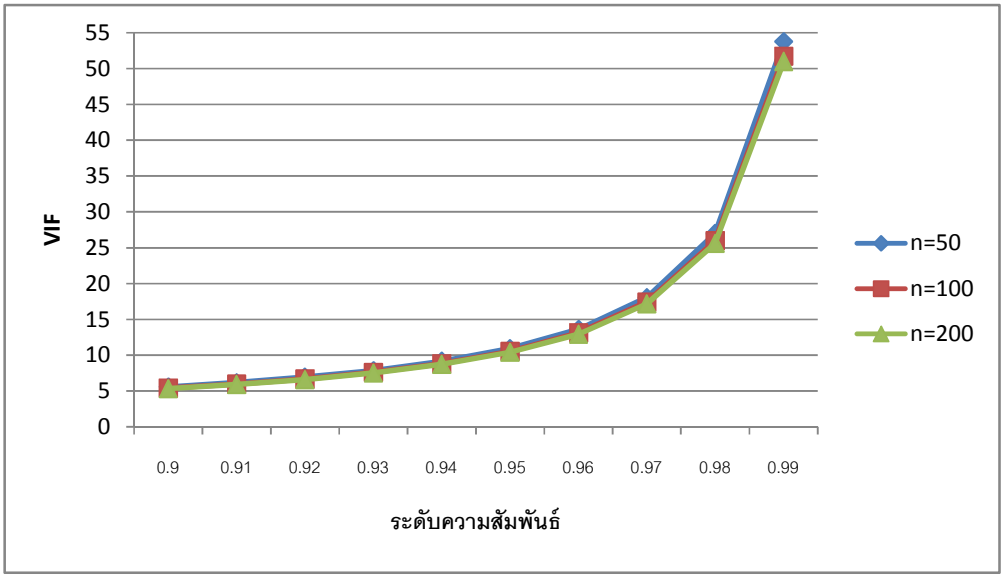
ρ	n=50		n=100		n=200	
	VIF1	VIF2	VIF1	VIF2	VIF1	VIF2
0.94	9.137819	9.137819	8.830469	8.830469	8.716101	8.716101
0.95	10.92063	10.92063	10.54533	10.54533	10.40596	10.40596
0.96	13.59586	13.59586	13.11823	13.11823	12.9412	12.9412
0.97	18.05608	18.05608	17.40714	17.40714	17.16704	17.16704
0.98	26.97917	26.97917	25.98592	25.98592	25.61905	25.61905
0.99	53.75592	53.75592	51.72352	51.72352	50.97387	50.97387

โดยที่ ρ คือ ระดับความสัมพันธ์ของตัวแปรอิสระของตัวที่ 1 และ ตัวที่ 2

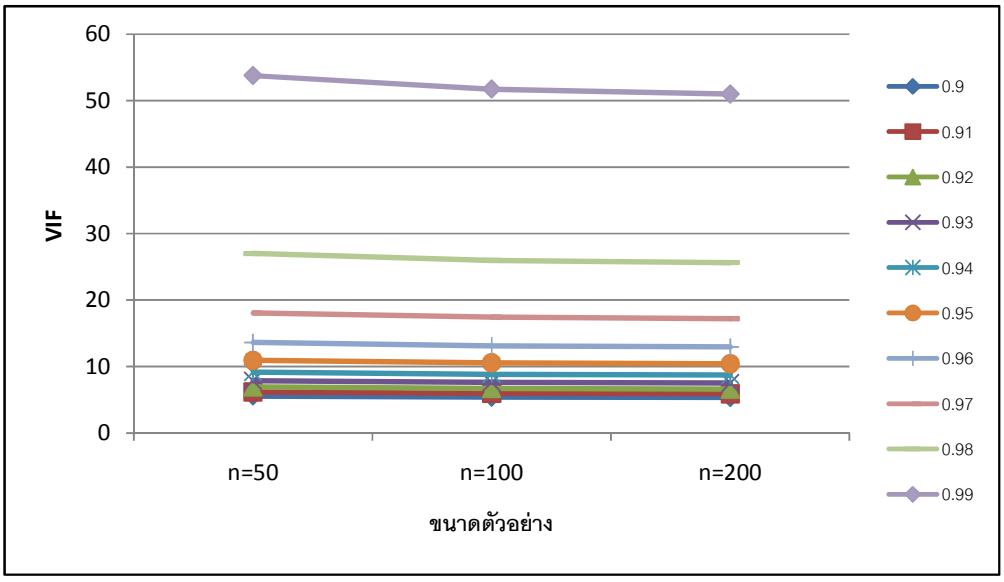
n คือ ขนาดตัวอย่าง

VIF คือ ค่าที่ใช้วัดในการเกิดปัญหาความสัมพันธ์ของตัวแปรอิสระ

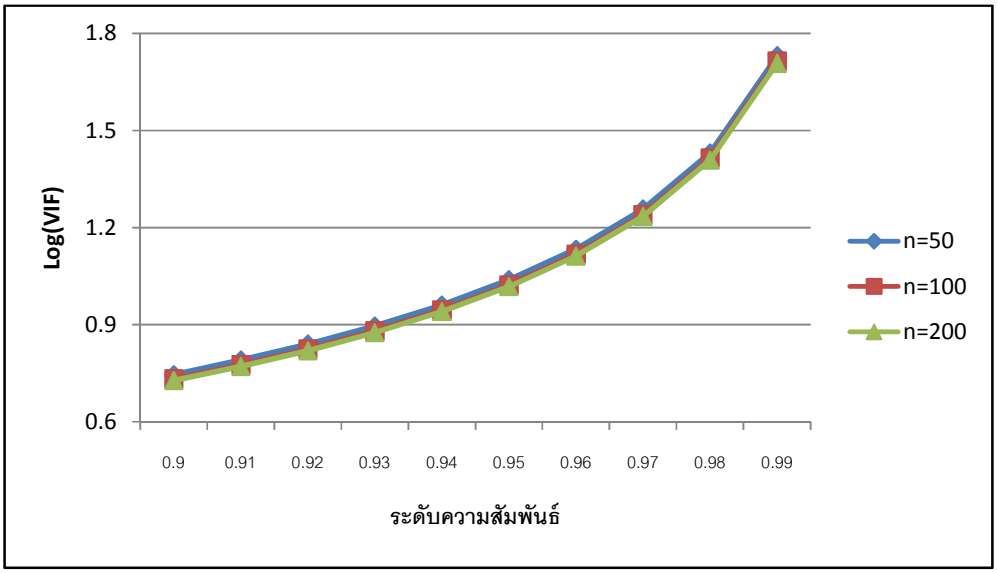
รูปที่ 4.1.1 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF ในแต่ละระดับความสัมพันธ์ เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



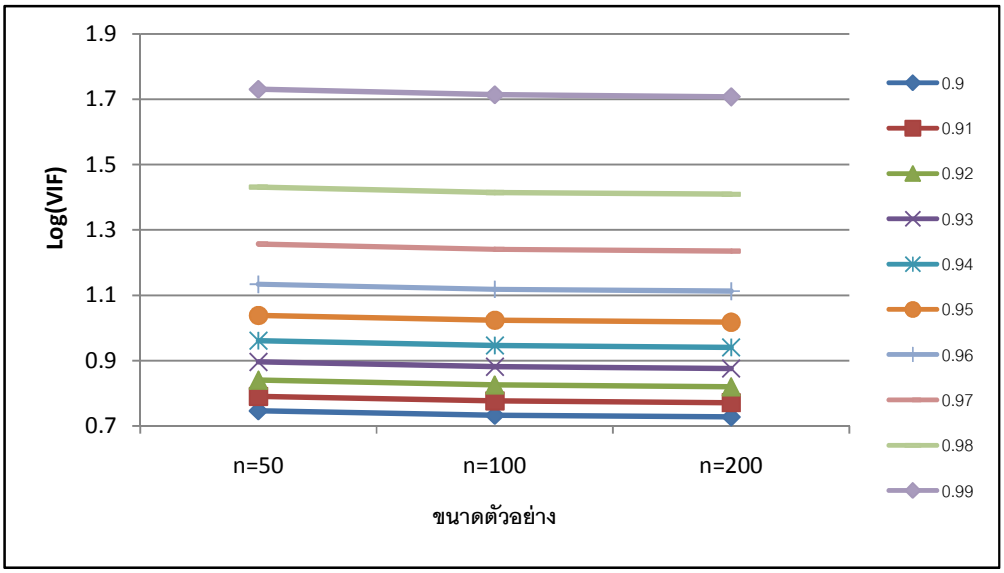
รูปที่ 4.1.2 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย (VIF) ในแต่ละขนาดตัวอย่าง เมื่อระดับความสัมพันธ์มีค่าตั้งแต่ 0.90 – 0.99



รูปที่ 4.1.3 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF) ในแต่ละระดับความสัมพันธ์ เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



รูปที่ 4.1.4 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF) ในแต่ละขนาดตัวอย่าง เมื่อระดับความสัมพันธ์มีค่าตั้งแต่ 0.90 – 0.99



จากตารางที่ 4.1.1 และกราฟที่ 4.1.1- 4.1.4 สามารถสรุปได้ว่า

แนวโน้มค่าเฉลี่ย VIF สามารถพิจารณาได้ 2 ลักษณะ คือ

1) พิจารณาจากระดับความสัมพันธ์ เมื่อขนาดตัวอย่างคงที่ พบว่าระดับความสัมพันธ์เท่ากับ 0.90 มีค่าเฉลี่ย VIF น้อยที่สุด แล้วค่าระดับความสัมพันธ์ลำดับถัดไปที่มีค่าเฉลี่ย VIF มากกว่า คือ 0.91 และค่าเฉลี่ย VIF จะมีค่าเพิ่มขึ้นไปเรื่อยๆจนถึงระดับความสัมพันธ์ 0.99 มีค่าเฉลี่ย VIF มากที่สุด

2) พิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าขนาดตัวอย่างเท่ากับ 50 มีค่าเฉลี่ย VIF มากที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าเฉลี่ย VIF น้อยกว่าคือ 100 และขนาดตัวอย่างที่มีค่าเฉลี่ย VIF น้อยที่สุดคือ 200

ดังนั้นสามารถสรุปแนวโน้มค่าเฉลี่ย VIF ได้ว่าเมื่อพิจารณาจากระดับความสัมพันธ์ พบว่าถ้าระดับความสัมพันธ์มีค่าเพิ่มขึ้น แล้วค่าเฉลี่ย VIF จะมีแนวโน้มเพิ่มขึ้น และเมื่อพิจารณาจากขนาดตัวอย่าง พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้น แล้วค่าเฉลี่ย VIF จะมีแนวโน้มลดลง

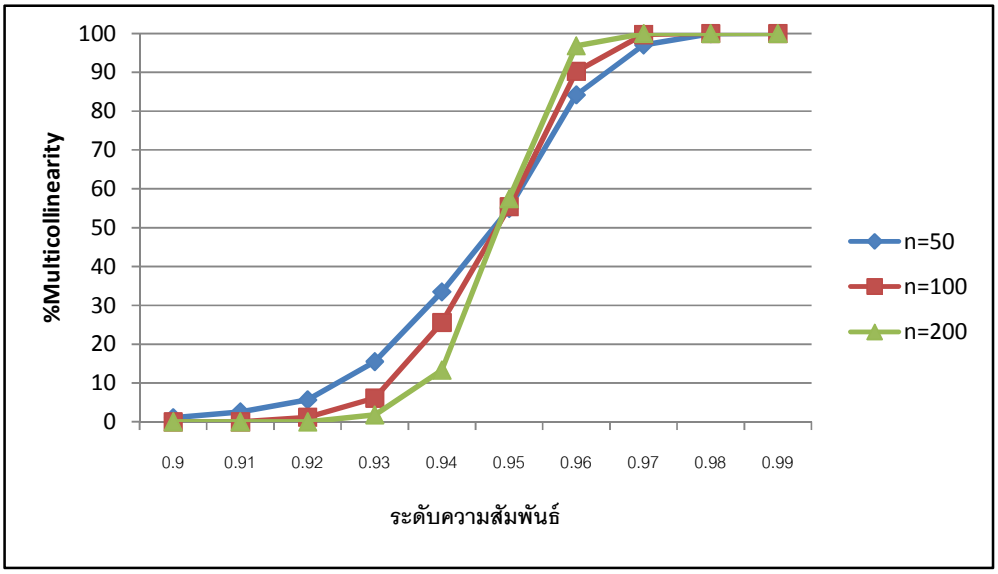
จากตารางแสดงค่าเฉลี่ย VIF จำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์ พบว่าระดับความสัมพันธ์ตั้งแต่ 0.95 ขึ้นไป ของทุกๆขนาดตัวอย่างที่ได้ศึกษา จะมีค่าเฉลี่ย VIF มากกว่า 10 แสดงว่าค่าเฉลี่ย VIFในระดับความสัมพันธ์ดังกล่าว เกิดปัญหาพหุสัมพันธ์รุนแรง ดังนั้น ที่

ระดับความสัมพันธ์ 0.95 คือจุดเปลี่ยนที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง

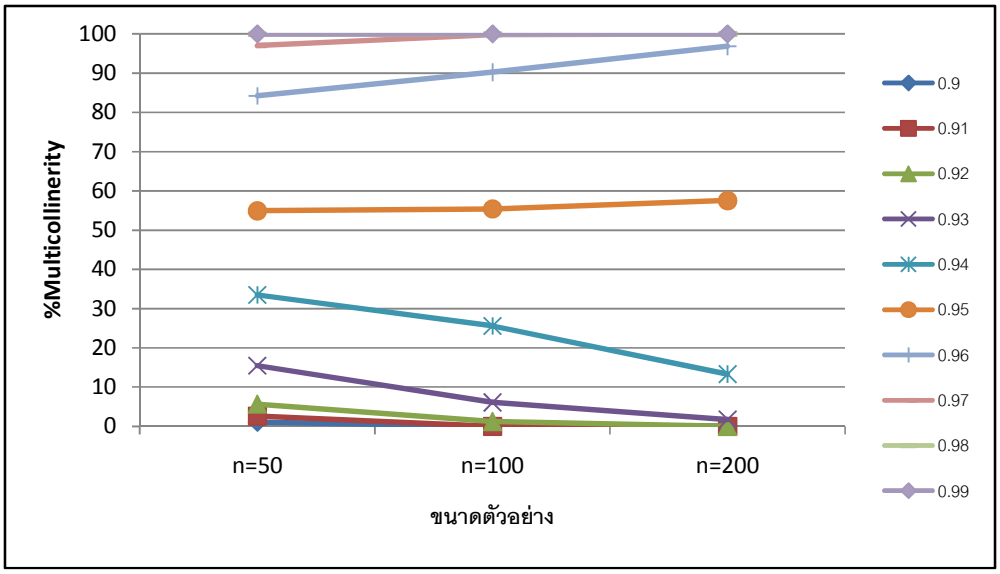
ตารางที่ 4.1.2 แสดงค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

rho	0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
n=50	1.1	2.6	5.7	15.5	33.5	54.9	84.2	97	99.9	100
n=100	0	0	1.2	6.1	25.6	55.4	90.2	99.7	100	100
n=200	0	0	0	1.8	13.3	57.5	96.8	100	100	100

รูปที่ 4.1.5 กราฟแสดงการเปรียบเทียบค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์ เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



รูปที่ 4.1.6 กราฟแสดงการเปรียบเทียบค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ในแต่ละขนาดตัวอย่าง เมื่อระดับความสัมพันธ์มีค่าตั้งแต่ 0.90 – 0.99



จากตารางที่ 4.1.2 และกราฟที่ 4.1.5- 4.1.6 สามารถสรุปได้ว่า

แนวโน้มค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงสามารถพิจารณาได้2ลักษณะ คือ

1. พิจารณาจากระดับความสัมพันธ์ เมื่อขนาดตัวอย่างคงที่ พบว่าระดับความสัมพันธ์เท่ากับ 0.90 จะมีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงน้อยที่สุด แล้วค่าระดับความสัมพันธ์ลำดับถัดไปที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากกว่าคือ 0.91 และค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีค่าเพิ่มขึ้นไปเรื่อยๆจนถึงระดับความสัมพันธ์ 0.99 มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากที่สุด

2. พิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าสามารถแบ่งแนวโน้มออกเป็น 3 ลักษณะ คือ

2.1) ระดับความสัมพันธ์ 0.90 – 0.94 พบว่าขนาดตัวอย่าง เท่ากับ 50 ค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมีค่ามากที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงน้อยกว่าคือ 100 และขนาดตัวอย่างที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงน้อยที่สุด คือ 200

2.2) ระดับความสัมพันธ์ 0.95 – 0.98 พบว่าขนาดตัวอย่างเท่ากับ 50 ค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมีค่าน้อยที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าร้อยละ

ของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากกว่าคือ 100 และขนาดตัวอย่างที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากที่สุดคือ 200

2.3) ระดับความสัมพันธ์ 0.99 พบว่าทุกๆขนาดตัวอย่างที่ทำการศึกษา จะมีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงเท่ากับ 100

ดังนั้นสามารถสรุปแนวโน้มค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ได้ว่า เมื่อพิจารณาจากระดับความสัมพันธ์ พบว่าถ้าระดับความสัมพันธ์มีค่าเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มเพิ่มขึ้น ซึ่งมีลักษณะคล้ายตัว S และเมื่อพิจารณาจากขนาดตัวอย่างสามารถแบ่งแนวโน้มออกเป็น 3 ลักษณะ คือ 1) ระดับความสัมพันธ์ 0.90 – 0.94 พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มลดลง 2) ระดับความสัมพันธ์ 0.95 – 0.99 พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มเพิ่มขึ้น 3) ระดับความสัมพันธ์ 0.99 พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มคงที่

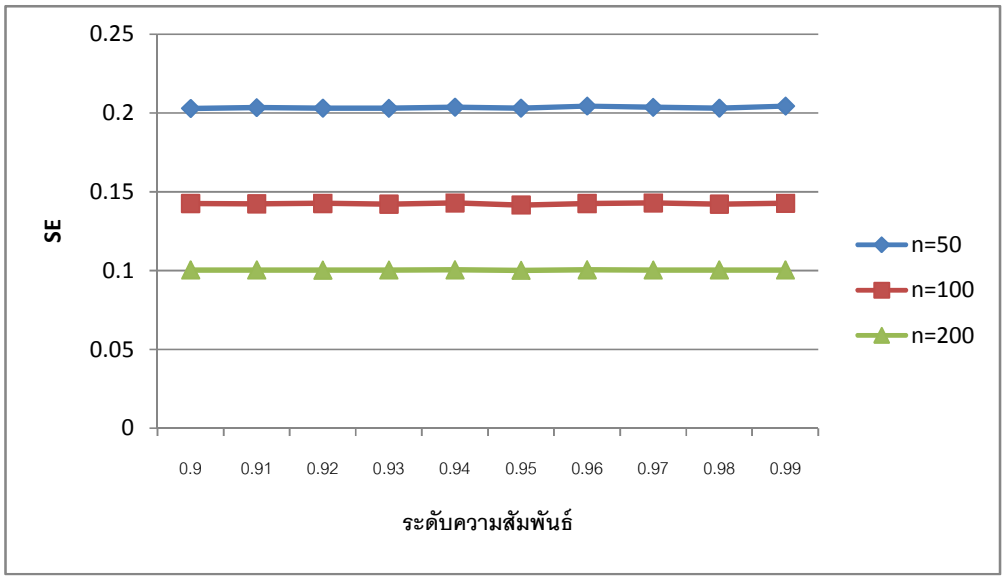
ตารางที่ 4.1.3 แสดงค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยจำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์

rho	n=50		n=100		n=200	
	X_OLS	PC_OLS	X_OLS	PC_OLS	X_OLS	PC_OLS
0.9	0.202743	0.198464	0.142525	0.141009	0.100362	0.099827
0.91	0.203281	0.198763	0.14231	0.140743	0.100336	0.099815
0.92	0.20295	0.198677	0.142585	0.14113	0.100257	0.099745
0.93	0.202859	0.198678	0.142161	0.140747	0.100322	0.099821
0.94	0.203589	0.199487	0.142843	0.141388	0.100432	0.099939
0.95	0.203002	0.198734	0.141547	0.140141	0.100113	0.099624
0.96	0.204248	0.199753	0.142565	0.141061	0.100464	0.099957
0.97	0.203481	0.199239	0.142857	0.141415	0.10032	0.099826
0.98	0.202979	0.198743	0.142182	0.140733	0.100354	0.099828
0.99	0.204233	0.20001	0.142678	0.141225	0.100357	0.099842

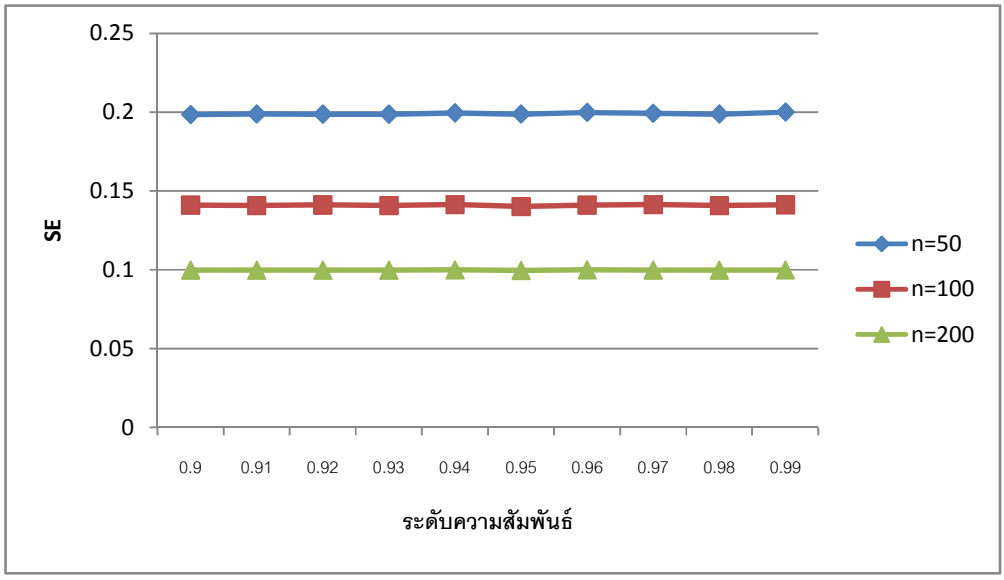
โดยที่ X_OLS คือ ค่าก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรง

PC_OLS คือ ค่าหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก

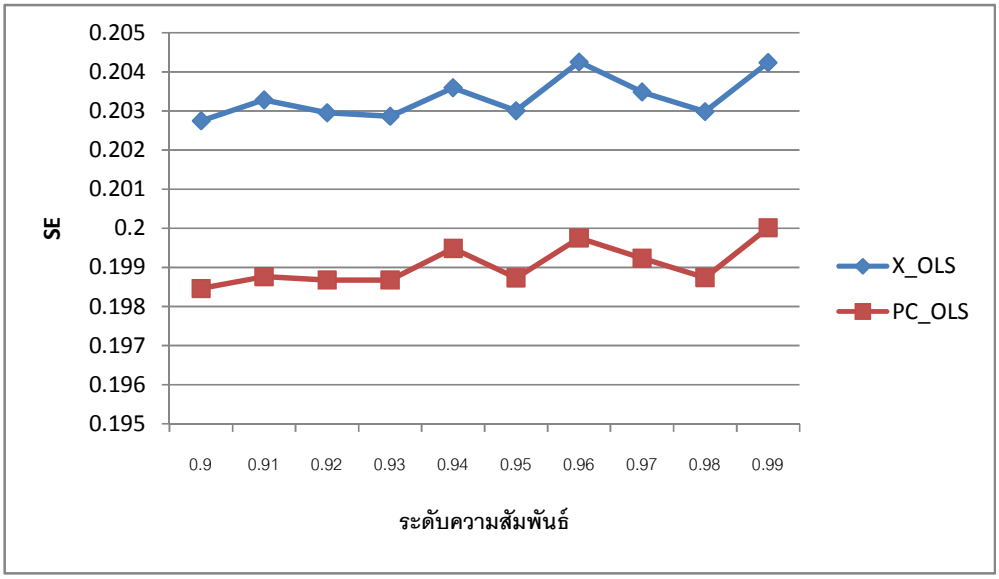
รูปที่ 4.1.7 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย ก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์ เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



