การประมาณค่าพารามิเตอร์ในตัวแบบความถดถอยโลจิสติกเมื่อมีค่าสูญหาย

นายประลองพล ประสงค์พร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2551

Estimation of Parameters in Logistic Regression with Missing of Independent Variables

Mr. Pralongpol Prasongporn

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Statistics Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2008

หัวข้อวิทยานิพนธ์	การประมาณค่าพารามิเตอร์ในตัวแบบความถดถอยโลจิสติก เมื่อมีค่าสูญหาย
โดย	นายประลองพล ประสงค์พร
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

……………………………….คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร. อรรณพ ตันละมัย)

คณะกรรมการสอบวิทยานิพนธ์

…………………………………………… ประธานกรรมการ
(รองศาสตราจารย์ ดร. ธีระพร วีระถาวร)

…………………………………………... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา)

…………………………………………… กรรมการ
(รองศาสตราจารย์ผกาวดี ศิริรังษี)

…………………………………………… กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ชูศักดิ์ อุดมศรี)

ประลองพล ประสงค์พร : การประมาณค่าพารามิเตอร์ในตัวแบบความถดถอยโลจิสติกเมื่อมีค่าสูญหาย. (ESTIMATION OF PARAMETERS IN LOGISTIC REGRESSION WITH MISSING OF INDEPENDENT VARIABLES) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ.ดร.กัลยา วานิชย์บัญชา, 74 หน้า.

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระของตัวประมาณตัวแบบความถดถอยโลจิสติก ที่ประมาณค่าสูญหายด้วยวิธี Mean Imputation (MEAN) วิธี Maximum Likelihood Estimation (MLE) วิธี Pseudo Maximum Likelihood Estimation (PMLE) และวิธี The Filling Method (FILL) เมื่อมีค่าสูญหายเกิดขึ้นในกรณีที่มีตัวแปรอิสระ 2 ตัว และเกิดค่าสูญหายในตัวแปรอิสระตัวใดตัวหนึ่ง โดยการจำลองข้อมูลกำหนดขนาดตัวอย่าง 40, 70 , 90 , 100 , 200 และ 400 ร้อยละการสูญหายของตัวแปรอิสระคือร้อยละ 5 ,10 และ 15 และค่าสหสัมพันธ์ระหว่างตัวแปรอิสระคือ 0 ,0.1 และ 0.2 โดยกำหนดค่าพารามิเตอร์เริ่มต้นแบ่งเป็น 2 กรณี คือ $\beta_0, \beta_1, \beta_2$ = 0.2 และ β_0, β_1 = 0.2 และ β_2 = 1 พิจารณาเปรียบเทียบค่าเฉลี่ยของความแตกต่างระหว่างค่าจริงและค่าประมาณ และระยะห่างมาหาลาโนบิสเฉลี่ย (Average Mahalanobis Distance) เป็นเกณฑ์ประกอบในการตัดสินใจ ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองด้วยเทคนิคมอนติคาโลโดยการกระทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์ ผลการวิจัยสรุปได้ดังนี้

การเปรียบเทียบค่า BIAS และค่า AMH ของทั้ง 4 วิธีพบว่า ในกรณีขนาดตัวอย่าง น้อยกว่า 90 วิธี MEAN จะให้ค่า BIAS และค่า AMH น้อยที่สุด แต่ในกรณีขนาดตัวอย่าง มากกว่า 90 วิธี FILL จะให้ค่า BIAS และค่า AMH น้อยที่สุด โดยที่ค่า BIAS และค่า AMH มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น เนื่องจากจำนวนข้อมูลที่เพิ่มขึ้นจะทำให้ความคลาดเคลื่อนลดลง ค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้นเนื่องจากเมื่อร้อยละการสูญหายเพิ่มขึ้นจะส่งผลให้ข้อมูลที่มีอยู่ลดลง ทำให้ค่าความคลาดเคลื่อนเพิ่มขึ้น เมื่อค่าพารามิเตอร์เริ่มต้นของตัวแปรอิสระเพิ่มขึ้นค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น และเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้นค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น

ภาควิชา............... สถิติ................. ลายมือชื่อนิสิต..

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

ปีการศึกษา ....2551..................

##4982194526 : MAJOR STATISTICS
KEYWORDS : LOGIRTIC REGRESSION / MISSING DATA / MEAN IMPUTATION / MLE ESTIMATION / PMLE ESTIMATION / FILLING METHOD

PRALONGPOL PRASONGPORN : ESTIMATION OF PARAMETERS IN LOGISTIC REGRESSION WITH MISSING OF INDEPENDENT VARIABLES. ADVISOR : ASSOC. PROF. KANLAYA VANICHBANCHA , Ph.D., 74 pp.

This research is to study and compare estimation method for missing data of the independent variables in logistic regression. The methods used to estimate missing data are Mean Imputation (MEAN) , Maximum Likelihood Estimation (MLE) , Pseudo Maximum Likelihood Estimation (PMLE) and The Filling Method (FILL) in the case of two independent variables where only one independent variable is affected by missing values. The comparisons are done under condition of sample size of 40, 70 , 90 ,100 , 200 and 400 ; percentage of missing data of 5%,10% and 15% ; correlation in independent variables of 0 , 0.1 and 0.2 . Initial parameter at $\beta_0, \beta_1, \beta_2$ = 0.2 and β_0, β_1 = 0.2, β_2 = 1. The criteria of determination is the average difference between the estimates and the true parameter and Average Mahalanobis Distance (AMH). The data for this research is simulated by using the Monte Carlo simulation technique with 1,000 repetitions for each case. The results of this research are as follows :

According to the comparison of Bias and AMH from four referred methods , it is found that when sample size is less than 90, MEAN method has a smallest BIAS and AMH. In case sample size is more than 90 , FILL method has smallest BIAS and AMH. The BIAS and AMH decreases when sample size increases because more sample decreases error but BIAS and AMH increases when the proportions of missing data in independents variables increase because increasing missing will decrease sample size. In case of initial parameters of independent variables increases, BIAS and AMH increase. In case of the level of correlation among independent variables increases, BIAS and AMH increases.

Department : …. Statistics…… Student's Signature : ....
Field of Study : ...Statistics…… Advisor's Signature : ....
Academic Year : …2008……....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือ และเอาใจใส่เป็นอย่างดีจาก รองศาสตราจารย์ ดร. กัลยา วานิชย์บัญชา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ปรึกษาตลอดจนช่วยเหลือตรวจสอบ แก้ไขข้อบกพร่องต่างๆ เป็นอย่างดีจนกระทั่งวิทยานิพนธ์ เสร็จสมบูรณ์ ผู้วิจัยขอกราบขอบพระคุณไว้ ณ โอกาสนี้

ผู้วิจัยขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. ธีระพร วีระถาวร ประธานกรรมการ สอบวิทยานิพนธ์ และ รองศาสตราจารย์ผกาวดี ศิริรังษี กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้ คำแนะนำตรวจสอบแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น และขอกราบขอบคุณคณาจารย์ ทุกท่านที่ถ่ายทอดความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา ที่ให้การส่งเสริม สนับสนุน รวมทั้งให้ ความรักและให้กำลังใจแก่ผู้วิจัยจนสำเร็จการศึกษา รวมทั้งขอขอบคุณญาติพี่น้อง และเพื่อน ๆ ทุกท่านที่คอยเป็นที่ปรึกษาและให้กำลังใจผู้วิจัยมาโดยตลอด

สารบัญ

สารบัญตาราง

หน้า

ตารางที่

ตารางที่ หน้า

สารบัญภาพ

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในสังคมปัจจุบัน ข้อมูลเป็นสิ่งที่สำคัญมากสำหรับทุกสาขาอาชีพ ไม่ว่าจะด้านการเงิน การธนาคาร สังคมศาสตร์ การแพทย์ รวมถึงงานด้านวิจัยต่าง ๆ ที่จะต้องหาวิธีทางสถิติในหลากหลายรูปแบบ มาอธิบายและวิเคราะห์ข้อมูลเหล่านี้ เพื่อนำไปประยุกต์ใช้กับงานของตนให้ตรงกับวัตถุประสงค์ได้อย่างมีประสิทธิภาพ

ในการวิเคราะห์ข้อมูลส่วนมากนั้น ผู้วิเคราะห์จึงมักใช้วิธีการวิเคราะห์ความถดถอยเชิงเส้น เนื่องจากมีรูปแบบการวิเคราะห์ที่ง่าย ไม่ซับซ้อน ซึ่งมีความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ โดยที่ตัวแปรทั้งคู่เป็นตัวแปรเชิงปริมาณหรืออาจจะมีตัวแปรบางตัวเป็นตัวแปรเชิงปริมาณและตัวแปรบางตัวเป็นตัวแปรเชิงกลุ่ม แต่ก็มีงานวิจัยอีกไม่น้อย เช่นงานด้านการแพทย์หรือวิทยาศาสตร์ รวมถึงงานด้านธุรกิจต่าง ๆ ที่ผู้วิจัยมักจะต้องการตัวแปรตามที่เป็นตัวแปรเชิงคุณภาพ เช่นในด้านการแพทย์ ที่หมออาจต้องการแบ่งกลุ่มคนไข้ว่า คนไข้คนใดคนหนึ่งเป็นโรคหรือไม่ โดยดูจากข้อมูลเบื้องต้น ซึ่งจะทำให้ประหยัดเวลาในการรักษาขั้นต้นรวมถึงสามารถดูแลรักษาคนไข้ได้มีประสิทธิภาพยิ่งขึ้น หรือในงานด้านธุรกิจที่อาจต้องการจัดกลุ่มบริษัทว่าเป็นระดับย่อย หรือระดับกลาง เพื่อไปวิเคราะห์ส่วนแบ่งตลาดหรือการลงทุนต่าง ๆ ซึ่งการใช้วิธีการวิเคราะห์ความถดถอยเชิงเส้นนั้นอาจไม่สามารถตอบปัญหาเหล่านี้ได้ดีนัก

ดังนั้นในกรณีที่ต้องการให้ตัวแปรตามเป็นตัวแปรคุณภาพที่สามารถมีค่าได้ 2 ค่า เช่น เป็นโรคและไม่เป็นโรค หรือบริษัท A อยู่ในกลุ่มธุรกิจรูปแบบใดนั้น จึงควรใช้การวิเคราะห์ที่เป็นที่นิยมอีกวิธีหนึ่งคือ การวิเคราะห์ความถดถอยโลจิสติก (Logistic Regression Analysis) ซึ่งมีวัตถุประสงค์คือเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ คือดูว่าตัวแปรอิสระตัวใดที่มีผลกระทบต่อตัวแปรตามมากหรือน้อย เพื่อนำผลที่ได้ไปพยากรณ์โอกาสที่จะเกิดเหตุการณ์ที่สนใจ ทำให้สามารถแบ่งกลุ่มให้หน่วยตัวอย่างได้ต่อไป

แต่ในการวิเคราะห์ข้อมูลจริงนั้น อาจมีบางครั้งที่ข้อมูลตัวแปรอิสระสูญหายไปทำให้หน่วยตัวอย่างนั้นไม่สามารถนำไปจัดกลุ่มได้ ซึ่งเป็นปัญหาสำคัญสำหรับวงการธุรกิจและวงการแพทย์เป็นอย่างมาก ตัวอย่างเช่น การที่ผู้ตอบแบบสอบถามไม่ตอบคำถามที่เกี่ยวกับน้ำหนักตัว ซึ่งทำให้ผู้วิจัยไม่สามารถนำผลไปวิเคราะห์ข้อมูลได้ ดังนั้นในการงานวิจัยฉบับนี้จึงสนใจที่จะแก้ไขปัญหาดังกล่าว โดยการเปรียบเทียบวิธีประมาณค่าพารามิเตอร์วิธีการต่าง ๆ ในการจัดการกับข้อมูลใน

ตัวแปรอิสระที่หายไปนั้นให้สามารถนำไปประมาณค่าต่อไปได้อย่างมีประสิทธิภาพตามความเหมาะสมกับลักษณะของตัวแปรอิสระนั้นและเงื่อนไขต่าง ๆ ที่เกิดขึ้น

มีนักสถิติหลายท่านได้ทำการคิดวิธีประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในการวิเคราะห์ความถดถอยโลจิสติก เช่น

Wilks (1932) ได้เสนอวิธีการประมาณ ที่ไม่ซับซ้อนและเป็นวิธีที่เก่าแก่วิธีหนึ่งสำหรับตัวแปรอิสระที่เป็นแบบต่อเนื่อง เรียกว่า Mean Imputation โดยมีหลักการคือ ทำการประมาณค่าเฉลี่ยของข้อมูลที่ไม่ได้สูญหายแล้วนำมาประมาณให้กับตัวแปรที่ข้อมูลมีค่าสูญหาย และนำข้อมูลที่ได้นี้มาประมาณค่าพารามิเตอร์ต่อไป

Rubin (1976) ได้เสนอวิธี Maximum Likelihood Estimation (MLE) มาประยุกต์ใช้ในการประมาณค่าสูญหายกับสถิติพื้นฐานทั่วไป จนกระทั่ง Little & Schluchter (1985) นำวิธี ML มาประยุกต์ใช้ในวิธีการวิเคราะห์ความถดถอยโลจิสติกในกรณีที่มีค่าสูญหายเกิดขึ้น

Gong & Samaniego (1981) ได้เสนอการประมาณวิธีใหม่ขึ้นมา Pseudo Maximum Likelihood Estimation (PMLE) ซึ่งเป็นวิธีการประมาณที่ประยุกต์มาจากวิธี ML แต่มีข้อดีตรงที่วิธีนี้ไม่จำเป็นต้องทำภายใต้ MAR assumption (Missing at random) เท่านั้น เพราะ Pepe & Fleming (1991) สามารถประมาณค่าภายใต้ MDX assumption (Missing dependent on X) และ Vach & Schumacher (1992,1993) สามารถขยายไปทำใน MDY (Missing dependent on Y) และ MDXY (Missing dependent on X and Y) ในเวลาต่อมา

Vach & Blettner (1991) ได้เสนอวิธีที่ไม่ยุ่งยากในการประมาณค่าพารามิเตอร์ในกรณีที่มีค่าสูญหายเกิดขึ้น นั่นคือจะพยายามสร้างค่าจริงขึ้นมาจากค่าที่สูญหายในรูปแบบตาราง จากนั้นใช้วิธีการวิเคราะห์ความถดถอยโลจิสติกเข้ามาช่วยในการประมาณพารามิเตอร์ ซึ่งวิธีการนี้มีชื่อว่า The Filling Method (FILL)

ซึ่งจากการศึกษาวิธีการประมาณค่าพารามิเตอร์ข้างต้นนั้น ในแต่ละวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นนั้น ก็มีข้อดีข้อเสียแตกต่างกันไปตามสถานการณ์ต่าง ๆ ดังนั้นผู้วิจัยสนใจที่จะทำการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในการวิเคราะห์ความถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression) ในกรณีที่มีตัวแปรอิสระ 2 ตัว (Two Independent Variables) โดยการสูญหายเกิดขึ้นกับตัวแปรอิสระตัวที่ 1 หรือตัวแปรอิสระที่ 2 ตัวใดตัวหนึ่ง และการสูญหายเป็นไปอย่างสุ่ม (Missing at Random) (MAR) โดยใช้วิธีการประมาณค่าพารามิเตอร์ ดังต่อไปนี้

วิธีที่ 1 คือ วิธี Mean Imputation (MEAN)

วิธีที่ 2 คือ วิธี Maximum Likelihood Estimation (MLE)

วิธีที่ 3 คือ วิธี Pseudo Maximum Likelihood Estimation (PMLE)

วิธีที่ 4 คือ วิธี The Filling Method (FILL)

เพื่อทำการศึกษาว่าวิธีการใดมีประสิทธิภาพมากที่สุดในการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น ภายใต้สถานการณ์ที่แตกต่างกัน

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในการวิเคราะห์ความถดถอยโลจิสติก ด้วยวิธี MEAN วิธี MLE วิธี PMLE และวิธี FILL เมื่อมีค่าสูญหายเกิดขึ้นในกรณีที่มีตัวแปรอิสระ 2 ตัว ภายใต้สถานการณ์ที่แตกต่างกัน
2. เพื่อหาข้อเสนอแนะและวิธีการประมาณที่เหมาะสม ในการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในการวิเคราะห์ความถดถอยโลจิสติก ในกรณีที่มีตัวแปรอิสระ 2 ตัว ภายใต้สถานการณ์ที่แตกต่างกัน

1.3 ข้อตกลงเบื้องต้น

1. ตัวแปรที่สนใจศึกษา y และ x มีความสัมพันธ์ภายใต้ตัวแบบความถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression) คือ

$$P(Y=1)=\frac{e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}}}{1+e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}}} \qquad ;i=1,2,\dots,n$$

เมื่อ $P(Y=1)$ เป็นความน่าจะเป็นที่ Y=1 ณ ระดับ X_{i1} และ X_{i2}

X_{i1} และ X_{i2} เป็นตัวแปรอิสระตัวที่ 1 และ 2

$\beta_0, \beta_1, \beta_2$ เป็นพารามิเตอร์ที่ไม่ทราบค่า

n เป็นขนาดตัวอย่าง

2. ความคลาดเคลื่อนเป็นอิสระกัน
3. ตัวแปร X_1, X_2 เป็นตัวแปรเชิงปริมาณ
4. การสูญหายเกิดขึ้นกับตัวแปรอิสระตัวที่ 1 หรือตัวแปรอิสระที่ 2 ตัวใดตัวหนึ่ง โดยในที่นี้สมมติให้เกิดขึ้นในตัวแปร X_2 และการสูญหายเป็นการสูญหายแบบสุ่ม

1.4 ขอบเขตของการวิจัย

1. การสูญหายเกิดขึ้นกับตัวแปรอิสระตัวที่ 1 หรือตัวแปรอิสระที่ 2 ตัวใดตัวหนึ่ง โดยการสูญหายเป็นการสูญหายแบบสุ่ม
2. ลักษณะของตัวแปรอิสระที่นำมาศึกษามีรูปแบบการแจกแจงปกติ (Normal Distribution)

