วารสารวิชาการ **มหาวิทยาลัยหอการค้าไทย**
ปีที่ 31 ฉบับที่ 4 เดือนตุลาคม - ธันวาคม 2554

การโปรแกรมเชิงเส้นแบบ 2 ขั้นสำหรับการวิเคราะห์การจำแนกกลุ่ม: การคำนวณด้วยโปรแกรม Microsoft Excel

A Two-stage Linear Programming for Discriminant Analysis: Calculated with the 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 ผู้เขียนต้องการนำเสนอการโปรแกรมเชิงเส้นแบบ 2 ขั้น สำหรับการวิเคราะห์การจำแนกตามตัวแบบที่เสนอโดย Lam, Choo, and Moy (1996) ซึ่งเป็นการจำแนกข้อมูล 2 กลุ่มพร้อมการคำนวณด้วยโปรแกรม Microsoft Excel (เวอร์ชั่น 2003) โดยใช้ตัวอย่างข้อมูลของ Freed and Glover (1981) แสดงการคำนวณผลจากคำนวณการโปรแกรมเชิงเส้นด้วยโปรแกรม Microsoft Excel พบว่า ตัวแบบที่เสนอโดย Lam, Choo, and Moy (1996) สามารถแก้ปัญหา trivial solution ซึ่งเป็นผลลัพธ์ที่เกิดจากตัวแบบที่เสนอโดย Freed and Glover (1986) ได้

คำสำคัญ: การวิเคราะห์การจำแนก การโปรแกรมเชิงเส้นแบบ 2 ขั้น โปรแกรม Microsoft Excel (เวอร์ชั่น 2003) การวิเคราะห์ตัวแปรพหุ

Abstract

In this article, the writers present a two-stage linear programming model proposed by Lam, Choo and Moy (Lam, Choo, and Moy, 1996) for two-group discriminant analysis calculated with Microsoft Excel (version 2003) program. Data from the paper presented by Freed and Glover (1981) was used, the results were as follows: the model proposed by Lam, Choo and Moy (Lam, Choo, and Moy, 1996) can be used to solve trivial solutions, which agreed with results from the model presented by Freed and Glover (1986).

Keywords: Discriminant Analysis, Two-stage Linear Programming, Microsoft Excel (Version 2003) Program, Multivariate Analysis.

บทนำ

การวิเคราะห์การจำแนก (Discriminant Analysis) โดยการโปรแกรมเชิงเส้น (Linear Programming) ได้รับการพัฒนาเป็นเวลาร่วม 30 ปีมาแล้ว แต่ความนิยมในการใช้วิธีการดังกล่าวยังน้อยมากในประเทศไทย ไม่ว่าจะเป็นในแง่การวิจัย การนำไปประยุกต์ใช้ หรือการเรียนการสอน ซึ่งอาจจะเนื่องมาจากความยุ่งยากในการทำความเข้าใจขั้นตอนวิธีการคำนวณการโปรแกรมเชิงเส้นตลอดจนโปรแกรมสำเร็จรูปในการคำนวณมีขั้นตอนที่ซับซ้อนต่อการทำความเข้าใจ บทความนี้เป็นบทความที่ต่อเนื่องจากบทความ เรื่อง “การโปรแกรมเชิงเส้นสำหรับการวิเคราะห์การจำแนกกลุ่มโดยการคำนวณด้วยโปรแกรม Microsoft Excel” ซึ่งได้เสนอการโปรแกรมเชิงเส้นสำหรับการวิเคราะห์การจำแนกตามตัวแบบที่เสนอโดย Freed, and Glover (1986) คือ ตัวแบบ MSD (Minimize the Sum of Deviation) พร้อมการคำนวณด้วยโปรแกรม Microsoft Excel (อาฟีฟี ลาเต๊ะ และ ประสพชัย พสุนนท์, 2552)

การวิเคราะห์การจำแนกด้วยวิธีการโปรแกรมเชิงเส้นนั้น มีข้อได้เปรียบจากวิธีการทางสถิติทั้งวิธีการวิเคราะห์การจำแนกโดยใช้ฟังก์ชันเชิงเส้นของฟิชเชอร์ (Fisher’s Linear Discriminant Function) และวิธีการวิเคราะห์การจำแนกโดยใช้การถดถอยแบบโลจีสติก (Logistic Regression Discriminant Analysis) นั่นคือ วิธีการแรกมีข้อสมมติในเรื่องการแจกแจงแบบปกติของตัวแปรพหุและการเท่ากันของเมทริกซ์ความแปรปรวนร่วมของตัวแปรอิสระระหว่างกลุ่ม ส่วนวิธีการที่ 2 มีข้อสมมติในเรื่องการแจกแจงแบบปกติของความคลาดเคลื่อนของตัวแปรอิสระ ส่วนการวิเคราะห์การจำแนกโดยการโปรแกรมเชิงเส้นนั้น ข้อมูลในการวิเคราะห์ไม่จำเป็นต้องสอดคล้องกับข้อสมมติข้างต้น เนื่องจากว่าเป็นการวิเคราะห์ที่อาศัยกระบวนการทำซ้ำด้วยขั้นตอนวิธีทางคณิตศาสตร์ซึ่งเป็นการสร้างตัวแบบแทนระบบปัญหา โดยตัวแบบจะประกอบด้วย ฟังก์ชันจุดประสงค์ หรือฟังก์ชันเป้าหมาย (Objective Function) และเงื่อนไขบังคับ หรือข้อจำกัด (Constraints) โดยมีจุดมุ่งหมายเพื่อแก้ปัญหาและตัดสินใจให้เกิดผลตาม

แนวทางการดำเนินงานที่ดีที่สุดภายใต้เงื่อนไขที่กำหนด

บทความนี้ ผู้เขียนจะเสนอตัวแบบการโปรแกรมเชิงเส้นที่เสนอโดย Lam, Choo, and Moy (1996) ซึ่งเป็นอีกตัวแบบหนึ่งที่ไม่ซับซ้อน เพื่อเป็นอีกหนึ่งทางเลือกในการนำไปใช้ในการวิจัย การนำไปประยุกต์ใช้ หรือใช้ในการเรียนการสอนหัวข้อการวิเคราะห์การจำแนกในวิชาการวิเคราะห์ตัวแปรพหุ และได้เลือกใช้โปรแกรม Microsoft Excel (เวอร์ชั่น 2003) เป็นเครื่องมือในการคำนวณ เนื่องจากเป็นโปรแกรมสำเร็จรูปที่มีความนิยมในการใช้งานอย่างแพร่หลาย และสามารถประยุกต์ใช้งานได้ค่อนข้างสะดวก

การวิเคราะห์การจำแนกโดยการโปรแกรมเชิงเส้นแบบ 2 ขั้น

ตัวแบบการโปรแกรมเชิงเส้นแบบ 2 ขั้นที่จะนำเสนอเป็นตัวแบบของ Lam, Choo and Moy (จะเรียกย่อๆ ว่า ตัวแบบ LCM) เสนอในปี ค.ศ. 1996 เป็นตัวแบบที่ใช้ในการจำแนกข้อมูลกรณี 2 กลุ่ม และมีพื้นฐานอยู่บนแนวคิดของการวิเคราะห์กลุ่ม (Cluster Analysis) ซึ่งเป็นเทคนิควิธีของการวิเคราะห์ตัวแปรพหุเพื่อใช้ในการแบ่งกลุ่มค่าสังเกต โดยมีหลักเกณฑ์ให้ค่าสังเกตที่มีลักษณะเหมือนกันหรือคล้ายกันอยู่กลุ่มเดียวกัน และค่าสังเกตที่มีลักษณะต่างกันอยู่ต่างกลุ่มกัน สำหรับขั้นตอนของตัวแบบ LCM เป็นดังนี้

