ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา

นางสาว สุกัญญา บุญมา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2551

LOGISTIC REGRESSION MODEL WITH GAUSSIAN COPULA

Miss Sukanya Bunma

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 2008

หัวข้อวิทยานิพนธ์	ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา
โดย	นางสาวสุกัญญา บุญมา
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.... คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร. อรรณพ ตันละมัย)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(รองศาสตราจารย์ ดร. ธีระพร วีระถาวร)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์)

.... กรรมการภายนอกมหาวิทยาลัย

(ดร. โอวาท สุนันท์)

.... กรรมการ

(รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา)

.... กรรมการ

(รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา)

สุกัญญา บุญมา : ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา.
(LOGISTIC REGRESSION MODEL WITH GAUSSIAN COPULA)
อ. ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ. ดร. เสกสรร เกียรติสุไพบูลย์, 102 หน้า.

งานวิจัยนี้มีวัตถุประสงค์เพื่อประมาณค่าพารามิเตอร์ของสัมประสิทธิ์ความถดถอยในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาที่มีปัจจัยเดียว สำหรับในกรณีทราบค่าปัจจัยของคอพพูลา Z พบว่า ตัวแบบคอพพูลาโลจิสติคนี้คือ ตัวแบบโพรบิท ที่ต้องมีการปรับค่าตัวประมาณด้วย $\sqrt{1-\rho}$ เมื่อ ρ คือ ค่าสหสัมพันธ์ในตัวแบบเกาซ์เซียนคอพพูลา ซึ่งข้อมูลได้ถูกจำลองเพื่อทดสอบความถูกต้องในการประมาณค่าสัมประสิทธิ์ความถดถอยในกรณีที่ทราบค่าปัจจัยของคอพพูลา Z และทำการทดสอบเพิ่มเติมในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z โดยการจำลองอยู่ภายใต้เงื่อนไขต่อไปนี้ ตัวแปรตามมีความสัมพันธ์กันด้วยเกาซ์เซียนคอพพูลาที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ ตัวแปรอิสระจำนวน 1 ตัวแปร มีการแจกแจงแบบเบอร์นูลลี จำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม จำนวนขนาดตัวอย่างในแต่ละกลุ่มเท่ากับ 100, 500 และ 1000 และเกณฑ์ที่ใช้ในการเปรียบเทียบคือ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ในการทดลองซ้ำจำนวน 100 รอบ ผลการวิจัยสรุปได้ดังนี้

สำหรับกรณีทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณด้วย $\sqrt{1-\rho}$ เป็นกรณีเดียวที่ให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยถูกต้องและใกล้เคียงค่าพารามิเตอร์มากที่สุดในทุกระดับความสัมพันธ์ โดยที่จำนวนกลุ่มตัวอย่างต้องมากกว่า 1 กลุ่ม ไม่เช่นนั้นทำให้ผลการประมาณมีค่าสูงกว่าค่าพารามิเตอร์ และสำหรับในกรณีศึกษาอื่นๆ พบว่าผลการประมาณค่าสัมประสิทธิ์การถดถอยมีทั้งค่าที่สูงกว่าค่าพารามิเตอร์หรือต่ำกว่าพารามิเตอร์ นอกจากนี้เมื่อพิจารณาปัจจัยที่มีผลต่อค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าสัมประสิทธิ์การถดถอยพบว่า เมื่อระดับความสัมพันธ์เพิ่มขึ้นส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้น ในขณะที่จำนวนกลุ่มตัวอย่างที่เพิ่มขึ้น ส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลง

ภาควิชา............สถิติ........... ลายมือชื่อนิสิต.......สุกัญญา บุญมา.......
สาขาวิชา...........สถิติ............ ลายมือชื่อ อ. ที่ปรึกษาวิทยานิพนธ์หลัก.......เสกสรร เกียรติสุไพบูลย์.......
ปีการศึกษา.........2551..........

4982242026 : MAJOR STATISTICS

KEY WORD: GAUSSIAN COPULA / LOGIT MODEL / PROBIT MODEL / LOGISTIC MODEL WITH GAUSSIAN COPULA

SUKANYA BUNMA: LOGISTIC REGRESSION MODEL WITH GAUSSIAN COPULA. THESIS PRINCIPAL ADVISOR: ASST. PROF. SEKSAN KIATSUPAIBUL, Ph.D. 102 pp.

The objective of this research is to estimate the parameters of a logistic regression model with one-factor Gaussian copula. When the copula factor Z is known, this copula logistic regression model becomes a probit model. Our study finds that this copula logistic regression is a regular probit model whose parameters are adjusted by the factor of $\sqrt{1-\rho}$ where ρ is the correlation parameter of the Gaussian copula. A simulated data is generated to test the corrected estimation technique when the copula factor value is known, and to test the traditional estimation technique when the copula factor is unknown. The experiment is done under the following conditions. The dependent variables are correlated with Gaussian copula at $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ and $\rho = 0.8$. The independent variable is generated from Bernoulli distribution. The number of the sample groups is varied from to 1, 5 to 10. The number of data points in each sample group varies from 100, 500 to 1000. The performance measurement is the mean square error (MSE) in 100 repetitions. The result of this research are as follows:

We find that the estimation method is correct only when the copula factor value is known and the parameter estimates are adjusted by $\sqrt{1-\rho}$ and the number of sample group is greater than 1. For other cases the traditional regression estimation technique yields either overestimated or underestimated parameter values. And other factors that affect the performance of the estimation include the level of the correlation parameter, the sample group size and the number of the sample groups. We find that the higher the correlation, the higher the MSE. We also find that the higher the sample group size, the lower the MSE.

Department ....Statistics......... Student's signature ....สุกัญญา บุญมา........................

Field of study ....Statistics......... Principal Advisor's signature ..เสกสรร เกียรติสุไพบูลย์

Academic year ....2008..............

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความกรุณา และความเอาใจใส่อย่างดียิ่งของผู้ช่วยศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้ให้คำปรึกษา คำแนะนำ และข้อคิดเห็นต่างๆ ตลอดจนช่วยเหลือแก้ไขข้อบกพร่อง จนกระทั่งวิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ ผู้วิจัยใคร่ขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. ธีระพร วีระถาวร รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา รองศาสตราจารย์ ดร. สุพล ดุรงค์วัฒนา และ ดร. โอวาท สุนันท์ ในฐานะประธานกรรมการและกรรมการสอบวิทยานิพนธ์ ที่กรุณาตรวจสอบและให้คำแนะนำอันเป็นประโยชน์ในการแก้ไขวิทยานิพนธ์ให้สมบูรณ์ยิ่งขึ้น และขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา และขอขอบคุณจุฬาลงกรณ์มหาวิทยาลัยที่ให้ผู้วิจัยมีโอกาสเข้ามาศึกษา ณ ที่อันทรงเกียรติแห่งนี้

ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา มารดา และครอบครัว ที่ช่วยส่งเสริมและสนับสนุนในทุกๆ ด้านให้ผู้วิจัยได้มีโอกาสทางการศึกษาเสมอมาจนสำเร็จการศึกษา สุดท้ายนี้ขอขอบคุณเพื่อนๆ ที่คอยให้คำแนะนำ และให้กำลังใจในการทำวิทยานิพนธ์เป็นอย่างดีตลอดมา

สารบัญ

หน้า

สารบัญตาราง

ตาราง หน้า

สารบัญภาพ

ภาพประกอบ หน้า

ภาพประกอบ หน้า

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในยุคปัจจุบันถือได้ว่าเป็นยุคของข้อมูล สารสนเทศ ซึ่งมีความเกี่ยวข้องอย่างมากกับศาสตร์ต่างๆ ทุกแขนง ทั้งทางด้านเศรษฐศาสตร์ สังคมศาสตร์ การแพทย์ วิทยาศาสตร์ ธุรกิจ การเงิน เป็นต้น ล้วนแต่ได้มีการนำวิธีวิเคราะห์ข้อมูลทางสถิติต่างๆ มาใช้เพื่อใช้ในการวางแผนและตัดสินใจตามวัตถุประสงค์การใช้งาน โดยวิธีการหนึ่งที่เป็นที่นิยมในปัจจุบัน คือ การวิเคราะห์ข้อมูลด้วยความถดถอยโลจิสติค (Logistic regression analysis) มีวัตถุประสงค์เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ และนำสมการความถดถอยที่ได้ไปพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ โดยที่ตัวแปรตามเป็นตัวแปรจำแนกพวกหรือเป็นข้อมูลเชิงคุณภาพ เช่น ในด้านสังคมศาสตร์มีการสำรวจอัตราการว่างงานในวัยแรงงาน อาจแบ่งตัวแปรตามเป็นวัยแรงงานที่ว่างงานและมีงานทำ ด้านการแพทย์ต้องการศึกษาปัจจัยที่มีผลต่อการเกิดโรคเบาหวาน ซึ่งตัวแปรตามเป็นคนไข้ที่เป็นโรคเบาหวานและไม่เป็นโรคเบาหวาน หรือในทางธุรกิจการเช่าซื้อต้องการหาปัจจัยเสี่ยงต่อการเกิดหนี้เสียของลูกหนี้ จึงแบ่งลูกหนี้เป็นลูกหนี้ที่มีโอกาสเกิดหนี้เสียและไม่มีโอกาสเกิดหนี้เสีย เป็นต้น ส่วนตัวแปรอิสระอาจเป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ อย่างใดอย่างหนึ่งหรือทั้งสองอย่าง

โดยทั่วไป การวิเคราะห์ความถดถอยโลจิสติค มักมีการสมมติให้ข้อมูลตัวแปรตามเป็นอิสระกัน นั่นคือ ตัวแปรตามในแต่ละค่าสังเกต (Observation) จะมีการแจกแจงแบบเบอร์นูลลีที่เป็นอิสระกัน โดยค่าของตัวแปรตามแต่ละค่าสังเกตไม่ได้ให้ข้อมูลเพิ่มเติมกับตัวแปรตามในค่าสังเกตอื่น ดังนั้นฟังก์ชันความน่าจะเป็นร่วมจะเท่ากับผลคูณของฟังก์ชันความน่าจะเป็นของตัวแปรแต่ละค่าสังเกตนั่นเอง

จากงานวิจัยของสันติและเสกสรร (2007) ได้ศึกษาการพยากรณ์มูลค่าความเสี่ยง (Credit Value at Risk: Credit VaR) ซึ่งเป็นการวัดความเสี่ยงที่มีความสำคัญในการจัดการด้านความเสี่ยงของสถาบันทางการเงิน ด้วยความถดถอยโลจิสติคพบว่า การประมาณค่ามูลค่าความเสี่ยงด้วยความถดถอยโลจิสติคมีค่าประมาณต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Underestimate) ซึ่งอาจเป็นไปได้ว่ามีการระบุตัวแปรอิสระไม่ครบถ้วนหรือไม่เพียงพอ หรืออาจเกิดเงื่อนไขความผิดปกติบางอย่างที่มีผลต่อความคลาดเคลื่อนของข้อมูลและอาจส่งผลต่อความเป็นอิสระของตัวแปรตาม ซึ่งทำให้ตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กันได้ หากเกิดกรณีเช่นนี้ขึ้นแล้วจะมีวิธีการใดที่สามารถนำมาแก้ไขปัญหาตรงจุดนี้ได้ ฟังก์ชันความน่าจะเป็นร่วมของตัวแปรตามที่มี

ความสัมพันธ์กันนั้นจะเป็นอย่างไร และการวิเคราะห์และการสรุปผลจะเปลี่ยนแปลงไปจากกรณีตัวแปรตามเป็นอิสระกันหรือไม่

จากปัญหาเหล่านี้ ทำให้เกิดแนวคิดในการนำความสัมพันธ์เข้ามาประกอบในตัวแปรตาม หรือสร้างความสัมพันธ์ระหว่างตัวแปรตามในแต่ละค่าสังเกตขึ้น โดยความสัมพันธ์ที่จะทำการศึกษาในครั้งนี้ ได้แก่ เกาซ์เซียนคอพพูลา (Gaussian copula) โดยความสัมพันธ์แบบเกาซ์เซียนคอพพูลานี้ เป็นความสัมพันธ์ผ่านทางการแจกแจงร่วมของควอนไทล์ (quantile) ของตัวแปรร่วมที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal variables) และเรียกตัวแบบความถดถอยโลจิสติคที่มีความสัมพันธ์ด้วยเกาซ์เซียนคอพพูลานี้ว่า ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา ที่สามารถนำไปใช้ในการวิเคราะห์และสรุปผลข้อมูลได้อย่างถูกต้องและเหมาะสมต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาและประมาณค่าพารามิเตอร์ของสัมประสิทธิ์ความถดถอยในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา จากข้อมูลการจำลอง
2. เพื่อเปรียบเทียบความถูกต้องในการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์การถดถอยจากตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา จากข้อมูลการจำลอง

ขอบเขตของการวิจัย

ขอบเขตการวิจัยในวิทยานิพนธ์ฉบับนี้จะครอบคลุมถึงการศึกษาในเรื่องต่อไปนี้

1. คุณสมบัติของตัวแบบโลจิสติคอย่างง่ายแบบ 2 กลุ่ม (Simple binary logistic model) โดยกำหนด
   - ตัวแปรตาม (Y) เป็นข้อมูลเชิงคุณภาพ โดยมีการแจกแจงแบบคอพพูลาเบอร์นูลลี (Copula Bernoulli distribution) นั่นคือ ค่าสังเกตในแต่ละค่าของตัวแปรตามมีความสัมพันธ์กันด้วยเกาซ์เซียนคอพพูลา และกำหนดค่าตัวแปรตามมีค่าเพียง 2 ค่า คือ 0 และ 1
   - ตัวแปรอิสระ (X) เป็นข้อมูลเชิงคุณภาพ โดยมีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution) ด้วยพารามิเตอร์ p นั่นคือ $X \sim Ber(p)$ โดยมีฟังก์ชันความน่าจะเป็นอยู่ในรูปของ

$$p_X(k) = p^k(1-p)^{1-k}, k = 0,1$$

ในงานวิจัยครั้งนี้จะศึกษาที่ $p = 0.5$

2. เกาซ์เซียนคอพพูลา (Gaussian copula) ที่ระดับความสัมพันธ์ $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$
3. จำนวนกลุ่มตัวอย่างที่ใช้ในแต่ละรอบของการทดลอง (m) เป็น 1 , 5 และ 10 กลุ่ม
4. จำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เป็น 100, 500 และ 1,000
5. กำหนดจำนวนการกระทำซ้ำในแต่ละสถานการณ์เป็น 100 รอบ
6. ทำการจำลองข้อมูล (Simulation) ตามสถานการณ์ที่กำหนด โดยใช้โปรแกรม R

สมมติฐานการวิจัย

1. ความสัมพันธ์ในระดับที่สูง สามารถทำให้ผลการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์ความถดถอยและผลการวิเคราะห์ข้อมูลจากตัวแบบนี้มีความถูกต้องมากขึ้น
2. จำนวนกลุ่มตัวอย่างในการทดลองและจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลองที่มากขึ้น ทำให้ผลการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์ความถดถอยและผลการวิเคราะห์ข้อมูลจากตัวแบบนี้มีความเหมาะสมมากขึ้น

วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเชิงทฤษฎี (Theoretical Research) มุ่งศึกษาถึงการวิเคราะห์ข้อมูลด้วยความถดถอยโลจิสติคกรณีข้อมูลตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กันแบบเกาซ์เซียนคอพพูลา ซึ่งได้ดำเนินการวิจัย ดังนี้

1. ศึกษาความสัมพันธ์เกาซ์เซียนคอพพูลา (Gaussian copula) ด้วยเมทริกซ์สหสัมพันธ์ Σ
2. ศึกษาฟังก์ชันความน่าจะเป็นของตัวแปรตามในตัวแบบโลจิสติค (Logistic model)
3. ทำการหารูปแบบฟังก์ชันภาวะน่าจะเป็น (Likelihood function) ของตัวแปรตามในตัวแบบความถดถอยโลจิสติค โดยนำความสัมพันธ์แบบเกาซ์เซียนคอพพูลา (Gaussian copula) เข้ามาประกอบในการแจกแจงความน่าจะเป็นของตัวแบบความถดถอยโลจิสติค ทำให้ได้รูปแบบฟังก์ชันภาวะน่าจะเป็นของตัวแบบความถดถอยโลจิสติคที่มีความสัมพันธ์กันด้วยเกาซ์เซียนคอพพูลา
4. ทำการจำลองข้อมูลตัวแปรตามให้มีการแจกแจงแบบคอพพูลาเบอร์นูลลี ที่มีความสัมพันธ์กันที่ระดับความสัมพันธ์ $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$ และตัวแปรอิสระจำนวน 1 ตัวแปร ให้มีการแจกแจงแบบเบอร์นูลลีด้วยพารามิเตอร์ $p=0.5$ นั่นคือ $X \sim Ber(0.5)$ โดยกำหนดจำนวนขนาดตัวอย่างในแต่ละกลุ่ม (n) เป็น 100, 500 และ 1000 โดยจำนวนกลุ่มตัวอย่างในแต่ละรอบของการทดลอง (m) เป็น 1 , 5 และ 10 กลุ่ม เพื่อนำมาใช้ในการวิเคราะห์ข้อมูลจากตัวแบบโลจิสติคแบบเกาซ์เซียนคอพพูลา

5. ทำการประมาณค่าสัมประสิทธิ์การถดถอยของพารามิเตอร์ β_0, β_1
6. ทำการวิเคราะห์ข้อมูลและเปรียบเทียบความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอยด้วยการวิเคราะห์ความถดถอยโลจิสติคในกรณีที่ข้อมูลของตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กันในระดับความสัมพันธ์ต่างๆ จากข้อมูลที่ได้การจำลอง
7. สรุปผลการวิจัย

เกณฑ์การตัดสินใจ

สำหรับเกณฑ์ที่ใช้ในการเปรียบเทียบความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอยของตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาว่า ที่ระดับความสัมพันธ์และจำนวนกลุ่มข้อมูลใดที่ทำให้ได้ค่าประมาณของสัมประสิทธิ์การถดถอยมีค่าใกล้เคียงกับค่าพารามิเตอร์มากที่สุดนั้น จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) และพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบประกอบการตัดสินใจ

จากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์การถดถอยของพารามิเตอร์ β_0, β_1 ดังต่อไปนี้

$$MSE_i = \frac{\sum_{k=0}^{K}(\hat{\beta}_k - \beta_k)^2}{K+1}$$

เมื่อ MSE_i แทน ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการทดลองครั้งที่ i

β_k แทน ตัวประมาณสัมประสิทธิ์การถดถอยตัวที่ k ที่ประมาณได้ในครั้งที่ i

K แทน จำนวนตัวแปรอิสระ โดยที่ $k = 0,1,2,\ldots,K$

ดังนั้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์การถดถอยจากการทดลองซ้ำ เป็นดังนี้

$$MSE = \frac{\sum_{i=1}^{L} MSE_i}{L}$$

โดยที่ L แทน จำนวนรอบในการทดลองซ้ำ

และ MSE แทน ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณสัมประสิทธิ์การถดถอยจากการทดลองซ้ำ

ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ทราบรูปแบบตัวแบบความถดถอยโลจิสติคในกรณีที่ข้อมูลมีความสัมพันธ์กันแบบเกาซ์เซียนคอพพูลา เพิ่มเติมจากกรณีที่ข้อมูลเป็นอิสระกันซึ่งใช้กันอยู่ในปัจจุบัน
2. สามารถนำมาเป็นทางเลือกหนึ่งในการแก้ปัญหาในกรณีที่มีปัจจัยหรือตัวแปรต้นไม่ครบถ้วนหรือไม่เพียงพอในการนำมาวิเคราะห์ข้อมูล ซึ่งอาจส่งผลต่อการเกิดความสัมพันธ์กันของตัวแปรตามได้
3. ทำให้ผลการวิเคราะห์ข้อมูลด้วยความถดถอยโลจิสติคนี้มีความถูกต้อง และมีประสิทธิภาพมากขึ้น
4. ทำให้เกิดความเข้าใจเกี่ยวกับลักษณะของเกาซ์เซียนคอพพูลา เพื่อสามารถนำไปประยุกต์ใช้ในงานด้านอื่นๆ
5. เพื่อเป็นเอกสารค้นคว้าและข้อมูลอ้างอิงสำหรับผู้ที่สนใจศึกษาทฤษฎี หลักการ แนวทางในการนำงานวิจัยนี้ไปใช้หรือนำไปศึกษาในประเด็นอื่น ๆ ที่เกี่ยวข้อง

บทที่ 2
ทฤษฎีและสถิติที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงแบบจำลองที่ตัวแปรตามมีค่าเพียง 2 ค่า (Binary Dependent Variable Model) ซึ่งได้แก่ แบบจำลองโลจิท (Logit model) และแบบจำลองโพรบิท (Probit model) ตัวแบบเกาซ์เซียนคอพพูลา (Gaussian copula model) ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา (Logistic regression model with Gaussian copula) และตัวอย่างการประยุกต์ใช้งานตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา

2.1 แบบจำลองที่ตัวแปรตามมีค่าเพียง 2 ค่า (Binary Dependent Variable Model)

แบบจำลองที่นิยมนำมาใช้ในการนำมาใช้วิเคราะห์ข้อมูลที่มีลักษณะของตัวแปรตามเป็นเชิงคุณภาพที่มีค่าเป็นไปได้ 2 ค่า (Binary response) คือ แบบจำลองโลจิท (Logit Model) และแบบจำลองโพรบิท (Probit Model) แบบจำลองทั้งสองนี้ สมมติว่าควอนไทล์ของความน่าจะเป็นที่จะสังเกตค่าของการเกิดเหตุการณ์ในตัวอย่างสัมพันธ์เป็นเชิงเส้นกับตัวแปรอิสระ โดยแบบจำลองโพรบิท ใช้ควอนไทล์ของการแจกแจงแบบปกติ แต่แบบจำลองโลจิท ใช้ควอนไทล์ของการแจกแจงแบบโลจิสติค นอกจากนี้การแจกแจงทั้งสองแตกต่างกันที่ช่วงปลายของการแจกแจง โดยส่วนปลายของการแจกแจงแบบโลจิสติคจะหนากว่าการแจกแจงแบบโพรบิท อย่างไรก็ตามผลการประมาณค่าพารามิเตอร์ในแบบจำลองโลจิทและแบบจำลองโพรบิทยังให้ผลที่ไม่แตกต่างกันมากนัก (ธาตรี จันทรโคลิกา, 2542: 16-24) สำหรับในงานวิจัยนี้ขอกล่าวอ้างถึงแบบจำลองโลจิทและแบบจำลองโพรบิท (Daniel A. Powers และ Yu Xie, 2000) ดังรายละเอียดต่อไปนี้

2.1.1 แบบจำลองโลจิท (Logit model)

แบบจำลองโลจิท มีรูปแบบเริ่มต้นดังนี้

กำหนดให้ Y_i^* เป็นตัวแปรลาเทนท์ (Latent variable) ของหน่วยสังเกตที่ i ; $i = 1,2,\dots,n$ ที่มีรูปแบบเป็นฟังก์ชันเชิงเส้น (Linear function) ของตัวแปรอิสระ $K+1$ ตัว $(1, X_{i1},\dots X_{iK})$ ที่มีสัมประสิทธิ์ $(\beta_0, \beta_1,\dots, \beta_K)$ และค่าความคลาดเคลื่อน (e_i) โดยที่ e_i มีการแจกแจงแบบโลจิสติค โดยรูปแบบฟังก์ชันเชิงเส้น เป็นดังนี้

$$Y_i^* = \sum_{k=0}^{K} \beta_k X_{ik} + e_i \tag{1}$$

สามารถเขียนให้อยู่ในรูปเมทริกซ์ได้ดังนี้

$$\mathbf{Y}^{*} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{2}$$

เมื่อ $\mathbf{X}_{(n\times K)}$ เป็นเมทริกซ์ของตัวแปรอิสระ, $\boldsymbol{\beta}_{(K\times 1)}$ เป็นเวกเตอร์ของพารามิเตอร์ของสัมประสิทธิ์การถดถอย และ $\mathbf{e}_{(n\times 1)}$ เป็นเวกเตอร์สุ่มซึ่งแทนความคลาดเคลื่อน

ในทางปฏิบัติตัวแปรลาเทนท์ Y_i^{*} ไม่สามารถเก็บข้อมูลหรือสังเกตค่าได้จริง ดังนั้นผลลัพธ์ซึ่งเป็นสิ่งที่สังเกตได้ต้องนำมาปรับให้เป็นตัวแปรหุ่น (Dummy Variable) แทนด้วยค่า Y_i ดังนี้

$$\begin{aligned} &Y_i = 1 \text{ ถ้า } Y_i^{*} > 0 \\ &Y_i = 0 \text{ ถ้า } Y_i^{*} \le 0 \end{aligned} \tag{3}$$

ปิยะลักษณ์ พุทธวงศ์ (2546) ได้ยกตัวอย่าง การตัดสินใจเข้าร่วมตลาดแรงงาน โดยสิ่งที่กำหนดการเข้าร่วมตลาดแรงงานคือค่าจ้างสำรอง (Reservation wage) หมายถึง ระดับค่าจ้างที่แรงงานจะเริ่มตัดสินใจเข้าร่วมตลาดแรงงาน หากนายจ้างเสนอค่าจ้างที่ต่ำกว่าค่าจ้างสำรอง แรงงานจะเลือกไม่เข้าร่วมตลาดแรงงาน แต่ถ้าหากนายจ้างเสนอค่าจ้างที่สูงกว่าหรือเท่ากับค่าจ้างสำรอง แรงงานจะเลือกเข้าร่วมตลาดแรงงาน ซึ่งค่าจ้างสำรองเป็นสิ่งที่แตกต่างไปตามลักษณะของแรงงานที่ผู้วิจัยไม่สามารถสังเกตได้ แต่สิ่งที่ผู้วิจัยสังเกตได้คือการตัดสินใจเข้าร่วมหรือไม่เข้าร่วมตลาดแรงงาน ดังนั้น ตัวแปรลาเทนท์ในตัวอย่างนี้ คือ ค่าจ้างสำรอง (Y^{*}) และตัวแปรที่สังเกตได้ คือการเข้าร่วมหรือไม่เข้าร่วมตลาดแรงงาน (Y) เป็นต้น

จากสมการ (2) และ (3) ข้างต้น จะได้ว่า

$$\begin{aligned} P[Y_i = 1] &= P[Y_i^{*} > 0] \\ &= P\left(\sum_{k=0}^{K} \beta_k X_{ik} + e_i > 0\right) \\ &= P\left(e_i > -\sum_{k=0}^{K} \beta_k X_{ik}\right) \\ &= 1 - F\left(-\sum_{k=0}^{K} \beta_k X_{ik}\right) \\ &= F\left(\sum_{k=0}^{K} \beta_k X_{ik}\right) \end{aligned} \tag{4}$$

โดยที่ F คือ ฟังก์ชันการแจกแจงสะสม (Cumulative distribution function) ของค่าความคลาดเคลื่อน (e_i) ที่มีรูปแบบการแจกแจงแบบโลจิสติค

<u>การประมาณค่าแบบจำลองโลจิท</u>

ในกรณีค่า Y_i ที่เก็บข้อมูลมาได้จะมีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution) ซึ่งค่าความน่าจะเป็นถูกกำหนดโดยสมการ (4) สำหรับวิธีที่ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยแบบจำลองโลจิทที่เหมาะสมและเป็นที่นิยม คือ วิธีภาวะน่าจะเป็นสูงสุด (Maximum likelihood) ซึ่งแนวคิดในการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีนี้ คือ ทำการประมาณโดยการหาค่าตัวประมาณของสัมประสิทธิ์การถดถอยที่ทำให้ความน่าจะเป็นร่วมของข้อมูลมีค่าสูงที่สุด และรูปแบบของสมการฟังก์ชันภาวะน่าจะเป็น (Likelihood function) มีรูปแบบดังนี้

$$L = \prod_i F\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)^{Y_i} \left[1 - F\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)\right]^{1-Y_i} \tag{5}$$

จากสมการ (4) เมื่อฟังก์ชันการแจกแจงสะสมของค่าความคลาดเคลื่อน (e_i) มีการแจกแจงแบบโลจิสติค จะได้ว่า

$$P\left[Y_i^* > 0\right] = P\left[Y_i = 1\right] = F\left(\sum_{k=0}^{K} \beta_k X_{ik}\right) = p_i = \frac{\exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)}{1 + \exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)} \tag{6}$$

โดยเรียกสมการ (6) ว่า Logistic Response Function

และพบว่า

$$P\left[Y_i^* \le 0\right] = P\left[Y_i = 0\right] = 1 - F\left(\sum_{k=0}^{K} \beta_k X_{ik}\right) = 1 - p_i = \frac{1}{1 + \exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)} \tag{7}$$

นั่นคือ จากสมการที่ (5) จะได้ว่า

$$L = \prod_i \left\{\frac{\exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)}{1 + \exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)}\right\}^{Y_i} \left\{\frac{1}{1 + \exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)}\right\}^{1-Y_i} \tag{8}$$

จากสมการ (8) ข้างต้น จะเห็นว่าความสัมพันธ์ในรูปแบบของสมการฟังก์ชันภาวะน่าจะเป็นไม่ได้อยู่ในรูปเชิงเส้น จึงได้มีการปรับให้อยู่ในรูปเชิงเส้น ดังนี้

กำหนด Odd Ratio (OR) เป็นอัตราส่วนระหว่างโอกาสที่จะเกิดเหตุการณ์และโอกาสที่จะไม่เกิดเหตุการณ์ ดังนี้

$$\text{Odd Ratio} = OR = \frac{P(Y=1)}{P(Y=0)} = \frac{p_i}{1-p_i}$$

$$= \left\{ \frac{\exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)}{1+\exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)} \right\} \Bigg/ \left\{ \frac{1}{1+\exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)} \right\}$$

$$\therefore OR = \exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right) \tag{9}$$

ถอด Natural logarithms ของสมการ (9) จะได้

$$\ln(OR) = \ln\left(\frac{p_i}{1-p_i}\right) = \sum_{k=0}^{K} \beta_k X_{ik} \tag{10}$$

