การหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมเพื่อแก้ปัญหาพหุสัมพันธ์ในการวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่ม

นางสาวสาวิตรี บุญพัชรนนท์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

OPTIMAL RIDGE PARAMETER FOR SOLVING MULTICOLLINEARITY PROBLEM IN BINARY LOGISTIC REGRESSION

Miss Sawitree Boonpatcharanon

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 2011

หัวข้อวิทยานิพนธ์	การหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมเพื่อแก้ปัญหาพหุสัมพันธ์ในการวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่ม
โดย	นางสาวสาวิตรี บุญพัชรนนท์
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

....คณบดีคณะพาณิชยศาสตร์และการบัญชี

(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

.... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา)

.... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์)

.... กรรมการภายนอกมหาวิทยาลัย

(อาจารย์ ดร.อรุณี กำลัง)

สาวิตรี บุญพัชรนนท์: การหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมเพื่อแก้ปัญหาพหุสัมพันธ์ในการวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่ม. (OPTIMAL RIDGE PARAMETER FOR SOLVING MULTICOLLINEARITY IN BINARY LOGISTIC REGRESSION) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: รศ.ดร.กัลยา วานิชย์บัญชา, 88 หน้า.

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมสำหรับแก้ปัญหาพหุสัมพันธ์ในการวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่มซึ่งดำเนินการภายใต้ขอบเขตของจำนวนตัวแปรอิสระที่ทำการศึกษาคือ 2 และ 3 ตัว ระดับความสัมพันธ์ของตัวแปรอิสระซึ่งมีทั้งระดับความสัมพันธ์สูง กลาง และต่ำ การแจกแจงของตัวแปรอิสระซึ่งทำการศึกษาภายใต้ 3 การแจกแจงคือ การแจกแจงปัวซองส์ การแจกแจงปกติ และการแจกแจงแกมมา ค่าสัมประสิทธิ์ความถดถอยมีค่า 2, 3 และ 4 และขนาดตัวอย่างเป็น 40 และ 70 สำหรับจำนวนตัวแปรอิสระเท่ากับ 2 ตัว ส่วนจำนวนตัวแปรอิสระเท่ากับ 3 ตัวจะทำการศึกษาที่ขนาดตัวอย่าง 60 และ 100 โดยทำการจำลองข้อมูลและวิเคราะห์ผลด้วยโปรแกรม R 2.9.0 และโปรแกรม SPSS for Windows ver.19 ทั้งนี้การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธีการแบบริดจ์ใช้หลักการของนิวตัน-ราฟสัน ส่วนเกณฑ์ที่ใช้ในการตัดสินใจเลือกค่าพารามิเตอร์ริดจ์ที่เหมาะสมมีด้วยกัน 2 เกณฑ์ คือ เกณฑ์ Mean Absolute Percentage Error (MAPE) และเกณฑ์ค่าส่วนเบี่ยงเบนมาตรฐาน (SD)

การศึกษาภายใต้ขอบเขตดังกล่าวผลปรากฏว่า ในกรณีของจำนวนตัวแปรอิสระเท่ากับ 2 และ 3 ตัว ให้ผลสอดคล้องไปในทิศทางเดียวกัน โดยถ้าทำการเพิ่มระดับความสัมพันธ์ของตัวแปรอิสระหรือเพิ่มขนาดตัวอย่างจะส่งผลให้ค่าพารามิเตอร์ริดจ์มีค่าสูงขึ้นซึ่งถ้าพิจารณาเปอร์เซ็นต์การเพิ่มขึ้นของค่าพารามิเตอร์ริดจ์ในแต่ละการแจกแจงของตัวแปรอิสระจะพบว่าการแจกแจงที่ให้เปอร์เซ็นต์การเพิ่มขึ้นเรียงลำดับจากมากไปน้อยเป็นดังนี้ การแจกแจงแกมมา การแจกแจงปกติ และการแจกแจงปัวซองส์ ตามลำดับ ส่วนถ้าทำการพิจารณาค่าสัมประสิทธิ์ความถดถอยที่เพิ่มขึ้นพบว่าค่าพารามิเตอร์ริดจ์มีค่าลดลง โดยเปอร์เซ็นต์การลดลงของค่าพารามิเตอร์ริดจ์เรียงลำดับตามการแจกแจงของตัวแปรอิสระสำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 และ 3 จากมากไปน้อยเป็นดังนี้ การแจกแจงแกมมา การแจกแจงปกติ และการแจกแจงปัวซองส์ ส่วนค่าสัมประสิทธิ์ความถดถอยเท่ากับ 4 สามารถเรียงลำดับได้ดังนี้ การแจกแจงปกติ การแจกแจงแกมมา และการแจกแจงปัวซองส์ ตามลำดับ

ภาควิชาสถิติ ..... สถิติ ..... ลายมือชื่อนิสิต...

สาขาวิชา ..... สถิติ ..... ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.................................

ปีการศึกษา ..... 2554 .....

5281926026: MAJOR STATISTICS
KEYWORDS: Multicollinearity / Ridge Parameter / Binary Logistic Regression

SAWITREE BOONPATCHARANON: OPTIMAL RIDGE PARAMETER FOR SOLVING MULTICOLLINEARITY IN BINARY LOGISTIC REGRESSION. ADVISOR: ASSOC.PROF.KANLAYA VANICHBUNCHA, Ph.D., 88 pp.

The purpose of this study is to find the optimal ridge parameter for solving multicollinearity problem between the independent variables in binary logistic regression. This study scopes on 2 and 3 independent variables, low, medium and high correlations, Poisson Normal and Gamma distribution and which the logistic regression coefficients, are 2, 3 and 4, sample sizes are 40 and 70 for the two independent variables, 60 and 100 for the three independent variables. Simulating and analyzing data in this study use R 2.9.0 and SPSS for Windows ver.19. The coefficient of logistic regression uses Newton-Raphson methods. Mean absolute percentage error (MAPE) and standard deviation (SD) are the criteria for selecting the optimal ridge parameter.

Studying under these assumptions can gain many useful results. In case of two and three independent variables, the results are the same. When we increase the correlation or the sample size, the value of ridge parameters are increase. If we consider on the increased percents of ridge parameter, the distributions that have the highest percent increased to the lowest percent are Gamma, Normal and Poisson distribution, respectively. On the other hand, when we consider from the view of the coefficient, the ridge parameters are decreased. If we focus on the decreased percents of ridge parameter, the distributions that have the highest percent to the lowest percent are Gamma, Normal and Poisson distribution for the coefficients that equal 2 and 3, but when the coefficients are equal 4, the highest percent increased to the lowest percent are Normal, Gamma and Poisson distribution.

Department: Statistics Student's Signature…………………………………...

Field of Study: Statistics Advisor's Signature…………………………………...

Academic Year : 2011

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ลงได้ด้วยความช่วยเหลือ และการเอาใจใส่จากอาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา ผู้วิจัยจึงขอกราบขอบพระคุณท่านอาจารย์เป็นอย่างสูงที่คอยให้คำปรึกษา อบรมสั่งสอน และเป็นกำลังใจที่ดีเสมอมา ทั้งนี้ผู้วิจัยขอกราบขอบพระคุณท่านประธานในการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา ท่านคณะกรรมการ ผู้ช่วยศาสตราจารย์ ดร.เสกสรร เกียรติสุไพบูลย์ และท่านคณะกรรมการภายนอก อาจารย์ ดร.อรุณี กำลัง เป็นอย่างสูงที่ท่านอาจารย์ทั้งสามท่านได้เสียสละเวลามาสอบและให้คำแนะนำที่ดีและมีประโยชน์ในการปรับปรุงงานของผู้วิจัยต่อไป

ผู้วิจัยต้องขอขอบพระคุณอาจารย์ทุกท่านในภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ได้ประสิทธิ์ประสาทวิชาความรู้ให้ผู้วิจัยตั้งแต่ระดับปริญญาตรีถึงปริญญาโท ทำให้ผู้วิจัยสามารถนำความรู้เหล่านั้นมาใช้ในวิทยานิพนธ์ได้อย่างเต็มที่

สุดท้ายนี้ผู้วิจัยต้องขอบพระคุณคุณพ่อ คุณแม่ และครอบครัว รวมทั้งเพื่อนๆ ร่วมหลักสูตรที่เป็นกำลังใจและคอยช่วยเหลือเรื่องต่างๆ มาโดยตลอด

สารบัญ

หน้า

หน้า

สารบัญตาราง

ตารางที่ หน้า

ตารางที่ หน้า

สารบัญภาพ

ภาพที่ หน้า

กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัว

ภาพที่ หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เนื่องจากในปัจจุบันการวิเคราะห์ความถดถอย (Regression Analysis) ได้รับความนิยมแพร่หลายในวงการต่างๆ อาทิเช่น วงการการเงิน การแพทย์ ชีวสถิติ เศรษฐศาสตร์ และสังคมศาสตร์ ทั้งนี้ใช้เพื่อการหาความสัมพันธ์ของตัวแปรที่สนใจและเพื่อการประมาณค่าหรือพยากรณ์สิ่งที่สนใจ ซึ่งหลักการของการวิเคราะห์ความถดถอยได้มีข้อกำหนดเบื้องต้นเกี่ยวกับการแจกแจงของความคลาดเคลื่อนไว้หลายประการ เช่น ความคลาดเคลื่อนต้องมีความเป็นอิสระจากกัน ความคลาดเคลื่อนมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และค่าแปรปรวนคงที่ เป็นต้น อีกทั้งตัวแปรตาม (Dependent Variable) ต้องเป็นตัวแปรเชิงปริมาณ ซึ่งจะเห็นได้ว่าในทางปฏิบัติ ตัวแปรตามที่สนใจเป็นได้ทั้งตัวแปรเชิงปริมาณและเชิงกลุ่ม อีกทั้งการแจกแจงของความคลาดเคลื่อนมีการแจกแจงแบบปกติได้ยาก จึงเป็นความน่าสนใจที่การวิเคราะห์ความถดถอยโลจิสติกสามารถช่วยแก้ปัญหาเหล่านี้ได้ กล่าวคือ การวิเคราะห์ความถดถอยโลจิสติกไม่มีเงื่อนไขเกี่ยวกับการแจกแจงของความคลาดเคลื่อนและตัวแปรตามเป็นตัวแปรเชิงกลุ่ม แต่เงื่อนไขที่ทั้งการวิเคราะห์ความถดถอยและการวิเคราะห์ความถดถอยโลจิสติกมีเหมือนกันคือ ความเป็นอิสระระหว่างตัวแปรอิสระด้วยกัน ซึ่งจะเกิดในกรณีที่มีตัวแปรอิสระตั้งแต่ 2 ตัวขึ้นไป

การเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) พบบ่อยในทางปฏิบัติ ซึ่งระดับความรุนแรงของปัญหาขึ้นอยู่กับระดับความสัมพันธ์ของตัวแปรอิสระว่ามากน้อยเพียงใด และปัญหานี้ยังส่งผลต่อการพิจารณาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม โดยปัญหานี้จะทำให้ค่าประมาณสัมประสิทธิ์ความถดถอยไม่เป็นไปตามความจริง คือ เครื่องหมายอาจกลับทิศทาง ค่าอาจมากหรือน้อยเกินความเป็นจริง เป็นต้น ดังนั้นจึงจำเป็นต้องแก้ไขปัญหาพหุสัมพันธ์ก่อนทำการวิเคราะห์ผล ซึ่งวิธีที่ใช้กันอย่างแพร่หลายวิธีหนึ่งในการแก้ไขปัญหาพหุสัมพันธ์คือ การวิเคราะห์ความถดถอยแบบริดจ์ (Ridge Regression) โดยมีหลักการคือ การหาค่าพารามิเตอร์ริดจ์ (Ridge Parameter) ที่เหมาะสม ซึ่งค่าพารามิเตอร์ริดจ์ (λ) จะแตกต่างกันไปในแต่ละกรณี ทางผู้เสนอจึงมีความเห็นว่าในการแก้ปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระในการวิเคราะห์ความถดถอยโลจิสติก

ด้วยการวิเคราะห์ความถดถอยแบบริดจ์นั้น สิ่งที่สำคัญที่สุดคือ การหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมในแต่ละกรณีของขนาดตัวอย่าง จำนวนตัวแปรอิสระ ระดับความสัมพันธ์ของตัวแปรอิสระ และการแจกแจงของตัวแปรอิสระ

ในงานวิจัย[1] ได้เสนอวิธีการประมาณค่าสัมประสิทธิ์ความถดถอยแบบริดจ์ในการวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่มด้วยวิธี Newton Raphson ไว้ ซึ่งจะใช้เป็นรูปแบบของสมการที่จะนำมาใช้ในการศึกษาครั้งนี้ ส่วนในแง่ของพหุสัมพันธ์ที่เกิดขึ้นและส่งผลกระทบต่อการประมาณค่าสัมประสิทธิ์ความถดถอยจะทำการศึกษาทั้งระดับพหุสัมพันธ์ที่ไม่สูงมากจนถึงระดับพหุสัมพันธ์ที่สูงมาก จากที่กล่าวมาทั้งหมดนี้เป็นเหตุผลให้การแก้ไขปัญหาพหุสัมพันธ์ (Multicollinearity) มีความจำเป็นอย่างยิ่งโดยเฉพาะในกรณีที่ระดับความสัมพันธ์ระหว่างตัวแปรอิสระสูง จึงเป็นที่มาของการศึกษาเพื่อหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมในกรณีต่างๆ

1.2 วัตถุประสงค์ของการศึกษา

เพื่อหาค่าพารามิเตอร์ริดจ์ (λ) ที่เหมาะสม สำหรับแก้ปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระของการวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่ม

1.3 ขอบเขตของการศึกษา

การศึกษาครั้งนี้ได้ทำการศึกษาภายใต้ขอบเขตดังนี้

1. ศึกษาการวิเคราะห์ความถดถอยโลจิสติกที่มีตัวแปรตาม (Y) มีค่าเพียง 2 ค่า (Binary Logistic Regression)

---

[1] Duffy, D. E. and Santner, T. J., "On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models.," Communs Statist. Theory Meth., 18, (1989): 959-980.

2. กำหนดตัวแปรอิสระ (X) เป็นตัวแปรเชิงปริมาณเท่านั้น และทำการศึกษาในกรณีที่มีจำนวนตัวแปรอิสระ (p) 2 และ 3 ตัว โดยที่ตัวแปรอิสระมีการแจกแจงแบบเดียวกันทุกตัว และมีความสัมพันธ์ระหว่างตัวแปรอิสระเป็นดังนี้

2.1 กรณีตัวแปรอิสระ 2 ตัว (p = 2)

ให้ ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 2 โดยทำการศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

2.1.1 ระดับความสัมพันธ์ 0.3

2.1.2 ระดับความสัมพันธ์ 0.5

2.1.3 ระดับความสัมพันธ์ 0.7

2.1.4 ระดับความสัมพันธ์ 0.8

2.2 กรณีตัวแปรอิสระ 3 ตัว (p = 3)

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

ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 2

ρ_{13} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 3

ρ_{23} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 2 และตัวแปรอิสระตัวที่ 3

ทำการศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

2.2.1 ระดับความสัมพันธ์ (0.3, 0.5, 0.8)

2.2.2 ระดับความสัมพันธ์ (0.5, 0.5, 0.8)

2.2.3 ระดับความสัมพันธ์ (0.5, 0.7, 0.8)

2.2.4 ระดับความสัมพันธ์ (0.7, 0.7, 0.8)

2.2.5 ระดับความสัมพันธ์ (0.8, 0.8, 0.8)

ซึ่งทำการศึกษาภายใต้การแจกแจง 3 การแจกแจง ดังนี้

2.3 การแจกแจงแบบต่อเนื่อง

2.3.1 การแจกแจงปกติหลายตัวแปร (Multivariate Normal) ที่มีพารามิเตอร์ค่าเฉลี่ย $(\tilde{\mu})$ เป็น $\tilde{0}$ และเมทริกซ์ความแปรปรวนร่วม (Σ) มีความแปรปรวนร่วมระหว่างตัวแปรอิสระตามระดับความสัมพันธ์ของตัวแปรอิสระ ดังที่ได้กล่าวมา

2.3.2 การแจกแจงแกมมา (Gamma Distribution) ที่มีพารามิเตอร์คือ (γ,α) เท่ากับ (1,2) สำหรับตัวแปรอิสระตัวที่ 1 (γ,α) เท่ากับ (1.5,2) สำหรับตัวแปรอิสระตัวที่ 2 และ (γ,α) เท่ากับ (1,2) สำหรับตัวแปรอิสระตัวที่ 3 โดยทำการศึกษาเมื่อตัวแปรอิสระมีระดับความสัมพันธ์ตามที่กล่าวมา

2.4 การแจกแจงแบบไม่ต่อเนื่อง

2.4.1 การแจกแจงปัวซองส์ (Poisson Distribution) ที่มีพารามิเตอร์คือค่าเฉลี่ย (θ) เท่ากับ 2 สำหรับตัวแปรอิสระตัวที่ 1 และ (θ) เท่ากับ 3 สำหรับตัวแปรอิสระตัวที่ 2 โดยทำการศึกษาเมื่อตัวแปรอิสระมีระดับความสัมพันธ์ตามที่กล่าวมา

3. ศึกษาภายใต้ขนาดตัวอย่างที่มีจำนวนตัวอย่างอย่างน้อย 20 เท่าของจำนวนตัวแปรอิสระ (20p) คือ

3.1 ถ้าจำนวนตัวแปรอิสระเท่ากับ 2 จำนวนตัวอย่างที่ทำการศึกษาคือ 40 และ 70

3.2 ถ้าจำนวนตัวแปรอิสระเท่ากับ 3 จำนวนตัวอย่างที่ทำการศึกษาคือ 60 และ100

4. กำหนดค่าเริ่มต้นของสัมประสิทธิ์ความถดถอย $(\beta_0,\beta_1,...,\beta_p)$ ดังนี้

4.1 กรณีตัวแปรอิสระ 2 ตัว (p = 2) ค่าสัมประสิทธิ์ความถดถอยเริ่มต้นมีค่าเป็น $(\beta_0,\beta_1,\beta_2)\in[2,4]$

4.2 กรณีตัวแปรอิสระ 3 ตัว (p = 3) ค่าสัมประสิทธิ์ความถดถอยเริ่มต้นมีค่าเป็น $(\beta_0,\beta_1,\beta_2,\beta_3)\in[2,4]$

5. จำลองค่า y_i โดยทำการพิจารณาจากตัวแปรสุ่มแบบยูนิฟอร์มที่มีพารามิเตอร์คือ 0 และ 1 โดย

$$\pi(x_i) \quad = \quad \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_p X_p}}$$

โดยที่ $y_i = 1$ ก็ต่อเมื่อ ค่าของตัวแปรสุ่มแบบยูนิฟอร์มมีค่าน้อยกว่าหรือเท่ากับ $\pi(x_i)$

และ $y_i = 0$ ก็ต่อเมื่อ ค่าของตัวแปรสุ่มแบบยูนิฟอร์มมีค่ามากกว่า $\pi(x_i)$

1.4 คำจำกัดความที่ใช้ในการศึกษา

ค่าพารามิเตอร์ริดจ์ (λ) คือ ค่าที่ต้องการพิจารณาซึ่งเป็นจุดประสงค์หลักของงานวิจัยครั้งนี้ โดยมีขอบเขตของค่าพารามิเตอร์ริดจ์มีค่าไม่เกิน 1

1.5 วิธีดำเนินการศึกษา

1. กำหนดค่าเริ่มต้น ดังนี้
   - 1.1 ระดับความสัมพันธ์ของตัวแปรอิสระ (ρ)
   - 1.2 จำนวนตัวแปรอิสระ (p)
   - 1.3 ขนาดตัวอย่าง (n)
   - 1.4 ค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $(\tilde{\beta})$
2. จำลองข้อมูลตามการแจกแจงโดยให้มีระดับความสัมพันธ์ตามที่กำหนด
3. กำหนดค่าพารามิเตอร์ริดจ์เริ่มต้น ดังนี้

   การแจกแจงแบบปัวซองส์ สำหรับตัวแปรอิสระ 2 ตัว

   - 3.1 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.00001

3.2 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.0000001

3.3 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.000000001

การแจกแจงแบบปกติ สำหรับตัวแปรอิสระ 2 และ 3 ตัว

3.4 สำหรับค่าสัมประสิทธิ์ความถดถอยทุกค่า กำหนดให้ค่าพารามิเตอร์ริดจ์เริ่มต้น มีค่า 0.001

การแจกแจงแบบแกมมา สำหรับตัวแปรอิสระ 2 ตัว

3.5 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.0001

3.6 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 และ 4 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.0000001

การแจกแจงแบบแกมมา สำหรับตัวแปรอิสระ 3 ตัว

3.7 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.000001

3.8 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.0000001

3.9 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.00000001

4. คำนวณค่าความน่าจะเป็น $\pi(x_i)$
5. สร้างตัวแปรสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม โดยให้พารามิเตอร์มีค่า 0 และ 1
6. จำลองค่า y โดยแปลงค่าความน่าจะเป็นจากข้อ 4 ให้เป็นค่า 0 หรือ 1 โดย

   y_i = 1 ก็ต่อเมื่อ ค่าของตัวแปรสุ่มแบบยูนิฟอร์มมีค่าน้อยกว่าหรือเท่ากับ $\pi(x_i)$ และ

   y_i = 0 ก็ต่อเมื่อ ค่าของตัวแปรสุ่มแบบยูนิฟอร์มมีค่ามากกว่า $\pi(x_i)$
7. ประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธี NR (Newton Raphson) โดยจะหยุดเมื่อ

$$\left|\hat{\tilde{\beta}}_{k+1} - \hat{\tilde{\beta}}_k\right| < 0.0000001$$

โดย $\hat{\hat{\beta}}_{k+1}$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยที่คำนวณได้จากรอบที่ k+1

$\hat{\hat{\beta}}_{k}$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยที่คำนวณได้จากรอบที่ k

8. คำนวณค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE)

$$\mathrm{MAPE} = \left(\frac{1}{p+1} \sum_{i=0}^{p} \left| \frac{\beta_i - \hat{\beta}_i}{\beta_i} \right| \right) \times 100 \quad ; i = 0, 1, \ldots, p$$

โดย p คือ จำนวนตัวแปรอิสระ

β_i คือ ค่าสัมประสิทธิ์ความถดถอยที่กำหนดขึ้นตัวที่ i

$\hat{\beta}_i$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยตัวที่ i

พิจารณาค่า MAPE ตามเกณฑ์ดังนี้

สำหรับการแจกแจงแบบปกติและการแจกแจงแบบปัวซองส์ เมื่อตัวแปรอิสระมีจำนวน 2 และ 3 ตัว

8.1 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่า MAPE ที่กำหนดไว้คือ 10%

8.2 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่า MAPE ที่กำหนดไว้คือ 7%

8.3 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่า MAPE ที่กำหนดไว้คือ 5%

สำหรับการแจกแจงแบบแกมมา เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่า 2, 3 และ 4

8.4 ถ้าจำนวนตัวแปรอิสระมี 2 ตัว ค่า MAPE ที่กำหนดไว้คือ 10%

8.5 ถ้าจำนวนตัวแปรอิสระมี 3 ตัว ค่า MAPE ที่กำหนดไว้คือ 15%

ค่า MAPE ต้องน้อยกว่าเกณฑ์ที่กำหนดไว้ จึงสามารถดำเนินการในขั้นตอนต่อไปได้

9. คำนวณค่าส่วนเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ริดจ์ (SD)

$$\mathrm{SD} = \sqrt{\frac{\sum_{j=1}^{r} \left(\lambda_i - \bar{\lambda}\right)^2}{r-1}}$$

โดยที่ SD คือ ค่าเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ริดจ์

r คือ จำนวนรอบที่ทำการวิเคราะห์ซึ่งนับจำนวนรอบตามจำนวนค่าพารามิเตอร์ริดจ์ที่ผ่านเกณฑ์ของ MAPE

λ_i คือ ค่าพารามิเตอร์ริดจ์ตัวที่ i ที่ผ่านเกณฑ์ของ MAPE

$\bar{\lambda}$ คือ ค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์จนถึงรอบที่ r

10. พิจารณาค่า SD ตามเกณฑ์ดังนี้

การแจกแจงแบบปัวซองส์ สำหรับตัวแปรอิสระ 2 ตัว

10.1 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่า SD ที่กำหนดไว้คือ 0.00001

10.2 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่า SD ที่กำหนดไว้คือ 0.0000001

10.3 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่า SD ที่กำหนดไว้คือ 0.000000001

การแจกแจงแบบปกติ สำหรับตัวแปรอิสระ 2 และ 3 ตัว

10.4 สำหรับค่าสัมประสิทธิ์ความถดถอยทุกค่า ค่า SD ที่กำหนดไว้คือ 0.001

การแจกแจงแบบแกมมา สำหรับตัวแปรอิสระ 2 ตัว

10.5 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่า SD ที่กำหนดไว้คือ 0.0001

10.6 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 และ 4 ค่า SD ที่กำหนดไว้คือ 0.0000001

การแจกแจงแบบแกมมา สำหรับตัวแปรอิสระ 3 ตัว

10.7 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่า SD ที่กำหนดไว้คือ 0.000001

10.8 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่า SD ที่กำหนดไว้คือ 0.0000001

10.9 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่า SD ที่กำหนดไว้คือ 0.00000001

ค่า SDต้องน้อยกว่าเกณฑ์ที่กำหนดไว้ จึงสามารถดำเนินการในขั้นตอนต่อไปได้

11. คำนวณค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์และค่า MAPE
12. สรุปผลการศึกษา

1.6 ประโยชน์ที่คาดว่าจะได้รับ

ได้ค่าพารามิเตอร์ริดจ์ (λ) ที่เหมาะสมในกรณีต่างๆ ที่ทำการพิจารณาเพื่อนำไปใช้แก้ปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระในการวิเคราะห์ความถดถอยโลจิสติกแบบ 2 กลุ่ม

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

การศึกษาครั้งนี้ทำการศึกษาเกี่ยวกับการแก้ปัญหาพหุสัมพันธ์ของตัวแปรอิสระที่เกิดขึ้นเมื่อตัวแปรอิสระมีความสัมพันธ์กัน ซึ่งขัดแย้งกับหลักการเบื้องต้นของการวิเคราะห์ความถดถอยว่า ตัวแปรอิสระต้องเป็นอิสระกัน ทั้งนี้ปัญหาพหุสัมพันธ์ถือเป็นปัญหาที่สำคัญมากที่ต้องได้รับการแก้ไขก่อนที่จะทำการวิเคราะห์และแปลผลเพื่อการใช้งานจริง ดังนั้นการศึกษาครั้งนี้จึงทำการศึกษาเกี่ยวกับการแก้ปัญหาพหุสัมพันธ์ของตัวแปรอิสระที่เกิดขึ้น ซึ่งวิธีการแก้ปัญหามีหลากหลายวิธีมาก แต่ทางผู้ศึกษาสนใจวิธีการริดจ์ซึ่งเป็นวิธีการที่ไม่ต้องลดหรือตัดตัวแปรอิสระใดๆ ออกจากการวิเคราะห์ ทำให้ตัวแปรอิสระทุกตัวยังคงมีบทบาทโดยตรงในการวิเคราะห์ต่อไป อีกทั้งผู้ศึกษายังสนใจการแก้ปัญหานี้ในการวิเคราะห์ความถดถอยแบบโลจิสติกซึ่งใช้กันอย่างแพร่หลายในขณะนี้อีกด้วย การศึกษาครั้งนี้จึงเกี่ยวข้องกับทฤษฎีทางสถิติที่สำคัญหลายทฤษฎี ซึ่งสามารถอธิบายรายละเอียดได้ดังนี้

2.1 **การวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่ม** (Binary Logistic Regression Analysis)[1]

การวิเคราะห์ความถดถอยโลจิสติกยังคงมีวัตถุประสงค์และแนวคิดเหมือนกับการวิเคราะห์ความถดถอยแบบปกติ คือ เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ (independent variable) กับตัวแปรตาม (dependent variable) และนำสมการถดถอยที่ได้ไปประมาณหรือพยากรณ์ค่าตัวแปรตาม เมื่อกำหนดค่าตัวแปรอิสระ แต่แตกต่างจากการวิเคราะห์ความถดถอยแบบปกติในส่วนของตัวแปรตามซึ่งเป็นตัวแปรเชิงกลุ่ม (categorical variable) สำหรับการวิเคราะห์ความถดถอยโลจิสติกแบบ 2 กลุ่ม (binary logistic regression) เป็นการวิเคราะห์ความถดถอยที่ตัวแปรตามเป็นตัวแปรเชิง

---

[1]กัลยา วานิชย์บัญชา, <u>การวิเคราะห์ข้อมูลหลายตัวแปร</u>, พิมพ์ครั้งที่ 4 (กรุงเทพมหานคร: บริษัทธรรมสาร, 2552), 424 – 484.

