บทที่ 3

การทดสอบสมมติฐานเกี่ยวกับเวคเตอร์ค่าเฉลี่ยและเมทริกซ์

ในการทดสอบสมมติฐานทางสถิติเพื่อสรุปลักษณะที่สำคัญของประชากร ถ้าหากในกรณีที่ตัวแปรที่ต้องการทดสอบเป็นข้อมูลเชิงปริมาณจะต้องทำการทดสอบค่าเฉลี่ยของประชากร ซึ่งในทางปฏิบัติในการศึกษาจะมีการศึกษาตัวแปรหลาย ๆ ตัว จากกลุ่มตัวอย่างหรือประชากรเดียวกัน ทำให้มีค่าข้อมูลหลาย ๆ ค่าจากตัวอย่างแต่ละหน่วย หากเราต้องการศึกษาตัวแปร p ตัว การทดสอบค่าเฉลี่ยประชากรครั้งละ 1 ตัวแปร เป็นจำนวน p ครั้ง จะทำให้ความน่าจะเป็นของการเกิดความคลาดเคลื่อนชนิดที่ 1 หรือระดับนัยสำคัญ α เพิ่มขึ้นตามจำนวนครั้งของการทดสอบ ดังนั้นในการทดสอบสมมติฐานจึงควรทำการทดสอบพร้อมกันในรูปของเวคเตอร์ค่าเฉลี่ย

การทดสอบค่าเฉลี่ยประชากรหนึ่งกลุ่มตัวแปรหนึ่งตัว

π

ก่อนที่จะกล่าวถึงการทดสอบสมมติฐานในรูปของเวคเตอร์ค่าเฉลี่ยและเมทริกซ์ จะขอกล่าวถึงการทดสอบค่าเฉลี่ยประชากรหนึ่งกลุ่มตัวแปรหนึ่งตัวเสียก่อน ดังนี้

1. กรณีทราบค่าความแปรปรวนของประชากร (σ^2)

สถิติทดสอบคือ $$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

เนื่องจาก $$Z^2 = \frac{\left(\bar{x} - \mu_0\right)^2}{\sigma^2/n}$$ จะมีการแจกแจงแบบไคกำลังสองที่มีค่า df = 1 หรือ $Z^2 \sim \chi^2_1$

โดยสามารถจัดอยู่ในรูปใหม่ได้ดังนี้

$$Z^2 = n\left(\frac{\bar{x} - \mu_0}{\sigma}\right)^2 = n\left(\bar{x} - \mu_0\right)\left(\sigma^2\right)^{-1}\left(\bar{x} - \mu_0\right)$$

2. กรณีไม่ทราบค่าความแปรปรวนของประชากรจะประมาณค่า σ^2 ด้วย S^2

สถิติทดสอบคือ $$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

หรือ

$$t^2 = n\left(\frac{\bar{x} - \mu_0}{S}\right)^2 = n\left(\bar{x} - \mu_0\right)\left(S^2\right)^{-1}\left(\bar{x} - \mu_0\right)$$

สาเหตุที่ไม่ควรใช้การทดสอบค่าเฉลี่ยประชากรครั้งละ 1 ตัวแปร ในการทดสอบค่าเฉลี่ยของตัวแปรหลายตัว

ในการทดสอบค่าเฉลี่ยของตัวแปรหลายตัวหากทำการทดสอบแยกครั้งละ 1 ตัวแปร จะสามารถทำได้ถ้าตัวแปรที่ศึกษาทุกตัวเป็นอิสระกัน แต่ในทางปฏิบัตินั้นตัวแปรที่เราศึกษาจะมีความสัมพันธ์กันไม่มากก็น้อย ดังนั้นในการที่จะทดสอบค่าเฉลี่ยของประชากรครั้งละ 1 ตัวแปร จำนวน p ตัวแปรจะทำให้ค่าความคลาดเคลื่อนประเภทที่ 1 (Type I Error) สูงขึ้นตามจำนวนครั้งของการทดสอบ

P(ปฏิเสธ H_0 เมื่อ H_0 เป็นจริง อย่างน้อย 1 ครั้ง จากการทดสอบ p ครั้ง)

