บทที่ 7
การวิเคราะห์จำแนกกลุ่ม
(Discriminant Analysis)

การวิเคราะห์จำแนกกลุ่มเป็นวิธีการทางสถิติที่ใช้วิเคราะห์จำแนกกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป โดยการวิเคราะห์จำแนกกลุ่มจะมีลักษณะเหมือนกับการวิเคราะห์ถดถอยคือ มีตัวแปรตาม 1 ตัว และมีตัวแปรอิสระอย่างน้อย 1 ตัว แต่จะมีข้อแตกต่างจาการวิเคราะห์ถดถอยที่ลักษณะหรือชนิดของตัวแปร โดยที่การวิเคราะห์การถดถอย ตัวแปรตามจะเป็นตัวแปรเชิงปริมาณ (Quantitative variable) แต่การวิเคราะห์จำแนกกลุ่ม ตัวแปรตามจะเป็นตัวแปรเชิงคุณภาพ (Qualitative variable) หรือตัวแปรเชิงกลุ่ม

วัตถุประสงค์ของการวิเคราะห์จำแนกกลุ่ม

1. เพื่อศึกษาปัจจัยหรือตัวแปรที่ทำให้เกิดความแตกต่างในแต่ละกลุ่ม ซึ่งเป็นการศึกษาว่าตัวแปรอิสระใดบ้างเป็นตัวแปรที่ทำให้กลุ่มแตกต่างกัน โดยการสร้างฟังก์ชันการวิเคราะห์จำแนกกลุ่มที่อยู่ในรูปเชิงเส้น

2. เพื่อนำสมการที่ได้จากการวิเคราะห์จำแนกกลุ่มมาใช้พยากรณ์การวิเคราะห์หน่วยใหม่ที่ยังไม่ทราบกลุ่มมาก่อนว่าสมควรจัดอยู่ในกลุ่มใด

ข้อตกลงเบื้องต้นของการวิเคราะห์จำแนกกลุ่ม

1. ตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร (Normality of Independent Variable)

2. เมทริกซ์ความแปรปรวนความแปรปรวนร่วมของตัวแปรอิสระของกลุ่มตัวอย่างต้องเท่ากัน (Equal Dispersion Matrices)

3. มีความสัมพันธ์เชิงเส้น (Linearity of Relationship)

4. ตัวแปรอิสระไม่มีความสัมพันธ์กันสูงเกินไป (multicollinearity)

ลักษณะข้อมูลและการเตรียมข้อมูลที่จะนำมาวิเคราะห์จำแนกกลุ่ม

1. มีการแบ่งกลุ่มประชากรหรือกลุ่มตัวอย่างออกเป็นกลุ่มอย่างน้อย 2 กลุ่ม (ต้องทราบมาก่อนการวิเคราะห์ว่าจะแบ่งเป็นกี่กลุ่ม)

2. ตัวแปรตาม (ตัวแปรที่ถูกจำแนก) ต้องเป็นตัวแปรที่แบ่งเป็นกลุ่ม ๆ ตั้งแต่ 2 กลุ่มขึ้นไป มีระดับการวัดในมาตรการวัดแบบนามบัญญัติ (Nominal Scale) หรือมาตราเรียงอันดับ (Ordinal Scale) และ ถ้าในกรณีที่มีข้อมูลอยู่ในระดับอื่นจะต้องแปลงข้อมูลให้เป็นข้อมูลเชิงกลุ่มก่อนการวิเคราะห์

3. ตัวแปรต้น (ตัวแปรอิสระหรือตัวแปรที่ใช้ในการจำแนกหรือทำนาย) เป็นตัวแปรต่อเนื่องที่มีระดับการวัดอยู่ในมาตราอันตรภาค (Internal Scale) หรือมาตราอัตราส่วน (Ratio Scale) และถ้าในกรณีที่มีข้อมูลอยู่ในระดับอื่นต้องแปลงข้อมูลให้เป็นตัวแปรหุ่น (Dummy Coding) ก่อนที่จะนำข้อมูลไปวิเคราะห์

หลักการของการวิเคราะห์จำแนกกลุ่ม

การวิเคราะห์จำแนกกลุ่มผู้ศึกษาจะต้องกำหนดจำนวนกลุ่มและแบ่งกลุ่มมาก่อนที่จะมาทำการวิเคราะห์จำแนกกลุ่ม โดยจะใช้การวิเคราะห์จำแนกกลุ่มเพื่อหาความสัมพันธ์หรือหาสาเหตุที่ทำให้หน่วยตัวอย่างมีความแตกต่างกันในรูปของสมการแสดงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระที่คาดว่าจะเป็นสาเหตุ แล้วนำสมการที่ได้ไปพยากรณ์ว่าหน่วยตัวอย่างใหม่นั้นควรจัดอยู่ในกลุ่มใด โดยความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระจะอยู่ในรูปเชิงเส้น (Linear) ดังนี้

ตัวแบบของประชากร (Model of Population)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

ตัวแบบของตัวอย่าง (Model of Sample)

$$Y = a + b_1 X_1 + b_2 X_2 + \dots + b_p X_p + \varepsilon$$

สำหรับการสร้างสมการจำแนกกลุ่มจะมี 2 วิธี คือ

1. วิธีตรง (Direct Method) เป็นวิธีการที่ผู้วิจัยต้องการตัวแปรทุกตัวตามที่ระบุไว้ด้วยเหตุผลทางทฤษฎีว่าจะแบ่งแยกได้กี่สมการ และมีลักษณะเป็นอย่างไร เพื่อพิสูจน์ตัวแปรที่คิดว่ามีความสำคัญต่อการจำแนกที่ระบุไว้ตามทฤษฎีนั้นแท้จริงแล้วมีความสำคัญหรือไม่

2. วิธีแบบขั้นตอน (Stepwise Method) เป็นวิธีการที่เลือกตัวแปรทีละตัวมาเข้าสมการ โดยหาตัวแปรที่ดีที่สุดในการจำแนกมาเข้าสมการเป็นตัวแรก จากนั้นก็จะหาตัวแปรที่ดีที่สุดตัวถัดมาเข้ามาในสมการเพื่อปรับปรุงแก้ไขทำให้สมการจำแนกกลุ่มดีขึ้น และในขั้นตอนต่อ ๆ ไปก็จะเป็นการนำตัวแปรที่ดีที่สุดแต่ละตัวที่เหลือมาเข้าสมการต่อไปเพื่อจะได้สมการจำแนกกลุ่มที่ดีที่สุด

สถิติที่สำคัญของการวิเคราะห์จำแนกกลุ่ม

1. ค่าไอเก้น (Eigenvalue) เป็นค่าที่แสดงอัตราส่วนการผันแปรระหว่างกลุ่มต่อการผันแปรภายในกลุ่ม ถ้าค่าไอเก้นมีค่าสูง ก็แสดงว่าสมการดีหรือมีค่าจำแนกสูงหรือก็คือ Eigenvalue ก็คือVariance ของคะแนนแปลงรูป Y ที่แปลงมาจาก X_1, X_2, ... X_p นั่นเอง (สมบัติ ท้ายเรือคำ. 2552 : 153)

2. ค่าสหสัมพันธ์แคนนอนิคอล (Canonical Correlation) เป็นตัวสถิติที่ใช้ในการตัดสินความสำคัญของสมการจำแนก เป็นมาตรวัดความสัมพันธ์ของสมการกับกลุ่มของตัวแปรซึ่งระบุการเป็นสมาชิกของกลุ่มนั้น ๆ ของตัวแปรตาม โดยเป็นการชี้ให้เห็นว่าการเป็นสมาชิกกลุ่มมีความสัมพันธ์กับสมการที่หาได้มากน้อยเพียงใด ด้วยเหตุนี้ ถ้าหากว่าค่าสหสัมพันธ์แคนนอนิคอลมีค่าสูง แสดงว่า การเป็นสมาชิกของกลุ่มสามารถอธิบายความผันแปรของตัวแปรกับสมการจำแนกได้มาก (สมบัติ ท้ายเรือคำ. 2552 : 153)

3. ค่าวิลค์แลมป์ดา (Wilks’ Lambda) เป็นมาตรวัดอำนาจจำแนก โดยถ้าค่าของวิลค์แลมป์ดามีค่ามาก ตัวแปรที่เหลือจะอธิบายการเป็นสมาชิกของกลุ่มโดยสมการใหม่น้อยลง (สมบัติ ท้ายเรือคำ. 2552 : 154)

การตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์จำแนกกลุ่ม

1. ตัวแปรอิสระมีการแจกแจงแบบปกติหลายตัวแปร (Normality of Independent Variable) การตรวจสอบดูจาก Normal Q-Q Plot หรือใช้สถิติทดสอบ เช่น Kolmogorov-Smirnov, Shapiro Wilk โดยใช้วิธีการตรวจสอบ Univariate Normality ทีละตัวแปร ถ้าทุกตัวแปรร่วมกันมีการแจกแจงแบบ Multivariate Normality แต่ละตัวแปรจะต้องมีการแจกแจงแบบ Univariate Normality ทุกตัว

2. เมทริกซ์ความแปรปรวนความแปรปรวนร่วมของตัวแปรอิสระของกลุ่มตัวอย่างต้องเท่ากัน (Equal Dispersion Matrices) มีวิธีการทดสอบ ดังนี้

1. กำหนดสมมติฐาน

$H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma$

$H_1 : \Sigma_i \neq \Sigma_j$ อย่างน้อย 1 คู่ ; $i \neq j$

2. สถิติทดสอบ

$$B = (1-c)\left\{\left[\sum_{i=1}^{k}(n_i - 1)\ln\left|S_p\right|\right] - \left[\sum_{i=1}^{k}(n_i - 1)\ln\left|S_i\right|\right]\right\}$$

$$c = \left[\sum_{i=1}^{k}\frac{1}{(n_i - 1)} - \frac{1}{\sum_{i=1}^{k}(n_i - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\right]$$

โดยที่

p คือ จำนวนตัวแปรอิสระ

n_i คือ จำนวนข้อมูลในกลุ่มที่ i

k คือ จำนวนกลุ่ม

S_i คือ เมทริกซ์ความแปรปรวนร่วมตัวอย่างของกลุ่มที่ i ; i = 1 , 2, ..., k

S_p คือ เมทริกซ์ความแปรปรวนร่วมของตัวอย่างทุกกลุ่ม

สถิติทดสอบ B จะมีการแจกแจงแบบไคกำลังสอง โดยมีจำนวนองศาความเป็นอิสระเท่ากับ $\frac{1}{2}p(p+1)(k-1)$

จะปฏิเสธ H_0 ถ้า $B > \chi^2_{\alpha,\frac{1}{2}p(p+1)(k-1)}$ โดยสถิติทดสอบไคสแควร์จะขึ้นอยู่กับขนาดตัวอย่าง ถ้าตัวอย่างมีขนาดใหญ่จะทำให้ค่าไคสแควร์มีค่ามาก ซึ่งจะทำให้ปฏิเสธ H_0 เสมอ

ตัวอย่างที่ 1 (สายชล สินสมบูรณ์ทอง, 2559, หน้า 420 – 421) ศูนย์หนังสือของมหาวิทยาลัยแห่งหนึ่งมีสมาชิก 50,000 คน ได้มีหนังสือออกใหม่เกี่ยวกับงานศิลปะ ผู้บริหารของศูนย์หนังสือสนใจส่งรายละเอียดหนังสือใหม่ให้แก่สมาชิกที่คาดว่าจะซื้อหนังสือ จึงสุ่มลูกค้าที่เป็นสมาชิกมา 1,000 คน แล้วส่งเอกสารให้ทางไปรษณีย์ พบว่ามีลูกค้าที่สั่งซื้อหนังสือ 83 คน ถ้าผู้บริหารศูนย์หนังสือต้องการศึกษาความแตกต่างของลูกค้ากลุ่มที่ซื้อและกลุ่มที่ไม่ซื้อหนังสือ