   ฟังก์ชันการแจกแจงคือ

   $$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

   ค่าคาดหวัง $E(X) = \mu$

   ความแปรปรวน $V(X) = \sigma^2$

3. กำหนดค่าพารามิเตอร์เริ่มต้น โดยแบ่งเป็น 2 กรณี คือ $\beta_0, \beta_1, \beta_2$ = 0.2 และ β_0, β_1 = 0.2 และ β_2 = 1
4. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 40 ,70 , 90 , 100 , 200 และ 400
5. ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ คือ 0 , 0.1 และ 0.2
6. ข้อมูลตัวแปรอิสระสูญหาย 3 ระดับ คือ ร้อยละ 5 , 10 และ 15
7. การวิจัยครั้งนี้ทำการจำลองข้อมูลให้มีสถานการณ์ที่แตกต่างกันตามข้อกำหนดข้างต้น โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte carlo Simulation Technique) ทำการจำลองในแต่ละสถานการณ์ 1,000 รอบ

1.5 เกณฑ์ที่ใช้ในการตัดสินใจ

เกณฑ์ที่ใช้ในการตัดสินใจว่าการประมาณค่าสูญหายด้วยวิธีใดใช้ได้ดีกว่าจากการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง Bias และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (Average Mahalanobis Distance) เป็นเกณฑ์ประกอบในการตัดสินใจ โดยมีวิธีการคำนวณคือ

Bias คำนวณจาก ค่าเฉลี่ยของความแตกต่างระหว่างค่าจริงและค่าประมาณ

Average Mahalanobis Distance คำนวณจาก

$$AMH = \sum_{k=1}^{1000} \frac{\sqrt{\left(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}\right)' \left(Cov\left(\hat{\underset{\sim}{\beta}}_i, \hat{\underset{\sim}{\beta}}_j\right)\right)^{-1} \left(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}\right)}}{1000}$$

เมื่อ $\underset{\sim}{\beta}_j$ หมายถึง ค่าจริงของพารามิเตอร์ในตัวแบบถดถอยโลจิสติกในการจำลองรอบที่ j

เมื่อ $\hat{\underset{\sim}{\beta}}_j$ หมายถึง ค่าประมาณของพารามิเตอร์ในตัวแบบถดถอยโลจิสติกในการจำลองรอบที่ j

1.6 วิธีดำเนินการวิจัย

1. ศึกษาโปรแกรม MATHLAB (Matrix Laboratory Version 7) ซึ่งเป็นโปรแกรมที่ผู้วิจัยใช้ในการศึกษา
2. สร้างโปรแกรมที่ใช้ในการสร้างข้อมูลที่ใช้ในการทดสอบ โดยใช้เทคนิควิธี การจำลองแบบมอนติคาร์โล (Monte carlo Simulation Technique)
3. ศึกษาวิธีการสร้างตัวประมาณที่ต้องการเปรียบเทียบ และเขียนโปรแกรมสร้างตัวประมาณทั้ง 4 วิธีคือ

   วิธีที่ 1 คือ วิธี Mean Imputation (MEAN)

   วิธีที่ 2 คือ วิธี Maximum Likelihood Estimation (MLE)

   วิธีที่ 3 คือ วิธี Pseudo Maximum Likelihood Estimation (PMLE)

   วิธีที่ 4 คือ วิธี The Filling Method (FILL)
4. เขียนโปรแกรมสำหรับใช้เป็นเกณฑ์ในการตัดสินใจ ได้แก่ Bias และ Average Mahalanobis Distance
5. วิเคราะห์และสรุปผลการเปรียบเทียบวิธีการทั้ง 4 วิธี

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการตัดสินใจเลือกวิธีการประมาณค่าพารามิเตอร์ในตัวแบบความถดถอยโลจิสติกเมื่อมีค่าสูญหายได้อย่างเหมาะสม
2. เพื่อทราบประสิทธิภาพของตัวประมาณในสถานการณ์ต่าง ๆ เช่น ขนาดตัวอย่างที่แตกต่างกัน ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระที่แตกต่างกัน , พารามิเตอร์เริ่มต้นและระดับการสูญหายที่แตกต่างกันเป็นต้น เพื่อสามารถนำไปประยุกต์ใช้กับข้อมูลจริงได้ต่อไป
3. เพื่อเป็นแนวทางวิเคราะห์วิธีประมาณพารามิเตอร์วิธีอื่น ๆ ต่อไป

บทที่ 2

ทฤษฎีและสถิติที่ใช้ในการวิจัย

ในกรณีที่เกิดปัญหาข้อมูลสูญหายขึ้นในตัวแปรอิสระ การประมาณค่าพารามิเตอร์ในตัวแบบนั้นมักจะตัดชุดข้อมูลที่สูญหายออกหรือประมาณค่าสูญหายด้วยวิธีการประมาณแบบต่าง ๆ ก่อน ซึ่งการตัดชุดข้อมูลที่สูญหายออกนั้นจะทำให้ค่าประมาณที่ได้อาจคลาดเคลื่อนจากความเป็นจริง ดังนั้นในบทนี้จะกล่าวถึงข้อสมมติของตัวแบบถดถอยโลจิสติก วิธีการประมาณค่าด้วยวิธี Mean Imputation , วิธี Maximum Likelihood Estimation , วิธี Pseudo Maximum Likelihood Estimation และวิธี The Filling Method เมื่อเกิดปัญหาข้อมูลสูญหายขึ้นในตัวแปรอิสระ ซึ่งมีรายละเอียดดังต่อไปนี้

2.1 ข้อสมมติทั่วไปของตัวแบบถดถอยโลจิสติก

ในการวิจัยครั้งนี้ข้อมูลที่นำมาพิจารณาอยู่ในรูป (x_i, y_i) ; i = 1,2 ,…, N ซึ่งตัวแปรตาม y_i มีการแจกแจงแบบแบร์นูลลี (Bernoulli Distribution) และ x_i เป็นตัวแปรอิสระร่วมกัน m ตัวแปร ซึ่งความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจคือ $\pi(x_i)$ นั่นคือ

$$Y_i = \begin{cases} 1 & \text{ด้วยความน่าจะเป็น} \quad \pi(x_i) \\ 0 & \text{ด้วยความน่าจะเป็น} \quad 1-\pi(x_i) \end{cases}$$

ดังนั้นจะทำการแปลงค่า $\pi(x_i)$ จากช่วง (0,1) เป็นค่าของ logit $(\pi(x_i))$ ที่อยู่ในช่วง $(-\infty,\infty)$ ซึ่งฟังก์ชันการแปลงโลจิสติกของ $\pi(x_i)$ คือ

$$\text{logit } (\pi(x_i)) = \ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}$$

$$\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}}$$

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_m x_{im}}}$$

ซึ่งสามารถเขียนสมการถดถอยโลจิสติกเชิงเส้น ในรูปเมตริกซ์คือ

H(X) = XB

เมื่อ $$H(X)=\begin{bmatrix}\log it(\pi(x_1))\\ \log it(\pi(x_2))\\ \vdots \\ \log it(\pi(x_N))\end{bmatrix}_{Nx1}$$

$$X = \begin{bmatrix}1 & x_{11} & x_{12}\cdots & x_{1m}\\ 1 & x_{21} & x_{22}\cdots & x_{2m}\\ \vdots & \vdots & \vdots \quad \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & x_{Nm}\end{bmatrix}_{N\,x\,(m+1)} \qquad B=\underset{\sim}{\beta}=\begin{bmatrix}\beta_0\\ \beta_1\\ \vdots\\ \beta_m\end{bmatrix}_{(m+1)\,x\,1}$$

ลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็นสูงสุดในสมการถดถอยโลจิสติก ฟังก์ชันความควรจะเป็นของ β คือ

$$\ell(\beta) = \prod_{i=1}^{N}(\pi(x_i))^{y_i}(1-\pi(x_i))^{1-y_i}$$

$$L(\beta) = \ln(\ell(\beta))$$

$$= \sum_{i=1}^{N}\{y_i\ln(\pi(x_i))+(1-y_i)\ln(1-\pi(x_i))\}$$

$$= \sum_{i=1}^{N}\{y_i\ln(\pi(x_i))+\ln(1-\pi(x_i))-y_i\ln(1-\pi(x_i))\}$$

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

$$= \sum_{i=1}^{N}\left\{y_i\ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right]+\ln(1-\pi(x_i))\right\}$$

$$= \sum_{i=1}^{N}\left\{y_i\ln\left[\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_mx_{im}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_mx_{im}}}\,x\,\frac{1}{1-\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_mx_{im}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_mx_{im}}}}\right]+\ln\left[1-\frac{e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_mx_{im}}}{1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_mx_{im}}}\right]\right\}$$

$$= \sum_{i=1}^{N}\left\{y_i(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_mx_{im})-\ln\left(1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\ldots+\beta_mx_{im}}\right)\right\}$$

2.2 การประมาณค่าพารามิเตอร์ด้วยวิธี Mean Imputation (MEAN)

วิธีนี้ใช้หลักการของค่าเฉลี่ย (Mean) เข้ามาช่วยในการแก้ปัญหาค่าสูญหาย โดยมีวิธีการ คือ การแทนค่าสูญหายด้วยค่าเฉลี่ยของข้อมูลที่ไม่ได้สูญหาย จากนั้นนำมาวิเคราะห์ด้วยสมการถดถอยโลจิสติก เพื่อประมาณค่าพารามิเตอร์ต่อไป เช่นในกรณีที่มีค่าสูญหายในตัวแปรอิสระตัวที่ 2 จะนำข้อมูลที่ไม่สูญหายของตัวแปรอิสระตัวที่ 2 มาหาค่าเฉลี่ย แล้วนำมาแทนในตัวแปรอิสระที่ข้อมูลสูญหาย ซึ่งวิธีการนี้จะส่งผลให้ค่าของข้อมูลที่สูญหายมีค่าเท่ากันในทุกค่าที่ข้อมูลสูญหาย โดยจะได้แบบจำลอง คือ

$$P(Y=1|X_1=j, X_2=k') \quad = \quad \mu_{j\bar{k}} \qquad \text{เมื่อ } \bar{k}=\sum_{i=1}^{N} k_i$$

และ $$P(Y=1|X_1=j', X_2=k) \quad = \quad \mu_{\bar{j}k} \qquad \text{เมื่อ } \bar{j}=\sum_{i=1}^{N} j_i$$

ในกรณีที่มีค่าสูญหายในตัวแปรอิสระตัวที่ 1

เมื่อ $\mu_{jk} = \dfrac{e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}{1+e^{\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}}}$

$X_1 = j'$ หมายถึง ข้อมูลมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระตัวที่ 1

$X_2 = k'$ หมายถึง ข้อมูลมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระตัวที่ 2

2.3 การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด Maximum Likelihood Estimation (MLE)

การประมาณค่าด้วยวิธีความควรจะเป็นสูงสุดนี้มีจุดประสงค์เพื่อหาตัวประมาณพารามิเตอร์ที่ทำให้ $L(\beta)$ มีค่ามากที่สุด โดยจะทำการหาอนุพันธ์ เทียบกับ β_i เมื่อ $i = 0,1,...,m$ และให้ผลลัพธ์มีค่าเท่ากับศูนย์ จะทำให้ได้สมการที่ไม่เป็นเชิงเส้น p+1 สมการ ซึ่งสามารถนำไปประมาณได้ด้วยวิธีการของนิวตัน – ราฟสัน (Newton – Raphson method)

โดยวิธีการ ของนิวตัน – ราฟสัน มีวิธีการคือ ทำการหาอนุพันธ์ย่อยของ $L(\beta)$ เทียบกับ β_i เมื่อ $i = 0,1,...,m$ ซึ่งจะเรียกว่า efficient scores จากนั้นนำมาเป็นสมาชิกของเวกเตอร์ U(B) ที่มีมิติ (m+1) x 1

$$U(B) = \begin{bmatrix} \dfrac{\partial L(B)}{\partial \beta_0} \\ \dfrac{\partial L(B)}{\partial \beta_1} \\ \vdots \\ \dfrac{\partial L(B)}{\partial \beta_m} \end{bmatrix}_{(m+1)x1}$$

กำหนดเมตริกซ์ H(B) ที่มีมิติ (m+1) x (m+1) โดยมีสมาชิกเป็นอนุพันธ์ย่อยอันดับที่สองของ $L(\beta)$ และเรียกเมตริกซ์ H(B) ว่า Hessian matrix

โดยที่สมาชิกตัวที่ (j,k) คือ $\dfrac{\partial^2 L(B)}{\partial \beta_j \partial \beta_k}$; j,k = 0 ,1,2, … , m

ทำให้เราสามารถหาเวกเตอร์ $U(\hat{B})$ ซึ่งเป็นเวกเตอร์ที่ประมาณด้วยความควรจะเป็นสูงสุดของ U(B) โดยใช้ Taylor series กระจาย U(B) รอบ B^* ทำให้ได้ว่า

$$U(\hat{B}) \approx U(B^*) + H(B^*)(\hat{B} - B^*)$$

โดยนิยามความควรจะเป็นสูงสุดของ B จะได้ว่า

$$\left.\frac{\partial L(B)}{\partial \beta_j}\right|_{\hat{B}} = 0$$ สำหรับ j = 0 ,1,2, … , m

และ $U(\hat{B}) = 0$

ดังนั้น $\hat{B} = B^* - H^{-1}(B^*)U(B^*)$

ซึ่งชี้นำให้มีการประมาณ $\hat{B}$ โดยการคำนวณซ้ำ ซึ่งค่าประมาณ ณ รอบที่ r+1 คือ

$$\hat{B}_{r+1} = \hat{B}_r - H^{-1}(B_r)\,U(B_r)$$

สำหรับ r = 0 ,1,2 ซึ่งเวกเตอร์ $\hat{B}_0$ เป็นเวกเตอร์ของตัวประมาณเริ่มต้น

ถ้าค่าผลต่างระหว่าง $\hat{B}$ ในรอบที่ r กับรอบที่ r+1 มีค่าน้อยมากจนถือว่าไม่แตกต่างกันซึ่งในที่นี้กำหนดเกณฑ์ว่า $\left|\hat{B}_{r+1} - \hat{B}_r\right| < 0.00001$ ค่า $\hat{B}_{r+1}$ จะถือว่าเป็นค่าที่ยอมรับได้

ลอการิทึมธรรมชาติของฟังก์ชันความควรจะเป็นสูงสุดในสมการถดถอยโลจิสติก ซึ่งมีฟังก์ชันความควรจะเป็นของ β_i คือ

$$L(\beta) = \sum_{r=1}^{n} L(\beta, Y_r, X_{1r}, X_{2r})$$

โดย $L(\beta, Y_r, X_{1r}, X_{2r}) = \log \mu_{jk}(\beta)^y (1 - \mu_{jk}(\beta))^{1-y}$

กำหนด joint distribution คือ

$$P(X_1 = j) = \tau_j$$

$$P(X_2 = k) = \tau_k$$

$$P(X_1 = j | X_2 = k) = \pi_{j|k} = \frac{n_{.jk}}{n_{..k}}$$

$$P(X_2 = k | X_1 = j) = \pi_{k|j} = \frac{n_{.jk}}{n_{.j.}}$$

ทำให้ได้ ฟังก์ชันความควรจะเป็นของ β_i ในกรณีที่ไม่มีค่าสูญหาย คือ

$$L(\beta, \pi) = \sum_{r=1}^{n} L(\beta, \pi; Y_r, X_{1r}, X_{2r})$$

โดย $L(\beta, \pi; Y_r, X_{1r}, X_{2r}) = \log\left[\mu_{jk}(\beta)^y (1 - \mu_{jk}(\beta))^{1-y}\right] + \log \pi_{k|j}$

และในกรณีที่มีค่าสูญหายจะได้ว่า

$$L(\beta, \pi) = \sum_{r=1}^{n} L(\beta, \pi; Y_r, X_{1r}, X_{2r})$$

โดย
$$L(\beta, \pi; y, j, k) = \begin{cases} \log\left[\sum_{j'=1}^{J} \mu_{j'k}(\beta)^y (1 - \mu_{j'k}(\beta))^{1-y} \pi_{j'|k}\right] & j = j' \\ \log\left[\sum_{k'=1}^{K} \mu_{jk'}(\beta)^y (1 - \mu_{jk'}(\beta))^{1-y} \pi_{k'|j}\right] & k = k' \end{cases}$$

เมื่อ $j = j'$ หมายถึง ข้อมูลมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระตัวที่ 1

เมื่อ $k = k'$ หมายถึง ข้อมูลมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระตัวที่ 2

2.4 การประมาณค่าพารามิเตอร์ด้วยวิธี Pseudo Maximum Likelihood Estimation (PMLE)

วิธีนี้มีหลักในการคำนวณ รวมถึงวัตถุประสงค์ที่คล้ายกับวิธี Maximum Likelihood คือ เพื่อหาตัวประมาณพารามิเตอร์ที่ทำให้ $L_{PMLE}(\beta)$ มีค่ามากที่สุด ซึ่งเมื่อได้ Log likelihood Function แล้วก็นำไปประมาณพารามิเตอร์โดยวิธี ML ต่อไป แต่วิธี PMLE นี้ ไม่จำเป็นต้องกระทำภายใต้ MAR Assumption เพราะสามารถกระทำได้ภายใน MDX Assumption (Missing dependent on X) และ MDY Assumption (Missing dependent on Y) โดยในงานวิจัยนี้ได้ทำการประมาณพารามิเตอร์ภายใต้ MDX Assumption ทำให้สามารถประมาณพารามิเตอร์รบกวน π ได้โดย

$$\hat{\pi}^n_{k|j} = \frac{n_{.jk}}{n_{+k}} \qquad \text{และ} \quad \hat{\pi}^n_{j|k} = \frac{n_{.jk}}{n_{.j+}}$$

และเรียกวิธีการประมาณนี้ว่า PMLEX ทำให้ได้ ฟังก์ชันความควรจะเป็นของ β_i ในกรณีที่ไม่มีค่าสูญหาย คือ

$$L_{PMLE}(\beta) = \sum_{r=1}^{n} L_{PMLE}(\beta, \hat{\pi}_n; Y_r, X_{1r}, X_{2r})$$

โดย $L_{PMLE}(\beta, \hat{\pi}_n; Y_r, X_{1r}, X_{2r}) = \log\left[\mu_{jk}(\beta)^y (1-\mu_{jk}(\beta))^{1-y}\right]$

และในกรณีที่มีค่าสูญหายจะได้ว่า

$$L_{PMLE}(\beta) = \sum_{r=1}^{n} L_{PMLE}(\beta, \hat{\pi}_n, Y_r, X_{1r}, X_{2r})$$

โดย $L_{PMLE}(\beta, \hat{\pi}_n; y, j, k) = \begin{cases} \log\left[\mu_{j'k}(\beta,\pi)^y (1-\mu_{j'k}(\beta,\pi))^{1-y}\right] & j = j' \\ \log\left[\mu_{jk'}(\beta,\pi)^y (1-\mu_{jk'}(\beta,\pi))^{1-y}\right] & k = k' \end{cases}$

