

การวิเคราะห์การจำแนกกลุ่ม

Discriminant Analysis

กอดาริยะห์ เสกเมธิ

การวิเคราะห์การจำแนกกลุ่มเป็นเทคนิคที่ใช้ในการแบ่งกลุ่มคน สัตว์ องค์กร หรือสิ่งของ ฯลฯ ออกเป็นกลุ่มย่อยๆ ตั้งแต่ 2 กลุ่มขึ้นไป ในที่นี้จะใช้คำว่า case แทน คน หรือองค์กรหรือสิ่งของ ฯลฯ โดย case ที่อยู่ในกลุ่มเดียวกันจะคล้ายกัน case ที่อยู่ต่างกลุ่มกันจะแตกต่างกัน และ case ใด case หนึ่งจะต้องอยู่ในกลุ่มใดกลุ่มหนึ่งเพียงกลุ่มเดียว เช่น แบ่งบริษัทผู้ผลิตสินค้าเป็น 2 กลุ่ม กลุ่มส่งออกและกลุ่มเฉพาะขายภายในประเทศ บริษัทที่อยู่ในกลุ่มผู้ส่งออก จะคล้ายกันหรือเหมือนกันในแง่การขาย คือ ทุกบริษัทส่งสินค้าออกไปจำหน่ายต่างประเทศ(หมายถึงบริษัทที่ส่งออกอย่างเดียว หรือบริษัททั้งขายภายในประเทศและส่งออก) ในขณะที่บริษัทที่อยู่ในกลุ่มขายในประเทศจะไม่ได้ทำการส่งออก ซึ่งจะเริ่มต้นด้วยการเก็บข้อมูลที่ทราบจำนวนกลุ่ม และทราบว่าแต่ละ case อยู่ในกลุ่มใด แล้วใช้ข้อมูลดังกล่าวในการสร้างสมการแสดงความสัมพันธ์ระหว่างกลุ่มตัวแปรที่คาดว่าจะทำให้ case ที่อยู่คนละกลุ่มมีความแตกต่างกัน และยังสามารถใช้สมการดังกล่าวจัดกลุ่มหรือพยากรณ์กลุ่มให้กับ case ที่ยังไม่ทราบกลุ่ม

ตัวอย่างการใช้เทคนิคการจำแนกกลุ่ม

1. การแบ่งกลุ่มลูกหนี้ธนาคารหรือสถาบันการเงินออกเป็น 2 กลุ่ม คือลูกหนี้ปกติ และลูกหนี้มีปัญหา ถ้าคิดว่าตัวแปรหรือปัจจัยที่ทำให้ลูกหนี้มีสถานภาพหนี้ต่างกันคือ รายได้ ปริมาณหนี้ อายุ อาชีพ สถานภาพสมรส เป็นต้น
2. แบ่งกลุ่มครอบครัวที่มีบุตรเป็น 2 กลุ่ม คือกลุ่มที่วางแผนครอบครัวกับกลุ่มที่ไม่วางแผนครอบครัว โดยคิดว่าตัวแปรหรือปัจจัยที่ทำให้ครอบครัวอยู่คนละกลุ่ม คือรายได้ ระดับการศึกษา และอาชีพของหัวหน้าครอบครัว เป็นต้น
3. แบ่งกลุ่มพนักงานขายออกเป็น 3 กลุ่ม คือกลุ่มที่ได้รับรางวัลยอดเยี่ยม กลุ่มที่ได้รับรางวัลชมเชย และกลุ่มที่ไม่ได้รับรางวัล ซึ่งคิดว่าปัจจัยหรือตัวแปรที่ทำให้พนักงานถูกจัดอยู่ในกลุ่มที่ต่างกันคือ ยอดขาย อายุงาน ระดับการศึกษา ผลงานด้านอื่นๆ เป็นต้น

เงื่อนไข

1. ต้องการทราบจำนวนกลุ่มมาก่อนว่ามีกี่กลุ่ม และต้องเป็นกลุ่มที่มีอยู่จริง
2. ต้องการทราบว่า case ใดอยู่กลุ่มใดมาก่อน(เป็นกลุ่มที่มีจริงอยู่แล้ว)และ case หนึ่งต้องอยู่กลุ่มใดกลุ่มหนึ่งมาก่อน หรือกลุ่มต้องไม่ซ้ำซ้อนกัน เช่น นายน้อย เป็นลูกหนี้ธนาคาร A และเป็นลูกหนี้ที่ดี (ผ่อนชำระตามกำหนด) จึงจัดนายน้อยให้อยู่ในกลุ่มที่หนึ่งคือลูกหนี้ดี

หลักเกณฑ์ของการวิเคราะห์จำนวนจำแนกกลุ่ม

การที่ต้องทราบจำนวนกลุ่มและต้องทราบว่าแต่ละ case อยู่กลุ่มใด จึงเป็นการกำหนดว่า ข้อมูลนั้นจะต้องเป็นข้อมูลจริงที่เกิดขึ้นแล้ว เช่น ธนาคาร A ทราบว่ามีลูกหนี้รายใดบ้างเป็นลูกหนี้ปกติ รายใดบ้างเป็นลูกหนี้มีปัญหา และมีลูกหนี้ปกติที่ราย ลูกหนี้มีปัญหาราย แล้วศึกษาถึงปัจจัยหรือตัวแปรที่คาดว่าจะมีผลทำให้ลูกหนี้มีสถานภาพนี้แตกต่างกัน ถ้าคาดว่าตัวแปรที่ทำให้ลูกหนี้มีสถานภาพต่างกันอาจจะเนื่องจากการมีรายได้ต่างกัน จำนวนหนี้ต่างกัน จากนั้นจึงนำข้อมูลเกี่ยวกับสถานภาพหนี้ รายได้ และจำนวนหนี้ มาสร้างสมการแสดงความสัมพันธ์ของสถานภาพการเป็นหนี้กับรายได้และจำนวนหนี้ โดยใช้หลักการของการวิเคราะห์ความถดถอย สำหรับสมการความถดถอยที่สร้างขึ้นใน Discriminant Analysis นั้น ตัวแปรตามจะต้องเป็นตัวแปรเชิงกลุ่ม เช่น ถ้าแบ่งลูกหนี้เป็น 2 กลุ่ม ตัวแปรตาม D จะมี 2 ค่าคือ

$$D = \begin{cases} 1 & \text{ถ้าเป็นลูกหนี้ปกติ} \\ 2 & \text{ถ้าเป็นลูกหนี้มีปัญหา} \end{cases}$$

หรือถ้าธนาคารแบ่งลูกหนี้เป็น 3 กลุ่ม ตัวแปรตาม D จะมีค่า 3 ค่าคือ

$$D = \begin{cases} 1 & \text{ถ้าเป็นลูกหนี้ปกติ (ไม่ค้างชำระเลย)} \\ 2 & \text{ถ้าเป็นลูกหนี้มีปัญหา (ค้างชำระไม่เกิน 3 งวด)} \\ 3 & \text{ถ้าเป็นลูกหนี้เสีย (ค้างชำระเกิน 3 เดือน)} \end{cases}$$

ส่วนตัวแปรอิสระหรือตัวแปรที่ทำให้กลุ่มแตกต่างกัน ในที่นี้จะเรียกว่า **ตัวแปรจำแนกกลุ่ม** (Discriminator Variable) หรือบางครั้งเรียกตัวแปรต้น ซึ่งเป็นตัวแปรที่นำมาใช้ในการแบ่งกลุ่ม case ส่วนตัวแปรอิสระหรือตัวแปรจำแนกกลุ่มควรเป็นตัวแปรเชิงปริมาณ กรณีที่ตัวแปรอิสระเป็นตัวแปรเชิงกลุ่ม (Categorical Variable) หรือตัวแปรเชิงคุณภาพจะต้องปรับให้อยู่ในระดับตัวแปรเทียม (Dummy Variable) ตัวแปรต้นที่ทำให้ case แตกต่างกัน อาจจะมีเพียง 1 ตัว หรือตั้งแต่ 2 ตัวขึ้นไป ความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรต้นจะอยู่ในรูปเชิงเส้น ดังนี้

$$D = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + e \quad \text{----- (3.1)}$$

โดยที่ D = เป็นตัวแปรตามและเป็นข้อมูลเชิงกลุ่ม

X_1, X_2, X_p เป็นตัวแปรอิสระหรือตัวแปรจำแนกกลุ่ม ; $p \geq 1$

e = ค่าความคลาดเคลื่อน

ใน Discriminant Analysis จะเรียกสมการที่ (3.1) ว่า ฟังก์ชันจำแนกกลุ่ม (Discriminant Function) บางครั้งเรียก สมการจำแนกกลุ่ม หรือเรียกว่า **Fisher Discriminant Function** ซึ่ง R.A.Fisher ได้เป็นผู้คิดค้นขึ้นตั้งแต่ปี ค.ศ. 1936

เมื่อใช้ข้อมูลอย่างประมาณสมการที่ (3.1) จะเป็น

$$D' = a + b_1X_1 + b_2X_2 \dots + b_pX_p \quad \text{-----}(3.2)$$

การประมาณสัมประสิทธิ์ $\beta_0, \beta_1, \dots, \beta_p$ ด้วย a, b_1, \dots, b_p ตามลำดับ โดยมีเป้าหมายที่ทำให้ความแตกต่างระหว่างกลุ่มมีมากที่สุด นั่นคือ ทำให้ค่า

ความผันแปรระหว่างกลุ่ม (Between-Groups Sum Square)

ความผันแปรภายในกลุ่ม (Within Groups Sum Square)

มีค่ามากที่สุด และค่า D' ที่ได้ เรียกว่า **Discriminant Score**

วัตถุประสงค์ของการวิเคราะห์จำแนกข้อมูล

1. เพื่อหาสมการเชิงเส้นหรือฟังก์ชันจำแนกกลุ่มซึ่งแสดงความสัมพันธ์ของตัวแปรแยกกลุ่ม หรือตัวแปรตาม (D) กับตัวแปรอิสระอย่างน้อย $X_1, X_2, \dots, X_p; p \geq 1$ โดยการสร้างสมการเชิงเส้นดังกล่าวจะต้องใช้ข้อมูลจริงที่ทราบกลุ่มอยู่แล้ว
2. เพื่อทดสอบความแตกต่างระหว่างกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป (Multivariate) โดยการเปรียบเทียบค่ากลางของกลุ่ม (Group Centroid)
3. เพื่อนำสมการ หรือฟังก์ชันจำแนกกลุ่มที่สร้างในข้อ 1 มาใช้ในการพยากรณ์ว่า case ใหม่ที่ยังไม่ทราบกลุ่มว่าควรจะอยู่ในกลุ่มใด เช่น ถ้ามีลูกค้ามาซื้อใหม่ อาจใช้ฟังก์ชันจำแนกกลุ่มพยากรณ์ว่าลูกค้ารายนี้จะเป็นลูกหนี้ดีหรือลูกหนี้มีปัญหา โดยใช้ตัวแปรอิสระของลูกค้ารายดังกล่าว เช่น รายได้ ปริมาณหนี้เป็น อายุ เป็นต้น
4. เพื่อพิจารณาว่าตัวแปรอิสระตัวใดบ้างเป็นตัวแปรที่สำคัญที่ใช้ในการแบ่งกลุ่ม เช่น จากตัวอย่าง ลูกหนี้ธนาคาร ถ้าจากการพิจารณาค่าสัมประสิทธิ์ที่อยู่ในรูปมาตรฐาน (Standardized Coefficient) และการทดสอบสมมติฐานพบว่า รายได้เป็นตัวแปรที่สำคัญในการแบ่งกลุ่ม หรือทำให้ลูกหนี้มีสถานภาพต่างกัน และตัวแปรปริมาณหนี้สำคัญรองลงมา
5. สามารถใช้ฟังก์ชันจำแนกกลุ่มที่สร้างในข้อ 1 มาใช้ในการประเมินเปอร์เซ็นต์ความถูกต้องของการจำแนกกลุ่ม

ขั้นตอนของการจำแนกกลุ่ม

ก่อนที่จะกล่าวถึงขั้นตอนการแบ่งกลุ่ม case นั้น จะต้องทราบจำนวนกลุ่มมาก่อน และต้องทราบว่าแต่ละ case เดิมอยู่กลุ่มใด จึงจะสามารถใช้ฟังก์ชันจำแนกกลุ่มในการพยากรณ์ว่า case ใหม่ หรือ case ที่ยังไม่ได้ถูกจัดกลุ่มควรอยู่ในกลุ่มใด และยังสามารถศึกษาได้ว่าตัวแปร หรือปัจจัยใดมีอิทธิพลต่อการแบ่งกลุ่ม หลังจากทราบจำนวนกลุ่มแล้ว ให้ทำตามขั้นตอนดังนี้

