การทดสอบช่วงห่างระหว่างกลุ่ม

ฉัตรศิริ ปิยะพิมลสิทธิ์*

ปัญหาของการวัดความแตกต่างระหว่างกลุ่ม 2 กลุ่ม ในแต่ละกลุ่มถูกวัดด้วยตัวแปรหลายตัว เราจะหาความแตกต่างระหว่างกลุ่มทั้ง 2 กลุ่มได้อย่างไร เช่น กลุ่มของสายพันธุ์สุนัข 7 ชนิดที่ถูกวัดด้วยตัวแปรต่าง ๆ 6 ตัวแปร สุนัขสายพันธุ์หนึ่งจะแตกต่างจากสุนัขสายพันธุ์อื่น ๆ หรือไม่ จากแนวคิดนี้ ถ้ามีสุนัข 2 กลุ่มที่มีค่าเฉลี่ยของตัวแปรต่าง ๆ ที่ถูกวัดใกล้เคียงกัน ย่อมหมายความว่า สุนัขทั้ง 2 สายพันธุ์นี้มีคุณลักษณะใกล้เคียงกัน

1. ช่วงห่างระหว่างกลุ่ม

ในการอธิบายเรื่องช่วงห่าง (distance) จะเริ่มต้นอธิบายด้วยตัวอย่างง่าย ๆ อย่างในกรณีที่เรามีกลุ่มทั้งหมด n กลุ่ม ซึ่งในแต่ละกลุ่มวัดตัวแปร p ตัว เขียนเป็นสัญลักษณ์ได้ว่า $X_1, X_2, \ldots, X_p$ ค่าในแต่ละตัวแปรสำหรับกลุ่ม i เขียนเป็นสัญลักษณ์ได้ว่า $X_{i1}, X_{i2}, \ldots, X_{ip}$ และค่าในแต่ละตัวแปรสำหรับกลุ่ม j เขียนเป็นสัญลักษณ์ได้ว่า $X_{j1}, X_{j2}, \ldots, X_{jp}$ ปัญหาอยู่ตรงที่ว่าเราจะวัดช่วงห่างระหว่างกลุ่มทั้งสองนี้ได้อย่างไร

ถ้าเรามีกลุ่มที่ต้องการวัด 2 กลุ่มและในแต่ละกลุ่มวัด 2 ตัวแปร ค่าที่ได้จะสามารถนำมาสร้างเป็นกราฟได้ดังรูปภาพ 1 ตามทฤษฎีบทของปีทากอรัส เราสามารถหาความยาว (d_{ij}) ระหว่างกลุ่ม i และกลุ่ม j ด้วยสูตรที่เรียกว่า Euclidean distance ดังนี้

$$d_{ij} = \sqrt{\{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2\}}$$

เมื่อ p = 3 ตัวแปร สามารถนำค่ามาสร้างเป็นกราฟได้ดังรูปภาพ 2 ตามทฤษฎีของปีทากอรัส ช่วงห่างระหว่างจุดทั้งสองจะเท่ากับ

$$d_{ij} = \sqrt{\left\{\sum_{k=1}^{3} (X_{ik} - X_{jk})^2\right\}}$$

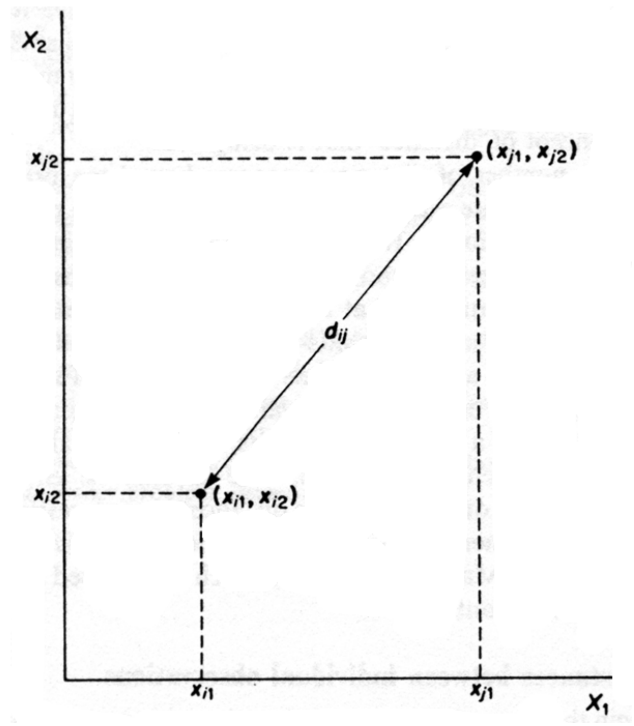
เมื่อมีตัวแปรมากกว่า 3 ตัวแปรเราไม่สามารถสร้างกราฟให้ดูได้ แต่เราสามารถคำนวณช่วงห่างระหว่างกลุ่ม 2 กลุ่มได้โดยใช้สูตรในรูปทั่วไปดังนี้

