การเปรียบเทียบวิธีการทางสถิติและวิธีการโปรแกรมเชิงเส้น สำหรับพยากรณ์บริษัทธุรกิจของประเทศไทยที่จะประสบปัญหาทางการเงิน
Comparison of Statistical and Linear Programming Approaches for Predicting Thai Listed Companies' Financial Problems

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

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

ดร.สุดา ตระการเถลิงศักดิ์
รองศาสตราจารย์ประจำภาควิชาสถิติ
คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร
E-mail: suda@su.ac.th

ดร.ปราณี นิลกรณ์
ผู้ช่วยศาสตราจารย์ประจำภาควิชาสถิติ
คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร
E-mail: pranee@su.ac.th

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบตัวแบบการวิเคราะห์การจำแนกกรณี 2 กลุ่มระหว่างวิธีการทางสถิติ คือ ฟังก์ชันจำแนกเชิงเส้นของฟิชเชอร์ (Fisher's Linear Discriminant Function:

FLDF) และการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis: LRA) และวิธีการโปรแกรมเชิงเส้น คือ ตัวแบบ MSD (Freed and Glover, 1986) และตัวแบบ LCM (Lam, Choo, and Moy, 1996) ในข้อมูลทุติยภูมิเกี่ยวกับอัตราส่วนทางการเงินของบริษัทธุรกิจที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยจำนวน 210 บริษัท ที่เก็บรวบรวมในช่วงปี พ.ศ. 2536-2539 แบ่งเป็นบริษัทที่ประสบปัญหาทางการเงิน 30 บริษัท และบริษัทที่ไม่ประสบปัญหาทางการเงิน 180 บริษัท ในการเปรียบเทียบประเมินจากความถูกต้องในการจัดเข้ากลุ่ม (Correct Classification) และความถูกต้องในการพยากรณ์ (Prediction Accuracy) ผลการวิจัยพบว่า ตัวแบบ LRA ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงสุด โดยที่ตัวแบบ MSD และตัวแบบ LCM ให้ผลของความถูกต้องรองลงมาตามลำดับ ในขณะที่ตัวแบบ LCM ให้ผลของความถูกต้องในการพยากรณ์สูงสุด โดยที่ตัวแบบ LRA และตัวแบบ MSD ให้ผลของความถูกต้องรองลงมาตามลำดับ

คำสำคัญ: การวิเคราะห์การจำแนก การโปรแกรมเชิงเส้น

Abstract

The objective of this research is to compare four discriminant models FLDF: Fisher's Linear Discriminant Function, LRA: Logistic Regression Analysis, MSD (Freed and Glover, 1986), and LCM (Lam, Choo, and Moy, 1996) for two-group discriminant analysis. The data were financial ratios of 210 Thai listed companies collected during 2536-2539 B.C., 30 bankrupts and 180 non-bankrupts. The performance of each model was assessed by correct classification and prediction accuracy. The result is the LRA model performed best, followed by the MSD and LCM models respectively. In terms of prediction accuracy, the LCM model performed best, followed by the LRA and MSD models respectively.

Keywords: Discriminant Analysis, Linear Programming

บทนำ

การวิเคราะห์การจำแนก (Discriminant Analysis) เป็นวิธีการทางสถิติที่ใช้สำหรับตัวแปรพหุประกอบด้วย 2 ขั้นตอนที่ต้องกระทำต่อเนื่องกัน คือ การจำแนกกลุ่มและการจัดเข้ากลุ่ม (Discrimination and Classification) โดยในขั้นตอนการจำแนกกลุ่มจะใช้ค่าสังเกตที่ทราบแล้วว่ามาจากกลุ่มใด คำนวณค่าเชิงตัวเลข ภาพ หรือสมการทางพีชคณิตให้เห็นถึงความแตกต่างระหว่างกลุ่มอย่างเด่นชัดที่สุดเท่าที่จะกระทำได้ โดยปกติแล้วจะเป็นการสร้างฟังก์ชันจำแนก (Discriminant Function) หรือผลบวกเชิงเส้น (Linear Combination) ของตัวแปรอิสระหรือตัวแปรทำนาย (Independent Variable or Predictor) เพื่ออธิบายลักษณะที่ต่างกันของตัวแปร

ตามหรือของกลุ่ม (Dependent Variable or Group) ส่วนขั้นตอนการจัดเข้ากลุ่มเป็นการจัดค่าสังเกตที่เพิ่งเข้ามาใหม่เข้ากลุ่มใดกลุ่มหนึ่ง โดยอาศัยผลลัพธ์จากฟังก์ชันจำแนกและเกณฑ์การจัดเข้ากลุ่ม (Classification Rule) การจัดเข้ากลุ่มจะทำได้ถูกต้องเพียงใดขึ้นกับฟังก์ชันจำแนกและกฎเกณฑ์การจัดเข้ากลุ่มว่าเหมาะสมหรือไม่

วิธีการวิเคราะห์การจำแนกที่ดีควรให้ฟังก์ชันจำแนกที่แสดงถึงลักษณะที่ต่างกันระหว่างตัวแปรตามหรือระหว่างกลุ่มอย่างเด่นชัด และกฎการจัดเข้ากลุ่มที่ใช้ควรมีความเหมาะสมเพื่อให้ความผิดพลาดในการจำแนกเกิดขึ้นน้อย หรืออีกนัยหนึ่ง คือ มีการจำแนกถูกสูง ในการประเมินประสิทธิภาพฟังก์ชันจำแนกนั้น ถ้านำฟังก์ชันจำแนกที่ได้ไปใช้กับตัวอย่างที่ใช้ในการสร้างฟังก์ชันจำแนก จะทำให้อัตราการจำแนกผิดต่ำกว่าที่ควรเป็น (Johnson and Wichern, 1988; Tabachnick and Fidell, 2001) จึงนิยมประเมินโดยการแบ่งตัวอย่างเป็น 2 ส่วน ตัวอย่างส่วนแรกนำมาใช้ในการสร้างฟังก์ชันจำแนก เรียกตัวอย่างส่วนนี้ว่า training sample หรือ design sample และตัวอย่างอีกส่วนหนึ่งนำมาใช้ประเมินฟังก์ชันจำแนกที่สร้างจากตัวอย่างส่วนแรก เรียกตัวอย่างส่วนหลังว่า holdout sample หรือ test sample (Berry and Linoff, 2004)

