

หัวข้อวิทยานิพนธ์

การคัดเลือกตัวแบบในการถดถอยเชิงเส้นพหุคูณ
โดยใช้วิธีดับเบิลเจเนติกอัลกอริทึม

ผู้เขียน

นางสาวศรินทิพย์ หมั่นจันทร์

ปริญญา

วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)

อาจารย์ที่ปรึกษา

ดร.วรา มินแสน

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประยุกต์วิธีดับเบิลเจเนติกอัลกอริทึม (Double Genetic Algorithms: DGA) ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ พร้อมทั้งทำการเปรียบเทียบกับวิธีแบบขั้นตอน (Stepwise Regression: SR) และวิธีเจเนติกอัลกอริทึม (Genetic Algorithm: GA) โดยใช้เกณฑ์ความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squares Error: MSE) ซึ่งประกอบด้วยค่า MSE ที่ต่ำที่สุดและค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Squares Error: AMSE) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของทั้ง 3 วิธี และใช้เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike's Information Criterion: AIC) เป็นเกณฑ์ในการคัดเลือกตัวแบบ ดำเนินการโดยโปรแกรม R Version 3.0.1 ภายใต้สถานการณ์ต่างๆ คือ กำหนดจำนวนตัวแปรอิสระเท่ากับ 10 และ 20 ตัวแปร ขนาดตัวอย่างเท่ากับ 50, 100 และ 500 เมื่อความคลาดเคลื่อนมีการแจกแจงแบบปกติ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนเท่ากับ 50, 80, 150 และ 250 เมื่อเกิดและไม่เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ โดยกำหนดจำนวนการทำซ้ำ 100 รอบในแต่ละสถานการณ์

ผลการวิจัยพบว่าในสถานการณ์ที่เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ วิธี DGA ให้ค่า AMSE น้อยที่สุด ส่วนวิธี SR และวิธี GA ให้ค่าที่ใกล้เคียงกัน สำหรับกรณีที่ไม่เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ วิธี SR วิธี GA และวิธี DGA ให้ค่า MSE ที่ต่ำที่สุดเท่ากันและมีบางกรณีที่วิธี DGA ให้ค่าที่น้อยกว่าวิธี SR และวิธี GA

Thesis Title

Model Selection in Multiple Linear Regression
Using Double Genetic Algorithms Method

Author

Ms. Sirintip Muenjan

Degree

Master of Science (Applied Statistics)

Advisor

Dr. Watha Minsan

ABSTRACT

This research aims to apply Double Genetic Algorithms (DGA) method for model selection in multiple linear regression. The results of DGA are compared with the results obtained by the Stepwise Regression (SR) and Genetic Algorithms (GA) based on Mean Squares Error (MSE) inclusive the lowest MSE and AMSE value. Model selection using Akaike's Information Criterion (AIC) every method. The data simulations via the R program Version 3.0.1, in with and without multicollinearity problems, the variables are 10 and 20, the sample sizes are 50, 100 and 500, the standard deviation of errors are 50, 80, 150 and 250. These data sets are generated by assuming the normal distributed error and repeated 100 times in each situation.

The results of this research showed with multicollinearity problem, DGA method results lowest AMSE from the simulation in all cases but SR and GA are similarly. Without multicollinearity problem, the lowest MSE of SR, GA, and DGA method are almost the same. However, there is some sample that DGA method brings lower value comparing with MSE value in the same range of others.