$$d_{ij} = \sqrt{\left\{\sum_{k=1}^{p} (X_{ik} - X_{jk})^2\right\}} \qquad (1)$$

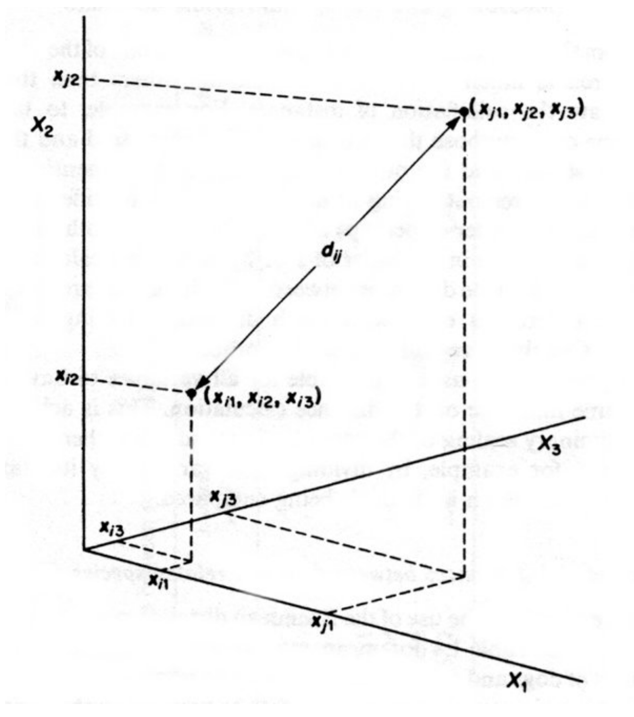
การคำนวณ Euclidian Distance นี้มีปัญหาตรงสเกลการวัดของแต่ละตัวแปรซึ่งอาจจะไม่เท่ากับ หรือสเกลต่างกัน วิธีแก้ปัญหาตรงนี้ก็คือแปลงค่าที่วัดได้นั้นให้เป็นคะแนนมาตรฐาน (Z-score) แล้วนำคะแนนมาตรฐานมาใช้ในการคำนวณหาช่วงห่าง เพื่อความเข้าใจลองศึกษาจากตัวอย่างต่อไปนี้

---

* กศ.ม. (การวัดผลการศึกษา) http://www.watpon.com/



ภาพประกอบ 1 แสดง Euclidean distance ระหว่างกลุ่มตัวอย่าง i และ j ที่วัดใน 2 ตัวแปร



ภาพประกอบ 2 แสดง Euclidean distance ระหว่างกลุ่มตัวอย่าง i และ j ที่วัดใน 3 ตัวแปร

ตัวอย่างที่ 1 ช่วงห่างระหว่างสายพันธุ์ของสุนัข

ตาราง 1 ค่าเฉลี่ยที่วัดได้จากคุณลักษณะต่าง ๆ ของสุนัขสายพันธ์ต่าง ๆ

กลุ่มพันธุ์สุนัข	X_1	X_2	X_3	X_4	X_5	X_6
Modern Thai dogs	9.7	21.0	19.4	7.7	32.0	36.5
Golden jackal	8.1	16.7	18.3	7.0	30.3	32.9
Chinese wolf	13.5	27.3	26.8	10.6	41.9	48.1
Indian wolf	11.5	24.3	24.5	9.3	40.0	44.6
Cuon	10.7	23.5	21.4	8.5	28.8	37.6
Dingo	9.6	22.6	21.1	8.3	34.4	43.1
Prehistoric dog	10.3	22.1	19.1	8.1	32.3	35.0

