วารสารวิชาการ **มหาวิทยาลัยหอการค้าไทย**
ปีที่ 29 ฉบับที่ 3 เดือนกรกฎาคม - กันยายน 2552

การโปรแกรมเชิงเส้นสำหรับการวิเคราะห์การจำแนกกลุ่มโดยคำนวณด้วยโปรแกรม Microsoft Excel

Linear Programming Model for Discriminant Analysis Calculated with Microsoft Excel Program

อาฟีฟี ลาเต๊ะ
อาจารย์ประจำภาควิชาประเมินผลและวิจัยทางการศึกษา
คณะศึกษาศาสตร์ มหาวิทยาลัยสงขลานครินทร์
E-mail: afifi-l@bunga.pn.psu.ac.th

ประสพชัย พสุนนท์
ผู้ช่วยศาสตราจารย์ประจำสาขาวิชาการจัดการธุรกิจทั่วไป
คณะวิทยาการจัดการ มหาวิทยาลัยศิลปากร
E-mail: pasunon@gmail.com

บทคัดย่อ

บทความนี้ ผู้เขียนต้องการนำเสนอการโปรแกรมเชิงเส้นสำหรับการวิเคราะห์การจำแนกตามตัวแบบที่เสนอโดย Freed and Glover (1986) พร้อมการคำนวณด้วยโปรแกรม Microsoft Excel ข้อดีของการวิเคราะห์การจำแนกด้วยการโปรแกรมเชิงเส้น คือ สามารถหลีกเลี่ยงข้อสมมติจากวิธีการทางสถิติ อาทิ ข้อมูลที่นำมาใช้ในการวิเคราะห์ต้องมีการแจกแจงแบบปกติของตัวแปรพหุ การเท่ากันของเมทริกซ์ ความแปรปรวนร่วมของตัวแปรอิสระระหว่างกลุ่ม การแจกแจงของความคลาดเคลื่อนของตัวแปรอิสระ เป็นต้น เนื้อหาในบทความฉบับนี้ประกอบไปด้วย 5 ส่วน คือ 1) บทนำเป็นการเกริ่นถึงลักษณะทั่วไปของการวิเคราะห์การจำแนก 2) การอธิบายหลักการวิเคราะห์การจำแนกโดยใช้การโปรแกรมเชิงเส้น 3) การคำนวณการโปรแกรมเชิงเส้นด้วยโปรแกรม Microsoft Excel

โดยเริ่มตั้งแต่ขั้นตอนแรกจนถึงได้ผลลัพธ์เป็นค่าเหมาะสุดในตัวแบบที่ใช้ในการคำนวณ 4) การแปลผลที่ได้จากการคำนวณ และ 5) การสรุป ผู้เขียนได้ให้ข้อเสนอแนะในการนำวิธีการไปใช้

คำสำคัญ: การวิเคราะห์การจำแนก การโปรแกรมเชิงเส้น Microsoft Excel

Abstract

In this article, the writers present a linear programming model by Freed and Glover (Freed and Glover, 1986) for discriminant analysis calculated with Microsoft Excel program. The strong point of linear programming for discriminant analysis is that it can avoid assumptions of the parametric statistics approach, such as data for analyzing from multivariate normal distributions, or equal covariance matrices of independent variables between groups, or normal distribution of residuals. There are 5 parts in this article: 1) prelude, which contains the general characteristics of discriminant analysis, 2) description of the concept of linear programming for discriminant analysis, 3) calculation of the linear programming using Microsoft Excel program, which begins with the first step and proceeds until arriving at the optimal results in the model, 4) interpretation of calculations, and 5) conclusion in which the writers give some suggestions for using this method.

Keywords: Discriminant Analysis, Linear Programming, Microsoft Excel

บทนำ

การจัดการข้อมูลในสถานการณ์ปัจจุบันมีความสำคัญมาก เนื่องจากข้อมูลถือเป็นวัตถุดิบในการสร้างสารสนเทศต่างๆ เพื่อใช้ประกอบการกำหนดนโยบาย วางแผน หรือตัดสินใจ ปัญหาบางอย่างต้องการคำตอบง่ายๆ เพียงแค่ว่า ใช่ หรือ ไม่ใช่ ไม่ว่าจะเป็นปัญหาในด้านธุรกิจ การเมือง การแพทย์ หรือการศึกษา เช่น การพิจารณาข้อมูลผู้กู้รายใหม่ของธนาคารพาณิชย์ การพิจารณานโยบายของผู้สมัครรับเลือกตั้งสมาชิกสภาผู้แทนราษฎร การพิจารณาแนวทางการรักษาของผู้ป่วยโรคมะเร็ง การพิจารณาคะแนนสอบระดับบัณฑิตศึกษาของมหาวิทยาลัย เป็นต้น ซึ่งผู้อยู่ในแวดวงดังกล่าวจะต้องมีการประเมินข้อมูลในแต่ละกลุ่มเพื่อการตัดสินใจในทางใดทางหนึ่ง โดยการประเมินจะใช้รูปแบบใดนั้นขึ้นอยู่กับความรู้ความสามารถของคนในกลุ่มนั้นๆ หรือขึ้นอยู่กับการคาดการณ์ด้วยเหตุผลส่วนตัวก็เป็นได้ อย่างไรก็ตาม หากต้องการประเมินด้วยวิธีการที่เป็นแบบแผนทางวิชาการก็สามารถกระทำได้ ซึ่งจะเรียกการประเมินดังกล่าวนั้นว่า การวิเคราะห์การจำแนก (Discriminant Analysis)

การวิเคราะห์การจำแนก เป็นวิธีการทางสถิติที่ใช้สำหรับตัวแปรพหุซึ่งประกอบด้วยขั้นตอนการจำแนกกลุ่ม และการจัดเข้ากลุ่ม (Discrimination and Classification) โดยในขั้นตอนการจำแนก