ในที่นี้ใช้การวิเคราะห์จำแนกกลุ่ม โดยแบ่งลูกค้าเป็น 2 กลุ่ม คือ กลุ่มที่ซื้อและกลุ่มที่ไม่ซื้อ ถ้าคาดว่าปัจจัยที่ส่งผลต่อพฤติกรรมการซื้อและไม่ซื้อคือ จำนวนเดือนนับจากที่ซื้อครั้งที่แล้วและจำนวนหนังสือด้านศิลปะที่เคยซื้อ ข้อมูลดังแสดงในตาราง

การซื้อ	จำนวน	จำนวนเดือนเฉลี่ย	จำนวนหนังสือเฉลี่ย
ไม่ซื้อ	917	12.73	0.33
ซื้อ	83	9.41	1.00

กำหนดให้

$$S_1 = \begin{bmatrix} 65.73 & 0.239 \\ 0.239 & 0.369 \end{bmatrix} \quad S_2 = \begin{bmatrix} 35.42 & -0.671 \\ -0.671 & 1.122 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 63.24 & 0.164 \\ 0.164 & 0.431 \end{bmatrix}$$

จงตรวจสอบเมทริกซ์ความแปรปรวนร่วมของตัวแปรจำนวนเดือนและจำนวนหนังสือของลูกค้ากลุ่มที่ไม่ซื้อและกลุ่มที่ซื้อหนังสือว่าเท่ากันหรือไม่ กำหนดให้ระดับนัยสำคัญเท่ากับ 0.05

วิธีทำ ให้ $\Sigma_{\text{ไม่ซื้อ}}$ แทน เมทริกซ์ความแปรปรวนร่วมของตัวแปรจำนวนเดือนและจำนวนหนังสือของกลุ่มที่ไม่ซื้อหนังสือ

$\Sigma_{\text{ซื้อ}}$ แทน เมทริกซ์ความแปรปรวนร่วมของตัวแปรจำนวนเดือนและจำนวนหนังสือของกลุ่มที่ซื้อหนังสือ

1. กำหนดสมมติฐาน

$$H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma$$

$$H_1 : \Sigma_i \neq \Sigma_j \text{ อย่างน้อย } 1 \text{ คู่ } ; i \neq j$$

หรือ H_0 : เมทริกซ์ความแปรปรวนร่วมของตัวแปรจำนวนเดือนและจำนวนหนังสือของลูกค้ากลุ่มที่ไม่ซื้อและกลุ่มที่ซื้อเท่ากัน

H_1 : เมทริกซ์ความแปรปรวนร่วมของตัวแปรจำนวนเดือนและจำนวนหนังสือของลูกค้ากลุ่มที่ไม่ซื้อและกลุ่มที่ซื้อแตกต่างกัน

2. สถิติทดสอบ

$$B = (1-c)\left\{\left[\sum_{i=1}^{k}(n_i - 1)\ln|S_p|\right] - \left[\sum_{i=1}^{k}(n_i - 1)\ln|S_i|\right]\right\}$$

โดยที่

$$c = \left[\sum_{i=1}^{k}\frac{1}{(n_i - 1)} - \frac{1}{\sum_{i=1}^{k}(n_i - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\right]$$

$$= \left[\left(\frac{1}{917}+\frac{1}{83}\right) - \frac{1}{917+83}\right]\left[\frac{2(2)^2 + 3(2) - 1}{6(2+1)(2-1)}\right]$$

$$= 0.010$$

ดังนั้น

$$B = (1 - 0.010)\left[998\ln\left|27.229\right| - 916\ln\left|24.197 - 82\ln\left|39.29\right|\right|\right]$$

$$= 0.99\left[998(3.304) - 916(3.186) - 82(3.671)\right]$$

$$= 77.21$$

จากตารางไคกำลังสอง จะได้ $\chi^2_{0.05,\frac{1}{2}(2)(3)(1)} = 7.81$

เนื่องจาก B = 77.21 > 7.81 จึงปฏิเสธ H_0 ดังนั้น เมทริกซ์ความแปรปรวนร่วมของตัวแปรจำนวนเดือนและจำนวนหนังสือของลูกค้ากลุ่มที่ไม่ซื้อและกลุ่มที่ซื้อหนังสือแตกต่างกันที่ระดับนัยสำคัญ 0.05

3. มีความสัมพันธ์เชิงเส้น (Linearity of Relationship) ตรวจสอบได้จากสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรของเพียร์สัน (Pearson Product Moment Correlation: rxy)

4. ตัวแปรอิสระไม่มีความสัมพันธ์กันสูงเกินไป (multicollinearity) ในการตรวจสอบความสัมพันธ์ร่วมเชิงพหุเชิงเส้น ตรวจสอบได้จากวิธีใช้สถิติสหสัมพันธ์เพียร์สัน (Pearson’s Product Moment Correlation) หรือใช้วิธีตรวจสอบโดยใช้สถิติ Collinearity โดยดูจากค่า Tolerance และ Variance Inflation Factor (VIF) หากค่า Tolerance เข้าใกล้ 0 มากเท่าใด ก็แสดงว่าระดับความสัมพันธ์ร่วมเชิงพหุเชิงเส้นของตัวแปรมีปัญหามาก ส่วนค่า VIF หากมีค่า เข้าใกล้ 10 มากเท่าใดก็แสดงว่าระดับความสัมพันธ์ร่วมเชิงพหุเชิงเส้นของตัวแปรมีปัญหามาก (ทรงศักดิ์ ภูสีอ่อน. 2551: 289)

การประมาณค่าสัมประสิทธิ์ของการวิเคราะห์จำแนกกลุ่ม

สำหรับหลักการของการวิเคราะห์จำแนกกลุ่มจะใช้วิธีของฟิชเชอร์ (Fisher's Approach) โดยถ้าตัวอย่างแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มที่ 1 และ 2 มีตัวแปรอิสระ p ตัว คือ $X_1, X_2, \ldots, X_p$ ซึ่งขนาดตัวอย่างคือ n หน่วย จัดอยู่ในกลุ่มที่ 1 จำนวน n_1 หน่วย และจัดอยู่ในกลุ่มที่ 2 จำนวน n_2 หน่วย จะได้ว่า $n = n_1 + n_2$ ซึ่งฟิชเชอร์ (Fisher : 1936) ได้ศึกษาในกรณีที่แบ่งเป็น 2 กลุ่ม ดังนี้

ให้ กลุ่มที่ 1 มีเวคเตอร์ค่าเฉลี่ย $\underline{\mu}_1$ และเมทริกซ์ความแปรปรวนร่วม Σ_1

กลุ่มที่ 2 มีเวคเตอร์ค่าเฉลี่ย $\underline{\mu}_2$ และเมทริกซ์ความแปรปรวนร่วม Σ_2

ถ้า $\Sigma_1 = \Sigma_2 = \Sigma_{p \times p}$

ฟังก์ชันจำแนกกลุ่มที่อยู่ในรูปเชิงเส้นคือ

$$\hat{Y} = b_1X_1 + b_2X_2 + \dots + b_pX_p$$

หรือ $\hat{Y} = \underline{b}^{/}\,\underline{X}$

โดยที่ $\underline{X}$ เป็นเวคเตอร์ของตัวแปรอิสระ p ตัว

$$\underline{X}^{/} = (X_1, X_2, \dots, X_p)$$

และ $\underline{b} = (b_1, b_2, \dots, b_p)$

ในการวิเคราะห์ต้องการหาค่าเวคเตอร์สัมประสิทธิ์จำแนกกลุ่ม $\underline{b}$ ที่ทำให้อัตราส่วนความแปรผันระหว่างกลุ่ม (SSB) กับความแปรผันภายในกลุ่ม (SSW) มีค่ามากที่สุด หรือหาค่า $\underline{b}$ ที่ทำให้ $L = \dfrac{SSB}{SSW}$ มีค่ามากที่สุด

โดยที่ SSB คือ ผลบวกกำลังสองของความแตกต่างระหว่างกลุ่ม

(Sum Square Between Group)

SSW คือ ผลบวกกำลังสองของความแตกต่างภายในกลุ่ม

(Sum Square Within Group)

ค่าเฉลี่ยของกลุ่มที่ 1 คือ $\bar{Y}_1 = \underline{b}' \underline{\bar{X}}_1$

ค่าเฉลี่ยของกลุ่มที่ 2 คือ $\bar{Y}_2 = \underline{b}' \underline{\bar{X}}_2$

โดยที่ X_1 เป็นเมทริกซ์ขนาด $n_1 \times p$ ของข้อมูลในกลุ่มที่ 1

X_2 เป็นเมทริกซ์ขนาด $n_2 \times p$ ของข้อมูลในกลุ่มที่ 2

$\underline{\bar{X}}_1$ เป็นเวคเตอร์ค่าเฉลี่ยของตัวแปรอิสระ p ตัว ในกลุ่มที่ 1

$\underline{\bar{X}}_2$ เป็นเวคเตอร์ค่าเฉลี่ยของตัวแปรอิสระ p ตัว ในกลุ่มที่ 2

$$X_1 = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n_1 1} & X_{n_1 2} & \cdots & X_{n_1 p} \end{bmatrix}_{n_1 \times p} \quad , \quad \underline{X}_1 = \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \\ \vdots \\ \bar{X}_{1p} \end{bmatrix}_{p \times 1}$$

$$X_2 = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n_2 1} & X_{n_2 2} & \cdots & X_{n_2 p} \end{bmatrix}_{n_1 \times p} \quad , \quad \underline{X}_2 = \begin{bmatrix} \bar{X}_{21} \\ \bar{X}_{22} \\ \vdots \\ \bar{X}_{2p} \end{bmatrix}_{p \times 1}$$

$$SSW = \sum_{i=1}^{n_1} \left(Y_{1i} - \bar{Y}_1\right)^2 + \sum_{j=1}^{n_2} \left(Y_{2j} - \bar{Y}_2\right)^2$$

แทนค่า $Y_i = \underline{b}' \underline{X}_i$ และ $\bar{Y}_i = \underline{b}' \underline{\bar{X}}_1$ จะได้

$$SSW = \sum_{i=1}^{n_1} b'\left(X_{1i} - \overline{X}_1\right)\left(X_{1i} - \overline{X}_1\right)' b + \sum_{j=1}^{n_2} b'\left(X_{2i} - \overline{X}_2\right)\left(X_{2i} - \overline{X}_2\right)' b$$

$$= b'\left[\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)\left(X_{1i} - \overline{X}_1\right)'\right] b + b'\left[\sum_{j=1}^{n_2}\left(X_{2i} - \overline{X}_2\right)\left(X_{2i} - \overline{X}_2\right)'\right] b$$

$$= b'\left[\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)\left(X_{1i} - \overline{X}_1\right)' + \sum_{j=1}^{n_2}\left(X_{2i} - \overline{X}_2\right)\left(X_{2i} - \overline{X}_2\right)'\right] b$$

เทอมที่ 1 ทางขวามือในวงเล็บของสมการนี้เป็นผลบวกกำลังสองของความแตกต่างภายในกลุ่มที่ 1 ส่วนเทอมที่ 2 ทางขวามือในวงเล็บเป็นผลบวกกำลังสองของความแตกต่างภายในกลุ่มที่ 2