จากข้อมูลของตัวแปรทั้ง 6 ตัวที่วัดจากกลุ่มสุนัข 7 สายพันธ์ มาคำนวณหา Euclidean distance ก่อนอื่นต้องแปลงผลการวัดให้เป็นคะแนนมาตรฐานเสียก่อน วิธีการหาคะแนนมาตรฐานจะต้องหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแต่ละตัวแปรแล้วแทนค่าในสูตร

$$Z = \frac{X - \overline{X}}{SD}$$

ตาราง 2 คะแนนมาตรฐานที่คำนวณจากข้อมูลในตาราง 1

กลุ่มพันธุ์สุนัข	X_1	X_2	X_3	X_4	X_5	X_6
Modern Thai dogs	-0.50	-0.50	-0.74	-0.74	-0.40	-0.61
Golden jackal	-1.52	-1.93	-1.12	-1.39	-0.86	-1.31
Chinese wolf	1.92	1.60	1.84	1.95	1.68	1.62
Indian wolf	0.65	0.60	1.04	0.74	1.26	0.95
Cuon	0.14	0.33	-0.04	0.00	-1.19	-0.40
Dingo	-0.56	0.03	-0.14	-0.19	0.03	0.66
Prehistoric dog	-0.12	-0.13	-0.84	-0.37	-0.43	-0.90

ในตัวแปร X_1 คำนวณได้ค่าเฉลี่ย 10.486 และมีส่วนเบี่ยงเบนมาตรฐาน 1.572 ดังนั้นคะแนนมาตรฐานของกลุ่ม Modern thai dog จะเท่ากับ (9.7-10.486)/1.572 = -0.50 กลุ่ม Golden jackal ได้คะแนนมาตรฐานเท่ากับ (8.1-10.486)/1.572 = -1.52 โดยคำนวณให้ครบทุกกลุ่ม และครบทั้ง 7 ตัวแปร จะได้ค่าในตาราง 2

จากนั้นคำนวณค่า Euclidean distances ระหว่างสุนัขสายพันธ์ต่าง ๆ โดยจับมาคำนวนทีละคู่ เช่นคู่ของ Modern Thai dog กับ Golden jackal คำนวณหาช่วงห่างได้ดังนี้

$$d_{12} = \sqrt{(-0.50+1.52)^2 + (-0.50+1.93)^2 + (-0.74+1.12)^2 +} \sqrt{(-0.74+1.39)^2 + (-0.49+0.86)^2 + (-0.61+1.31)^2}$$

$$= \sqrt{1.0404 + 2.0449 + 0.1444 + 0.4225 + 0.1369 + 0.49}$$

$$= \sqrt{4.2791}$$

$$= 2.068599$$

เมื่อคำนวนหาช่วงห่างครบทุกคู่แล้วจะได้ค่าดังตาราง 3

ตาราง 3 แสดง Euclidean distance ระหว่างสายพันธุ์สุนัขทั้ง 7 พันธุ์

	G1	G2	G3	G4	G5	G6	G7
Modern Thai dogs	-						
(G1)	2.07	-					
Golden jackal (G2)	5.81	7.69	-				
Chinese wolf (G3)	3.66	5.49	2.31	-			
Indian wolf (G4)	1.63	3.45	4.94	3.14	-		
Cuon (G5)	1.68	3.44	4.55	2.37	1.80	-	
Dingo (G6)	0.72	2.58	5.52	3.49	1.38	1.84	-
Prehistoric dog (G7)							

ผลที่ได้จากการคำนวณช่วงห่างในตาราง 3 จะเห็นได้ว่ากลุ่มสุนัขพันธุ์ Prehistoric dog มีคุณลักษณะคล้ายคลึงกับสุนัขสายพันธุ์ Modern thai dogs ทั้งนี้เพราะค่าช่วงห่างระหว่างสองกลุ่มนี้มีค่าน้อยที่สุดในตาราง ซึ่งสรุปได้ว่าสุนัขทั้งสองสายพันธุ์นี้มีคุณลักษณะของตัวแปรต่าง ๆ ใกล้เคียงกันมากที่สุด

2. ช่วงห่างระหว่างกลุ่มประชากรและกลุ่มตัวอย่าง

สมมติว่ามีประชากร g กลุ่มและแต่ละกลุ่มถูกวัด p ตัวแปร แล้ว เราจะใช้สัญลักษณ์ μ_{ki} แทนค่าเฉลี่ยของตัวแปร X_k ในกลุ่มประชากรที่ i และให้ V_k แทนความแปรปรวนของตัวแปร X_k