กลุ่มจะใช้ค่าสังเกต หรือตัวอย่างที่ทราบแล้วว่ามาจากกลุ่มใด คำนวณค่าเชิงตัวเลข หรือสมการทางพีชคณิตให้เห็นถึงความแตกต่างระหว่างกลุ่มอย่างเด่นชัดที่สุดเท่าที่จะกระทำได้ โดยปกติแล้วจะเป็นการสร้างฟังก์ชันจำแนก (Discriminant Function) หรือผลบวกเชิงเส้น (Linear Combination) ของตัวแปรอิสระหรือตัวแปรทำนาย (Independent Variable or Predictor) เพื่ออธิบายลักษณะที่ต่างกันของตัวแปรตามหรือของกลุ่ม (Dependent Variable or Group) ส่วนขั้นตอนการจัดเข้ากลุ่มเป็นการจัดค่าสังเกต หรือตัวอย่างที่เพิ่งเข้ามาใหม่เข้ากลุ่มใดกลุ่มหนึ่งโดยอาศัยผลลัพธ์จากฟังก์ชันจำแนกและเกณฑ์การจัดเข้ากลุ่ม (Classification Rule) การจัดเข้ากลุ่มจะทำได้ถูกต้องเพียงใดขึ้นกับฟังก์ชันจำแนกและกฎเกณฑ์การจัดเข้ากลุ่มว่าเหมาะสมหรือไม่ วิธีการวิเคราะห์การจำแนกที่ดีควรให้ฟังก์ชันจำแนกที่แสดงถึงลักษณะที่ต่างกันระหว่างตัวแปรตาม หรือระหว่างกลุ่มอย่างเด่นชัด และกฎการจัดเข้ากลุ่มที่ใช้ควรมีความเหมาะสมเพื่อให้ความผิดพลาดในการจำแนกเกิดขึ้นน้อย

วิธีการวิเคราะห์การจำแนกที่นิยมใช้มีด้วยกัน 2 วิธี คือ การวิเคราะห์การจำแนกโดยใช้ฟังก์ชันเชิงเส้นของฟิชเชอร์ (Fisher's Linear Discriminant Function) และการวิเคราะห์การจำแนกโดยใช้การถดถอยแบบโลจิสติก (Logistic Regression Discriminant Analysis) ซึ่งทั้งสองวิธีนี้เป็นวิธีการทางสถิติที่แพร่หลายมาก จะเห็นได้จากการมีโปรแกรมสำเร็จรูปหลายโปรแกรมที่สามารถวิเคราะห์ด้วยวิธีนี้ ไม่ว่าจะเป็นโปรแกรม SPSS MINITAB SAS หรือ FORTRAN อย่างไรก็ตาม ถึงแม้ว่ามีโปรแกรมที่สนับสนุนการใช้วิธีการดังกล่าว แต่หากผู้ใช้ไม่คำนึงถึงข้อสมมติดั้งเดิมของวิธีการวิเคราะห์ ผลที่ตามมา คือ ไม่สามารถให้ผลของการวิเคราะห์ที่ถูกต้องตามหลักวิชา รวมถึงความผิดพลาดในการใช้ผลดังกล่าวได้ ข้อสมมติเบื้องต้นของวิธีการแรกมีด้วยกัน 2 ข้อ กล่าวคือ การแจกแจงของตัวแปรพหุ หรือตัวแปรอิสระต้องมีการแจกแจงแบบปกติของตัวแปรพหุ (Multivariate Normal Distributions) และเมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระระหว่างกลุ่มต้องเท่ากัน (Equal Covariance Matrices between Groups) ส่วนข้อสมมติเบื้องต้นของวิธีการที่ 2 คือ ความคลาดเคลื่อนของตัวแปรอิสระต้องมีการแจกแจงแบบปกติ (Normal Distribution of Residuals)

ในทางปฏิบัติจะเห็นได้ว่าวิธีการที่นิยมใช้ทั้ง 2 วิธีนั้น ต่างมีข้อสมมติที่เป็นไปได้ยากที่ข้อมูลในการวิเคราะห์จะมีลักษณะตรงกับข้อสมมติข้างต้น อย่างไรก็ตาม ทางเลือกของวิธีการวิเคราะห์การจำแนกอีกวิธีหนึ่ง คือ การวิเคราะห์การจำแนกโดยใช้การโปรแกรมเชิงเส้น (Linear Programming) ซึ่งเป็นวิธีการทางสถิติที่ไม่ใช้พารามิเตอร์ (Non-parametric Statistics) และด้วยวิธีการนี้ไม่จำเป็นที่ข้อมูลของการวิเคราะห์ต้องสอดคล้องกับข้อสมมติเบื้องต้น เนื่องจากว่าเป็นการวิเคราะห์ที่อาศัยกระบวนการทำซ้ำด้วยขั้นตอนวิธีทางคณิตศาสตร์ ซึ่งเป็นการสร้างตัวแบบแทนระบบปัญหามาใช้ในการวิเคราะห์การจำแนก ตัวแบบจะประกอบด้วยฟังก์ชันจุดประสงค์ หรือฟังก์ชันเป้าหมาย (Objective Function) และเงื่อนไขบังคับ หรือข้อจำกัด (Constraints) โดยมีจุดมุ่งหมายเพื่อแก้ปัญหาและตัดสินใจให้เกิดผลตามแนวทางการดำเนินงานที่ดีที่สุดภายใต้เงื่อนไขที่กำหนด ซึ่งการวิเคราะห์ด้วยวิธีการนี้ได้เสนอครั้งแรกเมื่อปี ค.ศ. 1981 โดย Freed and Glover

และ Hand จากนั้นวิธีการนี้ได้พัฒนาอย่างต่อเนื่องจนถึงปัจจุบัน โดยมีตัวแบบที่เสนอแล้วทั้งสิ้นไม่ต่ำกว่า 30 ตัวแบบในรูปแบบต่างๆ กันตามฟังก์ชันจุดประสงค์ หรือเงื่อนไขบังคับ เช่น การโปรแกรมเป้าหมายเชิงเส้น การโปรแกรมจำนวนเต็มแบบผสมและการโปรแกรมไม่ใช่เชิงเส้น ในบทความนี้ผู้เขียนจะเสนอตัวแบบที่ไม่ซับซ้อน และเป็นที่นิยมใช้ในงานวิจัยต่างๆ เพียง 1 ตัวแบบ เพื่อเป็นพื้นฐานของการวิเคราะห์การจำแนกโดยใช้การโปรแกรมเชิงเส้น สำหรับตัวแบบอื่นๆ ผู้เขียนจะได้นำเสนอในโอกาสต่อไป

การวิเคราะห์การจำแนกโดยการโปรแกรมเชิงเส้น

ตัวแบบที่จะนำเสนอเป็นตัวแบบของ Freed and Glover ซึ่งเสนอในปี ค.ศ. 1986 ซึ่งเป็นตัวแบบที่พัฒนาจากตัวแบบที่เสนอในปี ค.ศ. 1981 เพื่อแก้ปัญหากรณีที่ค่าถ่วงน้ำหนักมีโอกาสเป็น 0 โดยจะเรียกวิธีนี้ย่อๆ ว่าตัวแบบ MSD (Minimize the Sum of Deviation) ซึ่งมีรูปแบบการโปรแกรมเชิงเส้นดังนี้