เมื่อ $j = j'$ หมายถึง ข้อมูลมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระตัวที่ 1

เมื่อ $k = k'$ หมายถึง ข้อมูลมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระตัวที่ 2

2.5 **การประมาณค่าพารามิเตอร์ด้วยวิธี** The Filling Method (FILL)

วิธีนี้มีวิธีการ 2 ขั้นตอนคือ ขั้นตอนแรกเราจะสร้างตารางค่าสูญหายขึ้นมา (unobservable table) ($\left(n_{ijk}\right)_{i=0,1;\, j=1,\dots,J;\, k=1,\dots,K}$) โดยการกระจาย หน่วย $n_{ij'k}$ และ $n_{ijk'}$ ไปยังเซล (i,j,k) สำหรับค่าสูญหายในตัวแปรอิสระตัวที่ 1 และค่าสูญหายในตัวแปรอิสระตัวที่ 2 ตามลำดับ ซึ่งจะพบว่า $n_{ij'k}$ และ $n_{ijk'}$ นั้น จะเป็นสัดส่วนโดยตรงกับความน่าจะเป็นที่ j และ k จะเป็นค่าจริงใน X_1 และ X_2 จากนั้นทำการประมาณความน่าจะเป็นโดย $\dfrac{n_{ijk}}{n_{i+k}}$ และ $\dfrac{n_{ijk}}{n_{ij+}}$ แล้วนำไปเติมเป็นตารางที่สมบูรณ์ นั่นคือ

ตารางที่สมบูรณ์ = ตารางที่ไม่สมบูรณ์ + (ส่วนที่สูญหาย x ความน่าจะเป็น)

$$\hat{n}_{ijk} = n_{ijk} + n_{ij'k}\frac{n_{ijk}}{n_{i+k}}$$; สำหรับค่าสูญหายในตัวแปรอิสระตัวที่ 1

$$\hat{n}_{ijk} = n_{ijk} + n_{ijk'}\frac{n_{ijk}}{n_{ij+}}$$; สำหรับค่าสูญหายในตัวแปรอิสระตัวที่ 2

เมื่อ $n_{i+k} = \sum_{j=1}^{J} n_{ijk}$ และ $n_{ij+} = \sum_{k=1}^{K} n_{ijk}$

ตารางที่สมบูรณ์

$(\hat{n}_{ijk})$

n_{011}	n_{012}		…	n_{01K}
n_{021}	n_{022}		…	n_{02K}
n_{0J1}	n_{0J2}		…	n_{0JK}
n_{111}	n_{112}		…	n_{11K}
n_{121}	n_{122}		…	n_{12K}
n_{1J1}	n_{1J2}		…	n_{1JK}

ตารางที่ไม่สมบูรณ์ (n_{ijk})

(สมมติให้หน่วยที่ $n_{02K'}$ มีค่าสูญหายในตัวแปรอิสระตัวที่ 2)

n_{011}	n_{012}		…	n_{01K}
n_{021}	$n_{02K'}$		…	n_{02K}
n_{0J1}	n_{0J2}		…	n_{0JK}
n_{111}	n_{112}		…	n_{11K}
n_{121}	n_{122}		…	n_{12K}
n_{1J1}	n_{1J2}		…	n_{1JK}

และจากตารางที่สมบูรณ์ที่ได้นี้ จะทำการประมาณ $\hat{\beta}_i$ โดยใช้สมการถดถอยโลจิสติกและวิธีความควรจะเป็นสูงสุดต่อไป

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลองเพื่อต้องการศึกษาและเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในการวิเคราะห์ความถดถอยโลจิสติก โดยใช้เกณฑ์ bias และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (Average Mahalanobis Distance) เป็นเกณฑ์ประกอบในการตัดสินใจ ซึ่งวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นที่นำมาใช้มี 4 วิธี ดังนี้

วิธีที่ 1 คือ วิธี Mean Imputation (MEAN)

วิธีที่ 2 คือ วิธี Maximum Likelihood Estimation (MLE)

วิธีที่ 3 คือ วิธี Pseudo Maximum Likelihood Estimation (PMLE)

วิธีที่ 4 คือ วิธี The Filling Method (FILL)

ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จาก การจำลองด้วยเทคนิควิธี แบบมอนติคาร์โล (Monte carlo Simulation Technique) และใช้โปรแกรม MATHLAB (Matrix Laboratory Version 7) เพื่อประมวลผลและวิเคราะห์ข้อมูล โดยมีขั้นตอนในการวิจัยดังนี้

3.1 แผนการทดลอง

ผู้วิจัยได้กำหนดสถานการณ์ต่าง ๆ สำหรับการวิจัยครั้งนี้ไว้ดังนี้

1. การสูญหายเกิดขึ้นกับตัวแปรอิสระตัวที่ 1 หรือตัวแปรอิสระที่ 2 ตัวใดตัวหนึ่ง โดยการสูญหายเป็นการสูญหายแบบสุ่ม
2. ลักษณะของตัวแปรอิสระที่นำมาศึกษามีรูปแบบการแจกแจงปกติ (Normal Distribution)

   ฟังก์ชันการแจกแจงคือ

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

   ค่าคาดหวัง $E(X) = \mu$

   ความแปรปรวน $V(X) = \sigma^2$

3. กำหนดค่า $\beta_0, \beta_1, \beta_2$ = 0.2 และ β_0, β_1 = 0.2 , β_2 = 1
4. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 40 ,70 , 90 , 100 , 200 และ 400
5. ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ คือ 0 , 0.1 และ 0.2

6. ข้อมูลตัวแปรอิสระสูญหาย 3 ระดับ คือ ร้อยละ 5 , 10 และ 15
7. การวิจัยครั้งนี้ทำการจำลองข้อมูลให้มีสถานการณ์ที่แตกต่างกันตามข้อกำหนดข้างต้น โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte carlo Simulation Technique) ทำการจำลองในแต่ละสถานการณ์ 1,000รอบ

3.2 ขั้นตอนการวิจัย

ในการวิจัยครั้งนี้ ได้ทำการจำลองการทดลองตามสถานการณ์ต่าง ๆ โดยการใช้โปรแกรมคอมพิวเตอร์ ด้วยโปรแกรมแมทแล็บ (MATHLAB) โดยอาศัยเทคนิคมอนติคาร์โล เพื่อสร้างข้อมูลให้เป็นไปตามการวิจัย โดยทำการจำลองในแต่ละสถานการณ์ 1,000 รอบตามขั้นตอนดังนี้

1.การสร้างข้อมูล

ตัวแปรที่สนใจศึกษา y และ x มีความสัมพันธ์ภายใต้ตัวแบบความถดถอยโลจิสติกแบบ 2 กลุ่ม (Binary Logistic Regression) คือ

$$P(Y=1)=\frac{e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}}}{1+e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}}} \qquad ;i=1,2,\ldots,n$$

เมื่อ $P(Y=1)$ เป็นความน่าจะเป็นที่ Y=1 ณ ระดับ X_{i1} และ X_{i2}

X_{i1} และ X_{i2} เป็นตัวแปรอิสระตัวที่ 1 และ 2

$\beta_0, \beta_1, \beta_2$ เป็นพารามิเตอร์ที่ไม่ทราบค่า

n เป็นขนาดตัวอย่าง

โดยตัวแปรอิสระที่นำมาศึกษามีรูปแบบการแจกแจงปกติ (Normal Distribution)

ฟังก์ชันการแจกแจงคือ

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

ค่าคาดหวัง $E(X)=\mu$

ความแปรปรวน $V(X)=\sigma^2$

2.สุ่มตำแหน่งการสูญหายของข้อมูล

ทำการสุ่มตำแหน่งที่สูญหายตามร้อยละการสูญหายที่กำหนดโดยใช้การวนซ้ำในการหาตำแหน่งการสูญหาย

3.ประมาณค่าสูญหายด้วยวิธีการทั้ง 4 วิธี

1. วิธี Mean Imputation (MEAN) โดยการแทนค่าสูญหายด้วยค่าเฉลี่ยของข้อมูลที่ไม่สูญหาย
2. วิธี Maximum Likelihood Estimation (MLE) โดยใช้หลักของ EM algorithm (Expectation Maximization) ในการประมาณค่าสูญหายโดยมีขั้นตอนคือ ประมาณพารามิเตอร์ β_i จากชุดข้อมูลที่ไม่สูญหาย ซึ่งถือเป็นพารามิเตอร์เริ่มต้น $\hat{\beta}^0$ ประมาณข้อมูลสูญหายด้วยค่าคาดหวังของค่าที่สูญหายภายใต้เงื่อนไขของชุดข้อมูลที่ไม่สูญหายและพารามิเตอร์ตัวปัจจุบัน จากนั้นแทนค่าข้อมูลสูญหายจากค่าประมาณที่ได้ แล้วหาพารามิเตอร์ปัจจุบันใหม่ $\hat{\beta}^t$ โดยวิธี MLE และทำซ้ำจนกระทั่งพารามิเตอร์ปัจจุบันคงที่จะได้ค่าประมาณค่าที่สูญหาย ซึ่งตัวประมาณที่ได้เป็นตัวประมาณภาวะน่าจะเป็นสูงสุด
3. วิธี Pseudo Maximum Likelihood Estimation (PMLE) ที่หลักในการประมาณค่าสูญหายเหมือนกับวิธี ML แต่มีรูปแบบของฟังก์ชันความควรจะเป็นที่แตกต่างกัน
4. วิธี The Filling Method (FILL) ใช้หลัก EM algorithm ในการประมาณค่าสูญหายเช่นกัน แต่มีการใช้ความน่าจะเป็นที่จะพบข้อมูลสูญหายมาใช้ในการประมาณค่าพารามิเตอร์ด้วย

4. ประมาณพารามิเตอร์ใหม่

จากวิธีการประมาณค่าสูญหายทั้ง 4 วิธี เมื่อแทนค่าข้อมูลสูญหายจากค่าประมาณที่ได้แล้วจะทำการประมาณค่าพารามิเตอร์ใหม่ด้วยวิธีความควรจะเป็นสูงสุด (MLE)

5.หาค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง

เปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง และค่าระยะห่างมาหาลาโนบิสเฉลี่ย (Average Mahalanobis Distance) เป็นเกณฑ์ประกอบในการตัดสินใจ โดยมีวิธีการคำนวณคือ

Bias คำนวณจาก ค่าเฉลี่ยของความแตกต่างระหว่างค่าจริงและค่าประมาณ

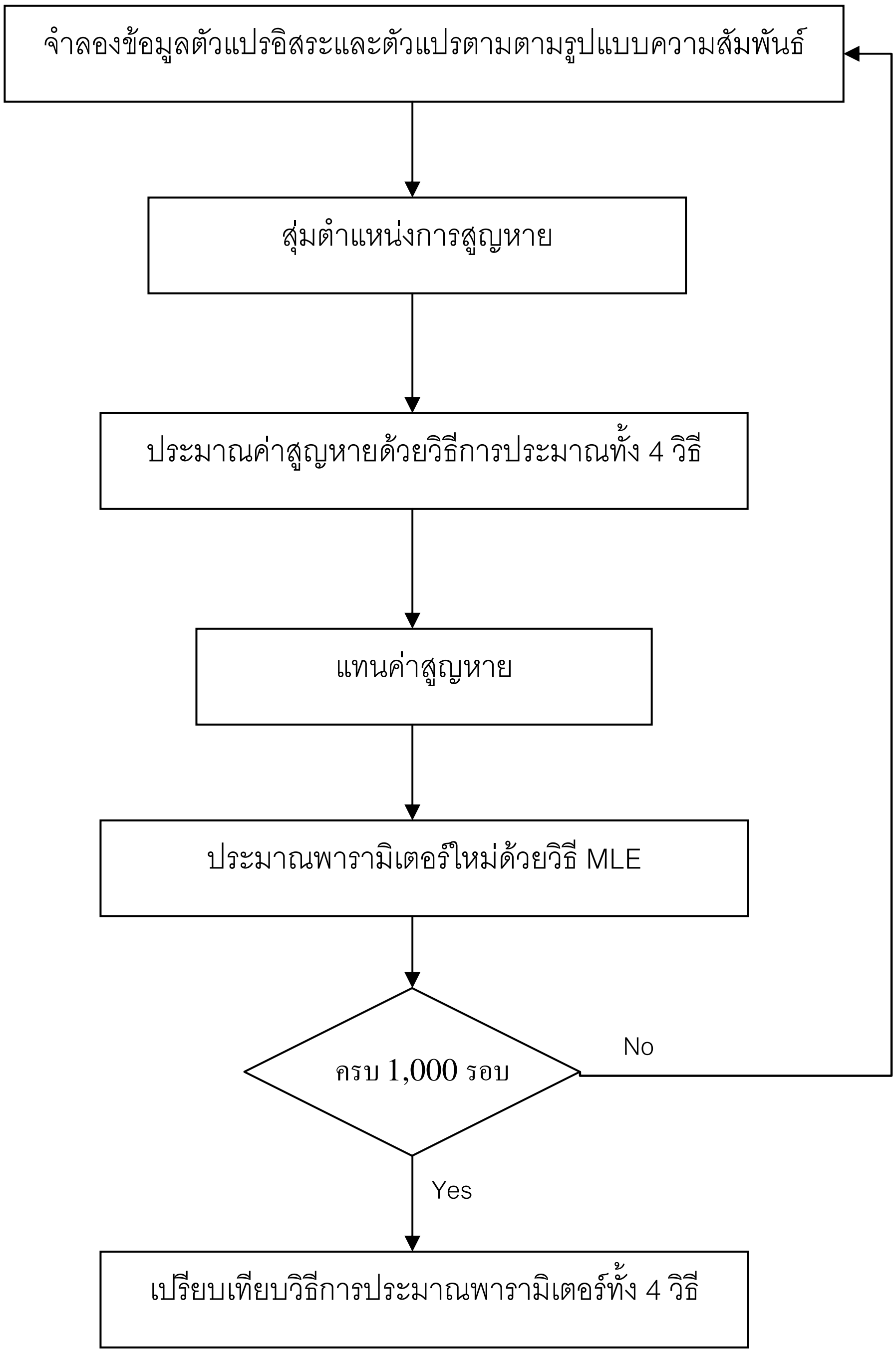
Average Mahalanobis Distance คำนวณจาก

$$AMH=\sum_{k=1}^{1000}\frac{\sqrt{\left(\underset{\sim}{\hat{\beta}}-\underset{\sim}{\beta}\right)'\left(Cov\left(\underset{\sim}{\hat{\beta}_i},\underset{\sim}{\hat{\beta}_j}\right)\right)^{-1}\left(\underset{\sim}{\hat{\beta}}-\underset{\sim}{\beta}\right)}}{1000}$$

เมื่อ $\underset{\sim}{\beta_j}$ หมายถึง ค่าจริงของพารามิเตอร์ในตัวแบบถดถอยโลจิสติกในการจำลองรอบที่ j

เมื่อ $\underset{\sim}{\hat{\beta}_j}$ หมายถึง ค่าประมาณของพารามิเตอร์ในตัวแบบถดถอยโลจิสติกในการจำลองรอบที่ j

ผู้วิจัยได้แสดงตารางแผนผังการเขียนโปรแกรม ดังรูป 3.1



ภาพที่ 3.1 แผนผังการเขียนโปรแกรม

บทที่ 4

ผลการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณของตัวแบบความถดถอยโลจิสติกแบบ 2 กลุ่ม ที่ประมาณค่าสูญหายด้วยวิธี MEAN Imputation (MEAN) วิธี Maximum Likelihood Estimation (MLE) วิธี Pseudo Maximum Likelihood Estimation (PMLE) และวิธี The Filling Method (FILL) เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระในกรณีที่มีตัวแปรอิสระ 2 ตัว ภายใต้สถานการณ์ที่แตกต่างกัน โดยเกณฑ์ที่ใช้ในการตัดสินใจว่าการประมาณค่าสูญหายด้วยวิธีใดใช้ได้ดีกว่าจากการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง BIAS, และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (Average Mahalanobis Distance) เป็นเกณฑ์ประกอบในการตัดสินใจ

การนำเสนอผลการวิจัยจะนำเสนอในรูปตารางและรูปกราฟ โดยใช้สัญลักษณ์แทนความหมายต่าง ๆ ดังนี้

n	หมายถึง	ขนาดตัวอย่าง
MEAN	หมายถึง	วิธี MEAN Imputation
MLE	หมายถึง	วิธี Maximum Likelihood Estimation
PMLE	หมายถึง	วิธี Pseudo Maximum Likelihood Estimation
FILL	หมายถึง	วิธี The Filling Method
BIAS	หมายถึง	ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง
AMH	หมายถึง	ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (Average Mahalanobis Distance)
SD	หมายถึง	ส่วนเบี่ยงเบนมาตรฐาน

การเสนอผลการวิเคราะห์ข้อมูลในบทนี้ เป็นการแสดงผลการวิเคราะห์ตามขั้นตอนการวิเคราะห์ข้อมูลที่ได้กล่าวไว้ในบทที่ 3 โดยผู้วิจัยได้เสนอผลการวิจัยโดยแบ่งออกเป็น 2 ส่วน คือ

ส่วนที่ 1 ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ กรณีที่ ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

ส่วนที่ 2 ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ กรณีที่ ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

4.1 ผลการเปรียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ กรณีที่ $\beta_0, \beta_1, \beta_2$ = 0.2

การวิจัยครั้งนี้ผู้วิจัยทำการศึกษา กรณีที่ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2 ความสัมพันธ์ระหว่างตัวแปรอิสระหรือค่าสหสัมพันธ์เท่ากับ 0 , 0.1 และ 0.2 ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ40, 70, 90, 100 ,200 และ 400 และระดับการสูญหายของตัวแปรอิสระ 3 ระดับ คือ ร้อยละ 5 , 10 และ 15 ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.1.1 – 4.1.3

ตารางที่	ค่าพารามิเตอร์เริ่มต้น	ระดับการสูญหายของตัวแปรอิสระ
4.1.1	$\beta_0, \beta_1, \beta_2$ = 0.2	ร้อยละ 5
4.1.2	$\beta_0, \beta_1, \beta_2$ = 0.2	ร้อยละ 10
4.1.3	$\beta_0, \beta_1, \beta_2$ = 0.2	ร้อยละ 15

ตารางที่ 4.1.1 แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือ ร้อยละ 5 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