นั่นคือ จากสมการ (10) จะได้การแปลงโลจิท (Logit transformation) หรือเรียกอีกอย่างหนึ่งว่า ฟังก์ชันตอบสนองโลจิท (Logit Response Function) คือ

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \sum_{k=0}^{K} \beta_k X_{ik} \tag{11}$$

โดยที่

$$p_i = \frac{\exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)}{1+\exp\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)} = \frac{1}{1+\exp\left(-\sum_{k=0}^{K} \beta_k X_{ik}\right)} \tag{12}$$

ดังนั้น สามารถเขียนสมการฟังก์ชันภาวะน่าจะเป็นให้อยู่ในรูปของการแจกแจงแบบโลจิสติค (Logistic) ได้ดังนี้

$$L = \prod_{i=1}^{n} p_i^{Y_i} (1-p_i)^{1-Y_i}$$

$$\ln(L) = \sum_{i=1}^{n} \{Y_i \ln(p_i) + (1-Y_i)\ln(1-p_i)\}$$

$$\ln L = \sum_{i=1}^{n}\left\{Y_i \ln\left(\frac{p_i}{1-p_i}\right) + \ln(1-p_i)\right\}$$
$$= \sum_{i=1}^{n}\left\{Y_i\left(\sum_{k=0}^{K}\beta_k X_{ik}\right) - \ln\left(1+\exp\left(\sum_{k=0}^{K}\beta_k X_{ik}\right)\right)\right\} \tag{13}$$

นั่นคือ จะสามารถหาค่าพารามิเตอร์ของสัมประสิทธิ์การถดถอย $(\beta_0, \beta_1, ..., \beta_K)$ โดยการหาค่า First order condition ของสมการฟังก์ชันภาวะน่าจะเป็น โดยกำหนด

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_k} = 0, \qquad k = 0,1,...,K \tag{14}$$

จากการแก้สมการหาค่า First order condition ของแบบจำลองโลจิท พบว่า

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{U}(\boldsymbol{\beta}) = \sum_i (Y_i - \Lambda_i)\mathbf{X}_i = 0 \tag{15}$$

เมื่อ $\Lambda_i = \Lambda\left(\sum_{k=0}^{K}\beta_k X_{ik}\right) = \exp\left(\sum_{k=0}^{K}\beta_k X_{ik}\right) \Big/ \left[1+\exp\left(\sum_{k=0}^{K}\beta_k X_{ik}\right)\right]$

2.1.2 แบบจำลองโพรบิท (Probit model)

แบบจำลองโพรบิท มีรูปแบบจำลองเริ่มต้นเช่นเดียวกันกับแบบจำลองโลจิท ดังแสดงในสมการ (1) คือ

$$Y_i^* = \sum_{k=0}^{K}\beta_k X_{ik} + e_i \tag{16}$$

โดย Y_i^* เป็นตัวแปรลาเทนท์ (Latent variable) ของหน่วยสังเกตที่ i; $i = 1,2,...,n$ ที่มีรูปแบบเป็นฟังก์ชันเชิงเส้น (Linear function) ของตัวแปรอิสระ $(1, X_{i1},... X_{iK})$ ที่มีพารามิเตอร์ของสัมประสิทธิ์การถดถอย $(\beta_0, \beta_1,..., \beta_K)$ และค่าความคลาดเคลื่อน (e_i) โดยที่ e_i มีรูปแบบการแจกแจงเป็นแบบปกติ

ในทางปฏิบัติตัวแปรลาเทนท์ Y_i^* ไม่สามารถเก็บข้อมูลหรือสังเกตค่าได้จริง ดังนั้นผลลัพธ์ซึ่งเป็นสิ่งที่สังเกตได้ต้องนำมาปรับให้เป็นตัวแปรหุ่น (Dummy Variable) แทนด้วยค่า Y_i ดังนี้

$$Y_i = 1 \text{ ถ้า } Y_i^* > 0$$
$$Y_i = 0 \text{ ถ้า } Y_i^* \le 0 \tag{17}$$

และจากสมการ (16) และ (17) ข้างต้น จะได้ว่า

$$
\begin{aligned}
P[Y_i = 1] &= P[Y_i^* > 0] \\
&= P\left(\sum_{k=0}^{K} \beta_k X_{ik} + e_i > 0\right) \\
&= P\left(e_i > -\sum_{k=0}^{K} \beta_k X_{ik}\right) \\
&= 1 - \Phi\left(-\sum_{k=0}^{K} \beta_k X_{ik}\right) \\
&= \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)
\end{aligned} \tag{18}
$$

โดยที่ Φ คือ ฟังก์ชันการแจกแจงสะสม (Cumulative distribution function) ของค่าความคลาดเคลื่อน (e_i) ที่มีรูปแบบการแจกแจงแบบปกติ

<u>การประมาณค่าแบบจำลองโพรบิท</u>

จากค่า Y_i ที่เก็บข้อมูลมาได้จะมีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution) ซึ่งค่าความน่าจะเป็นถูกกำหนดโดยสมการ (18) และจะแปรผันตามค่าตัวแปรอิสระ(X_i) ที่มีรูปแบบของสมการฟังก์ชันภาวะน่าจะเป็น (Green, 2000: 811-895) ดังนี้

$$
L = \prod_i \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)^{Y_i} \left[1 - \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)\right]^{1-Y_i} \tag{19}
$$

เมื่อ Φ เป็น การแจกแจงสะสมของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ

จากสมการ (18) เมื่อฟังก์ชันการแจกแจงสะสมของค่าความคลาดเคลื่อน (e_i) มีการแจกแจงแบบปกติ ดังนั้น รูปแบบของฟังก์ชันการแจกแจง เป็นดังนี้

$$
P[Y_i^* > 0] = P[Y_i = 1] = \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right) = p_i = \int_{-\infty}^{\eta_i} \phi(t)dt \tag{20}
$$

เมื่อ ϕ เป็นฟังก์ชันความหนาแน่นของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ โดยที่

$$\phi(t) = \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{t^2}{2}\right) \sim N(0,1)$$

และ จะได้ว่า

$$\eta_i = \Phi^{-1}[p_i] = \sum_{k=0}^{K} \beta_k X_{ik} \tag{22}$$

นั่นคือ การแปลงโพรบิท (Probit transformation) ซึ่งเป็นตัวผกผันของฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน (The inverse of the standard cumulative normal distribution function) สามารถเขียนได้เป็น

$$probit(p_i) = \Phi^{-1}(p_i) = \eta_i = \sum_{k=0}^{K} \beta_k X_{ik} \tag{23}$$

โดยที่

$$p_i = \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right) = \int_{-\infty}^{\eta_i} \left[\frac{1}{(2\pi)^{1/2}}\right] \exp\left[-\frac{t^2}{2}\right] dt$$

จากสมการ (20) จะพบว่าค่าของฟังก์ชันจะอยู่ระหว่าง 0 และ 1 โดยฟังก์ชันดังกล่าวเป็นฟังก์ชันการแจกแจงสะสม (Cumulative distribution function) ของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน (Standard normal distribution)

โดยปกติแล้วการประมาณแบบจำลองเชิงเส้น ผู้วิจัยจะใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares: OLS) วิธีการดังกล่าวไม่สามารถใช้ได้กับการประมาณแบบจำลองที่มีลักษณะไม่เป็นเส้นตรง (Nonlinear Model) วิธีการที่จะใช้ประมาณแบบจำลองโพรบิทจะใช้วิธีการเดียวกันกับแบบจำลองโลจิท คือใช้การประมาณด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood Estimation) โดยเริ่มจากรูปแบบฟังก์ชันภาวะน่าจะเป็น (Likelihood function) ดังนี้

$$L = \prod_i \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)^{Y_i} \left[1 - \Phi(\sum_{k=0}^{K} \beta_k X_{ik})\right]^{1-Y_i}$$

$$\ln L = \sum_i \left\{ Y_i \ln \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right) + (1 - Y_i) \ln\left[1 - \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)\right]\right\} \tag{24}$$

ซึ่งสามารถหาค่าพารามิเตอร์ของสัมประสิทธิ์การถดถอย $(\beta_0, \beta_1, ..., \beta_K)$ โดยการหาค่า First order condition ของฟังก์ชันภาวะน่าจะเป็น โดยกำหนด

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \beta_k} = 0, \qquad k = 0,1,...,K \tag{25}$$

และสามารถการแก้สมการหาค่า First order condition ของแบบจำลองโพรบิท ซึ่งพบว่า

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{U}(\boldsymbol{\beta}) = \sum_{y_i=0} -[\phi_i/(1-\Phi_i)]\mathbf{X}_i + \sum_{y_i=1}(\phi_i/\Phi_i)\mathbf{X}_i = 0 \tag{26}$$

เมื่อ ϕ_i แทน ฟังก์ชันความหนาแน่นของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน

Φ_i แทน ฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน

<u>การทดสอบสมมติฐานและพิจารณาความเหมาะสมของแบบจำลองโลจิทและแบบจำลองโพรบิท</u>

กัลยา วานิชย์บัญชา (2548) ได้กล่าวถึงการทดสอบสมมติฐานโดยแบ่งออกเป็นสองส่วน คือ การทดสอบความมีนัยสำคัญของสัมประสิทธิ์ความถดถอยโลจิสติคของตัวแปรอิสระแต่ละตัว ดังนี้

$$\begin{aligned} H_0 &: \beta_k = 0 \\ H_1 &: \beta_k \neq 0 \end{aligned} \qquad ; k = 0,1,...,K \tag{27}$$

การทดสอบสมมติฐานข้างต้นนี้ อาจใช้สถิติทดสอบวอลด์ (Wald test) ซึ่งมีการแจกแจงแบบไคกำลังสอง โดยที่องศาอิสระเท่ากับ 1 โดยสถิติทดสอบวอลด์ เป็นดังนี้

$$\text{สถิติทดสอบ} \quad Wald = \left[\frac{b_k}{SE(b_k)}\right]^2 \tag{28}$$

แต่ในกรณีที่ค่าสัมประสิทธิ์การถดถอยมีค่ามากจะทำให้ค่าคลาดเคลื่อนมาตรฐานมีค่ามากด้วย ซึ่งจะทำให้ค่าสถิติทดสอบวอลด์มีค่าน้อยและทำให้เกิดความผิดพลาดในการทดสอบประเภทที่ 2 ขึ้น ดังนั้นในกรณีที่ค่าสัมประสิทธิ์มีค่ามากหรือตัวแปรอิสระเป็นตัวแปรเชิงคุณภาพ ซึ่งต้องแปลงให้อยู่ในรูปตัวแปรหุ่น (Dummy variable) ไม่ควรใช้ค่าสถิติทดสอบวอลด์ แต่ควรใช้สถิติทดสอบอัตราส่วนความควรจะเป็น ที่มีองศาอิสระเท่ากับ 1 โดยจะใช้ Block Chi-Square ที่ block 2 ไม่รวมตัวแปรอิสระ X_k ในขณะที่ block 1 รวมตัวแปรอิสระ X_k ไว้ในสมการ ดังนี้

$$\begin{aligned} \text{Block Chi-Square} &= [-2LL(block2)] - [-2LL(block1)] \\ &= [-2LL(X_1, X_2, ..., X_{k-1}, X_{k+1}, .., X_K)] - [-2LL(X_1, X_2, ..., X_K)] \end{aligned} \tag{29}$$

<u>ตัวอย่าง</u> ถ้ามีตัวแปร 4 ตัว (X_1, X_2, X_3, X_4) และต้องการทดสอบ

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

สถิติทดสอบ คือ

$$\text{Block Chi-Square} = [-2LL(X_1, X_3, X_4)] - [-2LL(X_1, X_2, X_3, X_4)]$$

โดย Block Chi-Square จะมีการแจกแจงแบบไคกำลังสองที่องศาอิสระเท่ากับ 1

จากการศึกษาเปรียบเทียบสถิติทดสอบวอลด์และอัตราส่วนความควรจะเป็นโดย Agresti (2002) ได้สรุปว่า ในกรณีที่ขนาดตัวอย่างเล็กสถิติทดสอบอัตราส่วนความควรจะเป็นจะเชื่อถือได้มากกว่าสถิติทดสอบวอลด์

สำหรับการทดสอบในกรณีที่ตัวแปรอิสระ K ตัว

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0$$
$$H_1 : \text{มี } \beta_k \neq 0 \text{ อย่างน้อย 1 ค่า, } k = 1,2,\ldots,K \tag{30}$$

การทดสอบสมมติฐานนี้จะใช้ฟังก์ชันความควรจะเป็น โดยพิจารณาจากสถิติทดสอบอัตราส่วนความควรจะเป็น ซึ่งเป็นอัตราส่วนของค่าที่ทำให้ฟังก์ชันความควรจะเป็นเมื่อมีตัวแปรอิสระ K ตัว (L_1) กับค่าที่ทำให้ฟังก์ชันความควรจะเป็นเมื่อมีเฉพาะค่าคงที่ (L_0) มีค่ามากที่สุด ดังนั้น จะได้ว่า

$$\begin{aligned} \text{สถิติทดสอบอัตราส่วนความควรจะเป็น } &= -2\log\left(\frac{L_0}{L_1}\right) \\ &= -2[\log(L_0) - \log(L_0)] \\ &= -2[LL(0) - LL(1)] \end{aligned} \tag{31}$$

ซึ่งมีการแจกแจงแบบไคกำลังสอง

ค่าสถิติทดสอบอัตราส่วนความควรจะเป็น คือค่า $-2LL$ ที่เปลี่ยนแปลงไป ซึ่งเท่ากับผลต่างของค่า $-2LL(0)$ และ $-2LL(X_1, X_2, \ldots, X_K)$ นั่นหมายถึงการเปลี่ยนแปลงของค่า $-2LL$ ที่ลดลงเมื่อมีตัวแปรอิสระในสมการ K ตัว เมื่อเทียบกับเมื่อมีเฉพาะค่าคงที่ ซึ่งถ้าผลต่างมีค่ามาก แสดงว่าเมื่อเพิ่มตัวแปรอิสระเข้าไปในสมการแล้วทำให้ $-2LL$ ลดลงอย่างมากส่งผลให้ปฏิเสธ H_0 นั่นคือค่า $-2LL$ ใช้วัดความเหมาะสมของสมการโลจิสติค ถ้าสมการโลจิสติคเหมาะสม ค่า $-2LL$ จะมีค่าต่ำ นั่นคือ $-2LL$ เป็นความคลาดเคลื่อนในการพยากรณ์ของสมการความถดถอยโลจิสติค ซึ่งแสดงถึงผลต่างของค่าความผิดพลาดในการพยากรณ์ของสมการ

ความถดถอยโลจิสติคเมื่อไม่มีตัวแปรอิสระ กับค่าความผิดพลาดในการพยากรณ์ของสมการความถดถอยโลจิสติคเมื่อมีตัวแปรอิสระ K ตัว หรืออาจเรียกค่าสถิติทดสอบอัตราส่วนความควรจะเป็นว่า Model Chi-Square นั่นคือ

$$\begin{aligned}\text{Model Chi-Square} &= [-2LL(\beta_0)] - [-2LL(\beta_0, \beta_1, \ldots, \beta_K)] \\ &= [-2LL(0)] - [-2LL(X_0, X_1, \ldots, X_K)] \qquad (32)\end{aligned}$$

ที่มีองศาอิสระของ Chi-Square = K

และในการทดสอบสมมติฐานอีกส่วนเป็นการพิจารณาตรวจสอบความเหมาะสมของสมการความถดถอยโลจิสติค โดยมีสมมติฐานการทดสอบ คือ

$$H_0 : \text{Model: } p = \frac{1}{1+\exp(-\sum_{k=0}^{K}\beta_k X_{ik})} \quad \text{เหมาะสมกับข้อมูล}$$

$$H_1 : \text{Model: } p = \frac{1}{1+\exp(-\sum_{k=0}^{K}\beta_k X_{ik})} \quad \text{ไม่เหมาะสมกับข้อมูล} \qquad (33)$$

สำหรับสมมติฐานนี้จะใช้สถิติทดสอบไคกำลังสอง ในการตรวจสอบคล้องจอง (Goodness of fit test) ของสมการความถดถอยโลจิสติค ในกรณีตัวแปรอิสระเป็นตัวแปรเชิงคุณภาพ ส่วนกรณีที่ตัวแปรอิสระเป็นตัวแปรเชิงปริมาณ เมื่อจะใช้สถิติไคกำลังสองตรวจสอบความเหมาะสมของรูปแบบความสัมพันธ์จะต้องแบ่งค่าตัวแปรอิสระออกเป็นช่วงๆ ก่อน นอกจากนี้ยังสามารถใช้สถิติทดสอบความเหมาะสมของ Hosmer and Lemeshow's goodness of fit test (H-L) ได้เช่นกัน โดย H-L มีการแจกแจงโดยประมาณแบบไคกำลังสอง

สำหรับการวัดระดับความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ จะใช้สัมประสิทธิ์การตัดสินใจ (R^2) โดยในการวิเคราะห์ความถดถอยโลจิสติคค่า R^2 ที่ได้จะไม่ใช่ค่าสัดส่วนที่แท้จริงของความผันแปรของตัวแปรตามที่อธิบายได้ด้วยตัวแปรอิสระ อย่างไรก็ตามค่า R^2 สำหรับวัดระดับความสัมพันธ์ในการวิเคราะห์ความถดถอยโลจิสติค ได้แก่ Cox & Snell R^2 (R^2_{CS})และ Nagelkerki's R^2 (R^2_N) ซึ่งเรียกค่า R^2_N และ R^2_{CS} ว่า Pseudo R^2 โดย R^2_N จะมากกว่า R^2_{CS} เสมอ แต่มักจะต่ำกว่าค่า R^2 ของเทคนิควิเคราะห์ความถดถอยเชิงเส้น

2.2 เกาซ์เซียนคอพพูลา (Gaussian copula)

เกาซ์เซียนคอพพูลา (Gaussian copula) เป็นความสัมพันธ์ที่ผ่านทางการแจกแจงร่วมของควอนไทล์ (quantile) $\left(P\left[Y_1 < y_1, Y_2 < y_2, \ldots, Y_n < y_n\right]\right)$ ของตัวแปรสุ่มร่วมที่มีการแจกแจงแบบปกติ (Multivariate normal variables) ด้วยเมทริกซ์สหสัมพันธ์เชิงเส้น Σ

โดย
$$\Sigma = \begin{bmatrix} Var(Y_1) & Cov(Y_1, Y_2) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_2, Y_1) & Var(Y_2) & \cdots & Cov(Y_2, Y_n) \\ \vdots & \ddots & \ddots & \vdots \\ Cov(Y_n, Y_1) & Cov(Y_n, Y_2) & \cdots & Var(Y_n) \end{bmatrix}$$

และในกรณีของเกาซ์เซียนคอพพูลาในตัวแบบนี้ ได้กำหนด

$$\tilde{\rho} = \Sigma = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \ddots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

และกำหนด $U = (U_1, U_2, \ldots U_n)$ แทน เกาซ์เซียนคอพพูลา n ตัวแปร ดังนั้น ฟังก์ชันการแจกแจงร่วมของเกาซ์เซียนคอพพูลา (Filip Lindskog, 2001) มีรูปแบบดังนี้

$$C_\Sigma^{Ga}(u) = \Phi_\Sigma^n\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)\right) \tag{34}$$
$$= \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}\varsigma'(\boldsymbol{\Sigma}^{-1} - \mathbf{I})\varsigma\right)$$

โดย $C_\Sigma^{Ga}(\boldsymbol{u})$ คือ ฟังก์ชันการแจกแจงร่วมของเกาซ์เซียนคอพพูลา $U = (U_1, U_2, \ldots U_n)$ ที่มีเมทริกซ์สหสัมพันธ์เชิงเส้น Σ เมื่อค่า $U = (U_1, U_2, \ldots U_n)$ เท่ากับ $u = (u_1, u_2, \ldots u_n)$

Φ_Σ^n คือ ฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน n ตัวแปร ด้วยเมทริกซ์ สหสัมพันธ์เชิงเส้น Σ

Φ^{-1} คือ ตัวผกผันของฟังก์ชันการแจกแจงสะสมของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ

Σ คือ เมทริกซ์ความแปรปรวนร่วมที่สมมาตร (Symmetric) และเป็นบวกแน่นอน (Positive definite)

และ $\varsigma = \left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)\right)'$

สำหรับกรณีเกาซ์เซียนคอพพูลา (Gaussian copula) 2 ตัวแปร (Alexander Mcneil et al., 1999) ได้ให้นิยามไว้ ดังนี้

$$C_{\rho}^{Ga}(u_1,u_2) = \Phi_{\rho}\left(\Phi^{-1}(u_1),\Phi^{-1}(u_2)\right)$$

$$= \int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(-\frac{s^2-2\rho st+t^2}{2(1-\rho^2)}\right)dsdt \qquad (35)$$

ฐิติมา จิรเศรษฐสิริ (2548) ได้กล่าวไว้ว่า เทคนิคเกาซ์เซียนคอพพูลาเป็นเทคนิคที่นิยมเป็นอย่างมาก สำหรับการจำลองตัวแปรสุ่มร่วมเมื่อทราบการแจกแจงส่วนริมและสหสัมพันธ์ เนื่องจากเทคนิคนี้เป็นเทคนิคที่มีขั้นตอนในการทำที่ง่าย จึงได้มีการนำเทคนิคนี้มาใช้งานกันอย่างแพร่หลายกว่าเทคนิคอื่นๆ

2.3 ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา (Logistic regression model with Gaussian copula)

ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา เป็นตัวแบบโลจิสติคที่มีความสัมพันธ์แบบคอพพูลาเข้ามาเกี่ยวข้องในตัวแบบ กล่าวคือ เป็นตัวแบบที่ตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กันด้วยเกาซ์เซียนคอพพูลา ซึ่งเป็นความสัมพันธ์ผ่านทางการแจกแจงร่วมของควอนไทล์ (Quantile) ของตัวแปรสุ่มร่วมที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal variables) โดยมีวัตถุประสงค์เช่นเดียวกับตัวแบบความถดถอยโลจิสติคแบบปกติ เพียงแต่ตัวแปรตามในตัวแบบนี้มีการแจกแจงแบบคอพพูลาเบอร์นูลลี (Copula Bernoulli distribution) กล่าวคือ ตัวแปรตามในแต่ละค่าสังเกตมีการแจกแจงแบบเบอร์นูลลีที่มีความสัมพันธ์กันด้วยเกาซ์เซียนคอพพูลา ดังนั้น ตัวแปรตาม (Y) มีค่าได้เพียง 2 ค่า คือ 0 และ 1

จาก $Y_i \sim Ber(p)$ ด้วยความน่าจะเป็น

$$P[Y_i = y_i] = p_i^{y_i}(1-p_i)^{1-y_i};\;\; y_i = 0,1 \qquad (36)$$

โดยที่ p_i มาจากแบบจำลองโพรบิทในสมการ (20) คือ

$$p_i = P[Y=1] = \Phi\left(\sum_{k=0}^{K}\beta_k X_{ik}\right)$$

และเกาซ์เซียนคอพพูลา $U = (U_1, U_2,...U_n)$ จากสมการ (34) ซึ่งมีรูปแบบ ดังนี้

$$C_{\Sigma}^{Ga}(u) = \Phi_{\Sigma}^{n}\left(\Phi^{-1}(u_1),\Phi^{-1}(u_2),...,\Phi^{-1}(u_n)\right)$$

ดังนั้น รูปแบบของสมการฟังก์ชันภาวะน่าจะเป็น (Likelihood function) ของตัวแปรตามในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา เป็นดังนี้

$$L = \int \int_{\min(y_i, 1-p_i)}^{\max(1-p_i, y_i)} \ldots\ldots \int C_{\Sigma}^{Ga}(u_1, u_2, \ldots, u_n)\, du_1 du_2 \ldots du_n \tag{37}$$

จะเห็นว่าในสมการ (37) มีรูปแบบฟังก์ชันที่ค่อนข้างซับซ้อน ซึ่งการประมาณค่าสัมประสิทธิ์ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum likelihood) ทำได้ยาก จึงได้เพิ่มสมมติฐานบางประการเพื่อให้สามารถวิเคราะห์ภาวะความน่าจะเป็นของตัวแบบได้

จากความสัมพันธ์แบบเกาซ์เซียนคอพพูลา ดังแสดงไว้ในสมการ (34) นั้น เรากำหนดให้ค่าสหสัมพันธ์รวม (ρ) ในเมทริกซ์สหสัมพันธ์เชิงเส้น Σ เป็นค่า ρ เดียวกันหมด

นั่นคือ
$$\widetilde{\rho} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \ddots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

ดังนั้น จึงสามารถเขียนความสัมพันธ์แบบเกาซ์เซียนคอพพูลา e_i' ซึ่งเริ่มจากความคลาดเคลื่อนหน่วยสังเกตที่ i; $i = 1,2,\ldots,n$ จากในสมการ (16) โดยได้มีการนำความสัมพันธ์แบบเกาซ์เซียนคอพพูลาเข้ามาประกอบในความคลาดเคลื่อนนั้น ด้วยวิธีของตัวแบบปัจจัยเชิงเดี่ยว (one factor model) ดังนี้

$$e_i' = \sqrt{\rho} Z + \sqrt{1-\rho}\,\varepsilon_i \tag{38}$$

นั่นคือ จะได้ตัวแปรเกาซ์เซียนคอพพูลา

$$U_i = \Phi(e_i') \tag{39}$$

เมื่อ e_i' แทน ตำแหน่งของเกาซ์เซียนคอพพูลาหน่วยสังเกตที่ i; $i = 1,2,\ldots,n$ ด้วยสหสัมพันธ์ ρ

Z แทน ค่าปัจจัยของคอพพูลา โดยที่ $Z \sim N(0,1)$

ε_i แทน ความคลาดเคลื่อนของหน่วยสังเกตที่ i, $i = 1,2,\ldots,n$ โดยที่ $\varepsilon_i \sim N(0,1)$

และจากแบบจำลองโพรบิทในสมการ (20) จะได้ว่า

จาก
$$P[Y=1] = p_i = \Phi\left(\sum_{k=0}^{K} \beta_k X_{ik}\right)$$

นั่นคือ Probit model:

$$\eta_i = \Phi^{-1}(p_i) = \sum_{k=0}^{K} \beta_k X_{ik} \tag{40}$$

นอกจากนี้ เมื่อกำหนดเงื่อนไขให้ค่าปัจจัยของคอพพูลา Z เป็นตัวแปรที่ทราบค่า พบว่า โอกาสที่จะเกิดเหตุการณ์เมื่อทราบค่า Z เป็นดังนี้

$$\begin{aligned}
P[Y_i = 1 \mid z] &= P\left[e_i' \le \Phi^{-1}(p_i)\right] \\
&= P\left[\sqrt{\rho}z + \sqrt{1-\rho}\,\varepsilon_i \le \Phi^{-1}(p_i)\right] \\
&= P\left(\varepsilon_i \le \frac{\Phi^{-1}(p_i) - \sqrt{\rho}z}{\sqrt{1-\rho}}\right) \\
&= P\left(\varepsilon_i \le \frac{\sum_{k=0}^{K} \beta_k X_{ik} - \sqrt{\rho}z}{\sqrt{1-\rho}}\right) \\
&= \Phi\left(\frac{\sum_{k=0}^{K} \beta_k X_{ik} - \sqrt{\rho}z}{\sqrt{1-\rho}}\right)
\end{aligned} \tag{41}$$

จากสมการ (41) จะเห็นว่า เมื่อกำหนดให้ค่าปัจจัยของคอพพูลา Z เป็นตัวแปรที่ทราบค่า จะส่งผลให้เกิดความเป็นอิสระกันของข้อมูล และพบว่าตัวแบบคอพพูลานี้ก็คือ ตัวแบบโพรบิท นั่นเอง สำหรับเหตุผลในการนำตัวแบบโพรบิทมาใช้เนื่องจากมีรูปแบบที่สอดคล้องกันกับเกาซ์เซียนคอพพูลามากกว่าตัวแบบโลจิท ซึ่งทำให้ง่ายต่อการแก้สมการหาตัวประมาณสัมประสิทธิ์ความถดถอย และจากที่ได้กล่าวมาแล้ว ตัวแบบทั้งสองนี้ให้ผลการศึกษาที่ใกล้เคียงกัน ดังนั้นจึงสามารถนำมาใช้แทนกันได้ นั่นคือ เราสามารถประมาณค่าพารามิเตอร์จากตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาด้วยการวิเคราะห์ความถดถอยโพรบิท เมื่อเราทราบค่าปัจจัยของคอพพูลา Z ในการศึกษา นอกจากนี้ค่าประมาณสัมประสิทธิ์ความถดถอยจากที่ได้ควรมีการนำมาปรับค่าโดยการคูณด้วย $\sqrt{1-\rho}$ จึงจะทำให้ได้ตัวประมาณค่าสัมประสิทธิ์ความถดถอยที่เหมาะสมสำหรับตัวแบบนี้

2.4 ตัวอย่างการประยุกต์ใช้งานตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา

พิจารณาการวิเคราะห์สินเชื่อรถยนต์ของบริษัทแห่งหนึ่ง

กำหนด Y แทน สถานะของลูกค้า โดยแบ่งหน่วยข้อมูลออกเป็น 2 กลุ่ม คือ

$Y=0$ ถ้าเป็นลูกหนี้ชั้นดี หรือ

$Y=1$ ถ้าเป็นลูกหนี้ NPL

ดังนั้น $Y \sim Ber(p)$ ที่มีฟังก์ชันความน่าจะเป็นดังแสดงไว้ในสมการ (36)

X แทน ชั้นรายได้ของลูกหนี้ โดยแบ่งหน่วยข้อมูลออกเป็น 2 กลุ่ม

$X=0$ ถ้าลูกหนี้มีรายได้ในระดับต่ำ (มีรายได้ต่อเดือนต่ำกว่า 15,000 บาท)

$X=1$ ถ้าลูกหนี้มีรายได้ในระดับสูง (มีรายได้ต่อเดือนสูงกว่า 15,000 บาท)

