



การประมาณค่าสัมประสิทธิ์การถดถอยสำหรับข้อมูลที่มีปัญหาสหสัมพันธ์เชิงเส้นพหุ
ด้วยวิธีการถดถอยแบบริดจ์และวิธีการค้นหาแบบนกกาเหว่า

Estimation of Multiple Linear Regression Analysis Coefficients with
Multicollinearity by Ridge Regression and Cuckoo Search

พรทิwa ธรรมชัยหลง^{1*} และกุลจิรา กิ่งไพร¹

¹ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเชียงใหม่ เชียงใหม่ 50300

Corresponding Author, E-mail: phornthiwar2539@gmail.com

Received: 21 November 2018 | Revised: 11 January 2019 | Accepted: 10 August 2019

บทคัดย่อ

การศึกษาค้นคว้ามีวัตถุประสงค์เพื่อประมาณค่าสัมประสิทธิ์การถดถอยในการวิเคราะห์การถดถอยเชิงเส้นพหุ กรณีที่มีปัญหาสหสัมพันธ์เชิงเส้นพหุ (Multicollinearity) โดยใช้วิธีการถดถอยแบบริดจ์ (Ridge Regression) ที่ใช้วิธีการประมาณค่าพารามิเตอร์ริดจ์ของ โฮเอิร์ล เคนนาร์ด และ บาลด์วิน (Hoerl, Kennard and Baldwin) เปรียบเทียบกับวิธีการประมาณค่าพารามิเตอร์ดังกล่าวโดยวิธีการค้นหาแบบนกกาเหว่า (Cuckoo Search) ที่ใช้ฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE) เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของสัมประสิทธิ์การถดถอยที่ได้จากวิธีการประมาณค่า 2 วิธีดังกล่าว คือค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าปัจจัยการขยายตัวความแปรปรวน (Variance Inflation Factor : VIF) ที่ต่ำที่สุด การศึกษาทำการเปรียบเทียบโดยใช้ข้อมูลจากการจำลอง ซึ่งประกอบด้วยตัวแปรอิสระ 3 ตัว และมีจำนวน 30 รายการ และข้อมูลจากสถานการณ์จริงประกอบด้วยตัวแปรอิสระ 14 ตัว และมีจำนวน 252 รายการ ในขั้นตอนของวิธีการค้นหาแบบนกกาเหว่ากำหนดจำนวนรอบในการสุ่มค่า r เท่ากับ 5 10 และ 15 รอบ จำนวนรังเท่ากับ 20 รัง และกระทำซ้ำ 100 500 1,000 5,000 และ 10,000 ครั้ง ผลการศึกษาพบว่าสำหรับวิธีการถดถอยแบบริดจ์ที่ใช้วิธีการประมาณค่าพารามิเตอร์ริดจ์ โดยวิธีการค้นหาแบบนกกาเหว่า ที่มีจำนวนรอบในการสุ่มค่า r เท่ากับ 15 รอบ และทำซ้ำทั้งหมด 10,000 ครั้ง ทำให้ได้ค่าประมาณสัมประสิทธิ์การถดถอยที่เหมาะสมที่สุด นอกจากนั้น วิธีการค้นหาแบบนกกาเหว่ายังทำให้ได้สัมประสิทธิ์การถดถอยที่มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าปัจจัยการขยายตัวความแปรปรวนต่ำกว่าวิธีการประมาณค่าพารามิเตอร์ริดจ์ของ โฮเอิร์ล เคนนาร์ด และบาลด์วิน ทั้ง 2 ชุดข้อมูล

ABSTRACT

The purpose of this study is to estimate the coefficients in multiple linear regression analysis that has multicollinearity problems by the method of Ridge Regression by Hoerl, Kennard and Baldwin compare with those method by the Cuckoo Search using Mean Squared Error (MSE) as the objective function. The efficiency criteria are MSE and Variance Inflation Factor (VIF). The comparison, using a simulation data composes of 3 independent variables with a sample size 30 and a real data set consists of 14 independent variables with 252 observations. In the Cuckoo Search Algorithms, 5, 10 and 15 iterations of searching r constant in ridge, 20 nests and 100, 500, 1,000,

5,000 and 10,000 repeated times are proposed. The result revealed that, the Ridge Regression by the Cuckoo Search with 15 iterations of searching r constant and 10,000 times of repetition gives the most appropriate regression coefficients. Moreover, those method by the Cuckoo Search also provides regression coefficients with lower MSE and VIF values than those method by Hoerl, Kennard and Baldwin in both data sets.

คำสำคัญ: สัมประสิทธิ์การถดถอย สหสัมพันธ์เชิงเส้นพหุ การถดถอยแบบบริดจ์ ค่าพารามิเตอร์ริดจ์ การค้นหาแบบนกกาเหว่า

Keywords: Regression Coefficients, Multicollinearity, Ridge Regression, Ridge Parameter, Cuckoo Search

บทนำ

การวิเคราะห์การถดถอยเป็นวิธีการทางสถิติอย่างหนึ่งที่ใช้ในการศึกษาและอธิบายลักษณะความสัมพันธ์ระหว่างตัวแปรหนึ่งกับตัวแปรตัวอื่น ๆ และคาดคะเนค่าของตัวแปรนั้นโดยอาศัยหลักความสัมพันธ์เบื้องต้น ซึ่งตัวแปรที่นิยมใช้กันอย่างแพร่หลาย คือ ตัวแบบการถดถอยเชิงเส้นพหุคูณ โดยตัวแบบนี้เป็นความสัมพันธ์ระหว่างตัวแปรอิสระที่มีมากกว่า 1 ตัว กับตัวแปรตาม 1 ตัว ทั้งนี้ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม อยู่ในรูปเชิงเส้น การใช้ตัวแบบการถดถอยเชิงเส้นในการทำนายให้มีประสิทธิภาพนั้นจะขึ้นอยู่กับวิธีการเลือกตัวแปรที่มีความเหมาะสมที่สุด ซึ่งเกิดจากการคัดเลือกตัวแปรอิสระที่มีอิทธิพลมากที่สุดต่อตัวแปรตาม และจำนวนตัวแปรอิสระที่อยู่ในตัวแบบต้องไม่มากและไม่น้อยเกินไป เนื่องจากตัวแปรอิสระที่มีมากเกินไปจะทำให้ค่าทำนายที่ได้มีความคลาดเคลื่อนสูงและอาจทำให้เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ นอกจากนั้นแล้วหากตัวแปรที่สร้างขึ้นขาดตัวแปรอิสระที่สำคัญไปจะทำให้ค่าทำนายมีความคลาดเคลื่อนสูงได้เช่นเดียวกัน (รัฐพงศ์ และ ปริม, 2560)