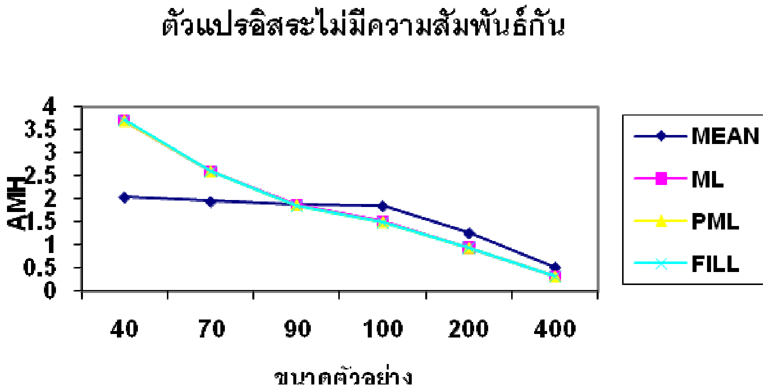
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
ตัวแปรอิสระไม่มีความสัมพันธ์กัน	40	BIAS	0.1884	0.3391	0.3398	0.3400
		SD	(0.0860)	(0.1429)	(0.1468)	(0.1553)
		AMH	2.0389	3.6894	3.6929	3.6992
		SD	(0.5366)	(0.9484)	(0.9493)	(0.9506)
	70	BIAS	0.1755	0.2345	0.2348	0.2350
		SD	(0.0818)	(0.1001)	(0.1005)	(0.1040)
		AMH	1.9438	2.5889	2.5962	2.5993
		SD	(0.4994)	(0.6644)	(0.6648)	(0.6650)
	90	BIAS	0.1642	0.1648	0.1656	0.1669
		SD	(0.0696)	(0.0747)	(0.0754)	(0.0661)
		AMH	1.8493	1.8651	1.8725	1.8803
		SD	(0.4735)	(0.4746)	(0.4751)	(0.4776)
	100	BIAS	0.1626	0.1309	0.1298	0.1290
		SD	(0.0775)	(0.0572)	(0.0541)	(0.0526)
		AMH	1.8486	1.5091	1.4995	1.4785
		SD	(0.4622)	(0.3815)	(0.3802)	(0.3781)
	200	BIAS	0.1070	0.0791	0.0787	0.0787
		SD	(0.0403)	(0.0364)	(0.0331)	(0.0324)
		AMH	1.2607	0.9397	0.9368	0.9360
		SD	(0.3160)	(0.2419)	(0.2414)	(0.2412)
	400	BIAS	0.0340	0.0208	0.0207	0.0207
		SD	(0.0129)	(0.0026)	(0.0022)	(0.0021)
		AMH	0.5173	0.3174	0.3167	0.3165
		SD	(0.1330)	(0.0789)	(0.0786)	(0.0785)

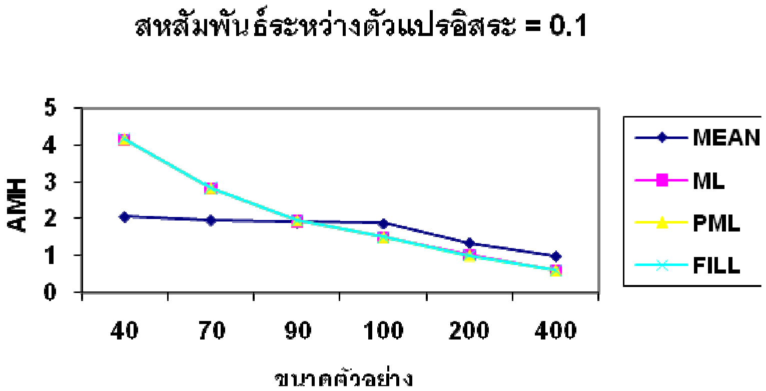
ตารางที่ 4.1.1 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 5 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

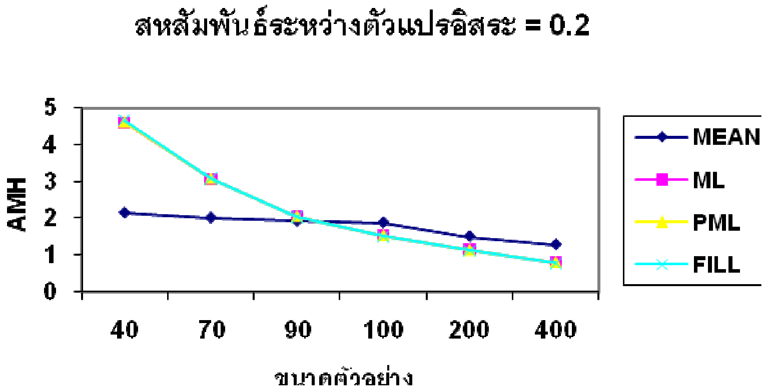
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.1	40	BIAS	0.2205	0.4460	0.4468	0.4476
		SD	(0.1095)	(0.2029)	(0.2063)	(0.2082)
		AMH	2.0587	4.1431	4.1510	4.1588
		SD	(0.5160)	(1.0332)	(1.0341)	(1.0349)
	70	BIAS	0.2070	0.3009	0.3010	0.3010
		SD	(0.1008)	(0.1350)	(0.1363)	(0.1370)
		AMH	1.9658	2.8271	2.8296	2.8308
		SD	(0.4939)	(0.7086)	(0.7088)	(0.7091)
	90	BIAS	0.1978	0.2032	0.2038	0.2043
		SD	(0.0868)	(0.0895)	(0.0896)	(0.0897)
		AMH	1.9040	1.9445	1.9454	1.9538
		SD	(0.4792)	(0.4919)	(0.4919)	(0.4922)
	100	BIAS	0.1932	0.1560	0.1552	0.1543
		SD	(0.0920)	(0.0671)	(0.0662)	(0.0657)
		AMH	1.8731	1.5159	1.5032	1.5027
		SD	(0.4718)	(0.3840)	(0.3835)	(0.3833)
	200	BIAS	0.1199	0.0893	0.0891	0.0890
		SD	(0.0459)	(0.0388)	(0.0385)	(0.0384)
		AMH	1.3460	1.0077	1.0053	1.0045
		SD	(0.3545)	(0.2621)	(0.2618)	(0.2616)
	400	BIAS	0.0875	0.0537	0.0536	0.0536
		SD	(0.0380)	(0.0235)	(0.0233)	(0.0232)
		AMH	0.9841	0.6079	0.6066	0.6064
		SD	(0.2460)	(0.1603)	(0.1597)	(0.1596)

ตารางที่ 4.1.1 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือ ร้อยละ 5 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.2	40	BIAS	0.2368	0.5104	0.5109	0.5156
		SD	(0.1121)	(0.2084)	(0.2095)	(0.2122)
		AMH	2.1522	4.5921	4.6152	4.6539
		SD	(0.5533)	(1.1602)	(1.1618)	(1.1635)
	70	BIAS	0.2169	0.3352	0.3354	0.3371
		SD	(0.1009)	(0.1393)	(0.1396)	(0.1407)
		AMH	2.0184	3.0656	3.0783	3.0891
		SD	(0.5183)	(0.7810)	(0.7816)	(0.7823)
	90	BIAS	0.2037	0.2180	0.2180	0.2186
		SD	(0.0929)	(0.0930)	(0.0932)	(0.0934)
		AMH	1.9292	2.0453	2.0459	2.0480
		SD	(0.4949)	(0.5281)	(0.5281)	(0.5282)
	100	BIAS	0.1971	0.1602	0.1594	0.1585
		SD	(0.0896)	(0.0701)	(0.0696)	(0.0692)
		AMH	1.8846	1.5392	1.5313	1.5243
		SD	(0.4832)	(0.4017)	(0.4014)	(0.4011)
	200	BIAS	0.1349	0.1014	0.1011	0.1008
		SD	(0.0587)	(0.0441)	(0.0439)	(0.0438)
		AMH	1.5061	1.1378	1.1344	1.1319
		SD	(0.3775)	(0.2986)	(0.2982)	(0.2979)
	400	BIAS	0.1112	0.0687	0.0687	0.0686
		SD	(0.0525)	(0.0328)	(0.0327)	(0.0327)
		AMH	1.2870	0.8000	0.7995	0.7985
		SD	(0.3218)	(0.2077)	(0.2075)	(0.2074)







<u>รูปที่ 4.1.1</u> แสดงการเปรียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ ระหว่างค่า AMH และขนาดตัวอย่าง กรณีที่ $\beta_0, \beta_1, \beta_2$ = 0.2 เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 5

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 5 กรณีที่ ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2 พบว่า

<u>ตัวแปรอิสระไม่มีความสัมพันธ์กัน</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

<u>สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.1</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40 , 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

<u>สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.2</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สรุปผล การที่สหสัมพันธ์ระหว่างตัวแปรอิสระมีค่าเพิ่มขึ้น นั่นคือตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น มีผลทำให้ค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น โดยเมื่อขนาดตัวอย่างน้อยกว่า 90 วิธี MEAN จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และในกรณีขนาดตัวอย่างมากกว่า 90 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE , MLE และ MEAN ตามลำดับ

ตารางที่ 4.1.2 แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 10 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

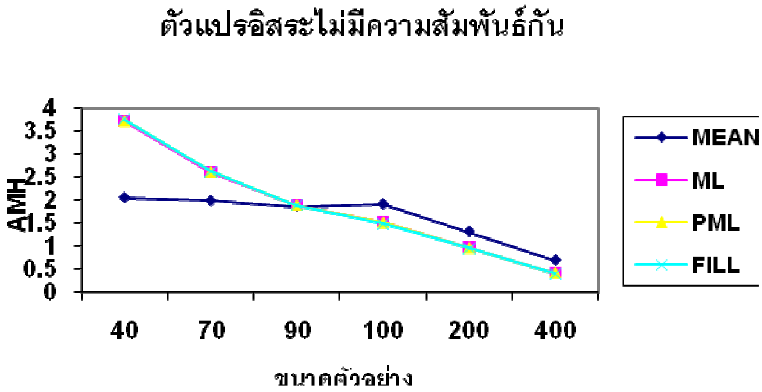
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
ตัวแปรอิสระไม่มีความสัมพันธ์กัน	40	BIAS	0.1963	0.3599	0.3609	0.3616
		SD	(0.0888)	(0.1519)	(0.1523)	(0.1536)
		AMH	2.0472	3.7142	3.7284	3.7431
		SD	(0.5373)	(0.9440)	(0.9445)	(0.9452)
	70	BIAS	0.1907	0.2524	0.2536	0.2538
		SD	(0.0830)	(0.1076)	(0.1078)	(0.1079)
		AMH	1.9801	2.6169	2.6191	2.6226
		SD	(0.5158)	(0.6608)	(0.6609)	(0.6611)
	90	BIAS	0.1795	0.1820	0.1830	0.1869
		SD	(0.0774)	(0.0778)	(0.0784)	(0.0791)
		AMH	1.8660	1.8854	1.8890	1.8954
		SD	(0.5015)	(0.5217)	(0.5225)	(0.5245)
	100	BIAS	0.1850	0.1476	0.1462	0.1431
		SD	(0.0771)	(0.0637)	(0.0629)	(0.0622)
		AMH	1.9130	1.5240	1.5168	1.4906
		SD	(0.4943)	(0.3782)	(0.3771)	(0.3764)
	200	BIAS	0.1199	0.0890	0.0887	0.0873
		SD	(0.0547)	(0.0419)	(0.0415)	(0.0410)
		AMH	1.3146	0.9695	0.9665	0.9600
		SD	(0.3287)	(0.2486)	(0.2480)	(0.2474)
	400	BIAS	0.0431	0.0259	0.0258	0.0256
		SD	(0.0183)	(0.0126)	(0.0124)	(0.0123)
		AMH	0.6957	0.4181	0.4174	0.4145
		SD	(0.1739)	(0.1068)	(0.1065)	(0.1063)

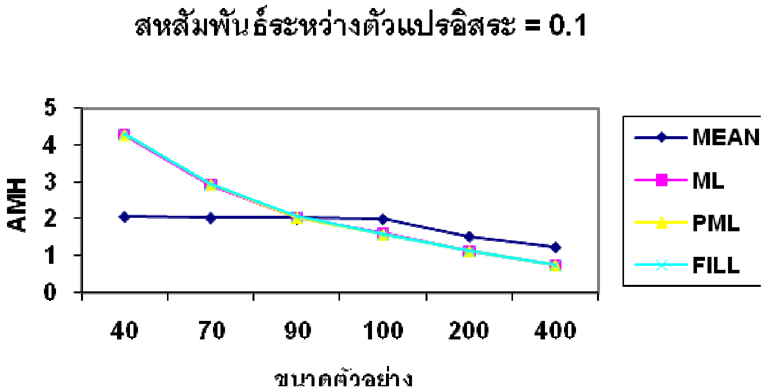
ตารางที่ 4.1.2 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือ ร้อยละ 10 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

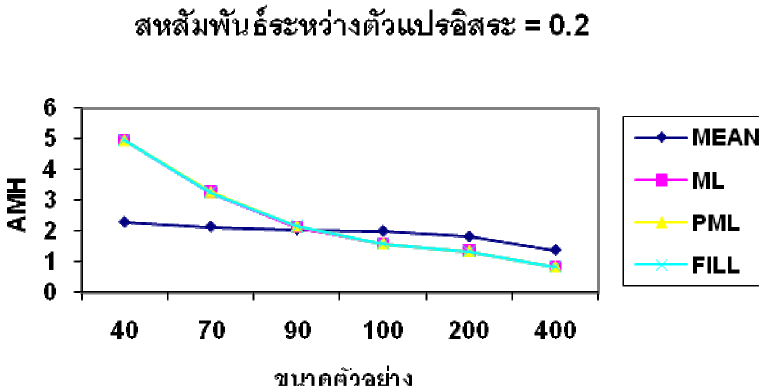
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.1	40	BIAS	0.2239	0.4685	0.4701	0.4702
		SD	(0.0969)	(0.2011)	(0.2017)	(0.2019)
		AMH	2.0588	4.2683	4.2800	4.2804
		SD	(0.5379)	(1.0671)	(1.0678)	(1.0681)
	70	BIAS	0.2186	0.3183	0.3197	0.3198
		SD	(0.0995)	(0.1371)	(0.1375)	(0.1376)
		AMH	2.0287	2.9243	2.9331	2.9361
		SD	(0.5319)	(0.7300)	(0.7301)	(0.7301)
	90	BIAS	0.2151	0.2170	0.2196	0.2207
		SD	(0.1012)	(0.0947)	(0.0949)	(0.0953)
		AMH	2.0086	2.0202	2.0351	2.0479
		SD	(0.5279)	(0.5047)	(0.5049)	(0.5054)
	100	BIAS	0.2133	0.1711	0.1695	0.1664
		SD	(0.1021)	(0.0741)	(0.0735)	(0.0733)
		AMH	1.9985	1.6038	1.5861	1.5681
		SD	(0.5259)	(0.3931)	(0.3923)	(0.3920)
	200	BIAS	0.1318	0.0982	0.0978	0.0968
		SD	(0.0563)	(0.0457)	(0.0456)	(0.0452)
		AMH	1.5177	1.1315	1.1191	1.1184
		SD	(0.3902)	(0.2874)	(0.2865)	(0.2860)
	400	BIAS	0.0940	0.0565	0.0562	0.0560
		SD	(0.0414)	(0.0254)	(0.0253)	(0.0252)
		AMH	1.2303	0.7453	0.7429	0.7423
		SD	(0.3204)	(0.1851)	(0.1849)	(0.1847)

ตารางที่ 4.1.2 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 10 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.2	40	BIAS	0.2415	0.5286	0.5297	0.5301
		SD	(0.0836)	(0.1888)	(0.1891)	(0.1893)
		AMH	2.2807	4.9441	4.9503	4.9612
		SD	(0.5702)	(1.2809)	(1.2814)	(1.2823)
	70	BIAS	0.2279	0.3490	0.3502	0.3503
		SD	(0.0783)	(0.1238)	(0.1240)	(0.1241)
		AMH	2.1371	3.2654	3.2693	3.2720
		SD	(0.5545)	(0.9203)	(0.9205)	(0.9209)
	90	BIAS	0.2188	0.2282	0.2305	0.2282
		SD	(0.0747)	(0.0801)	(0.0806)	(0.0809)
		AMH	2.0413	2.1348	2.1528	2.1531
		SD	(0.5440)	(0.6855)	(0.6861)	(0.6863)
	100	BIAS	0.2143	0.1719	0.1707	0.1678
		SD	(0.0729)	(0.0593)	(0.0589)	(0.0582)
		AMH	1.9934	1.5945	1.5937	1.5695
		SD	(0.5388)	(0.5674)	(0.5670)	(0.5661)
	200	BIAS	0.1471	0.1094	0.1088	0.1085
		SD	(0.0497)	(0.0398)	(0.0396)	(0.0395)
		AMH	1.8158	1.3539	1.3481	1.3453
		SD	(0.4620)	(0.3385)	(0.3381)	(0.3379)
	400	BIAS	0.1236	0.0747	0.0746	0.0743
		SD	(0.0413)	(0.0249)	(0.0248)	(0.0247)
		AMH	1.3855	0.8420	0.8406	0.8385
		SD	(0.3580)	(0.2148)	(0.2145)	(0.2142)







<u>รูปที่ 4.1.2</u> แสดงการเปรียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ ระหว่างค่า AMH และขนาดตัวอย่าง กรณีที่ $\beta_0, \beta_1, \beta_2$ = 0.2 เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 10

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 10 กรณีที่ ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2 พบว่า

ตัวแปรอิสระไม่มีความสัมพันธ์กัน ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.1 ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40 , 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.2 ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สรุปผล การที่สหสัมพันธ์ระหว่างตัวแปรอิสระมีค่าเพิ่มขึ้น นั่นคือตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น มีผลทำให้ค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น โดยเมื่อขนาดตัวอย่างน้อยกว่า 90 วิธี MEAN จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และในกรณีขนาดตัวอย่างมากกว่า 90 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE , MLE และ MEAN ตามลำดับ

ตารางที่ 4.1.3 แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือ ร้อยละ 15 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