กลุ่มที่มีได้เพียง 2 ค่า (dichotomy or binary variable) ส่วนตัวแปรอิสระอาจเป็นตัวแปรเชิงกลุ่มหรือตัวแปรเชิงปริมาณก็ได้ และตัวแปรอิสระอาจมีเพียง 1 ตัว (Simple Logistic Regression) หรืออาจมีได้ตั้งแต่ 2 ตัวขึ้นไป (Multiple Logistic Regression) โดยในการศึกษาครั้งนี้ กำหนดให้ ตัวแปรตามมีได้ 2 ค่าคือ

$$Y = \begin{cases} 1 & \text{ถ้าเกิดเหตุการณ์ที่สนใจ} \\ 0 & \text{ถ้าไม่เกิดเหตุการณ์ที่สนใจ} \end{cases}$$

และตัวแปรอิสระเป็นตัวแปรเชิงปริมาณทั้งหมดมีตั้งแต่ 2 ตัวขึ้นไป ซึ่งเป็นการวิเคราะห์ความถดถอยโลจิสติกเชิงพหุ (Multiple Logistic Regression Analysis) เป็นการวิเคราะห์ความถดถอยโลจิสติกที่มีตัวแปรอิสระมากกว่า 1 ตัว ในรูปทั่วไป กำหนดให้มี p ตัว คือ $X_1, X_2, \ldots, X_p$ และ Y มีการแจกแจงแบบเบอร์นูลีที่มีความน่าจะเป็น p

$$P(Y = y) = p^y (1-p)^{1-y} \qquad ; y = 0,1$$

สำหรับหน่วยตัวอย่างที่ i จะได้ว่า

$$P(Y = y_i) = p^{y_i} (1-p)^{1-y_i} \qquad ; y_i = 0,1$$

และเมื่อ $y_i = 0$ จะได้ว่า

$$P(Y_i = 0) = p^0 (1-p)^{1-0} = 1-p$$

และเมื่อ $y_i = 1$ จะได้ว่า

$$\begin{aligned} P(Y_i = 1) &= p^1 (1-p)^{1-1} = p \\ E(Y_i) &= \sum_{i=0}^{1} y_i P(Y_i = y_i) \\ &= 0 \times P(Y_i = 0) + 1 \times P(Y_i = 1) \\ &= 0 \times (1-p) + 1 \times p \\ &= p \end{aligned}$$

ซึ่งทำให้ $0 \leq E(Y) \leq 1$

เนื่องจากค่า Y มีได้เพียง 2 ค่าคือ 0 และ 1 จึงทำให้ความสัมพันธ์ระหว่าง X และY ไม่ได้อยู่ในรูปเชิงเส้น แต่อยู่ในรูป

$$E(Y) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}} = \pi(x)$$

โดยเรียกสมการนี้ว่า Logistic Response Function โดยที่ $0 \leq E(Y) \leq 1$

ดังนั้น P(เกิดเหตุการณ์ที่สนใจ) = $P(Y=1) = \dfrac{e^{\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p}}$

2.2 การประมาณค่าสัมประสิทธิ์ความถดถอยโลจิสติกด้วยวิธีแบบริดจ์[2]

การประมาณค่าสัมประสิทธิ์ความถดถอยจะใช้หลักการของความควรจะเป็นสูงสุด (Maximum Likelihood) แต่ไม่สามารถหาค่าสัมประสิทธิ์ความถดถอยได้โดยตรงจึงใช้เทคนิคการทำซ้ำ (Iteration techniques) โดยใช้วิธีการของนิวตัน-ราฟสัน (Newton Raphson method) และเรียกค่าประมาณสัมประสิทธิ์ความถดถอยว่า ตัวประมาณความควรจะเป็นสูงสุด (Maximum Likelihood Estimator: MLE)

พิจารณาตัวแปรอิสระ X ที่ประกอบด้วยเวกเตอร์ $\tilde{x}' = (X_1, X_2, \ldots, X_p)$ ซึ่งมีตัวแปรอิสระจำนวน p ตัว และตัวแปรตาม Y มีการแจกแจงทวินาม กำหนดโดยความน่าจะเป็นแบบมีเงื่อนไขของตัวแปรตาม Y คือ $P(Y = 1|X = x) = \pi(x)$ และ $P(Y = 0|X = x) = 1 - \pi(x)$ ทั้งนี้ทำการศึกษาความสัมพันธ์ระหว่าง $\pi(x) = E(Y|X = x)$ ซึ่งคือค่าเฉลี่ยของ Y เมื่อกำหนด X กับตัวแปรอิสระ X โดยรูปแบบฟังก์ชันโลจิสติก คือ

$$\pi(x_i) = \frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+...+\beta_pX_p}}$$

การแจกแจงความน่าจะเป็นคือ

$$f(y_i) = [\pi(x_i)]^{y_i}[1-\pi(x_i)]^{1-y_i} \qquad \text{เมื่อ } y_i = 0, 1 \,;\, i = 1, 2, \ldots, n$$

---

[2] S.LE Cessie and J.C.Van Houwelinge, "Ridge Regression in Logistic Regression," Journal of the Royal Statistical Society. Series C (Applied Statistics) Vol.41, No.1 (1990): 191-201

ดังนั้น เมื่อ x_i (i = 1,2,…,n) เป็นอิสระจากกัน จะได้ ฟังก์ชันความน่าจะเป็นร่วม คือ

$$\begin{aligned} L &= \prod_{i=1}^{n} (\pi(x_i))^{y_i} (1-\pi(x_i))^{y_i} \\ &= \prod_{i=1}^{n} \left(\frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}} \right)^{y_i} \left(1-\frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}} \right)^{1-y_i} \\ &= \prod_{i=1}^{n} \left(\frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}} \right)^{y_i} \left(\frac{1}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}} \right)^{1-y_i} \end{aligned}$$

ให้ $W = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$

ดังนั้น $L = \prod_{i=1}^{n} \left(\frac{e^{w}}{1+e^{w}} \right)^{y_i} \left(\frac{1}{1+e^{w}} \right)^{1-y_i}$

จะได้ ลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น (Log-likelihood function) คือ

$$\ln L = \sum_{i=1}^{n} \{ y_i \ln[\pi(x_i)] + (1-y_i) \ln[1-\pi(x_i)] \}$$

ทั้งนี้การประมาณค่าแบบบริดจ์ได้อ้างอิงงานวิจัยของ เรวดี เรืองอยู่ (2547) ซึ่งการประมาณค่าตามหลักการของวิธีการประมาณแบบบริดจ์จะให้ค่าประมาณสัมประสิทธิ์ความถดถอยมีค่าใกล้เคียงความเป็นจริงมากกว่าการประมาณค่าด้วยวิธีกำลังสองน้อยที่สุด และส่งผลให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าต่ำกว่าวิธีกำลังสองน้อยที่สุด อีกทั้งยังช่วยแก้ปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระอีกด้วย ซึ่งจากลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น (Log-likelihood function)

$$\ln L = \sum_{i=1}^{n} \{ y_i \ln[\pi(x_i)] + (1-y_i) \ln[1-\pi(x_i)] \}$$

Duffy and Santer (1989) ได้ทำการพิจารณาค่าสูงสุดของลอการิทึมธรรมชาติของฟังก์ชันภาวะน่าจะเป็น ไว้ดังนี้

$$l^{\lambda}(\tilde{\beta}) = l(\tilde{\beta}) - \lambda \|\tilde{\beta}\|^2 \quad \text{-------------- (1)}$$

โดย $\|\tilde{\beta}\| = \left(\sum_{j=0}^{p}\beta_j^2\right)^{\frac{1}{2}}$

กำหนดให้ $X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1i} & x_{2i} & \cdots & x_{pi} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix}_{n\times(p+1)}$ และ $\tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}_{(p+1)\times 1}$

ค่าสูงสุดของสมการ (1) จะถูกกำหนดโดยค่า $\hat{\tilde{\beta}}^{\lambda}$ โดยมีค่า λ เป็นตัวควบคุมการลดลงของพารามิเตอร์ $\tilde{\beta}$ ซึ่งสามารถประมาณค่า $\hat{\tilde{\beta}}^{\lambda}$ โดยวิธี Newton – Raphson ซึ่งตามวิธีของ Newton – Raphson จะทำการหาอนุพันธ์ย่อยเทียบกับ $\beta_0, \beta_1, ..., \beta_p$ ของ $l^{\lambda}(\tilde{\beta})$ แล้วนำมาเป็นสมาชิกของเวกเตอร์ $U^{\lambda}(\tilde{\beta})$ ซึ่งมีมิติเท่ากับ (p+1)x1 ดังนี้

$$U^{\lambda}(\tilde{\beta}) = \sum_{i=1}^{n} x_i'\{y_i - \pi(x_i)\} - 2\lambda\tilde{\beta}$$

$$= U(\tilde{\beta}) - 2\lambda\tilde{\beta} \qquad \text{-----------------(2)}$$

โดยที่ $U(\tilde{\beta}) = \begin{bmatrix} \dfrac{\partial L(\tilde{\beta})}{\partial \beta_0} \\ \dfrac{\partial L(\tilde{\beta})}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial L(\tilde{\beta})}{\partial \beta_p} \end{bmatrix}$

กำหนด $\Omega^{\lambda}(\tilde{\beta}) = \Omega(\tilde{\beta}) + 2\lambda I$ ----------------(3)

โดยที่ $\Omega = X'V(\tilde{\beta})X$ และ

$V(\tilde{\beta})$ คือ nxn diagonal matrix ที่มี $v_{ii} = \pi(x_i)(1-\pi(x_i))$ จะได้

$$V(\tilde{\beta}) = \begin{bmatrix} \pi(x_1)(1-\pi(x_1)) & 0 & \cdots & 0 \\ 0 & \pi(x_2)(1-\pi(x_2)) & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \pi(x_n)(1-\pi(x_n)) \end{bmatrix}$$

พิจารณาหาเวกเตอร์

$$U^{\lambda}(\hat{\hat{\beta}}^{\lambda}) = U^{\lambda}(\tilde{\beta}) - (\hat{\hat{\beta}}^{\lambda} - \tilde{\beta})'\Omega^{\lambda}(\tilde{\beta}) + 0\left(\left\|\hat{\hat{\beta}}^{\lambda} - \tilde{\beta}\right\|\right)$$

จาก (2) และ (3) และให้ $U^{\lambda}(\hat{\hat{\beta}}^{\lambda}) = 0$ จะได้ว่า

$$\hat{\hat{\beta}}^{\lambda} = \hat{\hat{\beta}} + \{\Omega(\hat{\hat{\beta}}) + 2\lambda I\}^{-1}\{U(\hat{\hat{\beta}}) - 2\lambda\hat{\hat{\beta}}\}$$

$$= \{\Omega(\hat{\hat{\beta}}) + 2\lambda I\}^{-1}\{U(\hat{\hat{\beta}}) + \Omega(\hat{\hat{\beta}})\hat{\hat{\beta}}\}$$

ดังนั้น $\hat{\hat{\beta}}^{\lambda} = \{\Omega(\hat{\hat{\beta}}) + 2\lambda I\}^{-1}\{U(\hat{\hat{\beta}}) + \Omega(\hat{\hat{\beta}})\hat{\hat{\beta}}\}$

เพราะฉะนั้นในการประมาณค่าสัมประสิทธิ์ความถดถอยแบบริดจ์ด้วยวิธี Newton – Raphson ในรอบที่ k+1 เป็นดังนี้

$$\hat{\hat{\beta}}^{\lambda}_{k+1} = \{\Omega(\hat{\hat{\beta}}_k) + 2\lambda I\}^{-1}\{U(\hat{\hat{\beta}}_k) + \Omega(\hat{\hat{\beta}}_k)\hat{\hat{\beta}}_k\}$$

2.3 การแจกแจงของตัวแปรอิสระ

ในการศึกษาครั้งนี้ได้ทำการศึกษาตัวแปรอิสระที่มีการแจกแจงในลักษณะต่างๆ ทั้งการแจกแจงแบบต่อเนื่องและการแจกแจงแบบไม่ต่อเนื่อง ซึ่งได้ทำการศึกษาทั้งสิ้น 3 การแจกแจง ดังนี้

2.3.1 การแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution)

ในการวิเคราะห์สถิติแบบใช้พารามิเตอร์เกือบทุกเทคนิคต้องมีเงื่อนไขหรือข้อสมมติเกี่ยวกับการแจกแจงของข้อมูลว่าต้องมาจากการแจกแจงแบบปกติซึ่งถ้าข้อมูลมีหลายตัวแปร

การแจกแจงแบบปกติหลายตัวแปร[3] จะถูกนำมาพิจารณา โดยให้เวกเตอร์ของตัวแปร x หรือ $\tilde{x}' = (x_1, x_2, ..., x_p)$ มีการแจกแจงแบบปกติที่มีเวกเตอร์ของค่าเฉลี่ยเป็น μ และเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix) คือ Σ โดยที่ฟังก์ชันความหนาแน่น[3] เป็นดังนี้

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

ในการศึกษาครั้งนี้ได้ทำการศึกษาตัวแปรอิสระที่มีการแจกแจงแบบปกติหลายตัวแปรที่มีพารามิเตอร์ค่าเฉลี่ยคือ $\tilde{\mu} = (0,0,...,0)$ และมีเมทริกซ์ความแปรปรวนร่วมตามระดับความสัมพันธ์ของตัวแปรอิสระที่กำหนดไว้

2.3.2 การแจกแจงแบบแกมมา (Gamma Distribution)

(Furman E. (2008))[4] ข้อมูลที่มีการแจกแจงแบบแกมมาส่วนใหญ่จะเกี่ยวข้องกับข้อมูลด้านเวลา และการรอคอย ซึ่งสามารถเขียนฟังก์ชันความหนาแน่นได้ ดังนี้

$$f(y) \quad = \quad e^{-\alpha y} \frac{y^{\gamma-1}\alpha^{y}}{\tau(\gamma)} \qquad ; y > 0, \gamma > 0, \alpha > 0$$

โดย γ คือ Shape Parameter และ α คือ Rate Parameter

แต่เนื่องจากในการศึกษาครั้งนี้ต้องการศึกษาเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระจึงต้องกำหนดให้ ตัวแปรอิสระมีความสัมพันธ์กันเพราะฉะนั้นจึงต้องทำการพิจารณาการแจกแจงแกมมาหลายตัวแปร (Multivariate Gamma) ที่มีความสัมพันธ์กัน โดย

ให้ $\tilde{Y} \quad = \quad (Y_0, Y_1, Y_2, .., Y_n)^T \qquad \tilde{Y}_i \sim Ga(\gamma_i, \alpha_i) \quad ; i = 0,1,2,...,n$

ซึ่งเป็นอิสระกัน โดยที่ การแจกแจงแบบแกมมาหลายตัวแปรที่มีความสัมพันธ์กันจะให้เป็น

---

[3]กัลยา วานิชย์บัญชา, การวิเคราะห์ข้อมูลหลายตัวแปร, พิมพ์ครั้งที่ 4 (กรุงเทพมหานคร: บริษัทธรรมสาร, 2552), 424 – 484.

[4]Furman E., “On a multivariate gamma distribution,” Statistics and Probability Letters (2008).

$\tilde{X} \sim MG(\bar{\gamma},\alpha)$ โดย $\tilde{X} = AY$

$$\bar{\gamma} = (\gamma_0+\gamma_1, \gamma_0+\gamma_1+\gamma_2, ..., \gamma_0+\gamma_1+...+\gamma_n)^T$$

$$\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)^T$$

$$A = \begin{bmatrix} \alpha_0/\alpha_1 & 1 & 0 & 0 & \cdots & 0 \\ \alpha_0/\alpha_2 & \alpha_1/\alpha_2 & 1 & 0 & \cdots & 0 \\ \alpha_0/\alpha_3 & \alpha_1/\alpha_3 & \alpha_2/\alpha_3 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_0/\alpha_n & \alpha_1/\alpha_n & \alpha_2/\alpha_n & \alpha_3/\alpha_n & \cdots & 1 \end{bmatrix}$$

โดยที่

$$E(X_j) = \frac{\bar{\gamma}_j}{\alpha_j}$$

$$Var(X_j) = \frac{\bar{\gamma}_j}{\alpha_j^2}$$

เมื่อ $i < j$ จะได้

$$Cov(X_i, X_j) = Cov(X_i, X_i) = Var(X_i) = \frac{\bar{\gamma}_j}{\alpha_j^2}$$

เพราะฉะนั้น เมทริกซ์ความแปรปรวนร่วมคือ

$$\Sigma = \begin{bmatrix} \frac{\bar{\gamma}_1}{\alpha_1^2} & \frac{\bar{\gamma}_1}{\alpha_1^2} & \cdots & \frac{\bar{\gamma}_1}{\alpha_1^2} \\ \frac{\bar{\gamma}_1}{\alpha_1^2} & \frac{\bar{\gamma}_2}{\alpha_2^2} & \cdots & \frac{\bar{\gamma}_2}{\alpha_2^2} \\ \vdots & \vdots & \ddots & \\ \frac{\bar{\gamma}_1}{\alpha_1^2} & \frac{\bar{\gamma}_2}{\alpha_2^2} & \cdots & \frac{\bar{\gamma}_n}{\alpha_n^2} \end{bmatrix}$$

ระดับความสัมพันธ์ (correlation) ระหว่าง (X_i, X_j) ; i < j คือ

$$\rho_{ij} = \mathrm{Corr}(X_i, X_j) = \sqrt{\frac{\mathrm{Var}(X_i)}{\mathrm{Var}(X_j)}}$$

2.3.3 **การแจกแจงแบบปัวส์ซอง** (Poisson Distribution)

เป็นการแจกแจงที่อธิบายถึงจำนวนครั้งของเหตุการณ์หรือจำนวนสิ่งที่สนใจที่เกิดขึ้นในช่วงเวลาที่กำหนดหรือในพื้นที่ที่กำหนด โดยให้ Y คือ จำนวนเหตุการณ์หรือจำนวนสิ่งที่สนใจที่เกิดขึ้นในช่วงเวลาหนึ่ง หรือในพื้นที่หนึ่ง และมีพารามิเตอร์คือ θ โดยที่ฟังก์ชันความน่าจะเป็น เป็นดังนี้

$$p(y) = \frac{e^{-\theta}\theta^{y}}{y!} \qquad ; y = 0,1,2,\ldots\ , \theta > 0$$

โดย

$$E(Y) = \theta, \quad \mathrm{Var}(Y) = \theta$$

แต่เนื่องจากในการศึกษาครั้งนี้ต้องการศึกษาเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ จึงต้องกำหนดให้ ตัวแปรอิสระมีความสัมพันธ์กันเพราะฉะนั้นจึงต้องทำการพิจารณาการแจกแจงปัวซองส์หลายตัวแปร[5] (Multivariate Poisson) ที่มีความสัมพันธ์กัน โดย

ให้ $\tilde{Y} = (Y_1, Y_2, .., Y_m)'$ และ $Y_i \sim \mathrm{Poi}(\theta_i)$ $; i = 1,2,...,m$ ซึ่งเป็นอิสระกัน ทั้งนี้การแจกแจงแบบปัวซองส์หลายตัวแปรที่มีความสัมพันธ์กันจะให้เป็น

$\tilde{X} = (X_1, X_2, .., X_k)'$ โดย $\tilde{X}$ = AY

โดยที่ $A = [A_1 \quad A_2 \quad \cdots \quad A_k]$

A_i คือ เมทริกซ์ที่มีมิติเท่ากับ $k \times \binom{k}{i}$ ซึ่งแต่ละคอลัมน์ประกอบไปด้วย 1 จำนวน i ค่า และ 0 จำนวน k-i ค่า และ ไม่มีคอลัมน์ใดซ้ำกัน

[5]Dimitris Karlis, Multivariate Poisson models (2002).

<u>ยกตัวอย่าง ถ้า k=3</u>

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad A_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

โดยมีพารามิเตอร์คือ

ดังนั้น $$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$\theta_i \,; i \in (\{1\},\{2\},\{3\})$ และ $\theta_{ij} \,; i \in (\{12\},\{13\},\{23\},\{123\})$

โดยที่ θ_{ij} คือ ความแปรปรวนร่วมระหว่าง X_i และ X_j และ θ_{123} คือ ความแปรปรวนร่วมระหว่างตัวแปรอิสระทั้งสามตัว แต่ในที่นี้จะทำการพิจารณาเฉพาะความแปรปรวนร่วมสองทาง θ_{ij} เท่านั้น ดังนั้น เมทริกซ์ A ที่นำมาพิจารณาจะเป็นดังนี้

$$A = [A_1 \quad A_2]$$

ในกรณีที่ k = 3

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$X_1 = Y_1 + Y_{12} + Y_{13}$$

$$X_2 = Y_2 + Y_{12} + Y_{23}$$

$$X_3 = Y_3 + Y_{13} + Y_{23}$$

โดยที่ Y_i's เป็นตัวแปรสุ่มที่มีการแจกแจงปัวซองส์และเป็นอิสระกัน ซึ่งมีพารามิเตอร์คือ $\theta_i \,; i \in (\{1\},\{2\},\{3\},\{12\},\{13\},\{23\})$ เพราะฉะนั้นจะได้เมทริกซ์ความแปรปรวนร่วมของ (X_1, X_2, X_3) เป็นดังนี้

$$\mathrm{Var}(X) = \begin{bmatrix} \theta_1 + \theta_{12} + \theta_{13} & \theta_{12} & \theta_{13} \\ \theta_{12} & \theta_2 + \theta_{12} + \theta_{23} & \theta_{23} \\ \theta_{13} & \theta_{23} & \theta_3 + \theta_{13} + \theta_{23} \end{bmatrix}$$

สำหรับกรณีทั่วๆ ไป ค่าเฉลี่ย และ ความแปรปรวนเป็นดังนี้

$$E(X) = AM$$

$$Var(X) = A\Sigma A^T$$

โดยที่ M คือ เวกเตอร์ของค่าเฉลี่ย และ Σ เป็นเมทริกซ์ความแปรปรวนร่วมสำหรับตัวแปร Y_i

Σ เป็นเมทริกซ์ทแยง (diagonal matrix) = $\mathrm{diag}(\theta_1, \theta_2, ..., \theta_m)$

2.4 สถิติที่ใช้เพื่อการตัดสินใจ

ในการศึกษานี้มี 3 ขั้นตอนที่ต้องทำการตัดสินใจเพื่อให้ได้ค่าพารามิเตอร์ริดจ์ที่เหมาะสมในการแก้ปัญหาพหุสัมพันธ์ของตัวแปรอิสระ

2.4.1 <u>ขั้นตอนที่ 1</u>: การประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธี Newton Raphson

วิธีการของ Newton Raphson เป็นวิธีการแก้สมการที่ไม่เป็นเชิงเส้น (nonlinear equations) แบบมีการย้อนซ้ำ (iterative) และใช้ร่วมกับการประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด โดยเริ่มจากการกำหนดค่าพารามิเตอร์เริ่มต้น แล้วจึงทำการย้อนซ้ำโดยการหาอนุพันธ์เทียบกับค่าพารามิเตอร์ของฟังก์ชันความควรจะเป็นสูงสุด จนกระทั่งได้ค่าประมาณที่ลู่เข้าหาศูนย์ จึงเสร็จสิ้นกระบวนการประมาณค่าพารามิเตอร์ ซึ่งในการศึกษาครั้งนี้ได้ใช้เกณฑ์ในการหยุดประมาณค่าพารามิเตอร์เมื่อ

$$\left|\hat{\hat{\beta}}_{k+1} - \hat{\hat{\beta}}_k\right| < 0.0000001$$

โดย $\hat{\hat{\beta}}_{k+1}$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยที่คำนวณได้จากรอบที่ k+1

$\hat{\hat{\beta}}_k$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยที่คำนวณได้จากรอบที่ k

2.4.2 <u>ขั้นตอนที่ 2</u>: การตัดสินใจเลือกค่าพารามิเตอร์ริดจ์ (λ) ที่เหมาะสม (เกณฑ์ MAPE)

ตัวประมาณที่ได้จากการประมาณค่าด้วยการวิเคราะห์ความถดถอยแบบริดจ์จะเป็นตัวประมาณที่เอนเอียงแต่ถึงอย่างไรก็ตามยังคงให้ค่าสัมประสิทธิ์ความถดถอยที่ดีกว่าวิธีกำลังสองน้อยสุด (OLS) และค่าพารามิเตอร์ริดจ์ (λ) เมื่อมีค่าเพิ่มสูงขึ้นยิ่งทำให้เกิด

ความเอนเอียงมากขึ้น ดังนั้น การเลือกค่าพารามิเตอร์ริดจ์จึงควรเลือกค่าที่น้อยที่สุดที่ทำให้ค่าประมาณสัมประสิทธิ์ความถดถอยที่ได้กับค่าสัมประสิทธิ์ความถดถอยที่กำหนดขึ้นมีค่าต่างกันน้อยในระดับที่ผู้ทำการศึกษายอมรับได้ โดยตัวสถิติที่นำมาใช้คือ MAPE (Mean Absolute Percentage Error) ซึ่งสูตรที่ใช้ในการคำนวณคือ

$$\text{MAPE} = \left(\frac{1}{p+1} \sum_{i=0}^{p} \left| \frac{\beta_i - \hat{\beta}_i}{\beta_i} \right| \right) \times 100 \quad ; i = 0, 1, \ldots, p$$

โดยที่ p คือ จำนวนตัวแปรอิสระ

β_i คือ ค่าสัมประสิทธิ์ความถดถอยที่กำหนดขึ้นตัวที่ i

$\hat{\beta}_i$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยตัวที่ i

2.4.3 <u>ขั้นตอนที่ 3</u>: การตัดสินใจเลือกค่าพารามิเตอร์ริดจ์ (λ) ที่เหมาะสม (เกณฑ์ SD)

ค่าพารามิเตอร์ริดจ์ที่ผ่านเกณฑ์ในขั้นตอนที่ 1 และขั้นตอนที่ 2 ถือเป็นค่าพารามิเตอร์ที่สามารถแก้ปัญหาพหุสัมพันธ์ได้ แต่เนื่องจากการจำลองข้อมูลไม่ได้มีการทำซ้ำหลายๆ รอบ จึงมีการตั้งเกณฑ์ในขั้นตอนที่ *3* ขึ้น เพื่อทำการพิจารณาว่า ค่าพารามิเตอร์ริดจ์ที่ได้ไม่มีความแตกต่างกันมากนัก ซึ่งแสดงถึงความน่าเชื่อถือของค่าพารามิเตอร์ริดจ์ในกรณีนั้นๆ ดังนั้นเกณฑ์ที่ใช้ในการพิจารณาในขั้นตอนที่ 3 คือ ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ซึ่งสามารถคำนวณได้จากสูตร

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^{r} (\lambda_i - \bar{\lambda})^2}{r-1}}$$

