วารสารวิชาการ **มหาวิทยาลัยหอการค้าไทย**
ปีที่ 28 ฉบับที่ 1 เดือนมกราคม - มีนาคม 2551

การเปรียบเทียบประสิทธิภาพการทำนายการเป็นสมาชิกกลุ่มโดยใช้วิธี OLS Regression วิธี Logistic Regression และวิธี Discriminant Analysis ในกรณีที่ตัวแปรตอบสนองมี 2 ค่า

Comparison of Efficiency of OLS Regression, Logistic Regression and Discriminant Analysis for Classification Binary Outcomes

ดวงพร ทัชชะวณิช
ผู้ช่วยศาสตราจารย์ประจำสาขาสถิติประยุกต์
มหาวิทยาลัยหอการค้าไทย
E-mail: doungporn__hat@utcc.ac.th

บทคัดย่อ

วิธีทางสถิติที่นิยมใช้ในการทำนายการเป็นสมาชิกกลุ่มที่ถูกต้องในกรณีที่ตัวแปรตอบสนองมี 2 ค่า คือ วิธี Discriminant Analysis วิธี Logistic Regression และวิธี OLS Regression ซึ่งข้อสมมติในการใช้วิธี Discriminant Analysis และวิธี OLS Regression จะคล้ายคลึงกัน ในขณะที่วิธี Logistic Regression ไม่มีข้อสมมติเกี่ยวกับการแจกแจงปกติ ความสัมพันธ์เชิงเส้นและความแปรปรวนที่เท่ากันของตัวแปรที่ใช้ทำนาย การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพในการทำนายการเป็นสมาชิกกลุ่มที่ถูกต้องด้วยวิธีทั้งสาม เพื่อหาแนวทางในการเลือกใช้วิธีที่เหมาะสม จากการศึกษาพบว่าค่าร้อยละของผลการทำนายที่ถูกต้องซึ่งได้จากวิธีทั้งสาม มีความสัมพันธ์เชิงบวกกับค่าสัมประสิทธิ์การตัดสินใจ ถึงแม้จะไม่ได้มีการตรวจสอบการแจกแจงปกติของตัวแปรที่ใช้ทำนายก็ตาม วิธี Discriminant Analysis จะให้ค่าเฉลี่ยของผลการทำนายที่ถูกต้องมากที่สุด ในทุกระดับ

ของค่า VIF และค่าร้อยละของตัวแปรเชิงคุณภาพ แต่สำหรับกรณีที่ข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจมากกว่า 0.5 วิธีทั้งสามให้ค่าเฉลี่ยของผลการทำนายที่ถูกต้องไม่แตกต่างกัน

คำสำคัญ: OLS Regression, Logistic Regression, Discriminant Analysis ตัวแปรตอบสนอง ตัวแปรเชิงคุณภาพ

Abstract

The most widely used statistical methods for analyzing categorical outcome variables are Discriminant Analysis and Logistic Regression. If a dependent variable is a binary outcome, an analyst can choose among Discriminant Analysis, Logistic and OLS Regression. The statistical assumptions required for Discriminant Analysis are essentially the same as for OLS Regression. Logistic Regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group. The purpose of this study was to compare the predication accuracy of the three statistical methods mentioned above in order to highlight the most appropriate method. The results showed that the percentage of predicted value resulting from these three methods was highly correlated to the coefficient of determination. However, the normal distribution of the predicted variables was not examined. For every level of VIF and the percentage of qualitative variables, the Discriminant Analysis method gave the most accurate result. Nevertheless, for data having coefficient determination of more than 0.5, all methods gave the same performance in the average of predicted value.

Keywords: OLS Regression, Logistic Regression, Discriminant Analysis, Response Variable, Categorical Variable

บทนำ

โดยทั่วไปในการศึกษาความสัมพันธ์โดยใช้การวิเคราะห์การถดถอย ตัวแปรตอบสนองต้องเป็นตัวแปรเชิงปริมาณที่มีค่าต่อเนื่อง ส่วนตัวแปรที่ใช้ทำนายอาจเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพ ในกรณีที่ผู้วิจัยต้องการศึกษาปัจจัยที่มีผลกระทบกับตัวแปรตอบสนองที่มี 2 ค่า ซึ่งพบมากในงานวิจัยทางการแพทย์และทางสังคมศาสตร์ โดยสามารถเลือกใช้ วิธี Discriminant Analysis วิธี Ordinary Least Square (OLS) Regression วิธี Logistic Regression หรือวิธี Probit Regression

จากการศึกษาเปรียบเทียบวิธี Discriminant Analysis และวิธี Logistic Regression พบว่าวิธีทั้งสองสามารถใช้ประมาณค่าความน่าจะเป็นของตัวแปรตอบสนองและสร้างตัวแบบที่ใช้ในการจำแนกกลุ่มได้ แม้ว่าวิธีทั้งสองนี้มีแนวคิดที่แตกต่างกัน ซึ่งตัวแปรที่ใช้ทำนายในวิธี Logistic Regression ไม่จำเป็นต้องมีการแจกแจงปกติ ส่วนวิธี Discriminant Analysis ได้ถูกพัฒนาขึ้นสำหรับตัวแปรที่ใช้ทำนายที่มีการแจกแจงปกติ จึงเป็นที่คาดหวังว่าเมื่อตัวแปรที่ใช้ทำนายมีการแจกแจงปกติ วิธีนี้จะให้ผลการทำนายที่ถูกต้องมากกว่า (Montgomer, White, and Martin, 1987: 495-498) แต่อย่างไรก็ตามในทางปฏิบัติ ข้อสมมตินี้อาจฝ่าฝืนได้และยังพบว่าเมื่อใช้วิธี Discriminant Analysis จะให้เครื่องหมายของค่าสัมประสิทธิ์ที่ถูกต้องเสมอ (Pampel, 2000: 30) และจากการเปรียบเทียบความสามารถในการทำนายการเป็นสมาชิกกลุ่มด้วยวิธี Discriminant Analysis และวิธี Logistic Regression โดยตัวแปรที่ใช้ทำนายมีการแจกแจงปกติพบว่า วิธี Logistic Regression ให้ค่าประมาณที่ดีกว่าเล็กน้อย (Kleinbaum, Kupper and Morgenstern, 1982: 461-470)

ดังนั้นจึงเป็นที่น่าสนใจว่าหากไม่ได้มีการตรวจสอบว่าตัวแปรที่ใช้ทำนายมีการแจกแจงปกติหรือไม่ วิธีใดจะให้ผลการทำนายการเป็นสมาชิกกลุ่มที่ถูกต้องมากที่สุด และปัจจัยใดที่มีผลต่อค่าร้อยละของผลการทำนายที่ถูกต้องที่ได้จากวิธีทั้งสาม

วัตถุประสงค์การวิจัย

เพื่อเปรียบเทียบประสิทธิภาพการทำนายการเป็นสมาชิกกลุ่มโดยใช้วิธี OLS Regression วิธี Logistic Regression และวิธี Discriminant Analysis ในกรณีที่ตัวแปรตอบสนองมี 2 ค่า

ทฤษฎีที่เกี่ยวข้อง

โดยทั่วไปในการทำนายการเป็นสมาชิกกลุ่มของตัวแปรตอบสนองที่มี 2 ค่า ผู้วิจัยอาจใช้วิธี OLS Regression ในการสร้างตัวแบบความน่าจะเป็นเชิงเส้น (Menard, 1995: 98) ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรตอบสนองและตัวแปรที่ใช้ทำนาย โดยตัวแปรตอบสนองเป็นผลรวมเชิงเส้นของตัวแปรที่ใช้ทำนาย $Y_i = \beta_0 + \beta_1 X_{il} + ... + \beta_k X_{ik} + \varepsilon_i$ ซึ่ง β_i เป็นค่าสัมประสิทธิ์การถดถอย X_i เป็นตัวแปรที่ใช้ทำนาย และ ε_i เป็นค่าความคลาดเคลื่อนในการประมาณค่าสัมประสิทธิ์การถดถอยซึ่งอธิบายการเปลี่ยนแปลงของค่าเฉลี่ยของตัวแปรตอบสนองเมื่อตัวแปรที่ใช้ทำนายตัวที่สนใจเพิ่มขึ้น 1 หน่วย ในขณะที่ตัวแปรที่ใช้ทำนายตัวอื่นๆ มีค่าคงที่ โดยมีข้อสมมติว่าความคลาดเคลื่อนมีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับศูนย์และมีค่าความแปรปรวนคงที่ ค่าประมาณตัวแปรตอบสนองจากตัวอย่างสุ่มเขียน

เป็นสมการได้ดังนี้ $\hat{y}_i = b_0 + b_1x_{il} + ...+ b_kX_{ik}$ โดย $\hat{y}_i$ เป็นค่าประมาณความน่าจะเป็นเมื่อ $y_i = 1$ โดยทั่วไปความน่าจะเป็นที่ประมาณได้ควรมีค่าอยู่ในช่วง (0,1) แต่ในการประมาณค่าความน่าจะเป็นด้วยวิธีนี้อาจจะให้ค่าประมาณที่อยู่นอกช่วง (0,1) ได้ ซึ่งอาจเป็นเพราะความสัมพันธ์ระหว่างตัวแปรตอบสนองและตัวแปรที่ใช้ในการทำนายไม่เป็นเส้นตรง (Hocking, 1985: 127)

