

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาเกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพและน่าเชื่อถือ ซึ่งควรนำไปสู่การได้ตัวแบบที่ถูกต้องเหมาะสมที่สุดภายใต้ข้อมูลชุดเดียวกัน เกณฑ์ที่ใช้คือร้อยละของความผิดพลาดโดยเฉลี่ยของค่าพยากรณ์กับค่าสังเกต (MAPE) ซึ่งได้ศึกษาในกรณีที่มีจำนวนตัวแปรอิสระเท่ากับ 3 และ 4 การแจกแจงของความคลาดเคลื่อนที่ศึกษาคือการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 2 3 และ 5 ตามลำดับ ขนาดตัวอย่างเท่ากับ 20 35 และ 50 ตามลำดับ โดยกำหนดระดับความสัมพันธ์ของตัวแปรอิสระเป็น 3 ระดับคือ ระดับต่ำ ($\rho = 0.05 \rightarrow 0.35$) ระดับกลาง ($\rho = 0.40 \rightarrow 0.65$) และระดับสูง ($\rho = 0.70 \rightarrow 0.95$) ในการวิจัยนี้ได้ทำการจำลองด้วยเทคนิคมอนติคาร์โลซึ่งกระทำซ้ำ 500 ครั้งในแต่ละสถานการณ์ ซึ่งผลของการวิจัยสามารถสรุปได้ดังนี้

กรณีที่ 1 จำนวนตัวแปรอิสระเท่ากับ 3

1. เมื่อขนาดตัวอย่างเท่ากับ 20 ก) ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1, 2 และตัวแปรอิสระมีความสัมพันธ์ในระดับกลางตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกหนึ่งตัว แต่ถ้าตัวแปรอิสระมีความสัมพันธ์ในระดับสูงตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกสองตัว ข) ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3, 5 และตัวแปรอิสระมีความสัมพันธ์ในระดับต่ำขึ้นไปตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกหนึ่งตัว แต่ถ้าตัวแปรอิสระมีความสัมพันธ์ในระดับสูงตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกสองตัว

2. เมื่อขนาดตัวอย่างเท่ากับ 35 ก) ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1, 2 และตัวแปรอิสระมีความสัมพันธ์ในระดับกลางขึ้นไปตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกหนึ่งตัว ข) ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3, 5 และตัวแปรอิสระมีความสัมพันธ์ในระดับกลางขึ้นไปตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกหนึ่งตัว แต่ถ้าตัวแปรอิสระมีความสัมพันธ์ในระดับสูงตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกสองตัว

3. เมื่อขนาดตัวอย่างเท่ากับ 50 ทุกค่าของส่วนเบี่ยงเบนมาตรฐาน และตัวแปรอิสระมีความสัมพันธ์ในระดับสูงตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกหนึ่งตัว

กรณีอื่น ๆ ภายใต้ขอบเขตการวิจัยตัวแบบที่ได้รับการคัดเลือกคือตัวแบบเต็มรูป

กรณีที่ 2 จำนวนตัวแปรอิสระเท่ากับ 4

1. เมื่อขนาดตัวอย่างเท่ากับ 20 ก) ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1, 2 ให้ผลลัพธ์เช่นเดียวกับกรณีที่ 1 ($n = 20$) ข) ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3, 5 และตัวแปรอิสระมีความสัมพันธ์ในระดับต่ำตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกหนึ่งตัว แต่ถ้าตัวแปรอิสระมีความสัมพันธ์ในระดับกลางตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกสองตัว และถ้าตัวแปรอิสระมีความสัมพันธ์ในระดับสูงตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกสามตัว

2. เมื่อขนาดตัวอย่างเท่ากับ 35 ให้ผลลัพธ์เช่นเดียวกับกรณีที่ 1 ($n = 35$)