วิธีการทางสถิติแบบมีพารามิเตอร์ที่นิยมใช้ในการวิเคราะห์การจำแนก เช่น ฟังก์ชันจำแนกเชิงเส้นของฟิชเชอร์ (Fisher's Linear Discriminant Function: FLDF) และการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis: LRA) มีข้อจำกัดในเรื่องข้อสมมติเกี่ยวกับการแจกแจงของข้อมูล จึงได้มีการนำวิธีการโปรแกรมเชิงคณิตศาสตร์มาใช้ในการจำแนก วิธีการโปรแกรมเชิงคณิตศาสตร์ที่เริ่มใช้ในการวิเคราะห์การจำแนก คือ วิธีการโปรแกรมเชิงเส้น เช่น ตัวแบบ MSD (Freed and Glover, 1986) ตัวแบบ LCM (Lam, Choo, and Moy, 1996) ตัวแบบ LAD (Lee and Ord, 1990) หลังจากนั้นได้มีการพัฒนาตัวแบบทางคณิตศาสตร์มาใช้ในการวิเคราะห์การจำแนกในรูปแบบต่างๆ กัน ได้แก่ วิธีการโปรแกรมเป้าหมายเชิงเส้น การโปรแกรมจำนวนเต็มแบบผสม และการโปรแกรมไม่ใช่เชิงเส้น เช่น ตัวแบบ MIP (Stam and Joachimsthaler, 1990) ตัวแบบ Two-stage MIP (Stam and Ragsdale, 1992)

อาฟีฟี ลาเต๊ะ, สุดา ตระการเถลิงศักดิ์ และปราณี นิลกรณ์ (2548) ได้วิเคราะห์อัตราการจำแนกผิดจาก 4 ตัวแบบ คือ ตัวแบบ FLDF ตัวแบบ MSD ตัวแบบ LCM และตัวแบบ Extended DEA-DA (Sueyoshi, 2001) จากข้อมูลทางการเงินของบริษัทจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย ผลการศึกษาพบว่า ตัวแบบ LCM ให้อัตราการจำแนกผิดต่ำที่สุด รองลงมา คือ ตัวแบบ FLDF ตัวแบบ Extended DEA-DA และตัวแบบ MSD ตามลำดับ

ประสพชัย พสุนนท์ และคณะ (2550) ได้ศึกษาตัวแบบวิธีการพยากรณ์เพื่อจำแนกกลุ่มสหกรณ์การเกษตรที่มีผลการดำเนินงานกำไรและขาดทุน โดยจำแนกจากตัวแปรอัตราส่วนทางการเงินของการวิเคราะห์แบบ CAMEL (CAMEL Analysis) และใช้ข้อมูลย้อนหลัง 3 ปี ที่ผ่านการตรวจสอบบัญชีแล้ว แบ่งเป็นสหกรณ์ที่มีกำไร 218 สหกรณ์ และสหกรณ์ที่ขาดทุน 46 สหกรณ์ เป็นการเปรียบเทียบระหว่าง 2 ตัวแบบ คือ ตัวแบบ LRA และตัว

แบบ LCM ผลการวิจัยพบว่าตัวแบบ LRA สามารถพยากรณ์กลุ่มสหกรณ์ได้ถูกต้องมากกว่าตัวแบบ LCM ประมาณร้อยละ 3.42

Hua, et al. (2007) ได้บูรณาการตัวแบบ LRA ร่วมกับวิธีการ Support Vector Machine (SVM) ซึ่งเป็นตัวแบบการโปรแกรมเชิงเส้น เทียบกับวิธีการทำเหมืองข้อมูล (Data Mining) ที่ชื่อว่า Integreted Binary Discriminant Rule (IBDR) ในการจำแนกระหว่างบริษัทที่ปรึกษาทางการเงินที่ไม่มีปัญหาผลการดำเนินงานและบริษัทที่มีปัญหาผลการดำเนินงานอย่างละ 60 บริษัท ข้อมูลที่ใช้ในการวิเคราะห์มาจากตลาดหลักทรัพย์เซี่ยงไฮ้ (Shanghai Stock Exchange) โดยสร้าง training sample และ holdout sample ในการทดสอบประสิทธิภาพการจำแนก ผลลัพธ์ที่ได้พบว่าตัวแบบ LRA ร่วมกับตัวแบบ SVM มีประสิทธิภาพสูงกว่าวิธีการ IBDR ที่ระดับนัยสำคัญ 0.01

บทความนี้เสนอการศึกษาเปรียบเทียบระหว่างตัวแบบ FLDF ตัวแบบ LRA ตัวแบบ MSD ตัวแบบ LCM ซึ่งในการศึกษาเปรียบเทียบจะใช้ข้อมูลอัตราส่วนทางการเงินของบริษัทธุรกิจที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยจำนวน 210 บริษัท แบ่งเป็นบริษัทที่ประสบปัญหาทางการเงินจำนวน 30 บริษัทและบริษัทที่ไม่ประสบปัญหาทางการเงินจำนวน 180 บริษัท (สุภัทรา วงศ์สืบชาติ, 2545) โดยข้อมูลดังกล่าวเป็นข้อมูลในช่วงวิกฤติการณ์เศรษฐกิจฟองสบู่ในปี พ.ศ. 2540 ซึ่งส่งผลเสียหายต่อระบบเศรษฐกิจของประเทศอย่างรุนแรง หากได้มีการทบทวนถึงมูลเหตุของวิกฤตินั้นย่อมเป็นประโยชน์ในการจัดการเพื่อไม่ให้เกิดความเสียหายซ้ำ อย่างไรก็ตามข้อมูลดังกล่าวมีจำนวนบริษัทน้อยจึงไม่เหมาะในการนำมาแยกประเภทธุรกิจหรืออุตสาหกรรม ซึ่งหากแยกประเภทแล้วอาจจะทำให้ไม่มีสารสนเทศเพียงพอในการหาฟังก์ชันจำแนกที่เหมาะสมได้

