



การใช้การถดถอยกำลังสองน้อยที่สุดบางส่วน
ในกรณีการละเมิดข้อสมมุติของตัวแบบการถดถอย
Using partial least square regression
in the case of violations of regression model assumptions

นพมาศ อัครจันทโชติ^{1*} และ กิดากการ สายธนู²

¹ สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ
สมุทรปราการ 10540

² ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ชลบุรี 20131

Noppamas Akarachantachote^{1*} and Kidakan Saithanu²

¹ Division of Mathematics and Statistics, Faculty of Science and Technology,
Huachiew Chalermprakiet University, Samutprakarn 10540

² Department of Mathematics, Faculty of Science, Burapha University, Chonburi 20131

บทคัดย่อ

การถดถอยกำลังสองน้อยที่สุดบางส่วนเป็นการสร้างตัวแบบแสดงความสัมพันธ์ระหว่างตัวแปรตามและกลุ่มของตัวแปรอิสระหรือตัวแปรทำนาย โดยอยู่บนพื้นฐานของตัวแปรแฝงที่สร้างจากผลรวมเชิงเส้นของตัวแปรทำนาย โดยตัวแปรแฝงที่ได้ไม่มีความสัมพันธ์กันและมีความแปรปรวนร่วมกับตัวแปรตามสูง จึงเหมาะสมที่จะใช้ในการวิเคราะห์ข้อมูลกรณีที่มีจำนวนตัวแปรทำนายมาก ซึ่งตัวแปรทำนายเหล่านี้มักมีความสัมพันธ์เชิงเส้นแบบพหุ เช่น ข้อมูลไมโครอาร์เรย์และข้อมูลจากการวัดสเปกตรัม ซึ่งหากใช้การถดถอยวิเคราะห์ข้อมูลดังกล่าวนี้จะเป็นการละเมิดข้อสมมุติของการวิเคราะห์การถดถอยเนื่องจากทำให้ความแปรปรวนของสัมประสิทธิ์การถดถอยมีค่าสูงและนำไปสู่การทำนายที่ไม่แม่นยำ ดังนั้นการถดถอยกำลังสองน้อยที่สุดบางส่วนจึงเป็นอีกทางเลือกหนึ่งในการศึกษาความสัมพันธ์ของตัวแปรตามและตัวแปรทำนาย

คำสำคัญ: การถดถอยกำลังสองน้อยที่สุดบางส่วน ความสัมพันธ์เชิงเส้นแบบพหุ ตัวแปรแฝง



Abstract

Partial least squares regression (PLS-regression) is a related modeling between dependent variable and group of independent or predictor variables basing on latent variables built from linear combination of predictor variables. These latent variables are uncorrelated and high covariance with the dependent variable so it is appropriate to the case of large number of predictor variables which having multicollinearity such as microarray and spectrum data. If regression analysis is analyzed in such those cases, it violates the assumptions of regression analysis because it leads to the high variance of regression coefficient also inaccurate prediction. Partial least squares regression is therefore the alternative for studying relationship between dependent and predictor variables.

Keywords: PLS-regression, Multicollinearity, Latent variable

บทนำ

การวิเคราะห์การถดถอย (regression analysis) เป็นวิธีเชิงสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรตาม (dependent variable) และตัวแปรอิสระ (independent variable) หรือตัวแปรทำนาย (predictor variable) เพื่อพยากรณ์หรือทำนายค่าของตัวแปรตามโดยใช้ตัวแปรอิสระหรือตัวแปรทำนาย ซึ่งในการวิเคราะห์การถดถอยนั้นมีข้อสมมติ (assumptions) ของตัวแบบการถดถอยดังนี้

1. ความคลาดเคลื่อนมีการแจกแจงปกติ (normal distribution for the error term)
2. ความคลาดเคลื่อนเป็นอิสระกัน (independence of the error term)
3. ความคลาดเคลื่อนมีความแปรปรวนคงที่ (constant variance of the error term)
4. ตัวแปรอิสระต้องไม่มีความสัมพันธ์เชิงเส้นแบบพหุ (multicollinearity)