โดยที่ SD คือ ค่าส่วนเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ริดจ์

r คือ จำนวนรอบที่ทำการวิเคราะห์ซึ่งนับจำนวนรอบตามจำนวนค่าพารามิเตอร์ริดจ์ที่ผ่านเกณฑ์ MAPE

λ_i คือ ค่าพารามิเตอร์ริดจ์ตัวที่ i ที่ผ่านเกณฑ์ตามเงื่อนไข MAPE

$\bar{\lambda}$ คือ ค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์จนถึงรอบที่ r

บทที่ 3

วิธีดำเนินการศึกษา

การศึกษาครั้งนี้มีเป้าหมายเพื่อหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมในกรณีต่างๆ ซึ่งดำเนินการด้วยการจำลองข้อมูลโดยใช้โปรแกรม R และทำการวิเคราะห์ผลบางส่วนโดยใช้โปรมแกรม SPSS for Windows ver.19 ซึ่งการศึกษาได้ดำเนินการภายใต้เงื่อนไขต่างๆ ดังนี้

3.1 ขอบเขตการศึกษา

1. ศึกษาการวิเคราะห์ความถดถอยโลจิสติกที่มีตัวแปรตาม (Y) มีค่าเพียง 2 ค่า (Binary Logistic Regression)
2. กำหนดตัวแปรอิสระ (X) เป็นตัวแปรเชิงปริมาณเท่านั้น และทำการศึกษาในกรณีที่มีตัวแปรอิสระ (p) จำนวน 2 และ 3 ตัว โดยที่ตัวแปรอิสระมีการแจกแจงแบบเดียวกันทุกตัว และมีความสัมพันธ์ระหว่างตัวแปรอิสระเป็นดังนี้

   2.1 กรณีตัวแปรอิสระ ตัว 2(p = 2)

   ให้ ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 2 โดยทำการศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

   2.1.1 ระดับความสัมพันธ์ 0.3

   2.1.2 ระดับความสัมพันธ์ 0.5

   2.1.3 ระดับความสัมพันธ์ 0.7

   2.1.4 ระดับความสัมพันธ์ 0.8

   2.2 กรณีตัวแปรอิสระ 3 ตัว (p = 3)

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

   ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 2

   ρ_{13} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 3

   ρ_{23} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 2 และตัวแปรอิสระตัวที่ 3

ทำการศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

2.2.1 ระดับความสัมพันธ์ (0.3, 0.5, 0.8)

2.2.2 ระดับความสัมพันธ์ (0.5, 0.5, 0.8)

2.2.3 ระดับความสัมพันธ์ (0.5, 0.7, 0.8)

2.2.4 ระดับความสัมพันธ์ (0.7, 0.7, 0.8)

2.2.5 ระดับความสัมพันธ์ (0.8, 0.8, 0.8)

ซึ่งทำการศึกษาภายใต้การแจกแจง 3 การแจกแจง ดังนี้

2.3 การแจกแจงแบบต่อเนื่อง

2.3.1 การแจกแจงปกติหลายตัวแปร (Multivariate Normal) ที่มีพารามิเตอร์ค่าเฉลี่ย $(\tilde{\mu})$ เป็น $\tilde{0}$ และเมทริกซ์ความแปรปรวนร่วม (Σ) มีความแปรปรวนร่วมระหว่างตัวแปรอิสระตามระดับความสัมพันธ์ของตัวแปรอิสระ ดังที่ได้กล่าวมา

2.3.2 การแจกแจงแกมมา (Gamma Distribution) ที่มีพารามิเตอร์คือ (γ,α) เท่ากับ (1,2) สำหรับตัวแปรอิสระตัวที่ 1 (γ,α) เท่ากับ (1.5,2) สำหรับตัวแปรอิสระตัวที่ 2 และ (γ,α) เท่ากับ (1,2) สำหรับตัวแปรอิสระตัวที่ 3 โดยทำการศึกษาเมื่อตัวแปรอิสระมีระดับความสัมพันธ์ตามที่กล่าวมา

2.4 การแจกแจงแบบไม่ต่อเนื่อง

2.4.1 การแจกแจงปัวซองส์ (Poisson Distribution) ที่มีพารามิเตอร์คือค่าเฉลี่ย (θ) เท่ากับ 2 สำหรับตัวแปรอิสระตัวที่ 1 และ (θ) เท่ากับ 3 สำหรับตัวแปรอิสระตัวที่ 2 โดยทำการศึกษาเมื่อตัวแปรอิสระมีระดับความสัมพันธ์ตามที่กล่าวมา

3. ศึกษาภายใต้ขนาดตัวอย่างที่มีจำนวนตัวอย่างอย่างน้อย 20 เท่าของจำนวนตัวแปรอิสระ (20p) คือ

3.1 ถ้าจำนวนตัวแปรอิสระเท่ากับ 2 จำนวนตัวอย่างที่ทำการศึกษาคือ 40 และ 70

3.2 ถ้าจำนวนตัวแปรอิสระเท่ากับ 3 จำนวนตัวอย่างที่ทำการศึกษาคือ 60 และ100

4. กำหนดค่าเริ่มต้นของสัมประสิทธิ์ความถดถอย $(\beta_0, \beta_1, ..., \beta_p)$ ดังนี้

   4.1 กรณีตัวแปรอิสระ 2 ตัว (p = 2) ค่าสัมประสิทธิ์ความถดถอยเริ่มต้นมีค่าเป็น $(\beta_0, \beta_1, \beta_2) \in [2,4]$

   4.2 กรณีตัวแปรอิสระ 3 ตัว (p = 3) ค่าสัมประสิทธิ์ความถดถอยเริ่มต้นมีค่าเป็น $(\beta_0, \beta_1, \beta_2, \beta_3) \in [2,4]$

5. จำลองค่า y_i โดยทำการพิจารณาจากตัวแปรสุ่มแบบยูนิฟอร์มที่มีพารามิเตอร์คือ 0 และ 1 โดย

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p}}$$

   โดยที่ $y_i = 1$ ก็ต่อเมื่อ ค่าของตัวแปรสุ่มแบบยูนิฟอร์มมีค่าน้อยกว่าหรือเท่ากับ $\pi(x_i)$

   และ $y_i = 0$ ก็ต่อเมื่อ ค่าของตัวแปรสุ่มแบบยูนิฟอร์มมีค่ามากกว่า $\pi(x_i)$

3.1 วิธีดำเนินการศึกษา

1. กำหนดค่าเริ่มต้น ดังนี้
   - ระดับความสัมพันธ์ของตัวแปรอิสระ (ρ)
   - จำนวนตัวแปรอิสระ (p)
   - ขนาดตัวอย่าง (n)
   - ค่าสัมประสิทธิ์ความถดถอยเริ่มต้น $(\tilde{\beta})$
2. จำลองข้อมูลตามการแจกแจงโดยให้มีระดับความสัมพันธ์ตามที่กำหนด
3. กำหนดค่าพารามิเตอร์ริดจ์เริ่มต้น ดังนี้

   การแจกแจงแบบปัวซองส์ สำหรับตัวแปรอิสระ 2 ตัว

   3.1 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.00001

3.2 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.0000001

3.3 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.000000001

การแจกแจงแบบปกติ สำหรับตัวแปรอิสระ 2 และ 3 ตัว

3.4 สำหรับค่าสัมประสิทธิ์ความถดถอยทุกค่า กำหนดให้ค่าพารามิเตอร์ริดจ์เริ่มต้น มีค่า 0.001

การแจกแจงแบบแกมมา สำหรับตัวแปรอิสระ 2 ตัว

3.5 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.0001

3.6 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 และ 4 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.0000001

การแจกแจงแบบแกมมา สำหรับตัวแปรอิสระ 3 ตัว

3.7 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.000001

3.8 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.0000001

3.9 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่าพารามิเตอร์ริดจ์เริ่มต้นมีค่า 0.00000001

4. คำนวณค่าความน่าจะเป็น $\pi(x_i)$
5. สร้างตัวแปรสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม โดยให้พารามิเตอร์มีค่า 0 และ 1
6. จำลองค่า y โดยแปลงค่าความน่าจะเป็นจากข้อ 4 ให้เป็นค่า 0 หรือ 1 โดย

   $y_i = 1$ ก็ต่อเมื่อ ค่าของตัวแปรสุ่มแบบยูนิฟอร์มมีค่าน้อยกว่าหรือเท่ากับ $\pi(x_i)$ และ

   $y_i = 0$ ก็ต่อเมื่อ ค่าของตัวแปรสุ่มแบบยูนิฟอร์มมีค่ามากกว่า $\pi(x_i)$

7. ประมาณค่าสัมประสิทธิ์ความถดถอยด้วยวิธี NR (Newton Raphson) โดยจะหยุดเมื่อ

$$\left|\hat{\hat{\beta}}_{k+1} - \hat{\hat{\beta}}_{k}\right| < 0.0000001$$

โดย $\hat{\hat{\beta}}_{k+1}$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยที่คำนวณได้จากรอบที่ k+1

$\hat{\hat{\beta}}_{k}$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยที่คำนวณได้จากรอบที่ k

8. คำนวณค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE)

$$\mathrm{MAPE} = \left(\frac{1}{p+1}\sum_{i=0}^{p}\left|\frac{\beta_i - \hat{\beta}_i}{\beta_i}\right|\right)\times 100 \quad ; i = 0,1,\ldots,p$$

โดย p คือ จำนวนตัวแปรอิสระ

β_i คือ ค่าสัมประสิทธิ์ความถดถอยที่กำหนดขึ้นตัวที่ i

$\hat{\beta}_i$ คือ ค่าประมาณสัมประสิทธิ์ความถดถอยตัวที่ i

พิจารณาค่า MAPE ตามเกณฑ์ดังนี้

สำหรับการแจกแจงแบบปกติและการแจกแจงแบบปัวซองส์ เมื่อตัวแปรอิสระมีจำนวน 2 และ 3 ตัว

8.1 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่า MAPE ที่กำหนดไว้คือ 10%

8.2 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่า MAPE ที่กำหนดไว้คือ 7%

8.3 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่า MAPE ที่กำหนดไว้คือ 5%

สำหรับการแจกแจงแบบแกมมา เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่า 2, 3 และ 4

8.4 ถ้าจำนวนตัวแปรอิสระมี 2 ตัว ค่า MAPE ที่กำหนดไว้คือ 10%

8.5 ถ้าจำนวนตัวแปรอิสระมี 3 ตัว ค่า MAPE ที่กำหนดไว้คือ 15%

ค่า MAPE ต้องน้อยกว่าเกณฑ์ที่กำหนดไว้ จึงสามารถดำเนินการในขั้นตอนต่อไปได้

9. คำนวณค่าส่วนเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ริดจ์ (SD)

$$SD = \sqrt{\frac{\sum_{j=1}^{r}\left(\lambda_i - \overline{\lambda}\right)^2}{r-1}}$$

โดยที่ SD คือ ค่าเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ริดจ์

r คือ จำนวนรอบที่ทำการวิเคราะห์ซึ่งนับจำนวนรอบตามจำนวนค่าพารามิเตอร์ริดจ์ที่ผ่านเกณฑ์ของ MAPE

λ_i คือ ค่าพารามิเตอร์ริดจ์ตัวที่ i ที่ผ่านเกณฑ์ของ MAPE

$\overline{\lambda}$ คือ ค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์จนถึงรอบที่ r

10. พิจารณาค่า SD ตามเกณฑ์ดังนี้

การแจกแจงแบบปัวซองส์ สำหรับตัวแปรอิสระ 2 ตัว

10.1 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่า SD ที่กำหนดไว้คือ 0.00001

10.2 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่า SD ที่กำหนดไว้คือ 0.0000001

10.3 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่า SD ที่กำหนดไว้คือ 0.000000001

การแจกแจงแบบปกติ สำหรับตัวแปรอิสระ 2 และ 3 ตัว

10.4 สำหรับค่าสัมประสิทธิ์ความถดถอยทุกค่า ค่า SD ที่กำหนดไว้คือ 0.001

การแจกแจงแบบแกมมา สำหรับตัวแปรอิสระ 2 ตัว

10.5 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่า SD ที่กำหนดไว้คือ 0.0001

10.6 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 และ 4 ค่า SD ที่กำหนดไว้คือ 0.0000001

การแจกแจงแบบแกมมา สำหรับตัวแปรอิสระ 3 ตัว

10.7 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ค่า SD ที่กำหนดไว้คือ 0.000001

10.8 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 ค่า SD ที่กำหนดไว้คือ 0.0000001

10.9 ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ค่า SD ที่กำหนดไว้คือ 0.00000001

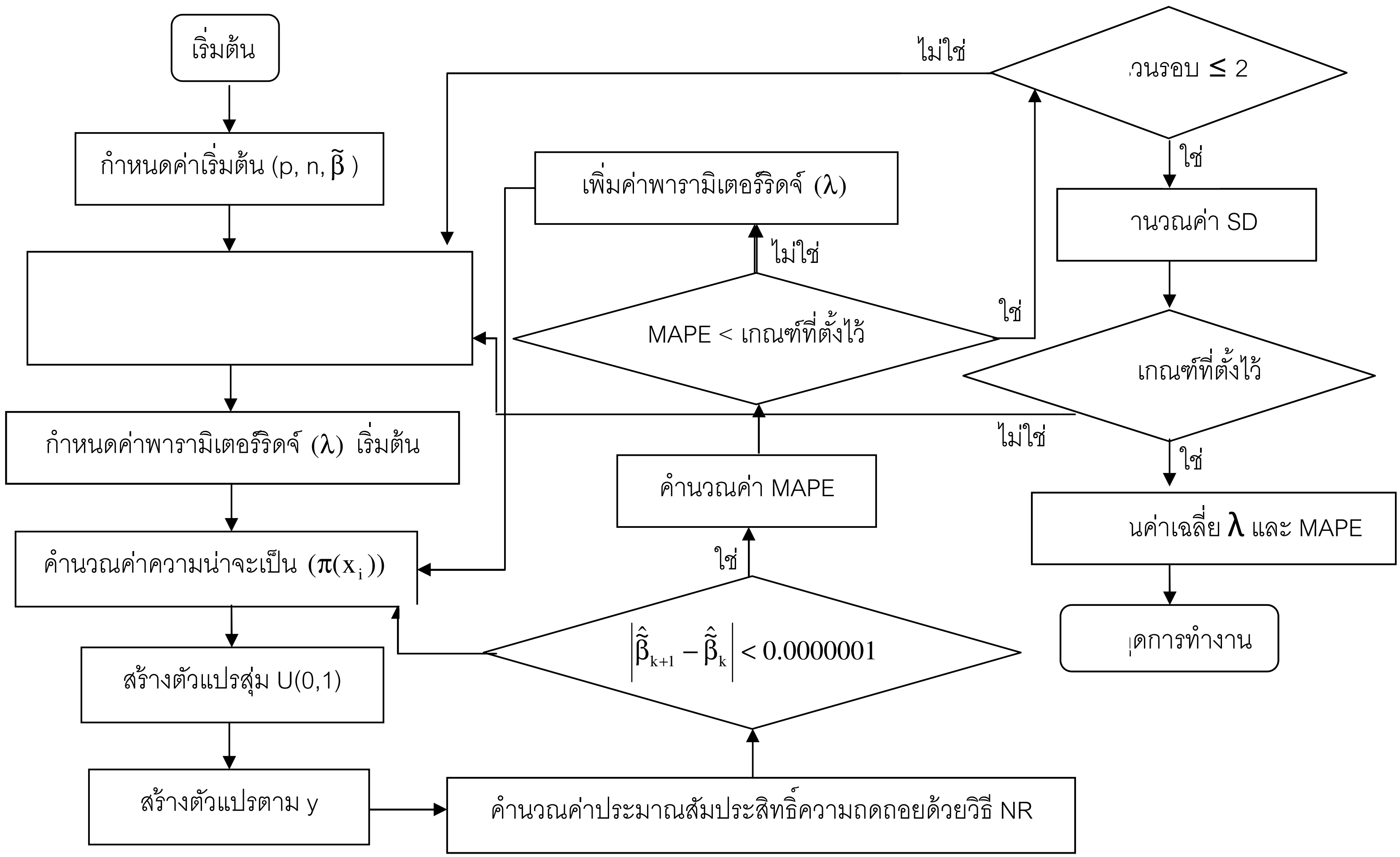
ค่า SDต้องน้อยกว่าเกณฑ์ที่กำหนดไว้ จึงสามารถดำเนินการในขั้นตอนต่อไปได้

11. คำนวณค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์และค่า MAPE

12. สรุปผลการศึกษา

จากขั้นตอนการศึกษาด้านบนสามารถเขียนแผนผังการดำเนินงานได้ดังนี้

แผนผังการเขียนโปรแกรมในแต่ละกรณีที่ทำการศึกษา



บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมในแต่ละกรณีของการศึกษาซึ่งขึ้นกับระดับความสัมพันธ์ จำนวนตัวแปรอิสระ การแจกแจงของตัวแปรอิสระ ขนาดตัวอย่าง และค่าพารามิเตอร์เริ่มต้น ซึ่งทำการศึกษาภายใต้การวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่ม (Binary Logistic Regression) โดยทำการจำลองข้อมูลต่างๆ ที่ใช้สำหรับการศึกษาครั้งนี้ด้วยโปรแกรม R

การวิเคราะห์ผลการศึกษาดำเนินการภายใต้กรณีต่างๆ ที่สนใจ ดังนี้

1. ในกรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัว มีทั้งสิ้น 72 กรณี

ค่าเริ่มต้นของสัมประสิทธิ์ความถดถอยโลจิสติก ($\tilde{\beta}'$)	การแจกแจงของตัวแปรอิสระ					
	Poisson		Normal		Gamma	
	n = 40	n = 70	n = 40	n = 70	n = 40	n = 70
(2,2,2)	ρ = 0.3, 0.5, 0.7, 0.8					
(3,3,3)						
(4,4,4)						

2. ในกรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัว มีทั้งสิ้น 60 กรณี

ค่าเริ่มต้นของสัมประสิทธิ์ความถดถอยโลจิสติก ($\tilde{\beta}$)	การแจกแจงของตัวแปรอิสระ			
	Normal		Gamma	
	n = 60	n = 100	n = 60	n = 100
(2,2,2,2)	ρ = (0.3, 0.5, 0.8), (0.5, 0.5, 0.8), (0.5, 0.7, 0.8), (0.7, 0.7, 0.8), (0.8, 0.8, 0.8)			
(3,3,3,3)				
(4,4,4,4)				

ดังนั้นกรณีที่ทำการศึกษามีทั้งสิ้น 132 กรณี ทำการวิเคราะห์ผลในแง่มุมต่างๆ ดังนี้

1. แสดงผลค่าพารามิเตอร์ริดจ์ในแต่ละกรณีที่ทำการศึกษา
2. ทดสอบค่าพารามิเตอร์ริดจ์ที่มีค่าน้อยใกล้ศูนย์ว่ามีค่าแตกต่างจากศูนย์หรือไม่
3. ทดสอบว่าการแจกแจงของตัวแปรอิสระส่งผลต่อค่าพารามิเตอร์ริดจ์หรือไม่
4. ทดสอบว่าค่าสัมประสิทธิ์เริ่มต้นส่งผลต่อค่าพารามิเตอร์ริดจ์หรือไม่
5. แสดงผลเปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้น
6. แสดงผลเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์สูงขึ้นในแต่ละขนาดตัวอย่างเทียบกับระดับความสัมพันธ์ 0.3
7. แสดงผลเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 40 เป็น 70

ผลการวิเคราะห์เป็นดังนี้

4.1 ผลการศึกษาในกรณีจำนวนตัวแปรอิสระ 2 ตัว

แสดงผลค่าพารามิเตอร์ริดจ์ในแต่ละกรณีที่ทำการศึกษา ได้ผลการวิเคราะห์ดังตารางด้านล่าง

ตารางที่ 4.1.1 ค่าพารามิเตอร์ริดจ์ในกรณีจำนวนตัวแปรอิสระ 2 ตัว จำแนกตามการแจกแจง

ค่าสัมประสิทธิ์ความถดถอย	**ρ**	n	Poisson		Normal		Gamma	
			λ	MAPE (%)	**λ**	MAPE (%)	**λ**	MAPE (%)
2	0.3	40	0.000315	9.895	0.0250	9.481	0.0004	9.875
		70	0.000320	9.710	0.0700	8.786	0.0007	9.304
	0.5	40	0.000320	9.830	0.0600	8.119	0.001	8.589
		70	0.000330	9.750	0.0860	7.246	0.0025	8.529
	0.7	40	0.000335	9.736	0.0850	8.975	0.002	9.288
		70	0.000365	9.650	0.1020	6.182	0.005	9.139
	0.8	40	0.000345	9.863	0.1150	9.743	0.0035	9.911

<u>ตารางที่ 4.1.1</u> (ต่อ) ค่าพารามิเตอร์ริดจ์ในกรณีจำนวนตัวแปรอิสระ 2 ตัว จำแนกตามการแจกแจง

ค่าสัมประสิทธิ์ความถดถอย	**ρ**	n	Poisson		Normal		Gamma	
			λ	MAPE (%)	**λ**	MAPE (%)	**λ**	MAPE (%)
2	0.8	70	0.000365	9.785	0.1385	7.663	0.006	9.838
3	0.3	40	0.0000103	6.966	0.02975	6.524	0.0000007	9.600
		70	0.000011	6.732	0.037	5.416	0.0000080	9.481
	0.5	40	0.0000108	6.974	0.052	6.847	0.0000175	9.077
		70	0.0000115	6.250	0.0575	6.730	0.00003	8.897
	0.7	40	0.000011	6.526	0.059	6.820	0.000035	9.211
		70	0.00001151	6.126	0.062	6.353	0.00005	7.959
	0.8	40	0.00001113	6.572	0.0675	6.833	0.000045	9.671
		70	0.0000116	6.305	0.07	6.694	0.00007	9.581
4	0.3	40	4.0155E-07	4.999	0.0135	4.446	0.0000004	9.829
		70	0.00000041	4.838	0.021	3.641	0.0000005	8.920
	0.5	40	4.016E-07	5.000	0.0185	4.598	0.0000015	9.685
		70	0.000000415	4.888	0.044	4.574	0.000003	9.429
	0.7	40	0.000000415	4.966	0.0235	4.691	0.000003	9.234
		70	0.000000418	4.893	0.05067	3.870	0.0000045	8.668
	0.8	40	0.00000043	4.547	0.0245	4.237	0.000006	9.498
		70	0.000000451	4.187	0.052	3.198	0.000007	7.642

โดยที่ ρ คือ ระดับความสัมพันธ์ของตัวแปรอิสระในที่นี้ตัวแปรอิสระมี 2 ตัว จึงเป็นระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 กับตัวแปรอิสระตัวที่ 2

n คือ ขนาดตัวอย่างที่ทำการศึกษา

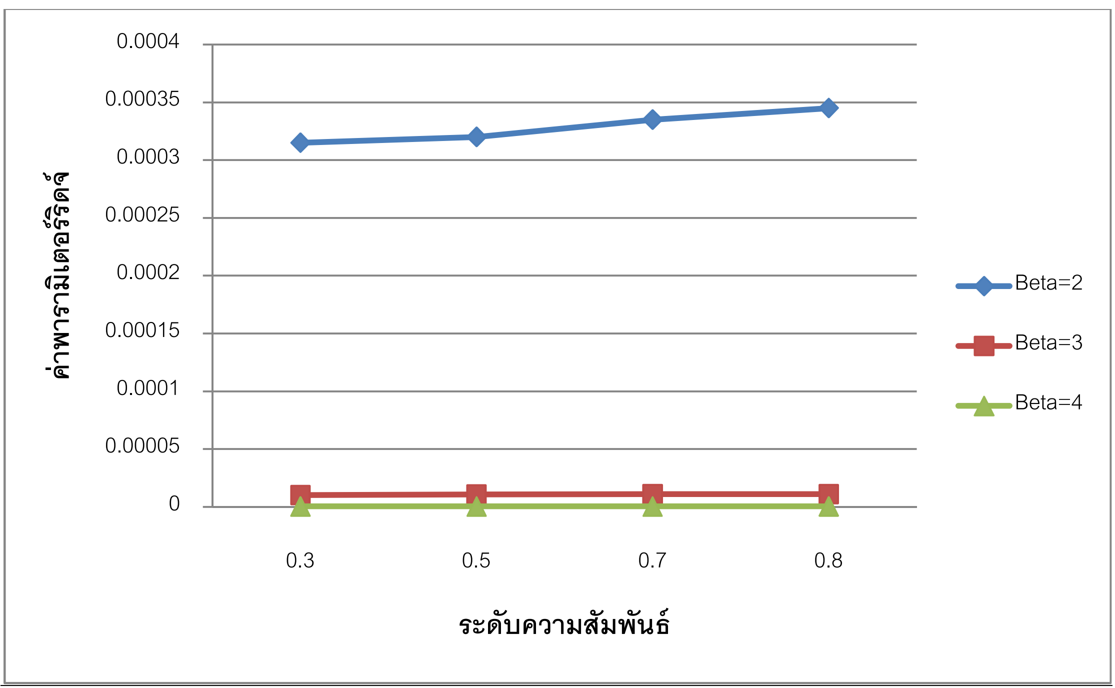
λ คือ ค่าพารามิเตอร์ริดจ์ที่เหมาะสม

MAPE คือ ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งกำหนดไว้ดังนี้

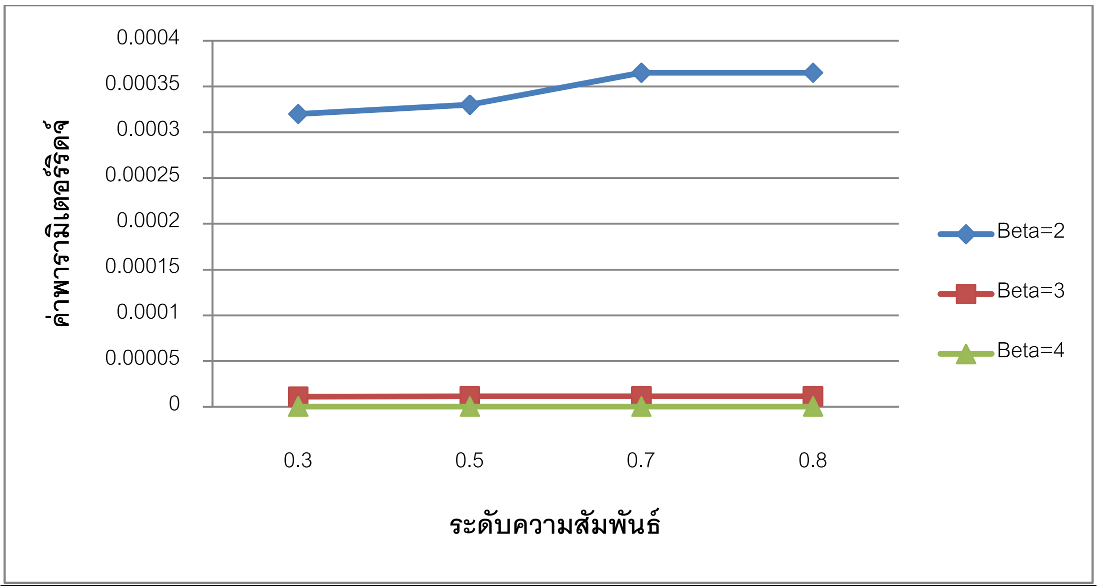
- สำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 ให้ MAPE มีค่าไม่เกิน 10%
- สำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 3 ให้ MAPE มีค่าไม่เกิน 7%
- สำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 4 ให้ MAPE มีค่าไม่เกิน 5%