ตัวแบบที่ใช้ในงานวิจัย

ตัวแบบที่ใช้ในการวิจัยนี้มี 4 ตัวแบบ ดังนี้

1. ตัวแบบ FLDF แนวคิดในการจำแนกของฟิชเชอร์ คือ การสร้างฟังก์ชันผลบวกเชิงเส้นของตัวแปรอิสระเพื่อทำให้อัตราส่วนระหว่างความแปรผันระหว่างกลุ่มต่อความแปรผันภายในกลุ่มมีค่าสูงสุด ซึ่งการจำแนกโดยวิธีการของฟิชเชอร์นี้จะมีความเหมาะสม ถ้าตัวแปรอิสระมีการแจกแจงร่วมกันแบบปกติ และมีเมทริกซ์ความแปรปรวนร่วมของทุกกลุ่มเท่ากัน ในกรณี 2 กลุ่ม กฎของฟิชเชอร์ในการจัดค่าสังเกต x_0 เข้ากลุ่ม 1 คือ

$$\mathbf{x}_0'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)\geq\frac{1}{2}(\bar{\mathbf{x}}_1-\bar{\mathbf{x}}_2)'\mathbf{S}^{-1}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_2)=c \quad (1)$$

และจัด x_0 อยู่ในกลุ่ม 2 ถ้าไม่สอดคล้องเงื่อนไขนี้ โดย $\bar{x}_1$, $\bar{x}_2$ เป็นเวกเตอร์ค่าเฉลี่ยของตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ และ **S** เป็นเมทริกซ์ความแปรปรวนร่วมของตัวอย่าง

2. ตัวแบบ LRA ตัวแบบการพยากรณ์การวิเคราะห์การถดถอยโลจิสติก เป็นตัวแบบที่อธิบายตัวแปรตามที่เป็นข้อมูลเชิงกลุ่มในลักษณะนามบัญญัติ (Nominal) แบบ 2 ผลลัพธ์ (Binary) และมีตัวแปรอิสระเป็นข้อมูลที่เป็นข้อมูลที่ต่อเนื่องหรือไม่ก็ได้ วิธีการนี้ไม่มีข้อสมมติเกี่ยวกับการแจกแจงของตัวแปรอิสระ การประมาณค่าพารามิเตอร์ใช้วิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) ตัวแบบการถดถอยโลจิสติกมีรูปแบบ คือ

$$\text{Prob}(Y=1|x_1, x_2, \ldots, x_p) = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)}{1+\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p)} \quad (2)$$

เมื่อ Y = 1 คือ เหตุการณ์ที่บริษัทไม่ประสบปัญหาทางการเงิน และ $x_1, x_2, \ldots, x_p$ เป็นตัวแปรพยากรณ์

3. ตัวแบบ MSD (Minimize the Sum of Deviation) เป็นตัวแบบการโปรแกรมเชิงเส้นแบบขั้นตอนเดียวที่เสนอโดย Freed and Glover (1986) ตัวแบบนี้มีประสิทธิภาพในการจำแนกทั้งในข้อมูลที่มีการแจกแจงแบบปกติหรือข้อมูลที่เบี่ยงเบนไปจากการแจกแจงแบบปกติ (Joachimsthaler and Stam, 1988) รวมทั้งในข้อมูลที่มีค่าผิดปกติปนอยู่ (Lee and Ord, 1990) และเป็นตัวแบบที่ได้รับการยอมรับว่ามีประสิทธิภาพในการพยากรณ์ค่าสังเกตที่เพิ่งเข้ามาใหม่ได้ดีที่สุดตัวแบบหนึ่ง มีรูปแบบดังนี้

ให้ β_i เป็นค่าถ่วงน้ำหนักที่ให้กับตัวแปรอิสระ i เมื่อ $i = 1,\ldots p$

x_{ij} เป็นค่าสังเกตลำดับที่ j ของตัวแปรอิสระที่ i

p เป็นจำนวนตัวแปรอิสระ

c เป็นจุดตัด (cutoff)

N เป็นค่าที่กำหนดขึ้นเพื่อให้สเกลของผลลัพธ์อยู่ในช่วงที่ต้องการ โดยกำหนดให้ N = 10

α_j เป็นตัวแปรเบี่ยงเบนของค่าสังเกตลำดับที่ j

$$\text{Minimize} \quad \sum_{j=1}^{m} \alpha_j \quad (3)$$

Subject to

$$\sum_{i=1}^{p} \beta_i x_{ij} - \alpha_j - c \le 0 \quad \text{สำหรับทุกค่า j ในกลุ่มที่ 1} \quad (4)$$

$$\sum_{i=1}^{p} \beta_i x_{ij} + \alpha_j - c \ge 0 \quad \text{สำหรับทุกค่า j ในกลุ่มที่ 2} \quad (5)$$

$$c + \sum_{i=1}^{p} \beta_i = N \quad (6)$$

β_i และ c ไม่จำกัดเครื่องหมาย

$\alpha_j \geqslant 0$

ตัวแบบ MSD เป็นตัวแบบที่มีเป้าหมายเพื่อหาค่าต่ำสุดของผลรวมของส่วนเบี่ยงเบนของค่าสังเกตกับจุดตัด ภายใต้เงื่อนไขเกณฑ์การจัดเข้ากลุ่มของกลุ่มที่ 1 และกลุ่มที่ 2 ตามเงื่อนไข (4) และ (5) โดยมีข้อสมมติว่ากลุ่ม 1 อยู่ใต้จุดตัด กลุ่ม 2 อยู่เหนือจุดตัด โดยตัวแปรเบี่ยงเบน α_j จะเป็นศูนย์เมื่อค่าสังเกตลำดับที่ j จัดถูกกลุ่ม สำหรับเงื่อนไข (6) 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 อย่างไรก็ตาม อาจจะกำหนดให้ N เป็น 1 หรือ 100 สามารถทำได้เช่นเดียวกัน 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