ขั้นที่หนึ่ง เป็นการหาค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนระหว่างคะแนนของค่าสังเกตและค่าเฉลี่ยของคะแนนในแต่ละกลุ่ม (Group Mean) ซึ่งเป็นแนวคิดที่แตกต่างจากแนวคิดเดิมของวิธีการโปรแกรมเชิงเส้นที่มักใช้ค่าของจุดตัด

ขั้นที่สอง เป็นการหาค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนระหว่างคะแนนการจำแนกและจุดตัด

ตัวแบบ LCM ได้รับการพัฒนาเพื่อแก้ปัญหาเกี่ยวกับผลลัพธ์ที่ไม่สามารถยอมรับได้ (Unacceptable Solution) หรือที่เรียกว่าเกิดปัญหา trivial solution ปัญหาเกี่ยวกับผลลัพธ์ที่ไม่มีขอบเขต (Unbounded Solution) หรือผลลัพธ์ที่ได้ไม่ลู่เข้าค่าใดค่าหนึ่ง รวมทั้งปัญหาเกี่ยวกับการได้ผลลัพธ์ที่ไม่เหมาะสม (Improper Solution) หรือผลลัพธ์ที่ได้จากการโปรแกรมเชิงเส้นทำให้ได้คะแนนการจำแนกอยู่บนเส้นของระนาบเกิน (Hyperplane) ซึ่งเป็นระนาบหรือเส้นแบ่งระหว่างกลุ่มของค่าสังเกต

กำหนดสัญลักษณ์ของตัวแบบ LCM ดังนี้

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

x_{ij} เป็นค่าสังเกตลำดับที่ j ของตัวแปรอิสระที่ i

p เป็นจำนวนตัวแปรอิสระ

μ_{1i} และ μ_{2i} เป็นค่าเฉลี่ยของตัวแปรอิสระที่ i ในกลุ่มที่ 1 และ 2 ตามลำดับ

d_j^- และ d_j^+ เป็นตัวแปรเบี่ยงเบนทางซ้ายและทางขวาของค่าสังเกตลำดับที่ j ในกลุ่มที่ 1

e_j^- และ e_j^+ เป็นตัวแปรเบี่ยงเบนทางซ้ายและทางขวาของค่าสังเกตลำดับที่ j ในกลุ่มที่ 2

n_1 และ n_2 เป็นจำนวนค่าสังเกตในกลุ่มที่ 1 และ 2 ตามลำดับ

S_j เป็นคะแนนการจำแนกของค่าสังเกตลำดับที่ j

h_j เป็นตัวแปรเบี่ยงเบนของค่าสังเกตลำดับที่ j

c เป็นจุดตัด (cut off)

m เป็นจำนวนค่าสังเกตทั้งหมด

ขั้นที่หนึ่ง Minimize $\sum_{j=1}^{n_1}(d_j^- + d_j^+) + \sum_{j=1}^{n_2}(e_j^- + e_j^+)$

Subject to

$$\sum_{i=1}^{p}(x_{ij} - \mu_{1i})\beta_i + d_j^- - d_j^+ = 0$$

สำหรับทุกค่า j ในกลุ่มที่ 1 .....(1)

$$\sum_{i=1}^{p}(x_{ij} - \mu_{2i})\beta_i + e_j^- - e_j^+ = 0$$

สำหรับทุกค่า j ในกลุ่มที่ 2 .....(2)

$$\sum_{i=1}^{p}(\mu_{1i} - \mu_{2i})\beta_i \geq 10 \quad .....(3)$$

β_i ไม่จำกัดเครื่องหมาย

$d_j^-, d_j^+, e_j^-, e_j^+ \geq 0$

ขั้นที่สอง Minimize $\sum_{j=1}^{m} h_j$

Subject to

$S_j - h_j - c \leq 0$ สำหรับทุกค่า j ในกลุ่มที่ 1 .....(4)

$S_j + h_j - c \geq 0$ สำหรับทุกค่า j ในกลุ่มที่ 2 .....(5)

c ไม่จำกัดเครื่องหมาย

$h_j \geq 0$

ในขั้นแรกของตัวแบบ LCM มีเป้าหมายเพื่อหาค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนระหว่างคะแนนของค่าสังเกตและค่าเฉลี่ยของคะแนนในแต่ละกลุ่ม ภายใต้เงื่อนไข (1) และ (2) สำหรับเงื่อนไข (3) เป็นเงื่อนไขที่เพิ่มในตัวแบบเพื่อไม่ให้เกิดกรณีที่ $\beta_i = 0$ ในทุกค่าของ i ซึ่งเรียกว่า trivial solution หรือเกิดกรณีที่ได้คะแนนการจำแนกอยู่บนเส้นของระนาบเกิน โดยค่าคงที่อาจจะกำหนดเป็นค่าใดก็ได้เพื่อให้ผลลัพธ์ของ $\beta_i, d_j^-, d_j^+, e_j^-, e_j^+$ ซึ่งเป็นค่าถ่วงน้ำหนักของตัวแปรอิสระและค่าของตัวแปรเบี่ยงเบนในแต่ละค่าสังเกตให้ค่าของจำนวนที่ได้ไม่น้อยหรือมากจนเกินไป ในที่นี้จะกำหนดให้เป็น 10 เพื่อสอดคล้องกับการกำหนดค่า N ในตัวแบบ MSD ของ Freed, and Glover (1986) ซึ่งได้เสนอในบทความก่อนหน้านี้

เมื่อแก้ปัญหาในขั้นแรกแล้ว นำค่าถ่วงน้ำหนักที่ได้ไปคูณกับค่าสังเกตเพื่อเป็นคะแนนการจำแนก (Classification Score: S_j) เมื่อได้คะแนนการจำแนกของแต่ละค่าสังเกตแล้วจะใช้ตัวแบบในขั้น 2 เพื่อหาค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนระหว่างคะแนนการจำแนกและจุดตัดภายใต้เกณฑ์การจัดเข้ากลุ่มของกลุ่มที่ 1 และกลุ่มที่ 2 ตามเงื่อนไข (4) และ (5) โดยมีข้อสมมติว่ากลุ่ม 1 อยู่ใต้จุดตัด กลุ่ม 2 อยู่เหนือจุดตัด การแก้ปัญหาในขั้น 2 นี้จะได้ค่า h_j และ c ซึ่งเป็นค่าของตัวแปรเบี่ยงเบนและค่าจุดตัดของแต่ละค่าสังเกต จากนั้นนำค่า S_j และ c มาใช้จัดกลุ่มค่าสังเกต นั่นคือ หาก $S_j \leq c$ แสดงว่าค่าสังเกตอยู่กลุ่ม 1 และถ้า $S_j > c$ แสดงว่าค่าสังเกตอยู่กลุ่ม 2

การคำนวณโปรแกรมเชิงเส้นแบบ 2 ขั้นด้วยโปรแกรม Microsoft Excel

สำหรับขั้นตอนการคำนวณตัวแบบ LCM ด้วยโปรแกรม Microsoft Excel นั้น ผู้เขียนเลือกใช้เวอร์ชั่น 2003 เนื่องจากฟังก์ชันการทำงานในเวอร์ชั่นนี้มีความคล่องตัวต่อการใช้งานและสามารถใช้งานได้อย่างไม่ยากนัก ส่วนข้อมูลที่ใช้แสดงตัวอย่างการคำนวณ ผู้เขียนเลือกใช้ข้อมูลของ Freed, and Glover (1981) มาใช้เพื่อแสดงการคำนวณดังได้เสนอไปแล้วในตัวแบบ MSD (อาฬีฬี ลาเต๊ะ และ ประสพชัย พสุนนท์, 2552) ซึ่งเป็นข้อมูลของลูกจ้างจำนวน 10 คน แบ่งเป็นกลุ่มไม่ประสบผลสำเร็จ และกลุ่มประสบผลสำเร็จในการทำงานซึ่งวัดจากคุณลักษณะของตัวแปร 2 ตัว นั่น