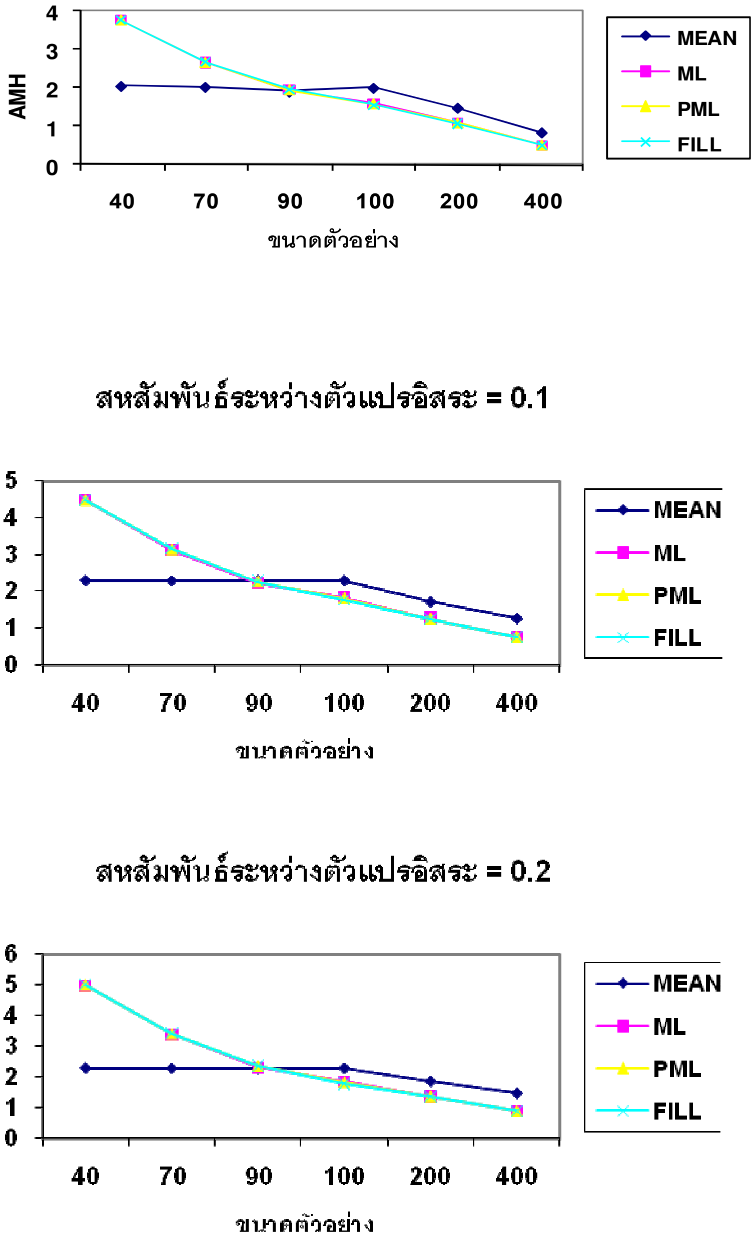
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
ตัวแปรอิสระไม่มีความสัมพันธ์กัน	40	BIAS	0.2107	0.3880	0.3881	0.3899
		SD	(0.0712)	(0.1311)	(0.1312)	(0.1315)
		AMH	2.0527	3.7484	3.7492	3.7688
		SD	(0.5132)	(0.9864)	(0.9866)	(0.9870)
	70	BIAS	0.2099	0.2745	0.2770	0.2779
		SD	(0.0712)	(0.0932)	(0.0933)	(0.0934)
		AMH	2.0290	2.6509	2.6663	2.6702
		SD	(0.5144)	(0.6924)	(0.6926)	(0.6927)
	90	BIAS	0.1975	0.2029	0.2039	0.2048
		SD	(0.0711)	(0.0677)	(0.0680)	(0.0682)
		AMH	1.9056	1.9443	1.9513	1.9532
		SD	(0.5151)	(0.5259)	(0.5265)	(0.5268)
	100	BIAS	0.2091	0.1669	0.1658	0.1590
		SD	(0.0711)	(0.0556)	(0.0553)	(0.0549)
		AMH	2.0053	1.5919	1.5833	1.5329
		SD	(0.5155)	(0.3990)	(0.3985)	(0.3977)
	200	BIAS	0.1327	0.0990	0.0977	0.0947
		SD	(0.0458)	(0.0345)	(0.0342)	(0.0339)
		AMH	1.4846	1.0999	1.0848	1.0614
		SD	(0.3866)	(0.2750)	(0.2746)	(0.2741)
	400	BIAS	0.0515	0.0310	0.0309	0.0303
		SD	(0.0172)	(0.0106)	(0.0105)	(0.0103)
		AMH	0.8454	0.5095	0.5079	0.5002
		SD	(0.2162)	(0.1293)	(0.1291)	(0.1287)

ตารางที่ 4.1.3 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 15 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.1	40	BIAS	0.2316	0.5043	0.5045	0.5048
		SD	(0.0844)	(0.1795)	(0.1797)	(0.1799)
		AMH	2.2832	4.4658	4.4701	4.4745
		SD	(0.5839)	(1.1569)	(1.1573)	(1.1576)
	70	BIAS	0.2311	0.3410	0.3440	0.3448
		SD	(0.0807)	(0.1205)	(0.1207)	(0.1208)
		AMH	2.2826	3.1228	3.1438	3.1498
		SD	(0.5876)	(0.8000)	(0.8003)	(0.8004)
	90	BIAS	0.2308	0.2317	0.2370	0.2384
		SD	(0.0782)	(0.0809)	(0.0814)	(0.0817)
		AMH	2.2821	2.2216	2.2595	2.2724
		SD	(0.5900)	(0.5615)	(0.5622)	(0.5627)
	100	BIAS	0.2306	0.1852	0.1835	0.1771
		SD	(0.0769)	(0.0621)	(0.0617)	(0.0611)
		AMH	2.2819	1.8337	1.8174	1.7710
		SD	(0.5912)	(0.4439)	(0.4432)	(0.4423)
	200	BIAS	0.1450	0.1072	0.1065	0.1044
		SD	(0.0480)	(0.0357)	(0.0355)	(0.0349)
		AMH	1.7177	1.2666	1.2577	1.2448
		SD	(0.4327)	(0.3256)	(0.3251)	(0.3245)
	400	BIAS	0.0987	0.0595	0.0588	0.0587
		SD	(0.0330)	(0.0203)	(0.0201)	(0.0200)
		AMH	1.2802	0.7714	0.7693	0.7658
		SD	(0.3241)	(0.1924)	(0.1918)	(0.1914)

ตารางที่ 4.1.3 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 15 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.2	40	BIAS	0.2487	0.5474	0.5484	0.5489
		SD	(0.0832)	(0.1862)	(0.1864)	(0.1866)
		AMH	2.2905	4.9727	4.9789	4.9857
		SD	(0.5726)	(1.2621)	(1.2624)	(1.2627)
	70	BIAS	0.2396	0.3638	0.3668	0.3638
		SD	(0.0801)	(0.1246)	(0.1247)	(0.1248)
		AMH	2.2889	3.3863	3.4060	3.4098
		SD	(0.5796)	(0.8713)	(0.8715)	(0.8716)
	90	BIAS	0.2335	0.2404	0.2457	0.2467
		SD	(0.0780)	(0.0832)	(0.0836)	(0.0839)
		AMH	2.2878	2.3200	2.3573	2.3679
		SD	(0.5842)	(0.6103)	(0.6108)	(0.6112)
	100	BIAS	0.2304	0.1865	0.1851	0.1787
		SD	(0.0769)	(0.0634)	(0.0630)	(0.0625)
		AMH	2.2872	1.8469	1.8330	1.7869
		SD	(0.5865)	(0.4810)	(0.4805)	(0.4798)
	200	BIAS	0.1568	0.1159	0.1156	0.1138
		SD	(0.0570)	(0.0414)	(0.0413)	(0.0409)
		AMH	1.8684	1.3799	1.3759	1.3639
		SD	(0.4815)	(0.3485)	(0.3482)	(0.3477)
	400	BIAS	0.1289	0.0781	0.0776	0.0770
		SD	(0.0449)	(0.0271)	(0.0269)	(0.0265)
		AMH	1.4961	0.9055	0.9017	0.8977
		SD	(0.3927)	(0.2287)	(0.2283)	(0.2278)



<u>รูปที่ 4.1.3</u> แสดงการเปรียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ ระหว่างค่า AMH และขนาดตัวอย่าง กรณีที่ $\beta_0, \beta_1, \beta_2$ = 0.2 เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 15

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 15 กรณีที่ ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2 พบว่า

<u>ตัวแปรอิสระไม่มีความสัมพันธ์กัน</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

<u>สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.1</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40 , 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

<u>สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.2</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สรุปผล การที่สหสัมพันธ์ระหว่างตัวแปรอิสระมีค่าเพิ่มขึ้น นั่นคือตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น มีผลทำให้ค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น โดยเมื่อขนาดตัวอย่างน้อยกว่า 90 วิธี MEAN จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และในกรณีขนาดตัวอย่างมากกว่า 90 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE , MLE และ MEAN ตามลำดับ

สรุปผลตอนที่ 4.1

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ กรณีที่ ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2 พบว่า วิธี MEAN จะเป็นวิธีที่ดีที่สุด ในกรณีที่ขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100 , 200 และ 400 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE MLE และ MEAN ตามลำดับ ในทุกค่าสหสัมพันธ์และทุกร้อยละการสูญหายของตัวแปรอิสระ

สามารถสรุปผลเกี่ยวกับปัจจัยที่มีผลต่อค่า BIAS และ AMH ได้ดังนี้

1. ขนาดตัวอย่าง

เมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH ลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ลง

2. ค่าสหสัมพันธ์

เมื่อค่าสหสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น ซึ่งจะส่งผลให้ความถูกต้องในการประมาณค่าพารามิเตอร์ลดลง

3. ร้อยละการสูญหายของตัวแปรอิสระ

เมื่อร้อยละการสูญหายของตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น เนื่องจากเมื่อร้อยละการสูญหายเพิ่มขึ้นจะส่งผลให้ข้อมูลลดลง ทำให้ค่าความคลาดเคลื่อนที่เกิดขึ้นจากการประมาณค่าพารามิเตอร์เพิ่มขึ้น

4.2 การเปรียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ กรณีที่ β_0, β_1 = 0.2 และ β_2 = 1

การวิจัยครั้งนี้ผู้วิจัยทำการศึกษา กรณีที่ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1 ความสัมพันธ์ระหว่างตัวแปรอิสระหรือค่าสหสัมพันธ์เท่ากับ 0 , 0.1 และ 0.2 ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ40, 70, 90, 100 ,200 และ 400 และระดับการสูญหายของตัวแปรอิสระ 3 ระดับ คือ ร้อยละ 5 , 10 และ 15 ซึ่งผลการวิจัยส่วนนี้นำเสนอในตารางที่ 4.2.1 – 4.2.3

ตารางที่	ค่าพารามิเตอร์เริ่มต้น	ระดับการสูญหายของตัวแปรอิสระ
4.2.1	β_0, β_1 = 0.2 และ β_2 = 1	ร้อยละ 5
4.2.2	β_0, β_1 = 0.2 และ β_2 = 1	ร้อยละ 10
4.2.3	β_0, β_1 = 0.2 และ β_2 = 1	ร้อยละ 15

ตารางที่ 4.2.1 แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือ ร้อยละ 5 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

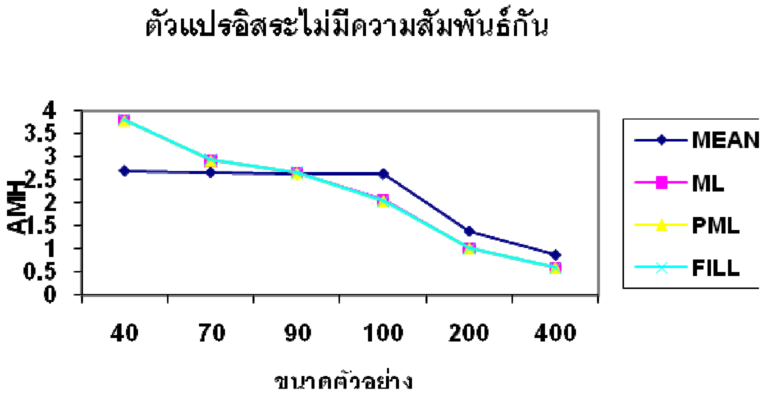
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
ตัวแปรอิสระไม่มีความสัมพันธ์กัน	40	BIAS	0.2951	0.3542	0.3549	0.3552
		SD	(0.1054)	(0.1247)	(0.1249)	(0.1251)
		AMH	2.6826	3.7791	3.7871	3.7895
		SD	(0.6809)	(1.0024)	(1.0029)	(1.0032)
	70	BIAS	0.2664	0.2693	0.2698	0.2703
		SD	(0.0930)	(0.0939)	(0.0940)	(0.0941)
		AMH	2.6518	2.9122	2.9177	2.9229
		SD	(0.6774)	(0.7779)	(0.7781)	(0.7783)
	90	BIAS	0.2473	0.2520	0.2530	0.2543
		SD	(0.0847)	(0.0881)	(0.8883)	(0.8887)
		AMH	2.6313	2.6473	2.6480	2.6512
		SD	(0.6750)	(0.6776)	(0.6782)	(0.6788)
	100	BIAS	0.2377	0.1863	0.1846	0.1833
		SD	(0.0806)	(0.0634)	(0.0631)	(0.0627)
		AMH	2.6210	2.0667	2.0482	2.0348
		SD	(0.6738)	(0.5541)	(0.5532)	(0.5525)
	200	BIAS	0.1195	0.0868	0.0864	0.0863
		SD	(0.0425)	(0.0300)	(0.0298)	(0.0297)
		AMH	1.3935	1.0198	1.0157	1.0146
		SD	(0.3484)	(0.2582)	(0.2577)	(0.2575)
	400	BIAS	0.0409	0.0276	0.0275	0.0275
		SD	(0.0150)	(0.0097)	(0.0096)	(0.0095)
		AMH	0.8822	0.6007	0.5991	0.5985
		SD	(0.2322)	(0.1624)	(0.1622)	(0.1620)

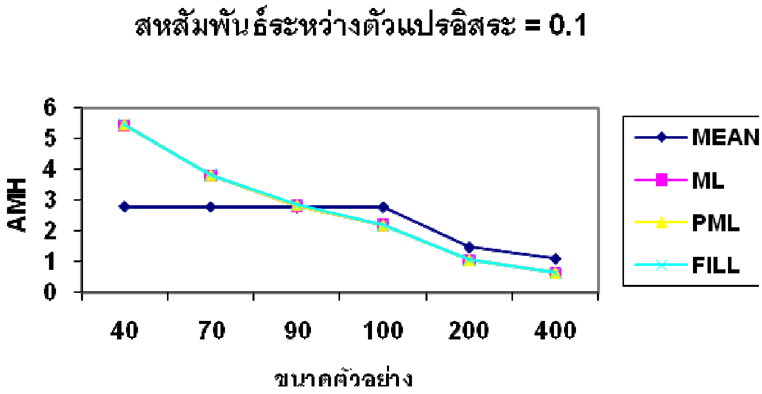
ตารางที่ 4.2.1 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 5 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

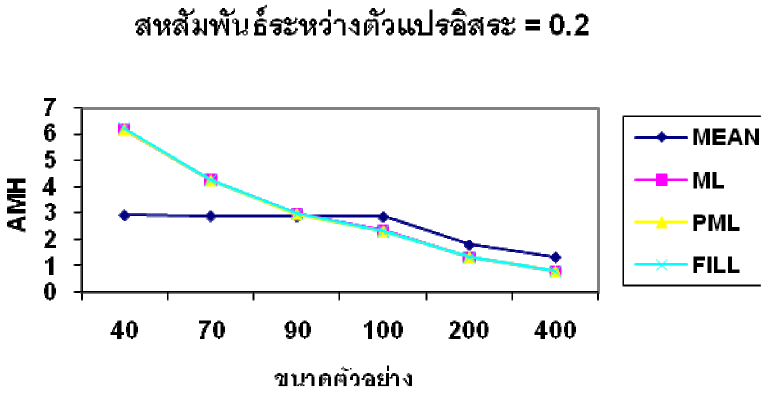
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.1	40	BIAS	0.3349	0.6841	0.6854	0.6867
		SD	(0.1179)	(0.2288)	(0.2291)	(0.2294)
		AMH	2.7831	5.4279	5.4377	5.4475
		SD	(0.7177)	(1.4591)	(1.4599)	(1.4608)
	70	BIAS	0.2952	0.4432	0.4433	0.4434
		SD	(0.1032)	(0.1512)	(0.1513)	(0.1513)
		AMH	2.7767	3.8205	3.8209	3.8221
		SD	(0.7148)	(1.0087)	(1.0088)	(1.0090)
	90	BIAS	0.2687	0.2812	0.2818	0.2825
		SD	(0.0933)	(0.0991)	(0.0994)	(0.0996)
		AMH	2.7724	2.8358	2.8430	2.8515
		SD	(0.7128)	(0.7277)	(0.7281)	(0.7284)
	100	BIAS	0.2555	0.2022	0.2011	0.2001
		SD	(0.0884)	(0.0738)	(0.0734)	(0.0730)
		AMH	2.7702	2.2162	2.2040	2.1935
		SD	(0.7118)	(0.5582)	(0.5577)	(0.5571)
	200	BIAS	0.2103	0.1528	0.1524	0.1522
		SD	(0.0743)	(0.0509)	(0.0506)	(0.0505)
		AMH	1.4759	1.0812	1.0786	1.0773
		SD	(0.3784)	(0.2794)	(0.2789)	(0.2787)
	400	BIAS	0.0849	0.0502	0.0502	0.0501
		SD	(0.0308)	(0.0177)	(0.0176)	(0.0176)
		AMH	1.1128	0.6659	0.6653	0.6646
		SD	(0.2868)	(0.1712)	(0.1710)	(0.1708)

ตารางที่ 4.2.1 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือ ร้อยละ 5 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.2	40	BIAS	0.3417	0.7446	0.7474	0.7533
		SD	(0.9875)	(0.2688)	(0.2693)	(0.2698)
		AMH	2.9284	6.1502	6.1733	6.2216
		SD	(0.7426)	(1.6314)	(1.6354)	(1.6392)
	70	BIAS	0.3123	0.4852	0.4860	0.4878
		SD	(0.5409)	(0.1747)	(0.1748)	(0.1749)
		AMH	2.8988	4.2505	4.2559	4.2730
		SD	(0.7411)	(1.1187)	(1.1203)	(1.1218)
	90	BIAS	0.2927	0.3108	0.3117	0.3122
		SD	(0.2431)	(0.1116)	(0.1117)	(0.1118)
		AMH	2.8790	2.9739	2.9776	2.9840
		SD	(0.7400)	(0.7768)	(0.7769)	(0.7769)
	100	BIAS	0.2829	0.2257	0.2245	0.2231
		SD	(0.0942)	(0.0806)	(0.0802)	(0.0799)
		AMH	2.8691	2.3508	2.3384	2.3243
		SD	(0.7395)	(0.6059)	(0.6052)	(0.6044)
	200	BIAS	0.2277	0.1670	0.1664	0.1659
		SD	(0.0774)	(0.0576)	(0.0574)	(0.0572)
		AMH	1.8194	1.3451	1.3408	1.3365
		SD	(0.4653)	(0.3494)	(0.3487)	(0.3480)
	400	BIAS	0.1256	0.0748	0.0747	0.0746
		SD	(0.0429)	(0.0263)	(0.0263)	(0.0262)
		AMH	1.3515	0.8144	0.8141	0.8128
		SD	(0.3557)	(0.2138)	(0.2137)	(0.2132)