4. ตัวแบบ LCM เป็นตัวแบบการโปรแกรมเชิงเส้นแบบ 2 ขั้นตอนที่เสนอโดย Lam *et al.* (1996) ตัวแบบนี้อยู่บนแนวคิดพื้นฐานจากการวิเคราะห์กลุ่ม (Cluster analysis) ที่เป็นแบบ 2 ขั้น โดยหาค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนระหว่างคะแนนการจำแนกกับค่าเฉลี่ยของคะแนนการจำแนกของค่าสังเกตในแต่ละกลุ่ม ซึ่งต่างจากแนวคิดเดิมที่ใช้ค่าส่วนเบี่ยงเบนของคะแนนการจำแนกกับจุดตัด จากการศึกษายังพบว่าตัวแบบนี้มีประสิทธิภาพในกรณีข้อมูลมีการแจกแจงที่ปนเปื้อนค่าผิดปกติ (Lam, Choo, and Moy, 1996) และสามารถพยากรณ์ค่าสังเกตใหม่ได้เป็นอย่างดีทั้งในข้อมูลที่มีการแจกแจงที่เบี่ยงเบนจากการแจกแจงแบบปกติ และข้อมูลที่มีการปนเปื้อนค่าผิดปกติ (Lam, Choo, and Moy, 1996; Lam and Moy, 2003) โดยเป้าหมายของขั้นตอนแรกเป็นการหาค่าต่ำสุดของผลรวมส่วนเบี่ยงเบนระหว่างคะแนนการจำแนกกับค่าเฉลี่ยของคะแนนการจำแนกของค่าสังเกตในแต่ละกลุ่ม ภายใต้เงื่อนไข (8) และ (9) โดยตัวแปรเบี่ยงเบนทางซ้ายและทางขวาของค่าสังเกตในแต่ละกลุ่มต้องมีตัวใดตัวหนึ่งเป็นศูนย์หรือเป็นศูนย์ทั้งคู่ ส่วนเงื่อนไข (10) เป็นเงื่อนไขที่เพิ่มในตัวแบบเพื่อไม่ให้เกิดกรณี trivial solution และยังตรึงให้ความแตกต่างระหว่างค่าเฉลี่ยของคะแนนการจำแนกของ 2 กลุ่มมีค่ามากกว่าหรือเท่ากับ 1 รูปแบบการโปรแกรมเชิงเส้นของขั้นตอนแรกเป็นดังนี้

ให้ μ_{ki} เป็นค่าเฉลี่ยของตัวแปรอิสระที่ i ในกลุ่มที่ k เมื่อ k = 1,2

n_i เป็นจำนวนค่าสังเกตในกลุ่มที่ i เมื่อ i = 1,2

m เป็นจำนวนค่าสังเกตรวมทั้ง 2 กลุ่ม

d_j^-, d_j^+ เป็นตัวแปรเบี่ยงเบนทางซ้ายและทางขวาของค่าสังเกตในกลุ่มที่ 1

e_j^-, e_j^+ เป็นตัวแปรเบี่ยงเบนทางซ้ายและทางขวาของค่าสังเกตในกลุ่มที่ 2

S_j เป็นคะแนนของค่าสังเกตที่ j

h_j เป็นระยะทางระหว่างคะแนนของค่าสังเกตกับจุดตัด

phase 1 Minimize $\sum_{j=1}^{n_1}(d_j^+ + d_j^-) + \sum_{j=1}^{n_2}(e_j^+ + e_j^-)$ (7)

Subject to

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

สำหรับทุกค่า j ในกลุ่มที่ 1 (8)

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

สำหรับทุกค่า j ในกลุ่มที่ 2 (9)

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

β_i ไม่จำกัดเครื่องหมาย

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

หลังจากแก้ปัญหาในขั้นตอนแรกแล้วใช้ค่า β_i มาคำนวณคะแนนการจำแนก (Classification Score, $S_j = \sum_{i=1}^{p}\beta_i x_{ij}$) เพื่อนำค่าไปใช้ในขั้นตอนที่ 2 ต่อไป

ขั้นตอนที่ 2 มีเป้าหมายเพื่อหาค่าต่ำสุดของผลรวมของระยะห่างระหว่างคะแนนของค่าสังเกตกับค่าของจุดตัด ภายใต้เงื่อนไข (12) และ (13) โดยมีข้อสมมติว่ากลุ่ม 1 อยู่เหนือจุดตัด กลุ่ม 2 อยู่ใต้จุดตัด โดยตัวแปรเบี่ยงเบน h_j ซึ่งเป็นระยะทางระหว่างคะแนนของค่าสังเกตลำดับที่ j กับจุดตัดจะเป็นศูนย์เมื่อค่าสังเกตนั้นจัดถูกกลุ่ม รูปแบบการ

โปรแกรมเชิงเส้นในขั้นตอนที่ 2 เป็นดังนี้

phase 2 Minimize $\sum_{j=1}^{m} h_j$ (11)

Subject to

$S_j + h_j \geq c$ สำหรับทุกค่า j ในกลุ่มที่ 1 (12)

$S_j - h_j \leq c$ สำหรับทุกค่า j ในกลุ่มที่ 2 (13)

c ไม่จำกัดเครื่องหมาย

$h_j \geq 0$

หลังจากแก้ปัญหาแล้วนำค่า S_j และ c ซึ่งเป็นผลลัพธ์จากขั้นตอนที่ 1 และขั้นตอนที่ 2 ตามลำดับ มาใช้จัดกลุ่มค่าสังเกต ถ้า $S_j \geq c$ แสดงว่าค่าสังเกตอยู่กลุ่มที่ 1 และถ้า $S_j < c$ แสดงว่าค่าสังเกตอยู่กลุ่มที่ 2

วิธีดำเนินการวิจัย

การศึกษาเปรียบเทียบระหว่างตัวแบบ FLDF ตัวแบบ LRA ตัวแบบ MSD และตัวแบบ LCM ในการศึกษาเปรียบเทียบ จะใช้ข้อมูลอัตราส่วนทางการเงินของบริษัทธุรกิจที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยจำนวน 210 บริษัท แบ่งเป็นบริษัทที่ประสบปัญหาทางการเงินในช่วงวิกฤติเศรษฐกิจปี พ.ศ. 2540 จำนวน 30 บริษัทและบริษัทที่ไม่ประสบปัญหาทางการเงินจำนวน 180 บริษัท โดยพิจารณาจากตัวแปรอิสระทั้งหมด 5 ตัวแปร ซึ่งเป็นอัตราส่วนที่แสดงถึงสภาพคล่องทางการเงินและความสามารถในการทำกำไรของบริษัท ค่าอัตราส่วนที่ใช้ได้จากการหาค่าเฉลี่ยตั้งแต่ปี พ.ศ. 2536-2539 (สุภัทรา วงศ์สืบชาติ, 2545) ซึ่งประกอบด้วย