ในปัจจุบันวิธีการที่นิยมนำมาใช้ในการแก้ปัญหาสหสัมพันธ์เชิงเส้นพหุ คือ การวิเคราะห์การถดถอยแบบบริดจ์ ซึ่งวิธีนี้ไม่จำเป็นที่จะต้องตัดตัวแปรอิสระที่มีพหุสัมพันธ์ เชิงเส้นออกจากตัวแบบ แต่ตัวประมาณที่ได้จากวิธีนี้จะเป็นตัวประมาณที่เอนเอียงของ β หลักการของวิธีนี้คือการบวกค่าคงตัว r ค่าหนึ่งที่มีค่ามากกว่าศูนย์ให้กับสมาชิกทุก ๆ ตัว บนแนวทแยงมุมของเมทริกซ์ $W'W$ เพื่อให้สมาชิกบนแนวทแยงมุมมีค่ามากขึ้นจะทำให้สภาวะการป่วย (Ill Condition) ของเมทริกซ์ $W'W$ ลดลงและเพื่อหาเมทริกซ์ผกผันของ $W'W$ ที่มีความเสถียรมากขึ้นทำให้สมาชิกบนแนวทแยงมุมของเมทริกซ์ผกผัน $W'W$ มีค่าน้อยลงเป็นการลดความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณพารามิเตอร์ β ซึ่งจะได้ตัวประมาณค่าสัมประสิทธิ์การถดถอยแบบบริดจ์ คือ $\hat{\beta}_r = (W'W + rI)^{-1}W'y$ เมื่อ $r > 0$ และในการวิเคราะห์การถดถอยแบบบริดจ์นั้นจะต้องมีการประมาณค่าคงตัว r ซึ่งมีผู้ศึกษาวิธีการประมาณค่าคงตัว r หลายวิธี เช่น วิธีของไฮเออร์ล เคนนาร์ด และ บาลด์วิน วิธีของโนมูระ (Nomura) และวิธีของคาลาฟ และ ชูเกอร์ (Khalaf and Shukur) เป็นต้น

นอกจากวิธีการที่กล่าวมาข้างต้นยังมีวิธีการแก้ปัญหาสหสัมพันธ์เชิงเส้นพหุที่ประยุกต์มาจากวิธีการหาคำตอบที่ดีที่สุดโดยอาศัยหลักการประมาณเป็นกลุ่มวิธีการที่เรียกว่า เมตาฮิวริสติก (Metaheuristic) ซึ่งเป็นวิธีการหาคำตอบที่ดีที่สุดที่เข้ามาช่วยในการหาคำตอบสำหรับบางปัญหาที่ไม่สามารถใช้วิธีการทางทฤษฎีแบบดั้งเดิม หรือถ้าหากทำได้ก็จะใช้เวลานาน (Babak et al., 2011) เช่น วิธีเชิงพันธุกรรม (Genetic Algorithm) วิธีการจำลองการอบเหนียว (Simulate Annealing) วิธีค้นหาแบบต้องห้าม (Tabu Search) รวมถึงวิธีการค้นหาแบบนกกาเหว่า (Cuckoo Search) ที่นำมาใช้ในการประมาณค่าสัมประสิทธิ์การถดถอยในงานวิจัยนี้ โดยหลัก ๆ แล้วหลักการทำงานของเมตาฮิวริสติกในทุกวิธีที่กล่าวมานั้นมีส่วนที่เหมือนกันคือ การหาคำตอบด้วยการสุ่มและมีการวนรอบการทำซ้ำ คำตอบที่ได้แต่ละครั้งและแต่ละรอบการทำซ้ำจึงไม่เท่ากัน เนื่องจากการค้นหาคำตอบที่มีพื้นฐานมาจากการสุ่ม (อารีรัตน์, 2556)

ในงานวิจัยต่าง ๆ ที่ผ่านมามีการนำวิธีการทางเมตาฮิวริสติกมาแก้ปัญหาในหลากหลายรูปแบบ อาทิเช่น วิธีการค้นหาแบบกลุ่มอนุภาคและวิธีเชิงพันธุกรรมถูกนำมาประยุกต์ใช้ในการแก้ปัญหาการจัดเรียงเครื่องจักรในระบบการผลิตแบบยืดหยุ่น (ณัฐพงศ์, 2551) วิธีการค้นหาแบบนกกาเหว่าถูกนำมาประยุกต์ใช้ในการจัดวางตำแหน่งติดตั้งอุปกรณ์ขึ้นบอกความผิดปกติ (อรธพล, 2557) และวิธีการค้นหาแบบต้องห้ามถูกนำมาประยุกต์ใช้ในการคัดเลือกตัวแปรในตัวแบบการถดถอยเชิงเส้นพหุ (กานต์ณัฐ, 2554) การศึกษาในครั้งนี้ผู้วิจัยจึงได้สนใจนำวิธีการค้นหาแบบนกกาเหว่ามาประยุกต์ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอย เนื่องจากเป็นกระบวนการทำงานที่ไม่ซับซ้อนและไม่ซับซ้อนหรือมีเงื่อนไขทางสถิติ และจากการทบทวนวรรณกรรมที่ผ่านมา ยังไม่พบงานวิจัยใดที่นำเอาวิธีการ

ค้นหาแบบนกกาเหว่ามาประยุกต์ใช้ในการประมาณค่าสัมประสิทธิ์การถดถอย ในกรณีที่เกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ จากที่กล่าวมา จึงทำให้เห็นว่าวิธีการค้นหาแบบนกกาเหว่า เป็นวิธีการใหม่และมีความน่าสนใจในการนำมาประยุกต์ใช้ในงานวิจัยนี้

ดังนั้น งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นพหุคูณ โดยวิธีการถดถอยแบบบริดจ์และวิธีการค้นหาแบบนกกาเหว่า และเปรียบเทียบประสิทธิภาพของทั้ง 2 วิธี

การประมาณค่าคงตัว r ในการถดถอยแบบบริดจ์

ในการศึกษาครั้งนี้ได้เปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีการถดถอยแบบบริดจ์ที่มีการประมาณค่า r ที่เป็นค่าคงตัว ในการประมาณค่าพารามิเตอร์ของตัวแบบการถดถอย โดยวิธีของโฮเอิร์ล เคนนาร์ต และ บาลด์วิน (r_{HKB})