คือ ประสบการณ์ในการทำงาน (x_1) และระยะเวลาที่จบการศึกษา (x_2) โดยขั้นแรกในการคำนวณจากตัวแบบ LCM นั้นจะเป็นการหาค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนระหว่างคะแนนของค่าสังเกตและค่าเฉลี่ยของคะแนนในแต่ละกลุ่ม จึงต้องหาค่าเฉลี่ยของแต่ละตัวแปรอิสระจากกลุ่มที่ 1 (μ_{1i}) ซึ่งมีค่าเท่ากับ 3.0 และ 2.2 ตามลำดับ และค่าเฉลี่ยของแต่ละตัวแปรอิสระจากกลุ่มที่ 2 (μ_{2i}) ซึ่งมีค่าเท่ากับ 6.6 และ 2.0 ตามลำดับ ต่อไปหาผลต่างระหว่างคุณลักษณะกับค่าเฉลี่ยแต่ละตัวแปรอิสระจากแต่ละกลุ่ม ($x_1 - \mu_{ki}$ และ $x_2 - \mu_{ki}$) ผลแสดงดังตารางที่ 1

ตารางที่ 1 คุณลักษณะของลูกจ้างทั้ง 10 คนจากกลุ่มไม่ประสบผลสำเร็จ และกลุ่มประสบผลสำเร็จ

กลุ่ม	ลูกจ้างคนที่	x_1	x_2	$x_1 - \mu_{ki}, k = 1,2$	$x_2 - \mu_{ki}, k = 1,2$
ไม่ประสบผลสำเร็จ	1	1	1	-2	-1.2
	2	3	1	0	-1.2
	3	2	2	-1	-0.2
	4	6	4	3	1.8
	5	3	3	0	0.8
ประสบผลสำเร็จ	6	4	1	-2.6	-1
	7	5	2	-1.6	0
	8	7	2	0.4	0
	9	9	1	2.4	-1
	10	8	4	1.4	2

จากตารางที่ 1 สามารถเขียนตัวแบบ LCM ขั้นที่หนึ่งเพื่อคำนวณหาค่าเหมาะสุดจากการโปรแกรมเชิงเส้นของลูกจ้างที่ไม่ประสบผลสำเร็จและประสบผลสำเร็จได้ ดังนี้

ขั้นที่หนึ่ง **Minimize**

$$(d_1^- + d_1^+ + d_2^- + d_2^+ + d_3^- + d_3^+ + d_4^- + d_4^+ + d_5^- + d_5^+)$$

$$+(e_1^- + e_1^+ + e_2^- + e_2^+ + e_3^- + e_3^+ + e_4^- + e_4^+ + e_5^- + e_5^+)$$

Subject to

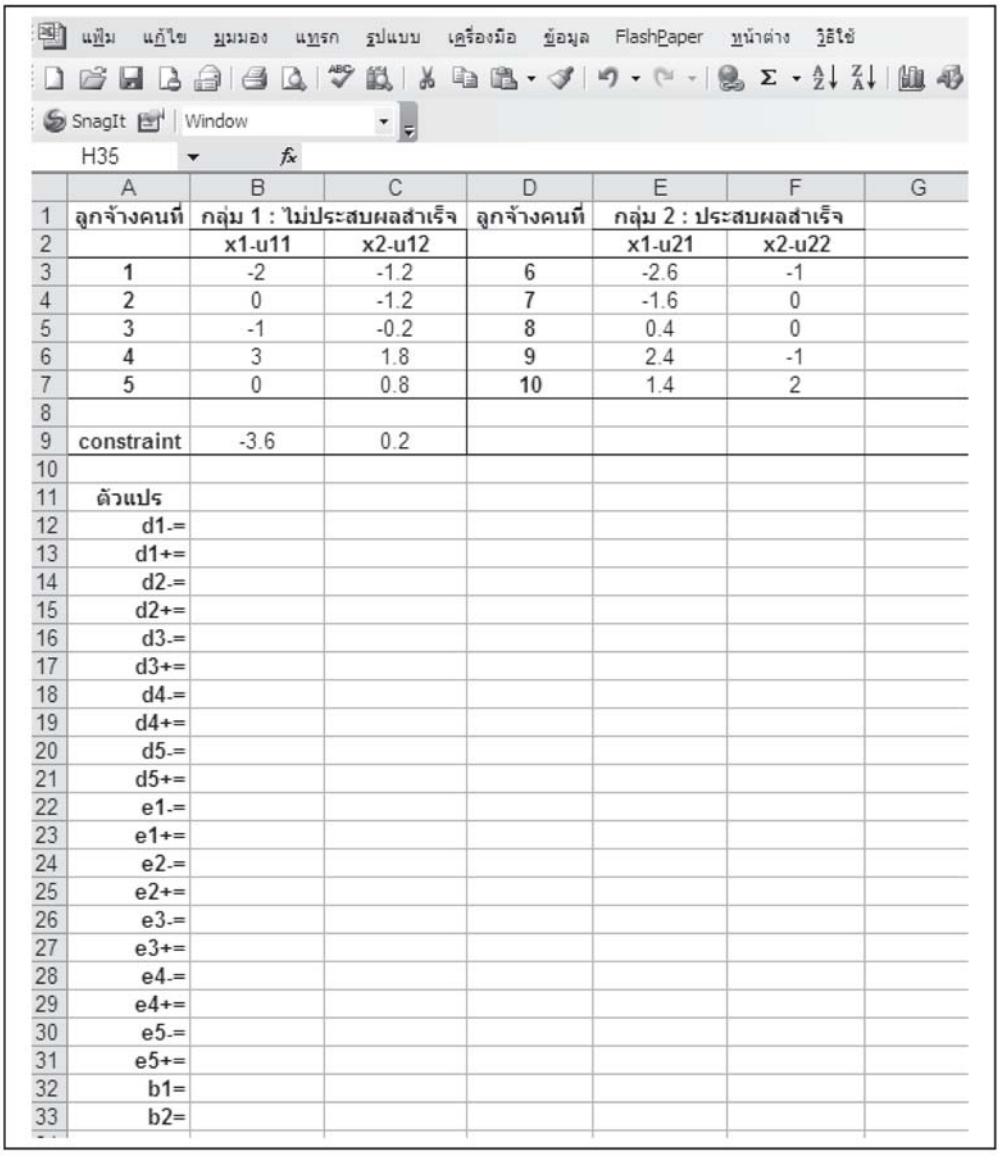
$$-2\beta_1 - 1.2\beta_2 + d_1^- - d_1^+ = 0$$
$$-1.2\beta_2 + d_2^- - d_2^+ = 0$$
$$-\beta_1 - 0.2\beta_2 + d_3^- - d_3^+ = 0$$
$$3\beta_1 + 1.8\beta_2 + d_4^- - d_4^+ = 0$$
$$0.8\beta_2 + d_5^- - d_5^+ = 0$$
$$-2.6\beta_1 - 1\beta_2 + e_1^- - e_1^+ = 0$$
$$-1.6\beta_1 + e_2^- - e_2^+ = 0$$
$$0.4\beta_1 + e_3^- - e_3^+ = 0$$
$$2.4\beta_1 - \beta_2 + e_4^- - e_4^+ = 0$$
$$1.4\beta_1 + 2\beta_2 + e_5^- - e_5^+ = 0$$
$$-3.6\beta_1 + 0.2\beta_2 \geq 10$$