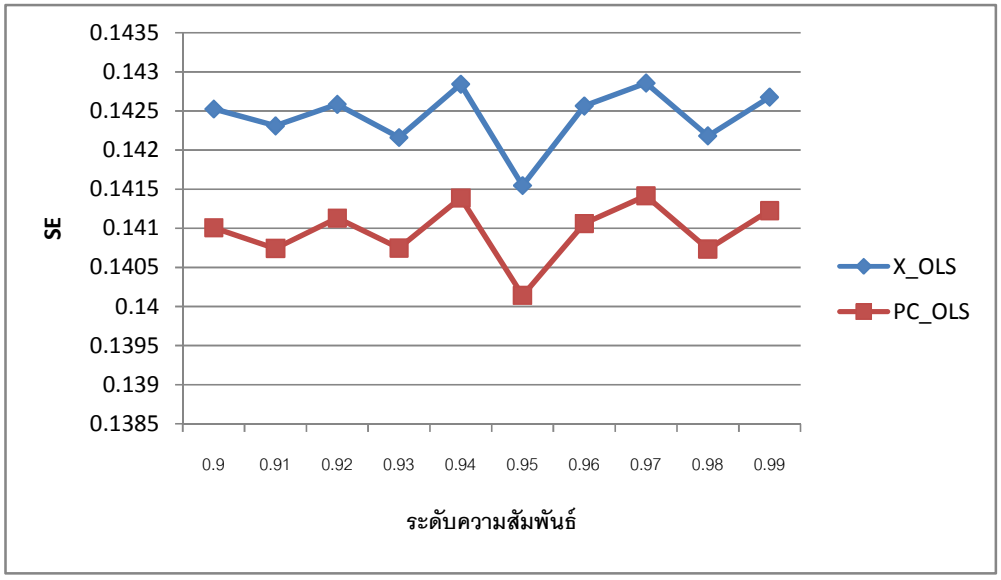
รูปที่ 4.1.8 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย หลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



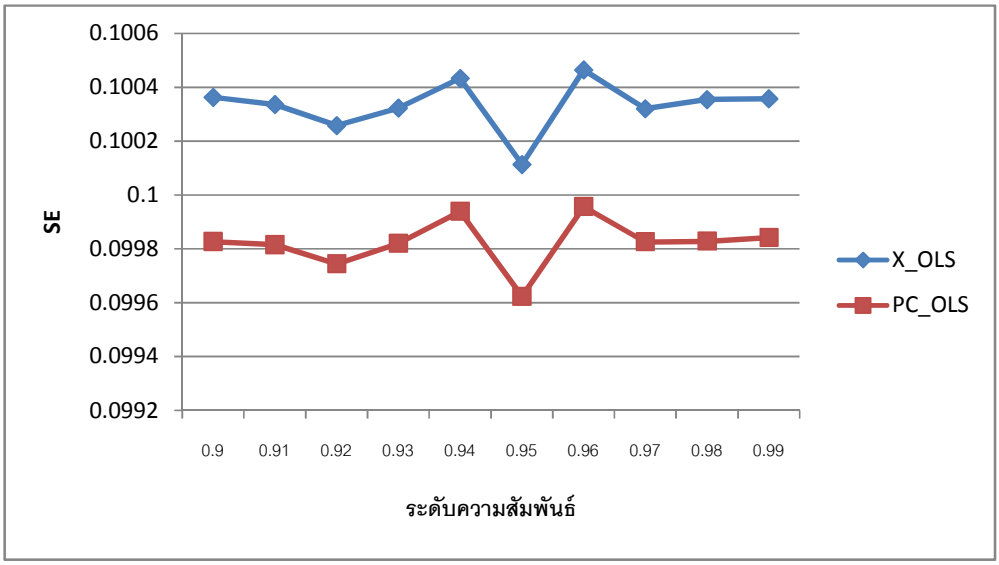
รูปที่ 4.1.9 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50



รูปที่ 4.1.10 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 100



รูปที่ 4.1.11 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 200



จากตารางที่ 4.1.3 และกราฟที่ 4.1.7- 4.1.11 สามารถสรุปได้ว่า

แนวโน้มค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย สามารถพิจารณาได้ 2 ลักษณะ คือ

1. พิจารณาจากระดับความสัมพันธ์ เมื่อขนาดตัวอย่างคงที่ พบว่าเมื่อระดับความสัมพันธ์มีค่าเพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีค่าใกล้เคียงกัน

2. พิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานมากที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานน้อยกว่าคือ 100 และขนาดตัวอย่างที่มีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานน้อยที่สุดคือ 200 ทั้งก่อนแก้ไขปัญหาพหุสัมพันธ์รุนแรงและหลังแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก

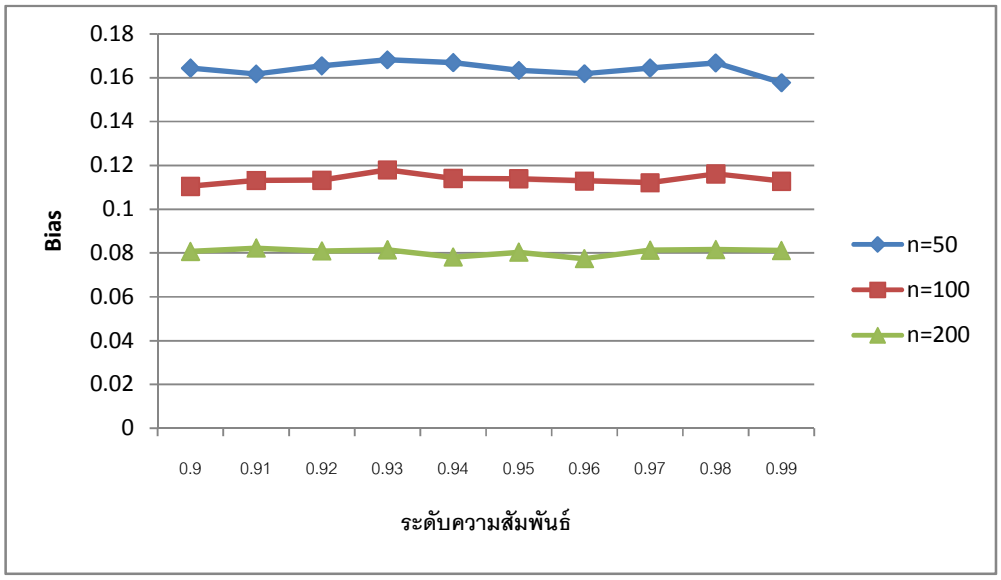
ดังนั้นสามารถสรุปแนวโน้มค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย ได้ว่าเมื่อพิจารณาจากระดับความสัมพันธ์ พบว่าถ้าระดับความสัมพันธ์เพิ่มขึ้นแล้วค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีแนวโน้มค่อนข้างคงที่ และเมื่อพิจารณาจากขนาดตัวอย่าง พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานมีแนวโน้มลดลง

จากการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงเมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงจะมีค่ามากกว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ทุกๆขนาดตัวอย่างที่ได้ศึกษา

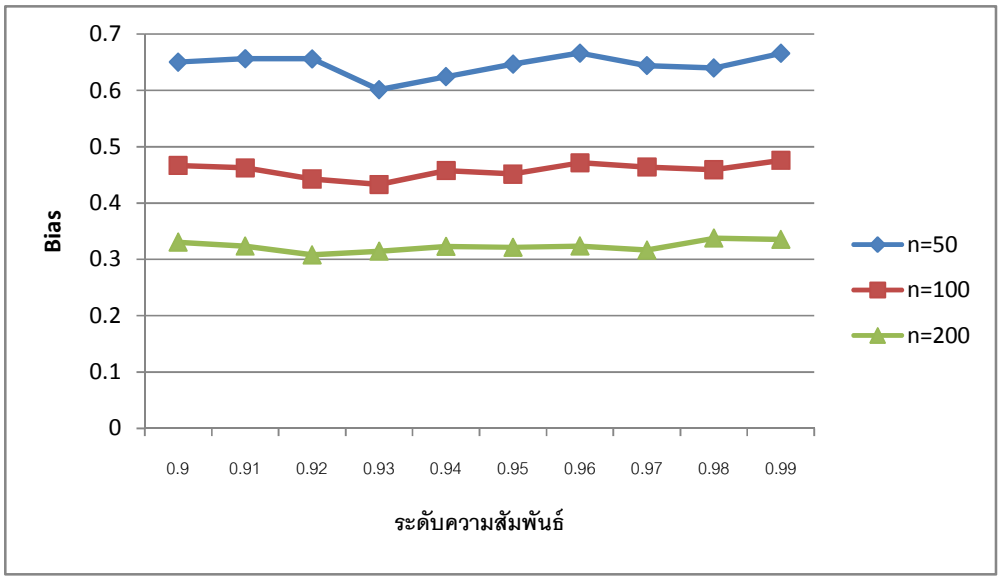
ตารางที่ 4.1.4 แสดงค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยจำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

rho	n=50		n=100		n=200	
	X_OLS	PC_OLS	X_OLS	PC_OLS	X_OLS	PC_OLS
0.9	0.164405	0.649746	0.110427	0.46643	0.080678	0.330397
0.91	0.161738	0.655774	0.113115	0.462346	0.082252	0.3235
0.92	0.165485	0.655611	0.113231	0.442705	0.080898	0.307767
0.93	0.168163	0.601078	0.117915	0.4329	0.081423	0.314368
0.94	0.166884	0.62408	0.114055	0.457528	0.0781	0.322861
0.95	0.163345	0.646207	0.113889	0.451345	0.080344	0.321146
0.96	0.161867	0.665853	0.112908	0.471443	0.077408	0.323584
0.97	0.164483	0.643798	0.112054	0.463969	0.081238	0.316281
0.98	0.166825	0.639513	0.116085	0.458867	0.081522	0.33751
0.99	0.157761	0.66558	0.112766	0.475605	0.081083	0.335154

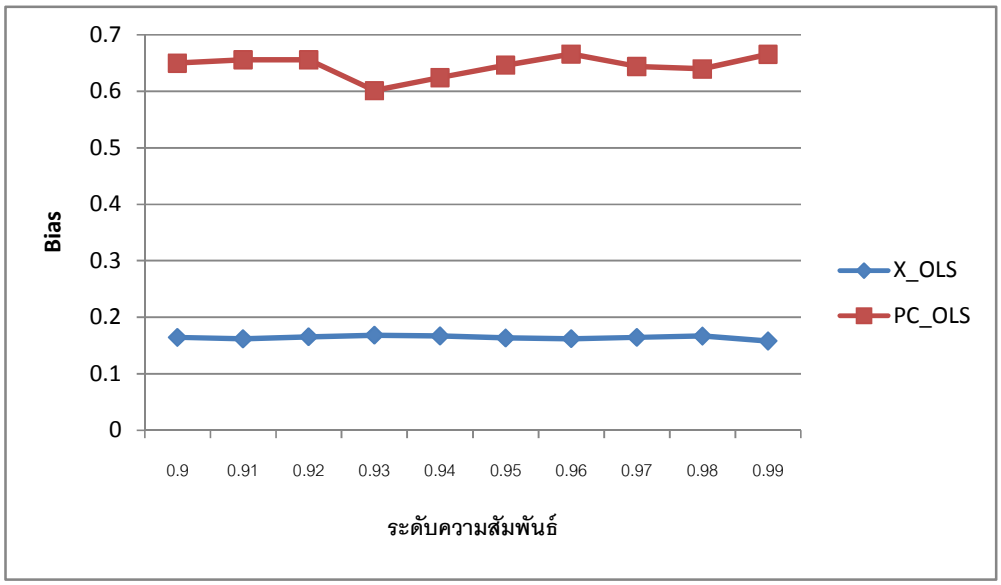
รูปที่ 4.1.12 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย ก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



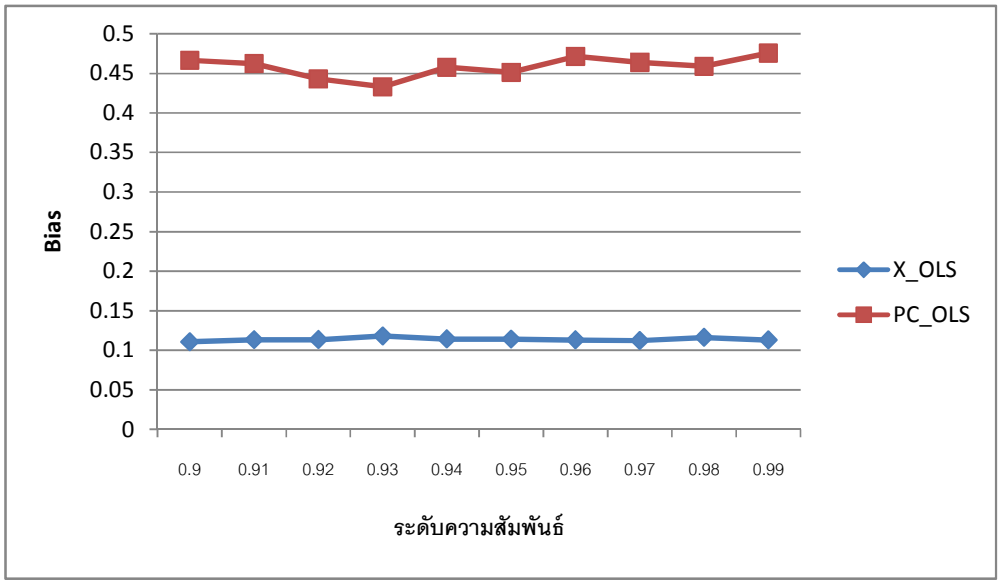
รูปที่4.1.13 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย หลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



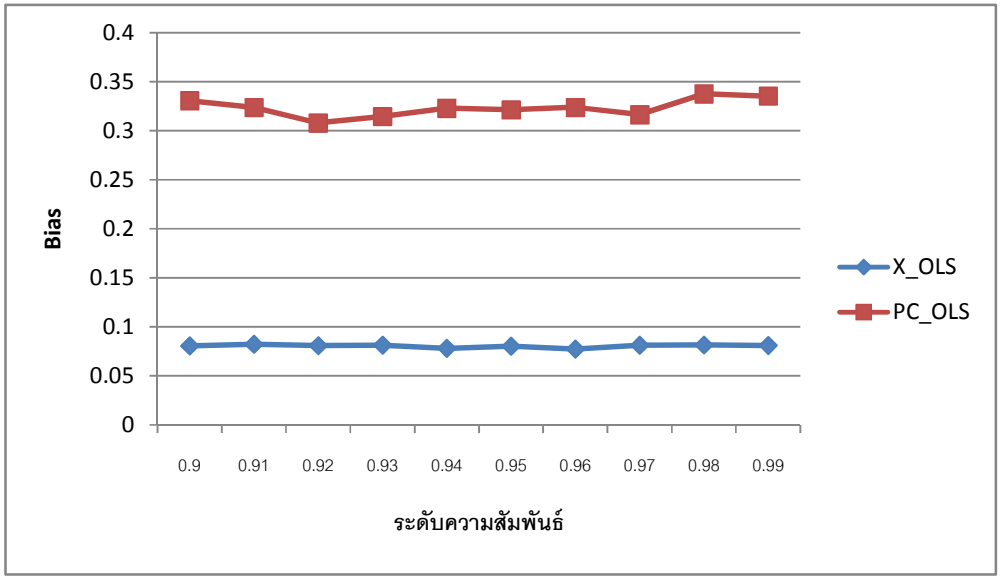
รูปที่ 4.1.14 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย ก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50



รูปที่ 4.1.15 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย ก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 100



รูปที่ 4.1.16 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย ก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 200



จากตารางที่ 4.1.4 และกราฟที่ 4.1.12- 4.1.16 สามารถสรุปได้ว่า

แนวโน้มค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย สามารถพิจารณาได้ 2 ลักษณะ คือ

1. พิจารณาจากระดับความสัมพันธ์ เมื่อขนาดตัวอย่างคงที่ พบว่าเมื่อระดับความสัมพันธ์มีค่าเพิ่มขึ้น ค่าเฉลี่ยความเอนเอียงจะมีค่าใกล้เคียงกัน

2. พิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าขนาดตัวอย่างเท่ากับ 50 มีค่าเฉลี่ยความเอนเอียงมากที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าเฉลี่ยความเอนเอียงน้อยกว่าคือ 100 และขนาดตัวอย่างที่มีค่าเฉลี่ยความเอนเอียงน้อยที่สุดคือ 200 ทั้งก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงและหลังแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก

ดังนั้นสามารถสรุปแนวโน้มค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย ได้ว่า เมื่อพิจารณาจากระดับความสัมพันธ์ พบว่าถ้าระดับความสัมพันธ์มีค่าเพิ่มขึ้นแล้วค่าเฉลี่ยความเอนเอียงจะมีแนวโน้มค่อนข้างคงที่ และเมื่อพิจารณาจากขนาดตัวอย่าง พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ยความเอนเอียงจะมีแนวโน้มลดลง

จากการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าค่าเฉลี่ยความเอนเอียงก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงมีค่าน้อยกว่าค่าเฉลี่ยความเอนเอียงหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ทุกๆขนาดตัวอย่างที่ได้ศึกษา

4.2 ผลการศึกษาในกรณีจำนวนตัวแปรอิสระ 3 ตัวแปร

จากการจำลองข้อมูล ในการศึกษากรณีจำนวนตัวแปรอิสระ 3 ตัว จะแสดงผลการศึกษาในแต่ละหัวข้อ โดยพิจารณาจากระดับความสัมพันธ์สูงสุดในแต่ละกรณีที่ทำการศึกษาได้ดังนี้

1. แสดงค่าเฉลี่ย VIF จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์ เพื่อหาค่าระดับความสัมพันธ์ที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง

2. แสดงค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

3. แสดงค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

4. แสดงค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

ตารางที่ 4.1.5 แสดงค่าเฉลี่ย VIF 1, VIF2, VIF3 จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

$\rho_{\max}$	rho	n=50			n=100			n=200		
		VIF1	VIF2	VIF3	VIF1	VIF2	VIF3	VIF1	VIF2	VIF3
0.91	0.91,0.91,0.91	8.189978	8.118997	8.068525	7.826317	7.844011	7.866966	7.685394	7.697018	7.730533
0.93	0.93,0.93,0.93	10.50157	10.41471	10.33464	10.02825	10.05254	10.07942	9.84591	9.860884	9.90604
0.95	0.91,0.93,0.95	8.506994	11.63423	14.76422	8.132379	11.31425	14.42042	7.986516	11.10853	14.17754
0.95	0.95,0.95,0.95	14.66289	14.54924	14.41254	13.99242	14.02897	14.06194	13.73563	13.75656	13.82273
0.97	0.91,0.93,0.97	8.119236	18.24984	23.21949	7.758066	17.83089	22.70518	7.618246	17.52	22.33131
0.97	0.91,0.95,0.97	11.39445	18.53091	32.6629	10.92887	18.10757	31.94096	10.73822	17.79228	31.41964
0.97	0.93,0.95,0.97	11.30733	18.38111	25.41152	10.80704	17.90896	24.83722	10.6124	17.58661	24.42735
0.97	0.97,0.97,0.97	24.3721	24.20085	23.92301	23.24243	23.30865	23.35327	22.81226	22.84691	22.96225
0.99	0.91,0.93,0.99	8.41229	55.9686	71.34675	8.040915	54.92162	69.87723	7.896526	54.03493	68.77373
0.99	0.91,0.95,0.99	21.27883	102.8766	181.6914	20.53971	101.0123	177.9994	20.19662	99.42311	175.2372
0.99	0.93,0.95,0.99	11.83324	56.85003	78.87559	11.31598	55.74156	77.22348	11.11331	54.82637	76.00108
0.99	0.93,0.97,0.99	82.80208	244.0725	558.721	80.4799	239.6242	547.3579	79.19306	235.912	538.9419
0.99	0.95,0.97,0.99	20.20591	59.00559	97.36081	19.36481	57.77578	95.28133	19.02495	56.80241	93.76955
0.99	0.99,0.99,0.99	72.90427	72.4795	71.43435	69.48502	69.70765	69.79286	68.1906	68.29294	68.65519

ตารางที่ 4.1.6 แสดงค่าเฉลี่ย VIF 1, VIF2, VIF3 จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์สูงสุด

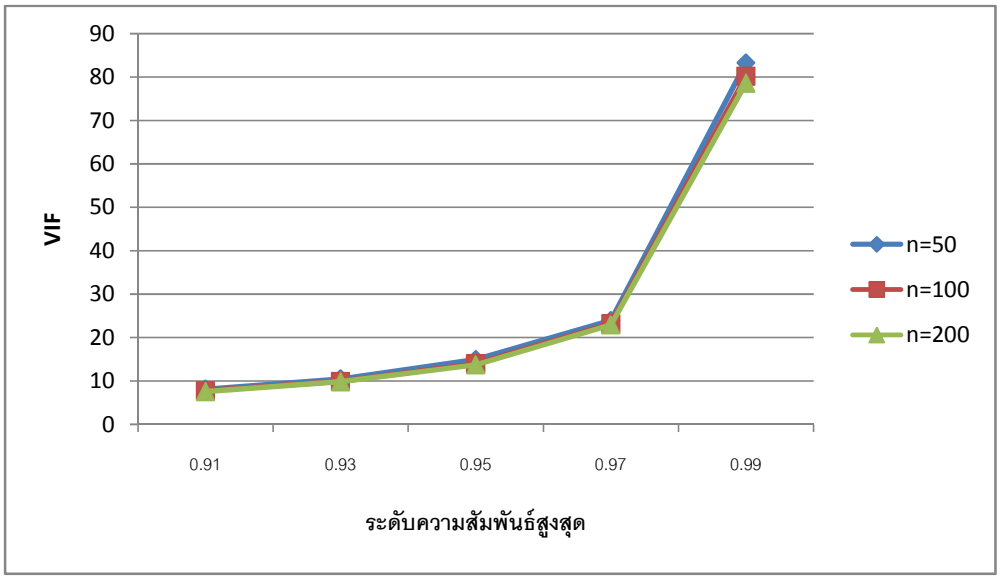
	$\rho_{\max}$	VIF1	VIF2	VIF3
n=50	0.91	8.189978	8.118997	8.068525
	0.93	10.50157	10.41471	10.33464
	0.95	14.66289	14.54924	14.76422
	0.97	24.3721	24.20085	32.6629
	0.99	82.80208	244.0725	558.721
n=100	0.91	7.826317	7.844011	7.866966
	0.93	10.02825	10.05254	10.07942
	0.95	13.99242	14.02897	14.42042
	0.97	23.24243	23.30865	31.94096
	0.99	80.4799	239.6242	547.3579
n=200	0.91	7.685394	7.697018	7.730533
	0.93	9.84591	9.860884	9.90604
	0.95	13.73563	13.75656	14.17754
	0.97	22.81226	22.84691	31.41964
	0.99	79.19306	235.912	538.9419

โดยที่ $\rho_{\max}$ คือ ค่าระดับความสัมพันธ์สูงสุด

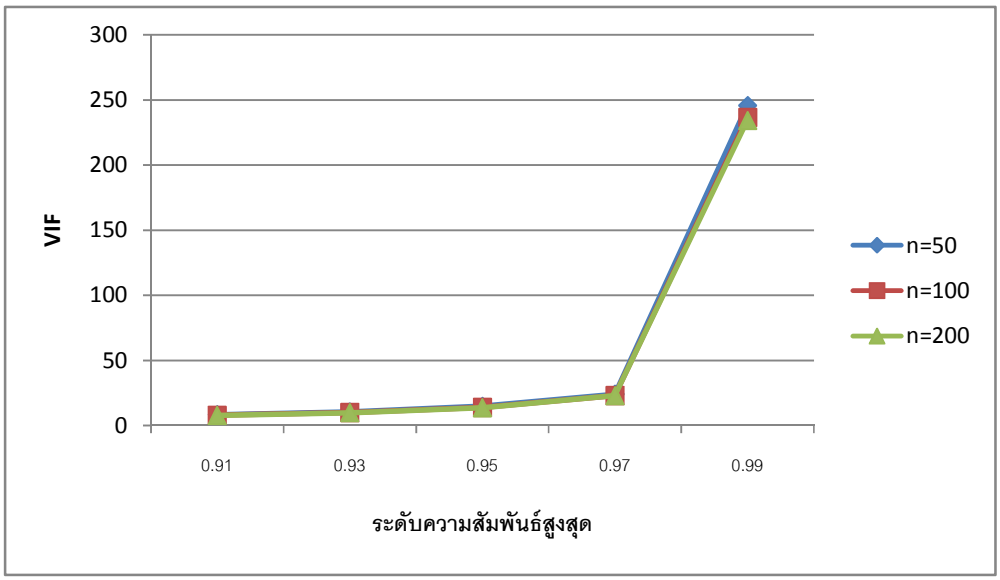
n คือ ขนาดตัวอย่าง

VIF คือ ค่าที่ใช้วัดในการเกิดปัญหาความสัมพันธ์ของตัวแปรอิสระ

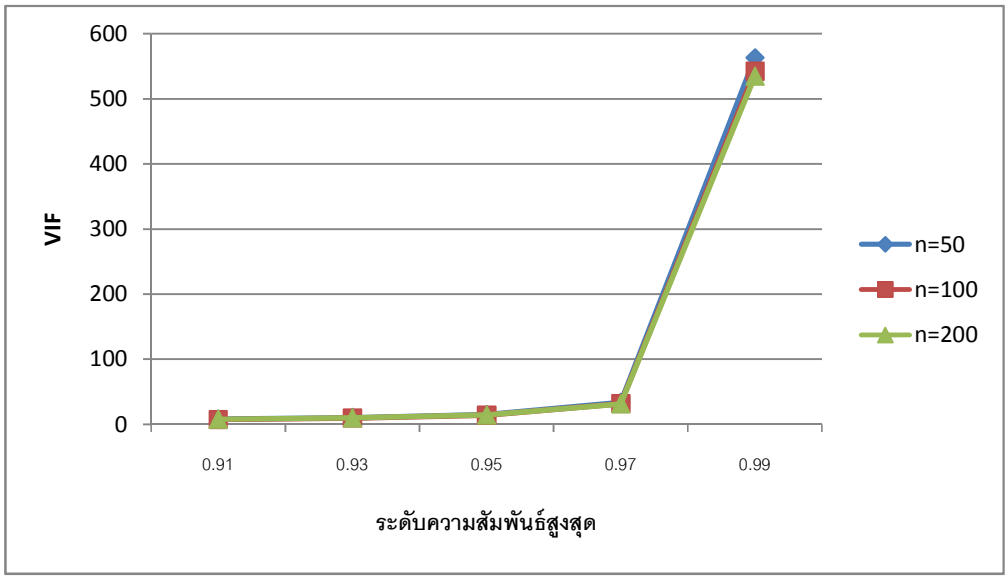
รูปที่ 4.1.17 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF 1 ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



