

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  $\beta_0, \beta_1 = 0.2$  และ  $\beta_2 = 1$  พิจารณาเปรียบเทียบค่าเฉลี่ยของความแตกต่างระหว่างค่าจริงและค่าประมาณ และระยะห่างมาหาลาโนบิสเจเลีย (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 มีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระเพิ่มขึ้นเนื่องจากเมื่อร้อยละการสูญหายเพิ่มขึ้น จะส่งผลให้ข้อมูลที่มีอยู่ลดลง ทำให้ค่าความคลาดเคลื่อนเพิ่มขึ้น เมื่อค่าพารามิเตอร์เริ่มต้นของตัวแปรอิสระเพิ่มขึ้นค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น และเมื่อระดับความสัมพันธ์ระหว่างตัวแปรอิสระเพิ่มขึ้นค่า BIAS และค่า AMH มีแนวโน้มเพิ่มขึ้น

##4982194526 : MAJOR STATISTICS

KEYWORDS : LOGISTIC 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  $\beta_0, \beta_1 = 0.2, \beta_2 = 1$ . 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.