ยกเว้นการแจกแจงแกมมา ให้ MAPE มีค่าไม่เกิน 10% ทุกค่าสัมประสิทธิ์ความถดถอย

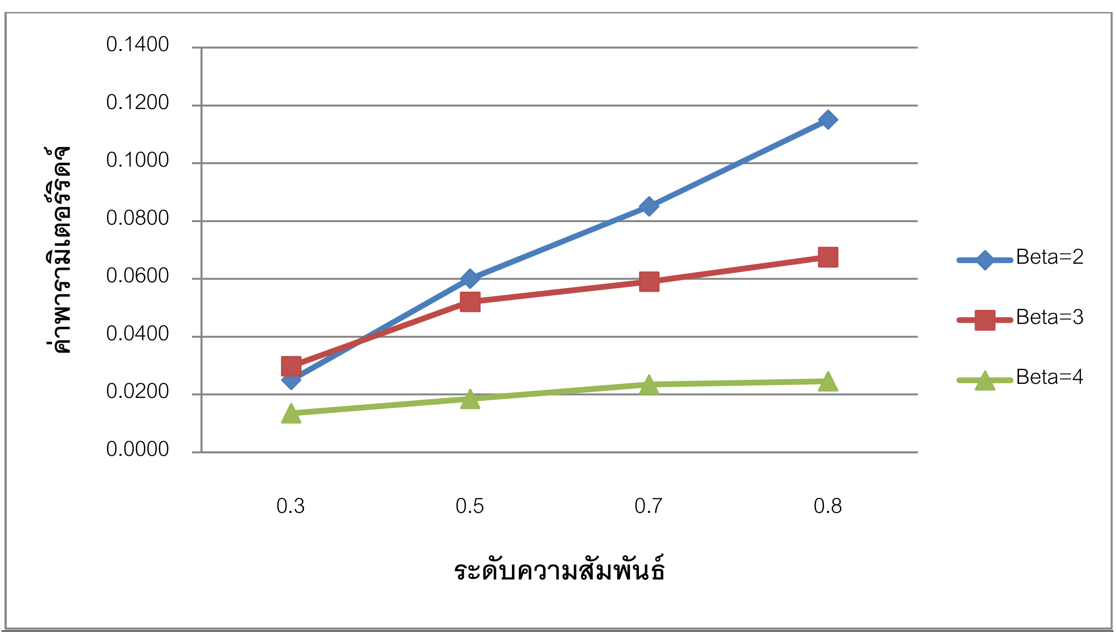
รูปที่ 4.1.1 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงปัวซองส์เมื่อขนาดตัวอย่างเท่ากับ 40



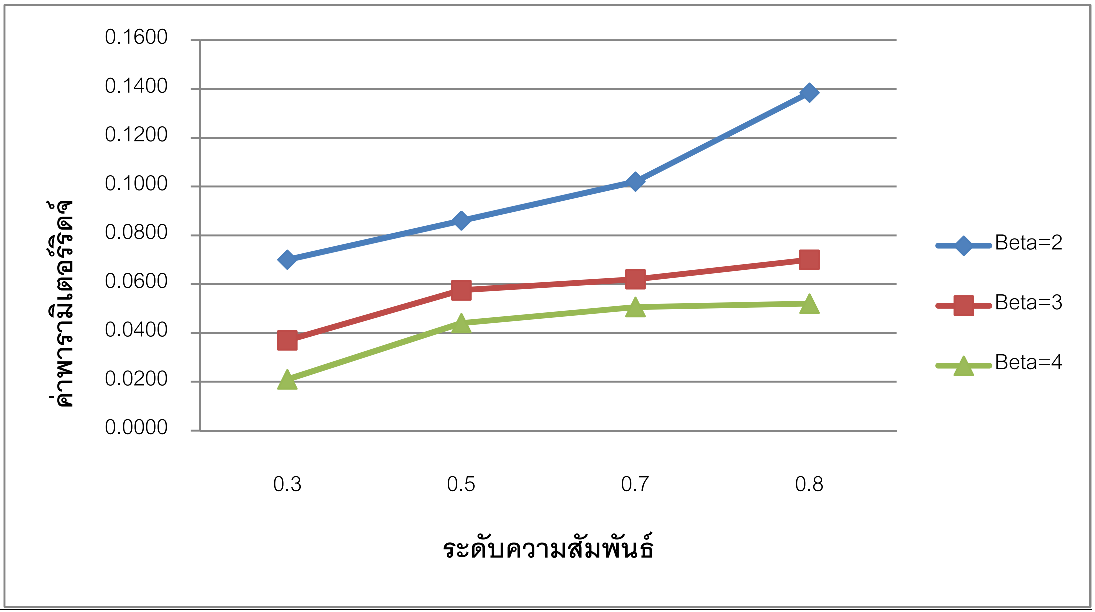
รูปที่ 4.1.2 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงปัวซองส์เมื่อขนาดตัวอย่างเท่ากับ 70



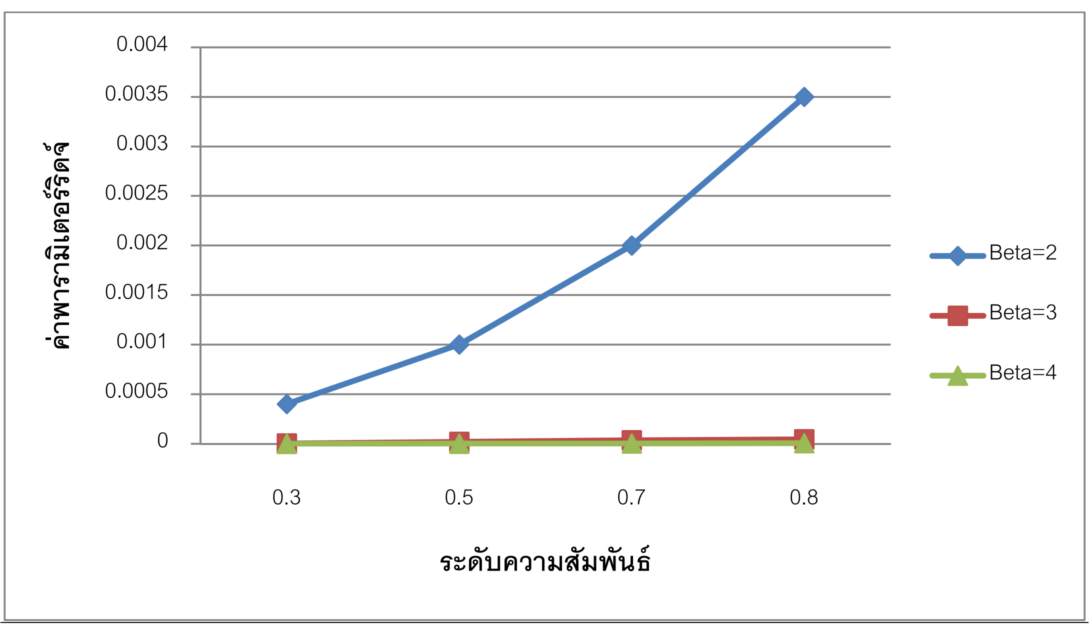
รูปที่ 4.1.3 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงปกติเมื่อขนาดตัวอย่างเท่ากับ 40



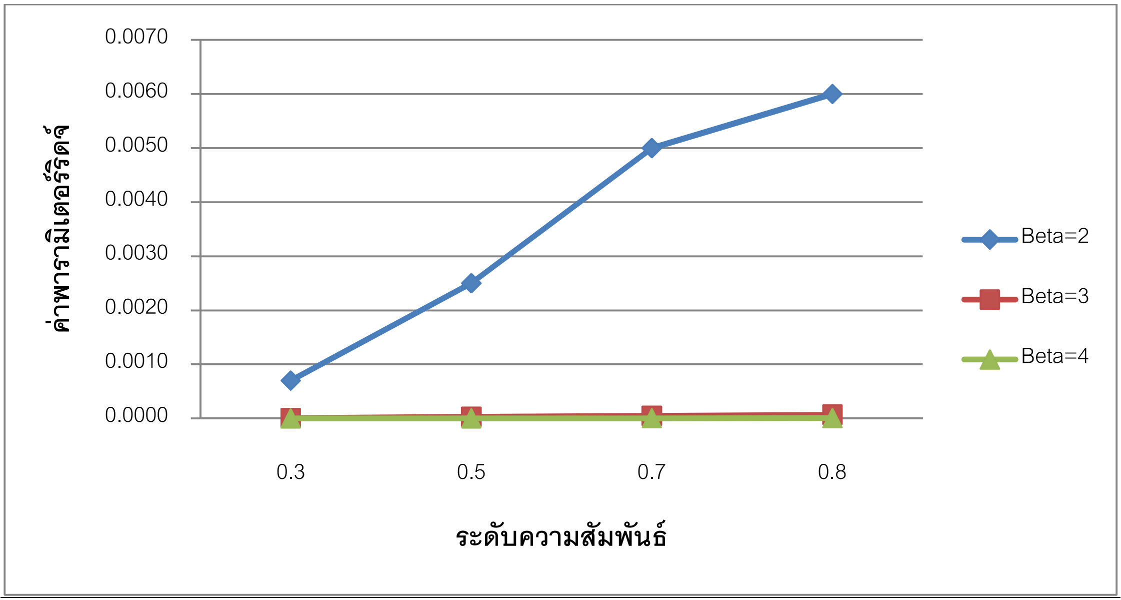
รูปที่ 4.1.4 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงปกติเมื่อขนาดตัวอย่างเท่ากับ 70



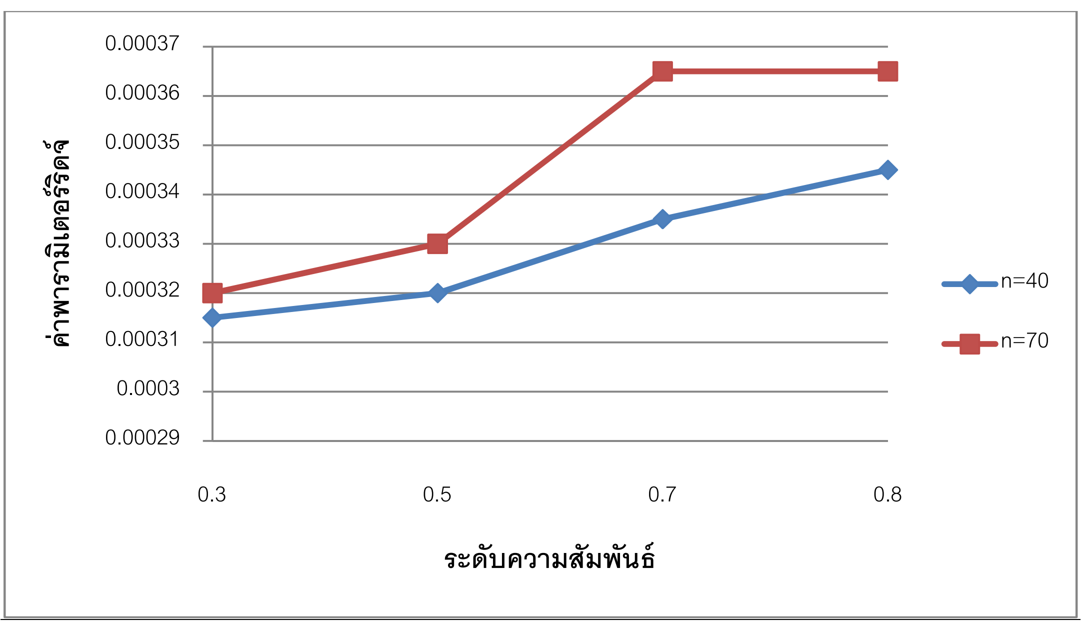
รูปที่ 4.1.5 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงแกมมาเมื่อขนาดตัวอย่างเท่ากับ 40



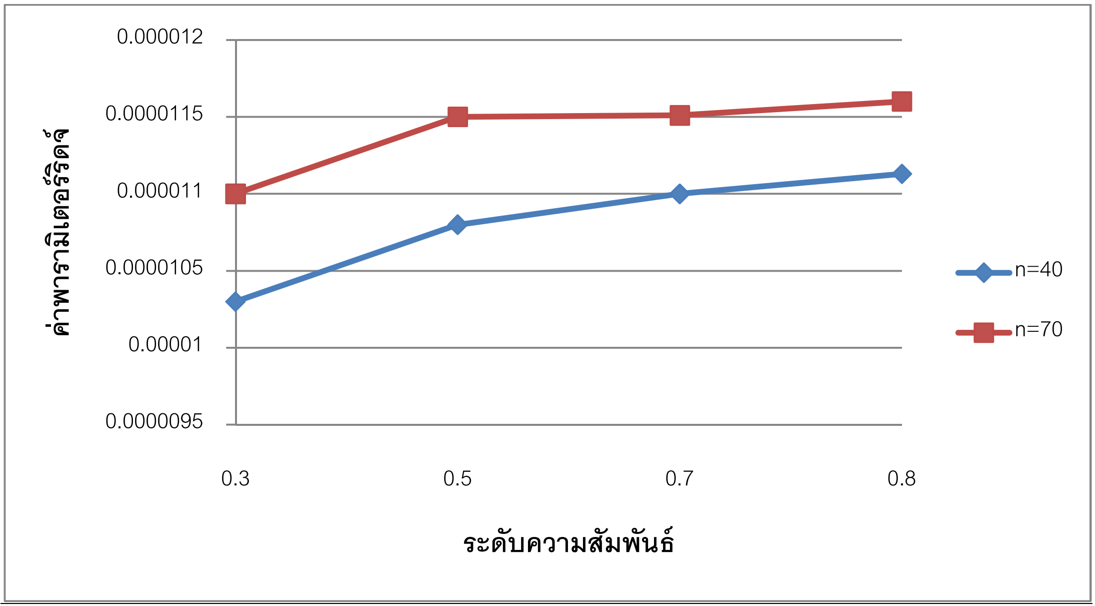
รูปที่ 4.1.6 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงแกมมาเมื่อขนาดตัวอย่างเท่ากับ 70



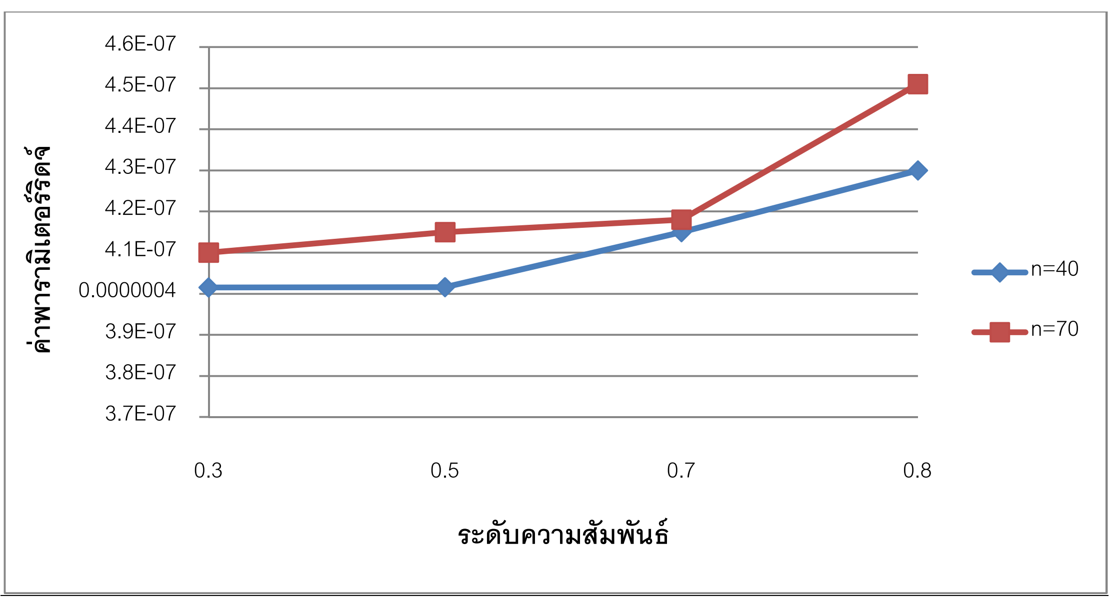
รูปที่ 4.1.7 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปัวซองส์เมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 2



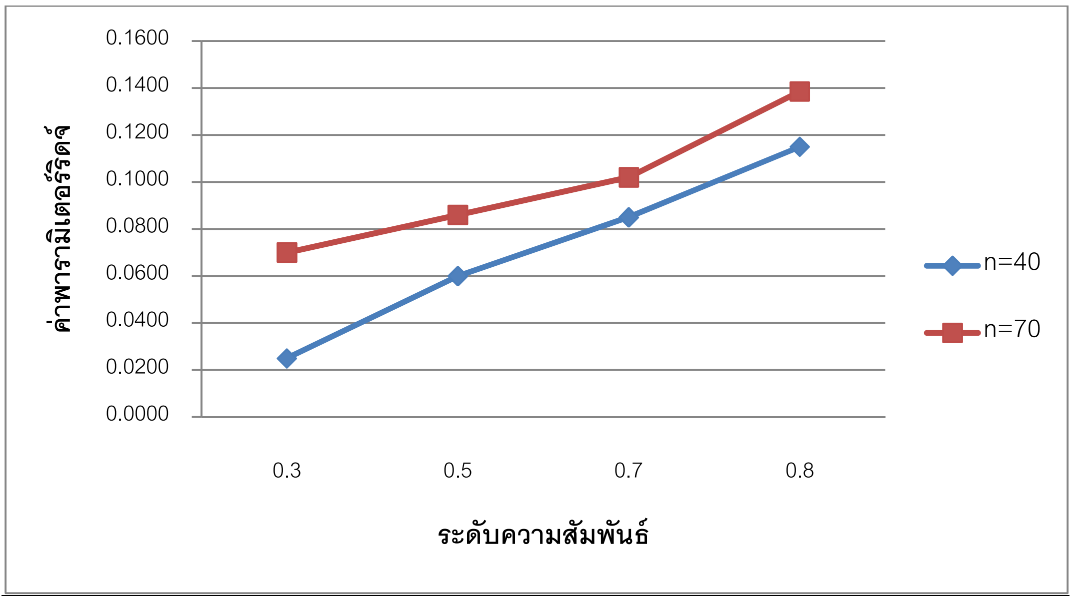
รูปที่ 4.1.8 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปัวซองส์เมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 3



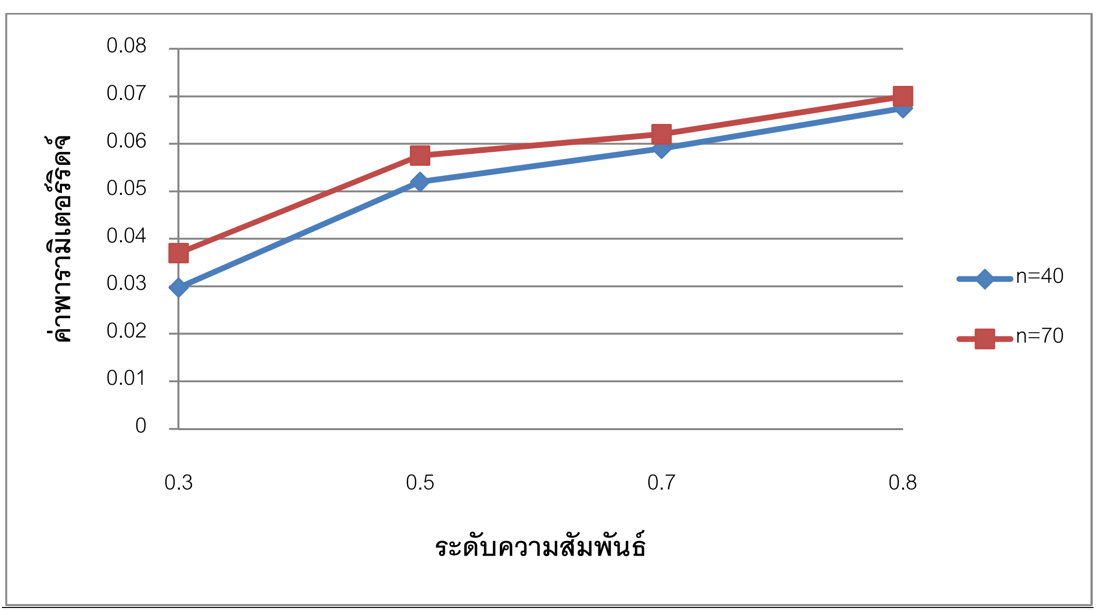
รูปที่ 4.1.9 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปัวซองส์เมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 4



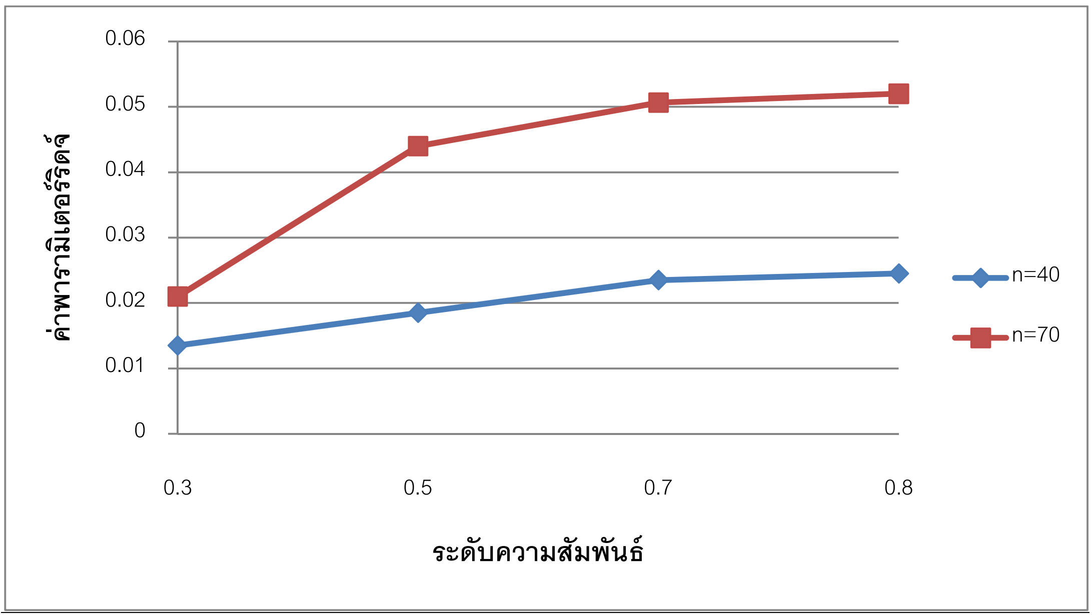
รูปที่ 4.1.10 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปกติเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 2



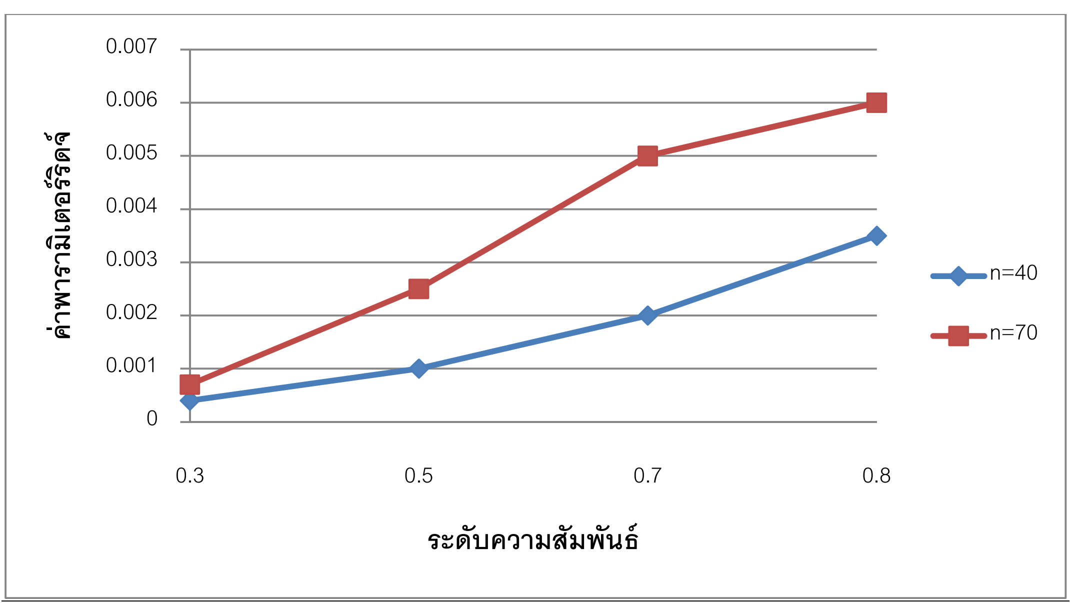
รูปที่ 4.1.11 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปกติเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 3



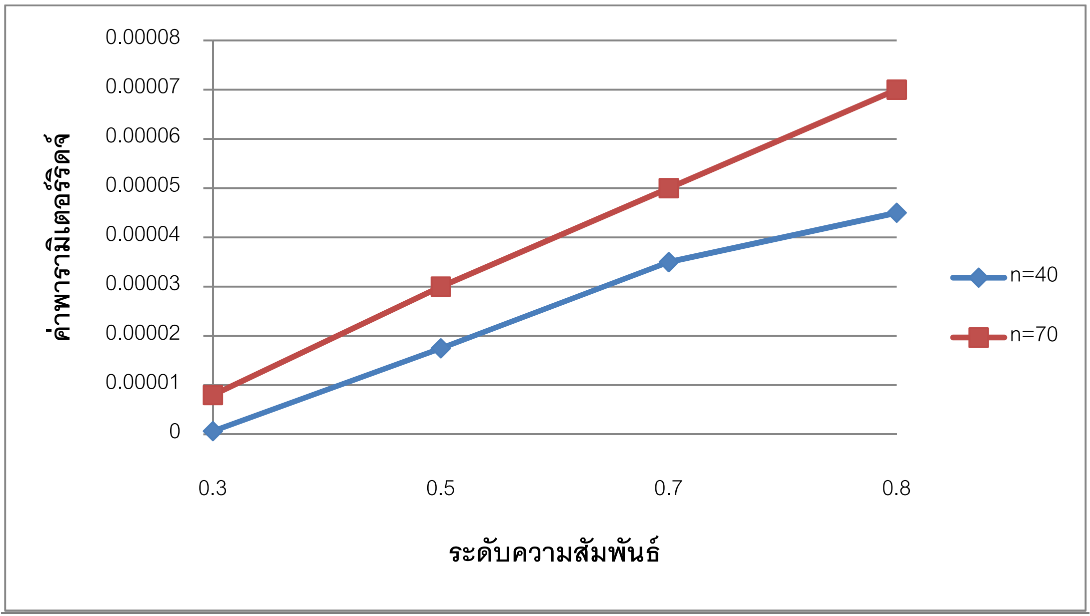
รูปที่ 4.1.12 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปกติเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 4



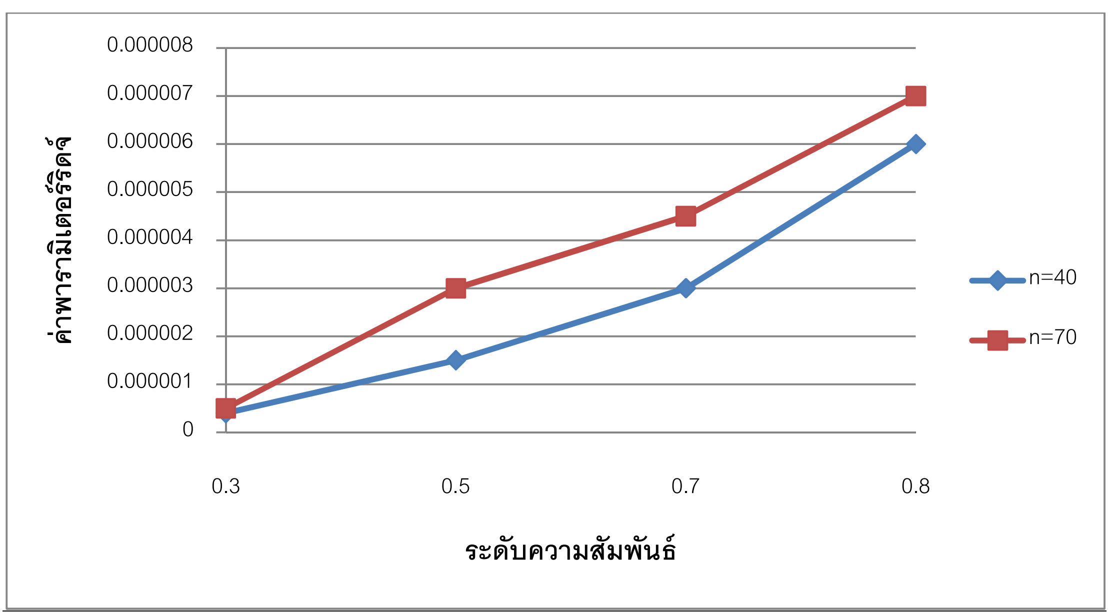
รูปที่ 4.1.13 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงแกมมาเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 2



