การเปรียบเทียบประสิทธิภาพของวิธีการอนุมานในตัวแบบถดถอยลอจิสติก

โดย

นางสาวจารุวรรณ เหมือนเงิน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาภาควิชาสถิติ
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
ปีการศึกษา 2559

การเปรียบเทียบประสิทธิภาพของวิธีการอนุมานในตัวแบบถดถอยลอจิสติก

โดย
นางสาวจารุวรรณ เหมือนเงิน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
ภาควิชาภาควิชาสถิติ
บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร
ปีการศึกษา 2559

EFFICIENT COMPARISON OF INFERENCE METHODS IN LOGISTIC REGRESSION MODEL

By

MISS Jaruwan MERNNGURN

A Thesis Submitted in partial Fulfillment of Requirements
for Master of Science (APPLIED STATISTICS)
Department of STATISTICS
Graduate School, Silpakorn University
Academic Year 2016

หัวข้อ	การเปรียบเทียบประสิทธิภาพของวิธีการอนุมานในตัวแบบถดถอยลอจิสติก
โดย	จารุวรรณ เหมือนเงิน
สาขาวิชา	สถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
อาจารย์ที่ปรึกษาหลัก	ผู้ช่วยศาสตราจารย์ ดร. ประหยัด แสงงาม

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร ได้รับพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

.... คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.ปานใจ ธารทัศนวงศ์)

พิจารณาเห็นชอบโดย

.... ประธานกรรมการ

(ดร. วิภาวรรณ เล้าอรุณ)

.... อาจารย์ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร. ประหยัด แสงงาม)

.... ผู้ทรงคุณวุฒิภายนอก

(ผู้ช่วยศาสตราจารย์ ดร. มานะชัย รอดชื่น)

58304201 : สถิติประยุกต์ แผน ก แบบ ก 2 ระดับปริญญามหาบัณฑิต
คำสำคัญ : การถดถอยลอจิสติก, การถดถอยลอจิสติกแบบแม่นตรง, ลูกโซ่มาร์คอฟมอนติคาร์โล, ข้อมูลแบบแบ่งแยกสมบูรณ์

นางสาว จารุวรรณ เหมือนเงิน: การเปรียบเทียบประสิทธิภาพของวิธีการอนุมานในตัวแบบถดถอยลอจิสติก อาจารย์ที่ปรึกษาวิทยานิพนธ์ : ผู้ช่วยศาสตราจารย์ ดร. ประหยัด แสงงาม

ตัวแบบการถดถอยลอจิสติกได้นำมาใช้อย่างกว้างขวางในงานวิจัยหลายแขนง เพื่ออธิบายความเกี่ยวพันระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง ในกรณีที่ตัวแปรตอบสนองเป็นตัวแปรจำแนกประเภท โดยทั่วไปวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกที่นิยมใช้ คือ วิธีภาวะน่าจะเป็นสูงสุด (ML) แต่จากการศึกษาพบว่าวิธีดังกล่าวไม่สามารถใช้ในการประมาณค่าพารามิเตอร์ในกรณีข้อมูลเป็นแบบแบ่งแยกสมบูรณ์ และข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ งานวิจัยนี้ศึกษาวิธีการประมาณค่าพารามิเตอร์วิธีอื่นที่สามารถแก้ไขข้อบกพร่องของวิธีภาวะน่าจะเป็นสูงสุด ได้แก่ วิธีการถดถอยลอจิสติกแบบแม่นตรง (ELR) และวิธีลูกโซ่มาร์คอฟมอนติคาร์โล (MCMC) สำหรับการถดถอยลอจิสติกแบบแม่นตรง จากนั้นทำการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ทั้งสามวิธี เมื่อข้อมูลมีจำนวนค่าสังเกตทับซ้อนต่างกัน และกรณีที่ข้อมูลมีร้อยละของค่าตอบสนองต่างกัน จากการศึกษาพบว่า

กรณีตัวแปรอธิบายแบบต่อเนื่อง 1 ตัวแปร: สำหรับทุกร้อยละของค่าตอบสนอง เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมีค่าต่ำกว่าหรือเท่ากับ 4 วิธีการประมาณค่าพารามิเตอร์ทั้ง 3 วิธี ยังมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ต่ำ เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมีค่าสูงกว่า 4 วิธี ML และวิธี ELR มีประสิทธิภาพใกล้เคียงกัน ส่วนวิธี MCMC มีประสิทธิภาพต่ำสุด

กรณีตัวแปรอธิบายแบบไม่ต่อเนื่องทั้ง 2 ตัวแปร: เมื่อค่าประมาณร้อยละของค่าตอบสนองต่ำกว่า 50 วิธี ML และ ELR ยังมีประสิทธิภาพต่ำในการประมาณค่าพารามิเตอร์ เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 50 ที่ขนาดตัวอย่างน้อยกว่าหรือเท่ากับ 28 วิธี ELR มีประสิทธิภาพสูงกว่าวิธี ML ส่วนที่ขนาดตัวอย่างมากกว่าหรือเท่ากับ 48 ทั้งสองวิธีมีประสิทธิภาพใกล้เคียงกัน

การประมาณค่าพารามิเตอร์แบบช่วงของทั้ง 2 กรณี พบว่า วิธีที่ให้ค่าประมาณค่าความน่าจะเป็นครอบคลุมไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนดและให้ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุด คือวิธี ML เนื่องจากวิธี ELR และวิธี MCMC โดยส่วนใหญ่ให้ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นเป็นอนันต์

58304201 : Major (APPLIED STATISTICS)
Keyword : Logistic Regression, Exact Logistic Regression, Markov Chain Monte Carlo, Completely Separate

MISS Jaruwan MERNNGURN: Efficient Comparison of Inference Methods in Logistic Regression Model Thesis advisor : Assistant Professor Dr. Prayad Sangngam, Ph.D.

Logistic regression model is used to explain the relationship between explanatory variables and categorical response variable in many research fields. The maximum likelihood (ML) was generally used to estimate parameters, but the ML shows very poor results in the case of separate data. Exact Logistic Regression (ELR) and Markov chain monte carlo (MCMC) exact inference are the logical alternative to the ML. This research offers a comparison of the property of three methods for estimation. The study found that;

In case of one continuous explanatory variable, when the percentage of the overlapping is less than or equal to 4, the three estimation methods have low efficient in parameter estimation. When the percentage of overlapping is higher than 4, the ML and ELR are equally efficient but the MCMC has the lowest efficiency.

In case of two discrete explanatory variables, when percentage of responses is less than 50, the ML and ELR poorly perform in parameter estimation. However, when the percentage of response is equal to 50 at sample size less than or equal to 28, the ELR is more effective than the ML, but at sample size greater than or equal to 48, both methods are similarly effective.

In interval estimation of two cases the results showed that, when the probability estimation coverage does not differ from the confidence level, the least confidence interval width estimation is ML. The ELR and MCMC methods commonly provided the infinite width of the confidence interval on average.

กิตติกรรมประกาศ

การศึกษาและเรียบเรียงวิทยานิพนธ์ฉบับนี้สามารถสำเร็จเป็นรูปเล่ม และเสร็จลุล่วงไปได้ด้วยดี ผู้วิจัยได้รับความอนุเคราะห์และความกรุณาจาก ผู้ช่วยศาตราจารย์ ดร.ประหยัด แสงงาม อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ให้คำแนะนำและการช่วยเหลือในการเรียบเรียงวิทยานิพนธ์ และแก้ไขข้อบกพร่องต่าง ๆ ด้วยความเมตตาเอาใจใส่อย่างสม่ำเสมอ

กราบขอบพระคุณ อาจารย์ ดร.วิภาวรรณ เล้าอรุณ ประธานกรรมการสอบวิทยานิพนธ์ และผู้ช่วยศาตราจารย์ ดร.มานะชัย รอดชื่น ผู้ทรงคุณวุฒิภายนอก ที่กรุณาสละเวลา และชี้แนะข้อบกพร่อง ทำให้วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ด้วยความเรียบร้อย

กราบขอบพระคุณ อาจารย์ในภาควิชาสถิติทุกท่าน ที่คอยประสาทวิชาความรู้ ส่งเสริมและเป็นกำลังใจให้อย่างสม่ำเสมอ และเจ้าหน้าที่ภาควิชาสถิติทุกท่านที่อำนวยความสะดวกในการดำเนินการต่าง ๆ ด้วยดีเสมอมา

กราบขอบพระคุณ ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย ผู้สนับสนุนทุนในการศึกษาและงานวิจัย

กราบขอบพระคุณ ครอบครัวเหมือนเงิน และเพื่อน ๆ ผู้คอยช่วยเหลือ ให้กำลังใจยามท้อถอย และการสนับสนุนอย่างดีโดยตลอด

ผู้วิจัยขออำนาจคุณพระศรีรัตนตรัย ดลบันดาลให้ทุกท่านจงมีแต่ความสุข ความเจริญ ประสบแต่ความสำเร็จ

จารุวรรณ เหมือนเงิน

สารบัญ

หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์การถดถอยเชิงเส้น เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง ในกรณีที่ตัวแปรตอบสนองเป็นตัวแปรเชิงปริมาณ แต่การวิเคราะห์การถดถอยลอจิสติกเป็นการศึกษาความเกี่ยวพันระหว่างตัวแปรอธิบายและตัวแปรตอบสนอง ในกรณีที่ตัวแปรตอบสนองเป็นตัวแปรจำแนกประเภทหรือตัวแปรเชิงกลุ่ม (กลุ่มที่สนใจ/กลุ่มที่ไม่สนใจ) โดยตัวแบบการถดถอยลอจิสติก (Logistic Regression Model) เป็นตัวแบบหนึ่งในกลุ่มของตัวแบบเชิงเส้นนัยทั่วไป (Generalized Linear Models : GLMs) ที่มีการนำไปใช้อย่างแพร่หลาย เช่น ทางชีว-การแพทย์ สาธารณสุข พันธุศาสตร์ สังคมศาสตร์ วิศวกรรมศาสตร์ การควบคุมคุณภาพ และด้านอื่นๆ [1]

โดยส่วนใหญ่แล้วการวิเคราะห์การถดถอยเชิงเส้น จะประมาณค่าพารามิเตอร์ในตัวแบบโดยใช้วิธีกำลังสองน้อยที่สุด ส่วนการวิเคราะห์การถดถอยลอจิสติก จะประมาณค่าพารามิเตอร์ในตัวแบบด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood: ML) ที่มีการคำนวณทวนซ้ำ (Iterative Algorithm) แต่เมื่อข้อมูลที่นำมาทำการวิเคราะห์การถดถอยลอจิสติกมีลักษณะเป็นแบบแบ่งแยกสมบูรณ์ (Completely Separate) หรือข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ (Quasi-Complete Separation) ซึ่งเกิดขึ้นเมื่อตัวแปรอธิบายสามารถจำแนกกลุ่มของตัวแปรตอบสนองได้อย่างถูกต้อง ยกตัวอย่างเช่น การศึกษาปัจจัยที่ทำให้เกิดโรคมะเร็งเต้านม (เกิด/ไม่เกิด) โดยตัวแปรอธิบายที่ทำการศึกษาคือ เพศ (หญิง/ชาย) ซึ่งโดยส่วนใหญ่พบว่าโรคมะเร็งเต้านมจะเกิดขึ้นในเพศหญิง กรณีดังกล่าวนี้เป็นตัวอย่างของข้อมูลที่เป็นแบบแบ่งแยกกึ่งสมบูรณ์ และพบว่าหากข้อมูลมีลักษณะเป็นแบบแบ่งแยกสมบูรณ์ หรือแบบแบ่งแยกกึ่งสมบูรณ์แล้ว การประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุดในการวิเคราะห์การถดถอยลอจิสติกจะไม่สามารถกระทำได้ [2] เนื่องจากในการคำนวณทวนซ้ำนั้นไม่สามารถหาค่าประมาณพารามิเตอร์ที่มีค่าลู่เข้าได้ นอกจากนี้ยังพบว่าการประมาณค่าพารามิเตอร์ในการวิเคราะห์การถดถอยลอจิสติกด้วยวิธีภาวะน่าจะเป็นสูงสุดจะมีประสิทธิภาพต่ำ ในกรณีที่ข้อมูลมีร้อยละของค่าตอบสนองต่ำ (กรณีที่ข้อมูลมีร้อยละของค่าสังเกตของตัวแปรตอบสนองที่อยู่ในกลุ่มที่สนใจต่ำ) [3]

ด้วยเหตุนี้ในงานวิจัยจึงได้ทำการศึกษาวิธีการประมาณค่าพารามิเตอร์วิธีอื่นที่สามารถใช้ประมาณค่าพารามิเตอร์ในกรณีที่ข้อมูลมีลักษณะเป็นแบบแบ่งแยกสมบูรณ์ หรือแบบแบ่งแยกกึ่งสมบูรณ์ได้ ซึ่งได้แก่ การถดถอยลอจิสติกแบบแม่นตรง (Exact Logistic Regression: ELR) โดยแนวคิดพื้นฐานของวิธี ELR คือ การหาค่าสถิติพอเพียงสำหรับพารามิเตอร์ที่สนใจที่เป็นไปได้ทั้งหมด และควบคุมค่าสถิติพอเพียงของพารามิเตอร์ตัวอื่นๆให้คงที่ [2] แต่ยังพบว่าการถดถอยลอจิสติกแบบแม่นตรงนั้นมีข้อจำกัดในด้านเวลา คือเป็นวิธีที่ใช้เวลาในการประมาณค่าพารามิเตอร์นาน [4] ดังนั้นหากข้อมูลที่นำมาวิเคราะห์มีขนาดใหญ่อาจกระทำได้ยาก จึงได้มีการศึกษาวิธีการประมาณค่าพารามิเตอร์อีกหนึ่งวิธีที่มีการนำเทคนิคลูกโซ่มาร์คอฟมอนติคาร์โล (Markov chain Monte Carlo) มาประยุกต์ใช้กับใช้วิธีการถดถอยลอจิสติกแบบแม่นตรง ได้แก่ วิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง หรือ Markov chain Monte Carlo (MCMC) Exact Inference เพื่อลดเวลาในการประมาณค่าพารามิเตอร์ แต่อย่างไรก็ตามวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง จะให้เพียงค่าประมาณพารามิเตอร์ที่ใกล้เคียงกับวิธีการถดถอยลอจิสติกแบบแม่นตรง เนื่องจากเทคนิคของวิธีลูกโซ่มาร์คอฟมอนติคาร์โล คือ การสุ่มตัวอย่างที่มีลักษณะลูกโซ่มาร์คอฟและมีการแจกแจงของจุดตัวอย่างลู่เข้าสู่การแจกแจงเป้าหมาย แต่มิได้สุ่มตัวอย่างจากการแจกแจงเป้าหมายที่แท้จริง และใช้ข้อมูลตัวอย่างนี้ไปใช้ในการประมาณค่าพารามิเตอร์ [5]

ดังนั้นงานวิจัยนี้จึงได้ทำการศึกษาและเปรียบเทียบประสิทธิภาพของวิธีการอนุมาน โดยวิธีการอนุมานที่ใช้ในงานวิจัยนี้คือ การประมาณค่าพารามิเตอร์แบบจุดและแบบช่วงในตัวแบบการถดถอยลอจิสติกทั้ง 3 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง ในกรณีที่ข้อมูลมีจำนวนค่าสังเกตทับซ้อน (Degree Overlap) ต่างกัน กล่าวคือกรณีที่จำนวนค่าสังเกตของตัวแปรอธิบายที่ทับซ้อนกันระหว่างกลุ่มของตัวแปรตอบสนองที่เราสนใจและกลุ่มที่เราไม่สนใจต่างกัน เพราะหากข้อมูลมีจำนวนค่าสังเกตที่ทับซ้อนน้อยแล้วข้อมูลจะมีลักษณะใกล้เคียงกับข้อมูลแบบแบ่งแยกสมบูรณ์ จึงทำการศึกษาว่าวิธี ML ยังมีประสิทธิภาพในการประมาณค่าพารามิเตอร์อยู่หรือไม่ และจำนวนค่าสังเกตทับซ้อนที่ต่างกันจะมีผลต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ของทั้ง 3 วิธี แตกต่างกันอย่างไร นอกจากนี้ยังพิจารณาในกรณีที่ข้อมูลมีร้อยละของค่าตอบสนอง (Percent of Responses) ต่างกัน เนื่องจากต้องการศึกษาว่าข้อมูลที่มีร้อยละของค่าตอบสนองต่างกันจะมีผลต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ของทั้ง 3 วิธี แตกต่างกันอย่างไร และจะทำการศึกษาเฉพาะการถดถอยลอจิสติกพื้นฐานอย่างง่าย

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อศึกษาวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก 3 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง

1.2.2 เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก 3 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง ในกรณีที่ข้อมูลมีจำนวนค่าสังเกตทับซ้อนต่างกัน และกรณีที่ข้อมูลมีร้อยละของค่าตอบสนองต่างกัน

1.3 สมมติฐานของการวิจัย

วิธีภาวะน่าจะเป็นสูงสุด มีประสิทธิภาพต่ำในการประยุกต์ใช้กับข้อมูลที่มีร้อยละของค่าสังเกตทับซ้อนน้อย และข้อมูลมีร้อยละของค่าตอบสนองต่ำ โดยที่วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ดีกว่าวิธี ML ในกรณีข้อมูลที่มีร้อยละของค่าสังเกตทับซ้อนน้อย และกรณีข้อมูลมีร้อยละของค่าตอบสนองต่ำ โดยวิธี ELR มีประสิทธิภาพสูงกว่าวิธี MCMC แต่วิธี ELR จะใช้เวลาในการวิเคราะห์มากกว่าวิธี MCMC

1.4 ขอบเขตของการวิจัย

ศึกษาวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก 3 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรงโดยวิธีของ Zamar และเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกทั้ง 3 วิธี โดยพิจารณาในกรณีที่ข้อมูลมีจำนวนค่าสังเกตทับซ้อนต่างกัน และกรณีที่ข้อมูลมีร้อยละของค่าตอบสนองต่างกัน โดยกำหนดของเขตของการวิจัยดังนี้

1.4.1 สร้างตัวแปรอธิบาย โดยแบ่งเป็น 2 กรณี ดังนี้

1.) กรณีตัวแปรอธิบาย 1 ตัวแปร: สร้างตัวแปรอธิบายแบบต่อเนื่อง (X) ที่มีการแจกแจงแบบ Uniform (6, 25)

2.) กรณีตัวแปรอธิบาย 2 ตัวแปร: สร้างตัวแปรอธิบายแบบไม่ต่อเนื่องหรือตัวแปรเชิงกลุ่ม X_1และ X_2 โดยตัวแปรแต่ละตัวมีค่าที่เป็นไปได้ 2 ค่า คือ 0 และ 1

1.4.2 กำหนดค่าพารามิเตอร์ โดยแบ่งเป็น 2 กรณี ดังนี้

1.) กรณีตัวแปรอธิบาย 1 ตัวแปร กำหนดค่าพารามิเตอร์ $\beta_0 = -18.49$ และ $\beta_1 = 0.88$

2.) กรณีตัวแปรอธิบาย 2 ตัวแปร กำหนดค่าพารามิเตอร์ขึ้นอยู่กับค่าประมาณร้อยละของค่าตอบสนอง ดังนี้

เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 10

- กำหนดค่า $\beta_0 = -2.53, \beta_1 = 0.45$ และ $\beta_2 = 0.19$

เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 20

- กำหนดค่า $\beta_0 = -1.99, \beta_1 = 0.01$ และ $\beta_2 = 1.04$

เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 30

- กำหนดค่า $\beta_0 = -1.99, \beta_1 = 0.07$ และ $\beta_2 = 1.86$

เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 40

- กำหนดค่า $\beta_0 = -1.49, \beta_1 = 0.09$ และ $\beta_2 = 1.89$

เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 50

- กำหนดค่า $\beta_0 = -1.49, \beta_1 = 1.09$ และ $\beta_2 = 1.89$

1.4.3 กรณีตัวแปรอธิบาย 1 ตัวแปร: กำหนดจำนวนค่าสังเกตทับซ้อน (Degree Overlap) ที่ทำการศึกษามีค่าเท่ากับ $1,2,3,4,6$ และ 8 ร้อยละของค่าตอบสนอง (Percent of Responses) ที่ทำการศึกษาคือ $20,\ 30,\ 40$ และ 50 และขนาดตัวอย่างที่ใช้ในการศึกษาคือ $n = 10, 20, 30, 50, 100$ และ 200

1.4.4 กรณีตัวแปรอธิบาย 2 ตัวแปร: กำหนดร้อยละของค่าตอบสนองมีค่าประมาณ $10, 20,30,40$ และ 50 และขนาดตัวอย่างที่ใช้ในการศึกษาคือ $n = 12, 20, 28, 48, 100$ และ 200

1.4.5 สร้างตัวแปร Y ที่มีค่าเท่ากับ 0 หรือ 1 จากความน่าจะเป็นที่ Y จะอยู่ในกลุ่มที่สนใจ หรือ $P(Y = 1|X = x) = \pi(x)$

1.4.6 จำนวนการทำซ้ำของการทดลองในการศึกษาคือ $m = 200$ ครั้ง

1.4.7 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ โดยจำแนกเป็น

1.) กรณีประมาณค่าพารามิเตอร์แบบจุด: พิจารณาจากค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย หรือ $\widehat{\mathrm{MSE}}(\widehat{\boldsymbol{\beta}})$

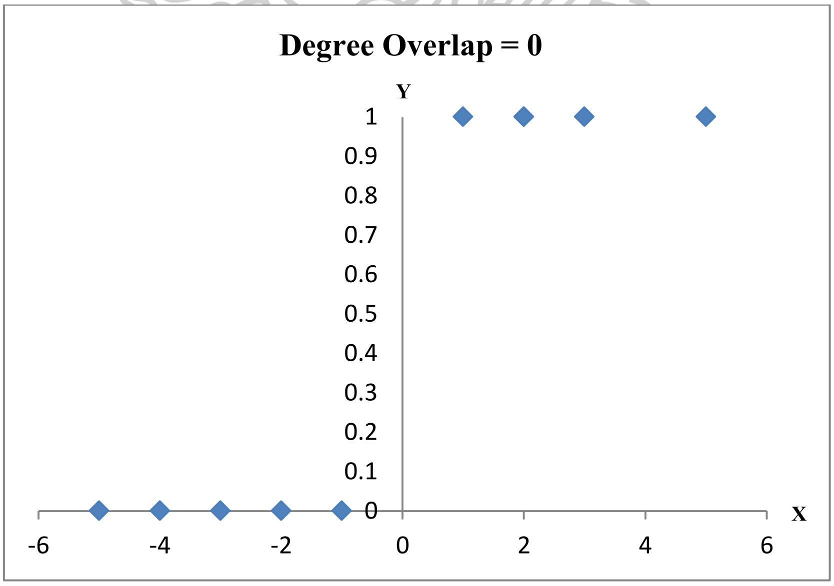
2.) กรณีประมาณค่าพารามิเตอร์แบบช่วง: พิจารณาจากค่าประมาณค่าความน่าจะเป็นครอบคลุม ($\widehat{\mathrm{CP}}$) และค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น ($\widehat{\mathrm{AW}}$)

1.5 นิยามศัพท์เฉพาะ

ตัวแบบการถดถอยลอจิสติก (Logistic Regression Model) เป็นตัวแบบที่ใช้อธิบายความเกี่ยวพันระหว่างตัวแปรอธิบาย (X) และตัวแปรตอบสนอง (Y) ในกรณีที่ตัวแปรตอบสนองเป็นตัวแปรจำแนกประเภท หรือตัวแปรเชิงกลุ่ม โดยในที่นี้ศึกษาเฉพาะกรณีที่ตัวแปรตอบสนองเป็นตัวแปรจำแนกประเภทแบบ 2 กลุ่ม (Binary Logistic Regression) [1]

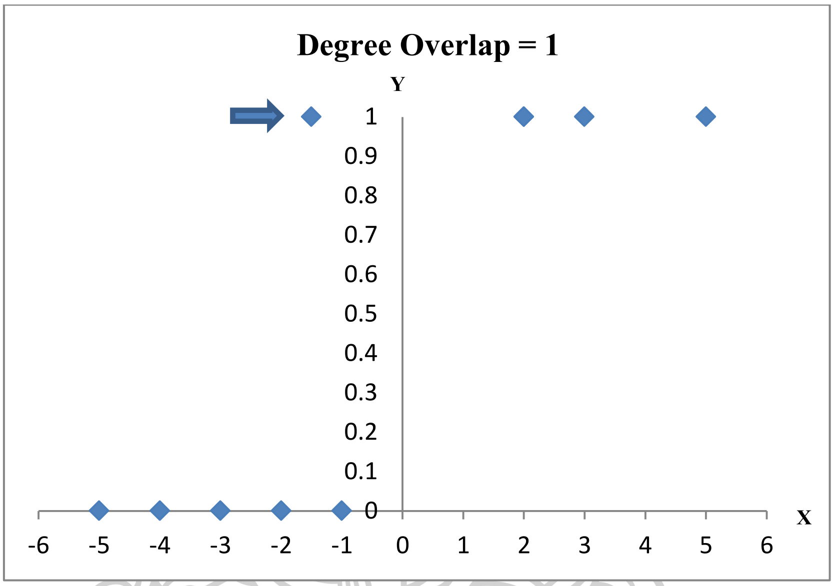
จำนวนค่าสังเกตทับซ้อน (Degree Overlap) คือ จำนวนค่าสังเกตของตัวแปรอธิบายที่ทับซ้อนกันระหว่างกลุ่มของตัวแปรตอบสนองที่สนใจ และไม่สนใจ ยกตัวอย่างเช่น

กรณีไม่มีจำนวนค่าสังเกตทับซ้อน (Degree Overlap = 0) สำหรับตัวแปรอธิบาย 1 ตัวแปร : หมายถึงค่าสังเกตของตัวแปรอธิบายที่ให้ค่าตัวแปรตอบสนองอยู่ในกลุ่มที่ไม่สนใจ $(Y = 0)$ และกลุ่มที่สนใจ $(Y = 1)$ ไม่ซ้อนทับกัน โดยข้อมูลที่ไม่มีจำนวนค่าสังเกตทับซ้อนนี้เรียกว่า ข้อมูลแบบแบ่งแยกสมบูรณ์ (Completely Separate) ดังแสดงในภาพที่ 1 จะพบว่าค่าของตัวแปรอธิบายที่น้อยกว่า 0 จะให้ค่า $\mathrm{Y} = 0$ และค่าของตัวแปรอธิบายที่มากกว่า 0 จะให้ค่า $Y = 1$ นั่นคือ ไม่มีการทับซ้อนกันระหว่างค่าของตัวแปรอธิบายที่ให้ $Y = 0$ และ $Y = 1$ [3]



ภาพที่ 1 ข้อมูลที่ไม่มีจำนวนค่าสังเกตทับซ้อน (Degree Overlap = 0)
สำหรับตัวแปรอธิบาย 1 ตัว

กรณีมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 (Degree Overlap = 1) สำหรับตัวแปรอธิบาย 1 ตัวแปร : หมายถึงค่าสังเกตของตัวแปรอธิบายที่ให้ค่าตัวแปรตอบสนองอยู่ในกลุ่มที่ไม่สนใจ $(Y = 0)$ และกลุ่มที่สนใจ $(Y = 1)$ ซ้อนทับกันอยู่ 1 ค่า ดังแสดงในภาพที่ 2 จะพบว่าค่าของตัวแปรอธิบายน้อยกว่า 0 จะให้ค่า $Y = 0$ และค่าของตัวแปรอธิบายที่มากกว่า 0 จะให้ค่า $Y = 1$ ยกเว้นที่ค่า $X = -1.5$ ซึ่งมีค่าน้อยกว่า 0 แต่ให้ค่า $Y = 1$ นั่นคือมีการทับซ้อนกันอยู่ 1 ค่า [3]



ภาพที่ 2 ข้อมูลที่มีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 (Degree Overlap = 1) สำหรับตัวแปรอธิบาย 1 ตัว

ร้อยละของค่าตอบสนอง (Percent of Responses) คือ ร้อยละของข้อมูลที่ตัวแปรตอบสนองมีค่าอยู่ในกลุ่มที่สนใจ $(Y = 1)$

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.) เพื่อให้ผู้ที่ทำการศึกษาความเกี่ยวพันระหว่างตัวแปรด้วยตัวแบบการถดถอยลอจิสติกได้ตระหนักถึงลักษณะของข้อมูลก่อนที่ทำการวิเคราะห์การถดถอย และเลือกใช้วิธีการประมาณค่าพารามิเตอร์ที่เหมาะสมกับลักษณะของข้อมูล

2.) เพื่อหาแนวทางพัฒนาในการศึกษาถึงวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการวิเคราะห์การถดถอยลอจิสติกโดยทั่วไปนิยมประมาณค่าพารามิเตอร์โดย วิธีภาวะน่าจะเป็นสูงสุด ที่มีการคำนวณทวนซ้ำ แต่การประมาณค่าพารามิเตอร์ด้วยวิธี ML จะไม่สามารถกระทำได้ ในกรณีที่ข้อมูลมีลักษณะเป็นแบบแบ่งแยกสมบูรณ์ หรือแบบแบ่งแยกกึ่งสมบูรณ์ ในงานวิจัยนี้จึงทำการศึกษาวิธีประมาณค่าพารามิเตอร์วิธีอื่นที่สามารถประมาณค่าพารามิเตอร์ในกรณีที่ข้อมูลมีลักษณะเป็นแบบแบ่งแยกสมบูรณ์ หรือแบบแบ่งแยกกึ่งสมบูรณ์ได้ ซึ่งได้แก่ วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง ในบทนี้จึงกล่าวถึงวิธีการประมาณค่าพารามิเตอร์ทั้ง 3 วิธี โดยเนื้อหาจะแบ่งเป็น 4 ส่วน ประกอบด้วย ส่วนที่ 1 ทฤษฎีที่เกี่ยวข้อง ส่วนที่ 2 วิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก ส่วนที่ 3 ข้อมูลแบบแบ่งแยกสมบูรณ์ ข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ และข้อมูลแบบทับซ้อน และส่วนที่ 4 งานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การถดถอยลอจิสติกพื้นฐาน (Binary Logistic Regression)

การถดถอยลอจิสติกพื้นฐาน (Binary Logistic Regression) เป็นการวิเคราะห์ความเกี่ยวพันระหว่างตัวแปรอธิบาย (X) และตัวแปรตอบสนอง (Y) ในกรณีที่ตัวแปรตอบสนองเป็นตัวแปรจำแนกประเภทแบบ 2 กลุ่ม คือ กลุ่มที่สนใจ $(Y = 1)$ และ กลุ่มที่ไม่สนใจ $(Y = 0)$ โดยที่ค่าสังเกตของตัวแปร Y ขนาด n หน่วย เป็นอิสระต่อกัน เนื่องจากตัวแปรตอบสนองมีค่าที่เป็นไปได้ 2 ค่า ดังนั้นตัวแปรตอบสนองจึงมีฟังก์ชันการแจกแจงแบบแบร์นูลลี (Bernoulli Distribution) โดยกำหนดให้ความน่าจะเป็นที่ตัวแปรตอบสนองจะอยู่ในกลุ่มที่สนใจ หรือ $P(Y_i = 1) = \pi_i$ และความน่าจะเป็นที่ตัวแปรตอบสนองจะอยู่ในกลุ่มที่ไม่สนใจ หรือ $P(Y_i = 0) = 1 - \pi_i$ ที่ $i = 1,2,\dots,n$ จะได้รูปแบบของฟังก์ชันมวลความน่าจะเป็น คือ [6]

$$f(y_i\,;\,\pi_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \qquad \text{เมื่อ } y_i = 0 \text{ หรือ } 1$$

$$= (1-\pi_i)\left(\frac{\pi_i}{1-\pi_i}\right)^{y_i} \tag{1}$$

2.1.2 ตัวแบบการถดถอยลอจิสติกพื้นฐานอย่างง่าย (Simple Logistic Regression Model)

ตัวแบบการถดถอยลอจิสติกพื้นฐานอย่างง่าย (Simple Logistic Regression Model) คือตัวแบบการถดถอยลอจิสติกที่มีตัวแปรอธิบาย (X) แบบต่อเนื่อง 1 ตัว ซึ่งแสดงได้ดังฟังก์ชันต่อไปนี้

$$\log\left[\frac{\pi(x)}{1-\pi(x)}\right] = \mathrm{logit}[\pi(x)] = \beta_0 + \beta_1 x$$

$$e^{\log\left[\frac{\pi(x)}{1-\pi(x)}\right]} = e^{(\beta_0+\beta_1 x)}$$

$$\frac{\pi(x)}{1-\pi(x)} = e^{(\beta_0+\beta_1 x)}$$

$$\pi(x) = e^{(\beta_0+\beta_1 x)}[1-\pi(x)]$$

$$\pi(x) = e^{(\beta_0+\beta_1 x)} - \left[\pi(x)e^{(\beta_0+\beta_1 x)}\right]$$

$$\pi(x) + \left[\pi(x)e^{(\beta_0+\beta_1 x)}\right] = e^{(\beta_0+\beta_1 x)}$$

$$\pi(x)\left[1 + e^{(\beta_0+\beta_1 x)}\right] = e^{(\beta_0+\beta_1 x)}$$

ดังนั้น

$$\pi(x) = P(Y=1|X=x) = \frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)} \qquad (2)$$

จากสมการที่ (2) พบว่าค่าความน่าจะเป็นที่ตัวแปรตอบสนองจะอยู่ในกลุ่มที่สนใจ หรือ $\pi(x)$ อยู่ในรูปแบบไม่เชิงเส้นในเทอมของพารามิเตอร์ เมื่อ β_0, β_1 และ x แทน จุดตัด (Intercept) ความชัน (Slope) และตัวแปรอธิบาย ตามลำดับ

และจะได้ว่า

$$1-\pi(x) = P(Y=0|X=x) = 1 - \frac{\exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)}$$

$$= \frac{1+\exp(\beta_0+\beta_1 x) - \exp(\beta_0+\beta_1 x)}{1+\exp(\beta_0+\beta_1 x)}$$

$$= \frac{1}{1+\exp(\beta_0+\beta_1 x)} \qquad (3)$$

จากสมการที่ (3) พบว่าค่าความน่าจะเป็นที่ตัวแปรตอบสนองจะอยู่ในกลุ่มที่ไม่สนใจ หรือ $1-\pi(x)$ อยู่ในรูปแบบไม่เชิงเส้นในเทอมของพารามิเตอร์เช่นกัน

2.1.3 ตัวแบบการถดถอยลอจิสติกพื้นฐานพหุคูณ (Multiple Logistic Regression Model)

ตัวแบบการถดถอยลอจิสติกพื้นฐานพหุคูณ (Multiple Logistic Regression Model) คือ ตัวแบบการถดถอยลอจิสติกที่มีตัวแปรอธิบาย (X) แบบต่อเนื่องอย่างน้อย 1 ตัว และตัวแปรอธิบายอื่นๆ 1 ตัวขึ้นไป ซึ่งแสดงได้ดังฟังก์ชันต่อไปนี้

$$\log\left[\frac{\pi(x_1,x_2,\dots,x_p)}{1-\pi(x_1,x_2,\dots,x_p)}\right] = \operatorname{logit}[\pi(\boldsymbol{x})] = \beta_0 + \sum_{j=1}^{p}\beta_j x_j$$

$$= \operatorname{logit}[\pi(\boldsymbol{x})] = \beta_0 + \sum_{j=1}^{p}\beta_j x_j$$

และในทำนองเดียวกันกับตัวแบบการถดถอยลอจิสติกพื้นฐานอย่างง่าย จะได้ว่า

$$\pi(x_1,x_2,\dots,x_p) = P(Y=1|\boldsymbol{X}=\boldsymbol{x}) = \frac{\exp(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p)}{1+\exp(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_px_p)}$$

$$= \frac{\exp(\beta_0+\sum_{j=1}^{p}\beta_jx_j)}{1+\exp(\beta_0+\sum_{j=1}^{p}\beta_jx_j)} \qquad (4)$$

จากสมการที่ (4) พบว่าค่าความน่าจะเป็นที่ตัวแปรตอบสนองจะอยู่ในกลุ่มที่สนใจ หรือ $\pi(x_1,x_2,\dots,x_p)$ อยู่ในรูปแบบไม่เชิงเส้นในเทอมของพารามิเตอร์ เมื่อ β_0,β_j และ x_j แทน จุดตัด (Intercept) ความชัน (Slope) ที่ j และตัวแปรอธิบายที่ j ตามลำดับ

และ

$$1-\pi(x_1,x_2,\dots,x_p) = P(Y=0|\boldsymbol{X}=\boldsymbol{x}) = \frac{1}{1+\exp(\beta_0+\sum_{j=1}^{p}\beta_jx_j)} \qquad (5)$$

จากสมการที่ (5) พบว่าค่าความน่าจะเป็นที่ตัวแปรตอบสนองจะอยู่ในกลุ่มที่ไม่สนใจ หรือ $1-\pi(x_1,x_2,\dots,x_p)$ อยู่ในรูปแบบไม่เชิงเส้นในเทอมของพารามิเตอร์เช่นกัน

2.2 วิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก

2.2.1 วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood: ML)

การประมาณค่าพารามิเตอร์

วิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกที่นิยมใช้โดยทั่วไป คือวิธีภาวะน่าจะเป็นสูงสุด ที่มีการคำนวณทวนซ้ำ (Iterative Algorithm) และการคำนวณทวนซ้ำที่ใช้ในงานวิจัยนี้ คือ การทวนซ้ำแบบ Fisher's Scoring Method โดยจะหาค่าประมาณของ $\boldsymbol{\beta}_{[(\boldsymbol{p+1})\times\mathbf{1}]} = \left[\beta_0, \dots, \beta_p\right]^T$ ที่ทำให้ ฟังก์ชันภาวะน่าจะเป็น (Likelihood Function) หรือ $\boldsymbol{l(\beta)}$ มีค่าสูงสุด ซึ่งมีขั้นตอนดังนี้

1. หาฟังก์ชันภาวะน่าจะเป็น [Likelihood Function: $\boldsymbol{l(\beta)}$]

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i}$$

$$= \prod_{i=1}^{n} (1-\pi_i) exp\left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right)\right]$$

จะได้ลอการิทึมของฟังก์ชันภาวะน่าจะเป็น [Log likelihood Function: $\boldsymbol{L(\beta)}$]

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \log\left[\textstyle\prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i}\right] \\
&= \textstyle\sum_{i=1}^{n} [y_i \log\pi_i + (1-y_i)\log(1-\pi_i)] \\
&= \textstyle\sum_{i=1}^{n} [y_i[\boldsymbol{x_i\beta} + \log(1-\pi_i)] + (1-y_i)\log(1-\pi_i)] \\
&= \textstyle\sum_{i=1}^{n} [y_i\boldsymbol{x_i\beta} + y_i\log(1-\pi_i) + \log(1-\pi_i) - y_i\log(1-\pi_i)] \\
&= \textstyle\sum_{i=1}^{n} [y_i\boldsymbol{x_i\beta} + \log(1-\pi_i)] \\
&= \textstyle\sum_{i=1}^{n} \left[y_i\boldsymbol{x_i\beta} + \log\left(\frac{1}{1+\exp(\boldsymbol{x_i\beta})}\right)\right] \\
&= \textstyle\sum_{i=1}^{n} [y_i\boldsymbol{x_i\beta} - \log(1+\exp(\boldsymbol{x_i\beta}))] \qquad (6)
\end{aligned}$$

เมื่อ $\boldsymbol{x_i}_{[\mathbf{1}\times(\boldsymbol{p+1})]} = \left[x_{i0}, x_{i1}, x_{i2}, \dots, x_{ip}\right]$

2. หาตัวประมาณภาวะน่าจะเป็นสูงสุดโดยการหาอนุพันธ์อันดับที่ 1 ของลอการิทึมของฟังก์ชันภาวะน่าจะเป็น หรือ $L(\boldsymbol{\beta})$ เทียบกับ $\boldsymbol{\beta}$ และกำหนดให้เท่ากับศูนย์ ดังนี้ [1]

$$\boldsymbol{U}(\boldsymbol{\beta})_{[\mathbf{1}\times(\boldsymbol{p}+\mathbf{1})]} = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right) = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_0}, \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_1}, \ \dots, \frac{\partial L(\boldsymbol{\beta})}{\partial \beta_P}\right)$$

$$= \ \sum_{i=1}^{n} y_i \boldsymbol{x_i} - \sum_{i=1}^{n} \boldsymbol{x_i} \left(\frac{e^{\boldsymbol{x_i\beta}}}{1+e^{\boldsymbol{x_i\beta}}}\right) = \mathbf{0} \tag{7}$$

3. ตรวจสอบว่าฟังก์ชัน $L(\boldsymbol{\beta})$ มีค่าสูงสุดหรือไม่ โดยการหาอนุพันธ์อันดับที่ 2 ของลอการิทึมของฟังก์ชันภาวะน่าจะเป็น หรือ $L(\boldsymbol{\beta})$ เทียบกับ $\boldsymbol{\beta}$ ดังนี้ [1]

$$\boldsymbol{I}(\boldsymbol{\beta})_{[(\boldsymbol{p}+\mathbf{1})\times(\boldsymbol{p}+\mathbf{1})]} = \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$$

$$= -\sum_{i=1}^{n} \left(\frac{\boldsymbol{x_i}^T \boldsymbol{x_i} e^{\boldsymbol{x_i\beta}}}{(1+e^{\boldsymbol{x_i\beta}})^2}\right) \tag{8}$$

4. กำหนดค่าประมาณเริ่มต้น $(\boldsymbol{\beta}^{(0)})$

5. ประมาณค่า $\boldsymbol{\beta}$ จากสมการ

$$\boldsymbol{\beta}^{new}{}_{[\mathbf{1}\times(\boldsymbol{p}+\mathbf{1})]} = \boldsymbol{\beta}^{old} - \boldsymbol{U}(\boldsymbol{\beta}^{old})\boldsymbol{E}[\boldsymbol{I}^{-1}(\boldsymbol{\beta}^{old})] \tag{9}$$

เมื่อ $\boldsymbol{E}[\boldsymbol{I}^{-1}(\boldsymbol{\beta})]$ คือ ค่าคาดหวังของเมทริกซ์ผกผัน (Inverse Matrix) ของ $\boldsymbol{I}(\boldsymbol{\beta})$

โดยการวนซ้ำในรอบที่ 1 จะใช้ค่า $\boldsymbol{\beta}^{(0)}$ แทนค่า $\boldsymbol{\beta}^{old}$ ในสมการ (9) เพื่อหาค่าประมาณ $\boldsymbol{\beta}^{new}$ ในรอบที่ 1 และการวนซ้ำในรอบที่ 2 จะใช้ค่า $\boldsymbol{\beta}^{new}$ ในรอบที่ 1 แทนค่า $\boldsymbol{\beta}^{old}$ ในสมการ (9) เพื่อหาค่าประมาณ $\boldsymbol{\beta}^{new}$ ใหม่ จากนั้นทำการวนซ้ำแบบนี้ไปเรื่อยๆ จนกระทั่งได้ค่าประมาณสัมประสิทธิ์การถดถอย $(\boldsymbol{\beta})$ ที่มีค่าลู่เข้า (ค่าผลต่างระหว่าง $\boldsymbol{\beta}^{new}$และ $\boldsymbol{\beta}^{old}$ มีค่าน้อยมาก ซึ่งบางวิธีใช้ผลต่างที่ใกล้เคียงกับเวกเตอร์ศูนย์) [1]

สำหรับข้อมูลที่อยู่ในรูปของตารางแจกแจงความถี่ที่มีตัวแปรตอบสนองจำแนกประเภทแบบ 2 กลุ่มและตัวแปรอธิบาย 1 ตัว (ที่มีค่าที่เป็นไปได้ 2 ค่า) สามารถหาค่าประมาณสัมประสิทธิ์การถดถอย $(\hat{\beta}_1)$ โดยวิธี ML ได้ดังนี้ [2]

ตารางที่ 1 ตารางแจกแจงความถี่สำหรับตัวแปรตอบสนองจำแนกประเภทแบบ 2 กลุ่มและตัวแปรอธิบาย 1 ตัว

ตัวแปร X	**ตัวแปร Y**	
	0	**1**
0	f_{11}	f_{12}
1	f_{21}	f_{22}

$$\hat{\beta}_1 = \log \frac{f_{11}f_{22}}{f_{12}f_{21}} \tag{10}$$

เมื่อ f_{ij} แทนความถี่ในแถวที่ i^{th} คอลัมภ์ที่ j^{th} ในตารางแจกแจงความถี่แบบ 2×2

การทดสอบสมมติฐาน

สมมติฐานการทดสอบ:

$$H_0\colon \beta_j = 0$$

$$H_1\colon \beta_j \neq 0$$

ตัวสถิติทดสอบ: การทดสอบอัตราส่วนภาวะน่าจะเป็นสูงสุด (Likelihood Ratio Test: LR)

$$\mathrm{LR} = -2\log\frac{\hat{L}_c}{\hat{L}_s} \tag{11}$$

เมื่อ $\hat{L}_c$ คือค่าฟังก์ชันภาวะน่าจะเป็นที่มีค่าสูงสุดภายใต้ H_0
$\hat{L}_s$ คือค่าฟังก์ชันภาวะน่าจะเป็นที่มีค่าสูงสุดภายใต้ H_1

โดยที่ตัวสถิติทดสอบ LR จะมีการแจกแจงเชิงเส้นกำกับไคสแควร์ ที่มีองศาแห่งความอิสระ (Degree of Freedom) เท่ากับผลต่างของจำนวนตัวแปรอธิบายภายใต้ H_0 และ H_1

ตัวสถิติทดสอบ: ตัวสถิติทดสอบของ Wald (Wald Test)

$$\mathrm{Wald} = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \tag{12}$$

โดยที่ตัวสถิติทดสอบของ Wald จะมีการแจกแจงเชิงเส้นกำกับไคสแควร์ ที่มีองศาแห่งความอิสระ (Degree of Freedom) เท่ากับ 1

การหาช่วงความเชื่อมั่น

Wald ได้แสดงผลลัพธ์เชิงเส้นกำกับหรือเมื่อใกล้อนันต์ (Asymptotic Results) ของตัวประมาณภาวะน่าจะเป็นสูงสุดในตัวแบบลอจิสติกว่าลู่เข้าสู่การแจกแจงปกติเมื่อตัวอย่างขนาดใหญ่ (Large-Sample Normal Distributions) และสามารถประมาณค่าช่วงความเชื่อมั่นสำหรับ β_j ได้ดังนี้ [1]

$$\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}}\,\widehat{SE}(\hat{\beta}_j) \tag{13}$$

เมื่อ $\widehat{SE}(\hat{\beta}_j)$ แทนค่าประมาณความคลาดเคลื่อนมาตรฐาน (Standard Error) ของ $\hat{\beta}_j$

2.2.2 การถดถอยลอจิสติกแบบแม่นตรง (Exact Logistic Regression: ELR)

เนื่องจากการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด ไม่สามารถกระทำได้ในกรณีที่ข้อมูลมีลักษณะเป็นแบบแบ่งแยกสมบูรณ์ หรือแบบแบ่งแยกกึ่งสมบูรณ์ จึงได้มีการศึกษาวิธีการประมาณค่าพารามิเตอร์วิธีอื่นที่สามารถประมาณค่าพารามิเตอร์ในกรณีที่วิธี ML ไม่สามารถกระทำได้ ซึ่งวิธีดังกล่าว คือ วิธีการถดถอยลอจิสติกแบบแม่นตรง โดยในปี 1995 Mehta และ Patel [7] ได้เสนอทฤษฎีที่เกี่ยวข้องกับการถดถอยลอจิสติกแบบแม่นตรง ไว้ดังนี้

การประมาณค่าพารามิเตอร์

จากฟังก์ชันภาวะน่าจะเป็น หรือ $l(\boldsymbol{\beta})$ จะได้ฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไขของพารามิเตอร์ตัวที่สนใจ (The Conditional Likelihood) ดังนี้

$$\begin{aligned} l(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \pi_i^{y_i}\,(1-\pi_i)^{1-y_i} \\ &= \prod_{i=1}^{n}(1-\pi_i)\left(\frac{\pi_i}{1-\pi_i}\right)^{y_i} \\ &= \prod_{i=1}^{n}\left(1+e^{\boldsymbol{x_i\beta}}\right)^{-1} e^{\boldsymbol{x_i\beta} y_i} \\ &= \frac{e^{\sum_{i=1}^{n}\sum_{j=0}^{p}\beta_j x_{ij} y_i}}{\prod_{i=1}^{n}(1+e^{\boldsymbol{x_i\beta}})} \end{aligned} \tag{14}$$

ให้ $t_j = \sum_{i=1}^{n} x_{ij} y_i$ ซึ่ง t_j เป็นค่าสถิติพอเพียงสำหรับ β_j โดยที่ $j = 0,1, \ldots, p$ ดังนั้นจาก (14) สามารถแสดงให้อยู่ในรูปของค่าสถิติพอเพียงได้ดังนี้

$$l(\boldsymbol{\beta}) = \frac{e^{\sum_{j=0}^{p} \beta_j t_j}}{\prod_{i=1}^{n}(1 + e^{\boldsymbol{x_i \beta}})} \tag{15}$$

กำหนดให้พารามิเตอร์ตัวที่สนใจคือ β_p จะเรียกพารามิเตอร์ตัวอื่นที่เหลือ $(\beta_0, \beta_1, \ldots, \beta_{p-1})$ ว่าพารามิเตอร์รบกวน (Nuisance Parameters) และกำหนดให้ $t_0, t_1, \ldots, t_p$ คือ ค่าของตัวสถิติพอเพียงจากตัวแปรสุ่ม $T_0, T_1, \ldots, T_p$ จะได้ ฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไขสำหรับ β_p เมื่อกำหนด $T_0 = t_0, \ldots, T_{p-1} = t_{p-1}$ ดังนี้

$$\begin{aligned} l_c(\beta_p) &= P(T_p = t_p | T_0 = t_0, T_1 = t_1, \ldots, T_{p-1} = t_{p-1}) \\ &= \frac{P(T_0 = t_0, T_1 = t_1, \ldots, T_p = t_p)}{P(T_0 = t_0, T_1 = t_1, \ldots, T_{p-1} = t_{p-1})} \end{aligned} \tag{16}$$

โดยที่

$$\mathrm{P}(T_0 = t_0, T_1 = t_1, \ldots, T_p = t_p) = \frac{c(t_0, \ldots, t_p) e^{\sum_{j=0}^{p} \beta_j t_j}}{\prod_{i=1}^{n}(1 + e^{\boldsymbol{x_i \beta}})} \tag{17}$$

เมื่อ $c(t_0, \ldots, t_p)$ คือ จำนวนการเรียงสับเปลี่ยนค่าของตัวแปรตอบสนองเมื่อกำหนดค่า $t_0, t_1, \ldots, t_p$

และ

$$\mathrm{P}(T_0 = t_0, T_1 = t_1, \ldots, T_{p-1} = t_{p-1}) = \frac{\sum_u c(t_0, \ldots, t_{p-1}, u) e^{\beta_p u + \sum_{j=0}^{p-1} \beta_j t_j}}{\prod_{i=1}^{n}(1 + e^{\boldsymbol{x_i \beta}})} \tag{18}$$

เมื่อ $c(t_0, \ldots, t_{p-1}, u)$ คือจำนวนการเรียงสับเปลี่ยนค่าของตัวแปรตอบสนอง เมื่อกำหนดค่า $t_0, t_1, \ldots, t_{p-1}$ และ u คือค่าสถิติพอเพียงของพารามิเตอร์ที่สนใจที่เป็นไปได้ทั้งหมด โดยที่ $c(t_0, \ldots, t_{p-1}, u) \geq 1$

ดังนั้นจะได้ว่า ฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไขสำหรับ β_p เมื่อกำหนด $T_0 = t_0, \ldots, T_{p-1} = t_{p-1}$ มีค่าดังนี้

$$l_c(\beta_p) = \frac{c(t_0,\ldots,t_p)e^{\beta_p t_p}}{\sum_u c(t_0,\ldots,t_{p-1},u)e^{\beta_p u}} \quad (19)$$

จากสมการ (19) พบว่า ฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไขข้างต้น ไม่ได้ขึ้นอยู่กับพารามิเตอร์รบกวน $(\beta_0, \beta_1, \ldots, \beta_{p-1})$ แต่ขึ้นอยู่กับพารามิเตอร์ตัวที่สนใจ (β_p) เพียงตัวเดียว ดังนั้นฟังก์ชันนี้จึงเป็นฟังก์ชันที่ใช้อนุมานค่าของพารามิเตอร์ตัวที่สนใจ (β_p) ได้อย่างแท้จริง โดยสามารถหาค่าประมาณของพารามิเตอร์ β_p ได้ด้วยวิธีภาวะน่าจะเป็นสูงสุด (ML)

การทดสอบสมมติฐาน

ให้ $f(t_p|\beta_p)$ คือ ความน่าจะเป็นแบบมีเงื่อนไข หรือก็คือฟังก์ชันภาวะน่าจะเป็นแบบมีเงื่อนไขสำหรับ β_p จะได้ว่า

$$f(t_p|\beta_p) = \frac{c(t_0,\ldots,t_p)e^{\beta_p t_p}}{\sum_u c(t_0,\ldots,t_{p-1},u)e^{\beta_p u}} \quad (20)$$

สมมติฐานการทดสอบ:

$$H_0\colon \beta_p = 0$$

$$H_1\colon \beta_p \neq 0$$

การคำนวณค่า exact p-value สามารถคำนวณได้จากการหาผลรวมของสมการที่ (20) ที่อยู่ภายใต้บริเวณวิกฤติ (E) ดังนี้

$$p = \sum_{v\in E} f(v|\beta_p = 0) \quad (21)$$

การกำหนดบริเวณวิกฤติ (E) ที่นิยมใช้มีอยู่ 2 วิธี คือ The Conditional Probabilities Test และ The Conditional Scores Test โดยที่