หากมีการวิเคราะห์ข้อมูลด้วยการถดถอยโดยมีการละเมิด (violation) ข้อสมมติของตัวแบบการถดถอยเพียงข้อใดข้อหนึ่ง แล้วอาจนำไปสู่ผลการวิเคราะห์ที่ไม่น่าเชื่อถือได้ ซึ่งข้อมูลในการวิเคราะห์การถดถอยนั้นบ่อยครั้งที่มีเกี่ยวข้องกับตัวแปรทำนายเป็นจำนวนมาก อาทิ ข้อมูลไมโครอาร์เรย์ (microarray data) ซึ่งเป็นข้อมูลที่ได้จากการศึกษารูปแบบการแสดงออกของยีนของสิ่งมีชีวิตหลาย ๆ ยีน

พร้อม ๆ กัน โดยยีนที่ศึกษามีจำนวนเป็นหลักพันหรือหลักหมื่น ยีนเหล่านี้สามารถนำมาใช้เป็นตัวแปรทำนายในการจำแนกเนื้อเยื่อมะเร็งและเนื้อเยื่อปกติได้ [1] ข้อมูลจากการวัดสเปกตรัมซึ่งประกอบด้วยการวัดความถี่คลื่นที่มีจำนวนนับพัน สามารถใช้ในการทำนายค่าคุณลักษณะคุณภาพของผลิตภัณฑ์ เช่น การทำนายปริมาณความชื้นและปริมาณไขมันในรำข้าว [2] การทำนายปริมาณแบคทีเรียในน้ำมันโคคิบบ [3] และการทำนายปริมาณน้ำมันปาล์มในผลปาล์มแบบไม่ทำลายผลปาล์ม [4] เป็นต้น ข้อมูลคุณสมบัติของโครงสร้างโมเลกุลของสารประกอบซึ่งมีคุณสมบัติเป็นจำนวนมากที่ใช้เป็นตัวแปรในการทำนายการออกฤทธิ์ทางชีวภาพของยา [5] ซึ่งในการศึกษาข้อมูลดังกล่าวนี้มีตัวแปรอิสระหรือตัวแปรทำนายอยู่เป็นจำนวนมาก และตัวแปรเหล่านี้มักมีความสัมพันธ์กัน การใช้การวิเคราะห์การถดถอยในการวิเคราะห์ข้อมูลลักษณะนี้ทำให้ผลลัพธ์ของการวิเคราะห์การถดถอยไม่แม่นยำ เนื่องจากความแปรปรวนของสัมประสิทธิ์การถดถอยจะมีค่าสูง อีกทั้งในบางครั้งขนาดตัวอย่างของข้อมูลที่ศึกษาอาจมีจำนวนน้อยกว่าจำนวนตัวแปรอิสระหรือตัวแปรทำนาย จึงทำให้ไม่สามารถวิเคราะห์ข้อมูลด้วยการถดถอยได้

การใช้การวิเคราะห์การถดถอยส่วนประกอบหลัก (principal component regression: PCR) ในการวิเคราะห์ข้อมูล เป็นทางเลือกหนึ่งของการแก้ไขปัญหาการเกิดความ



สัมพันธ์เชิงเส้นแบบพหุ โดยสร้างส่วนประกอบหลัก (principal component: PC) หรือตัวแปรแฝง (latent variable) ซึ่งเป็นการรวมกลุ่มของตัวแปรทำนายที่มีความสัมพันธ์กัน โดยตัวแปรแฝงที่สร้างขึ้นไม่มีความสัมพันธ์เชิงเส้นแบบพหุ ตัวแปรแฝงที่สร้างขึ้นแทนตัวแปรทำนายจากการวิเคราะห์การถดถอยส่วนประกอบหลักนี้จะรวมความแปรปรวนของตัวแปรทำนายเดิมไว้ให้มากที่สุด แต่ไม่ได้ยืนยันว่ามีความเหมาะสมสำหรับอธิบายตัวแปรตามได้มากที่สุด นอกจากนี้ในการวิเคราะห์จะต้องใช้ตัวอย่างจำนวนมาก การถดถอยกำลังสองน้อยที่สุดบางส่วน (partial least square regression: PLS-regression) ใช้หลักการคล้ายกับการวิเคราะห์การถดถอยส่วนประกอบหลักในด้านการสร้างตัวแปรแฝงเพื่อลดขนาดของตัวแปรทำนาย แต่ตัวแปรแฝงที่สร้างขึ้นโดยการถดถอยกำลังสองน้อยที่สุดบางส่วน มุ่งให้เกิดความแปรปรวนร่วมระหว่างตัวแปรแฝงและตัวแปรตามสูงที่สุด ด้วยเหตุผลของข้อจำกัดเกี่ยวกับข้อสมมติของตัวแบบการถดถอยและแนวทางการสร้างตัวแปรแฝงดังกล่าว ทำให้การถดถอยกำลังสองน้อยที่สุดบางส่วน เป็นวิธีที่ได้รับความนิยมจากนักวิจัยเป็นอย่างมาก ในกว่า 20 ปีที่ผ่านมา [6, 7] เนื่องจากการถดถอยกำลังสองน้อยที่สุดบางส่วนไม่มีข้อสมมติเกี่ยวกับการแจกแจงของประชากร และสามารถจัดการกับข้อมูลที่มีขนาดตัวอย่างเล็ก แต่เกี่ยวข้องกับตัวแปรเป็นจำนวนมากได้เป็นอย่างดี การถดถอยกำลังสองน้อยที่สุดบางส่วนจึงถูกเรียกว่าเป็น soft modeling [8]