นั่นคือ กำหนดให้ $X \sim Ber(p)$ เพื่อความสะดวกและ ง่ายต่อการศึกษา กล่าวคือ

ในการวิเคราะห์ความถดถอยโลจิสติคโดยทั่วไป เมื่อทราบค่าตัวแปรที่สนใจศึกษาดังแสดงไว้ข้างต้นแล้ว สามารถนำตัวแปรเหล่านี้มาสร้างตัวแบบเพื่อใช้ในการพยากรณ์โอกาสการเกิดหนี้เสียของลูกหนี้ต่อไปได้ อย่างไรก็ตามในบางครั้งตัวแปรในข้างต้นอาจยังไม่เพียงพอในการอธิบายข้อมูลได้ครบถ้วน เนื่องจากยังมีปัจจัยอื่นที่มีความสำคัญและมักมีการละเลยหรือไม่ได้นำมาพิจารณา ซึ่งอาจได้แก่ ปัจจัยทางเศรษฐศาสตร์ระดับมหภาคที่มีผลกระทบต่อทุกคนในกลุ่มนั้นๆ

โดยกำหนด Z แทน อัตราดอกเบี้ยเงินกู้ ที่มีการปรับข้อมูลให้อยู่ในรูปมาตรฐานแล้ว
ซึ่ง Z ดังกล่าวนี้ เรียกว่า ปัจจัยคอพพูลา ดังแสดงไว้ในสมการ (38)

กล่าวคือ อัตราดอกเบี้ยเงินกู้ (Z) นี้ ถือว่าเป็นปัจจัยทางเศรษฐกิจในระดับมหภาคที่มีผลกระทบต่อลูกหนี้ทุกราย ซึ่งถ้าอัตราดอกเบี้ยสูง มีผลให้ลูกหนี้แต่ละคนมีโอกาสเกิดหนี้เสีย $(P[Y_i=1])$ มากขึ้นพร้อมๆ กัน ดังนั้นปัจจัย Z ที่มากระทบนี้ ส่งผลให้ Y มีความสัมพันธ์กันได้ ดังนั้น ในการวิเคราะห์ตัวแบบที่ทราบค่าปัจจัยอัตราดอกเบี้ยเงินกู้ (Z) ซึ่งมีผลทำให้โอกาสที่ลูกหนี้เกิดหนี้เสียพร้อมๆ กันได้นั้น $(P[Y_i=1 \mid z])$ ได้แสดงการพิสูจน์ไว้ในสมการ (41) ข้างต้น

นั่นคือ การประมาณค่าสัมประสิทธิ์ความถดถอยสำหรับตัวแบบนี้ สามารถใช้การวิเคราะห์ความถดถอยแบบโพรบิท ที่ทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ซึ่งหากละเลยปัจจัยคอพพูลานี้รวมถึงการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ผลการประมาณจะเป็นเช่นไร ดังนั้น ในงานวิจัยนี้จึงได้ทำการจำลองข้อมูลเพื่อประมาณค่าสัมประสิทธิ์ในกรณีดังกล่าวเหล่านี้ โดยการดำเนินงานวิจัยดังแสดงไว้ในบทต่อไป

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อต้องการศึกษาและประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา โดยอาศัยวิธีการวิเคราะห์ความถดถอยโพรบิทในการหาตัวประมาณค่าสัมประสิทธิ์การถดถอย โดยได้พิจารณากรณีศึกษาเป็น 4 กรณี ดังนี้

1. กรณีไม่ทราบค่าปัจจัยของคอพพูลา Z
2. กรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$
3. กรณีทราบค่าปัจจัยของคอพพูลา Z
4. กรณีทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

ในการทดลองจากกรณีศึกษาในกรณีต่างๆ ข้างต้นนั้น เป็นการทดลองเพื่อทดสอบความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ เมื่อ ρ คือ ค่าสหสัมพันธ์ในตัวแบบเกาซ์เซียนคอพพูลา ดังได้แสดงไว้ในบทที่ 2 ในหัวข้อตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาข้างต้น และทำการทดสอบเพิ่มเติมในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z เนื่องจากต้องการทราบว่าผลการประมาณค่าสัมประสิทธิ์การถดถอยเป็นเช่นไร หากละเลยปัจจัยคอพพูลา Z ดังกล่าว จึงได้ทำการทดลองในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และกรณีทราบค่าปัจจัยของคอพพูลา Z ซึ่งได้ทำการทดลองทั้งแบบที่มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ และแบบที่ไม่มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยใดๆ และสำหรับค่าปัจจัยของคอพพูลา Z สำหรับงานวิจัยนี้ถือว่าเป็นตัวแปรที่มีบทบาทในการประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีวิเคราะห์ความถดถอยแบบโพรบิท ซึ่งนำไปสู่ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาต่อไป

สำหรับข้อมูลที่ใช้ในงานวิจัยนี้ เป็นข้อมูลจากการจำลองโดยใช้เทคนิคการจำลองโดยวิธีมอนติคาร์โล (Monte Carlo Simulation Method) และทำการวิเคราะห์ข้อมูลด้วยโปรแกรม R โดยมีแผนการทดลองและขั้นตอนในการวิจัย ดังต่อไปนี้

3.1 แผนการดำเนินงานวิจัย

สำหรับการดำเนินงานวิจัยครั้งนี้ได้กำหนดสถานการณ์การจำลองต่างๆ ในการศึกษาและเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยของตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา ในกรณีต่างๆ ข้างต้น ดังต่อไปนี้

3.1.1 ตัวแปรอิสระ (X) จำนวน 1 ตัวแปร โดยเป็นตัวแปรเชิงคุณภาพ ที่มีการแจกแจงแบบเบอร์นูลลี (Bernoulli distribution) ด้วยพารามิเตอร์ p นั่นคือ $X \sim Ber(p)$ โดยมีฟังก์ชันความน่าจะเป็นอยู่ในรูปของ

$$p_X(k) = p^k (1-p)^{1-k}, k = 0,1$$

ในงานวิจัยครั้งนี้จะศึกษาที่ $p = 0.5$

3.1.2 ตัวแปรตาม (Y) เป็นตัวแปรเชิงคุณภาพ โดยมีการแจกแจงแบบคอพพูลาเบอร์นูลลี (Copula Bernoulli distribution) $(P[Y = y \mid z])$ นั่นคือ ค่าสังเกตในแต่ละค่าของตัวแปรตามมีความสัมพันธ์กันด้วย เกาซ์เซียนคอพพูลา (Gaussian copula) และกำหนดค่าตัวแปรตามมีค่าเพียง 2 ค่า คือ 0 และ 1

3.1.3 เกาซ์เซียนคอพพูลา (Gaussian copula) ที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

3.1.4 กำหนดพารามิเตอร์สัมประสิทธิ์การถดถอย $\beta_1 = 1$ และ $\beta_0 = -0.47$ โดยคำนวณค่า β_0 ที่ให้ความน่าจะเป็นในการเกิดเหตุการณ์เป็น 0.5

3.1.5 จำนวนกลุ่มตัวอย่างที่ใช้ในแต่ละรอบของการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม

3.1.6 จำนวนขนาดตัวอย่างในแต่ละกลุ่ม (n) เป็น 100, 500 และ 1,000

3.1.7 กำหนดจำนวนการกระทำซ้ำในแต่ละสถานการณ์เป็น 100 รอบ

3.1.8 ทำการจำลองข้อมูล (Simulation) ตามสถานการณ์ที่กำหนด โดยใช้โปรแกรม R

3.2 ขั้นตอนในการดำเนินงานวิจัย

ขั้นตอนในการดำเนินงานวิจัยในการศึกษาครั้งนี้ มีดังนี้

3.2.1 ศึกษาความสัมพันธ์แบบเกาซ์เซียนคอพพูลา (Gaussian copula) ด้วยเมทริกซ์สหสัมพันธ์ ρ

3.2.2 ศึกษาแบบจำลองโลจิทและแบบจำลองโพรบิท พร้อมทั้งศึกษารูปแบบของฟังก์ชันความน่าจะเป็นของตัวแปรตามในตัวแบบโลจิสติค เพื่อนำไปใช้สำหรับการประมาณค่าสัมประสิทธิ์การถดถอย

3.2.3 ทำการหารูปแบบฟังก์ชันภาวะน่าจะเป็น (Likelihood function) ของตัวแปรตามในตัวแบบความถดถอยโลจิสติคโดยนำความสัมพันธ์แบบเกาซ์เซียนคอพพูลาเข้ามาประกอบในการแจกแจงความน่าจะเป็นของตัวแบบความถดถอยโลจิสติค ทำให้ได้รูปแบบฟังก์ชันภาวะน่าจะเป็นของตัวแบบความถดถอยโลจิสติคที่ตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กัน

3.2.4 ทำการจำลองข้อมูลที่ใช้ในการวิจัย

3.2.5 ทำการประมาณสัมประสิทธิ์การถดถอย โดยอาศัยวิธีการวิเคราะห์ความถดถอยแบบโพรบิท (Probit regression) ใน 4 กรณีศึกษาข้างต้น

3.2.6 ทำการวิเคราะห์เปรียบเทียบผลการทดลองและสรุปผลการวิจัย

เนื่องจากขั้นตอนที่ 3.2.1 - 3.2.3 ในวิธีดำเนินงานวิจัยเป็นการศึกษาเชิงวิเคราะห์ ซึ่งรายละเอียดของการดำเนินงานวิจัยในขั้นตอนดังกล่าว ผู้วิจัยจึงขอเสนอไปในบทที่ 2 ในหัวข้อตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา ส่วนในขั้นตอนอื่นๆ ที่เหลือ (ขั้นตอนที่ 3.2.4 - 3.2.6) มีรายละเอียดการดำเนินการวิจัย ดังต่อไปนี้

การจำลองข้อมูลที่ใช้ในการวิจัย

ในการจำลองข้อมูลที่ใช้ในงานวิจัยมีขั้นตอน ดังนี้

1. กำหนดจำนวนกลุ่มตัวอย่างที่ใช้ในแต่ละรอบของการทดลอง (m) เป็น 1 , 5 และ 10 กลุ่มโดยที่กลุ่มข้อมูลในแต่ละกลุ่มเป็นอิสระต่อกัน
2. กำหนดจำนวนขนาดตัวอย่างในแต่ละกลุ่ม (n) เป็น 100, 500 และ 1,000
3. สร้างตัวแปรอิสระ (X) จำนวน 1 ตัวแปร เป็นตัวแปรเชิงคุณภาพ โดยกำหนดให้มีการแจกแจงแบบเบอร์นูลลี $(X \sim Ber(0.5))$
4. สร้างความสัมพันธ์เกาซ์เซียนคอพพูลา e'_{ij} ในหน่วยสังเกตที่ i; $i = 1,2,...,n$ และกลุ่มตัวอย่างที่ j; $j = 1,2,...,m$ โดยข้อมูลในแต่ละกลุ่มตัวอย่างเป็นอิสระต่อกัน แต่ข้อมูลในแต่ละค่าสังเกตมีความสัมพันธ์กันด้วยเกาซ์เซียนคอพพูลา โดยกำหนดให้ค่าสหสัมพันธ์รวม (ρ) ในเมทริกซ์สหสัมพันธ์เชิงเส้น Σ เป็นค่า ρ เดียวกันหมด ดังนั้นสามารถสร้างความสัมพันธ์แบบเกาซ์เซียนคอพพูลาด้วยวิธี one factor model ดังนี้

$$e'_{ij} = \sqrt{\rho} Z_j + \sqrt{1-\rho}\,\varepsilon_{ij} \tag{42}$$

โดยที่ $Z_j \sim N(0,1)$ และ $\varepsilon_{ij} \sim N(0,1)$

และจะได้ตัวแปรเกาซ์เซียนคอพพูลา $U_{ij} = \Phi(e'_{ij})$

คำนวณค่าจากแบบจำลองโพรบิทจากสมการ (20)

$$p_{ij} = \Phi(\beta_0 + \beta_1 X_{ij})$$

ดังนั้น จากสมการ (40); $$\eta_{ij} = \Phi^{-1}\left(p_{ij}\right) = \beta_0 + \beta_1 X_{ij}$$

และ กำหนด $\beta_1 = 1$ และหาค่า β_0 ที่ให้ความน่าจะเป็นในการเกิดเหตุการณ์เป็น 0.5 ซึ่ง คำนวณได้ค่า $\beta_0 = -0.47$

5. สร้างตัวแปรตาม (Y) ที่มีการแจกแจงแบบคอพพูลาเบอร์นูลลี (Copula Bernoulli distribution) โดยคำนวณจากขั้นตอนที่ 3 และ 4 เปรียบเทียบกัน โดยตัวแปรตาม (Y) มีค่าได้เพียง 2 ค่า คือ 0 และ 1

การประมาณค่าพารามิเตอร์สัมประสิทธิ์การถดถอย

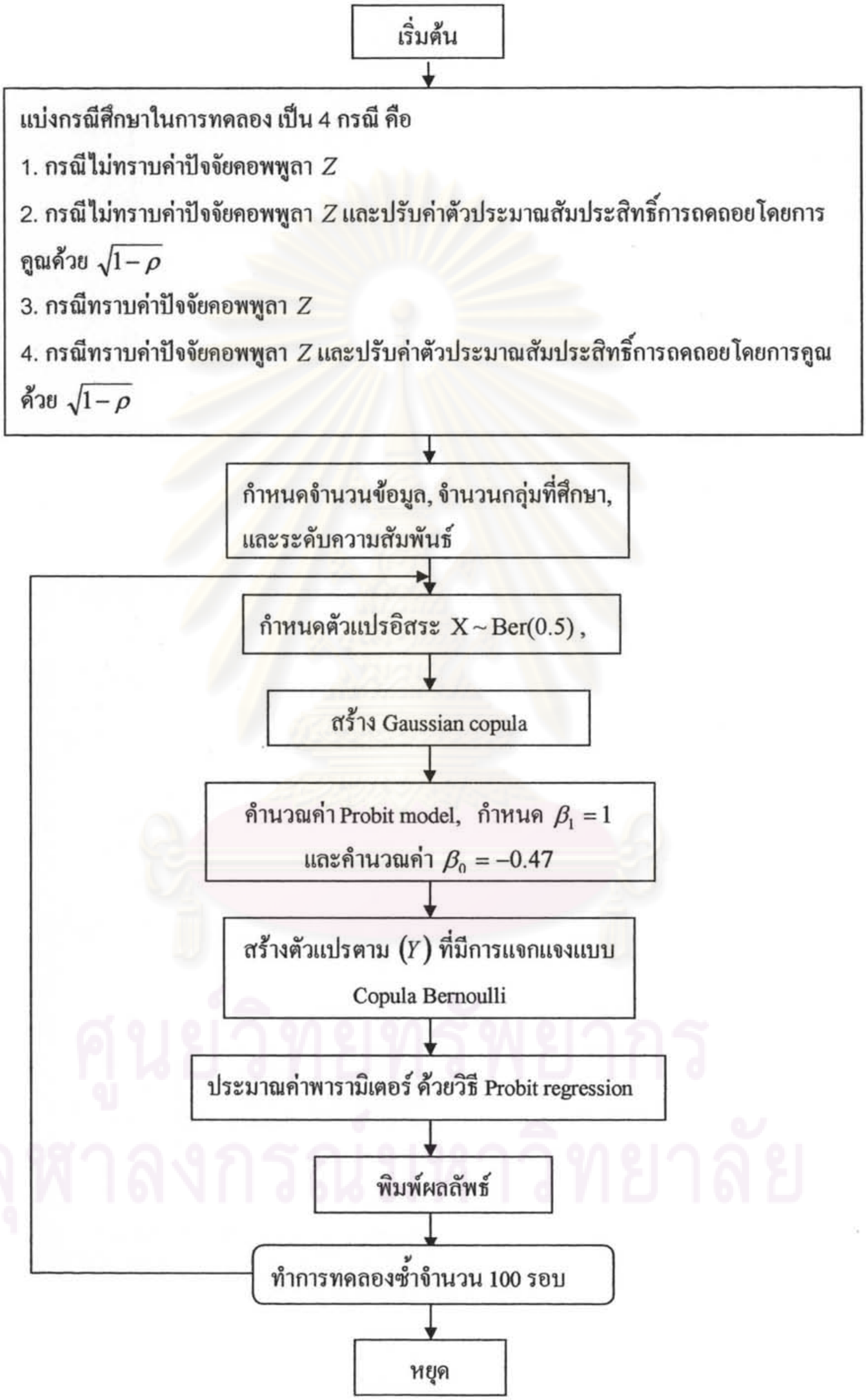
สำหรับการประมาณค่าสัมประสิทธิ์การถดถอยของตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลานั้น จะเห็นรูปแบบของสมการฟังก์ชันภาวะน่าจะเป็นค่อนข้างซับซ้อน ซึ่งการวิเคราะห์ภาวะความน่าจะเป็นสูงสุดทำได้ยาก อย่างไรก็ตาม หากทราบค่าปัจจัยคอพพูลา Z ทำให้ตัวแบบที่ซับซ้อนจะลดรูปเป็นตัวแบบโพรบิทดังที่แสดงไว้ในสมการ (41) จึงได้อาศัยแบบจำลองโพรบิทมาใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับตัวแบบนี้ ซึ่งแบบจำลองโพรบิทได้ให้ผลการวิเคราะห์ข้อมูลใกล้เคียงกับแบบจำลองโลจิท ดังนั้นในการประมาณค่าสัมประสิทธิ์การถดถอยของตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาในงานวิจัยนี้จะทำการประมาณค่าพารามิเตอร์ด้วยการวิเคราะห์ความถดถอยโพรบิท และทำการจำลองสถานการณ์เพื่อทดสอบความถูกต้องในการประมาณค่าพารามิเตอร์ในกรณีที่ทราบค่าปัจจัยของคอพพูลา Z ที่มีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ และทำการทดสอบเพิ่มเติมในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z โดยแบ่งเป็นกรณีศึกษาทั้งหมด 4 กรณี แล้วนำมาเปรียบเทียบกับค่าพารามิเตอร์ที่กำหนดจากข้อมูลในการจำลอง

การเปรียบเทียบและสรุปผลการทดลอง

เมื่อทำการประมาณค่าสัมประสิทธิ์การถดถอยในแต่ละกรณีศึกษาแล้ว นำผลการทดลองที่ได้มาสรุปผลในรูปแบบตารางและรูปภาพเพื่อแสดงการเปรียบเทียบของผลการศึกษาในแต่ละกรณีศึกษา

สำหรับขั้นตอนการจำลองในงานวิจัยข้างต้นนั้น ได้แสดงเป็นแผนภาพแสดงขั้นตอนในงานวิจัย ดังต่อไปนี้

แผนภาพที่ 3.2 แผนภาพแสดงขั้นตอนในการดำเนินงานวิจัย



บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา โดยอาศัยวิธีการวิเคราะห์ความถดถอยโพรบิทในการหาค่าตัวประมาณสัมประสิทธิ์การถดถอย และได้ทำการจำลองข้อมูลเพื่อทดสอบความถูกต้องในการประมาณสัมประสิทธิ์การถดถอยในกรณีที่ทราบค่าปัจจัยของคอพพูลา Z และทำการทดสอบเพิ่มเติมในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z โดยแบ่งเป็นกรณีศึกษาทั้งหมด 4 กรณี ดังต่อไปนี้

1. กรณีไม่ทราบค่าปัจจัยของคอพพูลา Z
2. กรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอย โดยการคูณด้วย $\sqrt{1-\rho}$
3. กรณีทราบค่าปัจจัยของคอพพูลา Z
4. กรณีทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอย โดยการคูณด้วย $\sqrt{1-\rho}$

ในการเปรียบเทียบผลการทดลองนั้น จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) เป็นเกณฑ์ในการตัดสินใจว่า ระดับความสัมพันธ์และจำนวนกลุ่มข้อมูลใดในแต่ละกรณีศึกษาที่ทำให้ได้ค่าประมาณของสัมประสิทธิ์การถดถอยมีค่าใกล้เคียงกับค่าพารามิเตอร์มากที่สุด นอกจากนี้ยังพิจารณาค่าเฉลี่ย และค่าความคลาดเคลื่อนมาตรฐาน ในการทดลองซ้ำจำนวน 100 รอบ ประกอบการตัดสินใจ เพื่อให้ได้ผลการทดลองที่ตัวประมาณมีค่าใกล้เคียงกับพารามิเตอร์ให้มากที่สุด

การนำเสนอผลการวิจัยนี้ ได้มีการนำเสนอในรูปแบบตารางและรูปภาพ โดยมีการใช้สัญลักษณ์แทนความหมายต่างๆ ดังนี้

n	แทน	ขนาดตัวอย่างในแต่ละกลุ่มตัวอย่าง
m	แทน	กลุ่มตัวอย่างที่ใช้ในแต่ละรอบของการทดลอง
ρ	แทน	ระดับความสัมพันธ์ในตัวแปรตาม
MSE	แทน	ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอย
Mean	แทน	ค่าเฉลี่ยของสัมประสิทธิ์การถดถอย
SD	แทน	ส่วนเบี่ยงเบนมาตรฐานของสัมประสิทธิ์การถดถอย
b_0, b_1	แทน	ตัวประมาณค่าสัมประสิทธิ์การถดถอยของพารามิเตอร์ β_0, β_1

การนำเสนอผลการวิจัยในการเปรียบเทียบค่าประมาณของสัมประสิทธิ์การถดถอยในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา โดยอาศัยวิธีการวิเคราะห์ความถดถอยโพรบิทนั้นได้จำแนกออกเป็น 4 กรณีศึกษา ดังนี้

กรณีศึกษาที่ 4.1 ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z

4.1.1 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 100 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.1.2 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 500 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.1.3 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 1000 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.1.4 กรณีจำนวนตัวอย่างทั้งหมดในแต่ละชุด (mn) เท่ากับ 500, 1000 และ 5000 และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

กรณีศึกษาที่ 4.2 ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

4.2.1 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 100 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.2.2 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 500 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.2.3 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 1000 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.2.4 กรณีจำนวนตัวอย่างทั้งหมดในแต่ละชุด (mn) เท่ากับ 500, 1000 และ 5000 และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

กรณีศึกษาที่ 4.3 ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีทราบค่าปัจจัยของคอพพูลา Z

4.3.1 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 100 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.3.2 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 500 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.3.3 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 1000 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.3.4 กรณีจำนวนตัวอย่างทั้งหมดในแต่ละชุด (mn) เท่ากับ 500, 1000 และ 5000 และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

กรณีศึกษาที่ 4.4 ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

4.4.1 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 100 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.4.2 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 500 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.4.3 กรณีจำนวนขนาดตัวอย่างในแต่ละกลุ่มทดลอง (n) เท่ากับ 1000 โดยมีจำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

4.4.4 กรณีจำนวนตัวอย่างทั้งหมดในแต่ละชุด (mn) เท่ากับ 500, 1000 และ 5000 และกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

รูปแบบการนำเสนอผลการวิจัยในแต่ละกรณีศึกษานั้น เริ่มจากผลการวิจัยจากตาราง และแสดงรูปภาพในการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยทุกกรณีย่อย พร้อมทั้งอธิบายผลการวิจัยในแต่ละกรณีศึกษา โดยผลการวิเคราะห์ข้อมูลในแต่ละกรณีศึกษา เป็นดังนี้

<u>กรณีศึกษาที่ 4.1</u> ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z

การวิจัยในกรณีนี้ ได้ทำการศึกษากรณีที่จำนวนขนาดตัวอย่างตัวอย่างในการทดลองเป็น 100, 500 และ 1000 โดยในแต่ละขนาดตัวอย่าง ได้ทำการศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนกลุ่มตัวอย่างในการทดลองเป็น 1, 5 และ 10 กลุ่ม ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.1.1 – 4.1.3 ส่วนในกรณี 4.1.4 เป็นการนำเสนอรูปภาพแสดงกรณีจำนวนตัวอย่างทั้งหมดในแต่ละชุด (mn) ที่เท่ากันในแต่ละกรณีย่อย ซึ่งเท่ากับ 500, 1000 และ 5000 โดยกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

ตารางที่ 4.1.1 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 100

		$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0697		0.6131		3.2551		11.4260	
	Mean	-0.481	0.986	-0.578	1.212	-1.007	1.853	-2.031	3.344
	SD	0.174	0.332	0.873	0.646	1.964	1.299	3.393	1.886
5	MSE	0.01034		0.02735		0.07193		0.17677	
	Mean	-0.486	1.019	-0.483	1.024	-0.491	1.041	-0.516	1.069
	SD	0.081	0.117	0.199	0.122	0.318	0.206	0.466	0.364
10	MSE	0.00498		0.0147		0.03803		0.08651	
	Mean	-0.475	1.006	-0.492	1.015	-0.506	1.031	-0.524	1.051
	SD	0.053	0.085	0.147	0.085	0.237	0.136	0.329	0.247

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีไม่ทราบค่าปัจจัยคอพพูลา Z เมื่อขนาดตัวอย่างเท่ากับ 100 พบว่า เมื่อระดับความสัมพันธ์เพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง ซึ่งที่ระดับความสัมพันธ์ $\rho = 0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด โดยกลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมากโดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น แต่เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มเพิ่มขึ้นค่อนข้างน้อยในแต่ละระดับความสัมพันธ์ที่สูงขึ้น

เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ในทุกระดับความสัมพันธ์ และทุกกลุ่มตัวอย่างมีค่าต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Underestimate) โดยเฉพาะในกลุ่มตัวอย่างเท่ากับ 1 กลุ่ม และในระดับความสัมพันธ์ที่สูงขึ้น ส่วนค่าประมาณสัมประสิทธิ์การถดถอย b_1 มีค่าสูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Overestimate) สำหรับส่วนเบี่ยงเบนมาตรฐานในการทดลอง 100 รอบ พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากในระดับความสัมพันธ์สูงและในกลุ่มตัวอย่างน้อย

ตารางที่ 4.1.2 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 500

		$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0479		0.3943		1.3491		8.0767	
	Mean	-0.512	1.023	-0.630	1.170	-0.860	1.554	-1.590	3.035
	SD	0.265	0.155	0.778	0.367	1.369	0.623	2.928	1.513
5	MSE	0.0024		0.0310		0.0881		0.2302	
	Mean	-0.466	1.005	-0.482	1.021	-0.504	1.054	-0.531	1.102
	SD	0.036	0.060	0.235	0.084	0.382	0.166	0.567	0.359
10	MSE	0.0012		0.0140		0.0401		0.0879	
	Mean	-0.468	1.004	-0.484	1.008	-0.498	1.020	-0.511	1.032
	SD	0.026	0.041	0.159	0.053	0.258	0.115	0.353	0.224

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีไม่ทราบค่าปัจจัยคอพพูลา Z เมื่อขนาดตัวอย่างเท่ากับ 500 พบว่า เมื่อระดับความสัมพันธ์เพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง ซึ่งที่ระดับความสัมพันธ์ $\rho = 0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด โดยกลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมากตามระดับความสัมพันธ์ที่สูงขึ้น เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มเพิ่มขึ้นค่อนข้างน้อยในแต่ละระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์

เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ในทุกระดับความสัมพันธ์ และทุกกลุ่มตัวอย่างมีค่าต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Underestimate) โดยเฉพาะในกลุ่มตัวอย่างเท่ากับ 1 กลุ่ม และในระดับความสัมพันธ์ที่สูงขึ้น ส่วนค่าประมาณสัมประสิทธิ์การถดถอย b_1 มีค่าสูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Overestimate) สำหรับส่วนเบี่ยงเบนมาตรฐานในการทดลอง 100 รอบ พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากในระดับความสัมพันธ์สูงและในจำนวนกลุ่มตัวอย่างที่น้อย

ตารางที่ 4.1.3 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 1000

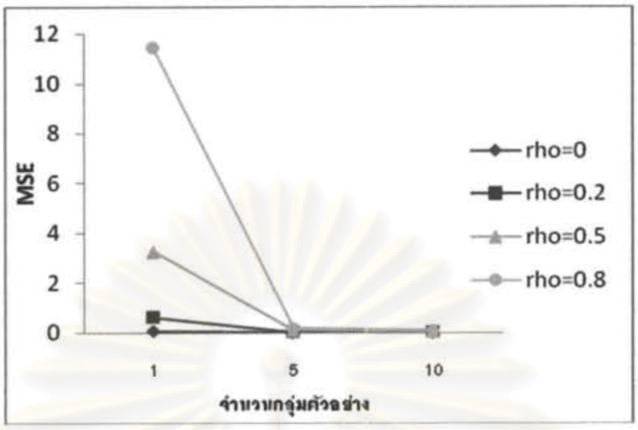
		$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0098		0.1540		1.1070		7.0681	
	Mean	-0.471	1.003	-0.550	1.111	-0.761	1.511	-1.506	2.955
	SD	0.108	0.090	0.514	0.169	1.261	0.544	2.713	1.404
5	MSE	0.0010		0.0198		0.0634		0.1736	
	Mean	-0.468	1.005	-0.47	1.018	-0.488	1.044	-0.500	1.089
	SD	0.025	0.038	0.180	0.084	0.302	0.186	0.442	0.382
10	MSE	0.0005		0.0105		0.0327		0.0796	
	Mean	-0.467	0.999	-0.470	1.004	-0.472	1.014	-0.474	1.029
	SD	0.017	0.025	0.134	0.058	0.222	0.130	0.313	0.249

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีไม่ทราบค่าปัจจัยคอพพูลา Z เมื่อขนาดตัวอย่างเท่ากับ 1000 พบว่า เมื่อระดับความสัมพันธ์เพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง ซึ่งที่ระดับความสัมพันธ์ $\rho = 0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ ตามลำดับ โดยที่กลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมากตามระดับความสัมพันธ์ที่สูงขึ้น เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มเพิ่มขึ้นค่อนข้างน้อยในแต่ละระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ นอกจากนี้ยังพบว่าในขนาดตัวอย่างในกรณีย่อยนี้ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าขนาดตัวอย่างที่กล่าวมาแล้วข้างต้น ในแต่ละระดับความสัมพันธ์ และในแต่ละกลุ่มตัวอย่าง