$= 1 – P($ยอมรับ H_0 เมื่อ H_0 เป็นจริง ทั้ง p ครั้ง)

$= 1 – (1 - \alpha)^p$

ซึ่งจะมีค่ามากกว่าระดับนัยสำคัญที่กำหนดไว้ในครั้งแรกคือ α

เช่น ถ้ากำหนดระดับนัยสำคัญในการทดสอบแต่ละครั้งเป็น 0.01 โดยมีตัวแปรที่ต้องการศึกษา 4 ตัวแปร (p = 4) โอกาสในการปฏิเสธ H_0 เมื่อ H_0 เป็นจริงอย่างน้อย 1 ครั้ง จะมากกว่า 0.01 ดังนี้

P(ปฏิเสธ H_0 เมื่อ H_0 เป็นจริง อย่างน้อย 1 ครั้ง จากการทดสอบ 4 ครั้ง)

$= 1 – P($ยอมรับ H_0 เมื่อ H_0 เป็นจริง ทั้ง 4 ครั้ง)

$= 1 – (1 – 0.01)^4$

$= 1 – (0.99)^4$

$= 1 – 0.9606$

$= 0.0394$ ซึ่งจะมีค่ามากกว่าระดับนัยสำคัญที่กำหนดไว้ในครั้งแรกคือ 0.01

การทดสอบสมมติฐานเวคเตอร์ค่าเฉลี่ยประชากร 1 กลุ่ม ($\underline{\mu}$) ตัวแปรหลายตัว

1. เมื่อทราบค่าเมตริกซ์ความแปรปรวนร่วมของประชากร Σ มีขั้นตอนการทดสอบ ดังนี้

1. กำหนดสมมติฐาน

$H_0 : \underline{\mu} = \underline{\mu}_0$

$H_1 : \underline{\mu} \neq \underline{\mu}_0$

หรือ

$$H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix} \qquad H_1 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \neq \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

โดยที่ $\underline{\mu}$ เป็นเวคเตอร์ของค่าคงที่ที่ต้องการทดสอบ

$\underline{\mu}_0$ เป็นเวคเตอร์ของค่าคงที่ที่กำหนด

ถ้ายอมรับ H_0 หมายถึง $\mu_i = \mu_{0i}$ ทุกค่า $i = 1, 2, \ldots, p$

ถ้าปฏิเสธ H_0 หมายถึง มี μ_i อย่างน้อย 1 ค่าที่ไม่เท่ากับ μ_{0i}

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

$$Z^2 = n(\overline{X} - \mu_0)' \Sigma^{-1} (\overline{X} - \mu_0)$$

โดยที่ $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$

3. คำนวณค่าสถิติทดสอบ

4. หาบริเวณวิกฤต ภายใต้ H_0 จะพบว่า Z^2 มีการแจกแจงแบบไคกำลังสองที่มีองศาความเป็นอิสระ (df) = p หรือกล่าวคือ $Z^2 \sim \chi_p^2$

5. สรุปผล

จะปฏิเสธสมมติฐาน H_0 เมื่อ $Z^2 > \chi^2_{\alpha;\, p}$

ตัวอย่างที่ 1 จากข้อมูลน้ำหนักและความสูงในตัวอย่างของนักศึกษาชายในมหาวิทยาลัยแห่งหนึ่งจำนวน 20 คน แสดงดังตาราง

คนที่	น้ำหนัก (X_1)	ความสูง (X_2)
1	69	153
2	74	175
3	68	155
4	70	135
5	72	172
6	67	150
7	66	115
8	70	137
9	76	200
10	68	130
11	72	140
12	79	265
13	74	185
14	67	112
15	66	140
16	71	150
17	74	165
18	75	185
19	75	210
20	76	220

สมมติว่าข้อมูลจากตัวอย่างนี้มาจากประชากรที่มีการแจกแจงแบบปกติสองตัวแปร $N_2(\underline{\mu}, \Sigma)$