Penrose (1953) ได้เสนอสูตรในการคำนวณช่วงห่างระหว่างประชากร i และประชากร j ไว้ดังนี้

$$P_{ij} = \sum_{k=1}^{p} \frac{(\mu_{ki} - \mu_{kj})^2}{pV_k} \qquad (2)$$

ข้อเสียในการวัดของ Penrose นั้นไม่ได้นำสหสัมพันธ์ระหว่างตัวแปรเข้ามาใช้ในการคำนวณด้วย นั่นคือเมื่อตัวแปร 2 ตัวถูกวัดกับกลุ่มตัวอย่างกลุ่มเดียวกันแล้ว อาจจะเป็นไปได้ว่าตัวแปร 2 ตัวนั้นมีความสัมพันธ์กัน

การวัดช่วงห่างที่ได้นำสหสัมพันธ์ระหว่างตัวแปรมาใช้ในการคำนวณด้วยก็คือ Mahalanobis distance (1948) มีสูตรอยู่ว่า

$$D_{ij}^2 = \sum_{r=1}^{p}\sum_{s=1}^{p}(\mu_{ri} - \mu_{rj})\upsilon^{rs}(\mu_{si} - \mu_{sj}) \qquad (3)$$

เมื่อ υ^{rs} คือสมาชิกในแถวที่ r และสดมภ์ที่ s ในอินเวอร์สของเมตริกซ์ความแปรปรวนร่วมระหว่างตัวแปร p ตัว จากสมการ (3) เขียนอยู่ในรูปของเมตริกซ์ได้ดังนี้

$$D_{ij}^2 = (\mu_i - \mu_j)'V^{-1}(\mu_i - \mu_j) \qquad (4)$$

เมื่อ

$$\mu_i = \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \\ \vdots \\ \mu_{pi} \end{bmatrix}$$

คือเวกเตอร์ค่าเฉลี่ยของประชากรกลุ่ม i และ V คือเมตริกซ์ความแปรปรวนร่วม

Mahalanobis distance เป็นสถิติที่ใช้กันมากสำหรับวัดช่วงห่างของกลุ่มตัวอย่างกลุ่มเดียวที่ถูกวัดหลายตัวแปรว่าแตกต่างไปจากค่าเฉลี่ยของกลุ่มประชากรหรือไม่ ถ้าให้ $X_1, X_2, \ldots, X_p$ เป็นค่าที่ได้จากการวัดกลุ่มตัวอย่างกลุ่มเดียวใน p ตัวแปร และให้ $\mu_1, \mu_2, \ldots \mu_p$ เป็นค่าเฉลี่ยของประชากรแล้ว

$$\begin{aligned} D^2 &= \sum_{r=1}^{p}\sum_{s=1}^{p}(X_r - \mu_r)\upsilon^{rs}(X_s - \mu_s) \\ &= (X - \mu)'V^{-1}(X - \mu) \qquad (5) \end{aligned}$$

สมการ (2) ถึงสมการ (5) เห็นได้ชัดเจนในการนำไปใช้กับกลุ่มตัวอย่างเพื่อประมาณค่าเฉลี่ยของประชากร ค่าความแปรปรวน และค่าความแปรปรวนร่วมจะถูกใช้ในฐานะของค่าแท้จริง (true values) ในกรณีของเมตริกซ์ความแปรปรวนร่วม V จะใช้ในสมการ (3) และ (4) ควรจะแทนที่ด้วยการรวมเมตริกซ์ความแปรปรวนร่วมของทุกกลุ่ม สมมติว่ามีกลุ่มตัวอย่าง m กลุ่ม และกลุ่มตัวอย่างที่ i มีขนาด n_i กับเมตริกซ์ความแปรปรวนร่วมคือ C_i แล้วแทนค่าในสูตร

$$C = \frac{\sum_{i=1}^{m}(n_i - 1)C_i}{\sum_{i=1}^{m}(n_i - 1)} \qquad (6)$$

C คือเมตริกซ์ความแปรปรวนร่วมรวมทุกกลุ่ม ในหนึ่งกลุ่มตัวอย่างเมตริกซ์ความแปรปรวนร่วม C_i จะมี df = n_i – 1 ขณะที่ C รวมทุกกลุ่มจะมี df = $\Sigma(n_i - 1)$