ให้ β_i เป็นค่าถ่วงน้ำหนักที่ให้กับตัวแปรอิสระ i เมื่อ $i = 1,...,p$

x_{ij} เป็นค่าสังเกตลำดับที่ j ของตัวแปรอิสระที่ i

p เป็นจำนวนตัวแปรอิสระ

c เป็นจุดตัด (cut off)

N เป็นค่าที่กำหนดขึ้นเพื่อให้สเกลของผลลัพธ์อยู่ในช่วงที่ต้องการ โดยกำหนดให้ N = 10

α_j เป็นตัวแปรเบี่ยงเบนของค่าสังเกตลำดับที่ j

Minimize $\sum_{j=1}^{m} \alpha_j$

Subject to

$$\sum_{i=1}^{p} \beta_i x_{ij} - \alpha_j - c \le 0$$

สำหรับทุกค่า j ในกลุ่มที่ 1 ..... (1)

$$\sum_{i=1}^{p} \beta_i x_{ij} + \alpha_j - c \ge 0$$

สำหรับทุกค่า j ในกลุ่มที่ 2 ..... (2)

$$c + \sum_{i=1}^{p} \beta_i = N \qquad \text{.....} \quad (3)$$

β_i และ c ไม่จำกัดเครื่องหมาย

$a_j \ge 0$

ตัวแบบ MSD เป็นตัวแบบที่มีเป้าหมายเพื่อหาค่าต่ำสุดของผลรวมของส่วนเบี่ยงเบนของค่าสังเกตกับจุดตัด ภายใต้เงื่อนไขเกณฑ์การจัดเข้ากลุ่มของกลุ่มที่ 1 และกลุ่มที่ 2 ตามเงื่อนไข (1) และ (2) โดยมีข้อสมมติว่ากลุ่ม 1 อยู่ใต้จุดตัด กลุ่ม 2 อยู่เหนือจุดตัด โดยตัวแปรเบี่ยงเบน α_i จะเป็นศูนย์เมื่อค่าสังเกตลำดับที่ j จัดถูกกลุ่ม สำหรับเงื่อนไข (3) เป็นเงื่อนไขที่เพิ่มในตัวแบบเพื่อไม่ให้เกิดกรณีที่ $\beta_i = 0$ ทุกค่าของ i ซึ่งเรียกว่า Trivial Solution ส่วนการกำหนดค่า N จะกำหนดให้เป็น 10 ซึ่งจากการศึกษาของ Freed and Glover (1986) พบว่าค่า N ที่กำหนดนั้นเพียงแต่จะทำให้ค่า β_i หรือ c ที่ได้หลังจากแก้ปัญหาอยู่ในช่วงของค่าที่มีขนาดไม่เล็กหรือใหญ่จนเกินไป อย่างไรก็ตามอาจจะกำหนดให้ N เป็น 1 หรือ 100 ก็สามารถทำได้เช่นเดียวกัน หลังจากแก้ปัญหาแล้ว นำค่า $\sum_{i=1}^{p} \beta_i x_{ij}$ และ c มาใช้จัดกลุ่มค่าสังเกต นั่นคือ ถ้า $\sum_{i=1}^{p} \beta_i x_{ij} \le c$ แสดงว่าค่าสังเกตอยู่กลุ่ม 1

และถ้า $\sum_{i=1}^{p} \beta_i x_{ij} > c$ แสดงว่าค่าสังเกตอยู่กลุ่ม 2

การคำนวณโปรแกรมเชิงเส้นด้วยโปรแกรม Microsoft Excel

โดยทั่วไปแล้ว การคำนวณการโปรแกรมเชิงเส้นมีโปรแกรมคอมพิวเตอร์โดยเฉพาะ อาทิ FORTRAN LINDO LINGO เป็นต้น อย่างไรก็ตาม การทำความเข้าใจเกี่ยวกับวิธีการคำนวณการโปรแกรมเชิงเส้นค่อนข้างยุ่งยาก และต้องใช้เวลาค่อนข้างมาก ดังนั้น บทความนี้ต้องการนำเสนอโปรแกรมสำเร็จรูป Microsoft Excel เพื่อใช้คำนวณตัวแบบ MSD เนื่องจากเป็นโปรแกรมสำเร็จรูปที่เป็นที่รู้จักคุ้นเคยกันทั่วไป ราคาไม่แพง และค่อนข้างง่ายต่อการทำความเข้าใจ

สำหรับขั้นตอนการคำนวณตัวแบบ MSD ด้วยโปรแกรม Microsoft Excel ผู้เขียนได้นำตัวอย่างของ Freed and Glover (1981) มาใช้เพื่อแสดงการคำนวณ แสดงดังตารางที่ 1 ซึ่งค่าในตารางเป็นข้อมูลของลูกจ้างจำนวน 10 คน ที่ไม่ประสบผลสำเร็จและประสบผลสำเร็จในการทำงาน ซึ่งวัดจากคุณลักษณะของตัวแปร 2 ตัว นั่นคือ ประสบการณ์ในการทำงาน (x_1) และระยะเวลาที่จบการศึกษา (x_2)

ตารางที่ 1 คุณลักษณะของลูกจ้างทั้ง 10 คนจากกลุ่มไม่ประสบผลสำเร็จ และกลุ่มประสบผลสำเร็จ

ลูกจ้างคนที่	กลุ่ม 1: ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2: ประสบผลสำเร็จ	
	x_1	x_2		x_1	x_2
1	1	1	**6**	4	1
2	3	1	**7**	5	2
3	2	2	**8**	7	2
4	6	4	**9**	9	1
5	3	3	**10**	8	4

ข้อมูลของ Freed and Glover (1981) สามารถเขียนตัวแบบ MSD เพื่อวิเคราะห์การจำแนกลูกจ้างที่ไม่ประสบผลสำเร็จและประสบผลสำเร็จ ได้ดังนี้