รูปที่ 4.1.4 แสดงการเปรียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ ระหว่างค่า AMH และขนาดตัวอย่าง กรณีที่ β_0, β_1 = 0.2 และ β_2 = 1 เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 5

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 5 กรณีที่ ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1 พบว่า

<u>ตัวแปรอิสระไม่มีความสัมพันธ์กัน</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

<u>สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.1</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40 , 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

<u>สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.2</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สรุปผล การที่สหสัมพันธ์ระหว่างตัวแปรอิสระมีค่าเพิ่มขึ้น นั่นคือตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น มีผลทำให้ค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น โดยเมื่อขนาดตัวอย่างน้อยกว่า 90 วิธี MEAN จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และในกรณีขนาดตัวอย่างมากกว่า 90 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE , MLE และ MEAN ตามลำดับ

ตารางที่ 4.2.2 แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือ ร้อยละ 10 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
ตัวแปรอิสระไม่มีความสัมพันธ์กัน	40	BIAS	0.3189	0.5977	0.5980	0.6009
		SD	(0.1092)	(0.2033)	(0.2034)	(0.2039)
		AMH	2.7751	5.1127	5.1271	5.1372
		SD	(0.7259)	(1.3109)	(1.3136)	(1.3159)
	70	BIAS	0.2887	0.3982	0.3991	0.3999
		SD	(0.1025)	(0.1361)	(0.1363)	(0.1364)
		AMH	2.7703	3.6488	3.6450	3.6488
		SD	(0.7130)	(0.9419)	(0.9432)	(0.9437)
	90	BIAS	0.2685	0.2730	0.2764	0.2780
		SD	(0.0980)	(0.1008)	(0.1015)	(0.1018)
		AMH	2.7670	2.7779	2.7828	2.7870
		SD	(0.7044)	(0.7059)	(0.7081)	(0.7092)
	100	BIAS	0.2584	0.2021	0.2001	0.1954
		SD	(0.0957)	(0.0695)	(0.0691)	(0.0682)
		AMH	2.7654	2.1848	2.1629	2.1140
		SD	(0.7001)	(0.5765)	(0.5727)	(0.5678)
	200	BIAS	0.1919	0.1390	0.1384	0.1361
		SD	(0.0711)	(0.0376)	(0.0374)	(0.0367)
		AMH	1.4875	1.0862	1.0818	1.0641
		SD	(0.5313)	(0.2858)	(0.2849)	(0.2811)
	400	BIAS	0.0595	0.0344	0.0343	0.0340
		SD	(0.0205)	(0.0117)	(0.0117)	(0.0115)
		AMH	1.0823	0.6333	0.6315	0.6261
		SD	(0.2775)	(0.1616)	(0.1614)	(0.1602)

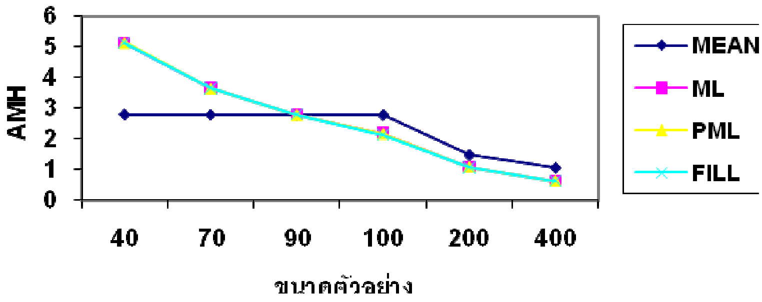
ตารางที่ 4.2.2 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 10 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.1	40	BIAS	0.3375	0.7098	0.7124	0.7125
		SD	(0.1241)	(0.2591)	(0.2596)	(0.2597)
		AMH	3.1372	5.7583	5.7787	5.7798
		SD	(0.8299)	(1.5397)	(1.5441)	(1.5444)
	70	BIAS	0.3290	0.4786	0.4810	0.4811
		SD	(0.1201)	(0.1746)	(0.1748)	(0.1749)
		AMH	2.9987	4.0960	4.1193	4.1213
		SD	(0.7888)	(1.0955)	(1.0971)	(1.0976)
	90	BIAS	0.3226	0.3233	0.3267	0.3283
		SD	(0.1174)	(0.1178)	(0.1184)	(0.1188)
		AMH	2.8709	2.9735	3.0130	3.0300
		SD	(0.7614)	(0.7962)	(0.7990)	(0.8029)
	100	BIAS	0.3204	0.2521	0.2496	0.2446
		SD	(0.1161)	(0.0907)	(0.0901)	(0.0894)
		AMH	2.7816	2.4843	2.4598	2.4122
		SD	(0.7477)	(0.6555)	(0.6500)	(0.6465)
	200	BIAS	0.2301	0.1673	0.1664	0.1646
		SD	(0.0767)	(0.0558)	(0.0555)	(0.0552)
		AMH	1.5909	1.1660	1.1604	1.1480
		SD	(0.3977)	(0.2915)	(0.2906)	(0.2887)
	400	BIAS	0.0970	0.0560	0.0557	0.0555
		SD	(0.0337)	(0.0194)	(0.0193)	(0.0192)
		AMH	1.3068	0.7642	0.7606	0.7577
		SD	(0.3368)	(0.2022)	(0.2015)	(0.2008)

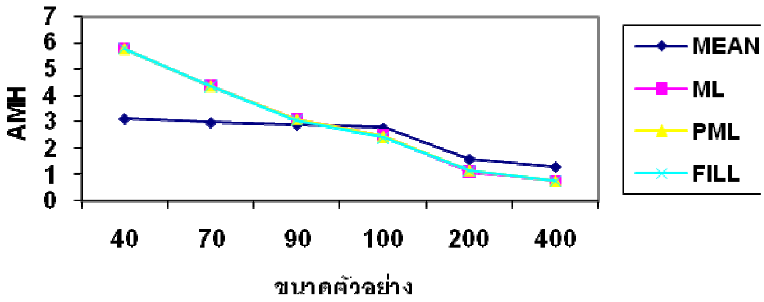
ตารางที่ 4.2.2 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 10 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.2	40	BIAS	0.3421	0.7528	0.7550	0.7562
		SD	(0.1222)	(0.2561)	(0.2566)	(0.2570)
		AMH	3.1289	6.2695	6.2873	6.2973
		SD	(0.7881)	(1.6076)	(1.6101)	(1.6137)
	70	BIAS	0.3307	0.5005	0.5021	0.5023
		SD	(0.1202)	(0.1725)	(0.1731)	(0.1732)
		AMH	2.9987	4.3561	4.3735	4.3739
		SD	(0.7673)	(1.1148)	(1.1178)	(1.1183)
	90	BIAS	0.3231	0.3300	0.3337	0.3349
		SD	(0.1161)	(0.1175)	(0.1178)	(0.1189)
		AMH	3.0494	3.0620	3.0976	3.1102
		SD	(0.7534)	(0.7822)	(0.7895)	(0.7921)
	100	BIAS	0.3193	0.2513	0.2494	0.2447
		SD	(0.1182)	(0.0901)	(0.0897)	(0.0879)
		AMH	2.8684	2.4783	2.4597	2.4149
		SD	(0.7465)	(0.6290)	(0.6254)	(0.6159)
	200	BIAS	0.2358	0.1709	0.1700	0.1694
		SD	(0.0833)	(0.0583)	(0.0580)	(0.0578)
		AMH	2.2684	1.6584	1.6492	1.6438
		SD	(0.5685)	(0.4146)	(0.4123)	(0.4110)
	400	BIAS	0.1352	0.0787	0.0785	0.0782
		SD	(0.1118)	(0.0274)	(0.0273)	(0.0272)
		AMH	1.6744	0.9864	0.9845	0.9807
		SD	(0.4371)	(0.2460)	(0.2456)	(0.2449)

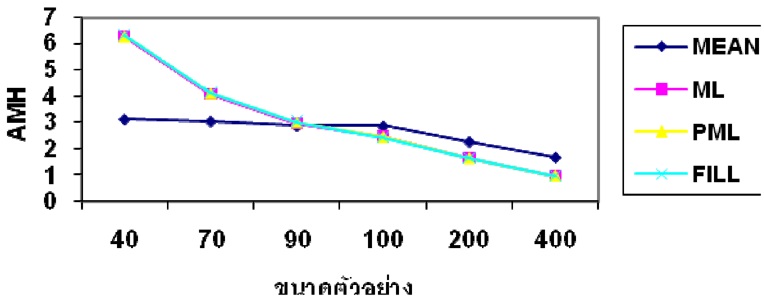
ตัวแปรอิสระไม่มีความสัมพันธ์กัน



สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.1



สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.2



<u>รูปที่ 4.1.5</u> แสดงการเปรียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ ระหว่างค่า AMH และขนาดตัวอย่าง กรณีที่ β_0, β_1 = 0.2 และ β_2 = 1 เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 10

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 10 กรณีที่ ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1 พบว่า

<u>ตัวแปรอิสระไม่มีความสัมพันธ์กัน</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

<u>สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.1</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40 , 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

<u>สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.2</u> ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สรุปผล การที่สหสัมพันธ์ระหว่างตัวแปรอิสระมีค่าเพิ่มขึ้น นั่นคือตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น มีผลทำให้ค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น โดยเมื่อขนาดตัวอย่างน้อยกว่า 90 วิธี MEAN จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และในกรณีขนาดตัวอย่างมากกว่า 90 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE , MLE และ MEAN ตามลำดับ

ตารางที่ 4.2.3 แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือ ร้อยละ 15 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

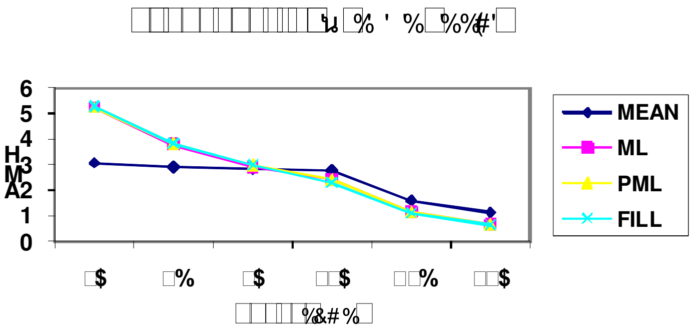
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
ตัวแปรอิสระไม่มีความสัมพันธ์กัน	40	BIAS	0.3282	0.6320	0.6322	0.6353
		SD	(0.1128)	(0.2135)	(0.2136)	(0.2145)
		AMH	3.0712	5.2546	5.2561	5.2812
		SD	(0.7679)	(1.3269)	(1.3272)	(1.3318)
	70	BIAS	0.3208	0.4340	0.4380	0.4387
		SD	(0.1121)	(0.2731)	(0.2747)	(0.2749)
		AMH	2.9199	3.7901	3.8302	3.8375
		SD	(0.7371)	(0.9596)	(0.9654)	(0.9668)
	90	BIAS	0.3158	0.3198	0.3265	0.3298
		SD	(0.1124)	(0.3122)	(0.3154)	(0.3158)
		AMH	2.8190	2.8960	2.9795	2.9928
		SD	(0.7166)	(0.7214)	(0.7242)	(0.7267)
	100	BIAS	0.3133	0.2454	0.2437	0.2327
		SD	(0.1123)	(0.3362)	(0.3358)	(0.3317)
		AMH	2.7685	2.4204	2.4042	2.2990
		SD	(0.7063)	(0.6066)	(0.6036)	(0.5873)
	200	BIAS	0.2366	0.1723	0.1699	0.1641
		SD	(0.0822)	(0.0592)	(0.0588)	(0.0579)
		AMH	1.5929	1.1690	1.1533	1.1151
		SD	(0.4064)	(0.3068)	(0.3035)	(0.2954)
	400	BIAS	0.0696	0.0403	0.0402	0.0393
		SD	(0.0237)	(0.0148)	(0.0147)	(0.0144)
		AMH	1.1338	0.6649	0.6631	0.6487
		SD	(0.2992)	(0.1696)	(0.1693)	(0.1663)

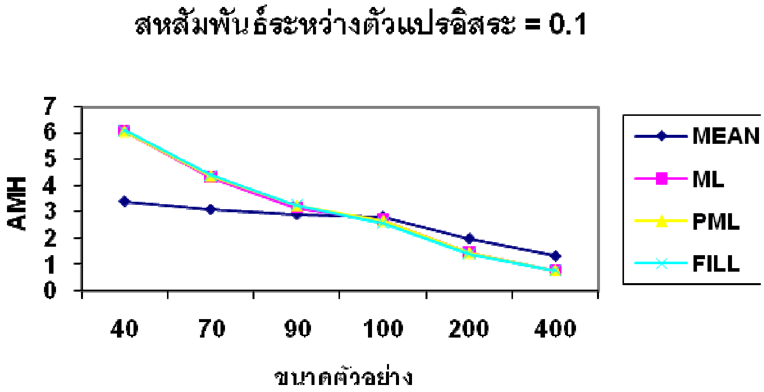
ตารางที่ 4.2.3 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 15 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

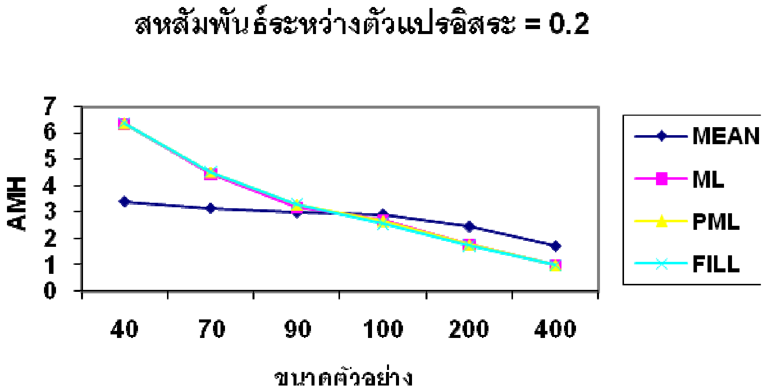
	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.1	40	BIAS	0.3381	0.7401	0.7405	0.7409
		SD	(0.1131)	(0.2588)	(0.2589)	(0.2590)
		AMH	3.4016	6.0758	6.0791	6.0825
		SD	(0.8590)	(1.5343)	(1.5349)	(1.5356)
	70	BIAS	0.3500	0.5063	0.5115	0.5128
		SD	(0.1181)	(0.1792)	(0.1803)	(0.1806)
		AMH	3.1116	4.3277	4.3755	4.3870
		SD	(0.7903)	(1.1122)	(1.1210)	(1.1231)
	90	BIAS	0.3579	0.3599	0.3612	0.3688
		SD	(0.1214)	(0.1260)	(0.1279)	(0.1284)
		AMH	2.9182	3.1578	3.2398	3.2611
		SD	(0.7475)	(0.8299)	(0.8450)	(0.8490)
	100	BIAS	0.3619	0.2854	0.2825	0.2717
		SD	(0.1231)	(0.1023)	(0.1017)	(0.0994)
		AMH	2.8215	2.6982	2.6719	2.5728
		SD	(0.7216)	(0.7119)	(0.7070)	(0.6888)
	200	BIAS	0.2399	0.1729	0.1717	0.1680
		SD	(0.0827)	(0.0576)	(0.0573)	(0.0564)
		AMH	2.0075	1.4589	1.4495	1.4189
		SD	(0.5019)	(0.3647)	(0.3626)	(0.3557)
	400	BIAS	0.1037	0.0602	0.0599	0.0593
		SD	(0.0346)	(0.0223)	(0.0222)	(0.0220)
		AMH	1.3486	0.7930	0.7882	0.7810
		SD	(0.3549)	(0.2087)	(0.2078)	(0.2061)

ตารางที่ 4.2.3 (ต่อ) แสดงผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้น โดยพิจารณาจาก ค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง (BIAS) และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) เมื่อระดับการสูญหายของตัวแปรอิสระคือร้อยละ 15 ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1

	ขนาดตัวอย่าง		MEAN	MLE	PMLE	FILL
สหสัมพันธ์ของตัวแปรอิสระ = 0.2	40	BIAS	0.3525	0.7801	0.7815	0.7823
		SD	(0.1216)	(0.2690)	(0.2694)	(0.2696)
		AMH	3.3914	6.3335	6.3446	6.3509
		SD	(0.9020)	(1.6667)	(1.6696)	(1.6712)
	70	BIAS	0.3574	0.5275	0.5324	0.5329
		SD	(0.1279)	(0.1848)	(0.1868)	(0.1870)
		AMH	3.1506	4.4609	4.5069	4.5120
		SD	(0.8241)	(1.1773)	(1.1833)	(1.1846)
	90	BIAS	0.3576	0.3607	0.3663	0.3681
		SD	(0.1282)	(0.1317)	(0.1321)	(0.1322)
		AMH	2.9901	3.2009	3.2817	3.2977
		SD	(0.7721)	(0.8480)	(0.8590)	(0.8632)
	100	BIAS	0.3623	0.2857	0.2833	0.2726
		SD	(0.1342)	(0.1047)	(0.1042)	(0.0999)
		AMH	2.9098	2.6905	2.6691	2.5709
		SD	(0.7461)	(0.7025)	(0.6969)	(0.6834)
	200	BIAS	0.2523	0.1819	0.1813	0.1782
		SD	(0.0888)	(0.0621)	(0.0619)	(0.0609)
		AMH	2.4474	1.7789	1.7735	1.7441
		SD	(0.6134)	(0.4538)	(0.4529)	(0.4449)
	400	BIAS	0.1445	0.0843	0.0838	0.0830
		SD	(0.0490)	(0.0288)	(0.0286)	(0.0284)
		AMH	1.7181	1.0143	1.0087	0.9997
		SD	(0.4394)	(0.2581)	(0.2568)	(0.2546)







<u>รูปที่ 4.1.6</u> แสดงการเปรียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ ระหว่างค่า AMH และขนาดตัวอย่าง กรณีที่ β_0, β_1 = 0.2 และ β_2 = 1 เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระร้อยละ 15

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ ร้อยละ 15 กรณีที่ ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1 พบว่า