1. อัตราส่วนเงินทุนหมุนเวียนต่อสินทรัพย์ทั้งหมด (x_1)
2. อัตราส่วนกำไรสะสมต่อสินทรัพย์ทั้งหมด (x_2)
3. อัตราส่วนกำไรก่อนหักดอกเบี้ยและภาษีต่อสินทรัพย์ทั้งหมด (x_3)
4. อัตราส่วนมูลค่าหุ้นสามัญต่อหนี้สินทั้งหมด (x_4)
5. อัตราส่วนยอดขายต่อสินทรัพย์ทั้งหมด (x_5)

การประเมินฟังก์ชันจำแนกและกฎการจัดเข้ากลุ่มประเมินจากความถูกต้องในการจัดเข้ากลุ่ม (Correct Classification) และความถูกต้องในการพยากรณ์ (Prediction Accuracy) โดยความถูกต้องในการจัดเข้ากลุ่มเป็นเปอร์เซ็นต์ของค่าสังเกตที่จัดกลุ่มถูกต้องใน training sample ส่วนความถูกต้องในการพยากรณ์เป็นเปอร์เซ็นต์ของค่าสังเกตที่จัดกลุ่มถูกต้องใน holdout sample ข้อมูลในการวิเคราะห์แบ่งเป็น 2 แบบ คือ แบบแรกการใช้ข้อมูลของค่าสังเกตทั้งหมดหาฟังก์ชันจำแนกและทดสอบประสิทธิภาพของการจำแนก ส่วนแบบที่ 2 เป็นข้อมูลเกิดจากการสุ่มค่าสังเกตจำนวนหนึ่งเพื่อใช้เป็น training sample สำหรับหาฟังก์ชันจำแนก ค่าสังเกตที่เหลือใช้เป็น holdout sample เพื่อทดสอบประสิทธิภาพของการจำแนก ซึ่งจากงานวิจัยของ Sueyoshi (1999, 2001) พบว่าการแบ่งกลุ่มตัวอย่างออกเป็น training sample และ holdout sample ด้วยเปอร์เซ็นต์ที่ต่างกันไม่ส่งผลต่อความถูกต้องในการจัดเข้ากลุ่มหรือความถูกต้องในการพยากรณ์ของตัวแบบ ซึ่งแน่นอนว่าหากจำนวน training sample มีจำนวนค่าสังเกตมากจะทำให้ได้ฟังก์ชันจำแนกที่ดีส่งผลต่อการพยากรณ์ที่ดีตามไปด้วย แต่ในหลายรายงานการวิจัยที่บางครั้งอาจสร้างฟังก์ชันจำแนกจากจำนวนค่าสังเกตที่น้อย เพื่อไปพยากรณ์จำนวนค่าสังเกตที่มาก เนื่องจากต้องการทดสอบข้อมูลลักษณะดังกล่าว ดังเช่นงานวิจัยของ Bajgier

and Hill (1982) Freed and Glover (1986) และ Pavur (2002) นอกจากนี้ในการจำแนกจริงเป็นไปได้ว่าอาจมีข้อมูลที่ใช้ในการจำแนกไม่มากพอ ดังนั้นจึงได้ออกแบบการวิจัยโดยแบ่งกลุ่มตัวอย่างของ training sample และ holdout sample ต่างๆ กัน เพื่อเปรียบเทียบผลที่ได้ว่าต่างกันมากน้อยเพียงใด โดยมีกลุ่มของจำนวนค่าสังเกตใน training sample และ holdout sample เป็นดังนี้

training sample	**holdout sample**
20% (6 ค่าจากกลุ่ม 1 และ 36 ค่าจากกลุ่ม 2)	80% (24 ค่าจากกลุ่ม 1 และ 144 ค่าจากกลุ่ม 2)
40% (12 ค่าจากกลุ่ม 1 และ 72 ค่าจากกลุ่ม 2)	60% (18 ค่าจากกลุ่ม 1 และ 108 ค่าจากกลุ่ม 2)
60% (18 ค่าจากกลุ่ม 1 และ 108 ค่าจากกลุ่ม 2)	40% (12 ค่าจากกลุ่ม 1 และ 72 ค่าจากกลุ่ม 2)

แต่ละกลุ่มทำซ้ำเพียง 30 ครั้ง เนื่องจากขั้นตอนการทำซ้ำมีเทคนิคที่ค่อนข้างซับซ้อนเกี่ยวกับการเลือกข้อมูลแบบไม่ใส่คืน และโดยทั่วไปในทางสถิติการทำซ้ำที่จำนวน 30 ครั้งในข้อมูลจริงถือว่าเป็นการพอเพียงแล้วในการอนุมานลักษณะข้อมูล นอกจากนี้ ในรายงานการวิจัยของ Bajgier and Hill (1982) ได้ทำซ้ำ 10 ครั้ง และผลที่ได้ไม่แตกต่างอย่างมีนัยสำคัญกับผลการวิจัยที่ได้มีการทำซ้ำจำนวนมาก การคำนวณใช้โปรแกรม FORTRAN Power Station 4.0 โดยใช้ตัวสถิติของ IMSL LIBRARY

ผลการวิจัย

ตารางที่ 1 และ 2 แสดงค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ความเบ้และความโด่งของตัวแปรแต่ละตัว รวมทั้งความเบ้ ความโด่งของตัวแปรทั้ง 5 ตัวแปร ระยะทาง Mahalanobis ระหว่างกลุ่ม 2 กลุ่มและตัวสถิติทดสอบ Box' M จะเห็นได้ว่าค่าเฉลี่ยของตัวแปรของทั้ง 2 กลุ่มมีความแตกต่างกันอย่างเห็นได้ชัด จากค่าความเบ้และความโด่งของตัวแปรแต่ละตัว แสดงว่าข้อมูลมีลักษณะการแจกแจงที่ไม่ใช่การแจกแจงแบบปกติ และเมื่อพิจารณาจากค่าความเบ้และความโด่งของตัวแปรทั้ง 5 ตัวแปร จะเห็นว่ามีลักษณะการแจกแจงที่ไม่ใช่การแจกแจงแบบปกติของตัวแปรพหุ และเมื่อพิจารณาระยะทาง Mahalanobis ระหว่างกลุ่ม 2 กลุ่มพบว่ามีค่ามากกว่า 1.5 แสดงให้เห็นว่าค่าสังเกตของ 2 กลุ่มมีความเหลื่อมล้ำกันค่อนข้างน้อยตามเกณฑ์ของ Joachimsthaler and Stam (1988) และ Stam and Joachimsthaler (1990) และเมื่อทดสอบการเท่ากันของเมทริกซ์ความแปรปรวนร่วมของทั้งสองกลุ่มด้วยตัวสถิติทดสอบ Box' M พบว่าความแปรปรวนร่วมแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

