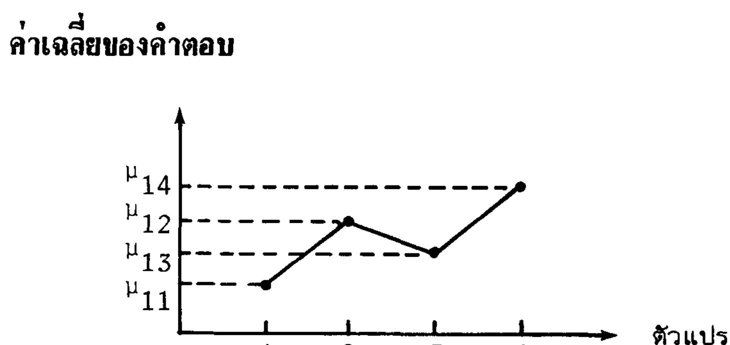
PROFILE ANALYSIS

คณิต ไข่มุกด์

การวิเคราะห์โปรไฟล์ (Profile analysis) เป็นสถานการณ์เมื่อชุดแบตเตอรี่ของทรีทเมนท์ p ตัว (อาจจะเป็น แบบสอบ คำถาม ก็ได้) ถูกจัดให้มีพลวิจัย (Subject) ตั้งแต่ 2 กลุ่มขึ้นไป โดยถือว่าการตอบของแต่ละกลุ่มเป็นอิสระกัน และคำตอบทั้งหมดจะต้องเป็นหน่วยที่เหมือนกัน โดยมีคำถามวิจัยว่า เวกเตอร์ประชากรเฉลี่ย (the population mean vectors) เท่ากันหรือไม่ ในการวิเคราะห์โปรไฟล์คำถามเกี่ยวกับการเท่ากันของเวกเตอร์ ค่าเฉลี่ยแบ่งได้หลายคำถาม

พิจารณาค่าเฉลี่ยของประชากร $\mu' = [\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}]$ แทนค่าเฉลี่ยของคำตอบจากทรีทเมนท์ 4 ตัว จากกลุ่มแรก ลองนำค่าไปลงจุด (plot) ดูดังรูปที่ 1



รูปที่ 1 ประชากรโปรไฟล์, p = 4

กราฟเส้นต่อไปนี้เป็นโปรไฟล์ของประชากรที่ 1 การโปรไฟล์อาจจะสร้างขึ้นจากแต่ละประชากร

ให้ $\mu'_1 = [\mu_{11}, \mu_{12}, \ldots, \mu_{1p}]$ และ $\mu'_2 = [\mu_{21}, \mu_{22}, \ldots \mu_{2p}]$ เป็นค่าเฉลี่ยของคำตอบของทรีทเมนท์ทั้ง p ตัว จากประชากรที่ 1 และ 2 ตามลำดับ

สมมติฐาน $H_0 : \mu_1 = \mu_2$ หมายถึงทรีทเมนท์ที่มีผลต่อประชากรทั้งสองเท่ากัน ในเทอมของโปรไฟล์ประชากร เราสามารถสร้างสูตรสำหรับคำถามของการเท่ากันดังนี้

1) โปรไฟล์ขนานกันหรือไม่ คือ $H_{01} : \mu_{1i} - \mu_{i1-1} = \mu_{2i} - \mu_{2i-1}, i = 2, 3, \ldots p,$

2) ถ้าโปรไฟล์ขนานกัน โปรไฟล์พ้องกัน (coincident) หรือไม่ คือ

$$H_{02} : \mu_{1i} = \mu_{2i},\ i = 1, 2, \ldots, p, \quad \text{ยอมรับได้หรือไม่}$$

3) ถ้าโปรไฟล์พ้องกัน ระดับโปรไฟล์ไหนที่ค่าเฉลี่ยทั้งหมดเท่ากัน คือ

$$H_{03} : \mu_{11} = \mu_{12} = \ldots = \mu_{1p} = \mu_{21} = \mu_{22} = \ldots = \mu_{2p} \quad \text{ยอมรับได้หรือไม่}$$

สมมติฐานศูนย์ในข้อที่ 1) เขียนได้เป็น

$$H_{01} : C\mu_1 = C\mu_2$$

เมื่อ C เป็นเมตริกซ์คอนทราส (contrast matrix)

$$\underset{((p-1)\times p)}{C} = \begin{bmatrix} -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots (1)$$