Logistic Regression เป็นหนึ่งในวิธีทางสถิติหลายๆ วิธีที่มีการนำมาใช้ในงานวิจัยทางการแพทย์และงานวิจัยทางสังคมศาสตร์ค่อนข้างมาก โดยใช้ในการประมาณค่าความน่าจะเป็นของตัวแปรตอบสนอง (King, and Ryan, 2002: 163-170) ให้ y_i แทนค่าของตัวแปรตอบสนองของหน่วยตัวอย่างที่ i โดยมีค่าเท่ากับ 1 ถ้าได้ผลลัพธ์ที่สนใจและมีค่าเท่ากับ 0 ถ้าไม่ได้ผลลัพธ์ที่สนใจ ให้ $\pi(x) = E(y|x)$ เป็นค่าเฉลี่ยแบบมีเงื่อนไขของ y เมื่อกำหนดค่า x ซึ่งค่าลอจิกของตัวแบบ Logistic Regression สามารถเขียนได้ดังนี้ $g(x) = \ln\frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1x_1 +...+\beta_k x_k$ โดย $\pi(x) = \frac{\exp[g(x)]}{1 + \exp[g(x)]}$ และ $\beta_1,\beta_2,...,\beta_k$ เป็นค่าสัมประสิทธิ์การถดถอยซึ่งประมาณได้จากวิธี Maximum Likelihood ค่าประมาณแต่ละตัวจะอธิบายการเปลี่ยนแปลงในรูปของค่าลอจิกเมื่อตัวแปรที่ใช้ทำนายตัวใดตัวหนึ่งมีค่าเพิ่มขึ้น 1 หน่วยโดยตัวแปรอื่นๆ มีค่าคงที่ ตัวแปรตอบสนองมีการแจกแจง Bernoulli ความน่าจะเป็นของความสำเร็จประมาณได้จากตัวแบบ ค่าลอจิกของแต่ละหน่วยตัวอย่างจะเป็นผลรวมเชิงเส้นของตัวแปรที่ใช้ทำนายซึ่งได้จากการแทนค่าตัวแปรที่ใช้ทำนายในสมการ $g(x_i) = b_0 + b_1x_{il} +... + b_kx_{ik}$ ค่าลอจิกที่คำนวณได้จะอยู่ในช่วง $(-\infty,\infty)$ ซึ่งจะทำให้ได้ค่าความน่าจะเป็นอยู่ในช่วง (0,1) การทดสอบค่าสัมประสิทธิ์การถดถอยด้วยวิธี Logistic Regression จะใช้ค่าสถิติทดสอบ Wald χ^2 (Menard, 1995: 39) โดยค่าสถิติที่ได้จะขึ้นกับการเปลี่ยนแปลงของ Likelihood Function เมื่อเพิ่มตัวแปรที่ใช้ทำนายเข้าไปในตัวแบบ การใช้ค่าสถิติทดสอบ Wald χ^2 มีกฎเกณฑ์เดียวกับการใช้ค่าสถิติทีและค่าสถิติเอฟจากวิธี OLS Regression ซึ่งสามารถใช้ค่าสถิติจาก Likelihood Function ในการทดสอบความเหมาะสมของตัวแบบเช่นเดียวกัน (Cox and Snell, 1989: 71) สำหรับข้อมูลบางชุดมีจำนวนตัวแปรที่ใช้ในทำนายมากเกินไปเมื่อเทียบกับจำนวนข้อมูลอาจให้ค่าประมาณพารามิเตอร์และค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนมีค่ามากจนทำให้ผลการทดสอบโดยรวมและผลการทดสอบค่าพารามิเตอร์แต่ละตัวไม่สอดคล้องกัน กล่าวคือ จากผลการทดสอบโดยรวมอาจให้ผลสรุปว่ามีตัวแปรที่ใช้ทำนายอย่างน้อย 1 ตัวที่สามารถอธิบายตัวแปรตอบสนองได้ แต่เมื่อทดสอบค่าพารามิเตอร์แต่ละตัวก็พบว่าไม่มีตัวแปรใดเลยที่สามารถอธิบายตัวแปรตอบสนองได้ ผู้วิจัยอาจจะแก้ปัญหาโดยการเพิ่มจำนวนข้อมูลหรือตัดตัวแปรที่ใช้ในการทำนายบางตัว และหากตัวแปรที่ใช้ทำนายมีความสัมพันธ์กันมากอาจทำให้เกิดความคลาดเคลื่อนได้

กรณีที่ข้อมูลในแต่ละกลุ่มมีจำนวนน้อยอาจจะทำให้ผลการทดสอบจากวิธี Hosmer-Lemeshow ไม่ถูกต้อง Pigeon and Heyse (1999: 847-853) ได้เสนอให้ใช้วิธีทดสอบความเหมาะสมของตัวแบบซึ่งให้ผลการทดสอบที่ถูกต้องกว่า โดยการลดค่า Likelihood ของความถี่ของค่าคาดหวังที่มีค่าน้อย

ในทางปฏิบัติเมื่อใช้วิธี Logistic Regression จำเป็นต้องใช้จำนวนตัวอย่างมากกว่าวิธี OLS Regression เนื่องจากการทดสอบค่าสัมประสิทธิ์การถดถอยที่ได้จากวิธี Maximum Likelihood อาจจะให้ผลการทดสอบผิดพลาดได้หากจำนวนตัวอย่างน้อยกว่า 100 (Pampel, 2000:30) เมื่อข้อมูลที่ต้องการวิเคราะห์มีตัวแปรที่ใช้ทำนายมากขึ้นจำเป็นต้องใช้จำนวนตัวอย่างมากขึ้น และอย่างน้อยที่สุดควรจะใช้จำนวนตัวอย่าง 50 รายต่อตัวแปรที่ใช้ทำนาย 1 ตัว (Wright, 1995: 221) โดยทั่วไปจำนวนตัวอย่างที่เหมาะสมจะขึ้นกับระดับความคลาดเคลื่อนชนิดที่ 1 และชนิดที่ 2 ที่ยอมรับได้ ระดับความสัมพันธ์โดยเฉลี่ยระหว่างตัวแปรที่ใช้ทำนายและตัวแปรตอบสนอง ความเชื่อถือได้ของการวัดและการแจกแจงความถี่ของตัวแปรตอบสนอง ในการใช้วิธี Logistic Regression นั้นหากมีจำนวนข้อมูลในแต่ละกลุ่มไม่เท่ากันก็จำเป็นต้องใช้จำนวนตัวอย่างมากขึ้น โดยตัวแปรที่ใช้ทำนายอาจเป็นตัวแปรชนิดใดก็ได้เช่นเดียวกับวิธี OLS Regression

วิธี Discriminant Analysis เป็นวิธีที่ได้นำหลักการของวิธีวิเคราะห์การถดถอยและวิธีวิเคราะห์ความแปรปรวนมาใช้หาสมการเชิงเส้น ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรตอบสนองและตัวแปรที่ใช้ทำนายที่ทำให้อัตราส่วนความผันแปรระหว่างกลุ่มและความผันแปรภายในกลุ่มมีค่าสูงสุด หรือทำให้ค่าร้อยละของทำนายสมาชิกกลุ่มที่ผิดพลาดมีค่าน้อยที่สุด โดยตัวแปรที่ใช้ทำนายเป็นตัวแปรเชิงปริมาณและมีการแจกแจงปกติ แต่อาจมีบางตัวที่เป็นตัวแปรเชิงคุณภาพได้ มีเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมของตัวแปรที่ใช้ทำนายแต่ละกลุ่มมีค่าเท่ากัน ส่วนตัวแปรตอบสนองต้องเป็นตัวแปรเชิงคุณภาพ ในการใช้วิธี Discriminant Analysis ทำนายสมาชิกกลุ่มด้วยโปรแกรมสำเร็จรูป SPSS สามารถเลือกใช้วิธีการทำนายได้ 2 รูปแบบ คือ แบบ Original ซึ่งทำนายสมาชิกกลุ่มโดยใช้ข้อมูลทั้งหมดในการสร้างสมการจำแนกกลุ่ม หรือแบบ Cross__Validation ซึ่งทำนายสมาชิกกลุ่มโดยใช้ข้อมูล n-1 ราย (เมื่อ n แทนจำนวนข้อมูลทั้งหมด) ในการสร้างสมการจำแนกกลุ่ม และเหลือข้อมูลอีก 1 รายไว้สำหรับทำนายการเป็นสมาชิกกลุ่มจากสมการที่สร้างขึ้น ในกรณีที่ค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้องซึ่งได้จากแบบ Cross__Validation มีค่าน้อยกว่าร้อยละ 25 ของผลรวมกำลังสองของความน่าจะเป็นเบื้องต้น (Prior Probability) ของตัวแปรตอบสนองแต่ละกลุ่มจะให้ผลการทำนายที่ผิดพลาดได้ และหากพบว่าอัตราส่วนของจำนวนข้อมูลทั้งหมดต่อจำนวนตัวแปรที่ใช้ทำนายน้อยกว่า 20 รายต่อ 1 ตัวแปร หรือจำนวนข้อมูลในกลุ่มน้อยของตัวแปรตอบสนองน้อยกว่า 20 หรือในกรณีที่ตัวแปรที่ใช้ทำนายมีมาตรการวัดในระดับช่วง ผู้วิจัยจำเป็นต้องระมัดระวังในการอธิบายผล (Schwab, 2003)

วิธีดำเนินการวิจัย

ในการศึกษาได้ใช้ข้อมูลจากงานวิจัย 3 เรื่อง และข้อมูลที่อยู่ในโปรแกรมสำเร็จรูป SPSS โดยกำหนดให้ตัวแปรที่ใช้ทำนายมีความสัมพันธ์กับตัวแปรตอบสนองในระดับต่างๆ ซึ่งได้แบ่งข้อมูลตามค่าสัมประสิทธิ์การตัดสินใจออกเป็น 3 ระดับ คือ 0-0.250 0.251-0.5 และมากกว่า 0.5 ตัวแปรที่ใช้ทำนายมีทั้งตัวแปรเชิงปริมาณและตัวแปรเชิงคุณภาพ ให้มีร้อยละของตัวแปรเชิงคุณภาพ 3

