บทที่ 2
ทบทวนความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในบทนี้จะกล่าวถึงความรู้พื้นฐานที่สำคัญซึ่งมีความจำเป็นต้องใช้ในงานวิจัยครั้งนี้ ได้แก่ การวิเคราะห์จำแนกกลุ่มเชิงเส้น, การวิเคราะห์ส่วนประกอบแกนมุขสำคัญ, การวิเคราะห์ความแปรปรวน, ขั้นตอนการปรับแนวแบบพีซไวส์, การตรวจสอบความสมเหตุสมผลแบบไขว้และการวิเคราะห์ความแปรปรวนแบบทางเดียวของค่าอันดับของครัสคาล-วอลิส นอกจากนี้ในท้ายบทจะกล่าวถึงงานวิจัย และข้อมูลเบื้องต้นของน้ำมันเบนซินที่เกี่ยวข้อง

2.1 ความรู้พื้นฐานที่เกี่ยวข้อง

2.1.1 การวิเคราะห์จำแนกกลุ่มเชิงเส้น (Linear Discriminant Analysis, LDA)

การวิเคราะห์จำแนกกลุ่มเชิงเส้นเป็นเทคนิควิเคราะห์ความสัมพันธ์ โดยประกอบด้วยตัวแปรตาม 1 ตัว ซึ่งเป็นตัวแปรเชิงคุณภาพ เรียกว่า ตัวแปรแสดงกลุ่ม และมีตัวแปรอิสระ p ตัว การจำแนกกลุ่มเชิงเส้นเป็นเทคนิคที่ใช้ในการแบ่งกลุ่ม คน สัตว์ สิ่งของ องค์กร ฯลฯ ออกเป็นกลุ่ม ๆ อย่างน้อย 2 กลุ่มขึ้นไป โดยผู้ศึกษาจะต้องทราบมาก่อนว่าแต่ละหน่วยอยู่กลุ่มใด โดยที่หน่วยตัวอย่างจัดอยู่ในกลุ่มใดกลุ่มหนึ่งเท่านั้น (Johnson,2002)

การวิเคราะห์จำแนกกลุ่มเป็นเทคนิคที่สร้างฟังก์ชันเชิงเส้นของตัวแปรอิสระ $(X_1, X_2, ..., X_p)$ ที่จะทำให้หน่วยที่อยู่กลุ่มต่างกันมีความแตกต่างกันมากที่สุด เมื่อเทียบกับความแตกต่างของหน่วยที่อยู่ภายในกลุ่มเดียวกัน

<u>สถิติสำคัญของการวิเคราะห์จำแนกกลุ่ม</u> (สมประสงค์, 2553)

1. ค่าไอเกน (Eigenvalue) เป็นค่าที่แสดงอัตราส่วนการผันแปรระหว่างกลุ่มต่อการผันแปรภายในกลุ่ม ถ้าค่าไอเกนมีค่าสูง ก็แสดงว่าสมการดีหรือมีค่าจำแนกสูงหรือกล่าวอีกนัยหนึ่งได้ว่า Eigenvalue ก็คือ Variance ของคะแนนแปลงรูป Y ที่แปลงมาจาก $X_1, X_2, ..., X_p$ นั่นเอง

2. ค่าสหสัมพันธ์คาโนนิคอล (Canonical Correlation) เป็นสถิติซึ่งสามารถใช้ในการตัดสินความสำคัญของสมการจำแนก เป็นมาตรวัดความสัมพันธ์ของสมการกับกลุ่มของตัวแปรซึ่งระบุการเป็นสมาชิกของกลุ่มนั้น ๆ ของตัวแปรตาม โดยชี้ให้เห็นว่าการเป็นสมาชิกกลุ่มมีความสัมพันธ์กับสมการที่หามาได้มากน้อยเพียงใด ดังนั้น ถ้าค่าสหสัมพันธ์คาโนนิคอลมีค่าสูง แสดงว่า การเป็นสมาชิกของกลุ่มสามารถอธิบายความผันแปรของตัวแปรกับสมการจำแนกได้มาก

3. ค่าวิลค์แลมบ์ดา (Wilks' Lambda) เป็นสถิติที่ใช้ทดสอบความแตกต่างระหว่างกลุ่ม และเป็นมาตรวัดอำนาจในการจำแนกกลุ่มของตัวแปรด้วย ถ้าค่าวิลค์แลมบ์ดามีค่ามาก ตัวแปรจะอธิบายการเป็นสมาชิกของกลุ่มได้น้อย ถ้าค่าวิลค์แลมบ์ดามีค่าน้อย ตัวแปรจะอธิบายการเป็นสมาชิกของกลุ่มได้มาก

<u>วัตถุประสงค์ของการวิเคราะห์จำแนกกลุ่ม</u> (นิรัติยา, 2551)

1. เพื่อหาสมการเชิงเส้น หรือฟังก์ชันจำแนกกลุ่ม ซึ่งแสดงความสัมพันธ์ของตัวแปรแยกกลุ่ม หรือตัวแปรตามกับตัวแปรอิสระอย่างน้อย 1 ตัว โดยการสร้างสมการเชิงเส้นดังกล่าวจะต้องใช้ข้อมูลจริงที่ทราบกลุ่มตัวอย่างแล้ว

2. เพื่อนำสมการ หรือฟังก์ชันจำแนกกลุ่มที่สร้างในข้อ 1 มาใช้ในการพยากรณ์ว่าข้อมูลใหม่ที่ยังไม่ทราบกลุ่ม ว่าควรจะอยู่ในกลุ่มใด

3. เพื่อพิจารณาว่าตัวแปรอิสระตัวใดบ้างเป็นตัวแปรที่สำคัญที่ใช้ในการแบ่งกลุ่ม

4. สามารถใช้ฟังก์ชันจำแนกกลุ่มที่สร้างในข้อ 1 มาใช้ในการประเมินร้อยละความถูกต้องของการจำแนกกลุ่ม

<u>ลักษณะของข้อมูลที่ใช้ในการวิเคราะห์จำแนกกลุ่ม</u>

1. ตัวแปรอิสระหรือตัวแปรจำแนกกลุ่มควรเป็นตัวแปรเชิงปริมาณ กรณีที่ตัวแปรอิสระเป็นตัวแปรเชิงกลุ่ม (Categorical Variable) หรือตัวแปรเชิงคุณภาพจะต้องปรับให้อยู่ในรูปตัวแปรหุ่น (Dummy Variable) ที่มีค่าเป็น 0 หรือ 1