รูปที่ 4.1.14 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงแกมมาเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 3



รูปที่ 4.1.15 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงแกมมาเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 4



จากตารางและกราฟด้านบนพบประเด็นที่น่าสนใจ ดังนี้

1. เมื่อขนาดตัวอย่างและค่าสัมประสิทธิ์ความถดถอยคงที่ ถ้าระดับความสัมพันธ์สูงขึ้น ค่าพารามิเตอร์ริดจ์จะมีค่าสูงขึ้น
2. เมื่อขนาดตัวอย่างและระดับความสัมพันธ์คงที่ ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้น ค่าพารามิเตอร์ริดจ์จะมีค่าลดลงในทุกการแจกแจงของตัวแปรอิสระ
3. เมื่อระดับความสัมพันธ์และค่าสัมประสิทธิ์ความถดถอยคงที่ ถ้าขนาดตัวอย่างเพิ่มขึ้น ค่าพารามิเตอร์ริดจ์จะมีค่าสูงขึ้น
4. เมื่อระดับความสัมพันธ์และค่าสัมประสิทธิ์ความถดถอยคงที่ ถ้าขนาดตัวอย่างเพิ่มขึ้น ค่า MAPE จะมีค่าลดลง

ทั้งนี้ ค่าพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์และการแจกแจงแกมมามีค่าน้อยใกล้ศูนย์จึงทำการทดสอบว่าค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์ดังกล่าวมีค่าต่างจากศูนย์หรือไม่ ซึ่งถ้าค่าพารามิเตอร์ริดจ์มีค่าไม่ต่างจากศูนย์จะเปรียบเสมือนยังไม่ได้มีการแก้ปัญหาพหุสัมพันธ์ จึงต้องทำการทดสอบค่าพารามิเตอร์ริดจ์ของทั้งสองการแจกแจงก่อนนำไปวิเคราะห์ผลต่อไป

ทดสอบค่าพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์และการแจกแจงแกมมาว่ามีค่าแตกต่างจากศูนย์หรือไม่

โดยขั้นตอนการทดสอบเป็นดังนี้

ขั้นตอนที่ 1 ทดสอบว่าค่าพารามิเตอร์ริดจ์มีการแจกแจงแบบปกติหรือไม่

H_0 : ค่าพารามิเตอร์ริดจ์มีการแจกแจงแบบปกติ

H_a : ค่าพารามิเตอร์ริดจ์ไม่ได้มีการแจกแจงแบบปกติ

ใช้สถิติทดสอบ Kolmogorov-Smirnov (KS) ได้ผลดังนี้

ตารางที่ 4.1.2 การทดสอบการแจกแจงปกติของค่าพารามิเตอร์ริดจ์

		Poisson Beta=2	Poisson Beta=3	Poisson Beta=4	Gamma Beta=2	Gamma Beta=3	Gamma Beta=4
N		8	8	8	8	8	8
Normal Parametersa,b	Mean	0.00034	0.00001	0.00000042	0.00264	0.000032	0.0000032
	Sd.	0.00002	0.00000044	0.00000002	0.00205	0.000023	0.0000025
Kolmogorov-Smirnov Z		0.503	0.545	0.691	0.460	0.313	0.463
Asymp. P-value		**0.962**	**0.928**	**0.726**	**0.984**	**1**	**0.983**

จากตารางพบว่าค่า Asymp. P-value ของทุกกรณีมีค่ามากกว่าระดับนัยสำคัญ 0.05 ดังนั้น ค่าพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์และการแจกแจงแกมมาในทุกค่าสัมประสิทธิ์ความถดถอยมีการแจกแจงแบบปกติ จึงดำเนินการในขั้นต่อไปคือ การทดสอบค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์ด้วยสถิติทดสอบ One Sample t-test

ขั้นตอนที่ 2

H_0 : ค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์มีค่าเท่ากับศูนย์

H_a : ค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์มีค่าไม่เท่ากับศูนย์

ได้ผลดังนี้

ตารางที่ 4.1.3 การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์และการแจกแจงแกมมาว่ามีค่าต่างจากศูนย์หรือไม่

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	P-value	Mean Difference	Lower	Upper
Poisson Beta=2	48.100	7	0.0000000004	0.0003369	0.0003203	0.0003534
Poisson Beta=3	72.091	7	0.00000000003	0.0000111	0.0000107	0.0000115
Poisson Beta=4	72.604	7	0.00000000002	0.0000004	0.0000004	0.0000004
Gamma Beta=2	3.634	7	0.008	0.0026375	0.000921	0.004354
Gamma Beta=3	3.931	7	0.006	0.0000320	0.000013	0.000051
Gamma Beta=4	3.736	7	0.007	0.0000032	0.00000119	0.0000053

จากตารางพบว่าค่า P-value ของทุกกรณีมีค่าน้อยกว่าค่าระดับนัยสำคัญ 0.05 ดังนั้น ค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์และการแจกแจงแกมมาในทุกค่าสัมประสิทธิ์ความถดถอยมีค่าแตกต่างจากศูนย์

เนื่องจากงานวิจัยครั้งนี้ได้ทำการศึกษาการแจกแจงของตัวแปรอิสระทั้งสิ้น 3 การแจกแจงคือ การแจกแจงปัวซองส์ การแจกแจงปกติ และการแจกแจงแกมมา ซึ่งครอบคลุมทั้งการแจกแจงแบบต่อเนื่องและการแจกแจงแบบไม่ต่อเนื่อง ซึ่งเป็นที่น่าสนใจว่าการแจกแจงจะมีอิทธิพลต่อค่าพารามิเตอร์ริดจ์หรือไม่

<u>ทดสอบว่าการแจกแจงของตัวแปรอิสระส่งผลต่อค่าพารามิเตอร์ริดจ์หรือไม่</u>

การทดสอบในขั้นตอนนี้ใช้สถิติทดสอบ One-Way ANOVA ซึ่งทำการทดสอบในแต่ละค่าสัมประสิทธิ์ความถดถอยโดยมีขั้นตอนการทดสอบเหมือนกัน ดังนี้

<u>ขั้นตอนที่</u> <u>1</u> ตรวจสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ของทั้ง 3 การแจกแจงว่ามีค่าเท่ากันหรือไม่

H_0 : ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์ = ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของการแจกแจงปกติ = ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมา

H_a : ความแปรปรวนของค่าพารามิเตอร์ริดจ์อย่างน้อย 2 การแจกแจงมีค่าไม่เท่ากัน

<u>ขั้นตอนที่</u> <u>2</u> ทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของทั้ง 3 การแจกแจงว่ามีค่าเท่ากันหรือไม่ โดยที่ถ้าในขั้นตอนที่ 1 ผลปรากฏว่าความแปรปรวนของทุกการแจกแจงมีค่าเท่ากัน ใช้สถิติ F ในการทดสอบขั้นตอนที่ 2 ส่วนถ้าผลปรากฏว่าความแปรปรวนอย่างน้อย 2 การแจกแจงมีค่าต่างกันให้ใช้สถิติ Welch ในการทดสอบ

H_0 : ค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์ = ค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงปกติ = ค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงแกมมา

H_a : ค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 การแจกแจงมีค่าไม่เท่ากัน

<u>ขั้นตอนที่</u> <u>3</u> ทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ทีละคู่

ถ้าในขั้นตอนที่ 2 ผลปรากฏว่า ค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 การแจกแจงมีค่าไม่เท่ากัน ให้ทำการทดสอบว่าค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงคู่ใดมีค่าเฉลี่ยที่แตกต่างกัน โดยถ้าในขั้นตอนที่ 1 ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของทั้ง 3 การแจกแจงมีค่าเท่ากันใช้สถิติ Bonferroni ส่วนถ้าความแปรปรวนอย่างน้อย 2 การแจกแจงมีค่าไม่เท่ากันใช้สถิติ Games-Howell ในการพิจารณา ซึ่งผลการวิเคราะห์เป็นดังนี้

ตารางที่ 4.1.4 การทดสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ในการแจกแจงปัวซองส์ การแจกแจงปกติ และการแจกแจงแกมมา

ค่าสัมประสิทธิ์ความถดถอย	Levene Statistic	df1	df2	P-value
2	9.705	2	21	0.00104
3	15.600	2	21	0.00007
4	47.992	2	21	0.000000015

จากตารางพบว่า ในทุกค่าสัมประสิทธิ์ความถดถอยให้ค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 เพราะฉะนั้นความแปรปรวนอย่างน้อย 2 การแจกแจงแตกต่างกัน ดังนั้นในขั้นตอนต่อไปจึงเลือกใช้สถิติ Welch ในการทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของทั้ง 3 การแจกแจง ซึ่งให้ผลการทดสอบดังนี้

ตารางที่ 4.1.5 การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์ การแจกแจงปกติ และการแจกแจงแกมมา

ค่าสัมประสิทธิ์ความถดถอย	Welch Statistic	df1	df2	P-value
2	26.819	2	9.334	0.00014
3	57.421	2	9.337	0.0000057
4	20.048	2	9.334	0.00042

จากตารางพบว่า ในทุกค่าสัมประสิทธิ์ความถดถอยให้ค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 ดังนั้น ค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 การแจกแจงมีค่าแตกต่างกัน จึงทำการพิจารณาในขั้นตอนต่อไปว่าการแจกแจงคู่ใดบ้างที่ค่าเฉลี่ยพารามิเตอร์ริดจ์มีความแตกต่างกัน โดยใช้ Multiple Comparison และวิเคราะห์ด้วยสถิติ Games-Howell ได้ผลการวิเคราะห์ดังนี้

เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2

ตารางที่ 4.1.6 การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงแต่ละคู่การแจกแจง เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2

	(I) dist	(J) dist	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Games-Howell	poisson	normal	-0.0848506	0.012322	0.001	-0.1211383	-0.0485629
		gamma	-0.0023006	0.000726	0.037	-0.0044381	-0.0001631
	normal	poisson	0.0848506	0.012322	0.001	0.0485629	0.1211383
		gamma	0.0825500	0.012343	0.001	0.0462603	0.1188397
	gamma	poisson	0.0023006	0.000726	0.037	0.0001631	0.0044381
		normal	-0.0825500	0.012343	0.001	-0.1188397	-0.0462603

<u>เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3</u>

<u>ตารางที่ 4.1.7</u> การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงแต่ละคู่การแจกแจง เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3

	(I) dist	(J) dist	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Games-Howell	poisson	normal	-0.0543326	0.005035	0.00003	-0.0691607	-0.0395045
		gamma	-0.0000209	0.000008	0.084	-0.0000449	0.0000031
	normal	poisson	0.0543326	0.005035	0.00003	0.0395045	0.0691607
		gamma	0.0543117	0.005035	0.00003	0.0394836	0.0691398
	gamma	poisson	0.0000209	0.000008	0.084	-0.0000031	0.0000449
		normal	-0.0543117	0.005035	0.00003	-0.0691398	-0.0394836

<u>เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4</u>

<u>ตารางที่ 4.1.8</u> การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงแต่ละคู่การแจกแจง เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4

	(I) dist	(J) dist	Mean Difference (I-J)	Std. Error	P-value	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Games-Howell	poisson	normal	-0.0309579	0.005441	0.002	-0.0469816	-0.01493427
		gamma	-0.0000028	0.000001	0.033	-0.0000054	-0.00000027
	normal	poisson	0.0309579	0.005441	0.002	0.0149343	0.04698156
		gamma	0.0309551	0.005441	0.002	0.0149315	0.04697874
	gamma	poisson	0.0000028	0.000001	0.033	0.0000003	0.00000537
		normal	-0.0309551	0.005441	0.002	-0.0469788	-0.01493145

จากตารางข้างต้นทำการวิเคราะห์โดยพิจารณาค่า P-value เทียบกับค่าระดับนัยสำคัญ 0.05 ซึ่งถ้าค่า P-value มีค่าน้อยกว่าค่าระดับนัยสำคัญจะทำให้ค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงคู่นั้นมีความแตกต่างกัน ส่วนถ้าค่า P-value มีค่าไม่น้อยกว่าค่าระดับนัยสำคัญจะทำให้ค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงคู่นั้นไม่มีความแตกต่างกัน ซึ่งจากตารางด้านบนพบว่า

1. การแจกแจงปัวซองส์และการแจกแจงปกติมีค่าเฉลี่ยพารามิเตอร์ริดจ์แตกต่างกันในทุกค่าสัมประสิทธิ์ความถดถอย
2. การแจกแจงแกมมาและการแจกแจงปกติมีค่าเฉลี่ยพารามิเตอร์ริดจ์แตกต่างกันในทุกค่าสัมประสิทธิ์ความถดถอย
3. การแจกแจงปัวซองส์และการแจกแจงแกมมามีค่าเฉลี่ยพารามิเตอร์ริดจ์แตกต่างกันเมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 และ 4 แต่ค่าเฉลี่ยพารามิเตอร์ริดจ์ไม่แตกต่างกันเมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3

ดังนั้นจึงสรุปได้ว่า การแจกแจงทั้ง 3 การแจกแจงของตัวแปรอิสระมีอิทธิพลต่อค่าเฉลี่ยพารามิเตอร์ริดจ์ ยกเว้นในกรณีของการแจกแจงปัวซองส์ที่ค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 และเมื่อตรวจสอบพบว่าการแจกแจงของตัวแปรอิสระมีอิทธิพลต่อค่าพารามิเตอร์ริดจ์ จึงทำการพิจารณาเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์ในแต่ละการแจกแจง เมื่อระดับความสัมพันธ์สูงขึ้นเทียบกับระดับความสัมพันธ์ 0.3

<u>เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์สูงขึ้นในแต่ละการแจกแจงและขนาดตัวอย่างเทียบกับระดับความสัมพันธ์ 0.3</u>

<u>ตารางที่ 4.1.9</u> เปอร์เซ็นต์การเพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์สูงขึ้น (เทียบกับระดับความสัมพันธ์ 0.3) ในแต่ละการแจกแจง

ค่าสัมประสิทธิ์	ρ	Poisson		Normal		Gamma	
		n=40	n=70	n=40	n=70	n=40	n=70
2	0.3 --> 0.5	1.587	3.125	140.000	22.857	150.000	257.143
	0.3 --> 0.7	6.349	14.063	240.000	45.714	400.000	614.286
	0.3 --> 0.8	9.524	14.063	360.000	97.857	775.000	757.143
	ค่าเฉลี่ย	**5.820**	**10.417**	**246.667**	**55.476**	**441.667**	**542.857**
3	0.3 --> 0.5	4.854	4.545	74.790	55.405	2400.000	275.000
	0.3 --> 0.7	6.796	4.636	98.319	67.568	4900.000	525.000
	0.3 --> 0.8	8.058	5.455	126.890	89.189	6328.571	775.000
	ค่าเฉลี่ย	**6.569**	**4.879**	**100.000**	**70.721**	**4542.857**	**525.000**
4	0.3 --> 0.5	0.012	1.220	37.037	109.524	275.000	500.000
	0.3 --> 0.7	3.350	1.951	74.074	141.270	650.000	800.000
	0.3 --> 0.8	7.085	10.000	81.481	147.619	1400.000	1300.000
	ค่าเฉลี่ย	**3.482**	**4.390**	**64.197**	**132.804**	**775.000**	**866.667**

การคำนวณค่าเปอร์เซ็นต์ที่เพิ่มขึ้นในตารางด้านบน สามารถคำนวณได้จาก

$$\text{\%เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์} = \left(\frac{\lambda_{\rho} - \lambda_{0.3}}{\lambda_{0.3}}\right) \times 100$$

โดยที่ λ_{ρ} คือ ค่าพารามิเตอร์ริดจ์ของระดับความสัมพันธ์ที่ต้องการเปรียบเทียบกับระดับความสัมพันธ์ 0.3 ในที่นี้มี 3 ค่า คือ $\lambda_{0.5}$, $\lambda_{0.7}$ และ $\lambda_{0.8}$

$\lambda_{0.3}$ คือ ค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์คือ 0.3

<u>ตัวอย่าง</u> สำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 ในการแจกแจงปัวซองส์ ที่ระดับความสัมพันธ์เพิ่มจาก 0.3 เป็น 0.5 ขนาดตัวอย่างเท่ากับ 40

$$\begin{aligned}\text{\%เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์} &= \left(\frac{\lambda_{0.5} - \lambda_{0.3}}{\lambda_{0.3}}\right) \times 100 \\ &= \left(\frac{0.00032 - 0.000315}{0.000315}\right) \times 100 \\ &\approx 1.587\%\end{aligned}$$

จากตารางพบว่า เมื่อพิจารณาในแต่ละการแจกแจงเปอร์เซ็นต์การเพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์สูงขึ้นเทียบกับระดับความสัมพันธ์ 0.3 เรียงตามลำดับจากมากไปน้อยได้ดังนี้ การแจกแจงแกมมา การแจกแจงปกติ และการแจกแจงปัวซองส์ ตามลำดับ ซึ่งเป็นลักษณะเดียวกันนี้ในทุกระดับความสัมพันธ์ที่เพิ่มขึ้น ทุกค่าสัมประสิทธิ์ความถดถอยและทุกขนาดตัวอย่าง

การศึกษาครั้งนี้ได้ทำการศึกษาภายใต้ค่าสัมประสิทธิ์ความถดถอยทั้งสิ้น 3 ค่า คือ 2, 3 และ 4 โดยค่าพารามิเตอร์ริดจ์ที่ได้ในแต่ละค่าสัมประสิทธิ์ความถดถอยมีความแตกต่างกัน จึงทำการทดสอบในลำดับต่อไปว่า ค่าสัมประสิทธิ์ความถดถอยมีอิทธิพลต่อค่าพารามิเตอร์ริดจ์หรือไม่

<u>ทดสอบว่าค่าสัมประสิทธิ์ความถดถอยโลจิสติกส่งผลต่อค่าพารามิเตอร์ริดจ์หรือไม่ โดยทำการพิจารณาในแต่ละการแจกแจง</u>

การทดสอบครั้งนี้ทำการทดสอบในแต่ละการแจกแจง โดยทดสอบค่าสัมประสิทธิ์ความถดถอยว่ามีอิทธิพลต่อค่าพารามิเตอร์ริดจ์หรือไม่ด้วยสถิติ One-Way ANOVA ซึ่งขั้นตอนการทดสอบในแต่ละการแจกแจงเหมือนกันดังนี้

<u>ขั้นตอนที่ 1</u> ตรวจสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ในกลุ่มของค่าสัมประสิทธิ์ความถดถอยทั้ง 3 ค่าว่ามีค่าเท่ากันหรือไม่

H_0 : ความแปรปรวนของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 = ความแปรปรวนของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 = ความแปรปรวนของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4

H_a : ความแปรปรวนของค่าพารามิเตอร์ริดจ์อย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าไม่เท่ากัน

<u>ขั้นตอนที่ 2</u> ทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของทั้ง 3 กลุ่มค่าสัมประสิทธิ์ความถดถอยว่ามีค่าเท่ากันหรือไม่ โดยที่

ถ้าในขั้นตอนที่ 1 ผลปรากฏว่าความแปรปรวนของทุกค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากัน ใช้สถิติ F ในการทดสอบขั้นตอนที่ 2 ส่วนถ้าผลปรากฏว่าความแปรปรวนอย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าต่างกันให้ใช้สถิติ Welch ในการทดสอบ

H_0 : ค่าเฉลี่ยพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 = ค่าเฉลี่ยพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 = ค่าเฉลี่ยพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4

H_a : ค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าไม่เท่ากัน

<u>ขั้นตอนที่ 3</u> ทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ทีละคู่

ถ้าในขั้นตอนที่ 2 ผลปรากฏว่า ค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าไม่เท่ากัน ให้ทำการทดสอบว่าค่าเฉลี่ยพารามิเตอร์ริดจ์ของค่าสัมประสิทธิ์ความถดถอยคู่ใดมีค่าเฉลี่ยที่แตกต่างกัน โดยถ้าในขั้นตอนที่ 1 ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของทั้ง 3 ค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากันใช้สถิติ Bonferroni ส่วนถ้าความแปรปรวนอย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าไม่เท่ากันใช้สถิติ Games-Howell ในการพิจารณา

ซึ่งผลการวิเคราะห์เป็นดังนี้

ตารางที่ 4.1.10 การทดสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ในแต่ละค่าสัมประสิทธิ์ความถดถอย

การแจกแจง	Levene Statistic	df1	df2	P-value
Poisson	21.039	2	21	0.0000097
Normal	2.354	2	21	0.11955
Gamma	19.233	2	21	0.000018

จากตารางพบว่า การแจกแจงปัวซองส์และการแจกแจงแกมมาให้ค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 เพราะฉะนั้นความแปรปรวนอย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยแตกต่างกัน ดังนั้นในขั้นตอนต่อไปจึงเลือกใช้สถิติ Welch ในการทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงปัวซองส์และการแจกแจงแกมมา ส่วนผลการทดสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ทั้ง 3 ค่าสัมประสิทธิ์ความถดถอยของการแจกแจงปกติให้ค่า P-value มากกว่าค่าระดับนัยสำคัญ 0.05 เพราะฉะนั้นความแปรปรวนของทั้ง 3 ค่าสัมประสิทธิ์ความถดถอยไม่มีความแตกต่างกัน ดังนั้นในขั้นตอนต่อไปจึงเลือกใช้สถิติ F ซึ่งให้ผลการทดสอบดังนี้

ตารางที่ 4.1.11 การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของทั้ง 3 ค่าสัมประสิทธิ์ความถดถอยในการแจกแจงปัวซองส์และแกมมา

การแจกแจง	Welch Statistic	df1	df2	P-value
Poisson	3333.192	2	9.346	0.000
Gamma	11.918	2	9.438	0.003

ตารางที่ 4.1.12 การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของทั้ง 3 ค่าสัมประสิทธิ์ความถดถอยในการแจกแจงปกติ

		Sum of Squares	df	Mean Square	F	P-value
Normal	Between Groups	0.012	2	0.006	10.734	0.001
	Within Groups	0.012	21	0.001		
	Total	0.023	23			

จากตารางพบว่า ในทุกการแจกแจงให้ค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 ดังนั้น ค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าแตกต่างกัน จึงทำการพิจารณาในขั้นตอนต่อไปว่าค่าสัมประสิทธิ์ความถดถอยคู่ใดบ้างที่ค่าเฉลี่ยพารามิเตอร์ริดจ์มีความแตกต่างกัน โดยใช้ Multiple Comparison และวิเคราะห์ด้วยสถิติ Bonferroni สำหรับการแจกแจงแบบปกติ ส่วนการแจกแจงแบบปัวซองส์และการแจกแจงแบบแกมมาวิเคราะห์ด้วยสถิติ Games-Howell ได้ผลการวิเคราะห์ดังนี้

<u>ตารางที่ 4.1.13</u> การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของค่าสัมประสิทธิ์ความถดถอยแต่ละคู่

Dependent Variable		(I) dist	(J) dist	Mean Difference (I-J)	Std. Error	P-value
Poisson	Games-Howell	Beta=2	Beta=3	0.00032578	0.00000701	0.0000001
			Beta=4	0.00033646	0.00000700	0.0000001
		Beta=3	Beta=2	-0.00032578	0.00000701	0.0000001
			Beta=4	0.00001068	0.00000015	0.0000001
		Beta=4	Beta=2	-0.00033646	0.00000700	0.0000001
			Beta=3	-0.00001068	0.00000015	0.0000001
Normal	Bonferroni	Beta=2	Beta=3	0.03084375	0.01174093	0.047
			Beta=4	0.05422917	0.01174093	0.00044
		Beta=3	Beta=2	-0.03084375	0.01174093	0.047
			Beta=4	0.02338542	0.01174093	0.179
		Beta=4	Beta=2	-0.05422917	0.01174093	0.00044
			Beta=3	-0.02338542	0.01174093	0.179
Gamma	Games-Howell	Beta=2	Beta=3	0.00260548	0.00072583	0.021
			Beta=4	0.00263426	0.00072579	0.020
		Beta=3	Beta=2	-0.00260548	0.00072583	0.021
			Beta=4	0.00002879	0.00000819	0.023
		Beta=4	Beta=2	-0.00263426	0.00072579	0.020
			Beta=3	-0.00002879	0.00000819	0.023

จากตารางพบว่า ในการแจกแจงปัวซองส์ การแจกแจงแกมมาในทุกคู่ของค่าสัมประสิทธิ์ความถดถอยและการแจกแจงปกติในคู่ของค่าสัมประสิทธิ์ความถดถอยที่มีค่า 2 กับ 3 และ 2 กับ 4 ค่า P-value ที่ได้มีค่าน้อยกว่าค่าระดับนัยสำคัญ 0.05 ส่วนการแจกแจงปกติในคู่ของค่าสัมประสิทธิ์ความถดถอยที่มีค่า 3 กับ 4 ค่า P-value ที่ได้มีค่ามากกว่าค่าระดับนัยสำคัญ 0.05 แสดงว่าค่า

สัมประสิทธิ์ความถดถอยมีอิทธิพลต่อค่าเฉลี่ยพารามิเตอร์ริดจ์ในทุกการแจกแจงยกเว้นการแจกแจงปกติที่ค่าสัมประสิทธิ์ความถดถอยเป็น 3 กับ 4 จะไม่มีความแตกต่างกัน

จากผลการทดสอบดังกล่าวแสดงให้เห็นว่าค่าสัมประสิทธิ์ความถดถอยส่งผลต่อค่าพารามิเตอร์ริดจ์ และเมื่อทำการพิจารณาค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยสูงขึ้นพบว่า ค่าพารามิเตอร์ริดจ์มีค่าลดลง จึงทำการพิจารณาเปอร์เซ็นต์ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าสูงขึ้นเทียบกับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2

<u>เปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้น</u>

<u>ตารางที่ 4.1.14</u> เปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้นจำแนกตามการแจกแจง

	ระดับความสัมพันธ์	ขนาดตัวอย่าง	Poisson	Normal	Gamma
ค่าสัมประสิทธิ์ความถดถอยจาก 2 ไป 3	0.3	40	96.730	-19.000*	99.825
		70	96.563	47.143	98.857
	0.5	40	96.625	13.333	98.250
		70	96.515	33.140	98.800
	0.7	40	96.716	30.588	98.250
		70	96.847	39.216	99.000
	0.8	40	96.774	41.304	98.714
		70	96.822	49.458	98.833
ค่าเฉลี่ย			**96.699**	**29.398**	**98.816**
ค่าสัมประสิทธิ์ความถดถอยจาก 2 ไป 4	0.3	40	99.873	46.000	99.900
		70	99.872	70.000	99.929
	0.5	40	99.875	69.167	99.850
		70	99.874	48.837	99.880
	0.7	40	99.876	72.353	99.850
		70	99.885	50.327	99.910
	0.8	40	99.875	78.696	99.829
		70	99.876	62.455	99.883
ค่าเฉลี่ย			**99.876**	**62.229**	**99.879**