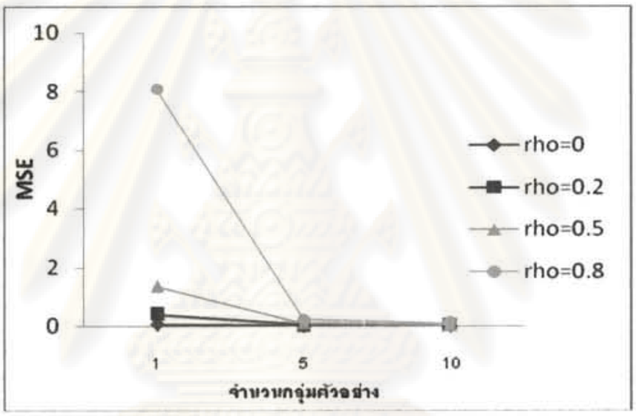
เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า ในกลุ่มตัวอย่างเท่ากับ 1 กลุ่ม มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็น แต่ค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น ส่วนจำนวนกลุ่มตัวอย่างเท่ากับ 5 กลุ่ม มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็นเล็กน้อยและค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์เล็กน้อยในระดับความสัมพันธ์ที่สูง และกลุ่มตัวอย่างเท่ากับ 10 กลุ่ม มีค่าเฉลี่ยของค่าประมาณค่อนข้างใกล้เคียงกับค่าพารามิเตอร์ และในทุกระดับความสัมพันธ์ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยลดลงเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น สำหรับส่วนเบี่ยงเบนมาตรฐานในการทดลอง 100 รอบ พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากในระดับความสัมพันธ์สูง และในจำนวนกลุ่มตัวอย่างที่น้อย และมีการกระจายลดลงเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น

จากตารางการวิเคราะห์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยใน 3 กรณีย่อยเมื่อเปลี่ยนขนาดตัวอย่างเป็น 100, 500 และ 1000 ข้างต้นนั้น สามารถพิจารณาจากรูปภาพแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ จำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม และขนาดตัวอย่างในแต่ละกลุ่มทดลองเป็น 100, 500 และ 1000 ดังรูปที่ 4.1.1 – 4.1.3 ดังต่อไปนี้

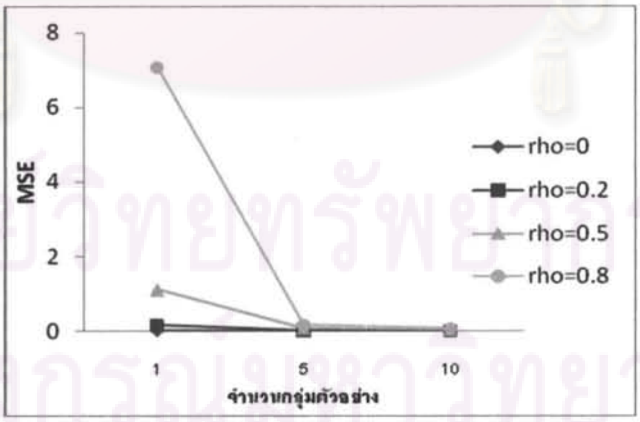
ก) กรณีจำนวนขนาดตัวอย่างเท่ากับ 100



ข) กรณีจำนวนขนาดตัวอย่างเท่ากับ 500

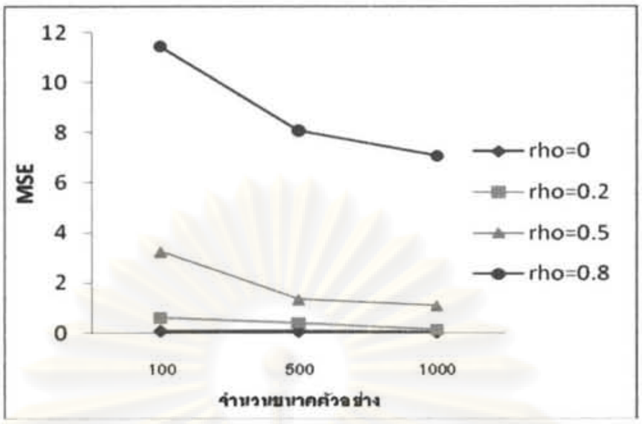


ค) กรณีจำนวนขนาดตัวอย่างเท่ากับ 1000

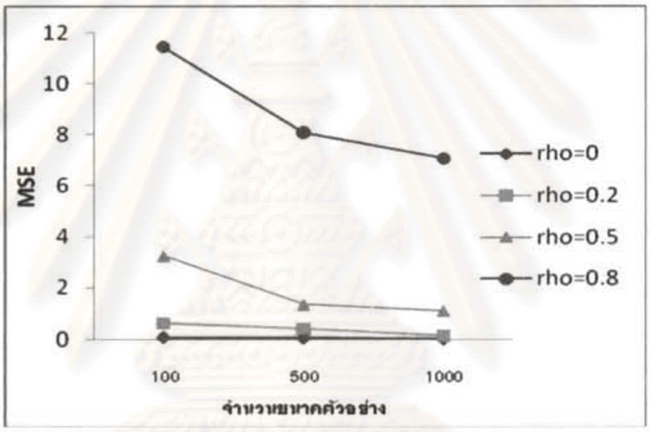


รูปที่ 4.1.1 แสดงการเปรียบค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ในแต่ละขนาดตัวอย่าง

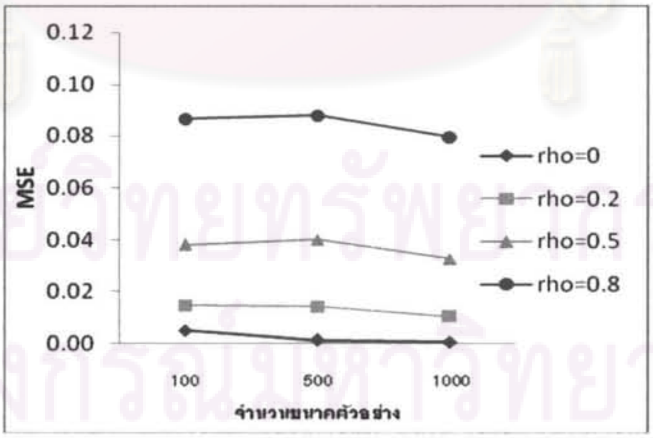
ก) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 1



ข) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 5

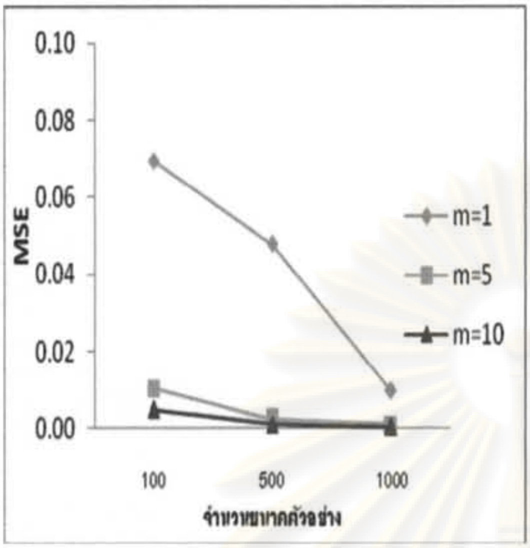


ค) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 10

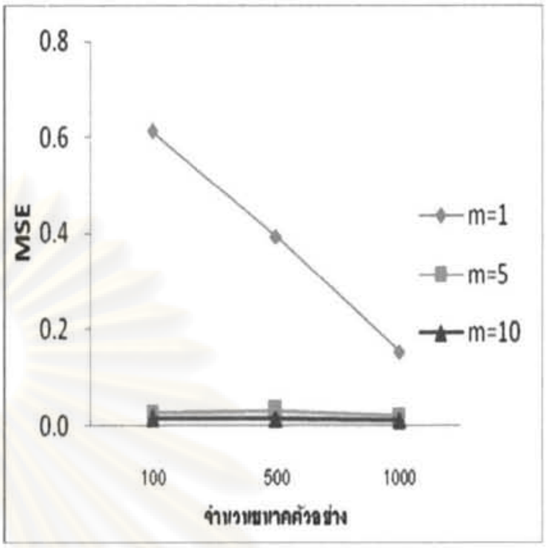


รูปที่ 4.1.2 แสดงการเปรียบค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนขนาดตัวอย่างเท่ากับ 100, 500 และ 1000 ในแต่ละกลุ่มตัวอย่าง

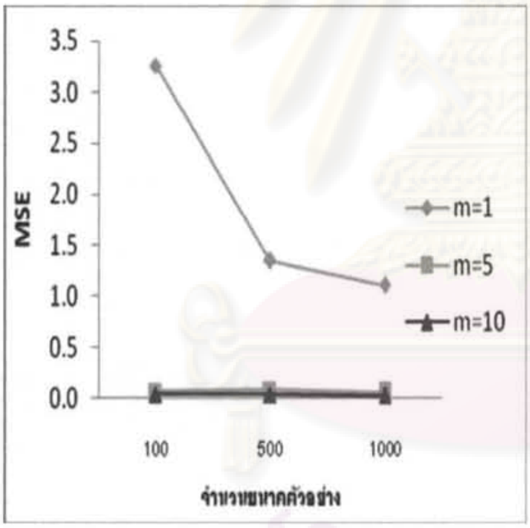
ก) $\rho = 0$



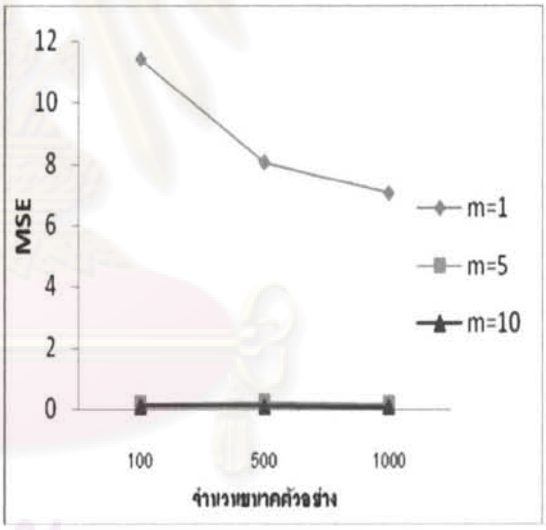
ข) $\rho = 0.2$



ค) $\rho = 0.5$



ง) $\rho = 0.8$



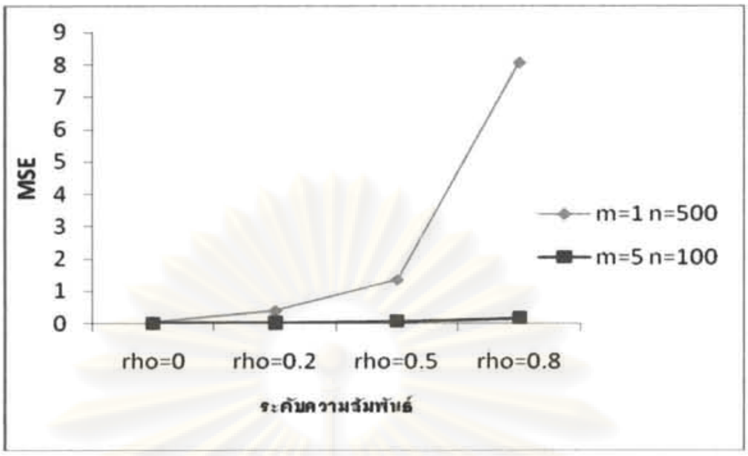
รูปที่ 4.1.3 แสดงการเปรียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่จำนวนขนาดตัวอย่างเท่ากับ 100, 500 และ 1000 และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ในแต่ละระดับความสัมพันธ์

จากรูป 4.1.1 – 4.1.3 สามารถสรุปผลการวิจัยเกี่ยวกับการเปลี่ยนแปลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) สำหรับในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z ดังนี้

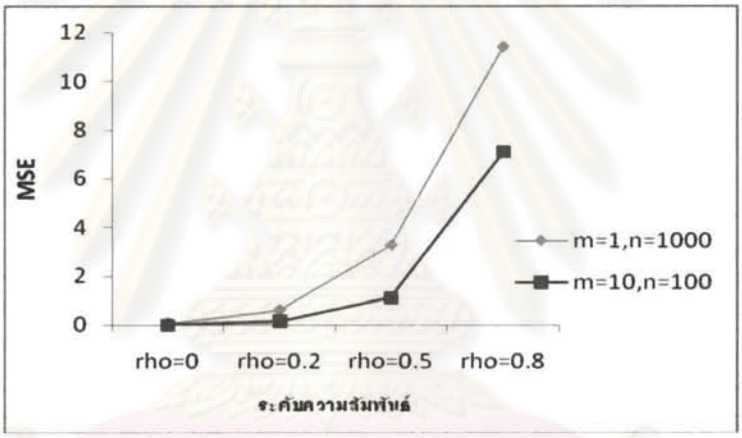
1. <u>เมื่อระดับความสัมพันธ์เพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นในทุกกลุ่มตัวอย่างและทุกขนาดตัวอย่าง โดยเฉพาะที่ระดับความสัมพันธ์ $\rho = 0.8$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมาก ซึ่งสังเกตได้ชัดเจนในกลุ่มตัวอย่างเท่ากับ 1 กลุ่ม ในรูปที่ 4.1.1 และรูปที่ 4.1.2 นอกจากนี้พบว่าในทุกกลุ่มตัวอย่างและในทุกขนาดตัวอย่าง ที่ระดับความสัมพันธ์ $\rho = 0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ ตามลำดับ
2. <u>เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง โดยมีแนวโน้มลดลงอย่างรวดเร็วในช่วงแรกที่มีจำนวนกลุ่มตัวอย่างน้อย และค่อยๆ ลดลงเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น และในทุกระดับความสัมพันธ์พบว่า จำนวนกลุ่มตัวอย่างเท่ากับ 10 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ จำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 1 กลุ่มตามลำดับ เนื่องจากจำนวนกลุ่มตัวอย่างที่เพิ่มขึ้นนี้จะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยให้น้อยลง
3. <u>เมื่อขนาดตัวอย่างเพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์และทุกกลุ่มตัวอย่าง โดยมีแนวโน้มในการลดลงค่อนข้างมากในช่วงที่มีจำนวนขนาดตัวอย่างน้อย และค่อยๆ ลดลงเมื่อจำนวนขนาดตัวอย่างเพิ่มมากขึ้น โดยเฉพาะจำนวนขนาดตัวอย่างเท่ากับ 1000 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด

จากในกรณีย่อยทั้ง 3 กรณีข้างต้นนั้น เป็นการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละกลุ่มและแต่ละขนาดทดลอง เช่น ในกรณีย่อย 4.1.1 เป็นกรณีที่ขนาดตัวอย่าง $n = 100$ โดยที่จำนวนกลุ่ม $m = 1,5$ และ 10 กลุ่ม เป็นการเปรียบเทียบในลักษณะจำนวนข้อมูลในแต่ละชุด (mn)ไม่เท่ากัน นั่นคือ $mn = 100$, 500 และ 1000 ซึ่งการเปรียบเทียบในลักษณะดังกล่าวอาจยังดูไม่ชัดเจน ทำให้เกิดข้อสงสัยในผลการเปรียบเทียบจากกรณีที่จำนวนข้อมูลทั้งหมดที่ไม่เท่ากันได้ ดังนั้นในกรณีดังต่อไปนี้ เป็นการแสดงภาพและอธิบายการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ สำหรับกรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn)ที่เท่ากันของแต่ละกรณีย่อย ซึ่งมีค่าเท่ากับ 500, 1000 และ 5000 ดังรูปที่ 4.1.4 ดังต่อไปนี้

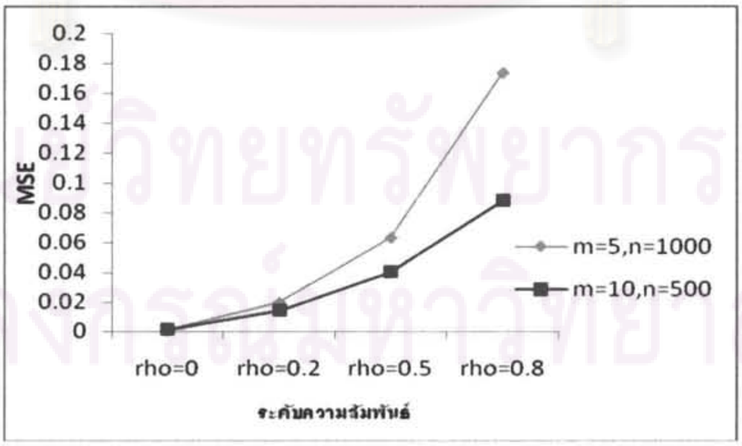
ก) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 500



ข) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 1000



ค) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 5000



รูปที่ 4.1.4 แสดงการเปรียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 500, 1000 และ 5000 ในแต่ละระดับความสัมพันธ์

จากรูป 4.1.4 ซึ่งแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีจำนวนข้อมูลตัวอย่างทั้งหมดในแต่ละชุด (mn) ที่เท่ากัน ซึ่งได้แก่จำนวนข้อมูล 500, 1000 และ 5000 ใน 3 กรณีข้างต้นนั้น พบว่า กลุ่มตัวอย่าง (m) ที่ต่างกันมีผลต่อค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมากกว่าขนาดตัวอย่าง (n) โดยเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกชุดข้อมูล โดยระดับการลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีที่กลุ่มตัวอย่าง 1 กลุ่มกับกลุ่มตัวอย่าง 5 กลุ่ม หรือ 10 กลุ่ม (ดังในกรณี ก, ข) มีอัตราการลดลงค่อนข้างมากในแต่ละระดับความสัมพันธ์ ส่วนในกรณีกลุ่มตัวอย่าง 5 กลุ่มกับ 10 กลุ่ม (ดังในกรณี ค) ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีอัตราการลดลงที่ไม่แตกต่างกันมากนัก สำหรับกรณีที่จำนวนชุดข้อมูลเพิ่มขึ้น จะเห็นว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงด้วย นั่นคือในกรณีที่มีจำนวนขนาดตัวอย่างน้อย ควรมีการเพิ่มจำนวนกลุ่มตัวอย่างให้มากขึ้นเพื่อส่งผลให้ข้อมูลมีจำนวนมากพอที่สามารถให้ผลการประมาณที่ถูกต้องมากขึ้น นอกจากนี้จะเห็นว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นตามระดับความสัมพันธ์ที่สูงขึ้นทุกชุดของข้อมูล ซึ่งแสดงให้เห็นว่าการประมาณค่าสัมประสิทธิ์โดยใช้การวิเคราะห์แบบโพรบิทนี้ยังไม่เหมาะสมสำหรับในกรณีไม่ทราบค่าปัจจัยคอพพูลา Z

สรุปผลการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z

สำหรับทุกขนาดตัวอย่าง พบว่า ในระดับความสัมพันธ์ที่สูงขึ้นส่งผลให้ค่าประมาณของสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Underestimate) และค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น (Overestimate) อีกทั้งการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยในการทดลอง 100 รอบ ค่อนข้างมากในทุกจำนวนกลุ่มตัวอย่าง แต่เมื่อจำนวนกลุ่มตัวอย่างและจำนวนขนาดตัวอย่างในแต่ละกลุ่มเพิ่มขึ้นทำให้ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยและส่วนเบี่ยงเบนมาตรฐานมีค่าลดลงในทุกระดับความสัมพันธ์ โดยเฉพาะจำนวนขนาดตัวอย่างเท่ากับ 1000 จำนวนกลุ่มตัวอย่างเท่ากับ 10 จะเห็นว่าค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยค่อนข้างใกล้เคียงกับค่าพารามิเตอร์มากที่สุดเมื่อเทียบกับกรณีย่อยอื่น และจากการพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสำหรับกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z นี้ พบว่า ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุดที่ระดับความสัมพันธ์ $\rho = 0$ จำนวนกลุ่มตัวอย่างเท่ากับ 10 และจำนวนขนาดตัวอย่างเท่ากับ 1000 ประกอบกับค่าประมาณของสัมประสิทธิ์การถดถอยสำหรับกรณีนี้สูงกว่าค่าพารามิเตอร์ที่ควรจะเป็น โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น จึงทำให้การประมาณค่าสัมประสิทธิ์การถดถอยโดยอาศัยแบบจำลองโพรบิทอาจยังไม่เหมาะสมสำหรับกรณีนี้มากนัก

<u>กรณีศึกษาที่ 4.2</u> ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

การวิจัยในกรณีนี้ ได้ทำการศึกษากรณีที่จำนวนขนาดตัวอย่างตัวอย่างในการทดลองเป็น 100, 500 และ 1000 โดยในแต่ละขนาดตัวอย่าง ได้ทำการศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยที่ระดับความสัมพันธ์ $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$ และจำนวนกลุ่มตัวอย่างในการทดลองเป็น 1, 5 และ 10 กลุ่ม ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.2.1 – 4.2.3 ส่วนในกรณี 4.2.4 เป็นการนำเสนอรูปภาพแสดงกรณีจำนวนตัวอย่างทั้งหมดในแต่ละชุด (mn) ที่เท่ากันในแต่ละกรณีย่อย ซึ่งเท่ากับ 500, 1000 และ 5000 โดยกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$

ตารางที่ 4.2.1 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 100

		$\rho=0$		$\rho=0.2$		$\rho=0.5$		$\rho=0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0670		0.4724		1.4509		1.7109	
	Mean	-0.481	0.986	-0.517	1.084	-0.712	1.311	-0.908	1.495
	SD	0.174	0.332	0.782	0.578	1.389	0.919	1.517	0.844
5	MSE	0.0103		0.0258		0.0777		0.1995	
	Mean	-0.486	1.019	-0.432	0.916	-0.347	0.737	-0.231	0.478
	SD	0.081	0.117	0.178	0.109	0.225	0.145	0.208	0.163
10	MSE	0.0050		0.0162		0.0614		0.1848	
	Mean	-0.476	1.006	-0.440	0.908	-0.358	0.729	-0.235	0.470
	SD	0.053	0.085	0.132	0.076	0.167	0.096	0.147	0.110

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีไม่ทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ เมื่อขนาดตัวอย่างเท่ากับ 100 พบว่า ในระดับความสัมพันธ์ที่สูงขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง ซึ่งที่ระดับความสัมพันธ์ $\rho=0$ มีค่าเฉลี่ยความ

คลาดเคลื่อนกำลังสองต่ำสุด โดยที่จำนวนกลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมากตามระดับความสัมพันธ์ที่สูงขึ้นโดยเฉพาะที่ระดับ $\rho = 0.5$ และ $\rho = 0.8$ สำหรับจำนวนกลุ่มตัวอย่างที่เหลือพบว่า ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มเพิ่มขึ้นค่อนข้างน้อยในแต่ละระดับความสัมพันธ์ที่สูงขึ้น นอกจากนี้ยังพบว่าเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงไม่มากนักในแต่ละระดับความสัมพันธ์

เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า สำหรับจำนวนกลุ่มตัวอย่าง 1 กลุ่ม มีค่าประมาณของสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ สำหรับจำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 10 กลุ่มนั้น ค่าประมาณของสัมประสิทธิ์การถดถอย b_0 สูงกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 ต่ำกว่าค่าพารามิเตอร์ โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น สำหรับส่วนเบี่ยงเบนมาตรฐาน พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากโดยเฉพาะในระดับความสัมพันธ์สูง

ตารางที่ 4.2.2 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 500

		$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0479		0.2991		0.5742		1.1685	
	Mean	-0.512	1.023	-0.564	1.046	-0.608	1.099	-0.711	1.357
	SD	0.265	0.155	0.697	0.328	0.968	0.441	1.310	0.677
5	MSE	0.0024		0.0291		0.0818		0.2001	
	Mean	-0.466	1.005	-0.431	0.913	-0.356	0.746	-0.238	0.493
	SD	0.036	0.060	0.210	0.075	0.270	0.118	0.254	0.161
10	MSE	0.0012		0.0167		0.0656		0.1914	
	Mean	-0.468	1.004	-0.433	0.902	-0.352	0.721	-0.229	0.462
	SD	0.026	0.041	0.143	0.047	0.182	0.081	0.158	0.100

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีไม่ทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ เมื่อขนาดตัวอย่างเท่ากับ 500 พบว่า เมื่อระดับความสัมพันธ์ที่สูงขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง โดยที่กลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมากในระดับความสัมพันธ์ $\rho = 0.5$ และ $\rho = 0.8$ ส่วนที่กลุ่มตัวอย่างที่เหลือมีแนวโน้มเพิ่มขึ้นค่อนข้างน้อยในแต่ละระดับความสัมพันธ์ที่สูงขึ้น นอกจากนี้ยังพบว่าเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์

เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า สำหรับจำนวนกลุ่มตัวอย่าง 1 กลุ่ม มีค่าประมาณของสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ สำหรับจำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 10 กลุ่มนั้น ค่าประมาณของสัมประสิทธิ์การถดถอย b_0 สูงกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 ต่ำกว่าค่าพารามิเตอร์ โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น สำหรับส่วนเบี่ยงเบนมาตรฐาน พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากในระดับความสัมพันธ์สูง

ตารางที่ 4.2.3 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 1000

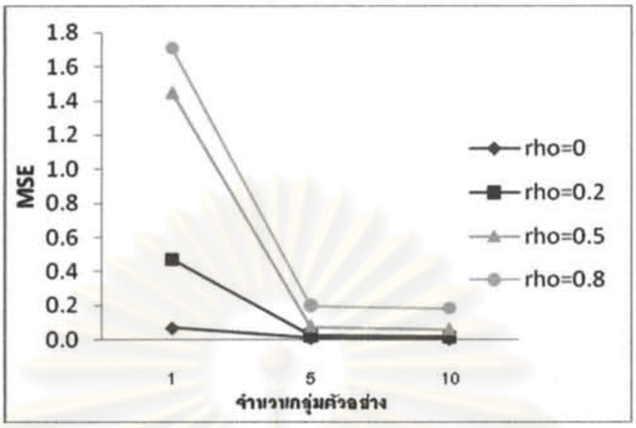
		$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0098		0.1160		0.4718		0.9965	
	Mean	-0.471	1.003	-0.492	0.993	-0.538	1.068	-0.674	1.321
	SD	0.108	0.090	0.460	0.149	0.892	0.385	1.213	0.628
5	MSE	0.0010		0.0205		0.0731		0.1960	
	Mean	-0.468	1.005	-0.428	0.911	-0.345	0.739	-0.223	0.487
	SD	0.025	0.038	0.161	0.075	0.214	0.132	0.198	0.171
10	MSE	0.0005		0.0148		0.0655		0.1948	
	Mean	-0.467	0.999	-0.420	0.898	-0.334	0.717	-0.212	0.460
	SD	0.017	0.025	0.120	0.052	0.1567	0.092	0.140	0.111

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีไม่ทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ เมื่อขนาดตัวอย่างเท่ากับ 1000 พบว่า เมื่อระดับความสัมพันธ์เพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง ซึ่งที่ระดับความสัมพันธ์ $\rho=0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$ ตามลำดับ โดยที่กลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมาก โดยเฉพาะในระดับความสัมพันธ์ $\rho=0.5$ และ $\rho=0.8$ ส่วนที่กลุ่มตัวอย่างที่เหลือมีแนวโน้มเพิ่มขึ้นค่อนข้างน้อยในแต่ละระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ นอกจากนี้ยังพบว่าในขนาดตัวอย่างในกรณีย่อยนี้ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าขนาดตัวอย่างในกรณีย่อยที่กล่าวมาแล้วข้างต้น ในแต่ละระดับความสัมพันธ์ และในแต่ละกลุ่มตัวอย่างในการทดลอง

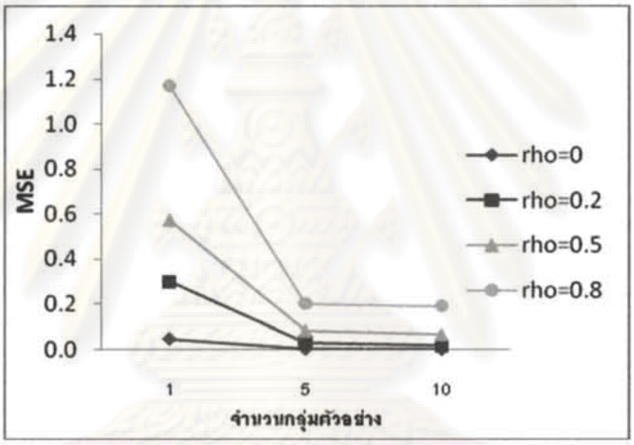
เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลอง 100 รอบ ประกอบการตัดสินใจพบว่า สำหรับจำนวนกลุ่มตัวอย่าง 1 กลุ่ม มีค่าประมาณของสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ สำหรับจำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 10 กลุ่มนั้น ค่าประมาณของสัมประสิทธิ์การถดถอย b_0 สูงกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 ต่ำกว่าค่าพารามิเตอร์ โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น สำหรับส่วนเบี่ยงเบนมาตรฐานในการทดลอง 100 รอบ พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากในระดับความสัมพันธ์สูง และในจำนวนกลุ่มตัวอย่างที่น้อย

จากตารางการวิเคราะห์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยใน 3 กรณีย่อยเมื่อเปลี่ยนขนาดตัวอย่างเป็น 100, 500 และ 1000 ข้างต้นนั้น สามารถพิจารณาจากรูปภาพแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$ จำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม และขนาดตัวอย่างในแต่ละกลุ่มทดลองเป็น 100, 500 และ 1000 ดังรูปที่ 4.2.1 – 4.2.3 ดังต่อไปนี้