Minimize $\alpha_1 + \alpha_2 + \alpha_3 + ... + \alpha_{10}$

Subject to

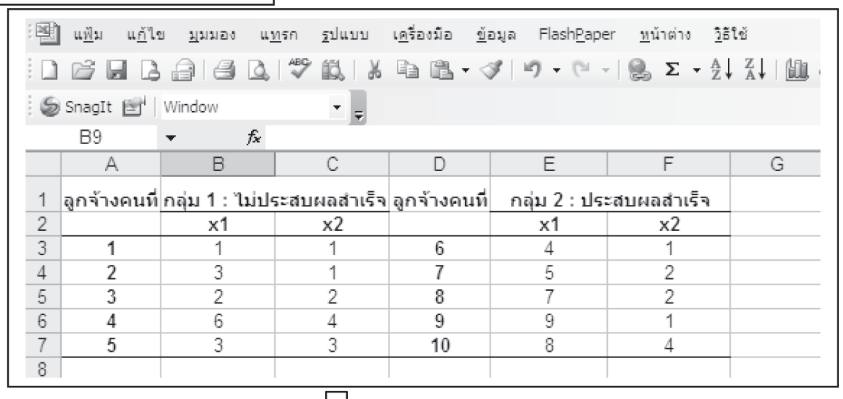
$$\beta_1 + \beta_2 - \alpha_1 - c \leq 0$$
$$3\beta_1 + \beta_2 - \alpha_2 - c \leq 0$$
$$2\beta_1 + 2\beta_2 - \alpha_3 - c \leq 0$$
$$6\beta_1 + 4\beta_2 - \alpha_4 - c \leq 0$$
$$3\beta_1 + 3\beta_2 - \alpha_5 - c \leq 0$$
$$4\beta_1 + \beta_2 + \alpha_6 - c \geq 0$$
$$5\beta_1 + 2\beta_2 + \alpha_7 - c \geq 0$$
$$7\beta_1 + 2\beta_2 + \alpha_8 - c \geq 0$$
$$9\beta_1 + \beta_2 + \alpha_9 - c \geq 0$$
$$8\beta_1 + 4\beta_2 + \alpha_{10} - c \geq 0$$
$$c + \beta_1 + \beta_2 = 10$$

β_1, β_2 และ c ไม่จำกัดเครื่องหมาย

α_1, α_2, α_3, ..., $\alpha_{10} \geq 0$

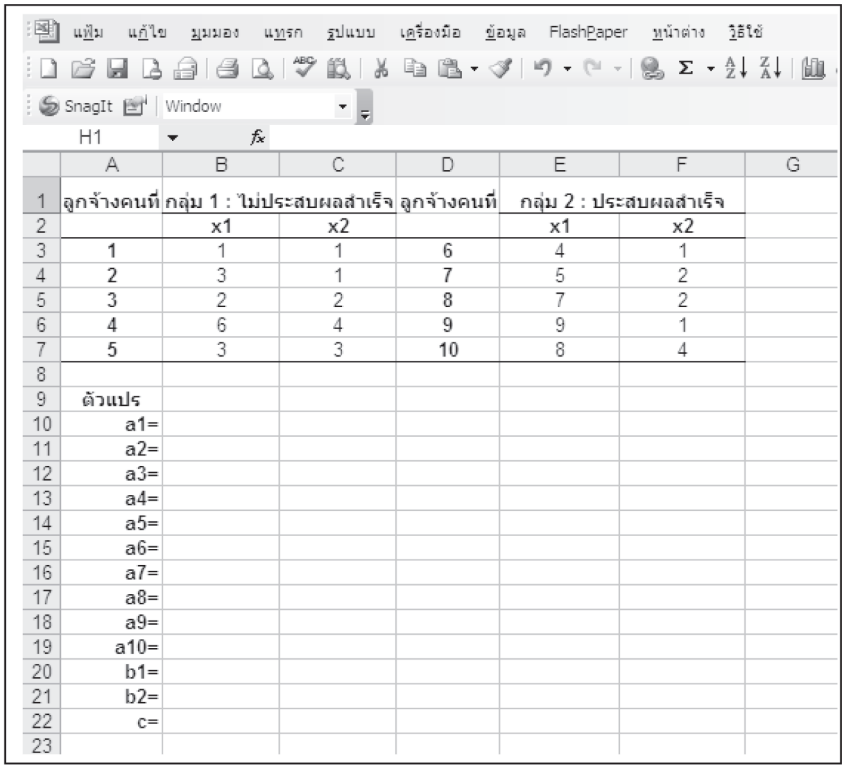
การคำนวณการโปรแกรมเชิงเส้นของตัวแบบ MSD ด้วยโปรแกรม Microsoft Excel จะกำหนดให้ α_1, α_2, α_3, ..., α_{10} แทนด้วย a_1, a_2, a_3, ..., a_{10} ตามลำดับ และให้ β_1, β_2 และ c แทนด้วย b_1, b_2 และ c ตามลำดับ สำหรับขั้นตอนการคำนวณมีทั้งหมด 7 ขั้นตอน แสดงดังรูปที่ 1-5 ดังนี้

ขั้น 1 เตรียมข้อมูลสำหรับการคำนวณ



	A	B	C	D	E	F	G
1	ลูกจ้างคนที่	กลุ่ม 1 : ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2 : ประสบผลสำเร็จ		
2		x1	x2		x1	x2	
3	1	1	1	6	4	1	
4	2	3	1	7	5	2	
5	3	2	2	8	7	2	
6	4	6	4	9	9	1	
7	5	3	3	10	8	4	
8							

ขั้น 2 กำหนดตัวแปรในการคำนวณโดยให้ cell ที่ B10, B11, B12, ..., B19 แทนค่าของตัวแปร a_1, a_2, a_3, ...,a_{10} ตามลำดับ และให้ cell ที่ B20, B21 และ B22 แทนค่าของตัวแปร b_1, b_2 และ c ตามลำดับ



	A	B	C	D	E	F	G
1	ลูกจ้างคนที่	กลุ่ม 1 : ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2 : ประสบผลสำเร็จ		
2		x1	x2		x1	x2	
3	1	1	1	6	4	1	
4	2	3	1	7	5	2	
5	3	2	2	8	7	2	
6	4	6	4	9	9	1	
7	5	3	3	10	8	4	
8							
9	ตัวแปร						
10	a1=						
11	a2=						
12	a3=						
13	a4=						
14	a5=						
15	a6=						
16	a7=						
17	a8=						
18	a9=						
19	a10=						
20	b1=						
21	b2=						
22	c=						
23							

รูปที่ 1 การเตรียมข้อมูล และการกำหนดตัวแปรในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ขั้น 3 กำหนด Subject to ของตัวแบบโดยให้ cell ที่ C10, C11, C12, ..., C20 ซึ่งมีค่าเป็น

