ควอแดรนท์ที่ 3 : เป็นบริเวณที่ค่าของตัวแปร $x_1 < 0$ และ $x_2 < 0$

ควอแดรนท์ที่ 4 : เป็นบริเวณที่ค่าของตัวแปร $x_1 > 0$ และ $x_2 < 0$

จะเห็นได้ว่า มีเพียงควอแดรนท์ที่ I เท่านั้นที่สอดคล้องกับฟังก์ชันข้อจำกัด ซึ่งกำหนดให้ค่าของตัวแปรต้องมีค่า x มากกว่าหรือเท่ากับศูนย์ (การผลิตจะผลิตจำนวนติดลบไม่ได้) นั่นคือ

$x_1 \geq 0$ และ $x_2 \geq 0$

ดังนั้น การหาคำตอบของปัญหา หรือการหาค่า x_1 และ x_2 ซึ่งทำให้ฟังก์ชันวัตถุประสงค์มีค่าสูงสุดจะหาเฉพาะค่าของ x_1 และ x_2 ในควอแดรนท์ที่ I เท่านั้น เราเรียกบริเวณที่สามารถหาคำตอบของปัญหาว่า "บริเวณที่หาคำตอบได้" (Feasible Region) ส่วนบริเวณซึ่งไม่สามารถหาคำตอบได้ เรียกว่า "บริเวณที่หาคำตอบไม่ได้" (Infeasible Region)

เมื่อเราสามารถหาบริเวณที่หาคำตอบได้ของปัญหาโดยการเขียนกราฟของฟังก์ชันจำกัดต่าง ๆ แล้ว ขั้นตอนถัดไปคือการหาคำตอบจากปัญหาของกราฟ ซึ่งสามารถทำได้โดยการเขียนเส้นตรงของฟังก์ชัน วัตถุประสงค์ลงบนกราฟ แล้วหาจุดของ x_1 และ x_2 ภายใต้ข้อจำกัดที่ทำให้ฟังก์ชันวัตถุประสงค์มีค่าสูงสุด

2. วิธีซิมเพล็กซ์ (Simplex Method) ในการแก้ปัญหาการโปรแกรมเชิงเส้นด้วยวิธีซิมเพล็กซ์ ต้องทำการแปลงฟังก์ชันข้อจำกัดทั้งหมดให้อยู่ในรูปสมการ โดยตัวแบบมาตรฐานสำหรับวิธีซิมเพล็กซ์มีคุณสมบัติดังต่อไปนี้

2.1. สมการข้อจำกัดทุกสมการจะต้องมีค่าทางขวามือของสมการเป็นค่าบวกเสมอ จะมีค่าลบไม่ได้

2.2. ตัวแปรทุกตัวจะต้องมีค่ามากกว่าหรือเท่ากับศูนย์

2.3. ฟังก์ชันวัตถุประสงค์อาจเป็นการหาค่าสูงสุดหรือต่ำสุดก็ได้ การปรับตัวแบบการโปรแกรมเชิงเส้นให้อยู่ในรูปตัวแบบมาตรฐานสำหรับวิธีซิมเพล็กซ์ มีวิธีการดังนี้

ฟังก์ชันข้อจำกัด

1. สำหรับฟังก์ชันซึ่งอยู่ในรูปของอสมการ คือมีเครื่องหมาย $\geq$ หรือ $\leq$ สามารถทำให้อยู่ในรูปของสมการ (=) โดยการลบหรือบวกด้วยตัวแปรตัวหนึ่ง

ถ้าเป็น อสมการ $\leq$ ให้บวกด้วยตัวแปรทางซ้ายของฟังก์ชัน

ถ้าเป็น อสมการ $\geq$ ให้ลบด้วยตัวแปรทางซ้ายของฟังก์ชัน

โดยที่ตัวแปรที่นำมาบวกหรือลบจะต้องมีค่ามากกว่าหรือเท่ากับศูนย์เสมอ

2. สำหรับกรณีที่ค่าคงที่ทางข้างขวาของสมการข้อจำกัดมีค่าเป็นลบ สามารถทำให้เป็นบวกได้โดยการเอาค่า -1 คูณตลอด

3. การเปลี่ยนข้างของอสมการจาก $\geq$ ไปเป็น $\leq$ หรือจาก $\leq$ ไปเป็น $\geq$ สามารถทำได้โดยการเอา -1 คูณตลอดทั้งฟังก์ชัน

ตัวแปรในตัวแบบมาตรฐานของวิธีซิมเพล็กซ์นั้นตัวแปรทุกตัวจะต้องมีค่าเป็นบวก จะมีค่าเป็นลบไม่ได้ แต่ในความเป็นจริงลักษณะของปัญหาที่นำไปสร้างเป็นตัวแบบการโปรแกรมเชิงเส้นค่าของตัวแปรอาจมีค่าเป็นบวกหรือลบก็ได้ หรืออีกนัยหนึ่งคือ มีค่าใด ๆ ก็ได้ ถ้าตัวแปรใดมีลักษณะเช่นนี้ จะต้องทำให้ตัวแปรนั้นอยู่ในรูปของตัวแปรที่มีค่าเป็นบวกเท่านั้น โดยการเขียนตัวแปรนั้นแทนด้วยตัวแปรสองตัว ตัวอย่างเช่น สมมติตัวแปร y มีค่าใด ๆ ก็ได้ ซึ่งสามารถเขียนความสัมพันธ์เป็น $-\infty \leq y \leq \infty$ หรืออาจเขียนด้วยข้อความว่า y มีค่าใด ๆ ก็ได้

ในกรณีนี้สามารถเขียนตัวแปร y แทนด้วยตัวแปรสองตัวคือ x_1 และ x_2 คือ

$y = x_1 - x_2$

$x_1, x_2 \geq 0$

จะเห็นได้ว่าจากความสัมพันธ์ข้างต้น สามารถแปลงตัวแปร y ซึ่งมีค่าใด ๆ ก็ได้ไปเป็นตัวแปรสองตัว x_1 และ x_2 โดยที่ตัวแปรทั้งสองจะร้องมีค่าเป็นบวกเท่านั้น

จากความสัมพันธ์ $y = x_1 - x_2$ ถ้า x_1 มีค่ามากกว่า x_2 จะได้ว่า y มีค่าเป็นบวก หรือ $y > 0$ ถ้า x_1 มีค่าน้อยกว่า x_2 จะได้ว่า มีค่าเป็นลบ หรือ $y < 0$ แต่ถ้า x_1 เท่ากับ x_2 จะได้ว่า $y = 0$ ดังนั้น $x_1 - x_2$ สามารถแทนตัวแปร y ได้

สรุปได้ว่า ถ้าปัญหาการโปรแกรมเชิงเส้นมีตัวแปรที่มีค่าใด ๆ ก็ได้รวมอยู่ด้วยในตัวแบบจะต้องทำการเปลี่ยนตัวแปรดังกล่าวให้เป็นตัวแปรที่มีค่ามากกว่าหรือเท่ากับศูนย์เท่านั้น โดยสร้างตัวแปรขึ้นใหม่สองตัวแล้วนำมาลบกัน ตัวแปรทั้งสองต้องมีค่ามากกว่าหรือเท่ากับศูนย์ สังเกตว่าการแปลงตัวแปรในกรณีนี้ทำให้จำนวนของตัวแปรในตัวแบบเพิ่มขึ้นอีก 1 ตัว

ฟังก์ชันวัตถุประสงค์ ถึงแม้ว่า ตัวแบบมาตรฐานสำหรับวิธีซิมเพล็กซ์ ฟังก์ชันวัตถุประสงค์อาจเป็นได้ทั้งการหาค่าสูงสุดหรือค่าต่ำสุด แต่ในบางครั้งการแก้ปัญหาอาจทำได้สะดวกขึ้นถ้าแปลงรูปจากปัญหาลักษณะหนึ่งไปเป็นอีกลักษณะหนึ่ง เช่น การแปลงจากปัญหาการหาค่าสูงสุดไปเป็นการหาค่าต่ำสุด หรือการแปลงปัญหาการหาค่าต่ำสุดไปเป็นการหาค่าสูงสุด ซึ่งในเชิงคณิตศาสตร์นั้น การหาค่าสูงสุดของฟังก์ชันใด ๆ มีความหมายเหมือนกับการหาค่าต่ำสุดของค่าลบของฟังก์ชั่นนั้น ๆ

แนวคิดเกี่ยวกับเทคนิคการวิเคราะห์วิธีเชิงโอบล้อมข้อมูล

Data Envelopment Analysis หรือเรียกโดยย่อว่า DEA เป็นวิธีการวัดประสิทธิภาพทางเทคนิค (Technical Effciency) ของหน่วยงานหรือองค์การ (Organization) โดยใช้หลักการและทฤษฎีของการโปรแกรมเชิงเส้น (Linear Programming หรือ LP) เป็นพื้นฐานในการกำหนดค่าดัชนีประสิทธิภาพ (Efficiency Index)

DEA เป็นเครื่องมือที่ใช้ประเมินประสิทธิภาพของหน่วยงานจำนวนหนึ่ง และหาแนวทางในการปรับปรุงประสิทธิภาพของหน่วยงานนั้น ๆ เช่น สหกรณ์ ธนาคาร โรงแรม โรงเรียน มหาวิทยาลัย โรงพยาบาล เป็นต้น และทุกหน่วยที่จะนำมาวิเคราะห์ต้องมีลักษณะการดำเนินงานเหมือนกัน หรือดำเนินธุรกิจเดียวกัน

องค์การ หรือ หน่วยงานที่มีความเหมาะสมสำหรับการนำเอาวิธี DEA มาใช้ในการศึกษาประสิทธิภาพนั้น ส่วนใหญ่เป็นองค์การของรัฐ โดยมีลักษณะพิเศษคือ เป็นองค์การที่ไม่มุ่งแสวงหากำไรเป็นหลัก (Non - Profit Organization) ประกอบไปด้วยหน่วยผลิตซึ่งมีชื่อแยกเฉพาะสำหรับวิธีการนี้ว่าหน่วยตัดสินใจ (Dicision Making Unit และต่อไปจะเรียกโดยย่อว่า DMU) หลาย ๆ หน่วยที่ใช้ปัจจัยการผลิต (Input) หลายชนิดในการผลิตผลผลิต (Output) หลายอย่าง โดยที่ปัจจัยการผลิตและผลผลิตของ DMU แต่ละหน่วยมีลักษณะเหมือนกัน ดังนั้น การกำหนดและเลือกว่าอะไรควรจะเป็นปัจจัยการผลิตและผลผลิตที่เหมาะสมขององค์การเป็นเรื่องมีความสำคัญมากที่สุดเป็นอันดับแรกของการใช้เทคนิค DEA (ประสพชัย พสุนนท์, 2551 ก, หน้า 30 - 42)

ดัชนีประสิทธิภาพ (Efficiency Index) ของ DMU ใด ๆ ที่ได้จาก DEA คือ ค่าอัตราส่วนระหว่างผลผลิตรวมถ่วงน้ำหนัก (Weighted Outputs) กับปัจจัยการผลิตรวมถ่วงน้ำหนัก (Weighted Inputs) ของ DMU นั้น ๆ ตัวถ่วงน้ำหนักที่ใช้ในการรวมผลผลิตหรือปัจจัยการผลิตไม่ใช่ราคาตลาดของผลผลิตหรือปัจจัยการผลิต แต่เป็นค่าที่ถูกกำหนดโดยอัตโนมัติในกระบวนการแก้ปัญหาของ Linear Programming ที่ใช้ในการหาค่าประสิทธิภาพของแต่ละ DMU ลักษณะการสร้างดัชนีประสิทธิภาพเช่นนี้ ทำให้ DEA เป็นเทคนิคที่เหมาะสมเป็นพิเศษสำหรับหน่วยผลิต หรือ หน่วยตัดสินใจขององค์การที่มีหน้าที่รับผิดชอบการผลิตสินค้าและบริการที่มุ่งประโยชน์เพื่อสาธารณะหรือส่วนรวมเป็นหลัก ซึ่งสินค้าและบริการขององค์การเหล่านี้ไม่มีราคาตลาดกำหนดหรือไม่สามารถกำหนดราคาตลาดได้โดยง่าย (Serrano, Mar, & Chaparro, 2002)

หลักการทำงานของ DEA จะใช้ข้อมูลจาก DMU ทั้งหมดที่นำมาศึกษาสร้าง Production Frontier หรือเรียกอีกอย่างหนึ่งว่า เส้นประสิทธิภาพ (Efficiency Frontier) ขึ้นมา การเชื่อมต่อกันของ DMU ต่าง ๆ เพื่อประกอบเป็น Frontier มีลักษณะเป็นการเชื่อมต่อกันแบบเส้นตรง (Linear Combination) DMU ใดที่มีตำแหน่งตั้งอยู่บน Frontier ก็จะถูกประเมินโดย DEA ว่ามีประสิทธิภาพ

100% ในการใช้ปัจจัยการผลิตจำนวนที่มีอยู่เพื่อผลิตผลผลิตที่มีอยู่หรือกำลังผลิตอยู่ ในทางตรงกันข้าม DMU ใดไม่ตั้งอยู่บน Frontier ก็จะถูก DEA ประเมินว่ามีประสิทธิภาพต่ำกว่า 100% โดยมีแบบจำลองพื้นฐาน ดังนี้

$$Max h_k = \frac{\sum_{r=1}^{s} U_{rk} Y_{rk}}{\sum_{i=1}^{m} V_{ik} X_{ik}} \qquad \text{..... (1)}$$

$$\text{Subject to} \qquad \frac{\sum_{r=1}^{s} U_{rk} Y_{rj}}{\sum_{i=1}^{m} V_{ik} X_{ij}} \leq 1$$

เมื่อ h_k แทนค่าดัชนีประสิทธิภาพของ DMU ที่ K

Y_{rk} แทนค่าน้ำหนักถ่วงของปัจจัยผลผลิต r จาก DMU ที่ K

X_{ik} แทนค่าน้ำหนักถ่วงของปัจจัยการผลิต i จาก DMU ที่ K

U_{rk} แทนปริมาณปัจจัยผลผลิตที่ r จาก DMU ที่ K

V_{ik} แทนปริมาณปัจจัยการผลิตที่ i จาก DMU ที่ K

DMU_j แทนหน่วยการวิเคราะห์

j แทนจำนวน DMU; j = 1,2,.....,k

r แทนปัจจัยผลผลิต; r = 1,2,...,S

i แทนปัจจัยการผลิต ; i = 1,2,...,m

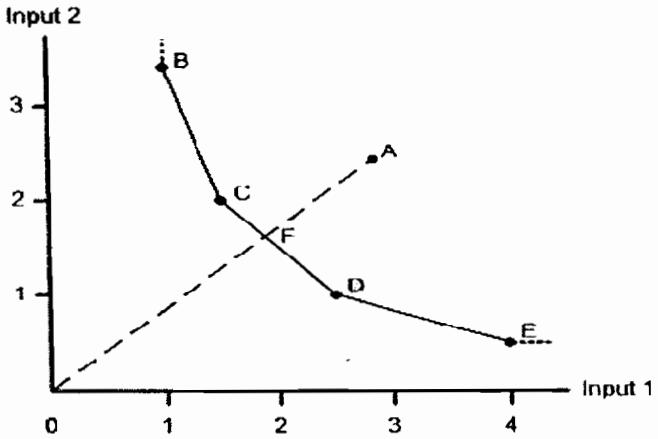
โดยที่ $Y_{rk}, X_{ik} U_{rk}$ และ $V_{ik} > 0$

ฟังก์ชั่นวัตถุประสงค์ของแบบจำลองนี้คือ การหาค่ามากที่สุดของอัตราส่วนระหว่างผลผลิตรวมถ่วงน้ำหนัก (Weighted Outputs) กับปัจจัยการผลิตรวมถ่วงน้ำหนัก (Weighted Inputs) ของ DMU k โดยมีข้อจำกัด 3 ประการใหญ่ ๆ คือ

1. ไม่มี DMU ใดมีค่าดัชนีประสิทธิภาพมากกว่า 1.00 ($h_k \leq 1$) เมื่อ DMU นั้น ๆ ใช้ค่าถ่วงน้ำหนัก ซึ่งได้ถูกกำหนดไว้สำหรับ DMU k หรือในอีกนัยหนึ่งข้อจำกัดนี้เป็นตัวบังคับให้ค่าดัชนีประสิทธิภาพของ DMU k มีค่าเป็นไปได้สูงสุดเท่ากับ 1.00 หรือ 100% เท่านั้นที่เป็นดังนี้เพราะแบบจำลองกำหนดให้ DMU k ซึ่งเป็น DMU ที่กำลังถูกประเมินประสิทธิภาพนั้นในขณะเดียวกันก็เป็นส่วนหนึ่งของสมการข้อจำกัดด้วย

2. ค่าน้ำหนักของผลผลิตทุกตัวของ DMU k มีค่ามากกว่าศูนย์

3. ค่าน้ำหนักของปัจจัยการผลิตทุกตัวของ DMU k มีค่ามากกว่าศูนย์ ค่าดัชนีประสิทธิภาพ h_k ซึ่งคำนวณได้จากแบบจำลอง สามารถแสดงให้เห็นได้โดยภาพที่ 5 ดังนี้



ภาพที่ 5 การประเมินประสิทธิภาพของหน่วยผลิต

จากภาพที่ 5 เราสมมุติให้มี DMU ที่ต้องการศึกษาอยู่ทั้งหมด 6 หน่วย คือ A, B, C, D และ E, F โดยที่แต่ละหน่วยใช้ปัจจัยการผลิต 2 ชนิด คือ Input 1 และ Input 2 เพื่อผลิตผล 1 ชนิด คือ Y จำนวน 1 หน่วย [นั่นคือ เส้นที่ลากเชื่อมต่อจุด A B C และ C หมายถึง เส้นผลผลิต 1 หน่วยเท่ากัน (A Unit Isoquant) จะเห็นว่า DMU A, B, C และ D ร่วมกันสร้าง Frontier จึงเป็นที่แน่นอนว่า DMU ทั้ง 4 หน่วยจะต้องอยู่บน frontier และเป็น DMU ที่ DEA จะให้ค่าดัชนีประสิทธิภาพเท่ากับ 1.00 และ DMU A เป็น DMU เดียวในรูปที่มีประสิทธิภาพไม่เต็ม 100%

DMU ที่ร่วมกันสร้างเส้น (Linear Segment) ของ Efficiency Frontier ซึ่งมีอิทธิพลโดยตรงมากที่สุดในการกำหนดค่าดัชนีประสิทธิภาพของ DMU k นั้น กลุ่มของ DMU เหล่านี้มีชื่อเรียกเฉพาะว่า Efficiency Reference Group แบบจำลองที่ใช้ในการคำนวณ การแก้ปัญหาเพื่อหาค่าดัชนีประสิทธิภาพ h_k ตามแบบจำลอง (1) ซึ่งมีรูปแบบเป็น Nonlinear สามารถทำให้ง่ายขึ้นได้โดย

การปรับเปลี่ยนแบบจำลอง (1) ให้อยู่ในรูปของ LP ธรรมดาซึ่งให้ค่าดัชนีประสิทธิภาพเท่ากับแบบจำลอง (1) รูปแบบของ LP ธรรมดาที่กล่าวถึงมีลักษณะดังนี้ คือ

$$Maxw_k = \sum_{r=1}^{s} u_{rk} Y_{rk}$$

Subject to $$\sum_{r=1}^{s} u_{rk} Y_{rk} - \sum_{i=1}^{m} v_{ik} X_{ij} \le 0$$

$$\sum_{i=1}^{m} v_{ik} X_{ik} = 1$$

$$y_{rk} x_{ik} \ge \epsilon > 0; \forall r,1$$

โดยที่ Y_{rk}, X_{ik}, U_{rk} และ $V_{ik} > 0$ และ $y_{rk.} = tY_{rk}$ และ $x_{ik} = tX_{ik}$

เมื่อ $t^{-1} = \sum {}_{i}V_{ik} X_{ik}$ และ $t > 0$ ส่วน ϵ คือค่าที่เล็กมาก ๆ แต่มากกว่าศูนย์ (เช่น 10^{-6})

DEA จะช่วยองค์การตัดสินใจเรื่องการวางนโยบายและระดับปฏิบัติการว่า องค์การควรจะจัดสรรทรัพยากรในกิจการอย่างไรให้เหมาะสม และผลที่ได้จากการวิเคราะห์ยังบอกด้วยว่าเมื่อเทียบเคียงกับกิจการที่มีประสิทธิภาพแล้ว กิจการที่ไม่มีประสิทธิภาพควรจะไปลดปัจจัยการผลิตตัวไหนเท่าไร หรือควรเพิ่มผลผลิตที่ตัวไหนเท่าไร งานศึกษาในต่างประเทศก็ใช้วิธี DEA นี้กันอย่างแพร่หลาย เพราะข้อดีของ DEA สามารถนำไปใช้ได้กับกิจการที่มี ผลผลิตและปัจจัยการผลิตที่หลากหลาย ถ้าเราใช้การหา อัตราส่วน เพียงอย่างเดียวก็จะมีข้อจำกัดคือ ประสิทธิภาพวัดได้ในรูปของตัวเงินเท่านั้น ถ้าหากกิจการขายสินค้าหรือบริการได้มากขึ้นจริง แต่เกิดของเสียจากการผลิตจำนวนมาก และพนักงานต้องทำงานล่วงเวลา ซึ่งสิ่งเหล่านี้เราไม่สามารถวัดออกมาได้ในรูปของตัวเงินได้ ก็จะต้องนำ DEA เข้ามาช่วย

ข้อดีอีกประการหนึ่งของ DEA คือ เป็นวิธีการที่เรียกว่า Nonparametric นั่นคือ ไม่ว่าข้อมูลจะอยู่ในรูปแบบใด ทั้งการกระจายแบบปกติหรือไม่ปกติ และไม่จำเป็นต้องรู้ว่าปัจจัยการผลิตมีความสัมพันธ์กับผลผลิตรูปแบบใด ก็สามารถวัดได้ทั้งสิ้น หน่วยงานในต่างประเทศหลายหน่วยงาน โดยเฉพาะอย่างยิ่งในภาคบริการ ไม่ว่าจะเป็นโรงพยาบาล ธนาคาร สถาบันการศึกษา ได้นำหลักการนี้ไปใช้ ตัวอย่างเช่น โรงพยาบาลก็ใช้ผลผลิตเป็นจำนวนเตียงคนไข้ ร่วมกับจำนวนผ่าตัดเล็ก ผ่าตัดใหญ่ มาเปรียบเทียบกับทรัพยากรที่โรงพยาบาลมีอยู่ไม่ว่าจะเป็นหมอ เภสัชกร พยาบาล ยา เครื่องมือเครื่องจักร ตึก พื้นที่ ในส่วนของธนาคารก็อาจใช้ผลผลิตเป็นจำนวนรายการเงินฝาก จำนวนรายการเงินกู้ และปัจจัยการผลิตเป็น พนักงาน ค่าใช้จ่ายที่เป็นดอกเบี้ย ค่าใช้จ่ายที่ไม่ใช่ดอกเบี้ย ค่าใช้จ่ายที่เป็นค่าเช่า ค่าสาธารณูปโภค ค่าโฆษณา ค่าซ่อมแซม ฯลฯ บางประเทศประยุกต์ใช้ DEA ในการศึกษาประสิทธิภาพของโรงพยาบาล โดยวัดเฉพาะด้านผลผลิตโดยใช้เพียง "อัตราส่วนทาง

การเงิน" ที่แสดงถึงผลการดำเนินงานด้านต่าง ๆ แล้วนำมาเปรียบเทียบกันระหว่างธนาคารก็ได้ (นิติพงษ์ ส่งศรีโรจน์ และ จารึก สิงหปรีชา, 2549)

นอกจากนี้ ยังมีงานศึกษางานที่นำเอา DEA ไปใช้ร่วมกับ Balanced Scorecard ก็สามาถทำได้ เนื่องจาก BSC เข้าไปเกี่ยวข้องกับตัวชี้วัดหลายด้าน ไม่ว่าจะเป็นด้านการเงิน ด้านลูกค้า ด้านกระบวนการภายใน และด้านการเรียนรู้และพัฒนา ซึ่งตัวชี้วัดแต่ละด้านค่อนข้างมีความหลากหลาย ซึ่ง DEA จะไปรวบรวมตัวชี้วัดที่หลากหลายเหล่านี้ออกมาเป็น Effficiency Score ให้ผู้บริหารพิจารณาได้โดยง่าย จะเห็นว่าสามารถนำวิธีการ DEA นี้ไปประยุกต์ใช้ได้อีกมาก อย่างไรก็ดี การนำ DEA ไปใช้ต้องคำนึงถึงข้อจำกัดของวิธีการนี้ด้วย จริงอยู่ที่ DEA สามารถวัดผลผลิตและปัจจัยการผลิตพร้อม ๆ กันได้ได้ทีละหลายตัว แต่ก็ต้องขึ้นอยู่กับการเลือกว่าจะใช้ผลผลิตและปัจจัยการผลิตตัวไหน ที่สามารถสะท้อนประสิทธิภาพการดำเนินงานของกิจการได้อย่างแท้จริง (พัชรศรี แดงทองดี, 2549, หน้า 90 - 96)

ตัวแบบในการประเมินประสิทธิภาพของ DEA

ตัวแบบแรก ๆ ของการประเมินประสิทธิภาพด้วยวิธีการ DEA ได้แก่ ตัวแบบ CCR (Charnes Cooper and Rhodes) และตัวแบบ BCC (Banker Charnes and Cooper) ก่อนที่จะมีการพัฒนาตัวแบบอื่น ๆ และยังได้นำวิธีการทางสถิติอ้างอิง เพื่อใช้ในการจัดกลุ่ม หรือจำแนกกลุ่มความมีประสิทธิภาพขององค์กร เช่น การวิเคราะห์ตัวประกอบ (Factor Analysis) การวิเคราะห์กลุ่ม (Cluster Analysis) การจำแนกกลุ่ม (Discriminant) เป็นต้น

Charnes, Cooper, and Rhodes (1978) ได้สร้าง ตัวแบบแรกของวิธีการ DEA ที่พิจารณามุมมองด้านผลผลิต (Output Oriented) คือ ตัวแบบ CCR ในรูป Linear Programming ดังนี้

ฟังชั่นวัตถุประสงค์ $\quad Max\tau = \sum_{r=1}^{s} v_r y_{rq}$

เงื่อนไขข้อจำกัด

$$\sum_{i=1}^{m} u_i x_{iq} = 1$$

$$\sum_{r=1}^{s} v_r y_{rj} - \sum_{i=1}^{m} u_i x_{ij} \le 0 \quad (j = 1,2,3,\ldots,n)$$

$$u_i =\ge \varepsilon \quad (i = 1,2,\ldots,m)$$

$$v_r =\ge \varepsilon \quad (i = 1,2,\ldots,s)$$

เมื่อ

$\tau =$ คะแนนประสิทธิภาพ

$x_{ij} =$ จำนวนปัจจัยนำเข้าที่ I จากองค์กรที่ j

y_{rj} = จำนวนผลผลิตที่ r จากองค์กรที่ j

u_i = ค่าถ่วงน้ำหนักของปัจจัยนำเข้าที่ i

v_r = ค่าถ่วงน้ำหนักของปัจจัยผลผลิตที่ r

s = จำนวนปัจจัยด้านผลผลิต

n = จำนวนขององค์กร

ε = ค่าบวกขนาดเล็ก

ส่วน Banker, Charnes, and Cooper (1984) ได้ปรับปรุงตัวแบบ CCR ไปเป็นตัวแบบ BCC ที่อยู่ในรูป Linear Programming ดังนี้

ฟังชั่นวัตถุประสงค์ $Max\tau = \sum_{r=1}^{s} v_r y_{rq} + w_q$

เงื่อนไขข้อจำกัด

$$\sum_{i=1}^{m} v_i x_{iq} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + w_j \leq 0 \qquad (j = 1,2,3, \ldots, n)$$

$$u_i \geq \varepsilon \qquad (i = 1,2,\ldots,m)$$

$$v_r \geq \varepsilon \qquad (i = 1,2,\ldots,s)$$

เมื่อ w_q แทนค่าการเปลี่ยนแปลงของปัจจัยนำเข้าหรือปัจจัยผลผลิต

ผลลัพธ์ที่ได้จากการคำนวณตัวแบบ CCR และตัวแบบ BCC สามารถแบ่งองค์กรที่นำมาประเมินประสิทธิภาพออกเป็น 2 ลักษณะ คือ 1) องค์กรที่มีประสิทธิภาพ และ 2) องค์การที่ไม่มีประสิทธิภาพ โดยที่องค์กรที่มีประสิทธิภาพจะมีค่า $\tau = 1$ ส่วนองค์กรที่ไม่มีประสิทธิภาพ จะมีค่า $\tau < 1$ ข้อจำกัด คือ ยังไม่สามารถเรียงลำดับความมีประสิทธิภาพได้ เนื่องจากมีค่า $\tau = 1$ เหมือนกันหมด และไม่สามารถระบุปัจจัยที่ส่งผลต่อคะแนนประสิทธิภาพของวิธีการ DEA ได้มีผู้พยายามพัฒนาตัวแบบของวิธีการ DEA ให้แก้ไขปัญหาดังกล่าว แต่ก็พบว่าเกิดปัญหาที่ทำให้ตัวแบบที่พัฒนาขึ้นไม่เป็นที่นิยม เพราะการคำนวณที่ยุ่งยากซับซ้อน ไม่มีโปรแกรมสำเร็จรูปที่ง่ายต่อการคำนวณตัวแบบ เพราะฉะนั้นจึงยังยึดกับตัวแบบ CCR และ BCC พร้อมกับหาแนวทางเพื่อแก้ปัญหาดังกล่าว สิ่งหนึ่งที่ผู้พัฒนาแนวทางแก้ปัญหาได้ตระหนัก คือ การคัดเลือกตัวแปรปัจจัยนำเข้าและปัจจัยผลผลิตเพราะจะลดปัญหาดังกล่าวได้ (ประสพชัย พสุนนท์, 2551 ข, หน้า 42)

วิธีการคัดเลือกตัวแปรสำหรับการประเมินประสิทธิภาพของ DEA

วิธีการ DEAจะสามารถประเมินประสิทธิภาพขององค์กรได้ตรงตามจุดประสงค์ ต้องสามารถแสดงถึงความสำคัญของปัจจัยนำเข้าและปัจจัยผลผลิต หากพบว่า ปัจจัยนำเข้าและปัจจัยผลผลิต

ไม่ส่งผลอย่างแท้จริงต่อความมีประสิทธิภาพหรือความไม่มีประสิทธิภาพ ก็ไม่มีความจำเป็นที่จะใช้ตัวแปรเหล่านั้นมาวิเคราะห์คะแนนประสิทธิภาพ ข้อดีประการหนึ่งของวิธีการ DEA คือ การไม่ต้องมีข้อตกลงทางสถิติ แต่ก็มีผู้พยายามใช้วิธีการทางสถิติมาใช้ร่วม โดยเฉพาะวิธีการคัดเลือกตัวแปร กล่าวได้ว่าการคัดเลือกตัวแปร แบ่งได้ 2 กลุ่ม กลุ่มแรกใช้วิธีการทางสถิติในการทดสอบความซ้ำซ้อนของตัวแปร กลุ่มที่สอง ไม่อิงวิธีการทดสอบทางสถิติ ได้มีผู้นำเสนอไว้หลายกรณี ในที่นี้จะกล่าวถึงแนวทางของ Wagner and Shimshak (2007) ซึ่งได้เสนอแนวทางการคัดเลือกตัวแปรโดยไม่อิงวิธีการทดสอบทางสถิติ มี 2 วิธี คือ 1) การคัดเลือกตัวแปรแบบ Backward และ 2) การคัดเลือกตัวแปรแบบ Forwrad ซึ่งแต่ละแนวทางมีขั้นตอน ดังต่อไปนี้

1. การคัดเลือกตัวแปรแบบ Backward มีขั้นตอนดังนี้

กำหนดให้ การประเมินประสิทธิภาพองค์กรมีปัจจัยนำเข้า J ปัจจัย เมื่อ j = 1, 2, 3, …. , J และมีปัจจัยผลผลิต K ปัจจัย เมื่อ k = 1, 2, 3, ... , K การคัดเลือกตัวแปร เริ่มต้นจาก คำนวณตัวแบบสมบูรณ์ (Full Model) หมายถึงตัวแบบที่ใช้ปัจจัยนำเข้าและปัจจัยผลผลิตทุกตัว (J, K) ในการคำนวณ แล้วบันทึกคะแนนประสิทธิภาพ มีค่าเท่ากับ $\hat{E}$

ขั้นที่ 1 เวียนตัดปัจจัยนำเข้าและปัจจัยผลผลิตออกที่ละปัจจัย โดยมีปัจจัยทั้งหมด J + K ปัจจัย และต้องตัดปัจจัยที่ i เมื่อ i = 1, 2, 3, …, J + K ออก จากนั้นนำ ปัจจัยนำเข้าและปัจจัยผลผลิตที่เหลือไปคำนวณคะแนนประสิทธิภาพ มีค่าเท่ากับ E_{ij} คำนวณผลต่างของคะแนนประสิทธิภาพแต่ละองค์กรจาก $\hat{E} - E_{ij}$ แล้วคำนวณค่าเฉลี่ยของผลต่างนั้น เลือกตัดปัจจัยที่ให้ค่าเฉลี่ยของผลต่างต่ำสุดออก มีค่าเท่ากับ $\hat{E}_1$ ซึ่งจะใช้คำนวณในขั้นต่อไป