1. พิจารณาหรือกำหนดตัวแปรอิสระที่คาดว่าจะมีผลหรืออิทธิพลต่อการจำแนกกลุ่ม เช่น การแบ่งกลุ่มลูกหนี้ อาจจะพิจารณาถึงรายได้ อายุ ปริมาณหนี้ อาชีพของลูกหนี้ เป็นตัวแปรอิสระที่คาดว่าจะทำให้เกิดความแตกต่างระหว่างกลุ่ม

2. เลือกตัวอย่างแต่ละกลุ่มเพื่อใช้เป็นตัวแทนของกลุ่ม เช่น เลือกตัวอย่างลูกหนี้ปกติและลูกหนี้ที่มีปัญหา จากลูกหนี้แต่ละกลุ่มที่มีอยู่จริง หรืออาจใช้ประชากรหรือลูกหนี้ทั้งหมดก็ได้

3. เก็บรวบรวมข้อมูลของตัวแปรอิสระที่เลือกไว้ในขั้นที่ 1 เช่น เก็บข้อมูลรายได้ อายุ ปริมาณหนี้ อาชีพของลูกหนี้ที่ถูกเลือกเป็นตัวอย่างในขั้นที่ 2

4. สร้างเกณฑ์หรือสมการจำแนกกลุ่มโดยใช้ข้อมูลที่เก็บจากขั้นที่ 2 และ 3 สำหรับหลักเกณฑ์ที่ใช้ จะนำหลักการของการวิเคราะห์ความถดถอย และการวิเคราะห์ความแปรปรวน (ANOVA) โดยการสร้างสมการเชิงเส้นที่เป็นสมการแสดงความสัมพันธ์ระหว่างตัวแปรแบ่งกลุ่ม (D) กับตัวแปรอิสระหรือประมาณค่า β ที่ทำให้ค่า

ความผันแปรระหว่างกลุ่ม (Between-Groups Sum Square)

ความผันแปรภายในกลุ่ม (Within Groups Sum Square)

มีค่าสูงสุด หรือทำให้มีเปอร์เซ็นต์การจัดกลุ่มผิดน้อยที่สุด สำหรับฟังก์ชันการจำแนกกลุ่มจะอยู่ในรูปเชิงเส้น ดังนี้

$$D' = a + b_1X_1 + b_2X_2 \dots + b_pX_p \quad \text{-----(3.2)}$$

โดยที่ D = ตัวแปรตามหรือเรียกว่า Discriminant Score

B_i = สัมประสิทธิ์ของสมการจำแนกกลุ่ม

X_i = ตัวแปรหรือเรียกว่า ตัวแปรจำแนกกลุ่ม (Discriminator Variable)

ตัวที่ i ; $i = 1, 2, \dots, p$

P = จำนวนตัวแปรจำนวนกลุ่ม

K = จำนวนกลุ่ม

$$\text{จำนวนสมการจำแนกกลุ่ม} = \min(p, k-1)$$

กรณีที่แบ่งเป็น 2 กลุ่ม จะมีสมการจำแนกกลุ่มเพียง 1 สมการ

5. นำเกณฑ์หรือฟังก์ชันจำแนกกลุ่มที่ได้จากขั้นที่ 4 มาพยากรณ์กลุ่มของคนหรือสิ่งของ

ที่เข้ามาใหม่

สรุปหลักเกณฑ์ของ Discriminant Analysis

ดังที่กล่าวมาแล้วว่า เทคนิค Discriminant Analysis ใช้หลักเกณฑ์ของการวิเคราะห์ความแปรปรวน (ANOVA) และการวิเคราะห์ความถดถอย (Regression) โดยต้องเริ่มจาก

1. มีข้อมูลจริงที่ทราบจำนวนกลุ่มและทราบว่าแต่ละ case อยู่กลุ่มใดบ้างซึ่งเป็นหลักเกณฑ์ของการวิเคราะห์ความแปรปรวน

2. หลังจากนั้นจึงใช้หลักเกณฑ์ของการวิเคราะห์ความถดถอยในการสร้างสมการจำแนกกลุ่ม โดยการประมาณค่าสัมประสิทธิ์ความถดถอย (β) ด้วย ใช้หลักการที่ว่า จะหาค่า β หรือ b ที่ทำให้

$$\frac{\text{ความผันแปรของกลุ่ม}}{\text{ความผันแปรในกลุ่ม}}$$

มีค่ามากที่สุด หรือกล่าวง่าย ๆ ว่า ให้อัตรา

ส่วนของความแตกต่างระหว่างกลุ่มต่อความแตกต่างภายในมีค่ามากที่สุดซึ่งแสดงว่า case ที่อยู่ต่างกลุ่มมีความแตกต่างกันมากกว่า case ที่อยู่ในกลุ่มเดียวกัน

3. ใช้สมการจำแนกกลุ่มในการจำแนกกลุ่มหรือกำหนดกลุ่มของ case ใหม่ เช่น ลูกค้ารายใหม่ พนักงานคนใหม่ หรือลูกหนี้รายใหม่

กรณีจำแนกออกเป็น 2 กลุ่ม ($K = 2$)

เมื่อแบ่ง case ออกเป็นสองกลุ่ม เช่น แบ่งลูกหนี้เป็น 2 กลุ่ม แบ่งบริษัทเป็น 2 กลุ่ม คือ กลุ่มที่จดทะเบียนในตลาดหลักทรัพย์ และกลุ่มที่ไม่ได้อยู่ในตลาดหลักทรัพย์ หรือแบ่งคนไข้โรคมะเร็งเป็น 2 กลุ่ม กลุ่มที่เป็นโรคมะเร็งขั้นที่ 2 ขึ้นไป กับกลุ่มที่เป็นระยะต้น (ขั้นที่ 1-2) กรณีเช่นนี้ตัวแปรตาม D มีเพียง 2 ค่า เช่น การแบ่งลูกหนี้ธนาคารเป็นลูกหนี้ปกติและลูกหนี้มีปัญหา โดยอาจจะกำหนดค่าตัวแปร D ดังนี้

$$D = \begin{cases} 1 & \text{ถ้าเป็นลูกหนี้ปกติ} \\ 2 & \text{ถ้าเป็นลูกหนี้มีปัญหา} \end{cases} \quad D = \begin{cases} 1 & \text{ถ้าเป็นลูกหนี้มีปัญหา} \\ 2 & \text{ถ้าเป็นลูกหนี้ปกติ} \end{cases}$$

โดยที่สมการแสดงความสัมพันธ์ซึ่งหมายถึง สมการแบ่งกลุ่มหรือเส้นแบ่งกลุ่มลูกหนี้ เป็นดังนี้

$$D = \beta_0 + \beta_1 \text{รายได้} + \beta_2 \text{ปริมาณหนี้} + e \text{-----} (3.3)$$

โดยที่ $D =$ สถานภาพหนี้

$$D = \begin{cases} 1 & \text{ถ้าเป็นลูกหนี้ปกติ} \\ 2 & \text{ถ้าเป็นลูกหนี้มีปัญหา} \end{cases}$$

จะพบว่าสามารถแบ่งลูกหนี้ออกเป็น 2 กลุ่มได้อย่างชัดเจน โดยที่กลุ่มที่ 1 ($D = 1$ ใช้เครื่องหมาย□) เป็นลูกหนี้ปกติ ส่วนกลุ่มที่ 2 ($D = 2$ ใช้เครื่องหมาย●) เป็นลูกหนี้ที่มีปัญหา อย่างไรก็ตาม จะใช้เส้นแบ่งกลุ่มซึ่งใช้สมการจำแนกกลุ่ม (สมการที่ (3.2)) ในการแบ่งกลุ่มโดยที่ส่วนที่อยู่ทางด้านขวาของเส้นจะถือเป็นลูกหนี้ปกติ (รายได้สูงแต่ปริมาณหนี้ต่ำ) แต่ถ้าอย่างด้านซ้ายของเส้นแบ่งกลุ่ม (รายได้ต่ำแต่ปริมาณหนี้สูง) จะหมายถึงลูกหนี้ที่มีปัญหา แต่จะพบว่า มีลูกหนี้ปกติบางรายที่ถูกจัดให้เป็นลูกหนี้ที่มีปัญหา ซึ่งแสดงด้วยส่วนที่แรเงา a โดยส่วนที่แรเงา a เป็นสัดส่วนของเปอร์เซ็นต์ของลูกหนี้ปกติ ที่ถูกจัดเป็นลูกหนี้ที่มีปัญหา หรือสัดส่วนของการจัดกลุ่มผิด ในขณะที่ส่วนที่แรเงา b หมายถึงสัดส่วนของลูกหนี้ที่มีปัญหา แต่ถูกจัดเป็นลูกหนี้ปกติด้วยการให้สมการจำแนกกลุ่ม (สมการที่ (3.2)) จึงเป็นสัดส่วนการการจัดกลุ่มผิดเช่นกัน หรืออาจเรียกส่วนที่แรเงาทั้ง 2 ส่วนว่าเป็นเปอร์เซ็นต์ความผิดพลาดในการจัดกลุ่ม หลักเกณฑ์ของ Discriminant Analysis จึงต้องการทำให้เปอร์เซ็นต์ความผิดพลาดในการจัดกลุ่มมีค่าต่ำสุด

การประมาณสัมประสิทธิ์ของสมการจำแนกกลุ่ม

จากฟังก์ชันจำแนกกลุ่มซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรตาม กับตัวแปรอิสระในสมการที่ (3.3) เมื่อใช้ข้อมูลตัวอย่างจะได้สมการที่ประมาณคะแนนจำแนกกลุ่ม (Discriminant Score) (D^*) ดังนี้

$$D^* = a + b_1 \text{รายได้} + b_2 \text{ปริมาณหนี้} \quad (3.4)$$

การประมาณสัมประสิทธิ์ β_0 , β_1 และ β_2 ด้วย a , b_1 และ b_2 ตามลำดับ โดยมีเป้าหมายทำให้ ความผันแปรระหว่างกลุ่มที่ 1 และ 2 มีค่ามากที่สุด นั่นคือให้มีความแตกต่างระหว่างกลุ่ม

ความผันแปรภายในกลุ่ม
มากกว่าความแตกต่างภายในกลุ่ม คือ กลุ่มลูกหนี้ปกติจะมีรายได้ และปริมาณหนี้ที่แตกต่างกันจากกลุ่มลูกหนี้ที่มีปัญหา มากกว่าความแตกต่างของรายได้และปริมาณหนี้ของลูกหนี้ที่อยู่ในกลุ่มเดียวกัน

เงื่อนไขของ Discriminant Analysis เมื่อมีตัวแปรอิสระ 2 ตัว และมีจำนวนกลุ่ม 2 กลุ่ม

1. ตัวแปรอิสระของแต่ละกลุ่มจะต้องมีการแจกแจงแบบ Multivariate Normal

$$\begin{bmatrix} \text{รายได้} \\ \text{ปริมาณหนี้} \end{bmatrix}_i \sim N \left[\begin{pmatrix} \mu \text{รายได้} \\ \mu \text{ปริมาณหนี้} \end{pmatrix}, \Sigma_{p \times p} \right]_i$$

โดยที่ $i =$ หมายถึงกลุ่มที่ I ; $I = 1, 2$

2. $\Sigma_{p \times p} =$ Variance-Covariance Matrix ของตัวแปรรายได้และปริมาณหนี้

$$= \begin{bmatrix} \text{Variance (รายได้)} & \text{Covariance (รายได้, ปริมาณหนี้)} \\ \text{Covariance (ปริมาณหนี้, รายได้)} & \text{Variance (ปริมาณหนี้)} \end{bmatrix}$$

ในที่นี้ $p = 2$

3. $\sum_1 = \sum_2$ นั่นคือ Vaiance-Covariance Matrix ของแต่ละกลุ่มต้องเท่ากัน
 โดยที่ $\sum_i =$ Variance-Covariance matrix ขนาด $p \times p$ ของกลุ่มที่ $i, i = 1, 2$

ในกรณีทั่วไป เมื่อต้องการแบ่ง case เป็น 2 กลุ่ม โดยมีตัวแปรอิสระ p ตัว จะได้สมการจำแนกกลุ่มเพียง 1 สมการ ดังนี้

$$D' = a + b_1 + X_1 + b_2 + X_2 + \dots + b_p X_p$$

โดยที่จะหาค่าสัมประสิทธิ์ a, b_1, b_2, \dots, b_p ที่ทำให้ ความผันแปรระหว่างกลุ่ม มีค่ามากที่สุด
ความผันแปรภายในกลุ่ม

กรณีที่จำแนกออกเป็น k กลุ่ม ($k \geq 3$)

เมื่อแบ่ง case ออกเป็น k กลุ่ม ($k \geq 3$) เช่น เป็นคนไข้โรคหัวใจเป็นสี่กลุ่ม ดังนี้