ระดับ คือ ไม่มีตัวแปรเชิงคุณภาพเลย มีตัวแปรเชิงคุณภาพไม่มากกว่าร้อยละ 50 และมีตัวแปรเชิงคุณภาพมากกว่าร้อยละ 50 ค่า VIF (Variance Inflation Factor) เป็นค่าที่แสดงความสัมพันธ์ระหว่างตัวแปรที่ใช้ทำนาย โดยหากค่า VIF ของตัวแปรที่มากที่สุดมีค่าตั้งแต่ 4 ขึ้นไปอาจจะทำให้เกิดความคลาดเคลื่อนในการทดสอบโดยรวม (Overall Significant Test) (Garson, 2006) จึงกำหนดให้ค่า VIF มี 2 ระดับ คือ VIF ไม่มากกว่า 4 และ VIF มากกว่า 4 สัดส่วนของจำนวนข้อมูลทั้งหมดต่อจำนวนข้อมูลในกลุ่มน้อยของตัวแปรตอบสนอง โดยแบ่งเป็น 3 ระดับ คือ ค่าสัดส่วนไม่มากกว่า 2 (แสดงว่าข้อมูลทั้งสองกลุ่มของตัวแปรตอบสนองมีจำนวนใกล้เคียงกัน) สัดส่วนมีค่ามากกว่า 2 แต่ไม่มากกว่า 3 และสัดส่วนมีค่ามากกว่า 3

จากข้อมูลที่ใช้ศึกษาได้สร้างชุดข้อมูลขึ้น 141 ชุด สามารถสรุปได้ดังแสดงในตารางที่ 1

ตารางที่ 1 รายละเอียดของข้อมูลที่ใช้ศึกษา

เรื่องที่ศึกษา	ตัวแปรที่ใช้ศึกษา
จากงานวิจัยเรื่อง “Probit Analysis and Economic Education” (Spector, 1989: 37-44) ใช้ศึกษาปัจจัยที่ทำให้นักศึกษามีเกรดเฉลี่ยเพิ่มขึ้น	เกรดเฉลี่ย คะแนนทดสอบก่อนเรียน และการใช้เทคนิคการสอนแบบใหม่
จาก http://www.oxfordjournals.org ใช้วิเคราะห์หาสาเหตุของโรคหลอดเลือดหัวใจ	การรับประทานยารักษาโรคหัวใจ การทำ Thallium Test ระดับความตีบตันของเส้นเลือด การใช้ยาในกลุ่ม Proprannolol อัตราการเต้นสูงสุดของหัวใจขณะที่ออกกำลังกาย การอุดตันของหลอดเลือด อาการเจ็บหน้าอกระหว่างออกกำลังกาย และเพศ
จาก http://www.stat.uga.edu ใช้ศึกษาปัจจัยเสี่ยงที่มีผลต่อการให้เครดิตนักศึกษา	เพศ อายุ เกรดเฉลี่ย จำนวนชั่วโมงทำงานต่อสัปดาห์ และวิชาเอกที่เลือกศึกษา
จาก file Employee* ใช้ศึกษาปัจจัยที่มีผลต่อรายได้คุณลักษณะของพนักงานที่มีสัญชาติอื่น	จำนวนปีที่เรียน เพศ ตำแหน่งงาน ประสบการณ์ทำงาน เงินเดือนปัจจุบัน เงินเดือนเริ่มต้น ช่วงเวลาที่ทำงาน
จาก file Car* ใช้ศึกษาปัจจัยที่มีผลทำให้รถยนต์มีอัตราการสิ้นเปลืองน้ำมันเชื้อเพลิงมากกว่า 25 ไมล์ต่อแกลลอน ปัจจัยที่มีผลในการจำแนกรถยนต์ที่มีความจุของถังน้ำมันมากกว่า 50 ลิตร ปัจจัยที่มีผลในการจำแนกรถยนต์ที่มีน้ำหนักมากกว่า 3 ตัน ปัจจัยที่มีผลทำให้รถยนต์มีราคามากกว่า 28,000 เหรียญ	ประเภทรถ ขนาดเครื่องยนต์ กำลังม้า ระยะห่างระหว่างล้อหน้าและล้อหลัง น้ำหนักรถ ความกว้างและความยาวของรถยนต์

ตารางที่ 1 รายละเอียดของข้อมูลที่ใช้ศึกษา (ต่อ)

เรื่องที่ศึกษา	ตัวแปรที่ใช้ศึกษา
จาก file World95* ใช้ศึกษาปัจจัยที่มีผลในการจำแนกประเทศตามกลุ่ม OCED การเป็นสมาชิกของประเทศในยุโรปตะวันออก เอเชีย แอฟริกา ตะวันออกกลาง กลุ่มลาติน กลุ่มทะเลทราย เขตแห้งแล้งกึ่งทะเลทราย และในเขตร้อน	สัดส่วนของประชาชนที่อยู่ในเมือง รายได้ประชาชาติ สัดส่วนของคนที่อ่านหนังสือออก อัตราการเกิด จำนวนประชาชนต่อพื้นที่ 1 ตารางกิโลเมตร อัตราการเพิ่มของประชาชนในแต่ละปี
จาก file Property-Assess* ใช้ศึกษาปัจจัยที่มีผลทำให้มีราคาประเมินที่สูงกว่าราคาเฉลี่ย ปัจจัยที่มีผลต่อราคาประเมินของที่ดินในภาคใต้ ภาคตะวันออก ภาคกลาง ภาคเหนือ และภาคตะวันตก	เขตพื้นที่ ผู้ประเมิน ราคาขาย ราคาประเมินครั้งล่าสุด ระยะเวลานับตั้งแต่การประเมินครั้งล่าสุด
จาก file Satisf* ใช้ศึกษาปัจจัยที่มีผลต่อการตัดสินใจซื้อ ความพึงพอใจในการใช้บริการ ความพึงพอใจโดยรวม ความพึงพอใจในเรื่องราคา ความพึงพอใจในเรื่องคุณภาพของสินค้า ความพึงพอใจในการใช้บริการ การเป็นห้างขนาดเล็ก ขนาดกลาง และห้างขนาดใหญ่	วิธีการชำระเงิน ช่วงอายุของผู้ตอบ ความถี่ในการซื้อ ระดับความพึงพอใจในเรื่องต่างๆ ความพึงพอใจโดยรวม การตัดสินใจซื้อ ระยะทาง
จาก file Breast and Cancer* ใช้ศึกษาปัจจัยที่มีผลต่อการมีต่อมน้ำเหลืองโต	อายุ เวลา ขนาดของ Pathological Tumor, Positive Axillary, Lymph Nodes, Histologic Grade, Status, Estrogen, Receptor Status, Progesterone Receptor Status
จาก file Grocery__ Coupons* ใช้ในการศึกษาปัจจัยที่ใช้ในการจำแนกคุณลักษณะของผู้ที่ใช้คูปอง ปัจจัยที่ใช้ในการจำแนกการเป็นผู้รับประทานอาหารมังสวิรัติ	ลักษณะการซื้อของ ประเภทของร้านค้า ขนาดของร้าน บุคคลที่ซื้ออาหารให้พฤติกรรมการซื้อ
จาก file Patlos__Sample* ใช้ศึกษาปัจจัยที่มีผลทำให้มีความดันโลหิตสูง โรคหัวใจ	ความดันโลหิต การสูบบุหรี่ การออกกำลังกาย ความอ้วน การใช้ยาปฏิชีวนะ ผลการตรวจ CPK ผลการตรวจ Troponin T blood
จาก file Cellular* ใช้ศึกษาปัจจัยที่ทำให้มีผู้ใช้บริการมากกว่า 3 ปี ปัจจัยที่ทำให้มีผู้ใช้งานในธุรกิจมากกว่าร้อยละ 30 ปัจจัยที่มีผลทำให้มีค่าใช้จ่ายเฉลี่ยมากกว่าเดือนละ 65 เหรียญ และปัจจัยที่มีผลต่อการได้คะแนนมากกว่าร้อยละ 50	ค่าใช้จ่ายเฉลี่ยต่อเดือน สัดส่วนการใช้งานในทางธุรกิจ จำนวนปีที่ใช้บริการ รายได้ของครอบครัว เวลาที่ใช้บริการโดยเฉลี่ย คะแนนแนวโน้มที่จะยกเลิกการใช้บริการ
จาก file Survey__Sample* ใช้ศึกษาปัจจัยที่ทำให้พนักงานไม่มีความสุขในการทำงาน ปัจจัยที่มีผลต่อการมีความสุขในชีวิตสมรส	การถูกลดค่าแรง คู่สมรสถูกไล่ออกจากงาน การเจ็บป่วยของบุตร การดื่มสุรา ความพึงพอใจในงาน ความคิดเห็นเกี่ยวกับตนเอง การใช้ยาเสพติด

ตารางที่ 1 รายละเอียดของข้อมูลที่ใช้ศึกษา (ต่อ)