2. ตัวแปรตามจะต้องเป็นตัวแปรเชิงกลุ่ม ซึ่งอาจแบ่งเป็น 2 กลุ่มหรือมากกว่า ซึ่งอาจมีรหัสเป็น 0, 1 หรือ 1, 2 หรือถ้ามี 3 กลุ่ม ก็อาจใช้รหัส 1, 2, 3 ก็ได้

<u>สมมติฐานเบื้องต้นของการวิเคราะห์จำแนกกลุ่ม</u>

1. ตัวแปรอิสระ $(X_1, X_2, \ldots, X_p)$ ของแต่ละกลุ่มต้องมีการแจกแจงแบบปกติพหุตัวแปร (Multivariate Normal Distribution) หรือ

$X_i \sim N(\mu_i, \Sigma_i)$; $i = 1, 2, \ldots, k$; $k \geq 2$

2. เมตริกซ์ค่าความแปรปรวนร่วมของตัวแปรอิสระของทุกกลุ่มเท่ากัน

$\Sigma_1 = \Sigma_2 = \ldots = \Sigma_k = \Sigma$ กรณีที่มี k กลุ่ม

โดย Σ_i = เมตริกซ์ค่าความแปรปรวนร่วมของกลุ่มที่ i ที่มีขนาด $p \times p$;

$i = 1, 2, \ldots, k$; $k \geq 2$

<u>วิธีการสร้างฟังก์ชันการจำแนกกลุ่ม</u>

การสร้างฟังก์ชันการจำแนกกลุ่มที่นิยมใช้มี 2 วิธี คือ วิธีตรง และวิธีแบบขั้นตอน

1. วิธีตรง (Direct Method) เป็นวิธีใช้ตัวแปรอิสระทุกตัวเข้าสมการพร้อมกัน หรือเป็นการสร้างฟังก์ชันจำแนกโดยไม่คำนึงว่าตัวแปรอิสระตัวใดจะมีอำนาจจำแนกเพียงใดต่อตัวแปรตาม

2. วิธีแบบขั้นตอน (Stepwise Method) เป็นวิธีที่ทำการเลือกตัวแปรอิสระทีละตัวเข้าสมการสร้างฟังก์ชันจำแนก การคัดเลือกตัวแปรอิสระอยู่บนพื้นฐานของการมีส่วนช่วยเพิ่มอำนาจจำแนกให้แก่ฟังก์ชันเป็นสำคัญ ดังนั้นตัวแปรอิสระจะถูกเลือกถ้าสามารถเพิ่มอำนาจการจำแนกของฟังก์ชันได้อย่างมีนัยสำคัญ วิธีนี้จะเป็นประโยชน์สำหรับการวิเคราะห์ที่มีจำนวนตัวแปรอิสระจำนวนมาก

<u>การสร้างฟังก์ชันการจำแนกกลุ่ม</u>

ฟังก์ชันการจำแนกกลุ่มซึ่งอยู่ในรูปของผลรวมเชิงเส้นของตัวแปรจะเขียนได้ดังสมการ

$$y_g = a_{1g}x_1 + a_{2g}x_2 + \ldots + a_{jg}x_j + \ldots + a_{pg}x_p \qquad 2.1$$

โดยที่ y_g คือ ค่าคะแนนจำแนกกลุ่ม (Discriminant Scores) ของฟังก์ชันการจำแนกกลุ่มที่ g

a_{jg} คือ ค่าสัมประสิทธิ์การจำแนกกลุ่ม (Discriminant Weight) ของตัวแปรตัวที่ j ฟังก์ชันการจำแนกกลุ่มที่ g

x_j คือ ค่าของตัวแปรตัวที่ j

เขียนในรูปเมตริกซ์ได้เป็น

$$Y_g = a_g^T X \qquad 2.2$$

เมื่อใช้ข้อมูลตัวอย่างมาประมาณสมการที่ 2.1 จะได้

$$y_g = a_{1g}x_1 + a_{2g}x_2 + \ldots + a_{pg}x_p \qquad 2.3$$

เขียนในรูปเมตริกซ์ได้เป็น

$$\hat{Y}_g = \hat{a}_g^T X \qquad 2.4$$

โดยที่

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1j} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2j} & \ldots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{l1} & x_{l2} & \ldots & x_{lj} & \ldots & x_{lp} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nj} & \ldots & x_{np} \end{bmatrix} \qquad 2.5$$

ซึ่ง X เป็นเมตริกซ์ข้อมูล มี p เป็นจำนวนของตัวแปร และ n เป็นจำนวนของกลุ่มตัวอย่าง และ $\hat{Y}_g$ เป็นเวกเตอร์ของกลุ่มข้อมูลที่ต้องการพยากรณ์กลุ่ม

ในการประมาณค่าสัมประสิทธิ์ a ด้วย $\hat{a}$ ใช้หลักการที่ว่าจะหาค่า $\hat{a}$ ที่ทำให้

$$\lambda = \frac{\text{Between group sum of square}}{\text{Within group sum of square}} = \frac{\hat{a}^T B\hat{a}}{\hat{a}^T W\hat{a}}$$ มีค่ามากที่สุด

หรือกล่าวว่า ให้อัตราส่วนของความแตกต่างระหว่างกลุ่มต่อความแตกต่างภายในกลุ่มมีค่ามากที่สุด ซึ่งแสดงว่า ข้อมูลที่อยู่ต่างกลุ่มมีความแตกต่างกันมากกว่าข้อมูลที่อยู่ในกลุ่มเดียวกัน

โดยที่ B เป็นเมตริกซ์ผลรวมกำลังสองของค่าความแตกต่างระหว่างค่าเฉลี่ยของกลุ่มกับค่าเฉลี่ยทั้งหมด (Sample Between Groups Matrix)

W เป็นเมตริกซ์ผลรวมกำลังสองของค่าความแตกต่างระหว่างค่าของข้อมูลกับค่าเฉลี่ยภายในกลุ่ม (Sample Within Groups Matrix)

สามารถคำนวณหาค่า B และ W ดังได้นี้

$$B = \sum_{i=1}^{k} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \qquad 2.6$$

และ

$$W = \sum_{i=1}^{k} \sum_{l=1}^{n_i} (\bar{x}_{li} - \bar{x}_i)(\bar{x}_{li} - \bar{x}_i)^T \qquad 2.7$$

โดยที่

$$\bar{x}_i = \frac{1}{n_i} \sum_{i=1}^{n_i} x_{il} \qquad 2.8$$

และ

$$\bar{x} = \frac{\sum_{i=1}^{n_i} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} \sum_{l=1}^{n_i} x_{il}}{\sum_{i=1}^{k} n_i} \qquad 2.9$$