สำหรับตัวอย่างที่เป็นอิสระกันขนาด n_1 และ n_2 จาก 2 ประชากร สมมติฐานศูนย์สามารถทดสอบโดยการสร้างค่าแปลงจากค่าสังเกต (transformed observation)

$$Cx_{1j},\quad j = 1, 2, \ldots, n_1$$

และ

$$Cx_{2j},\quad j = 1, 2, \ldots, n_2$$

โดยมีเวกเตอร์เฉลี่ยของตัวอย่าง $C\bar{x}_1$ และ $C\bar{x}_2$ ตามลำดับ และมีเมตริกซ์ความแปรปรวนร่วมรวม (pooled covariance matrix) $CS_{pooled}C'$

เพราะว่าค่าสังเกตที่ได้แปลงแล้วทั้งสองชุด มีการแจกแจงแบบ N_{p-1}, $(C\mu_1, C\Sigma C')$ และ N_{p-1} $(C\mu_2, C\Sigma C')$ ตามลำดับ การประยุกต์ใช้ให้การทดสอบสำหรับโปรไฟล์คู่ขนาน ดังนี้

การทดสอบโปรไฟล์คู่ขนานสำหรับประชากรปกติสองประชากร

ปฏิเสธ $H_{01}: C\mu_1 = C\mu_2$ (โปรไฟล์คู่ขนาน) ที่ระดับ α ถ้า

$$T^2 = (\bar{x}_1 - \bar{x}_2)'C'\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right) CS_{pooled}C'\right]^{-1} C(\bar{x}_1 - \bar{x}_2) > c^2$$

เมื่อ $c^2 = \dfrac{(n_1 + n_2 - 2)(p-1)}{n_1 + n_2 - p} F_{p-1, n_1 + n_2 - p}(\alpha)$

เมื่อโปรไฟล์ขนานกัน ค่าแรกมากกว่าค่าที่สอง $\mu_{1i} > \mu_{2i}$, ทุกๆ i หรือในทางกลับกัน ภายใต้เงื่อนไขนี้โปรโฟล์จะพ้องกันเพียงแต่ถ้าผลรวมของความสูง $\mu_{11} + \mu_{12} + \ldots + \mu_{1p} = 1'\mu_1$ และ $\mu_{21} + \mu_{22} + \ldots + \mu_{2p} = 1'\mu_2$ เท่ากัน ฉะนั้นสมมุติฐานศูนย์ในข้อ 2) สามารถเขียนในรูปเท่ากันดังนี้

$$H_{02};\ 1'\mu_1 = 1'\mu_2$$

เราสามารถทดสอบ H_{02} ด้วยสถิติ–ที สองตัวอย่างที่ใช้อยู่ในค่าสังเกตตัวแปรเดี่ยว (univariate observations) $1'x_{1j} = 1, 2, \ldots, n_1$, และ $1'x_{2j}, j = 1, 2, \ldots, n_2$

การทดสอบโปรไฟล์พ้องกันโดยกำหนดให้ว่าโปรไฟล์เท่ากัน สำหรับประชากรปกติสองประชากร : ปฏิเสธ $H_{02}\ 1'\mu_1 = 1'\mu_2$ (โปรไฟล์พ้องกัน) ที่ระดับ α ถ้า

$$T^2 = 1'(\bar{x}_1 - \bar{x}_2)\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)1'S_{pooled}1\right]^{-1} 1'(\bar{x}_1 - \bar{x}_2) \qquad \ldots\ldots\ldots\ldots(3)$$

$$= \left(\frac{1'(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)1'S_{pooled}1}}\right) > t^2_{n_1 + n_2 - 2}\left(\frac{\alpha}{2}\right) = F_{1, n_1 + n_2 - 2}(\propto)$$

สำหรับโปรไฟล์พ้องกัน $x_{11}, x_{12}, \ldots, x_{1n_1}$ และ $x_{21}, x_{22}, \ldots, x_{2n_2}$ ค่าสังเกตทั้งหมดได้จากประชากรปกติเดียวกัน ในขั้นต่อไปเป็นการดูว่าถ้าตัวแปรทั้งหมดมีค่าเฉลี่ยเดียวกัน ดังนั้นโปรโฟล์ร่วมเป็นระดับ

สมมติฐานศูนย์ในข้อ 3) สามารถเขียนเป็น

$$H_{03}: C(\mu_1 + \mu_2) = 0$$

เมื่อ C ถูกกำหนดโดย (1) เมื่อ H_{01} และ H_{02} ถูกกันไว้ได้ เวกเตอร์ค่าเฉลี่ยร่วมถูกประมาณโดย

$$\bar{x} = \frac{\sum_{j=1}^{n_1} x_{1j} + \sum_{j=1}^{n_2} x_{2j}}{n_1 + n_2} = \frac{n_1}{(n_1+n_2)}\bar{x}_1 + \frac{n_2}{(n_1+n_2)}\bar{x}_2$$

แล้วเราจะทดสอบต่อไปนี้

การทดสอบระดับโปรไฟล์ กำหนดว่าโปรไฟล์พ้องกัน สำหรับประชากรปกติสองประชากร :

