

วิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยเชิงเส้นพหุคูณ

Methods of Independent Variables Selection into Multiple Linear Regression Equations

สถัญญ์เทพ สุขแก้ว¹

¹ศษ.ด. (การวิจัยและประเมินทางการศึกษา) /ดร. สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช sarhistthep.suk@stou.ac.th

บทคัดย่อ

บทความนี้มุ่งนำเสนอวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยเชิงเส้นพหุคูณ โดยมีการให้รายละเอียดเกี่ยวกับขั้นตอนวิธีการคัดเลือกตัวแปรอิสระ ข้อดีและข้อจำกัดของวิธีการคัดเลือกตัวแปรอิสระ และแนวโน้มของการศึกษาและวิจัยที่เกี่ยวข้องกับวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอย วิธีการคัดเลือกตัวแปรอิสระสามารถแบ่งออกได้เป็น 2 กลุ่มวิธี คือ กลุ่มวิธีการทั่วไปที่นิยมใช้กันอย่างแพร่หลาย ซึ่งในบทความนี้จะกล่าวเฉพาะวิธีการถดถอยแบบเป็นขั้นตอน (Stepwise Regression) และกลุ่มวิธีการแบบ เมตาฮีริสติก (Metaheuristic) ซึ่งในบทความนี้จะกล่าวเฉพาะวิธีการค้นหาแบบต้องห้าม (Tabu Search) วิธีขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) และวิธีการจำลองการอบเหนียว (Simulated Annealing) โดยแต่ละวิธีมีความเหมาะสมสำหรับนำไปใช้ในสถานการณ์ที่มีความแตกต่างกัน

คำสำคัญ : วิธีการคัดเลือกตัวแปรอิสระ, วิธีการค้นหาแบบต้องห้าม, วิธีขั้นตอนวิธีเชิงพันธุกรรม, วิธีการจำลองการอบเหนียว

Abstract

This paper aims to present the methods of independent variables selection into multiple linear regression equations. It provides details about the independent variable selection algorithms, advantages, and limitations of independent variable selection methods, and trends in studies and research related to the independent variable selection methods into regression equations. The independent variable selection methods can be divided into two groups of methods: the general group of methods that is widely used. In this paper, we will discuss only Stepwise Regression and the metaheuristic group of methods which in this paper will discuss only Tabu Search, Genetic Algorithm, and Simulated Annealing. Each method is suitable for use in different situations.

Keywords: Methods of independent variables selection, Tabu Search, Genetic Algorithm, Simulated Annealing

บทนำ

การวิเคราะห์การถดถอย (Regression analysis) เป็นวิธีการวิเคราะห์ทางสถิติที่มีความเกี่ยวข้องกับการสร้างตัวแบบทางคณิตศาสตร์ เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรในการวิจัย โดยตัวแปรตัวหนึ่งที่เรียกว่า ตัวแปรตาม (Dependent variable) และตัวแปรอีกกลุ่มหนึ่งที่เรียกว่า ตัวแปรอิสระ (Independent variable) จำนวน 1 ตัวแปร หรือมากกว่า 1 ตัวแปรก็ได้ การสร้างตัวแบบทางคณิตศาสตร์เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรดังกล่าวมีวัตถุประสงค์เพื่อทำนายค่าของตัวแปรตามโดยอาศัยค่าของตัวแปรอิสระ และสถิติอื่น ๆ

การวิเคราะห์การถดถอยซึ่งเป็นที่รู้จักและนิยมนำไปใช้ในการศึกษา ค้นคว้า วิจัย แบ่งออกเป็น 2 ประเภท คือ ประเภทที่ 1 การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple linear regression analysis) เป็นการสร้างตัวแบบทางคณิตศาสตร์เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรอิสระ 1 ตัวแปรกับตัวแปรตาม 1 ตัวแปร โดยที่ตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์เชิงเส้นกัน และประเภทที่ 2 คือ การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple linear regression analysis) เป็นการสร้างตัวแบบทางคณิตศาสตร์ เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรอิสระมากกว่า 1 ตัวแปรกับตัวแปรตาม 1 ตัวแปร

โดยที่ตัวแปรอิสระแต่ละตัวแปรกับตัวแปรตามมีความสัมพันธ์เชิงเส้นกัน (วิชิต หล่อจิระขุนทดกุล, และจิราวัลย์ จิตรถเวช, 2548; วิรัช พานิชวงศ์, 2549)

การวิจัยที่มุ่งศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามที่ต้องใช้การวิเคราะห์การถดถอยในการวิเคราะห์ข้อมูลเพื่อตอบวัตถุประสงค์ของการวิจัยนั้น นักวิจัยต้องทำการทบทวนวรรณกรรมที่เกี่ยวข้อง เพื่อค้นหาตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามที่น่าสนใจศึกษา โดยส่วนใหญ่ นักวิจัยมักพบว่าตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามนั้นมีอยู่มากมายหลายตัวแปร เมื่อนักวิจัยต้องการสร้างตัวแบบทางคณิตศาสตร์เพื่อแสดงความสัมพันธ์ระหว่างตัวแปรตามและกลุ่มของตัวแปรอิสระจึงจำเป็นต้องอาศัยการวิเคราะห์ข้อมูลด้วยการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ ซึ่งมีความเสี่ยงที่อาจทำให้เกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง (Multicollinearity) ทำให้การประมาณค่าสัมประสิทธิ์การถดถอยที่ได้มีค่าไม่เสถียรและมีค่าความคลาดเคลื่อนสูง ในทางตรงกันข้ามหากมีตัวแปรอิสระจำนวนไม่มากนักที่มีความสัมพันธ์กันกับตัวแปรตาม อาจทำให้สมการถดถอยขาดตัวแปรอิสระที่สำคัญ ส่งผลให้ค่าทำนายที่ได้ มีความคลาดเคลื่อนสูงและเป็นปัญหาที่รุนแรงมากกว่าการมีตัวแปรอิสระที่ไม่สำคัญอยู่ในสมการถดถอย (นิสาชล งามประเสริฐสิทธิ์, และจิราวัลย์ จิตรถเวช, 2556; รัฐพงศ์ ชัยเอิก, ปริม ชูการ, และวรรณพร จันทโภาส, 2560) ส่งผลกระทบต่อความน่าเชื่อถือของผลการวิจัย นักวิจัยจึงจำเป็นต้องเลือกใช้วิธีการคัดเลือกตัวแปรอิสระเพื่อเข้าสู่สมการถดถอยให้เหมาะสม เพื่อให้ได้สมการถดถอยที่เหมาะสมที่สุด นั้นหมายความว่า ตัวแปรอิสระทุกตัวในสมการถดถอยต้องเป็นไปตามเงื่อนไขของการวิเคราะห์การถดถอย ประกอบด้วย ตัวแปรอิสระทั้งหลายที่อยู่ในสมการถดถอยต้องเป็นตัวแปรอิสระที่มีความสำคัญต่อตัวแปรตาม ตัวแปรอิสระเหล่านั้นจะต้องไม่มีความสัมพันธ์กันเองในสมการถดถอย และตัวแปรอิสระเหล่านั้นต้องร่วมกันอธิบายความแปรผันของตัวแปรตาม

วิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยเชิงเส้นพหุคูณที่นำเสนอในบทความนี้แบ่งออกเป็น 2 กลุ่มวิธี คือ กลุ่มวิธีที่ 1 เป็นกลุ่มวิธีการทั่วไปที่นักวิจัยคุ้นเคยเป็นอย่างดีและเป็นที่ยอมรับเลือกใช้ในการสร้างสมการถดถอย อีกทั้งยังเป็นวิธีที่มีให้เลือกใช้ในโปรแกรมสำเร็จรูปทางสถิติ ได้แก่ SPSS, Minitab, SAS และ R เป็นต้น ซึ่งมีอยู่มากมายหลายวิธี แต่ในบทความนี้จะขอกกล่าวเฉพาะวิธีการถดถอยแบบเป็นขั้นตอน (Stepwise regression) และกลุ่มวิธีที่ 2 เป็นกลุ่มวิธีการแบบเมตาฮีวิสติก ซึ่งประกอบด้วยวิธีการที่ถูกดัดแปลงมาจากวิธีการทางปัญญาประดิษฐ์ (Artificial intelligence) สำหรับนำมาใช้ในการค้นหาคำตอบที่เหมาะสมที่สุดของปัญหาที่มีความซับซ้อนและมีตัวแปรที่เข้ามาเกี่ยวข้องเป็นจำนวนมากซึ่งมีอยู่หลายวิธี สำหรับบทความนี้จะขอกกล่าวเฉพาะวิธีการค้นหาแบบต้องห้าม (Tabu Search) วิธีขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) และวิธีการจำลองการอบเหนียว (Simulated Annealing) ดังต่อไปนี้

1. วิธีการถดถอยแบบเป็นขั้นตอน (Stepwise regression)

วิธีการถดถอยแบบเป็นขั้นตอนเป็นวิธีการที่ผสมผสานขั้นตอนของวิธีการเลือกแบบไปข้างหน้า (Forward selection) และวิธีการกำจัดแบบถดถอยหลัง (Backward elimination) ดังนี้

ขั้นตอนวิธีของวิธีการถดถอยแบบเป็นขั้นตอน (ทรงศิริ แด่สมบัติ, 2548; วิชิต หล่อจิระขุนทดกุล, และจิราวัลย์ จิตรถเวช, 2548; วิรัช พานิชวงศ์, 2549; รัฐพงศ์ ชัยเอิก, ปริม ชูการ, และวรรณพร จันทโภาส, 2560; Montgomery, 2021) มีขั้นตอนดังต่อไปนี้

1. คำนวณสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple correlation coefficient) ระหว่างตัวแปรตามและตัวแปรอิสระทุกตัวแปร
2. เลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามสูงที่สุดเข้าสู่สมการถดถอยเป็นตัวแรก ตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามสูงที่สุดพิจารณาจากตัวแปรอิสระที่มีสัมประสิทธิ์สหสัมพันธ์อย่างง่ายสูงที่สุด โดยพิจารณาเฉพาะขนาดของความสัมพันธ์ซึ่งอยู่ในรูปของค่าที่เป็นตัวเลขของสัมประสิทธิ์สหสัมพันธ์อย่างง่ายเท่านั้น ไม่ต้องพิจารณาทิศทางของความสัมพันธ์ซึ่งแสดงด้วยเครื่องหมายบวก (+) หรือเครื่องหมายลบ (-)
3. ทดสอบสมมติฐานเพื่อหาข้อสรุปว่าตัวแปรอิสระตัวแรกที่เข้าสู่สมการถดถอยเหมาะสมในการทำนายตัวแปรตามหรือไม่ หากผลการทดสอบสมมติฐานพบว่าตัวแปรอิสระตัวแรกไม่เหมาะสมในการทำนายตัวแปรตาม ก็จะไม่นำตัวแปรอิสระตัวแรกเข้าสู่สมการถดถอย ขั้นตอนการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยจะสิ้นสุดลง แต่ถ้าผลการทดสอบสมมติฐานพบว่าตัวแปรอิสระตัวแรกเหมาะสมในการทำนายตัวแปรตาม ก็ให้นำตัวแปรอิสระตัวแรกเข้าสู่สมการถดถอย

4. คำนวณสัมประสิทธิ์สหสัมพันธ์บางส่วน (Partial correlation) ระหว่างตัวแปรตามกับตัวแปรอิสระที่เหลือทั้งหมด (ตัวแปรอิสระเหล่านั้นยังไม่ได้เข้าสู่สมการถดถอย) ในขณะที่ตัวแปรอิสระตัวแรกที่มีความเหมาะสมในการทำนายตัวแปรตาม (ที่ผ่านการทดสอบสมมติฐานเป็นที่เรียบร้อยแล้ว) เข้าไปอยู่ในสมการถดถอยแล้ว

5. เลือกตัวแปรอิสระที่มีสัมประสิทธิ์สหสัมพันธ์บางส่วนสูงที่สุดเข้าสู่สมการถดถอยเป็นตัวแปรต่อมา (โดยพิจารณาเฉพาะขนาดของความสัมพันธ์ ไม่ต้องพิจารณาทิศทางของความสัมพันธ์)

6. ทดสอบสมมติฐานเพื่อหาข้อสรุปว่าตัวแปรอิสระตัวต่อมา (ตัวใหม่) ที่เข้าสู่สมการถดถอยเหมาะสมในการทำนายตัวแปรตามหรือไม่ เมื่อตัวแปรอิสระตัวแรกอยู่ในสมการถดถอย หากผลการทดสอบสมมติฐานพบว่าตัวแปรอิสระตัวต่อมา (ตัวใหม่) ที่จะเข้าสู่สมการถดถอยไม่เหมาะสมในการทำนายตัวแปรตาม เมื่อตัวแปรอิสระตัวแรกอยู่ในสมการถดถอย ก็จะไม่นำตัวแปรอิสระตัวต่อมา (ตัวใหม่) เข้าสู่สมการถดถอย ถ้าผลการทดสอบสมมติฐานพบว่าตัวแปรอิสระตัวต่อมา (ตัวใหม่) ที่จะเข้าสู่สมการถดถอยเหมาะสมในการทำนายตัวแปรตาม เมื่อตัวแปรอิสระตัวแรกอยู่ในสมการถดถอย ก็จะนำตัวแปรอิสระตัวต่อมา (ตัวใหม่) เข้าสู่สมการถดถอย

7. ทดสอบสมมติฐานเพื่อหาข้อสรุปว่าตัวแปรอิสระตัวแรกที่เข้าสู่สมการถดถอยเหมาะสมในการทำนายตัวแปรตามหรือไม่ เมื่อตัวแปรอิสระตัวต่อมา (ตัวใหม่) อยู่ในสมการถดถอย หากผลการทดสอบสมมติฐานพบว่าตัวแปรอิสระตัวแรกที่เข้าสู่สมการถดถอยไม่เหมาะสมในการทำนายตัวแปรตาม เมื่อตัวแปรอิสระตัวต่อมา (ตัวใหม่) อยู่ในสมการถดถอย ก็จะนำตัวแปรอิสระตัวแรกออกจากสมการถดถอย ถ้าผลการทดสอบสมมติฐานพบว่าตัวแปรอิสระตัวแรกที่เข้าสู่สมการถดถอยเหมาะสมในการทำนายตัวแปรตาม เมื่อตัวแปรอิสระตัวต่อมา (ตัวใหม่) อยู่ในสมการถดถอย ตัวแปรอิสระตัวแรกจะยังคงอยู่ในสมการถดถอย

8. ทำซ้ำในขั้นตอนที่ 4 ถึงขั้นตอนที่ 7 จนกระทั่งไม่มีตัวแปรอิสระตัวใหม่ตัวใดที่จะสามารถเข้าสู่สมการถดถอย และไม่มีตัวแปรอิสระตัวใดในสมการถดถอยที่สามารถออกจากสมการถดถอย ขั้นตอนวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยโดยวิธีการถดถอยแบบเป็นขั้นตอนจึงจะสิ้นสุด

การคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยโดยวิธีการถดถอยแบบเป็นขั้นตอนเป็นวิธีที่มีการผสมผสานขั้นตอนของวิธีการเลือกแบบไปข้างหน้า (Forward selection) และวิธีการกำจัดแบบถดถอยหลัง (Backward elimination) ซึ่งเป็นการแก้ไขข้อบกพร่องของวิธีการเลือกแบบไปข้างหน้าที่มุ่งคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์กันสูงกับตัวแปรตามและเหมาะสมในการทำนายตัวแปรตามได้อย่างมีนัยสำคัญเข้าสู่สมการถดถอยเท่านั้น โดยไม่ได้ตรวจสอบผลกระทบที่เกิดจากการเพิ่มตัวแปรอิสระตัวใหม่เข้าไปในสมการถดถอยต่อตัวแปรอิสระที่เข้าไปในสมการถดถอยก่อนหน้านี้ (รัฐพงศ์ ชัยเอิก, ปริม ชูคาร และวรรณพร จันทโภาส, 2560) และแก้ไขข้อบกพร่องของวิธีการกำจัดแบบถดถอยหลังที่มุ่งคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามน้อยที่สุดและไม่เหมาะสมในการทำนายตัวแปรตามออกจากสมการถดถอย โดยวิธีการนี้จะไม่สามารถให้ข้อสรุปได้ว่าตัวแปรอิสระที่อยู่ในสมการแต่ละตัวสามารถทำนายตัวแปรตามได้หรือไม่ บอกได้เพียงว่าตัวแปรอิสระในสมการถดถอยนั้น สามารถร่วมกันทำนายตัวแปรตามได้ (ธิดาเดียว มยุรีสุวรรณ, 2559)

2. วิธีการค้นหาแบบต้องห้าม (Tabu Search)

วิธีการค้นหาแบบต้องห้ามเป็นวิธีการหาคำตอบวิธีการหนึ่งที่อยู่ในกลุ่มวิธีการแบบเมตาฮิวริสติกที่ถูกดัดแปลงมาจากวิธีการทางปัญญาประดิษฐ์ (Artificial intelligence) เพื่อค้นหาคำตอบที่มีความเหมาะสมที่สุดของปัญหาที่มีความซับซ้อนและมีตัวแปรที่เกี่ยวข้องเป็นจำนวนมาก โดยรายละเอียดเกี่ยวกับขั้นตอนในการทำงานของวิธีการค้นหาแบบต้องห้ามและขั้นตอนวิธีการค้นหาแบบต้องห้ามนั้น นิศาชล งามประเสริฐสิทธิ์, และจิราวัลย์ จิตรถเวช (2556) และรัฐพงศ์ ชัยเอิก, ปริม ชูคาร, และวรรณพร จันทโภาส (2560) กล่าวไว้ดังนี้

ขั้นตอนในการทำงานของวิธีการค้นหาแบบต้องห้ามสรุปได้ดังนี้ (1) หาคำตอบเริ่มต้น โดยจัดทำโครงสร้างหน่วยความจำมาใช้ในการเก็บค่า (2) สร้างคำตอบและเลือกคำตอบจากคำตอบที่สร้างขึ้น ภายใต้เงื่อนไขที่กำหนดเพื่อค้นหาคำตอบที่เหมาะสมที่สุด และ (3) ปรับปรุงคำตอบ โดยแทนคำตอบที่ดีที่สุดในปัจจุบันด้วยคำตอบที่ดีกว่า จากนั้นจึงเก็บคำตอบและบันทึกไว้

ขั้นตอนวิธีของวิธีการค้นหาแบบต้องห้าม มีขั้นตอนดังต่อไปนี้

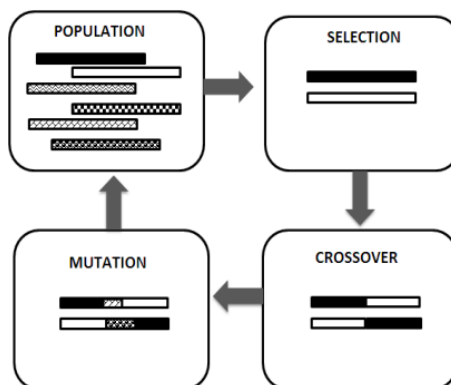
1. กำหนดค่าพิกัดเริ่มต้น โดยพิจารณาจากตัวประมาณค่าสัมประสิทธิ์การถดถอยที่ได้จากการประมาณค่าพารามิเตอร์โดยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS)

2. สร้างคำตอบเริ่มต้นโดยกำหนดค่าของตัวแปรอิสระต่าง ๆ จากการสุ่มค่าในพิสัยที่กำหนด พิสัยของตัวแปรอิสระแต่ละตัวอาจแตกต่างกันได้ตามการประมาณการเริ่มต้น

3. สร้างเขตคำตอบข้างเคียง (Neighborhood) ของค่าคำตอบปัจจุบันที่กำหนดขึ้น
 4. ตรวจสอบรายการต้องห้าม (Tabu list) ซึ่งเป็นรายการที่บันทึกข้อมูลกระบวนการค้นหาในอดีต ถ้าคำตอบข้างเคียงนั้นเป็นรายการต้องห้ามเกินระยะเวลาที่กำหนด ให้ยกเลิกการเป็นรายการต้องห้ามของคำตอบข้างเคียงนั้น
 5. ดำเนินการค้นหาคำตอบจากเขตคำตอบข้างเคียง ซึ่งต้องไม่เป็นรายการต้องห้ามและเป็นคำตอบที่ดีกว่าคำตอบที่มีค่าฟังก์ชันวัตถุประสงค์ที่สุดจากเขตคำตอบข้างเคียง
 6. กำหนดเกณฑ์ความปรารถนา (Aspiration criteria) ในการตรวจสอบคำตอบข้างเคียง ถ้าคำตอบข้างเคียงนั้นเป็นรายการต้องห้าม แต่สามารถให้คำตอบที่มีค่าฟังก์ชันวัตถุประสงค์ที่ดีกว่าก็สามารถเลือกมาเป็นคำตอบได้
 7. นำหน่วยความจำระยะยาว (Long term memory) มาใช้ในการหาคำตอบเริ่มต้นใหม่ ถ้าไม่สามารถหาคำตอบที่มีค่าฟังก์ชันวัตถุประสงค์ดีกว่าปัจจุบันได้
 8. ย้ายตำแหน่ง (Move) ไปยังคำตอบปัจจุบันที่ถูกเลือก แล้วปรับปรุงรายการต้องห้าม (Tabu list) ให้เป็นปัจจุบัน และเพิ่มจำนวนครั้งในการเลือกข้อมูลความถี่ (Frequency move) เป็นการใช้นิยามหน่วยความจำระยะยาวบันทึกการค้นหาทุกค่าตลอดช่วงของการค้นหา โดยใช้หลักการสร้างความหลากหลาย (Diversification) เพื่อค้นหาคำตอบในบริเวณที่แตกต่างจากที่ค้นพบมาแล้ว
 9. ทำซ้ำจนกระทั่งได้ค่าฟังก์ชันวัตถุประสงค์อยู่ในเกณฑ์ที่กำหนดขึ้นหรือครบตามจำนวนครั้งที่กำหนด
- การคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยโดยวิธีการค้นหาแบบต้องห้ามเหมาะกับกรณีที่ต้องการหาคำตอบที่มีความซับซ้อนและมีตัวแปรที่ใช้ในการศึกษาเป็นจำนวนมาก วิธีการนี้ประยุกต์ใช้กับการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยเชิงเส้นพหุคูณได้โดยการจัดการของขั้นตอนวิธี เพื่อให้ได้ตัวแปรอิสระที่เข้าสู่สมการถดถอยมีจำนวนน้อยที่สุดและมีฟังก์ชันวัตถุประสงค์คือ ความคลาดเคลื่อนต่ำสุด อีกทั้งยังประหยัดเวลาในการคัดเลือกตัวแปรอิสระ และสามารถใช้ได้กับข้อมูลทั้งที่ตัวแปรอิสระมีและไม่มีปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง (Multicollinearity)

3. วิธีขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm: GA)

วิธีขั้นตอนวิธีเชิงพันธุกรรมเป็นวิธีการแก้ปัญหาแบบหนึ่งเพื่อค้นหาคำตอบที่เหมาะสมที่สุด คิดค้นโดย John Holland ซึ่งได้รับแรงบันดาลใจจากทฤษฎีวิวัฒนาการของ Charles Darwin มีจุดมุ่งหมายเพื่ออธิบายการเปลี่ยนแปลงกระบวนการทางธรรมชาติของพันธุกรรมและนำกลไกการเปลี่ยนแปลงดังกล่าวมาประยุกต์ใช้ในการแก้ปัญหาการหาค่าเหมาะสมที่สุด หลักการทำงานของขั้นตอนวิธีเชิงพันธุกรรมถูกนำเสนอข้อมูลในรูปแบบโครโมโซม ขั้นตอนการทำงานของขั้นตอนวิธีเชิงพันธุกรรม เริ่มต้นจากการสร้างประชากรของคำตอบที่เรียกว่า โครโมโซมจากการสุ่ม จากนั้นจึงทำการถอดรหัสโครโมโซมและคำนวณค่าความเหมาะสม (Fitness value) ของแต่ละโครโมโซมจากฟังก์ชันวัตถุประสงค์ ประชากรเหล่านี้จะต้องผ่านตัวดำเนินการทางพันธุกรรม เพื่อให้เกิดการปรับเปลี่ยนสายพันธุ์ ซึ่งมีอยู่ด้วยกัน 3 วิธี ได้แก่ การคัดเลือกสายพันธุ์ (Selection) การสลับสายพันธุ์ (Cross over) และการกลายพันธุ์ (Mutation) (ศิรินทิพย์ หมั่นจัทร์ และวรา มินเสน, 2557; Deep & Thakur, 2007) ซึ่งกระบวนการดังกล่าวจะถูกทำซ้ำไปเรื่อย ๆ จนกว่าจะตรงกับเงื่อนไขในการหยุดค้นหา จะได้โครโมโซมที่เหมาะสมที่สุดหรือใกล้เคียงค่าที่ดีที่สุด (ศิรินทิพย์ หมั่นจัทร์ และวรา มินเสน, 2557; Luca, 2013)



ภาพที่ 1 ขั้นตอนการทำงานของขั้นตอนวิธีเชิงพันธุกรรม
ที่มา: ศิรินทิพย์ หมั่นจัทร์, และวรา มินเสน (2557)

ขั้นตอนวิธีของวิธีขั้นตอนวิธีเชิงพันธุกรรม (ศิริทิพย์ หมั่นจัทร์ และวรา มินเสน, 2557; รัฐพงศ์ ชัยเอิก, ปริม ชูคาร และวรรณพร จันทภาส, 2560; พีรภัทร์ ภาควิกมลเลิศ, จรัสศรี รุ่งรัตนอุบล และอนามัย นาคูม, 2563) มีขั้นตอนดังต่อไปนี้

1. สร้างประชากรของคำตอบ โดยอาศัยการสุ่ม หรือเรียกว่าโครโมโซมจากการสุ่ม เพื่อเป็นประชากรเริ่มต้น หรือประชากรรุ่นที่ 1 สมาชิกของประชากรเริ่มต้นจะอยู่ในรูปของสายโครโมโซม ใน 1 โครโมโซมจะประกอบด้วยยีนหลายยีนแบบเลขฐานสอง (Binary) และยีนแต่ละยีนแทนตัวแปรอิสระแต่ละตัวแปร ดังภาพที่ 2



ภาพที่ 2 สายโครโมโซมและยีนแบบเลขฐานสอง

ที่มา: พีรภัทร์ ภาควิกมลเลิศ, จรัสศรี รุ่งรัตนอุบล และอนามัย นาคูม (2563)

2. แปลงรหัสจากโครโมโซมให้อยู่ในรูปของตัวแปรอิสระที่จะอยู่ในสมการถดถอย ใน 1 สายโครโมโซมเปรียบได้กับ 1 สมการถดถอยที่มีจำนวนตัวแปรอิสระเท่ากับจำนวนยีน ณ ตำแหน่งของยีนใดที่ปรากฏเลข 1 หมายความว่า มีตัวแปรอิสระที่มีตัวเลขแสดงตำแหน่งของตัวแปรอิสระตรงกับตำแหน่งของยีนนั้น ๆ หาก ณ ตำแหน่งของยีนใดที่ปรากฏเลข 0 หมายความว่า ไม่มีตัวแปรอิสระที่มีตัวเลขแสดงตำแหน่งของตัวแปรอิสระตรงกับตำแหน่งของยีนนั้น ๆ จากภาพ 2 จะพบว่าตัวแปรอิสระ X_1 , X_4 , X_5 และ X_6 ในสมการถดถอย

3. คำนวณค่าฟังก์ชันวัตถุประสงค์ซึ่งก็คือ ค่าเกณฑ์สารสนเทศของอะไกกิ (Akaike's Information Criterion: AIC) ของแต่ละโครโมโซมที่ใช้เป็นเกณฑ์ในการคัดเลือกสมการถดถอยที่เหมาะสมที่สุด โดยสมการถดถอยที่เหมาะสมที่สุดจะมีค่า AIC ต่ำที่สุด

4. คัดเลือกประชากรเพื่อกำหนดเป็นโครโมโซมต้นแบบหรือโครโมโซมพ่อแม่ของประชากรรุ่นที่ 1 โดยพิจารณาจากโครโมโซมที่ให้ค่าเกณฑ์สารสนเทศของอะไกกิที่มีค่าน้อยที่สุด 2 อันดับ

5. สลับสายพันธุ์ โดยกำหนดความน่าจะเป็นในการสลับสายพันธุ์เท่ากับ 0.8 จากนั้นทำการสุ่มตัวเลขในช่วง 0 ถึง 1 ถ้าหากตัวเลขที่สุ่มมาได้มีค่าน้อยกว่าหรือเท่ากับความน่าจะเป็นในการสลับสายพันธุ์คือ 0.8 จึงจะทำการสลับสายพันธุ์

6. ทำการสลับสายพันธุ์แบบ 2 จุด

7. สุ่มเพื่อเลือกตำแหน่งของยีนที่จะทำการสลับสายพันธุ์ พร้อมทั้งทำการสลับสายพันธุ์

8. กลายพันธุ์ โดยกำหนดความน่าจะเป็นของการกลายพันธุ์เท่ากับ 0.2 จากนั้นทำการสุ่มตัวเลขในช่วง 0 ถึง 1 ถ้าหากตัวเลขที่สุ่มมาได้มีค่าน้อยกว่าหรือเท่ากับความน่าจะเป็นของการกลายพันธุ์คือ 0.2 จึงจะทำการกลายพันธุ์ ในกรณีที่มีการกลายพันธุ์ ตำแหน่งของยีนและจำนวนตำแหน่งที่จะทำการกลายพันธุ์ต้องเป็นไปอย่างสุ่ม จากนั้นจึงทำการกลายพันธุ์