- กลุ่มที่ 1 : ระยะรุนแรง
- กลุ่มที่ 2 : ระยะปานกลาง
- กลุ่มที่ 3 : ระยะเริ่มต้น
- กลุ่มที่ 4 : ไม่เป็นโรคหัวใจ

โดยคาดว่าตัวแปรที่มีผลต่อการเป็นโรคหัวใจ 4 ตัวแปรคือ อายุ, น้ำหนัก, ระดับ HDL, ระดับ Cholesterol ในที่นี้ $p = 4, k = 4$

จะได้จำนวนสมการจำแนกกลุ่ม = $\min(p, k-1) = 3$ สมการ คือ

$$D'_i = a + b_1 \text{ อายุ} + b_2 \text{ น้ำหนัก} + b_3 \text{ HDL} + b_4 \text{ cholesterol} \dots \dots \dots (3.5)$$

$i = 1, 2, 3$

กรณีที่มิฟังก์ชันจำแนกกลุ่มดังเช่น สมการที่ (3.5) ตั้งแต่ 2 สมการขึ้นไป เมื่อมีเงื่อนไขแต่ละฟังก์ชันจะต้องไม่มีความสัมพันธ์ หรือจะต้อง orthogonal กับฟังก์ชันก่อนหน้า เช่น ถ้ามี 3 ฟังก์ชัน ฟังก์ชันที่ 2 จะต้องไม่มีความสัมพันธ์กับฟังก์ชันที่ 1 ส่วนฟังก์ชันที่ 3 จะต้องไม่มีความสัมพันธ์กับฟังก์ชันที่ 1 และ 2 ทั้งนี้จะทำให้สามารถแบ่ง หรือแยกค่าแปรปรวนแต่ละฟังก์ชันได้อย่างชัดเจน

เงื่อนไขของ Discriminant Analysis เมื่อมีจำนวน 3 กลุ่มขึ้นไป

1. ตัวแปรอิสระ p ตัว (X_1, X_2, \dots, X_p) มีการแจกแจง Multivariate Nomal Distribution
2. Variance-Covariance matrix (ขนาด $p \times p$) ของแต่ละกลุ่มต้องเท่ากัน นั่นคือ

$$\sum_1 = \sum_2 = \dots = \sum_k$$

หมายเหตุ

1. กรณีที่เป็นไปตามเงื่อนไขที่ว่า Variance-Covariance ของทุกกลุ่มเท่ากัน การจัดการ case โดยวิธีที่ 1-3 จะให้ผลลัพธ์เหมือนกัน
2. SPSS จะแสดงค่า Posterior Probability ของวิธีการที่ 1 ให้ พร้อมกับให้ค่า Mahalanobis Distance ของวิธีการที่ 2 มาด้วย

เงื่อนไขของลักษณะตัวแปร

สำหรับเทคนิค Discriminant Analysis ซึ่งใช้หลักการของการวิเคราะห์ความถดถอย จึงประกอบด้วย ตัวแปรตาม และตัวแปรอิสระ หรือตัวแปรจำแนกกลุ่ม ซึ่งจะต้องมีชนิดของตัวแปรดังนี้

1. ตัวแปรตาม D ควรเป็นตัวแปรเชิงกลุ่ม นั่นคือ เป็นข้อมูลชนิดนามกำหนด (Nominal Scale) หรือสเกลอันดับ (Ordinal Scale)
2. ตัวแปรอิสระหรือตัวแปรจำแนกกลุ่ม (X_1, X_2, \dots, X_p) ควรเป็นตัวแปรเชิงปริมาณคือ เป็นข้อมูลชนิดสเกลอันดับ (Interval Scale) หรือสเกลอัตราส่วน (Ratio Scale) เนื่องจากตัวแปร (X_1, X_2, \dots, X_p) จะต้องมีการแจกแจงแบบ Multivariate Normal กรณีที่มีตัวแปรเชิงกลุ่มผสมอยู่ด้วย ต้องแปลงให้อยู่ในรูปตัวแปรเทียม

ตัวอย่างงานวิจัยที่ 1

การศึกษาตัวแปรที่สามารถจำแนกเจตคติต่ออาชีพครู

ความเป็นมา

การศึกษานับเป็นรากฐานที่สำคัญประการหนึ่งสำหรับการสร้างสรรค์ความเจริญก้าวหน้า และการแก้ไขปัญหาของประเทศในด้านต่างๆ เพราะการศึกษามีความสัมพันธ์โดยตรงต่อการพัฒนาทรัพยากรบุคคลของประเทศให้เป็นผู้รู้จักคิด รู้จักทำ รู้จักแก้ไขปัญหา ตลอดจนรู้จักใช้ทรัพยากรที่มีอยู่ให้เกิดประโยชน์สูงสุด ประเทศจะเจริญก้าวหน้าได้จำเป็นต้องมีทรัพยากรบุคคลที่มีความรู้ ความคิดและความสามารถจำนวนมาก การศึกษาเป็นกระบวนการในการเสริมสร้างพัฒนาคุณภาพของพลเมืองให้มีปัญญา คุณธรรม จริยธรรม และมีสุขภาพพลานามัยสมบูรณ์ ตลอดจนมีความรู้และทักษะในการประกอบอาชีพ สามารถพึ่งตนเองและดำรงชีวิตได้อย่างเป็นสุข (แผนพัฒนาการศึกษาแห่งชาติ ระยะที่ 7, 2535 : 1) การที่จะบรรลุจุดมุ่งหมายของการจัดการศึกษาได้นั้นจำเป็นต้องมีกระบวนการจัดการศึกษาที่มีประสิทธิภาพ ปัจจัยหนึ่งที่สำคัญต่อการจัดการศึกษาก็คือครู ครูเป็นบุคคลที่มีบทบาทสำคัญที่สุดในการพัฒนาสังคมและประเทศชาติ ทั้งนี้เพราะสังคมและประเทศชาติจะเจริญก้าวหน้ามั่นคง และเป็นไปในรูปแบบใดย่อมขึ้นกับคุณภาพและวิธีการที่บุคคลในสังคมได้รับจากการศึกษาเป็นปัจจัยสำคัญ พุทธทาสภิกขุได้กล่าวไว้ว่า “โลกเราประกอบด้วยมนุษย์แต่ละคน แล้วแต่ว่าครูจะสร้างเขาขึ้นมาอย่างไรตั้งแต่สมัยยังเป็นเด็ก ถ้าคนในโลกนี้ โลกเป็นโลกที่ดี คนในโลกไม่ดี โลกไม่ดี คนในโลกจะดีหรือไม่ดีก็แล้วแต่ว่าครูจะปั้นเขาขึ้นมาอย่างไรเป็นส่วนใหญ่” (ยนต์ ชุ่มจิต, 2531 : 22 อ้างอิงมาจาก พุทธทาสภิกขุ, 2529 : 92) คำกล่าวที่ว่า “ครูคือแม่พิมพ์ของชาติ” เป็นคำกล่าวที่ให้ความสำคัญกับครูมาก แต่ก็ก็เป็นคำกล่าวที่ไม่เกินความจริง เพราะคุณภาพของสังคมและประเทศชาติขึ้นอยู่กับคุณภาพของประชาชนคุณภาพของประชาชนขึ้นอยู่กับคุณภาพของการศึกษา คุณภาพของการศึกษาขึ้นอยู่กับคุณภาพของครู จึงเป็นเรื่องที่ทุกฝ่ายต้องตระหนักถึงการพัฒนาครูให้เป็นครูที่มีคุณภาพ ครูที่มีคุณภาพนั้นจะต้องประกอบด้วยองค์ประกอบหลายๆด้าน เช่น วิชาความรู้ในเนื้อหาวิชาที่สอน ทักษะในการสอน เจตคติต่ออาชีพครูเป็นต้น เจตคติต่ออาชีพครูผู้สอนมีความสำคัญในอันที่จะช่วยส่งเสริมหรือสกัดกั้นการเรียนรู้ของนักเรียน การสอนของครูจะได้ผลดีมีประสิทธิภาพหรือเป็นไปในรูปใดนั้น ขึ้นอยู่กับเจตคติต่ออาชีพของครูผู้สอนเป็นสำคัญ (สำรวย มีখনอน, 2516 : 2)เจตคติของครูผู้สอนมี

ความสำคัญยิ่งกว่าสิ่งที่ครูผู้สอนรู้มากมาย ทั้งนี้เพราะเจตคติของครูจะเป็นพื้นฐานของการเรียนรู้ เจตคติ และมีอิทธิพลต่อการเปลี่ยนแปลงพฤติกรรมของผู้เรียน ครูซึ่งมีเจตคติไม่ดีต่ออาชีพของตน เช่น ไม่เห็นคุณค่าและความสำคัญของอาชีพ ไม่มีความภูมิใจ ไม่คิดที่จะปรับปรุงอาชีพให้ก้าวหน้า และไม่กระทำตนเป็นครูที่ดีในสายตาของเด็กนักเรียน เหล่านี้ล้วนทำให้การเรียนการสอนเป็นไปอย่างไรประสิทธิภาพ ซึ่งกรมฝึกหัดครู (2527 : 201) ได้ศึกษาพบว่า เจตคติต่ออาชีพเป็น องค์ประกอบพื้นฐานที่สำคัญในการประกอบอาชีพ เจตคติต่ออาชีพเป็นตัวกำหนดพฤติกรรมของ บุคคลทั้งในแนวทางการทำงานและเป้าหมายของการทำงาน ตลอดจนทำให้เกิดความพึงพอใจใน การทำงาน

ในปัจจุบันนี้เราจะพบเห็นพฤติกรรมของครูที่ขาดความมุ่งมั่นในการทำงานได้โดยไม่ยากนัก เช่น ทำงานสาย หารายได้พิเศษในเวลาราชการ เป็นต้น การขาดความมุ่งมั่นนี้เป็นปัญหาที่ สำคัญในอันที่จะทำให้การพัฒนาในด้านอื่นๆ เป็นไปได้อีก พฤติกรรมการขาดความมุ่งมั่นในการ ทำงานนี้สะท้อนถึงเจตคติต่อการทำงาน การจะแก้ปัญหานี้ได้เพื่อไปสู่เป้าหมายของการศึกษานั้น จำเป็นต้องแก้ที่เจตคติต่องานในอาชีพให้ได้เสียก่อน

การศึกษาตัวแปรที่สามารถจำแนกเจตคติต่ออาชีพครูจึงมีความจำเป็นในเบื้องต้น ในอันที่ จะเป็นแนวทางในการศึกษากระบวนการในการพัฒนาเจตคติต่ออาชีพครูของครู เพราะการที่จะ ศึกษากระบวนการในการพัฒนาเจตคติต่ออาชีพครูได้นั้น จำเป็นที่จะต้องทราบถึงความสัมพันธ์ และความสำคัญของตัวแปรที่เกี่ยวข้องกับเจตคติต่ออาชีพครูเสียก่อน ฉะนั้นการศึกษาตัวแปรที่ สามารถจำแนกเจตคติต่ออาชีพครูจึงจำเป็นในการศึกษากระบวนการในการพัฒนาเจตคติต่ออาชีพ ครูในวงการศึกษาต่อไป

แนวคิดทฤษฎี

1. ความหมายของเจตคติ
2. องค์ประกอบของเจตคติ
3. การเกิดเจตคติ
4. การเปลี่ยนแปลงเจตคติ
5. การวัดเจตคติ
6. การวิเคราะห์จำแนกประเภท
7. ประโยชน์ของสมการจำแนกประเภท
8. วิธีการสร้างสมการวิเคราะห์จำแนกประเภท
9. สถิติที่สำคัญของการวิเคราะห์จำแนกประเภท
10. ลักษณะของข้อมูลที่ใช้ในการวิเคราะห์จำแนกประเภท
11. เจตคติของครูที่มีต่ออาชีพครู

วัตถุประสงค์

1. เพื่อหาตัวแปรที่สามารถจำแนกระหว่างกลุ่มครูที่มีเจตคติต่ออาชีพครูต่ำ กับกลุ่มครูที่มีเจตคติต่ออาชีพครูสูง

2. เพื่อหาลำดับความสำคัญของตัวแปรที่สามารถจำแนกระหว่างกลุ่มครูที่มีเจตคติต่ออาชีพครูต่ำ กับกลุ่มครูที่มีเจตคติต่ออาชีพครูสูง

ระเบียบวิธีวิจัย

การวิจัยเชิงบรรยาย

ประชากร/กลุ่มตัวอย่าง

ประชากรในการวิจัยเป็นครู สังกัดสำนักงานการประถมศึกษาจังหวัดอุดรธานี จำนวน 3,467 คน จากโรงเรียนทั้งหมด 312 โรงเรียน กลุ่มตัวอย่างได้จากการสุ่มหลายขั้นตอน โดยสุ่มมา 5 อำเภอ จากทั้งหมด 9 อำเภอ แล้วสุ่มแบบสัดส่วนตามขนาดโรงเรียนเพื่อได้โรงเรียนจำนวน 50 โรงเรียน ซึ่งมีครูทั้งหมดที่ใช้ในการรวบรวมข้อมูลจำนวน 750 คน แล้วคัดเลือกเป็นกลุ่มตัวอย่างในการวิเคราะห์ข้อมูลจากผู้ที่ได้คะแนนเจตคติต่ออาชีพครูต่ำ จำนวน 175 คน และผู้ที่ได้คะแนนเจตคติต่ออาชีพครูสูง จำนวน 175 คน รวมจำนวน 350 คน