ในกรณีที่ $\Sigma_1 = \Sigma_2 = \Sigma_{p\times p}$ สมการดังข้างต้นจะอยู่ในรูป

$$SSW \propto \underline{b}^{\prime}\Sigma\underline{b}$$

$$SSB = n_1\left(\overline{Y}_1 - \overline{Y}\right)^2 + n_2\left(\overline{Y}_2 - \overline{Y}\right)^2$$

แทนค่า $\overline{Y} = b^{\prime}\overline{X}$ และ $\overline{Y}_i = b^{\prime}\overline{X}_i$ จะได้

$$SSB = b^{\prime}\left[n_1\left(\overline{X}_1 - \overline{X}\right)\left(\overline{X}_1 - \overline{X}\right)^{\prime} + n_2\left(\overline{X}_2 - \overline{X}\right)\left(\overline{X}_2 - \overline{X}\right)^{\prime}\right]b$$

โดยที่ $\overline{X} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$

หรือ $\bar{X}_1 - \bar{X} = \dfrac{n_2(\bar{X}_1 - \bar{X}_2)}{n_1 + n_2} = \dfrac{n_2 d}{n_1 + n_2}$

$$\bar{X}_2 - \bar{X} = \frac{n_1(\bar{X}_2 - \bar{X}_1)}{n_1 + n_2} = \frac{n_1 d}{n_1 + n_2}$$

โดยที่ $d = \bar{X}_1 - \bar{X}_2$ เป็นเวคเตอร์ของผลต่างของเวคเตอร์ค่าเฉลี่ย 2 กลุ่ม

แทนค่า $\bar{X}_1 - \bar{X}$ และ $\bar{X}_2 - \bar{X}$ ลงในสมการของ SSB จะได้

$$SSB = b'\left[n_1\left(\frac{n_2}{n_1+n_2}\right)^2 dd' + n_2\left(\frac{n_1}{n_1+n_2}\right)^2 dd'\right]b$$

หรือ $SSB \propto \underline{b}'\underline{d}\;\underline{d}'\underline{b}$

ดังนั้น $L = \dfrac{SSB}{SSW} = \dfrac{b'd\,d'\,b}{b'\sum b}$

หาเวคเตอร์สัมประสิทธิ์จำแนกกลุ่ม $\underline{b}$ ที่ทำให้ L มีค่ามากที่สุด ดังนี้

$$\frac{\partial L}{\partial b} = \frac{\left[\left(b'\sum b\right)\left(2dd'b\right)-\left(b'dd'b\right)\left(2\sum b\right)\right]}{\left(b'\sum b\right)^2} = 0$$

นำ $\frac{1}{2}b'\sum b$ คูณทั้งสองข้างของสมการ จะได้

$$\frac{\left(b'\sum b\right)^2 dd'b}{\left(b'\sum b\right)^2} - \frac{b'\sum b\left(b'dd'b\right)}{\left(b'\sum b\right)^2} = 0$$

$$dd'b - \left(\frac{b'dd'b}{b'\sum b}\right)\sum b = 0$$

$$dd'b - L\sum b = 0$$

$$b = \frac{\sum^{-1} dd'b}{L}$$

หรือ $b \propto \Sigma^{-1} d$

โดยที่ L และ $d'b$ เป็นสเกลาร์

ในกรณีที่ไม่ทราบค่าเมทริกซ์ความแปรปรวนร่วมประชากร Σ จะประมาณค่าด้วยเมทริกซ์ความแปรปรวนร่วมตัวอย่างรวมกันหรือของทุกกลุ่ม S_p (Pool sample variance – covariance matrix)

ดังนั้น $b = S_p^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$

โดยที่ $S_p = \dfrac{1}{n_1 + n_2 - 2}\left(\bar{X}_1\bar{X}_1' - \bar{X}_2\bar{X}_2'\right)$

ตัวอย่างที่ 2 (สายชล สินสมบูรณ์ทอง, 2559, หน้า 418 – 419) ในการสร้างวิธีการตรวจสอบตัวนำพาโรคโลหิตจาง ตัวอย่างเลือดของผู้หญิงได้แบ่งเป็น 2 กลุ่ม และวัดตัวแปรอิสระคือ X_1 และ X_2 ผู้หญิงกลุ่มที่ 1 จำนวน 30 คน เลือกมาจากประชากรของผู้หญิงที่ไม่นำพาโรคโลหิตจาง เรียกว่ากลุ่มปกติ ส่วนผู้หญิงกลุ่มที่ 2 จำนวน 22 คน เลือกมาจากประชากรเพศหญิงที่นำพาโรคโลหิตจาง เรียกว่ากลุ่มนำพาโรค สมมติว่าเมทริกซ์ความแปรปรวนร่วมเท่ากัน แต่ไม่ทราบค่า $\sum$ จึงประมาณค่าด้วยเมทริกซ์ความแปรปรวนร่วมตัวอย่างรวมกัน S_{pooled} การแจกแจงแบบปกติหลายตัวเหมาะสมกับข้อมูลนี้

$$\overline{X}_1 = \begin{bmatrix} -0.0065 \\ -0.0390 \end{bmatrix}, \quad \overline{X}_2 = \begin{bmatrix} -0.2483 \\ 0.0262 \end{bmatrix}$$

$$S_{pooled}^{-1} = \begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix}$$

จงหาฟังก์ชันจำแนกกลุ่มเชิงกลุ่ม ตัวแปรอิสระตัวใดที่ทำให้กลุ่มแตกต่างกันมากกว่า

วิธีทำ

$$\bar{X}_1 - \bar{X}_1 = \begin{bmatrix} 0.2418 \\ -0.0652 \end{bmatrix}$$

$$b = S_p^{-1}(\bar{X}_1 - \bar{X}_2)$$

$$= \begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix} \begin{bmatrix} 0.2418 \\ -0.0652 \end{bmatrix}$$

$$= \begin{bmatrix} 37.610 \\ -28.915 \end{bmatrix}$$

ฟังก์ชันจำแนกประเภทเชิงเส้น คือ

$$\hat{Y} = 37.610X_1 - 28.915X_2$$

ดังนั้นตัวแปรอิสระ X_1 ทำให้กลุ่มแตกต่างกันมากกว่า

การทดสอบสมมติฐานของการวิเคราะห์จำแนกกลุ่ม

ในการวิเคราะห์จำแนกกลุ่ม ผู้วิเคราะห์ต้องแบ่งกลุ่มมาก่อนที่จะทำการวิเคราะห์ ดังนั้นกลุ่มที่แบ่งอาจจะไม่มีความแตกต่างกันก็สามารถเป็นไปได้ ดังนั้นจึงต้องทำการตรวจสอบก่อนว่ากลุ่มที่แบ่งมานั้นมีความแตกต่างกันหรือไม่ โดยการทดสอบค่าเฉลี่ยหรือจุดกลางของทั้ง 2 กลุ่ม ดังนี้

1. กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \underline{\mu}_1 = \underline{\mu}_2$$

$$H_1 : \underline{\mu}_1 \neq \underline{\mu}_2$$

2. กำหนดตัวสถิติทดสอบ

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\left(\overline{X}_1 - \overline{X}_2\right)' S_p^{-1}\left(\overline{X}_1 - \overline{X}_2\right)$$

$$= \frac{n_1 n_2}{n_1 + n_2} D^2$$

เมื่อ $D^2 = \left(\overline{X}_1 - \overline{X}_2\right)' S_p^{-1}\left(\overline{X}_1 - \overline{X}_2\right)$

โดยที่ T^2 จะมีการแจกแจงแบบโฮเทลลิ่งทีกำลังสอง (Hotelling T^2) ที่จำนวนองศาความเป็นอิสระ p และ $n_1 + n_2 - 2$ ตามลำดับ

3. กำหนดอาณาเขตวิกฤต

จะปฏิเสธ H_0 ถ้า $T^2 > T^2_{\alpha;n_1+n_2-2}$

ความสัมพันธ์ระหว่าง F กับ T^2 และ D^2

$H_0 : \underline{\mu}_1 = \underline{\mu}_2$

$H_1 : \underline{\mu}_1 \neq \underline{\mu}_2$

จาก

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\left(\overline{X}_1 - \overline{X}_2\right)' S_p^{-1}\left(\overline{X}_1 - \overline{X}_2\right)$$

$$\sim \frac{p(n_1 + n_2 - 2)}{n_1 + n_2 - p - 1} F_{p\,;\, n_1 + n_2 - p - 1}$$

จะได้

$$T^2 = \frac{p\left(n_1 + n_2 - 2\right)}{n_1 + n_2 - p - 1} F$$

ดังนั้นสถิติทดสอบ คือ

$$F = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

$$= \frac{n_1 n_2 \left(n_1 + n_2 - p - 1\right)}{p\left(n_1 + n_2\right)\left(n_1 + n_2 - 2\right)} D^2$$

โดยที่ F จะมีการแจกแจงแบบ F ที่องศาความเป็นอิสระ p และ $n_1 + n_2 - p - 1$ ตามลำดับ และจะปฏิเสธสมมติฐาน H_0 ถ้า $F = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2 > F_{\alpha;\, p, n_1 + n_2 - p - 1}$

ตัวอย่างที่ 3 (สายชล สินสมบูรณ์ทอง, 2559, หน้า 424) จากตัวอย่างที่ 1 พิจารณาข้อมูลตัวนำพาโรคโลหิตจาง ให้ทดสอบว่าค่าเฉลี่ยของกลุ่มปกติและกลุ่มนำพาโรคแตกต่างกันหรือไม่ที่ระดับนัยสำคัญ 0.05 โดยจากตัวอย่างที่ 1 ทราบว่า $\bar{X}_1 - \bar{X}_2 = \begin{bmatrix} 0.2418 \\ -0.0652 \end{bmatrix}$

วิธีทำ 1. กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \underline{\mu}_1 = \underline{\mu}_2$$

$$H_1 : \underline{\mu}_1 \neq \underline{\mu}_2$$

2. กำหนดตัวสถิติทดสอบโดยใช้โฮเทลลิ่งทีกำลังสอง

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}\left(\bar{X}_1 - \bar{X}_2\right)' S_p^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$$

$$= \frac{n_1 n_2}{n_1 + n_2} D^2$$

เมื่อ

$$D^2 = \left(\bar{X}_1 - \bar{X}_2\right)' S_p^{-1}\left(\bar{X}_1 - \bar{X}_2\right)$$

$$= \begin{bmatrix} 0.2418 & -0.0652 \end{bmatrix} \begin{bmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{bmatrix} \begin{bmatrix} 0.2418 \\ -0.0652 \end{bmatrix}$$

$$= 10.9793$$

ดังนั้น $T^2 = \dfrac{(30)(22)}{30+22}(10.9793) = 139.35$

3. กำหนดอาณาเขตวิกฤต

เปิดตารางโฮเทลลิ่งกำลังสอง จะได้ $T^2_{0.05\,;\,2,50} = 6.503$

4. สรุปผล

เนื่องจาก $T^2 = 139.35 > 6.503$ จึงปฏิเสธสมมติฐาน H_0 ดังนั้นค่าเฉลี่ยของกลุ่มปกติและกลุ่มนำพาโรคแตกต่างกัน ที่ระดับนัยสำคัญ 0.05

ถ้าใช้การแจกแจงแบบเอฟ

1. สถิติทดสอบ คือ

$$F = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)}T^2$$

$$= \frac{30 + 22 - 2 - 1}{2(30 + 22 - 2)}(139.35)$$

$$= 68.28$$

2. กำหนดอาณาเขตวิกฤต

เปิดตารางเอฟ จะได้ $F_{0.05\,;\,2\,,\,49} = 3.19$

3. สรุปผล

เนื่องจาก $F = 68.28 > 3.19$ จึงปฏิเสธสมมติฐาน H_0 ดังนั้นค่าเฉลี่ยของกลุ่มปกติและกลุ่มนำพาโรคแตกต่างกัน ที่ระดับนัยสำคัญ 0.05