บริเวณวิกฤติของ The Conditional Probabilities Test (E_{cp}) คือ

$$E_{cp} = \{v: f(v|\beta_p = 0) \leq f(t_p|\beta_p = 0)\} \quad (22)$$

บริเวณวิกฤติของ The Conditional Scores Test (E_{cs}) คือ

$$E_{cs} = \{v: (v - \mu_p)^2 \sigma_p^{-2} \geq (t_p - \mu_p)^2 \sigma_p^{-2}\} \quad (23)$$

เมื่อ μ_p และ σ_p^2 คือ ค่าเฉลี่ยและความแปรปรวนของ T_p ซึ่งขึ้นอยู่กับการแจกแจงแบบมีเงื่อนไขที่กำหนดโดยสมการที่ (20) ณ ที่ $\beta_p = 0$

การหาช่วงความเชื่อมั่น

กำหนดให้ (β_-, β_+) คือ ช่วงความเชื่อมั่น $100(1-\alpha)\%$ สำหรับ β_p โดยที่

$$F_1(t_p|\beta_p) = \sum_{v \geq t_p} f(v|\beta) \tag{24}$$

และ

$$F_2(t_p|\beta_p) = \sum_{v \leq t_p} f(v|\beta) \tag{25}$$

กำหนดให้ t_{min} และ t_{max} คือ ค่า u ที่มีค่าน้อยที่สุดและมากที่สุดสำหรับการแจกแจงแบบมีเงื่อนไขที่กำหนดโดยสมการที่ (20) ตามลำดับ โดยขอบเขตล่างของช่วงความเชื่อมั่น (The Lower Confidence Bound) หรือ β_- สามารถหาได้จาก

$$F_1(t_p|\beta_-) = \alpha/2 \text{ ถ้า } t_{min} < t_p \leq t_{max}, \tag{26}$$

$$\beta_- = -\infty \quad \text{ถ้า } t_p = t_{min} \tag{27}$$

และขอบเขตบนของช่วงความเชื่อมั่น (The Upper Confidence Bound) หรือ β_+ สามารถหาได้จาก

$$F_2(t_p|\beta_+) = \alpha/2 \text{ ถ้า } t_{min} \leq t_p < t_{max}, \tag{28}$$

$$\beta_+ = \infty \quad \text{ถ้า } t_p = t_{max} \tag{29}$$

2.2.3 ลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง (Markov chain Monte Carlo (MCMC) Exact Inference)

เนื่องจากวิธีการถดถอยลอจิสติกแบบแม่นตรง (ELR) มีข้อเสียเกี่ยวกับเวลาในการประมาณค่าพารามิเตอร์ ดังนั้นจึงได้มีการศึกษาเทคนิคที่ใช้ในการสุ่มข้อมูล ซึ่งได้แก่ เทคนิคลูกโซ่มาร์คอฟมอนติคาร์โล (Markov chain Monte Carlo) โดยมีแนวคิดคือ ทำการจำลองหรือสุ่มตัวแปรสุ่มร่วมที่ทราบฟังก์ชันการแจกแจง (f) และจะไม่ทำการสุ่มจาก f อย่างถูกต้อง แต่จะประมาณ f โดยสุ่มตัวอย่างที่มีลักษณะลูกโซ่มาร์คอฟซึ่งการแจกแจงของจุดตัวอย่างลู่เข้าสู่ f เพื่อแก้ไขปัญหาที่เกิดขึ้นเกี่ยวกับเวลาในการประมาณค่าพารามิเตอร์ของวิธีการถดถอยลอจิสติกแบบแม่นตรง [5]

เมโทรโปลิส-เฮสติง

เมโทรโปลิส-เฮสติง (Metropolis-Hastings Algorithms) เป็นวิธีการสร้างลูกโซ่มาร์คอฟวิธีหนึ่งที่มีขั้นตอนดังนี้ สร้างตัวอย่างแบบลูกโซ่มาร์คอฟที่มีการแจกแจงของจุดตัวอย่างลู่เข้าสู่การแจกแจงเป้าหมายในสองขั้น โดยขั้นแรกจะสร้างตัวเลือกจากความน่าจะเป็นในการเปลี่ยนสถานะ (Q) และขั้นที่สองคือการเลือกที่จะยอมรับหรือปฏิเสธตัวเลือกด้วยความน่าจะเป็น α หรือ $1-\alpha$ ตามลำดับ โดยสรุปขั้นตอนได้ดังนี้ [5]

กำหนด: ฟังก์ชันการแจกแจงเป้าหมาย (f)

1. กำหนดค่าเริ่มต้น y
2. สร้างตัวเลือกสำหรับตัวอย่างถัดไปเป็น y' จากความน่าจะเป็นเปลี่ยนสถานะของตัวเลือก (Q) ซึ่งมีฟังก์ชันการแจกแจงความน่าจะเป็นคือ q
3. เลือกตัวเลือกสำหรับตัวอย่างถัดไปเป็น y'ด้วยความน่าจะเป็น $\alpha(y,y')$ ซึ่ง

$$\alpha(y,y') = \begin{cases} min\left(1, \frac{f(y')q(y',y)}{f(y)q(y,y')}\right), & \text{ถ้า } f(y)q(y,y') > 0 \\ 1\,, & \text{ถ้า } f(y)q(y,y') = 0 \end{cases} \tag{30}$$

ผลลัพธ์: y ที่มีฟังก์ชันการแจกแจง f

เมโทรโปลิส-เฮสติง ที่เสนอโดย Forster และคณะ (2003)

ในปี 2003 Forster และคณะ [8] ได้พัฒนาวิธีเมโทรโปลิส-เฮสติง (Metropolis-Hastings Algorithms) มาประยุกต์ใช้กับวิธีการถดถอยลอจิสติกแบบแม่นตรง (ELR) เพื่อแก้ไขปัญหาที่เกิดขึ้นเกี่ยวกับเวลาในการวิเคราะห์ข้อมูลของวิธีการถดถอยลอจิสติกแบบแม่นตรง วิธีที่ Forster และคณะเสนอเรียกว่าวิธี เมโทรโปลิส-เฮสติงสำหรับการถดถอยลอจิสติกแบบแม่นตรง (Metropolis-Hastings Algorithms for Exact Conditional Inference) โดยมีวิธีการดังนี้

ให้

$$Y_i \sim Binomial(m_i, \pi_i)$$

และ

$$logit(\pi_i) = \boldsymbol{x_i\beta} + \boldsymbol{z_i\gamma} \quad \text{เมื่อ} \quad i = 1, \dots, n \qquad (31)$$

โดยที่ $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T$ เป็นเวกเตอร์ของพารามิเตอร์รบกวนที่สอดคล้องกับเวกเตอร์ของตัวแปรอธิบาย p ตัว คือ $\boldsymbol{x_i} = (x_{i0}, \dots, x_{ip})$ และ $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ เป็นเวกเตอร์ของพารามิเตอร์ที่สนใจที่สอดคล้องกับเวกเตอร์ของตัวแปรอธิบาย q ตัว คือ $\boldsymbol{z_i} = (z_{i1}, \dots, z_{iq})$

กำหนดให้ $\boldsymbol{y} = (y_1, \dots, y_n)^T$ $\boldsymbol{X}$ คือเมทริกซ์ขนาด $n \times (p+1)$ ที่มีสมาชิกในแถวที่ i เป็น $\boldsymbol{x_i}$ และ $\boldsymbol{Z}$ คือเมทริกซ์ขนาด $n \times q$ ที่มีสมาชิกในแถวที่ i เป็น $\boldsymbol{z_i}$ จะได้ $\boldsymbol{T} = \boldsymbol{Z^T y}$ เป็นตัวสถิติพอเพียงสำหรับ $\boldsymbol{\gamma}$ และ $\boldsymbol{S} = \boldsymbol{X^T y}$ เป็นตัวสถิติพอเพียงสำหรับ $\boldsymbol{\beta}$ และกำหนดให้ $\boldsymbol{S} = \boldsymbol{X^T y} = \boldsymbol{s}$ ดังนั้นจะได้ความน่าจะเป็นแบบมีเงื่อนไขของ $\boldsymbol{y} = (y_1, \dots, y_n)^T$ เมื่อ $\boldsymbol{S} = \boldsymbol{X^T y}$ เป็นตัวสถิติพอเพียงสำหรับ $\boldsymbol{\beta}$ ดังนี้

$$f(\boldsymbol{y}|\boldsymbol{S} = \boldsymbol{s}) \propto \exp(\gamma^T \boldsymbol{Z^T y}) \prod_{i=1}^{n} \binom{m_i}{y_i} \qquad (32)$$

วิธีเมโทรโปลิส-เฮสติง ที่เสนอโดย Forster และคณะ จะมีการสุ่ม $\boldsymbol{y'}$ จาก $\boldsymbol{y'} = \boldsymbol{y} + d\boldsymbol{v}$ โดยที่ d คือ จำนวนเต็ม และ $\boldsymbol{v}$ คือเวกเตอร์คอลัมภ์ของจำนวนเฉพาะสัมพัทธ์ หรือเวกเตอร์คอลัมภ์ของจำนวนเต็มที่ไม่มีตัวประกอบร่วมกัน ที่มีขนาด $n \times 1$ และเนื่องจากการประมาณค่าพารามิเตอร์ด้วยวิธีการถดถอยลอจิสติกแบบแม่นตรงต้องมีการควบคุมค่าสถิติพอเพียงของพารามิเตอร์รบกวน ดังนั้นจะได้ว่า

$$\boldsymbol{X^T v} = \boldsymbol{0} \tag{33}$$

จากนั้นแจกแจงทุก ๆ ค่า $\boldsymbol{v}$ ที่เป็นไปได้ภายใต้เงื่อนไข

$$\sum_{i=1}^{n}|v_i| \le r \tag{34}$$

สำหรับ r ที่กำหนด และเนื่องจากมีเวกเตอร์ $\mathbf{1}$ หรือเวกเตอร์ของค่าคงที่เป็นปริภูมิแถวตั้ง (Column Space) อยู่ในเมทริกซ์ $\boldsymbol{X}$ ดังนั้นจึงเกิดเงื่อนไขที่ว่า

$$\sum_{i=1}^{n} v_i = 0 \tag{35}$$

และจะได้ว่า $\sum_{i=1}^{n}|v_i|$ ต้องเป็นจำนวนคู่ และค่า r ที่ Forster และคณะ เสนอได้แก่ $r = 4, 6,$ หรือ 8

วิธีการของเมโทรโปลิส-เฮสติง คือ เมื่อเลือกค่า $\boldsymbol{v}$ ด้วยความน่าจะเป็นแบบ Uniform ที่เป็นไปได้ตามเงื่อนไขข้างต้น จากนั้นเลือกค่า d จากฟังก์ชันการแจกแจงความน่าจะเป็น

$$q(d|\boldsymbol{v}) \propto \exp\{\gamma^T \boldsymbol{Z^T}(\boldsymbol{y} + d\boldsymbol{v})\} \prod_{i=1}^{n} \binom{m_i}{y_i + dv_i} = \eta(d|\boldsymbol{v}, \boldsymbol{y}) \tag{36}$$

เมื่อ $0 \le y_i + dv_i \le m_i$ สำหรับทุกๆค่าของ i

เมื่อพิจารณาเปรียบเทียบ สมการที่ (32) และ (36) จะได้ว่า $q(d|\boldsymbol{v}) \propto f(\boldsymbol{y}')$ ดังนั้น $q(\boldsymbol{y}, \boldsymbol{y}') \propto f(\boldsymbol{y}')$ สำหรับทุก ๆ ค่า $\boldsymbol{y}, \boldsymbol{y}'$ และจะได้ว่า $q(\boldsymbol{y}', \boldsymbol{y}) \propto f(\boldsymbol{y})$ ดังนั้น $\alpha(\boldsymbol{y}, \boldsymbol{y}') = 1$ นั่นคือการปฏิเสธ $\boldsymbol{y}'$ จะไม่สามารถเกิดขึ้นได้

ขั้นตอนของวิธีเมโทรโปลิส-เฮสติง ที่เสนอโดย Forster และคณะ สามารถสรุปได้ดังนี้

1. สร้างเซ็ต $\boldsymbol{V}$ โดยที่ $\boldsymbol{V} = \{\boldsymbol{v}\colon \sum_{i=1}^{n}|v_i| \le r$ และ v_i คือจำนวนเฉพาะสัมพัทธ์ สำหรับ $i = 1, \dots, n,\ \boldsymbol{X^T v} = \boldsymbol{0}\}$
2. เลือก $\boldsymbol{v}$ จากเซ็ตของ $\boldsymbol{V}$ ด้วยความน่าจะเป็นแบบ Uniform
3. หาทุก ๆ ค่า d โดยที่ $0 \le y_i + dv_i \le m_i$ และกำหนดให้ k คือ จำนวนของค่า d ที่เป็นไปได้ทั้งหมด
4. คำนวณค่า $\sum_{i=1}^{k}\eta(d_i|\boldsymbol{v}, \boldsymbol{y})$ และ กำหนด $P(d_i) = \frac{\eta(d_i|\boldsymbol{v},\boldsymbol{y})}{\sum_{j=1}^{k}\eta(d_j|\boldsymbol{v},\boldsymbol{y})} \quad i, j = 1,2, \dots, k$
5. เลือก d ด้วยความน่าจะเป็นที่กำหนดในขั้นตอนที่ 4
6. กำหนดให้ $\boldsymbol{y}' = \boldsymbol{y} + d\boldsymbol{v}$
7. ให้ $\boldsymbol{y} = \boldsymbol{y}'$ และกลับไปที่ขั้นตอนที่ 2

เมโทรโปลิส-เฮสติง ที่มีการปรับโดย Zamar (2006)

วิธีการของเมโทรโปลิส-เฮสติงที่มีการปรับโดย Zamar [9] จะมีการปรับที่เซ็ต $\boldsymbol{V}$ เนื่องจากการคำนวณหาค่า $\boldsymbol{V}$ ด้วยวิธีที่เสนอโดย Forster และคณะ อาจใช้เวลาในการคำนวณนานโดยเฉพาะกรณีข้อมูลขนาดใหญ่ ดังนั้น Zamar จึงได้มีการปรับที่เซ็ต $\boldsymbol{V}$ โดยเพิ่มเงื่อนไขที่ว่า

$$|v_i| \leq m_i \qquad \text{สำหรับทุก ๆ ค่าของ } 1 \leq i \leq n \tag{37}$$

โดยเงื่อนไขที่เพิ่มมาข้างต้นจะสามารถแก้ปัญหาที่เกิดขึ้นในกรณีที่ $|v_i| > m_i$ เพราะถ้าหาก $|v_i| > m_i$ แล้วจะทำให้ค่า $d = 0$ และจะส่งผลทำให้ $\boldsymbol{y}' = \boldsymbol{y}$

ขั้นตอนของวิธีการของเมโทรโปลิส-เฮสติงที่มีการปรับโดย Zamar สามารถสรุปได้ดังนี้

1. สร้างเซ็ต $\boldsymbol{V}$ โดยที่ $\boldsymbol{V} = \{\boldsymbol{v}: \sum_{i=1}^{n} |v_i| \leq r$, $|v_i| \leq m_i$ และ v_i คือจำนวนเฉพาะสัมพัทธ์สำหรับ $i = 1, \dots, n$, $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{v} = \boldsymbol{0}\}$
2. เลือก $\boldsymbol{v}$ จากเซ็ตของ $\boldsymbol{V}$ ด้วยความน่าจะเป็นแบบ Uniform
3. หาทุก ๆ ค่า d โดยที่ $0 \leq y_i + dv_i \leq m_i$ และกำหนดให้ k คือจำนวนของค่า d ที่เป็นไปได้ทั้งหมด
4. คำนวณค่า $\sum_{i=1}^{k} \eta(d_i|\boldsymbol{v},\boldsymbol{y})$ และ กำหนด $P(d_i) = \frac{\eta(d_i|\boldsymbol{v},\boldsymbol{y})}{\sum_{j=1}^{k}\eta(d_j|\boldsymbol{v},\boldsymbol{y})}$ $\quad i,j = 1,2,\dots,k$
5. เลือก d ด้วยความน่าจะเป็นที่กำหนดในขั้นตอนที่ 4
6. กำหนดให้ $\boldsymbol{y}' = \boldsymbol{y} + d\boldsymbol{v}$
7. ให้ $\boldsymbol{y} = \boldsymbol{y}'$ และกลับไปที่ขั้นตอนที่ 2

เมื่อได้เซ็ตของ $\boldsymbol{y}$ จากขั้นตอนข้างต้น จากนั้นนำเซ็ตของ $\boldsymbol{y}$ ดังกล่าวไปคำนวณหาค่าประมาณพารามิเตอร์เช่นเดียวกับวิธีการถดถอยลอจิสติกแบบแม่นตรง โดยสามารถหาค่าประมาณของพารามิเตอร์ β_p ได้ด้วยวิธีภาวะน่าจะเป็นสูงสุด จากสมการที่ (38) ดังนี้

$$\widehat{l_c}(\beta_p) = \frac{\hat{c}(t_0,\dots,t_p)e^{\beta_p t_p}}{\sum_u \hat{c}(t_0,\dots,t_{p-1},u)e^{\beta_p u}} \tag{38}$$

โดยในงานวิจัยนี้ใช้วิธีการเมโทรโปลิส-เฮสติงที่มีการปรับโดย Zamar ในการสร้างลูกโซ่มาร์คอฟมอนติคาร์โล เพื่อลดเวลาในการสร้างเซ็ต $\boldsymbol{V}$ และศึกษาเฉพาะกรณีที่ตัวแปรตอบสนองเป็นตัวแปรจำแนกประเภทแบบ 2 กลุ่ม หรือ $Y_i \sim \mathrm{Bernoulli}(\pi_i)$

2.3 ข้อมูลแบบแบ่งแยกสมบูรณ์ ข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ และข้อมูลแบบทับซ้อน

2.3.1 ข้อมูลแบบแบ่งแยกสมบูรณ์ (Complete Separation)

ข้อมูลที่เป็นแบบแบ่งแยกสมบูรณ์ คือ ข้อมูลที่ค่าของตัวแปรอธิบาย (X) สามารถแบ่งกลุ่มของตัวแปรตอบสนองระหว่างกลุ่มที่สนใจ $(Y = 1)$ และ กลุ่มที่ไม่สนใจ $(Y = 0)$ ได้อย่างสมบูรณ์

ตัวอย่างที่ 1: กำหนดให้ $y_i = 0$ เมื่อ $x_{\mathrm{i}} < 0$ และ $\mathrm{y_i} = 1$ เมื่อ $x_{\mathrm{i}} \geq 0$ สำหรับ $i = 1,2, \dots ,10$

ตารางที่ 2 ตัวอย่างข้อมูลแบบแบ่งแยกสมบูรณ์

i	1	2	3	4	5	6	7	8	9	10
x_{i}	-5	-4	-3	-2	-1	1	2	3	2	5
$\mathrm{y_i}$	0	0	0	0	0	1	1	1	1	1

ตารางที่ 3 ตารางแจกแจงความถี่แบบสองทางสำหรับข้อมูลแบบแบ่งแยกสมบูรณ์

ตัวแปร X	ตัวแปร Y	
	0	1
0	5	0
1	0	5

จาก

$$\hat{\beta}_1 = \log \frac{f_{11} f_{22}}{f_{12} f_{21}}$$

จะได้ว่า

$$\hat{\beta}_1 = log \frac{(5)(5)}{(0)(0)}$$

จากตารางแจกแจงความถี่แบบสองทางสำหรับข้อมูลแบบแบ่งแยกสมบูรณ์ พบว่าค่าความถี่นอกแนวทแยงมุมมีค่าเป็น 0 สองค่า ดังนั้นพบว่าการคำนวณค่าประมาณสัมประสิทธิ์การถดถอย $(\hat{\beta}_1)$ โดยวิธี ML ไม่สามารถหาค่าได้เนื่องจากตัวเศษมีค่าเป็น 0 [2]

2.3.2 ข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ (Quasi-Complete Separation)

ข้อมูลที่เป็นแบบแบ่งแยกกึ่งสมบูรณ์ คือ ข้อมูลที่ค่าของตัวแปรอธิบาย (X) สามารถแบ่งกลุ่มของตัวแปรตอบสนองระหว่างกลุ่มที่สนใจ $(Y = 1)$ และ กลุ่มที่ไม่สนใจ $(Y = 0)$ ได้อย่างค่อนข้างสมบูรณ์

ตัวอย่างที่ 2: กำหนดให้ $y_i = 0$ เมื่อ $x_i < 0$, $\mathrm{y_i} = 1$ เมื่อ $\mathrm{x_i} > 0$ และที่ $x_i = 0$ ค่าสังเกต y_i จะมีค่าเท่ากับ 0 หรือ 1 สำหรับ $i = 1,2, \dots ,12$

ตารางที่ 4 ตัวอย่างข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์

$\boldsymbol{i}$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
$\boldsymbol{x}_{\mathbf{i}}$	-5	-4	-3	-2	-1	0	0	1	2	3	2	5
$\mathbf{y_i}$	0	0	0	0	0	0	1	1	1	1	1	1

ตารางที่ 5 ตารางแจกแจงความถี่แบบสองทางสำหรับข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์

ตัวแปร $\boldsymbol{X}$	**ตัวแปร $\boldsymbol{Y}$**	
	0	**1**
0	6	1
1	0	5

$$\hat{\beta}_1 = log \frac{(6)(5)}{(0)(1)}$$

จากตารางแจกแจงความถี่แบบสองทางสำหรับข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ พบว่า ค่าความถี่นอกแนวทแยงมุมมีค่าเป็น 0 อยู่หนึ่งค่า ดังนั้นพบว่าการคำนวณค่าประมาณสัมประสิทธิ์การถดถอย $(\hat{\beta}_1)$ โดยวิธี ML ไม่สามารถหาค่าได้เนื่องจากตัวเศษมีค่าเป็น 0 [2]

2.3.3 ข้อมูลแบบทับซ้อน (Overlap)

ข้อมูลที่เป็นแบบทับซ้อน คือ ข้อมูลที่ค่าของตัวแปรอธิบาย (X) ไม่สามารถแบ่งกลุ่มของตัวแปรตอบสนองระหว่างกลุ่มที่เราสนใจ $(Y = 1)$ และ กลุ่มที่เราไม่สนใจ $(Y = 0)$ ได้

ตารางที่ 6 ตัวอย่างข้อมูลแบบทับซ้อน

i	1	2	3	4	5	6	7	8	9	10
x_i	-5	2	-3	-2	-1	0	0	1	2	3
$\mathbf{y}_\mathbf{i}$	0	0	1	1	0	0	1	0	1	1

จากตารางที่ 6 พบว่าค่าของตัวแปรอธิบายไม่สามารถแบ่งแยกระหว่างกลุ่มของตัวแปรตอบสนองที่เราสนใจ และกลุ่มที่เราไม่สนใจได้ กล่าวคือค่าสังเกตุของตัวแปรอธิบายจะมีการทับซ้อนกันระหว่างกลุ่มของตัวแปรตอบสนองที่เราสนใจ และไม่สนใจ เช่น ที่ค่า $x_\mathrm{i} = 2$ จะเห็นว่า x_2 ให้ค่า $y_2 = 0$ แต่ที่ x_9 ให้ค่า $y_9 = 1$

ตารางที่ 7 ตารางแจกแจงความถี่แบบสองทางสำหรับข้อมูลแบบทับซ้อน

ตัวแปร X	ตัวแปร Y	
	0	1
0	3	3
1	1	3

$$\hat{\beta}_1 = log\frac{(3)(3)}{(1)(3)}$$

จากตารางแจกแจงความถี่แบบสองทางสำหรับข้อมูลแบบทับซ้อน ไม่พบค่าความถี่นอกแนวทแยงมุมที่มีค่าเป็น 0 ดังนั้นการคำนวณค่าประมาณสัมประสิทธิ์การถดถอย $(\hat{\beta}_1)$ โดยวิธี ML สามารถกระทำได้

2.4 งานวิจัยที่เกี่ยวข้อง

ในปี 2000 Mehta และคณะ [4] เสนอ วิธีโครงข่ายสำหรับการสุ่มตัวอย่างโดยตรงด้วยมอนติคาร์โล (Network Algorithm for Direct Monte Carlo Sampling) เพื่อแก้ไขปัญหาที่เกิดขึ้นจากการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด ที่ไม่สามารถใช้ในการประมาณค่าพารามิเตอร์ในกรณีข้อมูลแบบแบ่งแยกสมบูรณ์ หรือ ข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ได้ และแก้ไขปัญหาที่เกิดขึ้นจากวิธีการถดถอยลอจิสติกแบบแม่นตรง ที่มีการใช้เวลาในการวิเคราะห์ข้อมูลเป็นเวลานาน โดยวิธีโครงข่ายสำหรับการสุ่มตัวอย่างโดยตรงด้วยมอนติคาร์โล ได้มีการนำเทคนิคมอนติคาร์โลมาประยุกต์ใช้วิธีการถดถอยลอจิสติกแบบแม่นตรง เพื่อแก้ไขปัญหาดังกล่าวที่เกิดขึ้นกับการประมาณค่าพารามิเตอร์ด้วยวิธีภาวะน่าจะเป็นสูงสุด และวิธีการถดถอยลอจิสติกแบบแม่นตรง

ในปี 2002 King และ Ryan [3] ได้ทำการศึกษาเปรียบเทียบค่าประมาณพารามิเตอร์ที่ได้จาก วิธีภาวะน่าจะเป็นสูงสุด และวิธีการถดถอยลอจิสติกแบบแม่นตรง ในกรณีที่ข้อมูลมีจำนวนค่าสังเกตทับซ้อนต่างกัน กรณีร้อยละของค่าตอบสนองต่างกัน และกรณีเปรียบเทียบค่าพารามิเตอร์และค่าประมาณพารามิเตอร์ นอกจากนี้ยังมีการนำข้อมูลจริงหลายๆชุดมาวิเคราะห์ด้วยวิธีการประมาณค่าพารามิเตอร์ทั้ง 2 วิธี เพื่อเปรียบเทียบผลลัพธ์ที่ได้ ซึ่งจากงานวิจัยนี้พบว่า วิธี ML มีประสิทธิภาพต่ำในบางกรณีของข้อมูล ซึ่งเป็นการชี้ให้เห็นว่าก่อนที่จะทำการวิเคราะห์ข้อมูลควรตระหนักถึงลักษณะของข้อมูลก่อน เพื่อที่จะได้เลือกใช้วิธีการประมาณค่าพารามิเตอร์ที่เหมาะสมกับลักษณะของข้อมูล