นิยามศัพท์

เจตคติ หมายถึง ความรู้สึกซึ่งเกิดจากมวลประสบการณ์ของบุคคลที่ใช้เป็นเกณฑ์ในการประเมินต่อสิ่งเร้าต่าง ๆ ซึ่งพร้อมจะแสดงออกมาในสองลักษณะ คือ ทางบวก ได้แก่ ชอบ พอใจ เห็นด้วย หรือทางลบ คือ ไม่ชอบ ไม่พอใจ ไม่เห็นด้วย

เจตคติต่ออาชีพครูต่ำ หมายถึงคะแนนที่ได้จากผู้ตอบแบบวัดเจตคติต่ออาชีพครูที่มีคะแนนเรียงลำดับตั้งแต่ต่ำสุดขึ้นไป จำนวน 175 คน

เจตคติต่ออาชีพครูสูง หมายถึง คะแนนที่ได้จากผู้ตอบแบบวัดเจตคติต่ออาชีพครูที่มีคะแนนเรียงลำดับตั้งแต่สูงที่สุดลงมา จำนวน 175 คน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเป็นแบบสอบถาม ซึ่งแบ่งเป็น 2 ตอน ได้แก่

ตอนที่ 1 แบบสำรวจข้อมูลตัวแปรพยากรณ์

ตอนที่ 2 แบบวัดเจตคติต่ออาชีพครู ซึ่งเป็นทั้งแบบเทอร์ส โดนและแบบลิเคิร์ตในฉบับเดียวกัน จำนวน 36 ข้อ

การวิเคราะห์ข้อมูล

ผู้วิจัยนำข้อมูลจากกลุ่มตัวอย่างมาหาค่าสถิติพื้นฐานและวิเคราะห์แบบจำแนกประเภท (Discriminant Analysis) แบบขั้นตอน (Stepwise Method) ด้วยโปรแกรมสำเร็จรูป SPSS (Statistical Package for the Social Science) แล้วนำไปสร้างสมการจำแนกประเภทระหว่างกลุ่มครูที่มีเจตคติต่ออาชีพครูสูงกับกลุ่มครูที่มีเจตคติต่ออาชีพครูต่ำ

สรุปผลการวิจัย

ตัวแปรที่สามารถจำแนกกลุ่มเจตคติต่ออาชีพครูได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 12 ตัว (รวมทั้งตัวแปรหุ่น) ได้แก่ ตำแหน่ง (X2) สถานภาพสมรส (X7) จำนวนบุตร (X10) การอยู่ร่วมกับบิดามารดา (X12) รายได้พิเศษ (X17) การไถ่สองชั้น (X20) วิชาที่สอน (ภาษาไทย)(X23) วิชาที่สอน (การงานและพื้นฐานอาชีพ) (X26) วิชาที่สอน (ทุกวิชาประจำชั้น) (X28) ระดับชั้นที่สอน (ป.5-6)(X31) การสอนตรงตามวิชาเอก (X33) และการสอนตรงตามวิชาที่ ชอบ (X34)

ได้สมการจำแนกประเภทในรูปคะแนนมาตรฐานและคะแนนดิบ ดังต่อไปนี้ $Z_y = .18051Z_2 + .52348Z_7 - .35772Z_{10} - .19880Z_{12} + .18505Z_{17} + .27806Z_{20} - .40289Z_{23} - .17987Z_{26} - .30362Z_{28} - .25557Z_{31} + .27790Z_{33} + .54507Z_{34}$ $Y = .7173436X_2 + 1.469710X_7 - .3841373X_{10} - .3982716X_{12} + .00008817X_{17} + .07484294X_{20} - 1.784566X_{23} - .8599814X_{26} - .6104387X_{28} - .5437844X_{31} + .5739365X_{33} + 1.124536X_{34} - 1.920555$

สมการจำแนกประเภทสามารถจำแนกความเป็นสมาชิกกลุ่มได้ถูกต้อง ร้อยละ 64.86

ข้อเสนอแนะ

ข้อเสนอแนะทั่วไป

1. บางตัวแปรในงานวิจัยนี้มีความสัมพันธ์สูง เช่นอายุ ประสบการณ์ในอาชีพครู และ ระดับเงินเดือน ในการวิจัยประเภทเดียวกันนี้ต่อไปควรให้ระมัดระวังไม่ให้ตัวแปรอิสระมีความสัมพันธ์กันสูง
2. ตัวแปรขนาดโรงเรียนเป็นตัวแปรที่พึงสังเกตในการศึกษาครั้งต่อไป เพราะการทำวิจัยครั้งนี้เก็บรวบรวมข้อมูลกับครูทุกคนในโรงเรียนในกลุ่มตัวอย่าง ฉะนั้นค่าของตัวแปรขนาดโรงเรียนจึงมีค่าเท่ากันในกลุ่มตัวอย่าง โรงเรียนเดียวกัน
3. การศึกษานี้นอกจากจะทราบถึงตัวแปรที่สามารถจำแนกประเภท และลำดับความสำคัญของตัวแปรแล้ว ยังสามารถเป็นแนวทางจะศึกษาต่อไปถึงการทดลองควบคุมตัวแปรบางตัวที่สามารถควบคุมได้ในกลุ่มตัวแปรที่สามารถจำแนกเจตคติต่ออาชีพครูได้ เช่น การสอนตรงตามวิชาที่ชอบ การสอนตรงตามวิชาเอก เป็นต้น นอกจากนี้แล้วควรจะศึกษากระบวนการต่าง ๆ ในการที่จะเปลี่ยนแปลงเจตคติต่ออาชีพครู หากได้ผลการศึกษาทั้งสองส่วนที่ชัดเจนจะเป็นประโยชน์ต่อการศึกษามากยิ่งขึ้น
4. จากผลการศึกษาข้อค้นพบที่น่าจะเป็นประโยชน์ต่อผู้บริหารการศึกษาตลอดจนผู้เกี่ยวข้องเพื่อเป็นแนวทางในการจัดการให้สอดคล้องกับการส่งเสริมเจตคติต่ออาชีพครูที่ดี

ตัวอย่างงานวิจัยที่ 2

การวิเคราะห์ปัจจัยต่าง ๆ ที่มีความสัมพันธ์กับความนิยมดื่มนมหรือไม่ ดื่มนมของ นักเรียนและนักศึกษาในส่วนภูมิภาค

ความเป็นมา

นมเป็นอาหารที่มีคุณค่าของอาหารสูงและมีประโยชน์ต่อการเจริญเติบโตของร่างกาย ส่วนประกอบหลักของนม ประกอบด้วยน้ำ ซึ่งทำหน้าที่เป็นตัวกลาง สำหรับส่วนประกอบอื่นๆ ที่แขวนลอยอยู่ในน้ำนม ไขมัน หมายถึง ไขมันหรือน้ำมันเนย โปรตีน ที่มีคุณภาพสูงเนื่องจากมี กรดอะมิโน (Amino Acid) ที่จำเป็นอยู่อย่างครบถ้วนและแร่ ธาตุต่าง ๆ นอกจากส่วนประกอบหลักเหล่านี้แล้ว ในน้ำนมยังประกอบด้วยเอนไซม์ ฟอสโฟไลปิด และวิตามิน เป็นต้น แม้ว่านม จะเป็นของเหลวแต่ก็ประกอบด้วยของ แข็ง โดยเฉลี่ยประมาณ 13% ได้แก่ โปรตีน ไขมัน แล็กโทส และแร่ธาตุนั้นเอง ซึ่งจะ แตกต่างกันไปในสัตว์แต่ละชนิด และมนุษย์ก็นิยมดื่มนมวัวเป็นสำคัญ น้ำนมนี้ นอกจากจะใช้ดื่มในรูปของน้ำนมสดแล้วยังสามารถนำไปทำผลิตภัณฑ์ ต่าง ๆ ได้อีก เช่น นมผง เนยแข็ง ครีมและพวงนม ผลิตภัณฑ์นมข้นและนมเปรี้ยว เป็นต้น ในที่นี้สนใจ ในประเด็น ของนมพร้อมดื่มเท่านั้น นมพร้อมดื่ม ในที่นี้หมายถึงนมที่นำมาเตรียมสำเร็จ เพื่อใช้ดื่มได้ทันที กระบวนการในการผลิตนมชนิดนี้ ปกติมีอยู่ 3 วิธี คือ

1. กรรมวิธีพาสเจอร์ไรส์ (Pasteurised) เป็นการผ่านความร้อนแล้วต้องทำให้เย็น ลงอย่างรวดเร็ว การให้ความร้อนจะกำหนดอุณหภูมิและเวลาที่เพียงพอจะทำลายเชื้อ โครในนมได้ทั้งหมด และ ทำลายจุลินทรีย์อื่น ๆ ได้เกือบหมด โดยไม่ทำให้คุณสมบัติ ของนม เช่น กลิ่น รส เปลี่ยนแปลงไป มาก จุลินทรีย์ในนมร้อยละ 95-99 จะถูกทำลาย แต่ยังมีจุลินทรีย์พวกแบคทีเรียบางชนิด เช่น แลคโตบาซิลไล (Lactobacilli) สเตรปโตคอคไค (Streptococci) ซึ่งทำให้นมเสียได้ง่าย ยังสามารถทน ความร้อนและเจริญเติบโตต่อไปได้ จึงต้องใช้ความเย็น เพื่อชะงักการเจริญเติบโตของจุลินทรีย์ เหล่านี้ด้วย นม สดที่ได้จะเรียกว่า นมสดพาสเจอร์ไรส์
2. กรรมวิธีสเตอริไลส์ (Sterilized) เป็นการผ่านความร้อน โดยใช้อุณหภูมิสูงเพื่อ ทำลายจุลินทรีย์ใน นมได้ทั้งหมด จึงสามารถเก็บไว้ในอุณหภูมิปกติ (30-40 องศา เซลเซียส) ได้ เป็นเวลานาน กระบวนการมาตรฐานสุข กำหนดอุณหภูมิที่ใช้ในการ สเตอริไลส์ไว้ไม่ให้ต่ำกว่า 100 องศาเซลเซียส

โดยใช้เวลาที่เหมาะสม นมสดที่ได้จะ เรียกว่านมสดสเตอริไลส์ (Sterilized Milk) เนื่องจากการสเตอริไลส์ มีจุดมุ่งหมายเพื่อ ทำลายจุลินทรีย์ในนมทั้งหมด อุณหภูมิและเวลา ที่ใช้จึงทำให้นมสดมีคุณสมบัติ เปลี่ยนแปลงไปได้มาก เช่น นมจะมีกลิ่นและรสชาติขม (Cooked Flavor) การผ่านความร้อนในนมตามกรรมวิธีนี้ จึงนิยมใช้แบบ ยู.เอช.ที.