การทดสอบนัยสำคัญของสมการจำแนกกลุ่ม

ขั้นตอนนี้เป็นการทดสอบนัยสำคัญของสมการจำแนกกลุ่ม เมื่อได้สมการจำแนกกลุ่มแล้ว โดยจะทำการทดสอบนัยสำคัญของสมการที่ได้ เพื่อให้ทราบว่าสมการใดที่มีอำนาจจำแนกกลุ่มได้อย่างมีนัยสำคัญ โดยใช้วิธีของ Barlett test

จากสูตร $V_m = \left[N - 1 - 0.5(p+k)\right] \ln(1+\lambda_m)$

เมื่อ V_m แทน ค่าสถิติที่จะใช้เปรียบเทียบกับค่าวิกฤตเพื่อทราบความมีนัยสำคัญของสมการที่ m ค่าวิกฤต (Critical Value) ซึ่งหาได้จากการเปิดตารางไคสแควร์ที่ $df = p + k - 2m$

N แทน จำนวนสมาชิกในกลุ่มตัวอย่างทั้งหมด

p แทน จำนวนตัวแปร

k แทน จำนวนกลุ่ม

λ แทน ค่า Eigenvalue ของสมการที่ทดสอบ

สมการจำแนกจะมีนัยสำคัญ เมื่อค่า V_m ที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่าวิกฤต

ในกรณีที่ต้องการทราบว่าสมการจำแนกกลุ่มรวมกันแล้วสามารถจำแนกกลุ่มได้อย่างมีนัยสำคัญหรือไม่ ให้ทดสอบจากสูตร

$$v = \left[N - 1 - 0.5(p + k)\right] \sum_{m=1}^{r} \ln(1 + \lambda_m)$$

เมื่อ r แทน จำนวนสมการ

df แทน องศาความเป็นอิสระ ในกรณีนี้มีค่าเท่ากับ $p(k - 1)$

ในกรณีที่ต้องการค่าที่ใช้ทดสอบสมการจำแนกกลุ่มโดยภาพรวม V ให้นำค่า V_m ของแต่ละสมการมารวมกัน

สมการจำแนกกลุ่มแต่ละสมการจะไม่สัมพันธ์กัน นั่นคือ Y_1 , Y_2, ..., Y_p เป็นอิสระจากกัน สมการจำแนกกลุ่มที่อยู่หลัง ๆ จะส่งผลน้อยมากจนบางครั้งไม่จำเป็นต้องนำมาพิจารณาหรือนำเสนอในงานวิจัย ซึ่งในการพิจารณาจะสามารถพิจารณาได้จากการทดสอบนัยสำคัญของสมการ ถ้าพบว่าไม่มีนัยสำคัญแสดงว่าสมการนั้นส่งผลน้อยมาก

สมการจำแนกกลุ่มสมการแรก (V_1) จะมีอำนาจจำแนกสูงสุด สมการต่อมาจะมีอำนาจจำแนกรองลงมาตามลำดับ เมื่อต้องการเปรียบเทียบว่าสมการจำแนกกลุ่มแต่ละสมการมีสัดส่วนจำแนกได้มากน้อยอย่างไร สามารถพิจารณาได้จากสูตร

$$p_i = \frac{\lambda_i}{(\lambda_1 + \lambda_2 + \ldots + \lambda_m)}$$

เมื่อ p แทน ค่าอำนาจในการจำแนกตัวแปร

λ แทน ค่า Eigenvalue

ตัวอย่างที่ 4 (สมบัติ ท้ายเรือคำ. 2552 : 158) มีสมการจำแนกทั้งหมด 3 สมการ มีค่า λ จำนวน 3 ค่า สมมติให้ $\lambda_1 = 0.85$, $\lambda_2 = 0.27$ และ $\lambda_3 = 0.05$ สัดส่วนของการจำแนกของสมการทั้ง 3 เป็นดังนี้

$$p_1 = \frac{\lambda_1}{(\lambda_1+\lambda_2+\lambda_3)} = \frac{0.85}{0.85+0.27+0.05} = 0.73$$

$$p_2 = \frac{\lambda_2}{(\lambda_1+\lambda_2+\lambda_3)} = \frac{0.27}{0.85+0.27+0.05} = 0.23$$

$$p_3 = \frac{\lambda_3}{(\lambda_1+\lambda_2+\lambda_3)} = \frac{0.05}{0.85+0.27+0.05} = 0.04$$

จากผลการเปรียบเทียบ พบว่า สมการแรกมีอำนาจจำแนกกลุ่ม 73% สมการที่ 2 มีอำนาจจำแนกกลุ่ม 23% และสมการที่ 3 มีอำนาจจำแนกกลุ่ม 4% ซึ่งจะเห็นได้ว่า สมการแรกจะมีอำนาจจำแนกสูงสุด สมการต่อมามีอำนาจจำแนกรองลงมาตามลำดับ ส่วนสมการสุดท้ายมีอำนาจจำแนกน้อยมากหรือส่งผลน้อยมากนั่นเอง

การวิเคราะห์จำแนกกลุ่มด้วยโปรแกรมคอมพิวเตอร์

การคำนวณด้วยคอมพิวเตอร์มีขั้นตอนดังนี้

1. สร้างไฟล์ข้อมูล
2. ตรวจสอบข้อตกลงเบื้องต้น
3. วิเคราะห์ข้อมูลและกำหนดเงื่อนไข
4. แปลผลการวิเคราะห์ข้อมูล โดยนำผลลัพธ์จากการคำนวณมาสรุปผล โดยดูว่าตัวแปรอิสระตัวใดที่ทำให้กลุ่มมีความแตกต่างกัน และตัวใดที่ไม่ทำให้กลุ่มแตกต่างกัน
5. สร้างสมการจำแนกกลุ่ม
6. ตรวจสอบและพิจารณาความน่าเชื่อถือของสมการจำแนกกลุ่ม โดยพิจารณาจากความสามารถในการพยากรณ์กลุ่ม

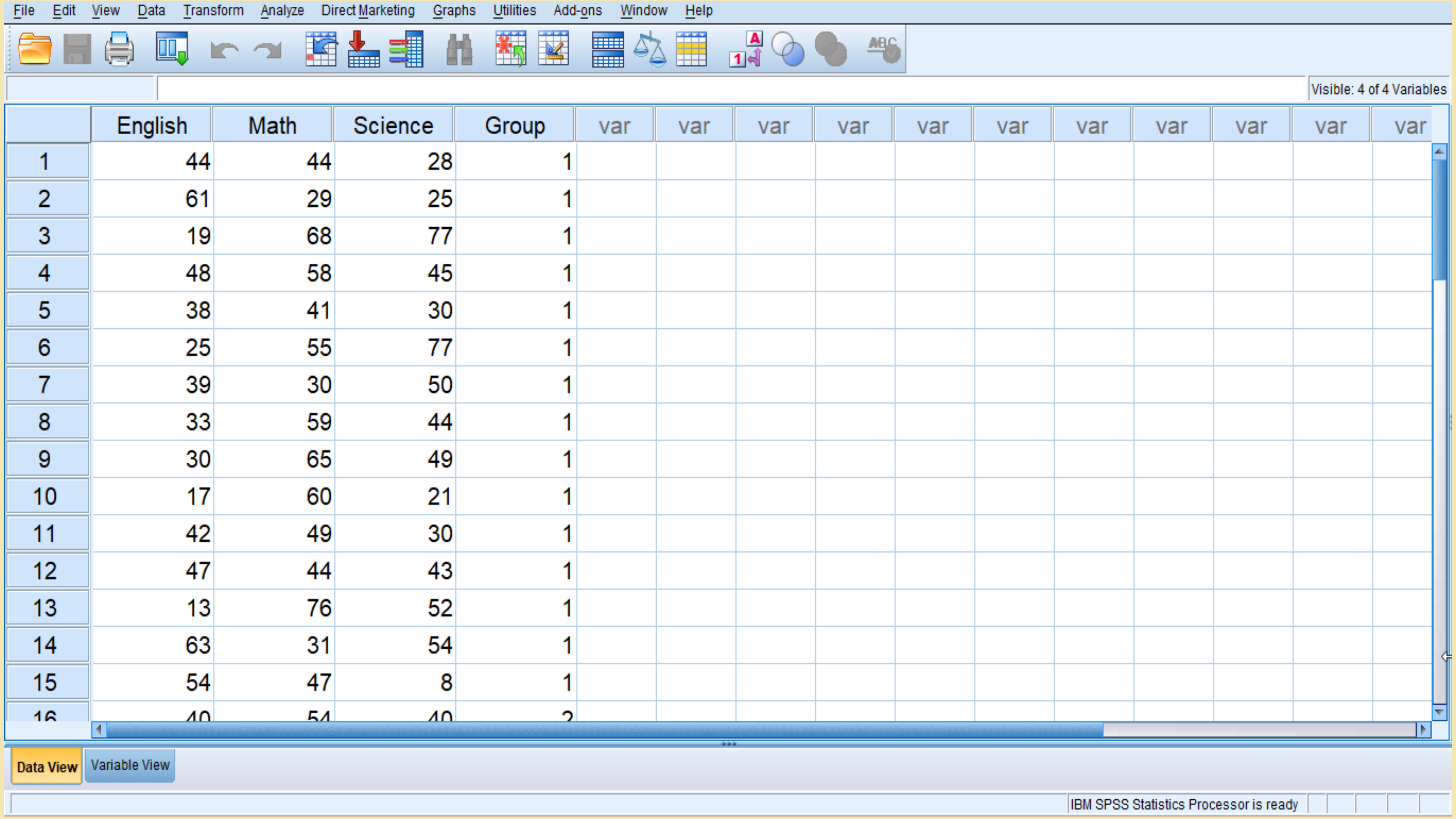
ตัวอย่างที่ 5 จากข้อมูลกลุ่มตัวอย่างที่แบ่งออกเป็น 2 กลุ่ม คือ กลุ่มที่ฉลาดมาก กับ ฉลาดปานกลาง โดยมีผลการเรียนวิชาภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์ ดังนี้

ผลการเรียนวิชาภาษาอังกฤษ	ผลการเรียนวิชาคณิตศาสตร์	ผลการเรียนวิชาวิทยาศาสตร์	กลุ่ม
44	44	28	1
61	29	25	1
19	68	77	1
48	58	45	1
38	41	30	1
25	55	77	1
39	30	50	1
33	59	44	1
30	65	49	1
17	60	21	1
42	49	30	1
47	44	43	1
13	76	52	1
63	31	54	1
54	47	8	1
40	54	40	2

ผลการเรียนวิชาภาษาอังกฤษ	ผลการเรียนวิชาคณิตศาสตร์	ผลการเรียนวิชาวิทยาศาสตร์	กลุ่ม
29	53	19	2
28	66	71	2
27	67	17	2
45	66	79	2
55	43	51	2
68	45	58	2
52	56	51	2
74	47	33	2
70	51	29	2
49	67	74	2
80	53	40	2
50	61	13	2
48	71	71	2
65	60	39	2