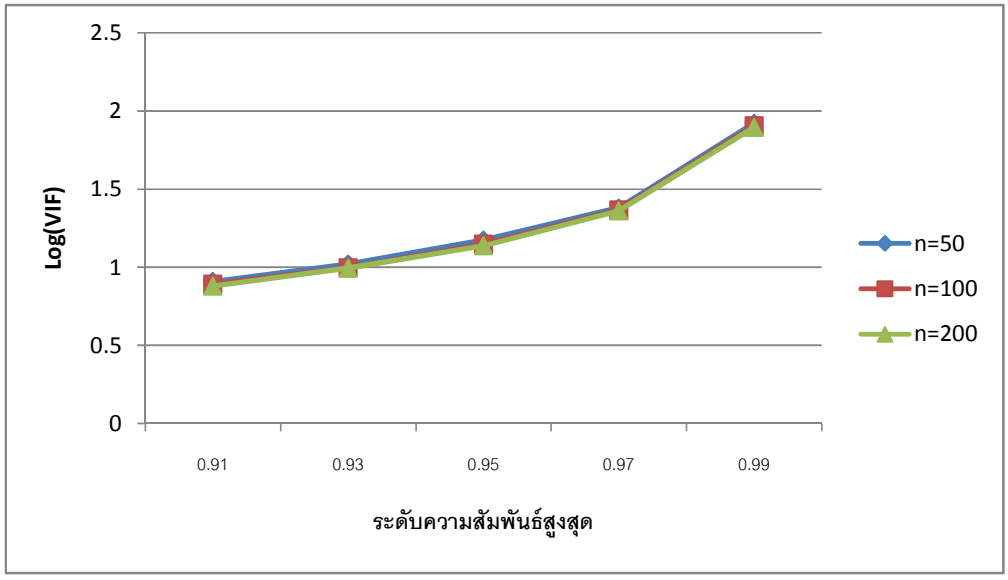
รูปที่ 4.1.18 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF 2 ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



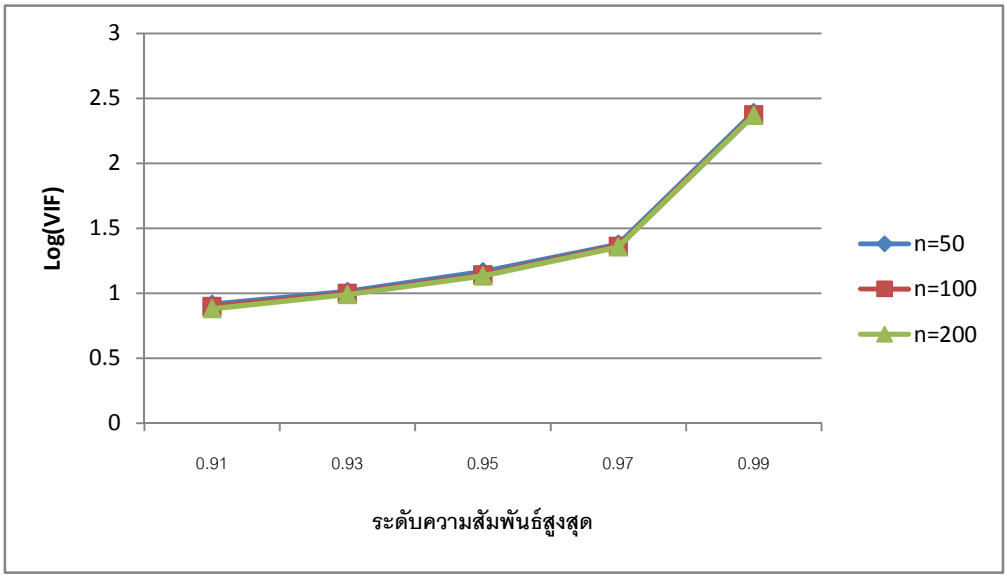
รูปที่ 4.1.19 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF 3 ในแต่ละระดับความสัมพันธ์สูงสุด
เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



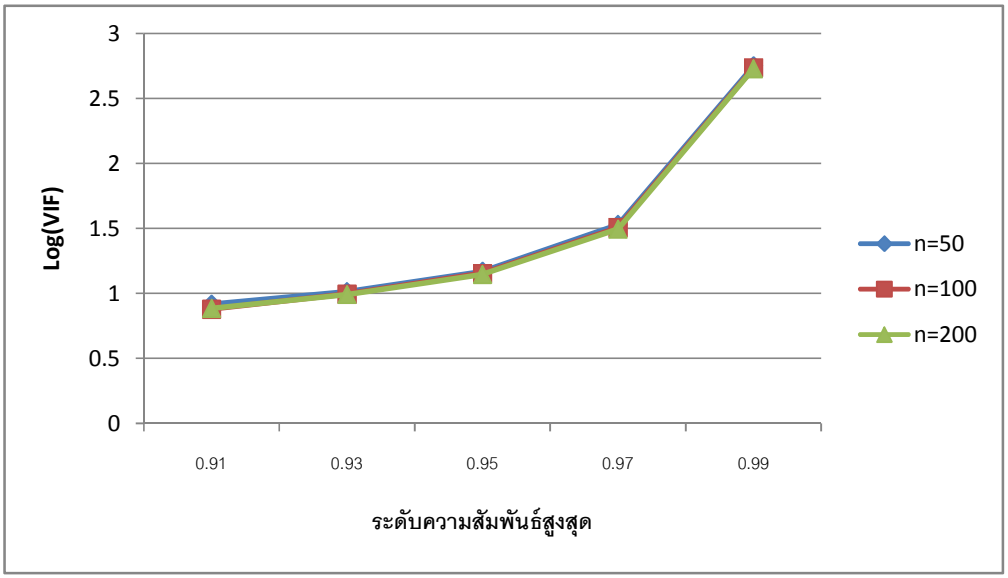
รูปที่ 4.1.20 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1) ในแต่ละระดับความสัมพันธ์สูงสุด
เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



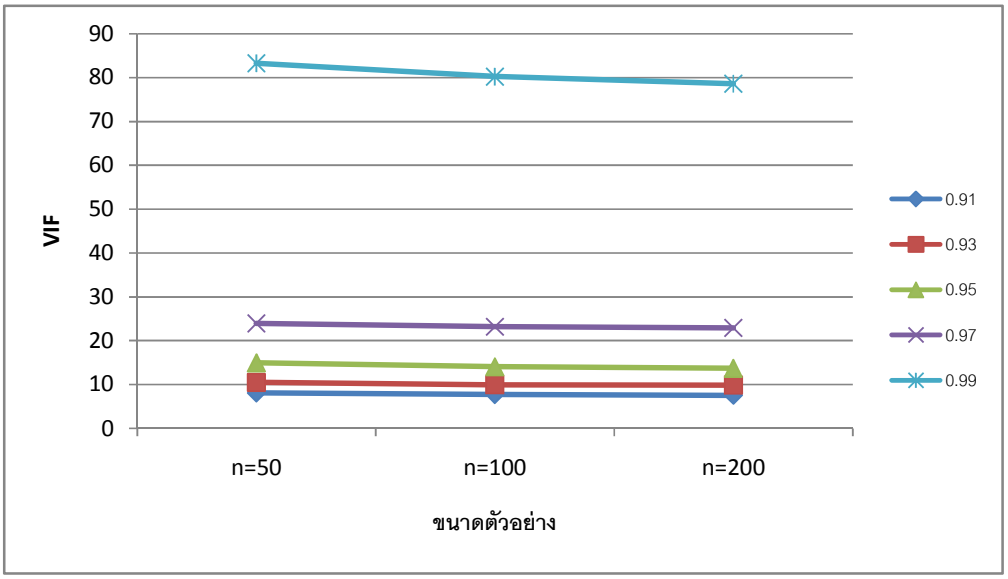
รูปที่ 4.1.21 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF2) ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



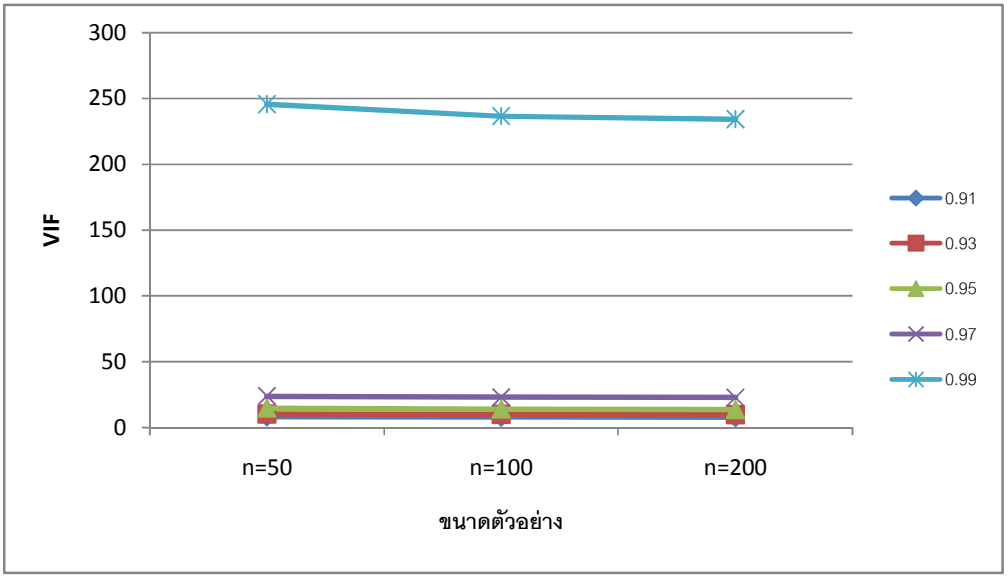
รูปที่ 4.1.22 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF3) ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



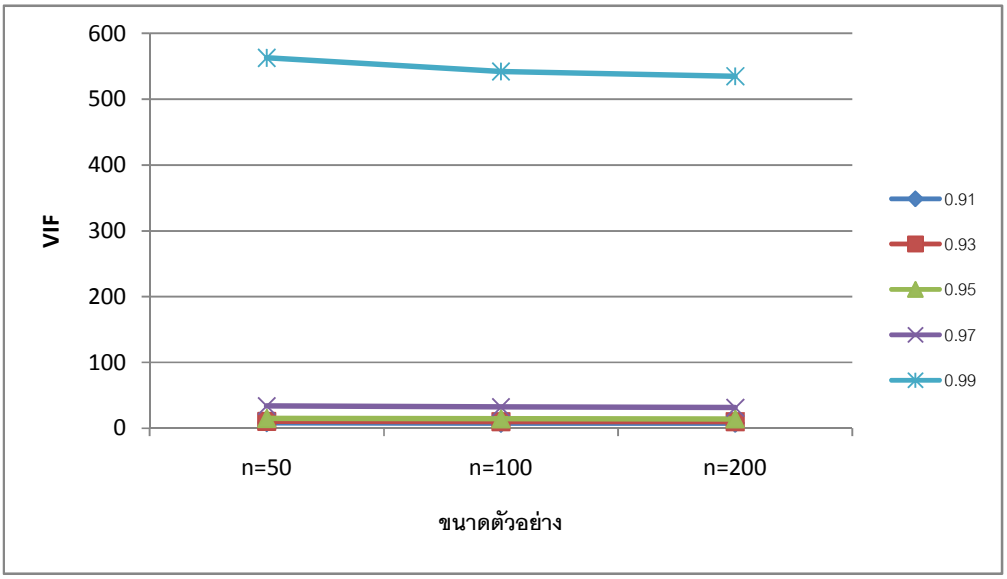
รูปที่ 4.1.23 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย (VIF1) ในแต่ละขนาดตัวอย่างเมื่อระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99



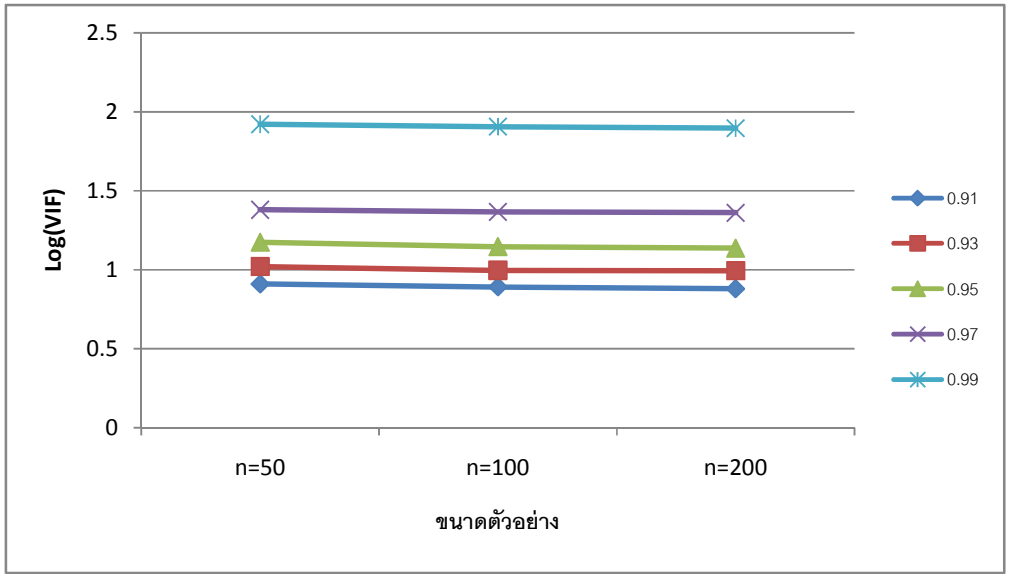
รูปที่ 4.1.24 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย (VIF2) ในแต่ละขนาดตัวอย่างเมื่อระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99



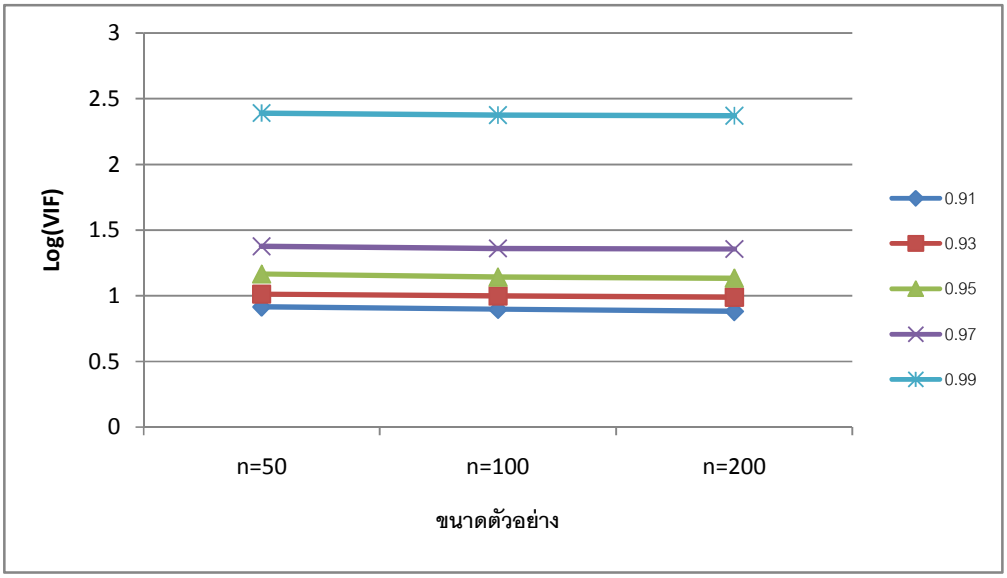
รูปที่ 4.1.25 กราฟแสดงการเปรียบเทียบค่า (VIF3) ในแต่ละขนาดตัวอย่างเมื่อระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99



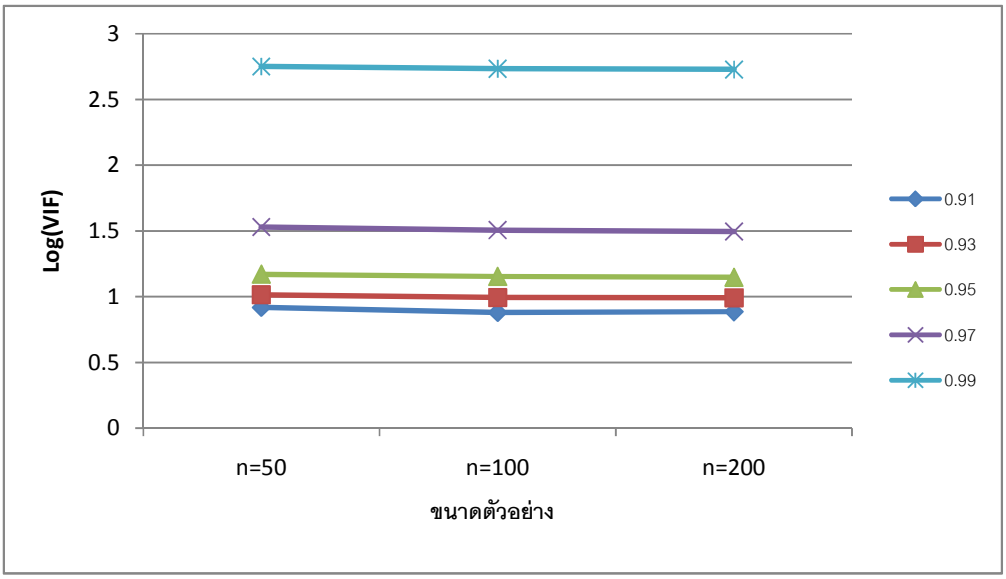
รูปที่ 4.1.26 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1) ในแต่ละขนาดตัวอย่างเมื่อระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99



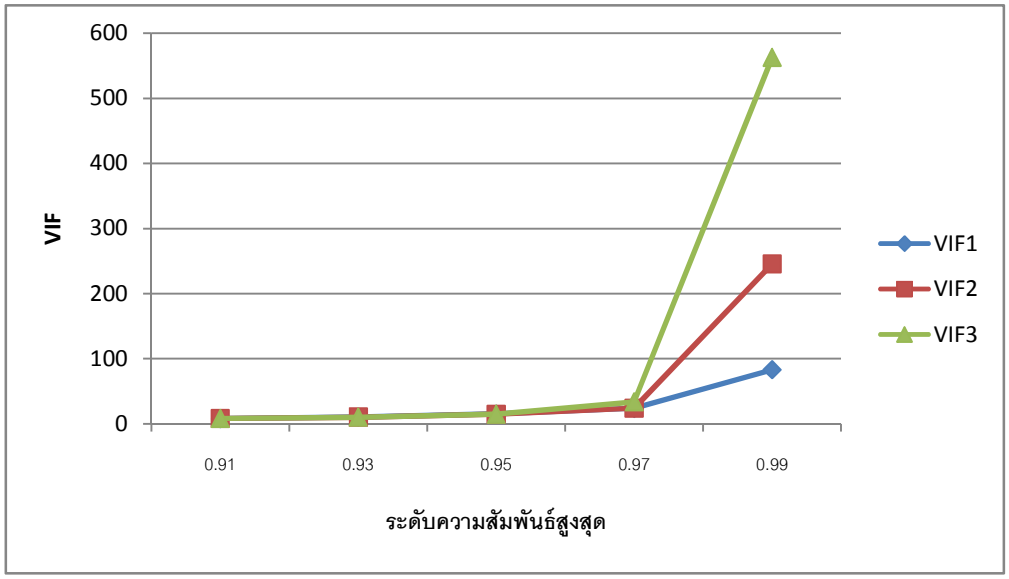
รูปที่ 4.1.27 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF2) ในแต่ละขนาดตัวอย่างเมื่อระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99



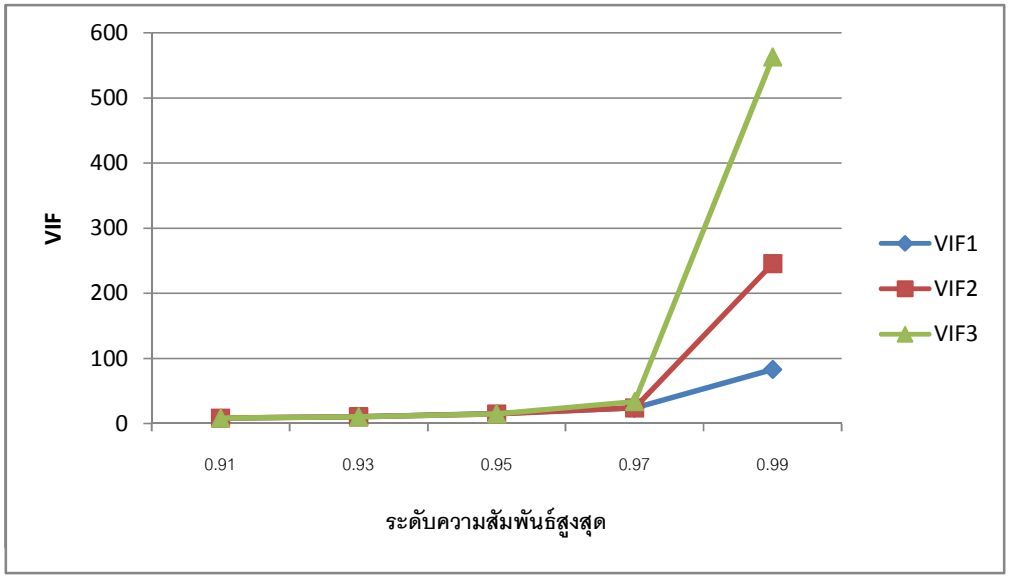
รูปที่ 4.1.28 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF3) ในแต่ละขนาดตัวอย่างเมื่อระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99



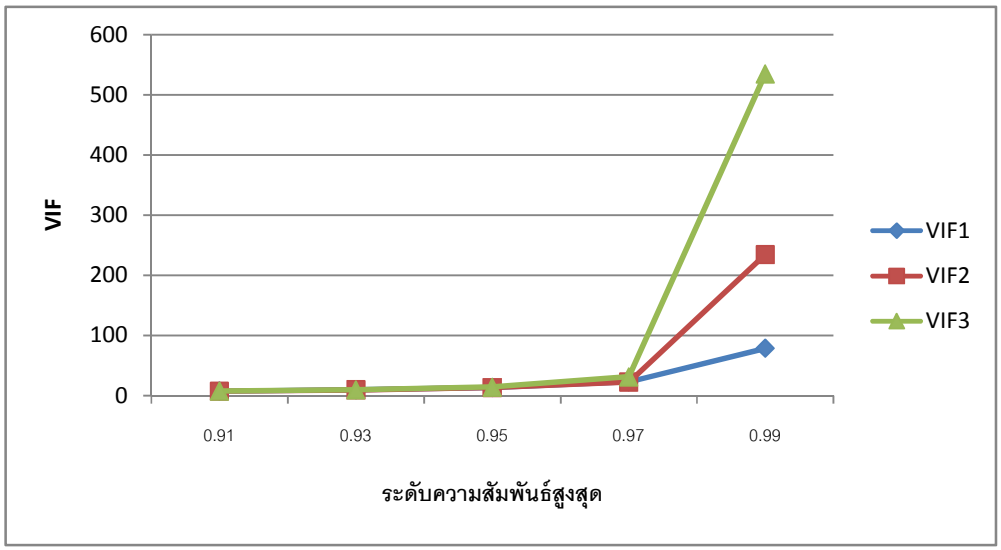
รูปที่ 4.1.29 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50