3. เมื่อขนาดตัวอย่างเท่ากับ 50 ก) ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1, 2 และตัวแปรอิสระมีความสัมพันธ์ในระดับสูงตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกหนึ่งตัว ข) ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3, 5 และตัวแปรอิสระมีความสัมพันธ์ในระดับกลางขึ้นไปตัวแบบที่ได้รับการคัดเลือกคือตัวแบบลดรูปที่มีการตัดตัวแปรออกหนึ่งตัว

กรณีอื่น ๆ ภายใต้ขอบเขตการวิจัยตัวแบบที่ได้รับการคัดเลือกคือตัวแบบเต็มรูป

ค่า MAPE แปรผันตามปัจจัยต่อไปนี้คือ ระดับความสัมพันธ์ ส่วนเบี่ยงเบนมาตรฐาน และจำนวนตัวแปรอิสระ แต่แปรผกผันกับขนาดตัวอย่าง

4682415026 : MAJOR STATISTICS

KEY WORD : Model Selection / Multicollinearity / Mean Absolute Percentage Error / Lattice

WALAITHIP BUNYATISAI : THE MODEL SELECTION UNDER LATTICE FOR THE BEST LINEAR REGRESSION EQUATION : ASSOC. PROF. THEERAPORN VERATHAWORN, Ph.D. 158 pp.

The objective of this research is to study efficiently criterion of model selection and reliability of search for the best suitably correctness of model within the same data. The Mean Absolute Percentage Error (MAPE), which the statistic tool used, act as the criteria in this research. This research used 3 and 4 independent variables. The distribution of error is normal distribution with mean equal to 0 and standard deviation equal to 1, 2, 3 and 5, respectively. The sample sizes are 20, 35 and 50, respectively. The levels of correlation among the independent variables could be classified into 3 levels for which low levels equal to 0.05 - 0.35, middle levels equal to 0.40 - 0.65 and high levels equal to 0.70 - 0.95. The study used the Monte Carlo Simulation method. The experiment was replicated 500 times under each situation. The analyzed results of the data were demonstrated as follow.

CASE 1 The number of independent variables is 3

1. a sample size = 20 a) standard deviation = 1, 2 and independent variables when they have correlation in middle levels of the selected model is reduced model which was subtracted one variable but if independent variables have correlation in high levels the selected model is reduced model that was subtracted two variables b) standard deviation = 3,5 and also independent variables when they have low levels of correlation and upper the selected model is reduced model that was subtracted one variable but if independent variables have correlation in high levels the selected model is reduced model that was subtracted two variables

2. a sample size = 35 a) standard deviation = 1, 2 and independent variables when they have correlation in middle levels and upper the selected model is reduced model which was subtracted one variable b) standard deviation = 3,5 and independent variables when they have correlation in middle levels and upper the selected model is reduced model which was subtracted one variable but if independent variables have correlation in high levels of the selected model is reduced model that was subtracted two variables

3. a sample size = 50 independent variables have correlation in high levels of the selected model is reduced model that was subtracted one variable

In the other cases the selected model is a full model.

CASE 2 The number of independent variables is 4

1. a sample size = 20 a) standard deviation = 1, 2 the result is the same as case 1 ($n = 20$) b) standard deviation = 3, 5 and independent variables when they have correlation in low levels of the selected model is reduced model that was subtracted one variable but if independent variables have correlation in middle levels of the selected model is reduced model that was subtracted two variables and if independent variables have correlation in high levels of the selected model is reduced model that was subtracted three variables

2. a sample size = 35 the result is the same as case 1 ($n = 35$)

3. a sample size = 50 a) standard deviation = 1, 2 and independent variables when they have correlation in high levels of the selected model is reduced model that was subtracted one variable b) standard deviation = 3, 5 and independent variables when they have correlation in middle levels and upper the selected model is reduced model that was subtracted one variable

In the other cases the selected model is a full model.

The MAPE varies with level of correlations, standard deviation, number of independent variables but converses to sample sizes.