โดยที่ $\bar{x}$ คือ ค่าเฉลี่ยของตัวแปรอิสระทั้งหมด

$\bar{x}_i$ คือ ค่าเฉลี่ยของตัวแปรอิสระกลุ่มที่ i

จากเงื่อนไขที่กำหนดไว้ในตอนต้น กล่าวว่าฟังก์ชันการจำแนกกลุ่มที่จะแยกกลุ่มข้อมูลได้ดีที่สุดนั้นจะต้องให้ค่า λ มากที่สุดซึ่ง

$$\lambda = \frac{\hat{a}^T B\hat{a}}{\hat{a}^T W\hat{a}} \qquad 2.10$$

λ เป็นค่าที่ใช้เป็นเกณฑ์ที่เรียกว่า เกณฑ์ของการจำแนกกลุ่ม (Discriminant Criterion) การคิดแก้สมการเพื่อหาค่า λ ใช้สมการ

$$\left|W^{-1}B - \lambda I\right| = 0 \qquad 2.11$$

เมื่อ W^{-1} คือ เมตริกซ์ผกผันของ W

I คือ เมตริกซ์เอกลักษณ์

λ เป็นค่าที่แทนค่าเฉพาะค่าที่มากที่สุดเมื่อมีค่าเฉพาะหลาย ๆ ค่า

เพื่อที่จะประมาณค่า $\hat{a}$ จึงต้องนำเมตริกซ์ $\hat{a}$ คูณเข้าไปในสมการ 2.11 จะได้

$$\left|W^{-1}B - \lambda I\right|\hat{a} = 0 \qquad 2.12$$

ต่อไปจะแก้สมการหาค่า $\hat{a}$ โดยที่ $\hat{a}$ จากสมการ 2.12 จะเป็นไอเกนเวกเตอร์ของเมตริกซ์ $W^{-1}B$ ที่สอดคล้องกับ λ ก็ต่อเมื่อ $\hat{a}$ มีคำตอบไม่เท่ากับศูนย์ (Nontrivial Solution)

ดังนั้น ค่าของ λ ก็คือค่าไอเกนของเมตริกซ์ $W^{-1}B$ และ $\hat{a}$ ก็คือไอเกนเวกเตอร์ที่ได้จากค่าไอเกนที่มีค่ามากที่สุดนั่นเอง

จากฟังก์ชันการจำแนกกลุ่มในสมการ 2.3 ค่าน้ำหนักการจำแนกกลุ่ม ($\hat{a}$) จะถูกนำไปแทนที่ในสมการ และผลจากผลรวมเชิงเส้นของตัวอย่างข้อมูลทั้งหมดจะทำให้ได้เวกเตอร์ของค่าคะแนนการจำแนกกลุ่ม $\hat{Y}_g$ ที่กลุ่มข้อมูลมีการจำแนกได้ดีที่สุด นอกจากนี้ค่าคะแนนการจำแนกกลุ่มดังกล่าวยังจะถูกนำไปใช้เป็นตัวตัดสินในการทำนายกลุ่มของข้อมูล ซึ่งถือเป็นอีกเป้าหมายหนึ่งของการวิเคราะห์

<u>เกณฑ์ในการพยากรณ์กลุ่ม</u>

สำหรับวิธีการที่ใช้ในการพยากรณ์กลุ่มของข้อมูลที่ไม่ทราบค่า เพื่อจำแนกกลุ่มข้อมูลนั้นจะใช้เทคนิค Linear classification functions โดยวิธีการนี้จะใช้ฟังก์ชันการจำแนกกลุ่ม โดยการแทนค่าของตัวแปรอิสระหรือตัวแปรจำแนกกลุ่มของตัวอย่างใหม่ แล้วคำนวณหาค่า Discriminant score (y_g) ถ้าค่า y_g ของกลุ่มใด มีค่ามากที่สุด จะจัดให้ตัวอย่างอยู่ในกลุ่มนั้น (กัลยา, 2546)

2.1.2 การวิเคราะห์ส่วนประกอบแกนมุขสำคัญ (Principal Component Analysis, PCA)

การวิเคราะห์ส่วนประกอบแกนมุขสำคัญเป็นเทคนิคในการลดตัวแปรเทคนิคหนึ่ง (Richard,2002) โดยการสร้างเซตของตัวแปรใหม่ให้เป็นฟังก์ชันเชิงเส้นของตัวแปรเดิม และเซตของตัวแปรใหม่จะมีรายละเอียด หรือข้อมูลของตัวแปรเดิม โดยจำนวนตัวแปรใหม่จะต้องไม่เกิน

จำนวนตัวแปรเดิม นั่นคือ กรณีที่มีตัวแปรเดิม p ตัว จำนวนตัวแปรใหม่เท่ากับ m ตัว จะได้ว่า $m \leq p$ และจะต้องดึงรายละเอียดหรือค่าความแปรปรวนมาไว้ในตัวแปรใหม่ให้มากที่สุด ซึ่งในตัวแปรใหม่จะประกอบด้วยตัวประกอบหลักแรก (PC_1) ที่มีความผันแปรของตัวแปรเดิมมากที่สุด และความผันแปรของตัวประกอบหลักตัวที่สอง, ตัวที่สาม, ..., จนถึงตัวที่ p จะมีความแปรผันของตัวแปรเดิมลดลงตามลำดับ นอกจากนั้นตัวประกอบหลักแต่ละตัวจะไม่มีความสัมพันธ์กันเอง และตัวประกอบหลักจึงมีจำนวนน้อยกว่าตัวแปรเดิม โดยตัวประกอบหลักที่สร้างขึ้นถือเป็นตัวแปรใหม่และสามารถนำตัวแปรใหม่ที่สร้างขึ้นไปทำการวิเคราะห์ด้วยเทคนิคอื่น ๆ ทางสถิติต่อไป

หลักการของการวิเคราะห์ส่วนประกอบแกนมุขสำคัญ

1. หาเมตริกซ์ความแปรปรวนร่วม S หรือเมตริกซ์สหสัมพันธ์ของกลุ่มตัวอย่าง R
2. หาค่าไอเกนและไอเกนเวกเตอร์ของเมตริกซ์ในข้อ 1
3. เลือกจำนวนองค์ประกอบหลักที่เหมาะสมเพื่อลดมิติของข้อมูล