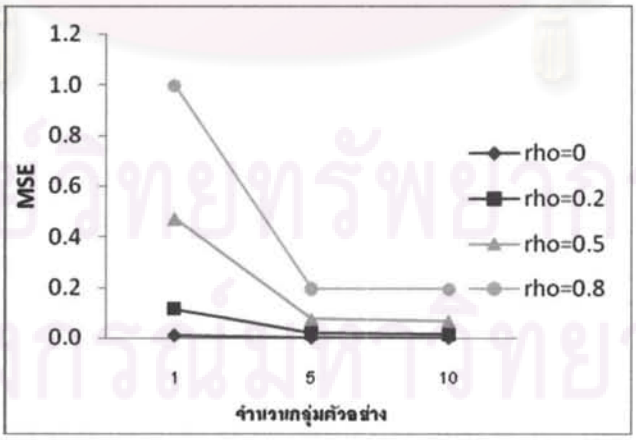
ก) กรณีจำนวนขนาดตัวอย่างเท่ากับ 100



ก) กรณีจำนวนขนาดตัวอย่างเท่ากับ 500

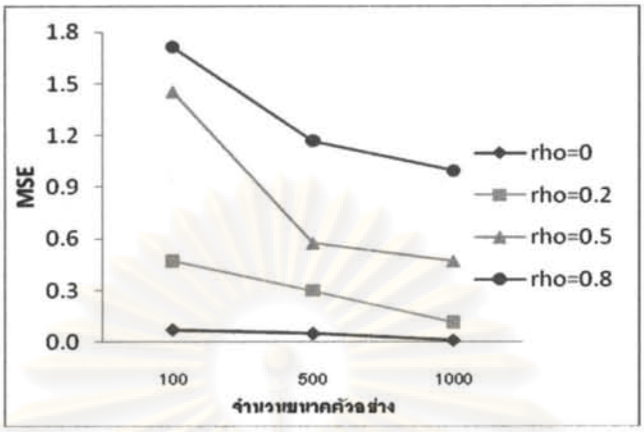


ข) กรณีจำนวนขนาดตัวอย่างเท่ากับ 1000

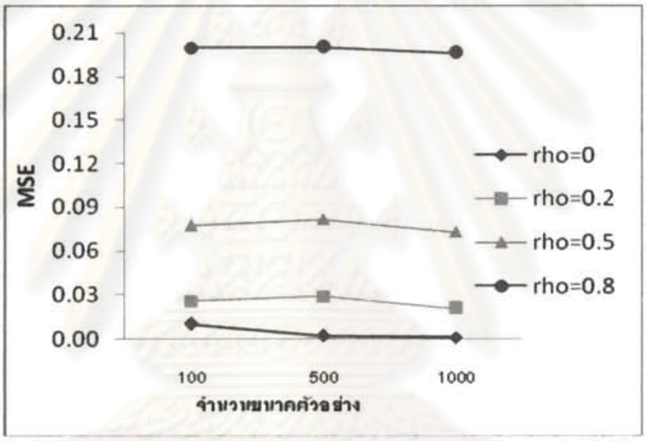


รูปที่ 4.2.1 แสดงการเปรียบค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ในแต่ละขนาดตัวอย่าง

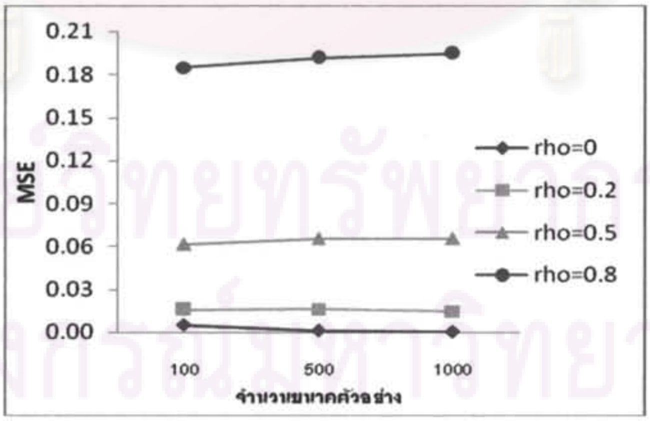
ก) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 1



ข) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 5

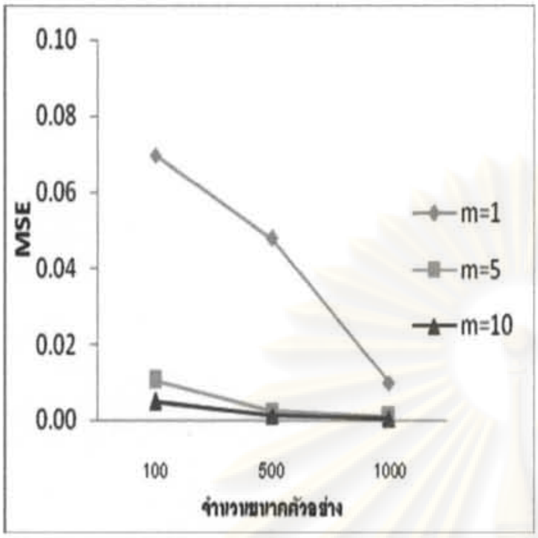


ค) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 10

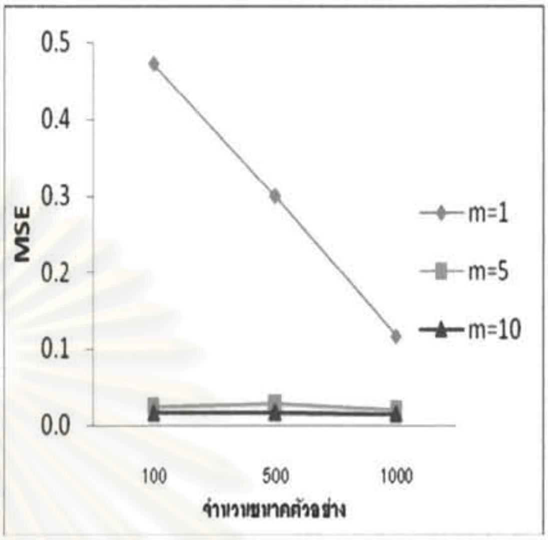


รูปที่ 4.2.2 แสดงการเปรียบค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนขนาดตัวอย่างเท่ากับ 100, 500 และ 1000 ในแต่ละกลุ่มตัวอย่าง

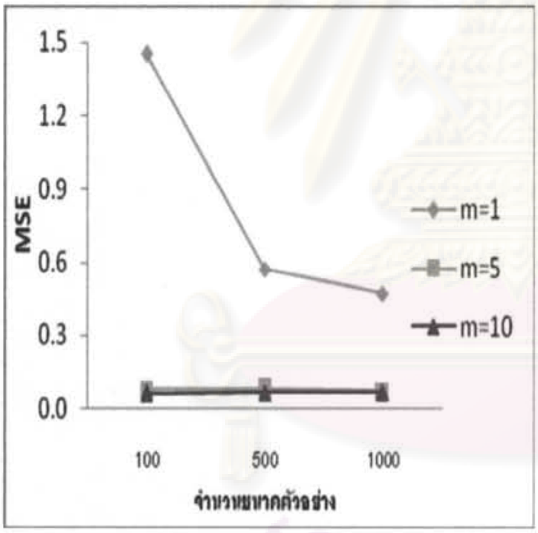
ก) $\rho = 0$



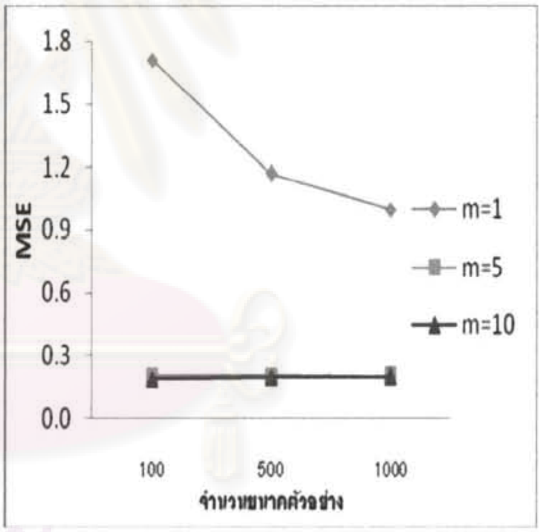
ข) $\rho = 0.2$



ค) $\rho = 0.5$



ง) $\rho = 0.8$



รูปที่ 4.2.3 แสดงการเปรียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่จำนวนขนาดตัวอย่างเท่ากับ 100, 500 และ 1000 และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ในแต่ละระดับความสัมพันธ์

จากรูป 4.2.1 – 4.2.3 สามารถสรุปผลการวิจัยเกี่ยวกับการเปลี่ยนแปลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) สำหรับในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ดังนี้

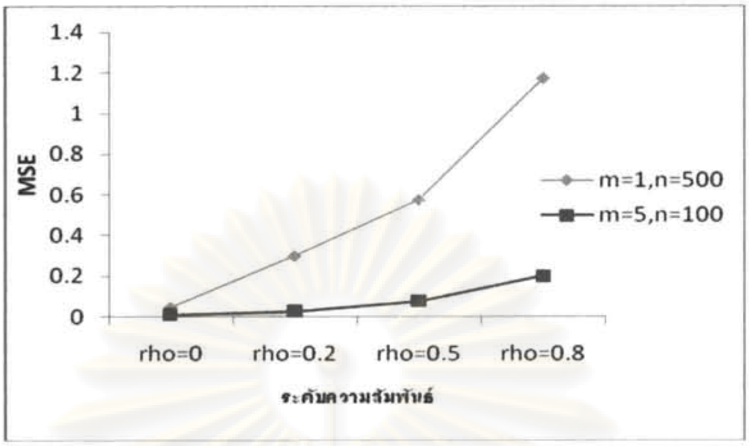
1. <u>เมื่อระดับความสัมพันธ์เพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นในทุกกลุ่มตัวอย่าง โดยเฉพาะที่ระดับความสัมพันธ์ $\rho = 0.8$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองค่อนข้างมาก ซึ่งสังเกตได้ชัดเจนในกลุ่มตัวอย่างเท่ากับ 1 กลุ่ม ในรูปที่ 4.1.1 และรูปที่ 4.1.2 นอกจากนี้พบว่าในทุกกลุ่มตัวอย่างและในทุกขนาดตัวอย่าง ที่ระดับความสัมพันธ์ $\rho = 0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ ตามลำดับ

2. <u>เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง โดยมีแนวโน้มลดลงอย่างรวดเร็วในช่วงแรกที่มีจำนวนกลุ่มตัวอย่างน้อย และค่อยๆ ลดลงเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น และในทุกระดับความสัมพันธ์พบว่า จำนวนกลุ่มตัวอย่างเท่ากับ 10 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ จำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 1 กลุ่มตามลำดับ เนื่องจากจำนวนกลุ่มตัวอย่างที่เพิ่มขึ้นนี้จะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยให้น้อยลง

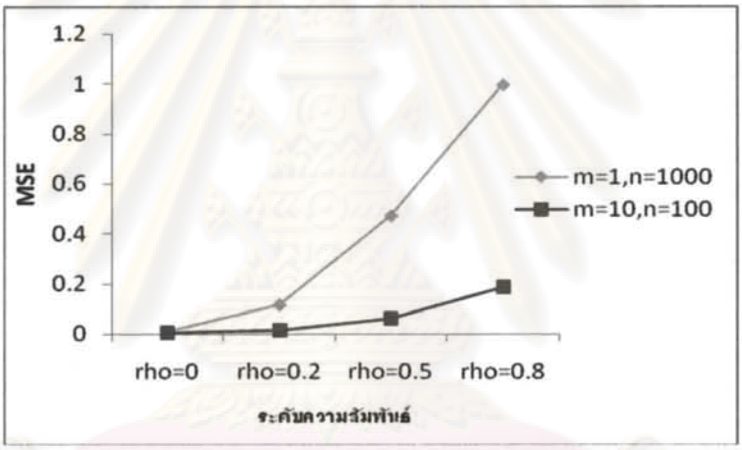
3. <u>เมื่อขนาดตัวอย่างเพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์และทุกกลุ่มตัวอย่าง โดยมีแนวโน้มในการลดลงค่อนข้างมากในช่วงที่มีจำนวนขนาดตัวอย่างน้อย และค่อยๆ ลดลงเมื่อจำนวนขนาดตัวอย่างเพิ่มมากขึ้น โดยเฉพาะจำนวนขนาดตัวอย่างเท่ากับ 1000 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด

จากในกรณีย่อยทั้ง 3 กรณีข้างต้นนั้น เป็นการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละกลุ่มและแต่ละขนาดทดลอง เช่น ในกรณีย่อย 4.2.1 เป็นกรณีที่ขนาดตัวอย่าง $n = 100$ โดยที่จำนวนกลุ่ม $m = 1,5$ และ 10 กลุ่ม เป็นการเปรียบเทียบในลักษณะจำนวนข้อมูลในแต่ละชุด (mn) ไม่เท่ากัน นั่นคือ $mn = 100$, 500 และ 1000 ซึ่งการเปรียบเทียบในลักษณะดังกล่าวอาจยังดูไม่ชัดเจน ทำให้เกิดข้อสงสัยในผลการเปรียบเทียบจากกรณีที่จำนวนข้อมูลทั้งหมดที่ไม่เท่ากันได้ ดังนั้นในกรณีดังต่อไปนี้ เป็นการแสดงภาพและอธิบายการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ สำหรับกรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) ที่เท่ากันของแต่ละกรณีย่อย ซึ่งมีค่าเท่ากับ 500, 1000 และ 5000 ดังรูปที่ 4.2.4 ดังต่อไปนี้

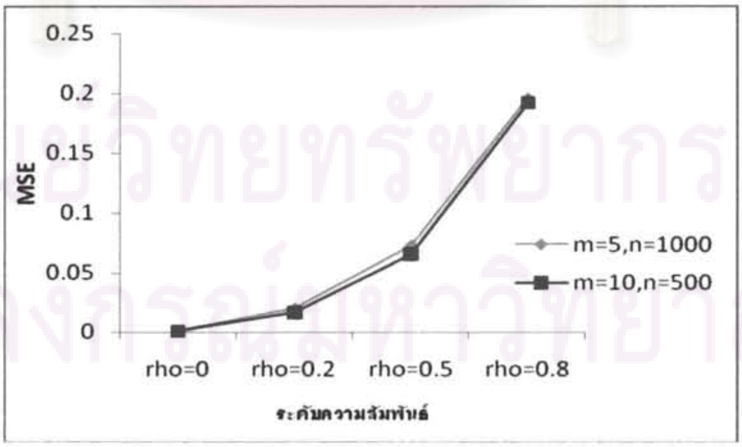
ก) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 500



ข) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 1000



ค) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 5000



รูปที่ 4.2.4 แสดงการเปรียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 500, 1000 และ 5000 ในแต่ละระดับความสัมพันธ์

จากรูป 4.2.4 ซึ่งแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีจำนวนข้อมูลตัวอย่างทั้งหมดในแต่ละชุด (mn) ที่เท่ากัน ซึ่งได้แก่จำนวนข้อมูล 500, 1000 และ 5000 ในกรณีข้างต้นนั้น พบว่า กลุ่มตัวอย่าง (m) ที่เพิ่มขึ้นส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลงโดยระดับการลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีที่กลุ่มตัวอย่าง 1 กลุ่มกับกลุ่มตัวอย่าง 5 กลุ่ม หรือ 10 กลุ่ม (ดังในกรณี ก, ข) มีอัตราการลดลงที่แตกต่างกันค่อนข้างมาก ส่วนในกรณีกลุ่มตัวอย่าง 5 กลุ่มกับ 10 กลุ่ม (ดังในกรณี ค) ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีอัตราการลดลงที่ไม่แตกต่างกันมากนัก และค่อนข้างมีค่าที่ใกล้เคียงกันมาก นั่นคือ จำนวนกลุ่มตัวอย่างมีบทบาทสำคัญต่อการลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมากกว่าจำนวนขนาดตัวอย่าง (n) ที่ไม่ส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลง ทั้งที่ขนาดตัวอย่างเพิ่มขึ้น นอกจากนี้จะเห็นว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นตามระดับความสัมพันธ์ที่สูงขึ้นทุกชุดของข้อมูล ซึ่งแสดงให้เห็นว่าการประมาณค่าสัมประสิทธิ์โดยใช้การวิเคราะห์แบบโพรบิทนี้ยังไม่เหมาะสมสำหรับในกรณีไม่ทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

สรุปผลการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

สำหรับทุกขนาดตัวอย่าง พบว่า จำนวนกลุ่มตัวอย่าง 1 กลุ่ม ให้ค่าประมาณของสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ (Overestimate) สำหรับจำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 10 กลุ่มนั้น ค่าประมาณของสัมประสิทธิ์การถดถอย b_0 สูงกว่าค่าพารามิเตอร์ (Overestimate) ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น อีกทั้งการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยในการทดลอง 100 รอบ ค่อนข้างมากในทุกจำนวนกลุ่มตัวอย่าง ประกอบกับการพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสำหรับกรณีศึกษานี้ พบว่า ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุดที่ระดับความสัมพันธ์ $\rho = 0$ จำนวนกลุ่มตัวอย่างเท่ากับ 10 และจำนวนขนาดตัวอย่างเท่ากับ 1000 และสังเกตได้ว่าในกรณีศึกษานี้ ได้ค่าประมาณสัมประสิทธิ์การถดถอยต่ำกว่าในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z เนื่องจากในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ นี้ เป็นการปรับค่าตัวประมาณจากกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z อีกที ดังนั้นในการประมาณค่าสัมประสิทธิ์การถดถอยโดยอาศัยแบบจำลองโพรบิทอาจยังไม่เหมาะสมสำหรับกรณีนี้

<u>กรณีศึกษาที่ 4.3</u> ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีทราบค่าปัจจัยของคอพพูลา Z

การวิจัยในกรณีนี้ ได้ทำการศึกษากรณีที่จำนวนขนาดตัวอย่างตัวอย่างในการทดลองเป็น 100, 500 และ 1000 โดยในแต่ละขนาดตัวอย่าง ได้ทำการศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนกลุ่มตัวอย่างในการทดลองเป็น 1, 5 และ 10 กลุ่ม ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.3.1 – 4.3.3 ส่วนในกรณี 4.3.4 เป็นการนำเสนอรูปภาพแสดงกรณีจำนวนตัวอย่างทั้งหมดในแต่ละชุด (mn) ที่เท่ากันในแต่ละกรณีย่อย ซึ่งเท่ากับ 500, 1000 และ 5000 โดยกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$

ตารางที่ 4.3.1 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 100

		$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0697		0.6131		3.2551		11.4260	
	Mean	-0.481	0.986	-0.578	1.212	-1.007	1.854	-2.031	3.344
	SD	0.174	0.332	0.874	0.646	1.965	1.299	3.393	1.888
5	MSE	0.0109		0.0245		0.1428		1.1036	
	Mean	-0.489	1.021	-0.541	1.139	-0.684	1.444	-1.108	2.305
	SD	0.088	0.116	0.097	0.125	0.127	0.165	0.166	0.266
10	MSE	0.0050		0.0149		0.1223		1.0606	
	Mean	-0.476	1.007	-0.532	1.124	-0.672	1.429	-1.088	2.299
	SD	0.054	0.085	0.056	0.087	0.080	0.117	0.120	0.195

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีทราบค่าปัจจัยคอพพูลา Z เมื่อขนาดตัวอย่างเท่ากับ 100 พบว่า เมื่อระดับความสัมพันธ์เพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง ซึ่งที่ระดับความสัมพันธ์ $\rho = 0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด สำหรับกลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความ

คลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมากตามระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น นอกจากนี้ยังพบว่าเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์

เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า ในทุกระดับความสัมพันธ์ และทุกกลุ่มตัวอย่างมีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ โดยเฉพาะในกลุ่มตัวอย่างเท่ากับ 1 กลุ่ม และในระดับความสัมพันธ์ที่สูงขึ้น สำหรับส่วนเบี่ยงเบนมาตรฐาน พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากในระดับความสัมพันธ์สูง และกลุ่มตัวอย่างน้อย และเมื่อจำนวนกลุ่มตัวอย่างมากขึ้นการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยลดลง

ตารางที่ 4.3.2 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 500

m		$\rho=0$		$\rho=0.2$		$\rho=0.5$		$\rho=0.8$	
		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0479		0.3943		1.3491		8.0767	
	Mean	-0.512	1.023	-0.630	1.170	-0.860	1.554	-1.590	3.036
	SD	0.265	0.155	0.779	0.367	1.369	0.623	2.928	1.513
5	MSE	0.0025		0.0121		0.1118		0.9385	
	Mean	-0.466	1.005	-0.524	1.125	-0.665	1.423	-1.043	2.240
	SD	0.039	0.060	0.043	0.063	0.052	0.066	0.066	0.085
10	MSE	0.0012		0.0102		0.1077		0.9373	
	Mean	-0.468	1.004	-0.525	1.122	-0.663	1.418	-1.048	2.238
	SD	0.027	0.040	0.029	0.040	0.034	0.044	0.050	0.075

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีทราบค่าปัจจัยคอพพูลา Z เมื่อขนาดตัวอย่างเท่ากับ 500 พบว่า เมื่อระดับความสัมพันธ์เพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง ซึ่งที่ระดับความสัมพันธ์ $\rho=0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด สำหรับกลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความ

คลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมากตามระดับความสัมพันธ์ที่สูงขึ้น เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น นอกจากนี้ยังพบว่าเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ โดยมีแนวโน้มการลดลงเร็วในกลุ่มตัวอย่างน้อยและลงลงอย่างช้าๆ เมื่อจำนวนกลุ่มตัวอย่างมากขึ้น

เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า ในทุกระดับความสัมพันธ์ และทุกกลุ่มตัวอย่างมีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ โดยเฉพาะในกลุ่มตัวอย่างเท่ากับ 1 กลุ่มและในระดับความสัมพันธ์ที่สูงให้ค่าเฉลี่ยของค่าประมาณสูงกว่าค่าพารามิเตอร์ค่อนข้างมาก สำหรับส่วนเบี่ยงเบนมาตรฐาน พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากในระดับความสัมพันธ์สูงและจำนวนกลุ่มตัวอย่างที่น้อย และเมื่อจำนวนกลุ่มตัวอย่างมากขึ้นการกระจายของข้อมูลลดลง

ตารางที่ 4.3.3 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 1000

m		$\rho=0$		$\rho=0.2$		$\rho=0.5$		$\rho=0.8$	
		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0098		0.1540		1.1070		7.0681	
	Mean	-0.471	1.003	-0.550	1.111	-0.761	1.511	-1.506	2.955
	SD	0.108	0.089	0.514	0.167	1.261	0.544	2.713	1.404
5	MSE	0.0011		0.0100		0.1088		0.9736	
	Mean	-0.469	1.005	-0.523	1.122	-0.662	1.422	-1.056	2.257
	SD	0.028	0.038	0.031	0.040	0.031	0.043	0.071	0.136
10	MSE	0.0005		0.0089		0.1058		0.9339	
	Mean	-0.467	0.999	-0.522	1.118	-0.662	1.416	-1.048	2.237
	SD	0.018	0.025	0.020	0.029	0.022	0.032	0.034	0.057

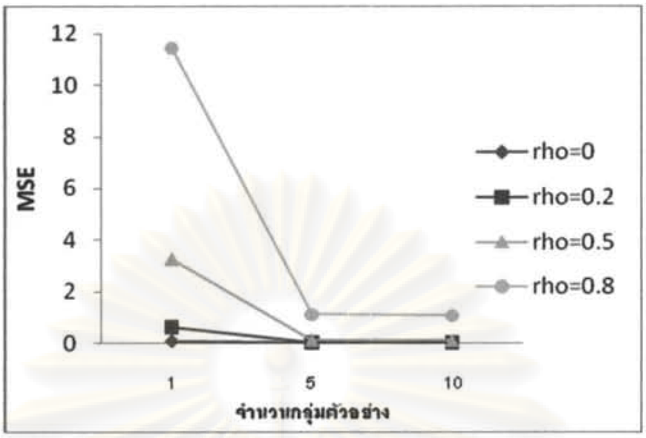
ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีทราบค่าปัจจัยคอพพูลา Z เมื่อขนาดตัวอย่างเท่ากับ 1000 พบว่า เมื่อระดับความสัมพันธ์เพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นทุกกลุ่มตัวอย่าง ซึ่งที่ระดับความสัมพันธ์ $\rho=0$ มี

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ ตามลำดับ โดยที่กลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมากตามระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มเพิ่มขึ้นค่อนข้างน้อยในแต่ละระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ โดยมีแนวโน้มการลดลงเร็วในกลุ่มตัวอย่างน้อย และลงลงอย่างช้าๆ เมื่อจำนวนกลุ่มตัวอย่างมากขึ้น นอกจากนี้ยังพบว่าในขนาดตัวอย่างในกรณีย่อยนี้ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่าขนาดตัวอย่างในกรณีย่อยที่กล่าวมาแล้วข้างต้น ในแต่ละระดับความสัมพันธ์ และในแต่ละกลุ่มตัวอย่างในการทดลอง

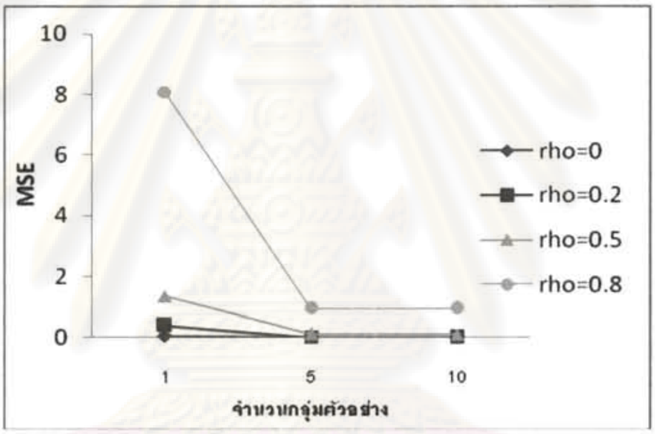
เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า ในทุกระดับความสัมพันธ์ และทุกกลุ่มตัวอย่างมีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ โดยเฉพาะในกลุ่มตัวอย่างเท่ากับ 1 กลุ่มและในระดับความสัมพันธ์ที่สูงขึ้นที่ให้ค่าเฉลี่ยของค่าประมาณสูงกว่าค่าพารามิเตอร์ค่อนข้างมาก สำหรับส่วนเบี่ยงเบนมาตรฐานในการทดลอง 100 รอบ พบว่าการกระจายของค่าประมาณของสัมประสิทธิ์การถดถอยค่อนข้างมากในระดับความสัมพันธ์สูง และในจำนวนกลุ่มตัวอย่างที่น้อย และเมื่อจำนวนกลุ่มตัวอย่างมากขึ้นการกระจายของข้อมูลลดลง

จากตารางการวิเคราะห์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยใน 3 กรณีย่อยเมื่อเปลี่ยนขนาดตัวอย่างเป็น 100, 500 และ 1000 ข้างต้นนั้น สามารถพิจารณาจากรูปภาพแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ดังรูปที่ 4.3.1 – 4.3.3 ดังต่อไปนี้

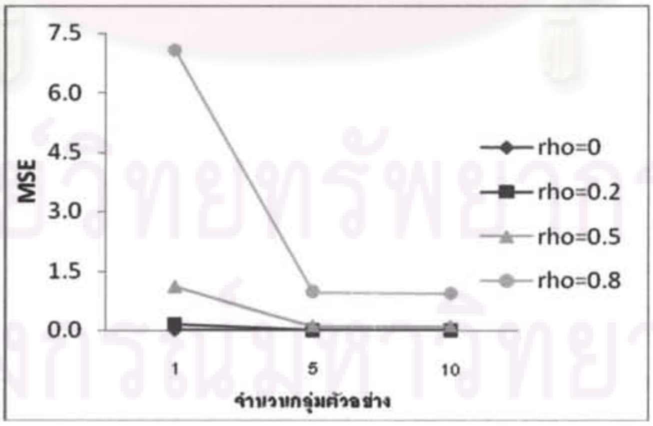
ก) กรณีจำนวนขนาดตัวอย่างเท่ากับ 100



ข) กรณีจำนวนขนาดตัวอย่างเท่ากับ 500

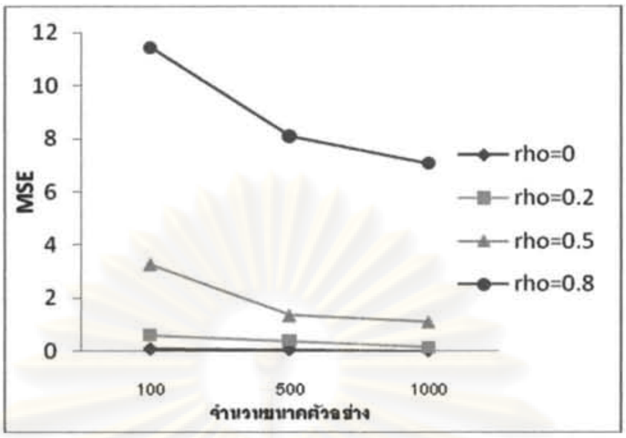


ค) กรณีจำนวนขนาดตัวอย่างเท่ากับ 1000

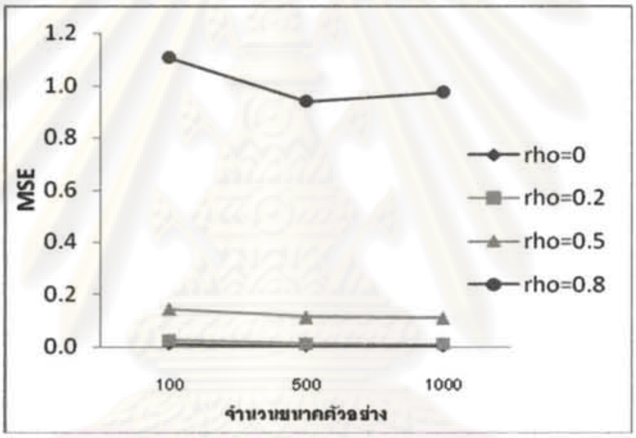


รูปที่ 4.3.1 แสดงการเปรียบค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ในแต่ละขนาดตัวอย่าง