โดยที่ $\Sigma = \begin{bmatrix} 20 & 100 \\ 100 & 1000 \end{bmatrix}$ ที่ระดับนัยสำคัญ 0.05 จงทดสอบว่า $\underline{\mu} = \begin{bmatrix} 70 \\ 170 \end{bmatrix}$ หรือไม่

วิธีทำ 1. กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \end{bmatrix} = \begin{bmatrix} 70 \\ 170 \end{bmatrix}$$

$$H_1 : \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \end{bmatrix} \neq \begin{bmatrix} 70 \\ 170 \end{bmatrix}$$

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

$$Z^2 = n(\overline{X} - \mu_0)' \Sigma^{-1} (\overline{X} - \mu_0)$$

3. คำนวณค่าสถิติทดสอบ

เวคเตอร์ค่าเฉลี่ยตัวอย่าง

$$\overline{X} = \begin{bmatrix} \dfrac{69+74+...+76}{20} \\ \dfrac{153+175+...+220}{20} \end{bmatrix} = \begin{bmatrix} 71.45 \\ 164.7 \end{bmatrix}$$

เมทริกซ์ความแปรปรวนร่วมประชากร

$$\Sigma = \begin{bmatrix} 20 & 100 \\ 100 & 1000 \end{bmatrix}$$

$$\Sigma^{-1} = \frac{1}{20000 - 10000} \begin{bmatrix} 1000 & -100 \\ -100 & 20 \end{bmatrix}$$

$$= \begin{bmatrix} 0.1 & -0.01 \\ -0.01 & 0.002 \end{bmatrix}$$

สถิติทดสอบ

$$Z^2 = n(\overline{X} - \mu_0)' \Sigma^{-1} (\overline{X} - \mu_0)$$

$$= 20\begin{bmatrix} 71.45-70 \\ 164.7-170 \end{bmatrix}' \begin{bmatrix} 0.1 & -0.01 \\ -0.01 & 0.002 \end{bmatrix} \begin{bmatrix} 71.45-70 \\ 164.7-170 \end{bmatrix}$$

$$= 20\begin{bmatrix} 1.45 & -5.3 \end{bmatrix} \begin{bmatrix} 0.1 & -0.01 \\ -0.01 & 0.002 \end{bmatrix} \begin{bmatrix} 1.45 \\ -5.3 \end{bmatrix}$$

$$= 8.40$$

4. หาบริเวณวิกฤต

ที่ระดับนัยสำคัญ 0.05 มีตัวแปร 2 ตัว จะได้ $\chi^2_{0.05;\,2} = 5.99$

5. สรุปผล

เนื่องจาก $Z^2 = 8.40$ มีค่ามากกว่า 5.99 จึงปฏิเสธสมมติฐาน H_0 นั่นคือที่ระดับนัยสำคัญ 0.05

$$\underline{\mu} \neq \begin{bmatrix} 70 \\ 170 \end{bmatrix}$$

2. เมื่อไม่ทราบค่าเมตริกซ์ความแปรปรวนร่วมของประชากร Σ

ถ้าสุ่มตัวอย่างขนาด n (X_1 , X_2 , … , X_n) จากประชากรที่มีการแจกแจงแบบปกติหลายตัวแปร $N_p(\underline{\mu}, \Sigma)$

โดยที่ X_i แทนข้อมูล p ตัวแปร จากหน่วยตัวอย่างที่ i ; i = 1,2, ..., n แต่ไม่ทราบค่า Σ จึงประมาณค่า Σ ด้วยเมทริกซ์ S^2 โดยมีขั้นตอนการทดสอบสมมติฐาน ดังนี้

1. กำหนดสมมติฐาน

$$H_0 : \underline{\mu} = \underline{\mu}_0$$

$$H_1 : \underline{\mu} \neq \underline{\mu}_0$$

หรือ

$$H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

$$H_1 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \neq \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}$$

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

$$T^2 = n\left(\bar{X} - \mu_0\right)' S^{-1}\left(\bar{X} - \mu_0\right)$$

3. คำนวณค่าสถิติทดสอบ

4. หาบริเวณวิกฤต

โดยที่ T^2 คือ โฮเทลลิ่งทีกำลังสอง (Hotelling’s T^2) ที่มีองศาความเป็นอิสระ (df) เท่ากับ p , n – 1