ขั้นที่ n + 1 ทำซ้ำในการคำนวณวิธีการ DEA จาก i = 1, 2, …, J+K-n โดยจะเหลือปัจจัยนำเข้าและปัจจัยผลผลิตทั้งหมด J+K – n ปัจจัย จากนั้น คำนวณค่าเฉลี่ยของผลต่างระหว่าง $E_{n+1,i}$ และ $\hat{E}_n$ และตัดตัวแปรปัจจัยที่ให้ค่าเฉลี่ยของผลต่างต่ำที่สุด

ขั้นหยุด จะหยุดขั้นตอน เมื่อเหลือปัจจัยนำเข้าและปัจจัยผลผลิตอย่างละ 1 ปัจจัย ประสพชัย พสุนนท์ (2551 ข, หน้า 42) ได้ตั้งข้อสังเกตว่า การคัดเลือกตัวแปรแบบBackward หากกำหนดให้ตัดตัวแปรออกจนกระทั่งเหลือปัจจัยอย่างละ 1 ปัจจัย อาจทำให้สูญเสียสารเทศที่มีความสำคัญต่อการประเมินประสิทธิภาพ และเสนอว่าควรกำหนดขั้นหยุดด้วยการกำหนดค่าเฉลี่ยผลต่างด้วยค่าคงที่ อาจใช้เกณฑ์ในการทดสอบสมมติฐาน (เช่น .01, .02, .03, .., .05 เป็นต้น) ด้วยเหตุผลที่ว่าการเปลี่ยนแปลงของค่าเฉลี่ยผลต่างที่มีค่าน้อยกว่าเกณฑ์ไม่มีนัยสำคัญ หากมีค่าเฉลี่ยผลต่างมากกว่าเกณฑ์แสดงว่าปัจจัยที่เหลือมีนัยสำคัญส่งผลต่อการมีประสิทธิภาพขององค์กร ทั้งนี้ การกำหนดค่าหยุดขึ้นอยู่กับความเหมาะสมของแต่ละองค์กร สิ่งแวดล้อม วัฒนธรรม และความจำเป็น จำนวนของปัจจัย สอดคล้องกับหลักการ ทฤษฎี (Wagner & Shimshak, 2007)

2. การคัดเลือกตัวแปรแบบ Forward จะเริ่มต้นจากเลือกปัจจัยนำเข้าและปัจจัยผลผลิตอย่างละ 1 ปัจจัย ซึ่งต้องเป็นปัจจัยที่มีความสำคัญ ถือเป็นปัจจัยเริ่มต้นในการคำนวณคะแนนประสิทธิภาพ เรียกว่าคะแนนประสิทธิภาพเริ่มต้น จากนั้นเลือกปัจจัยนำเข้าหรือปัจจัยผลผลิตทีละปัจจัยเพื่อนำไปคำนวณคะแนนประสิทธิภาพ เรียกว่าคะแนนประสิทธิภาพที่เพิ่ม 1 ปัจจัย ต่อมาคำนวณค่าเฉลี่ยผลต่างระหว่างคะแนนประสิทธิภาพเริ่มต้นกับคะแนนประสิทธิภาพที่เพิ่ม หากค่าเฉลี่ยผลต่างนั้นมีความแตกต่างกันมากพอก็เก็บไว้ แล้วเลือกปัจจัยนำเข้าหรือปัจจัยผลผลิตอีก 1 ปัจจัย เพื่อคำนวณคะแนนประสิทธิภาพที่เพิ่ม 2 ปัจจัย จากนั้นคำนวณค่าเฉลี่ยผลต่างระหว่างคะแนนประสิทธิภาพที่เพิ่ม 1 ปัจจัยและคะแนนประสิทธิภาพที่เพิ่ม 2 ปัจจัย ทำเช่นนี้ต่อไปจนกระทั่งค่าเฉลี่ยผลต่างมีค่าไม่มากนัก วิธีนี้เป็นเรื่องยากที่จะระบุปัจจัยที่มี ความสำคัญคู่แรก หรือแม้แต่ขั้นถัดไปก็ยากที่จะระบุว่าตัวแปรใดเหมาะสมในขั้นนั้น โดยเฉพาะอย่างยิ่งกรณีที่ตัวแปรมีความสำคัญใกล้เคียงกัน

การใช้โปรแกรม DEAP 2.1 สำหรับการประเมินประสิทธิภาพของ DEA

โปรแกรม DEAP 2.1 เป็นโปรแกรมประเภทไม่มีค่าใช้จ่ายในการนำมาใช้ (Free Ware) ผู้ใช้สามารถ Download โปรแกมได้ที่ Webside http://www.uq.edu/economics/cepa/deap.htm ข้อมูลจำเพาะของโปรแกรม เป็นโปรแกรมที่เขียนด้วยภาษา Fortran (Lahhey F77LEM/32) โดย Professor Tim Coelli ซึ่งอยู่ที่ Centre for Efficiency and Productivity Analysis (CEPA) School of Economics University of Queenland Australia โปรแกรมจะทำงานบนระบบปฏิบัติการ DOS หรือในลักษณะ Command Line ของเครื่องคอมพิวเตอร์ ต่อมาได้พัฒนาให้ใช้บนระบบปฏิบัติการ Windows ได้ ความต้องการระบบ เครื่องคอมพิวเตอร์ต้องมีหน่วยประมวลผลระดับ 386 หรือสูงกว่า มีขนาดหน่วยความจำ 4MB หรือสูงกว่า และใช้กับระบบปฏิบัติการ DOS 5.0 และหรือ Windows 3.1 หรือสูงกว่า ส่วนประกอบของโปรแกรม ภายใน Folders จะประกอบด้วย Executable file, DEAP.exe Start up file, DEAP.000 Usermanual, DEA.pdf Data file Instruction file Output file (Coeli, 1996, pp. 49 - 50; อัครพงศ์ อั้นทอง, 2547, หน้า 11 - 16)

ความสามารถของโปรแกรม DEAP 2.1 สามารถคำนวณประสิทธิภาพทางเทคนิคด้วยวิธี DEA ทั้งที่เป็นข้อสมติ CRS และ VRS ทั้งในด้าน Output - Oriented และInput - Oriented และในกรณี Multiple Output และ Input ได้ สามารถคำนวณประสิทธิภาพต้นทุน (Cost Efficiency) และประสิทธิภาพโดยรวมได้ นอกจากนี้ยังคำนวณ Malmquist DEA ในกรณีข้อมูลเป็น Panel data ได้ สำหรับโปรแกรม DEAP 2.1 นั้นจะอ่านข้อมูลเป็น ASCII หรือ Text File ในการจัดการข้อมูลผู้ใช้สามารถจัดการข้อมูลในโปรแกรมใดก็ได้ แล้วบันทึก File ให้อยู่ในรูปแบบดังกล่าว การจัดการข้อมูลมีข้อจำกัดของการจัด Column ของข้อมูลดังนี้ สำหรับข้อมูล Cross - Section Data การจัด

จะต้องกระทำดังนี้ Column แรกของข้อมูลแสดงถึงชุดของตัวแปรผลผลิต (Output) Column ต่อมาของข้อมูลแสดงถึงชุดของตัวแปรนำเข้า (Intput) ในการวิเคราะห์จะต้องมีการจัดการ File อยู่ 3 File คือ 1) Data file จัดการให้อยู่ในรูปแบบของ Text file (ASCII) จะมีนามสกุล .prn และ File ข้อมูลจะต้องไม่มีข้อความอยู่ 2) Output file จะมีนามสกุล .out และ 3) Instruction file จะมีนามสกุล .ins โดยที่ File ทั้ง 3 จะต้องอยู่ใน Directory เดียวกับ DEAP ที่ตั้งไว้ ใน File deap.exe

ผลลัพธ์ที่ได้จากโปรแกรม DEAP 2.1 ในกรณีวิเคราะห์ DEA ปกติจะประกอบไปด้วยค่า Technical Efficiency ค่า Output Slacks และ Output Targets และค่า Input Slacks และ Intput Targets (Coeli, 1996, pp. 10 - 22)

งานวิจัยที่ใช้การประเมินประสิทธิภาพของ DEA

Cinca Callen and Molineroc (2005) ได้ศึกษา ประสิทธิภาพของบริษัทที่ให้บริการอินเทอเน็ตความเร็วสูงจำนวน 40 บริษัท โดยกระจายทุกกรณีของการจัดหมู่ (Combination) ระหว่างปัจจัยด้านผลผลิต (ประกอบด้วย จำนวนลูกค้า และรายได้) และปัจจัยนำเข้า (ประกอบด้วย จำนวนพนักงาน ค่าใช้จ่าย และเงินลงทุน) จากนั้นใช้การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) ในการสกัดองค์ประกอบที่มีความสำคัญในการอธิบายผลการวิจัย พบว่าสามารถแบ่งประสิทธิภาพของบริษัทอินเทอเน็ตตามปัจจัยด้านผลผลิตได้ 3 กลุ่ม คือ 1) กลุ่มที่มีประสิทธิภาพด้านการดูแลลูกค้า 2) กลุ่มที่มีประสิทธิภาพด้านรายได้ และ 3) กลุ่มที่มีประสิทธิภาพโดยรวม นอกจากนี้ อาฟีฟี ลาเต๊ะ และคณะ (2550) ได้จัดกลุ่มประสิทธิภาพการดำเนินงานของห้องสมุดสถาบันอุดมศึกษาในเขตภาคใต้ของประเทศไทยทั้งหมด 13 แห่ง โดยใช้เทคนิคการวิเคราะห์องค์ประกอบหลักและการวิเคราะห์กลุ่ม (Cluster Analysis: CA) เป็นเครื่องมือในการจัดกลุ่มจากคะแนนประสิทธิภาพในทุกการกระจายการจัดหมู่ของปัจจัยนำเข้าและปัจจัยด้านผลผลิตที่ได้จากวิธีการ DEA ผลที่ได้พบว่ามีความสอดคล้องกันระหว่างผลลัพธ์จากวิธี PCA และวิธี CA โดยสามารถจำแนกห้องสมุดออกเป็น 3 กลุ่ม คือ กลุ่มห้องสมุดที่มีประสิทธิภาพของการดำเนินงานสูง ปานกลาง และต่ำ Serrano and Mar (2001) ได้ใช้การวิเคราะห์องค์ประกอบหลักร่วมกับการวิเคราะห์กลุ่มจากคะแนนประสิทธิภาพของวิธีการ DEA เพื่อจัดกลุ่มเมืองในประเทศจีน โดยใช้ข้อมูลจาก Zhu (1998) และ Premachandra (2001) ซึ่งเป็นข้อมูลที่พิจารณาเมืองแต่ละเมืองว่ามีความเป็นเมืองหรือชุมชนแค่ไหน พิจารณาจากปัจจัยนำเข้า 2 ปัจจัย และปัจจัยผลผลิต 3 ปัจจัย ใช้ค่าสังเกตของเมืองทั้งหมด 18 เมือง ใช้การจัดหมู่ของปัจจัยนำเข้าและปัจจัยผลผลิตที่ได้จากการวิเคราะห์ด้วยวิธีการ DEA โดยถือว่าการจัดหมู่เหล่านั้นเป็นตัวแปรที่ใช้ในการพิจารณา เช่นเดียวกับ Serrano, Mar, and Chaparro (2002) ได้ใช้ PCA ร่วมกับ CA เพื่อจัดกลุ่มธนาคารในประเทศสเปน พิจารณาจากปัจจัยนำเข้า 3 ปัจจัยและปัจจัยผลผลิต 3 ปัจจัย

ใช้ค่าสังเกตของธนาคารทั้งหมด 47 แห่ง ผลที่ได้จากวิธีนี้ สามารถจัดกลุ่มเมืองหรือธนาคารที่มีลักษณะคล้ายคลึงกันมากที่สุดไว้ในกลุ่มเดียวกัน ดิเรก ปัทมสิริวัฒน์ (2552) ประเมินประสิทธิภาพโรงพยาบาล ชุมชนขนาดกลาง จำนวน 166 แห่ง ในสังกัดกระทรวงสาธารณสุข ด้วยวิธีการ DEA จากตัวแบบ CCR และตัวแบบ BCC พบว่า มีโรงพยาบาลที่มีประสิทธิภาพระดับแนวหน้า จำนวน 17 แห่ง ส่วน จิราภรณ์ แซ่ตั้ง และ ประสพชัย พสุนนท์ (2551) ประเมินประสิทธิภาพท่าอากาศยานไทยจำนวน 6 แห่ง ระหว่างปี พ.ศ. 2549 - 2550 ด้วยวิธีการ DEA จากตัวแบบ CCR และตัวแบบ BCC โดยมีกรณีรวมปัจจัย พบว่า ในปี พ.ศ. 2549 มีท่าอากาศยาน 3 แห่ง มีประสิทธิภาพ และในปี พ.ศ. 2550 มีท่าอากาศยาน 4 แห่งที่มีประสิทธิภาพ เช่นเดียวกับ สลิลทิพย์ เหล่าไพโรจน์ และ พัชราภรณ์ เนียมมณี (2551) วัดประสิทธิภาพสำนักงานสาขา การประปานครหลวง จำนวน 15 สาขา ด้วยวิธีการ DEA จากตัวแบบ CCR และตัวแบบ BCC การวิเคราะห์ข้อมูลใช้โปรแกรม DEAP 2.1 นอกจากนี้มีงานวิจัยที่ใช้ข้อสมมติจากวิธีการ DEA เปรียบเทียบผลลัพท์ เช่น ดิเรก ปัทมสิริวัฒน์ (2550) ประเมินประสิทธิภาพการบริหารต้นทุนของสถานพยาบาลในสังกัดสำนักงานปลัดกระทรวงสาธารณสุข พบว่า ข้อสมมติ VRS มีสถานพยาบาล 19 แห่งอยู่ระดับแนวหน้า โดยมีระดับประสิทธิภาพเฉลี่ยเท่ากับ ร้อยละ 78 แต่ข้อสมมติ CRS มีค่าเฉลี่ย ร้อยละ 71 และ ก่อโชค ภูนิคม และสิทธิชัย แซ่เหล่ม (2547) ได้ศึกษาระบบการวัดและปรับปรุงสมรรถนะการดำเนินงานของโรงพยาบาล ในเขตจังหวัดอุบลราชธานี จากการพัฒนาตัวชี้วัด และวัดสมรรถนะโดยวิธีการ DEA พบว่า แบ่งกลุ่มสมรรถนะได้ 3 กลุ่ม และกรมตรวจบัญชีสหกรณ์ (2549) ประเมินประสิทธิภาพการดำเนินงานของสหกรณ์การเกษตร ปี พ.ศ. 2548 จำนวน 3,483 แห่ง โดยวิธีการ DEA พบว่า สหกรณ์การเกษตรโดยรวมมีประสิทธิภาพระดับปานกลาง ภาวะผลตอบแทนต่อขนาดลดลงร้อยละ 72.42

แนวคิดเกี่ยวกับเทคนิคการวัดอนุกรมวิธาน (Taxometric Method)

วิธีการวัดอนุกรมวิธาน หรือ อนุกรมวิธานเชิงตัวเลข (Numerical Taxonomy) เป็นสาขาวิชาที่ใช้วิธีการเชิงจำนวน (Numerical method) ในการจัดเข้ากลุ่ม (Classifications) วัตถุการจัดจำแนก (Classify) สิ่งต่าง ๆ รอบตัวคนเรานั้น เพื่อความสะดวกและความเข้าใจที่ตรงกันในการเรียกสิ่งนั้นๆ สำหรับนักชีววิทยาแล้ว นำพื้นฐานของรูปร่างลักษณะที่เหมือนกันของสิ่งมีชีวิตต่าง ๆ มาใช้ในการจัดจำแนก เราเรียกวิธีจัดจำแนกทางชีววิทยาว่า อนุกรมวิธาน (Taxonomy) ซึ่งเป็นส่วนหนึ่งในสาขาวิชาระบบวิทยา (Systematic หรือ Systematic Biology) เป็นสาขาทางวิทยาศาสตร์เพื่อตรวจวัด (Detect) พรรณนา (Describe) และอธิบาย (Explain)

ความหลากหลายทางชีววิทยาในโลกนี้ โดยมีวิธีการที่จะวิเคราะห์ข้อมูลที่ได้มา ที่เรียกว่า อนุกรมวิธานเชิงตัวเลข (Numerical Taxonomy หรือบางครั้งเรียกว่า Taxometric) จะอาศัยการวัดความคล้ายคลึงของลักษณะภายนอกโดยรวมทั้งหมด (Overall Similarity) ทั้งที่เป็นลักษณะที่เป็นค่าที่แยกออกจากกันเดี่ยว ๆ (Discrete Character) และลักษณะที่เป็นค่าต่อเนื่อง (Continuous Character) ของสิ่งมีชีวิตที่ศึกษา โดยการเปลี่ยนค่าความแตกต่างทางลักษณะให้กลายเป็นค่าระยะห่าง (Distance) ระหว่างสิ่งมีชีวิตแต่ละชนิด หลังจากนั้นจะใช้วิธีทางสถิติแบบต่าง ๆ อธิบายความแตกต่างของชนิด (Sneath & Sokal, 1973)

การจัดจำแนกสิ่งมีชีวิตออกเป็นหมวดหมู่ตามสายวิวัฒนาการอนุกรมวิธาน (Taxonomy) เป็นวิชาที่ว่าด้วยกฎเกณฑ์เกี่ยวกับ 1) การจัดจำแนกสิ่งมีชีวิต (Classification) ออกเป็นหมวดหมู่ต่าง ๆ 2) การกำหนดชื่อสากลของหมวดหมู่และชนิดของสิ่งมีชีวิต (Nomenclature) 3) การตรวจสอบชื่อวิทยาศาสตร์ของสิ่งมีชีวิต (Identification) โดยสามารถจัดจำแนกหมวดหมู่ของสิ่งมีชีวิต ได้ตั้งแต่ ระดับกลุ่มใหญ่ไปจนถึงระดับเล็ก ได้แก่ อาณาจักร (Kingdom) ไฟลัม, เดวิชั่น (Phylum, Division) คลาส (Class) ออร์เดอร์ (Order) แฟมิลี (Family) จีนัส (Genus), และสปีชีส์ (Species) อนุกรมวิธานเป็นศาสตร์ที่มีขอบเขตกว้างขวางในการศึกษาวิจัยเกี่ยวกับรูปพรรณสัณฐานของสิ่งใดสิ่งหนึ่ง โดยอาศัยข้อมูลหลาย ๆ ด้านของสิ่งใดสิ่งหนึ่งเหล่านั้น ดังเช่น เมื่อนำวิชาอนุกรมวิธานมาใช้ในวิชาพฤกษศาสตร์ จึงหมายถึงวิชา อนุกรมวิธานพืช เป็นวิชาที่ว่าด้วยการจัดเข้ากลุ่มพรรณพืชนั่นเอง ขอบเขตของวิชาอนุกรมวิธานพืช มีดังนี้ (Quicke, 1993)

I. การระบุ (Identification) คือ การพิสูจน์ชนิดพืช โดยการเปรียบเทียบ การตัดสินใจการใช้ความจำอันแม่นยำและประสบการณ์ การใช้อย่างใดอย่างหนึ่งหรือหลาย ๆ อย่างรวมกันนั้นจะมีลำดับขั้นตอน คือ ขั้นแรกใช้การจำ เหมือนกับการจำหน้าเพื่อนหรือบุคคลต่าง ๆ เราสามารถจดจำพืชได้ในขั้นวงศ์ สกุล หรือชนิดได้ แต่อย่างไรก็ตาม ถ้าจะให้มั่นใจในการวิเคราะห์ว่าถูกต้องหรือไม่ จะต้องทำการยืนยัน โดยนำพรรณไม้นั้นไปเปรียบเทียบกับตัวอย่างพรรณไม้ที่มีชื่อแล้วในหอพรรณไม้ต่อไป แต่ถ้าจำไม่ได้ก็มีวิธีต่อไป คือการใช้กุญแจหรือรูปวิธาน สามารถทำได้ถ้าตัวอย่างพืชนั้นสมบูรณ์ คือต้องเป็นกิ่งที่ประกอบด้วย ใบ ดอก และผล การใช้รูปวิธานเป็นวิธีที่ดีที่สุดที่สามารถใช้ได้ในทุกระดับของพืช และก็เช่นกันต้องตรวจสอบความถูกต้องในการใช้รูปวิธานนี้โดยการเปรียบเทียบ

สำหรับการเปรียบเทียบ (Comparison) ตัวอย่างพรรณไม้นั้น สามารถทำได้โดยวิธีการต่าง ๆ คือ เปรียบเทียบกับตัวอย่างพืชที่มีอยู่แล้วในหอพรรณไม้ เปรียบเทียบกับตัวอย่างสดที่มีชื่อที่ถูกต้องแล้วในสวนพฤกษศาสตร์ เปรียบเทียบกับภาพที่มีชื่อที่ถูกต้องแล้ว หรือเปรียบเทียบรายละเอียดของพืชนั้น ๆ ในหนังสือพรรณพฤกษชาติต่าง ๆ

2. การบัญญัติชื่อ (Nomenclature) คือการกำหนดตั้งชื่อพรรณพืช ภายหลังจากที่ได้ทำการวิเคราะห์พรรณไม้ชนิดใดชนิดหนึ่งแล้ว ก็จะต้องให้ชื่อที่ถูกต้องแก่พรรณไม้นั้น ๆ เป็นหน้าที่ของนักอนุกรมวิธานพืช ที่จะพิสูจน์ว่าชื่อใดเป็นที่ยอมรับ ชื่อใดเป็นชื่อพ้อง หรือชื่อใดที่ใช้ไม่ได้ เมื่อมีการสำรวจพรรณไม้มากขึ้น และพบพรรณไม้ใหม่ ๆ เพิ่มขึ้นก็เกิดปัญหาในการตั้งชื่อพรรณไม้ใหม่เหล่านี้ เพราะยังไม่มีกฎเกณฑ์ที่แน่นอนของการตั้งชื่อ สำหรับชื่อพรรณไม้ชนิดใดที่ได้ประกาศใช้เป็นชื่อแรกไปแล้วก็เป็นที่ยอมรับ และไม่ใช้ซ้ำกันอีก ผู้ที่ตั้งชื่อพรรณไม้ใหม่ ๆ ก็พยายามหลีกเลี่ยงไม่ใช้ชื่อที่ซ้ำกัน ด้วยเหตุนี้นักพฤกษศาสตร์จึงได้ตั้งกฎเกณฑ์ของการตั้งชื่อขึ้นไว้ให้เป็นระบบสากล เพื่อให้ทุก ๆ คนที่ปฏิบัติงานทางด้านอนุกรมวิธานพืช ได้ใช้ชื่อที่ถูกต้องไม่ไขว้เขว กฎนี้เรียกว่า International Code of Botanical Nomenclature

การตั้งชื่อจะต้องมีตัวอย่างพรรณไม้ ซึ่งผู้ตั้งได้ใช้เป็นตัวอย่างในการวิเคราะห์ และบรรยายลักษณะ กฎเกณฑ์มีดังนี้ คือ เมื่อพบพรรณไม้ชนิดใหม่ที่ยังไม่เคยมีชื่อมาก่อน เมื่อตั้งชื่อแล้วจะต้องเขียนบรรยายลักษณะ และตีพิมพ์ในเอกสารพฤกษศาสตร์เผยแพร่ไปทั่วโลก ในการบรรยายลักษณะก็ต้องดูตัวอย่างพรรณไม้ประกอบไปด้วย ตัวอย่างพรรณไม้ที่ใช้ดูประกอบนี้ เรียกว่าเป็น Type Speciment ในการเรียกชื่อหมวดหมู่ของพรรณไม้ใดต้องปฏิบัติตาม ลำดับก่อนหลังของการตีพิมพ์ คือ ตีพิมพ์ก่อน และถูกต้องตามกฎ คือ ยึดชื่อพรรณไม้ที่ได้ตั้งชื่อถูกต้องตามกฎเกณฑ์และได้ตีพิมพ์ในเอกสารก่อนเป็นอันถูกต้อง ชื่อพืชและทุก ๆ หมวดหมู่ของพืชจึงมีชื่อที่ถูกต้องเพียงชื่อเดียว

3. การจัดเข้ากลุ่ม (Classification) คือ การจัดเข้ากลุ่มพรรณพืชขึ้นเป็นหมวดหมู่ต่าง ๆ การจัดเข้ากลุ่มอย่างง่าย ๆ คือ จัดพืชเป็นพวกไม้ต้น ไม้พุ่ม ไม้เถา และไม้ล้มลุก หรือจัดเป็นจำพวกผักกูด จำพวกไม้สน จำพวกพืชใบเลี้ยงคู่ ใบเลี้ยงเดี่ยว เหล่านี้ พอจะกล่าวได้ว่าการจัดเข้ากลุ่ม คือการจัดหมวดหมู่พืชที่ลักษณะคล้ายคลึงกันเข้าไว้ด้วยกัน ตามหลักปฏิบัตินั้นพรรณพืชที่มีลักษณะคล้ายกันหลายประการนั้นก็จัดขึ้นเป็นสกุลหนึ่ง (Genus)

4. การบรรยายลักษณะ (Description) พืชแต่ละชนิดก็มีรูปพรรณสัณฐานลักษณะต่าง ๆ แตกต่างกันไป เมื่อวิเคราะห์พืชชนิดใดชนิดหนึ่งได้อย่างถูกต้อง ต่อไปจำเป็นจะต้องบรรยายลักษณะต่าง ๆ ของพืชชนิดนั้น ทั้งนี้เพื่อความกระจ่างในการถ่ายทอดข้อมูลตามหลักอนุกรมวิธานพืช นอกจาก Description ของ Species แล้ว ก็มี Description ของสกุล (Genus) และวงศ์ (Family) อื่น ๆ อีก

5. ความสัมพันธ์ (Relationships) ของพืช ช่วยให้เราจำแนกชื่อพรรณไม้ได้อย่างถูกต้องหรือใกล้เคียงมากที่สุด พืชในสกุลเดียวกัน จะมีความละม้ายคล้ายคลึงกันมากกว่าพืชในสกุลอื่น ๆ หรือพืชวงศ์อื่น ๆ อย่างไรก็ตาม นักพฤกษศาสตร์จำต้องใส่ใจอยู่เสมอว่าพืชที่มีลักษณะละม้าย

คล้ายคลึงกันไม่จำเป็นจะต้องมีความสัมพันธ์อย่างใกล้ชิดเสมอไป เมื่อความรู้เรื่องพรรณพฤกษชาติของโลกมีเพิ่มมากขึ้น นักอนุกรมวิธานพืชก็สามารถรู้ถึงความสัมพันธ์ซึ่งกันและกันของพรรณพืชต่าง ๆ และได้อาศัยความรู้นี้กำหนดวิธีการจัดเข้ากลุ่มพรรณพืชให้ดียิ่งขึ้น อีกทั้งการสืบสาวและหาพืชชนิดใหม่ ๆ ที่มีความสัมพันธ์ใกล้เคียงกับชนิดพืชที่ทำการศึกษาวิจัยอยู่เดิม อันอาจจะนำไปสู่การปรับปรุงพันธุ์ใหม่

ดังนั้น อนุกรมวิธานจึงเป็นการศึกษาวิจัยเกี่ยวกับรูปพรรณสัณฐานของสิ่งใดสิ่งหนึ่งโดยอาศัยข้อมูลหลาย ๆ ด้านของสิ่งใดสิ่งหนึ่งเหล่านั้น หรือการจัดเข้ากลุ่มชนิดของสิ่งมีชีวิตนั่นเอง จากแนวคิดดังกล่าวถูกนำมาใช้จำแนกชนิด ประเภท และจัดกลุ่ม คุณลักษณะแฝง (Latent Structure) ของภาวะสันนิษฐาน (Construct) เป็นวิธีการที่เรียกว่า “Taxometric” (การวัดอนุกรมวิธาน เป็นคำเหมือนของ Numerical Taxonomy = อนุกรมวิธานเชิงจำนวน) (Sneath & Sokal, 1973, p. 4) ซึ่งการวัดอนุกรมวิธานตามวิธีการของ Meehl และคณะ (Ruscio, Haslam, & Ruscio, 2006) ก็คือ ทุกอย่างที่เกี่ยวกับลักษณะของความแตกต่าง ถ้าเราเริ่มจากการสังเกตอย่างง่าย ๆ ถ้าสิ่งนั้นไม่มีความแตกต่างกันเลยนั่นคือ ความเหมือนกัน ยกตัวอย่างความแตกต่างระหว่าง แมวกับสุนัข ความแตกต่างระหว่างความร้อนกับความเย็น หรือความแตกต่างระหว่างทองกับเงินซึ่งมาจากความต่างจากหินก้อนใหญ่ (Boulders) กับหินก้อนเล็ก ก้อนกรวด (Pebbles) หรือความแตกต่างระหว่างการแยกประเภทของสิ่งมีชีวิต ในตัวแปรเชิงกลุ่ม หรือจำแนกชนิด ซึ่งอย่างน้อยที่สุดเมื่อจำแนกประเภทของระดับการรวมกลุ่มสูงจะจำแนกเป็น อาณาจักร ชั้น ลำดับ และในบางครั้งการจัดเข้ากลุ่มระดับการรวมกลุ่มต่ำ จะจำแนกเป็น สกุลหรือชนิด เช่นเดียวกับความแตกต่างระหว่างองค์ประกอบทางเคมี ในทางตรงกันข้าม ความแตกต่างของอุณหภูมิ หรือความแตกต่างของขนาดเป็นความต่างเชิงปริมาณหรือระดับ บางสิ่งในโลกนี้ ดูเหมือนจะมีการแยกประเภทขึ้น ยกตัวอย่างในสัตว์เช่น อาจจะเป็นปลา หรือแมลง แต่มันไม่สามารถเป็นได้ทั้งปลาและแมลง ในบางสิ่งจะไม่มีรอยตะเข็บแบ่งแยกความต่างในความเป็นมิติ (Dimensional) มีเพียงความผิดแผกในขนาดตัวอย่าง ถึงแม้ว่าเส้นจะสามารถแสดงถึงการจัดเข้ากลุ่มความแตกต่าง ของคนที่มีความสูงมากจากคนทั้งหมด ซึ่งภาพของเส้นนั้นไม่ได้แสดงเป็นอย่างอื่น แต่เป็นการแยกประเภทตามลำดับที่ต่อเนื่องของความสูงมนุษย์ (Meehl, 1992)

Meehl (1995) ได้อธิบายวิธีการวัดอนุกรมวิธานเป็นการตอบที่กว้างขวางสำหรับความแตกต่างระหว่างชนิด และระดับความแตกต่างที่ได้รับความสนใจของนักจิตวิทยา อย่างไรก็ตาม ที่ผ่านมามีลักษณะที่หลากหลายทางหลักวิชาโดยขึ้นอยู่กับบริบท ในความแตกต่างนี้ มีการเปรียบเทียบสองสิ่งระหว่าง การจำแนกประเภท (Categorical) กับความเป็นมิติ ระหว่างประเภทกับคุณสมบัติ ความผันแปรไม่ต่อเนื่องกับต่อเนื่อง หรือความต่างระหว่างตัวแปรเชิงกลุ่มกับ

เชิงปริมาณ โดยที่ Meehl (1992) ให้ความสำคัญกับการจัดเข้ากลุ่มทางชีววิทยา อ้างอิงการจัดเข้ากลุ่มประเภทในฐานะหน่วยอนุกรมวิธาน (Taxa) และความไม่สอดคล้องตัวแปรที่แสดงชั้นคุณลักษณะแฝง (Latent Class) ของหน่วยอนุกรมวิธานและส่วนที่เหลือ (ชั้นขององค์ประกอบความไม่สอดคล้องของสิ่งหนึ่งสิ่งใดซึ่งไม่เป็นสมาชิก (Complement) ของหน่วยอนุกรมวิธาน) เช่นเดียวกับอนุกรมวิธาน ในความรู้สึกกว้าง ๆ ของการจัดเข้ากลุ่มประเภท นั้นเป็นการคัดเลือกหน่วยอนุกรมวิธาน ขึ้นอยู่กับความแตกต่างระหว่างการเป็นสมาชิกและการไม่ได้เป็นสมาชิก