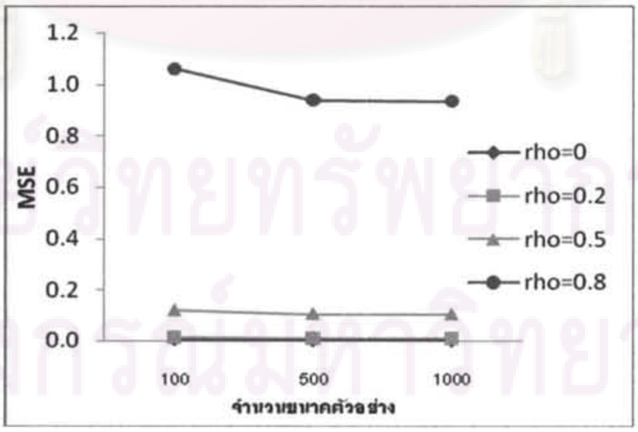
ก) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 1



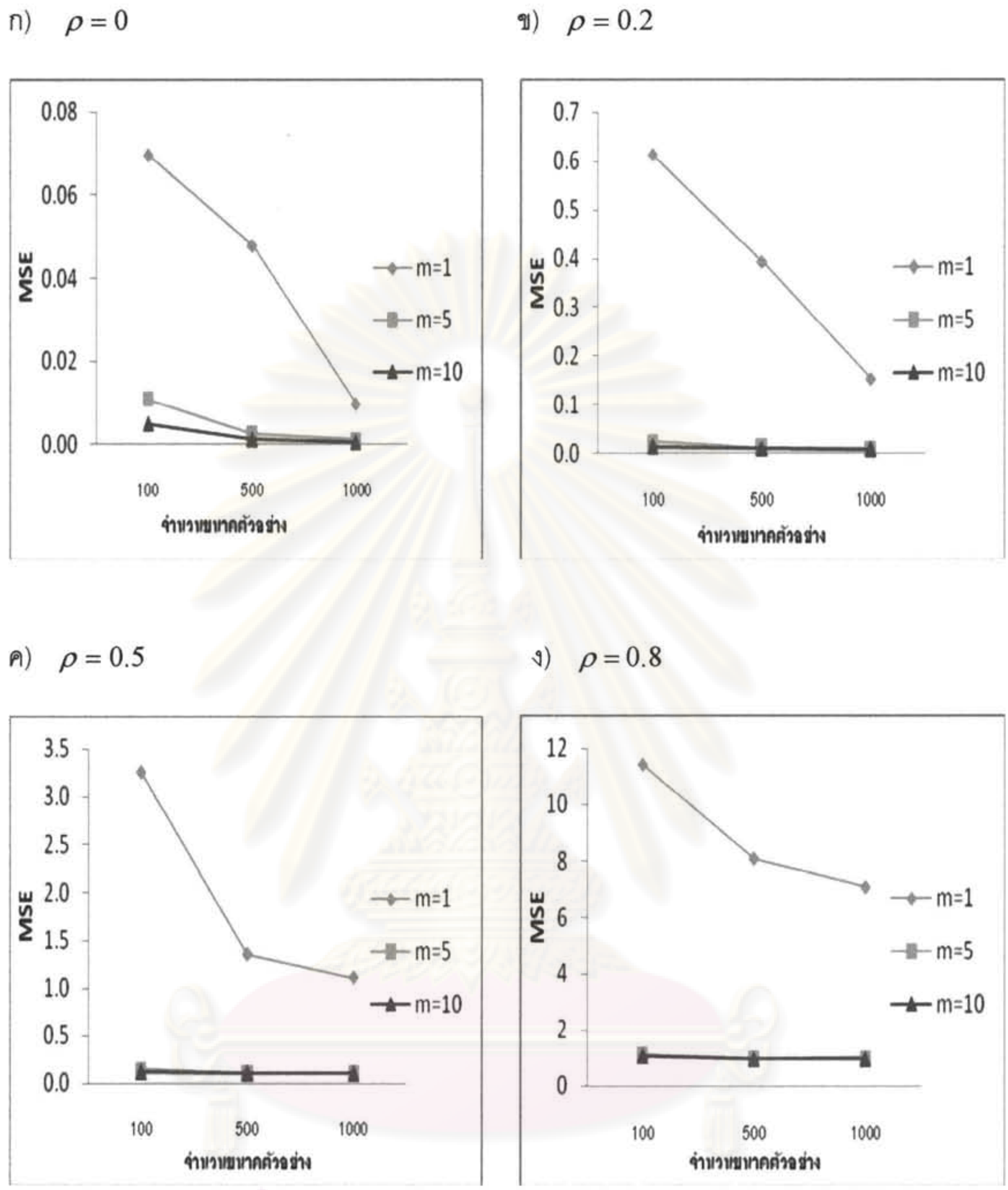
ข) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 5



ค) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 10



รูปที่ 4.3.2 แสดงการเปรียบค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนขนาดตัวอย่างเท่ากับ 100, 500 และ 1000 ในแต่ละกลุ่มตัวอย่าง



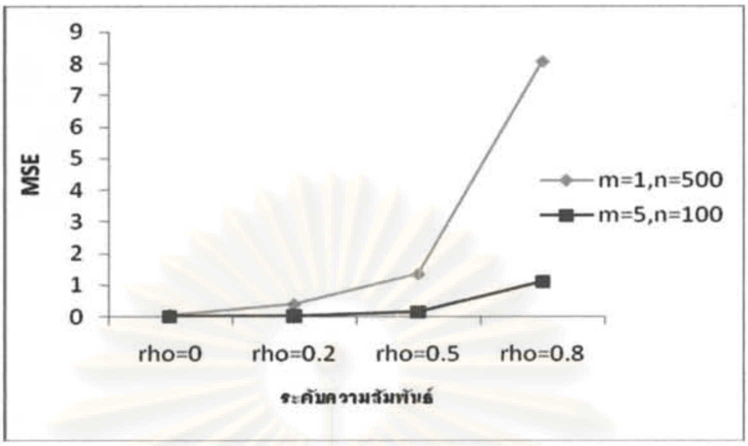
รูปที่ 4.3.3 แสดงการเปรียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่จำนวนขนาดตัวอย่างเท่ากับ 100, 500 และ 1000 และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ในแต่ละระดับความสัมพันธ์

จากรูป 4.3.1 – 4.3.3 สามารถสรุปผลการวิจัยเกี่ยวกับการเปลี่ยนแปลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) สำหรับในกรณีทราบค่าปัจจัยของคอพพูลา Z ดังนี้

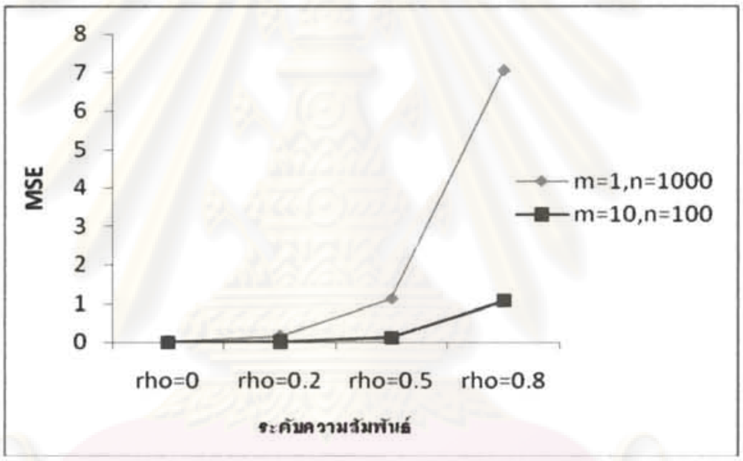
1. <u>เมื่อระดับความสัมพันธ์เพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นในทุกกลุ่มตัวอย่าง โดยเฉพาะที่ระดับความสัมพันธ์ $\rho = 0.5$ และ $\rho = 0.8$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นค่อนข้างมาก ซึ่งสังเกตได้ชัดในกลุ่มตัวอย่างเท่ากับ 1 กลุ่ม และสำหรับในทุกกลุ่มตัวอย่าง พบว่า ที่ระดับความสัมพันธ์ $\rho = 0$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ ตามลำดับ
2. <u>เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์และทุกขนาดตัวอย่าง โดยมีแนวโน้มลดลงอย่างรวดเร็วในช่วงแรกที่มีจำนวนกลุ่มตัวอย่างน้อย และค่อยๆ ลดลงเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น และในทุกระดับความสัมพันธ์พบว่า จำนวนกลุ่มตัวอย่างเท่ากับ 10 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ จำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 1 กลุ่มตามลำดับ เนื่องจากจำนวนกลุ่มตัวอย่างที่เพิ่มขึ้นนี้จะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยให้น้อยลง
3. <u>เมื่อขนาดตัวอย่างเพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์และทุกกลุ่มตัวอย่าง โดยมีแนวโน้มในการลดลงค่อนข้างมากในช่วงที่มีจำนวนขนาดตัวอย่างน้อย และค่อยๆ ลดลงเมื่อจำนวนขนาดตัวอย่างเพิ่มมากขึ้น โดยเฉพาะจำนวนขนาดตัวอย่างเท่ากับ 1000 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด

จากในกรณีย่อยทั้ง 3 กรณีข้างต้นนั้น เป็นการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละกลุ่มและแต่ละขนาดทดลอง เช่น ในกรณีย่อย 4.3.1 เป็นกรณีที่ขนาดตัวอย่าง $n = 100$ โดยที่จำนวนกลุ่ม $m = 1,5$ และ 10 กลุ่ม เป็นการเปรียบเทียบในลักษณะจำนวนข้อมูลในแต่ละชุด (mn) ไม่เท่ากัน นั่นคือ $mn = 100,\ 500$ และ 1000 ซึ่งการเปรียบเทียบในลักษณะดังกล่าวอาจยังดูไม่ชัดเจน ทำให้เกิดข้อสงสัยในผลการเปรียบเทียบจากกรณีที่จำนวนข้อมูลทั้งหมดที่ไม่เท่ากันได้ ดังนั้นในกรณีดังต่อไปนี้ เป็นการแสดงภาพและอธิบายการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ สำหรับกรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) ที่เท่ากันของแต่ละกรณีย่อย ซึ่งมีค่าเท่ากับ 500, 1000 และ 5000 ดังรูปที่ 4.3.4 ดังต่อไปนี้

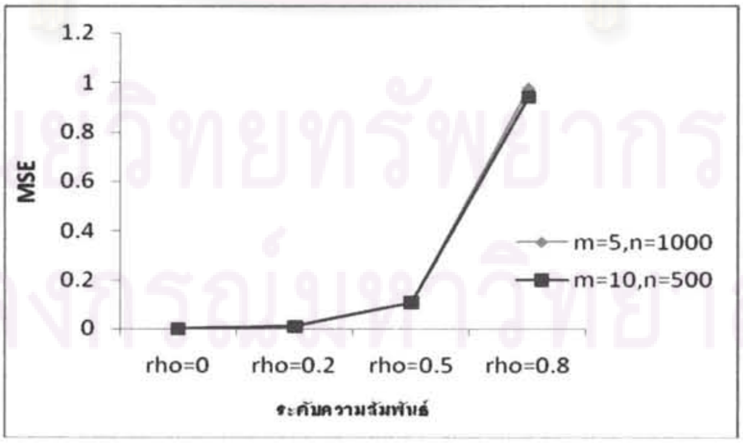
ก) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 500



ข) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 1000



ค) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 5000



รูปที่ 4.3.4 แสดงการเปรียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 500, 1000 และ 5000 ในแต่ละระดับความสัมพันธ์

จากรูป 4.3.4 ซึ่งแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีจำนวนข้อมูลตัวอย่างทั้งหมดในแต่ละชุด (mn) ที่เท่ากัน ซึ่งได้แก่จำนวนข้อมูล 500, 1000 และ 5000 ในกรณีข้างต้นนั้น พบว่า กลุ่มตัวอย่าง (m) ที่เพิ่มขึ้นส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลงโดยระดับการลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีที่กลุ่มตัวอย่าง 1 กลุ่มกับกลุ่มตัวอย่าง 5 กลุ่ม หรือ 10 กลุ่ม (ดังในกรณี ก, ข) มีอัตราการลดลงที่แตกต่างกันค่อนข้างมาก โดยค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกลุ่มตัวอย่าง 5 กลุ่ม หรือ 10 กลุ่มมีค่าลดลงค่อนข้างมากเมื่อเทียบกับกลุ่มตัวอย่าง 1 กลุ่ม ส่วนในกรณีกลุ่มตัวอย่าง 5 กลุ่มกับ 10 กลุ่ม (ดังในกรณี ค) ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีอัตราการลดลงที่ไม่แตกต่างกันมาก นั่นคือจำนวนกลุ่มตัวอย่างมีบทบาทสำคัญต่อการลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมากกว่าจำนวนขนาดตัวอย่าง (n) ที่ไม่ส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลง นอกจากนี้จะเห็นว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มขึ้นตามระดับความสัมพันธ์ที่สูงขึ้นทุกชุดของข้อมูลโดยเฉพาะในระดับความสัมพันธ์สูงๆ ซึ่งแสดงให้เห็นว่าการประมาณค่าสัมประสิทธิ์โดยใช้การวิเคราะห์แบบโพรบิทนี้ยังไม่เหมาะสมสำหรับในกรณีทราบค่าปัจจัยคอพพูลา Z

สรุปผลการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีทราบค่าปัจจัยของคอพพูลา Z

สำหรับทุกขนาดตัวอย่าง พบว่า ในระดับความสัมพันธ์ที่สูงขึ้นส่งผลให้ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ส่วนค่าประมาณของสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ (Overestimate) ในทุกกลุ่มการทดลองโดยเฉพาะเมื่อระดับความสัมพันธ์เพิ่มขึ้นส่งผลให้ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยที่ได้สูงขึ้นมาก อีกทั้งการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยในการทดลอง 100 รอบค่อนข้างมากในระดับความสัมพันธ์ที่สูง และในจำนวนกลุ่มตัวอย่างที่น้อย แต่เมื่อเพิ่มจำนวนกลุ่มตัวอย่างทำให้ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยและส่วนเบี่ยงเบนมาตรฐานมีค่าลดลงในทุกระดับความสัมพันธ์ และการพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสำหรับกรณีทราบค่าปัจจัยของคอพพูลา Z นี้ พบว่า ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าต่ำสุดที่ระดับความสัมพันธ์ $\rho = 0$ จำนวนกลุ่มตัวอย่างเท่ากับ 10 และจำนวนขนาดตัวอย่างเท่ากับ 1000 ประกอบกับค่าประมาณของสัมประสิทธิ์การถดถอยสำหรับกรณีนี้สูงกว่าค่าพารามิเตอร์ที่ควรจะเป็นค่อนข้างมาก โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น ทำให้การประมาณค่าโดยอาศัยแบบจำลองโพรบิทอาจยังไม่เหมาะสม ดังนั้นจึงควรมีการปรับค่าประมาณบางค่าเข้าไปเพื่อทำให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยมีความถูกต้องมากขึ้น ซึ่งจะแสดงไว้ในกรณีศึกษาต่อไปนี้

<u>กรณีศึกษาที่ 4.4</u> ผลการวิจัยเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

การวิจัยในกรณีนี้ ได้ทำการศึกษากรณีที่จำนวนขนาดตัวอย่างตัวอย่างในการทดลองเป็น 100, 500 และ 1000 โดยในแต่ละขนาดตัวอย่าง ได้ทำการศึกษาการประมาณค่าสัมประสิทธิ์การถดถอยที่ระดับความสัมพันธ์ $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$ และจำนวนกลุ่มตัวอย่างในการทดลองเป็น 1, 5 และ 10 กลุ่ม ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.4.1 – 4.4.3 สำหรับในกรณี 4.4.4 เป็นการนำเสนอรูปภาพแสดงกรณีจำนวนตัวอย่างทั้งหมดในแต่ละชุด (mn) ที่เท่ากันในแต่ละกรณีย่อย ซึ่งเท่ากับ 500, 1000 และ 5000 โดยกำหนดระดับความสัมพันธ์ในตัวแบบเป็น $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$

ตารางที่ 4.4.1 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 100

		$\rho=0$		$\rho=0.2$		$\rho=0.5$		$\rho=0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0697		0.4724		1.4509		1.7109	
	Mean	-0.481	0.986	-0.517	1.084	-0.712	1.311	-0.908	1.495
	SD	0.174	0.332	0.782	0.578	1.390	0.919	1.517	0.844
5	MSE	0.0108		0.0092		0.0092		0.0047	
	Mean	-0.487	1.012	-0.484	1.018	-0.481	1.015	-0.481	1.017
	SD	0.087	0.117	0.083	0.106	0.085	0.105	0.041	0.054
10	MSE	0.0050		0.0038		0.0041		0.0025	
	Mean	-0.476	1.007	-0.476	1.004	-0.474	1.008	-0.481	1.017
	SD	0.053	0.084	0.048	0.073	0.052	0.074	0.041	0.054

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ เมื่อขนาดตัวอย่างเท่ากับ 100 พบว่า สำหรับจำนวนกลุ่มตัวอย่าง 1 กลุ่ม ค่าเฉลี่ยความคลาด

เคลื่อนกำลังสองเพิ่มขึ้น ตามระดับความสัมพันธ์ที่สูงขึ้น โดยมีแนวโน้มในการเพิ่มขึ้นไม่มากนัก เมื่อเทียบกับในทุกกรณีก่อนหน้านี้ สำหรับจำนวนกลุ่มตัวอย่าง 5 และ 10 กลุ่ม จะเห็นว่าในแต่ละระดับความสัมพันธ์ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองใกล้เคียงกัน โดยที่ระดับความสัมพันธ์ $\rho = 0.2$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์

เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นส่งผลให้ในทุกระดับความสัมพันธ์มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย และส่วนเบี่ยงเบนมาตรฐานค่อนข้างใกล้เคียงกันมาก อีกทั้งค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยยังใกล้เคียงกับค่าพารามิเตอร์มากในทุกระดับความสัมพันธ์ และการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยค่อนข้างน้อย ยกเว้นกรณีจำนวนกลุ่มข้อมูล 1 กลุ่ม ที่ให้ค่าประมาณสัมประสิทธิ์การถดถอยสูงกว่าค่าพารามิเตอร์ โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น

ตารางที่ 4.4.2 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 500

		$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0479		0.2991		0.5742		1.1685	
	Mean	-0.512	1.023	-0.564	1.046	-0.608	1.099	-0.711	1.357
	SD	0.265	0.155	0.697	0.328	0.968	0.441	1.310	0.677
5	MSE	0.0025		0.0022		0.0017		0.0013	
	Mean	-0.466	1.004	-0.469	1.006	-0.471	1.007	-0.466	1.001
	SD	0.039	0.059	0.038	0.054	0.036	0.046	0.031	0.041
10	MSE	0.0012		0.0009		0.0007		0.0005	
	Mean	-0.468	1.004	-0.470	1.003	-0.470	1.004	-0.469	1.002
	SD	0.027	0.040	0.025	0.033	0.022	0.029	0.020	0.026

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

เมื่อขนาดตัวอย่างเท่ากับ 500 พบว่า สำหรับจำนวนกลุ่มตัวอย่าง 1 กลุ่ม ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้น ตามระดับความสัมพันธ์ที่สูงขึ้น โดยมีแนวโน้มในการเพิ่มขึ้นไม่มากนักเมื่อเทียบกับในทุกกรณีก่อนหน้านี้ สำหรับจำนวนกลุ่มตัวอย่าง 5 และ 10 กลุ่ม จะเห็นว่าในแต่ละระดับความสัมพันธ์ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองใกล้เคียงกัน และลดลงเล็กน้อยเมื่อระดับความสัมพันธ์สูงขึ้น โดยที่ระดับความสัมพันธ์ $\rho = 0.8$ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์

เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นส่งผลให้ในทุกระดับความสัมพันธ์มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย และส่วนเบี่ยงเบนมาตรฐานค่อนข้างใกล้เคียงกันมาก อีกทั้งค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยยังใกล้เคียงกับค่าพารามิเตอร์มากในทุกระดับความสัมพันธ์ และการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยค่อนข้างน้อย ยกเว้นกรณีจำนวนกลุ่มข้อมูล 1 กลุ่ม ที่ให้ค่าประมาณสัมประสิทธิ์การถดถอยสูงกว่าค่าพารามิเตอร์โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น

ตารางที่ 4.4.3 ตารางแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) เมื่อขนาดตัวอย่างในการทดลองเท่ากับ 1000

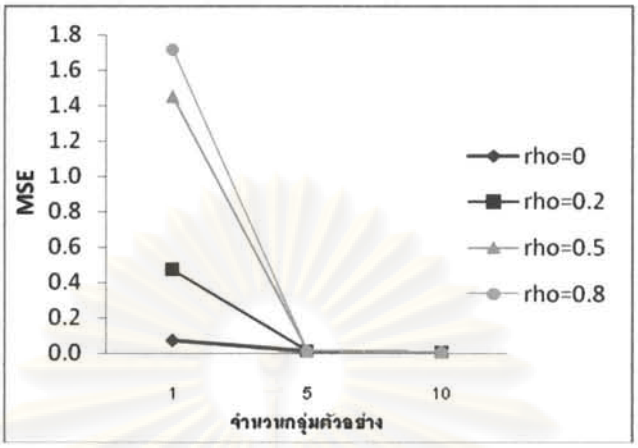
		$\rho = 0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
m		b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
1	MSE	0.0098		0.1160		0.4718		0.9965	
	Mean	-0.471	1.003	-0.492	0.993	-0.538	1.068	-0.674	1.321
	SD	0.108	0.090	0.460	0.149	0.892	0.385	1.213	0.629
5	MSE	0.0011		0.0011		0.0007		0.0007	
	Mean	-0.469	1.003	-0.467	1.003	-0.468	1.004	-0.469	1.004
	SD	0.028	0.038	0.028	0.037	0.024	0.029	0.020	0.025
10	MSE	0.0005		0.0005		0.0004		0.0003	
	Mean	-0.467	0.999	-0.467	0.999	-0.469	1.001	-0.469	1.001
	SD	0.018	0.025	0.019	0.026	0.016	0.022	0.012	0.018

ผลการวิจัยเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับกรณีทราบค่าปัจจัยคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ เมื่อขนาดตัวอย่างเท่ากับ 1000 พบว่า สำหรับจำนวนกลุ่มตัวอย่าง 1 กลุ่ม ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้น ตามระดับความสัมพันธ์ที่สูงขึ้น โดยมีแนวโน้มในการเพิ่มขึ้นไม่มากนักเมื่อเทียบกับในทุกกรณีก่อนหน้านี้ สำหรับจำนวนกลุ่มตัวอย่าง 5 และ 10 กลุ่ม จะเห็นว่าในแต่ละระดับความสัมพันธ์ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองใกล้เคียงกันมาก โดยจะเห็นได้ชัดในจำนวนกลุ่มตัวอย่างเท่ากับ 10 กลุ่ม และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลงเล็กน้อยในระดับความสัมพันธ์ที่สูงขึ้น และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์

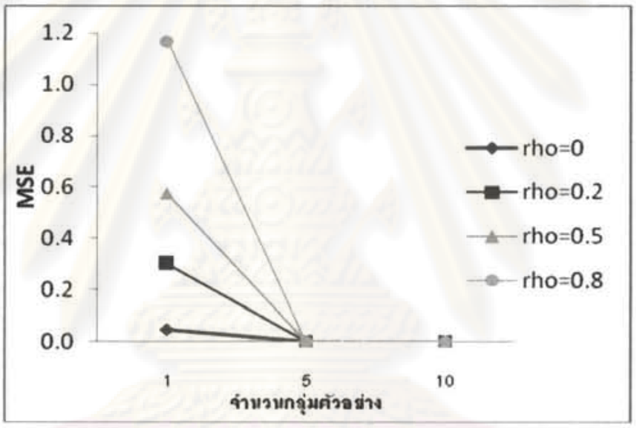
เมื่อพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ ประกอบการตัดสินใจพบว่า เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นส่งผลให้ในทุกระดับความสัมพันธ์มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย และส่วนเบี่ยงเบนมาตรฐานค่อนข้างใกล้เคียงกันมาก อีกทั้งค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยยังใกล้เคียงกับค่าพารามิเตอร์มากในทุกระดับความสัมพันธ์ โดยเฉพาะในจำนวนกลุ่มตัวอย่างที่มากขึ้น และการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยค่อนข้างน้อย ยกเว้นกรณีจำนวนกลุ่มข้อมูล 1 กลุ่ม ที่ให้ค่าประมาณสัมประสิทธิ์การถดถอยสูงกว่าค่าพารามิเตอร์ โดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น

จากตารางการวิเคราะห์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยใน 3 กรณีย่อยเมื่อเปลี่ยนขนาดตัวอย่างเป็น 100, 500 และ 1000 ข้างต้นนั้น สามารถพิจารณาจากรูปภาพแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ดังรูปที่ 4.4.1 – 4.4.3 ต่อไปนี้

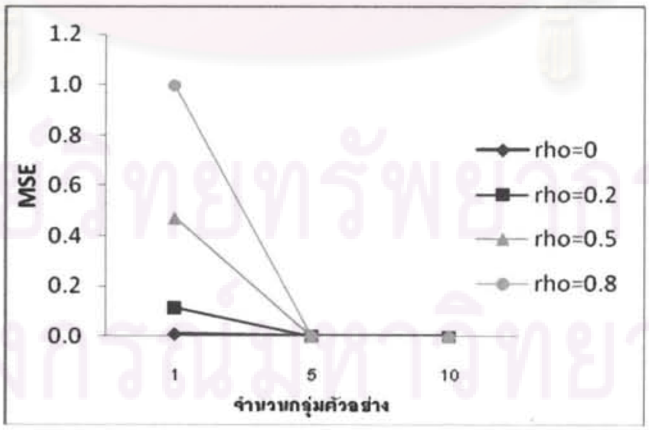
ก) กรณีจำนวนขนาดตัวอย่างเท่ากับ 100



ข) กรณีจำนวนขนาดตัวอย่างเท่ากับ 500

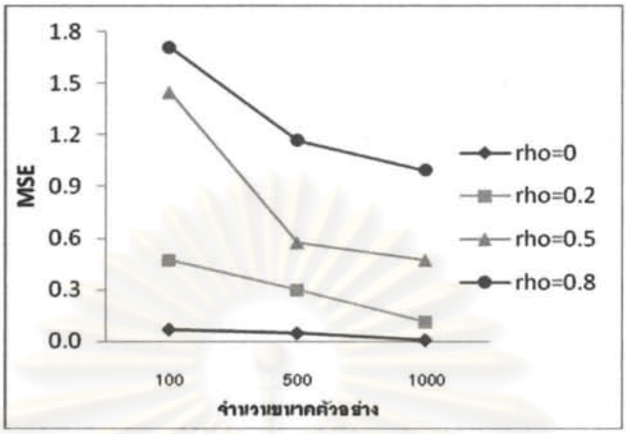


ค) กรณีจำนวนขนาดตัวอย่างเท่ากับ 1000

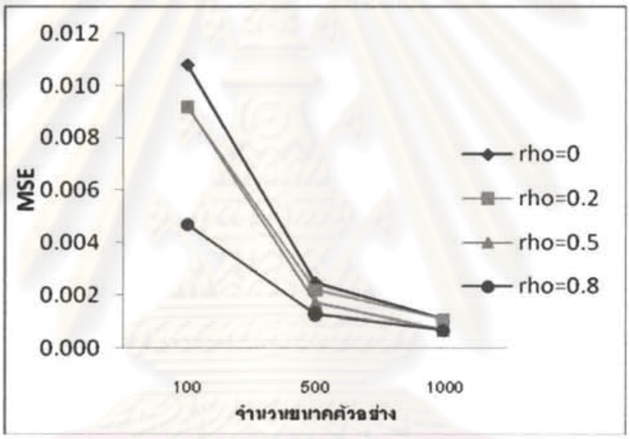


รูปที่ 4.4.1 แสดงการเปรียบค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ในแต่ละขนาดตัวอย่าง

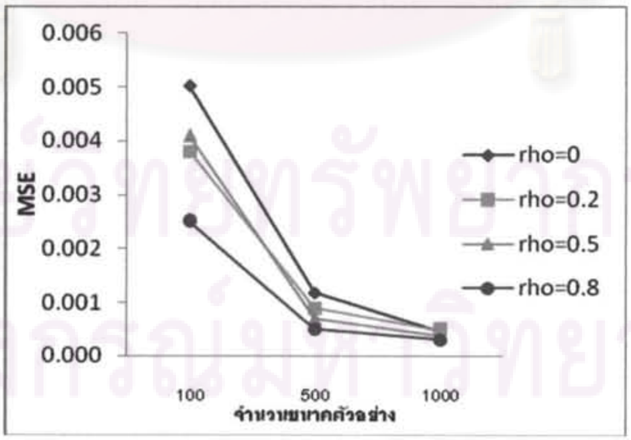
ก) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 1



ข) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 5

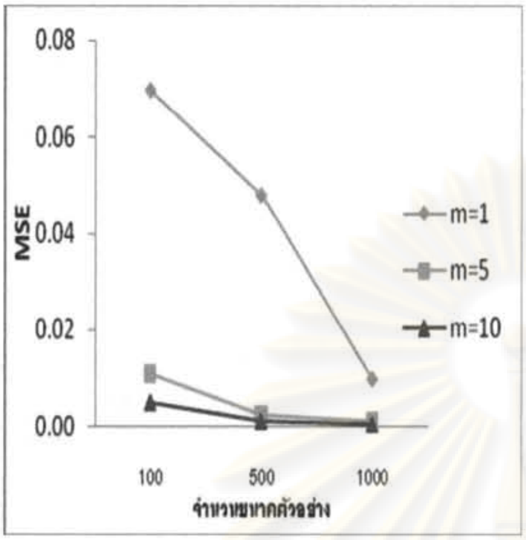


ค) กรณีจำนวนกลุ่มตัวอย่างเท่ากับ 10

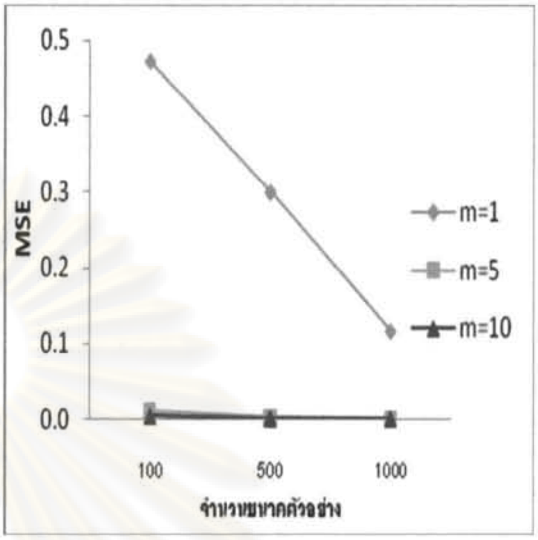


รูปที่ 4.4.2 แสดงการเปรียบค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$ และ $\rho = 0.8$ และจำนวนขนาดตัวอย่างเท่ากับ 100, 500 และ 1000 ในแต่ละกลุ่มตัวอย่าง