จากการประมาณค่า r ของโฮเอิร์ล และ เคนนาร์ต (r_{HK}) ที่ต้องการเลือกค่า r_{HK} ที่ทำให้ค่าความคลาดเคลื่อน กำลังสองเฉลี่ยของ $\hat{\alpha}_r$ มีค่าต่ำสุด โดยการหาอนุพันธ์บางส่วนเทียบกับค่า r_j จะได้ว่า

$$r_j = \frac{\hat{\sigma}^2}{\hat{\alpha}_j^2} \quad (1)$$

แต่เนื่องจาก $\alpha = T'\beta$ ดังนั้น อาจมีกรณีที่ α_j มีค่าเท่ากับศูนย์ ซึ่งจะทำให้ไม่สามารถหาค่า r ได้ และเนื่องจากค่า r มีผลต่อความเอนเอียงกำลังสองและค่าความแปรปรวนของตัวประมาณ ดังนั้น โฮเอิร์ล และ เคนนาร์ตจึงทำการกำหนดค่า r ที่เหมาะสมเพียงค่าเดียว โดยกำหนดให้ค่า $\hat{\alpha}_j$ จากสมการ (1) เป็นค่ามากที่สุดจะได้ว่า

$$r_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2}$$

$$\text{เมื่อ } \hat{\alpha}_{\max} = \max(\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p)$$

$$\text{และ } \hat{\sigma}^2 = \frac{y'y - \hat{\alpha}'T'W'y}{n - p}$$

จากสมการ (1) โฮเอิร์ล เคนนาร์ต และ บาลด์วิน ได้เสนอวิธีการกำหนดค่าคงตัว r ที่เหมาะสมเพื่อช่วยลดผลกระทบที่เกิดขึ้น ดังนี้

$$r_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

เมื่อ p คือจำนวนพารามิเตอร์ในตัวแบบการถดถอย

วิธีการค้นหาแบบนกกาเหว่า

วิธีการค้นหาแบบนกกาเหว่าถูกนำเสนอในปี 2009 โดย Yang และ Deb เพื่อนำมาใช้แก้ปัญหาการค้นหาค่าที่เหมาะสมที่สุด โดยวิธีการนี้ได้แรงบันดาลใจจากกลยุทธ์การสืบพันธุ์ของนกกาเหว่าตามธรรมชาติ ซึ่งมีความน่าสนใจคือ นกกาเหว่าจะสืบพันธุ์โดยการนำไข่ของตนเองไปฝากไว้ในรังนกชนิดอื่นที่มีลักษณะไข่คล้าย ๆ กัน โดยนกกาเหว่า จะทำการเลียนแบบไข่ของตนเอง เช่น สี ลวดลาย ให้เหมือนกับไข่ของนกเจ้าของรังที่มีอยู่แล้ว ถ้าเลียนแบบได้ดี ไข่นั้นก็จะถูกฟักโดยแม่เจ้าของรัง แต่ถ้าไข่ไม่ดีก็จะถูกนกเจ้าของรังโยนทิ้ง หรือทิ้งรังไปเพื่อสร้างรังใหม่ ดังนั้น การลอกเลียนแบบ สี ลวดลายไข่ จะเปรียบเสมือนการปรับปรุงค่าคำตอบ (Generate Solution) ซึ่งอาจมีโอกาสเป็นคำตอบที่ดีกว่า ก็จะได้รับการถ่ายทอดสู่รุ่นต่อไป แต่ถ้าเป็นคำตอบที่ไม่ดีก็จะมีโอกาสถูกคัดทิ้ง เพื่อให้เข้าใจง่ายในการอธิบายวิธีการค้นหาแบบนกกาเหว่า จะใช้กฎพื้นฐาน 3 ข้อ (Yang and Deb, 2009) คือ

1. นกกาเหว่าแต่ละตัวสามารถวางไข่ได้เพียง 1 ครั้ง ๆ ละ 1 ฟอง โดยจะทำการวางไข่ของตนเองลงไปรังที่ได้รับการสุ่มเลือกมา

2. รังที่ดีที่สุด ซึ่งมีไข่คุณภาพสูงจะได้รับการถ่ายทอดไปยังรุ่นต่อไป

3. จำนวนรังของนกเจ้าของรังที่มีอยู่จะคงที่ โดยนกกาเหว่าจะมีโอกาสถูกค้นพบโดยนกเจ้าของรังด้วยความน่าจะเป็น $p_a \in [0, 1]$ ซึ่งในกรณีนี้นกเจ้าของรังสามารถโยนไข่นกกาเหว่าทิ้งไปหรือทิ้งรังแล้วสร้างรังใหม่ที่ดีกว่า

จากกฎพื้นฐาน 3 ข้อ สามารถอธิบายขั้นตอนการทำงานของวิธีการค้นหาแบบนกกาเหว่าได้ดังนี้

ขั้นตอนที่ 1 : กำหนดสมการเป้าหมาย

ขั้นตอนที่ 2 : สร้างจำนวนประชากรของรังนกเพื่อให้เกิดความเรียงง่ายจะกำหนดรังนก 1 รังจะมีไข่ของนกเจ้าของรังเพียง 1 ฟอง โดยที่ไข่ในรังแต่ละฟองจะเป็นตัวแทนของคำตอบ

ขั้นตอนที่ 3 : นกกาเหว่าจะทำการสุ่มเลือกรังที่มีไข่ของนกเจ้าของรังวางอยู่แล้ว และทำการวางไข่ของตนเองลงในรังที่ทำการสุ่มเลือก โดยจะทำการเลียนแบบไข่ของตนเองจากไข่ของนกเจ้าของรังที่มีอยู่แล้ว ดังนั้น การเลียนแบบไข่ของนกกาเหว่าเปรียบเสมือนการปรับปรุงคำตอบ โดยใช้หลักการ Lévy Flights ซึ่งในการสร้างคำตอบใหม่ของ Lévy Flights จะอยู่ในรูปของสมการ (2)

$$x_i^{(t+1)} = x_i^t + \alpha \oplus Levy(\lambda) \quad (2)$$

เมื่อ x_i^t คือ คำตอบเดิม

$x_i^{(t+1)}$ คือ คำตอบใหม่ที่ทำการปรับปรุง

$\alpha \oplus Levy(\lambda)$ คือ ช่วงระยะห่างที่เกิดการเปลี่ยนแปลงตำแหน่งของคำตอบ

โดยที่ α ที่มีค่ามากกว่าศูนย์จะเป็นขนาดของช่วงการสุ่ม (Step Size) ซึ่งจะขึ้นอยู่กับขนาดของปัญหา ที่ต้องการจะหาคำตอบ ส่วนมากจะกำหนด $\alpha = 1$ จากสมการ (2) เป็นหลักการเดินสุ่มแบบ Markov Chain (Yang and Deb, 2010) คือตำแหน่งถัดไปจะขึ้นอยู่กับพารามิเตอร์ 2 ตัว คือ ตำแหน่งเดิม ณ ปัจจุบัน (x_i^t) และระยะทาง หรือความน่าจะเป็นที่มีการเปลี่ยนตำแหน่งใหม่

ในขั้นตอนการสุ่มแบบ Lévy Flights จะมีการกระจายตัวดังสมการ (3) มีค่าความแปรปรวนไปเรื่อย ๆ และมีค่าเฉลี่ยที่ไม่คงที่

$$Levy \sim u = t^{-\lambda} \quad (3)$$

โดยที่ $\lambda \in (0,3]$

ขั้นตอนที่ 4 : ทำการประเมินค่าความเหมาะสมของไข่ที่ได้รับการปรับปรุงและทำการเปรียบเทียบไข่เป็นคู่ที่อยู่ในรังเดียวกัน ในกรณีที่คำตอบใหม่ดีกว่าจะทำการแทนที่คำตอบเก่าด้วยคำตอบใหม่ โดยไข่ที่ดีมีคุณภาพที่สุดในแต่ละรังจะได้รับการถ่ายทอดสู่รุ่นต่อไป โดยจะกำหนดให้ไข่ใบนั้นเป็นคำตอบที่ดีที่สุดของแต่ละรัง

