

ชื่อเรื่องวิทยานิพนธ์	การประมาณค่าสัมประสิทธิ์การถดถอย กรณีปัญหาที่มีความสัมพันธ์ระหว่างตัวแปรอิสระและมีค่าผิดปกติ		
ชื่อผู้เขียน	นางสาวจินดา สุวินัยตระกูล		
วิทยาศาสตร์มหาบัณฑิต	สาขาวิชาสถิติประยุกต์		
คณะกรรมการสอบวิทยานิพนธ์	อาจารย์พิษณุ	เจียวคุณ	ประธานกรรมการ
	ผู้ช่วยศาสตราจารย์อังกาบ	บุญชัย	กรรมการ
	อาจารย์นพดล	เล็กสวัสดิ์	กรรมการ
	รองศาสตราจารย์รัชนี้	ติยพันธ์	กรรมการ

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงพหุเมื่อเกิดปัญหาที่มีความสัมพันธ์ระหว่างตัวแปรอิสระและมีค่าผิดปกติโดยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method : OLS) วิธีรีดจ์ที่มีค่าสัมบูรณ์น้อยที่สุด (Ridge Least Absolute Value Method : RLAV) และวิธีรีดจ์ชนิดตัวประมาณจีเอ็ม (Generalized M-estimator Ridge Method : GMR) ซึ่งใช้เกณฑ์ความแกร่งของ Huber (1964) ในการเปรียบเทียบใช้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) โดยศึกษาการแจกแจงของความคลาดเคลื่อนเมื่อมีการแจกแจงแบบปกติที่ $\mu = 1$ และระดับของสัมประสิทธิ์การแปรผัน (C.V.) เท่ากับ 5 % และ 10 % เมื่อจำนวนตัวแปรอิสระเท่ากับ 3 และ 5 กำหนดระดับความสัมพันธ์ของตัวแปรอิสระเป็น 3 ระดับคือระดับต่ำ $\rho = 0.10$ ระดับปานกลาง $\rho = 0.50$ และระดับสูง $\rho = 0.90$ กำหนดเปอร์เซ็นต์ค่าผิดปกติเท่ากับ 10 % 20 % และ 30 % ของขนาดตัวอย่าง โดยศึกษาขนาดตัวอย่างเท่ากับ 20 และ 50 กำหนดระดับนัยสำคัญการทดสอบสมมติฐาน α เท่ากับ 0.01 และ 0.05 สำหรับข้อมูลที่ใช้ในการวิจัยได้จากการจำลองโดยใช้เทคนิคมอนติคาร์โล และกระทำซ้ำ 1,000 ครั้งในแต่ละสถานการณ์ ซึ่งผลการวิจัยสามารถสรุปได้ดังนี้

- วิธีริคจ์ชนิดตัวประมาณจีเอ็มจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุดในทุกกรณี ยกเว้นเมื่อขนาดตัวอย่างเท่ากับ 20 สัมประสิทธิ์การแปรผันเท่ากับ 10 % จำนวนตัวแปรอิสระเท่ากับ 5 ระดับความสัมพันธ์สูง และเปอร์เซ็นต์ค่าผิดปกติเท่ากับ 30 % ของขนาดตัวอย่าง วิธีริคจ์ที่มีค่าสัมบูรณ์น้อยที่สุดจะให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยที่สุดเช่นกัน อย่างไรก็ตามค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของสัมประสิทธิ์การถดถอยเชิงพหุทั้งสามวิธีจะแปรผันตามจำนวนตัวแปรอิสระ ระดับความสัมพันธ์ เปอร์เซ็นต์ค่าผิดปกติ และสัมประสิทธิ์การแปรผัน แต่จะแปรผกผันกับขนาดตัวอย่าง

- โดยส่วนใหญ่วิธีกำลังสองน้อยที่สุด และวิธีริคจ์ที่มีค่าสัมบูรณ์น้อยที่สุดสามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ตามเกณฑ์ของ Bradley ขณะที่วิธีริคจ์ชนิดตัวประมาณจีเอ็มส่วนใหญ่ไม่สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ แต่การเพิ่มระดับนัยสำคัญ α จะทำให้สามารถควบคุมความน่าจะเป็นของความคลาดเคลื่อนประเภทที่ 1 ได้ดีขึ้น สำหรับวิธีกำลังสองน้อยที่สุดจะให้ค่าอำนาจการทดสอบสูงสุดในทุกกรณี โดยค่าอำนาจการทดสอบจากการประมาณค่าสัมประสิทธิ์การถดถอยเชิงพหุทั้งสามวิธีจะแปรผันตามจำนวนตัวแปรอิสระ ระดับความสัมพันธ์ ขนาดตัวอย่าง และระดับนัยสำคัญ แต่จะแปรผกผันกับเปอร์เซ็นต์ค่าผิดปกติ และสัมประสิทธิ์การแปรผัน

Thesis Title	Regression Coefficient Estimation with Multicollinearity and Outliers Problems		
Author	Miss Jinda	Suvainaitragool	
M.S.	Applied Statistics		
Examining Committee	Lect. Phisanu	Chiawkhun	Chairman
	Asst. Prof. Angkarb	Boonyoi	Member
	Lect. Nopadol	Leksawat	Member
	Assoc. Prof. Ratchanee	Tiyapun	Member

Abstract

The objective of this research was to study methods of estimating multiple regression coefficients with multicollinearity and outliers problems by comparing Ordinary Least Square method (OLS) , Ridge Least Absolute Value method (RLAV) and Generalized M – estimator Ridge method (GMR) under the criteria of Huber 's robustness (1964). The comparison of parameter efficiency was made by means of least MSE. In this study the residual distribution was assumed normally distributed with $\mu = 1$ and coefficients of variation (C.V.) of 5 % and 10 % with the number of 3 and 5 independent variables. Three levels of the strength of correlation coefficient were used : weak , moderate and strong which were identified by $\rho = 0.10$ $\rho = 0.50$ and $\rho = 0.90$, respectively. The percent of outliers were 10 % 20 % and 30 % of the sample sizes and the sample sizes were 20 and 50. Significant levels (α) used in the hypothesis testing were 0.01 and 0.05. The data were obtained by using Monte Carlo simulation technique repeating 1,000 times for each case. The results of this research are as follows :

- GMR method gave least MSE for all cases except when we use the sample sizes of 20 C.V. of 10 % , the number of independent variables of 5 , high level of correlation and the percent of the outliers of 30 % of the sample sizes RLAV method also gave the least MSE. However , the least MSE varied directly with the number of independent variables , the level of correlation , the percent of the outliers and coefficients of variation , but conversed to the sample size.

- Based on Bradley's test , probability of type I could be controlled by OLS and RLAV methods. But GMR method generally could not control probability of type I error except when we increase the significant level , its capability to control probability of type I error for all estimation methods was improved. OLS method gave the highest power of the test for all cases. The power of the test varied directly with the number of independent variables , the level of correlation , the sample size and significant level , but conversed to the percent of the outliers and coefficients of variation.