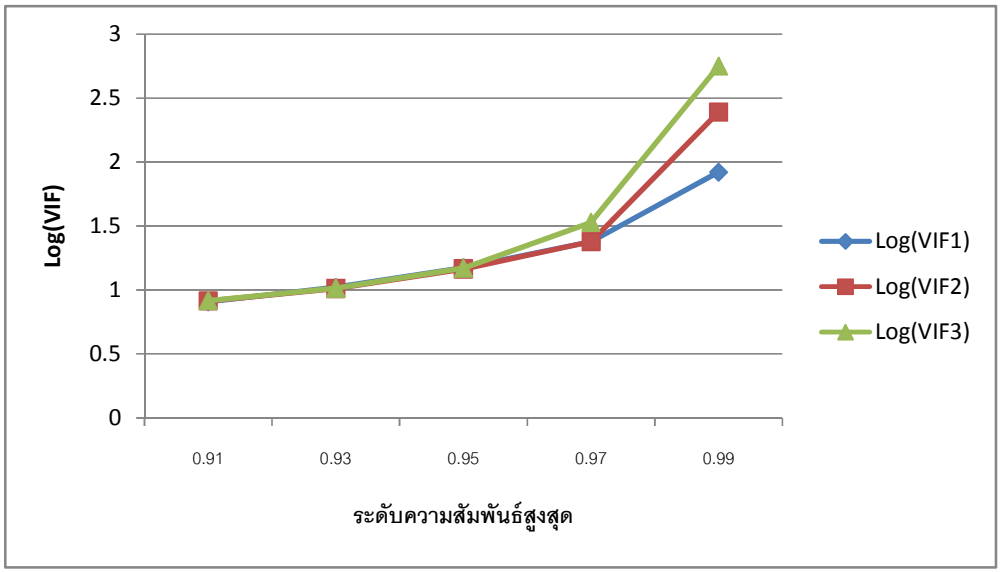
รูปที่ 4.1.30 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100



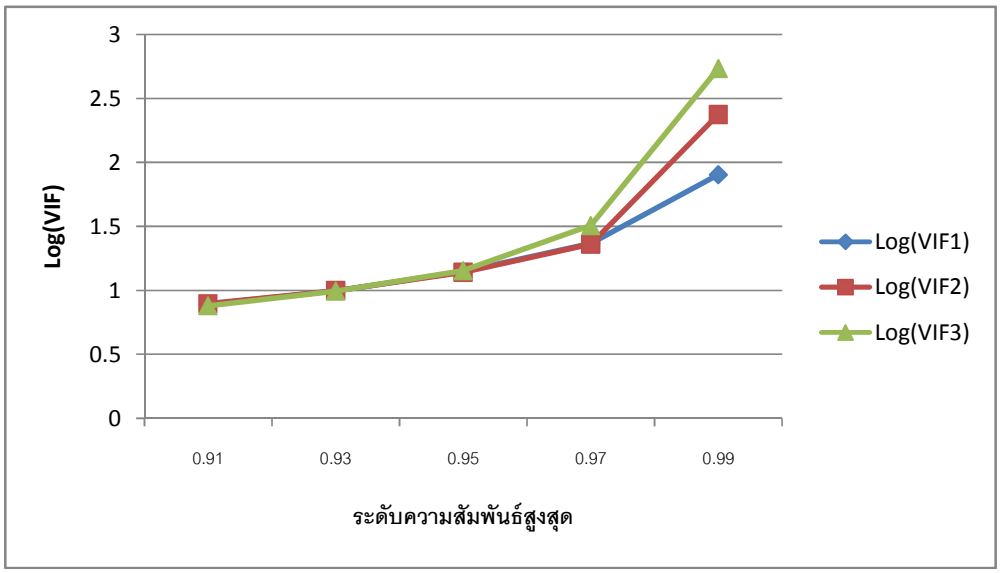
รูปที่ 4.1.31 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200



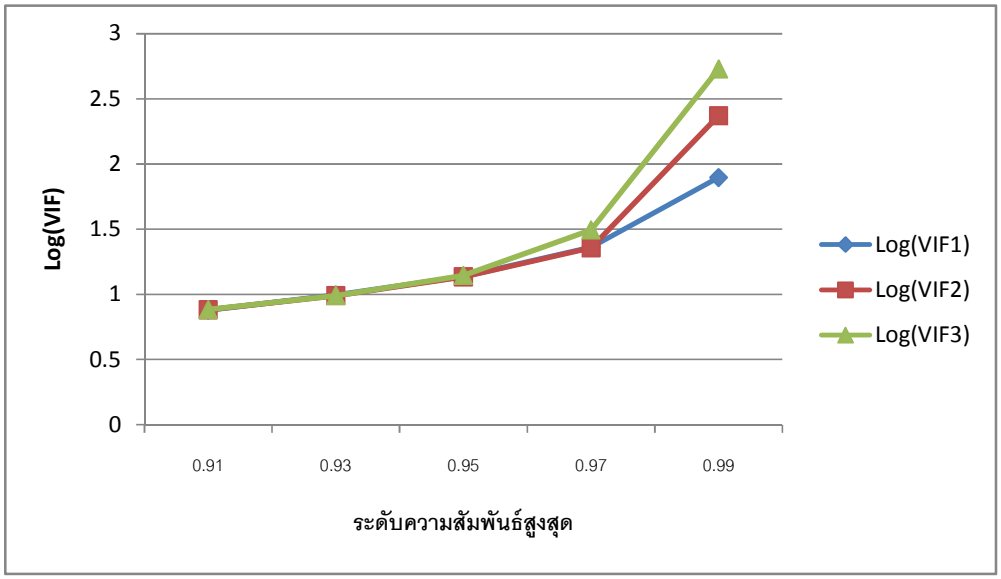
รูปที่ 4.1.32 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50



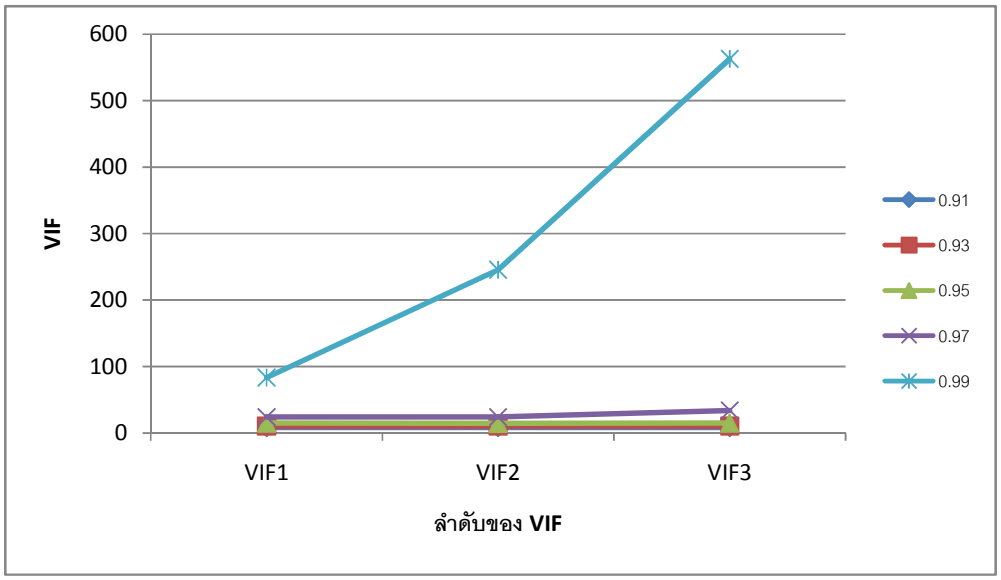
รูปที่ 4.1.33 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100



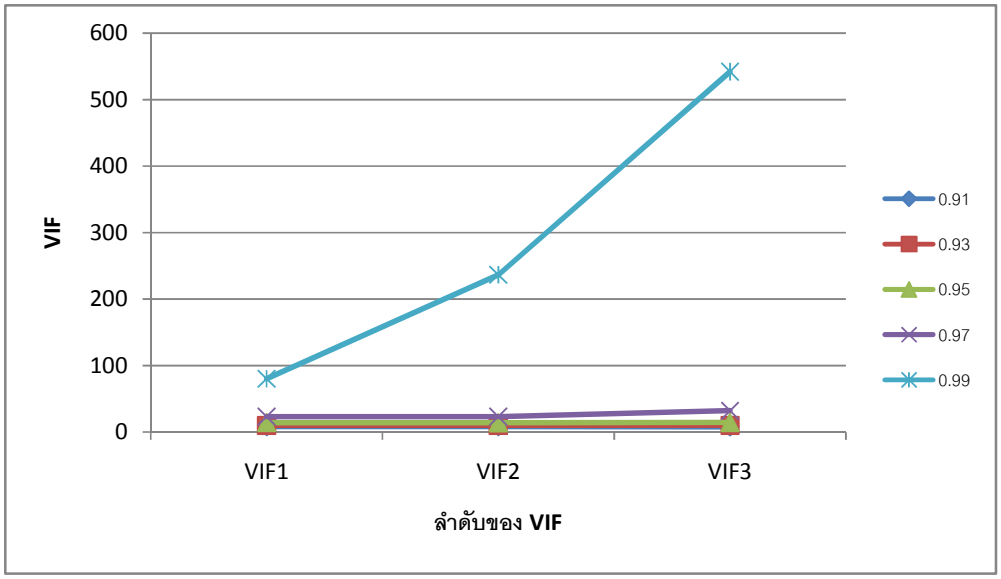
รูปที่ 4.1.34 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200



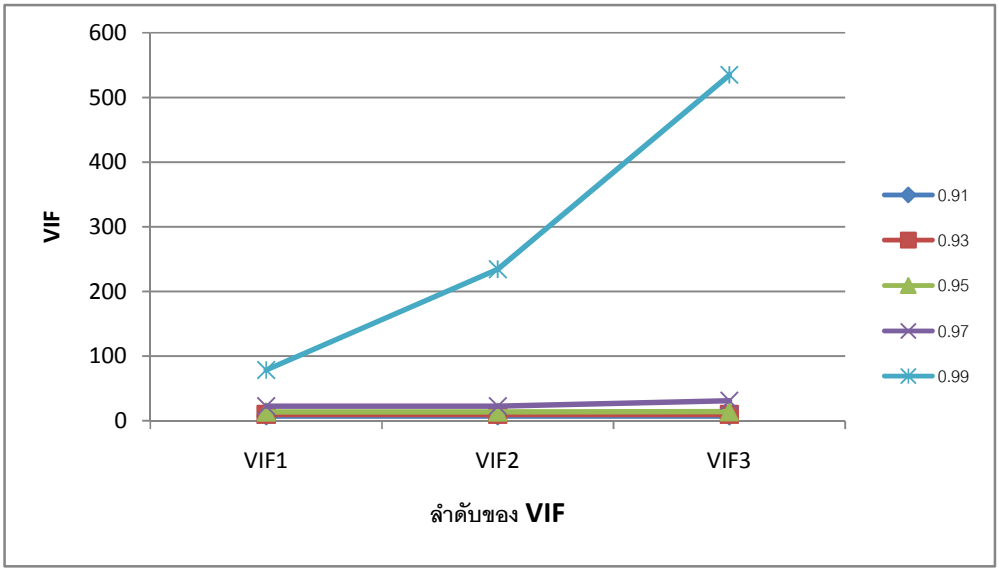
รูปที่ 4.1.35 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50



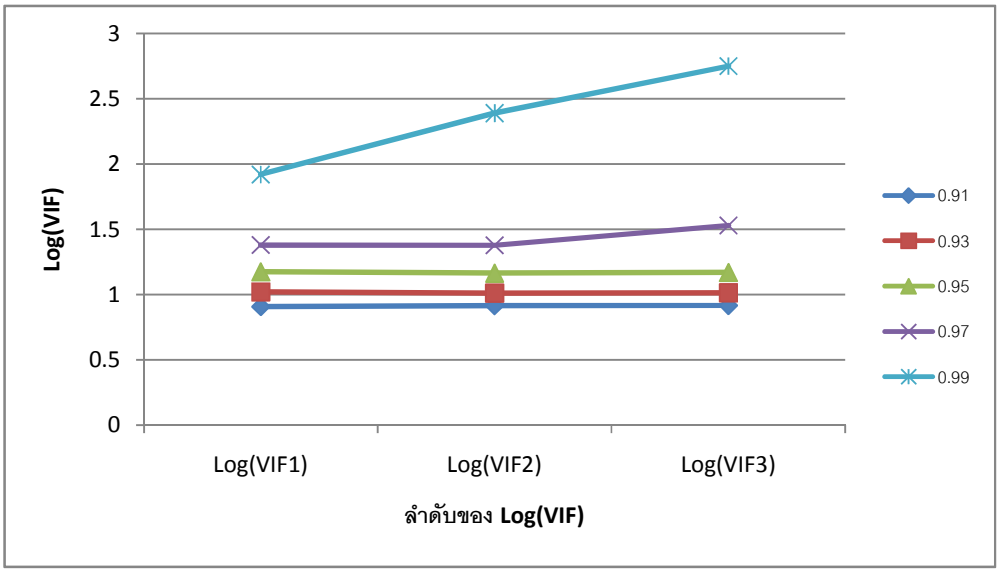
รูปที่ 4.1.36 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100



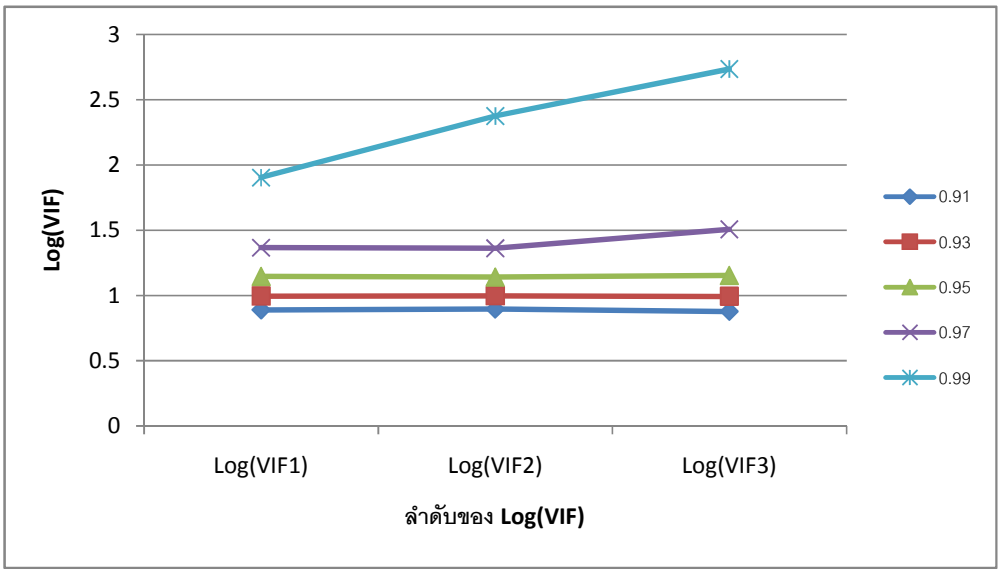
รูปที่ 4.1.37 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย VIF1, VIF2, VIF3 ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200



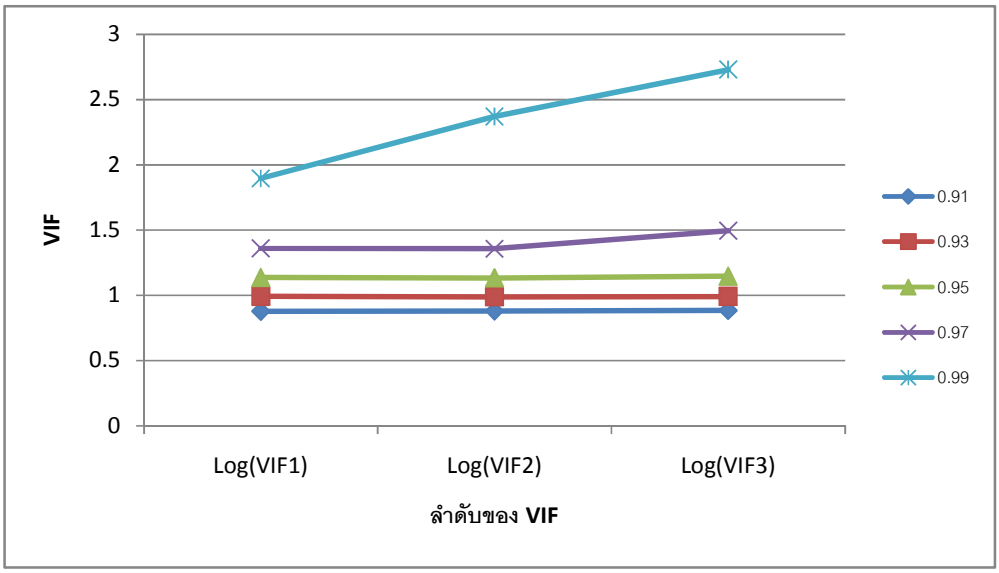
รูปที่ 4.1.38 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50



รูปที่ 4.1.39 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100



รูปที่ 4.1.40 กราฟแสดงการเปรียบเทียบค่าเฉลี่ย Log (VIF1), Log (VIF2), Log (VIF3) ในแต่ละค่าระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200



จากตารางที่ 4.1.6 และกราฟที่ 4.1.17- 4.1.40 สามารถสรุปได้ว่า

แนวโน้มค่าเฉลี่ย VIF1, VIF2, VIF3 โดยเรียงค่าระดับความสัมพันธ์จากน้อยไปมาก สามารถพิจารณาได้ 2 ลักษณะ คือ

1) พิจารณาจากระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างคงที่ พบว่าระดับความสัมพันธ์สูงสุดเท่ากับ 0.91 มีค่าเฉลี่ย VIF1, VIF2, VIF3 น้อยที่สุด แล้วค่าระดับความสัมพันธ์สูงสุดลำดับถัดไปที่มีค่าเฉลี่ย VIF1, VIF2, VIF3 มากกว่า คือ 0.93 และค่าเฉลี่ย VIF1, VIF2, VIF3 จะมีค่าเพิ่มขึ้นไปเรื่อยๆจนถึงระดับความสัมพันธ์สูงสุดที่ 0.99 มีค่าเฉลี่ย VIF1, VIF2, VIF3 มากที่สุด

2) พิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์สูงสุดคงที่ พบว่าขนาดตัวอย่างเท่ากับ 50 มีค่าเฉลี่ย VIF1, VIF2, VIF3 มากที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าเฉลี่ย VIF1, VIF2, VIF3 น้อยกว่าคือ 100 และขนาดตัวอย่างที่มีค่าเฉลี่ย VIF1, VIF2, VIF3 น้อยที่สุดคือ 200

สามารถสรุปแนวโน้มค่าเฉลี่ย VIF1, VIF2, VIF3 ได้ว่า เมื่อพิจารณาจากระดับความสัมพันธ์สูงสุด พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้นแล้วค่าเฉลี่ย VIF1, VIF2, VIF3 จะมีแนวโน้มเพิ่มขึ้น และเมื่อพิจารณาจากขนาดตัวอย่าง พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ย VIF1, VIF2, VIF3 จะมีแนวโน้มลดลง

จากตารางแสดงค่าเฉลี่ย VIF 1, VIF2, VIF3 จำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์สูงสุด พบว่าระดับความสัมพันธ์สูงสุดตั้งแต่ 0.93 ขึ้นไป เมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 มีค่าเฉลี่ย VIF1, VIF2, VIF3 มากกว่า 10 และระดับความสัมพันธ์สูงสุดตั้งแต่ 0.95 ขึ้นไปของทุกๆขนาดตัวอย่างที่ได้ศึกษา มีค่าเฉลี่ย VIF1, VIF2, VIF3 มากกว่า 10 แสดงว่าค่าเฉลี่ย VIF1, VIF2, VIF3 ในระดับความสัมพันธ์สูงสุดดังกล่าว เกิดปัญหาพหุสัมพันธ์รุนแรง สามารถสรุปได้ว่า ถ้าขนาดตัวอย่างเท่ากับ 50 และ 100 ค่าระดับความสัมพันธ์สูงสุดที่เป็นจุดเปลี่ยนทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง คือ 0.93 และ ถ้าขนาดตัวอย่าง เท่ากับ 200 ค่าระดับความสัมพันธ์สูงสุดที่เป็นจุดเปลี่ยนทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง คือ 0.95

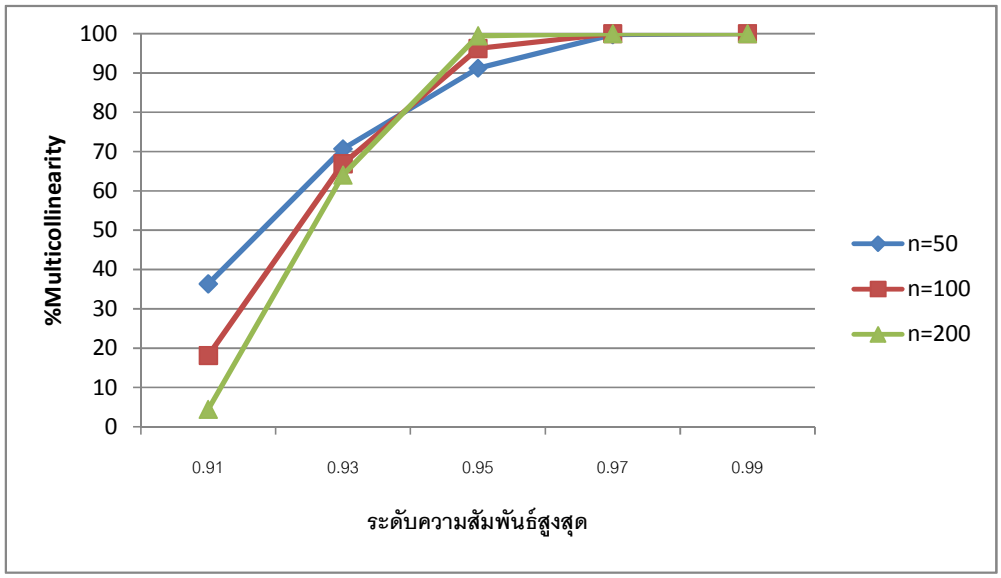
ตารางที่ 4.1.7 แสดงค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

$\rho_{\max}$	rho	n=50	n=100	n=200
0.91	0.91,0.91,0.91	36.3	18.1	4.4
0.93	0.93,0.93,0.93	70.7	66.9	64
0.95	0.91,0.93,0.95	91.2	96.2	99.4
0.95	0.95,0.95,0.95	96.7	99.5	99.5
0.97	0.91,0.93,0.97	99.7	100	100
0.97	0.91,0.95,0.97	100	100	100
0.97	0.93,0.95,0.97	100	100	100
0.97	0.97,0.97,0.97	100	100	100
0.99	0.91,0.93,0.99	100	100	100
0.99	0.91,0.95,0.99	100	100	100
0.99	0.93,0.95,0.99	100	100	100
0.99	0.93,0.97,0.99	100	100	100
0.99	0.95,0.97,0.99	100	100	100
0.99	0.99,0.99,0.99	100	100	100

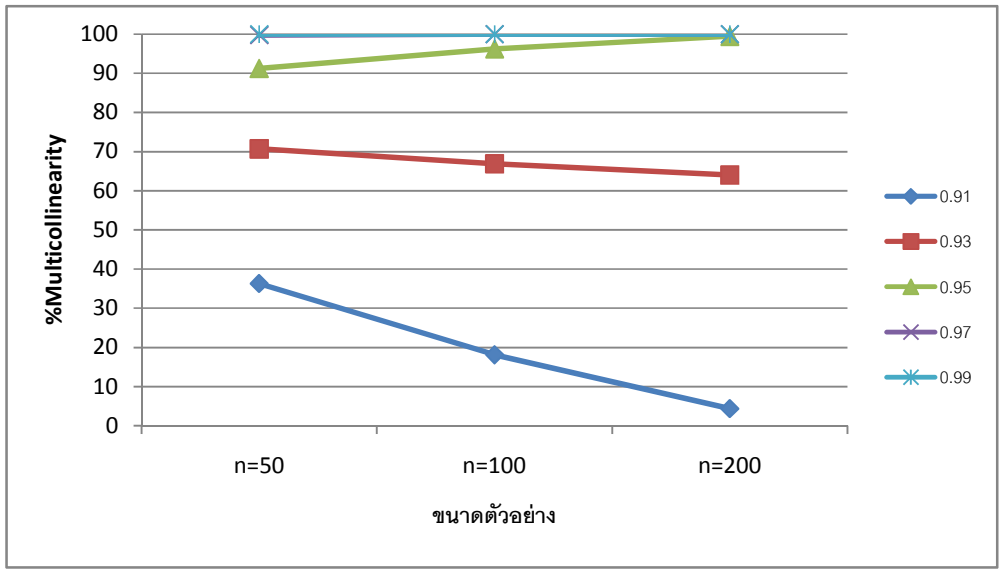
ตารางที่ 4.1.8 แสดงค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์สูงสุด

$\rho_{\max}$	0.91	0.93	0.95	0.97	0.99
n=50	36.3	70.7	91.2	99.7	100
n=100	18.1	66.9	96.2	100	100
n=200	4.4	64	99.4	100	100

รูปที่ 4.1.41 กราฟแสดงการเปรียบเทียบค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



รูปที่ 4.1.42 กราฟแสดงการเปรียบเทียบค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง ในแต่ละขนาดตัวอย่างเมื่อระดับความสัมพันธ์สูงสุดมีค่าเท่ากับ 0.91, 0.93, 0.95, 0.97, 0.99



จากตารางที่ 4.1.8 และกราฟที่ 4.1.41- 4.1.42 สามารถสรุปได้ว่า

แนวโน้มค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงสามารถพิจารณาได้ 2 ลักษณะ คือ

1. พิจารณาจากระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างคงที่ พบว่าระดับความสัมพันธ์สูงสุดเท่ากับ 0.91 มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงน้อยที่สุด แล้วค่าระดับความสัมพันธ์สูงสุดลำดับถัดไปที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากกว่าคือ 0.93 และค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีค่าเพิ่มขึ้นไปเรื่อยๆ จนถึง ระดับความสัมพันธ์สูงสุด 0.99 มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากที่สุด

2. พิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ เราสามารถแบ่งแนวโน้มออกเป็น 3 ลักษณะ คือ

2.1) ระดับความสัมพันธ์สูงสุด 0.91, 0.93 พบว่าที่ขนาดตัวอย่างเท่ากับ 50 มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงน้อยกว่าคือ 100 และขนาดตัวอย่างที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงน้อยที่สุดคือ 200

2.2) ระดับความสัมพันธ์สูงสุด 0.95, 0.97 พบว่าขนาดตัวอย่างเท่ากับ 50 มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงน้อยที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากกว่าคือ 100 และขนาดตัวอย่างที่มีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงมากที่สุดคือ 200

2.3) ระดับความสัมพันธ์สูงสุด 0.99 พบว่าทุกๆขนาดตัวอย่างที่ศึกษามีค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงเท่ากับ 100

ดังนั้นแนวโน้มค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จะได้ว่าเมื่อพิจารณาจากระดับความสัมพันธ์สูงสุด พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มเพิ่มขึ้น และเมื่อพิจารณาจากขนาดตัวอย่างสามารถแบ่งแนวโน้มออกเป็น 3 ลักษณะ คือ 1) ระดับความสัมพันธ์สูงสุด 0.91, 0.93 พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มลดลง 2) ระดับความสัมพันธ์สูงสุด 0.95, 0.97 พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มเพิ่มขึ้น 3) ระดับความสัมพันธ์สูงสุด 0.99 พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มคงที่

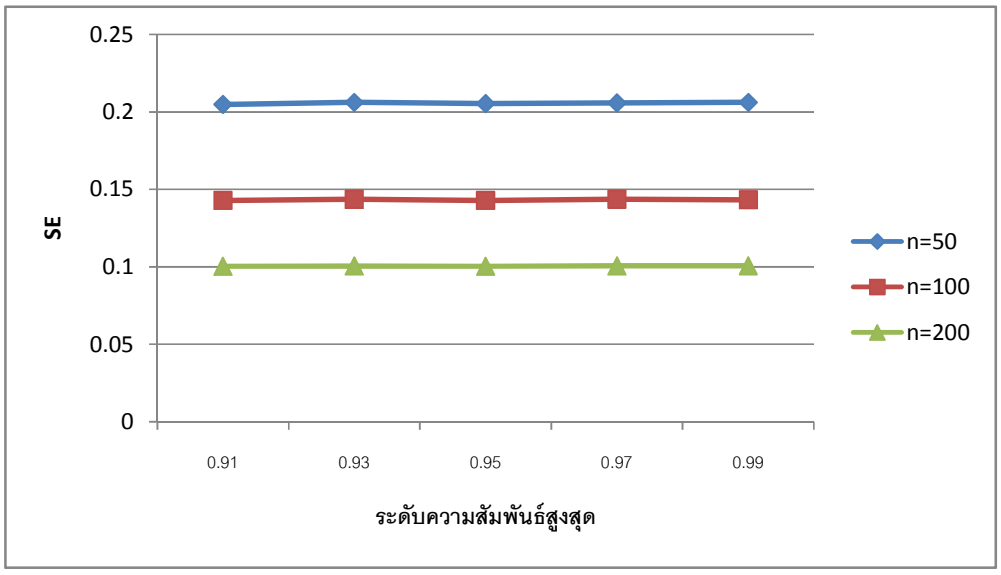
ตารางที่ 4.1.9 แสดงค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

$\rho_{\max}$	rho	n=50		n=100		n=200	
		X_OLS	PC_OLS	X_OLS	PC_OLS	X_OLS	PC_OLS
0.91	0.91,0.91,0.91	0.204705	0.198226	0.142856	0.140705	0.100445	0.099661
0.93	0.93,0.93,0.93	0.205985	0.199243	0.143511	0.141152	0.100644	0.099885
0.95	0.91,0.93,0.95	0.205213	0.198668	0.142856	0.140691	0.100436	0.099684
0.95	0.95,0.95,0.95	0.204682	0.198219	0.142855	0.140611	0.100335	0.099571
0.97	0.91,0.93,0.97	0.205547	0.198981	0.142943	0.140747	0.100592	0.09984
0.97	0.91,0.95,0.97	0.205664	0.199204	0.14358	0.141423	0.100699	0.099931
0.97	0.93,0.95,0.97	0.203503	0.197346	0.142815	0.140639	0.100362	0.099593
0.97	0.97,0.97,0.97	0.204214	0.197876	0.143207	0.140988	0.100645	0.099866
0.99	0.91,0.93,0.99	0.20488	0.198401	0.142909	0.140734	0.100742	0.099975
0.99	0.91,0.95,0.99	0.205438	0.19892	0.143019	0.140886	0.100476	0.099717
0.99	0.93,0.95,0.99	0.205571	0.198864	0.14322	0.140952	0.100617	0.099842
0.99	0.93,0.97,0.99	0.205791	0.199074	0.143194	0.140882	0.100468	0.099692
0.99	0.95,0.97,0.99	0.204826	0.1984	0.143016	0.140838	0.100645	0.099868
0.99	0.99,0.99,0.99	0.205983	0.199419	0.143117	0.14094	0.100765	0.099997

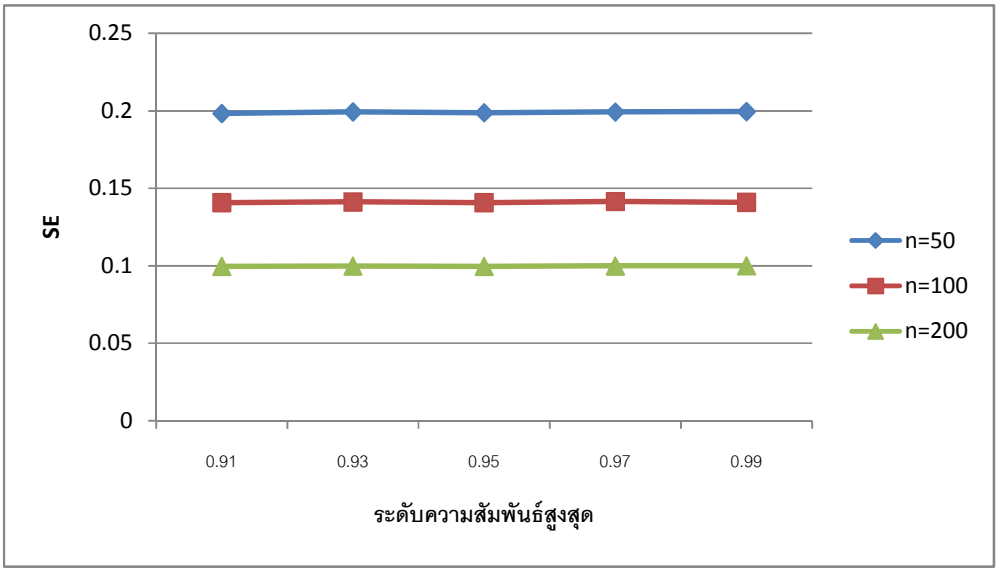
ตารางที่ 4.1.10 แสดงค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย จำแนกตามขนาดตัวอย่างและระดับความสัมพันธ์สูงสุด

ρ_{max}		0.91	0.93	0.95	0.97	0.99
n=50	X_OLS	0.204705	0.205985	0.205213	0.205664	0.205983
	PC_OLS	0.198226	0.199243	0.198668	0.199204	0.199419
n=100	X_OLS	0.142856	0.143511	0.142856	0.14358	0.14322
	PC_OLS	0.140705	0.141152	0.140691	0.141423	0.140952
n=200	X_OLS	0.100445	0.100644	0.100436	0.100699	0.100765
	PC_OLS	0.099661	0.099885	0.099684	0.099931	0.099997

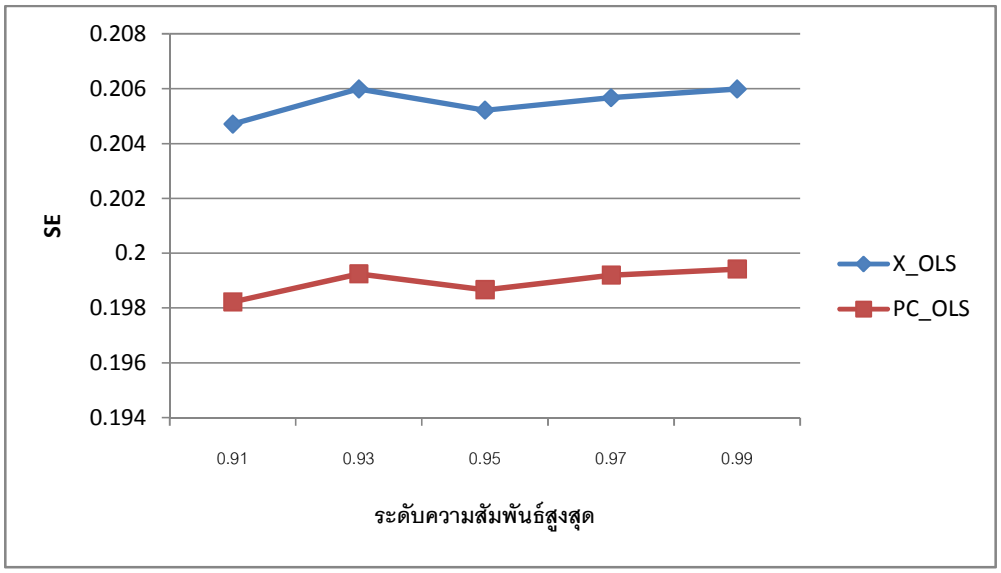
รูปที่ 4.1.43 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย ก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50,100, 200



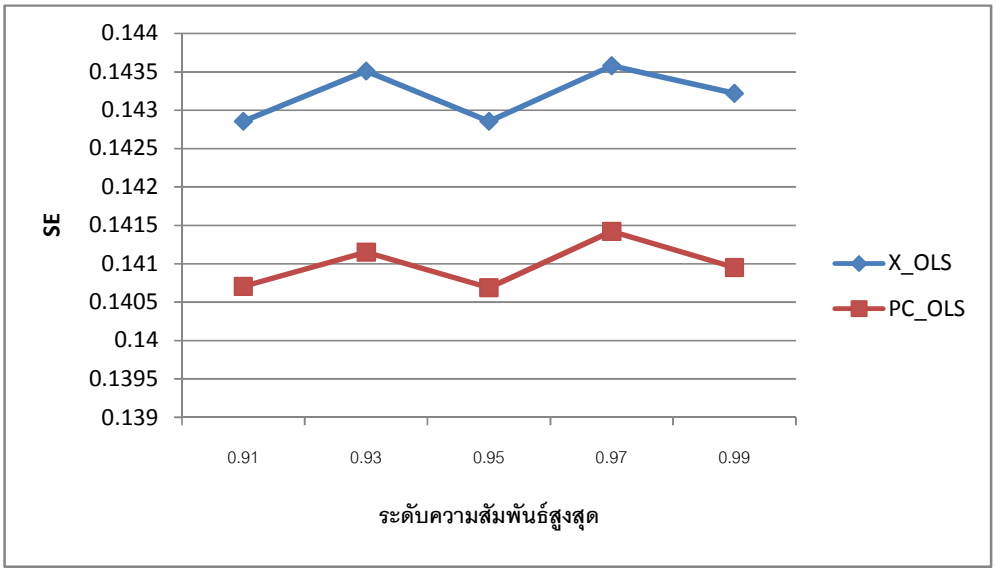
รูปที่ 4.1.44 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยมาตรฐาน หลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



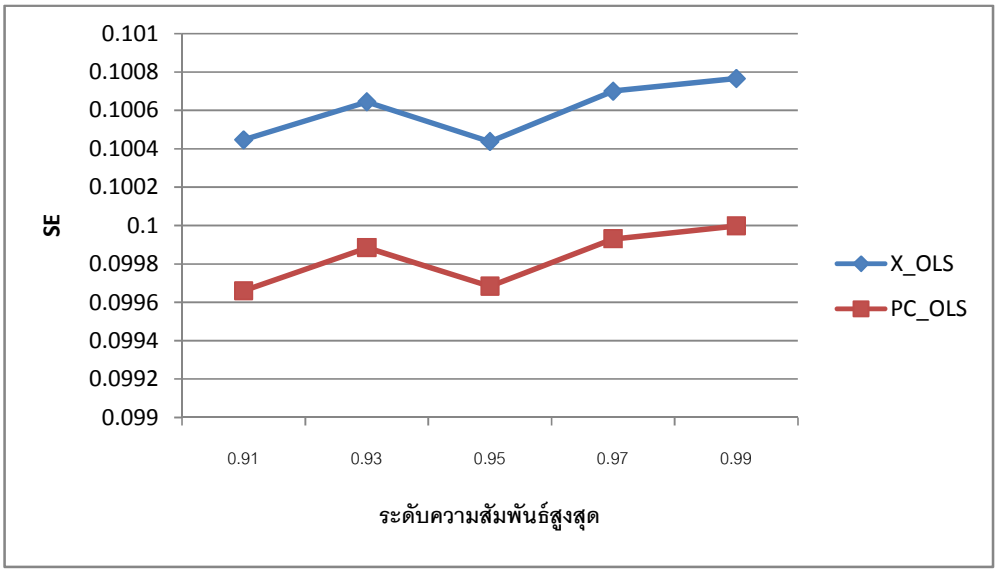
รูปที่ 4.1.45 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50



รูปที่ 4.1.46 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100



รูปที่ 4.1.47 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200



จากตารางที่ 4.1.10 และกราฟที่ 4.1.43- 4.1.47 สามารถสรุปได้ว่า

แนวโน้มค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย สามารถพิจารณาได้ 2 ลักษณะ คือ

1. พิจารณาจากระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีค่าใกล้เคียงกัน

2. พิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานมากที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานน้อยกว่าคือ 100 และขนาดตัวอย่างที่มีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานน้อยที่สุดคือ 200 ทั้งก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงและหลังแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก

ดังนั้นสามารถสรุปแนวโน้มค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย คือเมื่อพิจารณาจากระดับความสัมพันธ์สูงสุด พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้นแล้วค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีแนวโน้มค่อนข้างคงที่ และเมื่อพิจารณาจากขนาดตัวอย่าง พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีแนวโน้มลดลง

จากการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงจะมีค่ามากกว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ทุกๆขนาดตัวอย่างที่ได้ศึกษา

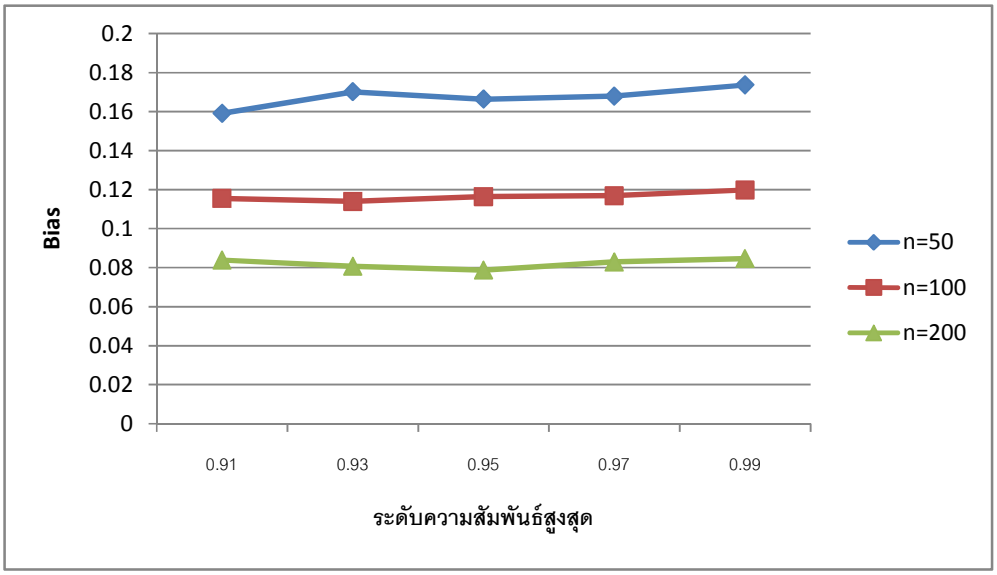
ตารางที่ 4.1.11 แสดงค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์

$\rho_{\max}$	rho	n=50		n=100		n=200	
		X_OLS	PC_OLS	X_OLS	PC_OLS	X_OLS	PC_OLS
0.91	0.91,0.91,0.91	0.159086	0.902121	0.11549	0.665956	0.083924	0.478384
0.93	0.93,0.93,0.93	0.170116	0.953831	0.113974	0.688742	0.080716	0.478786
0.95	0.91,0.93,0.95	0.163131	0.930703	0.111502	0.674844	0.078772	0.482029
0.95	0.95,0.95,0.95	0.1664	0.922237	0.116453	0.657476	0.078311	0.474507
0.97	0.91,0.93,0.97	0.167873	0.958871	0.116839	0.656217	0.082944	0.471419
0.97	0.91,0.95,0.97	0.163928	0.953734	0.112341	0.666411	0.079291	0.466807
0.97	0.93,0.95,0.97	0.164757	0.91683	0.109071	0.658947	0.078833	0.453766
0.97	0.97,0.97,0.97	0.167224	0.962224	0.112536	0.674114	0.078007	0.48364
0.99	0.91,0.93,0.99	0.164568	0.925318	0.117023	0.648385	0.084491	0.481493
0.99	0.91,0.95,0.99	0.170577	0.952913	0.114768	0.669241	0.080904	0.463114
0.99	0.93,0.95,0.99	0.169015	1.00448	0.116165	0.712519	0.080853	0.464713
0.99	0.93,0.97,0.99	0.173654	0.985682	0.119813	0.695595	0.084689	0.49252
0.99	0.95,0.97,0.99	0.164972	0.983502	0.113871	0.688829	0.08197	0.494079
0.99	0.99,0.99,0.99	0.165536	0.952807	0.116181	0.686079	0.080783	0.489203

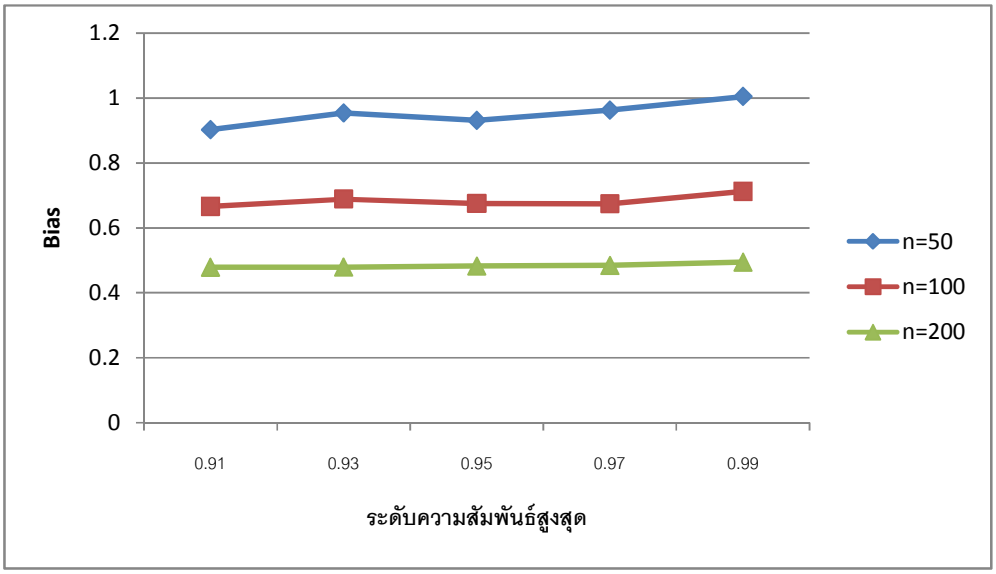
ตารางที่ 4.1.12 แสดงค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย จำแนกตามขนาดตัวอย่าง และระดับความสัมพันธ์สูงสุด

ρ_{max}		0.91	0.93	0.95	0.97	0.99
n=50	X_OLS	0.159086	0.170116	0.1664	0.167873	0.173654
	PC_OLS	0.902121	0.953831	0.930703	0.962224	1.00448
n=100	X_OLS	0.11549	0.113974	0.116453	0.116839	0.119813
	PC_OLS	0.665956	0.688742	0.674844	0.674114	0.712519
n=200	X_OLS	0.083924	0.080716	0.078772	0.082944	0.084689
	PC_OLS	0.478384	0.478786	0.482029	0.48364	0.494079

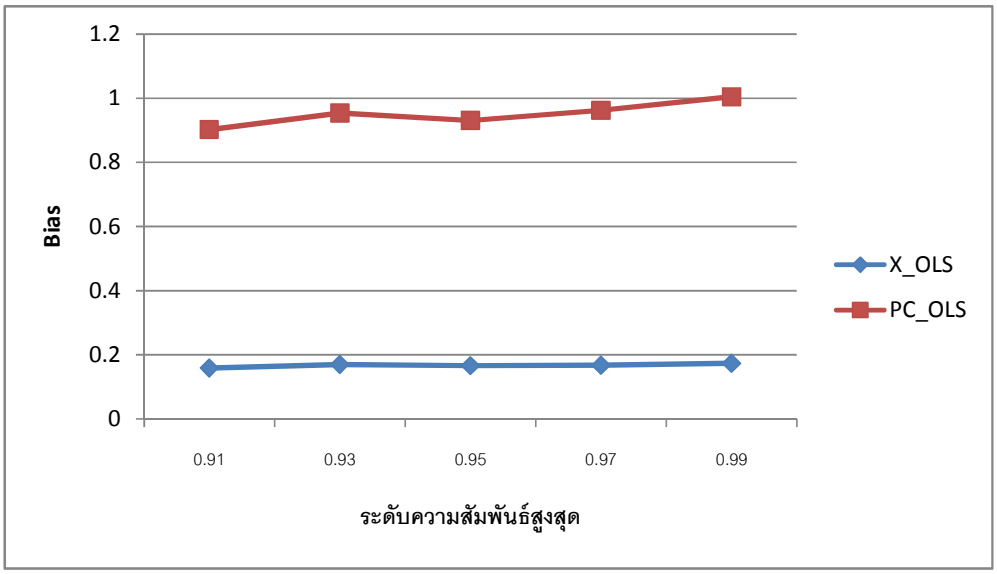
รูปที่ 4.1.48 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



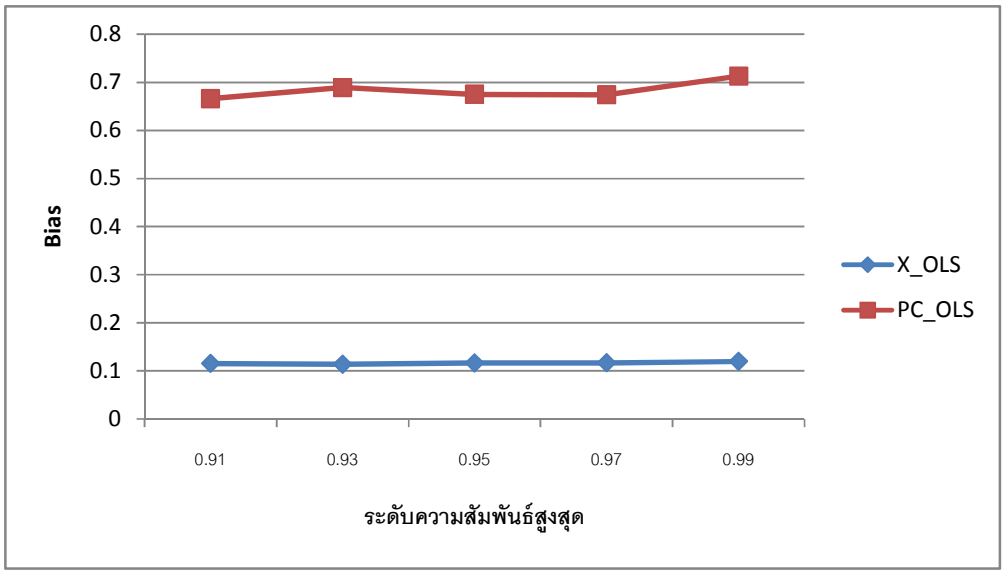
รูปที่ 4.1.49 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200