ขั้นตอนที่ 5 : หากจำนวนรังนกที่มีคำตอบไม่ดีและมีโอกาสสุมนกเจ้าของรังทำการทิ้งรังและไปสร้างรังใหม่ โดยพิจารณาความน่าจะเป็น p_a โดยที่ $p_a \in [0,1]$ เช่น มีรังนกทั้งหมด 100 รัง กำหนดค่า $p_a = 0.25$ ดังนั้น จะมีจำนวนรังที่มีคำตอบไม่ดีจำนวน 25 รังจะถูกทิ้งรัง โดยเริ่มนับจากรังที่แย่มากจนครบตามจำนวน จากนั้นทำการตัดชุดคำตอบที่ถูกเลือกทิ้งทิ้ง แล้วทำการสร้างรังขึ้นมาใหม่เพื่อแทนรังที่ถูกคัดออก

ขั้นตอนที่ 6 : นำผลเฉลี่ยที่ดีที่สุดของแต่ละรังมาเรียงลำดับ โดยพิจารณาจากค่าความเหมาะสมและกำหนดให้ไข่ที่มีคุณภาพมากที่สุด (คำตอบที่ดีที่สุด) ไว้ เป็นลำดับแรกแล้วเก็บค่าคำตอบที่ดีที่สุดไว้

ขั้นตอนที่ 7 : ทำการหาคำตอบในรอบต่อไป ทำจนครบรอบที่กำหนดไว้ แล้วเลือกคำตอบที่ดีที่สุด

ค่าสถิติที่ใช้ในการพิจารณาเลือกตัวแบบการถดถอย

เกณฑ์ที่ใช้ในการพิจารณาในการเลือกตัวแบบการถดถอยคือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (MSE) และค่าปัจจัยการขยายตัวความแปรปรวน (VIF) การหาค่าความคลาดเคลื่อนกำลังสองเฉลี่ยสำหรับสัมประสิทธิ์การถดถอยในแต่ละวิธี มีดังนี้

1. วิธีการถดถอยแบบบริดจ์

$$MSE_R = \frac{\mathbf{e}'_{ru} \mathbf{e}_{ru}}{n - k - 1}$$

2. วิธีการค้นหาแบบนกกาเหว่า

$$MSE_C = \frac{[(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})]}{n - k - 1}$$

และการหาค่าปัจจัยการขยายตัวความแปรปรวนของทั้งสองวิธีคือ

$$VIF = (\mathbf{W}\mathbf{W} + r\mathbf{I})^{-1} \mathbf{W}\mathbf{W}(\mathbf{W}\mathbf{W} + r\mathbf{I})^{-1}$$

วิธีการดำเนินการวิจัย

ขั้นตอนการดำเนินการวิจัย

งานวิจัยนี้ศึกษากรณีที่ตัวแปรอิสระเกิดปัญหาสหสัมพันธ์เชิงเส้นพหุ โดยใช้ข้อมูล 2 ชุด ดังนี้

ข้อมูลชุดที่ 1 คือข้อมูลที่ได้จากการจำลองโดยใช้ข้อมูลจากหนังสือการวิเคราะห์การถดถอย (Regression Analysis) (จิราวัลย์, 2558) ซึ่งประกอบด้วยตัวแปรอิสระ 3 ตัว จำนวน 30 รายการ

ข้อมูลชุดที่ 2 คือข้อมูลที่ได้จากสถานการณ์จริงโดยใช้ข้อมูลค่าประมาณเปอร์เซ็นต์ของไขมันในร่างกายที่ได้จากการชั่งน้ำหนักได้น้ำและการวัดขนาดส่วนต่าง ๆ ของร่างกาย (StatLib Datasets Archive, 2005) ซึ่งประกอบด้วยตัวแปรอิสระ 14 ตัว จำนวน 252 รายการ

โดยการศึกษาทั้ง 2 กรณีกำหนดจำนวนรอบในการสุ่มค่า r เท่ากับ 5 10 และ 15 รอบ จำนวนรังเท่ากับ 20 รัง และกระทำซ้ำ 100 500 1,000 5,000 และ 10,000 ครั้ง มีขั้นตอนดังนี้

1. ทำการแปลงรูปข้อมูลด้วยวิธีสเกลหน่วยความยาว (Unit Length Scaling) โดยเริ่มคำนวณค่า r_{ij} และค่า r_{jy} จากสูตร

$$r_{ij} = \frac{\sum_{u=1}^n (x_{ui} - \bar{x}_i)(x_{uj} - \bar{x}_j)}{(S_{ii} S_{jj})^{1/2}}$$

$$r_{jy} = \frac{\sum_{u=1}^n (x_{uj} - \bar{x}_j)(y_u - \bar{y})}{(S_{jj} SS_T)^{1/2}}$$

เมื่อ $S_{ii} = \sum_{u=1}^n (x_{ui} - \bar{x}_i)^2$ เป็นผลบวกกำลังสองของตัวแปรอิสระ x_i

$S_{jj} = \sum_{u=1}^n (x_{uj} - \bar{x}_j)^2$ เป็นผลบวกกำลังสองของตัวแปรอิสระ x_j

และ $SS_T = \sum_{u=1}^n (y_u - \bar{y})^2$ เป็นผลบวกกำลังสองของตัวแปรตาม y

โดยที่ x_{ui} เป็นค่าของข้อมูลลำดับที่ u ของตัวแปรอิสระตัวที่ $i; i = 1, 2, \dots, k$

x_{uj} เป็นค่าของข้อมูลลำดับที่ u ของตัวแปรอิสระตัวที่ $j; j = 1, 2, \dots, k$

จากนั้นทำการจัดค่า r_{ij} และค่า r_{jy} ในรูปของเมทริกซ์ความสัมพันธ์ ดังนี้

$$\mathbf{W}'\mathbf{W} = \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & 1 & r_{23} & \cdots & r_{2k} \\ r_{31} & r_{32} & 1 & \cdots & r_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & r_{k3} & \cdots & 1 \end{bmatrix}, \mathbf{W}'\mathbf{y} = \begin{bmatrix} r_{1y} \\ r_{2y} \\ r_{3y} \\ \vdots \\ r_{ky} \end{bmatrix}$$

2. คำนวณค่าสัมประสิทธิ์การถดถอย (β) โดยใช้วิธีประมาณค่าพารามิเตอร์รีดจ์ด้วยวิธีของไฮเออร์ล เคนนาร์ด และ บาลด์วิน และวิธีการค้นหาแบบนกเขา