ในปี 2007 Zamar และคณะ [10] ได้พัฒนาแพ็คเกจที่ชื่อว่า elrm หรือ โปรแกรมที่ใช้อนุมานค่าของตัวแบบถดถอยลอจิสติกแบบแม่นตรง (Software Implementing Exact-like Inference for Logistic Regression Models) ซึ่งเป็นแพ็คเกจที่สร้างขึ้นในโปรแกรม R เพื่อใช้ในการวิเคราะห์ข้อมูลด้วยตัวแบบการถดถอยลอจิสติก โดยใช้การสุ่มข้อมูลด้วยวิธีการลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับวิธีการถดถอยลอจิสติกแบบแม่นตรง ในการสร้างลูกโซ่มาร์คอฟแบบเมโทรโปลิส-เฮสติง ที่มีการปรับโดย Zamar เพื่อแก้ไขปัญหาที่เกิดขึ้นเกี่ยวกับเวลาในการวิเคราะห์ข้อมูลของวิธีการถดถอยลอจิสติกแบบแม่นตรง

บทที่ 3

วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก 3 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง ทำการเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบจุด และการประมาณค่าพารามิเตอร์แบบช่วง ในกรณีที่ข้อมูลมีจำนวนค่าสังเกตทับซ้อน และกรณีที่ข้อมูลมีร้อยละของค่าตอบสนองต่างกัน โดยมีวิธีดำเนินการศึกษา ดังนี้

3.1. วิธีการวิจัย

3.1.1 กรณีตัวแปรอธิบาย 1 ตัวแปร

1.) สุ่มตัวแปรอธิบายโดยกำหนดตัวแปรอธิบายแบบต่อเนื่อง (X) ที่มีการแจกแจงแบบ Uniform (6, 25) และสร้างค่าสังเกตของตัวแปรอธิบายขนาด $n = 10, 20, 30, 50, 100$ และ 200

2.) กำหนดค่าพารามิเตอร์ $\beta_0 = -18.49$ และ $\beta_1 = 0.88$

3.) นำค่าสังเกตของตัวแปรอธิบาย และพารามิเตอร์ ที่ได้จากขั้นตอนที่ 1 และ 2 ไปสร้างตัวแปร Y จากความสัมพันธ์ดังนี้

$$\pi(x) = P(Y = 1) = \frac{\exp(\beta_0 + \beta_1 x_1)}{1 + \exp(\beta_0 + \beta_1 x_1)}$$

4.) กำหนดให้ตัวแปร Y มีค่าเท่ากับ 0 หรือ 1 จากค่า $\pi(x)$ ที่ได้ ตามจำนวนค่าสังเกตทับซ้อน และร้อยละของค่าตอบสนอง โดยกำหนด จำนวนค่าสังเกตทับซ้อน (Degree Overlap) มีค่าเท่ากับ $1,2,3,4, 6$ และ 8 ร้อยละของค่าตอบสนอง (Percent of Responses) คือ $20,30,40$ และ 50

5.) ทำการประมาณค่าพารามิเตอร์ และหาช่วงความเชื่อมั่นของพารามิเตอร์ ด้วยวิธีการประมาณค่าพารามิเตอร์ 3 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับวิธีการถดถอยลอจิสติกแบบแม่นตรงโดยวิธีของ Zamar

6.) ทำซ้ำขั้นตอนที่ 1-5 จำนวน 200 ครั้ง

7.) เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ โดยจำแนกเป็น

7.1) กรณีประมาณค่าพารามิเตอร์แบบจุด: พิจารณาจากค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย หรือ $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$

7.2) กรณีประมาณค่าพารามิเตอร์แบบช่วง: พิจารณาจากค่าประมาณค่าความน่าจะเป็นครอบคลุม ($\widehat{\mathrm{CP}}$) และค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น ($\widehat{\mathrm{AW}}$)

3.1.2 กรณีตัวแปรอธิบาย 2 ตัวแปร

ในกรณีการศึกษานี้ ผู้วิจัยได้ทำการประมาณค่าพารามิเตอร์โดยใช้ตัวแปรอธิบายไม่ต่อเนื่อง 2 ตัวแปร ซึ่งไม่สอดคล้องกับเงื่อนไขของตัวแบบการถดถอยลอจิสติกพื้นฐานพหุคูณ (Multiple Logistic Regression Model) เนื่องจาก ตัวแบบการถดถอยลอจิสติกพื้นฐานพหุคูณ คือ ตัวแบบการถดถอยลอจิสติกที่มีตัวแปรอธิบาย (X) แบบต่อเนื่องอย่างน้อย 1 ตัว และตัวแปรอธิบายอื่นๆ 1 ตัวขึ้นไป แต่ในกรณีศึกษานี้จะเป็นการประมาณค่าพารามิเตอร์ในตัวแบบลอจิต (Logit Model) ซึ่งวิธีการประมาณค่าพารามิเตอร์ที่ศึกษาในงานวิจัยนี้สามารถใช้ในตัวแบบดังกล่าวได้ และมีวิธีดำเนินการศึกษา ดังนี้

1.) สุ่มตัวแปรอธิบายโดยกำหนดตัวแปรอธิบายไม่ต่อเนื่องหรือตัวแปรเชิงกลุ่ม X_1 และ X_2 และสร้างค่าสังเกตของตัวแปรอธิบายขนาด $n = 12, 20, 28, 48, 100$ และ 200 โดยที่

ข้อมูล 1 ใน 4 ส่วนแรก กำหนดให้ X_1 และ X_2 มีค่าเท่ากับ 0

ข้อมูล 1 ใน 4 ส่วนที่สอง กำหนดให้ X_1 มีค่าเท่ากับ 0 และ X_2 มีค่าเท่ากับ 1

ข้อมูล 1 ใน 4 ส่วนที่สาม กำหนดให้ X_1 มีค่าเท่ากับ 1 และ X_2 มีค่าเท่ากับ 0

ข้อมูล 1 ใน 4 ส่วนที่สี่ กำหนดให้ X_1 และ X_2 มีค่าเท่ากับ 1

2.) กำหนดค่าพารามิเตอร์โดยขึ้นอยู่กับค่าประมาณร้อยละของค่าตอบสนอง เมื่อกำหนดร้อยละของค่าตอบสนองมีค่าประมาณ $10, 20,30,40$ และ 50 ดังนี้

- ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 10

กำหนดค่า $\beta_0 = -2.53, \beta_1 = 0.45$ และ $\beta_2 = 0.19$

- ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 20

กำหนดค่า $\beta_0 = -1.99, \beta_1 = 0.01$ และ $\beta_2 = 1.04$

- ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 30

กำหนดค่า $\beta_0 = -1.99, \beta_1 = 0.07$ และ $\beta_2 = 1.86$

- ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 40

กำหนดค่า $\beta_0 = -1.49, \beta_1 = 0.09$ และ $\beta_2 = 1.89$

- ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 50

กำหนดค่า $\beta_0 = -1.49, \beta_1 = 1.09$ และ $\beta_2 = 1.89$

3.) นำค่าสังเกตของตัวแปรอธิบาย และพารามิเตอร์ ที่ได้จากขั้นตอนที่ 1 และ 2 ไปสร้างค่าความน่าจะเป็นที่ตัวแปรตอบสนอง (Y) จะอยู่ในกลุ่มที่สนใจ หรือ $\pi(x_1, x_2)$ ดังนี้

$$\pi(x_1, x_2) = P(Y = 1) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}$$

4.) สร้างค่าของตัวแปรสุ่ม Y โดยที่ $Y_i \sim Bernoulli(\pi_i(x_1, x_2))$

5.) ทำการประมาณค่าพารามิเตอร์ และหาช่วงความเชื่อมั่นของพารามิเตอร์ ด้วยวิธีการประมาณค่าพารามิเตอร์ 2 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด และวิธีการถดถอยลอจิสติกแบบแม่นตรง

6.) ทำซ้ำขั้นตอนที่ 1-5 จำนวน 200 ครั้ง

7.) เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ โดยจำแนกเป็น

7.1) กรณีประมาณค่าพารามิเตอร์แบบจุด: พิจารณาจากค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย หรือ $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$

7.2) กรณีประมาณค่าพารามิเตอร์แบบช่วง: พิจารณาจากค่าประมาณค่าความน่าจะเป็นครอบคลุม $(\widehat{\mathrm{CP}})$ และค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น $(\widehat{\mathrm{AW}})$

3.2. เกณฑ์การตัดสินใจ

เกณฑ์ที่ใช้พิจารณาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก 3 วิธี จำแนกเป็น 2 กรณี ได้แก่

3.2.1 กรณีประมาณค่าพารามิเตอร์แบบจุด

เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์จากค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย หรือ $\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\beta}})$ โดยวิธีการประมาณค่าพารามิเตอร์ที่ให้ค่า $\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\beta}})$ ต่ำสุด เป็นวิธีที่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์แบบจุดสูงสุด สำหรับการประมาณค่าพารามิเตอร์ทั้ง 3 วิธี โดยที่

ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error : MSE) คือ ค่าที่ใช้วัดความผิดพลาดของการประมาณค่าพารามิเตอร์ ซึ่งเป็นค่าเฉลี่ยของความแตกต่างระหว่างค่าประมาณพารามิเตอร์และค่าพารามิเตอร์ในตัวแบบยกกำลังสอง ค่าประมาณของ MSE สามารถคำนวณได้ดังนี้

$$\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\beta}}) = \sum_{k=1}^{m} \frac{(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)^T (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k)}{m}$$

เมื่อ $\hat{\beta}_k = [\hat{\beta}_{k1}, \dots, \hat{\beta}_{kp}]^T$ คือ เวกเตอร์ของค่าประมาณพารามิเตอร์ในการทำซ้ำครั้งที่ k และ $\beta_k = [\beta_{k1}, \dots, \beta_{kp}]^T$ คือ เวกเตอร์ของค่าพารามิเตอร์ในการทำซ้ำครั้งที่ k โดยที่ $k = 1, \dots, m$ และ m คือ จำนวนการทำซ้ำของการทดลอง ซึ่งมีค่าเท่ากับ 200

3.2.2 กรณีประมาณค่าพารามิเตอร์แบบช่วง

เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์จากค่าประมาณค่าความน่าจะเป็นครอบคลุม ($\widehat{\mathrm{CP}}$) และค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น ($\widehat{\mathrm{AW}}$) โดยที่

ความน่าจะเป็นครอบคลุม (Coverage Probability : CP) คือ ค่าความน่าจะเป็นที่ช่วงความเชื่อมั่นของค่าพารามิเตอร์ครอบคลุมค่าของพารามิเตอร์ ซึ่งค่าประมาณของ CP คือ

$$\widehat{\mathrm{CP}} = \frac{\text{จำนวนครั้งทั้งหมดที่ช่วงความเชื่อมั่นครอบคลุมค่าพารามิเตอร์}}{m}$$

เมื่อ m คือ จำนวนการทำซ้ำของการทดลอง ซึ่งมีค่าเท่ากับ 200

จากนั้นพิจารณาค่าประมาณค่าความน่าจะเป็นครอบคลุมที่เท่ากับระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% ซึ่งใช้เกณฑ์การพิจารณาจากการทดสอบสมมติฐาน โดยใช้ตัวสถิติทดสอบ Z ดังนี้

H_0: ค่าความน่าจะเป็นครอบคลุมเท่ากับระดับความเชื่อมั่นที่กำหนด $(\mathrm{cp} = cp_0)$
H_1: ค่าความน่าจะเป็นครอบคลุมไม่เท่ากับระดับความเชื่อมั่นที่กำหนด $(\mathrm{cp} \neq cp_0)$

ตัวสถิติที่ใช้ในการทดสอบคือ

$$Z = \frac{\widehat{cp} - cp_0}{\sqrt{\frac{cp_0(1 - cp_0)}{m}}}$$

บริเวณวิกฤตในการทดสอบคือ $Z < -Z_{\frac{\alpha}{2}}$ และ $Z > Z_{\frac{\alpha}{2}}$

และจะยอมรับสมมติฐาน H_0 เมื่อ $-Z_{\frac{\alpha}{2}} \leq Z \leq Z_{\frac{\alpha}{2}}$

นั่นคือจะสรุปว่า ค่าความน่าจะเป็นครอบคลุมไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ถ้า

$$cp_0 - Z_{\frac{\alpha}{2}}\,\mathrm{SE}(\widehat{cp}) \leq \widehat{cp} \leq cp_0 + Z_{\frac{\alpha}{2}}\,\mathrm{SE}(\widehat{cp})$$

เมื่อ cp_0 แทน ระดับความเชื่อมั่นที่กำหนด

cp แทน ค่าความน่าจะเป็นครอบคลุมของพารามิเตอร์

$\widehat{cp}$ แทน ค่าประมาณค่าความน่าจะเป็นครอบคลุม

α แทน ระดับนัยสำคัญของการทดสอบ ในที่นี้มีค่าเท่ากับ 0.05

$\mathrm{SE}(\widehat{cp})$ แทน ความคลาดเคลื่อนมาตรฐานของค่าประมาณค่าความน่าจะเป็นครอบคลุม ซึ่งประมาณได้จาก $\mathrm{SE}(\widehat{cp}) = \sqrt{\frac{cp_0(1-cp_0)}{m}}$

m แทน จำนวนการทำซ้ำของการทดลองซึ่งมีค่าเท่ากับ 200

ดังนั้น จะได้ว่าค่าความน่าจะเป็นครอบคลุมไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ถ้าค่าประมาณค่าความน่าจะเป็นครอบคลุมอยู่ในช่วง [0.9198, 0.9802]

จากนั้นหาค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น โดยที่

ความกว้างเฉลี่ยของช่วงความเชื่อมั่น (Average Width : AW) คือ ค่าเฉลี่ยของผลต่างระหว่างค่าขอบเขตบนและขอบเขตล่างของช่วงความเชื่อมั่น โดยค่าประมาณของ AW สามารถคำนวณได้ดังนี้

$$\widehat{\mathrm{AW}} = \frac{\sum_{k=1}^{m}\sum_{j=1}^{p}(U_{kj}-L_{kj})}{m}$$

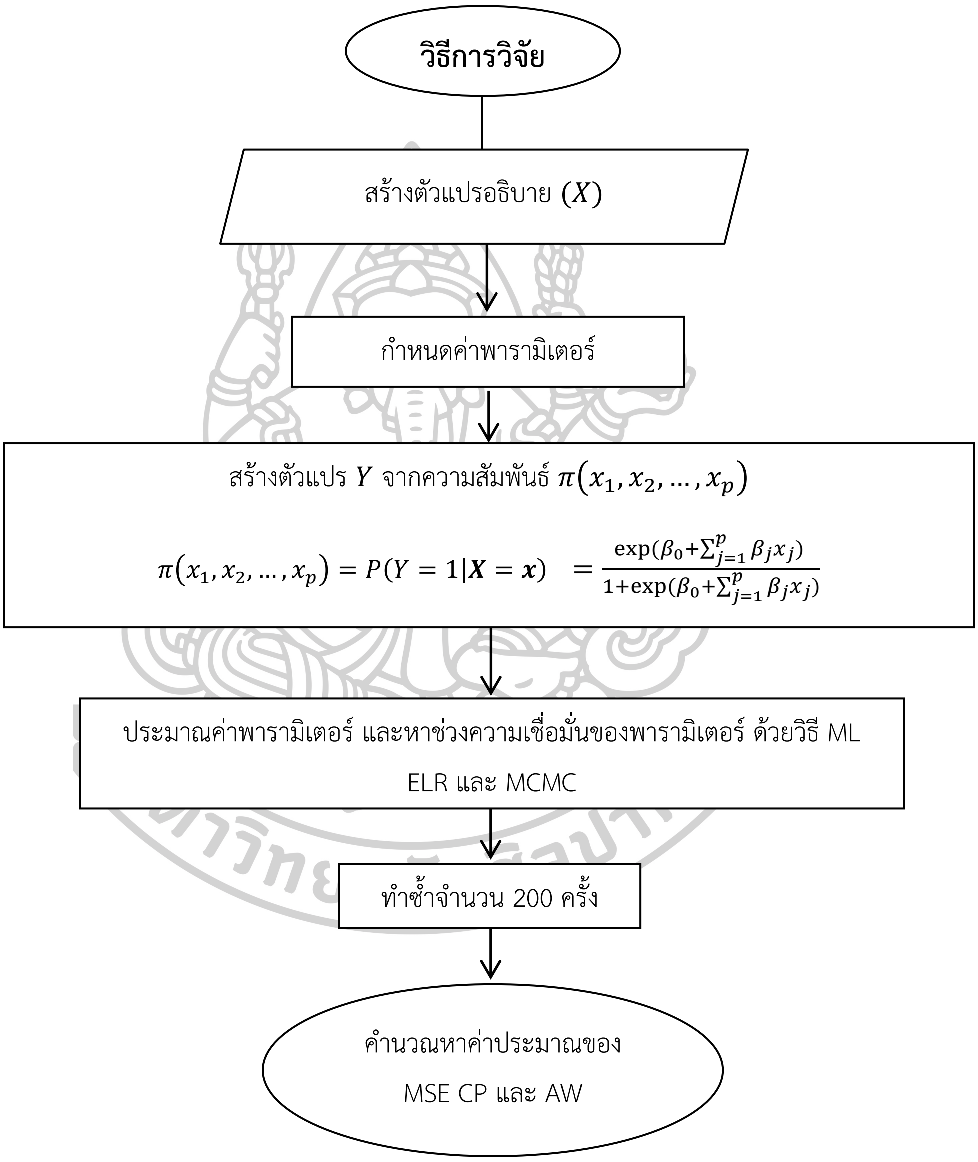
เมื่อ $\sum_{j=1}^{p}(U_{kj} - L_{kj}) = [(U_{k1} - L_{k1}) + \cdots + (U_{kp} - L_{kp})]$ คือผลรวมของผลต่างระหว่างค่าขอบเขตล่างและขอบเขตบนของช่วงความเชื่อมั่นของพารามิเตอร์ p ตัว ในการทำซ้ำครั้งที่ k โดยที่ $k = 1, \dots, m$ และ m คือ จำนวนการทำซ้ำของการทดลอง ซึ่งมีค่าเท่ากับ 200

และพิจารณาค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นเฉพาะกรณีที่ให้ค่าประมาณค่าความน่าจะเป็นครอบคลุม ($\widehat{\mathrm{CP}}$) อยู่ในช่วงความเชื่อมั่นที่กำหนด โดยวิธีการประมาณค่าพารามิเตอร์ที่ให้ค่า $\widehat{\mathrm{AW}}$ น้อยสุด เป็นวิธีที่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์แบบช่วงสูงสุด สำหรับการประมาณค่าพารามิเตอร์ทั้ง 3 วิธี

3.3. เครื่องมือที่ใช้ในการวิจัย

โปรแกรม R เวอร์ชั่น 3.4.0 และ โปรแกรม SAS เวอร์ชั่น 9.3

จากวิธีการวิจัยข้างต้น สามารถสรุปเป็นขั้นตอนโดยรวมได้ดัง ภาพที่ 3 ดังนี้



ภาพที่ 3 แผนผังวิธีการศึกษาในงานวิจัย

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูลนำเสนอเป็นผลการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก 3 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง โดยเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบจุด และการประมาณค่าพารามิเตอร์แบบช่วง การนำเสนอผลการวิเคราะห์ข้อมูลกำหนดสัญลักษณ์เพื่อประกอบการนำเสนอผลการวิเคราะห์ ดังนี้

n แทน ขนาดตัวอย่างที่ใช้ในการศึกษา

ML แทน วิธีภาวะน่าจะเป็นสูงสุด

ELR แทน วิธีการถดถอยลอจิสติกแบบแม่นตรง

MCMC แทน วิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง

สัญลักษณ์ของเกณฑ์การพิจารณา

$\widehat{\mathrm{MSE}}(\hat{\boldsymbol{\beta}})$ แทน ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$\widehat{\mathrm{CP}}$ แทน ค่าประมาณค่าความน่าจะเป็นครอบคลุม

$\widehat{\mathrm{AW}}$ แทน ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น

ผลการศึกษาการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก 3 วิธี นำเสนอเปรียบเทียบค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสำหรับการประมาณค่าพารามิเตอร์แบบจุด ค่าประมาณค่าความน่าจะเป็นครอบคลุมและค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นสำหรับการประมาณค่าพารามิเตอร์แบบช่วง โดยแบ่งเป็น 2 กรณี ได้แก่

- กรณีที่ 1 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 1 ตัวแปร และเป็นตัวแปรอธิบายแบบต่อเนื่อง
- กรณีที่ 2 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 2 ตัวแปร และเป็นตัวแปรอธิบายแบบไม่ต่อเนื่อง

4.1 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 1 ตัวแปร และเป็นตัวแปรอธิบายแบบต่อเนื่อง

จากการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 1 ตัวแปร เมื่อกำหนดจำนวนค่าสังเกตทับซ้อนเท่ากับ $1, 2, 3, 4, 6$ และ 8 กำหนดร้อยละของค่าตอบสนองคือ $20, 30, 40$ และ 50 และกำหนดขนาดตัวอย่างมีค่าเท่ากับ $10, 20, 30, 50, 100$ และ 200 โดยมีผลการศึกษาดังนี้

4.1.1 การประมาณค่าพารามิเตอร์แบบจุด

1.) ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ตารางที่ 8 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 20

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30		
	ML	ELR	MCMC	ML	ELR	MCMC	ML	ELR	MCMC
1	3.3240	2.1056	2.7649	20.4586	14.1038	1.1644	38.8787	22.2236	0.6258
2	0.4551	0.4844	0.4571	0.1457	0.1441	0.8928	0.2912	0.1961	0.6084
3	-	-	-	0.2620	0.2877	1.4738	0.1015	0.1156	0.6116
4	-	-	-	0.4690	0.4838	0.5237	0.2056	0.2248	0.6165
6	-	-	-	-	-	-	0.4730	0.4829	1.0171
8	-	-	-	-	-	-	-	-	-

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

จากตารางที่ 8 พบว่า ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 10 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 10) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 20 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 5) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 30 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 3.33) ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี ML และ ELR มีค่าสูง โดยค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ จากวิธี ML มีค่าสูงกว่าวิธี ELR ทั้งนี้เกิดจากข้อมูลในกรณีนี้มีร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำ จึงมีลักษณะใกล้เคียงกับข้อมูลแบบแบ่งแยกสมบูรณ์หรือข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ที่ได้จึงมีค่าสูง แต่ที่ขนาดตัวอย่างและจำนวนค่าสังเกตทับซ้อนอื่น ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ จากวิธี ML และ ELR มีค่าต่ำและมีค่าใกล้เคียงกัน และพบว่าค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ จากวิธี MCMC โดยส่วนใหญ่จะมีค่ามากที่สุด

ตารางที่ 9 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 20

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	34.2701	**0.6607**	348.8810	**0.6577**	2085.0590	**0.6788**
2	2.2698	**0.6446**	12.2432	**0.6674**	75.6261	**0.6742**
3	**0.2823**	0.6447	3.4847	**0.6907**	21.7324	**0.6658**
4	**0.0856**	0.6360	1.0853	**0.6685**	8.5617	**0.6671**
6	**0.1528**	0.6424	**0.1127**	0.6457	2.4931	**0.6676**
8	**0.3316**	0.6026	**0.0275**	0.6618	0.7990	**0.6731**

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ตัวพิมพ์หนา หมายถึง ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่มีค่าต่ำที่สุด

จากตารางที่ 9 พบว่า ที่ร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำกว่าหรือเท่ากับ 4 ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี MCMC มีค่าต่ำกว่าวิธี ML แต่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมากกว่า 4 ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี ML มีค่าต่ำกว่าวิธี MCMC เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนเพิ่มขึ้น ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี ML มีแนวโน้มลดลง แต่ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี MCMC มีแนวโน้มคงที่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมีการเปลี่ยนแปลง เนื่องด้วยขนาดตัวอย่างในกรณีนี้มีขนาดใหญ่ และการประมาณค่าพารามิเตอร์วิธี ELR ต้องใช้หน่วยความจำของเครื่องคอมพิวเตอร์เยอะ การประมาณค่าพารามิเตอร์โดยวิธี ELR จึงไม่สามารถกระทำได้เนื่องจากหน่วยความจำไม่เพียงพอต่อการคำนวณค่าประมาณพารามิเตอร์

ตารางที่ 10 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกรณีข้อมูลมีร้อยละของค่าตอบสนอง คือ 30

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30		
	ML	ELR	MCMC	ML	ELR	MCMC	ML	ELR	MCMC
1	2.9460	1.9059	2.6795	11.2767	7.4776	0.5162	16.4613	10.8810	0.5861
2	0.3618	0.3991	0.3549	0.1653	0.1443	0.5188	0.4309	0.2885	0.5797
3	0.6945	0.7018	0.6952	0.2030	0.2287	0.4973	0.0753	0.0861	0.5650
4	-	-	-	0.3662	0.3855	0.9089	0.1503	0.1685	0.4967
6	-	-	-	0.7016	0.7050	0.6816	0.3863	0.3990	0.5791
8	-	-	-	-	-	-	0.5939	0.5997	0.6059

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

จากตารางที่ 10 พบว่า ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 10 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 10) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 20 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 5) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 30 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 3.33) ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี ML และ ELR มีค่าสูง โดยค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ จากวิธี ML มีค่าสูงกว่าวิธี ELR ทั้งนี้เกิดจากข้อมูลในกรณีนี้มีร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำ จึงมีลักษณะใกล้เคียงกับข้อมูลแบบแบ่งแยกสมบูรณ์หรือข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ที่ได้จึงมีค่าสูง แต่ที่ขนาดตัวอย่างและจำนวนค่าสังเกตทับซ้อนอื่น ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ จากวิธี ML และ ELR มีค่าต่ำและมีค่าใกล้เคียงกัน และพบว่าค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ จากวิธี MCMC โดยส่วนใหญ่จะมีค่ามากที่สุด

ตารางที่ 11 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 30

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	71.5932	**0.6204**	397.2690	**0.6337**	1274.7390	**0.64440**
2	2.3366	**0.6203**	13.4060	**0.6446**	74.1510	**0.6516**
3	**0.2631**	0.6188	3.8013	**0.6461**	18.9114	**0.6592**
4	**0.0548**	0.6013	1.0569	**0.6331**	8.2940	**0.6412**
6	**0.1198**	0.5832	**0.1521**	0.6340	2.3694	**0.6549**
8	**0.2710**	0.6559	**0.0216**	0.6277	0.8880	**0.6301**

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้
ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ตัวพิมพ์หนา หมายถึง ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่มีค่าต่ำที่สุด

จากตารางที่ 11 พบว่า ที่ร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำกว่าหรือเท่ากับ 4 ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี MCMC มีค่าต่ำกว่าวิธี ML แต่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมากกว่า 4 ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี ML มีค่าต่ำกว่าวิธี MCMC เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนเพิ่มขึ้น ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี ML มีแนวโน้มลดลง แต่ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี MCMC มีแนวโน้มคงที่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมีการเปลี่ยนแปลง เนื่องด้วยขนาดตัวอย่างในกรณีนี้มีขนาดใหญ่ และการประมาณค่าพารามิเตอร์วิธี ELR ต้องใช้หน่วยความจำของเครื่องคอมพิวเตอร์เยอะ การประมาณค่าพารามิเตอร์โดยวิธี ELR จึงไม่สามารถกระทำได้เนื่องจากหน่วยความจำไม่เพียงพอต่อการคำนวณค่าประมาณพารามิเตอร์