3. กรรมวิธี ยู.เอช.ที. (Ultrahigh Temperature) เป็นแบบหนึ่งของการผ่านความร้อน โดยใช้ อุณหภูมิสูงมาก จำนวนจุลินทรีย์ในนมที่ถูกทำลายจะขึ้นอยู่กับระดับ อุณหภูมิ และเวลาที่ใช้ ถ้าจุลินทรีย์ถูกทำลายจนเหลือเพียงเล็กน้อย ก็จะได้ นมสดพาสเจอร์ไรส์ ยู.เอช.ที. ซึ่งยังคงต้องเก็บไว้ในความเย็น แต่ถ้าจุลินทรีย์ถูกทำลายจนหมดก็ จะเป็นการสเตอริไลส์แบบ ยู.เอช.ที. ซึ่งเป็นการผ่านความร้อนตามกรรมวิธี สเตอริไลส์ ที่นิยมใช้กันมากในการผลิตนมสด เพราะทำให้นมสดที่ได้ ยังคงคุณสมบัติ เดิมไว้ได้ เช่นเดียวกับนมสดพาสเจอร์ไรส์ แต่สามารถเก็บไว้ได้นานกว่าใน อุณหภูมิปกติ

ในปัจจุบันปรากฏว่าอัตราการดื่มนมของประชากรในประเทศไทยมีเพียง 2 ลิตร/คน/ปี ซึ่งมีอัตราที่ต่ำมากเมื่อเทียบกับประเทศที่พัฒนาแล้วและต่ำกว่าประเทศเพื่อนบ้านใกล้เคียง เช่น ฟิลิปปินส์ อินโดนีเซีย มาเลเซีย และสิงคโปร์ ทำให้รัฐบาลและเอกชน ตระหนักถึงความร่วมมือ เพื่อที่จะส่งเสริมและหาแนวทางในการจัดทำโครงการ ธรรมรงค์ เพื่อการบริโภคนม ซึ่งคณะรัฐมนตรีอนุมัติให้แต่งตั้งคณะกรรมการดังกล่าว เมื่อวันที่ 19 กุมภาพันธ์ 2528 โดยมี รัฐมนตรีประจำสำนักนายกรัฐมนตรี (นายกระมล ทองธรรมชาติ) เป็นประธานกรรมการ รัฐมนตรีช่วยว่าการกระทรวงเกษตรและสหกรณ์ (นายประยูทธ ศิริพานิชย์) เป็นรองประธานกรรมการ มีผู้แทนจากส่วน ราชการ รัฐวิสาหกิจ บริษัทผู้ผลิตนมแปรรูป และ จำหน่ายผลิตภัณฑ์นม สหกรณ์โคนม เกษตรกรและผู้ทรงคุณวุฒิ เป็นกรรมการ ผู้อำนวยการ อสค. เป็นกรรมการและ เลขานุการ คณะกรรมการดังกล่าวนี้ มีวัตถุประสงค์ 3 ประการ คือ

1. ทำการธรรมรงค์รูปแบบ เพื่อให้ประชาชน โดยเฉพาะเด็กและเยาวชนได้ บริโภคนมอย่างกว้างขวางและต่อเนื่อง เพื่อพัฒนาคุณภาพชีวิตทั้งร่างกายและกำลัง สมอง เนื่องจากนมมีคุณค่าอาหารสูง

2. เพื่อสนับสนุนนโยบายการส่งเสริมการเลี้ยงโคนมของรัฐบาล โดยสร้างฐาน การตลาดรองรับน้ำนมดิบของเกษตรกร อันจะสร้างความมั่นคงให้กับอาชีพการเลี้ยง โคนม

3. เพื่อดำเนินการให้มีการลดต้นทุนการผลิต การแปรรูป และการจำหน่าย ผลิตภัณฑ์นม เพื่อให้ผู้บริโภคสามารถซื้อนมได้ในราคาที่เหมาะสมและยุติธรรม

แนวคิดทฤษฎี

1. ขั้นตอนและวิธีการเก็บข้อมูล

การเก็บรวบรวมข้อมูล กำหนดแผนแบบการสุ่มตัวอย่าง โดยให้ประชากรกลุ่ม เป้าหมาย

ที่จะทำการศึกษา คือ

- 1.1 นักเรียนชั้นมัธยมศึกษาปีที่ 1-3 และ 4-6 จากโรงเรียนต่าง ๆ ในเขต-เทศบาลเมืองขอนแก่น
- 1.2 นักศึกษามหาวิทยาลัยขอนแก่น

ใช้วิธีการสุ่มตัวอย่างแบบง่าย (Simple Random Sampling) โดยกำหนดขนาด ตัวอย่างไว้ทั้งหมด จำนวน 583 คน การรวบรวมข้อมูลใช้วิธีสัมภาษณ์ โดยใช้แบบสอบถามที่ผ่านการทดสอบเรียบร้อยแล้ว สำหรับพนักงานสัมภาษณ์ ใช้นักศึกษาวิชา เอกสถิติ ชั้นปีที่ 3 คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

2. เทคนิคการวิเคราะห์จำแนกประเภท

การวิเคราะห์ปัจจัยที่มีผลต่อการดื่มนมพร้อมดื่มของประชากรกลุ่มเป้าหมาย ได้แก่ นักเรียนม.1-ม.3, ม.4-ม.6 และนักศึกษา โดยใช้เทคนิควิเคราะห์จำแนกประเภท โดยแบ่งประชากรกลุ่มเป้าหมายออกเป็น 2 กลุ่ม คือ กลุ่มที่ดื่มนมพร้อมดื่ม และกลุ่ม ที่ไม่ดื่มนมพร้อมดื่ม (รวมถึงพวกที่เคยดื่มนมแต่ปัจจุบันเลิกดื่มแล้วด้วย) ให้ภาวะการ ดื่มนมพร้อมดื่ม เป็นตัวแปรตาม ส่วนตัวแปรอิสระที่นำมาวิเคราะห์ มีจำนวน 8 ตัว คือ

1. อายุ
2. รายได้
3. ทักษะคิดเกี่ยวกับนมพร้อมดื่ม
4. เพศ
5. ภูมิลำเนา
6. เครื่องดื่มประเภททดแทน
7. เครื่องดื่มประเภทคู่แข่งอื่น
8. เครื่องดื่มประเภทอื่น ๆ

วัตถุประสงค์

1. เพื่อทำการวิจัยตลาดนมของนักเรียนและนักศึกษา ในส่วนภูมิภาค
 2. เพื่อทำการวิเคราะห์ปัจจัยต่าง ๆ ที่มีความสัมพันธ์กับความนิยมดื่มนมหรือไม่ ดื่มนมของนักเรียนและนักศึกษาในส่วนภูมิภาค โดยสร้างสมการจำแนกประเภท (Discriminant Function) จากเทคนิคทางสถิติที่เรียกว่า การวิเคราะห์จำแนกประเภท (Discriminant Analysis)
- เทคนิคการวิเคราะห์จำแนกประเภท เป็นเทคนิคทางสถิติเกี่ยวกับการวิเคราะห์ตัวแปรหลายตัว (Multivariate Analysis) ที่มีวัตถุประสงค์จะคัดเลือกตัวแปร (ปัจจัย) กลุ่ม หนึ่งหรือชุดหนึ่งที่คาดว่า จะมีความสัมพันธ์กับสิ่งที่ต้องการจะศึกษาจนสามารถแยก ประชากรออกเป็นกลุ่มต่าง ๆ ได้ เช่น กลุ่มที่ดื่มนมกับกลุ่มที่ไม่ดื่มนม เป็นต้น โดย การสร้างสมการจำแนกประเภท ซึ่งสมการดังกล่าวจะสามารถระบุ ว่า กรณีใดจะเป็น สมาชิกของกลุ่มใดได้ โดยอาศัยตัวแปรและค่าสัมประสิทธิ์ต่าง ๆ ที่

ปรากฏอยู่ใน สมการจำแนกประเภท เป็นเครื่องมือช่วยระบุจำแนกประเภท ความสามารถของ สมการในการจำแนกประเภท อาจทดสอบได้จากการจำแนกประเภทของกรณีที่เป็นสมาชิก ของกลุ่มใด และใช้สมการเปรียบเทียบว่า ตรงกับที่เป็นจริงหรือไม่ เทคนิคทางสถิติดังกล่าว จึง เหมาะสมในการวิเคราะห์ปัจจัยที่มีความสัมพันธ์กับ ความนิยมบริโภคขนมของกลุ่มประชากร เป้าหมายดังกล่าว ทั้งนี้เนื่องจากปัจจัยที่เกี่ยวข้องกับทั้ง 2 กลุ่ม อาจจะไม่เหมือนกัน หรือหาก เหมือนกัน น้ำหนักของปัจจัยต่าง ๆ ของแต่ละตัว ก็อาจจะแตกต่างกัน การวิเคราะห์จำแนกประเภท จึงสามารถนำมา ประยุกต์ใช้ได้เหมาะสมเกี่ยวกับเรื่องดังกล่าว

ประชากรและกลุ่มตัวอย่าง

1. **กลุ่มนักเรียน ม.1-ม.3** การศึกษาค่าเฉลี่ยของตัวแปร ของกลุ่มนักเรียน ม.1-ม.3 แสดงให้เห็นว่า กลุ่มที่ ดื่มนมและไม่ดื่มนม มีความแตกต่างกันในเรื่องของรายได้ที่ได้รับจากผู้ปกครอง และ อัตราการดื่ม เครื่องดื่มประเภททดแทน ส่วนตัวแปรอื่น ๆ ไม่แตกต่างกันนัก พิจารณาอิทธิพล ของตัวแปรต่าง ๆ จากขนาดของสัมประสิทธิ์คะแนนมาตรฐาน พบว่า มีเพียง 3 ตัว ที่สามารถนำมา สร้างสมการจำแนกประเภทได้ เรียงตามลำดับ ความสำคัญจากมากไปน้อยคือทัศนคติภูมิลำเนา และการดื่มเครื่องดื่มประเภททดแทน ซึ่งมีความสัมพันธ์เชิงบวกกับภาวะการดื่มนมพร้อมดื่ม กล่าวคือ ผู้ที่มีทัศนคติที่ดีต่อ นมพร้อมดื่ม มีความรู้ความเข้าใจและเห็นคุณค่าของนมพร้อมดื่ม เป็น ผู้ที่มีภูมิลำเนา อยู่ในเขตเทศบาลและสุขาภิบาลและเป็นผู้ที่ดื่มเครื่องดื่มประเภททดแทน ได้แก่ นมเปรี้ยว นมถั่วเหลือง และเครื่องดื่มประเภทอาหาร เป็นประจำ จะเป็นผู้ที่ดื่มนมพร้อม ดื่มสูงกว่า เมื่อ พิจารณาอำนาจในการจำแนกของสมการที่ได้ ปรากฏว่า สมการมีนัยสำคัญ ทาง สถิติที่ระดับ 0.0003 ค่า Canonical Correlation ต่ำ ในขณะที่ ค่า Wilk's Lamda สูง (0.9125) แสดงว่า อำนาจใน การจำแนกกลุ่มของสมการนี้มีไม่มากนัก อย่างไรก็ตาม ผลของการคาดคะเนความเป็นสมาชิกของ กลุ่ม มีความสามารถ ทำนาย ได้ถูกต้อง ร้อยละ 73.1

2. **กลุ่มนักเรียน ม.4-ม.6** การศึกษาค่าเฉลี่ยของตัวแปรของกลุ่มนักเรียน ม.4-ม.6 แสดงให้ เห็นว่า กลุ่มที่ดื่มนมและไม่ดื่มนม มีความแตกต่างกันในเรื่องรายได้ที่ได้รับจากผู้ปกครอง ซึ่งกลุ่ม ที่ไม่ดื่มนมกลับได้รับรายได้มากกว่า และอัตราการดื่มเครื่องดื่มประเภททดแทน ส่วนตัวแปรอื่น ๆ ไม่แตกต่างกันนัก พิจารณาอิทธิพลของตัวแปรต่างๆ จากขนาดของสัมประสิทธิ์คะแนนมาตรฐาน พบว่า มีเพียง 2 ตัว ที่สามารถนำมาสร้างสมการจำแนกประเภทได้ เรียงตามลำดับ ความสำคัญจาก มากไปน้อย คือ ทัศนคติและเพศ โดยทัศนคติมีความสัมพันธ์เชิงบวก กับภาวะการดื่มนม ในขณะที่ เพศ มีความสัมพันธ์เชิงลบ เมื่อพิจารณาอำนาจในการจำแนกของสมการที่ได้ ปรากฏว่า สมการไม่มี นัย สำคัญที่ 0.05 (มีนัยสำคัญที่ 0.228) ค่า Canonical Correlation ต่ำมาก ในขณะที่ ค่า Wilk's Lamda สูงมาก แสดงว่า จำนวนในการจำแนกกลุ่มของสมการนี้ มีไม่มากนัก ซึ่งมีความสามารถทำ นายได้ถูกต้องเพียงร้อยละ 61.0 เท่านั้น

3. **กลุ่มนักศึกษา** การศึกษาค่าเฉลี่ยของตัวแปร ของกลุ่มนักศึกษา แสดงให้เห็นว่า กลุ่มที่ดื่ม

นม และไม่ดื่มนมมีความแตกต่างกันในเรื่องของรายได้ที่ได้รับจากผู้ปกครอง อัตราการดื่มนม เครื่องดื่มประเภท ทดแทนและนมมีไขมัน ส่วนตัวแปรอื่นๆ ไม่แตกต่างกันนัก พิจารณาอิทธิพลของตัวแปรต่าง ๆ จากขนาดสัมประสิทธิ์คะแนนมาตรฐาน พบว่า มี เพียง 2 ตัว ที่สามารถนำมาสร้างสมการจำแนกประเภทได้ เรียงตามลำดับความสำคัญจากมากไปน้อย คือ นมมีไขมัน ที่มีความสัมพันธ์เชิงบวก และอัตราการดื่มนม เครื่องดื่มประเภททดแทนที่มีความสัมพันธ์เชิงลบ เมื่อพิจารณาอำนาจในการจำแนกของสมการที่ได้ ปรากฏว่า สมการมีนัย สำคัญที่ 0.0361 มีค่า Canonical Correlation ต่ำ ในขณะที่ ค่า Wilk's Lamda สูง แสดง ว่า อำนาจในการจำแนกกลุ่มของสมการนี้ มีไม่มากนัก ผลการคาดคะเนทำนายได้ถูก ต้องร้อยละ 69.6