เรื่องที่ศึกษา	ตัวแปรที่ใช้ศึกษา
จาก file Recidivism* ใช้ศึกษาปัจจัยที่มีผลทำให้มีการกระทำผิดและถูกจับกุมในครั้งที่ 2 ปัจจัยที่ผู้ซึ่งเคยถูกจับกุมการมีงานทำ	ฐานะทางสังคม สถานภาพการทำงาน ลักษณะการกระทำความผิดครั้งแรก การเคยได้รับการประกันตัว การปรับ เปลี่ยนอุปนิสัย ระดับการศึกษา การขัดขืนในการถูกจับกุมครั้งที่ 2
จาก file road construction bids* ใช้ศึกษาปัจจัยที่มีผลในการจำแนกสัญญาที่เป็น Competitive Contract กับ Fixed Contract	ค่าก่อสร้างโดยประมาณจากผู้เชี่ยวชาญ สัดส่วนของราคาที่มีผู้เสนอราคาต่ำเป็นลำดับที่สองต่อราคาต่ำสุด สัดส่วนราคาของผู้ที่เสนอราคาต่ำเป็นลำดับที่สามต่อราคาต่ำสุด สัดส่วนของราคาที่มีผู้เสนอสูงสุดต่อราคาต่ำสุดการอยู่ในเขต South Florida District หรือไม่ สัดส่วนของจำนวนผู้ที่เสนอราคาต่อผู้ที่วางแผนจะเสนอราคา จำนวนวันทำงาน
จาก file Verd1985* ใช้ศึกษาปัจจัยที่ทำให้ได้คะแนนคณิตศาสตร์ในระดับต่ำ และระดับสูง ปัจจัยที่มีผลในการเป็นผู้ที่ชอบดนตรี classic และดนตรี new wave	สถานภาพ อายุ จำนวนสัตว์เลี้ยง ความถี่ในการอ่านหนังสือพิมพ์ต่อสัปดาห์ ดนตรีที่ชอบ แหล่งที่อยู่ คะแนนทดสอบทางภาษา คะแนนทดสอบทางคณิตศาสตร์
จาก file Bankloan* ใช้ศึกษาปัจจัยที่มีผลทำให้ได้รับอนุมัติเงินกู้	จำนวนปีที่ทำงานปัจจุบัน อายุ ระดับการศึกษา จำนวนปีที่อาศัยอยู่ในที่อยู่ปัจจุบัน รายได้ของครอบครัว อัตราส่วนหนี้สินต่อรายได้ จำนวนหนี้ในบัตรเครดิต หนี้สินอื่นๆ
จาก file Car__Sales* ใช้ศึกษาปัจจัยในการจำแนกคุณลักษณะของรถ Truck และ Automobile จำแนกรถยนต์ที่มีความจุของถังน้ำมันมากกว่า 60 ลิตร จำแนกรถยนต์ที่มีน้ำหนักมากกว่า 3 ตัน มีราคามากกว่า 28,000 เหรียญ และปัจจัยที่ทำให้รถยนต์มีอัตราการสิ้นเปลืองน้ำมันเชื้อเพลิงมากกว่า 25 ไมล์ ต่อแกลลอน	ยอดขาย จำนวนรถที่ขายซึ่งใช้ไม่เกิน 4 ปี ราคา ขนาดของเครื่องยนต์ กำลังม้า ระบบการขับเคลื่อน ความกว้าง ความยาว น้ำหนักรถ ความจุของถังน้ำมัน อัตราการสิ้นเปลืองน้ำมัน
จาก file German__Credit* ใช้ศึกษาปัจจัยที่มีผลต่อการได้รับเครดิตของคนงานต่างชาติและคุณลักษณะของคนงานต่างชาติ	สถานภาพ ช่วงเวลาที่จ้าง ทรัพย์สินที่ครอบครอง ประเภทของที่อยู่ ลักษณะการว่าจ้าง เงินออม วงเงินเครดิต อายุ ยอดวงเงินเครดิตที่เหลืออยู่ เงินออมและหุ้น
จาก file Workprog* ใช้ศึกษาปัจจัยที่มีผลในการจำแนกผู้ที่สำเร็จการศึกษาระดับมัธยมปลาย ปัจจัยที่มีผลในการจำแนกสถานะภาพของโปรแกรม	อายุ สถานะภาพสมรส ระดับการศึกษา รายได้ก่อนเข้าโปรแกรม รายได้หลังเข้าโปรแกรม จำนวนบุตร
จาก file Car__Sale* ใช้ศึกษาปัจจัยที่ใช้ในการจำแนกรถยนต์ที่มี 6 กระบอกสูบ 8 กระบอกสูบ ปัจจัยที่ใช้ในการจำแนกรถที่ผลิตในอเมริกา ยุโรป ญี่ปุ่น รถที่ผลิตหลังปี 79 รถที่ผลิตก่อนปี 80 รถที่มีอัตราเร่งมากกว่า 16 นาที	อัตราการสิ้นเปลืองเชื้อเพลิง ชนิดเครื่องยนต์ กำลังม้า น้ำหนัก ความเร่ง ปีที่ผลิต ประเทศที่ผลิต จำนวนกระบอกสูบ

* เป็น file ข้อมูลที่อยู่ในโปรแกรมสำเร็จรูป SPSS version 11.5

ตารางที่ 2-ตารางที่ 5 แสดงจำนวนข้อมูลที่ใช้ศึกษาโดยจำแนกตามจำนวนตัวแปรที่ใช้ทำนาย ร้อยละของตัวแปรเชิงคุณภาพ ค่าสัมประสิทธิ์การตัดสินใจ ค่า VIF ของตัวแปรที่มีค่ามากที่สุด และสัดส่วนของจำนวนข้อมูลทั้งหมดต่อจำนวนข้อมูลในกลุ่มน้อย

ตารางที่ 2 จำนวนข้อมูลที่ใช้ศึกษาซึ่งจำแนกตามค่าร้อยละของตัวแปรเชิงคุณภาพและจำนวนตัวแปรที่ใช้ทำนาย

จำนวนตัวแปรที่ใช้ทำนาย	ร้อยละของตัวแปรเชิงคุณภาพ			ผลรวม
	0	1-50	มากกว่า 50	
น้อยกว่า 4 ตัว	11	6	5	22
4-5 ตัว	10	18	11	39
มากกว่า 5 ตัว	16	34	30	80
ผลรวม	37	58	46	141

ตารางที่ 3 จำนวนข้อมูลที่ใช้ศึกษาซึ่งจำแนกตามค่าสัมประสิทธิ์การตัดสินใจและจำนวนตัวแปรที่ใช้ทำนาย

จำนวนตัวแปรที่ใช้ทำนาย	ค่าสัมประสิทธิ์การตัดสินใจ			ผลรวม
	0-0.250	0.251-0.5	มากกว่า 0.5	
น้อยกว่า 4 ตัว	7	9	6	22
4-5 ตัว	20	11	8	39
มากกว่า 5 ตัว	26	30	24	80
ผลรวม	53	50	38	141

ตารางที่ 4 จำนวนข้อมูลที่ใช้ศึกษาซึ่งจำแนกตามค่า VIF ของตัวแปรที่มีค่ามากที่สุดและจำนวนตัวแปรที่ใช้ทำนาย

จำนวนตัวแปรที่ใช้ทำนาย	ค่า VIF ของตัวแปรที่มีค่ามากที่สุด		ผลรวม
	น้อยกว่า 4	ตั้งแต่ 4 ขึ้นไป	
น้อยกว่า 4 ตัว	18	4	22
4-5 ตัว	24	15	39
มากกว่า 5 ตัว	40	40	80
ผลรวม	82	59	141

ตารางที่ 5 จำนวนข้อมูลที่ใช้ศึกษาซึ่งจำแนกตามค่าสัดส่วนของจำนวนข้อมูลทั้งหมดต่อจำนวนข้อมูลในกลุ่มน้อยของตัวแปรตอบสนองและจำนวนตัวแปรที่ใช้ทำนาย

จำนวนตัวแปรที่ใช้ทำนาย	จำนวนข้อมูลทั้งหมด / จำนวนข้อมูลในกลุ่มน้อย			ผลรวม
	ไม่มากกว่า 2	2.01-3	มากกว่า 3	
น้อยกว่า 4 ตัว	7	6	9	22
4-5 ตัว	11	13	15	39
มากกว่า 5 ตัว	19	17	44	80
ผลรวม	37	36	68	141

ในการศึกษาได้ดำเนินการดังนี้

1. เปรียบเทียบผลการทดสอบโดยรวมซึ่งพิจารณาจากค่านัยสำคัญ เพื่อหาลักษณะข้อมูลที่ให้ผลการทดสอบแตกต่างกัน ใช้การทดสอบไคส แควร์ในการทดสอบความสัมพันธ์ระหว่างผลการทดสอบที่เหมือนกันของวิธีทั้งสามกับค่าร้อยละของตัวแปรเชิงคุณภาพและค่า VIF

2. หาค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้องที่ได้จากวิธี Discriminant Analysis โดยคำนวณแบบ Original ซึ่งใช้เฉพาะข้อมูลที่ให้ค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้องจากแบบ Cross__Validation มีค่าไม่น้อยกว่าร้อยละ 25 ของผลรวมกำลังสองของความน่าจะเป็นเบื้องต้นของตัวแปรตอบสนองแต่ละกลุ่ม

3. หาค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้องของข้อมูลที่มีความสอดคล้องกันระหว่างผลการทดสอบโดยรวมและผลการทดสอบพารามิเตอร์แต่ละตัวที่ได้จากวิธี OLS Regression และวิธี Logistic Regression

4. หาความสัมพันธ์ระหว่างค่าสัมประสิทธิ์การตัดสินใจกับช่วงความเชื่อมั่น 95% ของค่าประมาณความน่าจะเป็นเฉลี่ยที่ตัวแปรตอบสนองจะมีค่าเท่ากับ 0 และ 1 ซึ่งได้จากวิธีทั้งสาม

5. ทดสอบความแตกต่างระหว่างค่าเฉลี่ยของร้อยละของผลทำนายที่ถูกต้องของกลุ่มที่มีค่าน้อยซึ่งได้จากวิธีทั้งสาม (ใช้ 0.5 เป็นเกณฑ์ในการแบ่งกลุ่ม เนื่องจากในการศึกษาได้กำหนดให้ตัวแปรตอบสนองมีค่าเท่ากับ 0 ถ้าได้ผลลัพธ์ที่ไม่สนใจและมีค่าเท่ากับ 1 ถ้าได้ผลลัพธ์ที่สนใจ) โดยใช้วิธีวิเคราะห์ความแปรปรวน (Analysis of Variance) วิธีเปรียบเทียบเชิงซ้อน (Multiple Comparison) ซึ่งแบ่งกลุ่มข้อมูลตามค่า VIF ค่าสัดส่วนของจำนวนข้อมูลทั้งหมดต่อจำนวนข้อมูลในกลุ่มน้อย ค่าสัมประสิทธิ์การตัดสินใจและค่าร้อยละของตัวแปรเชิงคุณภาพ