3. ทำการคัดเลือกค่าสัมประสิทธิ์การถดถอย ที่เหมาะสมจากวิธีการประมาณค่าสัมประสิทธิ์การถดถอย 2 วิธี

การประมาณค่าสัมประสิทธิ์การถดถอย

1. ประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีการถดถอยแบบบริดจ์

1. หลังจากทำการแปลงรูปข้อมูลตามขั้นตอนที่กล่าวไว้ข้างต้น จะได้สัมประสิทธิ์การถดถอยมาตรฐาน คือ

$$\hat{\beta}_r = (\mathbf{W}'\mathbf{W} + r\mathbf{I})^{-1} \mathbf{W}'\mathbf{y}$$

จากนั้นเขียนให้ในรูปฟังก์ชันค่าเฉพาะโดยที่ $\sum_{i=1}^k \lambda_i = \text{tr}(\mathbf{W}'\mathbf{W})$ และ $i = 1, 2, \dots, k$

เมื่อ k เป็นจำนวนตัวแปรอิสระในตัวแบบ

2. คำนวณค่าสัมประสิทธิ์การถดถอยแบบบริดจ์มาตรฐาน (ภายใต้ตัวแบบแปลงรูป)

$$b_{r_{ij}} = b_{rj} \frac{SS_T^{1/2}}{(S_{jj})^{1/2}}$$

$$b_{r_{i0}} = \bar{y} - \sum_{j=1}^k b_{r_{ij}} \bar{x}_j$$

โดยที่ $j = 1, 2, \dots, k$

เมื่อ b_{rj} เป็นสัมประสิทธิ์การถดถอยบริดจ์มาตรฐาน

และ $b_{r_{ij}}$ เป็นสัมประสิทธิ์การถดถอยบริดจ์

3. เนื่องจากการกำหนดค่า r ที่เหมาะสมโดยวิธีของวิธีโฮเอิร์ล เคนนาร์ด และ บาลด์วิน (r_{HKB}) จะช่วยลดผลกระทบที่เกิดขึ้น

เมื่อ $\hat{\alpha}_j^2$ มีค่าน้อย ซึ่ง $\hat{\alpha}_j^2$ สามารถ คำนวณได้จาก $\boldsymbol{\alpha} = \mathbf{T}'\boldsymbol{\beta}$ จากนั้นทำการคำนวณค่า r_{HKB} จากสูตร $r_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}}$

4. หลังจากได้ค่า r_{HKB} จากการคำนวณในข้อ 3) แล้วทำการประมาณค่าสัมประสิทธิ์การถดถอยบริดจ์มาตรฐานตามข้อ 1. และ 2. หลังจากนั้นทำการคำนวณค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าปัจจัยการขยายตัวความแปรปรวน เพื่อนำไปเปรียบเทียบกับวิธีการค้นหาแบบนกกาเหว่าในขั้นตอนต่อไป

2. ประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีการค้นหาแบบนกกาเหว่า

1. ทำการสุ่มค่าคงตัว r จากช่วงที่กำหนด

2. คำนวณค่าปัจจัยการขยายตัวความแปรปรวน

3. กำหนดจำนวนรังเริ่มต้นและฟังก์ชันเป้าหมายซึ่งเป็นเกณฑ์ในการพิจารณาเลือกตัวแบบการถดถอย โดยในงานวิจัยนี้ใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

4. คำนวณค่าสัมประสิทธิ์การถดถอยเริ่มต้น โดยวิธีการถดถอยแบบบริดจ์ที่มีการประมาณค่าคงตัว r ด้วยวิธีการสุ่มจากข้อ 1.

5. สุ่มค่าสัมประสิทธิ์การถดถอยโดยใช้ลักษณะการเคลื่อนที่ของ Lévy Flights

6. นำค่าสัมประสิทธิ์การถดถอยที่ได้มาคำนวณในฟังก์ชันเป้าหมาย

7. ดำเนินการค้นหาค่าสัมประสิทธิ์การถดถอยเซตต่อไปแล้วนำค่าที่ได้มาคำนวณในฟังก์ชันเป้าหมาย

8. นำผลลัพธ์ที่ได้มาเรียงลำดับและเลือกเซตคำตอบที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด

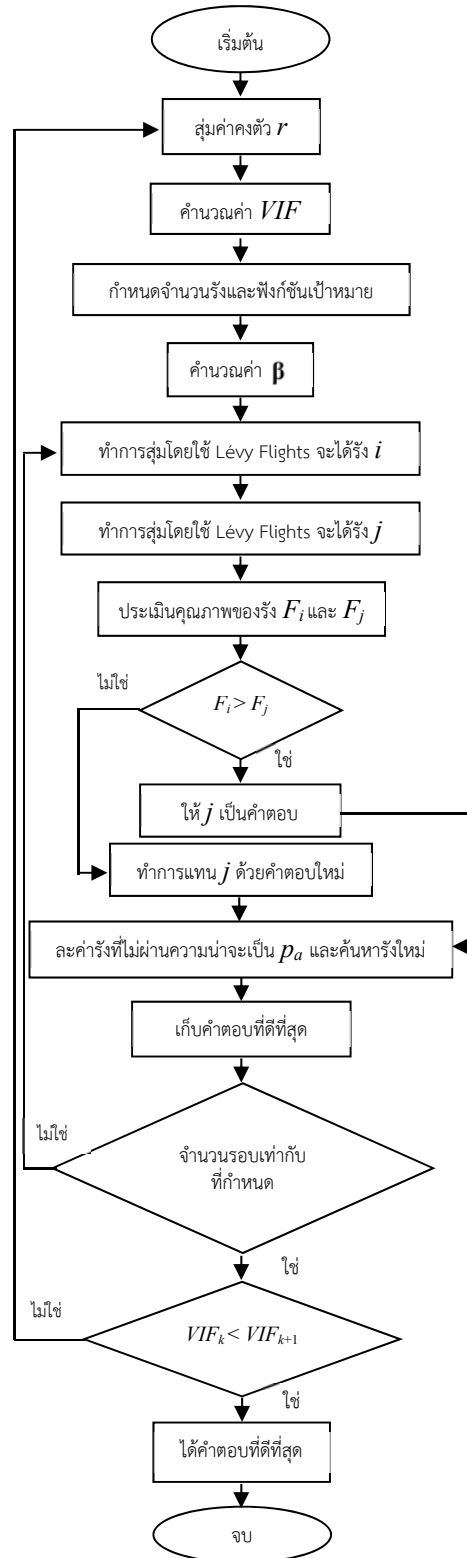
9. ค้นหารังใหม่ โดยพิจารณาจากค่าความน่าจะเป็นที่มากกว่าค่า p_a และมีเงื่อนไขว่ารังที่ไม่ผ่านค่าความน่าจะเป็น p_a ที่กำหนดจะถูกละทิ้งรังและรับค่ารังใหม่แบบสุ่ม