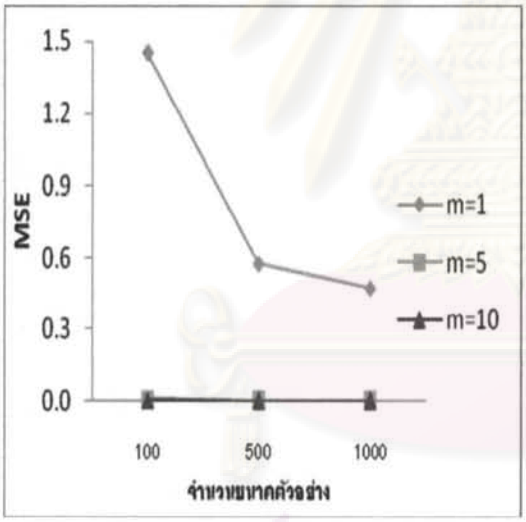
ข) $\rho = 0$



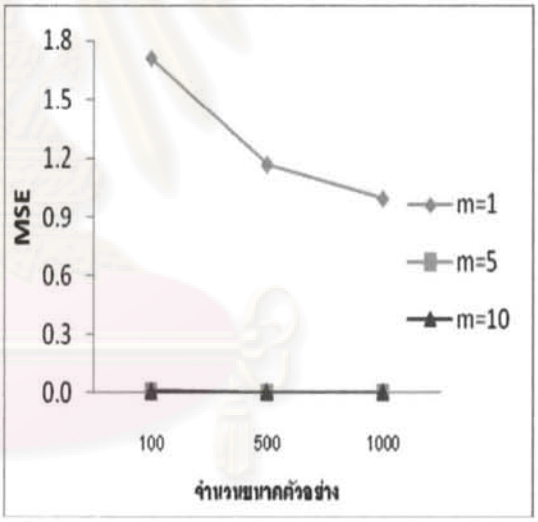
ข) $\rho = 0.2$



ค) $\rho = 0.5$



ง) $\rho = 0.8$



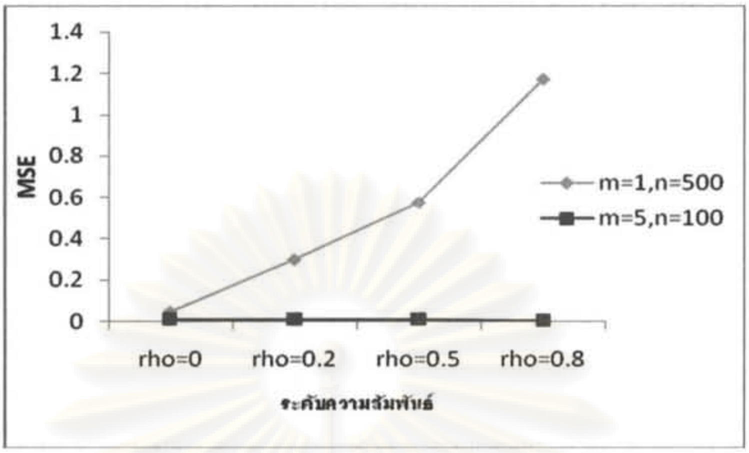
รูปที่ 4.4.3 แสดงการเปรียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่จำนวนขนาดตัวอย่างเท่ากับ 100, 500 และ 1000 และจำนวนกลุ่มตัวอย่างเท่ากับ 1, 5 และ 10 กลุ่ม ในแต่ละระดับความสัมพันธ์

จากรูป 4.4.1 – 4.4.3 สามารถสรุปผลการวิจัยเกี่ยวกับการเปลี่ยนแปลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) สำหรับในกรณีทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ดังนี้

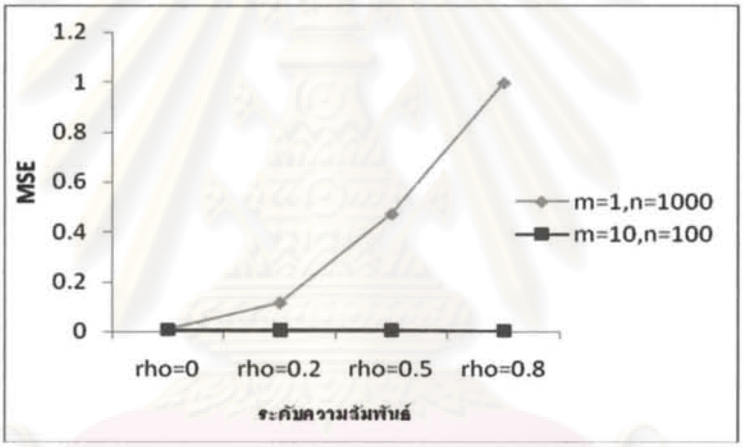
1. <u>เมื่อระดับความสัมพันธ์เพิ่มขึ้น</u> สำหรับจำนวนกลุ่มตัวอย่าง 1 กลุ่ม ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นตามระดับความสัมพันธ์ที่สูงขึ้น โดยมีแนวโน้มในการเพิ่มขึ้นไม่มากนักเมื่อเทียบกับในทุกกรณีก่อนหน้านี้ สำหรับจำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 10 กลุ่ม พบว่าในแต่ละระดับความสัมพันธ์ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองใกล้เคียงกันมาก และมีค่าลดลงเพียงเล็กน้อยในระดับความสัมพันธ์ที่สูงขึ้น
2. <u>เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ โดยมีแนวโน้มลดลงอย่างรวดเร็วในช่วงแรกที่มีจำนวนกลุ่มตัวอย่างน้อย และค่อยๆ ลดลงเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น และในทุกระดับความสัมพันธ์พบว่า จำนวนกลุ่มตัวอย่างเท่ากับ 10 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด รองลงมาคือ จำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 1 กลุ่มตามลำดับ เนื่องจากจำนวนกลุ่มตัวอย่างที่เพิ่มขึ้นนี้จะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยให้น้อยลง
3. <u>เมื่อขนาดตัวอย่างเพิ่มขึ้น</u> ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์และทุกกลุ่มตัวอย่าง โดยมีแนวโน้มในการลดลงค่อนข้างมากในช่วงที่มีจำนวนขนาดตัวอย่างน้อย และค่อยๆ ลดลงเมื่อจำนวนขนาดตัวอย่างเพิ่มมากขึ้น โดยเฉพาะจำนวนขนาดตัวอย่างเท่ากับ 1000 มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุด

จากในกรณีย่อยทั้ง 3 กรณีข้างต้นนั้น เป็นการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของแต่ละกลุ่มและแต่ละขนาดทดลอง เช่น ในกรณีย่อย 4.4.1 เป็นกรณีที่ขนาดตัวอย่าง $n=100$ โดยที่จำนวนกลุ่ม $m=1,5$ และ 10 กลุ่ม เป็นการเปรียบเทียบในลักษณะจำนวนข้อมูลในแต่ละชุด (mn) ไม่เท่ากัน นั่นคือ $mn=100$, 500 และ 1000 ซึ่งการเปรียบเทียบในลักษณะดังกล่าวอาจยังดูไม่ชัดเจน ทำให้เกิดข้อสงสัยในผลการเปรียบเทียบจากกรณีที่จำนวนข้อมูลทั้งหมดที่ไม่เท่ากันได้ ดังนั้นในกรณีดังต่อไปนี้ เป็นการแสดงภาพและอธิบายการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่ระดับความสัมพันธ์ $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$ สำหรับกรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) ที่เท่ากันของแต่ละกรณีย่อย ซึ่งมีค่าเท่ากับ 500, 1000 และ 5000 ดังรูปที่ 4.4.4 ดังต่อไปนี้

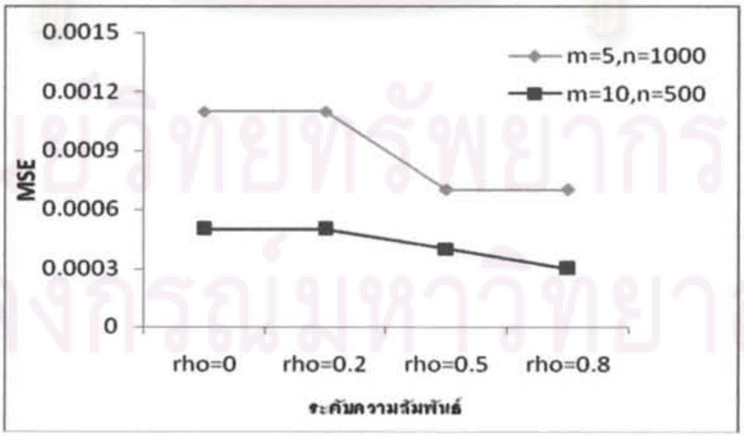
ก) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 500



ข) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 1000



ค) กรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 5000



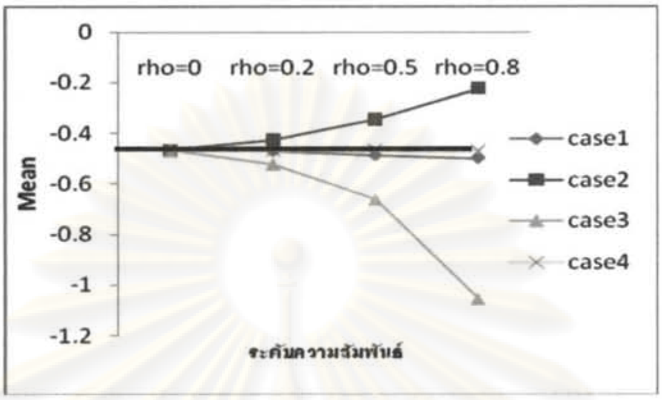
รูปที่ 4.4.4 แสดงการเปรียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีจำนวนข้อมูลทั้งหมดในแต่ละชุด (mn) เท่ากับ 500, 1000 และ 5000 ในแต่ละระดับความสัมพันธ์

จากรูป 4.4.4 ซึ่งแสดงการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีจำนวนข้อมูลตัวอย่างทั้งหมดในแต่ละชุด (mn) ที่เท่ากัน ซึ่งได้แก่จำนวนข้อมูล 500, 1000 และ 5000 ในกรณีข้างต้นนั้น พบว่า กลุ่มตัวอย่าง (m) ที่เพิ่มขึ้นส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลงโดยระดับการลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองโดยเห็นได้ชัดเจนจากกรณี ก) และ ข) พบว่ากลุ่มตัวอย่าง 1 กลุ่มค่าเฉลี่ยความคลาดเคลื่อนกำลังสองสูงขึ้นตามระดับความสัมพันธ์ที่สูงส่วนในกรณีกลุ่มตัวอย่าง 5 กลุ่มหรือ 10 กลุ่ม ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าใกล้เคียงกันมากในทุกระดับความสัมพันธ์ สำหรับในกรณี ค) พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าค่อนข้างใกล้เคียงกัน และเมื่อระดับความสัมพันธ์สูงขึ้นค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีแนวโน้มลดลงไม่มากนัก จากในกรณีข้างต้นนี้จะเห็นว่าในกลุ่มตัวอย่างที่มากกว่า 1 กลุ่ม ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าค่อนข้างน้อย และมีค่าใกล้เคียงกันในทุกระดับความสัมพันธ์

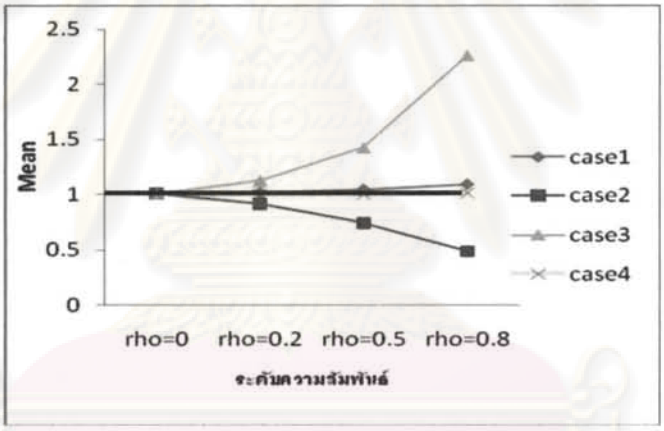
สรุปผลการประมาณค่าสัมประสิทธิ์การถดถอยในกรณีทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

สำหรับทุกขนาดตัวอย่าง พบว่า เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นส่งผลให้ในทุกระดับความสัมพันธ์มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย ค่อนข้างใกล้เคียงกันมาก อีกทั้งค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยยังใกล้เคียงกับค่าพารามิเตอร์มากในทุกระดับความสัมพันธ์ โดยมีการกระจายของข้อมูลค่าประมาณสัมประสิทธิ์การถดถอยค่อนข้างน้อย ยกเว้นกรณีจำนวนกลุ่มข้อมูล 1 กลุ่ม ที่ให้ค่าประมาณสัมประสิทธิ์การถดถอยที่สูงกว่าค่าพารามิเตอร์เล็กน้อยโดยเฉพาะในระดับความสัมพันธ์ที่สูงขึ้น เมื่อเทียบกับกรณีศึกษาก่อนหน้านี้ ประกอบกับการพิจารณาค่าเฉลี่ยความคลาดเคลื่อนกำลังสองที่มีค่าใกล้เคียงกันมากในแต่ละระดับความสัมพันธ์ที่มีจำนวนกลุ่มตัวอย่างและจำนวนขนาดตัวอย่างมาก นั่นคือ ในกรณีทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ส่งผลต่อความถูกต้องในการประมาณ โดยขึ้นอยู่กับจำนวนกลุ่มที่มากพอ ดังนั้น การประมาณค่าโดยอาศัยแบบจำลองโพรบิท ในกรณีทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์ความถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ มีความเหมาะสมสำหรับประมาณค่าในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาในกรณีที่กลุ่มตัวอย่างมากกว่า 1 กลุ่ม ซึ่งสอดคล้องกับทฤษฎีในหัวข้อ 2.3

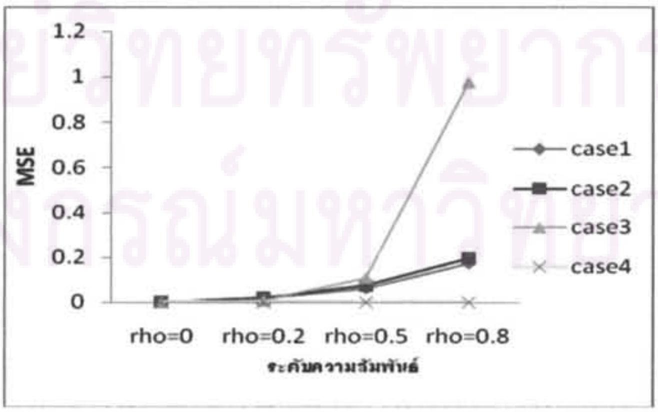
จากกรณีศึกษา 4.1 – 4.4 สามารถสรุปผลการประมาณค่าสัมประสิทธิ์การถดถอยที่เหมาะสมที่สุดในแต่ละกรณีศึกษา ซึ่งได้แก่จำนวนกลุ่มตัวอย่างเท่ากับ 10 และจำนวนขนาดตัวอย่างในแต่ละกลุ่มเท่ากับ 1000 ดังแสดงในรูปภาพต่อไปนี้



รูปที่ 4.5 แสดงค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์ $\hat{\beta}_0$ ที่ระดับความสัมพันธ์ต่างๆ ในทุกกรณีศึกษา



รูปที่ 4.6 แสดงค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์ $\hat{\beta}_1$ ที่ระดับความสัมพันธ์ต่างๆ ในทุกกรณีศึกษา



รูปที่ 4.7 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์ต่างๆ ในทุกกรณีศึกษา

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการศึกษาและประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา โดยอาศัยวิธีการวิเคราะห์ความถดถอยโพรบิทในการหาค่าตัวประมาณพารามิเตอร์ และได้ทำการจำลองสถานการณ์เพื่อทดสอบความถูกต้องในการประมาณค่าพารามิเตอร์ในกรณีที่ทราบค่าปัจจัยของคอพพูลา Z และทำการทดสอบเพิ่มเติมในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z โดยแบ่งเป็นกรณีศึกษาทั้งหมด 4 กรณี ดังต่อไปนี้

1. กรณีไม่ทราบค่าปัจจัยของคอพพูลา Z
2. กรณีไม่ทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$
3. กรณีทราบค่าปัจจัยของคอพพูลา Z
4. กรณีทราบค่าปัจจัยของคอพพูลา Z และปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$

โดยในงานวิจัยนี้ได้ทำการศึกษาในสถานการณ์ต่างๆ ดังนี้

- ตัวแบบโลจิสติคอย่างง่ายแบบ 2 กลุ่ม (Simple binary logistic model)
- ตัวแปรตาม (Y) มีการแจกแจงแบบคอพพูลาเบอร์นูลลี นั่นคือ ค่าสังเกตในแต่ละค่าของตัวแปรตามมีความสัมพันธ์กันด้วยเกาซ์เซียนคอพพูลา และมีค่าเพียง 2 ค่า คือ 0 กับ 1
- ตัวแปรอิสระ (X) ที่ศึกษาจำนวน 1 ตัวแปร โดยมีการแจกแจงแบบเบอร์นูลลี ด้วยพารามิเตอร์ $p=0.5$
- เกาซ์เซียนคอพพูลา ที่ระดับความสัมพันธ์ $\rho=0$, $\rho=0.2$, $\rho=0.5$ และ $\rho=0.8$
- จำนวนกลุ่มตัวอย่างในการทดลอง (m) เป็น 1, 5 และ 10 กลุ่ม
- จำนวนขนาดตัวอย่างในแต่ละกลุ่มตัวอย่าง (n) เป็น 100, 500 และ 1,000
- กำหนดจำนวนการกระทำซ้ำในแต่ละสถานการณ์เป็น 100 รอบ

เกณฑ์ในการเปรียบเทียบผลการวิเคราะห์ข้อมูลนั้น จะพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนมาตรฐาน (MSE) ประกอบกับการพิจารณาจากค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำจำนวน 100 รอบ ซึ่งสามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

ตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาเป็นตัวแบบที่ตัวแปรตามในแต่ละค่าสังเกตมีความสัมพันธ์กันด้วยเกาซ์เซียนคอพพูลา ซึ่งเป็นความสัมพันธ์ผ่านทางการแจกแจงร่วมของควอนไทล์ (Quantile) ของตัวแปรสุ่มร่วมที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal variables) โดยที่ตัวแปรตามเป็นตัวแปรเชิงคุณภาพ มีค่าเพียง 2 ค่า คือ 0 และ 1 และรูปแบบของสมการฟังก์ชันภาวะน่าจะเป็นของตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา ดังแสดงไว้ในสมการ (37) มีรูปแบบดังนี้

$$L = \int \int_{\min(y_i, 1-p_i)}^{\max(1-p_i, y_i)} \ldots\ldots \int C_{\Sigma}^{Ga}(u_1, u_2, \ldots, u_n)\, du_1 du_2 \ldots du_n$$

จากสมการฟังก์ชันภาวะน่าจะเป็นข้างต้น จะเห็นว่ามีรูปแบบฟังก์ชันที่ค่อนข้างซับซ้อน ซึ่งการวิเคราะห์ภาวะน่าจะเป็นสูงสุดทำได้ยาก อย่างไรก็ตาม หากทราบค่าปัจจัยคอพพูลา ตัวแบบที่ซับซ้อนจะลดรูปเป็นตัวแบบโพรบิทดังที่แสดงไว้ในสมการ (41) ดังนั้นจึงได้อาศัยแบบจำลองโพรบิทมาใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับตัวแบบนี้ นอกจากนี้ได้สังเกตว่าหากมีการละเลยปัจจัยคอพพูลานี้ ผลการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์การถดถอยจะมีลักษณะเป็นเช่นไร จึงได้ทำการจำลองสถานการณ์เพื่อทดสอบความถูกต้องในการประมาณค่าพารามิเตอร์ในกรณีที่ทราบค่าปัจจัยของคอพพูลา Z และทำการทดสอบเพิ่มเติมในกรณีไม่ทราบค่าปัจจัยของคอพพูลา Z โดยแบ่งเป็นกรณีศึกษาทั้งหมด 4 กรณีดังกล่าวไว้ข้างต้น โดยได้สรุปการเปรียบเทียบและการอภิปรายผลการวิจัย เป็น 5 ส่วน ดังนี้

5.1.1 <u>สรุปและอภิปรายผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างกรณีไม่ทราบค่าปัจจัยคอพพูลา Z และกรณีทราบค่าปัจจัยคอพพูลา Z</u>

5.1.1.1 เมื่อไม่มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ (กรณีศึกษา1: กรณีศึกษา3)

- การพิจารณาจากค่าเฉลี่ยความคลาดกำลังสองระหว่างกรณีศึกษา 1 กับกรณีศึกษา 3 นั้น พบว่า เมื่อจำนวนกลุ่มตัวอย่างในการทดลอง 1 กลุ่ม ทั้งสองกรณีศึกษามีค่าเฉลี่ยความคลาดกำลังสองเท่ากัน แต่เมื่อจำนวนกลุ่มตัวอย่างในการทดลองเพิ่มขึ้นค่าเฉลี่ยความคลาดกำลังสองในกรณีศึกษา 1 มีค่าต่ำกว่าในกรณีศึกษา 3 โดยเฉพาะในระดับความสัมพันธ์ที่สูง
- การพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ พบว่า ค่าเฉลี่ยของตัวประมาณสัมประสิทธิ์การถดถอยทั้งสองกรณีศึกษาอยู่ในทิศทาง

เดียวกัน กล่าวคือ ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ส่วนค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ (Overestimate) และเมื่อจำนวนข้อมูลมากขึ้น กรณีศึกษา 1 มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยใกล้เคียงกับพารามิเตอร์มากกว่าในกรณีศึกษา 3 สำหรับค่าส่วนเบี่ยงเบนมาตรฐานนั้น พบว่าในกรณีศึกษา 3 มีค่าส่วนเบี่ยงเบนมาตรฐานน้อยกว่ากรณีศึกษา 1 ค่อนข้างมาก โดยเฉพาะในกรณีที่จำนวนข้อมูลที่มากและในระดับความสัมพันธ์ที่สูง

การอภิปรายผล จากผลการทดลองที่ได้ข้างต้นนั้น สามารถสรุปได้ว่าทั้งกรณีทราบค่าปัจจัยคอพพูลา Z และไม่ทราบค่าปัจจัยคอพพูลา Z โดยไม่ได้มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ นั้น ให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยในทิศทางเดียวกันดังกล่าวไว้ข้างต้น และเมื่อจำนวนข้อมูลเพิ่มขึ้นแม้ว่ากรณีไม่ทราบค่าปัจจัยคอพพูลา Z มีค่าประมาณสัมประสิทธิ์การถดถอยใกล้เคียงพารามิเตอร์แต่การกระจายของข้อมูลก็มาก ที่เป็นเช่นนี้อาจเนื่องมาจากกรณีที่ไม่ทราบค่าปัจจัยคอพพูลาก็มีลักษณะเดียวกันกับการวิเคราะห์ความถดถอยโลจิสติคแบบปกติ ซึ่งความสัมพันธ์ที่เกิดขึ้นไม่ได้เป็นปัจจัยที่นำมาพิจารณาจึงทำให้ความสัมพันธ์นี้เข้าไปอยู่ในเทอมของความคลาดเคลื่อน ซึ่งส่งผลต่อความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย ถึงแม้ว่าในกรณีที่จำนวนข้อมูลที่มากขึ้น มีค่าเฉลี่ยของค่าประมาณใกล้เคียงกับพารามิเตอร์ก็ตาม แต่การกระจายของข้อมูลนั้นค่อนข้างมาก นั่นคืออัตราการแกว่งของข้อมูลในกรณีนี้ค่อนข้างสูง สำหรับในกรณีที่ทราบค่าปัจจัยคอพพูลา Z ที่มีการดึงเอาปัจจัยอาจที่ซ่อนอยู่ในเทอมความคลาดเคลื่อนออกมาบ้างแล้ว ซึ่งเป็นการชดเชยความผิดพลาดบางส่วนจากกรณีไม่ทราบค่าปัจจัยคอพพูลาที่อยู่ในเทอมความคลาดเคลื่อน แต่ที่ยังให้ผลการประมาณที่ไม่ถูกต้องนั้น เป็นผลมาจากค่าประมาณสัมประสิทธิ์การถดถอยที่ได้นั้นยังมีความสัมพันธ์เกี่ยวข้องในตัวแบบอยู่ดังแสดงไว้ในสมการ (41) ซึ่งถ้าไม่มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ จะทำให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยที่ได้นั้นมีค่าต่ำหรือสูงกว่าพารามิเตอร์อย่างใดอย่างหนึ่ง โดยเห็นได้ชัดจากผลการทดลองในกรณีศึกษา 4.3 ในบทที่ 4

5.1.1.2 เมื่อมีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ (กรณีศึกษา2: กรณีศึกษา4)

- การพิจารณาจากค่าเฉลี่ยความคลาดกำลังสองระหว่างกรณีศึกษา 2 กับ กรณีศึกษา 4 นั้น พบว่า เมื่อจำนวนกลุ่มตัวอย่าง 1 กลุ่มทั้งสองกรณีศึกษามีค่าเฉลี่ยความคลาดกำลังสองเท่ากัน แต่เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดกำลังสองในกรณีศึกษา 4 มีค่าต่ำกว่าในกรณีศึกษา 2 โดยเฉพาะในระดับความสัมพันธ์ที่สูง
- การพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ พบว่าในกรณีศึกษา 2 มีค่าเฉลี่ยของค่าประมาณของสัมประสิทธิ์การถดถอย b_0 สูงกว่าค่าพารามิเตอร์ (Overestimate) ส่วนค่าเฉลี่ยของค่าประมาณของสัมประสิทธิ์การถดถอย b_1 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ในทุกกลุ่มตัวอย่าง แต่กรณีศึกษา 4 นั้น มีเพียงกลุ่มตัวอย่าง 1 กลุ่มเท่านั้น ที่มีค่าเฉลี่ยของค่าประมาณสูงกว่าพารามิเตอร์ และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น พบว่าผลการประมาณค่าสัมประสิทธิ์การถดถอยมีค่าใกล้เคียงค่าพารามิเตอร์ในทุกระดับความสัมพันธ์ประกอบกับการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยค่อนข้างน้อยด้วย

<u>*การอภิปรายผล*</u> จากผลการทดลองข้างต้นนั้น สามารถสรุปได้ว่ากรณีไม่ทราบค่าปัจจัยคอพพูลา Z ที่มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ นั้น ให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยที่เบี่ยงเบนจากพารามิเตอร์ค่อนข้างมากดังกล่าวไว้ข้างต้น ที่เป็นเช่นนี้อาจเนื่องมาจากในกรณีศึกษานี้ เป็นการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ จากกรณีที่ไม่ทราบค่าปัจจัยที่แฝงอยู่ในเทอมของความคลาดเคลื่อนและไม่ได้นำปัจจัยนั้นเข้ามาพิจารณาตั้งแต่แรก และเมื่อไปปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยที่ได้ โดยการคูณด้วย $\sqrt{1-\rho}$ เข้าไป จะทำให้ผลการประมาณสัมประสิทธิ์การถดถอยที่ได้ไม่ถูกต้อง แต่สำหรับในกรณีที่ทราบค่าปัจจัยคอพพูลา Z ที่มีการดึงเอาปัจจัยอาจที่ซ่อนอยู่ในเทอมความคลาดเคลื่อนออกมาบ้างแล้ว และได้มีการพิสูจน์ตัวแบบดังสมการ (41) ย่อมทำให้การปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ นั้น ส่งผลต่อความถูกต้องของการประมาณค่าสัมประสิทธิ์ความถดถอย โดยจำนวนกลุ่มตัวอย่างควรมากกว่า 1 กลุ่มด้วย

5.1.2 <u>สรุปและอภิปรายผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างกรณีที่ไม่มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยและมีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$</u>

5.1.2.1 เมื่อไม่ทราบค่าปัจจัยคอพพูลา Z (กรณีศึกษา1: กรณีศึกษา2)

- การพิจารณาจากค่าเฉลี่ยความคลาดกำลังสองระหว่างกรณีศึกษา 1 และกรณีศึกษา 2 นั้น พบว่า เมื่อจำนวนกลุ่มตัวอย่าง 1 กลุ่มกรณีศึกษา 1 มีค่าเฉลี่ยความคลาดกำลังสองสูงกว่ากรณีศึกษา 2 ค่อนข้างมากโดยเฉพาะในระดับความสัมพันธ์ที่สูง และเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นค่าเฉลี่ยความคลาดกำลังสองในกรณีศึกษา 1 สูงกว่ากรณีศึกษา 2 เพียงเล็กน้อยเท่านั้นในแต่ละระดับความสัมพันธ์
- การพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ พบว่าทั้งสองกรณีศึกษานั้นมีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์ความถดถอยในทิศทางตรงข้ามกัน กล่าวคือ ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 สูงกว่าค่าพารามิเตอร์ (Overestimate) ส่วนค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_1 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) และเมื่อจำนวนข้อมูลมากขึ้นกรณีศึกษา 1 มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์ความถดถอยใกล้เคียงกับพารามิเตอร์มากกว่าในกรณีศึกษา 2 สำหรับค่าส่วนเบี่ยงเบนมาตรฐานที่ได้นั้นพบว่าเมื่อจำนวนกลุ่มตัวอย่างในการทดลองมากขึ้นกรณีศึกษา 2 มีลักษณะการกระจายของข้อมูลน้อยกว่ากรณีศึกษา 1 ไม่มากนัก