β_1 และ β_2 ไม่จำกัดเครื่องหมาย

$$d_1^-, d_1^+, d_2^-, d_2^+, ..., d_5^-, d_5^+,$$
$$e_1^-, e_1^+, e_2^-, e_2^+, ..., e_5^-, e_5^+ \geq 0$$

สำหรับการคำนวณการโปรแกรมเชิงเส้นของตัวแบบ LCM ในขั้นที่หนึ่งด้วยโปรแกรม Microsoft Excel มีขั้นตอนแสดงดังภาพที่ 1-4 ดังนี้

ขั้น 1 เตรียมข้อมูลสำหรับการคำนวณ และกำหนดตัวแปรในการคำนวณโดยให้ cell ที่ B12, B13, ..., B31 แทนค่าของตัวแปร $d_1^-, d_1^+, ..., e_5^+$ ตามลำดับ และให้ cell ที่ B32 และ B33 แทนค่าของตัวแปร b_1 (β_1) และ b_2 (β_2) ตามลำดับ



	A	B	C	D	E	F	G
1	ลูกจ้างคนที่	กลุ่ม 1 : ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2 : ประสบผลสำเร็จ		
2		x1-u11	x2-u12		x1-u21	x2-u22	
3	1	-2	-1.2	6	-2.6	-1	
4	2	0	-1.2	7	-1.6	0	
5	3	-1	-0.2	8	0.4	0	
6	4	3	1.8	9	2.4	-1	
7	5	0	0.8	10	1.4	2	
8							
9	constraint	-3.6	0.2				
10							
11	ตัวแปร						
12	d1-=						
13	d1+=						
14	d2-=						
15	d2+=						
16	d3-=						
17	d3+=						
18	d4-=						
19	d4+=						
20	d5-=						
21	d5+=						
22	e1-=						
23	e1+=						
24	e2-=						
25	e2+=						
26	e3-=						
27	e3+=						
28	e4-=						
29	e4+=						
30	e5-=						
31	e5+=						
32	b1=						
33	b2=						

ภาพที่ 1 การเตรียมข้อมูล และการกำหนดตัวแปรในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ขั้น 2 กำหนด Subject to ของตัวแบบโดยให้ cell ที่ C12, C13, ..., C22 ซึ่งมีค่าเป็น

$-2\beta_1 - 1.2\beta_2 + d_1^- - d_1^+$ แทนด้วย =B3*B32+C3*B33+B12-B13

$-1.2\beta_2 + d_2^- - d_2^+$ แทนด้วย =B4*B32+C4*B33+B14-B15

$-\beta_1 - 0.2\beta_2 + d_3^- - d_3^+$ แทนด้วย =B5*B32+C5*B33+B16-B17

$3\beta_1 + 1.8\beta_2 + d_4^- - d_4^+$ แทนด้วย =B6*B32+C6*B33+B18-B19

$0.8\beta_2 + d_5^- - d_5^+$ แทนด้วย =B7*B32+C7*B33+B20-B21

$-2.6\beta_1 - 1\beta_2 + e_1^- - e_1^+$ แทนด้วย =E3*B32+F3*B33+B22-B23

$-1.6\beta_1 + e_2^- - e_2^+$ แทนด้วย =E4*B32+F4*B33+B24-B25

$0.4\beta_1 + e_3^- - e_3^+$ แทนด้วย =E5*B32+F5*B33+B26-B27

$2.4\beta_1 - \beta_2 + e_4^- - e_4^+$ แทนด้วย =E6*B32+F6*B33+B28-B29

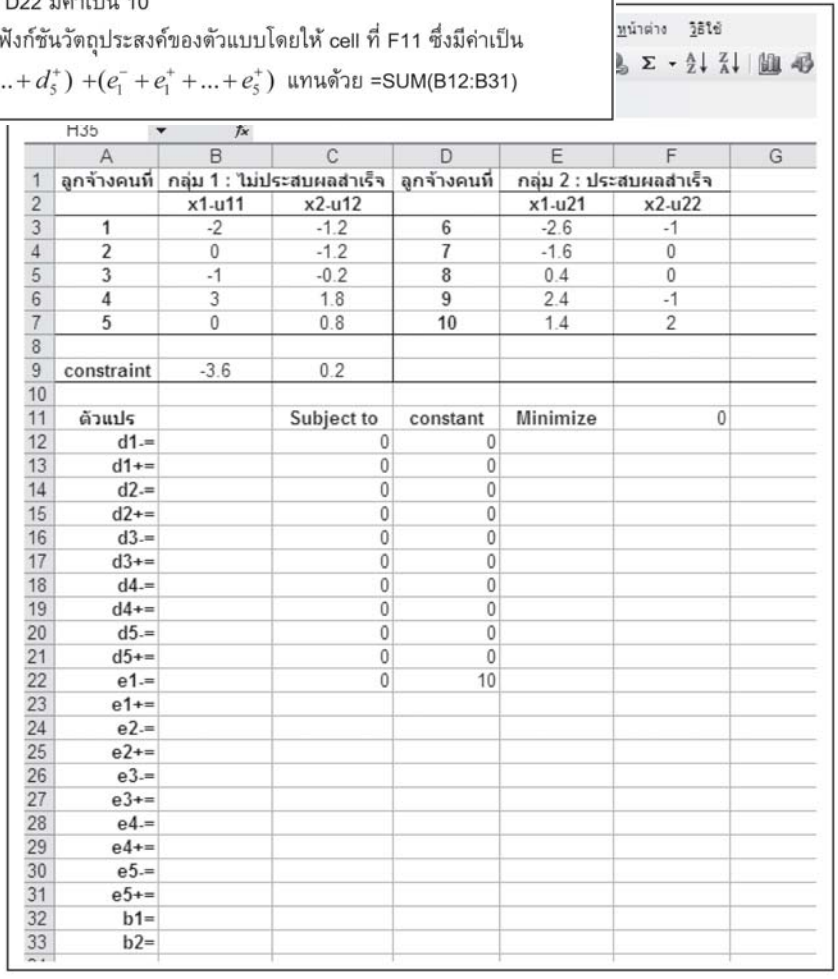
$1.4\beta_1 + 2\beta_2 + e_5^- - e_5^+$ แทนด้วย =E7*B32+F7*B33+B30-B31

$-3.6\beta_1 + 0.2\beta_2$ แทนด้วย =B9*B32+C9*B33 ตามลำดับ

กำหนด constant (ค่าคงที่) ของตัวแบบโดยให้ cell ที่ D12, D13, ..., D21 มีค่าเป็น 0 ส่วน D22 มีค่าเป็น 10

กำหนดฟังก์ชันวัตถุประสงค์ของตัวแบบโดยให้ cell ที่ F11 ซึ่งมีค่าเป็น

$(d_1^- + d_1^+ + ... + d_5^+) + (e_1^- + e_1^+ + ... + e_5^+)$ แทนด้วย =SUM(B12:B31)



H35

	A	B	C	D	E	F	G
1	ลูกจ้างคนที่	กลุ่ม 1 : ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2 : ประสบผลสำเร็จ		
2		x1-u11	x2-u12		x1-u21	x2-u22	
3	1	-2	-1.2	6	-2.6	-1	
4	2	0	-1.2	7	-1.6	0	
5	3	-1	-0.2	8	0.4	0	
6	4	3	1.8	9	2.4	-1	
7	5	0	0.8	10	1.4	2	
8							
9	constraint	-3.6	0.2				
10							
11	ตัวแปร		Subject to	constant	Minimize	0	
12	d1-=		0	0			
13	d1+=		0	0			
14	d2-=		0	0			
15	d2+=		0	0			
16	d3-=		0	0			
17	d3+=		0	0			
18	d4-=		0	0			
19	d4+=		0	0			
20	d5-=		0	0			
21	d5+=		0	0			
22	e1-=		0	10			
23	e1+=						
24	e2-=						
25	e2+=						
26	e3-=						
27	e3+=						
28	e4-=						
29	e4+=						
30	e5-=						
31	e5+=						
32	b1=						
33	b2=						