หมายเหตุ: *หมายถึง กรณีที่แตกต่างจากกรณีอื่นคือ ค่าพารามิเตอร์ริดจ์ของค่าสัมประสิทธิ์ความถดถอยเท่ากับ 3 มีค่าสูงกว่าค่าพารามิเตอร์ริดจ์ของค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2

การคำนวณค่าเปอร์เซ็นต์ที่ลดลงในตารางด้านบน สามารถคำนวณได้จาก

$$\text{\%ลดลงของค่าพารามิเตอร์ริดจ์} = \left(\frac{\lambda_{\text{beta}2} - \lambda_{\text{b}}}{\lambda_{\text{beta}2}}\right) \times 100$$

โดยที่ λ_{beta2} คือ ค่าพารามิเตอร์ริดจ์ของกรณีที่สนใจเมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2

λ_{b} คือ ค่าพารามิเตอร์ริดจ์ในกรณีที่สนใจของค่าสัมประสิทธิ์ความถดถอยที่ต้องการเปรียบเทียบกับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 ในที่นี้มี 2 ค่า คือ λ_{beta3} และ λ_{beta4}

<u>ตัวอย่าง</u> สำหรับการแจกแจงปัวซองส์ที่ระดับความสัมพันธ์ 0.3 ขนาดตัวอย่างมีค่า 40 โดยพิจารณาเปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์จากค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 ไปยังค่าสัมประสิทธิ์ความถดถอยเท่ากับ 3

$$\text{\%ลดลงของค่าพารามิเตอร์ริดจ์} = \left(\frac{0.000315 - 0.0000103}{0.000315}\right) \times 100$$

$$\approx 96.73\%$$

จากตารางด้านบนพบว่า เปอร์เซ็นต์การลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้นจาก 2 เป็น 3 และ จาก 2 เป็น 4 ของทั้ง 3 การแจกแจงเรียงลำดับจากการแจกแจงที่มีเปอร์เซ็นต์การลดลงมากที่สุดไปยังน้อยที่สุดได้ดังนี้ การแจกแจงแกมมา การแจกแจงปัวซองส์ และการแจกแจงปกติ ตามลำดับ โดยถ้ามองในแต่ละการแจกแจงพบว่า การแจกแจงแกมมามีเปอร์เซ็นต์การลดลงเฉลี่ยของค่าพารามิเตอร์ริดจ์ประมาณ 98.816% สำหรับค่าสัมประสิทธิ์ความถดถอยเพิ่มขึ้นจาก 2 เป็น 3 และเปอร์เซ็นต์การลดลงเฉลี่ยของค่าพารามิเตอร์ริดจ์มีค่าประมาณ 99.879% สำหรับค่าสัมประสิทธิ์ความถดถอยเพิ่มขึ้นจาก 2 เป็น 4 ส่วนการแจกแจงปัวซองส์มีเปอร์เซ็นต์การลดลงเฉลี่ยของค่าพารามิเตอร์ริดจ์ประมาณ 96.699% สำหรับค่าสัมประสิทธิ์ความถดถอยเพิ่มขึ้นจาก 2 เป็น 3 และเปอร์เซ็นต์การลดลงเฉลี่ยของค่าพารามิเตอร์ริดจ์มีค่าประมาณ 99.876% สำหรับค่าสัมประสิทธิ์ความถดถอยเพิ่มขึ้นจาก 2 เป็น 4 ซึ่งทั้งการแจกแจงแกมมาและการแจกแจงปัวซองส์มีเปอร์เซ็นต์การ

ลดลงของค่าพารามิเตอร์ริดจ์ในแต่ละระดับความสัมพันธ์และขนาดตัวอย่างใกล้เคียงกัน ซึ่งแตกต่างจากการแจกแจงปกติที่เปอร์เซ็นต์การลดลงของค่าพารามิเตอร์ริดจ์มีความแตกต่างกันทำให้เห็นแนวโน้มการลดลงที่ไม่ชัดเจน

การศึกษาครั้งนี้นอกจากทำการพิจารณาในมุมมองของระดับความสัมพันธ์และค่าสัมประสิทธิ์ความถดถอยแล้ว ยังมีการพิจารณาถึงขนาดตัวอย่าง ทั้งนี้มีการพิจารณาขนาดตัวอย่างใน 2 ระดับคือ 40 และ 70 ซึ่งเมื่อขนาดตัวอย่างเพิ่มขึ้นทำให้ค่าพารามิเตอร์ริดจ์ที่ได้มีค่าสูงขึ้น จึงทำการพิจารณาเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างมีค่าสูงขึ้นในแต่ละการแจกแจงได้ผลดังนี้

<u>เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 40 เป็น 70</u>

<u>ตารางที่ 4.1.15</u> เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างมีค่าสูงขึ้น

ค่าสัมประสิทธิ์ความถดถอย	ระดับความสัมพันธ์	Poisson	Normal	Gamma
2	0.3*	1.587	180.000	75.000
	0.5	3.125	43.333	150.000
	0.7	8.955	20.000	150.000
	0.8	5.797	20.435	71.429
	ค่าเฉลี่ย	**4.866**	**65.942**	**111.607**
3	0.3	6.796	24.370	1042.857
	0.5	6.481	10.577	71.429
	0.7	4.636	5.085	42.857
	0.8*	4.223	3.704	55.556
	ค่าเฉลี่ย	**5.534**	**10.934**	**303.175**
4	0.3	2.104	55.556	25.000
	0.5	3.337	137.838	100.000
	0.7	0.723	115.603	50.000
	0.8	4.884	112.245	16.667
	ค่าเฉลี่ย	**2.762**	**105.310**	**47.917**

หมายเหตุ: *คือ ค่าในตารางที่มีผลการวิเคราะห์แตกต่างจากกรณีอื่น

จากตารางด้านบนพบว่า ถ้าพิจารณาค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 และ 3 เปอร์เซ็นต์ของค่าพารามิเตอร์ริดจ์ที่เพิ่มขึ้นเมื่อขนาดตัวอย่างสูงขึ้นเรียงลำดับจากมากไปน้อยตามการแจกแจงได้ดังนี้ การแจกแจงแกมมา การแจกแจงปกติ และการแจกแจงปัวซองส์ ตามลำดับ ซึ่งเป็นลักษณะเช่นนี้ในทุกระดับความสัมพันธ์ยกเว้น ระดับความสัมพันธ์ 0.3 ของค่าสัมประสิทธิ์ความถดถอยเป็น 2 และระดับความสัมพันธ์ 0.8 ของค่าสัมประสิทธิ์ความถดถอยเป็น 3 ส่วนค่าสัมประสิทธิ์ความถดถอยเป็น 4 เปอร์เซ็นต์ของค่าพารามิเตอร์ริดจ์ที่เพิ่มขึ้นเมื่อขนาดตัวอย่างสูงขึ้นเรียงลำดับจากมากไปน้อยตามการแจกแจงเป็นดังนี้ การแจกแจงปกติ การแจกแจงแกมมา และการแจกแจงปัวซองส์ ตามลำดับ ซึ่งเป็นลักษณะเช่นนี้ในทุกระดับความสัมพันธ์

4.2 ผลการศึกษาในกรณีจำนวนตัวแปรอิสระ 3 ตัว

<u>แสดงผลค่าพารามิเตอร์ริดจ์ในแต่ละกรณีที่ทำการศึกษา</u> ได้ผลการวิเคราะห์ดังตารางด้านล่าง

<u>ตารางที่ 4.1.16</u> ค่าพารามิเตอร์ริดจ์ในกรณีจำนวนตัวแปรอิสระ 3 ตัว จำแนกตามการแจกแจง

ค่าสัมประสิทธิ์ความถดถอย	**ρ**	n	Normal		Gamma	
			λ	MAPE (%)	**λ**	MAPE (%)
2	(0.3, 0.5, 0.8)	60	0.0255	9.912	0.0000055	14.552
		100	0.027	6.438	0.00001	13.835
	(0.5, 0.5, 0.8)	60	0.046	9.783	0.000018	14.614
		100	0.0647	8.319	0.00003	13.963
	(0.5, 0.7, 0.8)	60	0.105	9.626	0.000019	13.237
		100	0.1125	9.024	0.000031	12.433
	(0.7, 0.7, 0.8)	60	0.111	9.093	0.000035	13.777
		100	0.118	7.067	0.00004	12.990
	(0.8, 0.8, 0.8)	60	0.124	9.893	0.00007	14.265
		100	0.125	6.336	0.0001169	13.428
3	(0.3, 0.5, 0.8)	60	0.026	6.660	0.0000005	14.065
		100	0.027	4.592	0.000001	13.257
	(0.5, 0.5, 0.8)	60	0.0445	6.873	0.000002	13.415
		100	0.046	5.472	0.0000021	9.907
	(0.5, 0.7, 0.8)	60	0.051	6.221	0.0000038	13.701
		100	0.0525	5.298	0.0000041	12.471
	(0.7, 0.7, 0.8)	60	0.0635	6.859	0.0000055	14.991

ตารางที่ 4.1.16(ต่อ) ค่าพารามิเตอร์ริดจ์ในกรณีจำนวนตัวแปรอิสระ 3 ตัวจำแนกตามการแจกแจง

ค่าสัมประสิทธิ์ความถดถอย	ρ	n	Normal		Gamma	
			λ	MAPE (%)	λ	MAPE (%)
3	(0.7, 0.7, 0.8)	100	0.066	5.503	0.0000061	13.467
	(0.8, 0.8, 0.8)	60	0.065	6.394	0.00000735	12.430
		100	0.0685	5.282	0.0000075	12.358
4	(0.3, 0.5, 0.8)	60	0.00967	4.689	0.00000007	14.530
		100	0.0285	4.658	0.00000009	13.740
	(0.5, 0.5, 0.8)	60	0.021	4.694	0.0000001	14.277
		100	0.033	4.602	0.0000002	13.187
	(0.5, 0.7, 0.8)	60	0.026	4.663	0.0000002	14.210
		100	0.0365	3.999	0.0000003	13.183
	(0.7, 0.7, 0.8)	60	0.027	3.889	0.00000035	14.384
		100	0.037	3.700	0.0000005	12.516
	(0.8, 0.8, 0.8)	60	0.028	4.294	0.0000005	12.724
		100	0.04	3.427	0.0000006	10.533

โดยที่ ρ คือ ระดับความสัมพันธ์ของตัวแปรอิสระในที่นี้ตัวแปรอิสระมี 3 ตัว จึงเป็นระดับความสัมพันธ์ของตัวแปรอิสระตัวที่ 1 กับตัวแปรอิสระตัวที่ 2 ตัวแปรอิสระตัวที่ 1 กับตัวแปรอิสระตัวที่ 3 และ ตัวแปรอิสระตัวที่ 2 ตัวแปรอิสระตัวที่ 3 ตามลำดับ

n คือ ขนาดตัวอย่างที่ทำการศึกษา

λ คือ ค่าพารามิเตอร์ริดจ์ที่เหมาะสม

MAPE คือ ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย ซึ่งกำหนดไว้ดังนี้

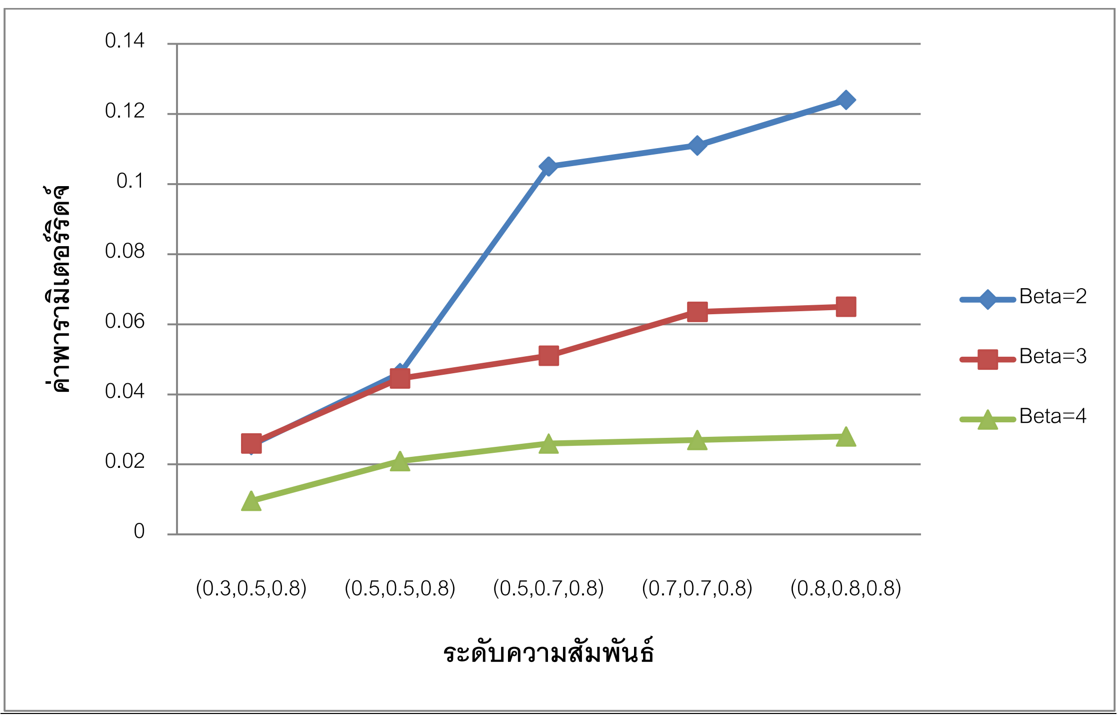
สำหรับการแจกแจงปกติ

- สำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 ให้ MAPE มีค่าไม่เกิน 10%
- สำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 3 ให้ MAPE มีค่าไม่เกิน 7%
- สำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 4 ให้ MAPE มีค่าไม่เกิน 5%

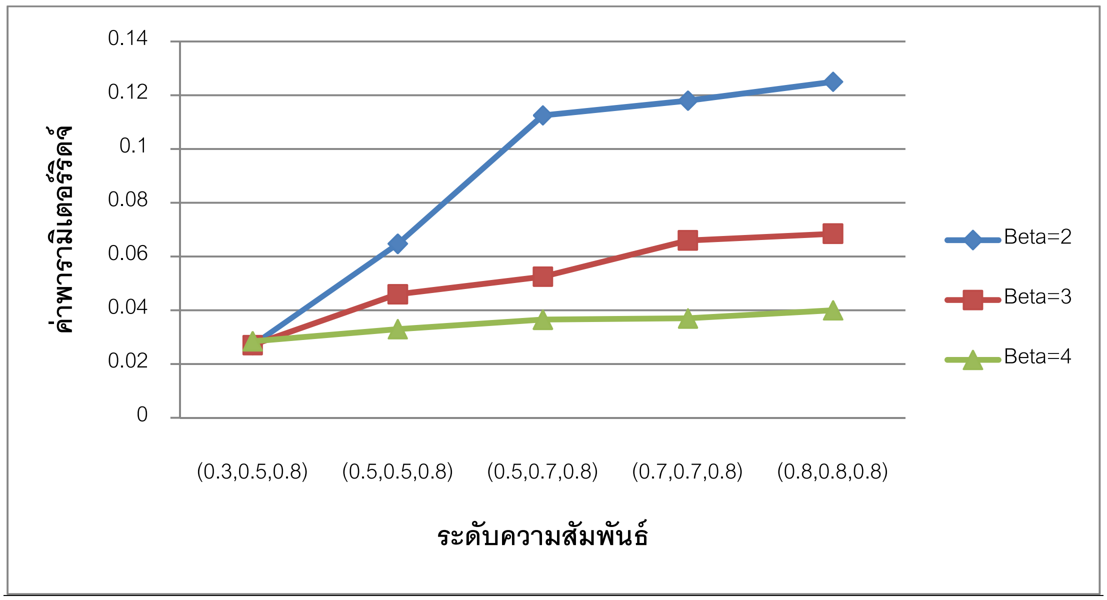
สำหรับการแจกแจงแกมมา

- ให้ MAPE มีค่าไม่เกิน 15% ทุกค่าสัมประสิทธิ์ความถดถอย

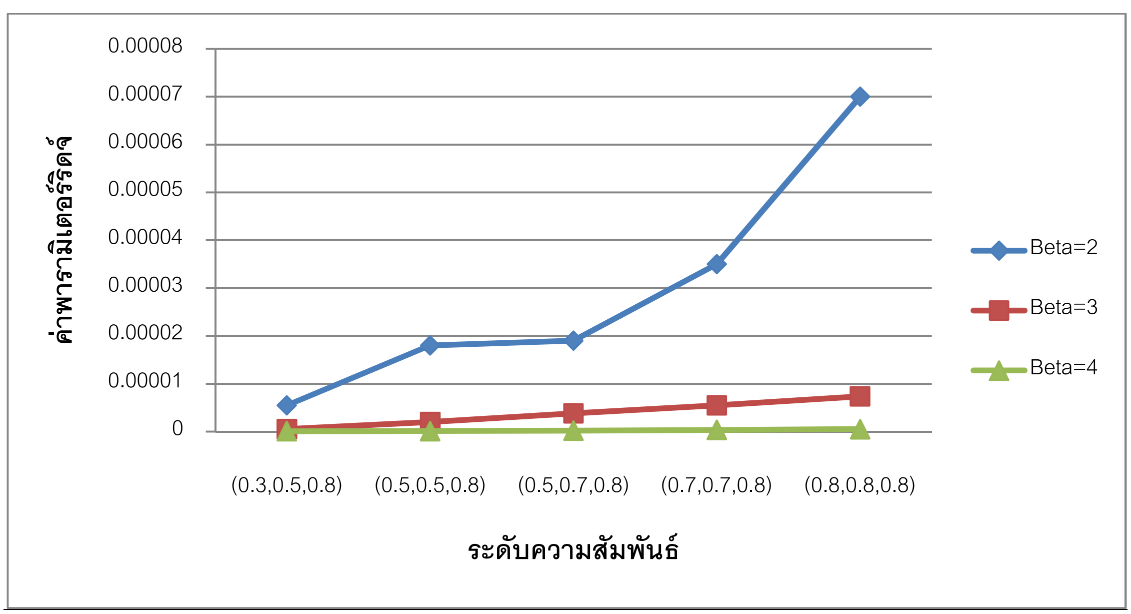
รูปที่ 4.1.16 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงปกติเมื่อขนาดตัวอย่างเท่ากับ 60



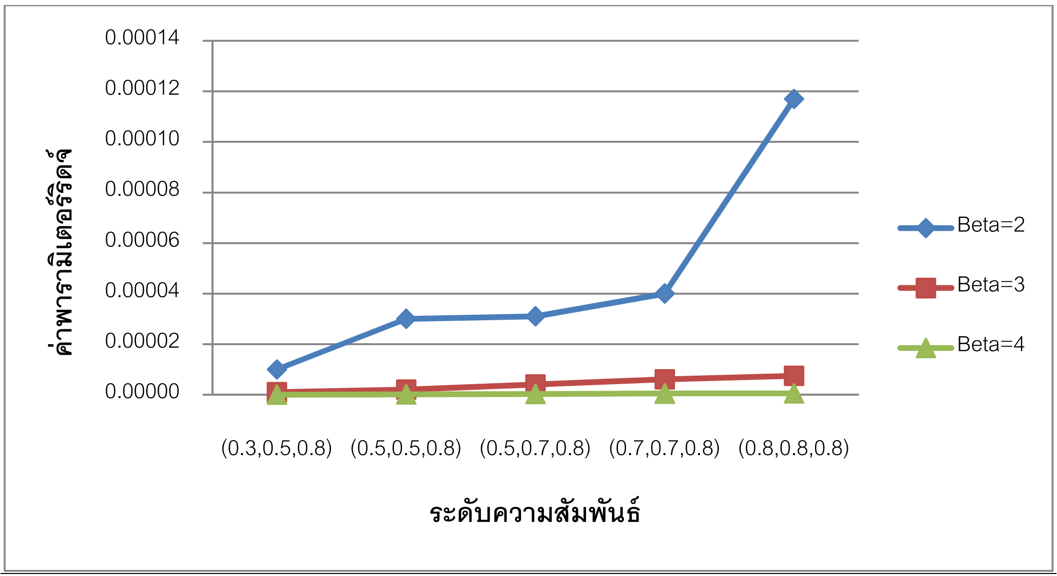
รูปที่ 4.1.17 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงปกติเมื่อขนาดตัวอย่างเท่ากับ 100



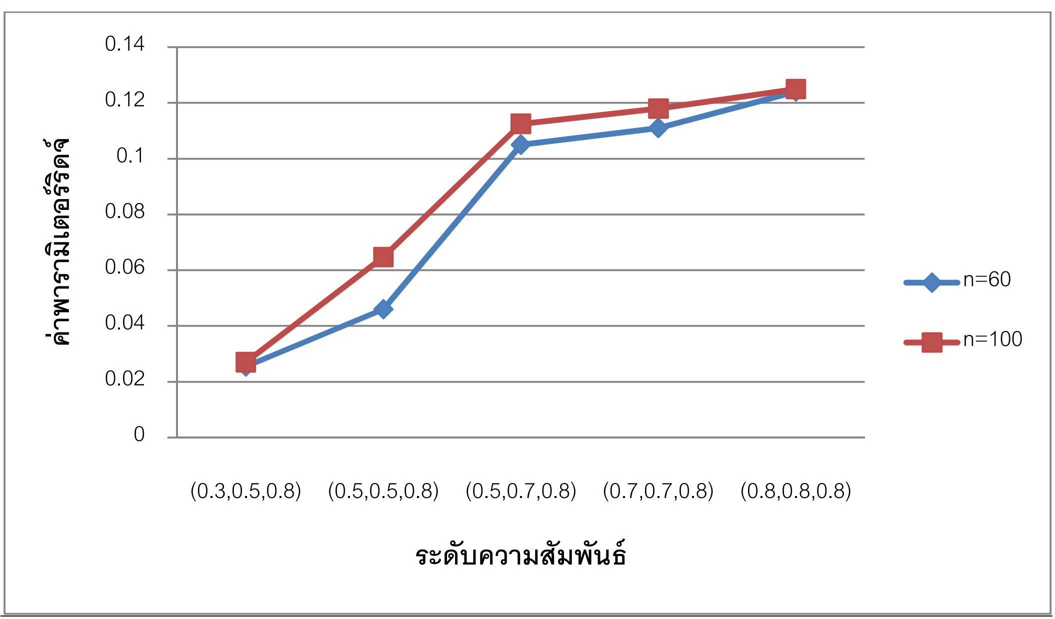
รูปที่ 4.1.18 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงแกมมาเมื่อขนาดตัวอย่างเท่ากับ 60



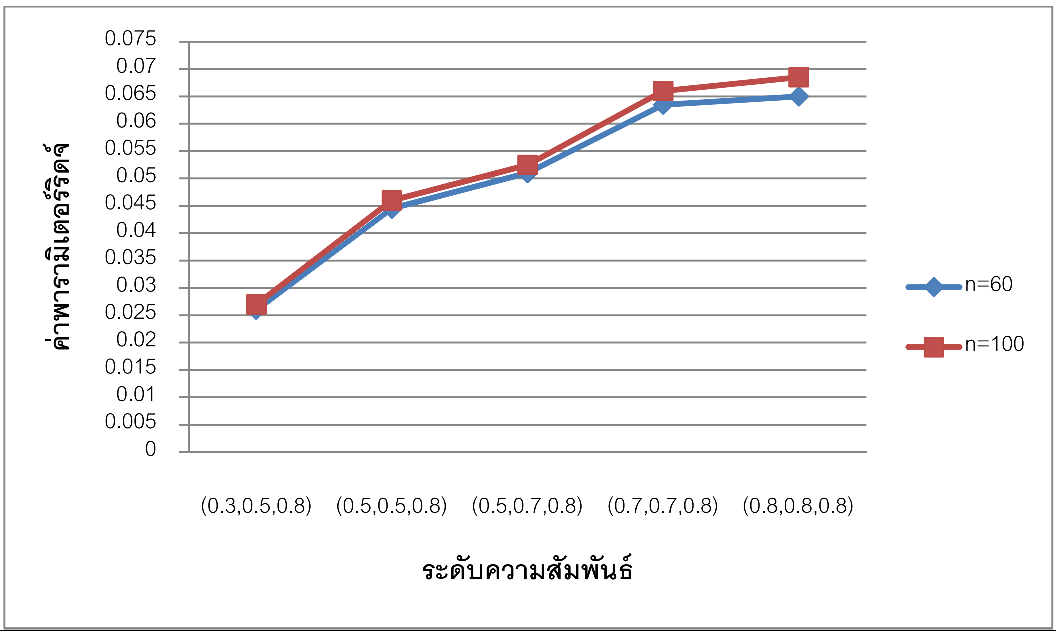
รูปที่ 4.1.19 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามค่าสัมประสิทธิ์ความถดถอยโลจิสติกของการแจกแจงแกมมาเมื่อขนาดตัวอย่างเท่ากับ 100



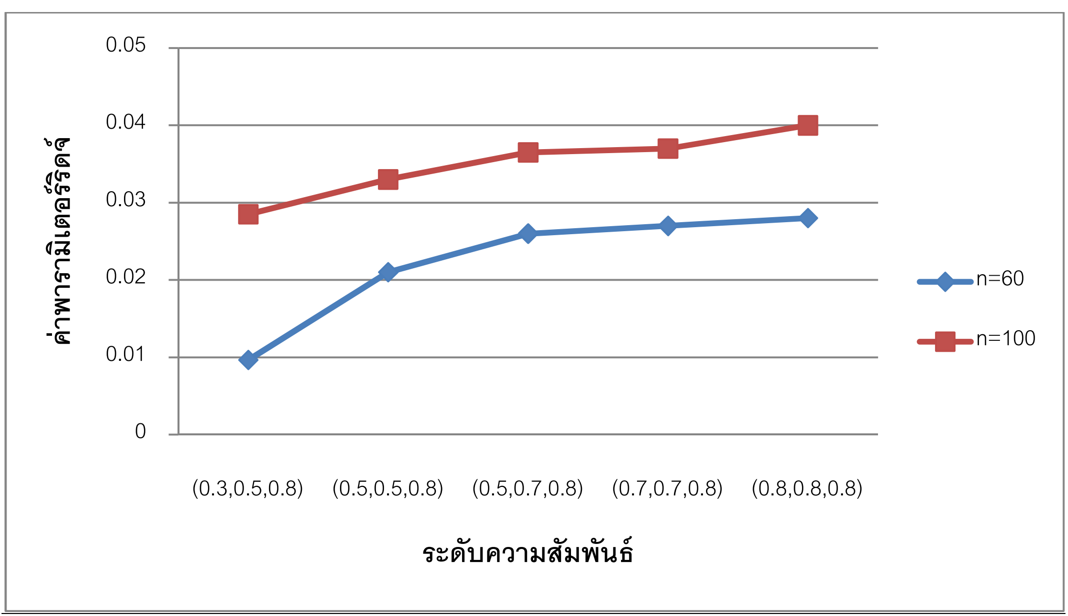
รูปที่ 4.1.20 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปกติเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 2