คุณลักษณะแฝง (Latent Structure) เป็นการแสดงผลของจำนวน ที่ให้ความสนใจในพฤติกรรมที่ผันแปรในตัวมนุษย์ และเป็นการแสดงผลเชิงประจักษ์ ระหว่างความแตกต่างของคุณลักษณะการแบ่งแยกประเภท (Toxonic) กับความเป็นมิติ ความรู้เกี่ยวกับคุณลักษณะแฝงได้รับการพิจารณาและทำให้ชัดเจน ในการจัดเข้ากลุ่มอย่างเป็นระบบ และถูกนำไปใช้กับการวินิจฉัย จัดประเภทบุคคล หรือการเข้าถึงตำแหน่งของคุณลักษณะแฝงที่เป็นค่าต่อเนื่อง (Continua) สามารถช่วยอธิบายทางจิตวิทยา และแสดงผลส่วนรวมของความคิดที่เกิดจาก ภาวะสันนิษฐานทางจิตและพฤติกรรม และที่สำคัญเป็นประโยชน์อย่างยิ่งต่อการปฏิบัติของนักวิจัยที่จะนำไปกำหนดคำถามวิจัย การออกแบบระเบียบวิธีวิจัย และวิธีการทางสถิติที่ใช้ตอบคำถามการวิจัยนั้น เหตุผลทั้งหมดนี้ (Ruscio, Haslam , & Ruscio, 2006 pp. 6 - 8) ได้กล่าวถึง วิธีการวัดอนุกรมวิธาน เป็นวิธีการที่สามารถอธิบายและมีอำนาจในการทดสอบ ความแตกต่างระหว่างหน่วยอนุกรมวิธาน (Taxa; ถ้าเป็นพหูพจน์ใช้ Taxon) ออกจากความเป็นมิติ ที่แสดงออกในทางปฏิบัติและทางพฤติกรรมศาสตร์ให้เห็นกัน ด้วยความพยายามทางพฤติกรรมศาสตร์ที่มีหลายเหตุผลหลายวิธีที่จะรู้ถึงคุณลักษณะแฝงของ ภาวะสันนิษฐานประกอบด้วยวิธีต่าง ๆ ได้แก่ การจัดเข้ากลุ่ม (Classification) การวินิจฉัย (Diagnosis) การประเมิน (Assessment) การวิจัย (Research) การอธิบาย (Explanation) และการสืบสวน (Investigation) เป็นต้น (Mccutcheon, 1987)

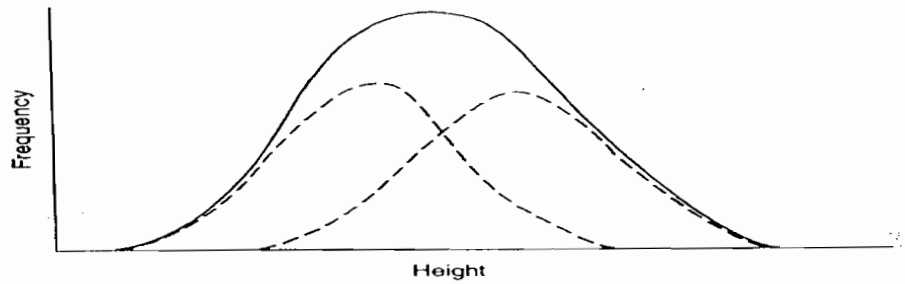
วิธีการวัดอนุกรมวิธานได้รับความสนใจในช่วงระยะเวลาประมาณ 30 ปีที่ผ่านมา โดยถูกนำมาใช้กับการศึกษาเกี่ยวกับบุคลิกภาพของบุคคล (Individual Personality) และการวินิจฉัยทางจิตวิทยา รวมทั้งการสืบสวนทางจิตพยาธิวิทยา (Psychopathology: การพิสูจน์หลักฐานทางจิต) (Ruscio & Ruscio, 2004) ซึ่งวิธีการวัดอนุกรมวิธาน โดยเฉพาะกรณีที่เป็นเทคนิควิธีการวิเคราะห์ทางสถิติ มีความสามารถที่จะแยกแยะการแบ่งแยกประเภท (การจัดจำแนกประเภท) ออกจากความเป็นมิติ (Non-Taxonic) มีความเที่ยงและความตรงตามเงื่อนไขของการศึกษา และยังมีความสามารถในการแยกแยะจากวิธีการทางเลือกสามารถเพิ่มความเชื่อมั่นในการลงสรุปผลอ้างอิงวิธีการวัดอนุกรมวิธานจึงถือว่าเป็นวิธีการเฉพาะที่จะตอบปัญหา ในคำถามวิจัยที่เกี่ยวกับการจัดเข้ากลุ่มประเภท (Powers & Xia Yu, 2000; Gordon, 1999; Agresti, 1990)

แนวคิดพื้นฐานเกี่ยวกับการวัดอนุกรมวิธาน

ในปี ค.ศ. 1962 Paul Meehl ได้ตีพิมพ์ครั้งแรก นำเสนอวิธีการใหม่ที่ใช้จำแนกความแตกต่าง เกี่ยวกับตัวแปรจำแนกประเภท (Categorical Variables) กับตัวแปรต่อเนื่อง (Continuous Variables) ได้พัฒนา ประเมิน และทำให้ชัดเจนของวิธีการวิเคราะห์ข้อมูล ซึ่งเรียกว่า วิธีการวัดอนุกรมวิธานของ Meehl (1992) นำไปใช้กับผู้มีหน้าที่สืบสวน ในการศึกษาคุณลักษณะแฝงของภาวะสันนิษฐาน จำนวนมาก ๆ โดยเฉพาะในขอบเขตของการศึกษาบุคลิกภาพ และจิตพยาธิวิทยา นักวิจัยใช้วิธีการวัดอนุกรมวิธานจัดกระทำต่อการทดสอบเชิงประจักษ์นำไปสู่การกำหนดเกณฑ์ของตัวแปรแฝงที่เกิดจากข้อมูลที่สังเกตได้ว่าจะเป็นข้อมูลจำแนกประเภทหรือข้อมูลต่อเนื่อง (Ruscio, Haslam, & Ruscio, 2006)

ตามแนวคิดของ Meehl (1992) การจำแนกคุณลักษณะแฝง โดยใช้วิธีวัดอนุกรมวิธานเป็นทั้งเทคนิควิธีทางสถิติ (Analytic Techniques) และระเบียบวิธี (Methodology) ที่ใช้เพื่อทดสอบคุณลักษณะว่าเป็นแบบจำแนกประเภท หรือแบบต่อเนื่อง (Continuous) การวิเคราะห์การวัดอนุกรมวิธานมีประโยชน์ในการบ่งบอกว่าความเชื่อ ทัศนคติ ความรู้สึก หรือพฤติกรรมมีความเชื่อมโยงกันในลักษณะที่เป็นหน่วยอนุกรมวิธานอันดับชั้นที่ปรากฏชัด (Manifest Class Taxa) ไม่ว่าความแตกต่างระหว่างบุคคลจะเป็นแบบจัดกลุ่มหรือแบบต่อเนื่อง ก็มีผลในทางทฤษฎีการประเมิน การจำแนกประเภท และการวิจัยในงานอาชญายุติธรรม (Criminal Justice) วิธีวัดอนุกรมวิธานของ Meehl (1995) ช่วยให้นักวิจัยสามารถทดสอบระหว่างโมเดลโครงสร้างที่คล้ายกัน 2 แบบ กรอบโครงสร้างเชิงอนุมานของวิธีนี้รวมทั้งวิธีวิเคราะห์ข้อมูล เนื่องจากแนวทางการดำเนินการวิเคราะห์ด้วยวิธีวัดอนุกรมวิธานและการแปลผลไม่ค่อยเป็นจุดสนใจในการวิจัย จึงกระตุ้นให้ผู้วิจัยนำวิธีฐานรากเชิงประจักษ์ (Empirically Grounded Approach) มาใช้ในการวิเคราะห์วิธีวัดอนุกรมวิธานมากกว่าทำตามแบบเดิม ๆ หรือตามความคิดเห็นส่วนตัว อาจใช้ข้อมูลเปรียบเทียบจำลอง (Simulating Comparison Data) มาประกอบกับแนวทางจากการศึกษาของวิธีการ Monte Carlo รวมทั้งอีกสองวิธีที่กล่าวข้างต้น วิธีฐานรากเชิงประจักษ์ช่วยให้การดำเนินวิธีวัดอนุกรมวิธานได้ผลดีและให้ข้อสรุปที่เที่ยงตรง (Ruscio, 2007) ในการวิเคราะห์การวัดอนุกรมวิธาน เพื่อระบุคุณลักษณะแฝง การวิเคราะห์โดยใช้วิธีวัดอนุกรมวิธาน จะประกอบด้วยชุดสถิติที่ใช้วิเคราะห์ ดังนี้ ค่าเฉลี่ย ส่วนต่างที่สูงกว่าและต่ำกว่าจุดตัด (Mean Above Minus Below a Cut, MAMBAC) ค่าไอเกนสูงสุด (Maximum Eigen Value, MAXEIG) การวิเคราะห์องค์ประกอบลักษณะแฝง (Latent Mode Factor Analysis, L - Mode) และการวิเคราะห์ความแปรปรวนร่วมสูงสุด (Maximum Covariance Analysis, MAXCOV)

หน่วยอนุกรมวิธาน (Taxon) คือ กลุ่มที่ต่างชนิดกันกับอีกกลุ่มหนึ่ง ตัวอย่างเช่น เพศชายเป็นอนุกรมวิธานที่ต่างจากอนุกรมวิธานของเพศหญิง ในทางตรงข้าม คนที่วิตกกังวลมักต่างจากคนที่ไม่ค่อยวิตกกังวลเพียงเพราะความจริงที่ว่า คนกลุ่มแรกมีความเป็นมิติของความวิตกกังวลสูงกว่า จึงเป็นประโยชน์หากทราบว่ามีความแตกต่างเชิงอนุกรมวิธานหรือความแตกต่างเชิงความเป็นมิติหรือไม่ ที่เป็นสิ่งจำแนกกลุ่มหนึ่งออกจากอีกกลุ่มหนึ่ง เป็นเรื่องง่ายที่จะบอกว่าปรากฏการณ์บางอย่าง (เช่น เพศ) เป็นเชิงอนุกรมวิธาน และปรากฏการณ์อื่นเป็นเชิงความเป็นมิติอย่างเห็นได้ชัด (เช่น เชาว์ปัญญา) โดยแจกแจงคนอย่างต่อเนื่องตามความเป็นมิติที่ต่างกันแล้วปรากฏการณ์อย่างเช่น โรคจิตเสื่อม หรือว่าโรคจิตเสื่อมเป็นกลุ่มหนึ่งที่แตกต่างอย่างเด่นชัดจากสภาวะที่ไม่มีอาการจิตเสื่อม หรือว่าความเป็นมิติที่คะแนนสูงมากซึ่งเข้าข่ายตามการวินิจฉัยหรือไม่ หากตอบคำถามนี้ได้ก็จะช่วยให้เราทำการวิจัยมุ่งไปที่สาเหตุของโรค หากโรคจิตเสื่อมเป็นเชิงอนุกรมวิธาน ซึ่งเป็นไปได้ว่ามียีน (Gene) ตั้งต้นตัวหนึ่ง ตัวใดที่เกี่ยวข้องกับความผิดปกติ วิธีค้นหาด้วยการวัดอนุกรมวิธานได้ออกแบบมาเพื่อตอบคำถามนี้ วิธีเหล่านี้นำมาประยุกต์ใช้เมื่อเราสงสัยว่าอาจเป็นอนุกรมวิธาน แต่ข้อบ่งชี้ที่เรามีแสดงถึงการแจกแจงแบบต่อเนื่อง ซึ่งจะเห็นภาพชัดเจนขึ้นถ้าเรามีตัวอย่างที่เราทราบอยู่แล้วว่าเป็นอนุกรมอภิธาน สมมติว่าเรามองไปที่คนและสงสัยว่ามีคนอยู่ 2 ประเภทจริง ๆใช่หรือไม่ (ชาย กับ หญิง) สมมุติว่าไม่มีคุณลักษณะที่แตกต่างกันอย่างเห็นได้ชัด เช่น อวัยวะเพศ ที่จำแนกคน แต่ที่ไม่เหมือนกับเพศ คือ อนุกรมวิธานส่วนใหญ่โดยธรรมชาติจะไม่มีวิธีที่ชัดเจนเช่นนั้นที่จะบ่งบอกว่าเป็นกลุ่มอนุกรมวิธาน บางคนสูงกว่าหรือหนักกว่า นั่นคือ รูปร่างมักต่างกันที่ความสูงเมื่อเทียบกับคนที่เตี้ย เราอาจสังเกตเห็นว่ากล้ามเนื้อใหญ่ขึ้นและมีขนตามตัวมากขึ้นโดยทั่วไปอยู่ในคนที่สูงกว่า และความสามารถทางสังคมมากกว่าส่วนใหญ่อยู่ในคนเตี้ย แต่ถ้าเราวัดคุณสมบัติอย่างใดก็ได้ที่กล่าวมานี้ เราพบว่าดูเหมือนเป็นการแจกแจงแบบต่อเนื่อง แต่เรายังสงสัยอยู่ว่าการแจกแจงแต่ละแบบเป็นผลรวมของการแจกแจง 2 แบบจริงหรือไม่ ในภาพที่ 6 แสดงให้เห็นสมมุติฐานดังกล่าว ซึ่งแสดงการแจกแจงความสูงของกลุ่มคนขนาดใหญ่ โดยที่เส้นประ หมายถึง การแจกแจงตามสมมุติฐานของหญิงและชาย ในตัวอย่างสมมุติฐานนี้ เราไม่ทราบว่าการแจกแจงกลุ่มย่อยเหล่านี้ (ชาย เปรียบเทียบกับหญิง) มีอยู่จริงหรือไม่ ถึงแม้จะมีอยู่จริง เราไม่มีทางบอกได้ว่าใครอยู่ในกลุ่มไหน หากเราจำแนกว่าทุก ๆ คนที่อยู่เหนือความสูงระดับหนึ่งเป็นเพศชาย และทุกคนที่อยู่ใต้ความสูงนั้นเป็นเพศหญิง อาจจะพบว่า ไม่ว่าเราจะเลือกความสูงระดับใด เราจะสร้างความผิดพลาดในการจำแนกหลายครั้ง เพราะการแจกแจงทับซ้อนกันอย่างมาก วิธีค้นหาด้วยการวัดอนุกรมวิธานเป็นการค้นหาความสัมพันธ์ระหว่างตัวแปรที่น่าจะมีอยู่ก็ต่อเมื่อข้อมูลเป็นหน่วยอนุกรมวิธาน หากพบความสัมพันธ์ดังกล่าว แสดงว่ามีกลุ่มที่ต่างกัน 2 กลุ่ม แต่ถ้าไม่พบความสัมพันธ์ แสดงว่ากลุ่มตัวอย่างแจกแจงแบบความเป็นมิติ (Waller & Meehl, 1998)



ภาพที่ 6 แสดงการแจกแจงความสูงของหน่วยอนุกรมวิธาน (เพศชาย และเพศหญิง)

คุณสมบัติของหน่วยอนุกรมวิธาน จะประกอบด้วย 1) เป็นค่าที่สังเกตได้จากคะแนนที่เกิดจากการสังเกตได้ (Manifest) หรือลักษณะที่บ่งชี้เฉพาะ 2) สามารถกำหนดความหมายที่แตกต่างจากความเป็นมิติ 3) สามารถจัดลำดับชั้นภายในกับหน่วยอนุกรมวิธานอื่น 4) สามารถระบุความแตกต่างระหว่างจากสิ่งหนึ่งกับสิ่งอื่นๆในจำนวนเดียวกัน 5)ไม่เกี่ยวข้องกับพื้นฐานทางชีววิทยาหรือเกี่ยวพันกับส่วนที่สำคัญร่วมระหว่างกันของสิ่งนั้น (Arnau, Thomson, & Cook, 2001)

ความแตกต่างระหว่างหน่วยอนุกรมวิธาน กับ ความเป็นมิติ ความเข้าใจในคุณสมบัติของหน่วยอนุกรมวิธานกับ ความเป็นมิติ นั้น ดังตัวอย่าง ถ้า "แมว" เป็นหน่วยอนุกรมวิธาน ก็เพราะว่า มีเอกลักษณ์เฉพาะตัวที่นับจำนวนได้ ที่มีชั้นคุณลักษณะแฝง มีลักษณะทางธรรมชาติที่จำแนกประเภทจากขอบเขตแยกระหว่างแมว กับที่ไม่ใช่แมว แมวไม่สามารถเปลี่ยนแปลงไปเป็นสิ่งอื่นได้ ในทางตรงกันข้ามสัตว์เลี้ยงไม่ได้เป็นหน่วยอนุกรมวิธาน เพราะเป็นขอบเขตระหว่างสัตว์และไม่ใช่สัตว์ขึ้นอยู่กับข้อตกลงทางสังคม นั่นคือสามารถผันแปรตามเวลาและสถานที่ ดังนั้นบางสิ่งอาจจะเรียกว่าสัตว์เลี้ยง ซึ่งไม่สามารถแยกแยะตามชั้นคุณลักษณะแฝง เหมือนกับความสูงของคนไม่ได้เป็นหน่วยของอนุกรมวิธาน เพราะไม่มีชั้นคุณลักษณะแฝงของคนที่มีความสูง ซึ่งเป็นการแสดงถึงความแตกต่างของประเภทจากประเภทอื่น ๆ ซึ่งไม่สามารถตรวจพบได้กับขอบเขตของค่าต่อเนื่องของความสูงที่ระดับของคุณลักษณะแฝง ความสูงของคนเป็นการทำความเข้าใจที่ดีต่อภาวะสันนิษฐานของความเป็นมิติ (Meehl, 1999)

ปัญหาของการจัดเข้ากลุ่ม (Classification Problem) วิธีการวิเคราะห์มีการพัฒนาในหลายวิธีที่พยายามจะแก้ปัญหาของการจัดเข้ากลุ่ม การที่จะกำหนดภาวะสันนิษฐาน ว่าเป็นการจำแนกประเภท หรือ ความเป็นมิติ ตามระดับของคุณลักษณะแฝง การแจกแจงความถี่จากหลักฐานของวิธี Bimodality จึงเป็นวิธีที่นิยมใช้ในการค้นหาหน่วยอนุกรมวิธาน วิธีการนี้มีข้อบกพร่องของค่าความเชื่อมั่น และล้าสมัย เมื่อมีการปรับปรุงวิธีการวิเคราะห์เชิงพหุ (Pampel,

2000) ที่มีความสามารถในการวิเคราะห์ได้ซับซ้อนมากกว่า มีการเปรียบเทียบความตรงกับ Model แบบผสม การวิเคราะห์กลุ่ม การวิเคราะห์ชั้นคุณลักษณะแฝง และการวัดอนุกรมวิธาน เป็นการจัดเข้ากลุ่มโครงสร้างของการจำแนกประเภทออกจาก ความเป็นมิติ การวิจัยเปรียบเทียบเป็นความต้องการที่สำคัญ ในทางปฏิบัติ คำถามในการศึกษายังเปิดโอกาสให้ แต่ละเทคนิควิธีที่สามารถจำแนกความแตกต่างระหว่างการจำแนกประเภทออกจากความเป็นมิติได้อย่างถูกต้อง (Meehl, 1999)

ในขณะเดียวกัน การทดลองในขอบเขตของ Model แบบผสม การวิเคราะห์กลุ่ม การวิเคราะห์ชั้นคุณลักษณะแฝง จะประสบปัญหาที่มีความยากที่จะตัดสินว่าอยู่ภายใต้ความเป็นภาวะสันนิษฐานอย่างถูกต้องหรือไม่ เป็นการยากที่จะแยกว่าเป็น Model สองกลุ่ม (Taxonic) จากกลุ่มเดียว (Dimensional) เมื่อการจัดเข้ากลุ่ม ซึ่งเป็นการนำวัตถุที่ได้มาใหม่มาวัดคุณสมบัติต่าง ๆ (วัดค่าตัวแปร) แล้ววิเคราะห์ว่า คุณสมบัติเช่นนั้นควรจัดวัตถุนั้นเข้าพวกใด ข้อจำกัดอยู่ที่ว่า (มนตรี พิริยะกุล, 2545, หน้า 187) ถ้าการแบ่งกลุ่มให้ผลลัพธ์ของการแบ่งกลุ่มที่ชัดเจนมาก (คือ ไม่เหลื่อมกัน; Redundancy) การจัดกลุ่มก็จะทำได้ง่าย แต่ผลการแบ่งกลุ่มปรากฏว่าได้กลุ่มที่เหลื่อมกัน วัตถุนั้นอาจถูกจัดเข้ากลุ่มผิด และอาจถูกจัดเข้าเป็นสมาชิกของประชากรมากกว่า 1 พวกได้ ในทางตรงกันข้ามวิธีการวัดอนุกรมวิธาน มีความสามารถที่จะแยกแยะความเป็นหน่วยอนุกรมวิธานออกจากความเป็นมิติ มีความเที่ยงและความตรงตามเงื่อนไขของการศึกษา และยังมีความสามารถในการแยกแยะจากวิธีการทางเลือกสามารถเพิ่มความเชื่อมั่นในการลงสรุปผลอ้างอิง ดังที่ Beauchaine & Waters (2003) กล่าวว่า วิธีการวัดอนุกรมวิธานจึงถือว่าเป็นวิธีการเฉพาะที่จะตอบปัญหา ในคำถามวิจัยที่เกี่ยวกับการจำแนกประเภท

วิธีการ Taxometric Analysis

วิธีการวัดอนุกรมวิธาน คือ วิธีการทางสถิติสำหรับระบุความสัมพันธ์ระหว่างค่าที่สังเกตได้ของหน่วยอนุกรมวิธานแฝง เช่น ประเภท ชนิด กลุ่ม หรือลักษณะของโรค (Meehl, 1999) วิธีการวัดอนุกรมวิธานประกอบด้วยวิธีการวัดที่สามารถใช้ประเมินคุณลักษณะแฝงภายใต้การสืบสวน เพราะเป็นวิธีการที่ได้มาจากการคำนวนระยะห่าง โดยหลักฐานไม่ซ้ำซ้อนกันสามารถตรวจสอบความสอดคล้องของผลลัพท์ได้ถูกต้อง และสร้างความเชื่อมั่นในการนำไปอ้างอิงของคุณลักษณะแฝง สำหรับวิธีการวัดอนุกรมวิธานจะนำเสนอสรุปดังนี้ (Ruscio, Haslam, & Ruscio, 2006)

1. ข้อมูลสำหรับการวัดอนุกรมวิธาน

วิธีการวัดอนุกรมวิธานมีความต้องการข้อมูลที่เป็นกลุ่มตัวอย่างขนาดใหญ่ (Ruscio, Haslam, & Ruscio, 2006) ซึ่งให้ความสนใจกับการหลีกเหลี่ยงการสร้างจากระเบียบวิธี คุณลักษณะของกลุ่มตัวอย่างที่สำคัญจะต้องประกอบด้วย ความเป็นตัวแทนที่เหมาะสมของหน่วยอนุกรมวิธาน

ตัวบ่งชี้มีความตรงสูง ตัวบ่งชี้มีความสัมพันธ์ภายในต่ำ และตัวบ่งชี้ซึ่งมีการแจกแจงจำนวนและคะแนนมีความเหมาะสมสำหรับวิธีการและการทดสอบความสอดคล้อง แม้ว่าการตัดสินเกี่ยวกับความเหมาะสมของชุดข้อมูลโดยเฉพาะสำหรับการวิเคราะห์เฉพาะเจาะจงแบบดั้งเดิม ซึ่งเกิดขึ้นบนพื้นฐานแนวทางหลัก จากการศึกษาแบบจำลองมอนติคาโล และการแนะนำเพิ่มต่อการตัดสินจากการสร้างข้อมูลเชิงประจักษ์ด้วยการแจกแจงแบบสุ่มของข้อมูลเชิงประจักษ์ผ่านการ Bootstrap ของการจำแนกกลุ่ม และวิเคราะห์ข้อมูลการเปรียบเทียบกับ ความเป็นมิติ

ความสำคัญที่ต้องไม่ลืมว่าประสิทธิภาพและข้อจำกัดของเทคนิควิธีนอกเหนือจากที่นำเสนอไว้ โดยเฉพาะถึงแม้ว่าทำการทดสอบความเหมาะสมจุดเชื่อมต่อของข้อมูล และวิเคราะห์แบบแผน การไม่ทำการตรวจสอบหรือไม่ยอมรับการเลือกตัวอย่างที่ถูกต้อง อย่างไรก็ตาม เมื่อใช้อย่างระมัดระวังอย่างเหมาะสม การแจกแจงแบบสุ่มของข้อมูลเชิงประจักษ์สามารถเกิดผลเพิ่มเติมในหลักทั่ว ๆ ไป ซึ่งได้มาจากการจำลองแบบมอนติคาโล ด้วยการพิจารณาองค์ประกอบที่เป็นเอกลักษณ์ของคุณลักษณะของข้อมูลนั้น และผลที่เกิดจากวิธีการ ในการสืบสวนด้วยการวัดอนุกรมวิธาน นักวิจัยควรใช้ Bootstrap ข้อมูลเปรียบเทียบระหว่างการแบ่งแยกประเภทกับความเป็นมิติ สำหรับแต่ละชุดของข้อมูลที่คาดว่าจะเป็นตัวบ่งชี้ และนำเสนอแต่ละชุดของข้อมูลเปรียบเทียบต่อทุกขั้นตอนและการทดสอบความสอดคล้องในแบบแผนการวิเคราะห์ เมื่อการแจกแจงแบบสุ่มของข้อมูลเชิงประจักษ์ ข้อมูลเชิงประจักษ์ของผลลัพธ์จากชุดข้อมูลของการแบ่งแยกประเภท กับ ความเป็นมิติ เป็นการแยกประเภท สามารถตัดสินชุดของตัวบ่งชี้มีความพอเพียงในการวิเคราะห์นี้ด้วยความเชื่อมั่นที่เพิ่มขึ้น เมื่อการแจกแจงแบบสุ่มของข้อมูลเชิงประจักษ์ ข้อมูลเชิงประจักษ์ไม่มีความสามารถแยกประเภทได้ ถ้ารับรองว่าตัวบ่งชี้นั้นสอดคล้องกับการวิจัยจะต้องใช้อย่างระมัดระวัง ถ้าทั้งหมดนี้ สำหรับการวิเคราะห์ และการปฏิบัติตามแนวทางนี้สามารถช่วยสนับสนุนข้อมูลเชิงประจักษ์ที่ความเหมาะสมของตัวบ่งชี้ และผลที่เกิดจากวิธีการ โดยขจัดความไม่เหมาะสมของข้อมูล หรือกลวิธีการวิเคราะห์เป็นการอธิบายทางเลือก ซึ่งใช้ปฏิบัติได้สำหรับผลลัพธ์ของการทดลอง สิ่งเหล่านี้มีความสำคัญอย่างยิ่งเมื่อต้องนำไปอ้างอิงถึงคุณลักษณะของความเป็นมิติ

นักวิจัยมีทางเลือกหลายทางเมื่อการแจกแจงแบบสุ่มของข้อมูลเชิงประจักษ์บอกเป็นนัยว่า Model ของการแบ่งแยกประเภท กับ ความเป็นมิติ ไม่สามรถแยกแยะได้ด้วยทางสถิติ เช่น จะประมาณค่าปรับแก้ Model Parameters (ตัวอย่าง ความตรงของตัวบ่งชี้ ความสัมพันธ์ภายในกลุ่ม) นำไปสู่ความพยายามกำหนดว่าอะไรเป็นลักษณะของปัญหาที่ยากจะแก้ไขได้ของข้อมูลอาจจะปรากฏขึ้นได้ การรวมรายการ (Items) จะทำให้ผลของความตรงของตัวบ่งชี้เพิ่มขึ้น หรือทำให้ความสัมพันธ์ในกลุ่มต่ำ หรือความพยายามใช้วิธีการการวัดอนุกรมวิธานที่แตกต่าง หรือแนวทางที่

เหมาะสมของการใช้เครื่องมือในกระบวนการ) นั่นอาจจะบอกความแตกต่างของการแบ่งแยกประเภท จากคุณลักษณะของ ความเป็นมิติ นอกจากนี้ ความละเอียดของชุดตัวบ่งชี้ และ/ หรือ แบบแผนของการวิเคราะห์อาจทำให้ผลสุดท้าย มีประสิทธิภาพมากในการจำแนก Model โครงสร้าง อย่างไรก็ตาม มีความสำคัญต่อการยอมรับ ในบางกรณี ข้อมูลที่สามารถใช้ประโยชน์ได้เป็นข้อมูลธรรมดาสำหรับเป็นสารสนเทศในการวิเคราะห์ด้วยการวัดอนุกรมวิธาน อย่างไรก็ตามเป็นความพยายามที่ต้องกระทำ ผลลัพท์นี้ต้องการมากกว่าแสดงข้อสรุปเชิงโครงสร้าง เมื่อข้อมูลมีความสามารถแยกแยกระหว่างคุณลักษณะแฝงสองกลุ่มนี้ ในแนวทางนี้ ความถูกต้องที่จะสรุปอ้างอิง และการใช้การแจกแจงแบบสุ่มเชิงประจักษ์อาจลดความเสี่ยงของส่วนสรุปของโครงสร้างที่ไม่แน่นอนบนพื้นฐานของผลลัพท์ที่ผิดทางไป

2. MAXSLOPE

วิธีการ MAXSLOPE เป็นแนวคิดอย่างง่าย ๆ ที่เป็นส่วนหนึ่งในวิธีการของการวัดอนุกรมวิธาน เพราะว่าใช้กับตัวบ่งชี้เพียงสองตัวของการวัดภาวะสันนิษฐาน เป็นการวิเคราะห์โดยใช้กราฟ (Grove, 2004)

ตรรกะของวิธีการ MAXSLOPE เริ่มด้วย Scatter plot แสดงความสัมพันธ์ระหว่างตัวบ่งชี้สองตัว ตัวหนึ่งเป็น Input อยู่บนแกน X และอีกตัวหนึ่งเป็น Output อยู่บนแกน Y เมื่อคุณลักษณะแฝงเป็นตัวแบ่งแยกประเภท และตัวบ่งชี้ทั้งสองแยก Taxon จาก Complement ซึ่งมีความตรงสูงพอเพียง และความสัมพันธ์ภายในกลุ่มมีสัดส่วนความแปรปรวนในตัวแปรที่อธิบายไม่ได้ด้วยตัวแปรอื่นมีค่าต่ำ ระยะห่างของจุดทั้งสองจะชัดเจน กลุ่มของจุดที่เป็นตัวแทนสมาชิกของ Taxon อยู่ที่ตำแหน่งที่เกือบจะถึงทางขวาบนของ Scatter Plot ทั้งนี้เพราะว่า แนวโน้มของคะแนนแต่ละคะแนนมีความสัมพันธ์กันสูงในทั้งสองตัวบ่งชี้นั้น และกลุ่มของจุดที่เป็นตัวแทนสมาชิกของ Complement อยู่ที่ตำแหน่งทางล่างซ้ายของ Scatter Plot เพราะว่าแนวโน้มของคะแนนแต่ละคะแนนมีความสัมพันธ์ต่ำในทั้งสองตัวบ่งชี้ ต่อจากนั้นใช้สมการถดถอยอย่างง่ายคำนวนเส้นตรงที่เหมาะสมของ Scatter Plot เส้นโค้งถดถอยเกิดจากการใช้จำนวนของเทคนิคการปรับเรียบ (Smooth Technique) เป็นการประมาณค่าความลาดชัน ภายในขอบเขตของ Scatter Plot เมื่อตัวบ่งชี้ของการแบ่งแยกประเภท เป็นความสัมพันธ์อิสระภายใน Taxon และชั้น Complement (อาจมีความสัมพันธ์เพียงเล็กน้อย) เส้นโค้งถดถอยมีรูปร่างเป็นตัว S (S - Shape) เส้นโค้งจะมีเส้นเกือบเป็นเส้นตรง ตรงกลางระหว่างที่เป็น Taxon และ Complement ในทางตรงกันข้าม ถ้าความสัมพันธ์ระหว่างตัวบ่งชี้หลักทั้งสองจากมิติของคุณลักษณะแฝง Scatter Plot จะมีรูปแบบที่ปกคลุมด้วยจุดที่มีลักษณะเหมือนกัน (Single Homogeneous) เส้นสมการถดถอยที่แสดงนี้จะชันขึ้น แต่จะยังคงเป็นเส้นตรงพาดขยายเต็มกราฟ ดังนั้นผลของโครงสร้างทางมิติจะเปรียบเทียบ

ด้วยการ Plot เส้น MAXSLOP ซึ่งเป็นระยะห่างจาก การ Plot S- Shaped ในการปรับแก้ของโครงสร้าง การแบ่งแยกประเภท (Meehl, 1995)

3. MAMBAC

วิธีการ MAMBAC มีความต้องการตัวบ่งชี้สองตัวแต่สามารถจะแสดงด้วย Complement หลายส่วนของตัวบ่งชี้นำเข้า (Input) นั่นคือ การใช้ข้อมูลทั้งหมดแต่ละเส้นโค้ง ด้วยเหตุนี้จะเพิ่มความชัดเจนของผลลัพท์ภายใต้เงื่อนไขของข้อกำหนด (Meehl, 1995)

ตรรกะของวิธีการ วิธีการ MAMBAC เป็นวิธีการที่นำข้อดีของข้อเท็จจริง ถ้าคะแนนจุดตัดเหมาะสมสำหรับการแยกทั้งสองกลุ่ม นั่นคือถ้าความแตกต่างของตัวบ่งชี้มีความตรงที่จะแยกเป็น Taxon หรือ Complement จำเป็นต้องมีค่าเฉพาะในตัวบ่งชี้ที่จะจำแนกกลุ่มด้วยจำนวนอย่างน้อยสุดของความผิดพลาดทางบวก และทางลบ ในการขาดหายของโครงสร้าง การแบ่งแยกประเภท ไม่มีกลุ่มที่จะแยก และไม่มีคะแนนจุดตัดที่เหมาะสมที่จะกระทำ ดังนั้น วิธีการ MAMBAC ขึ้นอยู่กับการค้นหาคะแนนจุดตัดที่เหมาะสม ถ้าสามารถพบคะแนนนั้น ก็จะเป็นการแบ่งแยกประเภท แต่ถ้าไม่เป็นอย่างนั้น ก็คือโครงสร้างของความเป็นมิติ