ภาพที่ 2 การกำหนด Subject to, constant และฟังก์ชันวัตถุประสงค์ของตัวแบบในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ขั้น 3 คำนวณโดยใช้ฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบ โดยคลิกที่ **เครื่องมือ (Tools)** จากแถบเมนู แล้วเลือก **Solver...** จากนั้นป้อนค่าฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบดังนี้

★ **Set Target Cell:** คือ การกำหนดฟังก์ชันวัตถุประสงค์ แทนด้วย cell ที่ F11 เมื่อคลิกที่ F11 จะปรากฏ $F$11 อัตโนมัติ

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

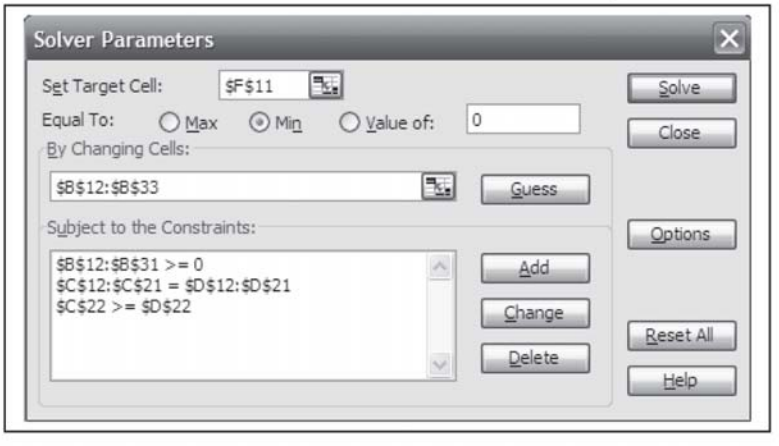
★ **By Changing Cells:** คือ การกำหนดการเปลี่ยนแปลงของตัวแปรในตัวแบบ ในกรณีนี้ตัวแปร คือ B12, B13, ..., B33 โดยสามารถกด Shift ที่แป้นพิมพ์ค้างไว้แล้วลากเมาส์เลือก cell ที่ B12, B13, ..., B33 ซึ่งจะปรากฏ $B$12:$B$33 อัตโนมัติ

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

เงื่อนไข $\sum_{i=1}^{p}(x_{ij}-\mu_{1i})\beta_i + d_j^- - d_j^+ = 0$ สำหรับทุกค่า j ในกลุ่มที่ 1 และเงื่อนไข $\sum_{i=1}^{p}(x_{ij}-\mu_{2i})\beta_i + e_j^- - e_j^+ = 0$ สำหรับทุกค่า j ในกลุ่มที่ 2 แทนด้วย $C$12:$C$21 =$D$12:$D$21

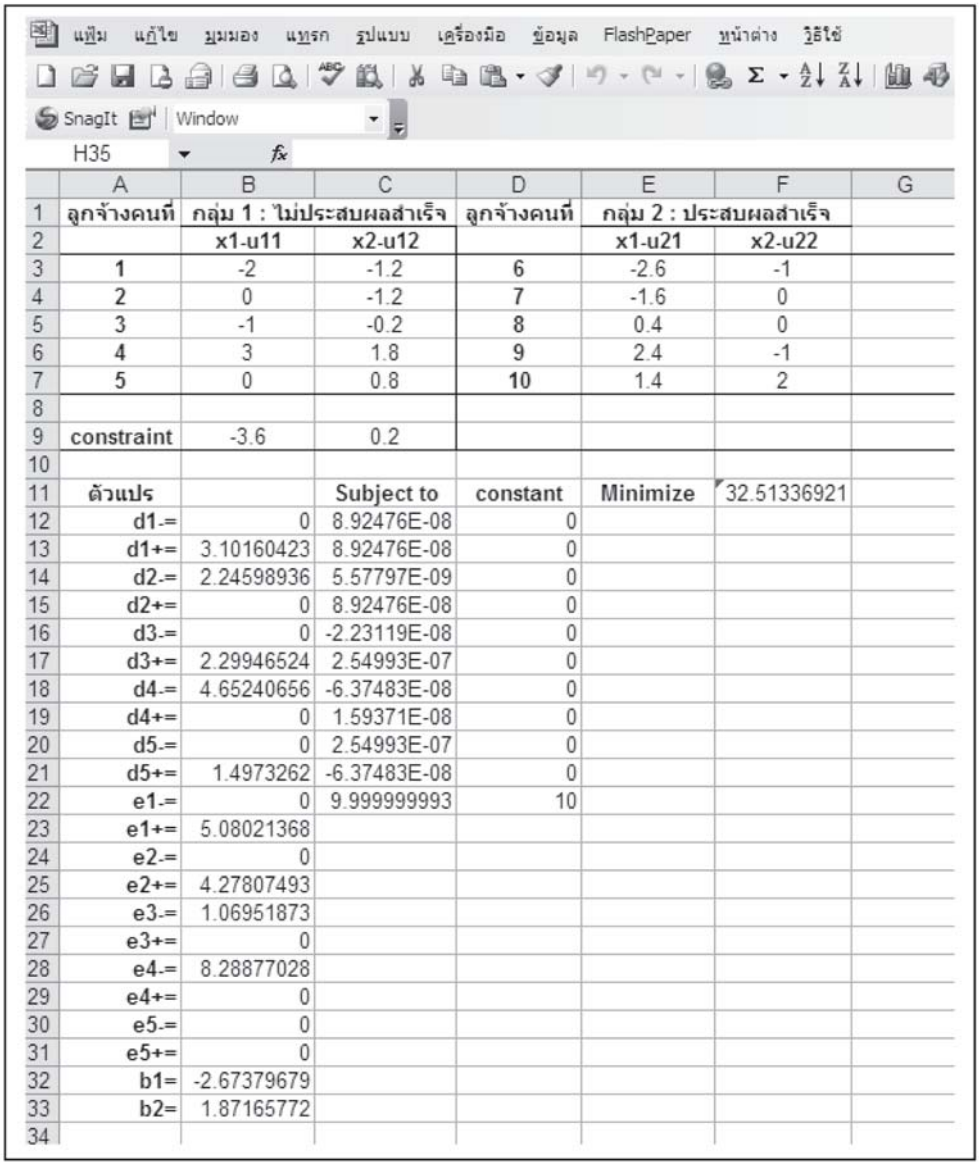
เงื่อนไข $\sum_{i=1}^{p}(\mu_{1i}-\mu_{2i})\beta_i \geq 10$ แทนด้วย $C$22 >= $D$22

เงื่อนไข $d_j^-, d_j^+, e_j^-, e_j^+ \geq 0$ แทนด้วย $B$12:$B$31 >= 0



ภาพที่ 3 การป้อนค่าฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ขั้น 4 คำนวณตามฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดที่กำหนดโดยการคลิก **Solve** ใน Dialog box ผลการคำนวณเป็นดังนี้