หลักของ Mahalanobis distance จะเหนือกว่า Penrose Distance เพราะว่าจะใช้ข้อมูลจากความแปรปรวนร่วม อย่างไรก็ตามก็มีประโยชน์เฉพาะเมื่อรู้ความแปรปรวนร่วมระหว่างตัวแปร สูตร Penrose distance จะเหมาะสมเมื่อจำนวนของ df น้อย ๆ ส่วน Mahalanobis distance จะเหมาะสมเมื่อ df มีจำนวนมากตั้งแต่ 100 ขึ้นไป

***ตัวอย่าง** 2* ช่วงห่างระหว่างกลุ่มตัวอย่างของกระโหลกศีรษะมนุษย์

ในการศึกษากลุ่มตัวอย่างเป็นกระโหลกศีรษะมนุษย์ 5 ยุคตั้งแต่ Early predynastic, Late predynastic, 12/13 th dynasties, Ptolematic และ Roman โดยแต่ละยุคจะถูกวัดใน 4 ตัวแปร ได้ค่าเฉลี่ยดังนี้

$$\bar{X}_1 = \begin{bmatrix} 131.37 \\ 133.60 \\ 99.17 \\ 50.53 \end{bmatrix}, \quad \bar{X}_2 = \begin{bmatrix} 132.37 \\ 132.70 \\ 99.07 \\ 50.23 \end{bmatrix}, \quad \bar{X}_3 = \begin{bmatrix} 134.437 \\ 133.80 \\ 96.03 \\ 50.57 \end{bmatrix}$$

$$\bar{X}_4 = \begin{bmatrix} 135.50 \\ 132.30 \\ 94.53 \\ 51.97 \end{bmatrix}, \quad \text{และ} \quad \bar{X}_5 = \begin{bmatrix} 136.17 \\ 130.33 \\ 93.50 \\ 51.37 \end{bmatrix}$$

และคำนวณหาเมตริกซ์ความแปรปรวนร่วมระหว่างตัวแปรของกลุ่มกระโหลกศีรษะมนุษย์ทั้ง 5 ยุคได้ดังนี้

$$C_1 = \begin{bmatrix} 26.31 & 4.45 & 0.45 & 7.25 \\ 4.15 & 19.97 & -0.79 & 0.39 \\ 0.45 & -0.79 & 34.63 & -1.92 \\ 7.25 & 0.39 & -1.92 & 7.64 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} 23.14 & 1.01 & 4.77 & 1.84 \\ 1.01 & 21.60 & 3.37 & 5.62 \\ 4.77 & 3.37 & 18.89 & 0.19 \\ 1.84 & 5.62 & 0.19 & 8.74 \end{bmatrix}$$

$$C_3 = \begin{bmatrix} 12.12 & 0.79 & -0.78 & 0.90 \\ 0.79 & 24.79 & 3.59 & -0.09 \\ -0.78 & 3.59 & 20.72 & 1.67 \\ 0.90 & -0.09 & 1.67 & 12.60 \end{bmatrix}$$

$$C_4 = \begin{bmatrix} 15.36 & -5.53 & -2.17 & 2.05 \\ -5.53 & 26.36 & 8.11 & 6.15 \\ -2.17 & 8.11 & 21.09 & 5.33 \\ 2.05 & 6.15 & 5.33 & 7.96 \end{bmatrix}$$

และ $$C_5 = \begin{bmatrix} 28.63 & -0.23 & -1.88 & -1.99 \\ -0.23 & 24.71 & 11.72 & 2.15 \\ -1.88 & 11.72 & 25.57 & 0.40 \\ -1.99 & 2.15 & 0.40 & 13.83 \end{bmatrix}$$

จากนั้นจะทำการรวมเมตริกซ์ความแปรปรวนร่วมโดยใช้สมการ (6) ขนาดของกลุ่มตัวอย่างทุกกลุ่มเป็น 30 ดังนั้นจะได้ค่าเฉลี่ยของเมตริกซ์ทั้ง 5 ดังนี้

$$C = \begin{bmatrix} 21.111 & 0.037 & 0.079 & 2.009 \\ 0.037 & 23.485 & 5.200 & 2.845 \\ 0.079 & 5.200 & 24.179 & 1.133 \\ 2.009 & 2.845 & 1.133 & 10.153 \end{bmatrix}$$