วิธีการ MAMBAC ใช้สองกลุ่มของตัวบ่งชี้ค้นหาคะแนนจุดตัดที่เหมาะสม เช่นเดียวกับในวิธีการ MAXSLOPE สิ่งหนึ่งคือ การกระทำกับตัวบ่งชี้นำเข้า และอยู่บนตำแหน่งของแกน X Case ในชุดข้อมูลเป็นการเรียงความเชื่อมโยงในส่วนตัวบ่งชี้ผล ซึ่งเป็นการใช้การสร้างชุดของคะแนนจุดตัด เริ่มต้นด้วยคะแนนจุดตัดไปยังจุดต่ำสุดของตัวบ่งชี้ผล คะแนนเฉลี่ยในตัวบ่งชี้ผลตัวอื่นสำหรับ Case ทั้งหมดจะอยู่ต่ำกว่าจุดตัดเป็นการลบออกจากคะแนนเฉลี่ยทุก Case ที่อยู่เหนือจุดตัด ความต่างของค่าเฉลี่ยเป็นการ Plot ค่าของ Y สำหรับคะแนนจุดตัด (ค่า X) การลบออกเป็นการกระทำซ้ำสำหรับแต่ละคะแนนจุดตัดบนแกน X และ แต่ละคะแนนความเฉลี่ยที่แตกต่างเป็นการ Plot ในฐานะความสอดคล้องของค่า Y ดังนั้น วิธีการ MAMBAC เกี่ยวข้องกับการ Plot ความต่างของคะแนนเฉลี่ยในตัวบ่งชี้ผล สำหรับคะแนนของ Case ที่อยู่เหนือแต่ละจุดตัดในตัวบ่งชี้ผลอยู่ต่ำกว่าคะแนนเป็นลบ (Beauchaine, & Waters, 2003)

รูปร่างของเส้นโค้งผลลัพท์ อ้างอิงไปสู่การแสดงผลเกี่ยวกับคุณลักษณะแฝง ถ้าจุดตัดคะแนนที่เหมาะสม การใช้ความต่างของค่าเฉลี่ยจะเข้าใกล้ตำแหน่งของคะแนนนี้ และจะค่อย ๆ เปลี่ยนไป จะเป็นค่าต่ำกว่า หรือสูงกว่าคะแนนจุดตัด ดังนั้น แนวโน้มโครงสร้างของ การแบ่งแยกประเภทเป็นค่าสูงสุดของเส้นโค้ง ด้วยค่าสูงสุดอธิบายขอบเขตของตัวบ่งชี้ผล ซึ่งคะแนนจุดตัดที่เหมาะสมตั้งอยู่ ในทางตรงกันข้าม โครงสร้างความเป็นมิติมีแนวโน้มเกิดผลเส้นโค้งเว้าด้านนอกจุดสูงสุดที่สังเกตได้ การขาดหายของจุดสูงสุดเป็นนัยของคุณลักษณะแฝงที่เป็นความเป็นมิติ (Haslam,1997)

4. L - MODE

วิธีการ L - MODE เกี่ยวข้องกับพื้นฐานของ Thurstone (Ruscio, Haslam, & Ruscio, 2006) โดยเชื่อว่า องค์ประกอบคุณลักษณะแฝงไม่จำเป็นต้องเป็นตัวแปรต่อเนื่องเสมอ และสามารถใช้ข้อมูลตัวแปรเชิงกลุ่มได้

ตรรกะของวิธีการ การวิเคราะห์องค์ประกอบ (Factor Analysis) มีการใช้เฉพาะต่อการสำรวจความเป็นมิติ ของภาวะสันนิษฐานทางจิตวิทยา ใช้ตัดสินลักษณะของ ความเป็นมิติคุณลักษณะแฝงภายใต้เหตุผลที่ว่าความแปรปรวนในคำตอบต่อข้อคำถาม ในตรงกันข้าม วิธีการ L - MODE จะพยายามอธิบายความแตกต่างของโครงสร้าง ความเป็นมิติ กับ การแบ่งแยกประเภทโดยใช้การแจกแจง (Distribution) จากกราฟ ของคะแนนประมาณค่าของ Case ในองค์ประกอบคุณลักษณะแฝงเชิงเดี่ยว คำนวณโดยใช้วิธีการ Bartlett ของการประมาณค่าคะแนนองค์ประกอบเหมือนบางส่วนที่ประกอบขึ้นของตัวแปร นั่นเป็นการวัดความตรงวิธีหนึ่งที่ใช้วัดคุณลักษณะแฝงนั้น เช่น คะแนนองค์ประกอบจะแยก Taxon และ Complement มีความตรงมากกว่าตัวบ่งชี้เดี่ยว ๆ ซึ่งตัวบ่งชี้ที่มีความตรงพอเพียง โครงสร้าง การแบ่งแยกประเภท มีผลการแจกแจงแบบ Bimodal ของคะแนนองค์ประกอบ ในขณะที่โครงสร้าง ความเป็นมิติ จะมีผลการแจกแจงเป็แบบ Unimodal ของคะแนนองค์ประกอบ เพราะฉะนั้น ผลลัพท์ของวิธีการ L - MODE จะเพิ่มความเหมาะสมด้วยจำนวนที่มากของตัวบ่งชี้ที่มีความตรง

5. MAXCOV และ MAXEIG

วิธีการ MAXCOV (Maximum Covariance) และ MAXEIG (Maximum Eigenvalue) เป็นวิธีการแยก โครงสร้าง การแบ่งแยกประเภท ออกจากโครงสร้าง ความเป็นมิติ ซึ่งใช้วิธีการทางคณิตศาสตร์ โดยแสดงสมการ General Covariance Mixture Theorem (GCMT) ดังนี้

$$\mathrm{cov}(xy) = P\,\mathrm{cov}_I(xy) + Q\,\mathrm{cov}_c(xy) + PQD_xD_y \quad \text{..... (1)}$$

แสดงค่าความแปรปรวนร่วมระหว่างสองตัวบ่งชี้ x และ y ในฐานะของฟังชัน (a) ความแปรปรวนร่วมภายใน Taxon [$cov_t(xy)$] (b) ความแปรปรวนร่วมภายใน Complement [$cov_c(xy)$] และ (c) ความตรงซึ่งสามารถแยกตัวบ่งชี้ออกเป็น Taxon และ Complement (D_x และ D_y) แสดงความแตกต่างของค่าเฉลี่ยระหว่าง Caxon กับ Complement ในตัวบ่งชี้ X และ Y แต่ละค่าเป็นการให้น้ำหนักโดย Base Rate (s) ของชั้นความสัมพันธ์ (es) ซึ่ง Base Rate ของ Taxon = P และ Base Rate ของ Complement คือ Q = 1 - P ถ้าตัวบ่งชี้ไม่เกิดความแปรปรวนร่วมภายในกลุ่ม ทำให้ง่ายขึ้นด้วยสมการ (2)

$$\text{cov}(xy) = PQD_xD_y \quad \text{..... (2)}$$

ความแปรปรวนร่วมของตัวบ่งชี้เป็นฟังชันของการ Base rates ของ Taxon และ Complement รวมทั้งความตรงของสองตัวบ่งชี้ X และ Y

ตรรกะของวิธีการ MAXCOV วิธีการของ MAXCOV เป็นการกระทำโดยข้อเท็จจริงของ GCMT ไม่เพียงเป็นการใช้ภายในกลุ่มตัวอย่างอย่างเต็มที่ของข้อมูลเท่านั้น แต่ภายในหน่วยย่อย (Subset) ของ Case ก็สามารถแสดงได้จากกลุ่มตัวอย่างนั้น จากสมการ (2) แสดงความแปรปรวนร่วมของตัวบ่งชี้จะแปรผันในฐานะฟังชันของ Base Rates ของ Taxon และ Complement เพราะว่าความตรงของตัวบ่งชี้เป็นข้อตกลงที่ไม่แปรผันไปตามกลุ่มย่อย สามารถแทนที่ด้วยค่าคงที่

$$\text{cov}_i(xy) = p_iq_iK \quad \text{..... (3)}$$

เมื่อ K แสดงผลของความตรงตัวบ่งชี้ และ p_i และ q_i แสดง Base Rates ของ Taxon และ Complement ภายในกลุ่มย่อย

ดังนั้น สามารถทดสอบคุณลักษณะแฝง โดยตรวจสอบความแปรปรวนร่วมของสองตัวบ่งชี้ ภายในชุดของกลุ่มย่อย ซึ่งเหมือนกับการกำหนดให้สัดส่วนของความแตกต่างระหว่างสมาชิกของ Taxon และComplement ในกลุ่มย่อยทั้งหมดมีความสอดคล้องกันของสมาชิก Taxon : $p_i = 1$, $q_i = 0$, $p_iq_i = 0$ และ $\text{cov}_i(xy) = 0$ นอกจากนั้น หน่วยย่อยของกลุ่มตัวอย่าง (Subsample) มีความสอดคล้องทั้งหมดในสมาชิกของ Complement : $p_i = 0$, $q_i = 1$, $p_iq_i = 0$ และ $\text{cov}_i(xy) = 0$ อย่างไรก็ตามในหน่วยย่อยของกลุ่มตัวอย่างประกอบด้วยการผสมของสมาชิกใน Taxon และ Complement: $p_i > 0$, $q_i > 1$, $p_iq_i > 0$ และ $\text{cov}_i(xy) > 0$ ความแปรปรวนร่วมจะไปถึงค่าสูงสุดในหน่วยย่อยของกลุ่มตัวอย่างมีความสอดคล้องในการผสมสมการของสมาชิก Taxon และ Complement เมื่อ $p_i = q_i = .5$, $p_iq_i = .25$ และ $\text{cov}_{max}(xy) = .25K$ ปรับใหม่กับสูตรที่ผ่านมาจะสามารถที่จะประมาณค่าผลของความตรงตัวบ่งชี้จะได้ค่า $K = 4\ \text{cov}_{max}(xy)$

ความชัดเจนที่แสดงในสูตรของสมการ (1) - (3) ค่าคาดหวังเช่นเดียวกับการวิเคราะห์ด้วยวิธี MAXSLOPE ในการปรากฏของโครงสร้างของ Taxon ภายในขอบเขตที่กำหนดคุณลักษณะที่เหมือนกันในหน่วยย่อยของกลุ่มตัวอย่างของสมาชิกของ Taxon และ Complement ค่าความลาดชันของสมการโค้งถดถอยจะแบนราบ นี่เป็นความคลายกับค่าศูนย์ ความแปรปรวนร่วมภายในกลุ่ม อย่างไรก็ตาม ภายในขอบเขตที่กำหนด การผสมของสมาชิกของ Taxon และ Complement ค่าความลาดชันของสมการโค้งถดถอยจะเป็นบวก การไปถึงค่าสูงสุดใกล้จุด เมื่อกลุ่มเกิดผสมปะปนกันมาก อย่างไรก็ตาม เมื่อคุณลักษณะแฝงเป็นความเป็นมิติ ความแปรปรวนร่วมระหว่าง ตัวบ่งชี้จะคงค่าคงที่ผ่านหน่วยย่อยของกลุ่มตัวอย่างต่อจากนั้นจะปรากฏขึ้นตามระบบต่อค่าสูงสุด

ตรรกะของวิธีการ MAXEIG ความเข้าใจกรอบแนวคิดและหลักการทางคณิตศาสตร์ของวิธีการ MAXCOV ปูพื้นในการทำความเข้าใจในหลักการวิเคราะห์ตัวแปรหลายตัว (Multivariate) ซึ่งเหมือนกับวิธีการ MAXEIG โดยความแตกต่างของทั้งสองวิธี ซึ่งประกอบด้วย ข้อแรก การคำนวนค่าความแปรปรวนร่วมระหว่างสองตัวบ่งชี้ผล (เหมือน MAXCOV) วิธีการ MEXEIG กำหนดว่า ความสัมพันธ์ระหว่างสองหรือมากกว่าของตัวบ่งชี้ผล โดยคำนวนค่า Eigenvalue ในรูปเมตริกซ์ความแปรปรวนร่วม เมตริกซ์ Diagonal ความแปรปรวน - ความแปรปรวนร่วม เป็นการแทนที่โดยค่าศูนย์ นำออกไปของความแปรปรวน และแยกจากความแปรปรวนร่วม การคำนวนของเงื่อนไข Eigenvalue แทนที่เงื่อนไขความแปรปรวนร่วม คือ ความแตกต่างวิธีการร่วมกัน ระหว่าง MEXEIG และ MAXCOV ความแตกต่างข้อที่สอง การใช้ทุกตัวแปรค่าที่เป็นไปได้หนึ่งในสามของตัวบ่งชี้นำเข้า และตัวบ่งชี้ผล ใช้แต่ละตัวแปรซึ่งเป็นตัวบ่งชี้นำเข้า ซึ่งตัวแปรที่เหลือยู่เกิดขึ้นพร้อมกันเป็นตัวบ่งชี้ผล วิธีการ MEXEIG เกี่ยวข้องกับตัวบ่งชี้ทั้งหมดในการคำนวนของแต่ละเส้นโค้ง ผลที่ได้ทำให้การแยกโครงสร้าง การแบ่งแยกประเภท และความเป็นมิติชัดเจนกว่า ข้อที่สาม เมื่อวิธีการ MAXCOV กระทำด้วยการแบ่งกลุ่มตัวอย่างไปสู่ความไม่ทับซ้อนคล้ายกับช่วงชั้นของตัวบ่งชี้ผล ส่วนวิธีการ MEXEIG กระทำด้วยการแบ่งกลุ่มตัวอย่างไปสู่ความทับซ้อน ผลทั้งหมดของกระบวนการหลังของจุดข้อมูล นั้นจะมาจากการใช้ขนาดช่วงชั้นเท่ากัน การเพิ่มจำนวนจุดข้อมูลซึ่งอยู่ตรงกลางจะเป็นสิ่งที่ดี ตัวอย่าง ถ้ากลุ่มตัวอย่าง 1,000 แบ่งไปเป็นช่วงชั้นผ่านวิธี เดไซล์ จะได้ 10 ช่วงชั้น ประกอบด้วย case 1 – 100, 101 – 200, 201 – 300, ... 901 - 1,000 ถ้ากรอบมีขนาดตัวอย่างเท่ากัน (n = 100) แต่ถ้าใช้การทับซ้อนกัน 90% จะประกอบด้วย 1 -100, 11 – 110, 21 - 120, ... 901 - 1,000 เดิมมีช่วงชั้นเท่ากับ 10 จะประกอบด้วย 9 กรอบ ระหว่างแต่ละช่วงชั้นทั้ง 10 ผลที่ได้คือ 10 + 9 x 9 = 91 กรอบ Waller and Meehl (1998) ได้อธิบาย ความสัมพันธ์หลักระหว่าง

จำนวนกรอบ W, ขนาดกลุ่มตัวอย่าง N, ขนาดของกลุ่มย่อย n_w, และสัดส่วนของการทับซ้อนของ O ระหว่างกรอบที่ติดกัน ดังสมการ

$$W = \frac{\frac{N}{n_w} - O}{1 - O} \qquad \text{..... (5)}$$

สำหรับการกำหนดขนาดของกลุ่มย่อย มากกว่าจำนวนกรอบ การตีความมากกว่าแนวโน้มของกราฟ เพราะการใช้ของกรอบความทับซ้อนเป็นเทคนิคการปรับเรียบ การใช้ของกรอบสามารถเพิ่มจำนวนของจุดข้อมูลบนเส้นโค้ง นอกจากตัดทิ้งจำนวนของ Case ที่ใช้คำนวนแต่ละจุด เพราะจำนวนมากกว่าของกรอบที่สามารถยืดขนาดของกลุ่มย่อยของจำนวนที่น้อยกว่าของช่วงชั้น (เช่น $n_i = n_w = 100$ ที่แสดงในตัวอย่าง) จัดกระทำใหม่ในสมการ (5)

$$n_w = \frac{N}{W \times (1 - O) + O} \qquad \text{..... (6)}$$

ตัวอย่าง ในขณะที่ การวิเคราะห์ MAXCOV ของ N = 1,000 ใช้เดไซล์ ได้ $n_i = 100$ ส่วนการวิเคราะห์ MAXEIG ด้วย 20 กรอบ นั้นทับซ้อน 90% ได้ $n_w = 345$ ดังนั้น เส้นกราฟของ MAXEIG จะไม่มีเพียงสองจุดข้อมูลเหมือนเส้นกราฟของ MAXCOV แต่จะเกี่ยวข้องกับจำนวนที่มากกว่าของ Case ในการคำนวนของแต่ละเส้นกราฟ (n_w [345] > n_i [100]) ผลประโยชน์ที่ได้รับจากจำนวนที่มากกว่าของจุดบน เส้นกราฟของ MAXEIG อาจจะแยกออกจากขอบเขต หรือลบล้าง ถ้าความคาดเคลื่อนแบบสุ่มของ Eigenvalue มีมากกว่า ความแปรปรวนร่วม ความเป็นไปได้ในที่นี้ ซึ่งเป็นเพียงการนำไปใช้ในการวิเคราะห์ตัวบ่งชี้ผลมากกว่าสองตัว

6. Consistency Test

สิ่งสำคัญพื้นฐานของวิธีการวัดอนุกรมวิธานคือ การประเมินความสอดคล้องระหว่างผลลัพท์ที่ได้มาด้วยการวิเคราะห์จำนวนไม่ซ้ำซ้อนที่เป็นไปได้ ความเชื่อมั่นในการอ้างอิงของการรวบรวมโครงสร้าง การแบ่งแยกประเภท หรือ ความเป็นมิติ ที่เป็นความสอดคล้องของผลลัพท์ ในความหมายของการทดสอบความสอดคล้อง ใช้อ้างอิงในหลายวิธีภายในวิธีการวัดอนุกรมวิธาน และเราจะใช้อ้างอิงอย่างกว้างขวางในบางกระบวนการ ที่นักวิจัยยอมรับต่อการตรวจสอบความสอดคล้องของผลลัพท์จากการวิเคราะห์ความแตกต่างของการวัดอนุกรมวิธานสำหรับตัวอย่าง ซึ่ง Meehl (1992) ได้พัฒนาวิธีการ MAXCOV เป็นการทดสอบความสอดคล้องต่อ Complement ที่ใช้วิธีการอื่น หลังจากนั้นจึงกลายเป็นวิธีการหนึ่งที่ใช้ในการศึกษาการวัดอนุกรมวิธาน ซึ่งเป็นการจัดระเบียบของความเกี่ยวข้องกับการทดสอบความสอดคล้อง Meehl ได้แนะนำว่า ความแตกต่างระหว่างวิธีการวัดอนุกรมวิธานกับการทดสอบความสอดคล้องของการวัดอนุกรมวิธานซึ่งมีการเลือกแบบไม่มีกฎเกณฑ์สูงมาก และตอบสนองต่อการจัดกระทำ แม้ว่า

มีความร่วมกันในบางวิธีสืบสวนทางการวัดอนุกรมวิธานที่เป็นความสามารถของวิธีการวัดอนุกรมวิธานเชิงพหุ การทดสอบวิธีอื่น มีการพัฒนาอย่างต่อเนื่องที่จะประเมินความสอดคล้องของผลจากการวัดอนุกรมวิธาน มีการศึกษาความสัมพันธ์บางส่วนกับระบบวิทยาของการทดสอบความสอดคล้องให้เกิดประโยชน์สูงสุด และความเป็นจริงไม่มีงานวิจัยที่มีการตรวจสอบการเพิ่มเติมความตรงเพิ่มขึ้นโดยแต่ละการทดสอบนำไปอ้างอิงจากวิธีการวัดอนุกรมวิธาน นอกจากนั้น ในขณะที่บางวิธีการทดสอบความสอดคล้อง จะเสนอการใช้ประโยชน์จากการตรวจรายการของการสรุปโครงสร้าง วิธีการอื่นที่พบต่อการจำแนกประเภทอย่างง่ายระหว่างโครงสร้างการแบ่งแยกประเภท และ ความเป็นมิติ เทคนิควิธีการทดสอบความสอดคล้อง มี 3 วิธี คือ 1) วิธีวัดอนุกรมวิธานเชิงพหุใช้กับหลายเวลา ในหลายทาง 2) ประมาณค่าผลการประเมินพารามิเตอร์คุณลักษณะแฝง และการจำแนก Case สำหรับตรวจรายการความสอดคล้อง และ 3) ประเมินความสอดคล้องโมเดล การแบ่งแยกประเภท และความเป็นมิติ ต่อผลโดยข้อมูลจากการวิจัย (Meehl, 1995)

7. การประยุกต์ใช้วิธีการ Taxometric Analysis

ภาวะสันนิษฐาน เป็นคุณภาพ เชิงนามธรรม เชิงทฤษฎี หรือคุณลักษณะซึ่งแตกต่างกันในแต่ละบุคคล ดังที่ Gregory (2007, p. 131) ได้ยกตัวอย่างว่า ภาวะสันนิษฐาน เช่น ความรู้ความสามารถในภาวะผู้นำ ความซึมเศร้า การมีเจตนาร้าย เป็นต้น ซึ่ง ภาวะสันนิษฐาน จะอ้างอิงจากพฤติกรรมแต่จะมีมากกว่าพฤติกรรมที่แสดงออก โดยทั่วไป ภาวะสันนิษฐาน เป็นการสร้างทฤษฎีที่มีบางรูปแบบของการมีอยู่จริงที่เป็นอิสระของพฤติกรรม และแสดงออกอย่างหลวม ๆ แต่มีความสามารถชี้วัดบางขอบเขตที่มีอิทธิพลต่อพฤติกรรมมนุษย์ได้ จากการศึกษาเอกสารที่เกี่ยวข้องของภาวะสันนิษฐาน ที่เกิดจากผลของการสืบสวนทาง การวัดอนุกรมวิธาน นั้น อย่างไรก็ตาม การพัฒนาในอนาคตมีแนวโน้มที่จะหยุดยั้งเพียงการขยายความของ ภาวะสันนิษฐาน ที่ถูกศึกษาด้วยวิธีการเหล่านั้น แต่จากการประยุกต์ใช้ของวิธีการ การวัดอนุกรมวิธาน ซึ่งนำไปสู่คำถามการวิจัยแบบใหม่ จนกระทั่งถึงปัจจุบันนักวิจัยทาง การวัดอนุกรมวิธาน เน้นคำถามที่ว่า ภาวะสันนิษฐานเหล่านี้เป็นการแบ่งแยกประเภท หรือ ความเป็นมิติ ตามธรรมชาติหรือไม่ มันสามารถเป็นประโยชน์ต่อจุดเริ่มต้นทางความคิดสำหรับการสืบสวนคำถามการวิจัยที่ซับซ้อนเพิ่มขึ้นตามมา มันเป็นสิ่งสำคัญสำหรับการพัฒนา การวัดอนุกรมวิธาน ที่นักวิจัยเริ่มที่จะสำรวจคำถามที่ซับซ้อนเหล่านี้ การกระทำเช่นนี้ จะโต้แย้งข้อวิจารณ์ที่มีศักยภาพ ซึ่ง การวัดอนุกรมวิธาน เป็นระเบียบวิธีที่จำกัดที่แก้ปัญหาคำถามพื้นฐานเกี่ยวกับคุณลักษณะแฝง แต่ไม่บูรณาการกับการวิจัยรูปแบบอื่น (Ruscio, Haslam, & Ruscio, 2006)

7.1 โครงร่างเต็มรูปของคุณลักษณะแฝง (Full Delineation of Latent Structure)

คุณลักษณะแฝงของภาวะสันนิษฐาน อาจจะพิจารณาซับซ้อนมากกว่า การแบ่งแยกประเภท (Two Latent Classes) หรือ ความเป็นมิติ (n Latent Factors) มีคำแนะนำว่า การทดสอบเบื้องต้นของขอบเขต การแบ่งแยกประเภทด้วยการวิเคราะห์ การวัดอนุกรมวิธาน ควรจะนำเสนอขั้นตอนแรกของโปรแกรมการออกแบบวิจัย ที่ทำให้คุณลักษณะแฝงสมบูรณ์ขึ้นเท่านั้น แนวทางของโปรแกรมวิจัยนี้ และเทคนิคการวิเคราะห์ที่ถูกใช้เป็นการควบคุม สามารถแนะนำด้วยการค้นหาการวัดอนุกรมวิธาน ของโครงสร้างว่าเป็น การแบ่งแยกประเภท หรือ ความเป็นมิติ

ถ้าการวิเคราะห์ การวัดอนุกรมวิธาน แสดงให้เห็นว่า ภาวะสันนิษฐานเป็นความเป็นมิติ เป็นโครงสร้างที่สามารถอธิบายเพิ่มขึ้นโดยการใช้เครื่องมือที่พัฒนาสำหรับใช้กับภาวะสันนิษฐานที่มีค่าต่อเนื่อง นักวิจัยสามารถใช้การวิเคราะห์องค์ประกอบเชิงสำรวจและเชิงยืนยัน (Factor Analysis) กำหนดจำนวนของ ความเป็นมิติ ที่เป็นพื้นฐานของ ภาวะสันนิษฐานที่ระบุตัวบ่งชี้เป็นการประเมินความตรงมากที่สุดและทำให้ ความเป็นมิติ แตกต่างหรือค้นหาลำดับชั้น สูงหรือต่ำของความเป็นมิติ และสามารถใช้การวิเคราะห์ Latent Trait Analysis เช่น IRT เพื่อระบุตัวบ่งชี้ที่เหมาะสมสำหรับการประเมินทั่วไป หรือการวิจัยของภาวะสันนิษฐาน ตัวอย่างเช่น ความสามารถของตัวบ่งชี้คล้ายกับการวัดความเที่ยงผ่านพิสัยเต็มของระดับคุณลักษณะ กับตัวบ่งชี้ที่อาจจะเหมาะสมมากกว่าของงานการตัดสินใจ นั่นคือเป็นความแม่นยำของการวัดที่ใกล้เคียงกับระดับคุณลักษณะเฉพาะของ ภาวะสันนิษฐาน

อย่างไรก็ตาม ถ้าการวิเคราะห์แสดงให้เห็นว่า ภาวะสันนิษฐานเป็นการแบ่งแยกประเภท มีจำนวนของวิธีการ สำหรับการสืบสวนที่เพิ่มขึ้น บางวิธีหรือทั้งหมดที่ทำให้ลุล่วงในความหลากหลายของเทคนิควิธีวิเคราะห์ Complement ดังนี้ ความเป็นไปได้ประการแรก คือการเสนอต่อความเหมือนหรือความคล้ายในชุดของตัวบ่งชี้เพิ่มเข้าในการวิเคราะห์ การวัดอนุกรมวิธาน เพื่อทดสอบขอบเขตภายใน Taxon หรือ Complement ในขณะที่กระบวนการในการวัดอนุกรมวิธาน สามารถค้นหาเพียงขอบเขตเดียว ระหว่างสองกลุ่มที่เวลาหนึ่ง นี่ไม่ได้หมายความว่าไม่มีกลุ่มอื่นนอกจาก Taxon หรือ Complement ที่ปรากฏอยู่ ความป็นไปได้ประการที่สอง เสนอต่อชุดใหม่ของตัวบ่งชี้ที่เพิ่มในการวิเคราะห์ การวัดอนุกรมวิธาน เพื่อทดสอบสมมติฐาน Subtype ภายใน Taxon หรือ Complement ในขั้นตอนนี้ การวิเคราะห์เป็นการกระทำใน Subsample ประกอบด้วยสมาชิกของกลุ่มเดียวเท่านั้น การใช้ตัวบ่งชี้ ที่ถูกเชื่อว่าการแบ่งแยกสิ่งเดียว ที่เป็น Subtype จากสิ่งที่ไม่เป็น Subtype เพราะว่าชุดของตัวบ่งชี้เดี่ยว ๆ ดูเหมือนว่าไม่สามารถให้คำจำกัดความได้เท่ากับ Multiple Subtype ได้ คำแนะนำนักวิจัยควรทดสอบ Subtype เดียวที่เวลาหนึ่งต้องเลือกอย่างระมัดระวัง ตัวบ่งชี้ที่ไม่ซ้ำซ้อน และครอบคลุมข้อเท็จจริงทั้งหมดของ

Subtype ได้ ถึงแม้ว่าบางครั้งจะกระทำจากความรู้สึกเพื่อค้นหา Type ก่อนการค้นหา Subtype นั้น เป็นการทำจากบนลงล่าง อาจจะจำเป็นที่ต้องเริ่มศึกษา Subtype ก่อน ซึ่งเป็นวิธีกระทำจากล่างไปบน ถ้า Type เป็นอิสระจะไม่สามารถจำแนก Subtype ได้ หรือถ้าไม่มีความสัมพันธ์กันมาก จนกระทั่งชุดตัวบ่งชี้เดี่ยว ๆ ไม่สามารถจะทดสอบโครงสร้างของ Type ที่ล้มเหลว ประการที่สาม ในการสืบสวนที่อันหนึ่งอาจจะลุล่วงหลังจาก Taxon เป็นการค้นหา คือเพื่อระบุที่เกิดประสิทธิภาพสูง และตัวบ่งชี้มีความซ้ำซ้อนน้อย แม้ว่าการวิเคราะห์องค์ประกอบ และ IRT จะสนับสนุนวัตถุประสงค์นี้ เมื่อภาวะสันนิษฐานเป็นความเป็นมิติตัวบ่งชี้ของภาวะสันนิษฐานที่เป็นการแบ่งแยกประเภท เป็นมากกว่าการประเมินที่เหมาะสมโดยค่าพารามิเตอร์ ยกตัวอย่างเช่น การประมาณค่าความตรง (Validity) เกิดผลด้วยการวิเคราะห์ การวัดอนุกรมวิธาน หรือโดยการกระทำด้วย Latent Class Analysis หรือ Latent Profile Analysis วิธีการที่เกิดขึ้น ในประการที่สี่ของการสืบสวนเป็นการประเมินแต่ละ Taxon ที่แสดงเป็น Type และ Subtype สำหรับความหลากหลายของ ความเป็นมิติ ที่มีความหมาย นักวิจัยสามารถใช้วิธีการ เช่น การวิเคราะห์องค์ประกอบ หรือ ความสอดคล้องภายใน เพื่อตรวจสอบว่าภายในแต่ละสิ่งของ Taxon หรือ Complement มีความตรงแตกต่างกัน ในความเข้มข้นระหว่าง หนึ่งหรือมากกว่านั้น ของความเป็นมิติ ถ้าความตรงคงเหลือถูกค้นพบ นักวิจัยสามารถ ใช้ IRT Model และใช้การวิเคราะห์องค์ประกอบเชิงสำรวจ หรือเชิงยืนยัน หรือวิธีการอื่นที่เหมาะสมสำหรับการทำด้วยความเป็นมิติ เพื่อชี้แจง ความผันแปรภายในของคุณลักษณะแฝง (Guay, Ruscio, Hare, & Knight, 2007)

7.2 ความตรงเชิงโครงสร้างของคุณลักษณะแฝง (Construct Validation of Latent Structure)

อย่างไรก็ตามข้อมูลที่มีความสัมพันธ์อย่างง่าย หรือเป็น Model โครงสร้างที่ซับซ้อน จะเป็นประโยชน์ต่อการทำตามด้วยเพิ่มการสำรวจความตรงเชิงโครงสร้าง เป็นความสำคัญที่จะสันนิษฐานและทดสอบความเชื่อมต่อระหว่างคุณลักษณะแฝง การกระทำเช่นนี้พร้อม ๆ กัน จัดหาหลักฐานที่พอเพียงของ Model ของทฤษฎีที่ถูกใช้เพื่อสร้างการทำนายและการวัดด้วย Model ที่เป็นการประเมิน ในบริบทของการวิจัยเชิงโครงสร้าง ความตรงเชิงโครงสร้าง ควรเน้นการทำนายและอยู่บนพื้นฐาน Model โครงสร้าง เช่น การยืนยันของสมมติฐาน จะสนับสนุน Model ในขณะที่การโต้แย้งของสมมติฐานจะเรียก Model เป็นคำถาม เราโต้แย้งว่า Taxon ควรมีความสัมพันธ์ที่เหมาะสมเหนือกรอบเวลา ดังนั้นการตรวจสอบของค่าคงที่ประกอบด้วย หนึ่งองค์ประกอบของความตรงเชิงโครงสร้าง ในผลลัพท์ของ การแบ่งแยกประเภท เช่น ในการวิเคราะห์ทางพยาธิวิทยา การระบุ Taxon ที่ไม่เกี่ยวข้องกัน การจัดเข้ากลุ่ม สิ่งเดี่ยว ๆ ในฐานะสมาชิกของ Taxon

หรือ ความเป็นมิติ การใช้ข้อสอบ และการทำอัลกอรึทึมของการให้คะแนน สมาชิกของชั้น ไม่มีค่าคงที่สูงเหนือเวลา และถูกโต้แย้งว่า

สิ่งนี้ท้าทายความตรงเชิงโครงสร้างการวิเคราะห์ทางพยาธิวิทยา การระบุ Taxon ที่ไม่เกี่ยวข้องกัน