ให้ PC_i แทนตัวประกอบหลักที่ i; i= 1, 2, …, p

$$PC_i = w_{i1}X_1 + w_{i2}X_2 + ... + w_{ip}X_p$$

หรือ $PCi = w_i^{/}X$ ที่ทำให้ $Var\left(w_i^{/}X\right)$ มีค่ามากที่สุด

ซึ่ง $w_i^{/}w_i = w_{i1}^2 + w_{i2}^2 + ... + w_{ip}^2 = 1$

โดยที่ ไอเกนเวกเตอร์ที่ i (w_i)เป็นน้ำหนักของตัวแปรเดิมในตัวประกอบหลักที่ i (PC_i) และ w_{ip} เป็นน้ำหนักของตัวแปรเดิมที่ i ในการสร้างตัวแปรหลักที่ i (PC_i)

การพิจารณาจำนวนตัวประกอบหลักที่เหมาะสม หรือพิจารณาค่า m ว่าควรเป็นเท่าใด จะพิจารณาจากร้อยละความผันแปรที่สามารถอธิบายตัวแปรเดิมได้ ซึ่งขึ้นอยู่กับค่าความสัมพันธ์ของตัวแปรเดิม ถ้าตัวแปรเดิม p ตัว มีความสัมพันธ์กันมาก จะทำให้ได้ตัวแปรใหม่ที่มีค่าไอเกนหรือค่าความแปรปรวนมากเพียงไม่กี่ค่า แต่ถ้าตัวแปรเดิม มีความสัมพันธ์กันน้อย จำนวนตัวประกอบหลักจะเท่ากับ หรือใกล้เคียงกับตัวแปรเดิม โดยมีแนวทางดังนี้

1. พิจารณาจำนวนร้อยละความแปรปรวนสะสม ถ้าร้อยละความแปรปรวนสะสมของตัวประกอบหลัก m ตัวแรก เป็นอย่างต่ำร้อยละ 80 ก็ควรให้จำนวนตัวประกอบหลักเท่ากับ m โดยที่ $m < p$

2. ใช้กราฟ Scree ในการพิจารณาจำนวนตัวประกอบหลักที่เหมาะสม โดยการพล็อตค่าไอเกน ถ้าตัวประกอบหลักที่ m+1 มีค่าไอเกนต่ำมากเมื่อเทียบกับตัวที่ m หรือค่าไอเกนลดลงอย่างรวดเร็ว ก็ไม่ควรพิจารณาตัวประกอบหลักที่ m+1, …, p หรือควรมีองค์ประกอบหลัก m ตัวเท่านั้น

3. พิจารณาเฉพาะตัวประกอบหลักที่มีค่าไอเกนมากกว่าหนึ่ง มีการศึกษาของ (Cliff , 1988) ที่พบว่าการใช้ค่าความแปรปรวนหรือค่าไอเกนของตัวประกอบหลักในการพิจารณาเฉพาะตัว

ประกอบหลักที่มีค่าไอเกนมากกว่าหนึ่งนั้น อาจทำให้จำนวนตัวประกอบหลักที่ได้มากหรือน้อยจนเกินไป ดังนั้นควรจะใช้หลักเกณฑ์อื่นๆ มาร่วม

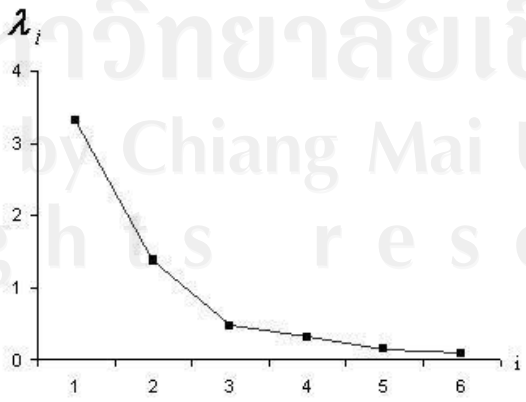
ตัวอย่างเช่น ถ้ามีจำนวนตัวแปรเดิม 6 ตัว (p=6) สามารถสร้างตัวประกอบหลัก 6 มิติ โดยมีค่าไอเกน และสัดส่วนค่าแปรปรวนของตัวประกอบหลักดังแสดงในตารางที่ 2.1

ตารางที่ 2.1 แสดงตัวอย่างค่าไอเกน, สัดส่วนค่าความแปรปรวน และสัดส่วนค่าแปรปรวนสะสมของตัวประกอบหลัก

ตัวประกอบหลัก	ค่าไอเกน	สัดส่วนค่าความแปรปรวน	สัดส่วนค่าแปรปรวนสะสม
PC_1	3.323	0.579	0.579
PC_2	1.374	0.239	**0.818**
PC_3	0.476	0.083	0.901
PC_4	0.325	0.057	0.957
PC_5	0.157	0.027	0.985
PC_6	0.088	0.015	1.000
รวม	5.743	1.000	

จากตารางที่ 2.1 จะพิจารณาได้ดังนี้

วิธีที่ 1 การพิจารณาจากสัดส่วน การพิจารณาจากสัดส่วนค่าแปรปรวนสะสม ในที่นี้ได้สัดส่วนค่าแปรปรวนสะสมของ PC_1 และ PC_2 รวม 0.818 หรือร้อยละ 81.8 จึงควรมีตัวประกอบหลัก 2 ตัว คือ PC_1 และ PC_2



รูปที่ 2.1 แสดงตัวอย่างScree Plot

วิธีที่ 2 การใช้กราฟ Scree ซึ่งพล็อตกราฟได้ดังนี้ จากรูปที่ 2.1 จะพบว่าควรมีตัวประกอบหลักเพียง 2 ตัว

วิธีที่ 3 พิจารณาเฉพาะตัวประกอบหลักที่มีค่าไอเกนมากกว่าหนึ่ง

2.1.3 การวิเคราะห์ความแปรปรวน (Analysis of variance, ANOVA)

การวิเคราะห์ความแปรปรวนเป็นวิธีการที่ใช้เปรียบเทียบค่าเฉลี่ยมากกว่า 2 กลุ่มขึ้นไป โดยอาศัยหลักการเปรียบเทียบค่าอัตราส่วนระหว่างความแปรปรวนระหว่างกลุ่มกับความแปรปรวนภายในกลุ่ม (ชลิฎา, 2551)

เงื่อนไขของการวิเคราะห์ความแปรปรวน

1. ตัวอย่างแต่ละชุดจะต้องเป็นอิสระกัน
2. สุ่มตัวอย่างจากประชากรที่มีการแจกแจงแบบปกติ
3. ค่าความแปรปรวนของแต่ละประชากรต้องเท่ากัน