กับ df $= \Sigma(n_i - 1) = 145$

เราจะใช้ penrose distance ในสมการ (2) มาคำนวณช่วงห่างระหว่างกลุ่มตัวอย่างแต่ละคู่ เมื่อตัวแปร p = 4 ตัวแปร กับความแปรปรวนที่คำนวณได้ $V_1 = 21.111$, $V_2 = 23.485$, $V_3 = 24.179$ และ $V_4 = 10.153$ ซึ่งเป็นค่าที่มาจากแนวทแยงของเมตริกซ์ความแปรปรวนร่วม (C) แทนค่าสูตรเพื่อคำนวณช่วงห่างระหว่างกลุ่มตัวอย่างที่ 1 และกลุ่มตัวอย่าง 2 ได้ดังนี้

$$P_{12} = \frac{(131.37-132.37)^2}{4\times 21.111} + \frac{(133.60-132.70)^2}{4\times 23.485} + \frac{(99.17-99.07)^2}{4\times 24.179} + \frac{(50.53-50.23)^2}{4\times 10.153}$$

$= \quad 0.023$

แล้วคำนวณช่วงห่างระหว่างกลุ่มตัวอย่างทุก ๆ คู่ ได้ค่าสถิติ Penrose distance ดังตาราง

	G1	G2	G3	G4	G5
Early predynastic (G1)	-				
Late predynastic (G2)	0.023	-			
12/13th dynasties (G3)	0.216	0.163	-		
Ptolemaic (G4)	0.493	0.404	0.108	-	
Roman (G5)	0.736	0.583	0.244	0.066	-

ผลจากการคำนวณแสดงว่า กระโหลกศีรษะมนุษย์ในยุคสมัยใกล้เคียงกันจะมีความแตกต่างกันระหว่างตัวแปรต่าง ๆ น้อยกว่ากระโหลกศีรษะมนุษย์ในยุคสมัยที่ห่างกัน นั่นแสดงว่าช่วงเวลามีผลต่อความเปลี่ยนแปลงคุณลักษณะต่าง ๆ ของกระโหลกศีรษะมนุษย์

ต่อมาจะใช้ Mahalanobis distance คำนวณหาช่วงห่างโดยใช้สูตรในสมการ (3) ซึ่งต้องหาเมตริกซ์อินเวอร์สของความแปรปรวนร่วมเสียก่อน

$$C^{-1} = \begin{bmatrix} 0.0483 & 0.0011 & 0.0001 & -0.0099 \\ 0.0011 & 0.0461 & -0.0094 & -0.0121 \\ 0.0001 & -0.0094 & 0.0435 & -0.0022 \\ -0.0099 & -0.0121 & -0.0022 & 0.1041 \end{bmatrix}$$

คำนวณหาช่วงห่างระหว่างกลุ่มตัวอย่างที่ 1 กับกลุ่มตัวอย่างที่ 2 ได้ดังนี้

$$\begin{aligned} D_{12}^2 = \; & (131.37 - 132.37)0.0483(131.37 - 132.37) \\ & + (131.37 - 132.37)0.0011(133.60 - 132.70) \\ & + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & - (50.53 - 50.23)0.0022(99.17 - 99.07) \\ & + (50.53 - 50.23)0.1041(50.53 - 50.23) \\ = \; & 0.091 \end{aligned}$$

คำนวณช่วงห่างระหว่างกลุ่มตัวอย่างคู่อื่น ๆ จะได้ผลดังตาราง

	G1	G2	G3	G4	G5
Early predynastic (G1)	-				
Late predynastic (G2)	0.091	-			
12/13th dynasties (G3)	0.903	0.729	-		
Ptolemaic (G4)	1.881	1.594	0.443	-	
Roman (G5)	2.697	2.176	0.911	0.219	-