ดำเนินการตามขั้นตอนดังนี้

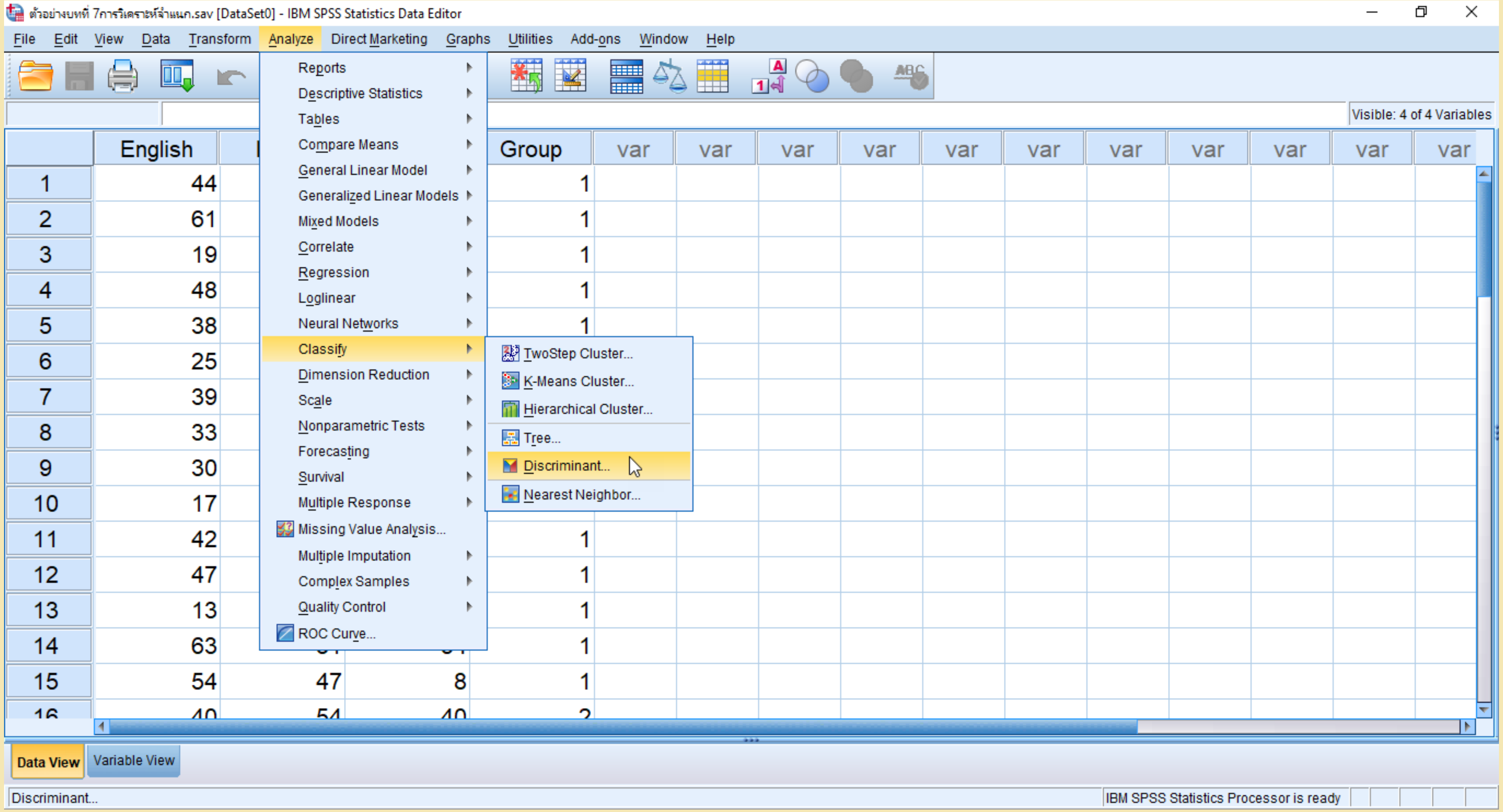
1. สร้างไฟล์ข้อมูลโดยกำหนดให้มีตัวแปรตาม 1 ตัว ชื่อว่า group และตัวแปรต้น 3 ตัว มีชื่อว่า English, Math และ Science ดังภาพประกอบที่ 7.1



	English	Math	Science	Group
1	44	44	28	1
2	61	29	25	1
3	19	68	77	1
4	48	58	45	1
5	38	41	30	1
6	25	55	77	1
7	39	30	50	1
8	33	59	44	1
9	30	65	49	1
10	17	60	21	1
11	42	49	30	1
12	47	44	43	1
13	13	76	52	1
14	63	31	54	1
15	54	47	8	1
16	40	54	40	2

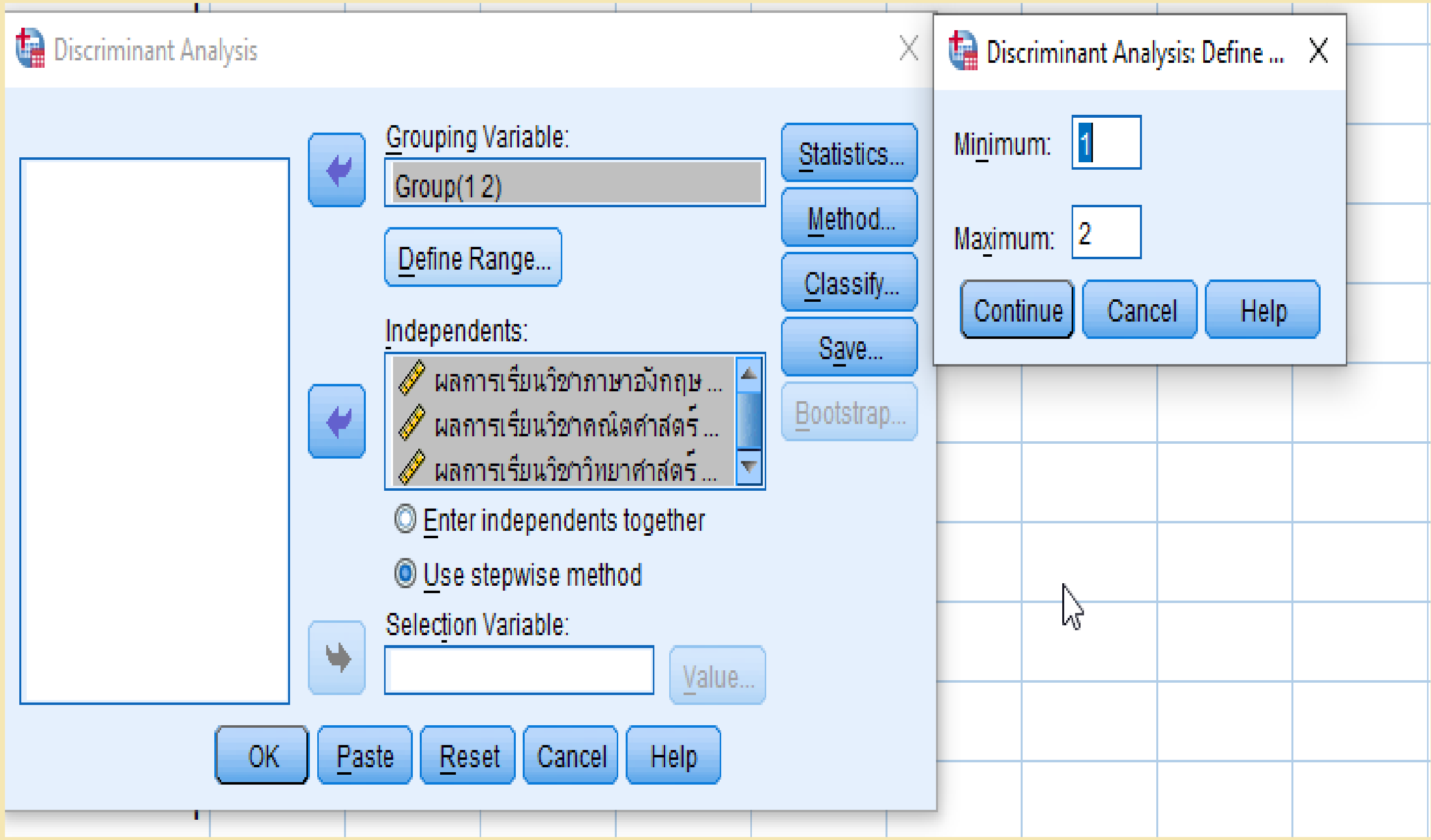
ภาพประกอบที่ 7.1 แสดงการสร้างไฟล์ข้อมูลลงในโปรแกรม SPSS

2. Click Analyze ⟶ Classify ⟶ Discriminant จะปรากฎหน้าต่าง Discriminant Analysis ดังภาพประกอบที่ 7.2



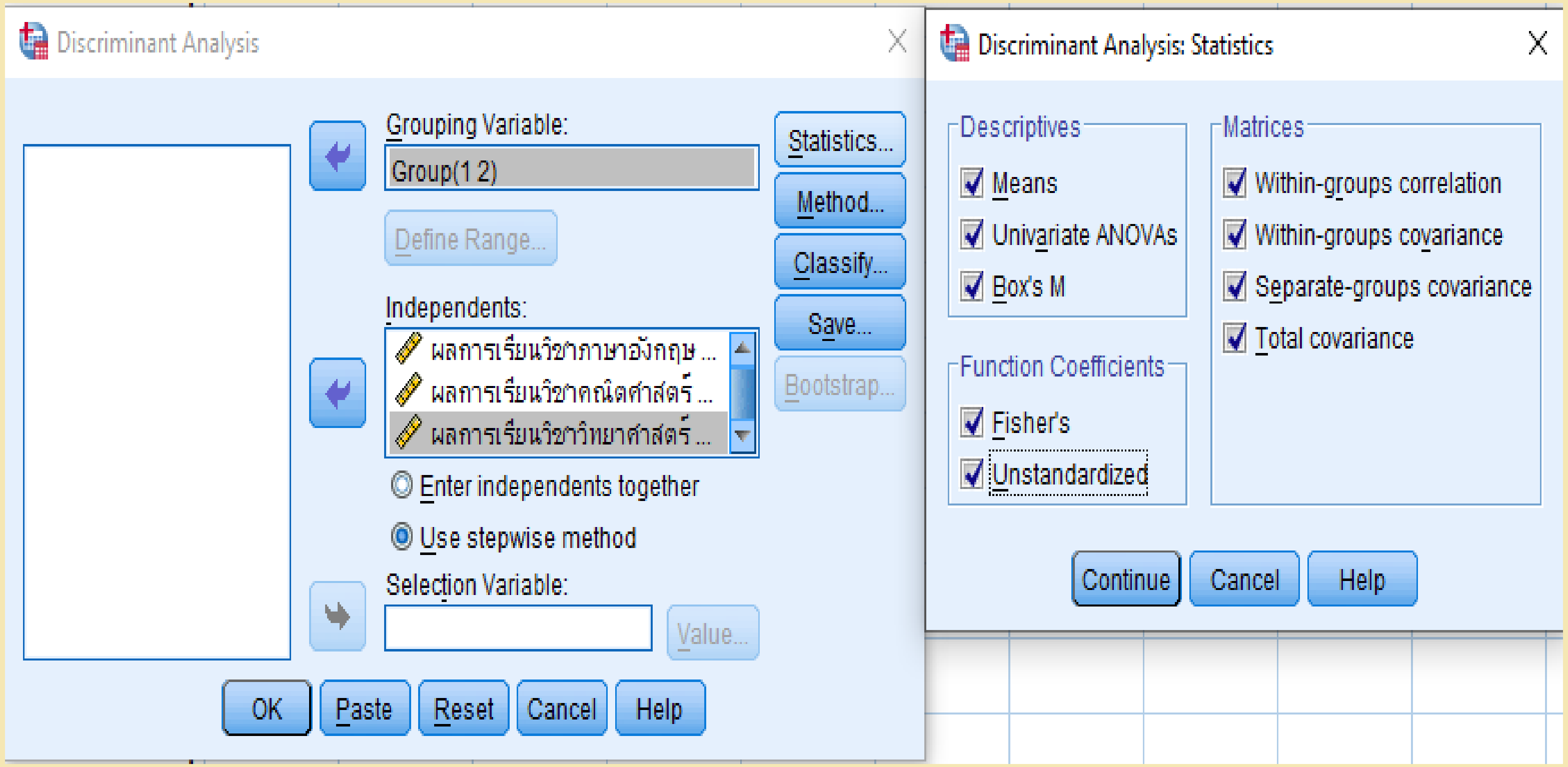
ภาพประกอบที่ 7.2 แสดงการเลือกใช้คำสั่ง Discriminant Analysis