โดยมีสมมติฐานคือ การทดสอบประชากร k กลุ่ม มีค่าเฉลี่ยเท่ากันหรือไม่ นั่นคือ

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

H_1: อย่างน้อย 2 กลุ่ม มีค่าเฉลี่ยไม่เท่ากัน

โดยทำการทดลองเพื่อทดสอบกรรมวิธี k กรรมวิธี ค่าวัดของผลการทดลองที่ได้ เรียกว่า ค่าสังเกต แทนด้วย X_{ij}

X_{ij} คือค่าสังเกตของผลการทดลองซ้ำที่ i ของกรรมวิธีที่ j

$\overline{X}_{..}$ คือ ค่าเฉลี่ยทั้งหมด

$\overline{X}_{.j}$ คือค่าเฉลี่ยของกรรมวิธีที่ j

$i = 1, 2, \ldots, n_j$ $\quad j = 1, 2, \ldots, k$

$\sum_{j=1}^{k} n_j = n$ โดยที่ n เป็นจำนวนค่าสังเกตทั้งหมด และ k เป็นจำนวนกลุ่มทั้งหมด

ในการวิเคราะห์เพื่อเปรียบเทียบอิทธิพลของกรรมวิธีต่าง ๆ ที่มีผลต่อการทดลองนั้นจะพิจารณาจากความแปรผันของข้อมูลโดยส่วนรวม ซึ่งเรียกว่า ผลบวกกำลังสองของการทดลองหรือของทั้งหมด (Sum square total, *SST*) นิยามดังนี้

$$SST = \sum_{i=1}^{n_j}\sum_{j=1}^{k}(X_{ij} - \overline{X}..)^2$$

ซึ่งแยกออกเป็นสองส่วนได้ดังนี้

$$\sum_{i=1}^{n_j}\sum_{j=1}^{k}(X_{ij} - \overline{X}..)^2 = \sum_{j=1}^{k} n_{.j}(\overline{X}_{.j} - \overline{X}..)^2 + \sum_{i=1}^{n_j}\sum_{j=1}^{k}(X_{ij} - \overline{X}_{.j})^2$$

พิจารณาพจน์ต่าง ๆ ได้ดังนี้

$\sum_{j=1}^{k} n_{.j}(\overline{X}_{.j} - \overline{X}..)^2$: เป็นผลรวมของความแตกต่างกำลังสองระหว่างค่าเฉลี่ยของกรรมวิธีที่ j กับค่าเฉลี่ยทั้งหมด เป็นความแปรผันของอิทธิพลอันเนื่องมาจากกรรมวิธี ต่อไปนี้จะเรียกพจน์ดังกล่าวว่า ผลบวกกำลังสองของกรรมวิธี (Sum square treatment, *SSTR*)

$\sum_{i=1}^{n_j}\sum_{j=1}^{k}(X_{ij} - \overline{X}_{.j})^2$: เป็นผลรวมของความแตกต่างกำลังสองระหว่างค่าสังเกตภายในกรรมวิธีกับค่าเฉลี่ยของกรรมวิธีนั้น ๆ เป็นค่าที่ใช้วัดความแปรผันของความคลาดเคลื่อน ของการทดลอง ต่อไปนี้จะเรียกพจน์ดังกล่าวว่า ผลบวกกำลังสองของความคลาดเคลื่อน (Sum square of error, *SSE*)

ดังนั้น $SST = SSTR + SSE$

ระดับความเป็นอิสระของความแปรผันของการทดลอง $df = n-1$

ระดับความเป็นอิสระของความแปรผันของกรรมวิธี $df = k-1$

ระดับความเป็นอิสระของความแปรผันของความคลาดเคลื่อน $df = n-k$

เมื่อหารผลบวกกำลังสองของความแปรผันดังกล่าวด้วย ระดับความเป็นอิสระของความแปรผันนั้น ๆ จะได้ความแปรผันเฉลี่ย (Mean square, MS) ซึ่งก็คือความแปรปรวน

ความแปรผันเฉลี่ยของการทดลอง (Mean square total , *MST*) $MST = \frac{SST}{n-1}$

ความแปรผันเฉลี่ยของกรรมวิธี (Mean square treatment , *MSTR*) $MSTR = \frac{SSTR}{k-1}$

ความแปรผันเฉลี่ยของความคลาดเคลื่อน (Mean square error , *MSE*) $MSE = \frac{SSE}{n-k}$

สถิติทดสอบ $F = \frac{MSTR}{MSE}$

การตัดสิน จะปฏิเสธสมมติฐาน H_0 ถ้า $F > F_{1-\alpha,k-1,n-k}$ ที่ระดับนัยสำคัญ α

2.1.4 การวิเคราะห์ความแปรปรวนแบบทางเดียวของค่าอันดับของครัสคาล-วอลิส (The Kruskal-Wallis One-Way Analysis of Variance by Rank)

การวิเคราะห์ความแปรปรวนแบบทางเดียวของค่าอันดับของครัสคาล-วอลิสเป็นวิธีการทดสอบค่าเฉลี่ยของอันดับของประชากรตั้งแต่ 3 กลุ่มขึ้นไป โดยข้อมูลจะประกอบด้วยข้อมูลที่เลือกโดยวิธีสุ่ม k ชุด ซึ่งแต่ละชุดอาจจะมีขนาดไม่เท่ากันก็ได้ (ชลิฎา, 2551) โดยจะแทนค่าของตัวอย่างชุด i ซึ่งมีขนาด n_i ซึ่งเขียนเป็น $X_{i1}, X_{i2}, \ldots, X_{kni}$

โดยจะรวมตัวอย่างทั้งหมดเข้าด้วยกันแล้วเรียงอันดับค่าของตัวอย่างเหล่านั้น โดยให้อันดับ 1 แก่ค่าน้อยที่สุด, อันดับ 2 ให้ค่าน้อยถัดมา...และอันดับที่ n ให้กับค่ามากที่สุด

ให้ $R(X_{ij})$ แทนค่าอันดับที่ให้ค่าแก่ X_{ij}

และ $R_i = \sum_{j=1}^{n_i} R(X_{ij})$, i = 1, 2, … , k (ถ้าบางตัวมีค่าเท่ากัน ก็ให้หาค่าเฉลี่ยของค่าอันดับ)