ตารางที่ 1 ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ความเบ้และความโด่งของตัวแปรแต่ละตัว

ตัวแปร	กลุ่ม 1				กลุ่ม 2			
	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความเบ้	ความโด่ง	ค่าเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน	ความเบ้	ความโด่ง
X_1	-4.9587	33.2682	-1.299	1.282	7.9108	19.8936	0.207	0.074
X_2	-22.6436	44.6685	-1.792	2.788	12.0727	13.0894	-0.104	3.792
X_3	-1.0496	17.2317	-1.706	1.941	8.6804	5.6625	1.697	9.477
X_4	139.1947	326.5882	4.416	20.709	318.0133	498.4370	4.881	33.807
X_5	0.3136	0.4125	1.627	4.004	0.5880	0.4559	3.037	10.128

ตารางที่ 2 ความเบ้ ความโด่งของตัวแปร 5 ตัวแปร ระยะทาง Mahalanobis และตัวสถิติทดสอบ Box' M

การทดสอบด้วยวิธีการของ Mardia		
ความเบ้	Group 1 :	54.89**
	Group 2 :	34.10**
ความโด่ง	Group 1 :	81.83**
	Group 2 :	90.54**
ระยะทาง Mahalanobis		3.7160
ตัวสถิติทดสอบ Box' M: 200.141	F = 12.487*	d.f. = 15, 10259

* มีนัยสำคัญที่ระดับ .001 ** มีนัยสำคัญที่ระดับ .0001

ตารางที่ 3 แสดงความถูกต้องในการจัดเข้ากลุ่มระหว่างตัวแบบ FLDF ตัวแบบ LRA ตัวแบบ MSD และตัวแบบ LCM จากตารางพบว่า ตัวแบบ LCM ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงที่สุด รองลงมา คือ ตัวแบบ FLDF ตัวแบบ LRA และตัวแบบ MSD ตามลำดับ เมื่อพิจารณากลุ่มบริษัทที่ประสบปัญหาทางการเงิน พบว่า ตัวแบบ FLDF ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงกว่าตัวแบบอื่น แต่ในกลุ่มบริษัทที่ไม่ประสบปัญหาทางการเงินตัวแบบ LRA ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงกว่าตัวแบบอื่น และเมื่อพิจารณาค่าสังเกตที่จัดผิดกลุ่มพบว่าส่วนใหญ่เป็นค่าสังเกตเดียวกันทั้ง 4 ตัวแบบ สำหรับตัวแบบ FLDF นั้นถึงแม้ข้อมูลนี้ไม่มีลักษณะที่สอดคล้องกับข้อสมมติเบื้องต้นแต่ตัวแบบ FLDF ยังคงสามารถจัดกลุ่มได้ถูกต้องโดยเฉพาะกลุ่มบริษัทที่ประสบปัญหาทางการเงิน

ตารางที่ 3 ความถูกต้องในการจัดเข้ากลุ่มระหว่างตัวแบบ FLDF ตัวแบบ LRA ตัวแบบ MSD และตัวแบบ LCM

ตัวแบบ	จำนวนบริษัทที่จัดเข้ากลุ่มได้ถูกต้อง		ความถูกต้องในการจัดเข้ากลุ่มโดยรวม (เปอร์เซ็นต์)
	กลุ่ม 1 (30 บริษัทที่ประสบปัญหา)	**กลุ่ม 2** (180 บริษัทที่ไม่ประสบปัญหา)	
FLDF	**28**	163	191 (90.95%)
LRA	9	**178**	187 (89.01%)
MSD	17	163	180 (85.71%)
LCM	22	172	**194 (92.38%)**

ตารางที่ 4 แสดงการเปรียบเทียบความถูกต้องในการจัดเข้ากลุ่มและความถูกต้องในการพยากรณ์ระหว่างตัวแบบ FLDF ตัวแบบ MSD ตัวแบบ LRA และตัวแบบ LCM จากตารางพบว่าในกรณี training sample ตัวแบบ LRA ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงกว่าตัวแบบอื่น ส่วนการประเมินจากความถูกต้องในการพยากรณ์ พบว่าตัวแบบ LCM ให้ผลของความถูกต้องในการพยากรณ์สูงกว่าตัวแบบอื่นๆ ในกรณีจำนวนค่าสังเกตน้อยและจำนวนค่าสังเกตมาก ในขณะที่ตัวแบบ LRA ให้ผลของความถูกต้องในการพยากรณ์สูงในกรณีจำนวนค่าสังเกตมีขนาดปานกลาง เมื่อพิจารณาค่าเฉลี่ยของความถูกต้องในการจัดเข้ากลุ่ม พบว่าตัวแบบ LRA ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงกว่าตัวแบบอื่นๆ และเมื่อพิจารณาค่าเฉลี่ยของความถูกต้องในการพยากรณ์ พบว่าตัวแบบ LCM ให้ผลของความถูกต้องในการพยากรณ์สูงกว่าตัวแบบอื่นๆ และสูงกว่าตัวแบบ FLDF อย่างเห็นได้ชัด

ตารางที่ 4 ความถูกต้องในการจัดเข้ากลุ่มและความถูกต้องในการพยากรณ์ระหว่างตัวแบบ FLDF ตัวแบบ MSD ตัวแบบ LRA และตัวแบบ LCM*