การตรวจสอบความคงที่ชั่วขณะนั้นไม่มีความจำเป็น ทางที่สำคัญที่สุด คือ การสำรวจความตรงเชิงโครงสร้างของการนำเสนอหน่วยอนุกรมวิธาน แต่มันจะแสดงให้เห็นการทดสอบที่มีศักยภาพ โดยเฉพาะ Model โครงสร้างที่ซับซ้อนมาก การเพิ่มขึ้นหรือชนิดที่แตกต่างของหลักฐานถูกต้องการ เหนือการแนะนำ ความสำคัญของการสาธิต ที่เป็นเหตุใน Model โครงสร้างหนึ่งต่อการทำความเข้าใจทฤษฎี ของเป้าหมาย ภาวะสันนิษฐาน เหมือนกับความสัมพันธ์ของ ภาวะสันนิษฐานสถานะปัจจุบันของทฤษฎีและการวิจัย ในความตรงเชิงโครงสร้างของ Model โครงสร้างไม่ได้จัดหารายละเอียดการแนะนำว่าจะรับผิดชอบงานนี้อย่างไร การยกประเด็นนี้มา เพื่อแนะนำว่าผู้สืบสวน ให้ความสนใจกับความตรงเชิงโครงสร้าง ในฐานะที่เป็นการไต่สวนการวัดทางจิตวิทยาในสายอื่น นักวิจัยในสาขาสังคมศาสตร์ และพฤติกรรมศาสตร์ ถูกฝึกในเรื่องความตรงเชิงโครงสร้างข่องเครื่องมือในการวัด แต่ความตรงของ Model โครงสร้าง จะต้องการความคิดสร้างสรรค์เพื่อคิดและทดสอบสมมติฐานซึ่งนำไปสู่ Model ทดสอบข้อสารสนเทศนั้น (Ruscio & Ruscio, 2004, pp. 403 - 447)

งานวิจัยที่ใช้การวัดอนุกรมวิธาน

คำถามการวิจัยที่เน้นการจัดเข้ากลุ่ม และอธิบายที่ว่า ภาวะสันนิษฐาน นั้นเป็นการแบ่งแยกประเภท หรือ ความเป็นมิติ ตามธรรมชาติหรือไม่ มีผู้ให้ความสนใจมากขึ้นในช่วงระยะเวลาที่ผ่านมา ดังเช่น Roisman, Fraley, and Belsky (2006) ได้ทำการศึกษาเรื่อง การศึกษาการวัดอนุกรมวิธานของการสัมภาษณ์เรื่อง ความรู้สึกผูกพันในผู้ใหญ่ การศึกษานี้เป็นการวิเคราะห์โครงสร้างองค์ประกอบคุณลักษณะแฝงของความแตกต่างระหว่างบุคคลที่เกิดจากผลสะท้อนในการสัมภาษณ์เรื่องความรู้สึกผูกพันในผู้ใหญ่ (Adult Attachment Interview; AAI) ซึ่งใช้กันทั่วไปและเป็นวิธีที่มีความตรงซึ่งออกแบบสำหรับประเมินสภาวะทางจิต ในขณะนั้นของผู้ใหญ่เกี่ยวกับประสบการณ์ในวัยเด็กกับพี่เลี้ยง โดยใช้วิธีวัดอนุกรมวิธานของ Meehl (1995) (MAXCOV - HITMAX) และข้อมูลจาก AAIs จำนวน 504 ราย ผลการวิเคราะห์แสดงให้เห็นว่าความแปรปรวนร่วมที่แฝงอยู่ในความรู้สึกมั่นคง (เมื่อเปรียบเทียบกับสภาวะทางจิตใจที่ไม่มั่นคง) มีความสอดคล้องกับโมเดลเชิง ความเป็นมิติ มากกว่าโมเดลเชิงกลุ่ม (การวิเคราะห์ด้วยการวัดอนุกรมวิธานของความกังวลที่มีอยู่ก่อนแสดงค่าไม่แน่นอน) นอกจากนี้ รายงานความแปรปรวนในผู้ใหญ่ที่มีความรู้สึกมั่นคง (n = 278) เกี่ยวกับประสบการณ์ในวัยเด็กชี้ให้เห็นข้อเท็จจริงมีเพียง

เล็กน้อยสำหรับกลุ่มตัวแปรเชิงกลุ่มระหว่างจิตใจมั่นคงที่ได้รับกับจิตใจมั่นคงแบบต่อเนื่อง (Qualitative Groups of Earned and Continuous - secures) แต่ ประสบการณ์ชีวิตของผู้ใหญ่ที่มีจิตใจมั่นคงดูเหมือนจะจำแนกแบบต่อเนื่อง ข้อค้นพบที่ได้จะอภิปรายในแง่ของการประยุกต์ใช้ทางทฤษฎีในเรื่องของปรากฏการณ์ของความมั่นคงที่ได้รับ (Earned - Security) ในลักษณะเฉพาะและความแปรปรวนร่วมที่แฝงอยู่ในสภาพจิตใจที่มั่นคงและไม่มั่นคงในลักษณะทั่ว ๆ ไป รวมทั้งวินิจฉัยผลที่ได้จากการวิเคราะห์ เพื่อการทดสอบความตรงและการจัดกลุ่ม AAI ในทำนองเดียวกัน Woodward, Lenzenweger, Kagan, Snidman and Arcus (2000) ศึกษาเรื่อง โครงสร้างเชิงอนุกรมวิธานของการแสดงปฏิกิริยาของทารก: ข้อค้นพบจากมุมมองการวัดอนุกรมวิธาน การวิจัยนี้ใช้วิธีการทางสถิติการวิเคราะห์ความแปรปรวนร่วมสูงสุด (Maximum Covariance Analysis หรือ MAXCOV) เพื่อทดสอบว่า จะพบโครงสร้างคุณลักษณะแฝงซึ่งสอดคล้องกับการทำนายเชิงทฤษฎีอยู่ในดัชนีชี้วัดของการแสดงปฏิกิริยาต่อสิ่งกระตุ้นหรือไม่ ในทารกวัย 4 เดือน จำนวน 599 ราย ผลการวิเคราะห์ MAXCOV แสดงข้อค้นพบที่ชัดเจนของคุณลักษณะแฝงความไม่ต่อเนื่องที่อยู่ในเกณฑ์วัดพฤติกรรมของการแสดงปฏิกิริยาของทารก ค่า Base Rates ของอันดับชั้นแฝง (หรือหน่วยอนุกรมวิธาน) คือร้อยละ 10 ทารกที่อยู่ในกลุ่มแสดงปฏิกิริยาสูงตามสมมุติฐานแสดงปฏิกิริยาสูงขึ้นที่ 4.5 ปี ตามเกณฑ์วัดการยับยั้งทางพฤติกรรม เมื่อเปรียบเทียบกับทารกที่ไม่ได้อยู่ในกลุ่มนี้ ผลการศึกษาแสดงข้อสนับสนุนเชิงประจักษ์ตามวัตถุประสงค์ของหลักการสำคัญในโมเดลเชิงทฤษฎีโดยสนับสนุนการวัดดอนุกรมวิธานของการแสดงปฏิกิริยาของทารก ในขณะที่ Walters (2008) ศึกษา คุณลักษณะแฝงของความผิดปกติจากการดื่มแอลกอฮอล์: การวิเคราะห์การวัดอนุกรมวิธานของข้อมูลการสัมภาษณ์เชิงโครงสร้างที่ได้จากผู้ต้องขังชาย วัตถุประสงค์ การศึกษาครั้งนี้ใช้ข้อมูลจากการสัมภาษณ์เชิงโครงสร้างผู้ต้องขังชาย จำนวน 1193 ราย ในการวิเคราะห์การวัดอนุกรมวิธาน เพื่อระบุคุณลักษณะแฝงของโครงสร้างองค์ประกอบความผิดปกติจากการดื่มสุรา วิธีการ: ทำการวิเคราะห์โดยใช้วิธีวัดอนุกรมวิธาน 3 แบบ คือ: ค่าเฉลี่ยส่วนต่างที่สูงกว่าและต่ำกว่าจุดตัด (Mean Above Minus Below a Cut หรือ MAMBAC) ค่าไอเกนสูงสุด (Maximum eigenvalue หรือ MAXEIG) และการวิเคราะห์องค์ประกอบคุณลักษณะแฝง (Latent Mode Factor Analysis หรือ L - Mode) ผลการศึกษาได้จากตัวชี้วัด 3 ตัว คือ: (1) เกณฑ์การพึ่งพาแอลกอฮอล์ DSM - IV Alcohol Dependence Criteria 1 และ 2 (ความอดทน/ การถอนตัว), (2) เกณฑ์การพึ่งพาแอลกอฮอล์ DSM - IV Alcohol Dependence Criteria 3 4 และ 5 และเกณฑ์การใช้แอลกอฮอล์ DSM - IV Alcohol Abuse Criterion 3 (สูญเสียการควบคุม I) และ (3) เกณฑ์การพึ่งพาแอลกอฮอล์ DSM - IV Alcohol Dependence Criteria 6 และ 7 และเกณฑ์การใช้แอลกอฮอล์ DSM - IV Alcohol Abuse Criteria 1 2 และ 4 (ผลลัพธ์ทางสังคม/ จิตในเชิงลบ)

ผลการศึกษาแสดงให้เห็นการสนับสนุนเชิงสอดคล้องกับการแปลความหมายหน่วยอนุกรมวิธาน (เชิงกลุ่ม - Categorical) ของความผิดปกติจากการดื่มแอลกอฮอล์ สรุป: อาจมีขอบเขตของหน่วยอนุกรมวิธานที่การแยกผู้ที่เข้าข่ายกับผู้ที่ไม่เข้าข่ายการวินิจฉัยว่าพึ่งพาหรือใช้แอลกอฮอล์ ซึ่งเป็นประโยชน์ต่อการวินิจฉัยและรักษา เช่นเดียวกับ Walters, Diamond, Magaletta, Geyer, and Duncan (2007) ทำการวิจัยเรื่อง การวิเคราะห์การวัดอนุกรมวิธานของมาตรวัดคุณลักษณะต่อต้านสังคม (ANT) ของแบบแสดงรายการประเมินบุคลิกภาพในกลุ่มผู้ต้องขัง การศึกษานี้นำมาตรวัดคุณลักษณะต่อต้านสังคมของแบบแสดงรายการประเมินบุคลิกภาพ (PAI) มาวิเคราะห์ด้วยวิธีวัดอนุกรมวิธาน ในกลุ่มผู้ต้องขัง 2,135 ราย โดยใช้คะแนนของมาตรวัดย่อย ANT ทั้ง 3 แบบ ซึ่งได้แก่ พฤติกรรมต่อต้านสังคม (Antisocial Behaviors - ANT - A), การให้ตัวเองเป็นศูนย์กลาง (Egocentricity - ANT - E), และการแสดงออกต่อสิ่งเร้า (Stimulus Seeking - ANT - S) เป็นตัวชี้วัดในการศึกษาครั้งนี้ และนำมาประเมินด้วยวิธีวัดอนุกรมวิธาน 3 แบบ คือ: ค่าเฉลี่ยส่วนต่างที่สูงและต่ำกว่าจุดตัด (Mean Above Minus Below a Cut - MAMBAC) ค่าไอเกนสูงสุด (Maximum Eigenvalue - MAXEIG) และการวิเคราะห์องค์ประกอบลักษณะแฝง (Latent Mode Factor Analysis - L - Mode) การประเมินแบบปรนัยและอัตนัยของผลการศึกษาแสดงให้เห็นการสนับสนุนเชิงสอดคล้องของการแปลความหมายเชิง ความเป็นมิติ ของโครงสร้างคุณลักษณะแฝงของวิธีวัดอนุกรมวิธานต่าง ๆ กัน รวมทั้ง เพศ เชื้อชาติ และระดับความมั่นคงต่าง ๆ กัน ปรากฏว่า ในลักษณะของโครงสร้างเชิงความเป็นมิติ นั้น ความผิดปกติด้านบุคลิกภาพต่อต้านสังคมเป็นการจัดเข้ากลุ่มผู้ตอบที่มีความแตกต่างกันในความเป็นมิติเชิงปริมาณ (ระดับการต่อต้านสังคม) 1 แบบขึ้นไป มากกว่าเป็นการจัดเข้ากลุ่มผู้ตอบเป็นกลุ่มที่ต่างกันตัวแปรเชิงกลุ่ม (ต่อต้านสังคมหรือไม่ต่อต้านสังคม) จะเห็นได้ว่าวิธีการ การวัดอนุกรมวิธาน s เป็นการตอบคำถามการวิจัยสำหรับการจัดเข้ากลุ่ม ระหว่าง การแบ่งแยกประเภท กับ ความเป็นมิติ ของ ภาวะสันนิษฐาน และนอกจากนี้ยังสามารถทดสอบความตรงเชิงโครงสร้างของเครื่องมือในการวัดภาวะสันนิษฐาน ดังที่ Franklin, Strong, and Greene (2002) ได้ศึกษาเกี่ยวกับ การวิเคราะห์การวัดอนุกรมวิธานของมาตรวัดภาวะซึมเศร้า MMPI - 2 ได้สรุปถึง มาตรวัด MMPI และ MMPI - 2 เป็นวิธีที่ใช้กันมานานสำหรับเป็นเกณฑ์วัดจิตพยาธิวิทยา ทั้งแพทย์และนักวิจัย ได้กล่าวถึง การแสดงอารมณ์ในทางลบที่พบได้ทั่วไปจากการใช้มาตรวัด MMPI และ MMPI - 2 ซึ่งอาจทำให้คะแนนมาตรวัดเพิ่มขึ้นและทำให้ลดความสามารถของแบบทดสอบในการจัดเข้ากลุ่มภาวะซึมเศร้าออกจากความผิดปกติทางการแพทย์อื่น ๆ การวิเคราะห์ด้วยวิธีวัดอนุกรมวิธานในการศึกษาครั้งนี้ พยายามทดสอบความตรงว่ามาตรวัดภาวะซึมเศร้า MMPI - 2 Depression Scales สามารถจำแนกผู้ป่วยที่มีอาการภาวะซึมเศร้าออกจากผู้ป่วยที่มีความผิดปกติอื่น ๆ ได้หรือไม่ โดยทำการวิเคราะห์กลุ่มตัวอย่าง

ขนาดใหญ่ที่มีความผิดปกติทางจิต (N = 2,000) แยกระหว่างชายกับหญิง การวิเคราะห์ด้วยวิธีวัดอนุกรมวิธานไม่พบจุดตัดของมาตรวัด MMPI - 2 ที่จำแนกผู้ป่วยที่มีอาการภาวะซึมเศร้าออกจากผู้ป่วยอื่น ๆ แต่ข้อค้นพบนี้สนับสนุนการศึกษาที่ผ่านมาซึ่งพบความเป็น ความเป็นมิติ แฝงของภาวะซึมเศร้าจากข้อค้นพบนี้ สามารถอภิปรายประโยชน์ของใช้มาตรวัด MMPI - 2 และการจำแนกภาวะซึมเศร้า

แนวคิดเกี่ยวกับเทคนิคการวิเคราะห์กลุ่ม

การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นกลุ่มของเทคนิควิธีวิเคราะห์ตัวแปรเพื่อจัดกลุ่มสิ่งต่าง ๆ เป็นกลุ่ม (Grouping) เช่น จัดนักเรียนออกเป็นกลุ่ม จัดองค์กรที่ทำกิจกรรมใด ๆ เหมือนกันเป็นกลุ่ม จัดสิ่งมีชีวิตออกเป็นหมวดหมู่ (Hair et al., 2010, pp. 508 - 510; Romesburg, 2004, p. 2; Fielding, 2007, p. 2; Norusis, & SPSS Inc., p. 361) การจัดหมวดหมู่ของสิ่งของมีจุดมุ่งหมายเพื่อจัดเข้าพวกหรือหาสิ่งที่เข้าพวกไม่ได้ (Outlier) และใช้เป็นแนวทางในการกำหนดสมมติฐานต่าง ๆ เกี่ยวกับความสัมพันธ์ระหว่างตัวแปร การจัดกลุ่มมีลักษณะต่างกันกับการจัดเข้ากลุ่ม (Classification) (Fielding, 2007, pp. 46 - 47) เช่น ในการวิเคราะห์สมการจำแนก (Discriminant Analysis) เป็นการจัดเข้ากลุ่มที่ต้องทราบจำนวนกลุ่มเสียก่อน เมื่อทราบจำนวนกลุ่ม และสร้างสมการจำแนกกลุ่ม (Discriminant Function) จึงจะสามารถจัดเข้ากลุ่มที่เหมาะสมได้ ขณะที่การจัดกลุ่มหรือแบ่งกลุ่ม (Grouping หรือ Cluster Analysis) จะไม่ทราบจำนวนกลุ่มเลย มีแต่คุณลักษณะ ที่แตกต่างกันของสิ่งที่ต้องการจัดกลุ่มปะปนคละกันอยู่ แล้วจึงหาหนทางหรือวิธีการที่เหมาะสมเพื่อแบ่งสิ่งเหล่านั้นออกเป็นหมวดหมู่ โดยอาศัยคุณลักษณะของสิ่งที่ต้องการจัดกลุ่มเป็นข้อมูล ซึ่งจำนวนกลุ่มที่ปรากฏภายหลังจากการจัดกลุ่มอาจมีกี่กลุ่มก็ได้ เมื่อพิจารณาให้ดีจะพบว่า การวิเคราะห์กลุ่ม นั้นคือ เครื่องมือสำหรับเสาะหาจำนวนกลุ่มที่เหมาะสมให้แก่การวิเคราะห์สมการจำแนกนั่นเอง ทั้งนี้การจัดกลุ่มพบว่า เทคนิควิธีวิเคราะห์กลุ่ม ใช้สำหรับการจัดกลุ่มสิ่งที่ต้องการจัดกลุ่ม มากกว่าการจัดกลุ่มตัวแปร การจัดกลุ่มตัวแปรจะใช้เทคนิควิเคราะห์องค์ประกอบมากกว่า (Factor Analysis) ซึ่งพิจารณาจากความสัมพันธ์ (Correlation) ในขณะที่การวิเคราะห์กลุ่ม จัดกลุ่มจากการพิจารณาระยะห่าง (Distance) สิ่งที่ต้องการจัดกลุ่ม ที่อยู่ในกลุ่มเดียวกันจะมีลักษณะที่เหมือนหรือคล้ายกัน ส่วนที่อยู่ต่างกลุ่มกันจะมีลักษณะที่แตกต่างกัน ดังนั้น ในการพิจารณาเลือกคุณลักษณะหรือตัวแปรที่จะนำมาใช้ในการแบ่งกลุ่ม จึงมีความสำคัญการกำหนดว่าต้องอยู่เพียงกลุ่มเดียว ในการพิจารณาจัดกลุ่มเราจะต้องมีวิจารณาญาณกำกับไว้ด้วยเสมอว่าควรรู้ว่าจัดอย่างไรจึงเป็นการจัดกลุ่มที่ดี จัดอย่างใดเป็นการจัดกลุ่มที่ไม่ดี เมื่อรู้เช่นนั้นจำนวนกลุ่มก็จะน้อยลงคือมีจำนวนที่พอเหมาะ (มนตรี พิริยะกุล, 2545, หน้า 69 - 70) การพิจารณาการจัดกลุ่มตามวิธีการของ สุชาต ประสิทธิ์รัฐสินธุ์ (2548, หน้า 404) ให้พิจารณาว่า เมื่อแบ่งกลุ่มแล้ว

กลุ่มมีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่ การทดสอบความแตกต่าง โดยใช้เทคนิคการวิเคราะห์ความแปรปรวนทางเดียว (One - Way Analysis of Variance) ซึ่งวิเคราะห์ค่าเฉลี่ยของตัวแปรที่ใช้ในการจัดกลุ่มว่ามีค่าแตกต่างระหว่างกลุ่ม (Between Group Difference) อย่างมีนัยสำคัญหรือไม่ ถ้าไม่มีนัยสำคัญ แสดงว่าการจัดกลุ่มมากไปต้องลดจำนวนกลุ่มลง ถ้าลดลงแล้วกลุ่มที่แบ่งมีความแตกต่างกันอย่างมีนัยสำคัญก็แสดงว่าเป็นจำนวนกลุ่มที่เหมาะสม สำหรับวิธีการสำหรับจัดกลุ่ม มีหลายวิธีเช่น การวัดความคล้ายกัน (Similarity Measure) การจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) เป็นต้น ในที่นี้จะกล่าวเฉพาะวิธีการจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering) ซึ่งการจัดกลุ่มแบบลำดับขั้นมีแนวคิดหลักถือเอา Distance Matrix หรือ Correlation Matrix เป็นเกณฑ์ สามารถระทำได้ 2 แนวทาง คือ (มนตรี พิริยะกุล, 2545, หน้า 85)

1. จากวัตถุ n หน่วย ให้จัดวัตถุที่คล้ายคลึงกัน (Similar) ที่สุดคือ มีระยะห่างกันน้อยที่สุดเป็นพวกเดียวกัน โดยพิจารณาทีละคู่ แล้วพิจารณานำวัตถุอื่นมาเข้ากลุ่มให้ดำเนินการเช่นนี้ไปเรื่อย ๆ เท่าที่พอจะจัดเข้ากลุ่มเดียวกันได้ เรียกวิธีนี้ว่า Aggromerative Hierarchical Method ซึ่งในทางปฏิบัติสามารถเลือกวิธีจัดกลุ่มได้หลายวิธี รวมเรียกว่า Linkage Method เป็นวิธีที่ใช้เส้นตรงโยงวัตถุที่คล้ายคลึงกันเข้าเป็นพวกเดียวกัน ที่เรียกว่า Dendogram เช่น Single Linkage จัดกลุ่มโดยอาศัยระยะห่างที่ใกล้ที่สุดเป็นเกณฑ์ Complete Linkage จัดกลุ่มโดยอาศัยระยะห่างที่ไกลที่สุดเป็นเกณฑ์ และ Average Linkage จัดกลุ่มโดยอาศัยระยะเฉลี่ยของทุกเส้นทาง (Fielding, 2007, pp. 59 - 65)

2. จากวัตถุ n หน่วย ให้แบ่งวัตถุเป็น 2 กลุ่มที่ต่างกัน จากนั้นในแต่ละกลุ่มให้แบ่งเป็น 2 กลุ่มที่ต่างกัน ดำเนินการต่อไปเรื่อย ๆ จนไม่สามารถแบ่งต่อได้ เรียกวิธีนี้ว่า Divisive Hierarchical Method (Fielding, 2007, pp. 65 - 67)

ในที่นี้จะเลือกกล่าวถึง วิธี Linkage Method เท่านั้น สำหรับวิธี Linkage Method เป็นการพิจารณาระยะห่างของวัตถุซึ่งทำได้ 3 วิธี คือ

1. Single Linkage เป็นวิธีการที่ใช้ระยะห่าง d_{ij} ใน Distance Matrix

โดยเริ่มต้นจัดวัตถุที่มีค่า d_{ij} ต่ำที่สุดเข้าเป็นกลุ่มเดียวกัน จากนั้นให้ปรับรูป Distance Matrix เดิมเป็นรูปใหม่ที่มีกลุ่มวัตถุที่ได้ ทำหน้าที่เป็นแถวและสดมภ์หนึ่งของเมตริกซ์ ซึ่งมีขั้นตอนดังต่อไปนี้

1.1 จากวัตถุ n หน่วย ให้ถือว่าหน่วยหนึ่ง ๆ ก็คือ กลุ่ม (Cluster) หนึ่งในชั้นนี้จึงมี Distance Matrix ซึ่งเป็น Symmetric Matrix ขนาด n x n (ถือว่า 1 หน่วยคือ 1 Cluster)

1.2 จากเมตริกซ์ $D = \left(d_{ij}\right)_{n\times n}$ ให้ตรวจดูว่าสมาชิกของ D ตัวใดมีค่าต่ำสุด สมมติว่า d_{UV} มีค่าต่ำสุด แสดงว่า Cluster U และ V คล้ายคลึงกันมากที่สุดสามารถรวมเป็นกลุ่มเดียวกันได้

1.3 ให้รวมกลุ่ม U และ V เป็นกลุ่มเดียวกัน เรียกว่ากลุ่ม (UV) แล้วปรับรูปเมตริกซ์ D ใหม่ ดังนี้

1.3.1 ตัดแถวและสดมภ์ที่สอดคล้องกับ d_{UV} และ d_{VU} ทิ้ง ($d_{UV} = d_{VU}$ เพราะเป็นระยะห่าง และ $D = D'$)

1.3.2 ผนวกแถวและสดมภ์ใหม่ที่สมาชิกของแถวและสดมภ์นี้คือระยะห่างที่สั้นที่สุดระหว่างกลุ่ม (UV) กับระยะห่างที่เหลือ เรียกแถวและสดมภ์ใหม่นี้ว่า แถว (UV) และสดมภ์ (UV)

1.4 ทำขั้นที่ 2 และ 3 เรื่อย ๆ ไป (มากที่สุด n - 1ครั้ง) แล้วบันทึกสมาชิกของกลุ่มต่าง ๆ

สำหรับกรณีที่ Distance Matrix มีขนาดใหญ่และมีระยะห่างต่ำสุดหลายค่าการวิเคราะห์จะซับซ้อนมากขึ้น หากค่าต่ำสุดที่เท่ากันอยู่ในแถวเดียวกันให้เลือกเพียงตัวเดียว มิเช่นนั้นจะเกิดปัญหาการตัดแถวและสดมภ์ที่สมนัยกับค่านั้น

2. Complete Linkage มีหลักการคล้ายกับ Single Linkage โดยจัดวัตถุที่คล้ายคลึงกันไว้เป็นกลุ่มเดียวกัน และวัตถุที่ต่างกันจะถูกแยกให้ห่างกัน ขั้นตอนการทำงานจึงเหมือนใน Single Linkage แต่ในขั้นที่ 3 ที่ Single Linkage หาระยะห่างระหว่างกลุ่ม (UV) กับกลุ่มอื่น สมมติว่าเป็น W โดยใช้

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\}$$

ใน Complete linkage ใช้เป็น

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\}$$

ทั้งนี้เพื่อให้กลุ่ม (UV) ถูกแยกห่างออกจากกลุ่มที่แตกต่างกับกลุ่ม (UV) ให้มากที่สุด

3. Average Linkage กระทำเช่นเดียวกับ Single Linkage และ Complete Linkage คือ เริ่มต้นด้วยการตรวจสอบเมตริกซ์ $D = (d_{ij})_{n\times n}$ เพื่อสำรวจค่า d_{ij} ที่น้อยที่สุด สมมติว่า d_{UV} มีค่าน้อยที่สุด ให้จัดวัตถุ U และ V เป็นกลุ่มเดียวกัน เรียกว่ากลุ่ม (UV) จากนั้นจึงหาระยะห่างระหว่างสมาชิกของกลุ่ม (UV) กับกลุ่มที่เหลือคือ W แล้วแทนที่จะหาระยะห่างที่ใกล้ที่สุดระหว่างสมาชิกของกลุ่ม (UV) กับ W ตามวิธี Single Linkage หรือระยะห่างที่สุดระหว่างสมาชิกของกลุ่ม (UV) กับ W กลับใช้ระยะห่างเฉลี่ยของทุกระยะห่างระหว่างสมาชิกในกลุ่ม (UV) กับ W คือ

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{n_{(UV)} n_W}$$

เมื่อ d_{ik} คือระยะระหว่างวัตถุ i ในกลุ่ม (UV) กับวัตถุที่ k ในกลุ่ม W

$n_{(UV)}$ และ n_W คือจำนวนวัตถุในกลุ่ม (UV) กับในกลุ่ม W ตามลำดับ

กัลยา วานิชย์บัญชา (2548,หน้า288) ซึ่งได้กล่าวถึง การนำเทคนิคการวิเคราะห์กลุ่มไปใช้ในงานด้านต่าง ๆ จะพบว่า การเลือกตัวแปรที่นำมาใช้ในการแบ่งกลุ่ม มีความสำคัญมาก ถ้าผู้วิจัยเลือกตัวแปรที่ไม่ได้ทำให้วัตถุที่อยู่ต่างกลุ่มกันมีความแตกต่างกันแล้วจะทำให้ไม่สามารถแบ่งกลุ่มได้ถูกต้อง และ เมื่อแบ่งกลุ่มแล้วควรจะศึกษาลักษณะวัตถุที่อยู่ในกลุ่มเดียวกันเพื่อนำมาใช้วางแผนงานต่อไป

ความเกี่ยวข้องของการวิเคราะห์กลุ่มกับการวัดอนุกรมวิธาน

การวัดอนุกรมวิธาน หรือ อนุกรมวิธานเชิงตัวเลข เป็นสาขาวิชาที่ใช้วิธีการเชิงจำนวนในการจัดเข้ากลุ่ม วัตถุ เช่น ชีววิทยาใช้จำแนกรูปร่างของสิ่งมีชีวิต ทางการแพทย์ใช้จำแนกอาการของโรคจากคนไข้ ทางธรณีวิทยาใช้จำแนกหินและแร่ เป็นต้น นอกจากนี้ การวิเคราะห์กลุ่ม การวัดอนุกรมวิธาน หรือ อนุกรมวิธานเชิงตัวเลข เป็นการใช้วิเคราะห์ตัวแปรหลายตัวแปร เช่นเดียวกับ Factor Analysis, Multidimensional Scaling, Principal Component Analysis และ Principal (Shepard, 1980; Aspey & Blankenship, 1978 cited in Romesburg, 2004, pp. 30 - 31) ข้อดีเมตริกซ์จะวิเคราะห์ด้วยการวิเคราะห์ตัวแปรหลายตัว ได้หลากหลายวิธีสำหรับการวิเคราะห์อนุกรมวิธานเชิงตัวเลขก็เช่นเดียวกับวัตถุประสงค์การวิจัยอื่น ๆ การวิเคราะห์กลุ่มถูกใช้เป็นวิธีการอธิบายสำหรับการวัดความคล้ายของวัตถุในกลุ่มตัวอย่าง (Romesburg, 2004, pp. 30 - 31)

แนวคิดเกี่ยวกับเทคนิคการวิเคราะห์องค์ประกอบหลัก

การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) เป็นเทคนิคการลดจำนวนตัวแปร โดยสร้างชุดของตัวแปรใหม่ให้เป็นฟังชั่นเชิงเส้นของตัวแปรเดิม และยังคงคุณลักษณะ หรือข้อมูลของตัวแปรเดิม จำนวนตัวแปรใหม่จะต้องไม่เกินจำนวนตัวแปรเดิม (กัลยา วานิชย์บัญชา, 2548, หน้า 183 - 186; Jolliffe, 2010, pp. 1 - 5; Dunteman, 1989, pp. 15 - 17)

หลักการของการวิเคราะห์องค์ประกอบหลัก

เมื่อ $Comp_i$ แทนองค์ประกอบหลักที่ i (i = 1, 2, 3, …, p) การสร้าง $Comp_i$ มีขั้นตอนดังนี้

ขั้นที่ 1 การสร้าง $Comp_1$ เป็นตัวแปรใหม่ตัวแรก โดยให้ $Comp_i$ เป็นฟังชั่นเชิงเส้นตัวแปรเดิม p ตัว และจะต้องสกัดหรือดึงค่าความแปรปรวนจากตัวแปรเดิมทั้งหมดมาไว้ที่ $Comp_i$ ให้มากที่สุด โดยที่

$Comp_1 = w_{11}X_1 + w_{12}X_2 + ... + w_{1p}X_p$

หรือ

$Comp_1 = w_1'x$ ที่ทำให้ $Var(w_1'x)$ มีค่ามากที่สุด และ $w_1'w_1 = 1$

ขั้นที่ 2 การสร้าง $Comp_2$ จะเป็นฟังชั่นเชิงเส้นของตัวแปรเดิม p ตัว และสกัด หรือ ดึงค่าความแปรปรวนที่เหลือจาก $Comp_1$ นำมาไว้ที่ $Comp_2$ ให้มากที่สุดเท่าที่จะทำได้ และ $Comp_2$จะต้องไม่มีความสัมพันธ์กับ $Comp_1$ หรือ Orthogonal กับ $Comp_1$ โดยที่

$Comp_2 = w_{21}X_1 + w_{21}X_2 + ... + w_{2p}X_p$

หรือ

$Comp_2 = w_2'x$ ที่ทำให้ $Var(w_2'x)$ มีค่ามากที่สุด

เงื่อนไข $w_2'w_2 = 1, w_1'w_2 = 0$ และ $Cov(w_1'x, w_2'x) = 0$

ขั้นที่ k การสร้าง $Comp_k$ จะเป็นฟังชั่นเชิงเส้นของตัวแปรเดิม p ตัว และสกัด หรือ ดึงค่าความแปรปรวนที่เหลือจาก $Comp_1, Comp_2, ..., Comp_{k-1}$ มาไว้ที่ $Comp_k$ ให้มากที่สุด และ $Comp_k$ จะต้องไม่มีความสัมพันธ์กับ $Comp_1, Comp_2, ..., Comp_{k-1}$ โดยที่

$Comp_k = w_{k1}X_1 + w_{k1}X_2 + ... + w_{kp}X_p$

หรือ

$Comp_k = w_k'x$ ที่ทำให้ $Var(w_k'x)$ มีค่ามากที่สุด

เงื่อนไข $w_k'w_k = 1, w_k'w_k = 0; i \neq k$ และ $Cov(w_k'x, w_j'x) = 0$ สำหรับ j < k

ขั้นที่ p การสร้าง $Comp_p$ จะเป็นฟังชั่นเชิงเส้นของตัวแปรเดิม p ตัว จะมีความผันแปรของตัวแปรเดิมที่เหลือจาก $Comp_1, Comp_2, ..., Comp_{p-}$ และ จะต้องไม่มีความสัมพันธ์กับ $Comp_1, Comp_2, ..., Comp_{p-1}$ โดยที่