สรุป ปัจจัยที่มีอิทธิพลต่อการดื่มนมพร้อมดื่มของกลุ่มเป้าหมายโดยใช้เทคนิค การวิเคราะห์ จำแนกประเภท ปรากฏว่า การมีทัศนคติที่ดีต่อนมพร้อมดื่ม เป็นปัจจัยที่ สำคัญที่สุด สำหรับ นักเรียน ม.1-ม.3 และนักเรียน ม.4-ม.6 แสดงว่า การเผยแพร่ ความรู้ ความเข้าใจเกี่ยวกับนมพร้อมดื่มต่อกลุ่มนักเรียน ยังมีความจำเป็นที่ต้องกระทำ อย่างต่อเนื่องไป สำหรับกลุ่มนักศึกษาปัจจัยนี้ไม่มีผล เนื่องจากมีความรู้ ความเข้าใจ เกี่ยวกับนมพร้อมดื่มเป็นอย่างดีแล้ว ส่วนปัจจัยเรื่องอัตราการดื่มนม เครื่องดื่มประเภท ทดแทน ได้แก่ นมเปรี้ยวนมถั่วเหลืองนั้น สำหรับนักเรียน ม.1-ม.3 จะมีผลเชิงบวก แต่ในกลุ่มนักศึกษา กลับมีผลเชิงลบ ซึ่งเป็นไปได้ว่า นักเรียน ม.1-ม.3 ได้ดื่มนม พร้อมดื่มและเครื่องดื่มประเภททดแทนควบคู่กันไป เนื่องจากเป็นสินค้าที่มีคุณภาพ และรสชาติใกล้เคียงกัน แต่ในกลุ่มนักศึกษา เมื่อดื่มนม เครื่องดื่มประเภททดแทนแล้ว จะ นิยมดื่มและติดเป็นนิสัยจนไม่ดื่มนมพร้อมดื่ม ส่วนปัจจัยนมมีไขมัน มีผลเชิงบวกต่อ การดื่มนมพร้อมดื่มในกลุ่มนักเรียน ม.1-ม.3 และ นักศึกษาซึ่งชี้ให้เห็นว่า ผู้ที่อยู่นอก เขตเทศบาลและสุขาภิบาล จะเป็นผู้ที่ไม่ดื่มนมพร้อมดื่มสูงกว่า ผู้ที่อยู่ในเขตเทศบาล และสุขาภิบาล ปัจจัยเรื่องเพศ ไม่มีผลต่อการดื่มนมพร้อมดื่ม ในกลุ่มนักเรียน ม.1-ม .3 และนักศึกษา ยกเว้นในกลุ่มนักเรียน ม.4-ม.6 ซึ่งมีผลในเชิงลบหมายความว่า นัก เรียนชายจะเป็นผู้ที่ไม่ดื่มนมมากกว่านักเรียนหญิง

ข้อเสนอแนะ

1. ควรมีการเผยแพร่ความรู้ ความเข้าใจเกี่ยวกับนมพร้อมดื่ม อย่างต่อเนื่องกันต่อไปตามสื่อโฆษณาต่าง ๆ โดยเฉพาะอย่างยิ่งทางโทรทัศน์ และหนังสือพิมพ์ ซึ่งนอกจากเป็นการเผยแพร่ให้เยาวชนได้รับทราบข้อเท็จจริงเพิ่มเติมแล้ว ยังมีผลต่อผู้ ปกครอง จะทำให้ค่าผลต่อการแนะนำให้เยาวชนในความปกครองดื่มนมพร้อมดื่ม เนื่องจากพบว่า ผู้ปกครอง ยังมีอิทธิพลต่อการแนะนำให้ดื่มนม โดยเฉพาะอย่างยิ่งใน กลุ่มนักเรียน

2. แม้ว่าในปัจจุบัน กลุ่มเป้าหมายจะหันมาดื่มนมพร้อมดื่มกันในอัตราที่สูง แต่ยังมีปัจจัยอื่นที่จะมีผลกระทบต่อ การส่งเสริมการดื่มนมในอนาคต คือ ปัจจัยด้านราคา ควรหาทางที่จะลดต้นทุนค่าใช้จ่ายด้านต่าง ๆ เช่น กล่องเตตราแพค เพื่อลดราคาของนมพร้อมดื่มลง อย่างน้อยที่สุดควรมีราคาใกล้เคียงกับเครื่องดื่มประเภทน้ำอัดลม มาตรการดังกล่าวนี้ ควรมีการวางแผนร่วมกันทั้ง

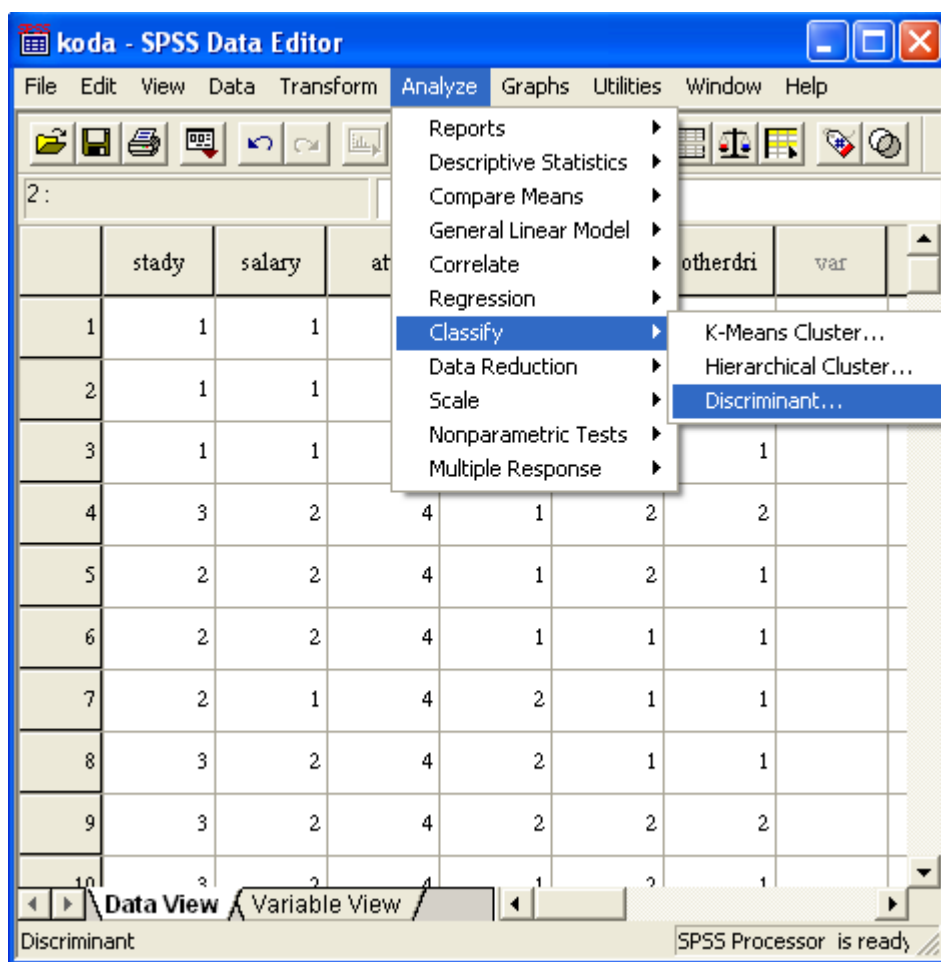
ภาครัฐและเอกชน

3. ควรพิจารณากับกลุ่มเยาวชนที่มีภูมิลำเนานอกเขตเทศบาลหรือสุขาภิบาล ด้วย ว่า จะมีวิธีการส่งเสริมการค้มนมพร้อมดื่มได้อย่างไร วิธีการหนึ่งที่น่าจะทำได้คือ ให้ทางโรงเรียนมีส่วนร่วมด้วย

4. ควรพิจารณาร่วมมือหรือส่งเสริมให้บริษัทผู้ผลิต จำหน่ายนมพร้อมดื่ม จัด จำหน่ายนมพร้อมดื่มในราคาพิเศษหรือมีส่วนลดแก่สถานศึกษาต่างๆ ในรูปของการ ขายตรงที่สถานศึกษานั้น ๆ

คำสั่งของ SPSS for Windows สำหรับการวิเคราะห์จำแนกกลุ่ม

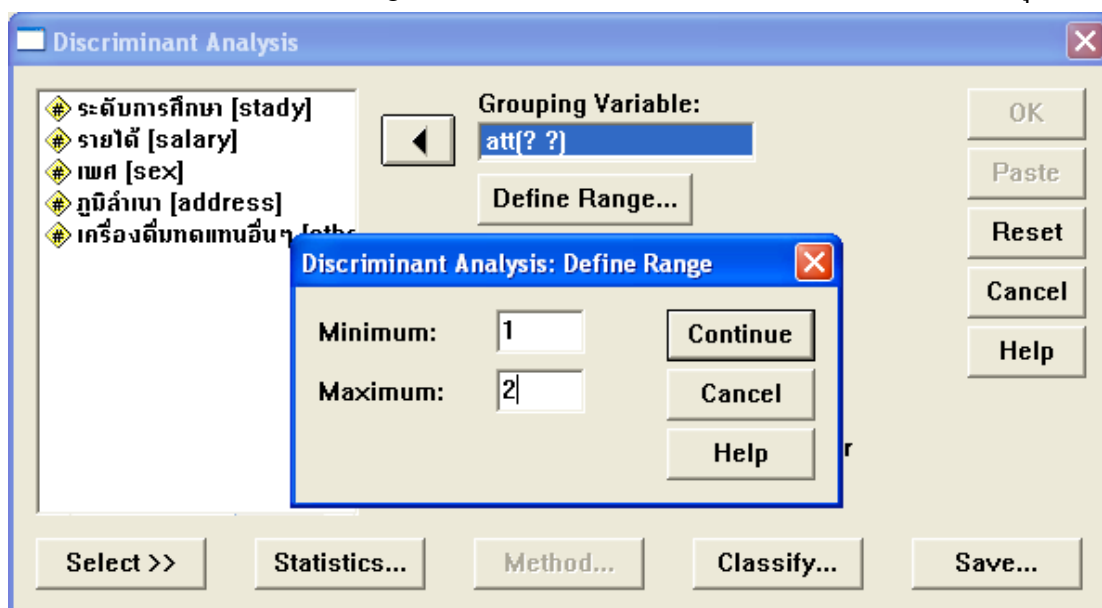
Analyze ⇒ Classify ⇒ Discriminant ...



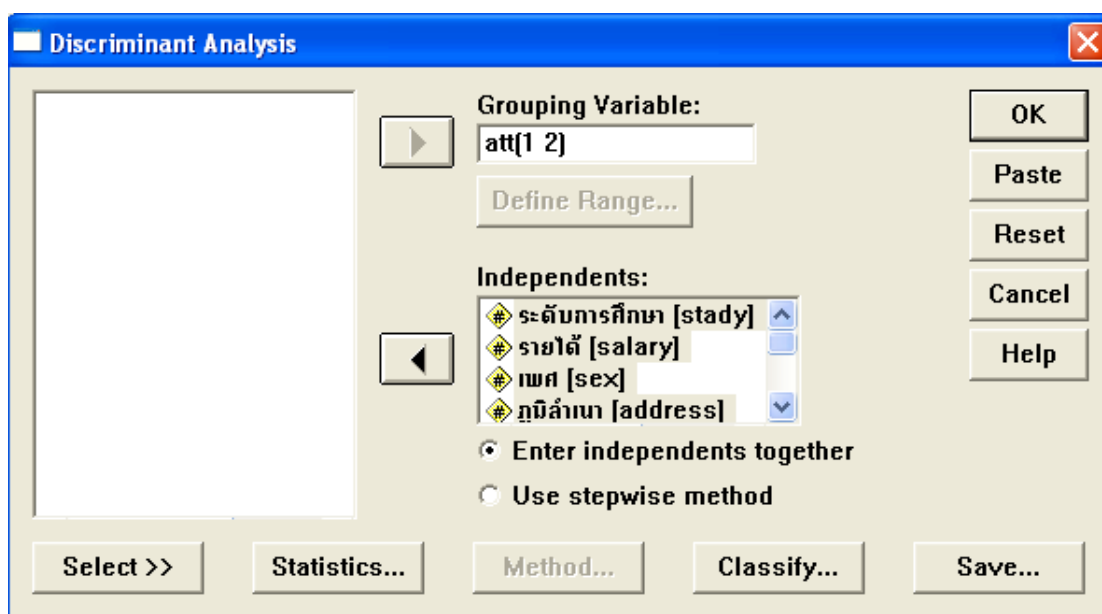
ผู้ใช้งานต้อง

1. เลือก **Define Range** ตัวแปรเชิงกลุ่ม 1 ตัวใส่ใน box ของ Grouping Variable เช่น ประเภทของลูกหนี้