ข้อตกลงเบื้องต้น

1. ตัวอย่างทุกชุดเป็นตัวอย่างสุ่ม ซึ่งถูกเลือกมาจากประชากรแต่ละชุด
2. มีความเป็นอิสระภายในแต่ละชุดของตัวอย่าง และระหว่างตัวอย่างแต่ละชุดด้วย
3. ตัวแปรสุ่ม X_{ij} เป็นตัวแปรสุ่มแบบต่อเนื่อง
4. มาตรวัดของข้อมูลตัวอย่างที่ใช้อย่างน้อยต้องเป็นมาตรเรียงลำดับ

โดยมีสมมติฐานคือ การทดสอบประชากร k กลุ่ม มีค่าเฉลี่ยของอันดับเท่ากันหรือไม่ นั่นคือ

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

H_1: อย่างน้อย 2 กลุ่ม มีค่าเฉลี่ยของอันดับไม่เท่ากัน

ค่าสถิติที่ใช้ คือ $$T = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

การตัดสิน เราจะปฏิเสธ H_0 ถ้าค่า T ที่คำนวณได้มากกว่าค่า $\chi^2_{1-\alpha,(\upsilon=k-1)}$ ที่ระดับนัยสำคัญ α

2.1.5 ขั้นตอนการปรับแนวแบบพีซไวส์ (Piecewise Alignment Algorithm)

ขั้นตอนการปรับแนวแบบพีซไวส์เป็นวิธีการหนึ่งที่ใช้เพื่อปรับแนวของข้อมูลในแต่ละหน่วยข้อมูลที่ถูกวัดค่าหลาย ๆ ครั้ง หรือถูกวัดค่า ณ เวลาต่าง ๆ ตามที่กำหนดไว้ ซึ่งค่าที่วัดได้ของแต่ละหน่วยข้อมูล อาจมีตำแหน่งที่วัดค่าไม่ตรงกันได้ การปรับแนวของข้อมูลให้มีตำแหน่งจุดเริ่มต้นของข้อมูลให้อยู่ในแนวเดียวกัน จึงมีความจำเป็นเพราะจะช่วยลดความคลาดเคลื่อนที่อาจเกิดขึ้นจากกระบวนการวัดค่าข้อมูลลงได้ โดยขั้นตอนจะเริ่มจากการเลือกช่วงเวลาหรือจุดเริ่มต้นของข้อมูลที่เป็น

เป้าหมายหรือค่าหลัก และทำการแบ่งช่วงของข้อมูลที่เลือกแล้วออกเป็นช่องเรียกว่า วินโดว์ (Window) ซึ่งแทนความกว้างของช่วงเวลาที่ทำการเลือก ซึ่งแทนด้วยสัญลักษณ์ W และนำตัวอย่าง ณ จุดเริ่มต้นและขนาดวินโดว์เดียวกัน ทำการเลื่อนวินโดว์ของตัวอย่างแบบจุดต่อจุดเทียบกับเป้าหมาย ทั้งเดินหน้าและถอยหลัง โดยในการเลื่อนวินโดว์ของตัวอย่างแต่ละครั้งจะมีการคำนวณค่า Pearson correlation coefficient ระหว่างตัวอย่างและเป้าหมาย เพื่อดูว่าตัวอย่างที่เลือกมานั้น เมื่อเปรียบเทียบกับเป้าหมายแล้วช่วงใดมีค่าสัมประสิทธิ์สหสัมพันธ์ที่สูงสุด ช่วงที่ได้นั้นจะถูกนำมาใช้เป็นวินโดว์ที่เหมาะสม (Pierce KM, 2005)

โดยมีขั้นตอนดังนี้ คือ

1. สุ่มเลือกตัวอย่าง 1 ตัวอย่าง เพื่อใช้เป็นค่าหลักในการเลื่อนของข้อมูล และกำหนดสุ่มจุดเริ่มต้นของข้อมูลและกำหนดขนาดความกว้างของวินโดว์
2. ตัดช่วงข้อมูลของค่าหลักจากจุดเริ่มต้นที่กำหนด และตามขนาดวินโดว์
3. เลื่อนข้อมูลของตัวอย่างทีละ 1 จุด ทำการเลื่อนข้อมูลนั้น ทั้งเดินหน้า และถอยหลัง ซึ่งในแต่ละจุดที่เลื่อนมีการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เปรียบเทียบกับค่าหลักหากช่วงที่เลื่อนช่วงใดมีค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson's Correlation Coefficient) แทนด้วยสัญลักษณ์ r มากที่สุดจะเลือกช่วงนั้นเป็นค่าที่เหมาะสมของข้อมูลตัวอย่างนั้น โดยค่า สัมประสิทธิ์สัมพันธ์แบบเพียร์สัน สามารถคำนวณได้ดังนี้ (สุรินทร์, 2545)

$$r = \frac{\sum_{i=1}^{w} x_i y_i - w\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{w} x_i^2 - w\bar{x}^2)(\sum_{i=1}^{w} y_i^2 - w\bar{y}^2)}}$$

โดยที่ x_i แทน ค่าของข้อมูลที่เป็นเป้าหมายหรือค่าหลัก ที่เวลา i เมื่อ $i = 1,2, \ldots, w$ และ

y_i แทน ค่าของข้อมูลตัวอย่าง ที่เวลา i

สัมประสิทธิ์สหสัมพันธ์ที่ได้เป็นสถิติไม่มีหน่วย จึงสามารถเปรียบเทียบความสัมพันธ์ระหว่างตัวแปรได้โดยตรง ซึ่งมีค่าอยู่ระหว่าง -1 และ 1 ถ้า r = -1 หรือ 1 แสดงว่าตัวแปรคู่นั้นมีความสัมพันธ์กันโดยสมบูรณ์ เป็นเชิงลบ และบวก ตามลำดับ ถ้า r = 0 แสดงว่าตัวแปรคู่นั้นไม่มีความสัมพันธ์กัน

4. ทำซ้ำกับตัวอย่างอื่น ๆ
5. ทำซ้ำกับจุดเริ่มต้นอื่น ๆ และขนาดความกว้างอื่น ๆ
6. เลือกจุดเริ่มต้นและขนาดความกว้างของวินโดว์ที่ให้ค่าเปอร์เซ็นต์การจำแนกสูงสุด

2.1.6 การตรวจสอบความสมเหตุสมผลแบบไขว้ (Cross Validation)

การตรวจสอบความสมเหตุสมผลแบบไขว้เป็นวิธีการในการประเมินค่าความถูกต้องของรูปแบบหรือวิธีการที่เราสนใจ โดยเริ่มจากการแบ่งชุดข้อมูลออกเป็นส่วน ๆ นำข้อมูลส่วนหนึ่งมาวิเคราะห์หรือสร้างรูปแบบ และนำข้อมูลอีกส่วนหนึ่งมาใช้ในการตรวจสอบ (Tutorial, 2006) โดยมีวิธีการในการแบ่งชุดข้อมูลเพื่อตรวจสอบความสมเหตุสมผลแบบไขว้เป็นดังนี้ คือ