$Comp_p = w_{p1}X_1 + w_{p1}X_2 + ... + w_{pp}X_p$

หรือ

$Comp_p = w_p'x$ ที่ทำให้ $Var(w_p'x)$ มีค่ามากที่สุด

เงื่อนไข $w_p'w_p = 1$ และ $Cov(w_p'x, w_j'x) = 0$ สำหรับ j < p

ความผันแปรร่วมของตัวแปรเดิม p ตัว = ความผันแปรขององค์ประกอบหลัก p ตัว หรือ

$$Var(X_1) + Var(X_2) + ... + Var(X_p) = Var(Comp_1) + Var(Comp_2) + ... + Var(Comp_p)$$

ความสัมพันธ์ระหว่างตัวแปรเดิมกับองค์ประกอบหลัก

องค์ประกอบหลักกับตัวแปรเดิมจะมีความสัมพันธ์ โดยการวัดความสัมพันธ์ด้วยค่าสัมประสิทธิ์สหสัมพันธ์ เรียกว่าค่า Loading ค่า Loading จึงเป็นค่าที่แสดงถึงอิทธิพลของตัวแปรเดิมที่มีต่อการสร้างองค์ประกอบหลัก ถ้าค่า Loading ของตัวแปรเดิมมีค่ามาก (ใกล้ +1 หรือใกล้ -1)

แสดงว่าตัวแปรเดิมนั้นมีความสำคัญหรือมีส่วนร่วมในการสร้างค์ประกอบหลักมาก ในทำนองเดียวกันการจะอธิบายว่าความหมายขององค์ประกอบหลัก ให้พิจารณาจากค่า Loading ของตัวแปรเดิม ถ้าตัวแปรเดิมตัวใดมีค่ามากความหมายขององค์ประกอบหลักควรเป็นความหมายของตัวแปรนั้น กำหนดค่า มากกว่าหรือเท่ากับ 0.50 (กัลยา วานิชย์บัญชา, 2548, หน้า 191, 200)

การพิจารณาจำนวนองค์ประกอบหลัก

จำนวนองค์ประกอบหลักจะเท่ากับจำนวนตัวแปรเดิม คือ p ตัว แต่องค์ประกอบหลักตัวท้าย ๆ จะมีสัดส่วนความแปรปรวนต่ำ โดยเฉพาะอย่างยิ่งกรณีที่ตัวแปรเดิมมีความสัมพันธ์กันมาก จะทำให้มีองค์ประกอบหลักเพียงไม่กี่ตัวที่มีสัดส่นความแปรปรวนสูง แนวทางในการพิจารณาจำนวนองค์ประกอบหลัก มีดังนี้ 1) พิจารณาจากร้อยละความแปรปรวนสะสม ถ้าร้อยละความแปรปรวนสะสมขององค์ประกอบหลัก k ตัวแรก เป็นอย่างต่ำร้อยละ 80 ก็ควรใช้จำนวนองค์ประกอบหลัก เท่ากับ k โดยที่ $k < p$ 2) ใช้กราฟ Scree ในการพิจารณาจำนวนองค์ประกอบหลัก ที่เหมาะสม โดยการพล็อตค่าไอเก็น ถ้าองค์ประกอบหลักตัวที่ $k + 1$ มีค่าไอเก็นต่ำมากเมื่อเทียบกับตัวที่ k หรือค่าลดลงอย่างรวดเร็ว ก็ไม่ควรพิจารณาองค์ประกอบตัวที่ $k+1, \ldots, p$ หรือควรมีองค์ประกอบหลักที่ k ตัวเท่านั้น 3) ให้พิจารณาค่าไอเก็นหรือค่าแปรปรวนขององค์ประกอบหลักแต่ละตัว ถ้าค่าแปรปรวนขององค์ประกอบหลักตัวใดน้อยกว่าค่าแปรปรวนเฉลี่ยจะต้องตัดทิ้ง (Jolliffe, 2010, pp. 112 - 115)

แนวคิดเกี่ยวกับการใช้โปรแกรม R วิเคราะห์ข้อมูล

ความเป็นมาของโปรแกรม R

โปรแกรม R เป็นโปรแกรมประเภทให้เปล่า (Free Software) เป็นทั้งโปรแกรมสำเร็จรูปที่ใช้คำนวณทางสถิติ และสร้างกราฟรายงานผล และภาษาคอมพิวเตอร์ (R Language) ที่พัฒนามาจาก ภาษา S แต่ โปรแกรม S เป็นซอฟท์แวร์ที่สร้างขึ้นมาเพื่อการค้าและมีลิขสิทธิ์ โปรแกรม R ได้เริ่มพัฒนาในปี พ.ศ. 2538 โดย Robert Gentleman และ Ross Ihaka แห่งภาควิชาสถิติ มหาวิทยาลัยโอคแลนด์ นิวซีแลนด์ ซึ่งเป็นที่มาของชื่อ โปรแกรม R ปัจจุบันโปรแกรม R อยู่ภายใต้การดูแลของมูลนิธิที่ไม่แสวงหากำไรชื่อ R Foundation โปรแกรม R เป็นลักษณะของซอฟท์แวร์ประเภท Open Source ที่เปิดโอกาสให้ผู้ใช้สามารถพัฒนา และแลกเปลี่ยน เพิ่มเติมเข้าไปผนวกไว้ในโปแกรมได้ สามารถนำไปใช้งานได้อย่างหลากหลาย (ศิริชัย พงษ์วิชัย, 2552 ก, หน้า 267 -268; Lewis, 2010, pp. 1 - 2)

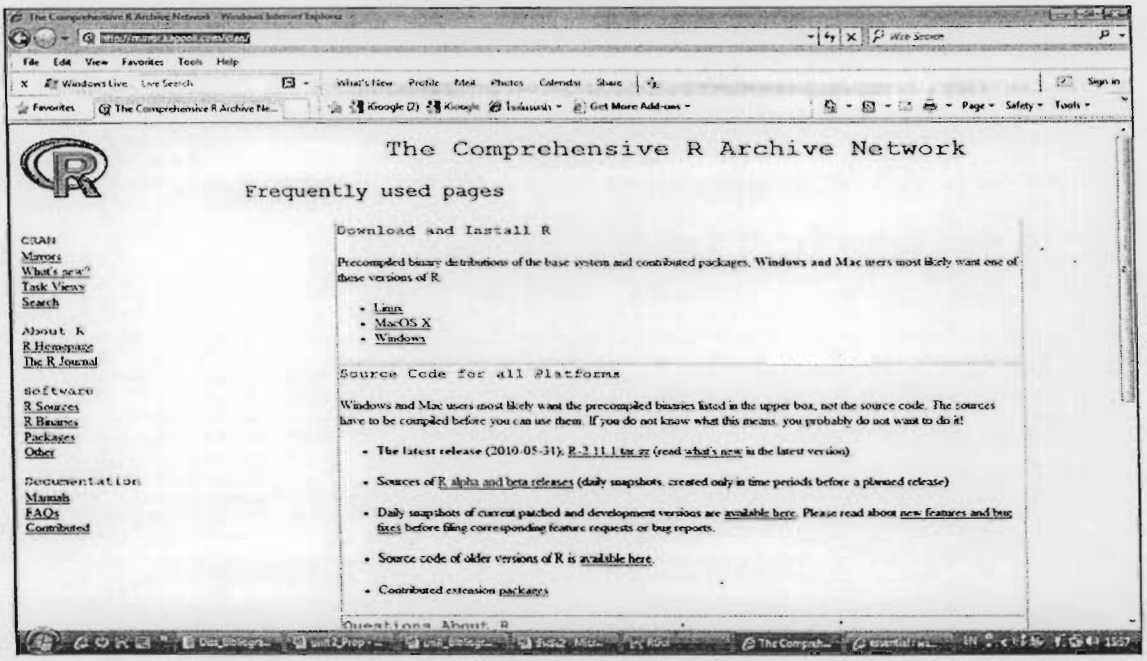
R ถูกพัฒนามาอย่างต่อเนื่องตั้งแต่กลางทศวรรตที่ 90 จนถึง พฤษภาคม 2010 เป็น R version 2.11.1 การพัฒนาจากกลุ่มบุคคลอิสระที่เรียกว่า R Core Team โดยทำงานพัฒนา R

เพื่อเป็นซอฟต์แวร์สาธารณะโดยไม่มีผลตอบแทนทางการค้า บุคคลเหล่านี้มีทั้งโปรแกรมเมอร์และนักสถิติที่ทำงานประจำอยู่ในสถาบันการศึกษาหรือสถาบันวิชาการที่มีชื่อเสียงต่าง ๆ (หัชชา ศรีปลั่ง, 2550, หน้า 3 - 4) โดยมีเว็ปไซต์อยู่ที่ http://www.r-project.org และได้รับการสนับสนุนทางการเงินจากมูลนิธิ R Foundation for Statistical Computing จนทำให้ R ถูกนำไปใช้อย่างกว้างขวางในการคำนวณทางสถิติและจุดเด่นอีกประการหนึ่ง คือความสามารถในด้านกราฟิกคุณภาพสูงในระดับตีพิมพ์ได้อย่าสวยงามด้วยคำสั่งที่ไม่ยุ่งยาก และสามารถใส่สัญลักษณ์ทางคณิตศาสตร์และสมการร่วมไปในกราฟได้ตามต้องการ แม้ว่าคำสั่งทางกราฟิกส่วนใหญ่ได้ตั้งค่า Default ไว้อย่างเหมาะสมแล้ว แต่ผู้ใช้ก็ยังมีอิสระที่จะเปลี่ยนแปลงได้ตามความต้องการ และจุดเด่นที่สำคัญอย่างยิ่ง เนื่องจาก R เป็นภาษาคอมพิวเตอร์ภาษาหนึ่ง จึงเปิดโอกาสให้ผู้ใช้สามารถสร้างคำสั่งหรือฟังชันใหม่ ๆ เพื่อทำงานได้ตรงตามความต้องการ และ R เปรียบเสมือนสำเนียงภาษาหนึ่งของภาษา S ชุดคำสั่งที่สร้างจากภาษา S จึงสามารถนำมาใช้กับ R ได้เกือบทั้งหมด นอกจากนี้ R ยังสามารถเชื่อมโยงกับภาษา C, C++, หรือ Fortrn ได้อีกด้วย ผู้ใช้สามารถเขียนคำสั่งมาจัดการวัตถุ (Object) ในสภาพแวดล้อม R (Enviroment) ได้ ปกติ R สามารถคำนวณพื้นฐานได้ โดยการทำงานของแพคเกจ พื้นฐาน (Faraway, 2004, pp. 217 - 219) แต่ที่พิเศษ คือสามารถเพิ่มความสามารถในการทำงานได้อย่างไม่จำกัดด้วยการติดตั้งแพคเกจเพิ่มเข้าไปให้กับระบบ หรือต้องการเรียกใช้สามารถดาวน์โหลดได้จากเว็บไซต์ของ CRAN (Collaborative R Archive Network) ซึ่งแพคเกจเหล่านั้นครอบคลุมการทำงานทางสถิติหลายสาขาได้รับการพัฒนา และเผยแพร่โดยนักวิชาการชั้นนำทั่วโลก จึงได้รับการตอบรับอย่างดีจากนักวิชาการทางสถิติจะเห็นได้จากมีผู้ร่วมผลิตแพคเกจและส่งให้เว็บไซต์ CRAN จำนวนมากและต่อเนื่องสม่ำเสมอ และนอกจากนี้ยังมีการจัดการประชุมทางวิชาการระดับโลก ที่เรียกว่า useR! 2004 และ useR! 2006 ที่ประเทศออสเตรีย (หัชชา ศรีปลั่ง, 2550, หน้า 5) และมีการจัดอย่างต่อเนื่องทุกปี ครั้งล่าสุด useR! 2010 จัดที่ National Institute of Standards and Technology (NIST), Gaithersburg, Maryland ประเทศสหรัฐอเมริกา ระหว่างวันที่ 20 - 23 กรกฎาคม 2553 สาระในการประชุมมุ่งเน้นให้ R เป็นภาษากลางของการวิเคราะห์ข้อมูลสถิติ เป็นเวทีแลกเปลี่ยนในการใช้ R ในด้านระเบียบวิธีทางสถิติ การวิเคราะห์ข้อมูล การแสดงผลและการนำไปใช้ และนำเสนอภาพรวมที่มีพัฒนาขึ้นจาก R นอกจากนี้ยังมีจัดฝึกอบรมระยะสั้น การบรรยาย การนำเสนอผลงาน เป็นต้น (R Development Core Team, 2010)

การเริ่มต้นโปรแกรม R

R ได้รับการพัฒนาสำหรับ Unix ใช้ได้กับทุกระบบปฏิบัติการ ได้แก่ Windows, Linux, Mac OS X เป็นต้น เมื่อเข้าไปในเว็บไซต์ CRAN จะสามารถดาวน์โหลดมาใช้ได้กับทุกระบบปฏิบัติการ การติดตั้ง R จึงทำได้ง่าย (Verzani, 2005, pp. 359 - 360; Lewis, 2010, pp. 2 - 4)

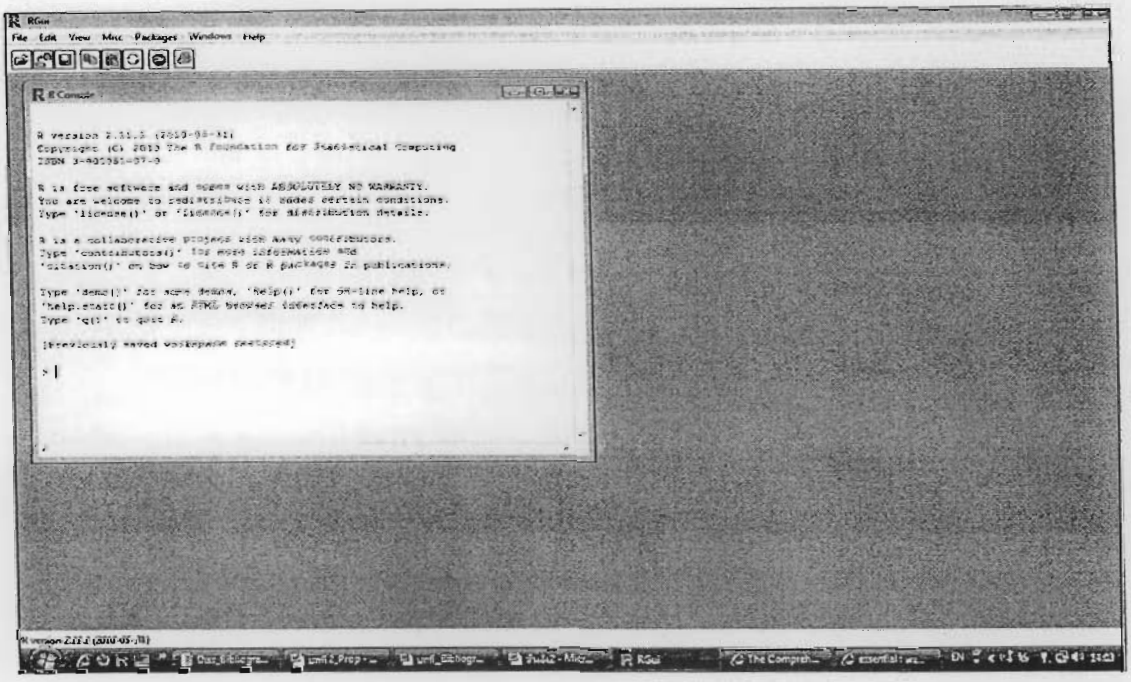
1. การติดตั้ง R สำหรับ Windows สามารถดาวน์โหลดได้ที่เว็บไซต์ CRAN ที่ http://mirror.kapook.com/cran/ จะปรากฏหน้าจอ ดังภาพที่ 7



ภาพที่ 7 เว็บไซต์ CRAN ที่ http://mirror.kapook.com/cran/

ให้เลือก Windows ซึ่งเมื่อเลือกเข้าไปแล้วเลือก Subdirectory base จะได้แฟ้มติดตั้งที่มีนามสกุลเป็น exe เช่น R-2.11.1 - win32.exe เมื่อคลิกติดตั้งจะได้โปรแกรมไปตามค่าที่ Default ไว้ โดยจะติดตั้งที่ ...\Program File\R\R-2.11.1 เมื่อติดตั้งเสร็จแล้วจะปรากฏ R ขึ้นบนรายการเมนู Start และมี R shortcut ขึ้นบน Desktop ด้วย

2. การเริ่มต้นใช้งาน R การเรียกใช้งานอาจเรียกจาก Desktop Shortcut หรือเรียกจาก Start >> All Programs >> R ก็ได้ ดังหน้าต่างตามภาพที่ 8



ภาพที่ 8 R ใน RGUI บนบทปฏิบัติการ Windows

ข้อความนำที่ปรากฏภายในหน้าต่าง R Console

R version 2.11.1 (2010-05-31)

Copyright (C) 2010 The R Foundation for Statistical Computing

ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

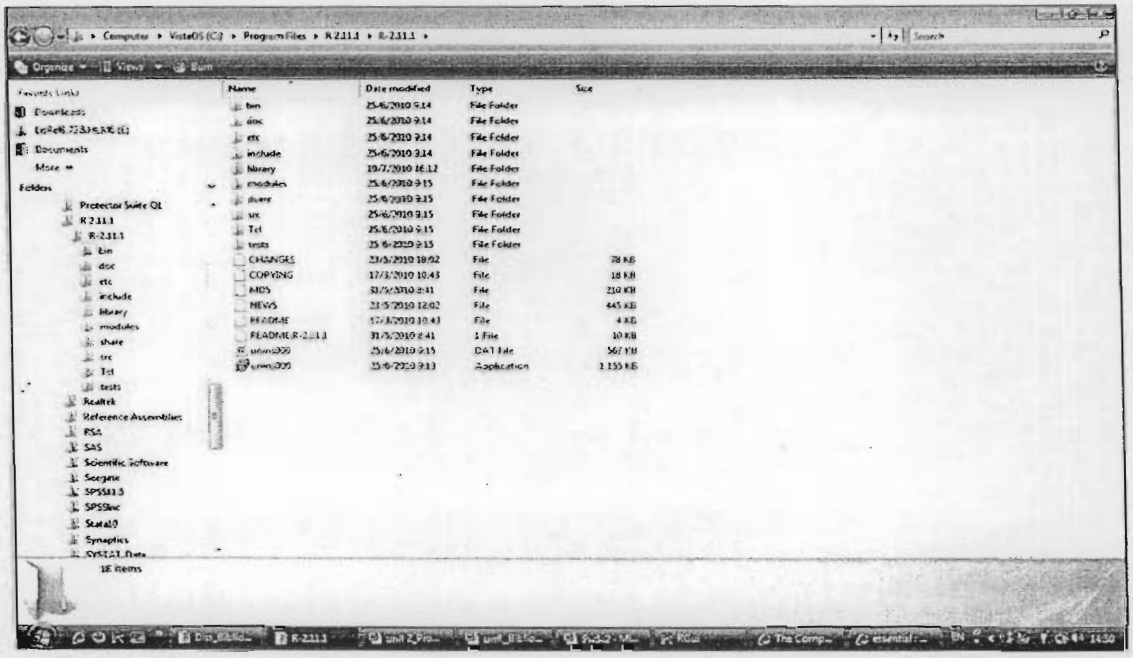
'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Previously saved workspace restored]

จะเห็นได้ว่าในการออกคำสั่งในสภาพแวดล้อม R นั้น จะต้องเขียนคำสั่งแล้วตามด้วยวงเล็บเสมอ ภายในวงเล็บอาจมี Argument หรือไม่ก็ได้ เช่น ต้องการความช่วยเหลือให้พิมพ์ Help ( ) หรือต้องการออกจากโปรแกรม ให้พิมพ์ q ( ) เป็นต้น และเครื่องหมาย > ที่ด้านล่าง คือ Prompt ของ R นั่นเอง เนื่องจากโดยปกติแล้ว R จะทำงานด้วยการพิมพ์บรรทัดคำสั่งหลัง Prompt สิ่งที่ต้องระวังคือ ตัวพิมพ์ใหญ่ หรือตัวพิมพ์เล็ก มีความแตกต่างกันในการเขียนคำสั่งและพิมพ์ข้อความในสภาพแวดล้อม R เหมือนการทำงานในสภาพแวดล้อม Unix ทั้งหลาย

3. การจัดระบบโฟลเดอร์และแฟ้มของโปรแกรม R บนระบบปฏิบัติการ Windows นั้น การติดตั้งโปรแกรมจะติดตั้งที่โฟลเดอร์ Program File โดยมีโฟลเดอร์หลักชื่อ R ภายในมีโฟลเดอร์ย่อย ชื่อ R-2.11.1 ประกอบด้วยโฟลเดอร์ย่อยหลายโฟลเดอร์ ดังหน้าต่างภาพที่ 9



ภาพที่ 9 หน้าต่างแสดงโฟลเดอร์ย่อย ในโฟลเดอร์ R-2.11.1

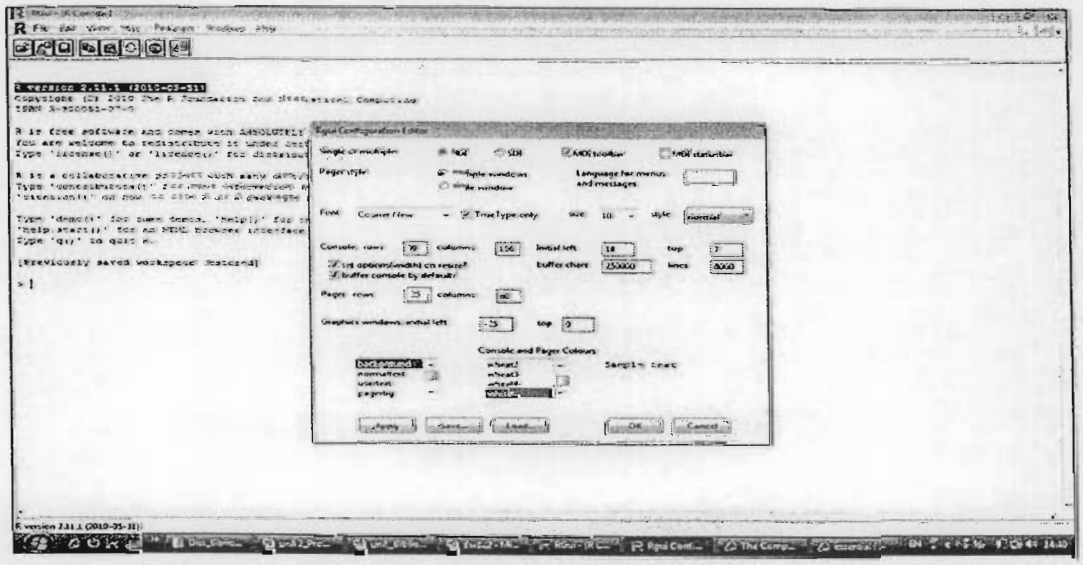
โฟลเดอร์ที่สำคัญ เช่น โฟลเดอร์ bin เป็นโฟลเดอร์เก็บตัวโปรแกรมหลักของ R โฟลเดอร์ doc เป็นโฟลเดอร์เก็บความช่วยเหลือ โฟลเดอร์ etc เก็บแฟ้มการตั้งค่าของสภาพแวดล้อม R โฟลเดอร์ library เก็บที่อยู่ของแพคเกจต่าง ๆ เป็นต้น

แฟ้มสำหรับการตั้งค่าเริ่มต้นเมื่อเริ่มเรียกใช้ R มีชื่อว่า Rprofile.site ในโฟลเดอร์ย่อย ..\R\R-2.11.1\etc\ ภายในประกอบด้วยข้อมูล ดังนี้

```
# Things you might want to change
# options(papersize="a4")
# options(editor="notepad")
# options(pager="internal")
```

```
# set the default help type
# options(help_type="text")
# .Library.site <- file.path(chartr("\\", "/", R.home()), "site-library")
```

เครื่องหมาย # ที่ปรากฏอยู่หน้าบรรทัด แสดงว่า R จะไม่ทำงานในบรรทัดคำสั่งเริ่มต้นเหล่านั้น หากต้องการให้ R ทำงานบรรทัดคำสั่งใด เมื่อเริ่มต้น R ให้ลบ # ที่หน้าบรรทัดคำสั่งนั้นออก (เฉพาะบรรทัดคำสั่งเท่านั้น สังเกตบรรทัดที่มีฟังชันและเครื่องหมายวงเล็บอยู่) เช่น หากต้องการตั้งค่ากระดาษให้เป็น A4 ทุกครั้งที่เรียกใช้ R ให้ลบเครื่องหมาย # หน้าบรรทัด options (papersize="a4") ออก (หัชชา ศรีปลั่ง, 2550, หน้า 5; Jones, Mallardet, & Robinson, 2009, pp. 3 - 5) โปรแกรมจะ Default อยู่แล้วโดยอัตโนมัติสำหรับเครื่องคอมพิวเตอร์ ที่ตั้งค่าสถานที่และเวลาประเทศไทย) หรือหากต้องการเรียกใช้ความช่วยเหลือเป็นแบบ Text ทุกครั้งที่เรียก (โปรแกรมจะ Default เป็นแบบ html สังเกตจะไม่ปรากฏเครื่องหมาย # อยู่หน้าบรรทัด Options (help_type="html"))ก็ให้ลบ # หน้าบรรทัด options(help_type="text") ออก นอกจากนี้หากต้องการให้ R ทำงานใดทุกครั้งที่เริ่มใช้ก็สามารถพิมพ์บรรทัดคำสั่งเพิ่มต่อท้ายได้ สำหรับแฟ้มการตั้งค่าของหน้าต่าง R บน Windows มีชื่อ Rconsole โดยอยู่ในโฟลเดอร์ My Document สามารถตั้งค่าผ่านเมนู Edite>> GUI preference ... บนหน้าต่าง Rconsole ดังภาพที่ 10



ภาพที่ 10 หน้าต่างเพื่อตั้งค่า Rgui บน Windows

4. การติดตั้งแพคเกจเพิ่มเติม เพื่อให้ R เพิ่มขีดความสามารถในการทำงานอย่างไม่มีขีดจำกัด (Sawitzki, 2009, pp. 54 - 57) ด้วยการติดตั้งแพคเกจที่ทำงานเฉพาะด้านเพิ่มเติม เช่น หากต้องการวิเคราะห์ข้อมูลด้วยการวัดอนุกรมวิธาน ก็ติดตั้งแพคเกจ TaxProg.R ต้องการวิเคราะห์

โมเดลสมการโครงสร้าง ติดตั้งแพคเกจ Sem การวิเคราะห์เชิงโอบล้อมข้อมูล แพคเกจ DEA หรือ หากต้องการเมนูที่เป็นกราฟิก อาจติดตั้งแพคเกจ Rcmdr หรือเมนูภาษาไทยติดตั้งแพคเกจ ICE เป็นต้น การติดตั้งแพคเกจอาจทำได้สองวิธี วิธีแรกติดตั้งโดยตรงจากเว็บไซต์ ของ CRAN กับวิธีที่สองติดตั้งจากเครื่องของผู้ใช้โดยตรง การติดตั้งวิธีแรก ต้องเข้าไปที่เว็บไซต์ของ CRAN และเลือก ตัวเลือก Package เพื่อหารายการแพคเกจที่ต้องการ เช่นต้องการหา Cluster เมื่อเลือกเข้าไปจะบอกรายละเอียดเกี่ยวกับแพคเกจ ดังตัวอย่าง

cluster: Cluster Analysis Extended Rousseeuw et al. Cluster Analysis, extended original from Peter Rousseeuw, Anja Struyf and Mia Hubert.

Version:	1.13.1
Priority:	recommended
Depends:	R ($\geq$ 2.9.0), stats, graphics, utils
Published:	2010-06-25
Author:	Martin Maechler, based on S original by Peter Rousseeuw, Anja.Struyf@uia.ua.ac.be and Mia.Hubert@uia.ua.ac.be, and initial R port by Kurt.Hornik@R-project.org
Maintainer:	Martin Maechler <maechler at stat.math.ethz.ch>
License:	GPL ($\geq$ 2)
Citation:	cluster citation info
In views:	Cluster, Environmetrics, Multivariate
CRAN checks:	cluster results

Downloads:

Package source:	cluster_1.13.1.tar.gz
MacOS X binary:	cluster_1.13.1.tgz
Windows binary:	cluster_1.13.1.zip
Reference manual:	cluster.pdf
News/ChangeLog:	ChangeLog
Old sources:	cluster archive

Reverse dependencies:

Reverse depends: ADaCGH, AdaptFit, FactoMineR, FunCluster, FunNet, GLDEX, GOSim, RPMM, SemiPar, StatDA, USPS, aqp, clValid, clustTool, clusterCons, clusterSim, clv, dpmixsim, fpc, gclus, geophys, hopach, hybridHclust, isopam, maptree, optpart, pARccs, pamr, qpcR, rEMM, rainbow, recommenderlab, sdcMicro, seriation, EBarrays, MLInterfaces, aCGH, flowStats, hopach, mdqc, nem, pamr

Reverse imports: Hmisc, clue, hopach, relations, sdcMicro, aCGH, adSplit, snapCGH

Reverse suggests: BiodiversityR, CORREP, Hmisc, R2HTML, RGraphics, TraMineR, agricolae, asbio, clue, clues, e1071, flexclust, hexbin, languageR, rrcov, vegan, BiocCaseStudies, CORREP, hexbin

Reverse enhances: clusterCons, sfsmisc

เมื่อเลือกได้แล้วให้กลับไปที่ R หรือเปิด R ขึ้นมาใช้งาน สำหรับบน Windows จะมีเมนู Package อยู่แล้ว เมื่อคลิกตัวเลือก Install Package (s) .. จะปรากฏรายการของ CRAN Mirror มาให้เลือกว่าจะดาวน์โหลดจาก mirror ที่ตั้งอยู่ในประเทศใด ถ้าเลือกประเทศไทย จะปรากฏ Mirror ดังนี้ Thailand

http://mirror.kapook.com/cran/ Kapook.com, Bangkok

http://mirrors.psu.ac.th/pub/cran/Prince of Songkla University, Hatyai

เมื่อเลือกได้แล้ว กดปุ่ม "OK" จะมีรายการแพคเกจให้เลือกต่อมา แล้วกดปุ่ม "OK" แพคเกจนั้นจะถูกดาวน์โหลดและติดตั้งเข้าสู่ R แม้ว่าแพคเกจจะถูกติดตั้งเข้าสู่โฟลเดอร์ library ของ R แล้วแต่แพคเกจนั้นจะยังไม่ถูกเรียกใช้งาน การเรียกเข้ามาในเพื่อใช้งานหน่วยความจำ ต้องทำอีกหนึ่งขั้นตอน คือ ออกคำสั่ง Library () แล้วใส่ชื่อของแพคเกจในวงเล็บ เช่น Library (Cluster) จากนั้น จึงจะเรียกใช้ฟังชันในแพคเกจได้ สำหรับการติดตั้งจากเครื่องคอมพิวเตอร์ อาจดาวน์โหลดจากที่ใดก็ได้ โดยต้องดาวน์โหลดจากแฟ้ม Windows Binary ที่มีนามสกุล Zip บน Windows ในรายการเมนู Package เมื่อเลือกแล้วจะมีตัวเลือก Install package (s) from local zip file... ซึ่งเมื่อเลือกจะมีหน้าต่าง Select File ปรากฏขึ้นให้เลือกแฟ้มแพคเกจที่มีนามสกุล zip เพื่อติดตั้ง เมื่อติดตั้งแล้วทุกครั้งที่เปิดใช้งาน R และต้องการใช้แพคเกจนั้น ให้สร้างคำสั่ง Library ( ) และใส่ชื่อของแพคเกจในวงเล็บ (Rizzo, 2008, pp. 18 - 19)

นอกจากแพคเกจแล้วบางครั้งอาจจำเป็นต้องเพิ่มโปรแกรมที่ใช้ร่วมกับ R เช่น โปรแกรมใช้เขียนสคริปต์ภาษา R หรือโปรแกรมสำหรับป้อนข้อมูลหรือสร้างแฟ้มข้อมูล ซึ่งโปรแกรมเหล่านี้จะเข้าใจฟังชัน คำสงวน และภาษาของ R ทำให้สะดวกต่อการใช้งาน

การทำงานและวัตถุพื้นฐานใน R

1. การคำนวณพื้นฐาน สามารถทำงานกับตัวเลขได้เหมือนเครื่องคิดเลข โดยพิมพ์คำสั่งหลัง Prompt แล้วกด Enter จะได้ผลลัพธ์ ดังตัวอย่าง

```
> 2+4
[1] 6
> 6^2
[1] 36
> (1-2)*4
> sqrt(2)          # the square root
[1] 1.414214
> sin(pi)          # the sin fuction
[1] 1.224606e-16     # this  is 0!
> exp(1)           # this is exp() = e^x
[1] 2.718282
> log(10)          # the log base e
[1] 2.302585
```

การเก็บค่าต่าง ๆ ไว้ใช้ภายหลัง

นอกจากโปรแกรม R จะทำหน้าที่เหมือนเครื่องคำนวณทั่ว ๆ ไปแล้ว โปรแกรม R ยังสามารถที่จะทำงานเหมือนภาษาคอมพิวเตอร์ทั่ว ๆ ไป ที่สามารถเก็บค่าต่าง ๆ ที่เกิดขึ้นไว้ใช้ภายหลัง (ปกติภาษาคอมพิวเตอร์ทั่ว ๆ ไปจะเรียกว่าตัวแปร) โดยกำหนดในลักษณะ ดังนี้ (ศิริชัย พงษ์วิชัย, 2552 ก, หน้า 287 -297)