$\beta_1+\beta_2-\alpha_1-c$ แทนด้วย =B3*B20+C3*B21-B10-B22

$3\beta_1+\beta_2-\alpha_2-c$ แทนด้วย =B4*B20+C4*B21-B11-B22

$2\beta_1+2\beta_2-\alpha_3-c$ แทนด้วย =B5*B20+C5*B21-B12-B22

$6\beta_1+4\beta_2-\alpha_4-c$ แทนด้วย =B6*B20+C6*B21-B13-B22

$3\beta_1+3\beta_2-\alpha_5-c$ แทนด้วย =B7*B20+C7*B21-B14-B22

$4\beta_1+\beta_2+\alpha_6-c$ แทนด้วย =E3*B20+F3*B21+B15-B22

$5\beta_1+2\beta_2+\alpha_7-c$ แทนด้วย =E4*B20+F4*B21+B16-B22

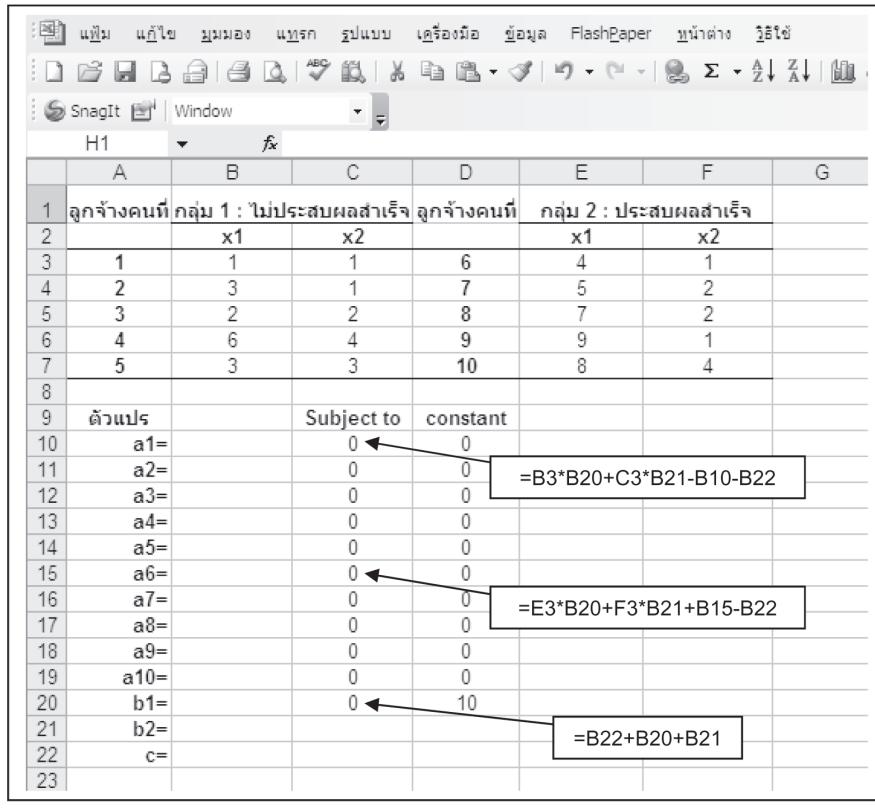
$7\beta_1+2\beta_2+\alpha_8-c$ แทนด้วย =E5*B20+F5*B21+B17-B22

$9\beta_1+\beta_2+\alpha_9-c$ แทนด้วย =E6*B20+F6*B21+B18-B22

$8\beta_1+4\beta_2+\alpha_{10}-c$ แทนด้วย =E7*B20+F7*B21+B19-B22

$c+\beta_1+\beta_2$ แทนด้วย =B22+B20+B21 ตามลำดับ

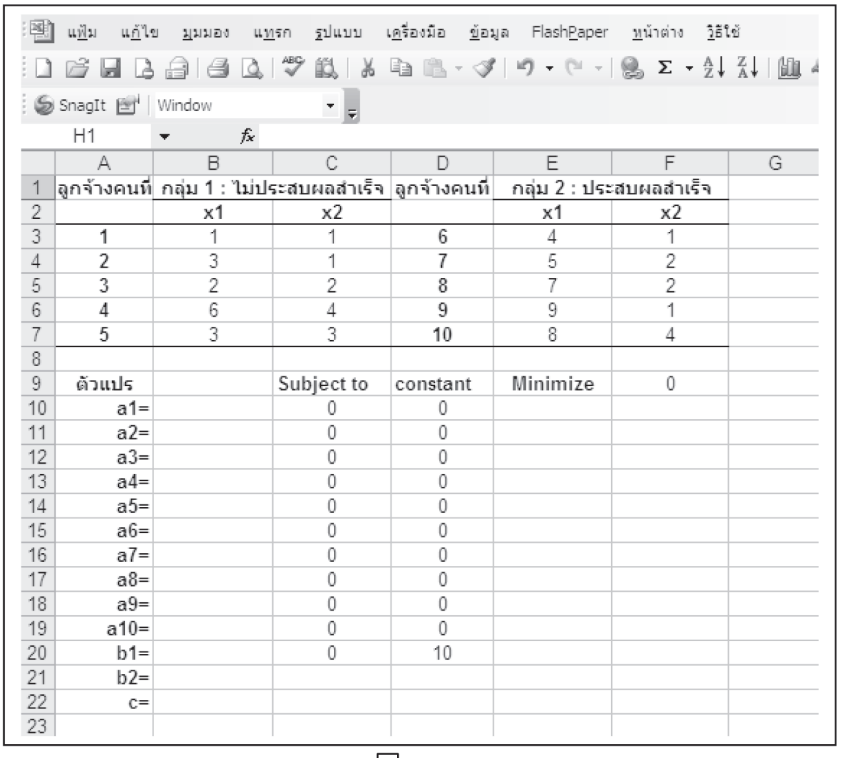
กำหนด Constant (ค่าคงที่) ของตัวแบบโดยให้ cell ที่ D10, D11, D12, ..., D19 มีค่าเป็น 0 ส่วน D20 มีค่าเป็น 10



	A	B	C	D	E	F	G
1	ลูกจ้างคนที่	กลุ่ม 1 : ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2 : ประสบผลสำเร็จ		
2		x1	x2		x1	x2	
3	1	1	1	6	4	1	
4	2	3	1	7	5	2	
5	3	2	2	8	7	2	
6	4	6	4	9	9	1	
7	5	3	3	10	8	4	
8							
9	ตัวแปร		Subject to	constant			
10	a1=		0	0			
11	a2=		0	0			
12	a3=		0	0			
13	a4=		0	0			
14	a5=		0	0			
15	a6=		0	0			
16	a7=		0	0			
17	a8=		0	0			
18	a9=		0	0			
19	a10=		0	0			
20	b1=		0	10			
21	b2=						
22	c=						
23							