วิธี Hold-out Method วิธีการนี้จะแบ่งข้อมูลออกเป็นสองส่วน ส่วนแรกเรียกว่ากลุ่มที่ใช้สร้างรูปแบบ (Training set) จะมีจำนวนประมาณสองในสามของข้อมูลทั้งหมด และกลุ่มทดสอบ (Test set) มีจำนวนประมาณหนึ่งในสามของข้อมูลทั้งหมด โดยกลุ่มที่ใช้สร้างรูปแบบจะถูกใช้เพื่อสร้างรูปแบบของข้อมูล จากนั้นจะใช้กลุ่มทดสอบประเมินค่าความถูกต้อง วิธีนี้จะเหมาะสมกับกรณีที่ข้อมูลมีจำนวนมาก อย่างไรก็ตามวิธีนี้มีข้อจำกัด ที่อาจจะได้ผลการทดลองที่คลาดเคลื่อนได้ถ้าการเลือกกลุ่มการสร้างรูปแบบของวิธีการ เลือกข้อมูลที่ไม่สามารถครอบคลุมทุกกรณีที่เกิดขึ้นในข้อมูลทั้งหมด จะส่งผลให้ค่าความถูกต้องได้ผลลัพธ์ไม่ดี ซึ่งอาจเลือกปรับมาใช้วิธีการสุ่มทำใหม่หลาย ๆ ครั้ง เพื่อจะหาค่าเฉลี่ยความถูกต้องของวิธีการ โดยใช้หลักการของวิธี k-fold cross-validation

วิธี k-fold cross-validation เป็นวิธีการที่พัฒนาจาก Hold-out Method โดยที่วิธีการนี้จะแบ่งข้อมูลออกเป็น k ส่วน และจะประเมินค่าความถูกต้องของรูปแบบ k รอบ (k เป็นเลขจำนวนเต็ม) โดยในแต่ละรอบหนึ่ง ๆ นั้น ใน 1 ส่วนจะถูกใช้เป็นกลุ่มทดสอบ และอีก k-1 ส่วนจะถูกใช้เป็น กลุ่มที่ใช้สร้างรูปแบบโดยทำการทดสอบ k รอบ วิธีนี้จะเหมาะกับกรณีข้อมูลมีจำนวนน้อย ข้อดีของวิธีการนี้คือ ข้อมูลในแต่ละชุดที่ทำการแบ่งจะถูกทดสอบอย่างน้อย 1 ครั้ง และถูกสร้างรูปแบบทั้งหมด k -1 ครั้ง โดยในแต่ละขั้นตอนเราสามารถกำหนดได้ว่าต้องการขนาดข้อมูลขนาดใด และต้องการทำการคำนวณเป็นจำนวนรอบเท่าใด แต่อย่างไรก็ตามข้อเสียของวิธีการนี้คือ ใช้เวลาในการคำนวณเป็น k เท่า

วิธี Leave-one-out cross-validation คือการทำ k - fold cross-validation เมื่อกำหนดให้ k มีค่าเท่ากับจำนวนข้อมูลทั้งหมด โดยที่กลุ่มที่ใช้สร้างรูปแบบถูกแบ่งเป็น N กลุ่มย่อย (N เป็นจำนวนตัวอย่าง) จาก N กลุ่มย่อย กลุ่มย่อยหนึ่งกลุ่มนำไปใช้เป็น กลุ่มทดสอบ เพื่อทดสอบรูปแบบ และเก็บ N-1 กลุ่มย่อย ใช้เป็นกลุ่มสร้างรูปแบบ จากนั้นจะใช้ข้อมูล กลุ่มทดสอบ ประเมินค่าความถูกต้องของรูปแบบวิธีการนี้เป็นการทำซ้ำ โดยที่จะเปลี่ยนกลุ่มตรวจสอบความถูกต้องไปเรื่อยๆ จนครบทั้งหมด N ครั้ง

2.1.7 ความรู้เบื้องต้นเกี่ยวกับน้ำมันเบนซิน

น้ำมันเบนซินเป็นเชื้อเพลิงที่ระเหยได้ง่าย ได้มาจากการกลั่นน้ำมันดิบในโรงกลั่น โดยกลั่นแล้วเอามาผสมกันและปรุงแต่งด้วยสารเพิ่มคุณภาพต่างๆ ไม่ว่าจะเป็น แนฟธา (Naphtha), Isomerate, Reformate และสารเติมแต่ง (Additives) เช่น MTBE (Methyl Tertiary Butyl Ether), เอ-ทานอล เป็นต้น เพื่อให้เหมาะสมแก่การใช้เป็นเชื้อเพลิงของเครื่องยนต์เบนซินชนิดสันดาปภายในโดยมีหัวเทียนเป็นเครื่องจุดระเบิด (Spark Ignition Internal Combustion Engine) ความสามารถในการระเหยน้ำมันต้องพอเหมาะกับการเผาไหม้ในกระบอกสูบและต้องเป็นไปอย่างสม่ำเสมอต่อเนื่อง (วาสนา, 2549)

2.1.7.1 ***คุณสมบัติของน้ำมันเบนซิน***

คุณสมบัติที่สำคัญของน้ำมันเบนซินที่ใช้กันอยู่ในปัจจุบันประกอบด้วย

1. การเกิดการเผาไหม้อย่างสม่ำเสมอภายในห้องเผาไหม้ของเครื่องยนต์
2. อัตราการละเหยต้องเหมาะสม เพราะจะมีผลต่อการติดเครื่องง่ายหรือยากของเครื่องยนต์ เวลาที่ใช้ในการอุ่นเครื่อง การกระตุกหรือสำลักในขณะรถวิ่ง อัตราการสิ้นเปลืองน้ำมัน และระยะเวลาในการเปลี่ยนถ่ายน้ำมันเครื่อง (นิติชัย, 2537)
3. ไม่ก่อให้เกิดสิ่งสกปรกสะสมในห้องเผาไหม้ และไม่ก่อให้เกิดยางเหนียวเกาะตามวาล์วไอดี