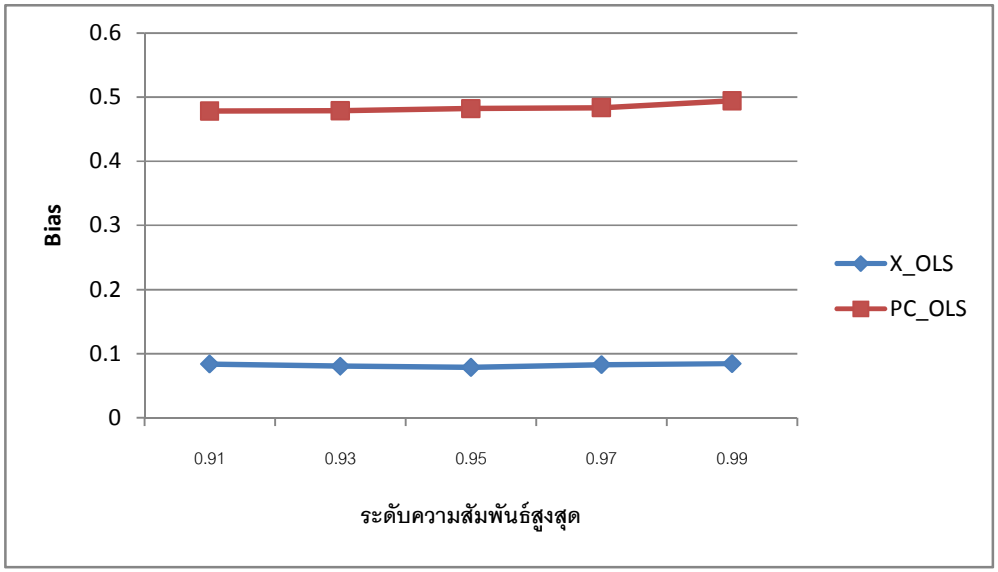
รูปที่ 4.1.50 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 50



รูปที่ 4.1.51 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 100



รูปที่ 4.1.52 กราฟแสดงการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ในแต่ละระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างเท่ากับ 200



จากตารางที่ 4.1.12 และกราฟที่ 4.1.48- 4.1.52 สามารถสรุปได้ว่า

แนวโน้มค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย สามารถพิจารณาได้ 2 ลักษณะ คือ

1. พิจารณาจากระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้น ค่าเฉลี่ยความเอนเอียงจะมีค่าใกล้เคียงกัน

2. พิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์สูงสุดคงที่ พบว่าขนาดตัวอย่างเท่ากับ 50 จะมีค่าเฉลี่ยความเอนเอียงมากที่สุด แล้วขนาดตัวอย่างลำดับถัดไปที่มีค่าเฉลี่ยความเอนเอียงน้อยกว่าคือ 100 และขนาดตัวอย่างที่มีค่าเฉลี่ยความเอนเอียงน้อยที่สุดคือ 200 ทั้งก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงและหลังแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก

ดังนั้นสามารถสรุปแนวโน้มค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย ได้ว่าเมื่อพิจารณาจากระดับความสัมพันธ์สูงสุด พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้นแล้วค่าเฉลี่ยความเอนเอียงจะมีแนวโน้มค่อนข้างคงที่ และเมื่อพิจารณาจากขนาดตัวอย่าง พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ยความเอนเอียงจะมีแนวโน้มลดลง

จากการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าค่าเฉลี่ยความเอนเอียงก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงมีค่าน้อยกว่าค่าเฉลี่ยความเอนเอียงหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ทุกๆขนาดตัวอย่างที่ได้ศึกษา

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาการแก้ไขปัญหาการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้นเมื่อเกิดปัญหาพหุสัมพันธ์รุนแรงด้วยวิธีการวิเคราะห์ตัวประกอบหลัก โดยมีขอบเขตการศึกษาดังนี้

1. กำหนดตัวแปรอิสระ(x) คือ 2, 3 ตัวแปร เป็นตัวแปรเชิงปริมาณ ที่มีการแจกแจงแบบปกติมีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 2 และตัวแปรอิสระมีความสัมพันธ์ดังนี้

1.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว (p=2)

ρ_{12} คือ ค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และตัวที่ 2

โดยทำการศึกษาระดับความสัมพันธ์ดังนี้

1.1.1 ระดับความสัมพันธ์ 0.90

1.1.2 ระดับความสัมพันธ์ 0.91

1.1.3 ระดับความสัมพันธ์ 0.92

1.1.4 ระดับความสัมพันธ์ 0.93

1.1.5 ระดับความสัมพันธ์ 0.94

1.1.6 ระดับความสัมพันธ์ 0.95

1.1.7 ระดับความสัมพันธ์ 0.96

1.1.8 ระดับความสัมพันธ์ 0.97

1.1.9 ระดับความสัมพันธ์ 0.98

1.1.10 ระดับความสัมพันธ์ 0.99

1.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว (p=3)

ρ_{12} คือ ค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และ ตัวที่ 2

ρ_{13} คือ ค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 และ ตัวที่ 3

ρ_{23} คือ ค่าระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 2 และ ตัวที่ 3

โดยทำการศึกษาระดับความสัมพันธ์ดังนี้

1.2.1 ระดับความสัมพันธ์ (0.91, 0.93, 0.95)

1.2.2 ระดับความสัมพันธ์ (0.91, 0.93, 0.97)

1.2.3 ระดับความสัมพันธ์ (0.91, 0.93, 0.99)

1.2.4 ระดับความสัมพันธ์ (0.91, 0.95, 0.97)

1.2.5 ระดับความสัมพันธ์ (0.91, 0.95, 0.99)

1.2.6 ระดับความสัมพันธ์ (0.93, 0.95, 0.97)

1.2.7 ระดับความสัมพันธ์ (0.93, 0.95, 0.99)

1.2.8 ระดับความสัมพันธ์ (0.93, 0.97, 0.99)

1.2.9 ระดับความสัมพันธ์ (0.95, 0.97, 0.99)

1.2.10 ระดับความสัมพันธ์ (0.91, 0.91, 0.91)

1.2.11 ระดับความสัมพันธ์ (0.93, 0.93, 0.93)

1.2.12 ระดับความสัมพันธ์ (0.95, 0.95, 0.95)

1.2.13 ระดับความสัมพันธ์ (0.97, 0.97, 0.97)

1.2.14 ระดับความสัมพันธ์ (0.99, 0.99, 0.99)

2. ศึกษาภายใต้ขนาดตัวอย่างดังนี้

2.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว จะศึกษาขนาดตัวอย่าง 50, 100, 200

2.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว จะศึกษาขนาดตัวอย่าง 50, 100, 200

3. กำหนดค่าสัมประสิทธิ์ความถดถอย $(\beta_0, \beta_1, \beta_2, ..., \beta_p)$ ได้ดังนี้

3.1 กรณีจำนวนตัวแปรอิสระ 2 ตัว มีค่าสัมประสิทธิ์ความถดถอย คือ

$$\beta_0 = \beta_1 = \beta_2 = 2$$

3.2 กรณีจำนวนตัวแปรอิสระ 3 ตัว มีค่าสัมประสิทธิ์ความถดถอย คือ

$$\beta_0 = \beta_1 = \beta_2 = \beta_3 = 2$$

4. จำลองข้อมูลความคลาดเคลื่อนที่มีการแจกแจงปกติ โดยมีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ 2

5. สร้างตัวแปรตาม y จากตัวแบบความถดถอยพหุเชิงเส้น $\widetilde{Y} = X\widetilde{\beta} + \widetilde{\varepsilon}$

5.1 **สรุปผลการศึกษา**

ในการสรุปผลการศึกษาการแก้ไขปัญหาการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้นเมื่อเกิดปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก สามารถสรุปผลการศึกษาได้ดังนี้

กรณีจำนวนตัวแปรอิสระ 2 ตัว

<u>พิจารณาค่าระดับความสัมพันธ์ที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง</u>

ค่าเฉลี่ย VIFตั้งแต่ระดับความสัมพันธ์ 0.95 ขึ้นไป ของทุกๆขนาดตัวอย่างที่ได้ศึกษา จะมีค่าเฉลี่ย VIF มากกว่า 10 ซึ่งแสดงว่าค่าเฉลี่ย VIF ในระดับความสัมพันธ์ดังกล่าว เกิดปัญหาพหุสัมพันธ์รุนแรง ดังนั้นระดับความสัมพันธ์ 0.95 คือจุดเปลี่ยนที่ทำให้เกิดปัญหาพหุสัมพันธ์รุนแรง โดยแนวโน้มค่าเฉลี่ย VIF จะได้ว่า ถ้าพิจารณาจากระดับความสัมพันธ์ เมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์มีค่าเพิ่มขึ้นแล้วค่าเฉลี่ย VIF จะมีแนวโน้มเพิ่มขึ้น และถ้าพิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ย VIF จะมีแนวโน้มลดลง

<u>พิจารณาค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์</u>

แนวโน้มค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จะได้ว่า ถ้าพิจารณาจากระดับความสัมพันธ์ เมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์มีค่าเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มเพิ่มขึ้น ซึ่งมีลักษณะคล้ายตัว S และ ถ้าพิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าสามารถแบ่งแนวโน้มออกเป็น 3 ลักษณะ คือ 1) ระดับความสัมพันธ์ 0.90 – 0.94 ถ้าขนาดตัวอย่างเพิ่มขึ้น แล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มลดลง 2) ระดับความสัมพันธ์ 0.95 – 0.98 ถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มเพิ่มขึ้น 3) ระดับความสัมพันธ์ 0.99 ถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มคงที่

<u>พิจารณาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย</u>

แนวโน้มค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย จะได้ว่าถ้าพิจารณาจากระดับความสัมพันธ์ เมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์มีค่าเพิ่มขึ้นแล้วค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีแนวโน้มค่อนข้างคงที่ และถ้าพิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีแนวโน้มลดลง

จากการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงมีค่ามากกว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัว

ประกอบหลัก ทุกๆขนาดตัวอย่างที่ได้ศึกษา

พิจารณาค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย

แนวโน้มค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย จะได้ว่าถ้าพิจารณาจากระดับความสัมพันธ์ เมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์มีค่าเพิ่มขึ้นแล้วค่าเฉลี่ยความเอนเอียงจะมีแนวโน้มค่อนข้างคงที่ และถ้าพิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ยความเอนเอียงจะมีแนวโน้มลดลง

จากการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าค่าเฉลี่ยความเอนเอียงก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงมีค่าน้อยกว่าค่าเฉลี่ยความเอนเอียงหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ทุกๆขนาดตัวอย่างที่ได้ศึกษา

กรณีจำนวนตัวแปรอิสระ 3 ตัว

พิจารณาค่าระดับความสัมพันธ์ที่ทำให้เกิดปัญหาพหุสัมพันธ์

เมื่อขนาดตัวอย่างเท่ากับ 50 และ 100 พบว่าระดับความสัมพันธ์สูงสุด 0.93 ขึ้นไป มีค่าเฉลี่ย VIF1, VIF2, VIF3 มากกว่า 10 และเมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าระดับความสัมพันธ์สูงสุด 0.95 ขึ้นไป มีค่าเฉลี่ย VIF1, VIF2, VIF3 มากกว่า 10 แสดงว่าค่าเฉลี่ย VIF1, VIF2, VIF3 ในระดับความสัมพันธ์สูงสุดดังกล่าว เกิดปัญหาพหุสัมพันธ์ จึงสามารถสรุปได้ว่า เมื่อขนาดตัวอย่างเท่ากับ 50, 100 ค่าระดับความสัมพันธ์สูงสุดที่เป็นจุดเปลี่ยนทำให้เกิดปัญหาพหุสัมพันธ์ คือ 0.93 และ เมื่อขนาดตัวอย่างเท่ากับ 200 ค่าระดับความสัมพันธ์สูงสุดที่เป็นจุดเปลี่ยนทำให้เกิดปัญหาพหุสัมพันธ์ คือ 0.95

แนวโน้มค่าเฉลี่ย VIF1, VIF2, VIF3 จะได้ว่าถ้าพิจารณาจากระดับความสัมพันธ์สูงสุดเมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้นแล้วค่าเฉลี่ย VIF1, VIF2, VIF3 จะมีแนวโน้มเพิ่มขึ้นและถ้าพิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์สูงสุดคงที่ พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ย VIF1, VIF2, VIF3 จะมีแนวโน้มลดลง

พิจารณาค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์

แนวโน้มค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรง จะได้ว่าถ้าพิจารณาจากระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มเพิ่มขึ้น และถ้าพิจารณาจากขนาด

ตัวอย่าง เมื่อระดับความสัมพันธ์สูงสุดคงที่ พบว่าสามารถแบ่งแนวโน้มออกเป็น 3 ลักษณะ คือ 1) ระดับความสัมพันธ์สูงสุดเท่ากับ 0.91, 0.93 ถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มลดลง 2) ระดับความสัมพันธ์สูงสุดเท่ากับ 0.95, 0.97 ถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มเพิ่มขึ้น 3) ระดับความสัมพันธ์สูงสุดเท่ากับ 0.99 ถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าร้อยละของการเกิดปัญหาพหุสัมพันธ์รุนแรงจะมีแนวโน้มคงที่

<u>พิจารณาค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย</u>

แนวโน้มค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอย จะได้ว่าถ้าพิจารณาจากระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างคงที่ พบว่า ถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้นแล้วค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีแนวโน้มค่อนข้างคงที่ และถ้าพิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์สูงสุดคงที่ พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้นแล้วค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานจะมีแนวโน้มลดลง

จากการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงมีค่ามากกว่าค่าเฉลี่ยความคลาดเคลื่อนมาตรฐานหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ทุกๆขนาดตัวอย่างที่ได้ศึกษา

<u>พิจารณาค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย</u>

แนวโน้มค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอย จะได้ว่าถ้าพิจารณาจากระดับความสัมพันธ์สูงสุด เมื่อขนาดตัวอย่างคงที่ พบว่าถ้าระดับความสัมพันธ์สูงสุดมีค่าเพิ่มขึ้นแล้วค่าเฉลี่ยความเอนเอียงจะมีแนวโน้มค่อนข้างคงที่ และถ้าพิจารณาจากขนาดตัวอย่าง เมื่อระดับความสัมพันธ์คงที่ พบว่าถ้าขนาดตัวอย่างเพิ่มขึ้น แล้วค่าเฉลี่ยความเอนเอียงจะมีแนวโน้มลดลง

จากการเปรียบเทียบค่าเฉลี่ยความเอนเอียงของสัมประสิทธิ์ความถดถอยก่อนและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง เมื่อขนาดตัวอย่างเท่ากับ 50, 100, 200 พบว่าค่าเฉลี่ยความเอนเอียงก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงมีค่าน้อยกว่าค่าเฉลี่ยความเอนเอียงหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรง ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก ทุกๆขนาดตัวอย่างที่ได้ศึกษา

5.2 ข้อเสนอแนะ

การแก้ไขปัญหาพหุสัมพันธ์รุนแรงด้วยการวิเคราะห์ตัวประกอบหลัก กรณีจำนวนตัวแปรอิสระ 2 ตัว พบว่า อาจจะไม่เหมาะสมด้วยวิธีดังกล่าวเมื่อตัวแปรอิสระมีความสัมพันธ์กันสูงมาก เนื่องจากเมื่อสร้างตัวแปรตัวใหม่แล้ว ตัวแปรใหม่ตัวที่ 1 จะดึงความแปรปรวนจากตัวแปรทั้งหมด 2 ตัวมาไว้ในตัวแปรใหม่ตัวที่ 1 มากที่สุด และ ตัวแปรใหม่ตัวที่ 2 ก็จะดึงความแปรปรวนที่เหลือจากตัวที่ 1 ดังนั้น ตัวแปรใหม่ตัวที่ 2 อาจจะไม่มีผลต่อการสร้างสมการถดถอย แต่ในการศึกษาครั้งนี้ไม่ได้ทำการตัดตัวแปรใดๆทิ้งไป เพื่อต้องการเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานและความเอนเอียง ก่อนการแก้ไขปัญหาพหุสัมพันธ์รุนแรงและหลังการแก้ไขปัญหาพหุสัมพันธ์รุนแรงด้วยวิธีการวิเคราะห์ตัวประกอบ

5.3 แนวทางการศึกษาต่อ

ในการศึกษาเรื่องการแก้ไขปัญหาการประมาณค่าพารามิเตอร์ของตัวแบบความถดถอยพหุเชิงเส้นเมื่อเกิดปัญหาพหุสัมพันธ์ด้วยวิธีการวิเคราะห์ตัวประกอบหลัก มีแนวทางในการศึกษาต่อดังนี้คือ

1. ศึกษาการแก้ไขปัญหาพหุสัมพันธ์เมื่อตัวแปรอิสระและตัวแปรตาม ที่เป็นทั้งข้อมูลเชิงคุณภาพและปริมาณ
2. ศึกษาวิธีการแก้ไขปัญหาพหุสัมพันธ์แบบอื่นๆ ที่ช่วยในการแก้ปัญหาตัวแปรอิสระมีความสัมพันธ์กัน
3. ศึกษาการแจงแจงแบบอื่นๆ ที่มีการแจกแจงแบบต่อเนื่องและไม่ต่อเนื่อง เพื่อนำไปประยุกต์ใช้ต่อไป

รายการอ้างอิง

ภาษาไทย

กรรณิกาณ์ หิรัญกสิ.การเปรียบเทียบการประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุดแบ่งส่วน วิธีการถดถอยองค์ประกอบ และวิธีกำลังสองน้อยที่สุดในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ,วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2540.

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

ธีระพร วีระถาวร. การอนุมานเชิงสถิติขั้นกลาง: โครงสร้างและความหมาย.กรุงเทพมหานคร: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2536.

เปรมวดี ชูไสว.การเปรียบเทียบวิธีการแก้ไขพหุสัมพันธ์ในการถดถอยเชิงเส้นพหุคูณ,วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2548.

วิชิต หล่อจีระชุณห์กุล และจิราวัลย์ จิตรถเวช. เทคนิคการพยากรณ์.พิมพ์ครั้งที่3.กรุงเทพมหานคร: โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์, 2548.

สุพล ดุรงค์วัฒนา. ตัวแบบและการวิเคราะห์ความถดถอย สำหรับการวิจัยขั้นสูง.กรุงเทพมหานคร: ม.ป.ท., 2549.

เสกสรร เกียรติสุไพบูลย์. การจำลอง. กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2555.

ภาษาอังกฤษ

S.Q.Lafi and J.B.Kaneene. An explanation of the use of principal components analysis to detect and correct for multicollinearity . Preventive Veterinary Medicine 13(1992): 261-275.

Gunst, R.F. and R.L.Mason. Biased estimation in regression: An evaluation using mean squared error. J.Am.Stat.Assoc. 72(1977) : 616-628.

ภาคผนวก

ภาคผนวก

โปรแกรมที่ใช้สร้างและวิเคราะห์ข้อมูล

1 การหาค่า VIF ของจำนวนตัวแปรอิสระ 2, 3 ตัว

```
        library(fmsb)
        N<-500
        vif2<-c()
        vif3<-c()
for(i in 1:N){
rho_x<-c(0.9,0.91,0.92,0.93,0.94,0.95,0.96,0.97,0.98,0.99)
VIF_TOTAL2<-c()
for(t in 1:length(rho_x)){
        sd_1<-sqrt(2)
        sd_2<-sqrt(2)
        sd_3<-sqrt(2)

        z1<-rnorm(50)
        z2<-rnorm(50)
        z3<-rnorm(50)
        z4<-rnorm(50)
        z5<-c(z1,z3)
        z6<-c(z2,z4)
        z7<-rnorm(100)
        z8<-rnorm(100)
        z9<- c(z1,z3,z7)
        z10<-c(z2,z4,z8)
##n=50##
        Rho_X12<-rho_x[t]
```

```
        x_501<-z1
        x_502<-((Rho_X12)*z1)+((sqrt(1-(Rho_X12^2))*z2))
        x50_1<-x_501*sd_1
        x50_2<-x_502*sd_2
        dat50X<-data.frame(x50_1,x50_2)
        vif2_x50<-VIF(lm(x50_1~x50_2,data=dat50X))
##n=100##
        x_1001<-z5
        x_1002<-((Rho_X12)*z5)+((sqrt(1-(Rho_X12^2))*z6))
        x100_1<-x_1001*sd_1
        x100_2<-x_1002*sd_2
        dat100X<-data.frame(x100_1,x100_2)
        vif2_x100<-VIF(lm(x100_1~x100_2,data=dat100X))
##n=200##
        x_2001<-z9
        x_2002<-((Rho_X12)*z9)+((sqrt(1-(Rho_X12^2))*z10))
        x200_1<-x_2001*sd_1
        x200_2<-x_2002*sd_2
        dat200X<-data.frame(x200_1,x200_2)
        vif2_x200<-VIF(lm(x200_1~x200_2,data=dat200X))
        VIF_TOTAL2<-c(VIF_TOTAL2,vif2_x50,vif2_x100,vif2_x200)
        }
        vif2<-c(vif2,VIF_TOTAL2)

        rho<-c(0.91,0.93,0.95,0.97,0.99)
        t_rho<-t(combn(rho,3))
        newrho<-t_rho[c(-6),]
```

```
        r_r<-c(rep(0.91,3),rep(0.93,3),rep(0.95,3),rep(0.97,3),rep(0.99,3))
        mr_r<-matrix(c(r_r),3)
        tr_r<-t(mr_r)
        total<-rbind(newrho,tr_r)
        VIF_TOTAL3 <- c()
for(p in 1:dim(total)[1]){
sd_1<-sqrt(2)
        sd_2<-sqrt(2)
        sd_3<-sqrt(2)
        z1<-rnorm(50)
        z2<-rnorm(50)
        z3<-rnorm(50)
        z4<-rnorm(50)
        z5<-rnorm(50)
        z6<-rnorm(50)
        z7<-c(z1,z4)
        z8<-c(z2,z5)
        z9<-c(z3,z6)
        z10<-rnorm(100)
        z11<-rnorm(100)
        z12<-rnorm(100)
        z13<-c(z1,z4,z10)
        z14<-c(z2,z5,z11)
        z15<-c(z3,z6,z12)

        RHo_X12<-total[p,1]
        RHo_X13<-total[p,2]
```

```
        RHo_X23<-total[p,3]
##n=50,p=3##
        x503_1<-z1
        x503_2<-((RHo_X12)*z1)+((sqrt(1-(RHo_X12^2))*z2))
        x503_3<-((RHo_X13)*z1)+((((RHo_X23-(RHo_X12*RHo_X13))/(sqrt(1-
(RHo_X12^2))))*z2)+((sqrt(1-(RHo_X13^2)-(((RHo_X23-(RHo_X12*RHo_X13))^2)/(1-
(RHo_X12^2))))*z3)))
        x5031<-x503_1*sd_1
        x5032<-x503_2*sd_2
        x5033<-x503_3*sd_3
        dat503X<-data.frame(x5031,x5032,x5033)
        vif_x5031<-VIF(lm(x5031~x5032+x5033,data=dat503X))
        vif_x5032<-VIF(lm(x5032~x5031+x5033,data=dat503X))
        vif_x5033<-VIF(lm(x5033~x5031+x5032,data=dat503X))
        vif503<-c(vif_x5031,vif_x5032,vif_x5033)
##n=100,p=3##
        x1003_1<-z7
        x1003_2<-((RHo_X12)*z7)+((sqrt(1-(RHo_X12^2))*z8))
        x1003_3<-((RHo_X13)*z7)+((((RHo_X23-(RHo_X12*RHo_X13))/(sqrt(1-
(RHo_X12^2))))*z8)+((sqrt(1-(RHo_X13^2)-(((RHo_X23-(RHo_X12*RHo_X13))^2)/(1-
(RHo_X12^2))))*z9)))
        x10031<-x1003_1*sd_1
        x10032<-x1003_2*sd_2
        x10033<-x1003_3*sd_3
        dat1003X<-data.frame(x10031,x10032,x10033)
        vif_x10031<-VIF(lm(x10031~x10032+x10033,data=dat1003X))
        vif_x10032<-VIF(lm(x10032~x10031+x10033,data=dat1003X))
```

```
        vif_x10033<-VIF(lm(x10033~x10031+x10032,data=dat1003X))
        vif1003<-c(vif_x10031,vif_x10032,vif_x10033)
##n=200,p=3##
        x2003_1<-z13
        x2003_2<-((RHo_X12)*z13)+((sqrt(1-(RHo_X12^2))*z14))
        x2003_3<-((RHo_X13)*z13)+((((RHo_X23-(RHo_X12*RHo_X13))/(sqrt(1-
(RHo_X12^2))))*z14)+((sqrt(1-(RHo_X13^2)-(((RHo_X23-(RHo_X12*RHo_X13))^2)/(1-
(RHo_X12^2))))*z15)))
        x20031<-x2003_1*sd_1
        x20032<-x2003_2*sd_2
        x20033<-x2003_3*sd_3
        dat2003X<-data.frame(x20031,x20032,x20033)
        vif_x20031<-VIF(lm(x20031~x20032+x20033,data=dat2003X))
        vif_x20032<-VIF(lm(x20032~x20031+x20033,data=dat2003X))
        vif_x20033<-VIF(lm(x20033~x20031+x20032,data=dat2003X))
        vif2003<-c(vif_x20031,vif_x20032,vif_x20033)
        VIF_TOTAL3<- c(VIF_TOTAL3,vif_x5031,vif_x5032,vif_x5033,vif_x10031,
                vif_x10032,vif_x10033,vif_x20031,vif_x20032,vif_x20033)
        }
        vif3<-c(vif3,VIF_TOTAL3)
        pie(c(i,N-i),c(i,N-i),radius=1,clockwise=T)
}
VIF กรณีตัวแปรอิสระ 2 ตัว
mvif2<-matrix(vif2,3,5000)
##Mean vif seq=i## i=1,2,…,10
ai<-seq(i,5000,10)
aai<-mvif2[,ai]
```

```
vi_50<-mean(aai[1,])
vi_100<-mean(aai[2,])
vi_200<-mean(aai[3,])
Vi<-cbind(vi_50,vi_100,vi_200)

VIF กรณีตัวแปรอิสระ 3 ตัว
mvif3<-matrix(vif3,9,63000)
##Mean vif seq=i## i=1,2,…,14
bi<-seq(i,7000,14)
bbi<-mvif3[,bi]
vi1_50<-mean(bbi[1,])
vi1_100<-mean(bbi[2,])
vi1_200<-mean(bbi[3,])
vi2_50<-mean(bbi[4,])
vi2_100<-mean(bbi[5,])
vi2_200<-mean(bbi[6,])
vi3_50<-mean(bbi[7,])
vi3_100<-mean(bbi[8,])
vi3_200<-mean(bbi[9,])
V3_i<-cbind(vi1_50,vi1_100,vi1_200,vi2_50,vi2_100,vi2_200,vi3_50,vi3_100,vi3_200)
VIF_2<-rbind(V1,V2,V3,V4,V5,V6,V7,V8,V9,V10)
VIF_3<-rbind(V3_1,V3_2,V3_3,V3_4,V3_5,V3_6,V3_7,V3_8,V3_9,V3_10,V3_11,V3_12,
V3_13,V3_14)
```