- คลิก Define Range ... ใส่ค่า Minimum และ Maximum ของตัวแปรเชิงกลุ่มดังนี้



2. เลือกตัวแปรอิสระอย่างน้อย 1 ตัว ใส่ใน box ของ Independents โดยที่ตัวแปรดังกล่าวต้องเป็นชนิด numeric เช่น อายุ รายได้ เป็นต้น ดังภาพ



3. เลือก **Select** >>

- จะพบว่าคล้ายกับหน้าจอ Discriminant Analysis แต่เพิ่มในส่วนล่างคือ เพิ่ม Selection Variable box เมื่อผู้ใช้เลือก **Select** >> หมายถึง ผู้ใช้ต้องการวิเคราะห์บางส่วนของเพิ่มข้อมูลไม่ใช่ทั้งเพิ่ม เช่น เลือกวิเคราะห์จำแนกกลุ่มลูกหนี้นี้เฉพาะเพศชาย ก็จะต้องเลือกตัวแปร Sex ใช้ใน box ของ Selection Variable เลือกคลิกที่ปุ่ม **Value ...**

- ในหน้าจอ Value ... ให้ใส่ค่าข้อมูลตัวแปรที่อยู่ใน box ของ Select for Variable เช่น ใส่ค่าเพศชาย = 1, หญิง = 2 จึงต้องใส่ 1 ใน box ของ Value ดังนั้นตัวแปรที่ใส่ใน box ของ Selection Variable ควรเป็นตัวแปรเชิงกลุ่ม

4. เลือก **Statistics** ประกอบด้วย 3 ส่วนใหญ่ ๆ ดังนี้

ส่วนที่ 1 : Descriptives ผู้ใช้สามารถเลือกทางเลือกต่อไปนี้อย่างน้อย 1 ทางเลือก

Mean จะแสดงค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐานของตัวแปรอิสระแต่ละตัวของแต่ละกลุ่มเช่น เรื่องสถานภาพหนี้ 2 กลุ่ม จะคำนวณหารายได้เฉลี่ย, รายได้เบี่ยงเบน มาตรฐาน, ปริมาณหนี้เบี่ยงเบนมาตรฐานของแต่ละกลุ่ม และของรวมทั้ง 2 กลุ่ม

Univariate ANOVAs เป็นการหาค่าสถิติทดสอบ F ใน 1-WAY ANOVA ในการทดสอบความเท่ากันของค่าเฉลี่ยของตัวแปรอิสระแต่ละตัวเมื่ออยู่ต่างกลุ่มกัน เช่น กรณีที่มีตัวแปรอิสระ 2 ตัวจะต้องทำการทดสอบ 2 ครั้ง ดังนี้

1) H_0 : รายได้เฉลี่ยของกลุ่มหนี้ปกติ = รายได้เฉลี่ยของกลุ่มหนี้มีปัญหา

H_1 : รายได้เฉลี่ยของกลุ่มหนี้ปกติ \neq รายได้เฉลี่ยของกลุ่มหนี้มีปัญหา

2) H_0 : ปริมาณหนี้เฉลี่ยของกลุ่มหนี้ปกติ = ปริมาณหนี้เฉลี่ยของกลุ่มหนี้มีปัญหา

H_1 : ปริมาณหนี้เฉลี่ยของกลุ่มหนี้ปกติ \neq ปริมาณหนี้เฉลี่ยของกลุ่มหนี้มีปัญหา

และยังให้ค่าเฉลี่ยทดสอบ Wilks' Lambda

Box's M จะให้ค่าสถิติ Box's M ของค่าระดับนัยสำคัญของการทดสอบความเท่ากันของ Variance-Covariance matrix ของแต่ละกลุ่ม $H_0 : \Sigma_1 = \Sigma_2 = \dots \Sigma_k$

ส่วนที่ 2 : Function Coefficients ผู้ใช้สามารถเลือกทางเลือกต่อไปนี้

Fisher's จะแสดงค่าสัมประสิทธิ์จำแนกกลุ่ม หรือสัมประสิทธิ์ความถดถอย b_1, \dots, b_p ของแต่ละกลุ่ม (Classification Function coefficients)

Unstandardized จะแสดงว่า Discriminant Function Coefficients

ส่วนที่ 3 : Matrices ผู้ใช้สามารถเลือกให้แสดง matrices ต่อไปนี้

Within-groups correlation จะแสดงค่าของ Pooled Within-groups Correlation

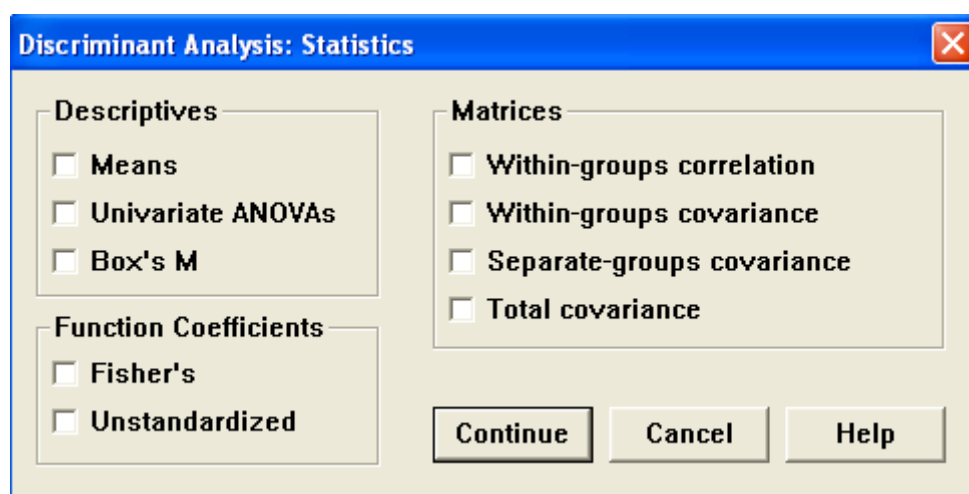
Matrix ขนาด $p \times p$ ของตัวแปรอิสระ p ตัว

- Within-groups covariance** จะแสดงค่าของ Pooled Within-groups Covariance

Matrix ขนาด $p \times p$ ของตัวแปรอิสระ p ตัว

- Swparate-groups covariance** จะแสดงค่า Covariance Matrix ขนาด $p \times p$ ของตัวแปรอิสระ p ตัว

- Total covariance** จะแสดงค่า Covariance Matrix ขนาด $p \times p$ ของตัวแปรอิสระ p ตัว โดยคำนวณจากทุกcaseซึ่งถือเสมือนว่ามีกลุ่มเดียว ดังภาพ



5. เลือก **Classify ...** ประกอบด้วยส่วนต่าง ๆ 5 ส่วน ดังนี้

ส่วนที่ 1 : Prior Probabilities เป็นการกำหนดความน่าจะเป็นเบื้องต้นในการคำนวณหา

Posterior probabilities ผู้ใช้ต้องเลือกทางเลือกใดทางเลือกหนึ่งจากทางเลือกต่อไปนี้

All group equal เป็นการกำหนดความน่าจะเป็นเบื้องต้นที่แต่ละ case จะอยู่ในกลุ่มใดกลุ่มหนึ่ง ทางเลือกนี้กำหนดให้เท่ากัน เช่น ถ้ามี 3 กลุ่ม ความน่าจะเป็นเบื้องต้นที่ case จะอยู่ในกลุ่มใดกลุ่มหนึ่ง = $1/3$

Compute from group sizes เป็นการให้คำนวณความน่าจะเป็นเบื้องต้นที่ case จะอยู่กลุ่มใดกลุ่มหนึ่งโดยใช้ข้อมูลจากตัวอย่าง เช่น จากการสุ่มตัวอย่างลูกหนี้ 2 สถานภาพ พบว่า ได้ลูกหนี้ปกติมา 40 % และลูกหนี้มีปัญหา 60 % ดังนั้นความน่าจะเป็นเบื้องต้นหรือ P (กลุ่มลูกหนี้ปกติ) = .4 และ P (ลูกหนี้มีปัญหา) = .6

การนำ Prior Probability ไปคำนวณหา Posterior Probability

ส่วนที่ 2 : Use Covariance Matrix เป็นการเลือกชนิดของ Covariance Matrix ที่จะใช้ในการจำแนกกลุ่ม case หรือใช้ใน Discriminant Functions โดยผู้ใช้งานจะต้องเลือกชนิดใดชนิดหนึ่งจากทางเลือกต่อไปนี้

Within-groups หมายถึง จะใช้ Pooled Within-groups Covariance Matrix ในการจำแนกกลุ่ม case

Separate-groups หมายถึง จะใช้ Separate-group Covariance Matrices ในการจำแนกกลุ่ม cases

ส่วนที่ 3 : Display เป็นการเลือกให้แสดงผลของการการจัดกลุ่มแต่ละ case โดยผู้ใช้สามารถเลือกได้อย่างน้อย 1 ทางเลือกต่อไปนี้

Casewise results จะใช้ตาราง Casewise Statistics ซึ่งในตารางจะแสดงกลุ่มที่แท้จริง และกลุ่มที่พยากรณ์โดยที่ใช้สมการจำแนกกลุ่มพร้อมทั้งแสดงค่า Posterior Probability และ Discriminant Score (D) ของแต่ละ case นอกจากนี้ยังสามารถเลือกได้ว่าจะให้แสดงทุก case หรือเฉพาะบางส่วน เช่น เฉพาะ 30 case แรก กรณีที่เลือกที่จะแสดงบางส่วน ให้เลือก Limit case to first แล้วใส่จำนวน case ใน box

Summary Table จะแสดงตาราง Classification results ซึ่งสรุปเปอร์เซ็นต์ของการจัดกลุ่มถูกและการจัดกลุ่มผิด

Leave-one-out Classification จากตาราง Classification Results ซึ่งได้จากการเลือก Summary table ข้างต้น ซึ่งแสดงเปอร์เซ็นต์ของการจัดกลุ่มและจัดกลุ่มผิด ซึ่งทำโดยการใช้อข้อมูลตัวอย่าง 1 ชุด ในการสร้างฟังก์ชันการจำแนกกลุ่ม แล้วใช้ฟังก์ชันการจำแนกกลุ่มที่ได้จัดข้อมูลตัวอย่างชุดเดิมว่าควรอยู่ในกลุ่มใด แล้วจึงเปรียบเทียบกลุ่มจริงที่แต่ละ case อยู่กับกลุ่มที่พยากรณ์โดยใช้ฟังก์ชันการจำแนกกลุ่มที่ สร้างขึ้น เพื่อคำนวณเปอร์เซ็นต์ความถูกต้องในการจัดกลุ่ม วิธีดังกล่าวข้างต้นอาจจะไม่ถูกต้อง ดังนั้น ถ้าเลือก Leave-one-out classification จะทำดังนี้

- จากตัวอย่าง 1 ชุด ขนาด n case ให้ใช้ข้อมูล n-1 case ในการสร้างฟังก์ชันจำแนกกลุ่ม ส่วนที่ case ที่เหลืออีก 1 Case ไว้ใช้ในการตรวจสอบความถูกต้องในการจัดกลุ่ม โดยมีขั้นตอนดังนี้

- การคำนวณครั้งที่ 1 จะใช้ข้อมูล case ที่ 2- n ในการสร้างฟังก์ชันจำแนกกลุ่มและใช้ case ที่ 1 ในการตรวจสอบความถูกต้องในการพยากรณ์กลุ่ม

- การคำนวณครั้งที่ 2 จะใช้ข้อมูล case ที่ 1 และ 3- n ในการสร้างฟังก์ชันจำแนกกลุ่ม แล้วใช้ case ที่ 2 ในการตรวจสอบความถูกต้องในการพยากรณ์กลุ่ม

- การคำนวณครั้งที่ n จะใช้ข้อมูล case ที่ 1 ถึง n - 1 ในการสร้างฟังก์ชันจำแนกกลุ่ม แล้วใช้ case ที่ n ในการตรวจสอบ

- จึงต้องมีการคำนวณทั้ง n ครั้งแล้วจึงจะคำนวณหาเปอร์เซ็นต์ความถูกต้องในการจัดกลุ่ม

ส่วนที่ 4 : Plots ผู้ใช้สามารถเลือกใช้ SPSS แสดงกราฟดังต่อไปนี้

Combined-groups จะแสดง scatterplot ของทุกกลุ่มในกราฟรูปเดียวกัน โดยที่

- กรณีที่มีจำนวนฟังก์ชันจำแนกกลุ่มมากกว่า 2 ฟังก์ชัน จะพล็อตระหว่าง

ฟังก์ชันจำแนกกลุ่ม (Discriminant function) 2 ฟังก์ชันแรก

- กรณีที่มีเพียง 2 กลุ่ม หรือฟังก์ชันกลุ่มจำแนกฟังก์ชันเดียว จะแสดง histogram

Separate-groups จะแสดง scatter ของแต่ละกลุ่มแยกกัน 1 กลุ่ม 1 รูป โดยที่

- กรณีที่มีจำนวนฟังก์ชันจำแนกกลุ่มมากกว่า 2 ฟังก์ชัน จะพล็อตระหว่าง 2

ฟังก์ชันแรก

- กรณีที่มี 2 กลุ่ม หรือมีฟังก์ชันจำแนกกลุ่มฟังก์ชันเดียว จะแสดง histogram

ของแต่ละกลุ่มแทน

Territorial map เป็นกราฟที่แสดงจุดกลางของกลุ่ม และขอบเขตของแต่ละกลุ่ม ดังแสดงในรูป กรณีที่มีเพียง 2 กลุ่มจะไม่แสดงกราฟนี้ให้

ส่วนที่ 5 : Replace missing values with mean เมื่อเลือกทางเลือกรนี้ หมายความว่า ในการจัดกลุ่ม ถ้ามีค่า missing สำหรับตัวแปรอิสระ จะใช้ค่าเฉลี่ยของตัวแปรนั้นแทนค่าที่ missing จึงสามารถจัดกลุ่มให้กับ case ที่มีค่า missing ได้ ดังภาพ

6. เลือก **Save ...** จะได้หน้าจอให้ผู้เลือกใช้ค่าที่จะ save ได้ในแฟ้มเดิม โดยเป็นตัวแปรใหม่ไว้ที่ท้ายแฟ้ม

Predicted group membership ถ้าเลือกทางเลือกรนี้ จะเป็นการบันทึกกลุ่มที่พยากรณ์ โดยพิจารณาว่ากลุ่มใดมีค่า Posterior Probabilities มากกว่ากลุ่มอื่น ๆ จะถือว่า case อยู่ในกลุ่มนั้น โดยจะบันทึกไว้ในตัวแปรชื่อ dis..