2.1.7.2 *กระบวนการกลั่นเพื่อให้ได้น้ำมันเบนซินที่มีคุณสมบัติตามต้องการ*

โดยปกติน้ำมันเบนซินที่กลั่นออกมาจากหอกลั่นบรรยากาศของโรงงานมีคุณภาพในการป้องกันการน็อกต่ำ มีจุดติดไฟเองต่ำ ยังไม่เหมาะสำหรับนำมาเป็นน้ำมันเชื้อเพลิง วิธีการในการเพิ่มคุณภาพในการป้องกันการน็อกทำได้ 2 วิธี คือ ผ่านขบวนการกลั่นในโรงกลั่นเพื่อเปลี่ยนแปลงโครงสร้างของโมเลกุลน้ำมันให้มาอยู่ในรูปที่จุดติดไฟเองสูงขึ้น และอีกวิธีหนึ่งคือเติมสารบางอย่างลงไปเพื่อเพิ่มค่า Octane เช่น Tetramethyl Lead (TML) และ Tetraethyl Lead (TEL) เพราะใช้ในปริมาณที่น้อยแต่ให้ผลต่อการเพิ่มค่า Octane ที่สูงขึ้น

2.1.7.3 *ความแตกต่างในน้ำมันแต่ละเครื่องหมายการค้า*

ปัจจุบันการจำหน่ายน้ำมันเบนซินมีอยู่ด้วยกัน 2 ชนิด คือ น้ำมันเบนซินชนิดธรรมดา ที่มีค่า Octane 91 และน้ำมันเบนซินชนิดพิเศษที่มีค่า Octane 95 ซึ่งมีบริษัทที่จัดจำหน่าย เช่น บริษัท ปตท. จำกัด (มหาชน) บริษัทบางจากปิโตเลี่ยม จำกัด (มหาชน) และ บริษัทเอสโซ่ (ประเทศไทย) จำกัด เป็นต้น

ความแตกต่างของน้ำมันในแต่ละเครื่องหมายการค้านั้น มาจากส่วนผสมที่เพิ่มเข้าไปในกระบวนการผลิต เช่นน้ำมันเบนซินของบริษัท ปตท. จำกัด (มหาชน) เป็นน้ำมันเบนซินที่มีสารเพิ่มคุณภาพ ฟริคชั่น โมดิฟายเออร์ สำหรับแก้ปัญหาความฝืดของอุปกรณ์ต่าง ๆ ที่เกิดขึ้น (ปตท. จำกัด (มหาชน), 2549) น้ำมันเบนซินของ บริษัทบางจากปิโตเลี่ยม จำกัด (มหาชน) มีสารเพิ่มคุณภาพ SCSS (Super Clean for Super Save) ที่ช่วยเพิ่มประสิทธิภาพการชะล้างทำความสะอาดส่วนประกอบสำคัญต่าง ๆ ในเครื่องยนต์และช่วยคงความสะอาดไว้ตลอดเวลา (บริษัทบางจากปิโตเลี่ยม จำกัด (มหาชน), 2549) และน้ำมันของบริษัทเอสโซ่ (ประเทศไทย) จำกัด มีสารเพิ่มคุณภาพพิเศษ สำหรับป้องกันสนิมกับการกัดกร่อนในถังน้ำมัน ท่อน้ำมัน ช่วยทำความสะอาดคาร์บิวเรเตอร์ และต้านทานการรวมตัวกับออกซิเจน (บริษัทเอสโซ่แสตนดาร์ด (ประเทศไทย) จำกัด, 2524)

2.2 งานวิจัยที่เกี่ยวข้อง

Kevin J. Johnson และ Robert E. Synovec (2002) ได้ศึกษาคุณลักษณะสารประกอบของน้ำมันเครื่องบินเจ็ททั้งหมด 5 ชนิด ชนิดละ 9 ตัวอย่างด้วยโดยใช้ผลจากการทำ Gas Chromatography (GC) เพื่อจำแนกชนิดของน้ำมัน โดยใช้การวิเคราะห์ความแปรปรวนการวิเคราะห์ส่วนประกอบแกนมุขสำคัญ พบว่าการใช้การวิเคราะห์ความแปรปรวน เป็นเครื่องมือที่มีประสิทธิภาพที่สามารถใช้ในการจำแนกชนิดของน้ำมันได้ โดยจะมีประสิทธิภาพมากขึ้นเมื่อมีการใช้ร่วมกับ การวิเคราะห์ส่วนประกอบแกนมุขสำคัญ

Karisa M. Pierce และคณะ (2005) ได้ทำการจำแนกชนิดของตัวอย่างน้ำมันทั้งหมด 5 เครื่องหมายการค้า โดยใช้ผลจากการทำ Gas Chromatography (GC) การวิเคราะห์มีพื้นฐานจากการจัดตำแหน่งของระยะเวลาโดยใช้วิธีขั้นตอนการปรับแนวแบบพีซไวส์ และใช้การวิเคราะห์ความแปรปรวนเพื่อเลือกคุณลักษณะของน้ำมันเบนซินก่อนทำการจำแนกร่วมกับการใช้การวิเคราะห์ส่วนประกอบแกนมุขสำคัญ โดยทำการวัดค่า degree of class separation ซึ่งใช้เป็นตัววัดว่าสามารถแบ่งกลุ่มได้ชัดเจนมากน้อยเพียงใด หลังจากนั้นจึงทำการเปรียบเทียบผลลัพธ์ที่ได้จากการเลือกใช้วิธีการทั้งสามอย่างร่วมกัน พบว่าผลลัพธ์ที่ได้จากวิธีการปรับแนวแบบพีซไวส์ร่วมกับการเลือกตัวแปรมีการจำแนกได้มีประสิทธิภาพถูกต้องมากที่สุด

Narong Lenghor และคณะ (2005) ได้พัฒนา Dynamic Interfacial Pressure Detector (DIPD) เพื่อใช้วัดความดันของพื้นผิวของหยดของเหลวที่ถูกหยดซ้ำต่อเนื่องกัน ซึ่งสัญญาณที่วัดได้นั้น เป็นแรงดันตึงผิว ที่ได้จากเทคนิค DIPD ซึ่งมีพื้นฐานมาจากเทคนิควิเคราะห์พื้นฐานด้านการไหล เช่น Flow Injection Analysis (FIA), Sequential Injection Analysis (SIA) และ High

Performance Liquid Chromatography (HPLC) โดย DIPD จะถูกใช้ร่วมกับ FIA เพื่อวิเคราะห์ลักษณะของพื้นผิวความดันตึงผิว และแสดงผลลัพธ์เป็นสัญญาณความดันที่วัดได้ และนำเทคนิค DIPD นี้ไปใช้ประโยชน์ เช่น การเลือกเครื่องมือจับสัญญาณสำหรับ Chromatographic separation เป็นต้น