3. เลือกตัวแปร Group ไปไว้ที่ช่อง Grouping Variable โดยคลิก Define Range จะปรากฏหน้าต่างเล็ก ๆ ขึ้นมาอีกให้ใส่เลข 1 ที่ช่อง Minimum และใส่เลข 2 ที่ช่อง Maximum จากนั้นให้นำตัวแปร English, Math และ Science ไปไว้ที่ช่อง Independents และคลิกเลือก Use Stepwise Method เพื่อเลือกสมการทำนายที่ดีที่สุด ดังภาพประกอบที่ 7.3



ภาพประกอบที่ 7.3 แสดงการเลือกตัวแปรต่าง ๆ เพื่อใช้ในคำสั่ง Discriminant Analysis

4. จากภาพประกอบที่ 7.3 คลิกเลือก Statistics จะปรากฏหน้าต่าง Discriminant Analysis Statistics ดังภาพประกอบที่ 7.4



ภาพประกอบที่ 7.4 แสดงผลจากการเลือกคำสั่ง Statistics

จากภาพประกอบที่ 7.4 สามารถอธิบายรายละเอียดของสถิติต่าง ๆ ได้ดังนี้ (ทรงศักดิ์ ภูสีอ่อน. 2551 : 312 ; กัลยา วาณิชย์บัญชา. 2551 : 43 – 44)

(1) Mean เป็นการนำเสนอเกี่ยวกับค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน จำนวนข้อมูลที่นำเสนอโดยแยกกลุ่มตัวแปรและนำเสนอในภาพรวมด้วย

(2) Univariate ANOVAs เป็นการนำเสนอเกี่ยวกับการทดสอบความเท่ากันของค่าเฉลี่ยของกลุ่มในแต่ละตัวแปร

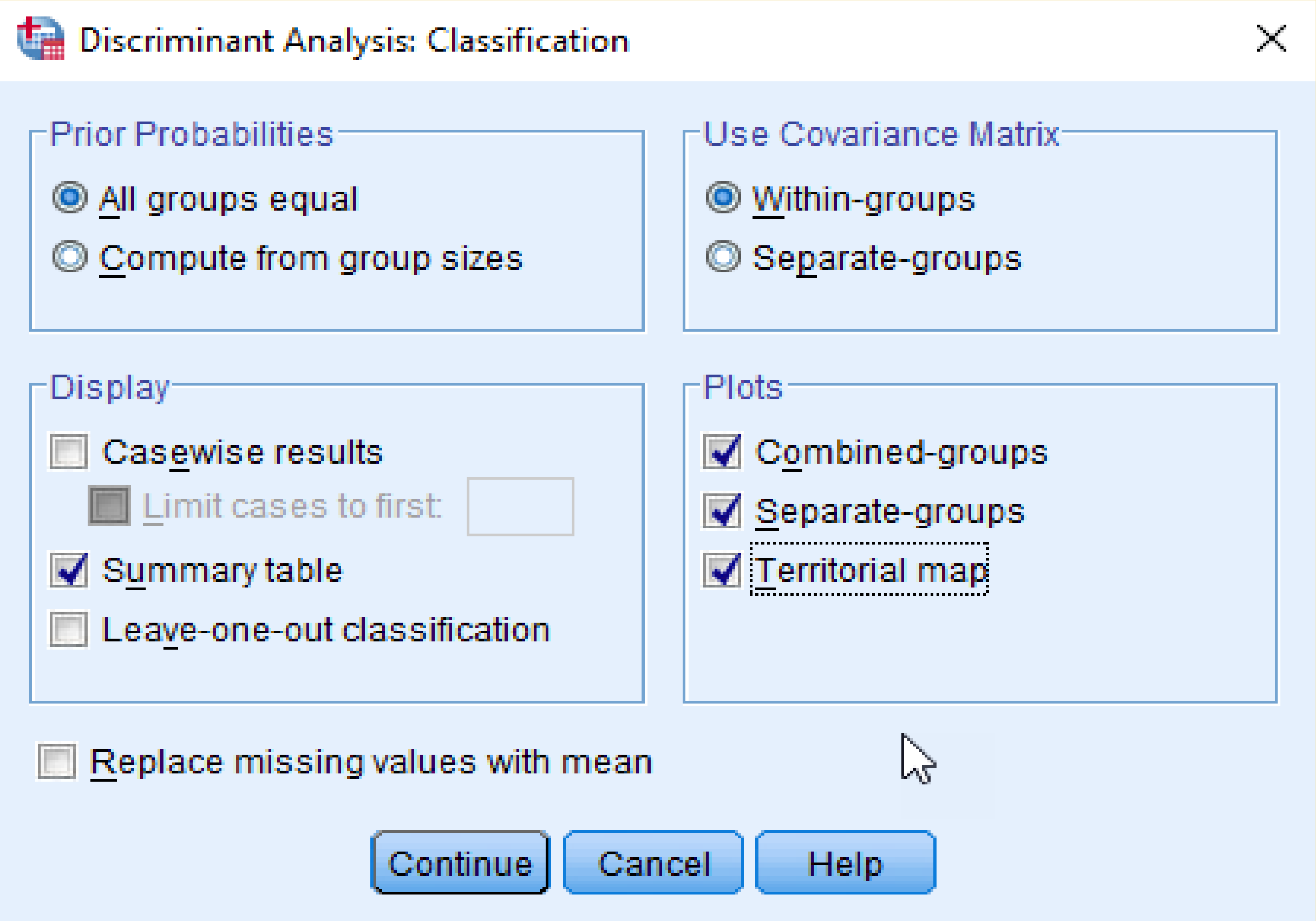
(3) Box’s M เป็นการนำเสนอผลการตรวจสอบข้อตกลงเบื้องต้น (Assumption) ในข้อที่ 2 ซึ่งเป็นการทดสอบเกี่ยวกับความเท่ากันของเมทริกซ์ความแปรปรวนร่วมของประชากร (Test of Population Covariance Matrices)

(4) ในช่องของ Fisher’s เป็นการนำเสนอค่าสัมประสิทธิ์ของตัวแปรในการจำแนกกลุ่ม (Classification Function Coefficient) หรือสัมประสิทธิ์การถดถอย ตามวิธีของ Fisher’s Linear Discriminant Function

(5) ในช่องของ Unstandardized เป็นการนำเสนอค่าสัมประสิทธิ์ของตัวแปรคาโนนิคอล (Canonical Discriminant Function Coefficients) ในรูปของคะแนนดิบ เพิ่มเติมจากการนำเสนอในรูปแบบของคะแนนมาตรฐาน

(6) ในช่องของ Matrices เป็นการนำเสนอเกี่ยวกับเมทริกซ์สหสัมพันธ์และเมทริกซ์ความแปรปรวนร่วมที่มีทั้งภายในกลุ่ม ระหว่างกลุ่ม และรวมกลุ่มเข้าด้วยกัน (Total)

5. จากภาพประกอบที่ 7.4 คลิก Continue จะกลับมาที่หน้าต่าง Discriminant Analysis ให้เลือก Classify โดยในเมนูย่อยของ Discriminant Analysis : Classification จะแบ่งออกเป็น 4 ส่วน ดังภาพประกอบที่ 7.5



ภาพประกอบที่ 7.5 แสดงผลจากการเลือกคำสั่ง Classify

จากภาพประกอบที่ 7.5 สามารถอธิบายรายละเอียดของสถิติต่าง ๆ ได้ดังนี้ (ทรงศักดิ์ ภูสีอ่อน. 2551 : 313 ; กัลยา วาณิชย์บัญชา. 2551 : 45)

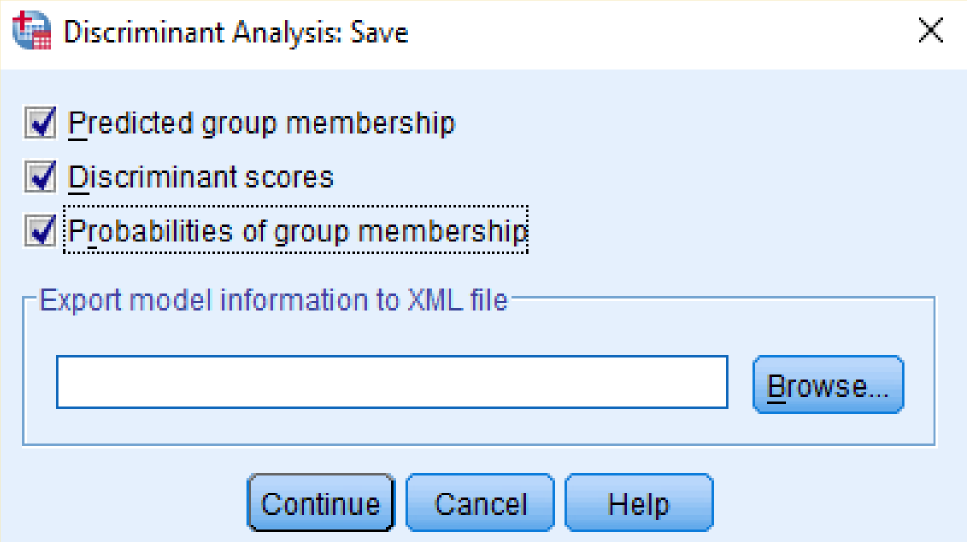
ส่วนที่ 1 Prior Probabilities เป็นการกำหนดโอกาสหรือความน่าจะเป็นของการเป็นกลุ่มต่าง ๆ ไว้ล่วงหน้าโดยแบ่งออกเป็น 2 กรณี คือ ให้โอกาสเท่ากันทุกกลุ่ม (All group equal) หรือคำนวณจากขนาดของกลุ่มตัวอย่าง (Compute from group sizes) ในที่นี้กำหนดให้ทุกกลุ่มมีโอกาสเท่ากัน

ส่วนที่ 2 Display เป็นส่วนของการนำเสนอผล โดยให้เลือกรูปแบบการนำเสนอผลที่มีหลายแบบให้เลือก

ส่วนที่ 3 Use Covariance Matrix เป็นการใช้เมทริกซ์ความแปรปรวนร่วม ซึ่งต้องเลือกใช้แบบใดแบบหนึ่งจาก Within – group กับ Separate – group

ส่วนที่ 4 Plots เป็นส่วนที่นำเสนอผลการวิเคราะห์ในรูปของแผนภาพ ซึ่งมีทั้งแบบรวมกลุ่ม (Combined – groups) แบบแยกกลุ่ม (Separate – group) และแบบเขตแดน (Territorial map) ซึ่งเป็นการแบ่งกลุ่มโดยใช้ตัวเลขกลุ่ม (เช่น 1, 2, 3..) เป็นสัญลักษณ์ในการแบ่งพรมแดนรายใดหรือหน่วยวิเคราะห์ใดตกอยู่ในพรมแดนก็เป็นสมาชิกของกลุ่มนั้น