Discriminant score จะแสดงค่า (D') (Discriminant score) ของแต่ละ case อยู่ในกลุ่มนั้น โดยจะบันทึกไว้ในตัวแปรชื่อ dis_1

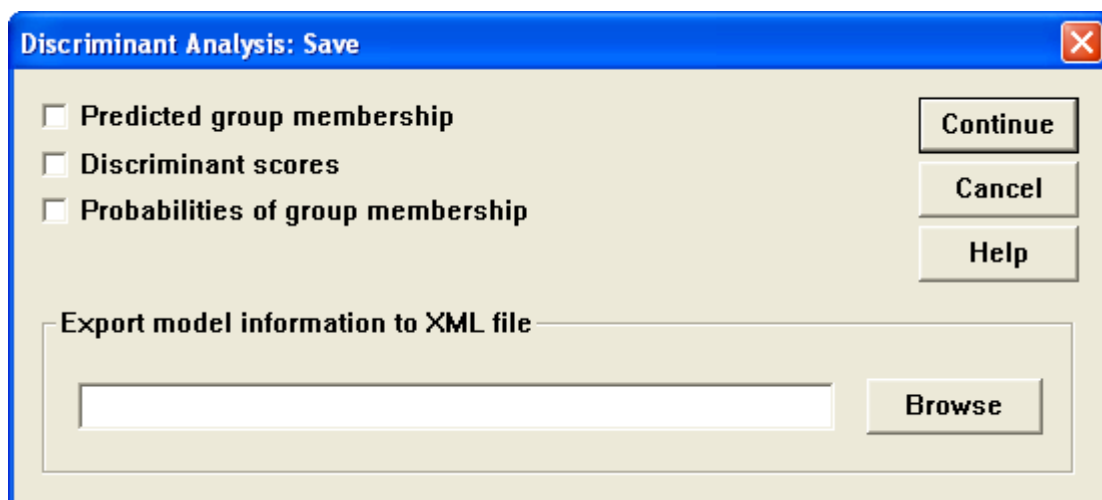
Probabilities of group membership จะแสดง Posterior Probabilities ที่แต่ละ case จะอยู่

ในกลุ่มที่ 1 จนถึงกลุ่มที่ k โดยจะบันทึกไว้ในตัวแปรชื่อ dis1_2 สำหรับกลุ่มที่ 1 และ dis2_2 สำหรับกลุ่มที่ 2 กรณีที่มี 2 กลุ่มนั่นคือ จะได้จำนวนตัวแปรเท่ากับจำนวนกลุ่ม

จากหน้าจอ มีวิธีการเลือกตัวแปรอิสระเข้าในสมการหรือฟังก์ชันจำแนกกลุ่ม โดยผู้ใช้ต้องเลือกทางเลือกหนึ่งทางเลือกเดียวจากต่อไปนี้

Enter independents together หมายถึง การนำตัวแปรอิสระทั้งหมด (p ตัว) เข้าในสมการจำแนกกลุ่ม

Use stepwise method เลือกวิธี Stepwise เป็นวิธีการเลือกตัวแปรอิสระเข้าสมการจำแนกกลุ่ม ดังภาพ



กรณีที่มีเงื่อนไขของเทคนิคการจำแนกกลุ่มไม่เป็นจริง

ดังที่ได้กล่าวมาแล้วว่า เทคนิค discriminant มีเงื่อนไขเกี่ยวกับตัวแปรอิสระ 2 ข้อ คือ

1. ตัวแปรอิสระของแต่ละกลุ่มต้องมีการแจกแจงแบบ Multivariate Normal
2. Variance-covariance Matrix ของแต่ละกลุ่มต้องเท่ากัน

ถ้าเงื่อนไขทั้ง 2 เงื่อนไขข้างต้นเป็นจริง แสดงว่าสัมประสิทธิ์ (b_1, b_2, \dots, b_p) ที่ได้จะทำให้โอกาสที่จะจัดกลุ่มผิดมีค่าต่ำสุด

- กรณีที่มีเงื่อนไขข้อ 2 ไม่จริง แต่ข้อ 1 เป็นจริง สมการจำแนกกลุ่มในรูปเชิงเส้น ดังสมการที่ (3.2) อาจจะไม่เหมาะสม จึงควรใช้สมการจำแนกกลุ่มในรูป Quadratic อย่างไรก็ตาม ถ้าขนาดตัวอย่างเล็ก และ Variance-covariance Matrix แตกต่างกันไปไม่มากนัก การใช้สมการเชิงเส้นจะให้ผลดีกว่า

- ถ้าเงื่อนไขข้อ 1 ไม่จริง และต้องการแบ่งออกเป็น 2 กลุ่ม ควรใช้เทคนิค Logistic Regression ซึ่งไม่มีเงื่อนไขเกี่ยวกับ Multivariate Normal

การเปรียบเทียบเทคนิคการวิเคราะห์ความถดถอยเชิงเส้นทั่วไปกับเทคนิคการจำแนกกลุ่ม

การวิเคราะห์ความถดถอย (Regression)	การจำแนกกลุ่ม (Discriminant Analysis)
<p>วัตถุประสงค์</p> <ol style="list-style-type: none"> พยากรณ์มีค่าเฉลี่ยของตัวแปรตาม โดยทราบตัวแปรอิสระ ศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม <p>เงื่อนไข</p> <ol style="list-style-type: none"> ตัวแปรตามเป็นตัวแปรเชิงปริมาณ ตัวแปรอิสระ เป็นตัวแปรเชิงปริมาณ และ ตัวแปรเชิงกลุ่ม กรณีที่เป็นตัวแปรเชิงกลุ่ม ต้องปรับให้เป็นตัวแปรเทียม 1 ตัวแปรอิสระเป็นค่าคงที่มีเงื่อนไขเกี่ยวกับค่าคลาดเคลื่อน (e) ควรตรวจสอบการเกิดปัญหา Multicollinearity 	<p>วัตถุประสงค์</p> <ol style="list-style-type: none"> พยากรณ์กลุ่มให้กับ case โดยการทราบค่าตัวแปรอิสระ ศึกษาว่าตัวแปรอิสระตัวใดเป็นตัวแปรสำคัญที่ทำให้กลุ่มต่างกัน <p>เงื่อนไข</p> <ol style="list-style-type: none"> ตัวแปรตามเป็นตัวแปรเชิงกลุ่ม ตัวแปรอิสระควรเป็นตัวแปรเชิงปริมาณหรืออาจมีตัวแปรเชิงกลุ่มผสมอยู่ด้วย และต้องปรับตัวแปรเชิงกลุ่มให้เป็นตัวแปรเทียม 1 ตัวแปรอิสระในแต่ละกลุ่มจะต้องมีการแจกแจงแบบ Multivariate Normal 2 Variance-Covariance Matrix ของตัวแปรอิสระของแต่ละกลุ่มต้องเท่ากัน 4. ควรตรวจสอบการเกิดปัญหา Multicollinearity
<p>หลักการในการประมาณค่าสัมประสิทธิ์ (b)</p> <p>จะหาค่าสัมประสิทธิ์ความถดถอย (b) ที่ทำให้ผลบวกของค่าคลาดเคลื่อนยกกำลังสองมีค่าต่ำสุด ($\min \sum e_i^2$)</p>	<p>หลักการในการประมาณค่าสัมประสิทธิ์ (b)</p> <p>จะหาค่า Discriminant Coefficient (b) ที่ทำให้ค่า Eigenvalue สูงสุด หรือที่ทำให้ความแตกต่างระหว่างกลุ่ม (Between-Group Sum Square) ความแตกต่างภายในกลุ่ม (Within-Group Sum Square) มีค่าสูงสุด หรือทำให้โอกาสในการจัดกลุ่มผิดต่ำสุด</p>

การเปรียบเทียบเทคนิคการจำแนกกลุ่มกับการวิเคราะห์ความถดถอยโลจิสติก

ส่วนที่เหมือนกัน

1. ทั้งเทคนิค Discriminant และ Logistic มีวัตถุประสงค์เหมือนกัน คือ ต้องการพยากรณ์กลุ่มให้กับ case ต่าง ๆ และศึกษาว่าตัวแปรอิสระตัวใดเป็นตัวแปรสำคัญที่ทำให้กลุ่มมีความแตกต่าง
2. ตัวแปรอิสระของทั้ง 2 วิธีควรเป็นตัวแปรเชิงปริมาณ (Interval หรือ Ratio Scale)กลุ่มได้ ซึ่งต้องปรับตัวแปรเชิงกลุ่มให้อยู่ในรูปตัวแปรเทียบ

ส่วนที่แตกต่างกัน

1. จะพบว่าเทคนิค Discriminant จะมีเงื่อนไขเกี่ยวกับตัวแปรอิสระมากกว่าเทคนิค Logistic ดังได้กล่าวแล้วถึงเงื่อนไข 2 ข้อของตัวแปรอิสระ ดังนั้น กรณีที่ตัวแปรอิสระไม่เป็นไปตามเงื่อนไขอาจใช้เทคนิค **Logistic Regression** แทน

หมายเหตุ :

- ก. จากการใช้ Monte Carlo Simulation พบว่าเงื่อนไขที่ว่าตัวแปรอิสระในแต่ละกลุ่มต้องการมีการแจกแจงแบบ Multivariate Normal ไม่ใช่เงื่อนไขสำคัญที่จะทำให้เทคนิค Discriminant ไม่มีประสิทธิภาพ แต่เทคนิค Discriminant จะมีประสิทธิภาพดีกว่าเทคนิค Logistic เมื่อตัวแปรอิสระเป็นไปตามเงื่อนไขทั้ง 2 ข้อ
- ข. กรณีที่ตัวแปรอิสระส่วนใหญ่เป็นตัวแปรเชิงกลุ่ม ควรใช้เทคนิค Logistic regression

การนำไปประยุกต์ใช้

- เทคนิค Discriminant มักใช้กับงานด้านธุรกิจ การเงิน และสังคมศาสตร์ เป็นต้น
- เทคนิค logistic Regression มักใช้กับงานด้านการแพทย์ วิทยาศาสตร์ เป็นต้น

บรรณานุกรม

กัลยา วานิชย์บัญชา. 2548. สถิติสำหรับงานวิจัย. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.

กัลยา วานิชย์บัญชา. 2548. การวิเคราะห์ข้อมูลด้วย SPSS for window. พิมพ์ครั้งที่ 4. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย

บุญชม ศรีสะอาด. 2540. การวิจัยทางการวัดผลและการประเมินผล. กรุงเทพฯ: สุวีริยาสาส์น.

บุญเรียง ขจรศิลป์. 2547. การวิเคราะห์และการแปลความหมายข้อมูลในการวิจัยโดยใช้โปรแกรมสำเร็จรูป SPSS. กรุงเทพฯ : เอส.พี.เอ็น. การพิมพ์.

อำนาจ เลิศขันธ์. 2545. สถิติขั้นสูง. กรุงเทพฯ: ศิลปสนองการพิมพ์.

<http://www.watpon.com>