การถดถอยกำลังสองน้อยที่สุดบางส่วน (partial least squares regression: PLS-regression)

วิธีกำลังสองน้อยที่สุดบางส่วน (partial least squares: PLS) เป็นวิธีเชิงสถิติที่ใช้สำหรับการวิเคราะห์หลายตัวแปรเพื่อสร้างตัวแบบความสัมพันธ์ระหว่างกลุ่มของตัวแปรทำนาย (predictor variable) โดยอาศัยตัวแปรแฝง (latent variable) ซึ่งมีจุดเริ่มต้นในปี ค.ศ. 1966 จากการพัฒนาของ Herman Wold [9] เพื่อสร้างตัวแบบเส้นทาง (path modeling) สำหรับการจัดการปัญหาทางด้านเศรษฐศาสตร์ ในปี ค.ศ. 1975 หลังจากนั้นในช่วงทศวรรษที่ 1980 Svante Wold บุตรชายของเขาและเพื่อนได้ร่วมกันพัฒนาวิธีกำลังสองน้อยที่สุดให้สามารถใช้ได้กับการวิเคราะห์การถดถอยสำหรับการสร้างตัวแบบเคมีเมทริกซ์ (chemometrics) และสเปกโตรเมตริก (spectrometric) [6] และเรียกวิธีเชิงสถิติดังกล่าวนี้ว่า การถดถอยกำลังสองน้อยที่สุดบางส่วน (partial least squares regression: PLS-regression)

ขั้นตอนวิธีการสร้างตัวแบบการถดถอยกำลังสองน้อยที่สุดบางส่วน

การถดถอยกำลังสองน้อยที่สุดบางส่วนเป็นการสร้างตัวแบบเชิงเส้นระหว่างกลุ่มของตัวแปรตามและกลุ่มของตัวแปรอิสระหรือตัวแปรทำนาย โดยจะสร้างตัวแปรใหม่ซึ่งเป็นผลรวมเชิงเส้นของตัวแปรเดิม และเรียกตัวแปรใหม่นี้ว่า ตัวแปรแฝง หรือส่วนประกอบ (component) โดยตัวแปรแฝงที่สร้างขึ้นใหม่นี้มีจำนวนน้อยกว่าหรือเท่ากับจำนวนตัวแปรเดิมและเป็นอิสระกัน ในบทความนี้จะกล่าวถึงเฉพาะกรณีที่กลุ่มของตัวแปรตามมีตัวแปรเพียง 1 ตัว เท่านั้น

กำหนดให้ $X = (x_1, x_2, \dots, x_n)'$ แทนเมทริกซ์ของตัวแปรอิสระหรือตัวแปรทำนายขนาด $n \times p$ โดยที่ n แทนขนาดตัวอย่าง และ p แทนจำนวนตัวแปรอิสระหรือตัวแปรทำนาย เมื่อ $x_j = (x_{j1}, x_{j2}, \dots, x_{jp})$