  <u>การอภิปรายผล</u> จากผลการทดลองที่ได้ข้างต้นนั้น สามารถสรุปได้ว่าทั้งกรณีที่มีการปรับและไม่ได้มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ โดยที่ไม่ทราบค่าปัจจัยคอพพูลา Z นั้น ให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยเบี่ยงเบนจากพารามิเตอร์ค่อนข้างมากดังกล่าวไว้ข้างต้น แต่เมื่อจำนวนข้อมูลเพิ่มขึ้นกรณีที่ไม่ได้มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ มีผลการประมาณค่าสัมประสิทธิ์การถดถอยที่ค่อนข้างใกล้เคียงพารามิเตอร์แต่การกระจายของข้อมูลก็มากด้วย ที่เป็นเช่นนี้อาจเนื่องมาจากกรณีที่ไม่ทราบค่าปัจจัยคอพพูลาและไม่ได้มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ยังคงเป็นปัจจัยแฝงที่เราไม่ได้นำมาพิจารณาในการวิเคราะห์ความถดถอย ดังนั้นปัจจัยดังกล่าวนี้จึงส่งผลต่อความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอย ถึงแม้ว่าจะมีจำนวนข้อมูลมากขึ้น หรือมี

ค่าเฉลี่ยของค่าประมาณใกล้เคียงกับพารามิเตอร์ก็ตาม แต่การกระจายของข้อมูลนั้นค่อนข้างมาก สำหรับกรณีที่ไม่ทราบค่าปัจจัยคอพพูลาและมีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ก็เป็นผลที่ได้จากการนำกรณีแรกมาปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ซึ่งเป็นการช่วยปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยที่มีค่าสูงกว่าพารามิเตอร์ให้มีค่าลดลงใกล้เคียงกับพารามิเตอร์มากขึ้น และส่งผลให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงจากกรณีแรก แต่สำหรับในกรณีที่ตัวประมาณสัมประสิทธิ์การถดถอยมีค่าไม่แตกต่างกับพารามิเตอร์มากนักเมื่อจำนวนข้อมูลเพิ่มขึ้น การปรับค่าประมาณสัมประสิทธิ์การถดถอยตรงนี้ ยิ่งส่งผลให้ค่าประมาณสัมประสิทธิ์การถดถอยที่ได้ต่ำกว่าพารามิเตอร์มากขึ้น นั่นย่อมแสดงว่าการปรับค่าประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ในกรณีที่ไม่ทราบค่าปัจจัยคอพพูลา Z นั้นยังไม่เหมาะสม

5.1.2.2 *เมื่อทราบค่าปัจจัยคอพพูลา Z (กรณีศึกษา3: กรณีศึกษา4)*

- การพิจารณาจากค่าเฉลี่ยความคลาดกำลังสองระหว่างกรณีศึกษา 3 กับกรณีศึกษา 4 นั้น พบว่าค่าเฉลี่ยความคลาดกำลังสองในกรณีศึกษา 4 มีค่าต่ำกว่ากรณีศึกษา 3 ในทุกระดับสัมพันธ์และทุกกลุ่มข้อมูล
- การพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ พบว่ากรณีศึกษา 3 มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ส่วนค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ (Overestimate) ในทุกกลุ่มตัวอย่าง สำหรับกรณีศึกษา 4 นั้น มีเพียงกลุ่มตัวอย่าง 1 กลุ่มเท่านั้นที่มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยสูงกว่าพารามิเตอร์ เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น ผลการประมาณค่าสัมประสิทธิ์การถดถอยใกล้เคียงค่าพารามิเตอร์ในทุกระดับความสัมพันธ์ โดยที่ลักษณะการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยค่อนข้างน้อยด้วย

<u>*การอภิปรายผล*</u> จากผลการทดลองที่ได้ข้างต้นนั้น สามารถสรุปได้ว่ากรณีที่ทราบค่าปัจจัยคอพพูลา Z นั้น การปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ มีบทบาทสำคัญต่อความถูกต้องของการประมาณค่าสัมประสิทธิ์ความถดถอยดังแสดงในสมการ (41) เนื่องจากกรณีทราบค่าปัจจัยคอพพูลา Z แต่ไม่ได้มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ จะ

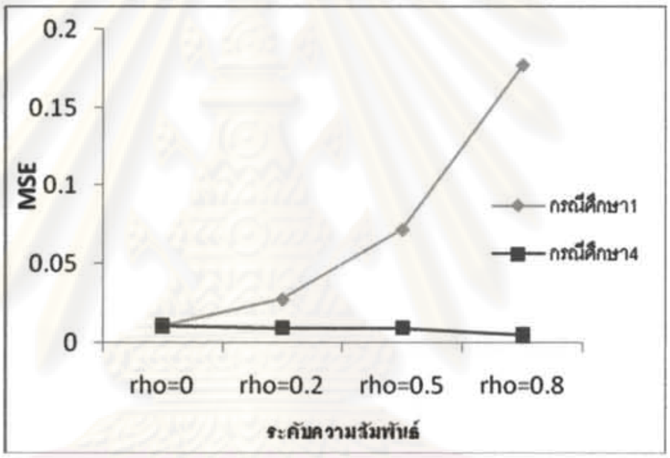
ส่งผลให้ตัวประมาณที่ได้นั้นยังมีความสัมพันธ์ปนอยู่ในค่าประมาณสัมประสิทธิ์ความถดถอยนั้น อาจส่งผลให้ตัวประมาณที่ได้ไม่มีคุณสมบัติเป็นตัวประมาณที่ดี ซึ่งส่งผลต่อการสรุปผลต่อไป แต่เมื่อมีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ ตามที่ได้พิสูจน์ในหัวข้อ 2.3 ย่อมทำให้ตัวประมาณที่ได้เป็นตัวประมาณมีคุณสมบัติที่ไม่เอนเอียงและมีความคงเส้นคงวา ซึ่งเป็นคุณสมบัติหนึ่งของตัวประมาณที่ดี โดยจะเห็นได้จากกรณีศึกษา 4 ที่ให้ผลการประมาณค่าสัมประสิทธิ์ความถดถอยใกล้เคียงค่าพารามิเตอร์ในทุกระดับความสัมพันธ์ โดยที่ค่าเฉลี่ยความคลาดกำลังสองมีค่าต่ำ

5.1.3 <u>สรุปและอภิปรายผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างกรณีไม่มีปัจจัยคอพพูลาและมีปัจจัยคอพพูลาเข้ามาเกี่ยวข้อง (กรณีศึกษา1: กรณีศึกษา4)</u>

- การพิจารณาค่าเฉลี่ยความคลาดกำลังสองระหว่างกรณีศึกษา 1 กับกรณีศึกษา 4 พบว่าค่าเฉลี่ยความคลาดกำลังสองในกรณีศึกษา 4 มีค่าต่ำกว่าในกรณีศึกษา 3 ในทุกระดับสัมพันธ์และทุกกลุ่มข้อมูล
- การพิจารณาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในการทดลองซ้ำ 100 รอบ พบว่าในกรณีศึกษา 1 มีค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ส่วนค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ (Overestimate) ในทุกกลุ่มตัวอย่าง แต่กรณีศึกษา 4 นั้นมีเพียงกลุ่มตัวอย่าง 1 กลุ่มเท่านั้นที่มีค่าเฉลี่ยของค่าประมาณสูงกว่าพารามิเตอร์ โดยที่ในจำนวนกลุ่มตัวอย่างที่เพิ่มขึ้น ผลการประมาณค่าสัมประสิทธิ์การถดถอยใกล้เคียงค่าพารามิเตอร์ในทุกระดับความสัมพันธ์ และลักษณะการกระจายของค่าประมาณสัมประสิทธิ์การถดถอยค่อนข้างน้อย

<u>*การอภิปรายผล*</u> จากผลการทดลองที่ได้ข้างต้นนั้น สามารถสรุปได้ว่าการที่มีปัจจัยคอพพูลาเข้ามาเกี่ยวข้อง ทำให้ผลการประมาณค่ามีความถูกต้องมากขึ้น เนื่องจากปัจจัยคอพพูลาที่ทราบค่านี้ ก็เปรียบเหมือนตัวแปรๆ หนึ่งเพิ่มเติม ซึ่งมีความสัมพันธ์หรือมีผลกระทบต่อตัวแปรที่สนใจศึกษา และส่งผลให้เทอมของความคลาดเคลื่อนมีค่าลดลง ทำให้ผลการวิเคราะห์ความถดถอยนี้มีความถูกต้องและมีประสิทธิภาพมากขึ้น แต่อย่างไรก็ตามปัจจัยคอพพูลานี้ถือเป็นปัจจัยที่กระทบต่อข้อมูลทั้งกลุ่ม ไม่ได้เป็นข้อมูลที่กระทบในแต่ละค่าสังเกต ดังนั้นยิ่งจำนวนกลุ่มข้อมูลมากขึ้น ย่อมส่งผลให้การประมาณค่าสัมประสิทธิ์ความถดถอยมีความถูกต้องมากขึ้นด้วย และการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย

$\sqrt{1-\rho}$ ก็เป็นอีกปัจจัยหนึ่งที่ส่งผลให้ตัวประมาณที่ได้มีคุณสมบัติเป็นตัวประมาณที่ดี สำหรับในการวิเคราะห์ความถดถอยโดยทั่วไปมักจะละเลยหรือไม่มีการนำปัจจัยคอพพูลาเข้ามาพิจารณา ซึ่งเป็นความผิดพลาดที่อาจเกิดขึ้นบ่อยและมักพบในการใช้งานจริงมากที่สุด ทั้งที่ในความเป็นจริงแล้วปัจจัยคอพพูลามีความสำคัญที่ควรนำมาพิจารณาโดยดูจากตัวอย่าง 2.4 ในบทที่ 2 ซึ่งการละเลยปัจจัยคอพพูลานี้อาจส่งผลต่อความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอยดังแสดงผลการทดลองในกรณีศึกษา 4.1 ซึ่งสามารถยกตัวอย่างเพื่อให้เห็นภาพอย่างชัดเจนในการเปรียบเทียบผลการประมาณค่าสัมประสิทธิ์การถดถอยระหว่างกรณีไม่มีปัจจัยคอพพูลา (กรณีศึกษา1) และมีปัจจัยคอพพูลา (กรณีศึกษา4) เข้ามาเกี่ยวข้องในกรณีที่มีจำนวนข้อมูลน้อย $(n=100, m=5)$ ดังนี้



รูปที่ 5.1 แสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ที่ระดับความสัมพันธ์ต่างๆ ระหว่างกรณีศึกษา1 กับกรณีศึกษา 4

จากรูปข้างต้นนั้น จะเห็นว่าการละเลยปัจจัยคอพพูลาสำหรับในกรณีข้อมูลมีไม่มาก มีผลต่อความถูกต้องในการประมาณค่าสัมประสิทธิ์การถดถอยค่อนข้างมากโดยเฉพาะในระดับความสัมพันธ์ที่สูง ดังนั้นสำหรับกรณีที่ข้อมูลไม่มากนี้ เราไม่ควรละเลยปัจจัยคอพพูลาดังกล่าว สำหรับในกรณีที่จำนวนข้อมูลมากนั้น เรายังพอยอมรับผลการประมาณค่าสัมประสิทธิ์ความถดถอยในกรณีไม่มีปัจจัยคอพพูลาได้บ้าง (ดังรูป 4.7) เนื่องจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในกรณีนี้สูงกว่าในกรณีที่มีปัจจัยคอพพูลาไม่มากนักเมื่อเทียบกับจำนวนข้อมูลที่น้อย อย่างไรก็ตามถ้าต้องการให้ได้ผลการประมาณค่าสัมประสิทธิ์การถดถอยที่เหมาะสม เราควรใช้ตัวแบบที่มีปัจจัยคอพพูลาเข้ามาเกี่ยวข้อง เนื่องจากสามารถ

ให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยใกล้เคียงพารามิเตอร์ ทั้งในกรณีมีจำนวนข้อมูลที่น้อยและจำนวนข้อมูลที่มาก อย่างไรก็ตามจำนวนกลุ่มข้อมูลควรมีมากกว่า 1 กลุ่มด้วย

5.1.4 <u>สรุปผลการวิจัยในการประมาณค่าสัมประสิทธิ์การถดถอยจาก 4 กรณีศึกษา</u>

จากผลการวิจัยทั้ง 4 กรณีศึกษาข้างต้นนั้น พบว่า สำหรับกรณีทราบค่าปัจจัยของคอพพูลา Z และมีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ เพียงกรณีเดียวที่ให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยถูกต้องและใกล้เคียงค่าพารามิเตอร์มากที่สุด โดยที่จำนวนกลุ่มตัวอย่างต้องมากกว่า 1 กลุ่ม ไม่เช่นนั้นทำให้ผลการประมาณสูงกว่าค่าพารามิเตอร์ สำหรับในกรณีที่ทราบค่าและไม่ทราบค่าปัจจัยของคอพพูลา Z ที่ยังไม่ได้มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ พบว่า ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ส่วนค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ (Overestimate) และในกรณีที่ไม่ทราบค่าปัจจัยของคอพพูลา Z ที่มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ พบว่า ค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_0 สูงกว่าค่าพารามิเตอร์ (Overestimate) ส่วนค่าเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอย b_1 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ซึ่งสามารถสรุปเป็นตารางแสดงผลการประมาณค่าสัมประสิทธิ์การถดถอยของทั้ง 4 กรณีศึกษา ได้ดังนี้

ตารางที่ 5.1 ตารางสรุปแสดงผลการประมาณค่าสัมประสิทธิ์การถดถอยสำหรับ 4 กรณีศึกษา

กรณีศึกษา	ไม่มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอย โดยการคูณด้วย $\sqrt{1-\rho}$	มีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอย โดยการคูณด้วย $\sqrt{1-\rho}$
ไม่ทราบค่าปัจจัยคอพพูลา Z	*(กรณีศึกษา1)* ✗ ผลการประมาณค่าสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ✗ ผลการประมาณสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ (Overestimate)	*(กรณีศึกษา2)* ✗ ผลการประมาณค่าสัมประสิทธิ์การถดถอย b_0 สูงกว่าค่าพารามิเตอร์ (Overestimate) ✗ ผลการประมาณสัมประสิทธิ์การถดถอย b_1 ต่ำกว่าค่าพารามิเตอร์ (Underestimate))
ทราบค่าปัจจัยคอพพูลา Z	*(กรณีศึกษา3)* ✗ ผลการประมาณค่าสัมประสิทธิ์การถดถอย b_0 ต่ำกว่าค่าพารามิเตอร์ (Underestimate) ✗ ผลการประมาณสัมประสิทธิ์การถดถอย b_1 สูงกว่าค่าพารามิเตอร์ (Overestimate)	*(กรณีศึกษา4)* ✗ <u>กลุ่มตัวอย่าง 1 กลุ่ม</u> ผลการประมาณค่าสัมประสิทธิ์การถดถอย สูงกว่าค่าพารามิเตอร์ (Overestimate) ✓ <u>กลุ่มตัวอย่าง 5 และ 10 กลุ่ม</u> ผลการประมาณถูกต้องใกล้เคียงค่าพารามิเตอร์

5.1.5 <u>สรุปผลการวิจัยเกี่ยวกับปัจจัยที่มีผลต่อค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าสัมประสิทธิ์การถดถอย</u> ดังนี้

- <u>ระดับความสัมพันธ์</u> สำหรับในกรณีศึกษา 1 – 3 พบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นตามระดับความสัมพันธ์ที่เพิ่มขึ้น โดยเฉพาะในระดับความสัมพันธ์ที่สูงทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าเพิ่มสูงขึ้นมากโดยเฉพาะจำนวนกลุ่มตัวอย่าง 1 กลุ่ม ส่วนในกรณีศึกษา 4 พบว่าในกลุ่มตัวอย่าง 1 กลุ่ม มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเพิ่มขึ้นตามระดับความสัมพันธ์ที่สูงขึ้น โดยมีแนวโน้มในการเพิ่มขึ้นไม่มากนักเมื่อเทียบกับในกรณีก่อนหน้านี้ สำหรับจำนวนกลุ่มตัวอย่างเท่ากับ 5 และ 10 กลุ่ม พบว่าในแต่ละระดับความสัมพันธ์ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองใกล้เคียงกันมาก และมีค่าลดลงเพียงเล็กน้อยในระดับความสัมพันธ์ที่สูงขึ้น

- <u>จำนวนกลุ่มตัวอย่าง</u> ในทุกกรณีศึกษาพบว่า เมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมีค่าลดลงในทุกระดับความสัมพันธ์ โดยมีแนวโน้มลดลงอย่างรวดเร็วในช่วงแรกที่มีจำนวนกลุ่มตัวอย่าง 1 กลุ่ม และค่อยๆ ลดลงเมื่อจำนวนกลุ่มตัวอย่างเพิ่มขึ้นเป็น 5 และ 10 กลุ่ม หรืออาจกล่าวได้ว่า จำนวนกลุ่มตัวอย่างแปรผกผันกับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองนั่นเอง และจำนวนกลุ่มตัวอย่างที่เพิ่มขึ้นนี้จะช่วยลดความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยให้น้อยลง

- <u>ขนาดตัวอย่าง</u> ในทุกกรณีศึกษาพบว่า จำนวนขนาดตัวอย่างไม่ได้มีบทบาทสำคัญต่อการลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองมากนักเมื่อเทียบกับจำนวนกลุ่มตัวอย่าง โดยเห็นได้ชัดเจนจากกรณีที่จำนวนชุดข้อมูลเท่ากัน (ดังรูป 4.1.4 – 4.4.4) ทั้งในกรณีศึกษา 1 – 4 ซึ่งพบว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเปลี่ยนแปลงตามกลุ่มที่เปลี่ยนไป โดยที่ขนาดตัวอย่างในการทดลองไม่ได้มีผลต่อการลดลงของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

ผลสรุปจากการศึกษาและประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลา โดยวิธีการประมาณสัมประสิทธิ์การถดถอยสำหรับกรณีที่ทราบค่าปัจจัยของคอพพูลา z ในแต่ละการศึกษานั้น พบว่าตัวแบบคอพพูลานี้ คือตัวแบบโพรบิท ดังนั้นจึงใช้วิธีการวิเคราะห์ความถดถอยโพรบิทในการหาค่าตัวประมาณของสัมประสิทธิ์การถดถอยได้ โดยตัวประมาณสัมประสิทธิ์การถดถอยที่ได้นั้นควรนำมาปรับค่าโดยการคูณด้วย $\sqrt{1-\rho}$ จึงทำให้ได้ตัวประมาณสัมประสิทธิ์การถดถอยที่เหมาะสมสำหรับตัวแบบนี้ในทุกระดับความสัมพันธ์ ที่จำนวนกลุ่มทดลองมากกว่า 1 กลุ่ม

5.2 ข้อเสนอแนะ

5.2.1 งานวิจัยครั้งนี้เหมาะสำหรับกรณีที่มีจำนวนกลุ่มทดลองมากกว่า 1 กลุ่ม และจำนวนข้อมูลในแต่ละกลุ่มควรมากด้วยจึงทำให้ผลการประมาณค่าสัมประสิทธิ์การถดถอยมีความถูกต้องมากขึ้น

5.2.2 ในการประมาณค่าสัมประสิทธิ์การถดถอยในตัวแบบความถดถอยโลจิสติคแบบเกาซ์เซียนคอพพูลาสำหรับงานวิจัยครั้งนี้ ใช้วิธีการวิเคราะห์ความถดถอยโพรบิทในการหาค่าตัวประมาณของสัมประสิทธิ์การถดถอย ซึ่งสามารถทำได้เฉพาะในกรณีที่ทราบค่าปัจจัยของคอพพูลา z และมีการปรับค่าตัวประมาณสัมประสิทธิ์การถดถอยโดยการคูณด้วย $\sqrt{1-\rho}$ เท่านั้น ซึ่งเป็นที่น่าสนใจที่จะทำการประมาณค่าด้วยวิธีอื่น ที่สามารถครอบคลุมกรณีไม่ทราบค่าปัจจัยของคอพพูลา z ต่อไป

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์ข้อมูลหลายตัวแปร. พิมพ์ครั้งที่ 1. กรุงเทพมหานคร: ธรรมสาร, 2548.

ฐิติมา จิรเศรษฐสิริ. การจำลองตัวแปรสุ่มร่วมด้วยเทคนิคคอพพูลาเมื่อทราบการแจกแจงส่วนริมและสหสัมพันธ์. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2548.

ธาตรี จันทรโคลิกา. "แบบจำลองตัวแปรไม่ต่อเนื่อง". เอกสารประกอบคำบรรยายเศรษฐมิติ 2. กรุงเทพฯ: คณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์, 2542.

ปิยะลักษณ์ พุทธวงศ์. "ตัวแปรตามเชิงคุณภาพและแบบจำลองที่ตัวแปรตามที่มีค่าจำกัด". เอกสารประกอบการบรรยายเศรษฐศาสตร์. เชียงใหม่: คณะเศรษฐศาสตร์ มหาวิทยาลัยเชียงใหม่, 2546.

ภาษาอังกฤษ

Alexander McNeil, Daniel Straumann and Paul Embrechts. Correlation and Dependence in Risk Management: Properties and Pitfalls. 5th ed. Switzerland, 1999.

Daniel A. Powers, Yu Xie. Statistical Methods for Categorical Data Analysis. USA: Academic Press, 2000.

Filip Lindskog. Modeling Dependence with Copulas and Applications to Risk Management. Switzerland, 2001.

Green, William H., Econometric Analysis. New Jersey, USA: Prentice Hall International Inc., 4th edition, 2000.

Sunti Tirapat and Seksan Kiatsupaibul. "Credit value at risk via credit scoring model". Simulation Society Research Workshop, 2007.

บรรณานุกรม

ภาษาไทย

กัลยา วานิชย์บัญชา. การวิเคราะห์สถิติขั้นสูงด้วย SPSS. พิมพ์ครั้งที่ 4. กรุงเทพมหานคร: ธรรมสาร, 2548.

ทัศนาพร จงเกตุกรณ์. การประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติคทวินาม. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2546.

ธีระพร วีระถาวร. การอนุมานสถิติขั้นกลาง: โครงสร้างและความหมาย. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2536.

เสกสรร เกียรติสุไพบูลย์. เอกสารประกอบการสอนวิชาการจำลองแบบเชิงสถิติ. กรุงเทพมหานคร: จุฬาลงกรณ์มหาวิทยาลัย, 2548.

ภาษาอังกฤษ

Daniel A. Powers, Yu Xie. Statistical Methods for Categorical Data Analysis. USA: Academic Press, 2000.

David W. Hosmer, Jr., Stanley Lemeshow. Applied Logistic Regression. 2nd ed. USA: John Wiley & Sons, Inc., 2000.

Elisa Luciano, Umberto Cherubini and Walter Vecchiato. Copula Method in Finance. England: John Wiley & Sons, Inc., 2004.

Paul D. Allison. Logistic Regression Using SAS: Theory and Application. USA: John Wiley & Sons, Inc., 2005.

Roger B. Nelsen. An Introduction to Copulas. USA, 1999.

ภาคผนวก

ภาคผนวก ก ตารางแสดงลักษณะการทำงานของฟังก์ชันในโปรแกรม R ที่ใช้ในการวิจัย

ลำดับที่	**ชื่อฟังก์ชัน**	**การทำงานของฟังก์ชัน**
1	rnorm(n)	สร้างเลขสุ่ม n ตัวให้มีการแจกแจงแบบปกติมาตรฐาน
2	rbinom(n, size, prob)	สร้างเลขสุ่มให้มีการแจกแจงแบบทวินาม โดย n แทน กำหนดจำนวนข้อมูล, size แทนจำนวนครั้งในการทดลอง และ prob แทนความน่าจะเป็นของเหตุการณ์ที่สนใจ
3	pnorm(q)	สร้างฟังก์ชันการแจกแจงของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน โดย q แทนเวกเตอร์ของ quantile
4	qnorm(p)	สร้าง quantile function ของตัวแปรสุ่มที่มีการแจกแจงแบบปกติมาตรฐาน โดย p แทนเวกเตอร์ของความน่าจะเป็น
5	data. frame(..., row. names = NULL)	สร้างกรอบข้อมูล ซึ่งประกอบด้วยตัวแปรเวกเตอร์ หลายๆ ตัว ซึ่งทุกตัวมีความยาวเท่ากัน โดยแต่ละสดมภ์ (Column) คือ หนึ่งตัวแปร และแต่ละแถว (Row) คือ ข้อมูลต่างๆ ของ 1 Case
6	glm(formula, family, data)	สร้างสมการตัวแบบเชิงเส้นทั่วไป (generalized linear models) ในการวิเคราะห์ข้อมูลโดย formula แทนลักษณะของตัวแบบที่ต้องการ fixed ค่า family แทน ลักษณะของฟังก์ชันการแจกแจงของความคลาดเคลื่อนหรือฟังก์ชันเชื่อมโยงที่ใช้ และ data แทนกรอบข้อมูลที่นำมาทำการวิเคราะห์ตัวแบบ
7	summary (object,...)	เป็นการสรุปข้อมูลในภาพรวม
8	rbind(x)	เก็บข้อมูลในรูปแบบของแถว

ภาคผนวก ข ตัวอย่างการใช้โปรแกรม R ในการดำเนินงานวิจัย

*** กรณีไม่ทราบค่า Z และกลุ่มตัวอย่าง 1 กลุ่ม ***

```
# Simulate one sample group and number of data 100.
Z <- rnorm(1)
E1 <- rnorm(100)
rho1 <- 0                #Fix rho1=0
U1 <- (sqrt(rho1)*Z)+(sqrt(1-rho1)*E1)
Beta0 <- (-0.47)
Beta1 <- (1)
X <- rbinom(100,1,0.5)
W1 <- Beta0+(Beta1*X)
P1 <- pnorm(W1)
InvP1 <- qnorm(P1)
Y <- ifelse(U1 <= InvP1, 1, 0)
#Test Probit Regression
Data <- data.frame(Y,X,Z)
Data <- edit(Data)
Probit <- glm(Y~X, family=binomial(link=probit),data=Data)
summary(Probit)
```

*** กรณีไม่ทราบค่า Z และกลุ่มตัวอย่าง 5 กลุ่ม ***

```
# Simulate sample group equal to 5 and number of data 100.
Temp0 <- data.frame(Y=5,X=5,round=0)
for (i in 1:5){
Z <- rnorm(1)
E1 <- rnorm(100)
rho1 <- 0.2                    #Fix rho1=0.2
U1 <- (sqrt(rho1)*Z)+(sqrt(1-rho1)*E1)
Beta0 <- (-0.47)
```

```
Beta1 <- (1)
X <- rbinom(100,1,0.5)
W1 <- Beta0+(Beta1*X)
P1 <- pnorm(W1)
InvP1 <- qnorm(P1)
Y <- ifelse(U1 <= InvP1, 1, 0)
round <- i
Temp <- data.frame(Y=Y,X=X,round=round)
Temp0 <- rbind(Temp0,Temp)
}
#Test Probit Regression
Data <- Temp0[Temp0$Y!=5 & Temp0$X!=5 & Temp0$round!=0,]
Data <- edit(Data)
Probit <- glm(Y~X, family=binomial(link=probit),data=Data)
summary(Probit)
```

*** กรณีทราบค่า Z และกลุ่มตัวอย่าง 1 กลุ่ม ***

```
# Simulate one sample group and number of data 500.
Z <- rnorm(1)
E1 <- rnorm(500)
rho1 <- 0.5                    #Fix rho1=0.5
U1 <- (sqrt(rho1)*Z)+(sqrt(1-rho1)*E1)
Beta0 <- (-0.47)
Beta1 <- (1)
X <- rbinom(500,1,0.5)
W1 <- Beta0+(Beta1*X)
P1 <- pnorm(W1)
InvP1 <- qnorm(P1)
Y <- ifelse(U1 <= InvP1, 1, 0)
```

```
#Test Probit Regression
Data <- data.frame(Y,X,Z)
#Data <- edit(Data)
Probit <- glm(Y~X+Z, family=binomial(link=probit),data=Data)
summary(Probit)
```

*** กรณีทราบค่า Z และกลุ่มตัวอย่าง 10 กลุ่ม ***

```
# Simulate one sample group and number of data 1,000.
Temp0 <- data.frame(Y=5,X=5,round=0)
for (i in 1:10){
Z <- rnorm(1)
E1 <- rnorm(1000)
rho1 <- 0.8                    #Fix rho1=0.8
U1 <- (sqrt(rho1)*Z)+(sqrt(1-rho1)*E1)
Beta0 <- (-0.47)          #Fixed Beta0=-0.47
Beta1 <- (1)              #Fixed Beta1=1
X <- rbinom(1000,1,0.5)
W1 <- Beta0+(Beta1*X)
P1 <- pnorm(W1)
InvP1 <- qnorm(P1)
Y <- ifelse(U1 <= InvP1, 1, 0)
round <- i
Temp <- data.frame(Y=Y,X=X,round=round)
Temp0 <- rbind(Temp0,Temp)
}
#Test Probit Regression
Data <- Temp0[Temp0$Y!=5 & Temp0$X!=5 & Temp0$round!=0,]
Data <- edit(Data)
Probit <- glm(Y~X+Z, family=binomial(link=probit),data=Data)
summary(Probit)
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาว สุกัญญา บุญมา เกิดเมื่อวันที่ 28 ธันวาคม พ.ศ. 2524 ที่จังหวัดลำปาง สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาสถิติ ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยธรรมศาสตร์ เมื่อปีการศึกษา 2547 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2549