10. ทำการค้นหาคำตอบ (ค่าสัมประสิทธิ์การถดถอย) ในรอบต่อไป ทำจนครบรอบที่กำหนดไว้แล้วเลือกคำตอบที่ดีที่สุด

11. หลังจากที่ได้คำตอบที่ดีที่สุดแล้ว ทำการ เปรียบเทียบคำตอบอีกครั้งโดยใช้ค่าปัจจัยการขยายตัวความแปรปรวนเป็นเกณฑ์รายละเอียดแสดงดังรูปที่ 1

การคัดเลือกตัวแบบ

นำข้อมูลที่ได้จากการคำนวณตามที่กล่าวมาข้างต้นมาคัดเลือกตัวแบบที่เหมาะสมจากวิธีการประมาณค่าสัมประสิทธิ์การถดถอย 2 วิธี คือ วิธีการถดถอยแบบ ริดจ์และวิธีการค้นหาแบบนกกาเหว่า โดยพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าปัจจัยการขยายตัวความแปรปรวน จากนั้นบันทึกผลการคัดเลือก



รูปที่ 1 กระบวนการทำงานของวิธีการค้นหาแบบนกกาเหว่า

ผลการวิจัย

ในส่วนนี้ผู้วิจัยจะนำเสนอผลการวิจัยเป็น 2 ส่วน คือ ความสามารถในการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีการค้นหาแบบนกกาเหว่า เมื่อกำหนดจำนวนรอบในการสุ่มค่า r ที่ 5 10 และ 15 รอบ จำนวนครั้งเท่ากับ 20 ครั้ง และกระทำซ้ำ 100 500 1,000 5,000 และ 10,000 ครั้ง และผลการเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยของวิธีการถดถอยแบบบริดจ์และวิธีการค้นหาแบบนกกาเหว่า โดยมีรายละเอียดดังนี้

1. ความสามารถในการประมาณค่าสัมประสิทธิ์การถดถอยโดยวิธีการค้นหาแบบนกกาเหว่า สรุปได้ดังตารางที่ 1

จากตารางที่ 1 พบว่าค่าประมาณค่าสัมประสิทธิ์การถดถอยที่ประมาณได้จากวิธีการค้นหาแบบนกกาเหว่าที่ใช้ฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ย สำหรับข้อมูลชุดที่ 1 ค่า MSE ที่ได้จากการสุ่มค่า r เท่ากับ 5 รอบ จำนวนรอบในการกระทำซ้ำ 100 ครั้ง คือ 13.56584 ซึ่งเท่ากับค่า MSE ที่ได้จากการสุ่มค่า r เท่ากับ 15 รอบ จำนวนรอบในการกระทำซ้ำ 10,000 ครั้ง เป็นค่า MSE ต่ำสุด ดังนั้น สำหรับข้อมูลชุดที่ 1 จำนวนรอบของการสุ่ม และการทำซ้ำจึงไม่ส่งผลต่อค่า MSE แต่เมื่อพิจารณาจากค่า VIF_{max} พบว่าจะมีค่าลดลงอย่างต่อเนื่อง เมื่อมีการเพิ่มจำนวนรอบในการสุ่ม r และจำนวนรอบในการทำซ้ำ และลดลงถึงค่าต่ำที่สุดเท่ากับ 282.1708 ที่ r เท่ากับ 15 รอบ และกระทำซ้ำ 10,000 ครั้ง

จากที่กล่าวมานั้นทำให้ได้ตัวแบบที่เหมาะสมที่สุด คือ $\hat{y} = 3.8109 + 8.3618x_1 + 2.9041x_2 + 6.5657x_3$

สำหรับข้อมูลชุดที่ 2 ค่า MSE ที่ได้จากการสุ่มค่า r เท่ากับ 5 รอบ จำนวนรอบในการกระทำซ้ำ 10,000 ครั้ง ได้ค่า MSE ต่ำสุดเท่ากับ 1.62467 รองลงมาคือ 1.62664 ซึ่งเป็นค่า MSE ที่ได้จากการสุ่มค่า r เท่ากับ 15 รอบจำนวนรอบในการกระทำซ้ำ 10,000 ครั้ง แต่เมื่อพิจารณาจากค่า VIF_{max} พบว่า ค่า VIF_{max} ที่ได้จากการสุ่มค่า r เท่ากับ 15 รอบ จำนวนรอบในการกระทำซ้ำ 10,000 ครั้ง มีค่าต่ำกว่ากรณีที่กำหนดจำนวนรอบในการสุ่มค่า r เท่ากับ 5 รอบ จำนวนรอบในการกระทำซ้ำ 10,000 ครั้ง จากที่กล่าวมานั้นทำให้ได้

$$\begin{aligned} \hat{y} = & 443.8594 - 405.9960x_1 + 0.0132x_2 + 0.0082x_3 \\ & - 0.0084x_4 - 0.0324x_5 + 0.0266x_6 + 0.0302x_7 \\ & + 0.0171x_8 - 0.0130x_9 - 0.0035x_{10} - 0.0815x_{11} \\ & - 0.0511x_{12} + 0.0385x_{13} - 0.0130x_{14} \end{aligned}$$

ในส่วนของเวลาในการประมาณค่าสัมประสิทธิ์การถดถอยจะแปรผันตรงกับจำนวนรอบในการทำซ้ำที่เพิ่มขึ้นในทั้ง 2 ชุดข้อมูล

2. เปรียบเทียบผลการประมาณค่าสัมประสิทธิ์การถดถอยแต่ละวิธีสรุปได้ดังตารางที่ 2

จากตารางที่ 2 พบว่าการประมาณค่าสัมประสิทธิ์ การถดถอยด้วยวิธีการค้นหาแบบนกกาเหว่าในข้อมูลทั้ง 2 ชุด ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าปัจจัยการขยายตัวความแปรปรวนมากที่สุด ต่ำกว่าวิธีการถดถอยแบบบริดจ์ที่ใช้วิธีการประมาณค่าพารามิเตอร์บริดจ์ โดยวิธีของโฮเอิร์ล เคนนาร์ด และ บาลด์วิน โดยข้อมูลชุดที่ 1 ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 13.56584 และค่าปัจจัยการขยายตัวความแปรปรวนมากที่สุดเท่ากับ 282.1708 เมื่อค่าคงตัว r เท่ากับ 0.00000112 สำหรับข้อมูลชุดที่ 2 ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ 1.62664 และค่าปัจจัยการขยายตัวความแปรปรวนมากที่สุดเท่ากับ 4.0018 เมื่อค่าคงตัว r เท่ากับ 0.00351594

ตารางที่ 1 ค่าสัมประสิทธิ์การถดถอย ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าปัจจัยการขยายตัวความแปรปรวน โดยวิธีการค้นหาแบบนกกาเหว่า เมื่อกำหนดจำนวนรอบในการสุ่ม r เท่ากับ 5 10 และ 15 รอบ