$y = (y_1, y_2, \dots, y_n)'$ แทนเวกเตอร์ของตัวแปรตาม

$T = (t_1, t_2, \dots, t_h)$ แทนเมทริกซ์ของตัวแปรแฝง h ตัว ซึ่งถูกสกัดจากตัวแปรอิสระหรือตัวแปรทำนายโดยความสัมพันธ์ของ X และตัวแปรแฝง แสดงได้ดังสมการที่ (1)

$$T = XW^* \tag{1}$$



โดยที่ $W^* = (w_1^*, w_2^*, \dots, w_h^*)$ เป็นเมทริกซ์ขนาด $p \times h$ ซึ่งมีสมาชิกที่ประกอบด้วยเวกเตอร์น้ำหนัก (weight vector) ของตัวแปรแฝงที่ถูกสกัดจากตัวแปรอิสระหรือตัวแปรทำนาย

X และ y ถูกแยกองค์ประกอบให้อยู่ในรูปแบบดังสมการที่ (2) และสมการที่ (3) ตามลำดับ

$$X = TP' + E \quad (2)$$

$$y = Tq' + f \quad (3)$$

โดยที่ $P = (p_1, p_2, \dots, p_h)$ เป็นเมทริกซ์ขนาด $p \times h$ แทนค่าถ่วง (Loading) ของตัวแปรอิสระหรือตัวแปรทำนาย และ $q = (q_1, q_2, \dots, q_h)$ เป็นเวกเตอร์ขนาด h แทนค่าถ่วงของตัวแปรตาม โดยทั่วไปแล้ว จะประมาณค่า P และ q ด้วยวิธีกำลังสองน้อยที่สุดสามัญ (ordinary least square: OLS) E เป็นเมทริกซ์ขนาด $n \times p$ และ f เป็นเวกเตอร์ขนาด n ซึ่งสมาชิกในเมทริกซ์หรือเวกเตอร์นั้นแสดงค่าส่วนเหลือ (residual) ของตัวแปรอิสระหรือตัวแปรทำนาย และตัวแปรตาม ตามลำดับ

ขั้นตอนวิธีการสร้างตัวแบบการถดถอยกำลังสองน้อยที่สุดบางส่วน (partial least squares model: PLS-model) มีหลายแนวทางซึ่งแตกต่างกัน โดยขึ้นอยู่กับปัจจัยหลายปัจจัย เช่น ฟังก์ชันเป้าหมาย (target function) สำหรับการหาค่าถ่วงและขั้นตอนวิธีในการหาค่าถ่วง เป็นต้น [6]

หลักการสร้างตัวแปรแฝงโดยการหาค่าถ่วงที่ทำให้ความแปรปรวนร่วมระหว่างตัวแปรตาม และตัวแปรแฝงมีค่าสูงสุด ทำได้โดยหาค่าพารามิเตอร์ของตัวแปรแฝงที่ละตัวตามลำดับ ดังนี้

กำหนดตัวแปรแฝงที่ 1 เป็น $t_1 = Xw_1$ แล้วจะหาค่า w_1 ที่ทำให้ $\max[\text{cov}(Xw_1, y)]^2$ โดยที่ $w_1'w_1 = 1$ และโดยวิธีตัวคูณลากรางจ์ จะได้ว่า

$$w_1 = X'y / \|X'y\| \quad (4)$$

สำหรับตัวแปรแฝงถัดไปจะลดค่า X และ y ลง โดยลบด้วยค่าประมาณของ X และ y ตามลำดับ ซึ่ง X และ y ลดค่าลงเป็นค่าส่วนเหลือ E และ f ตามลำดับ นั่นคือในการหาค่าพารามิเตอร์ของตัวแปรแฝงที่ 2 จะใช้ E และ f แทน X และ y จากนั้นดำเนินการในลักษณะเดียวกับการหาพารามิเตอร์ของตัวแปรแฝงที่ 1 โดยทำซ้ำในลักษณะเช่นเดียวกันนี้สำหรับตัวแปรแฝงอื่น ๆ จึงสามารถสรุปขั้นตอนวิธีการสร้างตัวแปรแฝงจำนวน h ตัว ได้ดังนี้ [10]