ผลการวิเคราะห์

ผลการทดสอบโดยรวม

จากการทดสอบโดยรวมพบว่าวิธีทั้งสามมีผลการทดสอบเหมือนกันร้อยละ 76.10 โดยวิธี OLS Regression และวิธี Discriminant Analysis ให้ผลการทดสอบเหมือนกันร้อยละ 97.85 วิธี OLS Regression และวิธี Logistic Regression ให้ผลการทดสอบเหมือนกันร้อยละ 77.42 วิธี Discriminant Analysis และวิธี Logistic Regression ให้ผลการทดสอบเหมือนกันร้อยละ 77.42 ซึ่งผลการทดสอบที่เหมือนกันนี้มีความสัมพันธ์กับร้อยละของตัวแปรเชิงคุณภาพที่ใช้ในการทำนาย (จากผลการทดสอบไคสแควร์ที่ระดับนัยสำคัญ 0.05) โดยเมื่อมีตัวแปรเชิงคุณภาพจำนวนมากขึ้นจะมีโอกาสที่วิธีทั้งสามจะให้การทดสอบโดยรวมที่แตกต่างกันมากขึ้น

การตรวจสอบประสิทธิภาพการทำนายสมาชิกกลุ่มที่ได้จากวิธี Discriminant Analysis

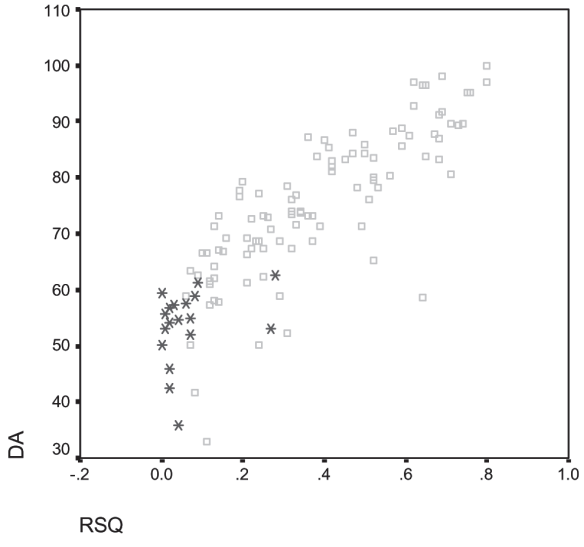
ในการใช้วิธี Discriminant Analysis ทำนายสมาชิกกลุ่มจะให้ผลการทำนายที่ผิดพลาดได้ หากค่าร้อยละของผลการทำนายที่ถูกต้องตามแบบ Cross__Validation มีค่าน้อยกว่าร้อยละ 25 ของผลรวมกำลังสองของความน่าจะเป็นเบื้องต้นของตัวแปรตอบสนองแต่ละกลุ่ม จึงไม่ควรใช้วิธีนี้ในการทำนายสมาชิกกลุ่ม โดยมักจะพบในกรณีที่ตัวแปรตอบสนองทั้งสองกลุ่มมีจำนวนข้อมูลแตกต่างกันมาก หรือเป็นข้อมูลที่มีค่าสัมประสิทธิ์การตัดสินใจน้อย (รูปที่ 1) เมื่อใช้วิธี OLS Regression และวิธี Logistic Regression จะได้ค่าร้อยละของผลการทำนายที่ถูกต้องไม่แตกต่างกัน (รูปที่ 3 และรูปที่ 4)

การตรวจสอบความสอดคล้องของผลการทดสอบโดยรวมและผลการทดสอบค่าพารามิเตอร์ที่ได้จากวิธี OLS Regression และวิธี Logistic Regression

โดยทั่วไปก่อนที่จะใช้ตัวแบบที่ได้จากวิธี OLS Regression และวิธี Logistic Regression ในการทำนายสมาชิกกลุ่ม ผู้วิจัยจำเป็นต้องตรวจสอบความสอดคล้องระหว่างผลการทดสอบโดยรวมและผลการทดสอบค่าพารามิเตอร์แต่ละตัว

จากการใช้วิธี OLS Regression พบว่ามีข้อมูล 2 ชุดให้ผลการทดสอบที่ไม่สอดคล้องกัน (ตารางที่ 8) โดยยอมรับสมมติฐานหลักของการทดสอบโดยรวมซึ่งแสดงว่าไม่มีตัวแปรใดที่สามารถอธิบายตัวแปรตอบสนองได้ แต่จากการทดสอบค่าพารามิเตอร์พบว่ามีตัวแปร 1 ตัวที่สามารถอธิบายตัวแปรตอบสนองได้ ซึ่งจะเกิดในกรณีที่ข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจน้อยและข้อมูลทั้งสองกลุ่มของตัวแปรตอบสนองมีจำนวนแตกต่างกันมาก เมื่อใช้วิธี Logistic Regression จะให้ค่าร้อยละของผลการทำนายที่ถูกต้องค่อนข้างน้อยและจากการตรวจสอบพบว่าไม่ควรใช้วิธี Discriminant Analysis เนื่องจากค่าร้อยละของผลการทำนายที่ถูกต้องตามแบบ Cross__Validation มีค่าน้อยกว่าร้อยละ 25 ของผลรวมกำลังสองของความน่าจะเป็นเบื้องต้นของตัวแปรตอบสนองแต่ละกลุ่ม

จากการใช้วิธี Logistic Regression พบว่ามีข้อมูล 25 ชุดที่มีผลการทดสอบไม่สอดคล้องกันซึ่งจากการทดสอบโดยรวมพบว่าสามารถใช้ตัวแบบทำนายสมาชิกกลุ่มได้ถูกต้องร้อยละ 100 แต่จากการทดสอบค่าพารามิเตอร์พบว่าไม่มีตัวแปรใดสามารถอธิบายการเปลี่ยนแปลงของตัวแปรตอบสนองได้ ซึ่งจะเกิดขึ้นในกรณีที่ตัวแปรตอบสนองมีความสัมพันธ์กับตัวแปรที่ใช้ทำนายบางตัวค่อนข้างมาก



* แทนข้อมูลที่ไม่ควรใช้วิธี Discriminant Analysis ในการทำนายสมาชิกกลุ่มเนื่องจากค่าร้อยละของผลการทำนายที่ถูกต้องตามแบบ Cross__Validation มีค่าน้อยกว่าร้อยละ 25 ของผลรวมกำลังสองของความน่าจะเป็นเบื้องต้นของตัวแปรตอบสนองแต่ละกลุ่ม

รูปที่ 1 การกระจายแสดงความสัมพันธ์ระหว่างค่าร้อยละของผลการทำนายที่ถูกต้องจากวิธี Discriminant Analysis (DA) และค่าสัมประสิทธิ์การตัดสินใจ (RSQ)

ตารางที่ 6 ค่าเฉลี่ย (ค่าส่วนเบี่ยงเบนมาตรฐาน; จำนวนข้อมูล) ของค่าร้อยละของผลการทำนายที่ถูกต้อง เมื่อใช้วิธี OLS Regression (OLS) วิธี Logistic Regression (LR) และวิธี Discriminant Analysis (DA)

ตัวแปรในการแบ่งกลุ่มข้อมูล	เกณฑ์ในการแบ่งกลุ่มข้อมูล	ค่าเฉลี่ย (ค่าส่วนเบี่ยงเบนมาตรฐาน ;จำนวนข้อมูล)		
		OLS	LR	DA
ค่า VIF	น้อยกว่า 4.00	55.88 (32.52;71)	44.72 (30.31;64)	**74.40** (14.11;49)
	ไม่น้อยกว่า 4.00	62.27 (31.53;48)	59.65 (27.69;39)	**76.98** (11.28;39)
ร้อยละของตัวแปรเชิงคุณภาพ	ไม่มีตัวแปรเชิงคุณภาพ	67.07 (32.07;25)	54.79 (25.43;20)	**79.90** (12.36;21)
	มีไม่มากกว่าร้อยละ 50	61.95 (32.20;53)	58.06 (29.27;44)	**77.34** (11.28;42)
	มากกว่าร้อยละ 50	48.70 (30.37;41)	39.43 (30.63;39)	**68.88** (13.88;25)
ค่าสัมประสิทธิ์การตัดสินใจ	0-0.250	33.66 (29.20;53)	27.15 (22.29;53)	**64.03** (9.49;32)
	0.251-0.500	68.61 (16.55;34)	68.63 (8.62;33)	**77.44** (6.07;26)
	มากกว่า 0.5	**88.74** (9.97;32)	**87.30** (9.53;17)	**86.20** (10.31;30)
ค่าสัดส่วนจำนวนข้อมูลทั้งหมดต่อจำนวนข้อมูลของตัวแปรตอบสนองในกลุ่มน้อย	ไม่มากกว่า 2*	**74.65** (15.82;36)	66.80 (22.31;32)	**75.91** (14.01;28)
	2.01-3.00**	61.77 (27.93;27)	49.76 (30.25;24)	**74.77** (10.54;21)
	มากกว่า 3***	46.45 (36.94;56)	39.50 (30.12;47)	**75.70** (13.58;39)