ข้อมูลชุดที่ 1 จำนวนตัวแปรอิสระเท่ากับ 3 ตัว

จำนวนรอบ ในการสุ่ม r	จำนวนรอบในการทำซ้ำ	ค่าสัมประสิทธิ์การถดถอย				MSE	VIF _{max}	เวลา (วินาที)
		b_0	b_1	b_2	b_3			
5	100	3.8168	8.3630	2.9040	6.5650	13.56584	285.1072	01.08
	500	3.8189	8.3633	2.9040	6.5648	13.56585	283.4198	03.03
	1000	3.8124	8.3621	2.9041	6.5655	13.56584	282.8767	05.64
	5000	3.8327	8.3661	2.9038	6.5633	13.56591	282.6167	24.96
	10000	3.8229	8.3642	2.9040	6.5644	13.56586	282.4940	49.15
10	100	3.8106	8.3618	2.9041	6.5657	13.56584	283.1074	01.53
	500	3.8132	8.3623	2.9041	6.5655	13.56584	283.0622	05.22
	1000	3.8295	8.3655	2.9039	6.5636	13.56589	282.5958	09.71
	5000	3.8144	8.3625	2.9041	6.5653	13.56584	282.4559	44.96
	10000	3.8217	8.3640	2.9040	6.5645	13.56586	282.2897	89.01
15	100	3.8157	8.3628	2.9040	6.5651	13.56584	282.6090	01.91
	500	3.8217	8.3639	2.9040	6.5645	13.56586	282.5239	07.14
	1000	3.8166	8.3629	2.9040	6.5650	13.56584	282.4854	13.19
	5000	3.8268	8.3649	2.9039	6.5639	13.56588	282.3485	62.66
	10000	3.8109	8.3618	2.9041	6.5657	13.56584	282.1708	124.33

ข้อมูลชุดที่ 2 จำนวนตัวแปรอิสระเท่ากับ 14 ตัว

จำนวนรอบในการสุ่ม r	จำนวนรอบในการทำซ้ำ	ค่าสัมประสิทธิ์การถดถอย								
		b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
5	100	147.5652	-	0.0526	0.0128	-	0.0377	0.0888	0.1152	0.0653
			150.0013			0.1221				
	500	186.5306	-	0.0545	0.0115	-	0.0048	0.0887	0.1255	0.0600
			182.6464			0.1204				
	1000	216.2084	-	0.0548	0.0102	-	-	0.0861	0.1304	0.0541
		207.6483			0.1126	0.0186				
	5000	437.6437	-	0.0140	0.0067	-	-	0.0267	0.0410	0.0151
		400.6372			0.0092	0.0366				
	10000	446.6796	-	0.0129	0.0090	-	-	0.0266	0.0250	0.0180
		408.4074			0.0081	0.0306				

ตารางที่ 1 ค่าสัมประสิทธิ์การถดถอย ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าปัจจัยการขยายตัวความแปรปรวน โดยวิธีการค้นหาแบบนกกาเหว่า เมื่อกำหนดจำนวนรอบในการสุ่ม r เท่ากับ 5 10 และ 15 รอบ (ต่อ)

ข้อมูลชุดที่ 2 จำนวนตัวแปรอิสระเท่ากับ 14 ตัว

จำนวนรอบในการสุ่ม r	จำนวนรอบในการทำซ้ำ	ค่าสัมประสิทธิ์การถดถอย								
		b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8
10	100	169.1664	-	0.0542	0.0121	-	0.0193	0.0893	0.1208	0.0628
			168.0890			0.1226				
	500	218.8121	-	0.0545	0.0100	-	-	0.0857	0.1305	0.0535
			209.8500			0.1117	0.0206			
	1000	361.9463	-	0.0305	0.0027	-	-	0.0454	0.1145	0.0145
			333.8356			0.0377	0.0720			
15	5000	262.2825	-	0.0499	0.0078	-	-	0.0775	0.1340	0.0423
			246.9079			0.0926	0.0493			
	10000	404.1153	-	0.0204	0.0028	-	-	0.0318	0.0847	0.0098
			371.1767			0.0185	0.0573			
	100	436.4210	-	0.0142	0.0064	-	-	0.0267	0.0431	0.0147
			399.5771			0.0094	0.0374			
15	500	360.4046	-	0.0309	0.0028	-	-	0.0460	0.1151	0.0148
			332.4735			0.0385	0.0722			
	1000	378.6614	-	0.0265	0.0024	-	-	0.0395	0.1050	0.0115
			348.6250			0.0294	0.0680			
	5000	383.7768	-	0.0252	0.0024	-	-	0.0378	0.1016	0.0109
			353.1568			0.0270	0.0663			
10000	443.8594	-	0.0132	0.0082	-	-	0.0266	0.0302	0.0171	
		405.9960			0.0084	0.0324				

ข้อมูลชุดที่ 2 จำนวนตัวแปรอิสระเท่ากับ 14 ตัว (ต่อ)

จำนวนรอบในการสุ่ม r	จำนวนรอบในการทำซ้ำ	ค่าสัมประสิทธิ์การถดถอย						MSE	VIF _{max}	เวลา (วินาที)
		b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}			
5	100	0.0747	0.0885	-0.0506	0.0683	0.0662	-0.2409	13.33844	0.2416	1.86
	500	0.0712	0.0688	-0.0663	0.0531	0.0660	-0.3415	9.85335	0.2591	7.40
	1000	0.0662	0.0533	-0.0717	0.0413	0.0688	-0.3969	7.77124	0.2644	14.04
	5000	-0.0092	-0.0028	-0.0778	-0.0474	0.0430	-0.0346	1.63524	17.3510	85.24
	10000	-0.0147	-0.0040	-0.0834	-0.0529	0.0364	-0.0035	1.62467	19.5022	169.39
10	100	0.0727	0.0782	-0.0606	0.0600	0.0655	-0.3003	11.28505	0.2372	3.22
	500	0.0658	0.0520	-0.0713	0.0402	0.0690	-0.4008	7.61126	0.4615	13.07
	1000	0.0281	-0.0004	-0.0606	-0.0149	0.0770	-0.3006	2.30391	2.3508	33.07
	5000	0.0572	0.0305	-0.0716	0.0228	0.0752	-0.4373	5.30111	0.2456	162.02
	10000	0.0095	-0.0023	-0.0647	-0.0313	0.0629	-0.1590	1.79257	4.2163	311.99

ตารางที่ 1 ค่าสัมประสิทธิ์การถดถอย ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าปัจจัยการขยายตัวความแปรปรวน โดยวิธีการค้นหาแบบนกกาเหว่า เมื่อกำหนดจำนวนรอบในการสุ่ม r เท่ากับ 5 10 และ 15 รอบ (ต่อ)