แฟ้ม แก้ไข มุมมอง แทรก รูปแบบ เครื่องมือ ข้อมูล FlashPaper หน้าต่าง วิธีใช้

SnagIt | Window

H35

	A	B	C	D	E	F	G
1	ลูกจ้างคนที่	กลุ่ม 1 : ไม่ประสบผลสำเร็จ		ลูกจ้างคนที่	กลุ่ม 2 : ประสบผลสำเร็จ		
2		x1-u11	x2-u12		x1-u21	x2-u22	
3	1	-2	-1.2	6	-2.6	-1	
4	2	0	-1.2	7	-1.6	0	
5	3	-1	-0.2	8	0.4	0	
6	4	3	1.8	9	2.4	-1	
7	5	0	0.8	10	1.4	2	
8							
9	constraint	-3.6	0.2				
10							
11	ตัวแปร		Subject to	constant	Minimize	32.51336921	
12	d1-=	0	8.92476E-08	0			
13	d1+=	3.10160423	8.92476E-08	0			
14	d2-=	2.24598936	5.57797E-09	0			
15	d2+=	0	8.92476E-08	0			
16	d3-=	0	-2.23119E-08	0			
17	d3+=	2.29946524	2.54993E-07	0			
18	d4-=	4.65240656	-6.37483E-08	0			
19	d4+=	0	1.59371E-08	0			
20	d5-=	0	2.54993E-07	0			
21	d5+=	1.4973262	-6.37483E-08	0			
22	e1-=	0	9.999999993	10			
23	e1+=	5.08021368					
24	e2-=	0					
25	e2+=	4.27807493					
26	e3-=	1.06951873					
27	e3+=	0					
28	e4-=	8.28877028					
29	e4+=	0					
30	e5-=	0					
31	e5+=	0					
32	b1=	-2.67379679					
33	b2=	1.87165772					
34							

ภาพที่ 4 ผลลัพธ์จากการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ผลที่ได้จากการคำนวณการโปรแกรมเชิงเส้นตามตัวแบบ LCM ขั้นที่หนึ่งด้วยโปรแกรม Microsoft Excel ได้ค่าเบี่ยงเบนของค่าสังเกตตั้งแต่ 0 ถึง 8.2888 มีผลให้ค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนระหว่างคะแนนของค่าสังเกตและค่าเฉลี่ยของคะแนนในแต่ละกลุ่มเท่ากับ 32.5134 โดยมีค่าถ่วงน้ำหนัก β_1 เท่ากับ -2.6738 และ β_2 เท่ากับ 1.8717 เพื่อการคำนวณการโปรแกรมเชิงเส้นตามตัวแบบ LCM ในขั้นที่สองด้วยโปรแกรม Microsoft Excel จึงต้องหาค่าของคะแนนการจำแนก ($S_j = \sum_{i=1}^{p} \beta_i x_{ij}$) ในแต่ละค่าสังเกตผลเป็นดังตารางที่ 2

ตารางที่ 2 คะแนนการจำแนกจากการหาค่าเหมาะสุดของตัวแบบ LCM ขั้น 1 ของลูกจ้างทั้ง 10 คน

ลูกจ้างคนที่	กลุ่ม 1: ไม่ประสบผลสำเร็จ		คะแนนการจำแนก ($S_j = \sum_{i=1}^{p} \beta_i x_{ij}$)	ลูกจ้างคนที่	กลุ่ม 2: ประสบผลสำเร็จ		คะแนนการจำแนก ($S_j = \sum_{i=1}^{p} \beta_i x_{ij}$)
	x_1	x_2			x_1	x_2	
1	1	1	-0.8021	**6**	4	1	-8.8235
2	3	1	-6.1497	**7**	5	2	-9.6257
3	2	2	-1.6043	**8**	7	2	-14.9733
4	6	4	-8.5562	**9**	9	1	-22.1925
5	3	3	-2.4064	**10**	8	4	-13.9037

เมื่อได้คะแนนการจำแนกของแต่ละค่าสังเกตดังตารางที่ 2 แล้วนั้นสามารถเขียนตัวแบบ LCM ขั้นที่สองเพื่อคำนวณหาค่าเหมาะสุดจากการโปรแกรมเชิงเส้นของลูกจ้างที่ไม่ประสบผลสำเร็จและประสบผลสำเร็จ ได้ดังนี้

ขั้นที่สอง **Minimize** $h_1 + h_2 + \ldots + h_{10}$

Subject to

$$-h_1 - c \le 0.8021$$
$$-h_2 - c \le 6.1497$$
$$-h_3 - c \le 1.6043$$
$$-h_4 - c \le 8.5562$$
$$-h_5 - c \le 2.4064$$
$$h_6 - c \ge 8.8235$$
$$h_7 - c \ge 9.6257$$
$$h_8 - c \ge 14.9733$$
$$h_9 - c \ge 22.1925$$
$$h_{10} - c \ge 13.9037$$
c ไม่จำกัดเครื่องหมาย
$$h_1, h_2, \ldots, h_{10} \ge 0$$

สำหรับการคำนวณการโปรแกรมเชิงเส้นของตัวแบบ LCM ในขั้นที่สองด้วยโปรแกรม Microsoft Excel มีขั้นตอนแสดงดังภาพที่ 5-7 ดังนี้

ขั้น 1 เตรียมข้อมูลสำหรับการคำนวณ และกำหนดตัวแปรในการคำนวณโดยให้ cell ที่ I12, I13, ..., I21 แทนค่าของตัวแปร $h_1, h_2, ..., h_{10}$ ตามลำดับ และให้ cell ที่ B22 แทนค่าของตัวแปร c

กำหนด Subject to ของตัวแบบโดยให้ cell ที่ J12, J13, ..., J21 ซึ่งมีค่าเป็น

$-h_1 - c$ แทนด้วย = -I12-I22, $-h_2 - c$ แทนด้วย = -I13-I22

$-h_3 - c$ แทนด้วย = -I14-I22, $-h_4 - c$ แทนด้วย = -I15-I22

$-h_5 - c$ แทนด้วย = -I16-I22, $h_6 - c$ แทนด้วย = I17-I22

$h_7 - c$ แทนด้วย = I18-I22, $h_8 - c$ แทนด้วย = I19-I22

$h_9 - c$ แทนด้วย = I20-I22, $h_{10} - c$ แทนด้วย = I21-I22

ตามลำดับ

กำหนด constant (ค่าคงที่) ของตัวแบบโดยให้ cell ที่ K12, K13, ..., K21 มีค่าเป็น 0.8021, 6.1497, ..., 13.9037 ตามลำดับ

กำหนดฟังก์ชันวัตถุประสงค์ของตัวแบบโดยให้ cell ที่ M11 ซึ่งมีค่าเป็น $h_1 + h_2 + ... + h_{10}$ แทนด้วย =SUM(I12:I21)

	H	I	J	K	L	M	N
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							
10							
11	ตัวแปร		Subject to	constant	Minimize	0	
12	h1=		0	0.80213907			
13	h2=		0	6.14973265			
14	h3=		0	1.60427814			
15	h4=		0	8.55614986			
16	h5=		0	2.40641721			
17	h6=		0	8.82352944			
18	h7=		0	9.62566851			
19	h8=		0	14.9732621			
20	h9=		0	22.1925134			
21	h10=		0	13.9037434			
22	c=						
23							

ภาพที่ 5 การกำหนดตัวแปร Subject to, constant และฟังก์ชันวัตถุประสงค์ของตัวแบบในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

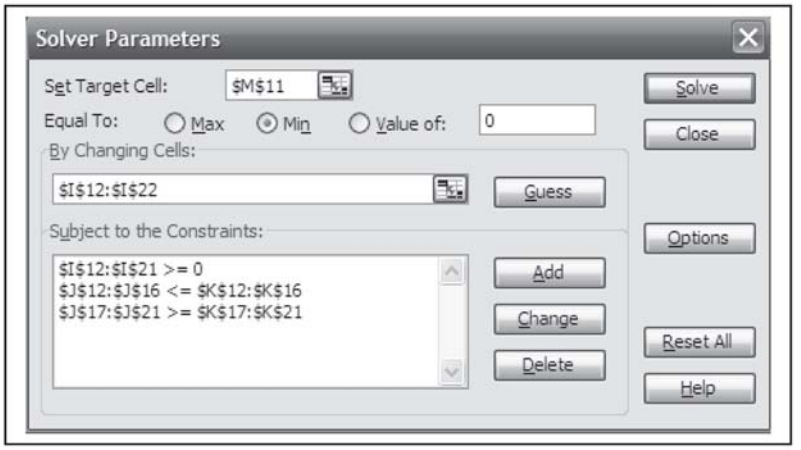
ขั้น 2 คำนวณโดยใช้ฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบ โดยคลิกที่ **เครื่องมือ (Tools)** จากแถบเมนู แล้วเลือก **Solver...** จากนั้นป้อนค่าฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบดังนี้

