

ภัทรสุดา สุตแดน : การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นเมื่อตัวอย่างมีขนาดเล็ก
(A COMPARISON OF MODEL SELECTION CRITERIA ON LINEAR REGRESSION WITH SMALL
SAMPLE SIZE) อ. ที่ปรึกษา : รศ.ดร.ธีระพร วีระถาวร , 118 หน้า. ISBN 974-14-2260-1

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นเมื่อตัวอย่างมีขนาดเล็กโดยจะเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอย 4 เกณฑ์ ได้แก่ เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้(Corrected Akaike's Information Criterion (AIC_c)) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้(Corrected Schwarz's Information Criterion (SIC_c)) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้(Corrected Hannan and Quinn's Information Criterion (HQ_c)) และเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลล์แบ็คที่ปรับแก้(Corrected Kullback's Information Criterion (KIC_c)) เกณฑ์การตัดสินใจคือค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Squares Error (AMSE)) และใช้อัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Squares Error (RDAMSE)) เพื่อเปรียบเทียบประสิทธิภาพของเกณฑ์การคัดเลือกตัวแบบทั้ง 4 เกณฑ์ การแจกแจงของความคลาดเคลื่อนสุ่มที่ศึกษาคือการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และความแปรปรวนเป็น 1, 2, 4, 8, และ 16 ขนาดตัวอย่างที่ใช้คือ 12, 15, 18, 21, 24, 27 และ 30 และจำนวนตัวแปรอิสระในตัวแบบการถดถอยคือ 3, 5 และ 7 ตัวแปร ข้อมูลที่ใช้ในการวิจัยใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โลกระทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ ซึ่งผลการวิจัยสรุปได้ดังนี้

ปัจจัยที่มีผลต่อค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) ของทุกเกณฑ์การคัดเลือกตัวแบบได้แก่ จำนวนตัวแปรอิสระและความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) และขนาดตัวอย่าง ตามลำดับ โดยที่ AMSE แปรผันตามจำนวนตัวแปรอิสระและความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) แต่ AMSE แปรผกผันกับขนาดตัวอย่าง ปัจจัยดังกล่าวข้างต้นส่งผลกระทบต่อเกณฑ์การคัดเลือกตัวแบบดังกรณีต่อไปนี้

1. กรณีขนาดตัวอย่างเท่ากับ 12 ถึง 15

เกณฑ์การคัดเลือกตัวแบบที่ดีที่สุดคือ เกณฑ์ HQ_c รองลงมาคือ เกณฑ์ AIC_c , SIC_c และ KIC_c ตามลำดับ สำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่มและจำนวนตัวแปรอิสระ

2. กรณีขนาดตัวอย่างเท่ากับ 18

เกณฑ์การคัดเลือกตัวแบบที่ดีที่สุดมีสองเกณฑ์คือ เกณฑ์ HQ_c และเกณฑ์ AIC_c และรองลงมาคือเกณฑ์ KIC_c และ SIC_c โดยที่เกณฑ์ KIC_c มีค่า AMSE ต่ำกว่า SIC_c เพียงเล็กน้อย สำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่มและจำนวนตัวแปรอิสระ

3. กรณีขนาดตัวอย่างเท่ากับ 21 ถึง 30

เกณฑ์การคัดเลือกตัวแบบที่ให้ค่า AMSE น้อยที่สุดคือ เกณฑ์ AIC_c รองลงมาคือ เกณฑ์ HQ_c , KIC_c และ SIC_c ตามลำดับ สำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่มและจำนวนตัวแปรอิสระ

เมื่อพิจารณาค่า RDMASE พบว่า ประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์การคัดเลือกตัวแบบจะแบ่งออกเป็นสองกลุ่มอย่างเห็นได้ชัด คือ กลุ่มที่ 1 ได้แก่ เกณฑ์ HQ_c และเกณฑ์ AIC_c มีประสิทธิภาพใกล้เคียงกันและดีที่สุดสำหรับทุกขนาดตัวอย่างและความแปรปรวนของความคลาดเคลื่อนสุ่ม ส่วนกลุ่มที่ 2 ได้แก่ เกณฑ์ KIC_c และ SIC_c ซึ่งมีประสิทธิภาพใกล้เคียงกัน แต่แตกต่างกันและต่ำกว่ากลุ่มแรกสำหรับทุกขนาดตัวอย่างและความแปรปรวนของความคลาดเคลื่อนสุ่ม

4682360026 : MAJOR STATISTICS

KEY WORD : AIC_C / HQ_C / KIC_C / LINEAR REGRESSION / MODEL SELECTION / SIC_C / SMALL SAMPLE SIZE

PATTARASUDA SUDSAEN : A COMPARISON OF MODEL SELECTION CRITERIA ON LINEAR REGRESSION WITH SMALL SAMPLE SIZE. THESIS ADVISOR : ASSOC. PROF. THEERAPORN

VERATHAWORN, Ph.D. 118 pp. ISBN 974-14-2260-1

The purpose of this research is to compare model selection criteria for linear regression with small sample size. The four model selection criteria in this comparison are composed of Corrected Akaike's Information Criterion (AIC_C), Corrected Schwarz's Information Criterion (SIC_C), Corrected Hannan and Quinn's Information Criterion (HQ_C) and Corrected Kullback's's Information Criterion (KIC_C). The criterion of comparison is the Average of Mean Squares Error (AMSE) using Ratio of Different Average Mean Squares Error (RDAMSE) to compare the efficiency of these model selection criteria. The distribution of random errors are normal distribution with zero mean and variances are 1, 2, 4, 8 and 16, respectively. The sample sizes used in this study are 12, 15, 18, 21, 24, 27 and 30, respectively. The number of independent variables in regression model are 3, 5, and 6, respectively. The data for this experiment are generated through the Monte Carlo simulation technique, repeating 1,000 times for each case. The results of this research are as follow :

The factors that affected AMSE of all model selection criteria are the number of independent variables, the variance of random errors and sample size, respectively. The AMSE vary with, most to least, respectively, the number of independent variables and the variance of random errors but AMSE is converse to sample size. These factors affected model selection criteria as follow.

1. In case of sample size is 12 to 15

For all the number of independent variables and all variance of random errors, the HQ_C is the best, the AIC_C , the SIC_C and the KIC_C , respectively.

2. In case of sample size is 18

For all the number of independent variables and all variance of random errors, the HQ_C and the AIC_C are the best, the KIC_C and the SIC_C , respectively. The KIC_C give AMSE slightly lower than the SIC_C .

3. In case of sample size is 21 to 30

For all the number of independent variables and all variance of random errors, the AIC_C is the best, the HQ_C , the SIC_C and the KIC_C , respectively.

Furthermore, the four model selection criteria efficiency is separated into two groups by RDAMSE. The first group is composed of the HQ_C and the AIC_C . Their RDAMSE are close and they are the best for all cases. The second group is composed of the SIC_C and the KIC_C . Their RDAMSE are close but their RDAMSE are clearly different from the first group for all cases.