5. สรุปผล

จะปฏิเสธสมมติฐาน H_0 เมื่อ $T^2 > T^2_{\alpha; p, n-1}$

ข้อจำกัดของการใช้ Hotelling's T^2

1. $n - 1 > p$ มิฉะนั้น S จะเป็นเมทริกซ์เอกฐาน (Singular Matrix) ซึ่งจะไม่สามารถหาค่า S^{-1} ได้ ทำให้ไม่สามารถหา T^2 ได้

2. T^2 จะใช้ในการทดสอบสมมติฐานแบบ 2 ทางเท่านั้น ไม่สามารถใช้กับการทดสอบแบบทางเดียวได้

ความสัมพันธ์ระหว่าง Hotelling's T^2 กับ F

ในกรณีที่มีตัวแปร 1 ตัว (Univariate) จะได้ว่า $t^2_{n-1} = F_{1,n-1}$

ดังนั้นจะสามารถปรับ T^2 ให้มีการแจกแจงแบบ F ดังนี้

$$T^2 = n\left(\bar{X} - \mu_0\right)' S^{-1}\left(\bar{X} - \mu_0\right) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

โดยที่ $F_{p,n-p}$ มีการแจกแจงแบบเอฟที่มี df $= p$ และ $n - p$ ตามลำดับ

โดยในการสรุปผล สามารถทำได้ดังนี้

จะปฏิเสธสมมติฐานหลัก (H_0) ถ้า $T^2 > \frac{(n-1)p}{n-p} F_{\alpha\,;\,p,n-p}$ หรือ จาก $T^2 = \frac{(n-1)p}{n-p} F$

จะได้สถิติทดสอบ $F = \frac{(n-p)}{(n-1)p} T^2$ ซึ่งมีการแจกแจงแบบเอฟที่จำนวนองศาความเป็นอิสระ p และ n – p

จะปฏิเสธสมมติฐานหลัก (H_0) ถ้า $F > F_{\alpha\,;\,p,n-p}$

ตัวอย่างที่ 2 กำหนดให้เมทริกซ์ของข้อมูลสำหรับตัวอย่างสุ่มขนาด n = 3 จากประชากรที่มีการแจกแจงแบบปกติสองตัวแปร (Bivariate Normal) คือ

$$X = \begin{bmatrix} 7 & 10 \\ 11 & 7 \\ 9 & 4 \end{bmatrix}$$

จงทดสอบ

$$H_0 : \mu = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$$

ที่ระดับนัยสำคัญ 0.05

ก. ใช้การแจกแจงแบบโฮเทลลิ่งทีกำลังสอง

ข. ใช้การแจกแจงแบบเอฟ

วิธีทำ กำหนดสมมติฐาน

$$H_0 : \boldsymbol{\mu} = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$$

$$H_1 : \boldsymbol{\mu} \neq \begin{bmatrix} 10 \\ 6 \end{bmatrix}$$

จากโจทย์ $n = 3 , p = 2$

$$\overline{X} = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{j=1}^{n} X_{1j} \\ \frac{1}{n}\sum_{j=1}^{n} X_{2j} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{7+11+9}{3} \\ \dfrac{10+7+4}{3} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{27}{3} \\ \dfrac{21}{3} \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \end{bmatrix}$$

$$S_{11} = \sum_{j=1}^{n} \frac{\left(X_{1j} - \overline{X}_1\right)^2}{n - 1}$$

$$= \frac{(7 - 9)^2 + (11 - 9)^2 + (9 - 9)^2}{3 - 1}$$

$$= \frac{4 + 4 + 0}{2} = \frac{8}{2} = 4$$

$$S_{12} = \sum_{j=1}^{n} \frac{\left(X_{1j} - \overline{X}_1\right)\left(X_{2j} - \overline{X}_2\right)}{n - 1}$$