1. กำหนดค่าเริ่มต้น $E_0 = X$ และ $f_0 = y$
2. สำหรับ $a=1, \dots, h$ ทำซ้ำ

$$2.1 \ w_a = E'_{a-1}f_{a-1} / \|E'_{a-1}f_{a-1}\|$$

$$2.2 \ t_a = E_{a-1}w_a$$

$$2.3 \ p_a = E'_{a-1}t_a / t'_a t_a$$

$$2.4 \ q_a = f'_{a-1}t_a / t'_a t_a$$

$$2.5 \ E_a = E_{a-1} - t_a p'_a$$

$$2.6 \ f_a = f_{a-1} - q_a t_a$$



ก่อนที่จะเริ่มเข้าสู่ขั้นตอนดังกล่าวนี้จะต้องมีการเตรียมข้อมูลด้วยการทำให้เป็นค่ามาตรฐานเพื่อจัดปัญหาในเรื่องหน่วยของตัวแปรที่มีค่าแตกต่างกัน โดยการลบค่าของแต่ละตัวแปรด้วยค่าเฉลี่ยของตัวแปรนั้น ๆ แล้วหารด้วยส่วนเบี่ยงเบนมาตรฐานของตัวแปรนั้น ๆ เพื่อปรับขนาดของข้อมูลในแต่ละตัวแปรให้มีความแปรปรวนเท่ากับ 1

สำหรับขั้นตอนวิธีการหาค่าตัวแปรแฝงที่ a หรือ t_o หาได้จากผลคูณของค่าส่วนเหลือกับค่าถ่วง w_o ความสัมพันธ์ระหว่างค่าถ่วง w_o และ w_o^* ซึ่งค่าหลังเป็นค่าถ่วงที่เกี่ยวข้องกับ X โดยตรง เป็นดังสมการที่ (5)

$$W^* = W(P'W)^{-1} \tag{5}$$

ส่วนการพยากรณ์หรือทำนายตัวแปรตามสามารถคำนวณได้ดังสมการที่ (6)

$$\begin{aligned} \hat{y} &= T(T'T)^{-1}T'y \\ &= Xb_{PLS} \end{aligned} \tag{6}$$

โดยที่ $b_{PLS} = W(P'W)^{-1}(T'T)^{-1}T'y$ เป็นเวกเตอร์สัมประสิทธิ์การถดถอยกำลังสองน้อยที่สุดบางส่วน

การกำหนดจำนวนตัวแปรแฝง

ถึงแม้ว่าจะสามารถสกัดจำนวนตัวแปรแฝงได้เท่ากับจำนวนลำดับที่ (rank) ของ X ได้ก็ตาม แต่ตัวแปรแฝงที่สำคัญซึ่งสกัดจากสารสนเทศของ X และ y มีเพียงตัวแปรแฝงแรก ๆ เท่านั้น ดังนั้นการกำหนดจำนวนตัวแปรแฝงจึงเป็นสิ่งสำคัญ ซึ่งในงานวิจัยแต่ละงานมีการกำหนดตัวแปรแฝงที่แตกต่างกันไปดังนี้

1. การกำหนดจำนวนตัวแปรแฝงที่แน่นอนเป็น 3 ถึง 5 ตัว เช่น งานของ Nguyen และ Rocke [1] และ Nguyen และ Rocke [11] กำหนดจำนวนแฝงของยีนเป็น 3 ตัวแปร
2. การกำหนดจำนวนตัวแปรแฝงด้วยการเปรียบเทียบความสอดคล้องของตัวแบบเมื่อมีตัวแปรแฝงจำนวน a ตัว บนข้อมูลทดสอบ โดยอาจทำได้ด้วยการใช้กระบวนการการตรวจสอบไขว้ (cross-validation: CV) ซึ่งเป็นการแบ่งส่วนข้อมูลออกเป็น ส่วน ๆ ที่เท่ากันอย่างสุ่ม เช่น ในการแบ่งข้อมูลออกเป็น k ส่วน จะใช้ข้อมูล $k - 1$ ส่วนในการสร้างตัวแบบ และตรวจสอบความสอดคล้องของตัวแบบบนข้อมูลส่วนที่เหลือหรือเรียกว่าข้อมูลทดสอบ เมื่อใช้ตัวแปรแฝง จำนวน a ตัว ในการทำนายค่าตัวแปรตาม โดย

วัดความสอดคล้องด้วยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (root mean squared error: RMSE) ทำจนข้อมูลทุกส่วนได้รับการทดสอบ จำนวนตัวแปรแฝงที่ทำให้รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าต่ำที่สุดจะได้รับการคัดเลือกเพื่อใช้เป็นพารามิเตอร์ งานวิจัยที่กำหนดจำนวนตัวแปรแฝงด้วยวิธีนี้ เช่น งานของ TeOfilos และคณะ [12] Gosselin และคณะ [13] เป็นต้น