ตารางที่ 12 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 40

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30	
	ML	ELR	MCMC	ML	ELR	MCMC	ML	MCMC
1	4.5255	2.5065	0.8583	14.3105	9.0552	0.5178	41.0706	0.5891
2	0.3095	0.3481	0.3268	0.1347	0.1236	0.5202	0.5615	0.5818
3	0.6248	0.6388	0.6299	0.1950	0.2204	0.4185	0.0688	0.5753
4	1.0204	0.9932	1.0223	0.3317	0.3523	0.5102	0.1422	0.5464
6	-	-	-	0.6329	0.6397	0.5958	0.3517	0.7605
8	-	-	-	0.9942	0.9821	1.2178	0.5385	0.4539

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

จากตารางที่ 12 พบว่า ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 10 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 10) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 20 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 5) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 30 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 3.33) ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี ML และ ELR มีค่าสูง โดยค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี ML มีค่าสูงกว่าวิธี ELR ทั้งนี้เกิดจากข้อมูลในกรณีนี้มีร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำ จึงมีลักษณะใกล้เคียงกับข้อมูลแบบแบ่งแยกสมบูรณ์หรือข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จึงมีค่าสูง แต่ที่ขนาดตัวอย่างและจำนวนค่าสังเกตทับซ้อนอื่น ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี ML และ ELR มีค่าต่ำและมีค่าใกล้เคียงกัน และพบว่าค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี MCMC โดยส่วนใหญ่จะมีค่ามากที่สุด

ตารางที่ 13 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 40

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	54.1925	**0.6343**	245.8246	**0.6386**	3902.7160	**0.6351**
2	2.6013	**0.6173**	12.3353	**0.6450**	74.9311	**0.6371**
3	**0.2477**	0.6131	3.4466	**0.6331**	21.2108	**0.6438**
4	**0.0679**	0.6086	1.0296	**0.6363**	9.5431	**0.6470**
6	**0.1018**	0.5861	**0.1395**	0.6336	2.5972	**0.6527**
8	**0.2343**	0.5956	**0.0225**	0.6521	0.8658	**0.6455**

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ตัวพิมพ์หนา หมายถึง ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่มีค่าต่ำที่สุด

จากตารางที่ 13 พบว่า ที่ร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำกว่าหรือเท่ากับ 4 ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี MCMC มีค่าต่ำกว่าวิธี ML แต่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมากกว่า 4 ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี ML มีค่าต่ำกว่าวิธี MCMC เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนเพิ่มขึ้น ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี ML มีแนวโน้มลดลง แต่ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี MCMC มีแนวโน้มคงที่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมีการเปลี่ยนแปลง เนื่องด้วยขนาดตัวอย่างในกรณีนี้มีขนาดใหญ่ และการประมาณค่าพารามิเตอร์วิธี ELR ต้องใช้หน่วยความจำของเครื่องคอมพิวเตอร์เยอะ การประมาณค่าพารามิเตอร์โดยวิธี ELR จึงไม่สามารถกระทำได้เนื่องจากหน่วยความจำไม่เพียงพอต่อการคำนวณค่าประมาณพารามิเตอร์

ตารางที่ 14 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 50

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30	
	ML	ELR	MCMC	ML	ELR	MCMC	ML	MCMC
1	1.8739	1.1563	1.0719	8.5114	4.9852	0.5390	16.3693	0.5795
2	0.2943	0.3332	0.8065	0.1756	0.1507	0.4925	0.3294	0.5912
3	0.5976	0.6139	0.6054	0.1783	0.2046	0.4164	0.0784	0.5692
4	0.9474	0.9284	0.9420	0.3295	0.3500	0.4871	0.1407	0.5477
6	-	-	-	0.6061	0.6140	0.7619	0.3425	0.7252
8	-	-	-	0.9448	0.9356	7.7933	0.5319	0.4616

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

จากตารางที่ 14 พบว่า ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 10 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 10) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 20 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 5) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 30 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 3.33) ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี ML และ ELR มีค่าสูง โดยค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี ML มีค่าสูงกว่าวิธี ELR ทั้งนี้เกิดจากข้อมูลในกรณีนี้มีร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำ จึงมีลักษณะใกล้เคียงกับข้อมูลแบบแบ่งแยกสมบูรณ์หรือข้อมูลแบบแบ่งแยกกึ่งสมบูรณ์ ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จึงมีค่าสูง แต่ที่ขนาดตัวอย่างและจำนวนค่าสังเกตทับซ้อนอื่น ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี ML และ ELR มีค่าต่ำและมีค่าใกล้เคียงกัน และพบว่าค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี MCMC โดยส่วนใหญ่จะมีค่ามากที่สุด

ตารางที่ 15 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 50

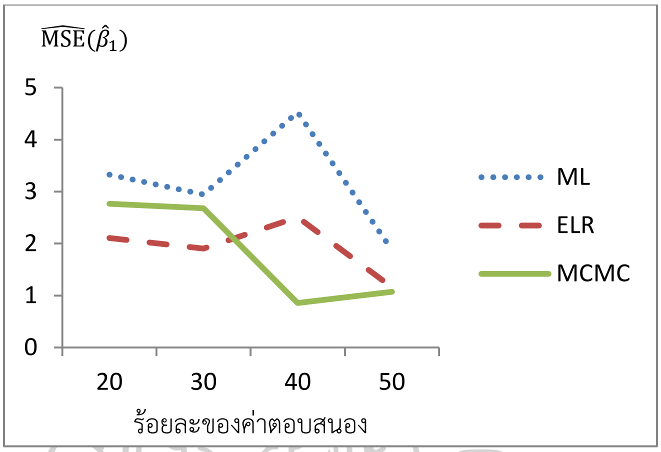
จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	35.5249	**0.6284**	325.5575	**0.6381**	1502.0590	**0.6405**
2	2.5447	**0.6237**	14.0725	**0.6354**	78.8282	**0.6425**
3	**0.2859**	0.6244	3.9567	**0.6288**	22.0534	**0.6471**
4	**0.0631**	0.6175	1.0870	**0.6395**	8.6244	**0.6385**
6	**0.0996**	0.5789	**0.1466**	0.6370	2.6752	**0.6436**
8	**0.2305**	0.5601	**0.0244**	0.6160	0.7841	**0.6371**

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ตัวพิมพ์หนา หมายถึง ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่มีค่าต่ำที่สุด

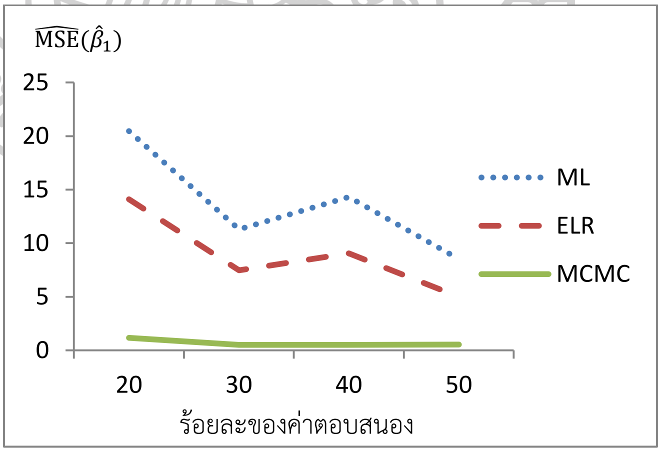
จากตารางที่ 15 พบว่า ที่ร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำกว่าหรือเท่ากับ 4 ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี MCMC มีค่าต่ำกว่าวิธี ML แต่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมากกว่า 4 ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธี ML มีค่าต่ำกว่าวิธี MCMC เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนเพิ่มขึ้น ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี ML มีแนวโน้มลดลง แต่ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ จากวิธี MCMC มีแนวโน้มคงที่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมีการเปลี่ยนแปลง เนื่องด้วยขนาดตัวอย่างในกรณีนี้มีขนาดใหญ่ และการประมาณค่าพารามิเตอร์วิธี ELR ต้องใช้หน่วยความจำของเครื่องคอมพิวเตอร์เยอะ การประมาณค่าพารามิเตอร์โดยวิธี ELR จึงไม่สามารถกระทำได้เนื่องจากหน่วยความจำไม่เพียงพอต่อการคำนวณค่าประมาณพารามิเตอร์

จากค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย หรือ $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ข้างต้น สามารถแสดงการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกทั้ง 3 วิธี ในเชิงของค่าร้อยละของค่าตอบสนองได้ดังนี้



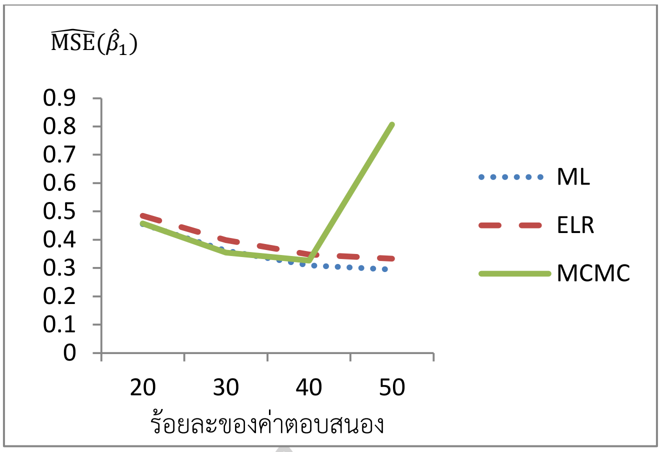
ภาพที่ 4 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ที่ขนาดตัวอย่าง 10

จากภาพที่ 4 พบว่า ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของทั้ง 3 วิธี มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น



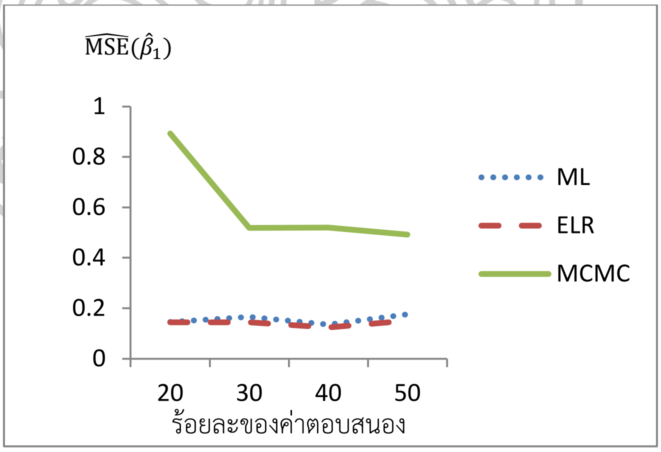
ภาพที่ 5 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ที่ขนาดตัวอย่าง 20

จากภาพที่ 5 พบว่า ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของทั้ง 3 วิธี มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น



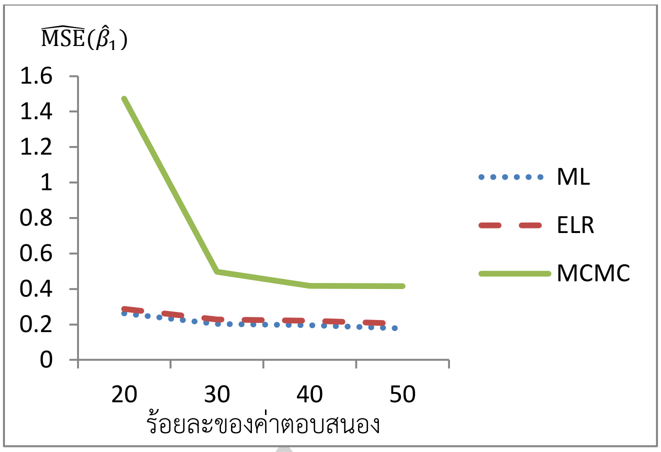
ภาพที่ 6 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 10

จากภาพที่ 6 พบว่า ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ของวิธี ML และ ELR มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น แต่ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ของวิธี MCMC มีแนวโน้มลดลงที่ร้อยละของค่าตอบสนองตั้งแต่ 20 ถึง 40 แต่จะมีค่าเพิ่มขึ้นที่ร้อยละของค่าตอบสนองมากกว่า 40



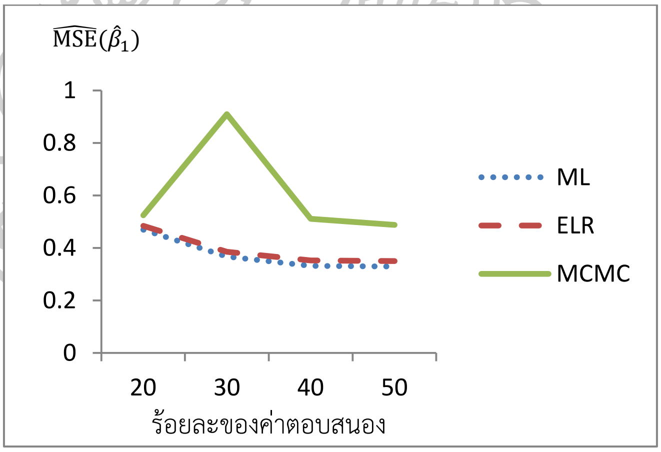
ภาพที่ 7 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 20

จากภาพที่ 7 พบว่า ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ของทั้ง 3 วิธี มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น



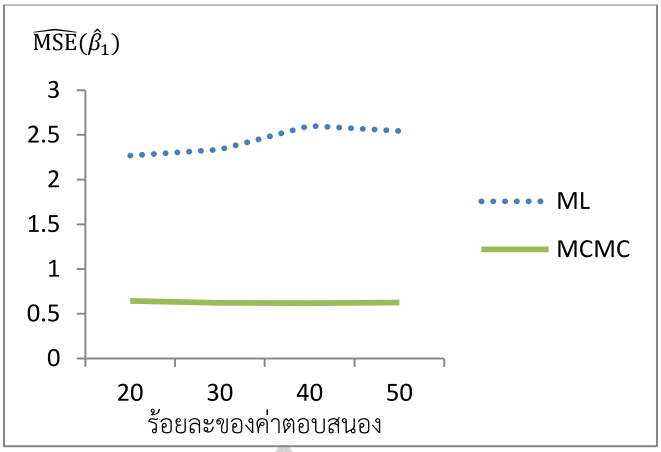
ภาพที่ 8 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 3 ที่ขนาดตัวอย่าง 20

จากภาพที่ 8 พบว่า ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของทั้ง 3 วิธี มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น



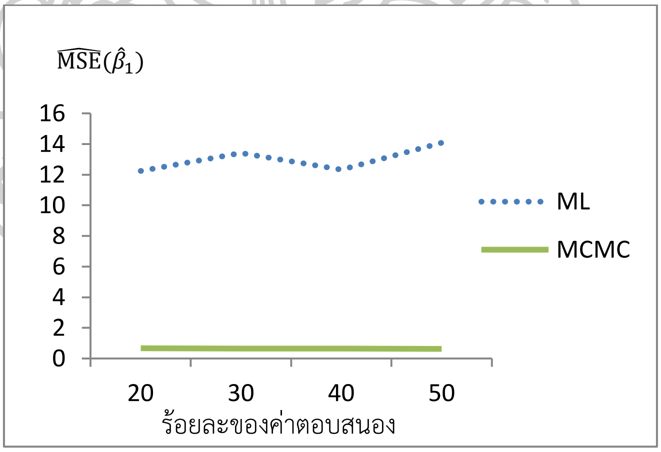
ภาพที่ 9 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 4 ที่ขนาดตัวอย่าง 20

จากภาพที่ 9 พบว่า ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของวิธี ML และ ELR มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น แต่ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของวิธี MCMC มีแนวโน้มเพิ่มขึ้นที่ร้อยละของค่าตอบสนองตั้งแต่ 20 ถึง 30 แต่จะมีค่าลดลงที่ร้อยละของค่าตอบสนองมากกว่า 30



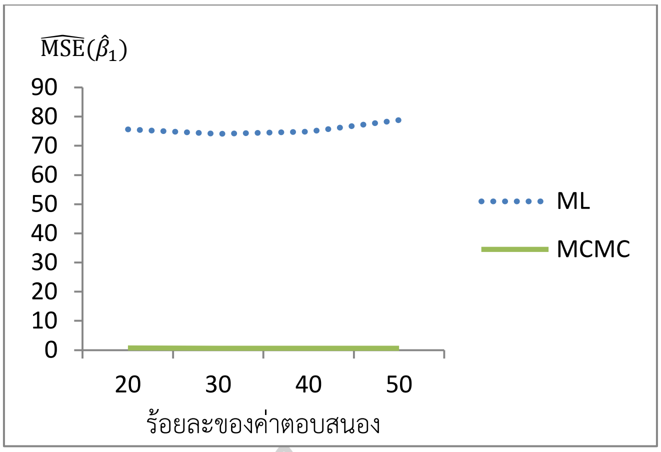
ภาพที่ 10 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 50

จากภาพที่ 10 พบว่า ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ของวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลง แต่ค่า $\widehat{\text{MSE}}$ จากวิธี ML มีแนวโน้มสูงขึ้น



ภาพที่ 11 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 100

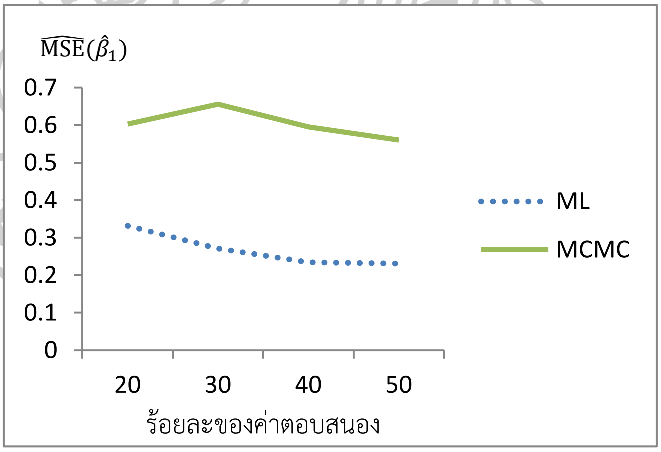
จากภาพที่ 11 พบว่า ค่า $\widehat{\text{MSE}}(\hat{\beta}_1)$ ของวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลง แต่ค่า $\widehat{\text{MSE}}$ จากวิธี ML มีแนวโน้มสูงขึ้น



ภาพที่ 12 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 200

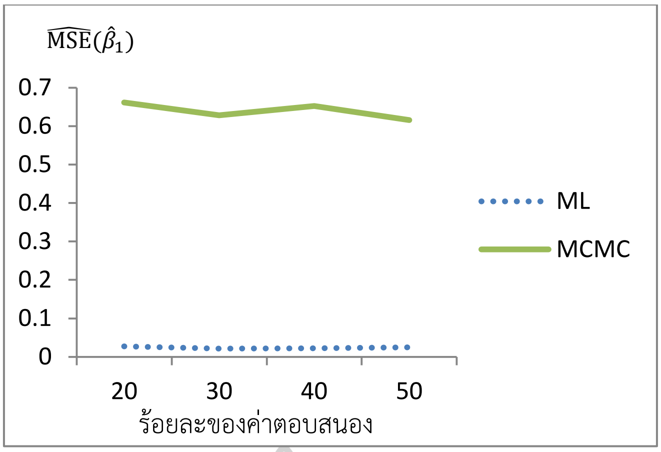
จากภาพที่ 12 พบว่า ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลง แต่ค่า $\widehat{\mathrm{MSE}}$ จากวิธี ML มีแนวโน้มสูงขึ้น



ภาพที่ 13 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

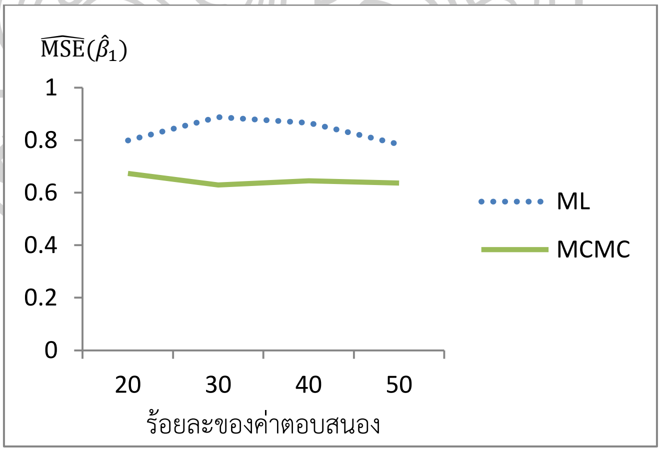
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 50

จากภาพที่ 13 พบว่า ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของทั้ง 2 วิธี มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น



ภาพที่ 14 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 100

จากภาพที่ 14 พบว่า ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของทั้ง 2 วิธี มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น



ภาพที่ 15 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 200

จากภาพที่ 15 พบว่า ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ของทั้ง 2 วิธี มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น

4.1.2 การประมาณค่าพารามิเตอร์แบบช่วง

1.) ค่าประมาณค่าความน่าจะเป็นครอบคลุม

ตารางที่ 16 ค่าประมาณค่าความน่าจะเป็นครอบคลุมกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 20 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30		
	ML	ELR	MCMC	ML	ELR	MCMC	ML	ELR	MCMC
1	0.9000	0.9300*	0.9350*	1.0000	1.0000	0.9950	1.0000	0.9950	1.0000
2	0.1550	0.2300	0.4650	0.8850	0.9100	1.0000	1.0000	1.0000	1.0000
3	-	-	-	0.2950	0.3600	0.9600*	0.8100	0.8300	1.0000
4	-	-	-	0.0050	0.0100	0.7250	0.3400	0.4000	0.9800*
6	-	-	-	-	-	-	0.0000	0.0000	0.7800
8	-	-	-	-	-	-	-	-	-

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

* หมายถึง กรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 16 พบว่า ที่ระดับความเชื่อมั่น 95% ที่ขนาดตัวอย่างเท่ากับ 10 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ELR และ MCMC ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 20 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 3 วิธี MCMC ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 30 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 4 วิธี MCMC ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด นอกจากนี้พบว่าทั้ง 3 วิธีจะให้ค่า $\widehat{\mathrm{CP}}$ ลดลงเมื่อจำนวนค่าสังเกตทับซ้อนเพิ่มขึ้น

ตารางที่ 17 ค่าประมาณค่าความน่าจะเป็นครอบคลุมกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 20 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	1.0000	1.0000	0.9850	1.0000	0.9200*	1.0000
2	1.0000	1.0000	0.9650*	1.0000	0.5150	1.0000
3	0.9900	1.0000	0.9650*	1.0000	0.1350	1.0000
4	0.9667*	1.0000	0.9950	1.0000	0.0200	1.0000
6	0.3600	0.9800*	1.0000	1.0000	0.0250	1.0000
8	0.0050	0.9350*	0.9500*	1.0000	0.2400	1.0000

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

* หมายถึง กรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 17 พบว่า ที่ระดับความเชื่อมั่น 95% ที่ขนาดตัวอย่างเท่ากับ 50 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 4 วิธี ML ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด และจำนวนค่าสังเกตทับซ้อนเท่ากับ 6 และ 8 วิธี MCMC ให้ค่า $\widehat{\mathrm{CP}}$ เท่ากับระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 100 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 2, 3 และ 8 วิธี ML ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 200 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ML ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

ตารางที่ 18 ค่าประมาณค่าความน่าจะเป็นครอบคลุมกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 30 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30		
	ML	ELR	MCMC	ML	ELR	MCMC	ML	ELR	MCMC
1	0.9550*	0.9700*	0.9800*	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.2400	0.2950	0.5500	0.9100*	0.9250*	1.0000	1.0000	1.0000	1.0000
3	0.0000	0.0000	0.0200	0.4000	0.4400	0.9400*	0.8700	0.8950	0.9950
4	-	-	-	0.0750	0.0800	0.8050	0.4950	0.5350	0.9900
6	-	-	-	0.0000	0.0000	0.0400	0.0000	0.0000	0.8500
8	-	-	-	-	-	-	0.0000	0.0000	0.3600

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

* หมายถึง กรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 18 พบว่า ที่ระดับความเชื่อมั่น 95% ที่ขนาดตัวอย่างเท่ากับ 10 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ทั้ง 3 วิธี ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 20 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 วิธี ML และ ELR ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด และที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 3 วิธี MCMC ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด นอกจากนี้พบว่าทั้ง 3 วิธีจะให้ค่า $\widehat{\mathrm{CP}}$ ลดลงเมื่อจำนวนค่าสังเกตทับซ้อนเพิ่มขึ้น

ตารางที่ 19 ค่าประมาณค่าความน่าจะเป็นครอบคลุมกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 30 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	1.0000	1.0000	1.0000	1.0000	0.9400*	1.0000
2	1.0000	1.0000	0.9400*	1.0000	0.5400	1.0000
3	1.0000	1.0000	0.9200*	1.0000	0.1800	1.0000
4	0.9700*	1.0000	0.9650*	1.0000	0.0250	1.0000
6	0.4600	0.9850	1.0000	1.0000	0.0200	1.0000
8	0.0100	0.9900	0.9750*	1.0000	0.2250	1.0000

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

* หมายถึง กรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 19 พบว่า ที่ระดับความเชื่อมั่น 95% ที่ขนาดตัวอย่างเท่ากับ 50 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 4 วิธี ML ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 100 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 2, 3, 4 และ 8 วิธี ML ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 200 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ML ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

ตารางที่ 20 ค่าประมาณค่าความน่าจะเป็นครอบคลุมกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 40 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30	
	ML	ELR	MCMC	ML	ELR	MCMC	ML	MCMC
1	0.9250*	0.9550*	0.9550*	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.3650	0.4150	0.5750	0.8850	0.9000	1.0000	1.0000	1.0000
3	0.0000	0.0000	0.0200	0.3650	0.3950	0.9700*	0.8750	1.0000
4	0.0000	0.0000	0.0000	0.0700	0.0900	0.8600	0.4600	1.0000
6	-	-	-	0.0000	0.0000	0.0600	0.0050	0.9100
8	-	-	-	0.0000	0.0000	0.0000	0.0000	0.4900

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

* หมายถึง กรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 20 พบว่า ที่ระดับความเชื่อมั่น 95% ที่ขนาดตัวอย่างเท่ากับ 10 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ทั้ง 3 วิธี ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 20 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 3 วิธี MCMC ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด นอกจากนี้พบว่าทั้ง 3 วิธีจะให้ค่า $\widehat{CP}$ ลดลงเมื่อจำนวนค่าสังเกตทับซ้อนเพิ่มขึ้น

ตารางที่ 21 ค่าประมาณค่าความน่าจะเป็นครอบคลุมกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 40 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	1.0000	1.0000	1.0000	1.0000	0.9650*	1.0000
2	1.0000	1.0000	0.9550*	1.0000	0.5400	1.0000
3	1.0000	1.0000	0.9250*	1.0000	0.1550	1.0000
4	0.9800*	1.0000	0.9600*	1.0000	0.0600	1.0000
6	0.5200	0.9950	1.0000	1.0000	0.0250	1.0000
8	0.0300	0.9650*	0.9900	1.0000	0.1700	1.0000