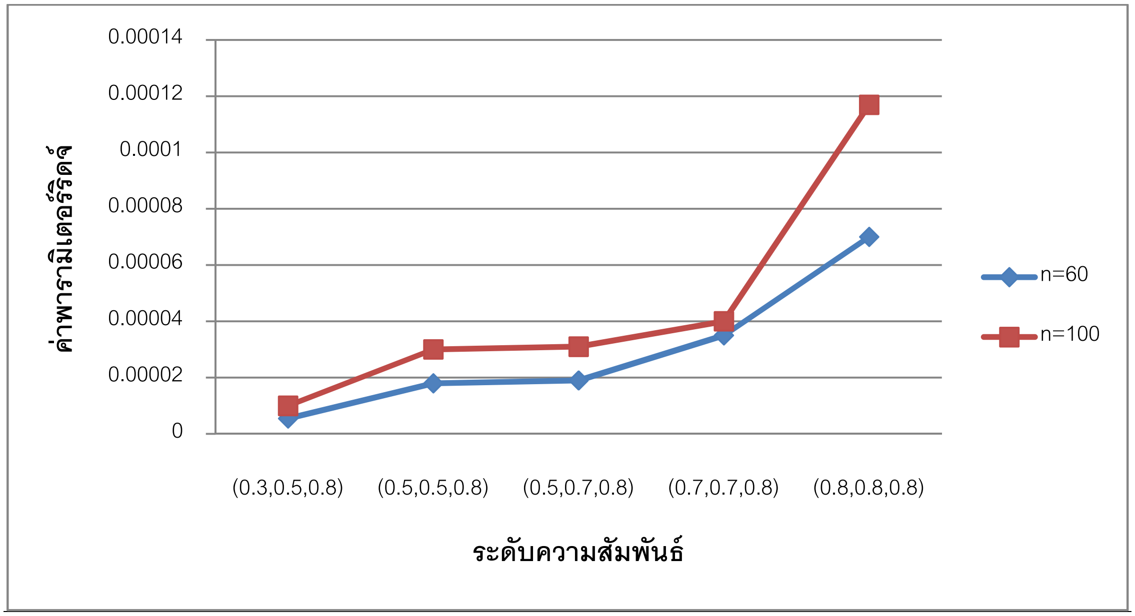
รูปที่ 4.1.21 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปกติเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 3



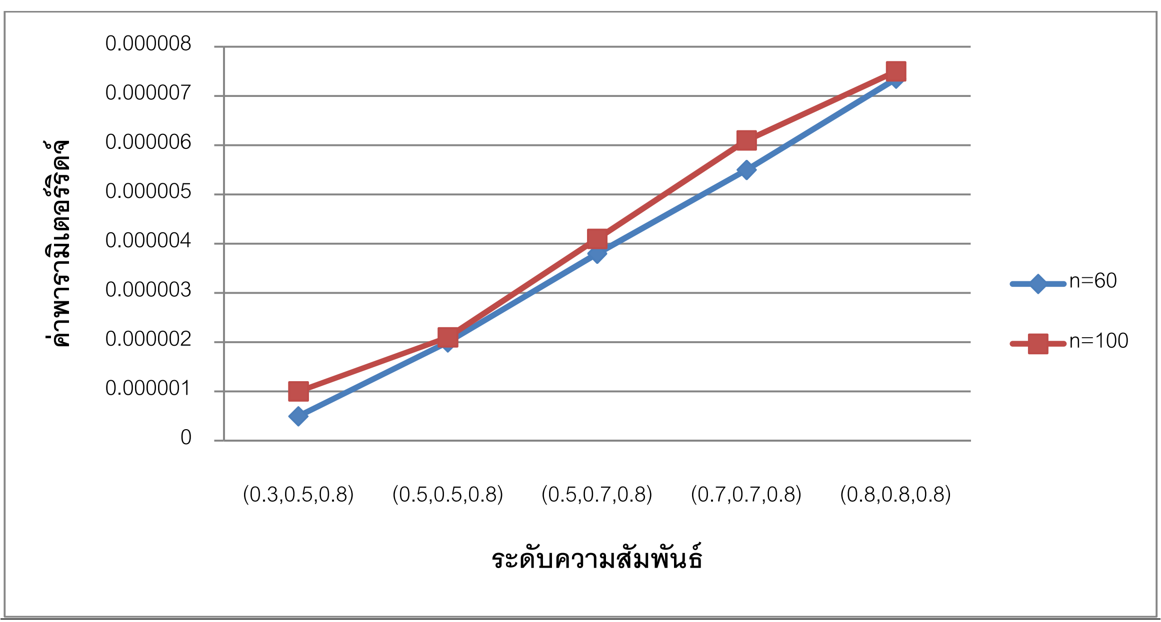
รูปที่ 4.1.22 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงปกติเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 4



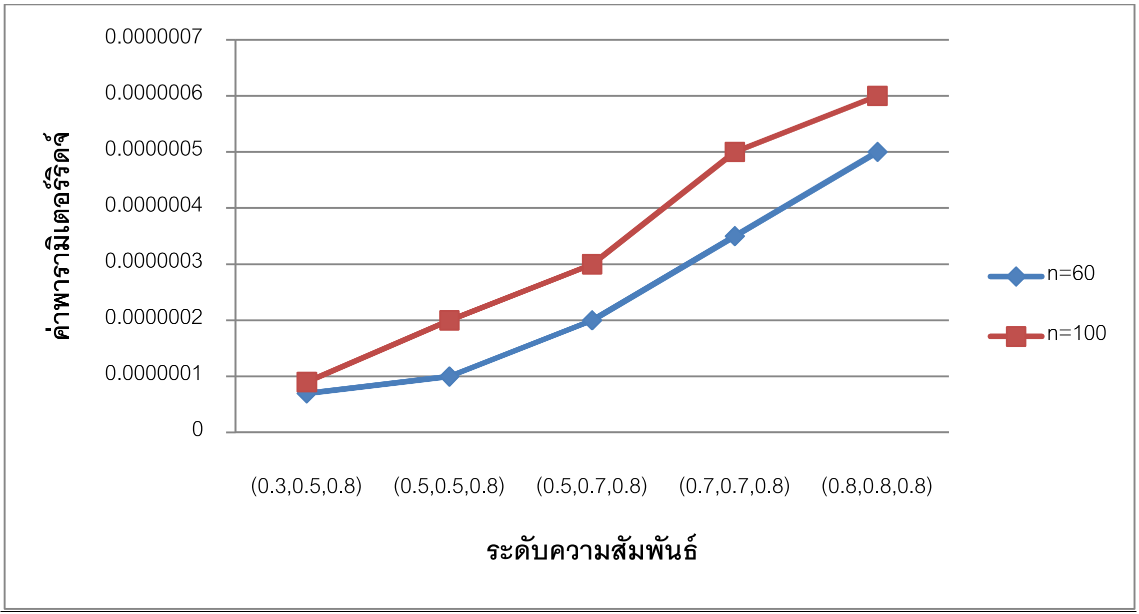
รูปที่ 4.1.23 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงแกมมาเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 2



รูปที่ 4.1.24 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงแกมมาเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 3



รูปที่ 4.1.25 กราฟแสดงค่าพารามิเตอร์ริดจ์แยกตามขนาดตัวอย่างของการแจกแจงแกมมาเมื่อค่าสัมประสิทธิ์ความถดถอยโลจิสติกมีค่าเท่ากับ 4



จากตารางและกราฟด้านบนพบประเด็นที่น่าสนใจ ดังนี้

1. เมื่อขนาดตัวอย่างและค่าสัมประสิทธิ์ความถดถอยคงที่ ถ้าระดับความสัมพันธ์สูงขึ้น ค่าพารามิเตอร์ริดจ์จะมีค่าสูงขึ้น
2. เมื่อขนาดตัวอย่างและระดับความสัมพันธ์คงที่ ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้น ค่าพารามิเตอร์ริดจ์จะมีค่าลดลงทั้งการแจกแจงปกติและการแจกแจงแกมมา
3. เมื่อระดับความสัมพันธ์และค่าสัมประสิทธิ์ความถดถอยคงที่ ถ้าขนาดตัวอย่างเพิ่มขึ้น ค่าพารามิเตอร์ริดจ์จะมีค่าสูงขึ้น
4. เมื่อระดับความสัมพันธ์และค่าสัมประสิทธิ์ความถดถอยคงที่ ถ้าขนาดตัวอย่างเพิ่มขึ้น ค่า MAPE จะมีค่าลดลง

ทั้งนี้ ค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมามีค่าน้อยใกล้ศูนย์จึงทำการทดสอบว่าค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์ดังกล่าวมีค่าต่างจากศูนย์หรือไม่ ซึ่งถ้าค่าพารามิเตอร์ริดจ์มีค่าไม่ต่างจากศูนย์จะเปรียบเสมือนยังไม่ได้มีการแก้ปัญหาพหุสัมพันธ์ จึงต้องทำการทดสอบค่าพารามิเตอร์ริดจ์ก่อนนำไปวิเคราะห์ผลต่อไป

<u>ทดสอบค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมาว่ามีค่าแตกต่างจากศูนย์หรือไม่</u>

โดยขั้นตอนการทดสอบเป็นดังนี้

<u>ขั้นตอนที่ 1</u> ทดสอบว่าค่าพารามิเตอร์ริดจ์มีการแจกแจงแบบปกติหรือไม่

H_0 : ค่าพารามิเตอร์ริดจ์มีการแจกแจงแบบปกติ

H_a : ค่าพารามิเตอร์ริดจ์ไม่ได้มีการแจกแจงแบบปกติ

ใช้สถิติทดสอบ Kolmolgorov-Smirnof (KS) ได้ผลดังนี้

<u>ตารางที่ 4.1.17</u> การทดสอบการแจกแจงปกติของค่าพารามิเตอร์ริดจ์

		Gamma Beta=2	Gamma Beta=3	Gamma Beta=4
N		10	10	10
Normal	Mean	0.0000375423	0.0000039950	0.0000002910
Parametersa,b	Std. Deviation	0.00003329664	0.00000256130	0.00000019110
Most Extreme	Absolute	0.271	0.170	0.183
Differences	Positive	0.271	0.170	0.183
	Negative	-0.168	-0.122	-0.163
Kolmogorov-Smirnov Z		0.856	0.539	0.579
Asymp. P-value		**0.457**	**0.934**	**0.891**

จากตารางพบว่าค่า Asymp. P-value ของทุกกรณีมีค่ามากกว่าระดับนัยสำคัญ 0.05 ดังนั้น ค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมาในทุกค่าสัมประสิทธิ์ความถดถอยมีการแจกแจงแบบปกติ จึงดำเนินการในขั้นต่อไปคือ การทดสอบค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์ด้วยสถิติทดสอบ One Sample t-test

ขั้นตอนที่ 2

H_0 : ค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์มีค่าเท่ากับศูนย์

H_a : ค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์มีค่าไม่เท่ากับศูนย์

ได้ผลดังนี้

ตารางที่ 4.1.18 การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแกมมาว่ามีค่าต่างจากศูนย์หรือไม่

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	P-value	Mean Difference	Lower	Upper
Gamma Beta=2	3.566	9	0.006	0.000037542	0.00001372	0.00006136
Gamma Beta=3	4.932	9	0.001	0.000003995	0.00000216	0.00000583
Gamma Beta=4	4.815	9	0.001	0.000000291	0.00000015	0.00000043

จากตารางพบว่าค่า P-value ของทุกกรณีมีค่าน้อยกว่าค่าระดับนัยสำคัญ 0.05 ดังนั้นค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงแกมมาในทุกค่าสัมประสิทธิ์ความถดถอยมีค่าแตกต่างจากศูนย์

เนื่องจากในกรณีของจำนวนตัวแปรอิสระเท่ากับ 3 ตัว ได้ทำการศึกษาการแจกแจงของตัวแปรอิสระทั้งสิ้น 2 การแจกแจงคือ การแจกแจงปกติ และการแจกแจงแกมมา ซึ่งในกรณีของจำนวนตัวแปรอิสระเท่ากับ 2 ตัวได้ทำการทดสอบแล้วว่าการแจกแจงของตัวแปรอิสระส่งผลต่อค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์ ดังนั้นจึงเป็นที่น่าสนใจว่าการแจกแจงจะมีอิทธิพลต่อค่าพารามิเตอร์ริดจ์ในกรณีของจำนวนตัวแปรอิสระเท่ากับ 3 หรือไม่

<u>ทดสอบว่าการแจกแจงของตัวแปรอิสระส่งผลต่อค่าพารามิเตอร์ริดจ์หรือไม่</u>

การทดสอบในขั้นตอนนี้ใช้สถิติทดสอบ Independent Sample T-test ซึ่งทำการทดสอบในแต่ละค่าสัมประสิทธิ์ความถดถอยโดยมีขั้นตอนการทดสอบเหมือนกัน ดังนี้

<u>ขั้นตอนที่ 1</u> ตรวจสอบความแปรปวนของค่าพารามิเตอร์ริดจ์ของทั้ง 2 การแจกแจงว่ามีค่าเท่ากันหรือไม่

H_0 : ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของการแจกแจงปกติ = ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมา

H_a : ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของการแจกแจงปกติ $\neq$ ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมา

<u>ขั้นตอนที่ 2</u> ทดสอบค่าเฉลี่ยของค่าพารามิเตอร์ริดจ์ของทั้ง 2 การแจกแจงว่ามีค่าเท่ากันหรือไม่ โดยที่ถ้าในขั้นตอนที่ 1 ผลปรากฏว่าความแปรปรวนของทั้ง 2 การแจกแจงมีค่าเท่ากัน ให้ใช้ equal variances assumed ในการทดสอบขั้นตอนที่ 2 ส่วนถ้าผลปรากฏว่าความแปรปรวนของ 2 การแจกแจงมีค่าต่างกันให้ใช้ equal variances not assumed ในการทดสอบ

H_0 : ค่าเฉลี่ยค่าพารามิเตอร์ริดจ์ของการแจกแจงปกติ = ค่าเฉลี่ยค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมา

H_a : ค่าเฉลี่ยค่าพารามิเตอร์ริดจ์ของการแจกแจงปกติ $\neq$ ค่าเฉลี่ยค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมา

ซึ่งผลการวิเคราะห์เป็นดังนี้

ตารางที่ 4.1.19 การทดสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ในการแจกแจงปกติและแกมมา

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	P-value	t	df	P-value	Mean Difference	Std. Error Difference
Beta=2	Equal variances assumed	62.433	0.0000003	6.677	18	0.000003	0.085832	0.012855
	Equal variances not assumed			6.677	9	0.000091	0.085832	0.012855
Beta=3	Equal variances assumed	19.093	0.00037	10.42	18	0.000	0.050996	0.004892
	Equal variances not assumed			10.42	9	0.000003	0.050996	0.004892
Beta=4	Equal variances assumed	12.046	0.003	10.22	18	0.000	0.028666	0.002805
	Equal variances not assumed			10.22	9	0.000003	0.028666	0.002805

จากตารางพบว่า ในขั้นตอนการทดสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ทั้ง 2 การแจกแจงด้วยสถิติ Levene's Test ซึ่งให้ค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 ในทุกค่าสัมประสิทธิ์ความถดถอยเพราะฉะนั้นความแปรปรวนของทั้ง 2 การแจกแจงแตกต่างกัน ดังนั้นในขั้นตอนต่อไปจึงเลือกใช้ผลการวิเคราะห์จาก Equal variances not assumed ซึ่งให้ค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 ในทุกค่าสัมประสิทธิ์ความถดถอย ดังนั้นค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงปกติและการแจกแจงแกมมามีค่าไม่เท่ากัน เพราะฉะนั้นการแจกแจงมีอิทธิพลต่อค่าพารามิเตอร์ริดจ์

ในขั้นตอนต่อไปจะทำการพิจารณาเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์ในแต่ละการแจกแจง เมื่อระดับความสัมพันธ์สูงขึ้นเทียบกับระดับความสัมพันธ์ (0.3,0.5,0.8)

<u>เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์สูงขึ้นในแต่ละขนาดตัวอย่างเทียบกับระดับความสัมพันธ์ (0.3,0.5,0.8)</u>

<u>ตารางที่ 4.1.20</u> เปอร์เซ็นต์การเพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์สูงขึ้น (เทียบกับระดับความสัมพันธ์ (0.3,0.5,0.8)) ในแต่ละการแจกแจง

ค่าสัมประสิทธิ์	ระดับความสัมพันธ์	Normal		Gamma	
		n=60	n=100	n=60	n=100
2	(0.3,0.5,0.8) --> (0.5,0.5,0.8)	80.392	139.630	227.273	200.000*
	(0.3,0.5,0.8) --> (0.5,0.7,0.8)	31.765	316.667	245.455	210.000*
	(0.3,0.5,0.8) --> (0.7,0.7,0.8)	335.294	337.037	536.364	300.000*
	(0.3,0.5,0.8) --> (0.8,0.8,0.8)	386.275	362.963	1172.727	1069.231
	ค่าเฉลี่ย	**208.432**	**289.074**	**545.455**	**444.808**
3	(0.3,0.5,0.8) --> (0.5,0.5,0.8)	71.154	70.370	300.000	110.000
	(0.3,0.5,0.8) --> (0.5,0.7,0.8)	96.154	94.444	660.000	310.000
	(0.3,0.5,0.8) --> (0.7,0.7,0.8)	144.231	144.444	1000.000	510.000
	(0.3,0.5,0.8) --> (0.8,0.8,0.8)	150	153.704	1370	650
	ค่าเฉลี่ย	**115.385**	**115.741**	**832.500**	**395.000**
4	(0.3,0.5,0.8) --> (0.5,0.5,0.8)	117.241	15.789	42.857*	122.222
	(0.3,0.5,0.8) --> (0.5,0.7,0.8)	168.966	28.070	185.714	233.333
	(0.3,0.5,0.8) --> (0.7,0.7,0.8)	179.310	29.825	400.000	455.556
	(0.3,0.5,0.8) --> (0.8,0.8,0.8)	189.655	40.351	614.286	566.667
	ค่าเฉลี่ย	**163.793**	**28.509**	**310.714**	**344.445**

หมายเหตุ: *คือ ค่าเปอร์เซ็นต์การลดลงของค่าพารามิเตอร์ริดจ์ที่มีผลการวิเคราะห์แตกต่างจากกรณีอื่น

การคำนวณค่าเปอร์เซ็นต์ที่เพิ่มขึ้นในตารางด้านบน สามารถคำนวณได้จาก

$$\text{\%เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์} = \left(\frac{\lambda_{\rho} - \lambda_{(0.3,0.5,0.8)}}{\lambda_{(0.3,0.5,0.8)}}\right) \times 100$$

โดยที่ λ_{ρ} คือ ค่าพารามิเตอร์ริดจ์ของระดับความสัมพันธ์ที่ต้องการเปรียบเทียบกับระดับความสัมพันธ์ (0.3,0.5,0.8) ในที่นี้มี 4 ค่า คือ $\lambda_{(0.5,0.5,0.8)}$, $\lambda_{(0.5,0.7,0.8)}$, $\lambda_{(0.7,0.7,0.8)}$ และ $\lambda_{(0.8,0.8,0.8)}$

$\lambda_{(0.3,0.5,0.8)}$ คือ ค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์คือ (0.3,0.5,0.8)

ตัวอย่าง สำหรับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 ในการแจกแจงปกติ ที่ระดับความสัมพันธ์เพิ่มจาก (0.3,0.5,0.8) เป็น (0.5,0.5,0.8) ขนาดตัวอย่างเท่ากับ 60

$$\begin{aligned}\text{\%เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์} &= \left(\frac{\lambda_{\rho} - \lambda_{(0.3,0.5,0.8)}}{\lambda_{(0.3,0.5,0.8)}}\right) \times 100 \\ &= \left(\frac{0.046 - 0.0255}{0.0255}\right) \times 100 \\ &\approx 80.392\%\end{aligned}$$

จากตารางพบว่า เมื่อพิจารณาในแต่ละการแจกแจงเปอร์เซ็นต์การเพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เฉลี่ยเมื่อระดับความสัมพันธ์สูงขึ้นเทียบกับระดับความสัมพันธ์ (0.3,0.5,0.8) ของการแจกแจงแกมมามีค่ามากกว่าการแจกแจงปกติทุกค่าสัมประสิทธิ์ความถดถอย ซึ่งถ้าพิจารณาในแต่ละกรณีย่อยจะพบว่า เปอร์เซ็นต์การเพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อระดับความสัมพันธ์สูงขึ้นเทียบกับระดับความสัมพันธ์ (0.3,0.5,0.8) ของการแจกแจงแกมมามีค่ามากกว่าการแจกแจงปกติเกือบทุกระดับความสัมพันธ์ที่เพิ่มขึ้น ทุกค่าสัมประสิทธิ์ความถดถอยและทุกขนาดตัวอย่าง ยกเว้นกรณีที่มีเครื่องหมาย (*) คือ กรณีที่ค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 ขนาดตัวอย่างเท่ากับ 100 ในทุกระดับความสัมพันธ์ ยกเว้นระดับความสัมพันธ์ที่เพิ่มขึ้นจาก (0.3,0.5,0.8) เป็น (0.8,0.8,0.8) รวมถึงกรณีที่ค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 ขนาดตัวอย่างเท่ากับ 60 และระดับความสัมพันธ์เพิ่มขึ้นจาก (0.3,0.5,0.8) เป็น (0.5,0.5,0.8)

การศึกษาครั้งนี้ได้ทำการศึกษาภายใต้ค่าสัมประสิทธิ์ความถดถอยทั้งสิ้น 3 ค่า คือ 2, 3 และ 4 โดยค่าพารามิเตอร์ริดจ์ที่ได้ในแต่ละค่าสัมประสิทธิ์ความถดถอยมีความแตกต่างกัน จึงทำการทดสอบในลำดับต่อไปว่า ค่าสัมประสิทธิ์ความถดถอยมีอิทธิพลต่อค่าพารามิเตอร์ริดจ์หรือไม่

<u>ทดสอบว่าค่าสัมประสิทธิ์ความถดถอยส่งผลต่อค่าพารามิเตอร์ริดจ์หรือไม่ โดยทำการพิจารณาในแต่ละการแจกแจง</u>

การทดสอบครั้งนี้ทำการทดสอบในแต่ละการแจกแจง โดยทดสอบค่าสัมประสิทธิ์ความถดถอยว่ามีอิทธิพลต่อค่าพารามิเตอร์ริดจ์หรือไม่ด้วยสถิติ One-Way ANOVA ซึ่งขั้นตอนการทดสอบในแต่ละการแจกแจงเหมือนกันดังนี้

<u>ขั้นตอนที่ 1</u> ตรวจสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ในกลุ่มของค่าสัมประสิทธิ์ความถดถอยทั้ง 3 ค่าว่ามีค่าเท่ากันหรือไม่

H_0 : ความแปรปรวนของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 = ความแปรปรวนของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 = ความแปรปรวนของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4

H_a : ความแปรปรวนของค่าพารามิเตอร์ริดจ์อย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าไม่เท่ากัน

<u>ขั้นตอนที่ 2</u> ทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของทั้ง 3 กลุ่มค่าสัมประสิทธิ์ความถดถอยว่ามีค่าเท่ากันหรือไม่ โดยที่

ถ้าในขั้นตอนที่ 1 ผลปรากฏว่าความแปรปรวนของทุกค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากัน ใช้สถิติ F ในการทดสอบขั้นตอนที่ 2 ส่วนถ้าผลปรากฏว่าความแปรปรวนอย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าต่างกันให้ใช้สถิติ Welch ในการทดสอบ

H_0 : ค่าเฉลี่ยพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 = ค่าเฉลี่ยพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 3 = ค่าเฉลี่ยพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4

H_a : ค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าไม่เท่ากัน

<u>ขั้นตอนที่ 3</u> ทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ทีละคู่

ถ้าในขั้นตอนที่ 2 ผลปรากฏว่า ค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าไม่เท่ากัน ให้ทำการทดสอบว่าค่าเฉลี่ยพารามิเตอร์ริดจ์ของค่าสัมประสิทธิ์ความถดถอยคู่ใดมีค่าเฉลี่ยที่แตกต่างกัน โดยถ้าในขั้นตอนที่ 1 ความแปรปรวนของค่าพารามิเตอร์ริดจ์ของทั้ง 3 ค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากันใช้สถิติ Bonferroni ส่วนถ้าความแปรปรวนอย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าไม่เท่ากันใช้สถิติ Games-Howell ในการพิจารณา

ซึ่งผลการวิเคราะห์เป็นดังนี้

ตารางที่ 4.1.21 การทดสอบความแปรปรวนของค่าพารามิเตอร์ริดจ์ในแต่ละค่าสัมประสิทธิ์ความถดถอย

การแจกแจง	Levene Statistic	df1	df2	P-value
Normal	23.380	2	27	0.0000013
Gamma	8.958	2	27	0.00104

จากตารางพบว่า การแจกแจงปกติและการแจกแจงแกมมาให้ค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 เพราะฉะนั้นความแปรปรวนอย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยแตกต่างกัน ดังนั้นในขั้นตอนต่อไปจึงเลือกใช้สถิติ Welch ในการทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของการแจกแจงปกติและการแจกแจงแกมมา ซึ่งให้ผลการทดสอบดังนี้

ตารางที่ 4.1.22 การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของทั้ง 3 ค่าสัมประสิทธิ์ความถดถอยในการแจกแจงปกติและแกมมา

การแจกแจง	Welch Statistic	df1	df2	P-value
Normal	14.983	2	15.131	0.00026
Gamma	15.773	2	12.067	0.00043

จากตารางพบว่า ทั้งสองการแจกแจงให้ค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 ดังนั้นค่าเฉลี่ยพารามิเตอร์ริดจ์อย่างน้อย 2 ค่าสัมประสิทธิ์ความถดถอยมีค่าแตกต่างกัน จึงทำการพิจารณาในขั้นตอนต่อไปว่าค่าสัมประสิทธิ์ความถดถอยคู่ใดบ้างที่ค่าเฉลี่ยพารามิเตอร์ริดจ์มีความแตกต่างกัน โดยใช้ Multiple Comparison และวิเคราะห์ด้วยสถิติ Games-Howell สำหรับทั้งสองการแจกแจงได้ผลการวิเคราะห์ดังนี้

ตารางที่ 4.1.23 การทดสอบค่าเฉลี่ยพารามิเตอร์ริดจ์ของค่าสัมประสิทธิ์ความถดถอยแต่ละคู่

Dependent Variable		(I) dist	(J) dist	Mean Difference (I-J)	Std. Error	P-value
Normal	Games-Howell	beta2	beta3	0.03487	0.013754	0.064
			beta4	0.057203	0.013157	0.004
		beta3	beta2	-0.03487	0.013754	0.064
			beta4	0.022333	0.005639	0.004
		beta4	beta2	-0.057203	0.013157	0.004
			beta3	-0.022333	0.005639	0.004
Gamma	Games-Howell	beta2	beta3	0.0000336	1.057E-05	0.027
			beta4	0.0000373	1.054E-05	0.016
		beta3	beta2	-0.0000336	1.057E-05	0.027
			beta4	0.0000037	8.122E-07	0.003
		beta4	beta2	-0.0000373	1.054E-05	0.016
			beta3	-0.0000037	8.122E-07	0.003

จากตารางพบว่า การแจกแกมมาในทุกคู่ของค่าสัมประสิทธิ์ความถดถอยมีค่า P-value น้อยกว่าค่าระดับนัยสำคัญ 0.05 แสดงว่าค่าเฉลี่ยพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยเป็น 2,3 และ 4 มีความแตกต่างกัน ส่วนการแจกแจงปกติในคู่ของค่าสัมประสิทธิ์ความถดถอยที่มีค่า 2 กับ 3 ค่า P-value ที่ได้มีค่ามากกว่าระดับนัยสำคัญ 0.05 แสดงว่าค่าสัมประสิทธิ์ความถดถอยมีอิทธิพลต่อค่าเฉลี่ยพารามิเตอร์ริดจ์ในการแจกแจงแกมมาและการแจกแจงปกติ ยกเว้นค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 และ 3 ของการแจกแจงปกติ

จากผลการทดสอบดังกล่าวแสดงให้เห็นว่าค่าสัมประสิทธิ์ความถดถอยส่งผลต่อค่าพารามิเตอร์ริดจ์ และเมื่อทำการพิจารณาค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยสูงขึ้น พบว่า ค่าพารามิเตอร์ริดจ์มีค่าลดลง จึงทำการพิจารณาเปอร์เซ็นต์ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าสูงขึ้นเทียบกับค่าสัมประสิทธิ์ความถดถอยเท่ากับ