6. จากภาพประกอบที่ 7.5 คลิก Continue จะกลับมาที่หน้าต่าง Discriminant Analysis ให้เลือก Save จะปรากฏหน้าต่าง Discriminant Analysis : Save ขึ้นมา จากนั้นให้เลือกว่าจะบันทึกข้อมูลอะไรบ้าง ดังภาพประกอบที่ 7.6



ภาพประกอบที่ 7.6 แสดงผลจากการเลือกคำสั่ง Save

จากภาพประกอบที่ 7.6 สามารถอธิบายรายละเอียดของรายการต่าง ๆ ในหน้าต่าง Discriminant Analysis : Save ได้ดังนี้คือ

(1) การเป็นสมาชิกของกลุ่มที่ได้จากการทำนาย (Predicted group membership) จะระบุความเป็นสมาชิกของกลุ่มต่าง ๆ โดยการทำนายจากการวิเคราะห์จำแนกกลุ่ม

(2) นำเสนอคะแนนจำแนก (Discriminant scores)

(3) หลังจากการวิเคราะห์จำแนกเรียบร้อยแล้วก็จะนำเสนอโอกาสหรือความน่าจะเป็นในการเป็นสมาชิกกลุ่มต่าง ๆ ของแต่ละบุคคลหรือแต่ละหน่วยวิเคราะห์ โดยผลการวิเคราะห์ทั้ง 3 รายการนั้นจะไม่ปรากฏใน Output แต่จะไปปรากฏใน SPSS for Windows Data Editor หรือใน file ข้อมูล ซึ่งจะสร้างตัวแรกต่อท้ายจากตัวแปรเดิม

7. จากภาพประกอบที่ 7.6 ให้คลิก Continue จะกลับไปที่หน้าต่าง Discriminant Analysis ดังภาพประกอบที่ 7.3 จากนั้นคลิกเลือก Ok จะปรากฏผลการวิเคราะห์ดังตาราง

ตารางที่ 7.1 แสดงตาราง Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
ผลการเรียนวิชาภาษาอังกฤษ	.836	5.474	1	28	.027
ผลการเรียนวิชาคณิตศาสตร์	.916	2.551	1	28	.121
ผลการเรียนวิชาวิทยาศาสตร์	.992	.215	1	28	.647

จากตารางที่ 7.1 ตาราง Tests of Equality of Group Means เป็นผลการวิเคราะห์ที่ได้จากคำสั่ง Univariate ANOVAs ใน Statistics Option ซึ่งเป็นการนำเสนอผลการทดสอบความเท่ากันของค่าเฉลี่ยของตัวแปรแต่ละกลุ่มในการทดสอบนั้น โดยการวิเคราะห์ความแปรปรวนทางเดียว (One – way ANOVA) และให้ค่าวิลค์แลมป์ดา ซึ่งทั้งสองการทดสอบต้องอาศัยค่าสถิติ F ทดสอบเช่นเดียวกัน สำหรับการพิจารณาจะพิจารณาจากค่านัยสำคัญ (Sig.) ของค่า F โดยถ้าค่า Sig. มีค่าต่ำกว่าระดับนัยสำคัญที่ตั้งไว้ จะสรุปได้ว่ากลุ่ม 3 กลุ่มนั้น มีค่าเฉลี่ยในตัวแปรนั้น ๆ แตกต่างกัน โดยจากตาราง Tests of Equality of Group Means นี้ พบว่า ที่ระดับนัยสำคัญ 0.05 มีค่าเฉลี่ยในตัวแปรผลการเรียนวิชาภาษาอังกฤษ : English (Sig = 0.027) แตกต่างกัน ส่วนตัวแปรผลการเรียนวิชาคณิตศาสตร์ : Math (Sig = 0.121) และ ผลการเรียนวิชาวิทยาศาสตร์ : Science (Sig = 0.647) มีค่าเฉลี่ยในตัวแปรไม่แตกต่างกัน

ตารางที่ 7.2 แสดงตาราง Pooled Within-Groups Matrices

		ผลการเรียนวิชาภาษาอังกฤษ	ผลการเรียนวิชาคณิตศาสตร์	ผลการเรียนวิชาวิทยาศาสตร์
Covariance	ผลการเรียนวิชาภาษาอังกฤษ	260.943	-128.686	-67.521
	ผลการเรียนวิชาคณิตศาสตร์	-128.686	141.319	76.552
	ผลการเรียนวิชาวิทยาศาสตร์	-67.521	76.552	419.419
Correlation	ผลการเรียนวิชาภาษาอังกฤษ	1.000	-.670	-.204
	ผลการเรียนวิชาคณิตศาสตร์	-.670	1.000	.314
	ผลการเรียนวิชาวิทยาศาสตร์	-.204	.314	1.000

a. The covariance matrix has 28 degrees of freedom.

จากตารางที่ 7.2 ตาราง Pooled Within-Groups Matrices เป็นตารางที่นำเสนอค่าความแปรปรวนร่วม (Covariance) ของทั้ง 2 กลุ่มรวมกันจาก 3 ตัวแปร และนำเสนอค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation) ซึ่งจากตาราง พบว่า ตัวแปรทั้ง 3 มีความสัมพันธ์กันทั้งทางบวกและทางลบ ดังนี้ คือ ตัวแปรผลการเรียนวิชาคณิตศาสตร์ : Math มีความสัมพันธ์ทางบวกกับ ตัวแปรผลการเรียนวิชาวิทยาศาสตร์ : Science (0.314) ส่วนตัวแปรผลการเรียนวิชาภาษาอังกฤษ : English กับตัวแปรผลการเรียนวิชาวิทยาศาสตร์ : Science มีความสัมพันธ์กันในทางลบ (-0.204) และ ตัวแปรผลการเรียนวิชาภาษาอังกฤษ : English กับตัวแปรผลการเรียนวิชาคณิตศาสตร์ : Math มีความสัมพันธ์กันในทางลบค่อนข้างสูง (-0.670)

ตารางที่ 7.3 แสดงตาราง Test Results

Box's M		5.757
F	Approx.	1.771
	df1	3
	df2	141120.000
	Sig.	.150

Tests null hypothesis of equal population covariance matrices.

จากตารางที่ 7.3 ตาราง Test Results เป็นตารางที่ให้ค่า Box’s M ใน Statistics Option ซึ่งเป็นค่าสถิติที่ใช้ทดสอบข้อตกลงเบื้องต้นที่ทดสอบเกี่ยวกับความแปรปรวนร่วมของประชากร (Test of population covariance matrices) โดยจะพิจารณาจากค่าความมีนัยสำคัญ (Sigs) ของ F สำหรับหลักเกณฑ์ในการพิจารณาก็จะเหมือนกับที่กล่าวมาแล้ว โดยจากตารางพบว่ามีค่า Sig. เท่ากับ 0.150 ซึ่งมีค่ามากกว่าระดับนัยสำคัญ 0.05 จึงยอมรับสมมติฐาน H_0 หมายความว่า ข้อมูลที่นำวิเคราะห์เป็นไปตามข้อตกลงเบื้องต้นข้อที่ 2 ที่ว่า เมทริกซ์ความแปรปรวนของประชากรทั้งสองกลุ่มเท่ากัน

ตารางที่ 7.4 แสดงตาราง Variables Entered/Removed[a,b,c,d]

Variables Entered/Removed[a,b,c,d]

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	ผลการเรียนวิชาภาษาอังกฤษ	.836	1	1	28.000	5.474	1	28.000	.027
2	ผลการเรียนวิชาคณิตศาสตร์	.542	2	1	28.000	11.406	2	27.000	.000

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a. Maximum number of steps is 6.

b. Minimum partial F to enter is 3.84.

c. Maximum partial F to remove is 2.71.

d. F level, tolerance, or VIN insufficient for further computation.

จากตารางที่ 7.4 ตาราง Variables Entered/Removed เป็นตารางที่แสดงตัวแปรที่อยู่ในสมการที่ดีที่สุดที่จะนำไปสร้างสมการวิเคราะห์ต่อไป คือ ตัวแปรผลการเรียนวิชาภาษาอังกฤษ : English กับตัวแปรผลการเรียนวิชาคณิตศาสตร์ : Math ที่มีค่า Wilks' Lambda เท่ากับ 0.836 และ 0.542 ตามลำดับ ซึ่งมีค่า Sig. เท่ากับ 0.027 และ 0.000

ตารางที่ 7.5 ตารางแสดง Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.845[a]	100.0	100.0	.677

a. First 1 canonical discriminant functions were used in the analysis.

จากตารางที่ 7.5 ตาราง Eigenvalues เป็นการนำเสนอค่า Eigenvalue และ Canonical Correlation โดยเนื่องจากข้อมูลที่กำหนดให้มี 2 กลุ่ม จึงมีเพียงสมการเดียว และมีค่า Eigenvalues เพียงค่าเดียว ซึ่งมีค่าเท่ากับ 0.845 และมีค่าความสัมพันธ์คาโนนิคอล (Canonical Correlation) เท่ากับ 0.698 เมื่อนำค่า Canonical Correlation มายกกำลังสอง จะเป็นค่าที่แสดงให้เห็นว่าตัวแปรในสมการจำแนกกลุ่ม สามารถอธิบายความแปรปรวนของตัวแปรตามได้ร้อยละเท่าไร ซึ่งในกรณีนี้อธิบายได้ $(0.677)^2$ เท่ากับ 45.8329%

ตารางที่ 7.6 ตารางแสดงค่า Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.542	16.535	2	.000

จากตารางที่ 7.6 ตาราง Wilks' Lambda เป็นการนำเสนอค่าสถิติที่ใช้ทดสอบสมการจำแนกกลุ่มที่วิเคราะห์ได้ว่ามีนัยสำคัญทางสถิติหรือไม่ โดยในการพิจารณาจะพิจารณาที่ค่าความมีนัยสำคัญ (Sig.) ถ้าค่า Sig. น้อยกว่านัยสำคัญที่กำหนดไว้ แสดงว่าสมการจำแนกกลุ่มสามารถจำแนกได้อย่างมีนัยสำคัญทางสถิติ ซึ่งจากการวิเคราะห์จะเห็นได้ว่าสมการมีนัยสำคัญ (0.000 > 0.05) นั่นแสดงว่า สมการที่ได้จากการวิเคราะห์จำแนกกลุ่มสามารถจำแนกออกเป็น 2 กลุ่มได้

ตารางที่ 7.7 ตารางแสดงค่า Standardized Canonical Discriminant Function Coefficients

	Function
	1
ผลการเรียนวิชาภาษาอังกฤษ	1.273
ผลการเรียนวิชาคณิตศาสตร์	1.181

จากตารางที่ 7.7 ตาราง Standardized Canonical Discriminant Function Coefficients แสดงค่าสัมประสิทธิ์ (ค่าน้ำหนัก) ของตัวแปรจำแนกในสมการจำแนกกลุ่มซึ่งเป็นสมการมาตรฐาน เพราะไม่มีน้ำหนัก (constant) เป็นการนำเสนอค่าน้ำหนักของตัวแปรแต่ละตัว ซึ่งจะเห็นว่าตัวแปรผลการเรียนวิชาภาษาอังกฤษ : English มีค่าน้ำหนัก (1.273) ซึ่งมากกว่าค่าน้ำหนักตัวแปรผลการเรียนวิชาคณิตศาสตร์ : Math (1.181) แสดงว่าตัวแปรผลการเรียนวิชาภาษาอังกฤษ : English มีความสำคัญในการจำแนกกลุ่มในสมการจำแนกกลุ่ม ซึ่งจากผลการวิเคราะห์สามารถเขียนเป็นสมการจำแนกกลุ่มได้ดังนี้

$$Z_y = 1.273(Z_{English}) + 1.181(Z_{Math})$$

ตารางที่ 7.8 ตารางแสดง Structure Matrix

	Function
	1
ผลการเรียนวิชาภาษาอังกฤษ	.481
ผลการเรียนวิชาคณิตศาสตร์	.328
ผลการเรียนวิชาวิทยาศาสตร์[a]	.112

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

a. This variable not used in the analysis.