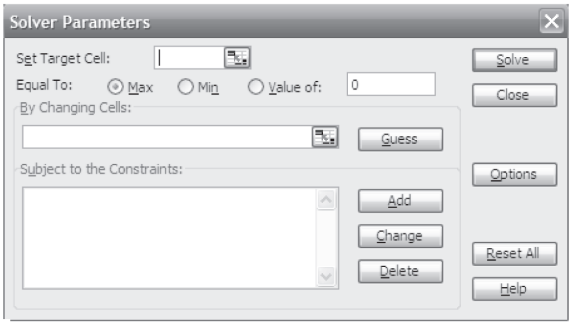
รูปที่ 2 การกำหนด Subject to และ constant ในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ขั้น 4 กำหนดฟังก์ชันวัตถุประสงค์ของตัวแบบโดยให้ cell ที่ F9 ซึ่งมีค่าเป็น $\alpha_1 + \alpha_2 + \alpha_3 + ... + \alpha_{10}$ แทนด้วย =SUM(B10:B19)



	A	B	C	D	E	F	G
1	ลูกจ้างคนที่	กลุ่ม 1 : ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2 : ประสบผลสำเร็จ		
2		x1	x2		x1	x2	
3	1	1	1	6	4	1	
4	2	3	1	7	5	2	
5	3	2	2	8	7	2	
6	4	6	4	9	9	1	
7	5	3	3	10	8	4	
8							
9	ตัวแปร		Subject to	constant	Minimize	0	
10	a1=		0	0			
11	a2=		0	0			
12	a3=		0	0			
13	a4=		0	0			
14	a5=		0	0			
15	a6=		0	0			
16	a7=		0	0			
17	a8=		0	0			
18	a9=		0	0			
19	a10=		0	0			
20	b1=		0	10			
21	b2=						
22	c=						
23							

ขั้น 5 คำนวณโดยใช้ฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบโดยคลิกที่ **เครื่องมือ (Tools)** จากแถบเมนู แล้วเลือก **Solver...** ซึ่งจะปรากฏ Dialog box ดังรูปนี้



รูปที่ 3 การกำหนดฟังก์ชันวัตถุประสงค์ของตัวแบบในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ขั้น 6 การป้อนค่าฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบ ดังนี้

★ **Set Target Cell:** คือ การกำหนดฟังก์ชันวัตถุประสงค์ แทนด้วย cell ที่ F9 เมื่อคลิกที่ F9 จะปรากฏ $F$9 อัตโนมัติ

★ **Equal To:** คือ การเลือกฟังก์ชันวัตถุประสงค์ว่าเป็นการหาค่าสูงสุด หรือ ค่าต่ำสุด กรณีนี้เป็นการหาค่าต่ำสุดจึงเลือก Min

★ **By Changing Cells:** คือ การกำหนดการเปลี่ยนแปลงของตัวแปรในตัวแบบ ในกรณีนี้ตัวแปร คือ B10, B11, B12, ..., B22 โดยสามารถกด Ctrl ที่แป้นพิมพ์ค้างไว้แล้วลากเมาส์เลือก cell ที่ B10, B11, B12, ..., B22 ซึ่งจะปรากฏ $B$10:$B$22 อัตโนมัติ

★ **Subject to the Constraints:** คือ การกำหนดเงื่อนไขข้อจำกัดในตัวแบบซึ่งจะกำหนด

เงื่อนไข $\sum_{i=1}^{p} \beta_i x_{ij} - \alpha_j - c \leq 0$ สำหรับทุกค่า j ในกลุ่มที่ 1

แทนด้วย $C$10:$C$14 <= $D$10:$D$14

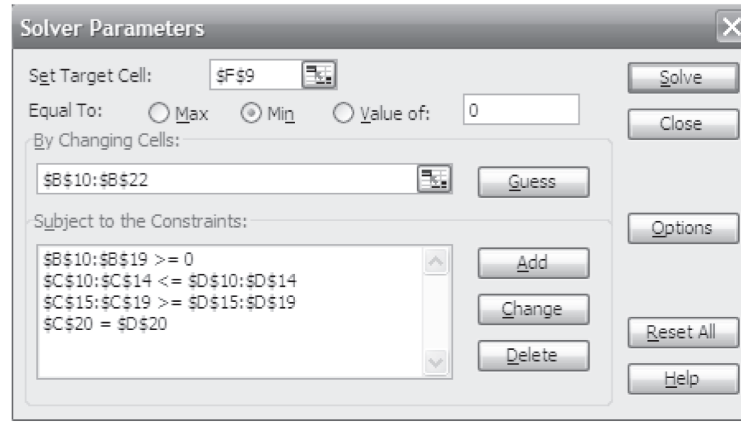
เงื่อนไข $\sum_{i=1}^{p} \beta_i x_{ij} + \alpha_j - c \geq 0$ สำหรับทุกค่า j ในกลุ่มที่ 2

แทนด้วย $C$15:$C$19 >= $D$15:$D$19

เงื่อนไข $c + \beta_1 + \beta_2 = 10$ แทนด้วย $C$20 = $D$20

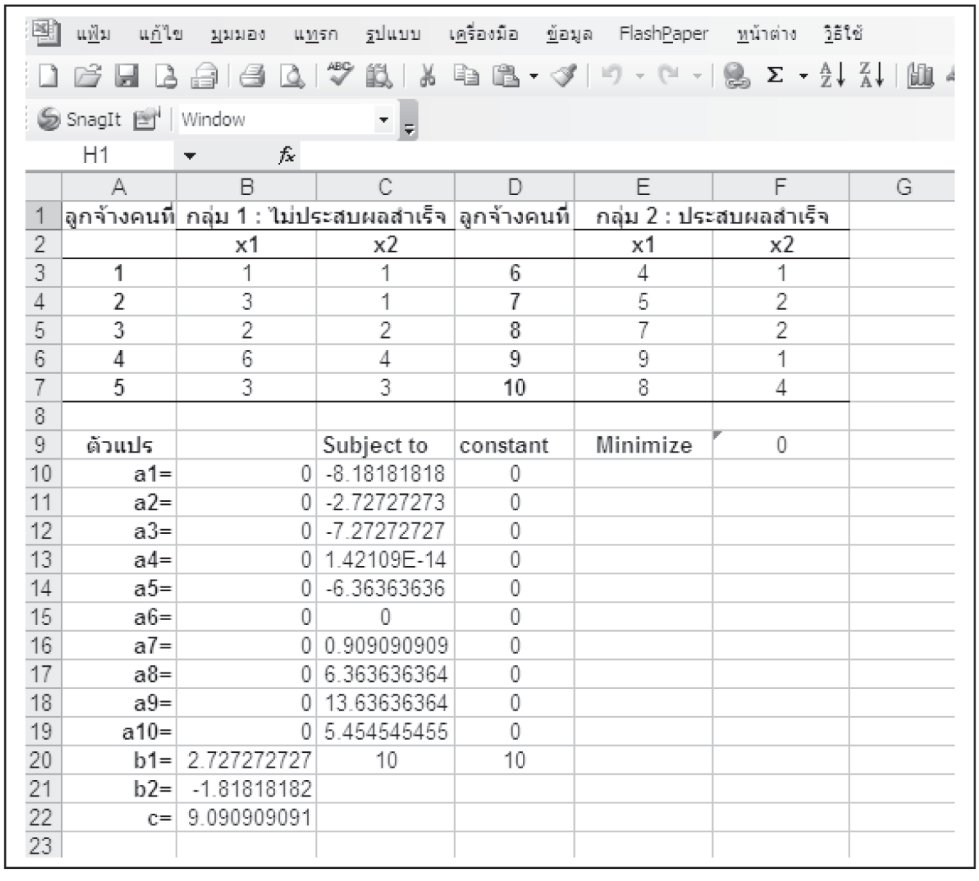
เงื่อนไข $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_{10} \geq 0$