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

* หมายถึง กรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 21 พบว่า ที่ระดับความเชื่อมั่น 95% ที่ขนาดตัวอย่างเท่ากับ 50 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 4 วิธี ML ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด และที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 8 วิธี MCMC ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 100 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 2, 3 และ 4 วิธี ML ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 200 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ML ให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

ตารางที่ 22 ค่าประมาณค่าความน่าจะเป็นครอบคลุมกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 50 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30	
	ML	ELR	MCMC	ML	ELR	MCMC	ML	MCMC
1	0.9500*	0.9650*	0.9700*	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.3250	0.3800	0.6500	0.9200*	0.9250*	1.0000	1.0000	1.0000
3	0.0050	0.0100	0.0400	0.4550	0.4950	0.9600*	0.8950	0.9950
4	0.0000	0.0000	0.0000	0.0850	0.1150	0.8500	0.4200	1.0000
6	-	-	-	0.0000	0.0000	0.0950	0.0050	0.9050
8	-	-	-	0.0000	0.0000	0.0000	0.0000	0.4800

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

* หมายถึง กรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 22 พบว่า ที่ระดับความเชื่อมั่น 95% ที่ขนาดตัวอย่างเท่ากับ 10 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ทั้ง 3 วิธี ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 20 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 วิธี ML และ ELR ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด และที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 3 วิธี MCMC ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด นอกจากนี้พบว่าทั้ง 3 วิธีจะให้ค่า $\widehat{\mathrm{CP}}$ ลดลงเมื่อจำนวนค่าสังเกตทับซ้อนเพิ่มขึ้น

ตารางที่ 23 ค่าประมาณค่าความน่าจะเป็นครอบคลุมกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 50 ที่ระดับความเชื่อมั่น 95%

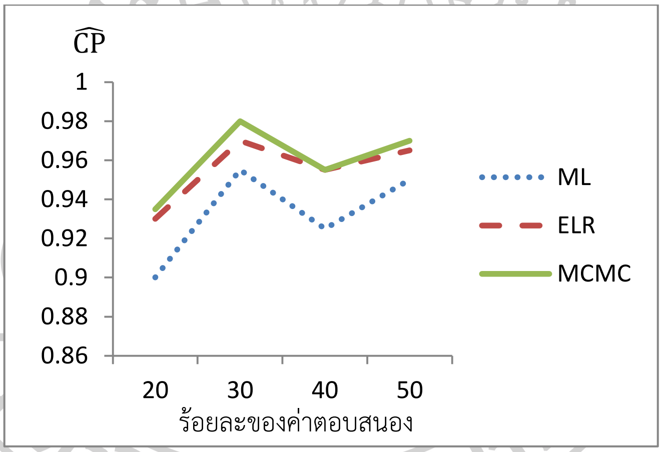
จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	1.0000	1.0000	1.0000	1.0000	0.9750*	1.0000
2	1.0000	1.0000	0.9800*	1.0000	0.5550	1.0000
3	1.0000	1.0000	0.9500*	1.0000	0.1950	1.0000
4	0.9850	1.0000	0.9850	1.0000	0.0100	1.0000
6	0.4750	1.0000	1.0000	1.0000	0.0150	1.0000
8	0.0500	0.9700*	0.9750*	1.0000	0.1900	1.0000

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

* หมายถึง กรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

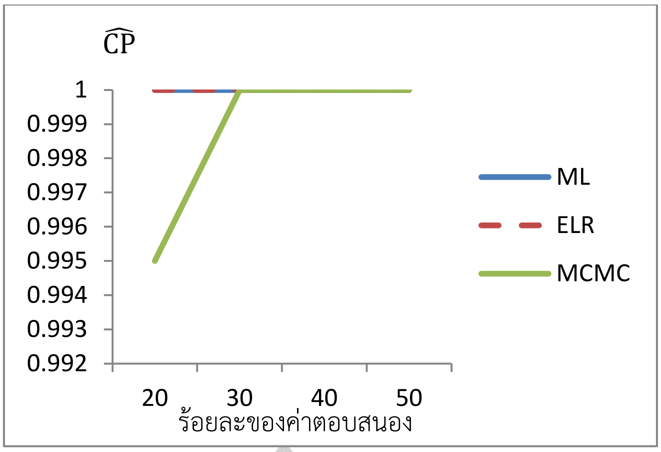
จากตารางที่ 23 พบว่า ที่ระดับความเชื่อมั่น 95% ที่ขนาดตัวอย่างเท่ากับ 50 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 วิธี MCMC ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 100 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 2, 3 และ 8 วิธี ML ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ขนาดตัวอย่างเท่ากับ 200 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ML ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากค่าประมาณค่าความน่าจะเป็นครอบคลุม ($\widehat{\mathrm{CP}}$) ข้างต้น สามารถแสดงการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกทั้ง 3 วิธี ในเชิงของค่าร้อยละของค่าตอบสนองได้ดังนี้



ภาพที่ 16 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ที่ขนาดตัวอย่าง 10

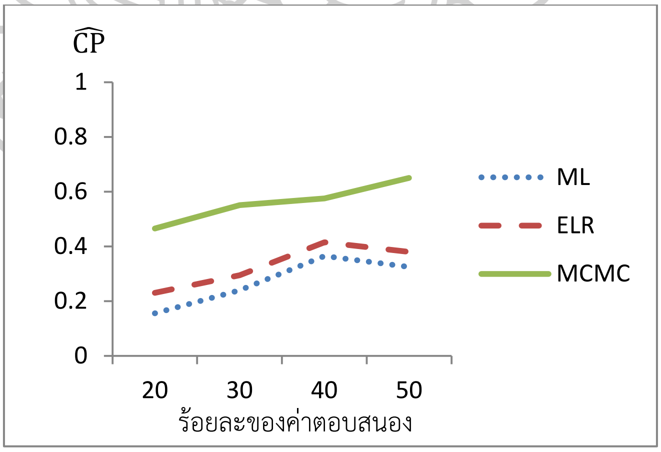
จากภาพที่ 16 พบว่า ค่า $\widehat{\mathrm{CP}}$ ของทั้ง 3 วิธี มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น และโดยส่วนใหญ่มีค่าไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด



ภาพที่ 17 ค่าประมาณค่าความน่าจะเป็นครอบคลุม

กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ที่ขนาดตัวอย่าง 20

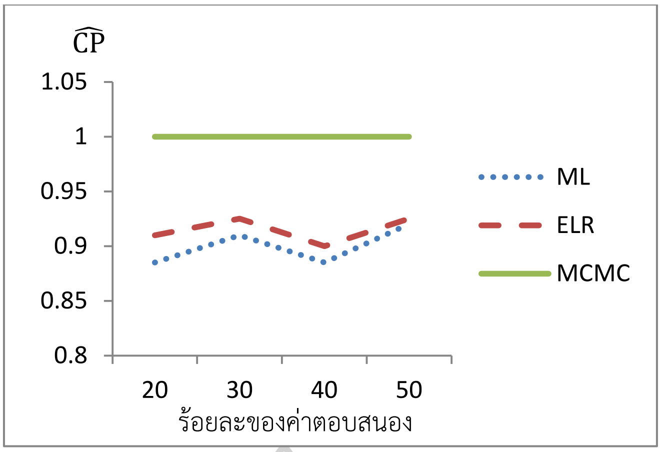
จากภาพที่ 17 พบว่า ทั้ง 3 วิธีจะให้ค่า $\widehat{CP}$ สูงกว่าระดับความเชื่อมั่นที่กำหนด เมื่อร้อยละของค่าตอบสนองเพิ่มขึ้น ค่า $\widehat{CP}$ ของวิธี ML และ ELR มีค่าคงที่เท่ากับ 1 และค่า $\widehat{CP}$ ของวิธี MCMC มีแนวโน้มเพิ่มขึ้น และที่ร้อยละของค่าตอบสนองตั้งแต่ 30 ขึ้นไปค่า $\widehat{CP}$ ของวิธี MCMC มีค่าคงที่เท่ากับ 1 และพบว่าทั้ง 3 วิธีจะให้ค่า $\widehat{CP}$ แตกต่างจากระดับความเชื่อมั่นที่กำหนด



ภาพที่ 18 ค่าประมาณค่าความน่าจะเป็นครอบคลุม

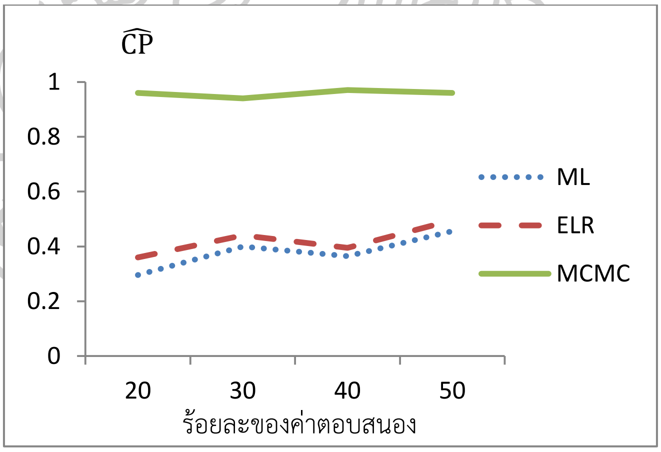
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 10

จากภาพที่ 18 พบว่า ค่า $\widehat{CP}$ ของทั้ง 3 วิธี มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น และมีค่าแตกต่างจากระดับความเชื่อมั่นที่กำหนด



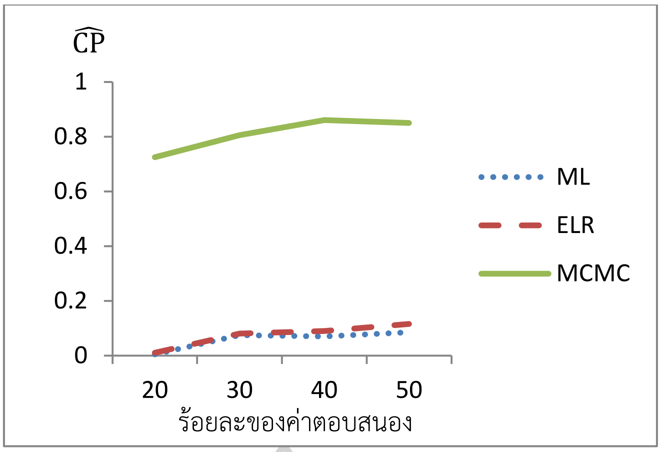
ภาพที่ 19 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 20

จากภาพที่ 19 พบว่า ค่า $\widehat{CP}$ ของวิธี ML และ ELR มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น แต่ค่า $\widehat{CP}$ ของวิธี MCMC มีค่าคงที่เท่ากับ 1 และพบว่าวิธี ELR มีค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนดที่ร้อยละของค่าตอบสนองเท่ากับ 30 และ 50



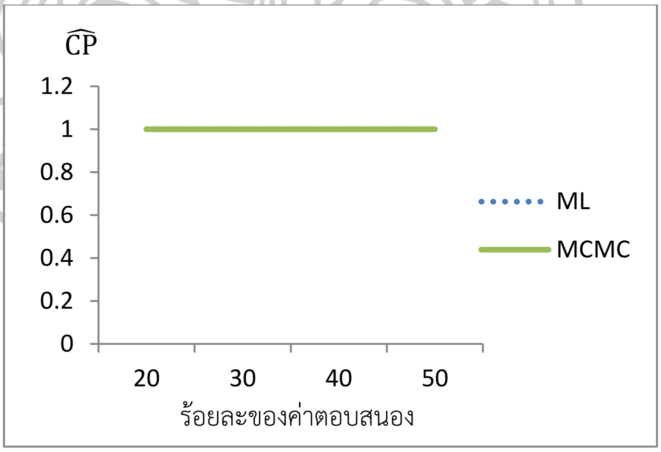
ภาพที่ 20 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 3 ที่ขนาดตัวอย่าง 20

จากภาพที่ 20 พบว่า เมื่อร้อยละของค่าตอบสนองเพิ่มขึ้น ค่า $\widehat{CP}$ ของวิธี ML และ ELR มีแนวโน้มเพิ่มขึ้น และให้ค่า $\widehat{CP}$ แตกต่างจากระดับความเชื่อมั่นที่กำหนด แต่ค่า $\widehat{CP}$ ของวิธี MCMC มีแนวโน้มค่อนข้างคงที่ และให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด



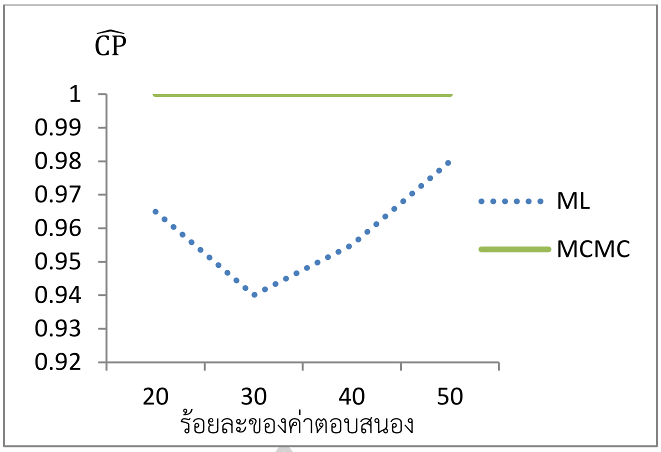
ภาพที่ 21 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 4 ที่ขนาดตัวอย่าง 20

จากภาพที่ 21 พบว่า ค่า $\widehat{CP}$ ของทั้ง 3 วิธี มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละของค่าตอบสนองมีค่าเพิ่มขึ้น และมีค่าแตกต่างจากระดับความเชื่อมั่นที่กำหนด



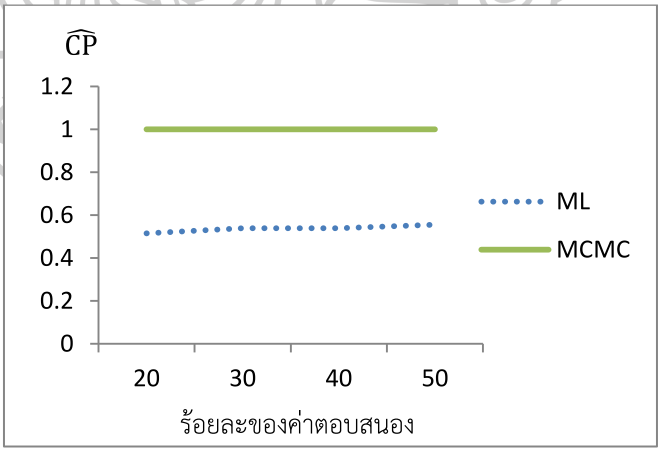
ภาพที่ 22 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 50

จากภาพที่ 22 พบว่า วิธี ML และ MCMC จะให้ค่า $\widehat{CP}$ สูงกว่าระดับความเชื่อมั่นที่กำหนด เมื่อร้อยละของค่าตอบสนองเพิ่มขึ้น ค่า $\widehat{CP}$ ของทั้ง 2 วิธี มีค่าคงที่เท่ากับ 1



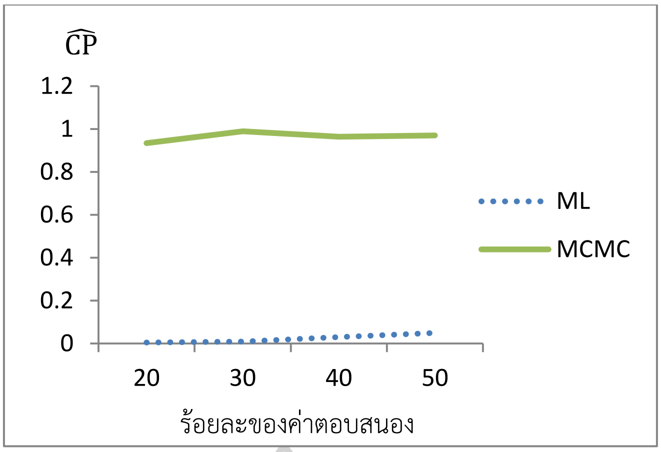
ภาพที่ 23 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 100

จากภาพที่ 23 พบว่า เมื่อร้อยละของค่าตอบสนองเพิ่มขึ้น ค่า $\widehat{CP}$ ของวิธี ML มีแนวโน้มเพิ่มขึ้น และให้ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด แต่ค่า $\widehat{CP}$ ของวิธี MCMC มีค่าคงที่เท่ากับ 1 และแตกต่างจากระดับความเชื่อมั่นที่กำหนด



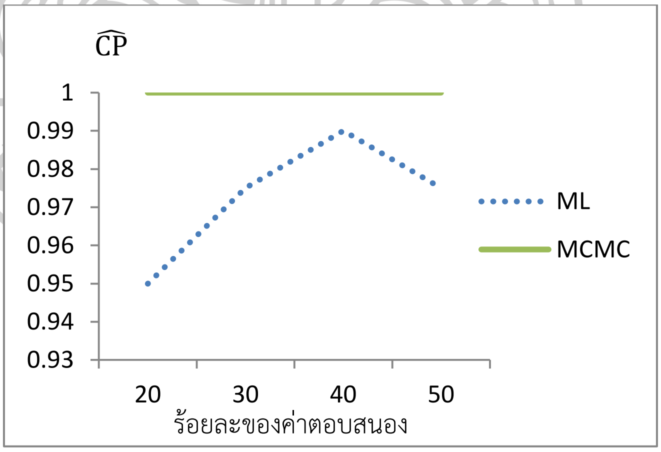
ภาพที่ 24 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 200

จากภาพที่ 24 พบว่า ค่า $\widehat{CP}$ ของวิธี ML และ MCMC มีแนวโน้มคงที่ เมื่อร้อยละของค่าตอบสนองเพิ่มขึ้น และมีค่าแตกต่างจากระดับความเชื่อมั่นที่กำหนด



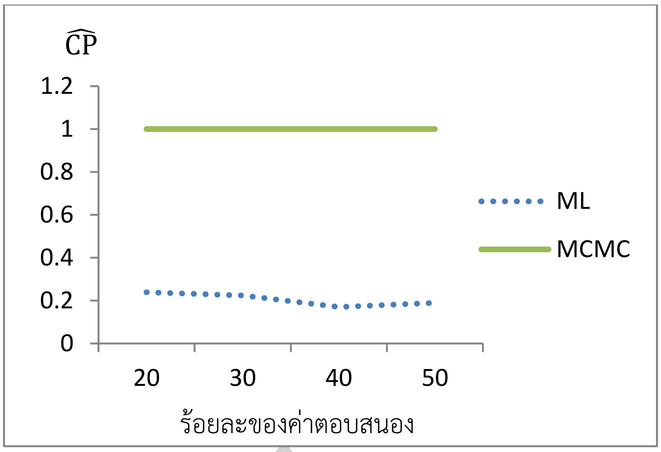
ภาพที่ 25 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 50

จากภาพที่ 25 พบว่า ค่า $\widehat{\text{CP}}$ ของวิธี ML และ MCMC มีแนวโน้มเพิ่มขึ้น เมื่อร้อยละของค่าตอบสนองเพิ่มขึ้น โดยค่า $\widehat{\text{CP}}$ ของวิธี ML มีค่าแตกต่างจากระดับความเชื่อมั่นที่กำหนด แต่ค่า $\widehat{\text{CP}}$ ของวิธี MCMC มีค่าไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด



ภาพที่ 26 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 100

จากภาพที่ 26 พบว่า เมื่อร้อยละของค่าตอบสนองเพิ่มขึ้น ค่า $\widehat{\text{CP}}$ ของวิธี ML มีแนวโน้มเพิ่มขึ้น และให้ค่า $\widehat{\text{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด แต่ค่า $\widehat{\text{CP}}$ ของวิธี MCMC มีค่าคงที่เท่ากับ 1 และแตกต่างจากระดับความเชื่อมั่นที่กำหนด



ภาพที่ 27 ค่าประมาณค่าความน่าจะเป็นครอบคลุม
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 200

จากภาพที่ 27 พบว่า เมื่อร้อยละของค่าตอบสนองเพิ่มขึ้น ค่า $\widehat{CP}$ ของวิธี ML มีแนวโน้มเพิ่มขึ้น และให้ค่า $\widehat{CP}$ แตกต่างจากระดับความเชื่อมั่นที่กำหนด แต่ค่า $\widehat{CP}$ ของวิธี MCMC มีค่าคงที่เท่ากับ 1 และแตกต่างจากระดับความเชื่อมั่นที่กำหนด

2.) ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น

ในการเปรียบเทียบค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น ($\widehat{\mathrm{AW}}$) จะพิจารณาเฉพาะกรณีที่ให้ค่าประมาณค่าความน่าจะเป็นครอบคลุม ($\widehat{\mathrm{CP}}$) ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

ตารางที่ 24 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 20 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30		
	ML	ELR	MCMC	ML	ELR	MCMC	ML	ELR	MCMC
1	7.5701	∞ *	∞*	12.2690	∞	∞	18.7191	∞	∞
2	0.8955	0.9197	∞	1.7696	∞	∞	2.4242	2.2965	∞
3	-	-	-	0.8743	0.8367	∞*	1.2168	1.1444	∞
4	-	-	-	0.5596	0.5547	∞	0.7961	0.7678	∞*
6	-	-	-	-	-	-	0.4389	0.4376	∞
8	-	-	-	-	-	-	-	-	-

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

* หมายถึง ค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 24 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% ขนาดตัวอย่างเท่ากับ 10 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ELR และวิธี MCMC ให้ค่า $\widehat{\mathrm{AW}}$ เท่ากัน ซึ่งมีค่าเป็นอนันต์ (∞)

เนื่องจากในการหาค่าขอบเขตล่างของช่วงความเชื่อมั่นของวิธี ELR และ วิธี MCMC ถ้าค่าสถิติพอเพียงของพารามิเตอร์ที่สนใจมีค่าเท่ากับค่า t_{min} จะได้ว่าค่าขอบเขตล่างของช่วงความเชื่อมั่นมีค่าเป็นอนันต์ทางลบ (-∞) หรือในการหาค่าขอบเขตบนของช่วงความเชื่อมั่นของวิธี ELR และ วิธี MCMC ถ้าค่าสถิติพอเพียงของพารามิเตอร์ที่สนใจมีค่าเท่ากับค่า t_{max} จะได้ว่าค่าขอบเขตบนของช่วงความเชื่อมั่นมีค่าเป็นอนันต์ทางบวก (∞) ดังนั้นค่า $\widehat{\mathrm{AW}}$ จึงมีค่าเป็นอนันต์ (∞)

ตารางที่ 25 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นกรณีข้อมูลมีร้อยละของค่าตอบสนอง คือ 20 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	22.7056	∞	47.8630	∞	111.7802*	∞
2	4.2951	∞	7.8351*	∞	17.7522	∞
3	2.0463	∞	4.0137*	∞	7.9540	∞
4	1.2964*	∞	2.3632	∞	4.6838	∞
6	0.6840	∞*	1.2176	∞	2.4885	∞
8	0.4348	∞*	0.8206*	∞	1.5489	∞

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

* หมายถึง ค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 25 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% โดยส่วนใหญ่วิธี ML จะให้ค่า $\widehat{\mathrm{AW}}$ น้อยที่สุด เนื่องจากวิธี ELR และวิธี MCMC ให้ค่า $\widehat{\mathrm{AW}}$ เป็นอนันต์ (∞)

ตารางที่ 26 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นกรณีข้อมูลมีร้อยละของค่าตอบสนอง คือ 30 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30		
	ML	ELR	MCMC	ML	ELR	MCMC	ML	ELR	MCMC
1	5.3586*	∞*	∞*	11.7545	∞	∞	20.3257	∞	∞
2	0.8912	0.8997	∞	1.8595*	∞*	∞	2.5758	∞	∞
3	0.5689	0.5711	0.6390	0.8821	0.8423	∞*	1.1704	1.1077	∞
4	-	-	-	0.5861	0.5734	∞	0.8125	0.7828	∞
6	-	-	-	0.3767	0.3771	∞	0.4352	0.4300	∞
8	-	-	-	-	-	-	0.3263	0.3259	∞

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

* หมายถึง ค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 26 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{AW}$ ในกรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% ขนาดตัวอย่างเท่ากับ 10 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ML ให้ค่า $\widehat{AW}$ น้อยสุด ที่ขนาดตัวอย่างเท่ากับ 20 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 วิธี ML ให้ค่า $\widehat{AW}$ น้อยสุด

ตารางที่ 27 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 30 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	26.3824	∞	62.6139	∞	113.3491*	∞
2	3.9731	∞	7.9349*	∞	17.4644	∞
3	1.8402	∞	4.0548*	∞	7.4310	∞
4	1.2456*	∞	2.3123*	∞	4.6650	∞
6	0.6633	∞	1.2564	∞	2.4197	∞
8	0.4234	∞	0.8070*	∞	1.5763	∞

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

* หมายถึง ค่า $\widehat{AW}$ ในกรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 27 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{AW}$ ในกรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% โดยส่วนใหญ่วิธี ML จะให้ค่า $\widehat{AW}$ น้อยที่สุด เนื่องจากวิธี ELR และวิธี MCMC ให้ค่า $\widehat{AW}$ เป็นอนันต์ (∞)

ตารางที่ 28 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 40 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30	
	ML	ELR	MCMC	ML	ELR	MCMC	ML	MCMC
1	6.5752*	∞*	∞*	10.6738	∞	∞	16.7258	∞
2	0.9514	0.9413	∞	1.7057	1.6290	∞	2.6773	∞
3	0.5467	0.5435	0.5930	0.8179	0.7860	∞*	1.1797	∞
4	0.5884	0.5846	0.6554	0.5701	0.5577	∞	0.7527	∞
6	-	-	-	0.3604	0.3589	∞	0.4312	∞
8	-	-	-	0.3800	0.3780	∞	0.3172	∞

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

* หมายถึง ค่า $\widehat{AW}$ ในกรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 28 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{AW}$ ในกรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% ขนาดตัวอย่างเท่ากับ 10 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ML ให้ค่า $\widehat{AW}$ น้อยสุด

ตารางที่ 29 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 40 ที่ระดับความเชื่อมั่น 95

จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	24.9139	∞	48.7368	∞	125.1009*	∞
2	4.2752	∞	7.7692*	∞	16.9496	∞
3	1.8791	∞	3.8502*	∞	7.7711	∞
4	1.2259*	∞	2.2805*	∞	4.9719	∞
6	0.6548	∞	1.2081	∞	2.4891	∞
8	0.4309	∞*	0.7885	∞	1.5681	∞

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

* หมายถึง ค่า $\widehat{AW}$ ในกรณีที่ค่า $\widehat{CP}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 29 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% โดยส่วนใหญ่วิธี ML จะให้ค่า $\widehat{\mathrm{AW}}$ น้อยที่สุด เนื่องจากวิธี ELR และวิธี MCMC ให้ค่า $\widehat{\mathrm{AW}}$ เป็นอนันต์ (∞)

ตารางที่ 30 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 50 ที่ระดับความเชื่อมั่น 95%