9. แปลงรหัสจากโครโมโซมที่ได้ให้อยู่ในรูปของตัวแปรอิสระที่มีความเหมาะสมที่จะอยู่ในสมการถดถอย พร้อมทั้งคำนวณค่าเกณฑ์สารสนเทศของอะไกกิ โดยสมการถดถอยที่เหมาะสมจะมีค่าเกณฑ์สารสนเทศของอะไกกีน้อยที่สุด ซึ่งจะเป็นสมการถดถอยที่เหมาะสมที่สุด

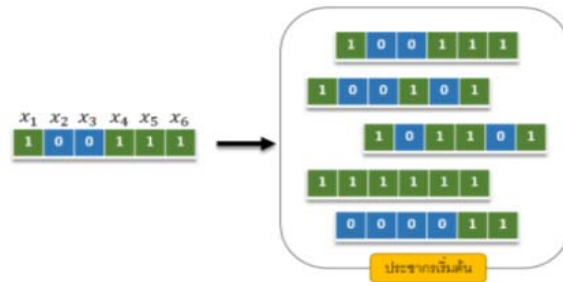
10. สำหรับการพิจารณาโครโมโซมรุ่นต่อไป จะต้องดำเนินการตามขั้นตอนที่ 1 ถึงขั้นตอนที่ 9 จนครบจำนวนรุ่นที่กำหนด

4. วิธีการจำลองการอบเหนียว (Simulated Annealing)

ขั้นตอนวิธีการจำลองการอบเหนียวเป็นเทคนิคการค้นหาคำตอบจากการสุ่มจุดตามที่สำรวจซ้ำจากการพัฒนาคำตอบเบื้องต้นเพื่อหาผลเฉลยอื่นที่ดีกว่า กระบวนการทำงานพื้นฐานหลัก ๆ ได้แก่ การสร้างคำตอบเบื้องต้น (Initial solution) โดยการวนสุ่มจุดขึ้นมาจนได้หนึ่งคำตอบ จากนั้นเข้าสู่กระบวนการปรับปรุงคำตอบด้วยการสุ่มจุด (Random swap) เพื่อค้นหาคำตอบใหม่ (Candidate solution) ที่คาดว่าจะดีขึ้นด้วยวิธีการค้นหาคำตอบใกล้เคียง (Neighborhood function) ทำซ้ำไปเรื่อย ๆ โดยใช้คำตอบที่ดีที่สุดที่รู้มาเป็นหลักฐานในการคำนวณ แต่อย่างไรก็ตามวิธีการค้นหาคำตอบใกล้เคียงที่ใช้การสุ่มจุดอาจไม่สามารถให้คำตอบที่ดีที่สุด (ปิยรัตน์ งามสนธิ, ธรา อังสกุล, และจิตมนต์ อังสกุล, 2560)

ขั้นตอนวิธีของวิธีการจำลองการอบเหี่ยว (ปิรัตน์ งามสนิท, ธรา อังสกุล และจิตมินต์ อังสกุล, 2560; พีรภัทร์ ภาควงมิมลเลิศ, จรัสศรี รุ่งรัตนอุบล, และอนามัย นาอุดม, 2563) มีขั้นตอนดังต่อไปนี้

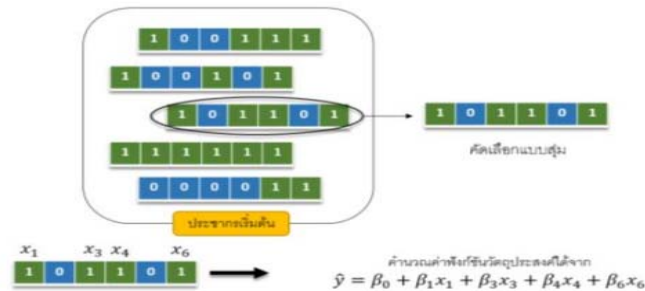
1. สุ่มประชากรเริ่มต้นจากเวกเตอร์ที่ใช้รหัสแบบไบนารีที่มีลักษณะเป็นเลขฐานสอง โดยมีความยาวเท่ากับจำนวนตัวแปรอิสระ ขึ้นมา 1 ชุด ดังภาพที่ 3 พร้อมทั้งกำหนดค่าพารามิเตอร์เริ่มต้น และกำหนดเกณฑ์การหยุดกระบวนการทำงานคือ การกำหนดรอบของการทำงาน



ภาพที่ 3 การสร้างประชากรเริ่มต้น

ที่มา: พีรภัทร์ ภาควงมิมลเลิศ, จรัสศรี รุ่งรัตนอุบล, และอนามัย นาอุดม (2563)

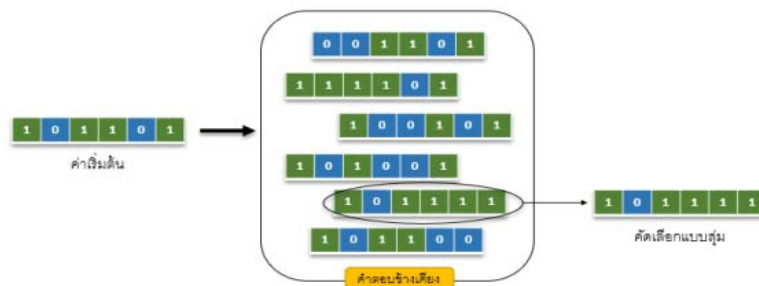
2. สุ่มเวกเตอร์ผลเฉลยเริ่มต้นจากประชากรเริ่มต้น มา 1 เวกเตอร์ ดังภาพที่ 4 พร้อมคำนวณค่าฟังก์ชันวัตถุประสงค์และกำหนดให้เป็นผลเฉลยที่ดีที่สุดที่สุด



ภาพที่ 4 การสุ่มผลเฉลยเริ่มต้นจากประชากรเริ่มต้น

ที่มา: พีรภัทร์ ภาควงมิมลเลิศ, จรัสศรี รุ่งรัตนอุบล, และอนามัย นาอุดม (2563)

3. สร้างเซตของเวกเตอร์ผลเฉลยข้างเคียงของผลเฉลยเริ่มต้น โดยเวกเตอร์ผลเฉลยข้างเคียงจะมีความแตกต่างจากเวกเตอร์ผลเฉลยเริ่มต้นเวกเตอร์ละ 1 ตำแหน่ง แล้วทำการสุ่มเวกเตอร์ ผลเฉลยข้างเคียงมา 1 เวกเตอร์ พร้อมทั้งคำนวณค่าฟังก์ชันวัตถุประสงค์ ดังภาพที่ 5



ภาพที่ 5 การสุ่มเวกเตอร์ผลเฉลยข้างเคียงจากเซตของเวกเตอร์ผลเฉลยข้างเคียง

ที่มา: พีรภัทร์ ภาควงมิมลเลิศ, จรัสศรี รุ่งรัตนอุบล, และอนามัย นาอุดม (2563)