* ข้อมูลทั้งสองกลุ่มของตัวแปรตอบสนองมีจำนวนใกล้เคียงกัน

** ข้อมูลทั้งหมดมีจำนวนมากกว่า 2 เท่าแต่น้อยกว่า 3 เท่าของจำนวนข้อมูลในกลุ่มน้อย

*** ข้อมูลทั้งหมดมีจำนวนมากกว่า 3 เท่าของจำนวนข้อมูลในกลุ่มน้อย

ตารางที่ 7 ค่าร้อยละของผลการทำนายการเป็นสมาชิกกลุ่มที่ถูกต้อง ซึ่งได้จากวิธี OLS Regression (OLS) วิธี Logistic Regression (LR) และวิธี Discriminant Analysis (DA) ค่าVIF ค่าสัมประสิทธิ์การตัดสินใจ (RSQ) และค่าสัดส่วนของจำนวนข้อมูลทั้งหมดต่อจำนวนข้อมูลในกลุ่มน้อย (RATIO) ของข้อมูลที่มีค่าสัมประสิทธิ์การตัดสินใจน้อยกว่า 0.15

OLS	LR	DA	VIF	RSQ	RATIO
0	0	54*	1	0.02	4
0	2.9	57.6*	1.65	0.06	5
0	42	57.8	40.8	0.14	5
0	0	53.1*	4.25	0.01	3
0	0	59	1.17	0.06	27
0	0	57.2*	1.1	0.03	3
0	0	56.8*	4.5	0.02	9
0	0	42.4*	1.01	0.02	3
0	15.4	69.2	1.12	0.16	20
0	0	46*	1	0.02	4
0	0	59.5*	2.29	0.003	9
0	12	73.3	1.09	0.14	14
0	5	50	3.49	0.07	16
0	42	67	40.8	0.14	5
0	4	62.7	1.57	0.09	5
0	10	66.5	2.92	0.11	5
0	18	51.9*	39.36	0.07	5
0	28	61.1	40.29	0.12	5
0	0	55.7*	7.94	0.01	8

* แสดงข้อมูลที่ไม่ควรใช้วิธี Discriminant Analysis

Classification Table[a]

	Observed		Predicted		
			NEWSIZE		Percentage Correct
			.00	1.00	
Step 1	NEWSIZE	.00	51	0	100.0
		1.00	0	19	100.0
	Overall Percentage				100.0

a The cut value is .500

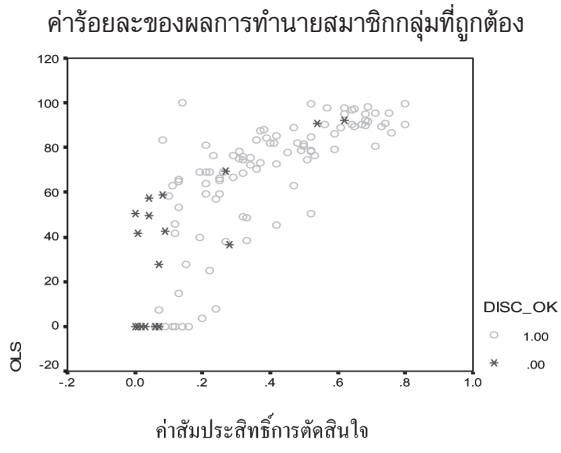
Variables in the Equation

		B	S.E.	Wald	df	Sig.
Step 1[a]	SALE	5.87	139.79	.002	1	.96
	TIME	4.41	181.73	.001	1	.98
	Constant	-335.33	8017.73	.002	1	.96

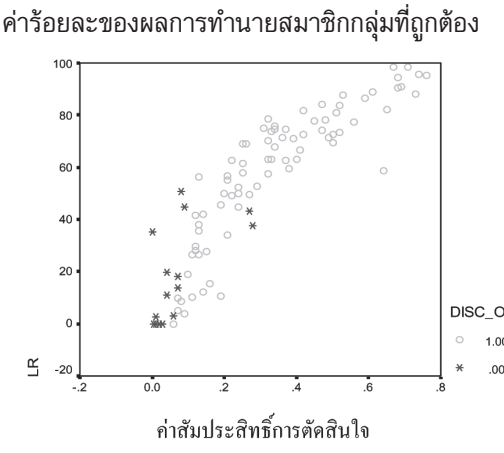
รูปที่ 2 ผลลัพธ์ที่ได้จากการประมวลผลด้วยโปรแกรม สำเร็จรูป SPSS ของข้อมูลที่มีผลการทดสอบโดยรวมและผลการทดสอบค่าพารามิเตอร์แต่ละตัวไม่สอดคล้องกัน

ตารางที่ 8 ค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้อง ซึ่งได้จากวิธี Logistic Regression (LR) และวิธี Discriminant Analysis (DA) ค่าสัมประสิทธิ์การตัดสินใจ (RSQ) ขนาดตัวอย่าง (n) และจำนวนข้อมูลของตัวแปรตอบสนองในกลุ่มน้อย (nsmall) ของข้อมูลที่มีผลการทดสอบโดยรวมและผลการทดสอบค่าพารามิเตอร์จากวิธี OLS Regression ไม่สอดคล้องกัน

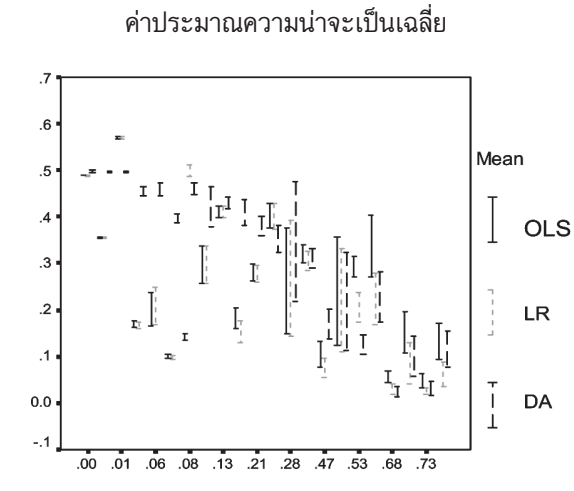
LR	DA	RSQ	n	nsmall
37.5	62.5	0.28	22	8
13.8	54.9	0.07	100	29



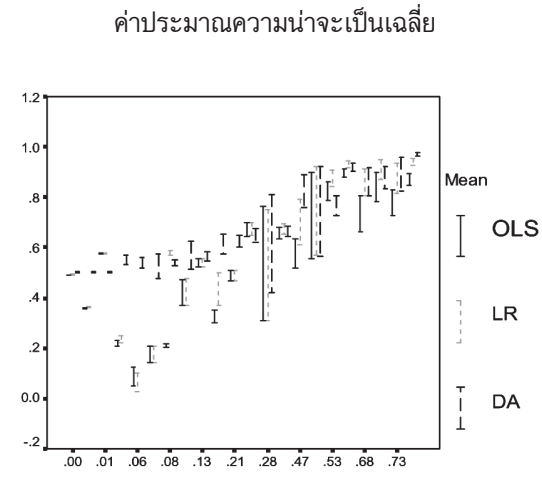
รูปที่ 3 แผนภาพการกระจายแสดงความสัมพันธ์ระหว่างค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้อง ซึ่งได้จากวิธี OLS Regression และค่าสัมประสิทธิ์การตัดสินใจ ซึ่งเครื่องหมาย * แสดงข้อมูลที่ไม่สามารถใช้วิธี



รูปที่ 4 แผนภาพการกระจายแสดงความสัมพันธ์ระหว่างค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้อง ซึ่งได้จากวิธี Logistic Regression และค่าสัมประสิทธิ์การตัดสินใจซึ่งเครื่องหมาย * แสดงข้อมูลที่ไม่สามารถใช้วิธี Discriminant Analysis ได้



(ก)



(ข)

รูปที่ 5 ความสัมพันธ์ระหว่างค่าสัมประสิทธิ์การตัดสินใจกับช่วงความเชื่อมั่น 95% ของค่าประมาณความน่าจะเป็นเฉลี่ยที่ตัวแปรตอบสนองจะมีค่าเท่ากับ 0 (ก) และ 1 (ข) ซึ่งได้จากวิธี OLS Regression (OLS) วิธี Logistic Regression (LR) และวิธี Discriminant Analysis (DA)

จากรูปที่ 2 แสดงผลลัพธ์ที่ได้จากการประมวลผลด้วยโปรแกรมสำเร็จรูป SPSS โดยใช้วิธี Logistic Regression ในการศึกษาปัจจัยที่ใช้จำแนกขนาดบริษัทโดยกำหนดให้ขนาดบริษัท (NEWSIZE) เป็นตัวแปรตอบสนองยอดขายครั้งล่าสุด (SALE) และช่วงเวลานับตั้งแต่มีการรายงานผล (TIME) เป็นตัวแปรที่ใช้ทำนายซึ่งให้ผลการทดสอบโดยรวมและผลการทดสอบค่าพารามิเตอร์ไม่สอดคล้องกัน แต่หากใช้วิธี OLS Regression จะให้ผลการทดสอบสอดคล้องกันโดยสามารถทำนายสมาชิกกลุ่มได้ถูกต้องร้อยละ 78.5 และร้อยละ 100 เมื่อศึกษาจากข้อมูลที่มีผลการทดสอบไม่สอดคล้องกันพบว่าการกำหนดให้ตัวแปรตอบสนองมีความสัมพันธ์กับตัวแปรที่ใช้ทำนายบางตัวมากหรือมีจำนวนข้อมูลน้อยเมื่อเทียบกับจำนวนตัวแปรที่ใช้ทำนาย อาจจะทำให้ผลการทดสอบโดยรวมและผลการทดสอบค่าพารามิเตอร์แต่ละตัวจากวิธี Logistic Regression ไม่สอดคล้องกันจึงไม่สามารถใช้วิธีนี้ในการทำนายสมาชิกกลุ่มได้ ผู้วิจัยสามารถใช้วิธี OLS Regression หรือ วิธี Discriminant Analysis แทนซึ่งทั้งสองวิธีจะให้ผลการทำนายถูกต้องโดยเฉลี่ยมากกว่าร้อยละ 85