- **Set Target Cell:** แทนด้วย cell M11
- **Equal To:** กรณีนี้เป็นการหาค่าต่ำสุดจึงเลือก Min
- **By Changing Cells:** แทนด้วยตัวแปร คือ I12, I13, ..., I22
- **Subject to the Constraints:**

โดยเงื่อนไข $S_j - h_j - c \leq 0$ สำหรับทุกค่า j ในกลุ่มที่ 1 แทนด้วย $J$12:$J$16 =$K$12:$K$16

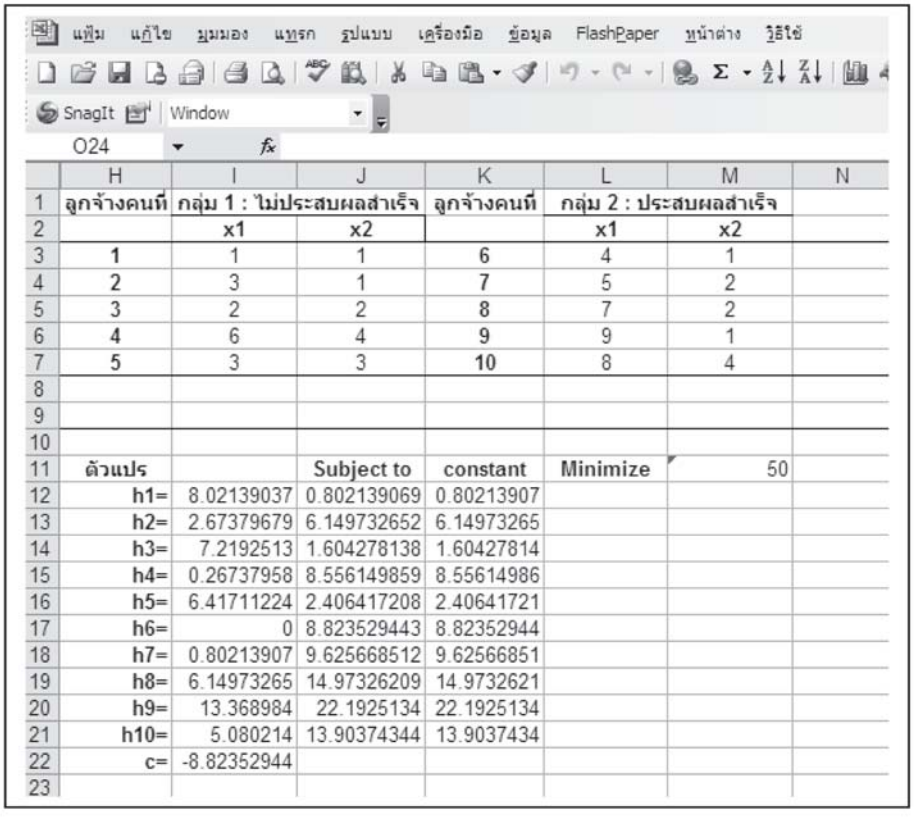
และเงื่อนไข $S_j + h_j - c \geq 0$ สำหรับทุกค่า j ในกลุ่มที่ 2 แทนด้วย $J$17:$J$21 =$K$17:$K$21

เงื่อนไข $h_j \geq 0$ แทนด้วย $I$12:$I$21 >= 0



ภาพที่ 6 การป้อนค่าฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดตามตัวแบบในการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

ขั้น 3 คำนวณตามฟังก์ชันวัตถุประสงค์ และเงื่อนไขข้อจำกัดที่กำหนด ผลการคำนวณเป็นดังนี้



แฟ้ม แก้ไข มุมมอง แทรก รูปแบบ เครื่องมือ ข้อมูล FlashPaper หน้าต่าง วิธีใช้

SnagIt Window

O24

	H	I	J	K	L	M	N
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							
10							
11	ตัวแปร		Subject to	constant	Minimize	50	
12	h1=	8.02139037	0.802139069	0.80213907			
13	h2=	2.67379679	6.149732652	6.14973265			
14	h3=	7.2192513	1.604278138	1.60427814			
15	h4=	0.26737958	8.556149859	8.55614986			
16	h5=	6.41711224	2.406417208	2.40641721			
17	h6=	0	8.823529443	8.82352944			
18	h7=	0.80213907	9.625668512	9.62566851			
19	h8=	6.14973265	14.97326209	14.9732621			
20	h9=	13.368984	22.1925134	22.1925134			
21	h10=	5.080214	13.90374344	13.9037434			
22	c=	-8.82352944					
23							

ภาพที่ 7 ผลลัพธ์จากการคำนวณการโปรแกรมเชิงเส้นด้วย Microsoft Excel

การแปลผล

ผลลัพธ์ที่ได้จากการคำนวณการโปรแกรมเชิงเส้นตามตัวแบบ LCM ด้วยโปรแกรม Microsoft Excel ในขั้นที่หนึ่งได้ค่าถ่วงน้ำหนัก β_1 เท่ากับ -2.6738 β_2 เท่ากับ 1.8717 ส่วนในขั้นที่สองได้ค่าเบี่ยงเบนของคะแนนการจำแนกในแต่ละสังเกตตั้งแต่ 0 ถึง 13.3690 (เมื่อค่าเบี่ยงเบนมีค่าสูงขึ้นการจำแนกของค่าสังเกตจะมีค่าชัดเจนสูงขึ้นตามไปด้วย) มีผลให้ค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนของคะแนนการจำแนกกับจุดตัดเท่ากับ 50 โดยมีค่าจุดตัด c เท่ากับ -8.8235 ดังนั้น ระนาบเกิน หรือฟังก์ชันเชิงเส้นของการจำแนกเป็นดังนี้

$$-2.6738x_1 + 1.8717x_2 = -8.8235 \text{ หรือ}$$
$$2.6738x_1 - 1.8717x_2 = 8.8235$$

หากเปรียบเทียบค่าถ่วงน้ำหนักและจุดตัดที่ได้ตามตัวแบบ LCM กับตัวแบบ MSD นั้น พบว่ามีค่าใกล้เคียงกันมาก นั่นคือ ค่าถ่วงน้ำหนัก β_1 เท่ากับ 2.7273 β_2 เท่ากับ -1.8182 และค่าจุดตัด c เท่ากับ 9.0909 ซึ่งระนาบเกิน หรือฟังก์ชันเชิงเส้นของการจำแนกตามตัวแบบ MSD เป็นดังนี้

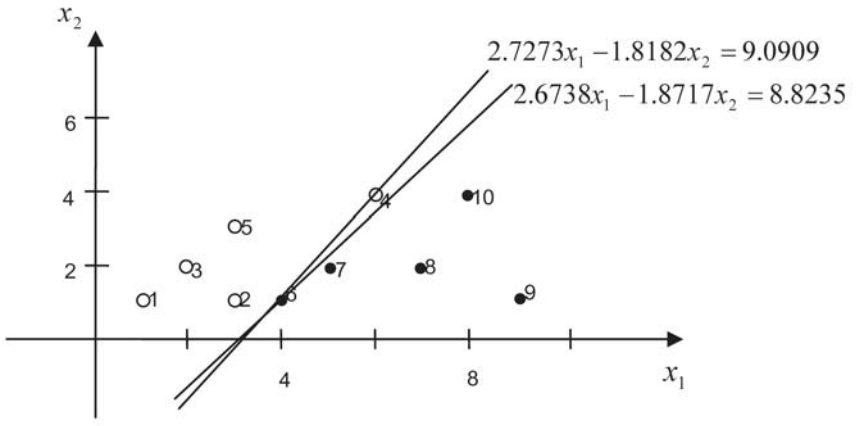
$$2.7273\beta_1 - 1.8182\beta_2 = 9.0909$$