$$= \frac{(7 - 9)(10 - 7) + (11 - 9)(7 - 7) + (9 - 9)(4 - 7)}{3 - 1}$$

$$= \frac{-6 + 0 + 0}{3 - 1} = -3$$

$$S_{22} = \sum_{j=1}^{n} \frac{\left(X_{2j} - \overline{X}_2\right)^2}{n - 1}$$

$$= \frac{(10 - 7)^2 + (7 - 7)^2 + (4 - 7)^2}{3 - 1}$$

$$= \frac{9 + 0 + 9}{2} = \frac{18}{2} = 9$$

จะได้

$$S = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$$

$$S^{-1} = \frac{1}{(4)(9) - (-3)(-3)} \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 0.3333 & 0.1111 \\ 0.1111 & 0.1481 \end{bmatrix}$$

ก. ใช้การแจกแจงแบบโฮเทลลิ่งทีกำลังสอง

คำนวณค่าสถิติทดสอบ

$$T^2 = n(\overline{X} - \mu_0)' S^{-1} (\overline{X} - \mu_0)$$

$$= 3\begin{bmatrix} 9-10 \\ 7-6 \end{bmatrix}' \begin{bmatrix} 0.3333 & 0.1111 \\ 0.1111 & 0.1481 \end{bmatrix} \begin{bmatrix} 9-10 \\ 7-6 \end{bmatrix}$$

$$= 3\begin{bmatrix} -1 \\ 1 \end{bmatrix}' \begin{bmatrix} 0.3333 & 0.1111 \\ 0.1111 & 0.1481 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= 3\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 0.3333 & 0.1111 \\ 0.1111 & 0.1481 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -3 & 3 \end{bmatrix} \begin{bmatrix} 0.3333 & 0.1111 \\ 0.1111 & 0.1481 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -0.6666 & 0.1110 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= 0.7776$$

หาบริเวณวิกฤต เปิดตาราง

สรุปผล

$$T^2_{\alpha; p,n-1} = T^2_{0.05; 2, 2} = 57.00$$

เนื่องจาก $T^2 = 0.7776$ มีค่าน้อยกว่า 57.00 จึงยอมรับสมมติฐาน H_0
นั่นคือ $\mu = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$

ข. ใช้การแจกแจงแบบเอฟ

1. คำนวณค่าสถิติทดสอบ

$$F = \frac{(n-p)}{(n-1)p} T^2$$

$$= \frac{3-2}{(3-1)2}(0.7776)$$

$$= \frac{1}{4}\left(0.7776\right) = 0.1944$$

2. หาบริเวณวิกฤต เปิดตาราง $F_{\alpha; p, n-p} = F_{0.05; 2, 1} = 199.5$

3. สรุปผล เนื่องจาก F = 0.1944 มีค่าน้อยกว่า 199.5 จึงยอมรับสมมติฐาน H_0 นั่นคือ $\mu = \begin{bmatrix} 10 \\ 6 \end{bmatrix}$

การทดสอบสมมติฐานเวคเตอร์ค่าเฉลี่ยประชากร 2 กลุ่ม ($\underline{\mu}$) ตัวแปรหลายตัว

1. การทดสอบสมมติฐานผลต่างของเวกเตอร์ค่าเฉลี่ยประชากร 2 กลุ่ม เมื่อสุ่มตัวอย่างเป็นอิสระกัน

ถ้าสุ่มตัวอย่าง $X_{11}, X_{12}, \ldots, X_{1n_1}$ ขนาด n_1 จากประชากรที่ 1 ที่มี p ตัวแปร ซึ่งมีเวคเตอร์ค่าเฉลี่ย $\mathbf{\mu}_1$ และเมตริกซ์ความแปรปรวนร่วม Σ_1 และสุ่มตัวอย่าง $X_{21}, X_{22}, \ldots, X_{2n_2}$ ขนาด n_2 จากประชากรที่ 2 ที่มี p ตัวแปร ซึ่งมีเวคเตอร์ค่าเฉลี่ย $\mathbf{\mu}_2$ และเมตริกซ์ความแปรปรวนร่วม Σ_2 โดยในการสุ่มจากตัวอย่างทั้ง 2 กลุ่ม ถูกสุ่มอย่างเป็นอิสระกัน

