

ปริยาภรณ์ เมืองพรหม : การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบย์เมื่อใช้การแจกแจงก่อนคู่สังยุคเบตา (A COMPARISON OF BAYESIAN SELECTION METHODS FOR BEST REGRESSION EQUATION WITH CONJUGATE BETA PRIOR DISTRIBUTION)

อ. ที่ปรึกษา : รศ.ดร. วีระพร วีระถาวร , 286 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบย์เมื่อใช้การแจกแจงก่อนคู่สังยุคเบตา โดยจะเปรียบเทียบวิธีการคัดเลือกสมการถดถอย 3 วิธี ได้แก่ วิธีการเฉลี่ยตัวแบบของเบย์ (Bayesian Model Averaging method) โดยการหาค่าประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition (MC³)) เมื่อพิจารณาการแปลงที่เหมาะสมของตัวแปรอิสระ (Model Uncertainty via Simultaneous Variable and Transformation Selection (BMA_{SVT})) วิธีการคัดเลือกตัวแบบที่เหมาะสมที่สุด (Optimal Predictive Model Selection: median probability model (OPM)) และวิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR)) ซึ่งสองวิธีแรกเป็นวิธีการภายใต้แนวทางของเบย์ เกณฑ์ที่ใช้ในการตัดสินใจ คือค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Square Error (AMSE)) และใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Square Error (RDAMSE)) เพื่อประกอบการเปรียบเทียบประสิทธิภาพของทั้งสามวิธี การแจกแจงของค่าคลาดเคลื่อนที่ศึกษาคือการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐาน (σ) เท่ากับ 0.25 0.50 และ 2.50 ตามลำดับ ขนาดตัวอย่างที่ศึกษา (n) คือ 10 30 50 และ 100 จำนวนตัวแปรอิสระที่ศึกษาคือ 3 5 10 และ 15 ตัวแปร ค่าคงที่ $\left(\frac{\sigma_\beta}{\tau}, c\right)$ สำหรับวิธี BMA_{SVT} และวิธี OPM ที่ศึกษามี 4 ระดับคือมีค่าเป็น (1,5) (1,10) (10,100) และ (10,500) ข้อมูลที่ใช้ในการวิจัยได้จากการจำลองแบบด้วยเทคนิคมอนติคาร์โล กระทำซ้ำ 500 รอบ ในแต่ละสถานการณ์ ซึ่งผลการวิจัยได้ข้อสรุปดังนี้

การเปรียบเทียบค่า AMSE ของทั้ง 3 วิธีเรียงลำดับจากน้อยไปมาก ได้แก่ วิธี BMA_{SVT} วิธี OPM และวิธี SR ตามลำดับสำหรับทุกสถานการณ์ วิธี OPM จะให้ค่า AMSE สูงกว่าวิธี BMA_{SVT} เพียงเล็กน้อยโดยเฉพาะอย่างยิ่งเมื่อค่าคงที่ $\left(\frac{\sigma_\beta}{\tau}, c\right)$ มีค่าต่ำๆ ส่วนวิธี SR มีค่า AMSE แตกต่างจากวิธี BMA_{SVT} และวิธี OPM อย่างชัดเจนในทุกสถานการณ์ ปัจจัยที่มีผลต่อค่า AMSE ของทุกวิธีคือขนาดตัวอย่าง ค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนและจำนวนตัวแปรอิสระ โดยที่ค่า AMSE จะแปรผันตามค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนและจำนวนตัวแปรอิสระ แต่จะแปรผกผันกับขนาดตัวอย่าง

นอกจากนั้นค่า AMSE ของ 2 วิธี คือ วิธี BMA_{SVT} และวิธี OPM ซึ่งเป็นวิธีการภายใต้แนวทางของเบย์ จะแปรผันตามค่าคงที่ $\left(\frac{\sigma_\beta}{\tau}, c\right)$ แต่ AMSE ของวิธี SR ไม่มีผลกระทบเนื่องจากไม่ได้นำค่าคงที่ $\left(\frac{\sigma_\beta}{\tau}, c\right)$

The objective of this research is to compare the Bayesian selection methods for best regression equation with conjugate beta prior distribution. The three selection methods which the first two methods are Bayesian methods for best regression equation in this comparison composed of Bayesian Model Averaging method using Markov Chain Monte Carlo model composition via Simultaneous Variable and Transformation Selection (BMA_{svt}), Optimal Predictive Model Selection (OPM) and Stepwise Regression method (SR). The criterion in this project is the average of mean square error (AMSE), and the ratio of different average mean square error (RDAMSE) which analyzed comparison to efficiency of the three methods. The details of the data were represented as follows. In this study, the distribution of random errors are normal distribution with mean equal to 0 and standard deviation equal to 0.25 0.50 and 2.50 , respectively. The sizes of the samples are varied, which composed of 10, 30, 50 and 100 samples. The numbers of independent variables in regression model are 3, 5, 10 and 15. The Bayesian hyperparameter $\left(\frac{\sigma_{\beta}}{\tau}, c \right)$ for BMA_{svt} and OPM are (1,5) (1,10) (10,100) and (10,500), respectively. Using the Monte Carlo simulation technique with 500 repetitions for each case generate all data.

The analyzed results of data were demonstrated as follow. The comparisons of the AMSE from the three methods ranging from minimum to maximum were BMA_{svt} , OPM and SR for all cases. The OPM gave AMSE slightly higher than BMA_{svt} especially when the Bayesian hyperparameters had low values. The SR gave AMSE clearly different from BMA_{svt} and OPM in all cases. The factors that affected AMSE of all methods are sample size, the standard deviation of random errors and the number of independent variables. The AMSE of all method were proportionate to the standard deviation of random errors and the number of independent variables although they were inversely proportionate to sample size.

Furthermore, the AMSE of two methods under Bayesian Analysis, BMA_{svt} and OPM were proportionate to the Bayesian hyperparameters, excluding AMSE of SR was not effected to this Bayesian hyperparameters case therefore we did not mention on it.