ปฏิเสธ $H_{03}: C(\mu_1 + \mu_2) = 0$ (ระดับโปรไฟล์) ที่ระดับ α ถ้า

$$(n_1+n_2)\,\bar{x}'\,C'\left[CS_{pooled}\,C'\right]^{-1} C\,\bar{x} > F_{p-1,\, n_1+n_2-p}(\alpha) \qquad \ldots (4)$$

ตัวอย่าง

ส่วนหนึ่งของการศึกษาเรื่องความรักและการแต่งงาน นักสังคมวิทยาได้สำรวจเกี่ยวกับการแต่งงานของผู้ใหญ่ในเรื่อง "การมีส่วนช่วยเหลือ" และ "ผลลัพธ์" และระดับของ "การหลงรัก" และ "ความเป็นเพื่อน" (ข้อมูลได้รับการเอื้อเฟื้อจาก E. Hatfield) ชายและหญิงที่แต่งงานใหม่ ๆ ตอบคำถามที่ใช้สเกล 8 จุด โดยมีลักษณะคำถามดังนี้

1. เมื่อพิจารณาทุกสิ่งแล้ว ท่านมีส่วนช่วยเหลือการแต่งงานอย่างไร
2. เมื่อพิจารณาทุกสิ่งแล้ว ท่านมีผลลัพธ์จากการแต่งงานอย่างไร

Extremely negative	Very negative	Moderately negative	Slightly negative	Slightly positive	Moderately positive	Very positive	Extremely positive
1	2	3	4	5	6	7	8

และให้ตอบคำถามต่อไปนี้โดยใช้สเกล 5 จุดดังนี้

3. ท่านรู้สึกหลงรักคู่ชีวิตของท่านระดับใด μ
4. ท่านรู้สึกเป็นเพื่อนกับคู่ชีวิตของท่านระดับใด μ

None at all	Very little	Some	A great deal	Tremendous amount
1	2	3	4	5

กำหนดให้ X_1 = คะแนนจากสเกล 8 จุดที่ตอบคำถามข้อที่ 1

X_2 = คะแนนจากสเกล 8 จุดที่ตอบคำถามข้อที่ 2

X_3 = คะแนนจากสเกล 5 จุดที่ตอบคำถามข้อที่ 3

X_4 = คะแนนจากสเกล 5 จุดที่ตอบคำถามข้อที่ 4

และประชากรสองประชากรคือ

ประชากรที่ 1 = ผู้ชายที่แต่งงานแล้ว

ประชากรที่ 2 = ผู้หญิงที่แต่งงานแล้ว

ค่าเฉลี่ยประชากร คือค่าเฉลี่ยของคำตอบจาก $p = 4$ คำถาม สำหรับประชากรผู้ชายและประชากรผู้หญิง สมมติแมตริกซ์ความแปรปรวนร่วมรวม เป็นที่สนใจที่จะดู ถ้าโปรไฟล์ของผู้ชายและผู้หญิงเป็นอันเดียวกัน

ตัวอย่าง $n_1 = 30$ คน และ $n_2 = 30$ คน ได้เวกเตอร์ค่าเฉลี่ยของตัวอย่าง

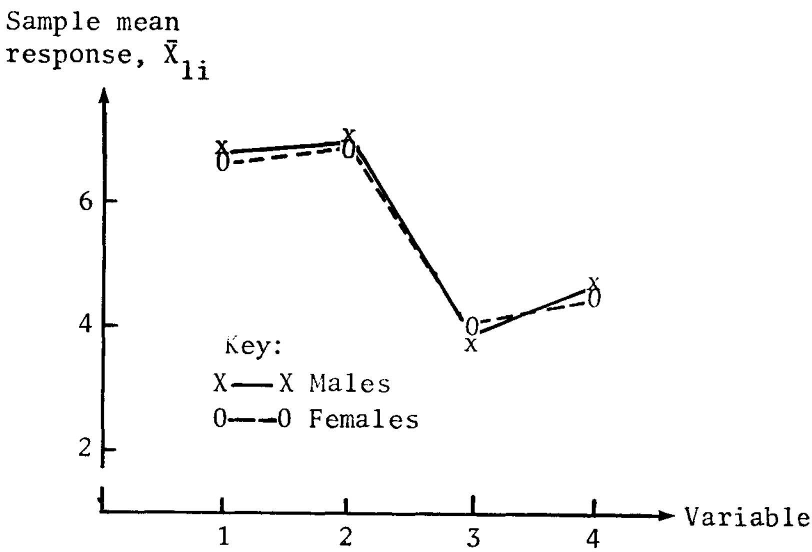
$$\bar{x}_1 = \begin{bmatrix} 6.833 \\ 7.033 \\ 3.967 \\ 4.700 \end{bmatrix}, \qquad \bar{x}_2 = \begin{bmatrix} 6.633 \\ 7.000 \\ 4.000 \\ 4.533 \end{bmatrix}$$