2.การแก้ไขปัญหาพหุสัมพันธ์โดยใช้วิธีตัวประกอบหลัก

```
library(fmsb)
library(MASS)
library(perturb)
library(psych)

VIF_2<-c()
SE_2<-c()
BIAS_2<-c()

VIF_3<-c()
SE_3<-c()
BIAS_3<-c()
N=1000
for (i in 1:N){
rho_x<-c(0.9,0.91,0.92,0.93,0.94,0.95,0.96,0.97,0.98,0.99)
VIF2<-c()
SE2<-c()
BIAS2<-c()
for(t in 1:length(rho_x)){
sd_1<-sqrt(2)
sd_2<-sqrt(2)
z1<-rnorm(50)
z2<-rnorm(50)
z3<-rnorm(50)
z4<-rnorm(50)
z5<-c(z1,z3)
```

```
z6<-c(z2,z4)
z7<-rnorm(100)
z8<-rnorm(100)
z9<-c(z1,z3,z7)
z10<-c(z2,z4,z8)
e1<-rnorm(50,0,sqrt(2))
e2<-rnorm(50,0,sqrt(2))
e3<-c(e1,e2)
e4<-rnorm(100,0,sqrt(2))
e5<-c(e1,e2,e4)

##n=50##
Rho_X12<-rho_x[t]
x2_501<-z1
x2_502<-((Rho_X12)*z1)+((sqrt(1-(Rho_X12^2))*z2))
x250_1<-x2_501*sd_1
x250_2<-x2_502*sd_2

y2.50<-2+2*x250_1+2*x250_2+e1
##OLS##
lmdata2.50<-lm(y2.50~x250_1+x250_2)
coef50_2ols<-coef(lmdata2.50)

##check mullticollinearity with VIF  ##
dat2.50X<-data.frame(x250_1,x250_2)
X2.50<-as.matrix(dat2.50X)
vif2_x501<-VIF(lm(x250_1~x250_2,data=dat2.50X))
```

```
        vif2_x502<-VIF(lm(x250_2~x250_1,data=dat2.50X))
        vif2.50<-ifelse(vif2_x501>10|vif2_x502>10,1,0)
##Find var-cov matrix b(ols)##
        bhat2.50_0<-coef50_2ols[1]
        bhat2.50_1<-coef50_2ols[2]
        bhat2.50_2<-coef50_2ols[3]
        yhat2.50<-bhat2.50_0+bhat2.50_1*x250_1+bhat2.50_2*x250_2

        diffy2.50<-y2.50-yhat2.50
        error2.50<-as.matrix(diffy2.50)
        mse2.50<-as.numeric(((t(error2.50)%*%error2.50))/(50-2-1))
        Dat2.50X<-data.frame(1,x250_1,x250_2)
        DX2.50<-as.matrix(Dat2.50X)
        inv50_2x<-ginv(t(DX2.50)%*%DX2.50)
        varb50_2ols<-mse2.50*inv50_2x
        varb50_2.0ols<-varb50_2ols[1,1]
        varb50_2.1ols<-varb50_2ols[2,2]
        varb50_2.2ols<-varb50_2ols[3,3]

        seb50_2.0ols<-sqrt(varb50_2.0ols)
        seb50_2.1ols<-sqrt(varb50_2.1ols)
        seb50_2.2ols<-sqrt(varb50_2.2ols)

##Bias ols##
        bias2.0_50ols<-abs(bhat2.50_0-2)
        bias2.1_50ols<-abs(bhat2.50_1-2)
        bias2.2_50ols<-abs(bhat2.50_2-2)
```

```
##PCA##
        co2.50<-cor(X2.50)
        e2.50 <- eigen(co2.50)
        w1_50.2<-e2.50$vectors[,1]
        w_1502<-matrix(c(w1_50.2),1,2)
        w2_50.2<-e2.50$vectors[,2]
        w_2502<-matrix(c(w2_50.2),1,2)
        standardize <- function(X) {(X - mean(X))/sd(X)}
        SX2.50<-apply(X2.50, MARGIN=2, FUN=standardize)

        St2.50_1<-SX2.50[,1]
        St2.50_2<-SX2.50[,2]

        pc2.50_1<-w_1502[1,1]*St2.50_1+w_1502[1,2]*St2.50_2
        pc2.50_2<-w_2502[1,1]*St2.50_1+w_2502[1,2]*St2.50_2

        lm.pca2.50<-lm(y2.50~pc2.50_1+pc2.50_2)
        coef50_2pca<-coef(lm.pca2.50)

##Find var-cov matrix b(pca)##
        bhatpc2.50_0<-coef50_2pca[1]
        bhatpc2.50_1<-coef50_2pca[2]
        bhatpc2.50_2<-coef50_2pca[3]

        yhatpc2.50<-bhatpc2.50_0+bhatpc2.50_1*pc2.50_1+bhatpc2.50_2*pc2.50_2
        diff_ypca2.50<-y2.50-yhatpc2.50
        error_pca2.50<-as.matrix(diff_ypca2.50)
```

```
mse2.50_pca<-as.numeric(((t(error_pca2.50)%*%error_pca2.50))/(50-2-1))
pc2.50<-data.frame(1,pc2.50_1,pc2.50_2)
PC2.50<-as.matrix(pc2.50)
inv50_2pca<-ginv(t(PC2.50)%*%PC2.50)
varb50_2pca<-mse2.50_pca*inv50_2pca
varb50_2.0pca<-varb50_2pca[1,1]
varb50_2.1pca<-varb50_2pca[2,2]
varb50_2.2pca<-varb50_2pca[3,3]

seb50_2.0pca<-sqrt(varb50_2.0pca)
seb50_2.1pca<-sqrt(varb50_2.1pca)
seb50_2.2pca<-sqrt(varb50_2.2pca)

##Bias pca##
bias2.0_50pca<-abs(bhatpc2.50_0-2)
bias2.1_50pca<-abs(bhatpc2.50_1-2)
bias2.2_50pca<-abs(bhatpc2.50_2-2)

##n=100##
x2_1001<-z5
x2_1002<-((Rho_X12)*z5)+((sqrt(1-(Rho_X12^2))*z6))
x2100_1<-x2_1001*sd_1
x2100_2<-x2_1002*sd_2

y2.100<-2+2*x2100_1+2*x2100_2+e3
```

```
##OLS##
        lmdata2.100<-lm(y2.100~x2100_1+x2100_2)
        coef100_2ols<-coef(lmdata2.100)

##check mullticollinearity with VIF  ##
        dat2.100X<-data.frame(x2100_1,x2100_2)
        X2.100<-as.matrix(dat2.100X)
        vif2_x1001<-VIF(lm(x2100_1~x2100_2,data=dat2.100X))
        vif2_x1002<-VIF(lm(x2100_2~x2100_1,data=dat2.100X))
        vif2.100<-ifelse(vif2_x1001>10|vif2_x1002>10,1,0)

##Find var-cov matrix b(ols)##
        bhat2.100_0<-coef100_2ols[1]
        bhat2.100_1<-coef100_2ols[2]
        bhat2.100_2<-coef100_2ols[3]

        yhat2.100<-bhat2.100_0+bhat2.100_1*x2100_1+bhat2.100_2*x2100_2

        diffy2.100<-y2.100-yhat2.100
        error2.100<-as.matrix(diffy2.100)
        mse2.100<-as.numeric(((t(error2.100)%*%error2.100))/(100-2-1))
        Dat2.100X<-data.frame(1,x2100_1,x2100_2)
        DX2.100<-as.matrix(Dat2.100X)
        inv100_2x<-ginv(t(DX2.100)%*%DX2.100)
        varb100_2ols<-mse2.100*inv100_2x
        varb100_2.0ols<-varb100_2ols[1,1]
        varb100_2.1ols<-varb100_2ols[2,2]
```

```
varb100_2.2ols<-varb100_2ols[3,3]


seb100_2.0ols<-sqrt(varb100_2.0ols)
seb100_2.1ols<-sqrt(varb100_2.1ols)
seb100_2.2ols<-sqrt(varb100_2.2ols)

##Bias ols##
bias2.0_100ols<-abs(bhat2.100_0-2)
bias2.1_100ols<-abs(bhat2.100_1-2)
bias2.2_100ols<-abs(bhat2.100_2-2)

##PCA##
co2.100<-cor(X2.100)
e2.100 <- eigen(co2.100)
w1_100.2<-e2.100$vectors[,1]
w_11002<-matrix(c(w1_100.2),1,2)
w2_100.2<-e2.100$vectors[,2]
w_21002<-matrix(c(w2_100.2),1,2)

standardize <- function(X) {(X - mean(X))/sd(X)}
SX2.100<-apply(X2.100, MARGIN=2, FUN=standardize)

St2.100_1<-SX2.100[,1]
St2.100_2<-SX2.100[,2]

pc2.100_1<-w_11002[1,1]*St2.100_1+w_11002[1,2]*St2.100_2
pc2.100_2<-w_21002[1,1]*St2.100_1+w_21002[1,2]*St2.100_2
```

```
        lm.pca2.100<-lm(y2.100~pc2.100_1+pc2.100_2)
        coef100_2pca<-coef(lm.pca2.100)

##Find var-cov matrix b(pca)##
        bhatpc2.100_0<-coef100_2pca[1]
        bhatpc2.100_1<-coef100_2pca[2]
        bhatpc2.100_2<-coef100_2pca[3]

        yhatpc2.100<-bhatpc2.100_0+bhatpc2.100_1*pc2.100_1+
                bhatpc2.100_2*pc2.100_2
        diff_ypca2.100<-y2.100-yhatpc2.100
        error_pca2.100<-as.matrix(diff_ypca2.100)
        mse2.100_pca<-as.numeric(((t(error_pca2.100)%*%error_pca2.100))/(100-2-1))
        pc2.100<-data.frame(1,pc2.100_1,pc2.100_2)
        PC2.100<-as.matrix(pc2.100)
        inv100_2pca<-ginv(t(PC2.100)%*%PC2.100)
        varb100_2pca<-mse2.100_pca*inv100_2pca
        varb100_2.0pca<-varb100_2pca[1,1]
        varb100_2.1pca<-varb100_2pca[2,2]
        varb100_2.2pca<-varb100_2pca[3,3]

        seb100_2.0pca<-sqrt(varb100_2.0pca)
        seb100_2.1pca<-sqrt(varb100_2.1pca)
        seb100_2.2pca<-sqrt(varb100_2.2pca)

##Bias pca##
        bias2.0_100pca<-abs(bhatpc2.100_0-2)
```

```
        bias2.1_100pca<-abs(bhatpc2.100_1-2)
        bias2.2_100pca<-abs(bhatpc2.100_2-2)

##n=200##
        x2_2001<-z9
        x2_2002<-((Rho_X12)*z9)+((sqrt(1-(Rho_X12^2))*z10))

x2200_1<-x2_2001*sd_1
        x2200_2<-x2_2002*sd_2

        y2.200<-2+2*x2200_1+2*x2200_2+e5

##OLS##
        lmdata2.200<-lm(y2.200~x2200_1+x2200_2)
        coef200_2ols<-coef(lmdata2.200)

##check mullticollinearity with VIF  ##
        dat2.200X<-data.frame(x2200_1,x2200_2)
        X2.200<-as.matrix(dat2.200X)
        vif2_x2001<-VIF(lm(x2200_1~x2200_2,data=dat2.200X))
        vif2_x2002<-VIF(lm(x2200_2~x2200_1,data=dat2.200X))
        vif2.200<-ifelse(vif2_x2001>10|vif2_x2002>10,1,0)


##Find var-cov matrix b(ols)##
        bhat2.200_0<-coef200_2ols[1]
        bhat2.200_1<-coef200_2ols[2]
```

```
bhat2.200_2<-coef200_2ols[3]

yhat2.200<-bhat2.200_0+bhat2.200_1*x2200_1+bhat2.200_2*x2200_2

diffy2.200<-y2.200-yhat2.200
error2.200<-as.matrix(diffy2.200)
mse2.200<-as.numeric(((t(error2.200)%*%error2.200))/(200-2-1))
Dat2.200X<-data.frame(1,x2200_1,x2200_2)
DX2.200<-as.matrix(Dat2.200X)
inv200_2x<-ginv(t(DX2.200)%*%DX2.200)
varb200_2ols<-mse2.200*inv200_2x
varb200_2.0ols<-varb200_2ols[1,1]
varb200_2.1ols<-varb200_2ols[2,2]
varb200_2.2ols<-varb200_2ols[3,3]

seb200_2.0ols<-sqrt(varb200_2.0ols)
seb200_2.1ols<-sqrt(varb200_2.1ols)
seb200_2.2ols<-sqrt(varb200_2.2ols)

##Bias ols##
bias2.0_200ols<-abs(bhat2.200_0-2)
bias2.1_200ols<-abs(bhat2.200_1-2)
bias2.2_200ols<-abs(bhat2.200_2-2)

##PCA##
co2.200<-cor(X2.200)
e2.200 <- eigen(co2.200)
```

```
w1_200.2<-e2.200$vectors[,1]
w_12002<-matrix(c(w1_200.2),1,2)
w2_200.2<-e2.200$vectors[,2]
w_22002<-matrix(c(w2_200.2),1,2)
standardize <- function(X) {(X - mean(X))/sd(X)}
SX2.200<-apply(X2.200, MARGIN=2, FUN=standardize)

St2.200_1<-SX2.200[,1]
St2.200_2<-SX2.200[,2]

pc2.200_1<-w_12002[1,1]*St2.200_1+w_12002[1,2]*St2.200_2
pc2.200_2<-w_22002[1,1]*St2.200_1+w_22002[1,2]*St2.200_2

lm.pca2.200<-lm(y2.200~pc2.200_1+pc2.200_2)
coef200_2pca<-coef(lm.pca2.200)

##Find var-cov matrix b(pca)##
bhatpc2.200_0<-coef200_2pca[1]
bhatpc2.200_1<-coef200_2pca[2]
bhatpc2.200_2<-coef200_2pca[3]

yhatpc2.200<-bhatpc2.200_0+bhatpc2.200_1*pc2.200_1+
        bhatpc2.200_2*pc2.200_2
diff_ypca2.200<-y2.200-yhatpc2.200
error_pca2.200<-as.matrix(diff_ypca2.200)
mse2.200_pca<-as.numeric(((t(error_pca2.200)%*%error_pca2.200))/(200-2-1))
pc2.200<-data.frame(1,pc2.200_1,pc2.200_2)
```

```
PC2.200<-as.matrix(pc2.200)

inv200_2pca<-ginv(t(PC2.200)%*%PC2.200)
varb200_2pca<-mse2.200_pca*inv200_2pca
varb200_2.0pca<-varb200_2pca[1,1]
varb200_2.1pca<-varb200_2pca[2,2]
varb200_2.2pca<-varb200_2pca[3,3]

seb200_2.0pca<-sqrt(varb200_2.0pca)
seb200_2.1pca<-sqrt(varb200_2.1pca)
seb200_2.2pca<-sqrt(varb200_2.2pca)

##Bias pca##
bias2.0_200pca<-abs(bhatpc2.200_0-2)
bias2.1_200pca<-abs(bhatpc2.200_1-2)
bias2.2_200pca<-abs(bhatpc2.200_2-2)

VIF2<-c(VIF2,vif2.50,vif2.100,vif2.200)

SE2<-c(SE2,seb50_2.0ols,seb50_2.1ols,seb50_2.2ols,seb50_2.0pca,
seb50_2.1pca,seb50_2.2pca,seb100_2.0ols,seb100_2.1ols,seb100_2.2ols,
seb100_2.0pca,seb100_2.1pca,seb100_2.2pca,seb200_2.0ols,seb200_2.1ols,
seb200_2.2ols,seb200_2.0pca,seb200_2.1pca,seb200_2.2pca)

BIAS2<-c(BIAS2,bias2.0_50ols,bias2.1_50ols,bias2.2_50ols,bias2.0_50pca,
bias2.1_50pca,bias2.2_50pca,bias2.0_100ols,bias2.1_100ols,bias2.2_100ols,
bias2.0_100pca,bias2.1_100pca,bias2.2_100pca,bias2.0_200ols,bias2.1_200ols,
```

```
bias2.2_200ols,bias2.0_200pca,bias2.1_200pca,bias2.2_200pca)
        }
        rho<-c(0.91,0.93,0.95,0.97,0.99)
        t_rho<-t(combn(rho,3))
        newrho<-t_rho[c(-6),]
        r_r<-c(rep(0.91,3),rep(0.93,3),rep(0.95,3),rep(0.97,3),rep(0.99,3))
        mr_r<-matrix(c(r_r),3)
        tr_r<-t(mr_r)
        total<-rbind(newrho,tr_r)
        VIF3<-c()
        SE3<-c()
        BIAS3<-c()
for(p in 1:dim(total)[1]){
z1<-rnorm(50)
        z2<-rnorm(50)
        z3<-rnorm(50)
        z4<-rnorm(50)
        z5<-rnorm(50)
        z6<-rnorm(50)
        z7<-c(z1,z4)
        z8<-c(z2,z5)
        z9<-c(z3,z6)
        z10<-rnorm(100)
        z11<-rnorm(100)
        z12<-rnorm(100)
        z13<-c(z1,z4,z10)
        z14<-c(z2,z5,z11)
```

```
        z15<-c(z3,z6,z12)
        e1<-rnorm(50,0,sqrt(2))
        e2<-rnorm(50,0,sqrt(2))
        e3<-c(e1,e2)
        e4<-rnorm(100,0,sqrt(2))
        e5<-c(e1,e2,e4)
        RHo_X12<-total[p,1]
        RHo_X13<-total[p,2]
        RHo_X23<-total[p,3]
##n=50,p=3##
        x3_501<-z1
        x3_502<-((RHo_X12)*z1)+((sqrt(1-(RHo_X12^2))*z2))
        x3_503<-((RHo_X13)*z1)+((((RHo_X23-(RHo_X12*RHo_X13))/(sqrt(1-
(RHo_X12^2))))*z2)+((sqrt(1-(RHo_X13^2)-(((RHo_X23-(RHo_X12*RHo_X13))^2)/(1-
(RHo_X12^2))))*z3)))

        x350_1<-x3_501*sd_1
        x350_2<-x3_502*sd_2
        x350_3<-x3_503*sd_3

        y3.50<-2+2*x350_1+2*x350_2+2*x350_3+e1

##OLS##
        lmdata3.50<-lm(y3.50~x350_1+x350_2+x350_3)
        coef50_3ols<-coef(lmdata3.50)
```

```
##check mullticollinearity with VIF  ##
        dat3.50X<-data.frame(x350_1,x350_2,x350_3)
        X3.50<-as.matrix(dat3.50X)
        vif3_x501<-VIF(lm(x350_1~x350_2+x350_3,data=dat3.50X))
        vif3_x502<-VIF(lm(x350_2~x350_1+x350_3,data=dat3.50X))
        vif3_x503<-VIF(lm(x350_3~x350_1+x350_2,data=dat3.50X))
        vif3.50<-ifelse(vif3_x501>10|vif3_x502>10|vif3_x503>10,1,0)

##Find var-cov matrix b(ols)##
        bhat3.50_0<-coef50_3ols[1]
        bhat3.50_1<-coef50_3ols[2]
        bhat3.50_2<-coef50_3ols[3]
        bhat3.50_3<-coef50_3ols[4]
        yhat3.50<-bhat3.50_0+bhat3.50_1*x350_1+bhat3.50_2*x350_2+
                bhat3.50_3*x350_3
        diffy3.50<-y3.50-yhat3.50
        error3.50<-as.matrix(diffy3.50)
        mse3.50<-as.numeric(((t(error3.50)%*%error3.50))/(50-3-1))
        Dat3.50X<-data.frame(1,x350_1,x350_2,x350_3)
        DX3.50<-as.matrix(Dat3.50X)
        inv50_3x<-ginv(t(DX3.50)%*%DX3.50)
        varb50_3ols<-mse3.50*inv50_3x
        varb50_3.0ols<-varb50_3ols[1,1]
        varb50_3.1ols<-varb50_3ols[2,2]
        varb50_3.2ols<-varb50_3ols[3,3]
        varb50_3.3ols<-varb50_3ols[4,4]
```

```
seb50_3.0ols<-sqrt(varb50_3.0ols)
seb50_3.1ols<-sqrt(varb50_3.1ols)
seb50_3.2ols<-sqrt(varb50_3.2ols)
seb50_3.3ols<-sqrt(varb50_3.3ols)

##Bias ols##
bias3.0_50ols<-abs(bhat3.50_0-2)
bias3.1_50ols<-abs(bhat3.50_1-2)
bias3.2_50ols<-abs(bhat3.50_2-2)
bias3.3_50ols<-abs(bhat3.50_3-2)

##PCA##
co3.50<-cor(X3.50)
e3.50 <- eigen(co3.50)
w1_50.3<-e3.50$vectors[,1]
w_1503<-matrix(c(w1_50.3),1,3)
w2_50.3<-e3.50$vectors[,2]
w_2503<-matrix(c(w2_50.3),1,3)
w3_50.3<-e3.50$vectors[,3]
w_3503<-matrix(c(w3_50.3),1,3)

standardize <- function(X) {(X - mean(X))/sd(X)}
SX3.50<-apply(X3.50, MARGIN=2, FUN=standardize)

St3.50_1<-SX3.50[,1]
St3.50_2<-SX3.50[,2]
St3.50_3<-SX3.50[,3]
```

```
pc3.50_1<-w_1503[1,1]*St3.50_1+w_1503[1,2]*St3.50_2+w_1503[1,3]*St3.50_3
pc3.50_2<-w_2503[1,1]*St3.50_1+w_2503[1,2]*St3.50_2+w_2503[1,3]*St3.50_3
pc3.50_3<-w_3503[1,1]*St3.50_1+w_3503[1,2]*St3.50_2+w_3503[1,3]*St3.50_3

lm.pca3.50<-lm(y3.50~pc3.50_1+pc3.50_2+pc3.50_3)
coef50_3pca<-coef(lm.pca3.50)

##Find var-cov matrix b(pca)##
bhatpc3.50_0<-coef50_3pca[1]
bhatpc3.50_1<-coef50_3pca[2]
bhatpc3.50_2<-coef50_3pca[3]
bhatpc3.50_3<-coef50_3pca[4]

yhatpc3.50<-bhatpc3.50_0+bhatpc3.50_1*pc3.50_1+bhatpc3.50_2*pc3.50_2
          +bhatpc3.50_3*pc3.50_3
diff_ypca3.50<-y3.50-yhatpc3.50
error_pca3.50<-as.matrix(diff_ypca3.50)
mse3.50_pca<-as.numeric(((t(error_pca3.50)%*%error_pca3.50))/(50-3-1))
pc3.50<-data.frame(1,pc3.50_1,pc3.50_2,pc3.50_3)
PC3.50<-as.matrix(pc3.50)
inv50_3pca<-ginv(t(PC3.50)%*%PC3.50)
varb50_3pca<-mse3.50_pca*inv50_3pca
varb50_3.0pca<-varb50_3pca[1,1]
varb50_3.1pca<-varb50_3pca[2,2]
varb50_3.2pca<-varb50_3pca[3,3]
varb50_3.3pca<-varb50_3pca[4,4]
```

```
seb50_3.0pca<-sqrt(varb50_3.0pca)
seb50_3.1pca<-sqrt(varb50_3.1pca)
seb50_3.2pca<-sqrt(varb50_3.2pca)
seb50_3.3pca<-sqrt(varb50_3.3pca)

##Bias pca##
bias3.0_50pca<-abs(bhatpc3.50_0-2)
bias3.1_50pca<-abs(bhatpc3.50_1-2)
bias3.2_50pca<-abs(bhatpc3.50_2-2)
bias3.3_50pca<-abs(bhatpc3.50_3-2)

##n=100,p=3##
x3_1001<-z7
x3_1002<-((RHo_X12)*z7)+((sqrt(1-(RHo_X12^2))*z8))
x3_1003<-((RHo_X13)*z7)+((((RHo_X23-(RHo_X12*RHo_X13))/(sqrt(1-
(RHo_X12^2))))*z8)+((sqrt(1-(RHo_X13^2)-(((RHo_X23-(RHo_X12*RHo_X13))^2)/(1-
(RHo_X12^2))))*z9)))

x3100_1<-x3_1001*sd_1
x3100_2<-x3_1002*sd_2
x3100_3<-x3_1003*sd_3

y3.100<-2+2*x3100_1+2*x3100_2+2*x3100_3+e3

##OLS##
lmdata3.100<-lm(y3.100~x3100_1+x3100_2+x3100_3)
coef100_3ols<-coef(lmdata3.100)
```

```
##check mullticollinearity with VIF  ##
        dat3.100X<-data.frame(x3100_1,x3100_2,x3100_3)
        X3.100<-as.matrix(dat3.100X)
        vif3_x1001<-VIF(lm(x3100_1~x3100_2+x3100_3,data=dat3.100X))
        vif3_x1002<-VIF(lm(x3100_2~x3100_1+x3100_3,data=dat3.100X))
        vif3_x1003<-VIF(lm(x3100_3~x3100_1+x3100_2,data=dat3.100X))
        vif3.100<-ifelse(vif3_x1001>10|vif3_x1002>10|vif3_x1003>10,1,0)

##Find var-cov matrix b(ols)##
        bhat3.100_0<-coef100_3ols[1]
        bhat3.100_1<-coef100_3ols[2]
        bhat3.100_2<-coef100_3ols[3]
        bhat3.100_3<-coef100_3ols[4]
        yhat3.100<-bhat3.100_0+bhat3.100_1*x3100_1+bhat3.100_2*x3100_2
            +bhat3.100_3*x3100_3
        diffy3.100<-y3.100-yhat3.100
        error3.100<-as.matrix(diffy3.100)
        mse3.100<-as.numeric(((t(error3.100)%*%error3.100))/(100-3-1))
        Dat3.100X<-data.frame(1,x3100_1,x3100_2,x3100_3)
        DX3.100<-as.matrix(Dat3.100X)
        inv100_3x<-ginv(t(DX3.100)%*%DX3.100)
        varb100_3ols<-mse3.100*inv100_3x
        varb100_3.0ols<-varb100_3ols[1,1]
        varb100_3.1ols<-varb100_3ols[2,2]
        varb100_3.2ols<-varb100_3ols[3,3]
        varb100_3.3ols<-varb100_3ols[4,4]
```

```
seb100_3.0ols<-sqrt(varb100_3.0ols)
seb100_3.1ols<-sqrt(varb100_3.1ols)
seb100_3.2ols<-sqrt(varb100_3.2ols)
seb100_3.3ols<-sqrt(varb100_3.3ols)

##Bias ols##
bias3.0_100ols<-abs(bhat3.100_0-2)
bias3.1_100ols<-abs(bhat3.100_1-2)
bias3.2_100ols<-abs(bhat3.100_2-2)
bias3.3_100ols<-abs(bhat3.100_3-2)

##PCA##
co3.100<-cor(X3.100)
e3.100 <- eigen(co3.100)
w1_100.3<-e3.100$vectors[,1]
w_11003<-matrix(c(w1_100.3),1,3)
w2_100.3<-e3.100$vectors[,2]
w_21003<-matrix(c(w2_100.3),1,3)
w3_100.3<-e3.100$vectors[,3]
w_31003<-matrix(c(w3_100.3),1,3)

standardize <- function(X) {(X - mean(X))/sd(X)}
SX3.100<-apply(X3.100, MARGIN=2, FUN=standardize)

St3.100_1<-SX3.100[,1]
St3.100_2<-SX3.100[,2]
St3.100_3<-SX3.100[,3]
```

```
        pc3.100_1<-w_11003[1,1]*St3.100_1+w_11003[1,2]*St3.100_2+
w_11003[1,3]*St3.100_3
        pc3.100_2<-w_21003[1,1]*St3.100_1+w_21003[1,2]*St3.100_2+
w_21003[1,3]*St3.100_3
        pc3.100_3<-w_31003[1,1]*St3.100_1+w_31003[1,2]*St3.100_2+
w_31003[1,3]*St3.100_3

        lm.pca3.100<-lm(y3.100~pc3.100_1+pc3.100_2+pc3.100_3)
        coef100_3pca<-coef(lm.pca3.100)

##Find var-cov matrix b(pca)##
        bhatpc3.100_0<-coef100_3pca[1]
        bhatpc3.100_1<-coef100_3pca[2]
        bhatpc3.100_2<-coef100_3pca[3]
        bhatpc3.100_3<-coef100_3pca[4]

        yhatpc3.100<-bhatpc3.100_0+bhatpc3.100_1*pc3.100_1+
bhatpc3.100_2*pc3.100_2+bhatpc3.100_3*pc3.100_3
        diff_ypca3.100<-y3.100-yhatpc3.100
        error_pca3.100<-as.matrix(diff_ypca3.100)
        mse3.100_pca<-as.numeric(((t(error_pca3.100)%*%error_pca3.100))/(100-3-1))

        pc3.100<-data.frame(1,pc3.100_1,pc3.100_2,pc3.100_3)
        PC3.100<-as.matrix(pc3.100)
        inv100_3pca<-ginv(t(PC3.100)%*%PC3.100)
        varb100_3pca<-mse3.100_pca*inv100_3pca
        varb100_3.0pca<-varb100_3pca[1,1]
```

```
varb100_3.1pca<-varb100_3pca[2,2]
varb100_3.2pca<-varb100_3pca[3,3]
varb100_3.3pca<-varb100_3pca[4,4]

seb100_3.0pca<-sqrt(varb100_3.0pca)
seb100_3.1pca<-sqrt(varb100_3.1pca)
seb100_3.2pca<-sqrt(varb100_3.2pca)
seb100_3.3pca<-sqrt(varb100_3.3pca)

##Bias pca##
bias3.0_100pca<-abs(bhatpc3.100_0-2)
bias3.1_100pca<-abs(bhatpc3.100_1-2)
bias3.2_100pca<-abs(bhatpc3.100_2-2)
bias3.3_100pca<-abs(bhatpc3.100_3-2)
##n=200,p=3##
x3_2001<-z13
x3_2002<-((RHo_X12)*z13)+((sqrt(1-(RHo_X12^2))*z14))
x3_2003<-((RHo_X13)*z13)+((((RHo_X23-(RHo_X12*RHo_X13))/(sqrt(1-
(RHo_X12^2))))*z14)+((sqrt(1-(RHo_X13^2)-(((RHo_X23-(RHo_X12*RHo_X13))^2)/(1-
(RHo_X12^2))))*z15)))

x3200_1<-x3_2001*sd_1
x3200_2<-x3_2002*sd_2
x3200_3<-x3_2003*sd_3

y3.200<-2+2*x3200_1+2*x3200_2+2*x3200_3+e5
```

```
##OLS##
        lmdata3.200<-lm(y3.200~x3200_1+x3200_2+x3200_3)
        coef200_3ols<-coef(lmdata3.200)

##check mullticollinearity with VIF##
        dat3.200X<-data.frame(x3200_1,x3200_2,x3200_3)
        X3.200<-as.matrix(dat3.200X)
        vif3_x2001<-VIF(lm(x3200_1~x3200_2+x3200_3,data=dat3.200X))
        vif3_x2002<-VIF(lm(x3200_2~x3200_1+x3200_3,data=dat3.200X))
        vif3_x2003<-VIF(lm(x3200_3~x3200_1+x3200_2,data=dat3.200X))
        vif3.200<-ifelse(vif3_x2001>10|vif3_x2002>10|vif3_x2003>10,1,0)

##Find var-cov matrix b(ols)##
        bhat3.200_0<-coef200_3ols[1]
        bhat3.200_1<-coef200_3ols[2]
        bhat3.200_2<-coef200_3ols[3]
        bhat3.200_3<-coef200_3ols[4]

yhat3.200<-bhat3.200_0+bhat3.200_1*x3200_1+bhat3.200_2*x3200_2
+bhat3.200_3*x3200_3
        diffy3.200<-y3.200-yhat3.200
        error3.200<-as.matrix(diffy3.200)
        mse3.200<-as.numeric(((t(error3.200)%*%error3.200))/(200-3-1))
        Dat3.200X<-data.frame(1,x3200_1,x3200_2,x3200_3)
        DX3.200<-as.matrix(Dat3.200X)
        inv200_3x<-ginv(t(DX3.200)%*%DX3.200)
        varb200_3ols<-mse3.200*inv200_3x
```

```
varb200_3.0ols<-varb200_3ols[1,1]
varb200_3.1ols<-varb200_3ols[2,2]
varb200_3.2ols<-varb200_3ols[3,3]
varb200_3.3ols<-varb200_3ols[4,4]
seb200_3.0ols<-sqrt(varb200_3.0ols)
seb200_3.1ols<-sqrt(varb200_3.1ols)
seb200_3.2ols<-sqrt(varb200_3.2ols)
seb200_3.3ols<-sqrt(varb200_3.3ols)

##Bias ols##
bias3.0_200ols<-abs(bhat3.200_0-2)
bias3.1_200ols<-abs(bhat3.200_1-2)
bias3.2_200ols<-abs(bhat3.200_2-2)
bias3.3_200ols<-abs(bhat3.200_3-2)
##PCA##
co3.200<-cor(X3.200)
e3.200 <- eigen(co3.200)
w1_200.3<-e3.200$vectors[,1]
w_12003<-matrix(c(w1_200.3),1,3)
w2_200.3<-e3.200$vectors[,2]
w_22003<-matrix(c(w2_200.3),1,3)
w3_200.3<-e3.200$vectors[,3]
w_32003<-matrix(c(w3_200.3),1,3)

standardize <- function(X) {(X - mean(X))/sd(X)}
SX3.200<-apply(X3.200, MARGIN=2, FUN=standardize)
```

```
St3.200_1<-SX3.200[,1]
St3.200_2<-SX3.200[,2]
St3.200_3<-SX3.200[,3]

pc3.200_1<-w_12003[1,1]*St3.200_1+w_12003[1,2]*St3.200_2+
w_12003[1,3]*St3.200_3
pc3.200_2<-w_22003[1,1]*St3.200_1+w_22003[1,2]*St3.200_2+
w_22003[1,3]*St3.200_3
pc3.200_3<-w_32003[1,1]*St3.200_1+w_32003[1,2]*St3.200_2+
w_32003[1,3]*St3.200_3
lm.pca3.200<-lm(y3.200~pc3.200_1+pc3.200_2+pc3.200_3)
coef200_3pca<-coef(lm.pca3.200)

##Find var-cov matrix b(pca)##
bhatpc3.200_0<-coef200_3pca[1]
bhatpc3.200_1<-coef200_3pca[2]
bhatpc3.200_2<-coef200_3pca[3]
bhatpc3.200_3<-coef200_3pca[4]

yhatpc3.200<-bhatpc3.200_0+bhatpc3.200_1*pc3.200_1+
bhatpc3.200_2*pc3.200_2+bhatpc3.200_3*pc3.200_3
diff_ypca3.200<-y3.200-yhatpc3.200
error_pca3.200<-as.matrix(diff_ypca3.200)
mse3.200_pca<-as.numeric(((t(error_pca3.200)%*%error_pca3.200))/(200-3-1))
pc3.200<-data.frame(1,pc3.200_1,pc3.200_2,pc3.200_3)
PC3.200<-as.matrix(pc3.200)
inv200_3pca<-ginv(t(PC3.200)%*%PC3.200)
```

```
varb200_3pca<-mse3.200_pca*inv200_3pca
varb200_3.0pca<-varb200_3pca[1,1]
varb200_3.1pca<-varb200_3pca[2,2]
varb200_3.2pca<-varb200_3pca[3,3]
varb200_3.3pca<-varb200_3pca[4,4]

seb200_3.0pca<-sqrt(varb200_3.0pca)
seb200_3.1pca<-sqrt(varb200_3.1pca)
seb200_3.2pca<-sqrt(varb200_3.2pca)
seb200_3.3pca<-sqrt(varb200_3.3pca)

##Bias pca##
bias3.0_200pca<-abs(bhatpc3.200_0-2)
bias3.1_200pca<-abs(bhatpc3.200_1-2)
bias3.2_200pca<-abs(bhatpc3.200_2-2)
bias3.3_200pca<-abs(bhatpc3.200_3-2)

VIF3<-c(VIF3,vif3.50,vif3.100,vif3.200)

SE3<-c(SE3,seb50_3.0ols,seb50_3.1ols,seb50_3.2ols,seb50_3.3ols,
seb50_3.0pca,seb50_3.1pca,seb50_3.2pca,seb50_3.3pca,seb100_3.0ols,
seb100_3.1ols,seb100_3.2ols,seb100_3.3ols,seb100_3.0pca,seb100_3.1pca,
seb100_3.2pca,seb100_3.3pca,seb200_3.0ols,seb200_3.1ols,seb200_3.2ols,
seb200_3.3ols,seb200_3.0pca,seb200_3.1pca,seb200_3.2pca,seb200_3.3pca)

BIAS3<-c(BIAS3,bias3.0_50ols,bias3.1_50ols,bias3.2_50ols,bias3.3_50ols,
bias3.0_50pca,bias3.1_50pca,bias3.2_50pca,bias3.3_50pca,bias3.0_100ols,
```

```
bias3.1_100ols,bias3.2_100ols,bias3.3_100ols,bias3.0_100pca,bias3.1_100pca,
bias3.2_100pca,bias3.3_100pca,bias3.0_200ols,bias3.1_200ols,bias3.2_200ols,
bias3.3_200ols,bias3.0_200pca,bias3.1_200pca,bias3.2_200pca,bias3.3_200pca)
}
        VIF_2<-c(VIF_2,VIF2)
        SE_2<-c(SE_2,SE2)
        BIAS_2<-c(BIAS_2,BIAS2)

        VIF_3<-c(VIF_3,VIF3)
        SE_3<-c(SE_3,SE3)
        BIAS_3<-c(BIAS_3,BIAS3)
        pie(c(i,N-i),c(i,N-i),radius=1,clockwise=T)
}
```