จากตารางที่ 7.8 ตาราง Structure Matrix สามารถนำไปใช้ตีความหมายสมการจำแนกกลุ่มอีกวิธีหนึ่ง โดยจะช่วยประเมินว่าตัวแปรแต่ละตัวมีผลอย่างไรต่อตัวแปรจำแนก โดยเป็นการดูความสัมพันธ์ระหว่างค่าของตัวแปรกับค่าคะแนนจำแนกที่คำนวณจากสมการจำแนก (ทรงศักดิ์ ภูสีอ่อน. 2551 : 317) ซึ่งจากตารางจะเห็นได้ว่าตัวแปรผลการเรียนวิชาภาษาวิทยาศาสตร์ : Science มีความสำคัญต่อสมการจำแนกน้อยมาก

ตารางที่ 7.9 ตารางแสดง Canonical Discriminant Function Coefficients

	Function
	1
ผลการเรียนวิชาภาษาอังกฤษ	.079
ผลการเรียนวิชาคณิตศาสตร์	.099
(Constant)	-8.905

Unstandardized coefficients

จากตารางที่ 7.9 ตาราง Canonical Discriminant Function Coefficients แสดงค่าเป็นสัมประสิทธิ์ของตัวแปรจำแนกในสมการจำแนกกลุ่ม ซึ่งเป็นสมการในรูปคะแนนดิบ ค่าน้ำหนักที่ได้จึงไม่อยู่ในรูปมาตรฐาน (Unstandardized coefficients) ซึ่งเป็นผลจากการเลือก Unstandardized ใน Statistics Option ผลที่ได้มีทั้งค่าน้ำหนักในแต่ละตัวแปร และค่าคงที่ (Constant) โดยจากผลการวิเคราะห์สามารถนำมาเขียนเป็นสมการในรูปคะแนนดิบได้ดังนี้

$$Y' = -8.905 + 0.079(\text{English}) + 0.099(\text{Math})$$

ตารางที่ 7.10 ตารางแสดง Functions at Group Centroids

กลุ่ม	Function
	1
1	-.888
2	.888

Unstandardized canonical discriminant functions evaluated at group means

จากตารางที่ 7.10 ตาราง Functions at Group Centroids เป็นตารางที่ให้ค่ากลางของกลุ่ม (Group Centroids) เป็นค่าที่สามารถใช้ประเมินสมการจำแนกคาโนนิคอลด้วยค่าเฉลี่ยของกลุ่ม (Canonical Discriminant Functions Evaluated at Group Means) ผลการวิเคราะห์เป็นคะแนนดิบ (Unstandardized) ซึ่งค่ากลางหาได้จากการหาคะแนนจำแนกแต่ละหน่วยวิเคราะห์ เมื่อได้คะแนนของแต่ละหน่วยจากสมการแล้วก็หาค่าเฉลี่ยหรือค่ากลางของแต่ละกลุ่มได้ โดยเอาผลรวมของค่าคะแนนจำแนกของแต่ละหน่วยในกลุ่มนั้นหารด้วนจำนวนหน่วยในกลุ่มนั้น ซึ่งจากตาราง พบว่า กลุ่ม 1 มีค่ากลางของกลุ่ม (Group Centroids) เท่ากับ -0.888 ส่วนกลุ่มที่ 2 มีค่า 0.888 ซึ่งแตกต่างกันมาก แสดงว่า สมการดังกล่าวสามารถจำแนกได้ดี ในกรณีที่มีหน่วยวิเคราะห์ใหม่ก็สามารถคำนวณหาคะแนนจำแนกแล้วนำเปรียบเทียบกับค่ากลางของแต่ละกลุ่ม ถ้ามีแนวโน้มเข้าใกล้ค่ากลางกลุ่มใดก็มีโอกาสในการเป็นสมาชิกกลุ่มนั้น ๆ

ตารางที่ 7.11 ตารางแสดง Classification Function Coefficients

	กลุ่ม	
	1	2
ผลการเรียนวิชาภาษาอังกฤษ	.585	.725
ผลการเรียนวิชาคณิตศาสตร์	.889	1.066
(Constant)	-34.276	-50.092

Fisher's linear discriminant functions

จากตารางที่ 7.11 ตาราง Classification Function Coefficients เป็นตารางแสดงผลการวิเคราะห์นำเสนอค่าสัมประสิทธิ์ (ค่าน้ำหนัก) และค่าคงที่ของสมการจำแนกโดยแยกเป็นกลุ่มตามวิธีของ Fisher’s จำนวนสมการจะมีเท่ากับจำนวนกลุ่ม โดยจากผลการวิเคราะห์จะได้สมการดังนี้

สมการของกลุ่ม 1

$$Y_1^{'} = -34.276 + 0.585English + 0.889Math$$

สมการของกลุ่ม 2

$$Y_2^{'} = -50.092 + 0.725English + 1.066Math$$

โดยจากผลการวิเคราะห์สามารถนำไปใช้ประโยชน์ในการพยากรณ์หน่วยวิเคราะห์ว่าควรจัดให้อยู่ในกลุ่มใด โดยการแทนค่าตัวแปรอิสระของหน่วยวิเคราะห์นั้น ๆ ลงในทั้ง 2 สมการ ถ้าสมการใดมีค่ามากกว่าก็จัดอยู่ในกลุ่มนั้น

ตารางที่ 7.12 ตารางแสดง Classification Results[a]

Original Count

Classification Results[a]

		กลุ่ม	Predicted Group Membership		Total
			1	2	
Original	Count	1	13	2	15
		2	5	10	15
	%	1	86.7	13.3	100.0
		2	33.3	66.7	100.0

a. 76.7% of original grouped cases correctly classified.

จากตารางที่ 7.12 แสดงประสิทธิภาพของสมการจำแนกว่าสามารถจำแนกกลุ่มได้ถูกต้องมากน้อยเพียงใด โดยเป็นการเปรียบเทียบกลุ่มที่แบ่งไว้เดิม (Original) กับการแบ่งกลุ่มที่ได้จากการทำนายจากสมการ (Predicted Group Membership) โดยจากตาราง 7.12 พบว่า ในกลุ่มที่ 1 เดิมมี 15 cases แต่จากการทำนายโดยใช้สมการจำแนกกลุ่ม พบว่า ทำนายได้ถูกต้อง 13 case คิดเป็นร้อยละ 86.7 ส่วนในกลุ่มที่ 2 เดิมมี 15 case แต่ทำนายโดยใช้สมการจำแนกกลุ่มได้ถูกต้อง 10 case คิดเป็นร้อยละ 33.3 เมื่อคิดรวมทั้งหมด (ทั้ง 30 case) พบว่าสมการจำแนกกลุ่มได้ถูกต้องร้อยละ 76.7

สำหรับผลจากการเลือกคำสั่ง Save จะไม่ปรากฏใน Output แต่จะปรากฏในส่วนของ Data View โดยสร้างเป็นตัวแปรต่อจากข้อมูลเดิมจำนวน 4 ตัวแปร ดังภาพประกอบที่ 7.7

	Group	Dis_1	Dis_2	Dis1_1	Dis1_2	Dis2_2	var
1	1	1	1	-1.06700	.86933	.13067	
2	1	1	1	-1.21818	.89692	.10308	
3	1	1	1	-.65180	.76090	.23910	
4	1	2	2	.63914	.24322	.75678	
5	1	1	1	-1.83774	.96317	.03683	
6	1	1	1	-1.47081	.93164	.06836	
7	1	1	1	-2.85192	.99373	.00627	
8	1	1	1	-.44316	.68720	.31280	
9	1	1	1	-.08333	.53693	.46307	
10	1	1	1	-1.60423	.94527	.05473	
11	1	1	1	-.72776	.78457	.21543	
12	1	1	1	-.83067	.81386	.18614	
13	1	1	1	-.32959	.64230	.35770	
14	1	1	1	-.86191	.82212	.17788	
15	1	2	2	.01885	.49163	.50837	
16	2	1	1	-.38852	.66597	.33403	

Data View | Variable View

ภาพประกอบที่ 7.7 แสดงผลจากการเลือกคำสั่ง Save

จากภาพประกอบที่ 7.7 สามารถอธิบายตัวแปรแต่ละตัวได้ดังนี้

1. ตัวแปร Dis_1 เป็นตัวแปรที่ระบุถึงการเป็นสมาชิกกลุ่มของหน่วยวิเคราะห์ที่ได้จากการทำนายของสมการจำแนกกลุ่ม

2. ตัวแปร Dis1_1 เป็นตัวแปรที่บอกถึงคะแนนจำแนกของหน่วยวิเคราะห์ โดยทำนายจากค่าน้ำหนักและค่าที่วัดได้จากตัวแปร (English และ Math) ในสมการจำแนก

3. ตัวแปร Dis1_2 เป็นตัวแปรที่บอกถึงโอกาสในการเป็นสมาชิกกลุ่ม 1 ของแต่ละหน่วยวิเคราะห์

4. ตัวแปร Dis2_2 เป็นตัวแปรที่บอกถึงโอกาสในการเป็นสมาชิกกลุ่ม 2 ของแต่ละหน่วยวิเคราะห์

บทสรุป

การวิเคราะห์จำแนกกลุ่ม (Discriminant Analysis) เป็นวิธีการทางสถิติ ที่ใช้วิเคราะห์จำแนกกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป โดยมีตัวแปรตาม 1 ตัวและตัวแปร อิสระตั้งแต่ 1 ตัวขึ้นไป ซึ่งการวิเคราะห์จำแนกกลุ่ม นอกจากจะสามารถจำแนกความแตกต่างระหว่าง กลุ่มได้แล้วยังสามารถบอกธรรมชาติบางอย่างของการจำแนกกลุ่มได้ด้วย เช่น บอกได้ว่าตัวแปรใดจำแนกได้ดีมากน้อยกว่ากัน สำหรับการสร้างสมการจำแนกกลุ่มจะ มี 2 วิธี คือ วิธีตรง (Direct Method) ซึ่งเป็นวิธีการ ที่ผู้วิจัยต้องการตัวแปรทุกตัว ตามที่ระบุไว้ด้วยเหตุผลทางทฤษฎีว่าจะแบ่งแยกได้กี่สมการมีลักษณะอย่างไร เพื่อพิสูจน์ตัวแปรที่คิดว่ามีความสำคัญต่อการจำแนกที่ระบุไว้ตามทฤษฎีนั้นแท้จริงแล้วมีความสำคัญหรือไม่และ วิธีแบบขั้นตอน (Stepwise Method) ซึ่งเป็นวิธีการที่เลือกตัวแปรทีละตัวมาเข้าสมการโดยหาตัวแปรที่ดีที่สุดในการจำแนกมาเข้าสมการเป็นตัวแรก จากนั้นก็จะหาตัวแปรที่ดีที่สุดตัวที่สองมาเข้าสมการเพื่อปรับปรุงแก้ไขทำให้สมการจำแนกกลุ่มดีขึ้น และในขั้นตอนต่อ ๆ ไปก็จะเป็นการนำตัวแปรที่ดีที่สุดแต่ละตัวที่เหลือมาเข้าสมการต่อไปเพื่อจะได้สมการจำแนกกลุ่มที่ดีที่สุด