4. คำนวณผลต่างของค่าฟังก์ชันวัตถุประสงค์ระหว่างผลเฉลยเริ่มต้นและผลเฉลยข้างเคียง
5. ตรวจสอบว่าควรมีการแทนที่ผลเฉลยเดิมหรือไม่ โดยใช้ความน่าจะเป็นในการตัดสินใจตามเงื่อนไขที่กำหนดไว้
6. นำผลเฉลยที่ได้จากขั้นตอนที่ 5 มาเปรียบเทียบกับผลเฉลยที่ดีที่สุดที่ได้ทำการบันทึกไว้เพื่อพิจารณาการแทนที่ผลเฉลยที่ดีที่สุด โดยจะทำการเปรียบเทียบค่าฟังก์ชันวัตถุประสงค์ของผลเฉลยที่ดีที่สุดที่บันทึกไว้กับค่าฟังก์ชันวัตถุประสงค์ของผลเฉลยที่ได้จากขั้นตอนที่ 5 หากค่าฟังก์ชันวัตถุประสงค์ของผลเฉลยที่ดีที่สุดที่บันทึกไว้มีค่าที่ต่ำกว่าจะทำการบันทึกผลเฉลยที่ดีที่สุดเดิมไว้ แต่ถ้าค่าฟังก์ชันวัตถุประสงค์ของผลเฉลยที่ได้จากขั้นตอนที่ 5 มีค่าที่ต่ำกว่าจะทำการแทนที่ให้เป็นผลเฉลยที่ดีที่สุดใหม่แทน
7. ปรับลดค่าพารามิเตอร์ ตามฟังก์ชันที่ได้กำหนดไว้
8. ตรวจสอบการหยุดกระบวนการทำงาน โดยพิจารณาจากการแทนที่ของผลเฉลยที่ดีที่สุด ถ้าผลเฉลยที่ดีที่สุดไม่ได้รับการแทนที่ จะหยุดกระบวนการทำงาน

แนวโน้มของการศึกษาและวิจัยที่เกี่ยวข้องกับวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอย

การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ ได้มีนักวิจัยนำมาใช้ในการวิจัยเป็นจำนวนมาก โดยประเด็นที่เกี่ยวข้องกับวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอย เป็นประเด็นที่ได้รับความสนใจและนำมาศึกษาอาทิเช่น

Hasan (2013) ทำการศึกษาการคัดเลือกสมการถดถอยในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยเสนอวิธีการแบบผสม (Hybrid GSA) ระหว่างวิธี GA และ SA ในการคัดเลือกตัวแปรของสมการถดถอยพร้อมทั้งเปรียบเทียบวิธีดังกล่าวกับวิธี GA FS และ BE ในกรณีที่ไม่เกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง และศึกษาตัวแปรอิสระตั้งแต่ 11 ถึง 25 ตัวแปรที่ขนาดตัวอย่าง 100 และ 200 เกณฑ์ที่ใช้ในการพิจารณาเลือกสมการถดถอยคือ ค่า AIC ผลการศึกษาพบว่า การเปรียบเทียบค่า AIC วิธี Hybrid GSA ให้ค่าต่ำกว่าวิธี FS และ BE แต่ให้ค่าที่ใกล้เคียงกับวิธี GA เมื่อตัวแปรเพิ่มขึ้น และให้ผลที่คล้ายกันทั้งสองขนาดตัวอย่าง

พัทธ์ชนก ศรีสุระเดชชัย และชญาณิน อินทวง (2563) ทำการวิจัยโดยใช้ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) กับตัวดำเนินการข้ามสายพันธุ์ทั้ง 6 วิธี ได้แก่ การข้ามสายพันธุ์แบบจุดเดียว การข้ามสายพันธุ์แบบสองจุด การข้ามสายพันธุ์แบบ $m/2$ จุด การข้ามสายพันธุ์แบบ $m - 1$ จุด การข้ามสายพันธุ์แบบเอกรูป และการข้ามสายพันธุ์แบบสับเปลี่ยน โดยใช้ทั้งข้อมูลจริงและข้อมูลที่ได้จากการจำลองจากทั้งตัวแบบการถดถอยเชิงเส้นและการถดถอยลอจิสติกทวิภาค โดยใช้เกณฑ์สารสนเทศของอะไกเกะ (Akaike's Information Criterion: AIC) ในการคัดเลือกตัวแปรอิสระ สำหรับข้อมูลที่จำลอง จะสมมติให้ตัวแปรอิสระไม่มีความสัมพันธ์กันและมีความสัมพันธ์แบบ AR1 (First – order autoregressive) ผลลัพธ์ที่ได้นำมาเปรียบเทียบกับวิธีคัดเลือกตัวแปรแบบการเลือกแบบไปข้างหน้า การกำจัดแบบถดถอยหลัง และการกำจัดแบบสองทิศทาง ผลการศึกษาพบว่าขั้นตอนเชิงพันธุกรรมสามารถหาตัวแบบที่มีค่า AIC ต่ำกว่าวิธีคัดเลือกตัวแปรแบบขั้นตอน และเมื่อเปรียบเทียบตัวดำเนินการข้ามสายพันธุ์ทั้ง 6 พบว่าการข้ามสายพันธุ์แบบ $m - 1$ จุด จะให้สมการถดถอยที่มีความเหมาะสมน้อยกว่าตัวดำเนินการข้ามสายพันธุ์แบบอื่น ๆ อย่างมีนัยสำคัญทางสถิติ

พิรภัทร ภาคภูมิภมิลเลิศ, จรัสศรี รุ่งรัตนอาบล และอนามัย นาอุดม (2563) ทำการวิจัยเพื่อศึกษาการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยเชิงเส้นพหุคูณด้วยวิธีการค้นหาแบบขั้นตอนเชิงพันธุกรรม (Genetic Algorithm: GA) และการค้นหาแบบจำลองการอบเหนียว (Simulated Annealing: SA) และทำการเปรียบเทียบประสิทธิภาพและความรวดเร็วในการทำงานของทั้ง GA และ SA ในสถานการณ์ที่เกิดและไม่เกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง ทำการศึกษาโดยการจำลองข้อมูล และกำหนดจำนวนตัวแปรอิสระเท่ากับ 5 10 และ 20 ตามลำดับและแต่ละขนาดของตัวแปรอิสระ กำหนดขนาดตัวอย่างเท่ากับ 50 100 200 และ 500 ตามลำดับ ทำการเปรียบเทียบประสิทธิภาพของ GA และ SA โดยพิจารณาจากค่าร้อยละความถูกต้องของการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอย และพิจารณาจำนวนรอบในการทำงานเป็นเกณฑ์ในการเปรียบเทียบความรวดเร็วในการทำงาน ผลการศึกษาพบว่า GA และ SA สามารถทำงานได้ดีทั้งในสถานการณ์ที่ไม่เกิดและเกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง โดยที่ GA ให้ค่าร้อยละความถูกต้องของการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยสูงกว่า SA ในกรณีที่ตัวแปรอิสระมีจำนวนน้อย และทั้ง 2 วิธีสามารถคัดเลือกตัวแปรอิสระได้ถูกต้องร้อยละ 100 เมื่อจำนวนตัวแปรอิสระมีค่าเพิ่มขึ้น นอกจากนี้ยังพบว่าทั้ง GA และ SA มีความรวดเร็วในการทำงานใกล้เคียงกันเมื่อตัวแปรอิสระมีจำนวนน้อย ในขณะที่ SA สามารถทำงานได้รวดเร็วกว่า GA เมื่อตัวแปรอิสระมีจำนวนมากขึ้น