ชื่อที่ผู้ใช้ตั้งขึ้นมา > นิพจน์ทางคณิตศาสตร์

หรือ

ชื่อที่ผู้ใช้ตั้งขึ้นมา = นิพจน์ทางคณิตศาสตร์

นิพจน์ทางคณิตศาสตร์ หมายถึง ประโยคทางคณิตศาสตร์ที่อาจจะประกอบไปด้วย ชื่อตัวแปรทั้งที่ผู้ใช้กำหนดค่ามาแล้วและชื่อที่โปรแกรมกำหนดไว้ให้ ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1 การกำหนดค่าที่เกิดจากการคำนวณ

>x = 5*7 <--------- คำสั่งกำหนดค่าผลลัพธ์ 5*7 ให้แก่ตัวแปร x

ค่า x จะถูกเก็บไว้ในโปรแกรม R แต่จะยังไม่แสดงผล ถ้าผู้ใช้ต้องการให้แสดงผลค่า x สามารถทำได้โดยการพิมพ์ชื่อตัวแปรหรือใช้คำสั่ง Print ดังนี้

>x <--------- คำสั่งให้แสดงค่าตัวแปร x

[1]<--------- แสดงค่าที่ตัวแปร x เก็บไว้ 5*7

>x^-8 <--------- คำสั่งให้แสดงค่าตัวแปร x^{-8} โดยไม่มีการเก็บผลลัพธ์ไว้

[1] 4.440743e-13 <--------- แสดงผลลัพธ์ xค่าที่ตัวแปร x^{-8}

ตัวอย่างที่ 2 การกำหนดค่าที่เกิดจากการฟังก์ชั่นในโปรแกรม R

>y=rnorm (9)<---------- กำหนดเลขสุ่ม 9 จำนวนที่มีการแจกแจงปกติ ให้แก่ตัวแปร y

>y <--------- คำสั่งให้แสดงค่าตัวแปร y

[1] 0.89354430	-0.41458481	-0.20926455
[4] -0.23448822	1.19834129	0.92363993
[7] 0.75461560	0.88909975	-0.41328879

การกำหนดค่าดังตัวอย่างในภาษาคอมพิวเตอร์เรียกว่า Assignments Statement

2. ข้อมูลแบบเวกเตอร์ (Vectors) ในโปรแกรม R

วกเตอร์ตัวเลข (Numeric Vecter) การคำนวณทางสถิติ ข้อมูลจะมีหลายค่าจากการวัดหลายครั้ง หลาย Case จะสร้างเป็นข้อมูลในลักษณะเวกเตอร์เป็นชุดของข้อมูลหลายค่าเรียงกัน ดังตัวอย่าง

> x<- c(42,57,12,39,1,3,4,7)

เป็นการสร้างเวกเตอร์ตัวเลข ด้วยคำสั่ง C() แล้วกำหนดให้ชื่อเป็น x โดยใช้เครื่องหมาย <- ความหมายเหมือนกับเครื่องหมาย "=" และยังใช้เครื่องหมาย -> ได้เช่นกันเพียงแต่กำหนดชื่อวัตถุไว้อีกด้าน ดังตัวอย่าง

> c(42,57,12,39,1,3,4,7) ->x

เมื่อกดแป้น Enter หน้าจอจะไม่แสดงผล เนื่องจากเวกเตอร์ที่สร้างขึ้นจะถูกเก็บไว้ในวัตถุชื่อ x แล้ว ในกรณีนี้เรียกว่าเวกเตอร์ x มีขนาด 8 หน่วย ถ้าเราเขียนคำสั่งบรรทัดถัดไปเพื่อหาค่าของ 1/x จะได้ผลลัพธ์ ดังนี้

```
> x<- c(42,57,12,39,1,3,4,7)
> 1/x
[1] 0.02380952 0.01754386 0.08333333 0.02564103 1.00000000 0.33333333 0.25000000
[8] 0.14285714
```

เวกเตอร์ตรรกะ (Logical vector) คือ ค่าที่เป็นไปได้ TRUE FALSE หรือ NA ค่า TRUE คือค่าที่เป็นจริง ค่า FALSE คือค่าที่เป็นเท็จ ส่วนค่า NA หมายถึงไม่มีค่า ซึ่งสามารถใช้ได้ทั้งในเวกเตอร์ตัวเลขได้ด้วย สัญลักษณ์ที่ใช้กับค่าตรรกะมีดังนี้ < (น้อยกว่า) <= (น้อยกว่าหรือเท่ากับ) > (มากกว่า) >= (มากกว่าหรือเท่ากับ) == (เท่ากับ) และ!= (ไม่เท่ากับ)

เวกเตอร์ตัวอักษร ค่าที่เป็นข้อความและเวกเตอร์ตัวอักษรใช้ในกรณี ระบุชื่อคน สิ่งของ ข้อความ และการเขียนกราฟ การสร้างข้อความโดยใส่ข้อความไว้ในเครื่องหมายคำพูด " หรือ ' หากไม่ใส่เครื่องหมาย R จะตีความเป็นชื่อของวัตถุในหน่วยความจำ ดังตัวอย่าง

```
> name1<-c("Donald","Edward","Marry")
> name2<-c(name1,"Robrt")
> name1
[1] "Donald" "Edward" "Marry"
> name2
[1] "Donald" "Edward" "Marry"  "Robrt"
```

3. ข้อมูลแบบแฟกเตอร์ (Factors) ในโปรแกรม R

แฟกเตอร์เป็นวัตถุชนิดเวกเตอร์แบบพิเศษที่ระบุการแบ่งกลุ่มของข้อมูลเป็นพวก ๆ เช่น เพศ เชื้อชาติ ศาสนา อาชีพ ซึ่งไม่มีลำดับ และยังใช้กับการแบ่งกลุ่มที่มีลำดับ เช่น ระดับการศึกษา เป็นประถมศึกษา มัธยมศึกษา อุดมศึกษา เป็นต้น การสร้างแฟคเตอร์ได้จาก เวกเตอร์ตัวอักษร หรือสร้างจากเวกเตอร์ตัวเลข ดังตัวอย่างต่อไปนี้

```
> sex<-c("male", "male","female","male","female","female")
> typeof(sex)
[1] "character"
> f.sex<- factor(sex)
> typeof(f.sex)
[1] "integer"
> is.factor(f.sex)
```

```
[1] TRUE
> f.sex
[1] male   male   female male   female female
Levels: female male
```

ในขั้นแรกเราสร้างเวกเตอร์ตัวอักษรระบุเพศของตัวอย่าง เมื่อตรวจสอบจะพบว่า R เก็บเวกเตอร์เป็นแบบ "character" จากนั้นหากเปลี่ยนเวกเตอร์ sex ให้เป็นแฟคเตอร์โดยใช้คำสั่ง factor() และนำผลไปเก็บในวัตถุชื่อ f.sex จากนั้นตรวจสอบวิธีที่ R เก็บข้อมูลจะเห็น ได้ว่าเปลี่ยนไปเก็บข้อมูลเป็นแบบตัวเลข แต่เมื่อเรียกดู f.sex พบว่า แสดงเป็นตัวอักษรไม่ใช่ตัวเลข แต่บอกเพิ่มเติมว่า Levels: female male ซึ่งแสดงว่า R เก็บค่า female เป็น 1 male เป็น 2 ตามลำดับตัวอักษร (ในกรณีที่ไม่ได้ระบุค่ารหัส R จะเก็บข้อมูลแฟคเตอร์โดยเรียงลำดับจากเลข 1, 2, 3, ... ไปเรื่อย ๆ จนหมดค่าที่เป็นไปได้ในเวกเตอร์นั้นโดยเรียงตามตัวอักษร) ในกรณีตัวอย่างดังกล่าว ในทางปฏิบัติในการเก็บข้อมูลมักจะระบุให้เพศชายเป็น 1 และเพศหญิงเป็น 2 เป็นรหัสสากลที่นิยมใช้กันทั่วโลก การที่ R ทำดังกล่าวจึงขัดกับความเข้าใจ ดังนั้นควรระมัดระวังในการกรอกข้อมูลเป็นตัวอักษร เพราะฉะนั้นควรสร้างวัตถุเป็นเวกเตอร์ตัวเลขก่อน โดย 1 ความหมายเป็นชาย 2 ความหมายเป็นหญิง แล้วจึงแปลงให้เป็นแฟกเตอร์และให้คำอธิบายค่า ดังตัวอย่าง

```
> sex<-c(1,1,2,1,2,2)
> typeof(sex)
[1] "double"
> f.sex<-factor(sex,label=c("male","female"))
> typeof(f.sex)
[1] "integer"
> f.sex
[1] male   male   female male   female female
Levels: male female
```

จะเห็นได้ว่าค่า 1 หมายถึง เพศชาย และค่า 2 หมายถึง เพศหญิง ตามที่ต้องการเนื่องจากในส่วนของ Levels คำว่า male มาก่อน female

4. ค่าว่าง (Missing Value) ในเวกเตอร์หนึ่งบางครั้งอาจมีค่าที่ไม่ทราบของบางตำแหน่งเกิดขึ้นได้ เช่น ข้อมูลจากแบบสอบถาม ผู้ตอบอาจไม่ตอบคำถามข้อใดข้อหนึ่งก็ได้ เมื่อเกิดค่าว่างโปรแกรมทางสถิติต่าง ๆ มีวิธีการใส่ค่าที่แตกต่างกัน แต่ห้ามใส่ 0 เพราะค่า 0 นั้นแสดงว่ามีค่า

ใน R จะใช้คำสงวน NA และสามารถใช้คำสั่ง is.na() เพื่อสร้างเวกเตอร์ตรรกะที่บอกตำแหน่งที่มีค่าเป็น NA ในเวกเตอร์ใด ๆ ได้ด้วย ดังตัวอย่าง

```
> z<-c(1:3,NA); ind<-is.na(z)
> ind
[1] FALSE FALSE FALSE  TRUE
```

ในบางครั้ง การหารค่าใด ๆ ด้วยค่า 0 หรือผลการคำนวณค่าใด ๆ กับ inf (ค่าอนันต์) ซึ่งจะได้ผลลัพธ์ที่หาค่าไม่ได้ ในกรณีนี้ R จะให้ค่า Na N ซึ่งหมายความว่า ตัวเลขที่หาค่าไม่ได้ (Not a number) ดังตัวอย่าง

```
> x<-0/0
> x
[1] NaN
> is.na(x)
[1] TRUE
> is.nan(x)
[1] TRUE
```

ซึ่งในกรณีนี้ เมื่อตรวจสอบด้วยฟังชัน is.na ก็จะพบว่า NaN ให้ผลลัพธ์เป็นจริงเช่นกัน นั่นคือ NaN เป็นค่าที่หาไม่ได้เช่นกัน แต่ไม่ได้เกิดจากการไม่ทราบค่า แต่เกิดจากการคำนวณที่ให้ผลลัพธ์หาค่าไม่ได้ สำหรับเวกเตอร์ตัวอักษรที่มีค่า NA อยู่ในกรณีที่พิมพ์ผลลัพธ์ของเวกเตอร์บนหน้าจอโดยไม่มีเครื่องหมายคำพูดแสดงข้อความ R จะแสดงผลลัพธ์ เป็น <NA> ทั้งนี้ เพื่อให้แตกต่าง จากข้อความว่า "NA" นั่นเอง

5. อาเรย์และเมตริกซ์ (Arrays and Metrix) การสร้างเวกเตอร์ที่ผ่านมา เป็นการสร้างเวกเตอร์มิติเดียว คือ การเรียงค่าไปตามลำดับ จะเห็นได้จาก R แสดงค่าในแนวบรรทัดเรียงต่อกันไป และสามารถระบุตำแหน่งในเวกเตอร์ได้โดยการระบุตัวเลขในมิติเดียว อาเรย์ ก็คือ เวกเตอร์ที่มีมิติหลายมิติ อาเรย์ที่มีสองมิติเรียกว่า เมตริกซ์ ดังตัวอย่าง

```
> x<-1:24
> x
 [1] 1 2 3 4 5 6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
> class(x)
[1] "integer"
> dim(x)<-c(6,4)
```

```
> x
     [,1] [,2] [,3] [,4]
[1,]    1    7   13   19
[2,]    2    8   14   20
[3,]    3    9   15   21
[4,]    4   10   16   22
[5,]    5   11   17   23
[6,]    6   12   18   24
> class(x)
[1] "matrix"
```

ในบรรทัดแรกเป็นการสร้างเวกเตอร์ตัวเลขที่มีค่าตั้งแต่ 1 ถึง 20 ขึ้นซึ่งตอนนี้ x จะมีคลาส (หรือวิธีเก็บข้อมูล) เป็น "integer" หรือเวกเตอร์ของเลขจำนวนเต็ม จากนั้นเราทำให้เวกเตอร์ x มีมิติเป็น 6 × 4 หรือมี 6 แถว และ 4 สดมภ์หรือคอลัมภ์ด้วยคำสั่ง dim( ) และใส่เวกเตอร์ระบุมิติและคลาสของ x ตอนนี้จะถูกเก็บข้อมูลแบบเมตริกซ์

ฟังชัน Transpose t( )สามารถทำได้ ดังนี้

```
> x<-
> x
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    2    3    4    5    6
[2,]    7    8    9   10   11   12
[3,]   13   14   15   16   17   18
[4,]   19   20   21   22   23   24
```

ฟังชัน cbind() เป็นคำสั่งให้ผนวกเวกเตอร์ที่มีขนาดเท่ากันเข้าด้วยกันโดยเชื่อม หรือเรียงลงมาตามแนวสดมภ์ สามารถทำได้ ดังนี้

```
> y<-1:10
> z<-log(1:10)
> m<-cbind(y,z)
> m
      y         z
 [1,] 1 0.0000000
```

```
[2,] 2 0.6931472
[3,] 3 1.0986123
[4,] 4 1.3862944
[5,] 5 1.6094379
[6,] 6 1.7917595
[7,] 7 1.9459101
[8,] 8 2.0794415
[9,] 9 2.1972246
[10,] 10 2.3025851
```

ด้วยวิธีนี้ จะมีความคล้ายคลึงกับชุดข้อมูลที่ใช้ทั่วไป และจะนำไปสู่วัตถุที่สำคัญอีกชนิดหนึ่ง คือ กรอบข้อมูล หรือ Data Frame นั่นเอง

6. ลิสต์และกรอบข้อมูล (List and Dataframe) ลิสต์เป็นวัตถุที่ประกอบด้วยวัตถุอื่น ๆ อยู่ภายใน ซึ่งอาจเป็นลิสต์เองก็ได้ โดยพื้นฐานแล้วลิสต์ถูกสร้างจากวัตถุพื้นฐาน เช่น เวกเตอร์ นั่นเอง ดังตัวอย่าง

```
> name<-c("Mary","Susan","Christina")
> y<-1:10
> z<-log(5:1)
> lst<-list(name,y,z)
> lst
[[1]]
[1] "Mary"     "Susan"    "Christina"
[[2]]
 [1]  1  2  3  4  5  6  7  8  9 10
[[3]]
[1] 1.6094379 1.3862944 1.0986123 0.6931472 0.0000000
```

จะเห็นได้ว่าสามารถผสมเวกเตอร์ที่มีขนาดไม่เท่ากันเข้าเป็นลิสต์ได้โดยใช้ฟังชัน list( ) หรือ as.list( ) จะมีประโยชน์มากเมื่อต้องการเขียนฟังชันใช้เอง เนื่องจากเป็นวัตถุที่สำคัญในการส่งข้อมูลออกจากฟังชัน สำหรับการผนวกเวกเตอร์เข้าด้วยกันทางด้านสดมภ์ สามารถเปลี่ยนให้เมตริกซ์ที่สร้างโดยวิธีนี้เป็นวัตถุชนิดกรอบข้อมูล โดยใช้ฟังชัน data.frame( ) เพื่อเชื่อมเวกเตอร์เข้ากันโดยตรง ดังตัวอย่าง

```
> x<-letters[1:10]
> y<-1:10
> z<-log(1:10)
> dat<-data.frame(x,y,z)
> dat
   x  y        z
1  a  1 0.0000000
2  b  2 0.6931472
3  c  3 1.0986123
4  d  4 1.3862944
5  e  5 1.6094379
6  f  6 1.7917595
7  g  7 1.9459101
8  h  8 2.0794415
9  i  9 2.1972246
10 j 10 2.3025851
```

วัตถุชนิดกรอบข้อมูลจะเป็นพื้นฐานในการทำงานคำนวณทางสถิติต่อไป

ในการอ้างอิงตำแหน่งในเมตริกซ์จะใช้เครื่องหมาย[]แต่มักใช้เครื่องหมาย $ และใช้ [ ] เพื่อระบุแถวในสดมภ์นั้น ๆ แทน ดังตัวอย่าง

```
> dat$x
 [1] a b c d e f g h i j
Levels: a b c d e f g h i j
> class(dat$x)
[1] "factor"
> dat$y
 [1]  1  2  3  4  5  6  7  8  9 10
> dat$z
 [1] 0.0000000 0.6931472 1.0986123 1.3862944 1.6094379 1.7917595 1.9459101
 [8] 2.0794415 2.1972246 2.3025851
> class(dat$y)
```

```
[1] "integer"
> class(dat$z)
[1] "numeric"
```

เมื่อเรียกชื่อสดมภ์เสมือนหนึ่งว่าเป็นเวกเตอร์หนึ่ง แล้วจึงระบุตำแหน่งของหน่วยย่อยในสดมภ์ด้วย เครื่องหมาย [ ]

อาจสรุปได้ว่า กรอบข้อมูลก้คล้ายกับเมตริกซ์นั่นเอง แต่จะมีความคล่องตัวมากกว่าตรงที่ R จะพยายามเก็บข้อมูลในรูปแบบตัวเลข ซึ่งสามารถนำมาคำนวณได้เลย และหากมีข้อมูลในสดมภ์ใด เป็นข้อความหรือตัวอักษร R จะ Default เป็นแฟกเตอร์ ซึ่งทำให้สามารถนำมาคำนวณทางสถิติได้ ยกเว้นระบุให้เก็บเป็นเวกเตอร์ตัวอักษร หรือในการนำเข้าข้อมูลโดยฟังชันอ่านแฟ้มข้อมูลบางชนิด Default เป็นเวกเตอร์ตัวอักษรแทนที่จะเป็นแฟกเตอร์ก็ได้

7. Attach และ Detach บางครั้งการทำงานกับกรอบข้อมูล เพียงกรอบข้อมูลเดียวที่ต้องการพิมพ์ชื่อกรอบข้อมูลทุกครั้งเมื่อเรียกชื่อเวกเตอร์หรือตัวแปรภายใน เป็นเรื่องยุ่งยาก จึงมีวิธีที่จะเลี่ยงการเรียกชื่อกรอบข้อมูลซ้ำ ๆ โดยใช้คำสั่ง attach() และ detach() ดังตัวอย่าง

```
> rm(x,y,z)
> attach(dat)
> x
 [1] a b c d e f g h i j
Levels: a b c d e f g h i j
> y
 [1]  1  2  3  4  5  6  7  8  9 10
> z
 [1] 0.0000000 0.6931472 1.0986123 1.3862944 1.6094379 1.7917595 1.9459101
 [8] 2.0794415 2.1972246 2.3025851
> table(x,y)
   y
    x  1 2 3 4 5 6 7 8 9 10
   a 1 0 0 0 0 0 0 0 0  0
   b 0 1 0 0 0 0 0 0 0  0
   c 0 0 1 0 0 0 0 0 0  0
   d 0 0 0 1 0 0 0 0 0  0
```

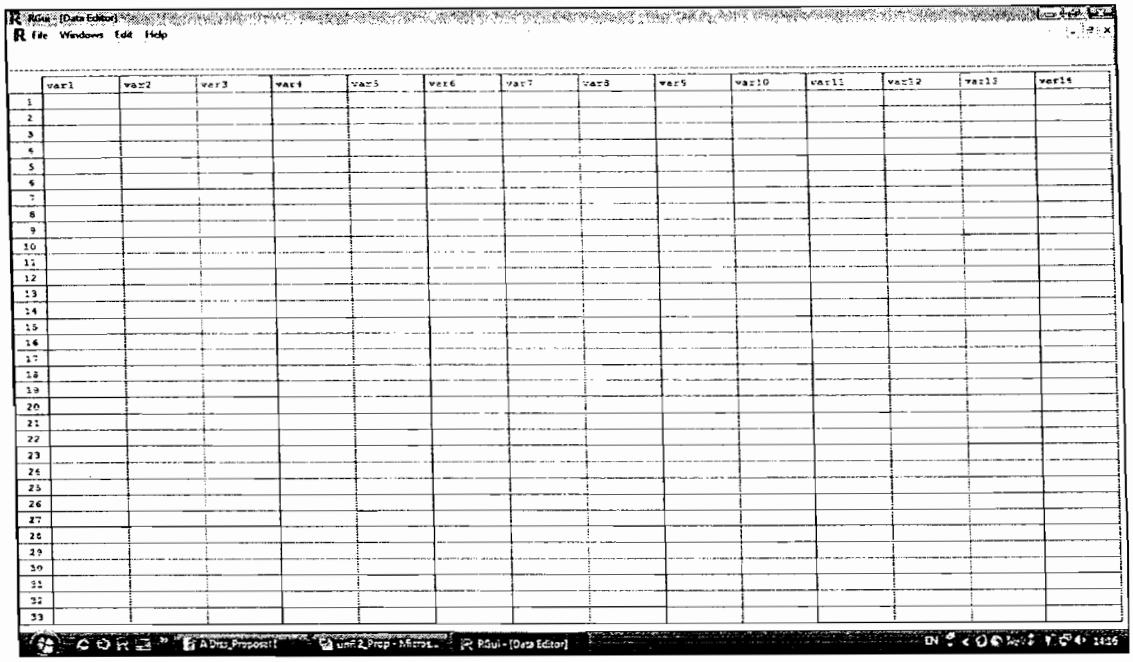
```
e 0 0 0 0 1 0 0 0 0  0
f 0 0 0 0 0 1 0 0 0  0
g 0 0 0 0 0 0 1 0 0  0
h 0 0 0 0 0 0 0 1 0  0
i 0 0 0 0 0 0 0 0 1  0
j 0 0 0 0 0 0 0 0 0  1
> detach (dat)
```

จะเห็นได้ว่า บรรทัดแรกสุดเป็นการใช้ฟังชัน rm( ) เพื่อลบวัตถุชื่อ x, y และ z ที่อยู่ในหน่วยความจำที่ได้สร้างไว้ก่อนหน้านี้ บรรทัดต่อมา attach(dat)เป็นการตรึงกรอบข้อมูล dat เข้ากับเส้นทางค้นหา เมื่อกรอบข้อมูล dat ได้ถูก attach แล้ว ต่อไปก็สามารถเรียกชื่อเวกเตอร์หรือตัวแปรภายใน dat ได้ โดยไม่ต้องเขียน dat$ นำหน้าอีก เมื่อทำงานกับกรอบข้อมูลนั้นเสร็จแล้ว จึงเลิกตรึงกรอบข้อมูลด้วยคำสั่ง detach(dat)

การนำข้อมูลเข้าและการบันทึกแฟ้มข้อมูลใน R

1. การกรอกข้อมูลโดยใช้ฟังชันภายใน R ในการใช้ R เพื่อการคำนวณทางสถิติที่มีข้อมูลจำนวนมาก มักไม่กรอกข้อมูลลงในโปรแกรม R โดยตรง แต่มักจะกรอกข้อมูลโดยใช้โปรแกรมสำหรับกรอกข้อมูล เช่น EpiData หรือโปรแกรมฐานข้อมูลอื่น ที่มีความสามารถในการตรวจสอบความถูกต้องของข้อมูลได้ดี หรืออาจเป็นโปรแกรม Spreadsheet เช่น Microsoft Exel Microsoft Access หรือ แฟ้มข้อมูลของโปรแกรมสถิติอื่น ๆ เช่น SPSS Stata SAS Minitab เป็นต้น ในการกรอกข้อมูลก็ได้ อย่างไรก็ตามสำหรับข้อมูลขนาดเล็ก R มีฟังชันในการสร้างกรอบข้อมูลในรูปแบบ Spreadsheet ให้ด้วย (Horgan, 2009, pp. 7 - 12) นั่นก็คือ ฟังชัน edit( ) ซึ่งในการออกคำสั่งนี้ จะต้องเขียนในรูปแบบ edit (data.frame( )) เพื่อบอก R ว่า กำลังจะสร้างกรอบข้อมูล และจะต้องส่งผลลัพธ์ให้กับวัตถุ ดังตัวอย่าง และภาพที่ 10

```
> dat<-edit(data.frame( ))
```



ภาพที่ 11 หน้าต่าง Data Editor ของคำสั่ง edit(data.frame( ))

สามารถกรอกข้อมูลในตารางนี้ได้ โดยที่แต่ละแถวในแนวนอนจะเป็นรายการของตัวอย่าง และแต่ละสดมภ์จะเป็นตัวแปร หรือเวกเตอร์ในภาษา R นั่นเอง นอกจากนี้ยังสามารถตั้งชื่อตัวแปรหรือเวกเตอร์ได้

2. การนำเข้าข้อมูลจากแฟ้มข้อมูลชนิดต่าง ๆ ข้อมูลที่สามารถนำมาใช้คำนวณทางสถิติในสภาพแวดล้อม R อาจสร้างขึ้นจากโปรแกรมต่าง ๆ ได้หลายโปรแกรม ดังที่ได้กล่าวมาแล้ว รูปแบบแฟ้มข้อมูลข้อความอย่างธรรมดาที่สุด คือ รูปแบบของ text ซึ่งมีรูปแบบที่แตกต่างกันไปที่สำคัญมี 4 รูปแบบ คือ แบบแรก Comma Separated Value (นามสกุล CSV) ซึ่งจะใช้เครื่องหมาย "," คั่นแยกระหว่างตัวแปร แต่ละสดมภ์ออกจากกัน สำหรับบางโปรแกรม ก็จะใช้เครื่องหมายคำพูด "..." ล้อมรอบด้วยเพื่อบอกขอบเขตของข้อมูลให้ชัดเจนยิ่งขึ้น แต่บางโปรแกรมจะใส่เครื่องหมายคำพูดเฉพาะเมื่อในข้อความนั้นมีเครื่องหมาย "," อยู่ด้วยซึ่งอาจทำให้สับสนว่าข้อความมีสองสดมภ์ได้ แบบที่สอง tab delimited (นามสกุล TXT หรือ TAB) ซึ่งแฟ้มนี้ จะใช้ เครื่องหมาย tab เพื่อแยกสดมภ์ออกจากกัน ดังนั้น ข้อควรระวังของรูปแบบนี้คือต้องไม่มีเครื่องหมาย tab อยู่ในข้อความที่เป็นเนื้อหา แบบที่สาม คือคั่นด้วยช่องว่าง ซึ่งอาจจะเป็นช่องว่างหนึ่งช่องหรือมากกว่าก็ได้ และรูปแบบสุดท้าย รูปแบบที่มีความกว้างของข้อความคงที่ สำหรับรูปแบบแฟ้มข้อมูลของโปรแกรมสถิติอื่น ๆ คำสั่งสำหรับอ่านแฟ้มที่มีรูปแบบพิเศษเหล่านี้ อยู่ใน Library ชื่อ foreign การเรียกใช้งาน เช่น ออกคำสั่ง read.dta( ) read.spss( ) read.epiinfo( ) มาใช้ได้เลย

3. การบันทึกแฟ้มข้อมูล โดยทั่วไปในงานวิเคราะห์ข้อมูล จะไม่ทำการบันทึกแฟ้มข้อมูลจากหน่วยความจำลงดิสก์ วิธีการคือ เตรียมข้อมูลให้ถูกต้อง จากนั้นเขียนแฟ้มสคริปต์เพื่อสร้างตัวแปรใหม่หรือปรับเปลี่ยนข้อมูลบางอย่าง วิคราะห์ข้อมูล แล้วบันทึกผลการวิเคราะห์เป็นขั้นสุดท้ายข้อดีของการทำงานลักษณะนี้ แฟ้มข้อมูลจะตรงกับแบบฟอร์มเก็บข้อมูล ไม่มีรหัสใดเปลี่ยนแปลงไป หรือไม่มีตัวแปรใดขาดหรือเพิ่ม และขั้นตอนต่าง ๆ ในการเปลี่ยนแปลงข้อมูลหรือสร้างตัวแปรใหม่อยู่ในแฟ้มข้อมูลให้ตรวจสอบในภายหลังได้ และหากผิดพลาดหรือต้องการเปลี่ยนแปลงการวิเคราะห์ ก็เพียงแก้ไขแฟ้มสคริปต์ แล้ว run กระบวนการทั้งหมดใหม่อีกครั้งก็จะได้ผลลัพธ์ใหม่ทันที

4. การบันทึกสภาพแวดล้อม R ลงแฟ้มและการเรียกกลับคืน สภาพแวดล้อมของ R ในที่นี้หมายถึง วัตถุต่าง ๆ และบรรทัดคำสั่งที่ให้ R ทำงาน คำสั่งในการบันทึกวัตถุต่าง ๆ ลงแฟ้มข้อมูลหรือคำสั่ง save( ) และ save.image( ) ทั้งสองคำสั่งนี้ต่างจากคำสั่ง write.table( ) และ write.dta( ) ตรงที่ save ( ) และ save.image( )สามารถบันทึกวัตถุชนิดต่าง ๆ ทุกชนิดได้ไม่จำเพาะต้องเป็นกรอบข้อมูล และสามารถบันทึกวัตถุจำนวนมากลงในแฟ้มเดียวกันได้ โดยคำสั่ง save.image( ) เป็นกรณีพิเศษที่บันทึกวัตถุทั้งหมดที่มีอยู่ในหน่วยความจำของ R ในขณะนั้นลงแฟ้ม แม้ว่าไม่มีข้อจำกัดในการระบุนามสกุล แต่เพื่อความจำเพาะและง่ายต่อการจดจำ นิยมให้แฟ้มบันทึกวัตถุของ R มีนามสกุล เป็น .Rdata เสมอ แต่ในบางกรณี ถ้าต้องการบันทึกบรรทัดคำสั่งต่าง ๆ ที่ทำไปแล้วเพื่อไว้ใช้ทำงานวิเคราะห์ข้อมูลต่าง ๆ ซ้ำอีก โปรแกรม R ได้สร้างคำสั่งไว้ให้ สำหรับการทำงานเช่นนี้ไว้ คือคำสั่ง savehistory ( ) ซึ่งนิยมกำหนดนามสกุลของแฟ้มเป็น Rhistory ดังตัวอย่าง

```
> rm(list=ls( ))
> setwd("c:/Rworkplace")
> x<-letters[1:10]
> y<-1:10
> z<-log(1:10)
> dat<-data.frame(x,y,z)
> save(x,y,z, file="test1.Rdata")
> save.image("test2.Rdata")
> savehistory("test.Rhistory")
```

คำสั่ง savehistory( ) จะบันทึกทุกคำสั่งบน R console แม้ว่าคำสั่งนั้นจะผิดพลาดก็ตาม

5. แฟ้มสคริปต์คำสั่ง R การทำแฟ้มสคริปต์คำสั่ง R แทนการพิมพ์คำสั่งบน R console สามารถพิมพ์และบันทึกไว้เป็นแฟ้มสคริปต์คำสั่ง R โดยใช้โปรแกรม Text editor เช่น Tinn-R , Crimson Editor, jEdit หรือโปรแกรมอื่น ๆ ก็ได้โดยไม่ต้องพิมพ์เครื่องหมาย prompt > นำหน้าบรรทัด ดังตัวอย่าง

```
rm(list=ls( ))
setwd("c:/Rworkplace")
x<-letters[1:10]
y<-1:10
z<-log(1:10)
dat<-data.frame(x,y,z)
```

จากนั้น บันทึกลงดิสก์ด้วยชื่อตามต้องการ ตามด้วยนามสกุล R เพื่อจะได้ทราบว่าเป็นแฟ้มสคริปต์คำสั่ง R เช่น ตั้งชื่อว่า test1.R และบันทึกในโฟลเดอร์ c:/Rworkplace แล้วเรียกคำสั่งในแฟ้มนี้ทำงานใน R โดยใช้คำสั่ง source ( ) โดยพิมพ์บนหน้าจอ R console ดังตัวอย่าง

```
> setwd("c:/Rworkplace")
> source("test.R")
```

การใช้งาน ภาษา เครื่องหมายและฟังชันพื้นฐานใน R

ดังที่ได้กล่าวมาแล้วโปรแกรม R เป็นทั้งภาษาและโปรแกรมสำเร็จรูป และเนื่องจากโปรแกรม R ถูกออกแบบมาสำหรับใช้งานด้วยคำสั่งแบบบรรทัด (Command Line) ถึงแม้ว่าจะสามารถใช้โปรแกรม R ด้วยเมนูใน R Commander ได้ก็ตาม เพราะเมนูไม่สามารถใช้งานได้ทุกอย่าง แต่คำสั่งสามารถทำงานได้ทุกอย่าง ดังนั้นถ้าผู้ใช้ต้องการใช้งานโปรแกรม R ได้อย่างมีประสิทธิภาพควรจะต้องเรียนรู้วิธีการใช้โปรแกรม R ด้วยการเขียนคำสั่งซึ่งมีกฎเกณฑ์เฉพาะของโปรแกรม R ซึ่งจะเรียกว่า ภาษา R (Rizzo, 2008, pp. 8 - 18)