จากผลดังกล่าวสามารถแสดงผลการจัดเข้ากลุ่มจากคุณลักษณะของลูกจ้างทั้ง 10 คน แสดงดังตารางที่ 3 ซึ่งจะเห็นได้ว่าคะแนนการจำแนกตามตัวแบบ LCM ของลูกจ้างในกลุ่มไม่ประสบผลสำเร็จมีค่าน้อยกว่า 8.8235 ทุกค่า และคะแนนการจำแนกของลูกจ้างในกลุ่มประสบผลสำเร็จมีค่ามากกว่า 8.8235 ทั้งหมด 4 ค่าโดยมีลูกจ้างคนที่ 6 ที่มีคะแนนการจำแนกเท่ากับค่าของจุดตัดจึงได้ผลการจัดเข้ากลุ่มผิดกลุ่ม ในขณะที่คะแนนการจำแนกตามตัวแบบ MSD ของลูกจ้างในกลุ่มไม่ประสบผลสำเร็จมีค่าน้อยกว่า 9.0909 ทั้งหมด 4 ค่า โดยมีลูกจ้างคนที่ 4 ที่มีคะแนนการจำแนกเท่ากับค่าของจุดตัด และคะแนนการจำแนกของลูกจ้างในกลุ่มประสบผลสำเร็จมีค่ามากกว่า 9.0909 ทั้งหมด 4 ค่าโดยมีลูกจ้างคนที่ 6 ที่มีคะแนนการจำแนกเท่ากับค่าของจุดตัดจึงได้ผลการจัดเข้ากลุ่มผิดกลุ่มเช่นเดียวกับตัวแบบ LCM

ตารางที่ 3 ผลการจัดเข้ากลุ่มจากคุณลักษณะของลูกจ้างทั้ง 10 คนในกลุ่มไม่ประสบผลสำเร็จ และประสบผลสำเร็จ

กลุ่ม	ลูกจ้างคนที่	ตัวแปร		ตัวแบบ MSD		ตัวแบบ LCM	
		x_1	x_2	คะแนนการจำแนก	ผลการจัดเข้ากลุ่ม	คะแนนการจำแนก	ผลการจัดเข้ากลุ่ม
ไม่ประสบผลสำเร็จ	1	1	1	0.9091	ถูก	0.8021	ถูก
	2	3	1	6.3636	ถูก	6.1497	ถูก
	3	2	2	1.8182	ถูก	1.6043	ถูก
	4	6	4	9.0909	ถูก	8.5562	ถูก
	5	3	3	2.7273	ถูก	2.4064	ถูก
ประสบผลสำเร็จ	6	4	1	9.0909	ผิด	8.8235	ผิด
	7	5	2	10.0000	ถูก	9.6257	ถูก
	8	7	2	15.4545	ถูก	14.9733	ถูก
	9	9	1	22.7273	ถูก	22.1925	ถูก
	10	8	4	14.5455	ถูก	13.9037	ถูก

เมื่อทำการพลอตกราฟ เพื่อแสดงฟังก์ชันจำแนกจากการคำนวณการโปรแกรมเชิงเส้นตามตัวแบบ LCM ด้วยโปรแกรม Microsoft Excel ได้ผลตามภาพที่ 8 ซึ่งจะเห็นได้ว่าเมื่อใช้ตัวแบบ MSD มีค่าสังเกตที่จัดเข้ากลุ่มถูกต้องอย่างชัดเจน 8 ค่า หากใช้ตัวแบบ LCM มีค่าสังเกตที่จัดเข้ากลุ่มถูกต้องอย่างชัดเจนถึง 9 ค่า โดยมีข้อน่าสังเกตเกี่ยวกับค่าสังเกต 1 ค่าที่อยู่บนระนาบจากตัวแบบ LCM นั่นคือ ค่าสังเกตที่ 6 ในกลุ่มของลูกจ้างที่ประสบผลสำเร็จ (กลุ่ม 2) ซึ่งจากกฎการจัดเข้ากลุ่มหาก $Sj > c$ แสดงว่าค่าสังเกตอยู่กลุ่ม 2 จึงสรุปได้ว่าค่าสังเกตที่ 6 จัดเข้ากลุ่มผิด หากพิจารณาจากตัวแบบ MSD นั้นมีค่าสังเกตที่อยู่บนระนาบถึง 2 ค่า คือ ค่าสังเกตที่ 4 ในกลุ่มลูกจ้างที่ไม่ประสบผลสำเร็จ (กลุ่ม 1) และค่าสังเกตที่ 6 ในกลุ่มของลูกจ้างที่ประสบผลสำเร็จ (กลุ่ม 2) แต่การสรุปผลให้ผลเช่นเดียวกับการคำนวณจากตัวแบบ LCM ผลการจัดเข้ากลุ่ม และการพลอตกราฟจะเห็นได้ว่าตัวแบบ LCM ให้ผลการจัดเข้ากลุ่มที่ชัดเจนกว่าตัวแบบ MSD ถึงแม้ว่าให้ผลของความถูกต้องในการจัดเข้ากลุ่มจะไม่แตกต่างกันก็ตาม



ภาพที่ 8 ฟังก์ชันการจำแนกจากคุณลักษณะของลูกจ้างทั้ง 10 คนในกลุ่มไม่ประสบผลสำเร็จ และกลุ่มประสบผลสำเร็จ

สรุป

การวิเคราะห์การจำแนกด้วยการโปรแกรมเชิงเส้นตามตัวแบบ LCM ที่เสนอโดย Lam, Choo, and Moy (1996) เป็นตัวแบบที่ซับซ้อนกว่าตัวแบบ MSD เล็กน้อย แต่ผลการคำนวณด้วยโปรแกรม Microsoft Excel ให้ผลที่ชัดเจนกว่าอย่างเห็นได้ชัด กล่าวคือ ตัวแบบ LCM นั้น ได้ทำการแก้ปัญหา trivial solution ซึ่งเป็นปัญหาหลักหรือข้อด้อยประการหนึ่งของตัวแบบ MSD ซึ่งจะส่งผลต่อการตีความในการจำแนกต่อไป ดังนั้นตัวแบบ LCM จึงน่าจะเป็นทางเลือกหนึ่งของการวิเคราะห์การจำแนกโดยการโปรแกรมเชิงเส้นได้เป็นอย่างดี

เอกสารอ้างอิง

อาฟีฟี ลาเต๊ะ และ ประสพชัย พสุนนท์. 2552. “การโปรแกรมเชิงเส้นสำหรับการวิเคราะห์การจำแนกกลุ่มโดยการคำนวณด้วยโปรแกรม Microsoft Excel.” **วารสารวิชาการ มหาวิทยาลัยหอการค้าไทย** 29, 3: 169-182.

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

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

Lam, K.F., Choo, E.U., and Moy, J.W. 1996. “Minimizing Deviations from the Group Mean: A New Linear Programming Approach for the Two-group Classification Problem.” **European Journal of Operational Research** Vol. 88: 358-367.

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

Asst.Prof. Prasopchai Pasunon received his Master of Science Degree in Applied Statistics from Silpakorn University, and Bachelor of Science in Mathematics Degree from Thaksin University. Currently, he is an Assistant Professor in 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.