ผู้ชาย ผู้หญิง

และแมตริกซ์ความแปรปรวนร่วมรวม

$$S_{pooled} = \begin{bmatrix} .606 & .262 & .066 & .161 \\ .262 & .637 & .173 & .143 \\ .066 & .173 & .810 & .029 \\ .161 & .143 & .029 & .306 \end{bmatrix}$$

เวกเตอร์ค่าเฉลี่ยของตัวอย่างถูกนำไปลงจุดเป็นตัวอย่างโปรไฟล์ ในรูปที่ 2



รูปที่ 2 โปรไฟล์ตัวอย่างสำหรับคำตอบ การแต่งงาน–ความรัก

เพราะว่าขนาดของตัวอย่างใหญ่สมเหตุสมผล เราจะใช้วิธีการทฤษฎีปกติแม้ว่าข้อมูลจะเป็นเลขจำนวนเต็มซึ่งเห็นได้ชัดว่าไม่เป็นปกติ เพื่อทดสอบความเป็นคู่ขนาน $(H_{01} : C\mu_1 = C\mu_2)$,

เราคำนวณ $$CS_{pooled}C' = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} S_{pooled} \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} .719 & -.268 & -.125 \\ -.268 & 1.101 & -.751 \\ -.125 & -.751 & 1.058 \end{bmatrix}$$

และ

$$C(\bar{x}_1 - \bar{x}_2) = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} .200 \\ .033 \\ -.033 \\ .167 \end{bmatrix} = \begin{bmatrix} -.167 \\ -.066 \\ .200 \end{bmatrix}$$

ดังนั้น

$$T^2 = [-.167. - .066,.200]\left(\frac{1}{30} + \frac{1}{30}\right)^{-1}\begin{bmatrix} .719 & -.268 & -.125 \\ -.268 & 1.101 & -.751 \\ -.125 & -.751 & 1.058 \end{bmatrix}^{-1}\begin{bmatrix} -.167 \\ -.066 \\ .200 \end{bmatrix}$$

$$= 15(.067) = 1.005$$

มากว่านั้น กับ $\alpha = .05,\ c^2 = \left[(30 + 30 - 2)(4 - 1)/(30 + 30 - 4)\right] F_{3,56}(.05)$

$$= 3.11\ (2.8) = 8.7$$

เพราะว่า $T^2 = 1.005 < 8.7$

เราสรุปสมมติฐานของโปรไฟล์คู่ขนานสำหรับผู้ชายและผู้หญิงเป็นจริง กำหนดการลงจุดในรูปที่ 2 ข้อค้นพบนี้ไม่น่าประหลาดใจ

เมื่อโปรไฟล์ขนานกันเราสามารถทดสอบความพ้องกันของโปรไฟล์ โดยทดสอบ

H_{02} ; $1'\mu_1 = 1'\mu_2$ (โปรไฟล์พ้องกัน)

เราต้องการผลบวกของสมาชิกใน $(\bar{x}_1 - \bar{x}_2) = 1'\ (\bar{x}_1 - \bar{x}_2) = .367$

ผลบวกของสมาชิกใน $S_{pooled} = 1'S_{pooled}1 = 4.027$ ใช้ สมการ (3)

$$T^2 = \left(\frac{.367}{\sqrt{\left(\frac{1}{30} + \frac{1}{30}\right)4.027}}\right)^2 = .708$$

ด้วย $\alpha = .05,\ F_{1,58}(.05) = 4.0,\quad T^2 = .708 < F_{1,58}(0.5) = 4.0,$

เราไม่สามารถปฏิเสธสมมติฐานที่ว่าโปรไฟล์พ้องกัน นั่นคือคำตอบของผู้ชายและผู้หญิงกับคำถามทั้ง 4 เหมือนกัน

เราควรจะทดสอบระดับโปรไฟล์ต่อไปได้ แต่อย่างไรก็ตามมันไม่มีความหมายที่จะทดสอบต่อไปเพราะตัวอย่างของเราคำถามข้อที่ 1 และ 2 ใช้สเกล 1–8 ขณะที่คำถามข้อที่ 3 และ 4 ใช้สเกล 1–5 สเกลการจัดที่ไม่เท่าเทียมกันทำให้การทดสอบระดับของโปรไฟล์ไม่มีความหมาย

หนังสืออ้างอิง

Johnson, Richard Arnold. Applied Multivariate Statistical Analysis Prentice-Hall, Inc., 1982.