3. การใช้เกณฑ์ในการประเมินว่าตัวแปรแฝงที่เพิ่มเข้ามาในตัวแบบสามารถลดค่าคลาดเคลื่อนของตัวแบบการถดถอยกำลังสองน้อยที่สุดบางส่วนได้อย่างมีนัยสำคัญหรือไม่ เช่น สถิติเอฟ (F statistic) เกณฑ์สารสนเทศของอะกะอิเกะ (Akaike information criterion) เกณฑ์ค่าอาร์ของ Wold (Wold's R criterion) สัมประสิทธิ์การกำหนดพหุคูณ (multiple coefficient of determination: R^2) และสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) [14-17]
- จากการกำหนดตัวแปรแฝงด้วยแนวทางข้างต้นนั้น ไม่มีวิธีการใดวิธีการหนึ่งที่ได้รับการยอมรับอย่างกว้างขวางสำหรับการกำหนดจำนวนตัวแปรแฝง [18] แต่อย่างไรก็ดี การกำหนดจำนวนตัวแปรแฝงโดยแนวทางที่ 2 และ 3 มีแนวโน้มที่จะได้จำนวนตัวแปรแฝงที่เหมาะสม เนื่องจาก



เป็นการกำหนดจำนวนตัวแปรแฝงที่ทำให้ตัวแบบมีความเหมาะสมที่สุดบนข้อมูลทดสอบ แต่การกำหนดจำนวนตัวแปรแฝงเป็นค่าคงที่ตามแนวทางที่ 1 จะทำให้ประหยัดเวลาในการวิเคราะห์ข้อมูล โดยเฉพาะข้อมูลที่มีขนาดใหญ่มาก

บทสรุป

การถดถอยกำลังสองน้อยที่สุดบางส่วนเป็นวิธีเชิงสถิติที่มีตัวแบบการถดถอยซึ่งสร้างตัวแปรแฝงจากตัวแปรทำนาย โดยตัวแปรแฝงที่สร้างขึ้นไม่มีความสัมพันธ์เชิงเส้นแบบพหุ ซึ่งมีลักษณะคล้ายกับการวิเคราะห์การถดถอยส่วนประกอบหลัก ที่เป็นการสร้างส่วนประกอบหลัก หรือตัวแปรแฝงที่ไม่มีความสัมพันธ์เชิงเส้นแบบพหุ แต่ตัวแปรแฝงที่สร้างขึ้นเป็นไปเพื่อรวมความแปรปรวนของตัวแปรทำนายเดิมไว้ให้มากที่สุด ซึ่งอาจจะไม่ใช่ตัวแปรแฝงที่เหมาะสมที่สุดสำหรับอธิบายตัวแปรตาม แตกต่างจากการถดถอยกำลังสองน้อยที่สุดบางส่วนที่สร้างตัวแปรแฝงเพื่ออธิบายตัวแปรทำนาย โดยเป็นตัวแปรแฝงที่ดีที่สุดในการอธิบายตัวแปรตาม การถดถอยกำลังสองน้อยที่สุดบางส่วนเป็นวิธีการที่พยายามสกัดตัวแปรแฝงเพื่อที่จะรวบรวมความแปรผันของตัวแปรทำนายให้ได้มากที่สุด อีกทั้งยังสามารถใช้ทำนายตัวแปรตามได้อย่างเหมาะสมด้วย นอกจากนี้การถดถอยกำลังสองน้อยที่สุดบางส่วนยังไม่มีข้อสมมติเกี่ยวกับการแจกแจงของประชากร และสามารถใช้วิเคราะห์ได้ในกรณีที่มีจำนวนตัวแปรอิสระหรือตัวแปรทำนายมาก โดยเฉพาะอย่างยิ่งมีจำนวนตัวแปรอิสระหรือตัวแปรทำนายมากกว่าขนาดตัวอย่าง ดังนั้นการถดถอยกำลังสองน้อยที่สุดบางส่วนจึงเป็นวิธีเชิงสถิติอีกทางเลือกหนึ่งที่ถูกนำมาใช้เมื่อไม่สามารถวิเคราะห์ข้อมูลด้วยการวิเคราะห์การถดถอยได้