ลักษณะข้อมูล		ตัวแบบ			
		FLDF	**LRA**	**MSD**	**LCM**
training sample [20%]	ความถูกต้องในการจัดเข้ากลุ่ม	91.67%	**91.85%**	91.03%	88.17%
		(2.50)	(3.31)	(2.03)	(1.69)
holdout sample [80%]	ความถูกต้องในการพยากรณ์	65.34%	85.72%	85.00%	**86.83%**
		(22.07)	(4.99)	(8.48)	(1.80)
training sample [40%]	ความถูกต้องในการจัดเข้ากลุ่ม	89.13%	**94.58%**	91.79%	91.39%
		(2.54)	(3.78)	(3.39)	(1.24)
holdout sample [60%]	ความถูกต้องในการพยากรณ์	70.61%	**91.63%**	88.52%	90.00%
		(14.69)	(3.13)	(3.93)	(3.30)
training sample [60%]	ความถูกต้องในการจัดเข้ากลุ่ม	89.58%	**92.78%**	89.63%	89.47%
		(2.47)	(3.00)	(3.82)	(2.65)
holdout sample [40%]	ความถูกต้องในการพยากรณ์	80.79%	87.55%	86.15%	**88.93%**
		(6.89)	(6.44)	(3.44)	(2.15)
ค่าเฉลี่ยของความถูกต้องในการจัดเข้ากลุ่ม		89.78%	**93.23**	90.58%	89.89%
ค่าเฉลี่ยของความถูกต้องในการพยากรณ์		70.53%	88.10	86.43%	**88.35%**

*ตัวเลขในวงเล็บเป็นค่าเบี่ยงเบนมาตรฐาน

สรุปผลและข้อเสนอแนะ

การศึกษาเปรียบเทียบตัวแบบการวิเคราะห์การจำแนกกรณี 2 กลุ่ม ระหว่างตัวแบบ FLDF ตัวแบบ LRA ตัวแบบ MSD และตัวแบบ LCM โดยเปรียบเทียบในข้อมูลอัตราส่วนทางการเงินของบริษัทธุรกิจที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยจำนวน 210 บริษัท แบ่งเป็นบริษัทที่ประสบปัญหาทางการเงิน 30 บริษัท และบริษัทที่ไม่ประสบปัญหาทางการเงิน 180 บริษัท ผลการวิจัยพบว่า กรณีการจัดเข้ากลุ่มบริษัทที่ไม่ประสบปัญหาทางการเงิน ตัวแบบ LRA ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงที่สุด รองลงมา คือ ตัวแบบ LCM ส่วนตัวแบบ FLDF และตัวแบบ MSD ให้ผลลัพธ์ที่เท่ากัน กรณีการจัดเข้ากลุ่มบริษัทที่ประสบปัญหาทางการเงิน ตัวแบบ FLDF ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงที่สุด รองลงมา คือ ตัวแบบ LCM ตัวแบบ MSD และตัวแบบ LRA ตามลำดับ และกรณีการจัดเข้ากลุ่มบริษัทโดยรวม ตัวแบบ LCM ให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงที่สุด รองลงมา คือ ตัวแบบ FLDF ตัวแบบ LRA และตัวแบบ MSD ตามลำดับ

สำหรับค่าเฉลี่ยของความถูกต้องในการจัดเข้ากลุ่ม พบว่าตัวแบบ LRA ให้ผลของความถูกต้องสูงที่สุด รองลงมา คือ ตัวแบบ MSD ตัวแบบ LCM และตัวแบบ FLDF ตามลำดับ และค่าเฉลี่ยของความถูกต้องในการพยากรณ์พบว่า ตัวแบบ LCM

ให้ผลของความถูกต้องสูงที่สุด รองลงมา คือ ตัวแบบ LRA ตัวแบบ MSD และตัวแบบ FLDF ตามลำดับ จากผลของความถูกต้องในการจัดเข้ากลุ่ม พบว่าตัวแบบ FLDF ให้ผลของความถูกต้องสูงเมื่อจำนวนค่าสังเกตน้อย ซึ่งสอดคล้องกับข้อมูลในงานวิจัยนี้ที่มีจำนวนบริษัทที่ประสบปัญหาทางการเงินน้อย และสอดคล้องกับงานวิจัยของ พนิดา คงแจ่ม, กมลชนก พานิชการ และยุพิน ศุพุทธมงคล (2549) และ ภาณุพงศ์ พนมวัน และคณะ (2549) อีกด้วย ในขณะที่ตัวแบบ MSD ให้ผลของความถูกต้องสูงขึ้นเมื่อจำนวนค่าสังเกตเพิ่มขึ้น และจากผลของความถูกต้องในการพยากรณ์ พบว่าตัวแบบ LRA ให้ผลของความถูกต้องสูงสุดเมื่อจำนวนค่าสังเกตมีขนาดปานกลาง แต่เมื่อจำนวนค่าสังเกตน้อยและมาก ตัวแบบ LCM ให้ผลของความถูกต้องสูงสุด

สำหรับวิธีการทางสถิติทั้ง 2 วิธี ให้ผลของความถูกต้องในการจัดเข้ากลุ่มไม่แตกต่างกันมากนักเมื่อประเมินจากข้อมูลบริษัททั้ง 210 บริษัท แต่เมื่อมีการสุ่มค่าสังเกตให้เป็น training sample และ holdout sample เห็นได้ชัดว่าตัวแบบ LRA ให้ผลของความถูกต้องสูงกว่าตัวแบบ FLDF ทั้งความถูกต้องในการจัดเข้ากลุ่มและความถูกต้องในการพยากรณ์ ส่วนตัวแบบการโปรแกรมเชิงเส้นทั้งแบบขั้นเดียวและแบบ 2 ขั้นให้ผลของความถูกต้องในการจัดเข้ากลุ่มและความถูกต้องในการพยากรณ์ที่สูงกว่าตัวแบบ FLDF อย่างเห็นได้ชัด โดยตัวแบบ MSD มีแนวโน้มที่จะให้ผลของความถูกต้องในการจัดเข้ากลุ่มสูงขึ้น ในขณะที่ตัวแบบ LCM ให้ผลของความถูกต้องในการพยากรณ์ที่สูงขึ้น ซึ่งจากงานวิจัยจะเห็นได้ว่าตัวแบบการโปรแกรมเชิงเส้นเป็นตัวแบบที่ให้ผลที่ดีทั้งในเทอมของความถูกต้องในการจัดเข้ากลุ่มและความถูกต้องในการพยากรณ์ จึงน่าจะเป็นอีกทางเลือกหนึ่งสำหรับการวิเคราะห์การจำแนกได้เป็นอย่างดี

บรรณานุกรม

ประสพชัย พสุนนท์ และคณะ. 2550. “การเปรียบเทียบวิธีการพยากรณ์กลุ่มกำไรขาดทุนของสหกรณ์การเกษตรจากการจำแนกด้วยวิธีการทางสถิติและวิธีการโปรแกรมเชิงเส้น.” ใน **เอกสารประกอบการประชุมวิชาการสถิติและสถิติประยุกต์ประจำปี 2550,** หน้า 103-107. นครปฐม: มหาวิทยาลัยศิลปากร.