จำนวนค่าสังเกตทับซ้อน	n = 10			n = 20			n = 30	
	ML	ELR	MCMC	ML	ELR	MCMC	ML	MCMC
1	5.3670*	∞*	∞*	10.8547	∞	∞	15.6243	∞
2	0.9769	0.9581	∞	1.7755*	∞*	∞	2.3412	∞
3	0.5448	0.5425	0.5860	0.8179	0.7824	∞*	1.2441	∞
4	0.5276	0.5224	0.5683	0.5588	0.5467	∞	0.7379	∞
6	-	-	-	0.3648	0.3628	∞	0.4294	∞
8	-	-	-	0.3548	0.3528	∞	0.3103	∞

หมายเหตุ: - หมายถึง ไม่สามารถสร้างข้อมูลในกรณีนั้นได้

* หมายถึง ค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 30 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% ขนาดตัวอย่างเท่ากับ 10 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 วิธี ML ให้ค่า $\widehat{\mathrm{AW}}$ น้อยสุด ที่ขนาดตัวอย่างเท่ากับ 20 และจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 วิธี ML ให้ค่า $\widehat{\mathrm{AW}}$ น้อยสุด

ตารางที่ 31 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นกรณีข้อมูลมีร้อยละของค่าตอบสนองคือ 50 ที่ระดับความเชื่อมั่น 95%

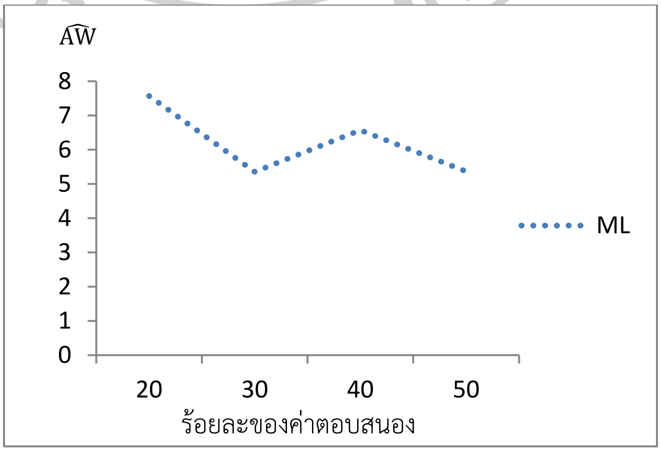
จำนวนค่าสังเกตทับซ้อน	n = 50		n = 100		n = 200	
	ML	MCMC	ML	MCMC	ML	MCMC
1	20.3824	∞	56.7283	∞	144.0579*	∞
2	4.3708	∞	8.4919*	∞	17.7538	∞
3	1.9144	∞	4.1111*	∞	7.9490	∞
4	1.1931	∞	2.3722	∞	4.7148	∞
6	0.6626	∞	1.2163	∞	2.5466	∞
8	0.4266	∞*	0.7845*	∞	1.5188	∞

หมายเหตุ: การประมาณค่าพารามิเตอร์โดยวิธีการถดถอยลอจิสติกแบบแน่นอน (ELR) ไม่สามารถกระทำได้ในกรณีนี้

* หมายถึง ค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

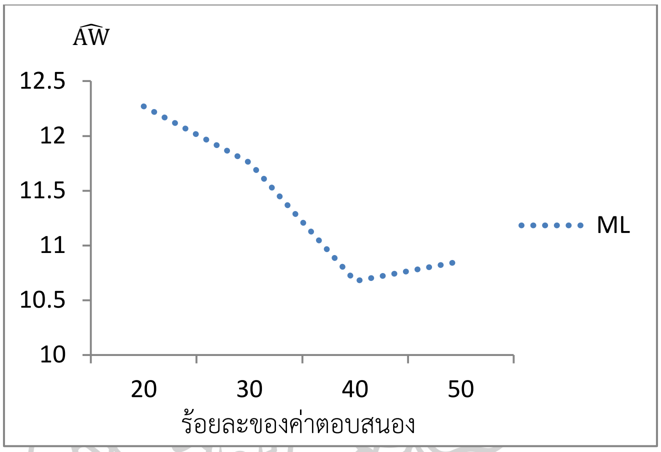
จากตารางที่ 31 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% โดยส่วนใหญ่วิธี ML จะให้ค่า $\widehat{\mathrm{AW}}$ น้อยที่สุด เนื่องจากวิธี ELR และวิธี MCMC ให้ค่า $\widehat{\mathrm{AW}}$ เป็นอนันต์ (∞)

จากค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น ($\widehat{\mathrm{AW}}$) ข้างต้น สามารถแสดงการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกทั้ง 3 วิธี ในเชิงของค่าร้อยละของค่าตอบสนองได้ดังนี้



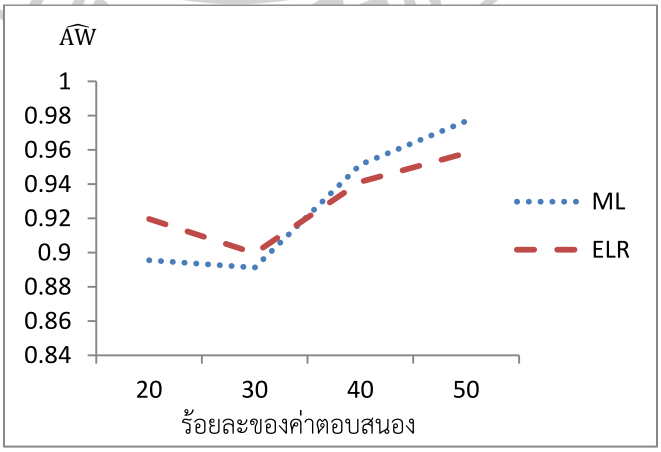
ภาพที่ 28 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ที่ขนาดตัวอย่าง 10

จากภาพที่ 28 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี ELR และวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



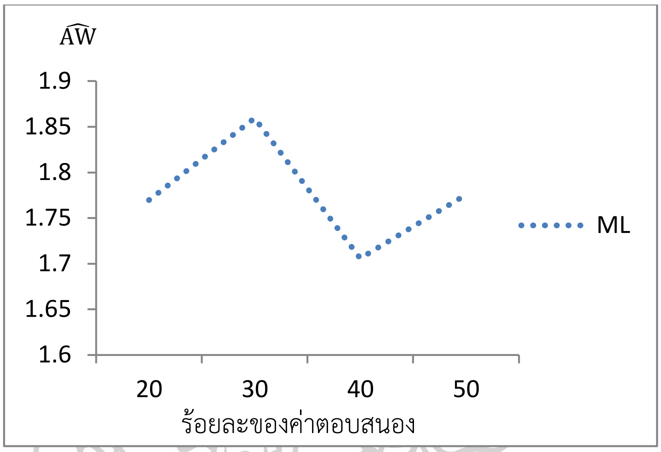
ภาพที่ 29 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 ที่ขนาดตัวอย่าง 20

จากภาพที่ 29 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี ELR และวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



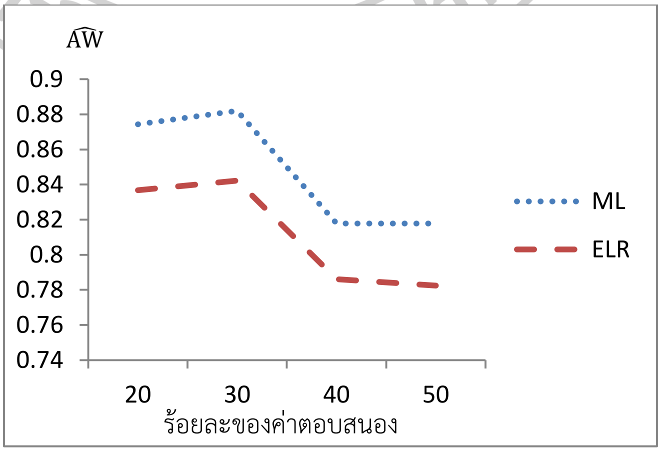
ภาพที่ 30 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 10

จากภาพที่ 30 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML และ ELR มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



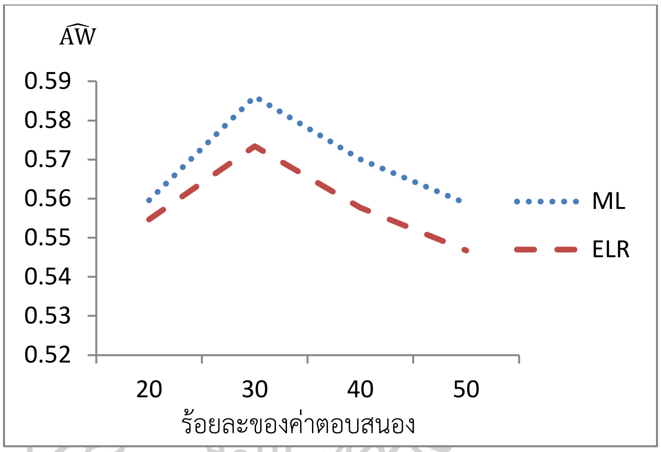
ภาพที่ 31 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 20

จากภาพที่ 31 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี ELR และวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



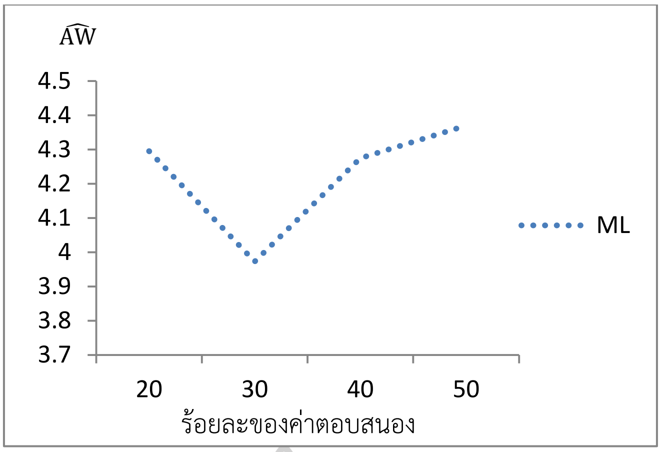
ภาพที่ 32 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 3 ที่ขนาดตัวอย่าง 20

จากภาพที่ 32 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML และ ELR มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



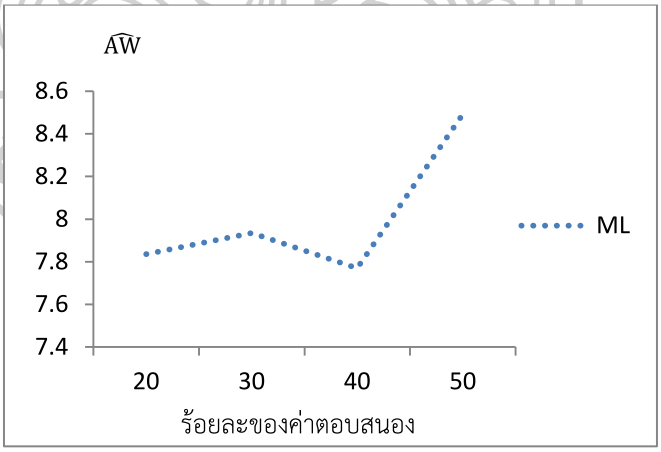
ภาพที่ 33 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 4 ที่ขนาดตัวอย่าง 20

จากภาพที่ 33 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML และ ELR มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



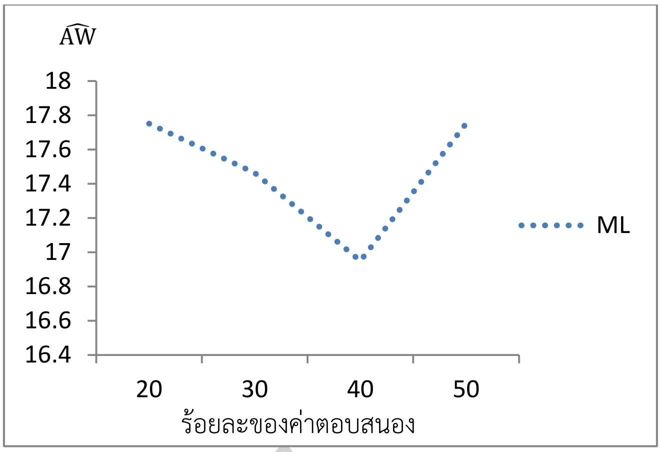
ภาพที่ 34 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 50

จากภาพที่ 34 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



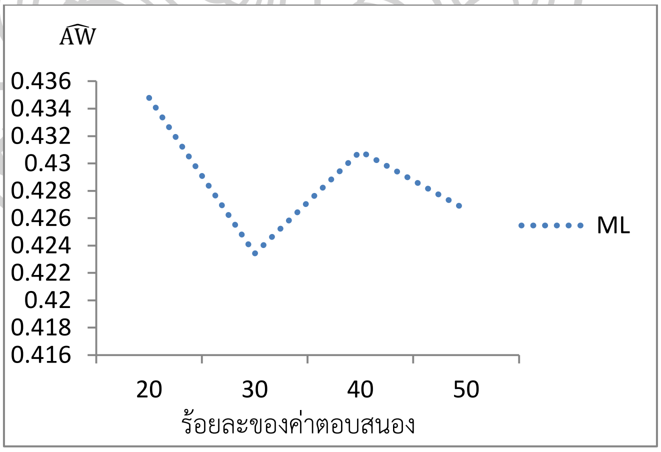
ภาพที่ 35 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 100

จากภาพที่ 35 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



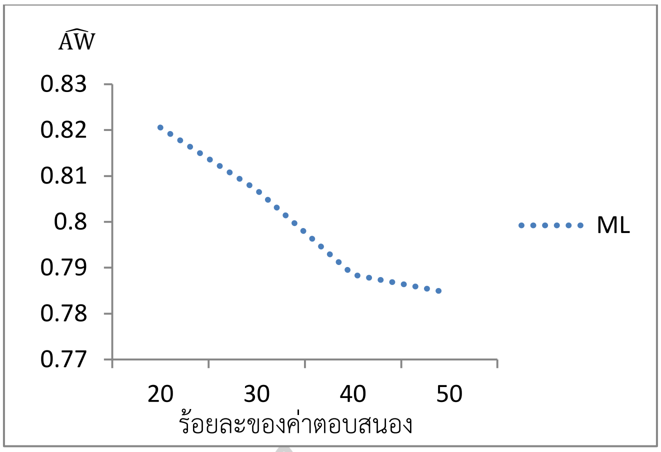
ภาพที่ 36 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 2 ที่ขนาดตัวอย่าง 200

จากภาพที่ 36 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มเพิ่มขึ้นเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



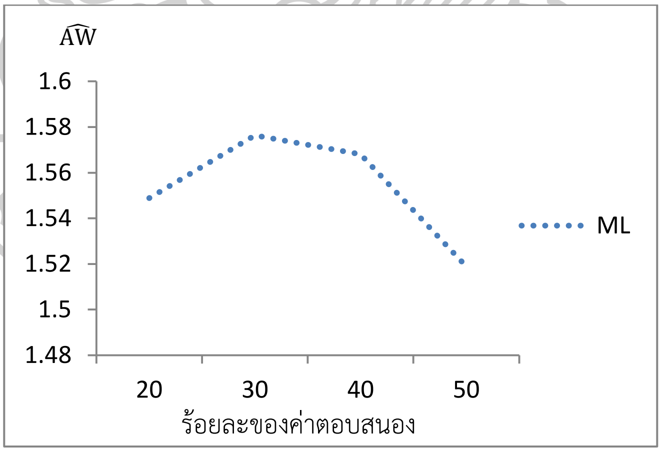
ภาพที่ 37 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 50

จากภาพที่ 37 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



ภาพที่ 38 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 100

จากภาพที่ 38 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)



ภาพที่ 39 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น
กรณีข้อมูลมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 8 ที่ขนาดตัวอย่าง 200

จากภาพที่ 39 พบว่า ค่า $\widehat{AW}$ ที่ได้จากวิธี ML มีแนวโน้มลดลงเมื่อร้อยละของค่าตอบสนองสูงขึ้น ส่วนค่า $\widehat{AW}$ ที่ได้จากวิธี MCMC มีค่าคงที่เมื่อร้อยละของค่าตอบสนองมีการเปลี่ยนแปลงและมีค่าเป็นอนันต์ (∞)

4.2 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 2 ตัวแปร และเป็นตัวแปรอธิบายแบบไม่ต่อเนื่อง

เนื่องจากการประมาณค่าพารามิเตอร์โดยวิธี MCMC ผู้วิจัยเลือกใช้แพ็กเกจสำเร็จรูปในโปรแกรม R ที่ชื่อว่า elrm และทำการประมาณค่าพารามิเตอร์โดยใช้ตัวแปรอธิบายไม่ต่อเนื่อง 2 ตัว แล้วพบว่าแพ็กเกจดังกล่าวต้องใช้เวลาในการประมาณค่าพารามิเตอร์ที่นานเกินกว่าเครื่องคอมพิวเตอร์จะประมวลผลได้ ผู้วิจัยจึงได้ทำการศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เพียง 2 วิธี ได้แก่ วิธี ML และวิธี ELR

จากการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 2 ตัวแปร เมื่อกำหนดค่าประมาณร้อยละของค่าตอบสนองคือ $10, 20, 30, 40$ และ 50 และกำหนดขนาดตัวอย่างมีค่าเท่ากับ $12, 20, 28, 48, 100$ และ 200 โดยมีผลการศึกษาดังนี้

4.2.1 การประมาณค่าพารามิเตอร์แบบจุด

1.) ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ตารางที่ 32 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ค่าประมาณร้อยละของค่าตอบสนอง	n = 12		n = 20		n = 28	
	ML	ELR	ML	ELR	ML	ELR
30	1947.5150	13.3652	605.7738	16.6492	121.8049	13.7836
40	1445.5450	11.2761	512.0542	11.0974	44.7516	9.6420
50	386.6863	**1.4490**	107.1283	**1.2551**	36.7074	**1.1534**

หมายเหตุ: ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ตัวพิมพ์หนา หมายถึง ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1)$ ที่มีค่าต่ำที่สุด

จากตารางที่ 32 พบว่าในทุกกรณีค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ ที่ได้จากทั้ง 2 วิธี มีค่ามาก โดยค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ จากวิธี ML มีค่าสูงกว่าวิธี ELR แต่ที่ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 50 วิธี ELR จะให้ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ น้อย เนื่องจากค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ที่ได้จาก วิธี ML มีค่าสูงมากโดยเฉพาะกรณีที่มีค่าประมาณร้อยละของค่าตอบสนองต่ำ และขนาดตัวอย่างน้อย เพราะในกรณีดังกล่าวมีค่าสังเกตของตัวแปรตอบสนองอยู่ในกลุ่มที่สนใจ (Y=1) ต่ำ ทำให้ค่าประมาณที่ได้มีค่า

แตกต่างจากค่าพารามิเตอร์มาก จึงส่งผลให้ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ มีค่าสูง นอกจากนี้พบว่าค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ของทั้ง 2 วิธีมีแนวโน้มลดลงเมื่อค่าประมาณร้อยละของค่าตอบสนองสูงขึ้น และมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

ตารางที่ 33 ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ค่าประมาณร้อยละของค่าตอบสนอง	n = 48		n = 100		n = 200	
	ML	ELR	ML	ELR	ML	ELR
10	360.2781	6.2185	37.0767	3.6864	2.0562	2.0160
20	64.1666	47.3211	30.9454	28.2165	13.8188	13.5442
30	12.0317	6.5083	4.5733	4.3842	1.4612	1.4304
40	6.1895	5.6508	2.5846	2.4737	1.1337	1.1089
50	10.6027	0.7883	0.3533	0.3323	0.1709	0.1650

จากตารางที่ 33 พบว่าค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ ที่ได้จากทั้ง 2 วิธี มีค่ามาก โดยค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ จากวิธี ML มีค่าสูงกว่าวิธี ELR แต่ที่ขนาดตัวอย่างเท่ากับ 48 ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 50 วิธี ELR จะให้ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ น้อย ขนาดตัวอย่างเท่ากับ 100 ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 50 ทั้ง 2 วิธี จะให้ค่า $\widehat{\mathrm{MSE}}$ น้อย และที่ขนาดตัวอย่างเท่ากับ 200 ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 30, 40 และ 50 ทั้ง 2 วิธี จะให้ค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ น้อย และพบว่าค่า $\widehat{\mathrm{MSE}}(\hat{\beta}_1, \hat{\beta}_2)$ ของทั้ง 2 วิธีมีแนวโน้มลดลงเมื่อค่าประมาณร้อยละของค่าตอบสนองสูงขึ้น และมีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น

4.2.2 การประมาณค่าพารามิเตอร์แบบช่วง

1.) ค่าประมาณค่าความน่าจะเป็นครอบคลุม

ตารางที่ 34 ค่าประมาณค่าความน่าจะเป็นครอบคลุมระดับความเชื่อมั่น 95%

ค่าประมาณร้อยละของค่าตอบสนอง	n = 12		n = 20		n = 28	
	ML	ELR	ML	ELR	ML	ELR
30	0.9800*	0.9750*	0.9750*	0.9350*	0.9500*	0.9700*
40	0.9800*	0.9750*	0.9700*	0.9600*	0.9350*	0.9650*
50	0.9800*	0.9550*	0.9650*	0.9550*	0.9500*	0.9400*

หมายเหตุ: * หมายถึง กรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 34 พบว่า ที่ระดับความเชื่อมั่น 95% ทั้ง 2 วิธี จะให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

ตารางที่ 35 ค่าประมาณค่าความน่าจะเป็นครอบคลุมที่ระดับความเชื่อมั่น 95%

ค่าประมาณร้อยละของค่าตอบสนอง	n = 48		n = 100		n = 200	
	ML	ELR	ML	ELR	ML	ELR
10	0.9950	0.9650*	0.9600*	0.9550*	0.9300*	0.9600*
20	0.9400*	0.9750*	0.9050	0.9450*	0.8950	0.9400*
30	0.9400*	0.9600*	0.9100	0.9200*	0.9450*	0.9550*
40	0.9100	0.9600*	0.9050	0.9500*	0.9150	0.9500*
50	0.9250*	0.9400*	0.9250*	0.9500*	0.8850	0.9200*

หมายเหตุ: * หมายถึง กรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด 95% ภายใต้ระดับนัยสำคัญ 0.05

จากตารางที่ 35 พบว่า ที่ระดับความเชื่อมั่น 95% และขนาดตัวอย่างเท่ากับ 48 วิธี ML ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด แต่ที่ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 10 และ 40 และวิธี ELR ให้ค่า $\widehat{\mathrm{CP}}$ แตกต่างจากระดับความเชื่อมั่นที่กำหนดสำหรับทุกค่าประมาณร้อยละของค่าตอบสนอง ที่ขนาดตัวอย่างเท่ากับ 100 วิธี ML ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 10 และ 50 และวิธี ELR ให้ค่า $\widehat{\mathrm{CP}}$

ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนดสำหรับทุกค่าประมาณร้อยละของค่าตอบสนอง ที่ขนาดตัวอย่างเท่ากับ 200 วิธี ML ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 10 และ 30 และวิธี ELR ให้ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนดสำหรับทุกค่าประมาณร้อยละของค่าตอบสนอง

2.) ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น

ในการเปรียบเทียบค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่น ($\widehat{\mathrm{AW}}$) จะพิจารณาเฉพาะกรณีที่ให้ค่าประมาณค่าความน่าจะเป็นครอบคลุม ($\widehat{\mathrm{CP}}$) ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

ตารางที่ 36 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่ระดับความเชื่อมั่น 95%

ค่าประมาณร้อยละของค่าตอบสนอง	n = 12		n = 20		n = 28	
	ML	ELR	ML	ELR	ML	ELR
30	142560.9000*	∞*	34596.9600*	∞*	5806.8100*	∞*
40	148186.1000*	∞*	30878.8400*	∞*	2049.0800*	∞*
50	37207.8700*	∞*	5122.9000*	∞*	1374.9000*	∞*

หมายเหตุ: * หมายถึง ค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 36 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% วิธี ML ให้ค่า $\widehat{\mathrm{AW}}$ น้อยกว่าวิธี ELR โดยค่า $\widehat{\mathrm{AW}}$ ที่ได้จากวิธี ML จะลดลงเมื่อค่าประมาณร้อยละของค่าตอบสนอง และขนาดตัวอย่างเพิ่มขึ้น

ตารางที่ 37 ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นที่ระดับความเชื่อมั่น 95%

ค่าประมาณร้อยละของค่าตอบสนอง	n = 48		n = 100		n = 200	
	ML	ELR	ML	ELR	ML	ELR
10	18496.7700	∞*	1190.2940*	∞*	14.5222*	16.4097*
20	773.0046*	∞*	259.3449	∞*	144.0891	156.6768*
30	179.4559*	∞*	28.4921	31.6844*	19.7702*	21.3510*
40	31.6112	36.4537*	21.2345	23.4734*	14.7976	15.9206*
50	285.1853*	∞*	2.7393*	3.0438*	1.9051	2.0548*

หมายเหตุ: * หมายถึง ค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด

จากตารางที่ 37 พบว่า เมื่อพิจารณาเฉพาะค่า $\widehat{\mathrm{AW}}$ ในกรณีที่ค่า $\widehat{\mathrm{CP}}$ ไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนด ที่ระดับความเชื่อมั่น 95% ขนาดตัวอย่างเท่ากับ 48 วิธี ML ให้ค่า $\widehat{\mathrm{AW}}$ น้อยกว่าวิธี ELR ยกเว้นที่ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 10 และ 40 ที่ขนาดตัวอย่างเท่ากับ 100 วิธี ML ให้ค่า $\widehat{\mathrm{AW}}$ น้อยกว่าวิธี ELR ที่ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 10 และ 50 ที่ขนาดตัวอย่างเท่ากับ 200 วิธี ML ให้ค่า $\widehat{\mathrm{AW}}$ น้อยกว่าวิธี ELR ที่ค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 10 และ 30 โดยค่า $\widehat{\mathrm{AW}}$ ที่ได้จากวิธี ML มีแนวโน้มลดลงเมื่อค่าประมาณร้อยละของค่าตอบสนอง และขนาดตัวอย่างเพิ่มขึ้น และจากตารางที่ 35 พบว่ากรณีที่ให้ค่า $\widehat{\mathrm{CP}}$ มาก ทั้ง 2 วิธี จะให้ค่า $\widehat{\mathrm{AW}}$ สูง

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติก 3 วิธี ได้แก่ วิธีภาวะน่าจะเป็นสูงสุด วิธีการถดถอยลอจิสติกแบบแม่นตรง และวิธีลูกโซ่มาร์คอฟมอนติคาร์โลสำหรับการถดถอยลอจิสติกแบบแม่นตรง โดยเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบจุด และการประมาณค่าพารามิเตอร์แบบช่วง และแบ่งเป็น 2 กรณี คือ เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 1 ตัวแปร และเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 2 ตัวแปร

จากผลการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกทั้ง 3 วิธี โดยใช้ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสำหรับการประมาณค่าพารามิเตอร์แบบจุด ค่าประมาณค่าความน่าจะเป็นครอบคลุมและค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นสำหรับการประมาณค่าพารามิเตอร์แบบช่วง สามารถสรุปผลการวิเคราะห์ได้ ดังนี้