แทนด้วย $B$10:$B$19 >= 0



รูปที่ 4 การป้อนค่าฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ขั้น 7 คำนวณตามฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดที่กำหนดโดยการคลิก **Solve** ใน Dialog box จากขั้น 6 ผลการคำนวณเป็นดังนี้



	A	B	C	D	E	F	G
1	ลูกจ้างคนที่	กลุ่ม 1 : ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2 : ประสบผลสำเร็จ		
2		x1	x2		x1	x2	
3	1	1	1	6	4	1	
4	2	3	1	7	5	2	
5	3	2	2	8	7	2	
6	4	6	4	9	9	1	
7	5	3	3	10	8	4	
8							
9	ตัวแปร		Subject to	constant	Minimize	0	
10	a1=	0	-8.18181818	0			
11	a2=	0	-2.72727273	0			
12	a3=	0	-7.27272727	0			
13	a4=	0	1.42109E-14	0			
14	a5=	0	-6.36363636	0			
15	a6=	0	0	0			
16	a7=	0	0.909090909	0			
17	a8=	0	6.363636364	0			
18	a9=	0	13.63636364	0			
19	a10=	0	5.454545455	0			
20	b1=	2.727272727	10	10			
21	b2=	-1.81818182					
22	c=	9.090909091					
23							

รูปที่ 5 ผลลัพธ์จากการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

การแปลผล

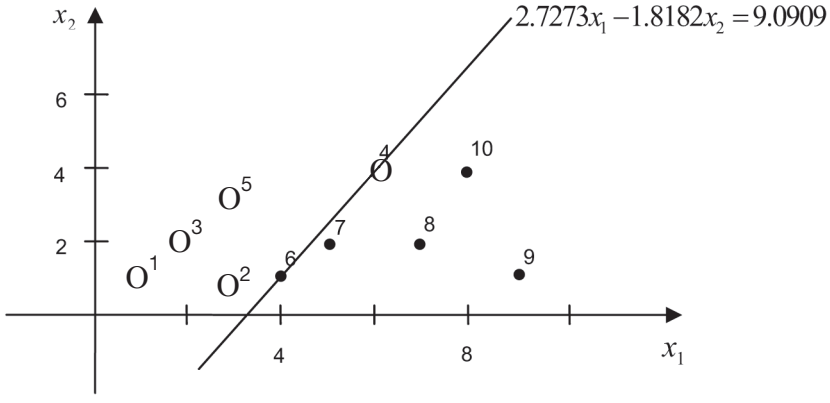
ผลลัพธ์ที่ได้จากการคำนวณโปรแกรมเชิงเส้นตามตัวแบบ MSD ด้วยโปรแกรม Microsoft Excel ได้ค่าเบี่ยงเบนของค่าสังเกตเท่ากับ 0 ทุกค่า ($\alpha_1, \alpha_2, \alpha_3, ..., \alpha_{10} = 0$) มีผลให้ค่าต่ำสุดของผลรวมของส่วนเบี่ยงเบนของค่าสังเกตกับจุดตัดเท่ากับ 0 ($\alpha_1 + \alpha_2 + \alpha_3 + ... + \alpha_{10} = 0$) โดยมีค่าถ่วงน้ำหนัก β_1 เท่ากับ 2.7273 β_2 เท่ากับ -1.8182 และค่าจุดตัด c เท่ากับ 9.0909 ซึ่งจะเห็นได้ว่าค่าถ่วงน้ำหนักทั้ง 2 ค่าจะไม่เป็นศูนย์ตามที่ Freed and Glover (1986) ได้อ้างไว้ และจะสังเกตได้ว่าทั้งค่า β_1, β_2 และ c จะมีค่าไม่แตกต่างกับค่าสังเกตมากนัก ซึ่งเกิดจากการกำหนดค่า N ให้เท่ากับ 10 ในตัวแบบ หากค่าสังเกตที่นำมาวิเคราะห์มีค่าน้อยกว่าหลักหน่วย หรือเป็นหลักร้อย อาจจะแทนค่าให้ N = 1 หรือ 100 จะให้ผลที่เหมาะสมกว่า ซึ่งผู้วิเคราะห์สามารถกำหนดได้ตามความเหมาะสม

จากนั้นคำนวณคะแนนการจำแนก (คะแนนการจำแนกพิจารณาจาก $\sum_{i=1}^{p} \beta_i x_{ij}$) ในแต่ละค่าสังเกตของลูกจ้างเพื่อใช้ในการจัดเข้ากลุ่มโดยการนำไปเทียบกับค่า c กล่าวคือ ถ้า $\sum_{i=1}^{p} \beta_i x_{ij} \le c$ แสดงว่าควรจัดค่าสังเกตเป็นกลุ่มลูกจ้างที่ไม่ประสบผลสำเร็จ (หรือกลุ่ม 1) และถ้า $\sum_{i=1}^{p} \beta_i x_{ij} > c$ แสดงว่าควรจัดค่าสังเกตเป็นกลุ่มลูกจ้างที่ประสบผลสำเร็จ (หรือกลุ่ม 2) ผลแสดงดังตารางที่ 2 ซึ่งจะเห็นได้ว่าคะแนนการจำแนกของลูกจ้างในกลุ่มไม่ประสบผลสำเร็จมีค่าน้อยกว่าหรือเท่ากับ 9.0909 ทุกค่า ในขณะที่คะแนนการจำแนกของลูกจ้างในกลุ่มประสบผลสำเร็จมีค่ามากกว่า 9.0909 ซึ่งเป็นค่าจุดตัดทั้งหมด 4 ค่าโดยมีลูกจ้างคนที่ 6 ที่มีคะแนนการจำแนกเท่ากับค่าของจุดตัดจึงได้ผลการจัดเข้ากลุ่มผิดกลุ่ม