ช่วงความเชื่อมั่น 95% ของค่าประมาณความน่าจะเป็นเฉลี่ย

จากรูปที่ 5 พบว่า ในกรณีที่ข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจน้อยกว่า 0.25 วิธี Logistic Regression และวิธี OLS Regression จะให้ช่วงความเชื่อมั่น 95% ของค่าประมาณความน่าจะเป็นที่ตัวแปรตอบสนองจะมีค่าเท่ากับ 0 และ 1 ไม่แตกต่างกัน และจะมีค่าแตกต่างจากค่าประมาณความน่าจะเป็นที่ได้จากวิธี Discriminant Analysis สำหรับในกรณีที่ข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจมากกว่า 0.25 ช่วงความเชื่อมั่น 95% ของค่าประมาณความน่าจะเป็นซึ่งได้จากวิธีทั้งสามมีค่าไม่แตกต่างกัน โดยในกลุ่มที่ตัวแปรตอบสนองมีค่าเท่ากับ 0 จะพบว่าเมื่อมีค่าสัมประสิทธิ์การตัดสินใจมากขึ้น ช่วงค่าประมาณความน่าจะเป็นจะแคบลงและเข้าใกล้ 0 มากขึ้น ส่วนในกลุ่มที่ตัวแปรตอบสนองมีค่าเท่ากับ 1 จะพบว่าเมื่อมีค่าสัมประสิทธิ์การตัดสินใจมากขึ้นช่วงค่าประมาณความน่าจะเป็นจะแคบลงและเข้าใกล้ 1 มากขึ้น เมื่อเปรียบเทียบช่วงการประมาณค่าความน่าจะเป็นเฉลี่ยที่ได้จากสามวิธีนี้พบว่าช่วงการประมาณค่าความน่าจะเป็นเฉลี่ยที่ได้จากวิธี Logistic Regression มีช่วงแคบกว่าวิธีอื่น และมีความสัมพันธ์กับค่าร้อยละของผลการทำนายที่ถูกต้องมากกว่าวิธีอื่นซึ่งสอดคล้องกับผลการศึกษาของ Pohlmann and Leitner (2003, 118-125) ก็พบว่าวิธี Logistic Regression ให้ค่าประมาณความน่าจะเป็นที่ของตัวแปรตอบสนองที่แม่นยำมากกว่า เมื่อใช้วิธี OLS Regression อาจจะให้ค่าประมาณตัวแปรตอบสนองที่อยู่นอกช่วง (0,1) ซึ่งค่าประมาณที่ติดลบนี้ไม่ค่อยมีความสัมพันธ์กับตัวแปรตอบสนองเหมือนกับค่าประมาณที่ได้จากวิธี Logistic Regression นอกจากนี้ยังพบว่าค่าประมาณที่ได้จากวิธี Logistic Regression จะมีค่าเข้าใกล้ค่าประมาณความน่าจะเป็นที่คำนวณได้จากค่าสังเกตมากกว่าวิธี OLS Regression ดังนั้นในงานวิจัยหากมีวัตถุประสงค์เพื่อต้องการประมาณค่าความน่าจะเป็นของตัวแปรตอบสนองวิธี Logistic Regression จะให้ตัวแบบที่ดีกว่า

การทดสอบความแตกต่างระหว่างค่าเฉลี่ยร้อยละของผลการทำนายที่ถูกต้องซึ่งได้จากวิธีทั้งสาม

จากค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้องด้วยวิธี OLS Regression และวิธี Logistic Regression ของข้อมูลที่มีผลการทดสอบโดยรวมสอดคล้องกับผลการทดสอบค่าพารามิเตอร์ และค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้องซึ่งได้จากวิธี Discriminant Analysis ของข้อมูลที่ให้ค่าร้อยละของผลการทำนายที่ถูกต้องตามแบบ Cross__Validation มีค่ามากกว่าร้อยละ 25 ของผลรวมกำลังสองของความน่าจะเป็นเบื้องต้นของตัวแปรตอบสนองแต่ละกลุ่ม (ตารางที่ 6) เมื่อนำมาทดสอบความแตกต่างระหว่างค่าเฉลี่ยที่ระดับนัยสำคัญ 0.05 สามารถสรุปได้ว่าวิธี Discriminant Analysis จะให้ค่าเฉลี่ยร้อยละของผลการทำนายถูกต้องมากที่สุดในทุกระดับของค่า VIF ของตัวแปรที่มีค่ามากที่สุด ค่าร้อยละของตัวแปรเชิงคุณภาพและค่าสัดส่วนของจำนวนข้อมูลทั้งหมดต่อจำนวนข้อมูลในกลุ่มน้อย แต่จะให้ค่าที่แตกต่างกันในทุกระดับของค่าสัมประสิทธิ์การตัดสินใจ โดยเมื่อข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจมากขึ้นจะให้ค่าเฉลี่ยร้อยละของผลการทำนายถูกต้องมากขึ้น ในกรณีที่ข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจมากกว่า 0.5 วิธีทั้งสามจะให้ค่าเฉลี่ยที่ไม่แตกต่างกัน สำหรับข้อมูลที่มีตัวแปรที่ใช้ทำนายซึ่งเป็นตัวแปรเชิงคุณภาพไม่มากกว่าร้อยละ 50 วิธี OLS Regression และวิธี Logistic Regression จะให้ค่าเฉลี่ยร้อยละของผลการทำนายที่ถูกต้องมากกว่ากรณีที่มีตัวแปรเชิงคุณภาพมากกว่าร้อยละ 50 และเมื่อพิจารณาจำนวนข้อมูลทั้งสองกลุ่มของตัวแปรตอบสนองก็พบว่าในกรณีที่จำนวนข้อมูลแตกต่างกันมากขึ้นวิธี Logistic Regression จะให้ค่าเฉลี่ยร้อยละของผลการทำนายที่ถูกต้องลดลง ในกรณีข้อมูลมีค่า VIF ของตัวแปรที่มากที่สุดมากกว่า 4.00 วิธี Logistic Regression จะให้ค่าเฉลี่ยร้อยละของผลการทำนายที่ถูกต้องมากกว่าข้อมูลที่มีค่า VIF ของตัวแปรที่มากที่สุดน้อยกว่า 4.00

ในกรณีที่มีค่าสัมประสิทธิ์การตัดสินใจน้อยกว่า 0.15 และมีจำนวนข้อมูลทั้งสองกลุ่มของตัวแปรตอบสนองแตกต่างกันมาก พบว่าวิธี OLS Regression และวิธี Logistic Regression มักจะให้ผลการทำนายถูกต้องเพียงกลุ่มเดียว (จากตารางที่ 7) ซึ่งเมื่อใช้วิธี Discriminant Analysis ก็อาจจะให้ผลการทำนายที่ผิดพลาดได้ เพราะให้ค่าร้อยละของผลการทำนายที่ถูกต้องตามแบบ Cross__Validation น้อยกว่าร้อยละ 25 ของผลรวมกำลังสองของความน่าจะเป็นเบื้องต้นของตัวแปรตอบสนองแต่ละกลุ่ม

การเปรียบเทียบข้อดีและข้อด้อยของวิธีทั้งสาม

จากการศึกษาพบว่ามีหลายกรณีที่วิธีทั้งสามให้ค่าร้อยละของผลการทำนายที่ถูกต้องไม่แตกต่างกัน ผู้วิจัยจึงสามารถเลือกใช้วิธีทางสถิติที่เหมาะสมกับข้อมูลที่มีอยู่ได้โดยพิจารณาจากข้อดีและข้อด้อยของแต่ละวิธี ซึ่งสามารถสรุปได้ดังนี้

- **วิธี Discriminant Analysis**

ข้อดี สามารถใช้ได้กับตัวแปรตอบสนองที่มีหลายค่า ทำให้ความคลาดเคลื่อนมีค่าลดลงและง่ายในการอธิบายความสัมพันธ์ของความแตกต่างระหว่างกลุ่ม

ข้อด้อย การอธิบายฟังก์ชันที่ได้มีความซับซ้อน

เช่นเดียวกับการอธิบายปัจจัยที่ได้จากการวิเคราะห์ปัจจัย (Factor Analysis) นอกจากนี้ยังมีความยุ่งยากในการตรวจสอบข้อสมมติของตัวแปรที่ใช้ทำนายแต่ละกลุ่มซึ่งต้องมีการแจกแจงปกติและเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมมีค่าเท่ากัน

- **วิธี Logistic Regression**

ข้อดี ไม่มีข้อจำกัดในเรื่องการแจกแจงปกติของตัวแปรที่ใช้ทำนาย การเท่ากันของค่าความแปรปรวน หรือความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ใช้ทำนาย ในการประมาณค่าสัมประสิทธิ์การถดถอยใช้วิธี Maximum Likelihood ซึ่งเป็นวิธีการแทนค่าซ้ำ (Iterative) ตัวแปรตอบสนองจึงไม่จำเป็นต้องมีการแจกแจงปกติ

ข้อด้อย หากการกำหนดค่าของตัวแปรตอบสนองไม่มีความหมายจะทำให้การอธิบายค่าสัมประสิทธิ์การถดถอยมีความยุ่งยากและจำเป็นต้องใช้จำนวนตัวอย่างมากกว่าวิธี OLS Regression โดยจำนวนข้อมูลของตัวแปรตอบสนองในกลุ่มน้อยต้องมีอย่างน้อยที่สุด 10 รายต่อพารามิเตอร์ 1 ตัว

- **วิธี OLS Regression**

ข้อดี เป็นวิธีที่รู้จักกันอย่างแพร่หลายใช้ง่าย ถึงแม้ตัวแปรที่ใช้ทำนายมีความสัมพันธ์กันมากก็ไม่ทำให้ค่าประมาณสัมประสิทธิ์การถดถอยมีความเอนเอียง