กรณีที่ 1 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 1 ตัวแปร สรุปแยกตามเกณฑ์การพิจารณาได้ดังนี้

1.) การประมาณค่าพารามิเตอร์แบบจุด

กรณีขนาดตัวอย่างเล็ก $(n \leq 30)$: ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 10 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 10) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 20 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 5) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 30 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 3.33) ที่ทุกค่าร้อยละของค่าตอบสนอง ทั้ง 3 วิธี ยังมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ต่ำ แต่ที่จำนวนค่าสังเกตทับซ้อน และขนาดตัวอย่างอื่นนอกเหนือจากที่กล่าวมาข้างต้นแล้ว ที่ทุกค่าร้อยละของค่าตอบสนอง วิธี ML และวิธี ELR มีประสิทธิภาพใกล้เคียงกัน วิธี MCMC มีประสิทธิภาพต่ำสุดสำหรับการประมาณค่าพารามิเตอร์ทั้ง 3 วิธี และเหมาะสมกับข้อมูลที่มีร้อยละของค่าตอบสนองมากกว่าหรือเท่ากับ 30

กรณีขนาดตัวอย่างใหญ่ $(n \geq 50)$: เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมีค่าต่ำกว่าหรือเท่ากับ 4 ที่ทุกค่าร้อยละของค่าตอบสนอง วิธี ML มีประสิทธิภาพต่ำกว่าวิธี MCMC แต่เมื่อร้อยละของจำนวนค่าสังเกตทับซ้อนมีค่าสูงกว่า 4 ที่ทุกค่าร้อยละของค่าตอบสนอง วิธี ML มีประสิทธิภาพสูงกว่าวิธี MCMC

2.) การประมาณค่าพารามิเตอร์แบบช่วง

เมื่อพิจารณาวิธีที่ให้ค่าประมาณค่าความน่าจะเป็นครอบคลุมไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนดและให้ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุด คือวิธี ML เนื่องจากวิธี ELR และวิธี MCMC โดยส่วนใหญ่ให้ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นเป็นอนันต์ โดยวิธี ML เหมาะสมกับข้อมูลที่มีร้อยละของค่าตอบสนองมากกว่าหรือเท่ากับ 40

กรณีที่ 2 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 2 ตัวแปร สามารถสรุปแยกตามเกณฑ์การพิจารณาได้ดังนี้

1.) การประมาณค่าพารามิเตอร์แบบจุด

กรณีขนาดตัวอย่างเล็ก $(n \leq 28)$ และค่าประมาณร้อยละของค่าตอบสนองต่ำกว่า 50 ทั้ง 2 วิธี (ML และ ELR) ยังมีประสิทธิภาพต่ำในการประมาณค่าพารามิเตอร์ แต่เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 50 วิธี ELR มีประสิทธิภาพในการประมาณค่าพารามิเตอร์แบบจุดดีกว่าวิธี ML

กรณีขนาดตัวอย่างใหญ่ $(n \geq 48)$ ที่ค่าประมาณร้อยละของค่าตอบสนองต่ำกว่า 50 ทั้ง 2 วิธี (ML และ ELR) ยังมีประสิทธิภาพต่ำในการประมาณค่าพารามิเตอร์ แต่เมื่อค่าประมาณร้อยละของค่าตอบสนองเท่ากับ 50 ทั้ง 2 วิธี มีประสิทธิภาพใกล้เคียงกัน ยกเว้นที่ขนาดตัวอย่างเท่ากับ 48 วิธี ML มีประสิทธิต่ำกว่าวิธี ELR

2.) การประมาณค่าพารามิเตอร์แบบช่วง

เมื่อพิจารณาวิธีที่ให้ค่าประมาณค่าความน่าจะเป็นครอบคลุมไม่แตกต่างจากระดับความเชื่อมั่นที่กำหนดและให้ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นน้อยที่สุด คือวิธี ML เนื่องจากวิธี ELR โดยส่วนใหญ่ให้ค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นเป็นอนันต์

จากผลการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกทั้ง 3 วิธี โดยใช้ค่าประมาณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสำหรับการประมาณค่าพารามิเตอร์แบบจุด ค่าประมาณค่าความน่าจะเป็นครอบคลุมและค่าประมาณความกว้างเฉลี่ยของช่วงความเชื่อมั่นสำหรับการประมาณค่าพารามิเตอร์แบบช่วง นอกจากนี้ผู้วิจัยได้ทดลองทำการศึกษา

ในเบื้องต้นว่าประเภทของตัวแปรอธิบาย (ตัวแปรแบบต่อเนื่องและไม่ต่อเนื่อง) จะส่งผลต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ของทั้ง 3 วิธี แตกต่างกันอย่างไร โดยสามารถสรุปผลการวิเคราะห์ได้ ดังนี้

ตารางที่ 38 สรุปผลการวิจัยกรณีตัวแปรอธิบาย 1 ตัวแปร

			ML	ELR	MCMC
ร้อยละของจำนวนค่าสังเกตทับซ้อนต่ำกว่าหรือเท่ากับ 4	$n \le 30$	**การประมาณค่าแบบจุด**	ประสิทธิภาพต่ำ	ประสิทธิภาพต่ำ	ประสิทธิภาพต่ำ
		ร้อยละของค่าตอบสนอง	ประสิทธิภาพต่ำที่ทุกค่าร้อยละค่าตอบสนอง	ประสิทธิภาพต่ำที่ทุกค่าร้อยละค่าตอบสนอง	ประสิทธิภาพต่ำที่ทุกค่าร้อยละค่าตอบสนอง
	$n \ge 50$	**การประมาณค่าแบบจุด**	ประสิทธิภาพต่ำกว่า MCMC	-	ประสิทธิภาพสูงกว่า ML
		ร้อยละของค่าตอบสนอง	ประสิทธิภาพต่ำที่ทุกค่าร้อยละค่าตอบสนอง	-	ประสิทธิภาพสูงที่ทุกค่าร้อยละค่าตอบสนอง
ร้อยละของจำนวนค่าสังเกตทับซ้อนมากกว่า 4	$n \le 30$	**การประมาณค่าแบบจุด**	ประสิทธิภาพต่ำ	ประสิทธิภาพต่ำ	ประสิทธิภาพต่ำ
		ร้อยละของค่าตอบสนอง	ประสิทธิภาพต่ำที่ทุกค่าร้อยละค่าตอบสนอง	ประสิทธิภาพต่ำที่ทุกค่าร้อยละค่าตอบสนอง	ประสิทธิภาพต่ำที่ทุกค่าร้อยละค่าตอบสนอง
	$n \ge 50$	**การประมาณค่าแบบจุด**	ประสิทธิภาพสูงกว่า MCMC	-	ประสิทธิภาพต่ำกว่า ML
		ร้อยละของค่าตอบสนอง	ประสิทธิภาพสูงกว่า MCMC ที่ทุกค่าร้อยละค่าตอบสนอง	-	ประสิทธิภาพต่ำกว่า ML ที่ทุกค่าร้อยละค่าตอบสนอง
การประมาณค่าแบบช่วง			ประสิทธิภาพสูงที่สุด	ประสิทธิภาพต่ำ	ประสิทธิภาพต่ำ
ตัวแปรอธิบายแบบต่อเนื่อง (กรณีศึกษาเบื้องต้น)			ใช้เวลาประมาณ 30 วินาที/กรณี	สำหรับตัวอย่างขนาดใหญ่ (n>30)ไม่สามารถประมวลผลได้	ใช้เวลาประมาณ 60 นาที/กรณีสำหรับตัวอย่างขนาดใหญ่
ตัวแปรอธิบายแบบไม่ต่อเนื่อง (กรณีศึกษาเบื้องต้น)			ใช้เวลาประมาณ 0.5 วินาที/กรณี	ใช้เวลาประมาณ 1 นาที/กรณี	สำหรับตัวอย่างขนาดใหญ่ (n>30)ไม่สามารถประมวลผลได้

ตารางที่ 39 สรุปผลการวิจัยกรณีตัวแปรอธิบาย 2 ตัวแปร

			ML	ELR	MCMC
การประมาณค่าแบบจุด	**ร้อยละของค่าตอบสนองต่ำกว่า 50**	$n \leq 28$	ประสิทธิภาพต่ำ	ประสิทธิภาพต่ำ	-
		$n \geq 48$	ประสิทธิภาพต่ำ	ประสิทธิภาพต่ำ	-
	ร้อยละของค่าตอบสนองเท่ากับ 50	$n \leq 28$	ประสิทธิภาพต่ำกว่า ELR	ประสิทธิภาพสูงกว่า ML	-
		$n \geq 48$	ประสิทธิภาพใกล้เคียงกับ ELR	ประสิทธิภาพใกล้เคียงกับ ML	-
การประมาณค่าแบบช่วง			ประสิทธิภาพสูง	ประสิทธิภาพต่ำ	-
ตัวแปรอธิบายแบบต่อเนื่อง 2 ตัว (กรณีศึกษาเบื้องต้น)			ใช้เวลาน้อยที่สุด	ไม่สามารถประมวลผลได้	ไม่สามารถประมวลผลได้
ตัวแปรอธิบายแบบไม่ต่อเนื่อง 2ตัว (กรณีศึกษาเบื้องต้น)			ใช้เวลาประมาณ 0.5 วินาที/กรณี	ใช้เวลาประมาณ 0.5 วินาที/กรณี	ไม่สามารถประมวลผลได้
ตัวแปรอธิบายแบบต่อเนื่อง 1 ตัว และตัวแปรอธิบายแบบไม่ต่อเนื่อง 1 ตัว (กรณีศึกษาเบื้องต้น)			ใช้เวลาประมาณ 0.5 วินาที/กรณี	ไม่สามารถประมวลผลได้	ไม่สามารถประมวลผลได้

5.2 อภิปรายผล

จากผลการเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกทั้ง 3 โดยพิจารณาจากค่าประมาณแบบจุด กรณีตัวแปรอธิบาย 1 ตัว พบว่า หากข้อมูลที่นำมาวิเคราะห์การถดถอยลอจิสติกมีจำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 10 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 10) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 20 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 5) ที่จำนวนค่าสังเกตทับซ้อนเท่ากับ 1 และขนาดตัวอย่างเท่ากับ 30 (ร้อยละของจำนวนค่าสังเกตทับซ้อนเท่ากับ 3.33) ทั้ง 3 วิธียังไม่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ ดังนั้นจึงอาจจะศึกษาวิธีการประมาณค่าพารามิเตอร์วิธีอื่น เช่น วิธี Firth’s method หรือ วิธี Hidden Logistic Regression เป็นต้น ถ้าขนาดตัวอย่างใหญ่ $(n \geq 30)$ วิธี ELR ไม่สามารถประมาณค่าพารามิเตอร์ได้เนื่องจากหน่วยความจำไม่เพียงพอ ดังนั้น วิธี MCMC จึงเป็นวิธีที่แนะนำ แต่ยังพบว่า วิธี MCMC ใช้เวลานานในการประมาณค่าพารามิเตอร์

กรณีตัวแปรอธิบาย 2 ตัว พบว่า วิธี ML และ ELR จะมีประสิทธิภาพเฉพาะกรณีขนาดตัวอย่างใหญ่ และค่าประมาณร้อยละของค่าตอบสนองมากกว่า 40 ดังนั้นหากตัวอย่างมีขนาดเล็กและค่าประมาณร้อยละของค่าตอบสนองต่ำกว่าหรือเท่ากับ 40 ควรเลือกใช้วิธีการประมาณค่าพารามิเตอร์วิธีอื่นที่นอกเหนือจากวิธีML และ ELR

เมื่อพิจารณาจากการประมาณค่าแบบช่วง สำหรับกรณีตัวแปรอธิบาย 1 ตัว และ 2 ตัว พบว่าวิธี ELR และวิธี MCMC ให้ค่าประมาณของความกว้างเฉลี่ยของช่วงความเชื่อมั่นมาก จึงถือว่าเป็นวิธีการที่ไม่เหมาะสมกับการประมาณค่าแบบช่วง

5.3 ข้อเสนอแนะ

1.) จากการศึกษาวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกทั้ง 3 วิธี พบว่าก่อนการประมาณค่าพารามิเตอร์ ควรตระหนักถึงลักษณะของข้อมูลก่อน และเลือกใช้วิธีการประมาณค่าพารามิเตอร์ที่เหมาะสมกับลักษณะของข้อมูล

2.) เนื่องจากการประมาณค่าพารามิเตอร์โดยวิธี ELR และ MCMC ใช้เวลาในการประมาณค่าพารามิเตอร์นาน ดังนั้นจึงแนะนำให้ใช้เครื่องคอมพิวเตอร์ที่มีหน่วยความจำขนาดใหญ่ เพื่อลดเวลาในการประมาณค่าพารามิเตอร์

3.) แม้ว่าวิธี MCMC จะสามารถแก้ปัญหาในกรณีที่วิธี ELR ไม่สามารถประมาณค่าพารามิเตอร์ได้ แต่ยังใช้เวลาในการประมาณค่าพารามิเตอร์นาน ผู้วิจัยจึงอยากเสนอให้มีการศึกษาถึงวิธีการประมาณค่าพารามิเตอร์ในตัวแบบการถดถอยลอจิสติกวิธีอื่น ที่สามารถลดเวลาในการประมาณค่าพารามิเตอร์ลงได้

รายการอ้างอิง

1] วีรานันท์ พงศาภักดี. (2555). **การวิเคราะห์ข้อมูลจำแนกประเภท: ทฤษฎีและการประยุกต์ด้วย GLIM, SPSS, SAS และ MTB (Analysis of Categorical Data: Theories and Application with GLIM, SPSS, SAS and MTB). 3.** นครปฐม: โรงพิมพ์มหาวิทยาลัยศิลปากร.

[2] Botes, M. (2013). "Comparing logistic regression methods for completely separated and quasi-separated data." Master of Science Mathematical Statistics. Faculty of Natural & Agricultural Sciences. University of Pretoria: Pretoria.

[3] King, E.N. and T.P. Ryan. (2002). "A preliminary investigation of maximum likelihood logistic regression versus exact logistic regression." **The American Statistician**, 56(3): 163-170.

[4] Mehta, C.R., N.R. Patel, and P. Senchaudhuri. (2000). "Efficient Monte Carlo Methods for Conditional Logistic Regression." **The American Statistical Association**, 95(449): 99–108.

[5] เสกสรร เกียรติสุไพบูลย์. (2555). **การจำลอง: Simulation.** กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์.

[6] กันยาพร หาญกล้า. (2556). "การเพิ่มกำลังการทดสอบภาวะสารูปดีในตัวแบบถดถอยลอจิสติกโดยใช้การวิเคราะห์องค์ประกอบหลัก." ปริญญาวิทยาศาสตร์มหาบัณฑิต. สาขาสถิติประยุกต์ ภาควิชาสถิติ คณะวิทยาศาสตร์. มหาวิทยาลัยศิลปากร.

[7] Mehta, C.R. and N.R. Patel. (1995). "Exact logistic regression: theory and examples." **Stat Med**, 14(19): 2143-2160.

[8] Forster, J.J., J.W. McDonald, and P.W.F. Smith. (2003). "Markov chain Monte Carlo exact inference for binomial and multinomial logistic regression models." **Statistics and Computing**, 13(2): 169-177.

[9] Zamar, D. (2006). "Monte Carlo Markov Chain Exact Inference for Binomial Regression Models." Master of Science. Faculty of Computer Science University of British Columbia.

[10] Zamar, D., B. McNeney, and J. Graham. (2007). "elrm: Software implementing exact-like inference for logistic regression models." **Statistical Software**, 21(3): 1-18.

ภาคผนวก

โปรแกรมคอมพิวเตอร์ที่ใช้ในงานวิจัย

กรณีที่ 1 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 1 ตัวแปร และเป็นตัวแปรอธิบายแบบต่อเนื่อง

วิธีภาวะน่าจะเป็นสูงสุด (ML) และวิธีลูกโซ่มาร์คอฟมอนติคาร์โล (MCMC) สำหรับการถดถอยลอจิสติกแบบแม่นตรง

```
rm(list= ls())
set.seed(110508)
N = 10
beta0 = -18.49
beta1 =  0.88
PercentNonResponse = 90
reps= 200
ML=NULL
MCMC=NULL
e=NULL

for (i in 1:reps){

xx = runif(N, min = 6, max = 25)       # some continuous
variables
X = sort(xx)
z = beta0 + beta1*X                    # linear combination
with a bias
pi = 1/(1+exp(-z))
M = quantile(pi,((PercentNonResponse)/100))

d=NULL
for (n in 1:N){
yy = if (pi[n] <= M)0                  # Generate y
     else 1

yy[n==((N*(PercentNonResponse/100))+1)]=0   # overlap1
yy[n==((N*(PercentNonResponse/100))-0)]=1   # overlap1

d = rbind(d,c(yy))
             }
d = data.frame(d)
Y = d[,1]
data = data.frame(X, Y, ntrials = rep(1,N) )

#ML
output1 <- glm(Y ~ X, family = binomial)
#summary(output)
b01 = coef(output1)[1]
b11 = coef(output1)[2]
#ConfidenceInterval
```

```
CI = confint.default(output1)
LCIB11 = CI[2,1]
UCIB11 = CI[2,2]

#PC
pihat = predict(output1, type="response")
yhat = as.numeric(pihat > 0.5)
y = as.numeric(Y > 0)
ClassiTable = table(y, yhat)
addmargins(table(y, yhat), 2)
prop.table(ClassiTable, 1)
PC = sum(diag(ClassiTable))/sum(ClassiTable)
ML = rbind(ML, c(i,b01,b11,LCIB11,UCIB11,PC))

#MCMC
output2 = elrm(Y/ntrials ~ X , interest= ~X, iter=1000,
burnIn=0, dataset= data)
#summary(output2)
b12 = output2$coeffs
#ConfidenceInterval
LCIB12 = output2$coeffs.ci[1]
UCIB12 = output2$coeffs.ci[2]
MCMC = rbind(MCMC, c(i,b12,LCIB12,UCIB12))

e = rbind(e,cbind(i,X,Y))

                    }
colnames(ML) = c("iteration","B0","B1","LCI","UCI","PC")
ML = data.frame(ML)
colnames(MCMC) = c("iteration2","B12","LCI2","UCI2")
MCMC = data.frame(MCMC)
e = data.frame(e)

write.table(e,"D:/Output/1_10_50_8.txt")


#MSE
MSE1  = (sum((ML$B1-beta1)^2))/reps
MSE1
MSE2 = (sum((as.numeric(MCMC$B12)-beta1)^2))/reps
MSE2

#CP
Cover1  = (ML$LCI <= beta1) & (beta1 <= ML$UCI)
CP1  = table(Cover1)/reps
CP1
Cover2  = ((MCMC$LCI2 <= beta1)|(MCMC$LCI2 == (-Inf))) &
((beta1 <= MCMC$UCI2)|(MCMC$UCI2 == Inf))
CP2   = table(Cover2)/reps
CP2
```

```
#AW
AW1 = (sum(ML$UCI - ML$LCI))/reps
AW1
AW2 = (sum(as.numeric(MCMC$UCI2) -
as.numeric(MCMC$LCI2)))/reps
AW2

#APC
APC = sum(ML$PC)/reps
APC

mydata = read.table("D:/Output/1_10_50_8.txt")
```

วิธีการถดถอยลอจิสติกแบบแม่นตรง (ELR)

```
PROC IMPORT OUT=WORK.sample
            DATAFILE="D:/Output/Degree8/8_20_50.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="Sheet1$";
run;
proc print data=WORK.sample;
run;

proc logistic data = WORK.sample descending; by i;
  model Y = X;
  exact  X / estimate = both;
  output out=pred p=phat predprob=(individual crossvalidate) ;
  ods output ExactParmEst = est ;
run;
quit;
data a;
set est;
Diff = (Estimate - 0.88)**2 ;
if LowerCL <= 0.88 & UpperCL >= 0.88
then CP=1;
else CP=0;
w = UpperCL -  LowerCL;
run;
proc print data = a;
run;
proc means data= a;
var  Diff CP W;
run;
```

กรณีที่ 2 เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์กรณีตัวแปรอธิบาย 2 ตัวแปร และเป็นตัวแปรอธิบายแบบไม่ต่อเนื่อง

วิธีภาวะน่าจะเป็นสูงสุด (ML)

```
set.seed(20010)

N = 10
beta0 = -2.53
beta1 =  0.45
beta2 = 0.19
reps = 200

ML = NULL
e = NULL
for (i in 1:reps){

x1 <- c(rep("0",0.5*N),rep("1",0.5*N))
X1 = as.numeric(x1)
x2 <-
c(rep("0",0.25*N),rep("1",0.25*N),rep("0",0.25*N),rep("1",0.25
*N))
X2 = as.numeric(x2)
z = beta0 + beta1*X1 + beta2*X2  # linear combination with a
bias
pi = 1/(1+exp(-z))
PercentResponse = mean(pi)
Y  = rbinom(N,1,pi)              # bernoulli response variable

data = data.frame(X1, X2, Y,ntrials = rep(1,N) )

#ML
output1 <- glm(Y ~ X1+X2, family = binomial)
#summary(output1)
b01 = coef(output1)[1]
b11 = coef(output1)[2]
b21 = coef(output1)[3]

#ConfidenceInterval
CI = confint.default(output1)
LCIB11 = CI[2,1]
UCIB11 = CI[2,2]
LCIB21 = CI[3,1]
UCIB21 = CI[3,2]

#PC
pihat = predict(output1, type="response")
yhat = as.numeric(pihat > 0.5)
y = as.numeric(Y > 0)
```

```
ClassiTable = table(y, yhat)
addmargins(table(y, yhat), 2)
prop.table(ClassiTable, 1)
PC = sum(diag(ClassiTable))/sum(ClassiTable)

ML = rbind(ML,
c(i,b01,b11,b21,LCIB11,UCIB11,LCIB21,UCIB21,PC))

e = rbind(e,cbind(i,X1,X2,Y))

                }
colnames(ML) =
c("iteration","B0","B1","B2","LCIB1","UCIB1","LCIB2","UCIB2","
PC")
ML = data.frame(ML)

e = data.frame(e)

write.table(e,"D:/Output_2X_nLarge/48_10_2.txt")

PercentResponse

#MSE
MSE  = sum(  (((ML$B1-beta1)^2)/beta1) + (((ML$B2-
beta2)^2)/beta2) )/reps
MSE

#CP
Cover  = (ML$LCIB1 <= beta1) & (beta1 <= ML$UCIB1)&(ML$LCIB2
<= beta2) & (beta2 <= ML$UCIB2)
CP     = table(Cover)/reps
CP
Cover1  = (ML$LCIB1 <= beta1) & (beta1 <= ML$UCIB1)
CP1     = table(Cover1)/reps
CP1
Cover2  =(ML$LCIB2 <= beta2) & (beta2 <= ML$UCIB2)
CP2     = table(Cover2)/reps
CP2

#AW
AW = sum(((ML$UCIB1 - ML$LCIB1)/beta1) + ((ML$UCIB2 -
ML$LCIB2)/beta2))/reps
AW
AW1 = sum(ML$UCIB1 - ML$LCIB1)/reps
AW1
AW2 = sum(ML$UCIB2 - ML$LCIB2)/reps
AW2

#APC
APC = (sum(ML$PC)/reps)*100
APC

mydata = read.table("D:/Output_2X_nLarge/48_10_2.txt")
```

วิธีการถดถอยลอจิสติกแบบแม่นตรง (ELR)

```
PROC IMPORT OUT=WORK.sample
            DATAFILE="D:/Output_2X/12_30.xlsx"
            DBMS=EXCEL REPLACE;
      RANGE="Sheet1$";
run;

proc logistic data = WORK.sample descending; by i;
  model Y = X1 X2;
  exact  INTERCEPT X1 X2 / estimate = both;
  output out=pred p=phat predprob=(individual crossvalidate) ;
  ods output ExactParmEst = est ;
run;
quit;

data a;
set est;
beta1 = 0.45 ;
beta2 = 0.19 ;

if Parameter = 'Intercept'
then D = 0;
else if Parameter = 'X1'
then D =((Estimate - (beta1))**2)/beta1;
else D =((Estimate - (beta2))**2)/beta2;

if Parameter = 'Intercept'
then C=0;
if Parameter = 'X1' & ( ((LowerCL <= (beta1)) & (UpperCL >=
(beta1))) or ((LowerCL = .M) & (UpperCL >= (beta1))) or
((LowerCL <= (beta1)) & (UpperCL = .I)) )
then C=1;
if Parameter = 'X2' & ( ((LowerCL <= (beta2)) & (UpperCL >=
(beta2))) or ((LowerCL = .M) & (UpperCL >= (beta2))) or
((LowerCL <= (beta2)) & (UpperCL = .I)) )
then C=1;

if Parameter = 'X1' & ( ((LowerCL <= (beta1)) & (UpperCL >=
(beta1))) or ((LowerCL = .M) & (UpperCL >= (beta1))) or
((LowerCL <= (beta1)) & (UpperCL = .I)) )
then C1=1;
if Parameter = 'X2' & ( ((LowerCL <= (beta2)) & (UpperCL >=
(beta2))) or ((LowerCL = .M) & (UpperCL >= (beta2))) or
((LowerCL <= (beta2)) & (UpperCL = .I)) )
then C2=1;

if Parameter = 'Intercept'
then w = 0;
else if Parameter = 'X1'
then w = (UpperCL -  LowerCL)/beta1;
else w = (UpperCL -  LowerCL)/beta2;
```

```
if Parameter = 'X1'
then w1 = (UpperCL - LowerCL);
else if Parameter = 'X2'
then w2 = (UpperCL - LowerCL);

run;

proc means data =a SUM N ; by i;
var D C C1 C2 w w1 w2;
ods output summary = means_summary;
run;

data b;
set means_summary;
if C_Sum = 2
then CP = 1;
else CP =0;
if w_N ^= 3
then W = 'Infinity' ;
if w1_N ^= 1
then W1 = 'Infinity' ;
if w2_N ^= 1
then W2 = 'Infinity' ;
run;
proc print data = b;
run;
proc means data = b MEAN N ;
var D_Sum CP w_Sum w1_Sum w2_Sum;
run;
```

ประวัติผู้เขียน

ชื่อ-สกุล	จารุวรรณ เหมือนเงิน
วัน เดือน ปี เกิด	11 เมษายน 2536
สถานที่เกิด	สุพรรณบุรี ประเทศไทย
วุฒิการศึกษา	วท.บ. สถิติ
ที่อยู่ปัจจุบัน	105/3 หมู่ 4 ตำบล เดิมบาง อำเภอ เดิมบางนางบวช จังหวัด สุพรรณบุรี 72120
ผลงานตีพิมพ์	Muenngun, J. and N. Meejun. (2015). A Comparison of Difference in Coefficients Methods to Test Mediation. The 4th ICADA 2015-SSIS, Bangkok, Thailand. Muenngun, J. and P. Paichit. (2016). A Comparison of Methods to Test Mediation by Difference in Coefficients Test. ICAS2016, Phuket, Thailand.