201447

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณเมื่อเกิดปัญหาข้อมูลสูญหายในตัวแปรอิสระ โดยจะเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณ 5 วิธี ได้แก่ วิธีข้อมูลสมบูรณ์(Complete Case method (CC)) วิธีการถดถอยอันดับที่หนึ่ง-วิธีผสมแบบถ่วงน้ำหนัก (First Order Regression method - Weight Mixed Estimator method(FOR-W)) วิธีการถดถอยอันดับที่หนึ่ง-วิธีผสมแบบสไตน์ (First Order Regression method - Stein Rule Mixed Estimator method (FOR-SR)) วิธีประยุกต์การถดถอยอันดับที่หนึ่ง-วิธีผสมแบบถ่วงน้ำหนัก (Modified First Order Regression method - Weight Mixed Estimator method (MFOR-W)) และวิธีประยุกต์การถดถอยอันดับที่หนึ่ง-วิธีผสมแบบสไตน์ (Modified First Order Regression method - Stein Rule Mixed Estimator method (MFOR-SR)) เกณฑ์ที่ใช้ในการตัดสินใจคือค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Average Root Mean Squares Error (ARMSE)) และส่วนที่ใช้ประกอบในการเปรียบเทียบประสิทธิภาพของวิธีการทั้ง 3 วิธีคือ ค่าอัตราส่วนผลต่างของค่าเฉลี่ยรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Ratio of Different Average Root Mean Squares Error (DIFF)) สถานการณ์ที่ศึกษาจะใช้จำนวนตัวแปรอิสระ(p) เท่ากับ 3 ขนาดตัวอย่างเท่ากับ 5p 10p 15p และ 20p ระดับความสัมพันธ์ระหว่างตัวแปรอิสระใช้ 3 ระดับคือ ระดับต่ำ (ρ = 0.1 0.2 0.3) ระดับปานกลาง (ρ =0.4 0.5 0.6) และระดับสูง (ρ =0.7 0.7 0.8) สัดส่วนข้อมูลสูญหายในตัวแปรอิสระใช้ 3 ระดับคือ 5 % 10 % และ 15 % และความคลาดเคลื่อนมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์ และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.5 1 และ 3 ตามลำดับ ข้อมูลที่ใช้ในการวิจัยครั้งนี้ได้จากการจำลองแบบโดยวิธีมอนติคาร์โลซึ่งกระทำซ้ำ 1,000 รอบ ในแต่ละสถานการณ์ ผลของการวิจัยสามารถสรุปได้ดังนี้

การเปรียบเทียบค่า ARMSE ของทั้ง 5 วิธี พบว่า ในทุกกรณี วิธี MFOR-SR จะให้ค่า ARMSE น้อยที่สุด และมีค่าใกล้เคียงกับวิธี MFOR-W โดยที่ค่า ARMSE มีแนวโน้มลดลงเมื่อขนาดตัวอย่างเพิ่มขึ้น และมีแนวโน้มเพิ่มขึ้นเมื่อสัดส่วนข้อมูลสูญหายในตัวแปรอิสระ ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเพิ่มขึ้น

ค่า ARMSE **แปรผันตาม**ปัจจัยต่อไปนี้จากมากไปน้อยคือ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน ระดับความสัมพันธ์ระหว่างตัวแปรอิสระ และสัดส่วนข้อมูลสูญหายในตัวแปรอิสระ **แต่แปรผกผัน**กับขนาดตัวอย่าง

The objective of this research is to compare multiple regression coefficients estimating ethods with missing data in independent variables. This research compares five multiple gression coefficient estimation methods, which comprise Complete Case (CC), First Order egression method - Weight Mixed Estimator method (FOR-W), First Order Regression method - ein Rule Mixed Estimator method (FOR-SR), Modified First Order Regression method - Weight ixed Estimator method (MFOR-W) and Modified First Order Regression method - Stein Rule Mixed stimator method (MFOR-SR). The criterion for making decision is Average Root Mean Squares Error (RMSE) and Ratio of Different Average Root Mean Squares Error (DIFF). As for the case study, e specify the number of independent variables (p) equal to 3. The sample sizes are 5p, 10p, 15p nd 20p. The level of correlations among independent variables could be classified into 3 levels for hich low levels equal to (0.1, 0.2, 0.3), middle levels equal to (0.4, 0.5, 0.6) and high levels equal to .7, 0.7, 0.8). The proportions of missing data in independent variables are 5%, 15% and 25%. ne distribution of error is normal distribution with mean equal to 0 and standard deviations equal to 5, 1 and 3, respectively. The data for this research is simulated by using the Monte Carlo simulation chnique with 1,000 repetitions for each case. The results of this research are as follows:

According to the comparison of ARMSE from five referred method, in all cases, MFOR-SR ethod has a smallest ARMSE and MFOR-W method is nearly to the same. The ARMSE decreases hen sample size increases but it increases when the proportions of missing data in independent ariables or the level of correlation among independent variables or the standard deviation of the rror increases.

The ARMSE varies with, most to least, respectively, the standard deviation of the error, the vel of correlation among independent variables and the proportions of missing data in independent ariables but converses to sample size.