

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ ซึ่งมีตัวแบบทั่วไปดังนี้

$$Y = X\beta + e$$

เมื่อ  $Y$  แทนเวกเตอร์ของค่าตัวแปรตามขนาด  $n \times 1$   $X$  แทนเมทริกซ์ของค่าตัวแปรอิสระที่ทราบค่าขนาด  $n \times p$   $\beta$  แทนเวกเตอร์ของค่าคงที่การถดถอยและสัมประสิทธิ์การถดถอยขนาด  $p$   $e$  แทนเวกเตอร์ความคลาดเคลื่อนสุ่มขนาด  $n \times 1$  โดยที่  $e \sim N(0, \sigma^2 I_n)$   $I_n$  แทนเมทริกซ์เอกลักษณ์ขนาด  $n \times n$   $n$  แทนขนาดตัวอย่าง เกณฑ์ที่นำมาเปรียบเทียบการคัดเลือกตัวแบบ มี 5 เกณฑ์ คือ เกณฑ์การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของอาโคเคะที่ปรับค่าความเสี่ยง โดยใช้ตัวประมาณของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ หรือ  $AIC_n$  เกณฑ์การคัดเลือกตัวแบบที่พัฒนาเกณฑ์  $AIC_C$  ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) หรือ  $KIC_C$  เกณฑ์การคัดเลือกตัวแบบที่พัฒนาเกณฑ์  $AIC_I$  ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) หรือ  $KIC_I$  การคัดเลือกตัวแบบ Cross-Validation ( $C_p$ ) ที่ใช้สำหรับตัวอย่างขนาดเล็ก หรือ  $G_n^{CV}$  ที่ใช้  $C_n^{(R)}$  และเกณฑ์การคัดเลือกตัวแบบวิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method) กำหนดค่า  $\alpha = 3/4$  และ  $\epsilon = 2/5$  ในเกณฑ์การคัดเลือกตัวแบบ  $AIC_n$  เกณฑ์ที่ใช้เปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Square Error : AMSE) และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average of Mean Square Error : RDAMSE) โดยศึกษาจากการจำลองข้อมูล ด้วยโปรแกรม S-plus 2000 โดยตัวแปรอิสระเริ่มต้นเท่ากับ 3, 6, 9 และ 12 ค่าคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติ ค่าเฉลี่ยเท่ากับ 0 และค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1, 5 และ 10 และขนาดตัวอย่างขึ้นกับจำนวนตัวแปรอิสระเริ่มต้น ทำซ้ำแต่ละกรณีเท่ากับ 500 ครั้ง โดยข้อมูลที่ทำการจำลองสมการถดถอยเชิงเส้นพหุคูณเป็นไปตามข้อตกลงเบื้องต้น ผลการวิจัยสรุปดังนี้

เมื่อขนาดตัวอย่างเล็ก วิธี  $AIC_n$  กับ วิธีการถดถอยแบบขั้นบันได เป็นเกณฑ์การคัดเลือกตัวแบบที่ดีที่สุด เมื่อเปรียบเทียบกับ วิธี  $KIC_I$ , วิธี  $KIC_C$ , และ วิธี  $G_n^{CV}$  ที่ใช้  $C_n^{(R)}$  และวิธี  $KIC_I$  เป็นเกณฑ์คัดเลือกตัวแบบที่ดีกว่า  $KIC_C$

เมื่อขนาดตัวอย่างปานกลาง วิธี  $AIC_n$  กับ วิธีการถดถอยแบบขั้นบันได เป็นเกณฑ์การคัดเลือกตัวแบบที่ดีที่สุดและให้ผลการคัดเลือกตัวแบบเหมือนกัน และ ใกล้เคียง วิธี  $KIC_C$  มาก และ วิธี  $KIC_C$  เป็นเกณฑ์คัดเลือกตัวแบบที่ดีกว่า  $KIC_I$

เมื่อขนาดตัวอย่างใหญ่ วิธี  $AIC_n$ , วิธีการถดถอยแบบขั้นบันได, วิธี  $KIC_C$  และ วิธี  $KIC_I$  ให้ผลการคัดเลือกตัวแบบเหมือนกันหรือใกล้เคียงกันมาก

The objective of this study is to compare model selection criterion for Multiple Linear Regression Model. The General Multiple Linear Regression Model is shown as follow.

$$Y = X\beta + e$$

where  $Y$  is a  $nx1$  vector of dependent variable,  $X$  is a  $nxp$  matrix of independent variables,  $\beta$  is an  $px1$  unknown parametric vector of regression coefficients,  $e$  is a  $nx1$  random vector of error and  $e \sim N(0, \sigma^2 I_n)$ ,  $I_n$  is an  $nxn$  identity matrix,  $n$  is the number of observed data. Model Selection Criteria for this research are Model Selection Criterion using adjusted Akaike Information Criterion by approximating Kullback Leibler Information for reduced Risk Function or  $AIC_{ri}$ , Model Selection Criterion using developed  $AIC_C$  Criterion by Kullback's Symmetric Divergence or  $KIC_C$ , Model Selection Criterion using developed  $AIC_I$  Criterion by Kullback's Symmetric Divergence or  $KIC_I$ , Cross-Validation for small sample sizes or  $G_n^{CV}$  using  $C_n^{(R)}$ , The Partial F-test Statistic using Stepwise Regression. Define  $\alpha = 3/4$  and  $\varepsilon = 2/5$  in  $AIC_{ri}$  Criterion. The criteria employed for the comparison are The Average of Mean Square Error (AMSE) and The Ratio of Different Average of Mean Square Error (RDAMSE). In this study, the data is simulated by S-plus 2000 Computer Program. The number of independent variables are 3, 6, 9 and 12. The error is normally distribution with mean 0 and standard deviation equals to 1, 5 and 10. The sample size based on the number of independent variables. In this study, we simulate data 500 times. The simulation data satisfy the assumption in Multiple Linear Regression Analysis. The results of the study are as follows:

In the case of small sample sizes:  $AIC_{ri}$  and Stepwise Regression methods are the best model selections comparing to  $KIC_C$ ,  $KIC_I$  and  $G_n^{CV}$  using  $C_n^{(R)}$  methods.  $KIC_I$  method is a better than  $KIC_C$  method.

In the case of moderate sample sizes:  $AIC_{ri}$  and Stepwise Regression methods are the best model selections and have the same result. Which also similar to  $KIC_C$  method.  $KIC_C$  method is a better than  $KIC_I$  method.

In the case of large sample sizes:  $AIC_{ri}$ , Stepwise Regression,  $KIC_C$  and  $KIC_I$  methods give the same or similar results.