พนิดา คงแจ่ม, กมลชนก พานิชการ และยุพิน ศุพุทธมงคล. 2549. “การศึกษาเปรียบเทียบตัวแบบในการวินิจฉัยโรคเลปโตสไปโรสีส.” ใน **เอกสารประกอบการประชุมวิชาการสถิติและสถิติประยุกต์ประจำปี 2549,** หน้า C124-C127. ปทุมธานี: มหาวิทยาลัยธรรมศาสตร์.

ภาณุพงศ์ พนมวัน และคณะ. 2549. “การจำแนกสมัยลูกปัดแก้วที่ได้จากการขุดค้นแหล่งโบราณคดีพรหมทินใต้ ลพบุรี โดยใช้เทคนิคการวิเคราะห์กลุ่มและการถดถอยโลจิสติก”. ใน **เอกสารประกอบการประชุมวิชาการสถิติและสถิติประยุกต์ประจำปี 2549,** หน้า C128-C131. ปทุมธานี: มหาวิทยาลัยธรรมศาสตร์.

สุภัทรา วงศ์สืบชาติ. 2545. “การพยากรณ์ความเป็นไปได้ของบริษัทธุรกิจที่จดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยจะประสบปัญหาทางการเงิน.” วิทยานิพนธ์ปริญญาบริหารธุรกิจมหาบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยธุรกิจบัณฑิตย์.

อาฟีฟี ลาเต๊ะ, สุดา ตระการเถลิงศักดิ์ และปราณี นิลกรณ์. 2548. "การเปรียบเทียบวิธีการทางสถิติและวิธีการโปรแกรมเชิงเส้นสำหรับการวิเคราะห์การจำแนกกรณี 2 กลุ่ม." ใน **เอกสารประกอบการประชุมวิชาการการวิจัยดำเนินงานประจำปี 2548,** หน้า 77-93. กรุงเทพมหานคร: สถาบันบัณฑิตพัฒนบริหารศาสตร์.

Bajgeir, S.M., and Hill, A.V. 1982. "An Experimental Comparison of Statistical and Linear Programming Approaches to the Discriminant Problem." **Decision Sciences** 13: 604-618.

Berry, Michael J.A., and Linoff, Gordon S. 2004. **Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management.** 2nd ed. Indianapolis: Wiley.

Freed, N., and Glover, F. 1986. "Evaluating Alternative Linear Programming Models to Solve the Two-Group Discriminant Problem." **Decision Sciences** 17: 151-162.

Hua, Z., et al. 2007. "Predicting Corporate Financial Distress Based on Integration of Support Vector Machine and Logistic Regression." **Expert Systems with Applications** 33: 434-440.

Joachimsthaler, E.A., and Stam, A. 1988. "Four Approaches to the Classification Problem in Discriminant Analysis : An Experimental Study." **Decision Sciences** 19: 322-333.

Johnson, Richard A., and Wichern, Dean W. 1988. **Applied Multivariate Statistical Analysis.** 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

Lam, K.F., and Moy, J.W. 2003. "A Simple Weighting Scheme for Classification in Two-Group Discriminant Problems." **Computers & Operations Research** 30: 155-164.

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** 88: 358-367.

Lee, C., and Ord, J.K. 1990. "Discriminant Analysis Using Least Absolute Deviations." **Decision Sciences** 21: 86-96.

Pavur, R. 2002. "A Comparative Study of the Effect of the Position of Outliers on Classical and Nontraditional Approaches to the Two-Group Classification Problem." **European Journal of Operational Research** 136: 603-615.

Stam, A., and Joachimsthaler, E.A. 1990. "A Comparison of a Robust Mixed-Integer Approach to Existing Methods for Establishing Classification Rules for the Discriminant Problem." **European Journal of Operational Research** 46: 113-122.

Stam, A., and Ragsdale, C.L. 1992. "On the Classification Gap in Mathematical Programming-Based Approaches to the Discriminant Analysis." **Naval Research Logistics** 39: 545-559.

Sueyoshi, T. 1999. "DEA-Discriminant Analysis in the Review of Goal Programming." **European Journal of Operational Research** 115: 564-582.

_____. 2001. "Extended DEA-Discriminant Analysis." **European Journal of Operational Research** 131: 324-351.

Tabachnick, Barbara G., and Fidell, Linda S. 2001. **Using Multivariate Statistics.** 4th ed. Boston: Allyn and Bacon.

Mr. Afifi Lateh received his Master of Science in Applied Statistics from Silpakorn University and his 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. Prof. Prasopchai Pasunon received his Master of Science in Applied Statistics from Silpakorn University and his Bachelor of Science in Mathematics from Thaksin University. He is currently an Assistant Professor at the Department of General Business Management, Faculty of Management Science, Silpakorn University. His research focuses on Outliers Detection, Data Envelopment Analysis, Multivariate Analysis and Analysis for Business Data.

Assoc. Prof. Dr. Suda Trakarnthalerngsak received her Doctoral Degree in Applied Mathematics from the University of Queensland, Australia, and her Master's Degree in Operations Research from the National Institute of Development Administration, Thailand. She is currently an Associate Professor at the Department of Statistics, Faculty of Science, Silpakorn University, Thailand.

Asst. Prof. Dr. Pranee Nilkorn received her Doctoral Degree in Statistics from Carnegie-Mellon University, U.S.A. and her Master's Degree in Statistics from the University of Michigan, U.S.A. She is currently an Assistant Professor at the Department of Statistics, Faculty of Science, Silpakorn University, Thailand.