ตัวแปรอิสระไม่มีความสัมพันธ์กัน ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธี ในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.1 ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40 , 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สหสัมพันธ์ระหว่างตัวแปรอิสระ = 0.2 ค่า BIAS และค่า AMH ของวิธี MEAN ต่ำกว่าทุกวิธีในขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และกรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 ค่า BIAS และค่า AMH ของวิธี FILL ต่ำกว่าทุกวิธี รองลงมาคือวิธี PMLE , MLE และ MEAN โดยเมื่อขนาดตัวอย่าง เพิ่มขึ้น ค่า BIAS และค่า AMH ของวิธี MEAN ,วิธี MLE ,วิธี PMLE และวิธี FILL มีแนวโน้มลดลง

สรุปผล การที่สหสัมพันธ์ระหว่างตัวแปรอิสระมีค่าเพิ่มขึ้น นั่นคือตัวแปรอิสระมีความสัมพันธ์กันมากขึ้น มีผลทำให้ค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น โดยเมื่อขนาดตัวอย่างน้อยกว่า 90 วิธี MEAN จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และในกรณีขนาดตัวอย่างมากกว่า 90 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE , MLE และ MEAN ตามลำดับ

สรุปผลตอนที่ 4.2

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ กรณีที่ ค่าพารามิเตอร์เริ่มต้น β_0, β_1 = 0.2 และ β_2 = 1 พบว่า วิธี MEAN จะเป็นวิธีที่ดีที่สุด ในกรณีที่ขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือวิธี MLE , PMLE และ FILL ตามลำดับ และ กรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE , MLE และ MEAN ตามลำดับ ในทุกค่าสหสัมพันธ์และทุกร้อยละการสูญหายของตัวแปรอิสระ

สามารถสรุปผลเกี่ยวกับปัจจัยที่มีผลต่อค่า BIAS และ AMH ได้ดังนี้

1. <u>ขนาดตัวอย่าง</u>

เมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH ลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ลง

2. <u>ค่าสหสัมพันธ์</u>

เมื่อค่าสหสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น ซึ่งจะส่งผลให้ความถูกต้องในการประมาณค่าพารามิเตอร์ลดลง

3. <u>ร้อยละการสูญหายของตัวแปรอิสระ</u>

เมื่อร้อยละการสูญหายของตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น เนื่องจากเมื่อร้อยละการสูญหายเพิ่มขึ้นจะส่งผลให้ข้อมูลลดลง ทำให้ค่าความคลาดเคลื่อนที่เกิดขึ้นจากการประมาณค่าพารามิเตอร์เพิ่มขึ้น

สรุปผลตอนที่ 4.1 – 4.2

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ พบว่า วิธี MEAN จะเป็นวิธีที่ดีที่สุด ในกรณีที่ขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือ วิธี MLE , PMLE และ FILL ตามลำดับ และ กรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE ,MLEและ MEAN ตามลำดับ ในทุกค่าสหสัมพันธ์ของตัวแปรอิสระ ทุกค่าพารามิเตอร์เริ่มต้นและทุกร้อยละการสูญหายของตัวแปรอิสระ

สามารถสรุปผลเกี่ยวกับปัจจัยที่มีผลต่อค่า BIAS และ AMH ได้ดังนี้

1. <u>ขนาดตัวอย่าง</u>

เมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH ลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ลง

2. <u>ค่าสหสัมพันธ์</u>

เมื่อค่าสหสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น ซึ่งจะส่งผลให้ความถูกต้องในการประมาณค่าพารามิเตอร์ลดลง

3. <u>ร้อยละการสูญหายของตัวแปรอิสระ</u>

เมื่อร้อยละการสูญหายของตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น เนื่องจากเมื่อร้อยละการสูญหายเพิ่มขึ้นจะส่งผลให้ข้อมูลลดลง ทำให้ค่าความคลาดเคลื่อนที่เกิดขึ้นจากการประมาณค่าพารามิเตอร์เพิ่มขึ้น

4. <u>ค่าพารามิเตอร์เริ่มต้น</u>

เมื่อค่าพารามิเตอร์ที่กำหนดตอนเริ่มต้นของตัวแปรอิสระตัวที่เกิดข้อมูลสูญหายเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น เนื่องจากพารามิเตอร์เริ่มต้นที่เพิ่มขึ้นทำให้การถ่วงน้ำหนักในตัวแปรอิสระเพิ่มขึ้น ซึ่งส่งผลให้เกิดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์เพิ่มขึ้นด้วย

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเปรียบเทียบประสิทธิภาพของตัวประมาณของตัวแบบความถดถอยโลจิสติกแบบ 2 กลุ่ม ที่ประมาณค่าสูญหายด้วยวิธี Mean Imputation (MEAN) วิธี Maximum Likelihood Estimation (MLE) วิธี Pseudo Maximum Likelihood Estimation (PMLE) และวิธี The Filling Method (FILL) เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระในกรณีที่มีตัวแปรอิสระ 2 ตัว ภายใต้สถานการณ์ที่แตกต่างกัน ซึ่งศึกษาในกรณีต่าง ๆ ดังนี้

1. การสูญหายเกิดขึ้นกับตัวแปรอิสระตัวที่ 1 หรือตัวแปรอิสระที่ 2 ตัวใดตัวหนึ่ง โดยการสูญหายเป็นการสูญหายแบบสุ่ม
2. ลักษณะของตัวแปรอิสระที่นำมาศึกษามีรูปแบบการแจกแจงปกติ (Normal Distribution)
3. กำหนดค่าพารามิเตอร์เริ่มต้น โดยแบ่งเป็น 2 กรณี คือ $\beta_0, \beta_1, \beta_2$ = 0.2 และ β_0, β_1 = 0.2 , β_2 = 1
4. ขนาดตัวอย่างที่ใช้ในการศึกษาเท่ากับ 40 ,70 , 90 , 100 , 200 และ 400
5. ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ คือ 0 , 0.1 และ 0.2
6. ข้อมูลตัวแปรอิสระสูญหาย 3 ระดับ คือ ร้อยละ 5 , 10 และ 15
7. การวิจัยครั้งนี้ทำการจำลองข้อมูลให้มีสถานการณ์ที่แตกต่างกันตามข้อกำหนดข้างต้น โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล (Monte carlo Simulation Technique) ทำการจำลองในแต่ละสถานการณ์ 1,000รอบ

เกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการประมาณค่าพารามิเตอร์วิธีใดเป็นวิธีการที่ดีที่สุด พิจารณาจากค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง BIAS และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH) วิธีใดที่ให้ค่า BIAS และค่า AMH ต่ำที่สุดจะเป็นวิธีที่ดีที่สุด

5.1 สรุปผลการวิจัย

5.1.1 ผลการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์ของตัวแปรตามกับค่าจริง BIAS และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (AMH)

ผลการเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในตัวแปรอิสระ พบว่า วิธี MEAN จะเป็นวิธีที่ดีที่สุด ในกรณีที่ขนาดตัวอย่างเท่ากับ 40, 70 และ 90 รองลงมาคือ วิธี MLE , PMLE และ FILL ตามลำดับ และ กรณีขนาดตัวอย่างเท่ากับ 100, 200 และ 400 วิธี FILL จะเป็นวิธีที่ดีที่สุด รองลงมาคือวิธี PMLE ,MLEและ MEAN ตามลำดับ เมื่อขนาดตัวอย่างมากกว่า 90 ในระดับสหสัมพันธ์ระหว่างตัวแปรอิสระ คือ 0 , 0.1 และ 0.2 ค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2 และ β_0, β_1 = 0.2 , β_2 = 1 และร้อยละการสูญหายของตัวแปรอิสระไม่เกินร้อยละ 15

5.1.2 พิจารณาปัจจัยที่มีผลต่อค่า BIAS และค่า AMH

1. <u>ขนาดตัวอย่าง</u>

เมื่อขนาดตัวอย่างเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH ลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นจะทำให้มีข้อมูลในการวิเคราะห์มากขึ้น จึงส่งผลให้ลดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ลงได้

2. <u>ค่าสหสัมพันธ์</u>

เมื่อค่าสหสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น เนื่องจากเมื่อตัวแปรอิสระมีความสัมพันธ์กันจะส่งผลให้ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์สูงขึ้น ทำให้ความถูกต้องในการประมาณค่าพารามิเตอร์ลดลง

3. <u>ร้อยละการสูญหายของตัวแปรอิสระ</u>

เมื่อร้อยละการสูญหายของตัวแปรอิสระเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น เนื่องจากเมื่อร้อยละการสูญหายเพิ่มขึ้นจะส่งผลให้ข้อมูลที่มีอยู่ลดลง ทำให้ค่าความคลาดเคลื่อนที่เกิดขึ้นจากการประมาณค่าพารามิเตอร์เพิ่มขึ้น

4. <u>ค่าพารามิเตอร์เริ่มต้น</u>

เมื่อค่าพารามิเตอร์ที่กำหนดตอนเริ่มต้นของตัวแปรอิสระตัวที่เกิดข้อมูลสูญหายเพิ่มขึ้น จะทำให้ค่า BIAS และ AMH เพิ่มขึ้น เนื่องจากพารามิเตอร์เริ่มต้นที่เพิ่มขึ้นทำให้การถ่วงน้ำหนักในตัวแปรอิสระเพิ่มขึ้น ซึ่งส่งผลให้เกิดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์เพิ่มขึ้นด้วย

5.1.3 ผลสรุปการเลือกวิธีการประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติกเมื่อเกิดข้อมูลสูญหายในตัวแปรอิสระ

ในการวิจัยครั้งนี้พบว่า ในกรณีที่ขนาดตัวอย่างเท่ากับ 40 ,70 และ 90 วิธีการประมาณค่าพารามิเตอร์ด้วยวิธี MEAN จะเป็นวิธีที่ดีที่สุด เนื่องจากวิธีนี้จะใช้ค่าเฉลี่ยของข้อมูลที่ไม่ได้สูญหายมาแทนในชุดข้อมูลที่สูญหาย โดยตัวแปรอิสระที่ใช้ในการวิจัยนี้มีการแจกแจงแบบปกติ ซึ่งมีการกระจายของข้อมูลที่ใกล้เคียงกัน ดังนั้นค่าเฉลี่ยที่คำนวณได้จึงสามารถเป็นตัวแทนของข้อมูลทั้งหมดได้ แต่ในกรณีที่ขนาดตัวอย่างมากกว่า 90 วิธี FILL จะเป็นวิธีที่มีประสิทธิภาพที่สุด เนื่องจากวิธี FILL เป็นวิธีที่มีการปรับปรุงมาจากวิธี MLE และขนาดตัวอย่างที่มีจำนวนมากพอ จึงทำให้วิธี MLE เป็นวิธีประมาณพารามิเตอร์ที่มีประสิทธิภาพ และทำให้วิธี FILL มีประสิทธิภาพด้วย โดยทั้งวิธี MEAN และวิธี FILL จะให้ผลดีเมื่อขนาดตัวอย่างมากกว่า 90 ในระดับสหสัมพันธ์ระหว่างตัวแปรอิสระ คือ 0 , 0.1 และ 0.2 ในค่าพารามิเตอร์เริ่มต้น $\beta_0, \beta_1, \beta_2$ = 0.2 และ β_0, β_1 = 0.2 , β_2 = 1 และร้อยละการสูญหายของตัวแปรอิสระไม่เกินร้อยละ 15 นอกจากนี้วิธี FILL ยังให้ผลในการประมาณพารามิเตอร์ที่ไม่แตกต่างกันกับวิธี MLE และวิธี PMLE เนื่องจากการทดสอบความแตกต่างของค่าเฉลี่ยของทั้งสามวิธีคือ วิธี MLE วิธี PMLE และวิธี FILL พบว่าทั้งสามวิธีให้ค่าเฉลี่ยที่ไม่แตกต่างกัน

5.2 ข้อเสนอแนะ

5.2.1 ด้านการนำไปใช้

5.2.1.1 จากผลการวิจัยจะเห็นได้ว่าวิธีการประมาณค่าพารามิเตอร์ด้วยวิธี MEAN จะเป็นวิธีที่ดีที่สุด ในกรณีที่ขนาดตัวอย่าง 40, 70 และ 90 แต่ในกรณีที่ขนาดตัวอย่างมากขึ้น วิธี FILL จะเป็นวิธีที่มีประสิทธิภาพที่สุดซึ่งมีค่าใกล้เคียงกันมากกับวิธี MLE และ วิธี PMLE ดังนั้นใน

กรณีที่มีขนาดตัวอย่างอยู่ระหว่าง 40 ถึง 90 และร้อยละการสูญหายของตัวแปรอิสระไม่เกินร้อยละ 15 ผู้ใช้งานควรเลือกวิธี MEAN มาใช้งานเนื่องจากเป็นวิธีที่คุ้นเคยและมีวิธีการที่ไม่ซับซ้อน แต่ในกรณีที่ขนาดตัวอย่างมีขนาดมากขึ้นตั้งแต่ 90 ขึ้นไป ผู้ใช้งานอาจเลือกใช้วิธี MLE ในการประมาณค่าพารามิเตอร์แทนวิธี FILL เนื่องจากเป็นวิธีที่ใช้กันอย่างแพร่หลายและผู้ใช้งานคุ้นเคยเป็นอย่างดี

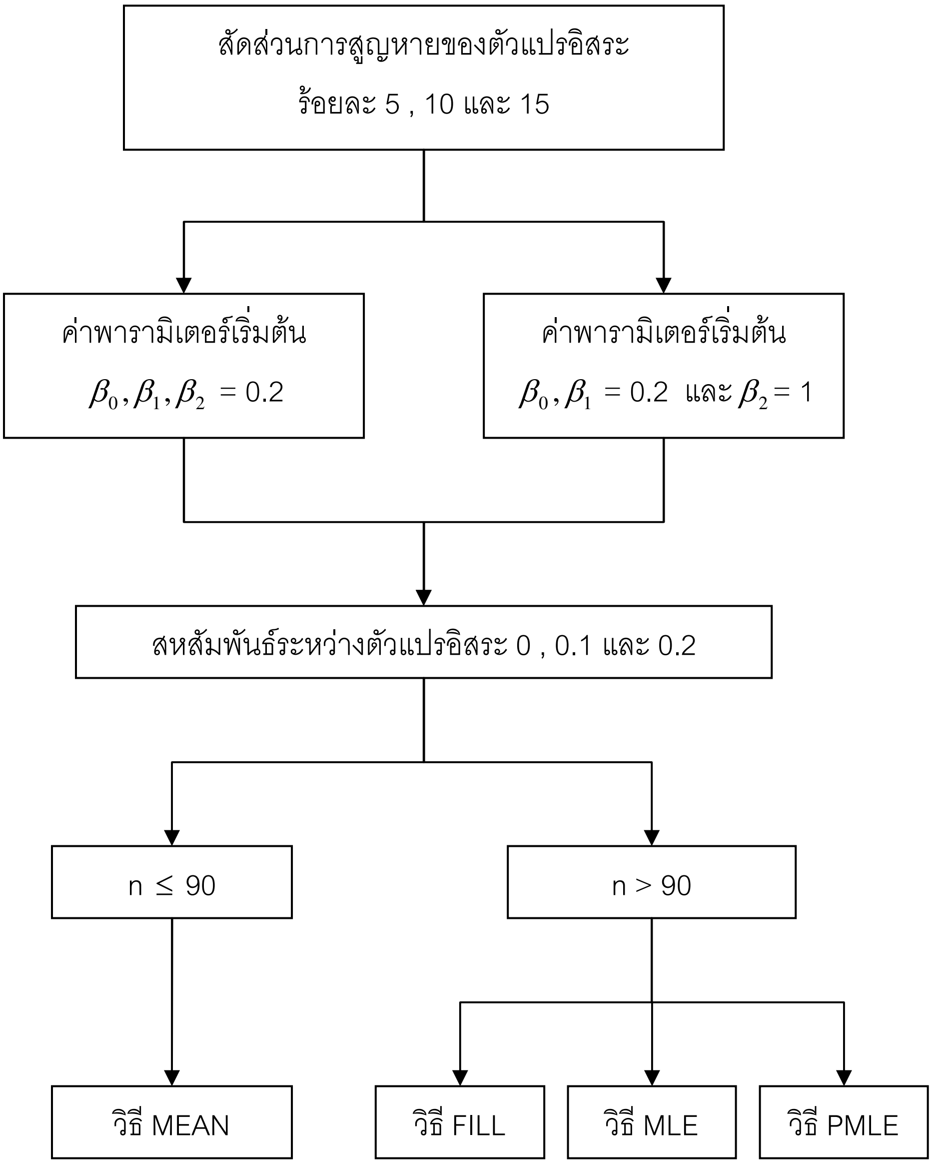
5.2.1.2 การวิจัยครั้งนี้ผู้วิจัยทำการจำลองข้อมูล โดยใช้เทคนิคการจำลองแบบมอนติคาร์โล ทำการจำลองในแต่ละสถานการณ์ 1,000 รอบ ซึ่งในการนำไปใช้งานจริง ผู้วิจัยไม่จำเป็นต้องจำลองครบ 1,000 รอบ ถ้าผลลัพธ์ที่ได้จากการจำลองมีการลู่เข้าสู่ค่าใดค่าหนึ่ง ซึ่งทำให้ประหยัดเวลาในการประมวลผลข้อมูลได้

5.2.2 ด้านการศึกษาวิจัย

5.2.2.1 สำหรับการวิจัยครั้งนี้ผู้วิจัยได้ศึกษาเฉพาะกรณีที่ตัวแปรอิสระมีการแจกแจงแบบปกติ สำหรับการวิจัยในครั้งต่อไปอาจทำการศึกษากรณีที่ตัวแปรอิสระมีการแจกแจงในรูปแบบอื่นที่ไม่เบ้มากเพื่อพิจารณาว่ารูปแบบการแจกแจงมีผลต่อความคลาดเคลื่อนมากน้อยเพียงใด

5.2.2.2 ในงานวิจัยครั้งต่อไปอาจจะทำการศึกษาโดยการเพิ่มสถานการณ์ต่าง ๆ เพิ่มขึ้น เช่น เพิ่มระดับการสูญหายของตัวแปรอิสระ เพิ่มจำนวนตัวแปรอิสระ เช่นกรณีที่มีตัวแปรอิสระ 3 ตัว ตัวแปรอิสระมีความสัมพันธ์กัน 2 ตัว หรือมีความสัมพันธ์กันทั้งสามตัว เป็นต้น

<u>รูปที่ 5.1</u> แผนผังแสดงการเลือกวิธีการประมาณค่าพารามิเตอร์ในตัวแบบความถดถอยโลจิสติกเมื่อมีค่าสูญหายในตัวแปรอิสระ



รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. 2548.<u>การวิเคราะห์ข้อมูลหลายตัวแปร</u>.กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

ทัศนาพร จงเกตุกรณ์ . 2546 . <u>การประมาณค่าพารามิเตอร์ของตัวแบบถดถอยโลจิสติกทวินาม</u>. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.