กรณีตัวแปรอิสระ 2 ตัว

```
mvif2<-matrix(VIF_2,3,10000)
ols2_se<-matrix(SE_2,18,10000)
bias2<-matrix(BIAS_2,18,10000)

##seq = i## ;i=1,2,3,…10
Ni<-seq(i,10000,10)
Vi<-mvif2[,Ni]
Oi<-ols2_se[,Ni]
Bi<-bias2[,Ni]

Pi_vif50<-mean(Vi[1,]>=1)*100
Pi_vif100<-mean(Vi[2,]>=1)*100
Pi_vif200<-mean(Vi[3,]>=1)*100
```

```
Oi1_50<-mean(Oi[1,])
Oi2_50<-mean(Oi[2,])
Oi3_50<-mean(Oi[3,])
MeOi_50<-mean(Oi1_50,Oi2_50,Oi3_50)
Pi1_50<-mean(Oi[4,])
Pi2_50<-mean(Oi[5,])
Pi3_50<-mean(Oi[6,])
MePi_50<-mean(Pi1_50,Pi2_50,Pi3_50)

Oi1_100<-mean(Oi[7,])
Oi2_100<-mean(Oi[8,])
Oi3_100<-mean(Oi[9,])
MeOi_100<-mean(Oi1_100,Oi2_100,Oi3_100)
Pi1_100<-mean(Oi[10,])
Pi2_100<-mean(Oi[11,])
Pi3_100<-mean(Oi[12,])
MePi_100<-mean(Pi1_100,Pi2_100,Pi3_100)

Oi1_200<-mean(Oi[13,])
Oi2_200<-mean(Oi[14,])
Oi3_200<-mean(Oi[15,])
MeOi_200<-mean(Oi1_200,Oi2_200,Oi3_200)
Pi1_200<-mean(Oi[16,])
Pi2_200<-mean(Oi[17,])
Pi3_200<-mean(Oi[18,])
MePi_200<-mean(Pi1_200,Pi2_200,Pi3_200)
```

```
Bi1_50<-mean(Bi[1,])
Bi2_50<-mean(Bi[2,])
Bi3_50<-mean(Bi[3,])
MeBi_50<-mean(Bi1_50,Bi2_50,Bi3_50)
BPi1_50<-mean(Bi[4,])
BPi2_50<-mean(Bi[5,])
BPi3_50<-mean(Bi[6,])
MeBPi_50<-mean(BPi1_50,BPi2_50,BPi3_50)

Bi1_100<-mean(Bi[7,])
Bi2_100<-mean(Bi[8,])
Bi3_100<-mean(Bi[9,])
MeBi_100<-mean(Bi1_100,Bi2_100,Bi3_100)
BPi1_100<-mean(Bi [10,])
BPi2_100<-mean(Bi [11,])
BPi3_100<-mean(Bi [12,])
MeBPi_100<-mean(BPi1_100,BPi2_100,BPi3_100)

Bi1_200<-mean(Bi [13,])
Bi2_200<-mean(Bi [14,])
Bi3_200<-mean(Bi [15,])
MeBi_200<-mean(Bi1_200,Bi2_200,Bi3_200)
BPi1_200<-mean(Bi [16,])
BPi2_200<-mean(Bi [17,])
BPi3_200<-mean(Bi [18,])
MeBPi_200<-mean(BPi1_200,BPi2_200,BPi3_200)
Pi_vif50<-mean(Vi[1,]>=1)*100
```

```
Pi_vif100<-mean(Vi[2,]>=1)*100
Pi_vif200<-mean(Vi[3,]>=1)*100

rowVifi<-c(Pi_vif50, Pi_vif100, Pi_vif200)
rowMeansei<-c(MeOi_50,MePi_50,MeOi_100,MePi_100,MeOi_200,MePi_200)
rowMbiasi<-c(MeBi_50,MeBPi_50,MeBi_100,MeBPi_100,MeBi_200,MeBPi_200)

Table1<-rbind(rowVif1,rowVif2,rowVif3,rowVif4,rowVif5,rowVif6,rowVif7,rowVif8,rowVif9,
rowVif10)
Table2<- rbind(rowMeanse1,rowMeanse2,rowMeanse3,rowMeanse4,rowMeanse5,
rowMeanse6,rowMeanse7,rowMeanse8,rowMeanse9,rowMeanse10)

Table3<- rbind(rowMbias1,rowMbias2,rowMbias3,rowMbias4,rowMbias5,rowMbias6,
rowMbias7,rowMbias8,rowMbias9,rowMbias10)

กรณีตัวแปรอิสระ 3 ตัว
mvif3<-matrix(VIF_3,3,14000)
ols3_se<-matrix(SE_3,24,14000)
bias3<-matrix(BIAS_3,24,14000)

##seq = j## j=1,2,…,14
NNj<-seq(j,14000,14)
VVj<-mvif3[,NNj]
OOj<-ols3_se[,NNj]
BBj<-bias3[,NNj]

Pj_vif50.3<-mean(VVj[1,]>=1)*100
```

```
Pj_vif100.3<-mean(VVj[2,]>=1)*100
Pj_vif200.3<-mean(VVj[3,]>=1)*100

Oj1_50.3<-mean(OOj[1,])
Oj2_50.3<-mean(OOj[2,])
Oj3_50.3<-mean(OOj[3,])
Oj4_50.3<-mean(OOj[4,])
MeOj_50.3<-mean(Oj1_50.3,Oj2_50.3,Oj3_50.3,Oj4_50.3)
Pj1_50.3<-mean(OOj[5,])
Pj2_50.3<-mean(OOj[6,])
Pj3_50.3<-mean(OOj[7,])
Pj4_50.3<-mean(OOj[8,])
MePj_50.3<-mean(Pj1_50.3,Pj2_50.3,Pj3_50.3,Pj4_50.3)

Oj1_100.3<-mean(OOj[9,])
Oj2_100.3<-mean(OOj[10,])
Oj3_100.3<-mean(OOj[11,])
Oj4_100.3<-mean(OOj[12,])
MeOj_100.3<-mean(Oj1_100.3,Oj2_100.3,Oj3_100.3,Oj4_100.3)
Pj1_100.3<-mean(OOj[13,])
Pj2_100.3<-mean(OOj[14,])
Pj3_100.3<-mean(OOj[15,])
Pj4_100.3<-mean(OOj[16,])
MePj_100.3<-mean(Pj1_100.3,Pj2_100.3,Pj3_100.3,Pj4_100.3)

Oj1_200.3<-mean(OOj[17,])
Oj2_200.3<-mean(OOj[18,])
```

```
Oj3_200.3<-mean(OOj[19,])
Oj4_200.3<-mean(OOj[20,])
MeOj_200.3<-mean(Oj1_200.3,Oj2_200.3,Oj3_200.3,Oj4_200.3)
Pj1_200.3<-mean(OOj[21,])
Pj2_200.3<-mean(OOj[22,])
Pj3_200.3<-mean(OOj[23,])
Pj4_200.3<-mean(OOj[24,])
MePj_200.3<-mean(Pj1_200.3,Pj2_200.3,Pj3_200.3,Pj4_200.3)

Bj1_50.3<-mean(BBj[1,])
Bj2_50.3<-mean(BBj[2,])
Bj3_50.3<-mean(BBj[3,])
Bj4_50.3<-mean(BBj[4,])
MeBj_50.3<-mean(Bj1_50.3,Bj2_50.3,Bj3_50.3,Bj4_50.3)
BPj1_50.3<-mean(BBj[5,])
BPj2_50.3<-mean(BBj[6,])
BPj3_50.3<-mean(BBj[7,])
BPj4_50.3<-mean(BBj[8,])
MeBPj_50.3<-mean(BPj1_50.3,BPj2_50.3,BPj3_50.3,BPj4_50.3)

Bj1_100.3<-mean(BBj[9,])
Bj2_100.3<-mean(BBj[10,])
Bj3_100.3<-mean(BBj[11,])
Bj4_100.3<-mean(BBj[12,])
MeBj_100.3<-mean(Bj1_100.3,Bj2_100.3,Bj3_100.3,Bj4_100.3)
BPj1_100.3<-mean(BBj[13,])
BPj2_100.3<-mean(BBj[14,])
```

```
BPj3_100.3<-mean(BBj[15,])
BPj4_100.3<-mean(BBj[16,])
MeBPj_100.3<-mean(BPj1_100.3,BPj2_100.3,BPj3_100.3,BPj4_100.3)

Bj1_200.3<-mean(BBj[17,])
Bj2_200.3<-mean(BBj[18,])
Bj3_200.3<-mean(BBj[19,])
Bj4_200.3<-mean(BBj[20,])
MeBj_200.3<-mean(Bj1_200.3,Bj2_200.3,Bj3_200.3,Bj4_200.3)
BPj1_200.3<-mean(BBj[21,])
BPj2_200.3<-mean(BBj[22,])
BPj3_200.3<-mean(BBj[23,])
BPj4_200.3<-mean(BBj[24,])
MeBPj_200.3<-mean(BPj1_200.3,BPj2_200.3,BPj3_200.3,BPj4_200.3)

rowVifj.3<-c(Pj_vif50.3, Pj_vif100.3, Pj_vif200.3)
rowMeansej.3<-c(MeOj_50.3,MePj_50.3,MeOj_100.3,MePj_100.3,MeOj_200.3,
MePj_200.3)
rowMbiasj.3<-c(MeBj_50.3,MeBPj_50.3,MeBj_100.3,MeBPj_100.3,MeBj_200.3,
MeBPj_200.3)

Table4<-rbind(rowVif1.3,rowVif2.3,rowVif3.3,rowVif4.3,rowVif5.3,rowVif6.3,rowVif7.3,
rowVif8.3, rowVif9.3,rowVif10.3,rowVif11.3,rowVif12.3,rowVif13.3,rowVif14.3)

Table5<- rbind(rowMeanse1.3,rowMeanse2.3,rowMeanse3.3,rowMeanse4.3,
rowMeanse5.3,rowMeanse6.3,rowMeanse7.3,rowMeanse8.3,rowMeanse9.3,
rowMeanse10.3,rowMeanse11.3,rowMeanse12.3,rowMeanse13.3,rowMeanse14.3)
```

```
Table6<- rbind(rowMbias1.3,rowMbias2.3,rowMbias3.3,rowMbias4.3,rowMbias5.3,
rowMbias6.3,rowMbias7.3,rowMbias8.3,rowMbias9.3,rowMbias10.3,rowMbias11.3,
rowMbias12.3,rowMbias13.3,rowMbias14.3)
```

ประวัติผู้เขียนวิทยานิพนธ์

นายกิตติ วารีรัตน์ เกิดวันที่ 21 สิงหาคม พ.ศ.2530 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยมหิดล ในปีการศึกษา 2552 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต (สต.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2553