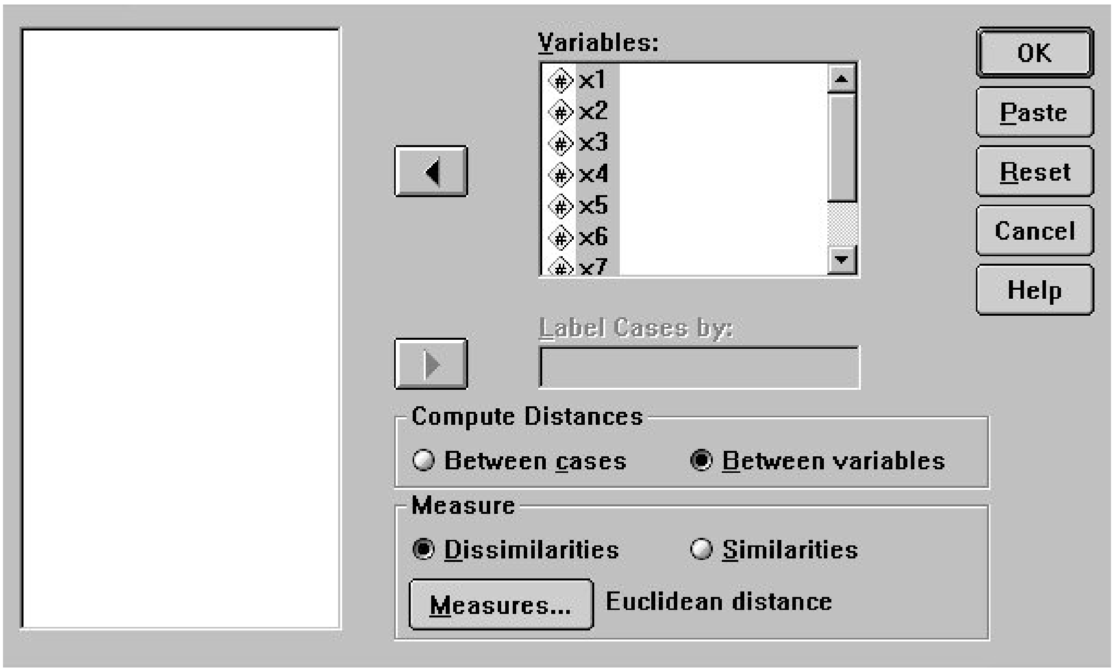
เปรียบเทียบการคำนวณช่วงห่างด้วยสูตรทั้งสองจะเห็นว่าสอดคล้องกัน โดยที่ Mahalanobis distance จะให้ค่าสูงกว่า Penrose distance ประมาณ 3-4 เท่า แต่ช่วงห่างระหว่างกลุ่มตัวอย่างแต่ละค่าที่คำนวณด้วยสูตรทั้งสองจะให้ผลทำนองเดียวกัน

3. **การใช้** SPSS 10.0 for Windows **วิเคราะห์** Euclidean Distance

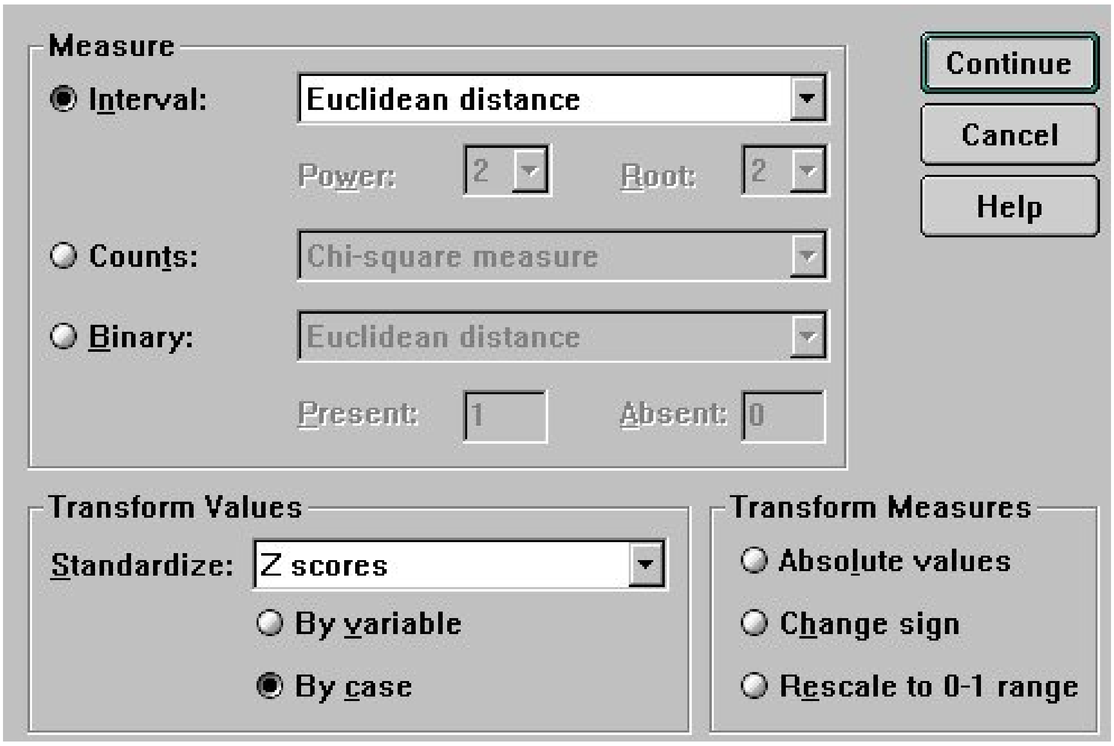
จากข้อมูลที่มีอยู่ในตาราง 1 นั้นป้อนใส่ในโปรแกรม SPSS for Windows ที่หน้าต่าง Data View โดยจะมีทั้งหมด 7 ตัวแปร ได้ดังภาพ