ธีระพร วีระถาวร. 2539.<u>ความน่าจะเป็น กับ การประยุกต์</u>. พิมพ์ครั้งที่ 2.กรุงเทพฯ : วิทยพัฒน์.

มนัส สังวรศิลป์ . 2543 . <u>คู่มือการใช้งาน MATLAB ฉบับสมบูรณ์</u> . พิมพ์ครั้งที่ 2. กรุงเทพฯ : อินโฟเพรส.

ภาษาอังกฤษ

Donald B. Rubin .1976. <u>Inference and Missing Data</u>. Biometrika 63, 3 :581-592.

Gail Gong; Francisco J. Samaniego .1981. <u>Pseudo Maximum Likelihood Estimation: Theory and Applications</u>. The Annals of Statistics 9, 4 : 861-869.

Vach Werner .1994. <u>Logistic Regression with Missing Values in the Covariates</u>. New York :Springer – Verlag.

ภาคผนวก

ภาคผนวก ก

โปรมแกรมที่ใช้ในงานวิจัย

ในงานวิจัยครั้งนี้ได้ใช้โปรแกรม MATLAB สำหรับหาประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นในการวิเคราะห์ความถดถอยโลจิสติก ซึ่งวิธีการประมาณค่าพารามิเตอร์เมื่อมีค่าสูญหายเกิดขึ้นที่นำมาใช้มี 4 วิธี ดังนี้

วิธีที่ 1 คือ วิธี Mean Imputation (MEAN)

วิธีที่ 2 คือ วิธี Maximum Likelihood Estimation (MLE)

วิธีที่ 3 คือ วิธี Pseudo Maximum Likelihood Estimation (PMLE)

วิธีที่ 4 คือ วิธี The Filling Method (FILL)

ใช้เกณฑ์ BIAS และ ค่าระยะห่างมาหาลาโนบิสเฉลี่ย (Average Mahalanobis Distance) เป็นเกณฑ์ประกอบในการตัดสินใจ รายละเอียดของโปรแกรมที่ใช้ในงานวิจัยมีดังนี้

กรณีขนาดตัวอย่าง 400 ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระ เท่ากับ 0.1 ร้อยละการสูญหายร้อยละ 15 และพารามิเตอร์เริ่มต้น เท่ากับ 0.2

```
n=400;
p=1;
stdevX1 = 1;
stdevX2 = 1;
varX1 = (stdevX1).^2;
varX2 = (stdevX2).^2;
CorrX1X2 = 0.1;
CovX1X2 = stdevX1*stdevX2*CorrX1X2;
SigmaX = [varX1 CovX1X2;CovX1X2 varX2 ]
CholSigmaX = chol(SigmaX)
CSigmaX = CholSigmaX'
```

```
%สร้างX1 X2 และ y ที่มีความสัมพันธ์กัน
z1 = normrnd(0,1,n,p)
z2 = normrnd(0,1,n,p)
X1 = CSigmaX(1,1)*z1
X2 = CSigmaX(2,1)*z1+CSigmaX(2,2)*z2

X=[ones([n 1]) X1 X2]
L=1000;
N=1000;

beta_start = [0.2 0.2 0.2]';
diff = 1; beta = beta_start; % initial values
beta_old = beta;
    Ldiff2=0
    Ldiff4=0
    Ldiff6=0
    Ldiff8=0
    Lma2=0
    Lma4=0
    Lma6=0
    Lma8=0

 while N > 0
pi = exp(beta(1)+beta(2)*z1+beta(3)*z2)./(1+exp(beta(1)+beta(2)*z1+beta(3)*z2))
i_round=1;
  y=(pi>=0.5);
    U(1,1)=sum(y-pi);
    U(2,1)=sum(z1.*(y-pi));
    U(3,1)=sum(z2.*(y-pi));
    U ;% scoring function
    H(1,1)=-sum(pi.*(1-pi));
    H(1,2)=-sum(z1.*pi.*(1-pi));
```

```
   H(1,3)=-sum(z2.*pi.*(1-pi));
   H(2,1)=-sum(z1.*pi.*(1-pi));
   H(2,2)=-sum(z1.^2.*pi.*(1-pi));
   H(2,3)=-sum(z1.*z2.*pi.*(1-pi));
   H(3,1)=-sum(z2.*pi.*(1-pi));
   H(3,2)=-sum(z1.*z2.*pi.*(1-pi));
   H(3,3)=-sum(z2.^2.*pi.*(1-pi));
   H;

beta = beta_old - inv(H) * U
   beta_start = beta;
   beta_mle = beta

%สมมติการสูญหาย
pmis = 15;
i_z=1;
Nmis=(n*pmis)/100;
Z_x=zeros(Nmis,1);
Z_y=zeros(Nmis,1);
Z_pi=zeros(Nmis,1);
pistill=pi;
x1still=X1;
x2still=X2;
ystill=y;
   M = randint(Nmis,1,[1,n])
   Z_pi(:,1) = pi(M,1)
   Z_x(:,1) = X2(M,1)
   Z_y(:,1) = y(M,1)
   pistill(M,:) = []
   x2still(M,:) = []
   x1still(M,:)=[]
   ystill(M,:) = []
```

```
  % วิธี Mean
X_mean=mean(x2still)
X2_method1=X2;
X2_method1(M,1)=X_mean;
X1_method1=X1;

diff = 1; beta = beta_start; % initial values
beta_1=beta_old  ;

while abs(beta_mean(:,1)- b_old(:,1))>0.00001
pi=exp(beta_1(1)+beta_1(2)*X1_method1+beta_1(2)*X2_method1)./(1+exp(beta_1(1)+beta_1
(2)*X1_method1+beta_1(2)*X2_method1));
 y=(pi>=0.5);
   U(1,1)=sum(y-pi);
   U(2,1)=sum(X1_method1.*(y-pi));
   U(3,1)=sum(X2_method1.*(y-pi));
   U % scoring function
   H(1,1)=-sum(pi.*(1-pi));
   H(1,2)=-sum(X1_method1.*pi.*(1-pi));
   H(1,3)=-sum(X2_method1.*pi.*(1-pi));
   H(2,1)=-sum(X1_method1.*pi.*(1-pi));
   H(2,2)=-sum(X1_method1.^2.*pi.*(1-pi));
   H(2,3)=-sum(X1_method1.*X2_method1.*pi.*(1-pi));
   H(3,1)=-sum(X2_method1.*pi.*(1-pi));
   H(3,2)=-sum(X1_method1.*X2_method1.*pi.*(1-pi));
   H(3,3)=-sum(X2_method1.^2.*pi.*(1-pi));
   H
beta_mean = beta_old - inv(H) * U
end
beta_mle
diff1=abs(beta_mle-beta_mean)
diff2=mean(diff1)
```

```
% วิธี ML

diff = 1; beta = beta_start; % initial values
beta_2=beta_old  ;

while abs(beta_2(:,1)- b_old(:,1))>0.00001
pi=exp(beta_2(1)+beta_2(2)*x1still+beta_2(2)*x2still)./(1+exp(beta_2(1)+beta_2(2)*x1still+bet
a_2(2)*x2still))
y=(pi>=0.5);
   U(1,1)=sum(y-pi)+log((100-pmis)/100);
   U(2,1)=sum(x1still.*(y-pi))+log((100-pmis)/100);
   U(3,1)=sum(x2still.*(y-pi))+log((100-pmis)/100);
   U % scoring function
H(1,1)=-sum(pi.*(1-pi))+(log((100-pmis)/100))^2;
   H(1,2)=-sum(x1still.*pi.*(1-pi))+(log((100-pmis)/100))^2;
   H(1,3)=-sum(x2still.*pi.*(1-pi))+(log((100-pmis)/100))^2;
   H(2,1)=-sum(x1still.*pi.*(1-pi))+(log((100-pmis)/100))^2;
   H(2,2)=-sum(x1still.^2.*pi.*(1-pi))+(log((100-pmis)/100))^2;
   H(2,3)=-sum(x1still.*x2still.*pi.*(1-pi))+(log((100-pmis)/100))^2;
   H(3,1)=-sum(x2still.*pi.*(1-pi))+(log((100-pmis)/100))^2;
   H(3,2)=-sum(x1still.*x2still.*pi.*(1-pi))+(log((100-pmis)/100))^2;
   H(3,3)=-sum(x2still.^2.*pi.*(1-pi))+(log((100-pmis)/100))^2;
   H
beta_2 = beta_old - inv(H) * U
end
b_still=beta_2(2,1)
x2new=X1(M,:)*b_still
X2_method2=X2
X2_method2(M,:)=x2new(:,:)
diff = 1; beta = beta_start; % initial values
beta_ml=beta_old  ;
```

```
while abs(beta_ml(:,1)- b_old(:,1))>0.00001
pi=exp(beta_2(1)+beta_2(2)*X1+beta_2(2)*X2_method2)./(1+exp(beta_2(1)+beta_2(2)*X1+b
eta_2(2)*X2_method2))
y=(pi>=0.5)
   U(1,1)=sum(y-pi);
   U(2,1)=sum(X1.*(y-pi));
   U(3,1)=sum(X2_method2.*(y-pi));
   U % scoring function
 H(1,1)=-sum(pi.*(1-pi));
   H(1,2)=-sum(X1.*pi.*(1-pi));
   H(1,3)=-sum(X2_method2.*pi.*(1-pi));
   H(2,1)=-sum(X1.*pi.*(1-pi));
   H(2,2)=-sum(X1.^2.*pi.*(1-pi));
   H(2,3)=-sum(X1.*X2_method2.*pi.*(1-pi));
   H(3,1)=-sum(X2_method2.*pi.*(1-pi));
   H(3,2)=-sum(X1.*X2_method2.*pi.*(1-pi));
   H(3,3)=-sum(X2_method2.^2.*pi.*(1-pi));
   H
beta_ml = beta_old - inv(H) * U
end
   beta_ml

% วิธี PML
diff = 1; beta = beta_start; % initial values
beta_3=beta_old  ;

while abs(beta_3(:,1)- b_old(:,1))>0.00001
pi=exp(beta_3(1)+beta_3(2)*x1still+beta_3(2)*x2still)./(1+exp(beta_3(1)+beta_3(2)*x1still+bet
a_3(2)*x2still))
y=(pi>=0.5);
   U(1,1)=sum(y-pi);
   U(2,1)=sum(x1still.*(y-pi));
```

```
    U(3,1)=sum(x2still.*(y-pi));
    U % scoring function
H(1,1)=-sum(pi.*(1-pi));
    H(1,2)=-sum(x1still.*pi.*(1-pi));
    H(1,3)=-sum(x2still.*pi.*(1-pi));
    H(2,1)=-sum(x1still.*pi.*(1-pi));
    H(2,2)=-sum(x1still.^2.*pi.*(1-pi));
    H(2,3)=-sum(x1still.*x2still.*pi.*(1-pi));
    H(3,1)=-sum(x2still.*pi.*(1-pi));
    H(3,2)=-sum(x1still.*x2still.*pi.*(1-pi));
    H(3,3)=-sum(x2still.^2.*pi.*(1-pi));
    H
beta_3 = beta_old - inv(H) * U
end
b_still=beta_3(2,1)
x2new3=X1(M,:)*b_still
X2_method3=X2
X2_method3(M,:)=x2new3(:,:)
diff = 1; beta = beta_start; % initial values
beta_pml=beta_old  ;

while abs(beta_pml(:,1)- b_old(:,1))>0.00001
pi=exp(beta_3(1)+beta_3(2)*X1+beta_3(2)*X2_method3)./(1+exp(beta_3(1)+beta_3(2)*X1+b
eta_3(2)*X2_method3))
y=(pi>=0.5)
    U(1,1)=sum(y-pi);
    U(2,1)=sum(X1.*(y-pi));
    U(3,1)=sum(X2_method3.*(y-pi));
    U % scoring function
    H(1,1)=-sum(pi.*(1-pi));
    H(1,2)=-sum(X1.*pi.*(1-pi));
    H(1,3)=-sum(X2_method3.*pi.*(1-pi));
```

```
   H(2,1)=-sum(X1.*pi.*(1-pi));
   H(2,2)=-sum(X1.^2.*pi.*(1-pi));
   H(2,3)=-sum(X1.*X2_method3.*pi.*(1-pi));
   H(3,1)=-sum(X2_method3.*pi.*(1-pi));
   H(3,2)=-sum(X1.*X2_method3.*pi.*(1-pi));
   H(3,3)=-sum(X2_method3.^2.*pi.*(1-pi));
   H
beta_pml = beta_old - inv(H) * U
end
   beta_pml
   beta_mle

% วิธี FILL
diff = 1; beta = beta_start; % initial values
beta_4=beta_old  ;

while abs(beta_4(:,1)- b_old(:,1))>0.00001
pi=exp(beta_4(1)+beta_4(2)*x1still+beta_4(2)*x2still)./(1+exp(beta_4(1)+beta_4(2)*x1still+bet
a_4(2)*x2still))
y=(pi>=0.5);
   U(1,1)=sum(y-pi);
   U(2,1)=sum(x1still.*(y-pi));
   U(3,1)=sum(x2still.*(y-pi));
   U % scoring function
H(1,1)=-sum(pi.*(1-pi));
   H(1,2)=-sum(x1still.*pi.*(1-pi));
   H(1,3)=-sum(x2still.*pi.*(1-pi));
   H(2,1)=-sum(x1still.*pi.*(1-pi));
   H(2,2)=-sum(x1still.^2.*pi.*(1-pi));
   H(2,3)=-sum(x1still.*x2still.*pi.*(1-pi));
   H(3,1)=-sum(x2still.*pi.*(1-pi));
   H(3,2)=-sum(x1still.*x2still.*pi.*(1-pi));
```

```
   H(3,3)=-sum(x2still.^2.*pi.*(1-pi));
   H
beta_4 = beta_old - inv(H) * U
end
b_still=beta_4(2,1)
x2new4=X1(M,:)*b_still
X2_method4=X2
X2_method4(M,:)=x2new4(:,:)*(100-pmis)/100

diff = 1; beta = beta_start; % initial values
beta_fill=beta_old  ;

while abs(beta_fill(:,1)- b_old(:,1))>0.00001
pi=exp(beta_4(1)+beta_4(2)*X1+beta_4(2)*X2_method4)./(1+exp(beta_4(1)+beta_4(2)*X1+b
eta_4(2)*X2_method4))
y=(pi>=0.5)
   U(1,1)=sum(y-pi);
   U(2,1)=sum(X1.*(y-pi));
   U(3,1)=sum(X2_method4.*(y-pi));
   U % scoring function
 H(1,1)=-sum(pi.*(1-pi));
   H(1,2)=-sum(X1.*pi.*(1-pi));
   H(1,3)=-sum(X2_method4.*pi.*(1-pi));
   H(2,1)=-sum(X1.*pi.*(1-pi));
   H(2,2)=-sum(X1.^2.*pi.*(1-pi));
   H(2,3)=-sum(X1.*X2_method4.*pi.*(1-pi));
   H(3,1)=-sum(X2_method4.*pi.*(1-pi));
   H(3,2)=-sum(X1.*X2_method4.*pi.*(1-pi));
   H(3,3)=-sum(X2_method4.^2.*pi.*(1-pi));
   H
beta_fill = beta_old - inv(H) * U
end
```

```
  beta_fill
  beta_mle

diff1=abs(beta_mle-beta_mean)
diff3=abs(beta_mle-beta_ml)
diff5=abs(beta_mle-beta_pml)
diff7=abs(beta_mle-beta_fill)
(exp(X1*b_still))./(1+exp(X1*b_still));

diff2=mean(diff1)
diff4=mean(diff3)
diff6=mean(diff5)
diff8=mean(diff7)

ma1=sqrt((diff2)'*inv(cov(beta_mean))*(diff2))
ma3=sqrt((diff4)'*inv(cov(beta_ml))*(diff4))
ma5=sqrt((diff6)'*inv(cov(beta_pml))*(diff6))
ma7=sqrt((diff8)'*inv(cov(beta_fill))*(diff8))
ma2=mean(ma1)
ma4=mean(ma3)
ma6=mean(ma5)
ma8=mean(ma7)

Adiff2(N) = diff2,
Adiff4(N) = diff4,
Adiff6(N) = diff6,
Adiff8(N) = diff8,

Ama2(N)=ma2
Ama4(N)=ma4
Ama6(N)=ma6
Ama8(N)=ma8
```

```
 if Adiff2(N) < 1000 & Adiff4(N) < 1000 & Adiff6(N) < 1000 & Adiff8(N) < 1000 & Ama2(N) <
1000 & Ama4(N) < 1000 & Ama6(N) < 1000 & Ama8(N) < 1000

   Ldiff2 = Ldiff2 + diff2;
   Ldiff4 = Ldiff4 + diff4;
   Ldiff6 = Ldiff6 + diff6;
   Ldiff8 = Ldiff8 + diff8;
   Lma2 = Lma2 + ma2
   Lma4 = Lma4 + ma4
   Lma6 = Lma6 + ma6
   Lma8 = Lma8 + ma8
   N=N-1
end
end
Mdiff2 = Ldiff2/L
Mdiff4 = Ldiff4/L
Mdiff6 = Ldiff6/L
Mdiff8 = Ldiff8/L
Mma2 = Lma2/L
Mma4 = Lma4/L
Mma6 = Lma6/L
Mma8 = Lma8/L

A=[Mdiff2 Mdiff4 Mdiff6 Mdiff8]'  B=[Mma2 Mma4 Mma6 Mma8]'

sAdiff2 = std(Adiff2)
sAdiff4 = std(Adiff4)
sAdiff6 = std(Adiff6)
sAdiff8 = std(Adiff8)
sAma2 = std(Ama2)
sAma4 = std(Ama4)
sAma6 = std(Ama6)
sAma8 = std(Ama8)
```

ประวัติผู้เขียนวิทยานิพนธ์

นายประลองพล ประสงค์พร เกิดวันเสาร์ที่ 7 เมษายน พ.ศ. 2527 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ในปีการศึกษา 2548 และสำเร็จการศึกษาปริญญาเศรษฐศาสตร์บัณฑิต (การพัฒนา) คณะเศรษฐศาสตร์ มหาวิทยาลัยรามคำแหง ในปีการศึกษา 2549 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต (สต.ม.) สาขาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปี พ.ศ. 2549