จากงานวิจัยที่เกี่ยวข้องกับวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยพบว่า แนวโน้มของงานวิจัยจะเป็นไปในทิศทางที่มุ่งศึกษาเปรียบเทียบวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยในวิธีการต่าง ๆ ทั้งวิธีการที่อยู่ในกลุ่มวิธีทั่วไปที่เป็นที่นิยมใช้กันอย่างแพร่หลาย และวิธีการในกลุ่มวิธีการแบบเมตาฮีริสติก ข้อมูลที่นำมาใช้ในการศึกษาทั้งข้อมูลจริงและข้อมูลที่ได้จากการจำลอง ในสถานการณ์ที่เกิดและไม่เกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง (Multicollinearity) จำนวนตัวแปรอิสระที่ใช้ในการศึกษาที่มีจำนวนน้อยไปจนถึงจำนวนมาก ขนาดตัวอย่างที่ใช้ในการศึกษาก็มีการศึกษาทั้งตัวอย่างขนาดเล็กไปจนถึงตัวอย่างขนาดใหญ่ เกณฑ์ที่ใช้ในการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยเพื่อให้ได้สมการถดถอยที่เหมาะสมที่สุด นิยมใช้เกณฑ์สารสนเทศของอะไกเก้ (Akaike's Information Criterion: AIC) รวมถึงค่าอื่น ๆ ที่ใช้ในการศึกษาประสิทธิภาพของวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอย เช่น ค่าร้อยละ ความถูกต้องของการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอย และจำนวนรอบในการทำงานซึ่งใช้ในการเปรียบเทียบความรวดเร็วในการทำงาน เป็นต้น ผลการวิจัยที่ได้ยังไม่สามารถนำไปสู่การหาข้อสรุปได้อย่างแน่ชัดว่าวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยวิธีการใดดีที่สุด เนื่องจากในแต่ละวิธีจะมีความเหมาะสมกับสถานการณ์ หรือเงื่อนไข หรือบริบทที่เกี่ยวข้องซึ่งมีความแตกต่างกัน และเป็นไปได้ว่าในอนาคตนักวิจัยอาจจะมีการพัฒนาหรือผสมผสานวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยวิธีการต่าง ๆ ที่มีอยู่ในปัจจุบันให้กลายเป็นวิธีการใหม่ ๆ ที่มีประสิทธิภาพมากขึ้น

สรุป

ในการวิจัยที่มุ่งศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระนั้น มักจะมีตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามเป็นจำนวนมาก ซึ่งตัวแปรอิสระทั้งหลายได้มาจากการที่นักวิจัยทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องการวิเคราะห์ข้อมูลในงานวิจัยที่มุ่งศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระทั้งหลายเพื่อค้นหาสมการที่อธิบายความสัมพันธ์ระหว่างตัวแปรตามและกลุ่มของตัวแปรอิสระมีข้อควรระวังอยู่ประการหนึ่งคือ ตัวแปรอิสระจำนวนมากที่อยู่ในสมการถดถอยอาจทำให้เกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง (Multicollinearity) ซึ่งขัดแย้งกับวัตถุประสงค์หลักของการวิเคราะห์การถดถอยที่มุ่งค้นหากลุ่มของตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตาม เพื่อให้ตัวแปรอิสระเหล่านั้นร่วมกันอธิบายความแปรผันในตัวแปรตาม แต่ถ้าปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเองเกิดขึ้น จะทำให้ตัวแปรอิสระเหล่านั้นอธิบายความผันแปรระหว่างตัวแปรอิสระด้วยกันเอง ตัวแปรอิสระเหล่านั้นจะไม่สามารถอธิบายความผันแปรในตัวแปรตาม ซึ่งขัดแย้งกับวัตถุประสงค์หลักของการวิเคราะห์การถดถอย การแก้ไขปัญหาดังกล่าววิธีการหนึ่งก็จะต้องพึ่งพาการคัดเลือกตัวแปรอิสระ ในอดีตที่ผ่านมามักจะใช้วิธีการคัดเลือก ตัวแปรอิสระด้วยวิธีการทั่วไป โดยเฉพาะวิธีการถดถอยแบบเป็นขั้นตอน (Stepwise regression) เพราะวิธีนี้เป็นวิธีที่คำนึงถึงตัวแปรอิสระที่เข้าและออกจากสมการถดถอย ตัวแปรอิสระแต่ละตัวที่เข้าสู่สมการถดถอยต้องสามารถอธิบายความแปรผันในตัวแปรตาม และในขณะเดียวกันตัวแปรอิสระทั้งหลายเหล่านั้นยังต้องร่วมกันกับตัวแปรอิสระตัวอื่น ๆ ในการอธิบายความแปรผันในตัวแปรตาม โดยที่ตัวแปรอิสระเหล่านั้นจะต้องไม่มีความสัมพันธ์กันเอง นักวิจัยบางท่านเลือกใช้วิธีการถดถอยแบบเป็นขั้นตอนในการสร้างสมการถดถอยตามประเพณีนิยม โดยไม่ทราบเหตุผลว่าเพราะเหตุใดจึงจะต้องเลือกใช้วิธีการนี้ ผู้เขียนหวังเป็นอย่างยิ่งว่าหากผู้อ่านได้อ่านบทความนี้ น่าจะช่วยให้ทราบถึงเหตุผลได้ไม่มากนัก เมื่อผู้อ่านศึกษาขั้นตอนวิธีการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยแบบเป็นขั้นตอนจะพบว่า หากคำนวณค่าต่าง ๆ ด้วยมือ เพื่อหาสมการถดถอย คงจะกระทำได้ด้วยความยุ่งยาก เนื่องจากขั้นตอนของวิธีการนี้ค่อนข้างซับซ้อน และอาจใช้ระยะเวลาในการค้นหาคำตอบ แต่เมื่อมีการพัฒนาโปรแกรมสำเร็จรูปทางสถิติซึ่งมีอยู่มากมายหลายโปรแกรมให้นักวิจัยได้เลือกใช้ อาทิเช่น SPSS, Minitab, SAS และ R นักวิจัยสามารถค้นหาสมการถดถอยด้วยวิธีการถดถอยแบบเป็นขั้นตอนได้อย่างรวดเร็วและสะดวกมากยิ่งขึ้น

ความก้าวหน้าของวิทยาการคอมพิวเตอร์ในปัจจุบันนี้ที่สามารถเรียนรู้และเข้าใจกลไกการทำงานของสมองของมนุษย์มีการพัฒนาอย่างรวดเร็วและถูกนำไปใช้ในการทำงานหลากหลายด้านด้วยกัน เพราะการทำงานในยุคปัจจุบันเต็มไปด้วยปัญหาที่มีความสลับซับซ้อน มีเงื่อนไข และข้อจำกัด รวมถึงกฎเกณฑ์ต่าง ๆ ที่เข้ามาเกี่ยวข้องเป็นจำนวนมาก ปัญหาที่เกิดขึ้นอาจมีลักษณะที่เป็นปัญหาแบบเดิม ๆ เกิดขึ้นซ้ำ ๆ กัน ที่ใช้วิธีการแก้ไขปัญหาแบบเดิม ๆ ก็สามารถแก้ปัญหาได้ และอาจพบปัญหาแบบใหม่ที่ไม่สามารถแก้ปัญหาด้วยวิธีการแบบเดิม ๆ ได้ จำเป็นต้องแก้ปัญหาด้วยวิธีการใหม่ ๆ ทางออกของการแก้ปัญหาที่อาจพบเจอในหลากหลายรูปแบบสามารถได้รับการแก้ปัญหาด้วยวิธีการที่หลากหลายซึ่งถูกพัฒนาขึ้น เพื่อค้นหาผลเฉลยที่เหมาะสมที่สุด เช่นเดียวกันกับปัญหาในการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอย นักวิจัยต้องการค้นหาสมการถดถอยที่มีความเหมาะสมที่สุดในการทำนายค่าของตัวแปรตามที่นักวิจัยสนใจศึกษา ภายใต้ข้อจำกัดที่ว่าตัวแปรอิสระทั้งหลายที่อยู่ในสมการถดถอยจะต้องไม่เกิด

ปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันเอง นักสถิติจึงได้มีการนำองค์ความรู้ทางวิทยาการคอมพิวเตอร์และองค์ความรู้ด้านวิธีการหาค่าที่เหมาะสมที่สุดมาพัฒนาเป็นขั้นตอนวิธีในการคัดเลือกตัวแปรอิสระ เช่น วิธีการค้นหาแบบต้องห้าม (Tabu search) วิธีขั้นตอนวิธีเชิงพันธุกรรม (Genetic algorithm) และวิธีการจำลองการอบเหนียว (Simulated annealing) เป็นต้น ซึ่งจัดอยู่ในกลุ่มวิธีการแบบเมตาฮิวริสติก (Metaheuristic) เพื่อช่วยค้นหาสมการถดถอยที่เหมาะสมและดีที่สุดได้อย่างแม่นยำ ในระยะเวลาที่รวดเร็ว แม้ตัวแปรอิสระที่เข้ามาเกี่ยวข้องมีจำนวนมาก ข้อจำกัดประการหนึ่งที่มีความสำคัญของกลุ่มวิธีการแบบเมตาฮิวริสติกคือ ยังไม่มีโปรแกรมสำเร็จรูปทางสถิติโปรแกรมใดที่มีคำสั่งที่ใช้ในการคัดเลือกตัวแปรอิสระด้วยวิธีการดังกล่าว จึงทำให้เป็นเรื่องยากในการที่นักวิจัยจะใช้กลุ่มวิธีการนี้ เพราะหากนักวิจัยต้องการค้นหาสมการถดถอยด้วยกลุ่มวิธีการนี้ นักวิจัยจำเป็นต้องเขียนชุดคำสั่งขึ้นมาใช้ในการวิเคราะห์ข้อมูล กลุ่มวิธีการแบบเมตาฮิวริสติกจึงอาจจะเหมาะสมกับนักวิจัยที่มีความสามารถในการเขียนชุดคำสั่งมากกว่านักวิจัยที่ไม่มีความรู้พื้นฐานในเรื่องของการเขียนชุดคำสั่ง แม้ว่ากลุ่มวิธีการแบบเมตาฮิวริสติกจะช่วยให้นักวิจัยได้สมการถดถอยที่เหมาะสมที่สุด แต่ก็ยังไม่แน่ชัดว่าวิธีใดเป็นวิธีการคัดเลือกตัวแปรอิสระที่ดีที่สุด เนื่องจากแต่ละวิธีมีความเหมาะสมกับสถานการณ์ที่มีความแตกต่างกัน

ข้อเสนอแนะ

1. กรณีที่มีตัวแปรอิสระในงานวิจัยเป็นจำนวนไม่มากและตัวแปรอิสระที่เกี่ยวข้องไม่มีความซับซ้อน นักวิจัยควรเลือกใช้วิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยด้วยวิธีการถดถอยแบบเป็นขั้นตอน
 2. กรณีที่มีตัวแปรอิสระในงานวิจัยเป็นจำนวนมาก นักวิจัยควรเลือกใช้วิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอยด้วยวิธีการที่อยู่ในกลุ่มวิธีการแบบเมตาฮิวริสติก โดยที่นักวิจัยต้องมีความสามารถในการเขียนชุดคำสั่ง
- ผู้เขียนหวังเป็นอย่างยิ่งว่าบทความนี้ที่ให้รายละเอียดเกี่ยวกับการคัดเลือกตัวแปรอิสระซึ่งมีอยู่มากมายหลากหลายวิธี ในแต่ละวิธีหรือกลุ่มวิธีก็มีข้อดี และข้อจำกัดที่แตกต่างกัน จะเป็นประโยชน์ต่อผู้อ่าน เพื่อจะได้นำไปใช้เป็นข้อมูลในการตัดสินใจเลือกวิธีการคัดเลือกตัวแปรอิสระเข้าสู่สมการถดถอย เพื่อให้ได้สมการถดถอยที่เหมาะสมที่สุด

เอกสารอ้างอิง

- ทรงศิริ แด่สมบัติ. (2548). *การวิเคราะห์การถดถอย*. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.
- ธิดาเดียว มยุรีสุวรรณ. (2559). *การวิเคราะห์การถดถอย*. ขอนแก่น : เพ็ญพรินต์.
- นิตาชล งามประเสริฐสุทธิ, และจิราวัลย์ จิตรถเวช. (2556). การเปรียบเทียบการคัดเลือกตัวแปรอิสระที่มีปัญหาสหสัมพันธ์เชิงเส้นพหุด้วยวิธีการถดถอยแบบบริดจ์และการค้นหาแบบต้องห้าม. *วารสารวิทยาศาสตร์บูรพา*, 18(2), 76 – 89.
- ปิยรัตน์ งามสนธิ, ธรา อังสกุล, และจิตินันต์ อังสกุล. (2560). ขั้นตอนวิธีการจำลองการอบเหนียวสำหรับการวางแผนแผนการเดินทางท่องเที่ยวภายใต้เงื่อนไขงบประมาณด้านเวลา. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม*, 36(6), 713 – 727.
- พัทธ์ชนก ศรีสุรเดชชัย, และชญานิษฐ์ อินทวง. (2563). ประสิทธิภาพของตัวดำเนินการข้ามสายพันธุ์ในขั้นตอนวิธีเชิงพันธุกรรมสำหรับการคัดเลือกตัวแปรในการวิเคราะห์การถดถอย. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์*, 28(9), 1517 – 1533.
- พีรภัทร ภาคภูมิภักดิ์, จรัสศรี รุ่งรัตนอุบล, และอนามัย นาอุดม. (2563). การเปรียบเทียบวิธีคัดเลือกตัวแปรอิสระในแบบการถดถอยเชิงเส้นพหุโดยใช้ขั้นตอนวิธีเชิงพันธุกรรมและการจำลองการอบเหนียว. *วารสารวิชาการวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครสวรรค์*, 12(16), 11 – 25.
- รัฐพงศ์ ชัยเอิก, ปริม ชูการ, และวรรณพร จันทโภาส. (2560). *วิธีการคัดเลือกตัวแปรในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ*. ขอนแก่น : มหาวิทยาลัยขอนแก่น.
- วิจิต หล่อจิระขุนทด, และจิราวัลย์ จิตรถเวช. (2548). *เทคนิคการพยากรณ์*. กรุงเทพฯ: โรงพิมพ์องค์การสงเคราะห์ทหารผ่านศึก.
- วิรัช พานิชวงศ์. (2549). *การวิเคราะห์การถดถอย*. กรุงเทพฯ: ศูนย์ผลิตตำราเรียน มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- ศรินทิพย์ หมั่นจันทร์, และวรา มินเสน. (2557). การคัดเลือกตัวแปรในการถดถอยเชิงเส้นพหุคูณโดยใช้วิธีดัดแปลงเจเนติกอัลกอริทึม. *วารสารวิทยาศาสตร์บูรพา*, 19(2), 139 – 153.
- Deep, K., & Thakur, M. (2007). A new mutation operator for real coded genetic algorithms. *Applied Mathematics and Computation*, 193(1), 211 – 230.

- Hasan, Örkücü, H. (2013). Subset selection in multiple linear regression models: A hybrid of genetic and simulated annealing algorithms. *Applied Mathematics and Computation*, 219(23), 11018 – 11028.
- Scrucca, L. (2013). GA : A Package for Genetic Algorithm in R. *Journal of Statistical Software*, 53(4), 1 – 37.
- Montgomery, D. C., Peck, E. A. & Vining, G. G. (2021). *Introduction to linear regression analysis*. New Jersey: John Wiley & Sons.