เอกสารอ้างอิง

1. Nguyen DV, Rocke DM. Multi-class cancer classification via partial least squares with gene expression profiles. *Bioinformatics* 2002; 18(9): 1216-26.
2. ภาณุวัฒน์ ทรัพย์ประจุ, จันทร์เฉิดฉาย สังเกตกิจ. การแบ่งกลุ่มและหาปริมาณความชื้นและไขมันในรำข้าวแบบรวดเร็วโดยใช้แสงเนียร์อินฟราเรดแบบฟูเรียร์ทรานส์ฟอร์ม. ใน: การประชุมวิชาการสมาคมวิศวกรรมเกษตรแห่งประเทศไทยระดับชาติ ครั้งที่ 16 และระดับนานาชาติ ครั้งที่ 8 วันที่ 17-19 มีนาคม 2558. ศูนย์นิทรรศการและการประชุมไบเทค. กรุงเทพฯ; 2558. หน้า 459-64.
3. จุฑาพร หงส์มาทอง, สนธยา นุ่มท้วม. การพัฒนาวิธีตรวจวัดปริมาณแบคทีเรียทั้งหมดในน้ำนมโคดิบด้วยเนียร์อินฟราเรดสเปกโตรสโกปี. ใน: เอกสารประกอบการประชุมวิชาการระดับชาติ มหาวิทยาลัยทักษิณ ครั้งที่ 26 ประจำปี 2559 วันที่ 26-29 พฤษภาคม 2559. โรงแรมบุรีศรีภูบูติก. สงขลา; 2559. หน้า 96-103.
4. ศุมาพร เกษมสำราญ, อาทิตย์ จันทร์หิรัญ, วาธุณี ธนะแพสย์, วิทยา ปันสุวรรณ. การวิเคราะห์ปริมาณน้ำมันปาล์มในผลปาล์มแบบไม่ทำลายด้วยเนียร์อินฟราเรดสเปกโตรสโกปี. *วารสารวิทยาศาสตร์เกษตร* 2553; 41(3/1)(พิเศษ):61-4.
5. ปราณีย์ คำแก้ว. การวิเคราะห์โครงสร้างยาโดยวิธีการถดถอยกำลังสองน้อยที่สุดบางส่วน. *วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติประยุกต์, บัณฑิตวิทยาลัยมหาวิทยาลัยเชียงใหม่*. เชียงใหม่; 2552.



6. Boulesteix AL, Strimmer K. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief Bioinform* 2006;8(1):32-44.
7. Mehmood T, Liland KH, Snipen L, Sæbø S. A review of variable selection methods in partial least squares regression. *Chemometr Intell Lab* 2012;118:62-9.
8. Manne R. Analysis of two partial-least-squares algorithms for multivariate calibration. *Chemometr Intell Lab* 1987;2:187-97.
9. Russolillo G. Partial least squares methods for non-metric data. Doctoral dissertation, Dipartimento di Matematica e Statistica, Università degli Studi di Napoli Federico II. Napoli; 2009.
10. Li GZ, Zeng XQ. Feature selection for partial least square based dimension reduction. In: Abraham A, Hassanien AE, Siarry P, Engelbrecht A, editors. *Foundations of computational intelligence*. Berlin: Springer-Verlag Berlin Heidelberg; 2009. p. 3-37.
11. Nguyen DV, Rocke DM. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics* 2002;18(1):39-50.
12. TeÓfilo RF, Martins JPA, Ferreira MMC. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. *J Chemometr* 2008;23:32-48.
13. Gosselin R, Rodrigue D, Duchesne C. A bootstrap-VIP approach for selecting wavelength intervals in spectral imaging application. *Chemometr Intell Lab* 2010;100:12-21.
14. Li B, Morris J, Martin EB. Model selection for partial least squares regression. *Chemometr Intell Lab* 2002;64:79-89.
15. Zeng XQ, Li GZ, Wu GF. On the number of partial least squares components in dimension reduction for tumor classification. [Internet]. 2012 [cited 2012 September 28]. Available from: <http://levis.tongji.edu.cn/gzli/publication/biodm07.pdf>
16. Lazraq A, Cl eroux R, Gauchi JP. Selecting both latent and explanatory variables in the PLS1 regression model. *Chemometr Intell Lab* 2003;66:117-26.
17. Gutkin M, Shamir R, Dror G. SlimPLS: a method for feature selection in gene expression-based disease classification. *PLoS ONE* 2009;4(7):e6416.
18. Boulesteix AL. PLS dimension reduction for classification with microarray data. *Stat Appl Genet Mo B* 2004;3(1):Article 33.