1. การเขียนภาษา R ภาษา R ที่ใช้ติดต่อกับโปรแกรม R เพื่อใช้งานนั้น ผู้ใช้สามารถพิมพ์จากวินโดวส์ของโปรแกรม R หรือจากโปรแกรมอื่นๆ ที่มีลักษณะเป็นโปรแกรมประเภท Editor ได้ แต่ในที่นี้จะกล่าวถึงเฉพาะพิมพ์จากวินโดวส์ของโปรแกรม R ซึ่งสามารถพิมพ์ได้ 3 ลักษณะ ดังนี้

พิมพ์ที่วินโดวส์ R Console

ซึ่งจะต้องพิมพ์หลังเครื่องหมาย Command Prompt > โปรแกรม R จะทำงานแปลคำสั่งนั้นทันทีที่ผู้ใช้พิมพ์คำสั่งนั้นเสร็จสิ้น และถ้าถูกต้องจะได้ผลลัพธ์ปรากฏที่วินโดวส์ R Console แต่ถ้าผิดโปรแกรมจะแสดงข้อความอธิบายให้ทราบ การพิมพ์ที่วินโดวส์นี้เหมาะกับการใช้งานที่ต้องการเห็นผลทีละคำสั่ง

พิมพ์ที่วินโดวส์ R Editor (เรียกจากเมนู File/New Script...ของวินโดวส์ R Gui)

ซึ่งเป็นโปรแกรมที่ทำหน้าที่เหมือนกับโปรแกรม Editor ทั่ว ๆ ไป การพิมพ์คำสั่งที่วินโดวส์นี้โปรแกรม R จะยังไม่ทำการแปลคำสั่งนั้นทันทีเมื่อพิมพ์คำสั่งเสร็จ เหมือนพิมพ์ในวินโดวส์ R Console แต่สามารถสั่งให้แปลคำสั่งด้วยเมนู Edit ที่วินโดวส์ Gui และเลือกเมนู Run ine or selection จึงจะได้ผลลัพธ์ปรากฏในวินโดวส์ R Console

พิมพ์ที่วินโดวส์ R Commander

จะต้องพิมพ์คำสั่งที่วินโดวส์ย่อย Script ที่อยู่ด้านบน ซึ่งจะยังไม่ทำการแปลคำสั่งนั้นทันทีเช่นเดียวกันกับพิมพ์ที่วินโดวส์R Editor แต่สามารถสั่งให้แปลคำสั่งด้วยการคลิกขวาที่วินโดวส์ย่อย Script นี้ และเลือกเมนู Submit จึงจะได้ผลลัพธ์ปรากฏในวินโดวส์ย่อยOutput

2. คำสั่งหรือฟังก์ชันเบื้องต้นในภาษา R คำสั่งของภาษา R ที่ผู้ใช้งานเบื้องต้นควรทราบมีดังนี้ (Trosset, 2009, pp. 437 - 443)

>help(ชื่อหัวข้อ) เป็นคำสั่งเพื่อขอคำอธิบายเกี่ยวกับหัวข้อนั้น ๆ หัวข้ออาจจะเป็นชื่อ Package หรือชื่อฟังก์ชั่นใน Package

>help( ) เป็นคำสั่งเพื่อขอดูรายละเอียดเกี่ยวกับโปรแกรม R ทั้งหมดซึ่งเป็นการขอดูผ่านทางอินเทอร์เน็ตที่เครื่องผู้ใช้จะต้องต่ออินเทอร์เน็ตไว้ด้วย

>?? ชื่อหัวข้อ เป็นคำสั่งให้แสดงข่าวสารที่เกี่ยวข้องกับหัวข้อที่ระบุ

>library(ชื่อ package) เป็นคำสั่งเรียก package มาเพิ่มในโปรแกรม R ถ้าไม่ใส่ชื่อ package จะเป็นการแสดง package ทั้งหมดที่มีการติดตั้งไว้แล้ว

>search(ชื่อ package) เป็นคำสั่งให้แสดง package ที่ใช้งานอยู่ในขณะนั้น

>data( ) เป็นคำสั่งให้แสดงข้อมูลตัวอย่างที่มีอยู่ขณะนั้น

3. เครื่องหมายที่ใช้ในโปรแกรม R โปรแกรม R มีการใช้เครื่องหมาย (Operator) หลายชนิด เช่น เครื่องหมาย > ที่เรียกว่า R Prompt ซึ่งเป็นเครื่องหมายชนิดแรกที่ผู้ใช้ต้องพบ ในที่นี้จะกล่าวถึงเครื่องหมายพื้นฐานทีผู้ใช้ควรทราบ 3 ชนิด ดังต่อไปนี้

3.1 Assignment Operators เป็นเครื่องหมายที่ใช้เกี่ยวกับการกำหนดค่าซึ่งมี 3 เครื่องหมายดังต่อไปนี้

เครื่องหมาย	ความหมาย	ตัวอย่างการใช้งาน
<- หรือ =	ใช้กำหนดค่า ให้แก่ Object ด้านซ้าย	x <- 2499 หรือ x = 2499
->	ใช้กำหนดค่า ให้แก่ Object ด้านขวา	2499 -> x
<<-	ใช้กำหนดค่าที่เกี่ยวกับฟังก์ชั่น	

3.2 Arithmetic Operators เป็นเครื่องหมายที่ใช้ในการคำนวณทางคณิตศาสตร์ ซึ่งมีด้วยกัน 7 ชนิดดังนี้

เครื่องหมาย	ความหมาย	ตัวอย่างการใช้งาน	ผลลัพธ์
+	ใช้สำหรับการบวก	x = 4+3	7
–	ใช้สำหรับการลบ	x = 4–3	1
*	ใช้สำหรับการคูณ	x = 4*3	12
/	ใช้สำหรับการหาร	x = 4/3	1.333333
^ หรือ **	ใช้สำหรับการยกกำลัง	x = 4^3	64
%%	ใช้สำหรับการหาเศษที่เหลือจากการหาร	x = 4%%3	1
		x = 6%%3	0
%/%	ใช้สำหรับการปัดเศษขึ้นถ้ามีค่าเกิน 0.5	x = 4%/%3	1
		x = 5%/%2	2

3.3 Logical Operators เป็นเครื่องหมายที่ใช้ในการเปรียบเทียบ ซึ่งจะให้ผลเป็นจริง (TRUE) และเท็จ (FLASE)

เครื่องหมาย	ความหมาย	ตัวอย่างการใช้งาน	ผลลัพธ์
<	น้อยกว่า	5< 3	FLASE
<=	น้อยกว่าหรือเท่ากับ	3 <= 5	TRUE
>	มากกว่า	5 > 3	TRUE
>=	มากกว่าหรือเท่ากับ	3 > = 5	FLASE
= =	เท่ากับ	5 = = 5	TRUE
! =	ไม่เท่ากับ	7 ! = 8	TRUE
เครื่องหมาย	ความหมาย	ตัวอย่างการใช้งาน	ผลลัพธ์
! ชื่อ Object	ไม่ใช่ชื่อ Object	! x	-
l	หรือ	x l y	เป็นจริงตัวใดตัวหนึ่ง
&	และ	x & y	เป็นจริงทั้งคู่
is TRUE (x)	ใช้ทดสอบเมื่อ ค่า x เป็นจริง		

4. ฟังก์ชั่นพื้นฐานที่ใช้ในโปรแกรม R ในการสั่งให้โปรแกรม R ทำงานตามวัตถุประสงค์ที่ต้องการ จะต้องใช้คำพิเศษที่เรียกว่า ฟังก์ชั่น (Functions) ซึ่งโปรแกรม R ได้กำหนดฟังก์ชั่นให้ผู้ใช้ได้เลือกใช้ตามวัตถุประสงค์หลายฟังก์ชั่น นอกจากนี้ผู้ใช้ยังสามารถสร้างฟังก์ชั่นขึ้นมาใช้เอง แต่ในที่นี้จะกล่าวถึงฟังก์ชั่นพื้นฐานที่มีอยู่โปรแกรม R เฉพาะ Package เบื้องต้นซึ่งสามารถจำแนกได้ 5 กลุ่มใหญ่ ๆ ดังต่อไปนี้

4.1 Numeric Functions

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
abs (x)	หาค่าสัมบูรณ์ของค่า x คือ ค่าที่ไม่คิดเครื่องหมาย
exp (x)	หาค่า e ยกกำลังค่า x โดยที่ e มีค่าเท่ากับ 2.718282
sqrt (x)	หาค่ารากที่ 2 (root) ของค่า x
log (x) log10 (x)	หาค่า logarithm ฐาน e และฐาน 10 ของค่า x
sin cos tan (x)	หาค่า sin หรือ cos หรือ tan ของค่า x
min (x) max (x)	หาค่าต่ำสุดหรือสูงสุดของ Vector x (กลุ่มข้อมูล)
sum (x)	หาค่าต่ำสุดหรือสูงสุดของ Vector x (กลุ่มข้อมูล)
length (x)	หาค่าต่ำสุดหรือสูงสุดของ Vector x (กลุ่มข้อมูล)
ceiling (x)	หาจำนวนเต็มที่มากกว่าและใกล้ค่า x เช่น ceiling (2.3) คือ 3
floor(x)	หาจำนวนเต็มที่น้อยกว่าและใกล้ค่า x เช่น floor (2.3) คือ 2
trunc (x)	หาจำนวนเต็มที่ปัดเศษทิ้งของค่า x เช่น trunc (2.9) คือ 2
round (x,digits = n)	หาค่าที่ได้จากการปัดจุดทศนิยมให้เป็น n ตำแหน่ง ของค่า เช่น round (2.765189,3) จะได้ 2.76

4.2 Character Functions เป็นฟังก์ชั่นประเภทที่ใช้งานเกี่ยวกับตัวอักษร ซึ่งมีฟังก์ชั่นที่มีประโยชน์ต่อการใช้งานเบื้องต้น ดังต่อไปนี้

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
substr (x,ตำแหน่งเริ่มต้น, ตำแหน่งหลัง)	ใช้ดึงตัวอักษรตามตำแหน่งเริ่มต้นถึงตำแหน่งหลังของค่า x คือ เช่น x = "THAILAND" substr(x , 3 , 5) จะได้ผล คือ "AIL" ที่ไม่คิดเครื่องหมาย substr(x , 3 , 5) = "12345" นำค่า 3 ตัวแรกจาก "12345" ไปแทนที่ในตำแหน่งที่ 3 ถึง 5 ของ Object x เมื่อสั่งให้แสดงค่า x จะได้ผลดังนี้ "TH123AND"

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
sub (ค่าที่ต้องการเปลี่ยน,ค่าที่ใช้	ใช้ค้นหาข้อความและเปลี่ยนเป็นค่าใหม่
เปลี่ยน, x ignore . case = FALSE	เช่น x = "THAILAND" y = sub("AILA", "z", "x")
	ให้เปลี่ยนค่า I ที่อยู่ใน x ให้เป็น z
fixed = FALSE)	เมื่อสั่งให้แสดงค่า y จะได้ผลดังนี้ "THzND"
paste (… , sep = "")	ใช้คัดลอกตัวอักษรและตัดช่องว่างออกด้วยคำสั่ง sep เช่น
	a = paste("x" , 1:2 , sep = "") สร้างค่า x ตามลำดับ 4 ตัว เมื่อ
	สั่งให้แสดงค่า x จะได้ผลดังนี้ "x1" "x2" "x3" "x4"
strsplit (x)	ใช้แยกตัวอักษรที่อยู่ในข้อความ x
toupper (x)	ใช้เปลี่ยนตัวอักษรที่อยู่ในข้อความ x ให้เป็นตัวพิมพ์ใหญ่
tolower (x)	ใช้เปลี่ยนตัวอักษรที่อยู่ในข้อความ x ให้เป็นตัวพิมพ์เล็ก

4.3 Graphical Functions เป็นฟังก์ชั่นประเภทที่ใช้งานเกี่ยวกับการสร้างกราฟ ซึ่งมีฟังก์ชั่นมาตรฐานที่ควรรู้ ดังต่อไปนี้

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน	
plot ()	ใช้สร้างกราฟแบบ Scatterplot	ของ 2 ตัวแปร
hist ()	ใช้สร้างกราฟแบบ แบบแท่งต่อเนื่อง	ของ 1 ตัวแปร
barplot ()	ใช้สร้างกราฟแบบ แท่งแบบแยกแท่ง	ของ1 ตัวแปร
boxplot ()	ใช้ใช้สร้างกราฟแบ กล่อง	ของ 1 ตัวแปร
dotplot ()	ใช้สร้างกราฟแบบ จุด	ของ 1 ตัวแปร
piechart ()	ใช้สร้างกราฟแบบ วงกลม	ของ 1 ตัวแปร

5. Staistical Functions ฟังก์ชั่นที่ใช้สำหรับการวิเคราะห์ข้อมูล ด้วยวิธีการซึ่งมีอยู่มากมายหลายฟังก์ชั่น โดยสามารถจัดเป็นประเภทต่าง ๆ ได้ 3 ประเภท ดังนี้

5.1 Statistical Standard Functions

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
mean (x , trim = 0 , na . rm	ใช้หาค่าเฉลี่ยของ object x โดยไม่นำค่า Missing ออก ดังนั้นถ้า
= False , …)	ข้อมูลมี Missing โปรแกรมจะให้ผลลัพธ์เป็น NA ดังตัวอย่าง
	ต่อไปนี้
	x = c(2, 3, NA, 4) ; mean(x) จะได้ผลคือ NA mean(x, na.rm =
	TRUE) จะได้ผลคือ 3

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
sd (x, na.rm = False)	ใช้หาค่าส่วนเบี่ยงแบบมาตรฐาน ของ object x โดยถือว่า Object เป็นข้อมูล ตัวอย่าง n และไม่นำค่า Missing ออก
var (x, na.rm = FALSE)	ใช้หาค่าความแปรปรวน ของ object โดยถือว่า Object เป็นข้อมูล ตัวอย่าง n และไม่นำค่า Missing ออก
COV (x, na.rm = False)	ใช้หาค่าความแปรปรวนร่วม ของ object x และ y โดยถือว่า Object เป็นข้อมูล ตัวอย่าง n และไม่นำค่า Missing ออก
median (x, na.rm = False)	ใช้หาค่ามัธยฐานของ object x โดยไม่นำค่า Missing ออก
range (x, na.rm = FALSE)	ใช้หาค่าพิสัยของ object x โดยไม่นำค่า Missing ออก
quantile (x, probs = seq (0, 1,0.25) , na.rm = FALSE)	ใช้หาค่า quantile ของ object x โดยไม่นำค่า Missing ออก ตัวอย่างเช่น x = c(9,7,11,10,13,7) 0% 25% 50% 75% 100% quantile(y) จะได้ผลดังนี้ 7.00 7.50 9.50 10.50 13.00

5.2 Statistical Probability Functions เป็นฟังก์ชั่นที่ใช้เกี่ยวกับการแจกแจงความน่าจะเป็นแบบต่าง ๆ 4 ลักษณะ โดยมีรูปแบบทั่ว ๆ ไป ดังนี้

d? หาค่าความน่าจะเป็นของ Object x ด้วยการแจกแจงแบบ ?

p? หาค่าความน่าจะเป็นสะสมของการแจกแจงแบบ ? ตั้งแต่ค่าต่ำสุดถึงค่า q

q? หาค่าสถิติของการแจกแจงแบบ ? ที่ตรงกับค่าความน่าจะเป็นสะสม ?

r? สร้างเลขสุ่มที่มีการแจกแจงแบบ ? ขึ้นมา n จำนวน

? เป็นชื่อของการแจกแจง ซึ่งมีหลายชนิด ตัวอย่างการแจกแจงปกติ ดังต่อไปนี้

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
dnorm (x)	ใช้หาค่าเฉลี่ยของ object x โดยไม่นำค่า Missing ออก ดังนั้นถ้าข้อมูลมี Missing โปรแกรมจะให้ผลลัพธ์เป็น NA ดังตัวอย่างต่อไปนี้ x = c(2, 3, NA, 4) ; mean(x) จะได้ผลคือ NA mean(x, na.rm = TRUE) จะได้ผลคือ 3

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
pnorm (q)	ใช้หาค่าความน่าจะเป็นสะสมของแจกแจงแบบปกติ ตั้งแต่ค่าต่ำสุดถึงค่า q ที่ระบุ pnorm (1.65) จะได้ผลเป็น 0.9505285
qnorm (p)	ใช้หา Z ที่ตรงกับความน่าจะเป็นสะสม p ของการแจกแจงแบบปกติ เช่น qnorm (0.95) จะได้ผลเป็น 1.644854
rnorm (n)	ใช้สร้างเลขสุ่มที่มีการแจกแจงแบบปกติขึ้นมา n จำนวน เช่น rnorm (50) จะได้ผลเป็นเลขสุ่มที่มีการแจกแบบปกติ 50 จำนวน

5.3 Statistical Standard Method Functions เป็นฟังก์ชั่นที่ใช้เกี่ยวกับการวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติ ซึ่งสามารถจำแนกได้ 3 กลุ่มใหญ่ ๆ ดังนี้คือ

5.3.1 ฟังก์ชั่นสำหรับการวิเคราะห์ข้อมูลแบบต่อเนื่อง (Continuous Response) โดยส่วนใหญ่ฟังก์ชั่นในกลุ่มนี้จะใช้กับข้อมูลที่สามารถนำมาวิเคราะห์แบบ Parametric

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
T . test()	ใช้ทดสอบค่าเฉลี่ยสำหรับ 1 หรือ 2 กลุ่มตัวอย่างด้วยค่าสถิติ *t*
Oneway . test() หรือ aov()	ใช้ทดสอบค่าเฉลี่ยสำหรับ 1 หลายกลุ่มตัวอย่างด้วยค่าสถิติ *F* โดยการวิเคราะห์ความแปรปรวน ซึ่งสามารถใช้วิเคราะห์ได้ทั้งแบบทางเดียวและแบบหลายทาง
var . test()	ใช้ทดสอบความแปรปรวนสำหรับ 2 กลุ่มตัวอย่างด้วยค่าสถิติ *F*
bartlett . test()	ใช้ทดสอบความแปรปรวนหลายกกลุ่มตัวอย่างด้วยค่าสถิติ Bartlett
levene . test()	ใช้ทดสอบความแปรปรวนหลายกกลุ่มตัวอย่างด้วยค่าสถิติ Levene
cor . test()	ใช้ทดสอบความสัมพันธ์ด้วยค่าสถิติ *t* แปรปรวน

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
levene . test()	ใช้ทดสอบความแปรปรวนหลายกลุ่มตัวอย่างด้วยค่า Levene
lm()	ใช้ในการวิเคราะห์การถดถอย ซึ่งสามารถใช้ได้ทั้ง Simple Regression และ Multiple Regression พร้อมค่าสถิติต่าง ๆ ที่ใช้ในการทดสอบ ซึ่งจะมีฟังก์ชั่นที่สามารถใช้ในการวิเคราะห์ร่วมกันอีกหลายฟังก์ชั่น เช่น coef() fitted() residuals() dwtest() และอื่น ๆ

5.3.2 ฟังก์ชั่นสำหรับการวิเคราะห์ข้อมูล แบบไม่ต่อเนื่อง (Discrete Response) โดยส่วนใหญ่ฟังก์ชั่นในกลุ่มนี้จะใช้กับข้อมูลแบบแจงนับในรูปของสัดส่วนหรือร้อยละ

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
binom . test()	ใช้ทดสอบค่าสัดส่วน 1 กลุ่ม ด้วยค่าสถิติ Binomial และ Sign Test
prop . test() หรือ aov()	ใช้ทดสอบค่าสัดส่วน หลายกลุ่ม ด้วยค่าสถิติ Chi-Square
fisher . test()	ใช้ทดสอบค่าสัดส่วนสำหรับกลุ่มตัวย่อย ด้วยค่าสถิติ Fisher
chisq . test()	ใช้ทดสอบที่เกี่ยวกับค่าสถิติ Chi-Square เช่น ทดสอบเกี่ยวกับ Independence และ Goodness of fit

5.3.3 การวิเคราะห์ข้อมูล สำหรับข้อมูล แบบ นอนพาราเมตริก (Nonparametric) โดยส่วนใหญ่ฟังก์ชั่นในกลุ่มนี้จะใช้กับข้อมูลที่ไม่สามารถใช้การทดสอบแบบพาราเมตริกได้

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
Wilcox . test ()	ใช้ทดสอบสำหรับ 1/ 2 กลุ่มตัวอย่าง ด้วยค่าสถิติ Wilcoxon
kruskal . test ()	ใช้ทดสอบสำหรับหลายกลุ่มตัวอย่าง ด้วยค่าสถิติ Kruskal - Wallis
friedmah . test ()	ใช้ทดสอบสำหรับหลายกลุ่มตัวอย่างแบบ Two - Ways Analysis of Varinance ด้วยค่าสถิติ Firedman
cor . test (. . . . , method = “kendall”) หรือ “spearman”	ใช้ทดสอบที่เกี่ยวกับความสัมพันธ์แบบ Nonparametric ซึ่งใช้ฟังก์ชั่นเดียวกันกับการหาความสัมพันธ์แบบ Parametric ที่ใช้ค่าของ Pearson แต่ด้วยวิธีของ Nonparametric จะใช้วิธีของ Speaman และ Kendall แต่ใช้ฟังก์ชั่น corr . test () เหมือนกัน

6. Useful Functions ฟังก์ชั่นอื่นๆ ที่ช่วยประโยชน์ในการวิเคราะห์และจัดการกับข้อมูล ซึ่งมีอยู่หลายกลุ่ม ดังต่อไปนี้

6.1 File Managemint Functions

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
getwd ( )	ใช้ดูตำแหน่งที่โปรแกรม R ทำงานอยู่ เช่น พิมพ์ getwd () จะได้ผลดังนี้ “C:/Documents and Settings/test_R/My Documents”
setwd (“ตำแหน่ง”)	ใช้กำหนดตำแหน่งที่จะให้โปรแกรม R ทำงาน เช่น setwd (“c:\test_R”) # ให้ทำงานที่ drive c ที่โฟล์เดอร์ชื่อ test_R
dir ( )	ใช้ดูรายชื่อไฟล์ที่อยู่ในตำแหน่งที่โปรแกรม R ทำงานอยู่
history (“ชื่อ”)	ให้แสดงคำสั่งที่ผู้ใช้พิมพ์ไปล่าสุด 215

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
savehistory ("ชื่อ")	ใช้บันทึกคำสั่งที่ป้อนไปล่าสุด 25 คำสั่ง # ได้ .Rhistory
loadhistory("ชื่อ")	ใช้เรียกไฟล์บันทึกคำสั่งที่ป้อนไปล่าสุด 25 คำสั่ง
save (object)	ใช้บันทึก object ที่ระบุและอยู่บนWorkspace ถ้าไม่ระบุถือว่าบันทึก Object ทั้งหมดที่มีอยู่ขณะนั้น # ได้ .RData
save . image ( )	ใช้บันทึก object ทั้งหมดที่มีอยู่ขณะนั้น ไฟล์ที่บันทึกแบบนี้จะถูกเรียกเข้ามาโดยอัตโนมัติ เมื่อมีการเรียกใช้โปรแกรม R
load ("ชื่อ.RData")	ใช้เรียกไฟล์ object มาไว้ใน Workspace ถ้าไม่ระบุถือว่าบันทึก Objectทั้งหมดที่มีอยู่ขณะนั้น
ls ( )	ใช้ดูรายชื่อ Object ที่มีอยู่บน Workspace ของโปรแกรม R ขณะนั้น
rm ( )	ใช้ลบ Object ออกจาก Workspace ถ้าต้องการลบหมดทุก Object ใช้คำสั่งดังนี้ rm(list=ls(all=TRUE))
q ( )	ใช้ออกจากโปรแกรม R

6.2 Package Functions เป็นฟังก์ชั่นที่ใช้งานเกี่ยวกับ Package มีฟังก์ชั่นที่สำคัญดังต่อไปนี้

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน

sample (เงื่อนไข, replace = False)	ใช้สร้างเลขสุ่มตามเงื่อนไขที่กำหนด โดยปกติโปรแกรมจะสุ่มเลขที่ไม่ซ้ำกันเลย (Default replace = False) ดังตัวอย่างต่อไปนี้
	sample(20) # สุ่มมา 20 จำนวนโดยไม่มีเงื่อนไขใดๆ
	sample(10:80) # สุ่มมา 81 จำนวนโดยมีค่าตั้งแต่ 10 ถึง 80
	sample(5,replace=TURE) # สุ่มมา 5 จำนวนโดยมีค่าซ้ำกันได้
	x = 1:20;sample(x[x>9])) # สุ่มเลขที่มากกว่า 9 แต่ไม่เกิน 20

search()	ใช้ดูรายชื่อ Package ที่มีอยู่ในโปรแกรม R ในขณะนั้น ซึ่งใช้ในการดู Object ที่อยู่บน Workspace ไปพร้อมๆ กันด้วย โดยโปรแกรม R จะแสดงชื่อ Package และ Object บนวินโดวส์ R Console
library()	ใช้ดูรายชื่อ Package ที่ติดตั้งแล้วในโปรแกรม R ขณะนั้น โดยจะแสดงเฉพาะชื่อ Package บนวินโดวส์ที่เปิดขึ้นมาใหม่

ฟังก์ชัน	วัตถุประสงค์การใช้งาน
library ("ชื่อ")	ใช้เรียก Package ที่ระบุเข้ามาใช้งาน
detach (Package "ชื่อ")	ใช้เรียกวินโดวส์ที่เป็น Editor ของโปรแกรม R ขึ้นมาใช้งาน

6.3 Generating Functions เป็นฟังก์ชั่นที่ใช้สร้างขึ้นมาในลักษณะต่าง ๆ

ฟังก์ชัน	วัตถุประสงค์การใช้งาน

sample (เงื่อนไข, replace = False)	ใช้สร้างเลขสุ่มตามเงื่อนไขที่กำหนด โดยปกติโปรแกรมจะสุ่มเลขที่ไม่ซ้ำกันเลย (Default replace = False) ดังตัวอย่างต่อไปนี้
	sample(20) # สุ่มมา 20 จำนวนโดยไม่มีเงื่อนไขใด ๆ
	sample(10:80) # สุ่มมา 81 จำนวนโดยมีค่าตั้งแต่ 10 ถึง 80
	sample(5,replace=TURE) # สุ่มมา 5 จำนวนโดยมีค่าซ้ำกันได้
	x = 1:20;sample(x[x>9])) # สุ่มเลขที่มากกว่า 9 แต่ไม่เกิน 20

seq (เงื่อนไข) หรือ ค่าเริ่มต้น: ค่าสุดท้าย	ใช้สร้างเลขตามลำดับที่ระบุในเงื่อนไข เช่น seq(-2,2,0.75) ได้ผลดังนี้ -2.00 -1.25 -0.50 -0.25 1.00 1.75
rep (ซ้ำ, n)	ใช้สร้างค่าที่ระบุขึ้นมาซ้ำกัน n ครั้ง เช่น rep ("x",3)

ผู้ใช้อาจสร้างเลขสุ่ม ให้มีการแจกแจงตามแบบที่ต้องการได้โดยใช้ฟังก์ชั่น *r* ชื่อการแจกแจง เช่น rnorm(100) เป็นการให้สร้างเลขสุ่มที่มีการแจกแจงแบบปกติขึ้นมา 100 จำนวน

6.4 Input/ Output Functions เป็นฟังก์ชั่นที่ใช้ในการอ่านและบันทึกข้อมูลเป็นไฟล์ประเภทต่าง ๆ

ฟังก์ชัน	วัตถุประสงค์การใช้งาน
read . table ("ชื่อไฟล์")	ใช้อ่านไฟล์ข้อมูลประเภท Text แบบกำหนดตำแหน่งคงที่ไว้และมีชื่อตัวแปรไว้บรรทัดบนสุด
read . csv ("ชื่อ")	ใช้อ่านไฟล์ข้อมูลประเภท Text ที่มีการแยกคำด้วย , (comma)
read delim ("ชื่อ")	ใช้อ่านไฟล์ข้อมูลประเภท Text ที่มีการแยกคำด้วย , (Tab Semicolon)
write . table ("ชื่อข้อมูล" มี "ชื่อไฟล์" , sep = "\")	ใช้เรียกชุดข้อมูลที่ระบุขึ้นมาแก้ไขแบบ Worksheet
Edit ("ชื่อข้อมูล")	ใช้เรียกวินโดวส์ที่เป็น Editor ของโปรแกรม R ขึ้นมาใช้งาน
sepSummary (object)	ใช้บันทึกข้อมูลที่ระบุชื่อ "ชื่อข้อมูล" ไว้ โดยบันทึกเป็น "ชื่อไฟล์" เป็นไฟล์ประเภท Text ที่มีการแยกค่าด้วย "\t" คือ Tab ถ้าไม่ระบุจะแยกค่าเครื่องหมาย ช่องว่าง

นอกจากนี้ ยังมีฟังก์ชั่นที่ใช้ในการอ่านและบันทึกข้อมูล File ประเภทอื่นอีก เช่น จาก Excel จะต้องใช้ Package ชื่อ xlsReadWrite SPSS SAS STATA SYSTAT จะต้องใช้ Package ชื่อ foreign

6.5 Data Management Functions เป็นฟังก์ชั่นที่ใช้จัดการเกี่ยวกับข้อมูล

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน
show (ชื่อข้อมูล) หรือระบุชื่อข้อมูล	ใช้ดูค่าของข้อมูลที่ระบุ ซึ่งสามารถระบุชื่อชุดข้อมูลที่ต้องการดูได้ทันที
attach (ชื่อข้อมูล)	ใช้ดึง "ชื่อข้อมูล" ที่อยู่บน Workspace มาใช้งาน
detach (ชื่อข้อมูล)	ใช้ดึง "ชื่อข้อมูล" ที่ใช้งานอยู่ออกมาจากการใช้งาน
fix ("ชื่อข้อมูล")	ใช้เรียกชุดข้อมูลที่ระบุขึ้นมาแก้ไขแบบ Worksheet
Edit ("ชื่อข้อมูล")	ใช้เรียกวินโดวส์ที่เป็น Editor ของโปรแกรม R ขึ้นมาใช้งาน

ฟังก์ชั่นที่ใช้ในการเขียนโปรแกรม (Programming Functions)

เนื่องจากโปรแกรม R เป็นทั้งภาษา (R language) และโปรแกรมสำเร็จรูปในส่วนที่เป็นภาษาเป็นส่วนที่ให้ผู้ใช้ที่มีความรู้ทางด้านการเขียนโปรแกรม สามารถสั่งงานโปรแกรม R แบบภาษาคอมพิวเตอร์ได้ โดยเฉพาะผู้ที่มีความรู้ทางด้านภาษา C เพราะโปรแกรม R มีลักษณะใกล้เคียงกับภาษา C โดยมีฟังก์ชั่นพื้นฐานสำหรับการเขียนโปรแกรมเบื้องต้น ดังต่อไปนี้

ฟังก์ชั่น	วัตถุประสงค์การใช้งาน

if (เงื่อนไข) นิพจน์
หรือ
if (เงื่อนไข) นิพจน์ 1
else นิพจน์ 2
	ใช้ให้ทำคำสั่งตามเงื่อนไขที่กำหนด เช่น
	if (x >)5000 ot = 20 else ot = 30)
	หมายถึง ถ้า x มากกว่า 5000 แล้ว ot = 20 ถ้าไม่ใช่ ot = 30

if else (เงื่อนไข) นิพจน์ 1, else นิพจน์ 2	ใช้กำหนดจำนวนครั้งที่ต้องการให้ทำซ้ำ เช่น if else (y = 3, sqrt (x), x*3)
While (เงื่อนไข) นิพจน์	ใช้กำหนดให้ทำซ้ำ ในขณะที่เงื่อนไขเป็นจริง While (y<8) print ("x") # พิมพ์ x เมื่อ y<8

ผู้ใช้งานสามารถใช้ฟังก์ชั่นเหล่านี้ร่วมกับฟังก์ชั่นอื่นได้ โดยสามารถเขียนโปรแกรมเป็นไฟล์ประเภท TEXT และนำมาใช้ได้ทันทีที่วินโดวส์ R Console

โปรแกรม R เป็นที่ยอมรับและใช้กันในหมู่นักวิชาการและนักสถิติ จากสถาบันการศึกษาและองค์การต่าง ๆ อย่างกว้างขวาง และวารสารทางวิชาการต่าง ๆ ยอมรับผลการวิเคราะห์จาก R เนื่องจากนักวิชาการที่อ่านประเมินผลงานส่วนที่ตีพิมพ์ในวารสารเหล่านั้น ก็ใช้ R หรือ S ในการทำงานอยู่แล้ว การพัฒนา R นั้นเป็นไปค่อนข้างรวดเร็ว จะออกรุ่นใหม่ทุก ๆ 3 เดือน และมีการแก้ไขข้อบกพร่องที่พบและออก Patch ใหม่ ๆ ทุกสัปดาห์ ดังนั้นคาดว่า R จะมีพัฒนาการอย่างรวดเร็วและต่อเนื่องไปอีกในระยะยาว เนื่องจากมีผู้ร่วมพัฒนาเป็นจำนวนมากจากทั่วทุกมุมโลก (หัชชา ศรีปลั่ง, 2550, หน้า 5 - 6)