ข้อมูลชุดที่ 2 จำนวนตัวแปรอิสระเท่ากับ 14 ตัว (ต่อ)

จำนวนรอบ ในการสุ่ม r	จำนวนรอบ ในการทำซ้ำ	ค่าสัมประสิทธิ์การถดถอย						MSE	VIF _{max}	เวลา (วินาที)
		b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}			
15	100	-0.0085	-0.0027	-0.0771	-0.0467	0.0438	-0.0390	1.63763	17.0928	7.34
	500	0.0287	-0.0002	-0.0607	-0.0144	0.0773	-0.3051	2.33016	2.2040	18.77
	1000	0.0214	-0.0019	-0.0610	-0.0212	0.0728	-0.2488	2.05511	2.2020	45.33
	5000	0.0191	-0.0021	-0.0614	-0.0232	0.0712	-0.2316	1.99143	2.1170	223.25
	10000	-0.0130	-0.0035	-0.0815	-0.0511	0.0385	-0.0130	1.62664	4.0018	411.84

ตารางที่ 2 เปรียบเทียบผลการประมาณค่าสัมประสิทธิ์การถดถอยแต่ละวิธี

ข้อมูลชุดที่ 1 จำนวนตัวแปรอิสระเท่ากับ 3 ตัว

วิธีการ	ค่าคงตัว r	ค่าสัมประสิทธิ์การถดถอย				MSE	VIF _{max}
		b_0	b_1	b_2	b_3		
วิธีการค้นหาแบบนกกาเหว่า	0.00000112	3.8109	8.3618	2.9041	6.5657	13.56584	282.1708
วิธีการถดถอยแบบบริดจ์	0.00002271	3.8265	8.3649	2.9039	6.5639	13.56588	285.9030

ข้อมูลชุดที่ 2 จำนวนตัวแปรอิสระเท่ากับ 14 ตัว

วิธีการ	ค่าคงตัว r	ค่าสัมประสิทธิ์การถดถอย							
		b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7
วิธีการค้นหาแบบนกกาเหว่า	0.00351594	443.8594	-405.9960	0.0132	0.0082	-0.0084	-0.0324	0.0266	0.0302
วิธีการถดถอยแบบบริดจ์	0.03085340	407.5415	-374.2087	0.0197	0.0030	-0.0172	-0.0555	0.0310	0.0812

ข้อมูลชุดที่ 2 จำนวนตัวแปรอิสระเท่ากับ 14 ตัว (ต่อ)

วิธีการ	ค่าสัมประสิทธิ์การถดถอย							MSE	VIF _{max}
	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}		
วิธีการค้นหาแบบนกกาเหว่า	0.0171	-0.0130	-0.0035	-0.0815	-0.0511	0.0385	-0.0130	1.62664	4.0018
วิธีการถดถอยแบบบริดจ์	0.0099	0.0077	-0.0023	-0.0656	-0.0327	0.0612	-0.1462	1.76721	6.6519

สรุปผลการวิจัย

งานวิจัยนี้ศึกษาการประมาณค่าสัมประสิทธิ์ การถดถอยสำหรับข้อมูลที่มีปัญหาสหสัมพันธ์เชิงเส้นพหุด้วยวิธีการถดถอยแบบบริดจ์และวิธีการค้นหาแบบนกกาเหว่า โดยใช้ข้อมูลที่ได้จากการจำลองและข้อมูลที่ได้จากสถานการณ์จริง และเปรียบเทียบผลการคำนวณโดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าปัจจัยการขยายตัวความแปรปรวนเป็นเกณฑ์ในการเลือกตัวแบบการถดถอยที่เหมาะสม ผลการวิจัยพบว่าวิธีการค้นหาแบบนกกาเหว่าที่ใช้ฟังก์ชันเป้าหมายเป็นค่าความคลาดเคลื่อนกำลังสองเฉลี่ย ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย และค่าปัจจัยการขยายตัวความแปรปรวนต่ำกว่าวิธีการถดถอยแบบบริดจ์ที่ใช้วิธีการประมาณค่าพารามิเตอร์ริคค์ โดยวิธีของ ไฮเอิร์ล เคนนาร์ต และบาลด์วิน

เอกสารอ้างอิง

- กานต์ณัฐ ณ บางช้าง. (2554). การคัดเลือกตัวแปรในตัวแบบการถดถอยเชิงเส้นพหุโดยใช้วิธีการค้นหาแบบต้องห้าม. วิทยาศาสตร์มหาบัณฑิต. คณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- จิราวัลย์ จิตถาวรเวช. (2558). การวิเคราะห์การถดถอย. โครงการส่งเสริมและพัฒนาเอกสารวิชาการ สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- ณัฐพงษ์ คำขาด. (2551). การศึกษาเปรียบเทียบประสิทธิภาพวิธีการพาร์ทีเคิลสวอม ออปติไมเซชันและเจเนติกอัลกอริทึม เพื่อการจัดเรียงเครื่องจักรในระบบการผลิตแบบยืดหยุ่น. วิทยาศาสตร์มหาบัณฑิต. มหาวิทยาลัยนเรศวร.
- รัฐพงศ์ ชัยเอิก และปริม ชูคาร. (2560). วิธีการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ. วิทยาศาสตร์บัณฑิต. สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.
- อรรถพล วัฒนสาครกุล. (2557). การจัดวางตำแหน่งการติดตั้งอุปกรณ์ซึ่งบอกความผิดพลาดที่เหมาะสมที่สุดในระบบจำหน่ายด้วยวิธีการค้นหาแบบนกกาเหว่า. วิศวกรรมศาสตรมหาบัณฑิต. มหาวิทยาลัยเกษตรศาสตร์.
- อารีรัตน์ เนื้อสีจัน. (2556). การประยุกต์ใช้วิธีการค้นหาแบบนกกาเหว่าเพื่อจัดเรียงเครื่องจักรในอุตสาหกรรม. วิศวกรรมศาสตรมหาบัณฑิต. สาขาวิชาวิศวกรรมการจัดการ มหาวิทยาลัยนเรศวร.
- Babak, A., S.T.A. Niakia, M.A. Khalifeb and Y. Faizeb. (2010). A Hybrid Variable Neighborhood Search and Simulated Annealing Algorithm to Estimate The Three Parameters of The Weibull Distribution. *Expert Systems with Applications* 38: 700-708.
- StatLib Datasets Archive. (2005). Body Fat. From: <http://lib.stat.cmu.edu/datasets/bodyfat>. Retrieved on August 6, 2018.
- Yang, X.-S., and Deb, S. (2009). Cuckoo Search via Levy flights. in *Proc of World Congress on Nature Biologically Inspired Computing* (Nabic 2009). USA: IEEE Publications. pp. 210-214.
- Yang, X.-S., and Deb, S. (2010). Engineering Optimization by Cuckoo Search. *Mathematical Modelling and Numerical Optimization* 1(4): 330-343.