<u>เปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้นในแต่ละระดับความสัมพันธ์และขนาดตัวอย่าง</u>

<u>ตารางที่ 4.1.24</u> เปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้นจำแนกตามการแจกแจง

	ระดับความสัมพันธ์	**ขนาดตัวอย่าง**	Normal	Gamma
ค่าสัมประสิทธิ์ความถดถอยจาก 2 ไป 3	(0.3, 0.5, 0.8)	60	-1.961*	90.909
		100	0.000	90.000
	(0.5, 0.5, 0.8)	60	3.261	88.889
		100	28.903	93.000
	(0.5, 0.7, 0.8)	60	51.429	80.000
		100	53.333	86.774
	(0.7, 0.7, 0.8)	60	42.793	84.286
		100	44.068	84.750
	(0.8, 0.8, 0.8)	60	47.581	89.500
		100	45.200	93.586
Mean			**31.461**	**88.169**
ค่าสัมประสิทธิ์ความถดถอยจาก 2 ไป 4	(0.3, 0.5, 0.8)	60	62.092	98.727
		100	-5.556	99.100
	(0.5, 0.5, 0.8)	60	54.348	99.444
		100	48.995	99.333
	(0.5, 0.7, 0.8)	60	75.238	98.947
		100	67.556	99.032
	(0.7, 0.7, 0.8)	60	75.676	99.000
		100	68.644	98.750
	(0.8, 0.8, 0.8)	60	77.419	99.286
		100	68.000	99.487
Mean			**59.241**	**99.111**

หมายเหตุ: *หมายถึง กรณีที่แตกต่างจากกรณีอื่นคือ ค่าพารามิเตอร์ริดจ์ของค่าสัมประสิทธิ์ความถดถอยเท่ากับ 3 มีค่าสูงกว่าค่าพารามิเตอร์ริดจ์ของค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2

การคำนวณค่าเปอร์เซ็นต์ที่ลดลงในตารางด้านบน สามารถคำนวณได้จาก

$$\%\text{ลดลงของค่าพารามิเตอร์ริดจ์} = \left(\frac{\lambda_{beta2} - \lambda_{b}}{\lambda_{beta2}}\right)\times 100$$

โดยที่ λ_{beta2} คือ ค่าพารามิเตอร์ริดจ์ของกรณีที่สนใจเมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2

λ_{b} คือ ค่าพารามิเตอร์ริดจ์ในกรณีที่สนใจของค่าสัมประสิทธิ์ความถดถอยที่ต้องการเปรียบเทียบกับค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 ในที่นี้มี 2 ค่า คือ λ_{beta3} และ λ_{beta4}

<u>ตัวอย่าง</u> สำหรับการแจกแจงแกมมาที่ระดับความสัมพันธ์ (0.3,0.5,0.8) ขนาดตัวอย่างมีค่า 60 โดยพิจารณาเปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์จากค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 ไปยังค่าสัมประสิทธิ์ความถดถอยเท่ากับ 3

$$\begin{aligned}\%\text{ลดลงของค่าพารามิเตอร์ริดจ์} &= \left(\frac{\lambda_{beta2} - \lambda_{beta3}}{\lambda_{beta2}}\right)\times 100 \\ &= \left(\frac{0.0000055 - 0.0000005}{0.0000055}\right)\times 100 \\ &\approx 90.909\%\end{aligned}$$

จากตารางด้านบนพบว่า เปอร์เซ็นต์การลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้นจาก 2 เป็น 3 และ จาก 2 เป็น 4 ของการแจกแจงแกมมามีค่ามากกว่าการแจกแจงปกติ โดยถ้ามองในแต่ละการแจกแจงพบว่า การแจกแจงแกมมามีเปอร์เซ็นต์การลดลงเฉลี่ยของค่าพารามิเตอร์ริดจ์ประมาณ 88.169% สำหรับค่าสัมประสิทธิ์ความถดถอยเพิ่มขึ้นจาก 2 เป็น 3 และเปอร์เซ็นต์การลดลงเฉลี่ยของค่าพารามิเตอร์ริดจ์มีค่าประมาณ 99.111% สำหรับค่าสัมประสิทธิ์ความถดถอยเพิ่มขึ้นจาก 2 เป็น 4 ซึ่งถ้าพิจารณาในแต่ละระดับความสัมพันธ์และขนาดตัวอย่างพบว่า การแจกแจงแกมมามีเปอร์เซ็นต์การลดลงเฉลี่ยของค่าพารามิเตอร์ริดจ์ใกล้เคียงกัน ส่วนการแจกแจงปกติ

มีเปอร์เซ็นต์การลดลงของค่าพารามิเตอร์ริดจ์ค่อนข้างแตกต่างกันทำให้เห็นแนวโน้มการลดลงที่ไม่ชัดเจน

การศึกษาครั้งนี้นอกจากทำการพิจารณาในมุมมองของระดับความสัมพันธ์และค่าสัมประสิทธิ์ความถดถอยแล้ว ยังมีการพิจารณาถึงขนาดตัวอย่าง ทั้งนี้มีการพิจารณาขนาดตัวอย่างใน 2 ระดับคือ 60 และ 100 ซึ่งเมื่อขนาดตัวอย่างเพิ่มขึ้นทำให้ค่าพารามิเตอร์ริดจ์ที่ได้มีค่าสูงขึ้น จึงทำการพิจารณาเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างมีค่าสูงขึ้นในแต่ละการแจกแจง ได้ผลดังนี้

เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 60 เป็น 100

ตารางที่ 4.1.25 เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างมีค่าสูงขึ้น

ค่าสัมประสิทธิ์ความถดถอย	ระดับความสัมพันธ์	Normal	Gamma
2	(0.3, 0.5, 0.8)	5.882	81.818
	(0.5, 0.5, 0.8)	40.652	66.667
	(0.5, 0.7, 0.8)	7.143	63.158
	(0.7, 0.7, 0.8)	6.306	14.286
	(0.8, 0.8, 0.8)	0.806	67.033
	ค่าเฉลี่ย	**12.158**	**58.592**
3	(0.3, 0.5, 0.8)	3.846	100.000
	(0.5, 0.5, 0.8)	3.371	5.000
	(0.5, 0.7, 0.8)	2.941	7.895
	(0.7, 0.7, 0.8)	3.937	10.909
	(0.8, 0.8, 0.8)*	5.385	2.041
	ค่าเฉลี่ย	**3.896**	**25.169**
4	(0.3, 0.5, 0.8)*	194.828	28.571
	(0.5, 0.5, 0.8)	57.143	100.000
	(0.5, 0.7, 0.8)	40.385	50.000
	(0.7, 0.7, 0.8)	37.037	42.857
	(0.8, 0.8, 0.8)*	42.857	20.000
	ค่าเฉลี่ย	**74.450**	**48.286**

หมายเหตุ: *คือ ค่าในตารางที่มีผลการวิเคราะห์แตกต่างจากกรณีอื่น

จากตารางด้านบนพบว่า ถ้าพิจารณาค่าสัมประสิทธิ์ความถดถอยเท่ากับ 2 และ 3 เปอร์เซ็นต์ของค่าพารามิเตอร์ริดจ์ที่เพิ่มขึ้นเมื่อขนาดตัวอย่างสูงขึ้นของการแจกแจงแกมมาสูงกว่าการแจกแจงปกติในทุกระดับความสัมพันธ์ ยกเว้นระดับความสัมพันธ์ (0.8, 0.8, 0.8) ส่วนกรณีค่าสัมประสิทธิ์ความถดถอยเท่ากับ 4 เปอร์เซ็นต์ของค่าพารามิเตอร์ริดจ์ของการแจกแจงแกมมาสูงกว่าการแจกแจงปกติในทุกระดับความสัมพันธ์ยกเว้น ระดับความสัมพันธ์ (0.3, 0.5, 0.8) และ (0.8, 0.8, 0.8) แต่ในภาพรวมแล้วเปอร์เซ็นต์เฉลี่ยของค่าพารามิเตอร์ริดจ์ที่เพิ่มขึ้นเมื่อขนาดตัวอย่างสูงขึ้นของการแจกแจงแกมมาสูงกว่าการแจกแจงปกติในกรณีค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 และ 3 และเปอร์เซ็นต์เฉลี่ยของค่าพารามิเตอร์ริดจ์ที่เพิ่มขึ้นเมื่อขนาดตัวอย่างสูงขึ้นของการแจกแจงปกติสูงกว่าการแจกแจงแกมมาในกรณีค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4

บทที่ 5

สรุปผลการวิจัย และข้อเสนอแนะ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อหาค่าพารามิเตอร์ริดจ์ที่เหมาะสมสำหรับแก้ปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระในการวิเคราะห์ความถดถอยโลจิสติกแบบสองกลุ่ม ซึ่งดำเนินการภายใต้เงื่อนไข และขอบเขตดังนี้

1. ศึกษาการวิเคราะห์ความถดถอยโลจิสติกที่มีตัวแปรตาม (Y) มีค่าเพียง 2 ค่า (Binary Logistic Regression)
2. กำหนดตัวแปรอิสระ (X) เป็นตัวแปรเชิงปริมาณเท่านั้น และทำการศึกษาในกรณีที่มีตัวแปรอิสระ (p) จำนวน 2 และ 3 ตัว โดยที่ตัวแปรอิสระมีการแจกแจงแบบเดียวกันทุกตัว และมีความสัมพันธ์ระหว่างตัวแปรอิสระเป็นดังนี้

2.1 กรณีตัวแปรอิสระ 2 ตัว (p = 2)

ให้ ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 2 โดยทำการศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

2.1.1 ระดับความสัมพันธ์ 0.3

2.1.2 ระดับความสัมพันธ์ 0.5

2.1.3 ระดับความสัมพันธ์ 0.7

2.1.4 ระดับความสัมพันธ์ 0.8

2.2 กรณีตัวแปรอิสระ 3 ตัว (p = 3)

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

ρ_{12} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 2

ρ_{13} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 1 และตัวแปรอิสระตัวที่ 3

ρ_{23} เป็นระดับความสัมพันธ์ระหว่างตัวแปรอิสระตัวที่ 2 และตัวแปรอิสระตัวที่ 3

ทำการศึกษาภายใต้ระดับความสัมพันธ์ ดังนี้

2.2.1 ระดับความสัมพันธ์ (0.3, 0.5, 0.8)

2.2.2 ระดับความสัมพันธ์ (0.5, 0.5, 0.8)

2.2.3 ระดับความสัมพันธ์ (0.5, 0.7, 0.8)

2.2.4 ระดับความสัมพันธ์ (0.7, 0.7, 0.8)

2.2.5 ระดับความสัมพันธ์ (0.8, 0.8, 0.8)

ซึ่งทำการศึกษาภายใต้การแจกแจง 3 การแจกแจง ดังนี้

2.3 การแจกแจงแบบต่อเนื่อง

2.3.1 การแจกแจงปกติหลายตัวแปร (Multivariate Normal) ที่มีพารามิเตอร์ค่าเฉลี่ย $(\tilde{\mu})$ เป็น $\tilde{0}$ และเมทริกซ์ความแปรปรวนร่วม (Σ) มีความแปรปรวนร่วมระหว่างตัวแปรอิสระตามระดับความสัมพันธ์ของตัวแปรอิสระ ดังที่ได้กล่าวมา

2.3.2 การแจกแจงแกมมา (Gamma Distribution) ที่มีพารามิเตอร์คือ (γ,α) เท่ากับ (1,2) สำหรับตัวแปรอิสระตัวที่ 1 (γ,α) เท่ากับ (1.5,2) สำหรับตัวแปรอิสระตัวที่ 2 และ (γ,α) เท่ากับ (1,2) สำหรับตัวแปรอิสระตัวที่ 3 โดยทำการศึกษาเมื่อตัวแปรอิสระมีระดับความสัมพันธ์ตามที่กล่าวมา

2.4 การแจกแจงแบบไม่ต่อเนื่อง

2.4.1 การแจกแจงปัวซองส์ (Poisson Distribution) ที่มีพารามิเตอร์คือค่าเฉลี่ย (θ) เท่ากับ 2 สำหรับตัวแปรอิสระตัวที่ 1 และ (θ) เท่ากับ 3 สำหรับตัวแปรอิสระตัวที่ 2 โดยทำการศึกษาเมื่อตัวแปรอิสระมีระดับความสัมพันธ์ตามที่กล่าวมา

3. ศึกษาภายใต้ขนาดตัวอย่างที่มีจำนวนตัวอย่างอย่างน้อย 20 เท่าของจำนวนตัวแปรอิสระ (20p) คือ

3.1 ถ้าจำนวนตัวแปรอิสระเท่ากับ 2 จำนวนตัวอย่างที่ทำการศึกษาคือ 40 และ 70

3.2 ถ้าจำนวนตัวแปรอิสระเท่ากับ 3 จำนวนตัวอย่างที่ทำการศึกษาคือ 60 และ100

4. กำหนดค่าเริ่มต้นของสัมประสิทธิ์ความถดถอย $(\beta_0, \beta_1, ..., \beta_p)$ ดังนี้
   4.1 กรณีตัวแปรอิสระ 2 ตัว (p = 2) ค่าสัมประสิทธิ์ความถดถอยเริ่มต้นมีค่าเป็น $(\beta_0, \beta_1, \beta_2) \in [2,4]$
   4.2 กรณีตัวแปรอิสระ 3 ตัว (p = 3) ค่าสัมประสิทธิ์ความถดถอยเริ่มต้นมีค่าเป็น $(\beta_0, \beta_1, \beta_2, \beta_3) \in [2,4]$

การวิเคราะห์ผลการศึกษาดำเนินการโดยใช้โปรแกรม R และสามารถสรุปผลการศึกษาในกรณีต่างๆ ได้ดังนี้

5.1 สรุปผลการศึกษา

ในการสรุปผลการศึกษาจะทำการพิจารณาค่าพารามิเตอร์ริดจ์โดยมุ่งประเด็นไปที่ขอบเขตการศึกษาที่มีอิทธิพลต่อค่าพารามิเตอร์ริดจ์ซึ่งได้แก่ ระดับความสัมพันธ์ การแจกแจงของตัวแปรอิสระ ค่าสัมประสิทธิ์ความถดถอย และขนาดตัวอย่าง ซึ่งสรุปผลการวิเคราะห์ได้ดังนี้

กรณีจำนวนตัวแปรอิสระเท่ากับ 2 ตัว

<u>พิจารณาตามระดับความสัมพันธ์</u>

ระดับความสัมพันธ์ของการศึกษาถูกแบ่งออกเป็น 4 ระดับความสัมพันธ์คือ 0.3, 0.5, 0.7 และ 0.8 ซึ่งเมื่อระดับความสัมพันธ์สูงขึ้น ค่าพารามิเตอร์ริดจ์มีค่าสูงขึ้นด้วย โดยมีแนวโน้มเช่นนี้ในทุกการแจกแจง ค่าสัมประสิทธิ์ความถดถอย และขนาดตัวอย่างที่ทำการศึกษา

<u>พิจารณาตามการแจกแจงของตัวแปรอิสระ</u>

การแจกแจงของตัวแปรอิสระที่ทำการศึกษามีทั้งสิ้น 3 การแจกแจงคือ การแจกแจงปัวซองส์ การแจกแจงปกติ และการแจกแจงแกมมา ซึ่งได้ทำการทดสอบแล้วว่าการแจกแจงมีอิทธิพลต่อค่าพารามิเตอร์ริดจ์ และได้ทำการพิจารณาเปอร์เซ็นต์ของค่าพารามิเตอร์ริดจ์ที่เพิ่มขึ้นเมื่อนำระดับ

ความสัมพันธ์ต่างๆ เทียบกับระดับความสัมพันธ์ 0.3 ผลปรากฏว่า เปอร์เซ็นต์การเพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เรียงลำดับจากมากไปน้อยได้ดังนี้ การแจกแจงแกมมา การแจกแจงปกติ และการแจกแจงปัวซองส์ ตามลำดับ

พิจารณาตามค่าสัมประสิทธิ์ความถดถอย

ค่าสัมประสิทธิ์ความถดถอยที่ใช้ในการศึกษาครั้งนี้มีทั้งสิ้น 3 ค่า คือ 2, 3 และ 4 ซึ่งผลการทดสอบพบว่า ค่าสัมประสิทธิ์ความถดถอยส่งผลต่อค่าพารามิเตอร์ริดจ์ อีกทั้งเมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าสูงขึ้นจะทำให้ค่าพารามิเตอร์ริดจ์มีค่าลดลง ดังนั้นจึงทำการพิจารณาเปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้นได้ผลว่า การแจกแจงที่มีเปอร์เซ็นต์ที่ลดลงเฉลี่ยเรียงตามลำดับจากมากไปน้อยได้ดังนี้ การแจกแจงแกมมา การแจกแจงปัวซองส์ และการแจกแจงปกติ ตามลำดับ

พิจารณาตามขนาดตัวอย่าง

การศึกษาครั้งนี้ขนาดตัวอย่างที่ทำการศึกษามีทั้งสิ้น 2 ระดับคือ 40 และ 70 โดยเมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้นทำให้ค่าพารามิเตอร์ริดจ์มีค่าสูงขึ้น เป็นลักษณะเช่นนี้ในทุกการแจกแจง และค่าสัมประสิทธิ์ความถดถอยที่ทำการศึกษา ทั้งนี้ได้ทำการพิจารณาเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 40 เป็น 70 ของแต่ละการแจกแจงซึ่งให้ผลการพิจารณาเปอร์เซ็นต์เรียงลำดับจากมากที่สุดไปน้อยที่สุด ได้ดังนี้

- ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 และ 3 เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เรียงลำดับจากมากไปน้อยตามการแจกแจงได้ดังนี้ การแจกแจงแกมมา การแจกแจงปกติ และการแจกแจงปัวซองส์ ตามลำดับ
- ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เรียงลำดับจากมากไปน้อยตามการแจกแจงได้ดังนี้ การแจกแจงปกติ การแจกแจงแกมมา และการแจกแจงปัวซองส์ ตามลำดับ

กรณีจำนวนตัวแปรอิสระเท่ากับ 3 ตัว

พิจารณาตามระดับความสัมพันธ์

ระดับความสัมพันธ์ของการศึกษาถูกแบ่งออกเป็น 5 ระดับความสัมพันธ์คือ (0.3, 0.5, 0.8), (0.5, 0.5, 0.8), (0.5, 0.7, 0.8), (0.7, 0.7, 0.8) และ (0.8, 0.8, 0.8) ซึ่งเมื่อระดับความสัมพันธ์สูงขึ้นค่าพารามิเตอร์ริดจ์มีค่าสูงขึ้นด้วย โดยมีแนวโน้มเช่นนี้ในทุกการแจกแจง ค่าสัมประสิทธิ์ความถดถอยและขนาดตัวอย่างที่ทำการศึกษา

พิจารณาตามการแจกแจงของตัวแปรอิสระ

การแจกแจงของตัวแปรอิสระที่ทำการศึกษามีทั้งสิ้น 2 การแจกแจงคือ การแจกแจงปกติ และการแจกแจงแกมมา ซึ่งได้ทำการทดสอบแล้วว่าการแจกแจงมีอิทธิพลต่อค่าพารามิเตอร์ริดจ์ และได้ทำการพิจารณาเปอร์เซ็นต์ของค่าพารามิเตอร์ริดจ์ที่เพิ่มขึ้นเมื่อนำระดับความสัมพันธ์ต่างๆ เทียบกับระดับความสัมพันธ์ (0.3,0.5,0.8) ผลปรากฏว่า เปอร์เซ็นต์การเพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เรียงลำดับจากมากไปน้อยได้ดังนี้ การแจกแจงแกมมา และการแจกแจงปกติ ตามลำดับ

พิจารณาตามค่าสัมประสิทธิ์ความถดถอย

ค่าสัมประสิทธิ์ความถดถอยที่ใช้ในการศึกษาครั้งนี้มีทั้งสิ้น 3 ค่า คือ 2, 3 และ 4 ซึ่งผลการทดสอบพบว่า ค่าสัมประสิทธิ์ความถดถอยส่งผลต่อค่าพารามิเตอร์ริดจ์ อีกทั้งเมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าสูงขึ้นจะทำให้ค่าพารามิเตอร์ริดจ์มีค่าลดลง ดังนั้นจึงทำการพิจารณาเปอร์เซ็นต์ที่ลดลงของค่าพารามิเตอร์ริดจ์เมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้นได้ผลว่า การแจกแจงที่มีเปอร์เซ็นต์ที่ลดลงเฉลี่ยเรียงตามลำดับจากมากไปน้อยได้ดังนี้ การแจกแจงแกมมา และการแจกแจงปกติ ตามลำดับ

<u>พิจารณาตามขนาดตัวอย่าง</u>

การศึกษาครั้งนี้ขนาดตัวอย่างที่ทำการศึกษามีทั้งสิ้น 2 ระดับคือ 60 และ 100 โดยเมื่อขนาดตัวอย่างมีค่าเพิ่มขึ้นทำให้ค่าพารามิเตอร์ริดจ์มีค่าสูงขึ้น เป็นลักษณะเช่นนี้ในทุกการแจกแจง และค่าสัมประสิทธิ์ความถดถอยที่ทำการศึกษา ทั้งนี้ได้ทำการพิจารณาเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เมื่อขนาดตัวอย่างเพิ่มขึ้นจาก 60 เป็น 100 ของแต่ละการแจกแจงซึ่งให้ผลการพิจารณาเปอร์เซ็นต์เรียงลำดับจากมากที่สุดไปน้อยที่สุด ได้ดังนี้

- ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 2 และ 3 เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เรียงลำดับจากมากไปน้อยตามการแจกแจงได้ดังนี้ การแจกแจงแกมมา และการแจกแจงปกติ ตามลำดับ
- ถ้าค่าสัมประสิทธิ์ความถดถอยมีค่าเท่ากับ 4 เปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์เรียงลำดับจากมากไปน้อยตามการแจกแจงได้ดังนี้ การแจกแจงปกติ และการแจกแจงแกมมา ตามลำดับ

จากผลสรุปดังกล่าวจะพบว่า กรณีของจำนวนตัวแปรอิสระเท่ากับ 2 ตัว และกรณีของจำนวนตัวแปรอิสระเท่ากับ 3 ตัว ให้ผลสอดคล้องไปในทิศทางเดียวกัน ดังนั้นจึงทำการสรุปผลการศึกษาทุกกรณีได้เป็นแผนภาพเพื่อการนำไปประยุกต์ใช้งานต่อ ได้ดังนี้


เกิดปัญหาพหุสัมพันธ์ระหว่างตัวแปรอิสระ
p = 2
p = 3
ระดับความสัมพันธ์ (ρ) ค่าพารามิเตอร์ริดจ์ (λ)
%ค่า λ ที่เพิ่มขึ้นในระดับความสัมพันธ์ต่างๆ เทียบกับ ρ = 0.3
→ Gamma > Normal > Poisson
%ค่า λ ที่เพิ่มขึ้นในระดับความสัมพันธ์ต่างๆ เทียบกับ ρ = (0.3,0.5,0.8)
→ Gamma > Normal
ขนาดตัวอย่าง (n) ค่าพารามิเตอร์ริดจ์ (λ)
%ค่า λ ที่เพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น
- ค่าสัมประสิทธิ์ความถดถอย = 2 และ 3
→ Gamma > Normal > Poisson
- ค่าสัมประสิทธิ์ความถดถอย = 4
→ Normal > Gamma > Poisson
%ค่า λ ที่เพิ่มขึ้นเมื่อขนาดตัวอย่างเพิ่มขึ้น
- ค่าสัมประสิทธิ์ความถดถอย= 2 และ 3
→ Gamma > Normal
- ค่าสัมประสิทธิ์ความถดถอย = 4
→ Normal > Gamma
ค่าสัมประสิทธิ์ความถดถอย ค่าพารามิเตอร์ริดจ์ (λ)
%ค่า λ ที่ลดลงเมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้น
→ Gamma > Poisson > Normal
%ค่า λ ที่ลดลงเมื่อค่าสัมประสิทธิ์ความถดถอยมีค่าเพิ่มขึ้น
→ Gamma > Normal

5.2 แนวทางการศึกษาต่อ

การศึกษาครั้งนี้ได้ทำการศึกษาภายใต้ขอบเขตของระดับความสัมพันธ์ ขนาดตัวอย่าง การแจกแจงของตัวแปรอิสระ และค่าสัมประสิทธิ์ความถดถอยเริ่มต้น แต่ได้ทำการศึกษาเพียงแค่บางกรณีของเงื่อนไขเหล่านี้ ดังนั้นเพื่อประโยชน์ในการใช้งานต่อไปในอนาคตจึงควรมีการศึกษาเพิ่มเติมโดยขยายขอบเขตให้กว้างขึ้น และครอบคลุมกรณีต่างๆ มากขึ้น เพื่อสามารถหาแนวโน้มที่ชัดเจนของค่าพารามิเตอร์ริดจ์ เพราะฉะนั้นทางผู้ศึกษาจึงได้เสนอข้อแนะนำในการศึกษาต่อในกรณีต่างๆ ดังนี้

5.2.1 ศึกษาระดับความสัมพันธ์ที่ละเอียดขึ้นเพื่อทำการพิจารณาค่าพารามิเตอร์ริดจ์ที่ได้ และจะเป็นประโยชน์ต่อการนำไปปรับใช้เมื่อข้อมูลเกิดระดับความสัมพันธ์นอกเหนือจากระดับความสัมพันธ์ที่ได้ศึกษาไว้

5.2.2 ศึกษาขนาดตัวอย่างที่สูงขึ้น เพื่อพิจารณาเปอร์เซ็นต์ที่เพิ่มขึ้นของค่าพารามิเตอร์ริดจ์รวมทั้งหาแนวโน้มการเพิ่มขึ้นว่ามากน้อยเพียงใดเพื่อการประยุกต์ใช้ในขนาดตัวอย่างอื่นๆ

5.2.3 ศึกษาการแจกแจงอื่นๆ ที่มีลักษณะเบ้ขวา สมมาตร และ ไม่ต่อเนื่อง เพื่อทำการเปรียบเทียบกับการแจกแจงแกมมา ปกติ และปัวซองส์ ว่ายังคงมีแนวโน้มการเพิ่มขึ้นหรือลดลงของค่าพารามิเตอร์ริดจ์ในกรณีต่างๆ เช่นเดียวกันอยู่หรือไม่ ซึ่งถ้ายังคงมีแนวโน้มเช่นเดียวกับกรณีที่ศึกษาในงานวิจัยครั้งนี้ ก็จะสามารถนำไปปรับใช้กับการแจกแจงอื่นๆ ที่มีลักษณะคล้ายกับการแจกแจงเหล่านี้ได้

รายการอ้างอิง

Duffy, D. E. and Santner, T. J., On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models., Communs Statist. Theory Meth., 18, (1989): 959-980.

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

S.LE Cessie and J.C.Van Houwelinge, “Ridge Regression in Logistic Regression,” Journal of the Royal Statistical Society. Series C (Applied Statistics) 41, 1 (1990): 191-201.

Furman E., On a multivariate gamma distribution, Statistics and Probability Letters (2008).

Dimitris Karlis, Multivariate Poisson models (2002).

บรรณานุกรม

กัลยา วานิชย์บัญชา, การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล, พิมพ์ครั้งที่ 16 กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2553.

ปราโมทย์ เดชะอำไพ, ระเบียบวิธีเชิงตัวเลขในงานวิศวกรรม, พิมพ์ครั้งที่ 5 กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2549.

กรรณิกา โกวเครือ, การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติก, วิทยานิพนธ์ปริญญามหาบัณฑิต, คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์, 2548.

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวสาวิตรี บุญพัชรนนท์ เกิดวันที่ 11 มิถุนายน พ.ศ.2529 สำเร็จการศึกษาปริญญาสถิติศาสตรบัณฑิต (สต.บ.) สาขาวิชาสถิติคณิตศาสตร์ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2550 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต (สต.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2552