ตารางที่ 2 ผลการจัดเข้ากลุ่มจากคุณลักษณะของลูกจ้างทั้ง 10 คนในกลุ่มไม่ประสบผลสำเร็จ และประสบผลสำเร็จ

ลูกจ้างคนที่	กลุ่ม 1: ไม่ประสบผลสำเร็จ		คะแนนการจำแนก ($\sum_{i=1}^{p} \beta_i x_{ij}$)	ผลการจัดเข้ากลุ่ม	ลูกจ้างคนที่	กลุ่ม 2: ประสบผลสำเร็จ		คะแนนการจำแนก ($\sum_{i=1}^{p} \beta_i x_{ij}$)	ผลการจัดเข้ากลุ่ม
	x_1	x_2				x_1	x_2		
1	1	1	0.9091	ถูก	**6**	4	1	9.0909	ผิด
2	3	1	6.3636	ถูก	**7**	5	2	10.0000	ถูก
3	2	2	1.8182	ถูก	**8**	7	2	15.4545	ถูก
4	6	4	9.0909	ถูก	**9**	9	1	22.7273	ถูก
5	3	3	2.7273	ถูก	**10**	8	4	14.5455	ถูก

เมื่อทำการพลอตกราฟ เพื่อแสดงฟังก์ชันจำแนกจากการคำนวณการโปรแกรมเชิงเส้นตามตัวแบบ MSD ด้วยโปรแกรม Microsoft Excel ได้ผลตามรูปที่ 6 ซึ่งจะเห็นได้ว่ามีค่าสังเกตที่จัดเข้ากลุ่มถูกต้องอย่างชัดเจนจำนวน 8 ค่า โดยมีข้อน่าสังเกตเกี่ยวกับค่าสังเกตที่อยู่บนระนาบ 2 ค่า คือ ค่าสังเกตที่ 4 ในกลุ่มของลูกจ้างที่ไม่ประสบผลสำเร็จ (กลุ่ม 1) และค่าสังเกตที่ 6 ในกลุ่มของลูกจ้างที่ประสบผลสำเร็จ (กลุ่ม 2) แต่จากกฎการจัดเข้ากลุ่ม (คือ ถ้า $\sum_{i=1}^{p}\beta_i x_{ij} \le c$ แสดงว่าค่าสังเกตอยู่กลุ่ม 1 และถ้า $\sum_{i=1}^{p}\beta_i x_{ij} > c$ แสดงว่าค่าสังเกตอยู่กลุ่ม 2) จึงสรุปได้ว่าค่าสังเกตที่ 4 จัดเข้ากลุ่มถูก ส่วนค่าสังเกตที่ 6 จัดเข้ากลุ่มผิด



รูปที่ 6 ฟังก์ชันการจำแนกจากคุณลักษณะของลูกจ้างทั้ง 10 คนในกลุ่มไม่ประสบผลสำเร็จ และกลุ่มประสบผลสำเร็จ

สรุป

การวิเคราะห์การจำแนกเป็นการค้นหาสารสนเทศที่สำคัญอย่างหนึ่ง เพื่อให้สามารถใช้ในการพิจารณาลักษณะที่ต้องการบางประการ เช่น ควรให้กู้หรือไม่ ควรเลือกหรือไม่ หายหรือไม่ ผ่านหรือไม่ เป็นต้น โดยทั่วไปแล้ว การวิเคราะห์การจำแนกจะใช้วิธีการของสถิติที่ใช้พารามิเตอร์ ได้แก่ การจำแนกโดยใช้ฟังก์ชันเชิงเส้นของฟิชเชอร์ และการวิเคราะห์การจำแนกโดยใช้การถดถอยแบบโลจิสติกเป็นอาทิ อย่างไรก็ตาม ในบางกรณีจะพบว่าการจำแนกด้วยวิธีการของสถิติที่ใช้พารามิเตอร์นั้นมีข้อจำกัดบางประการ ทำให้ไม่สามารถวิเคราะห์การจำแนกได้ แนวทางหนึ่งในการแก้ปัญหาดังกล่าว คือ การใช้ตัวแบบการโปรแกรมเชิงเส้นในการวิเคราะห์การจำแนก และเนื่องด้วยการโปรแกรมเชิงเส้นเป็นวิธีการทางสถิติที่ไม่ใช้พารามิเตอร์ จึงทำให้ไม่มีข้อจำกัดเหล่านั้น

การวิเคราะห์การจำแนกด้วยการโปรแกรมเชิงเส้นตามตัวแบบ MSD ที่เสนอโดย Freed and Glover (1986) เป็นตัวแบบที่ค่อนข้างง่าย และเห็นได้ว่าการคำนวณไม่ได้มีความซับซ้อนมากนัก โดยสามารถใช้โปรแกรมสำเร็จรูป Microsoft Excel ทำการคำนวณได้ ซึ่งเหมาะสำหรับบุคคลทั่วไปที่มีความคุ้นเคยกับโปรแกรมนี้ ทำให้ลด

ขั้นตอนความซับซ้อนของโปรแกรมคอมพิวเตอร์ที่ใช้ในการคำนวณโปรแกรมเชิงเส้น ซึ่งน่าจะเป็นทางเลือกหนึ่งของการวิเคราะห์การจำแนกได้เป็นอย่างดี

เอกสารอ้างอิง

Freed, N., and Glover, F. 1981. "A Linear Programming Approach to Discriminant Problem." **Decision Sciences** 12: 68-74.

_____. 1986. "Evaluating Alternative Linear Programming Models to Solve the Two-Group Discriminant Problem." **Decision Sciences** 17: 151-162.

Hand, D.J. 1981. **Discrimination and Classification.** New York: Wiley.

Mr. Afifi Lateh received his Master of Science in Applied Statistics from Silpakorn University, and Bachelor of Science in Education of Mathematics from Prince of Songkla University. He is currently a lecturer at the Department of Educational Evaluation and Research, Faculty of Education, Prince of Songkla University. His research focuses on Educational Research, Mathematical Programming, Data Envelopment Analysis and Multivariate Analysis.

Asst. Prasopchai Pasunon received his Master of Science in Applied Statistics from Silpakorn University, and Bachelor of Science in Mathematics from Thaksin University. Currently, he is an Assistant Professor in the General Business Management Department and an Associate Dean for Research and Graduate Studies, Faculty of Management Science, Silpakorn University. His research interests are Outliers Detection, Data Envelopment Analysis, Multivariate Analysis and Analysis for Business Data.