	x1	x2	x3	x4	x5	x6	x7	var
1	9.70	8.10	13.50	11.50	10.70	9.60	10.30	
2	21.00	16.70	27.30	24.30	23.50	22.60	22.10	
3	19.40	18.30	26.80	24.50	21.40	21.10	19.10	
4	7.70	7.00	10.60	9.30	8.50	8.30	8.10	
5	32.00	30.30	41.90	40.00	28.80	34.40	32.30	
6	36.50	32.90	48.10	44.60	37.60	43.10	35.00	
7								

เริ่มต้นวิเคราะห์โดยใช้เมนูหลัก “Analyze” เมนูรอง “Correlate” และเมนูย่อย “Distances…” จะปรากฏหน้าต่าง “Distances” คลิกตัวแปรทั้งหมดที่ต้องการเปรียบเทียบไปยังหน้าต่าง “Variables:” ด้านล่างในกรอบ “Compute Distance” เนื่องจากต้องเปรียบเทียบระหว่างตัวแปรให้คลิกที่ “Between Variables” ส่วนในกรอบ “Measure” นั้นให้เลือกที่ “Dissimiralities” ดังภาพ



คลิกปุ่ม “Measures...” จะปรากฏหน้าต่าง



ให้เลือกชนิดของข้อมูลมีทั้งข้อมูลแบบ “Interval”, “Counts” และ “Binary” แต่ละชนิดของข้อมูลก็มีสถิติให้เลือกคำนวณแตกต่างกันไป ในที่นี้เป็นคะแนน Interval และเลือกสถิติที่ต้องการทดสอบคือ “Euclidean Distance” ในกรอบ “Transform Value” ใช้ในการแปลงข้อมูล ในที่นี้ต้องการแปลงเป็นคะแนนมาตรฐาน ให้เลือก “Z-score” และแปลงทุก “Case” จากนั้นคลิกปุ่ม “Continue” และปุ่ม “OK” โปรแกรมจะทำการประมวลผลจนได้ผลลัพธ์ดังนี้

Proximity Matrix

	Euclidean Distance						
	X1	X2	X3	X4	X5	X6	X7
X1		1.912	5.382	3.386	1.512	1.559	.665
X2	1.912		7.121	5.059	3.190	3.183	2.389
X3	5.382	7.121		2.139	4.575	4.214	5.112
X4	3.386	5.059	2.139		2.911	2.197	3.228
X5	1.512	3.190	4.575	2.911		1.669	1.276
X6	1.559	3.183	4.214	2.197	1.669		1.704
X7	.665	2.389	5.112	3.228	1.276	1.704	

This is a dissimilarity matrix

สังเกตผลที่ได้จากการคำนวณด้วยมือในตาราง 3 จะได้ค่า Euclidean Distance แตกต่างจากการวิเคราะห์ด้วยโปรแกรมคอมพิวเตอร์เล็กน้อย ทั้งนี้เนื่องมาจากการปัดเศษทศนิยม

หนังสืออ้างอิง

Manly, Bryan F. J. ***Multivariate Statistical Methods***. London : Chapman & Hall, 1994.