ข้อด้อย ในกรณีที่มีตัวอย่างขนาดเล็กการใช้ค่าสถิติเอฟทดสอบโดยรวมอาจให้ผลสรุปว่าไม่มีนัยสำคัญในขณะที่การทดสอบค่าพารามิเตอร์ให้ผลสรุปว่ามีนัยสำคัญ

บทสรุป

จากการเปรียบเทียบผลการทดสอบโดยรวม พบว่า วิธี OLS Regression และวิธี Discriminant Analysis ให้ผลการทดสอบที่เหมือนกันมากที่สุด (ร้อยละ 97.5) และพบว่าผลการทดสอบที่เหมือนกันของวิธีทั้งสามมีความสัมพันธ์กับร้อยละของตัวแปรเชิงคุณภาพที่ใช้ในการทำนายโดยเมื่อมีตัวแปรเชิงคุณภาพจำนวนมากขึ้นจะมีโอกาสที่วิธีทั้งสามจะให้การทดสอบโดยรวมที่แตกต่างกันมากขึ้น

จากการศึกษาเปรียบเทียบประสิทธิภาพในการทำนายการเป็นสมาชิกกลุ่มที่ถูกต้องด้วยวิธีทั้งสามในกรณีที่ตัวแปรตอบสนองมี 2 ค่าโดยพิจารณาจากค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้องของตัวแปรตอบสนองจากกลุ่มที่มีค่าน้อยพบว่าถึงแม้จะไม่ได้มีการตรวจสอบว่าตัวแปรที่ใช้ทำนายมีการแจกแจงปกติหรือไม่ก็ตาม วิธี Discriminant Analysis จะให้ค่าเฉลี่ยของผลการทำนายที่ถูกต้องมากที่สุดในทุกระดับของค่า VIF และค่าร้อยละของตัวแปรเชิงคุณภาพ จากการพิจารณาที่ทุกระดับของค่าสัมประสิทธิ์การตัดสินใจพบว่าเมื่อข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจเพิ่มขึ้นวิธี OLS Regression และวิธี Logistic Regression จะให้ค่าเฉลี่ยของร้อยละของผลการทำนายที่ถูกต้องเพิ่มขึ้น ส่วนวิธี Discriminant Analysis ให้ค่าเฉลี่ยของร้อยละของผลการทำนายที่ถูกต้องไม่แตกต่างกัน แต่สำหรับกรณีที่ข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจมากกว่า 0.5 วิธีทั้งสามให้ค่าเฉลี่ยของร้อยละของผลการทำนายที่ถูกต้องไม่แตกต่างกัน นอกจากนี้ยังพบว่าไม่ควรนำข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจน้อยกว่า 0.15 มาใช้ในการทำนายการเป็นสมาชิกกลุ่ม เพราะเมื่อใช้วิธี OLS Regression และวิธี

Logistic Regression จะให้ผลการทำนายที่ถูกต้องเพียงกลุ่มเดียว และหากใช้วิธี Discriminant Analysis ก็มีโอกาสที่จะให้ผลการทำนายที่ผิดพลาดได้เนื่องจากค่าร้อยละของผลการทำนายที่ถูกต้องตามแบบ Cross__Validation มีค่าน้อยกว่าร้อยละ 25 ของผลรวมกำลังสองของความน่าจะเป็นเบื้องต้นของตัวแปรตอบสนองแต่ละกลุ่ม

ในกรณีที่มีจำนวนข้อมูลทั้งสองกลุ่มของตัวแปรตอบสนองใกล้เคียงกันวิธี OLS Regression และวิธี Discriminant Analysis จะให้ค่าเฉลี่ยของร้อยละของผลการทำนายที่ถูกต้องไม่แตกต่างกันและมีค่ามากกว่าค่าเฉลี่ยที่ได้จากวิธี Logistic Regression เมื่อมีจำนวนข้อมูลแตกต่างกันมากขึ้นวิธี OLS Regression และวิธี Logistic Regression จะให้ค่าเฉลี่ยของร้อยละของผลการทำนายที่ถูกต้องลดลง ในขณะที่วิธี Discriminant Analysis ให้ค่าเฉลี่ยของร้อยละของผลการทำนายที่ถูกต้องไม่แตกต่างกัน

ในการเปรียบเทียบความสัมพันธ์ระหว่างช่วงความเชื่อมั่น 95% ของการประมาณค่าความน่าจะเป็นเฉลี่ยกับค่าร้อยละของผลการทำนายสมาชิกกลุ่มที่ถูกต้องพบว่าวิธี Logistic Regression ให้ช่วงของค่าประมาณความน่าจะเป็นที่แคบกว่าวิธีอื่น โดยค่าประมาณความน่าจะเป็นเฉลี่ยมีความสัมพันธ์กับค่าร้อยละของผลการทำนายที่ถูกต้องมากกว่าวิธีอื่น ดังนั้นหากผู้วิจัยต้องการประมาณค่าความน่าจะเป็นของตัวแปรตอบสนองควรเลือกใช้วิธี Logistic Regression แต่ทั้งนี้ผู้วิจัยต้องระมัดระวังที่จะไม่กำหนดให้ตัวแปรตอบสนองเป็นผลรวมเชิงเส้นของตัวแปรที่ใช้ทำนายเพราะจะให้ผลการทดสอบโดยรวมไม่สอดคล้องกับผลการทดสอบค่าพารา-มิเตอร์

จากการศึกษาครั้งนี้จึงทำให้ทราบว่าถึงแม้วิธีทั้งสามจะมีวัตถุประสงค์และแนวคิดในการสร้างตัวแบบเพื่อประมาณค่าความน่าจะเป็นและข้อสมมติที่แตกต่างกันแต่สามารถใช้ในการทำนายการเป็นสมาชิกกลุ่มได้ ซึ่งพบว่าการใช้วิธี Discriminant Analysis มีความยุ่งยากในการตรวจสอบข้อสมมติของตัวแปรที่ใช้ทำนายซึ่งต้องมีการแจกแจง Multivariate Normal และมีค่าความแปรปรวนและความแปรปรวนร่วมเท่ากัน อีกทั้งการอธิบายฟังก์ชันที่ได้ยังมีความซับซ้อนกว่าวิธี OLS Regression และ วิธี Logistic Regression จากผลการศึกษาทำให้ทราบว่าในกรณี ที่ข้อมูลมีค่าสัมประสิทธิ์การตัดสินใจมากกว่า 0.5 ผู้วิจัยสามารถใช้วิธี OLS Regression หรือ วิธี Logistic Regression แทนวิธี Discriminant Analysis ได้ และในกรณีที่จำนวนข้อมูลทั้งสองกลุ่มใกล้เคียงกัน ผู้วิจัยสามารถใช้วิธี OLS Regression แทนเพราะมีความยุ่งยากน้อยกว่าวิธี Discriminant Analysis

ดังนั้นจากการศึกษาจึงทำให้ทราบเกณฑ์ในการเลือกใช้วิธีสถิติที่เหมาะสมกับข้อมูลที่ต้องการศึกษา ในบางกรณีผู้วิจัยสามารถเลือกใช้วิธีสถิติที่ไม่ยุ่งยากและไม่จำเป็นต้องมีการตรวจสอบการแจกแจงของตัวแปรที่ใช้ทำนายแต่ให้ค่าร้อยละของผลการทำนายการเป็นสมาชิกกลุ่มที่ถูกต้องมากที่สุด

บรรณานุกรม

Cox, D.R., and Snell, E.J. 1989. **The Analysis of Binary Data.** 2nd ed. London: Chapman and Hall.

Garson, David. 2006. **Logistic Regression** [On-line]. Available: http://www2.chass.

ncsu.edu/garson/PA765/regress.htm#toleranc

Hocking, R.R. 1985. **The Analysis of Linear Models.** Monterey: Brooks.

Hosmer, D.W., and Lemeshow, S. 1989. **Applied Logistic Regression.** New York: Wiley.

King, E.N., and Ryan T.P. 2002. "A Preliminary Investigation of Maximum Likelihood Logistic Regression Versus Exact Logistic Regression." **American Statistician** 56, 3: 163-170.

Kleinbaum, D.G., Kupper, L.L., and Morgenstern, H. 1982. **Epidermiologic Research: Principles and Quantitive Methods.** Toronto: Lifetime Learning.

Menard, S. 1995. **Applied Logistic Regression Analysis.** Sage Univ Paper Series on Quantitative Applications in the Social Sciences series no. 07-106. Thousand Oaks, CA: Sage.

Montgomer, M.E., White, M.E., and Martin, S.W. 1987. "A Comparison of Discriminant Analysis and Logistic Regression for the Prediction of Coliform Mastitis in Dairy Cows." **Journal of Veterinary Research** 51: 495-498.

Pampel, Fred. 2000. **Logistic Regression: A Primer.** Thousand Oaks, CA: Sage.

Pigeon, Joseph G., and Heyse, Joseph F. 1999. "A Cautionary Note about Assessing the Fit of Logistic Regression Models." **Journal of Applied Statistics** 26,7: 847-853.

Schwab, A.J. 2003. **Strategy for Complete Discriminant Analysis** [On-line]. http://www.utexas.edu/Courses/schwab/sw38847/SolvingProblems

Spector, L., and Mazzeo, M. 1980. "Probit Analysis and Economic Education." **Journal of Economic Education** 11, 2: 37-44.

Wright, R.E. 1995. "Logistic Regression." In L.G. Grimm and P.R. Yarnold (eds.), **Reading and Understanding Multivariate Statistics.** Washington, D.C: APA.

Asst. Prof. Doungporn Hatchavanich received her Master of Science in Statistics from Chulalongkorn University, Thailand. She is currently a lecturer at the School of Science, The University of the Thai Chamber of Commerce. Her main interest is in Sampling Techniques and Regression Analysis.