ข้อสมมติเบื้องต้น

1. ตัวอย่างทั้ง 2 ชุด ต้องมีขนาดใหญ่

2. ในกรณีที่ขนาดตัวอย่างของทั้ง 2 กลุ่ม มีขนาดเล็ก ประชากรของทั้ง 2 กลุ่มจะต้องมีการแจกแจงแบบปกติหลายตัวแปร

1.1 กรณีเมทริกซ์ค่าความแปรปรวนร่วมเท่ากัน ($\sum_1 = \sum_2$)

ในกรณีที่ไม่ทราบค่าเมทริกซ์ความแปรปรวนร่วมของประชากร $\sum_1 , \sum_2$ แต่ทราบว่าเมทริกซ์ความแปรปรวนร่วมของประชากรทั้ง 2 กลุ่มไม่แตกต่างกัน นั่นคือ $\sum_1 = \sum_2 = \sum_{pooled}$ แต่ไม่ทราบค่า $\sum_{pooled}$ จะประมาณค่า $\sum_{pooled}$ ด้วย S_p โดยที่ S_p เป็นเมทริกซ์ความแปรปรวนร่วมของตัวอย่าง

$$S_p = \frac{\sum_{j=1}^{n_1}\left(X_{1j}-\bar{X}_1\right)\left(X_{1j}-\bar{X}_1\right)' + \sum_{j=1}^{n_2}\left(X_{2j}-\bar{X}_2\right)\left(X_{2j}-\bar{X}_2\right)'}{n_1+n_2-2}$$

$$= \frac{(n_1 - 1)S_1}{n_1 + n_2 - 2} + \frac{(n_2 - 1)S_2}{n_1 + n_2 - 2}$$

โดยที่ S_1 แทน เมทริกซ์ความแปรปรวนร่วมของตัวอย่างในกลุ่มที่ 1

S_2 แทน เมทริกซ์ความแปรปรวนร่วมของตัวอย่างในกลุ่มที่ 2

P แทน จำนวนตัวแปร

โดยมีขั้นตอนการทดสอบสมมติฐาน ดังนี้

1. กำหนดสมมติฐาน

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

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

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

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

$$= \frac{1}{\frac{1}{n_1} + \frac{1}{n_2}} \left(\overline{X}_1 - \overline{X}_2\right)' S_p^{-1} \left(\overline{X}_1 - \overline{X}_2\right)$$

$$= \frac{n_1 n_2}{n_1 + n_2} \left(\overline{X}_1 - \overline{X}_2\right)' S_p^{-1} \left(\overline{X}_1 - \overline{X}_2\right)$$

$$= \frac{n_1 n_2}{n_1 + n_2} D^2$$ โดยที่ $D^2 = \left(\overline{X}_1 - \overline{X}_2\right)' S_p^{-1} \left(\overline{X}_1 - \overline{X}_2\right)$

3. คำนวณค่าสถิติทดสอบ

4. หาบริเวณวิกฤต

โดยที่ T^2 คือ โฮเทลลิ่งทีกำลังสอง (Hotelling's T^2) ที่มีองศาความเป็นอิสระ (df) เท่ากับ $p , n_1 + n_2 - 2$

5. สรุปผล

จะปฏิเสธสมมติฐาน H_0 เมื่อ $T^2 > T^2_{\alpha; p , n_1 + n_2 - 2}$

ความสัมพันธ์ระหว่าง Hotelling's T^2 กับ F

ถ้าสุ่มตัวอย่าง $X_{11}, X_{12}, \ldots, X_{1n_1}$ จากประชากรที่มีการแจกแจงแบบปกติตัวแปรหลายตัว $N_p(\mu_1 , \Sigma)$ และ $X_{21}, X_{22}, \ldots, X_{2n_1}$ ขนาด n_2 จากประชากรที่มีการแจกแจงแบบปกติตัวแปรหลายตัว $N_p(\mu_2 , \Sigma)$ โดย $X_{11}, X_{12}, \ldots, X_{1n_1}$ และ $X_{21}, X_{22}, \ldots, X_{2n_1}$ เป็นอิสระกัน

สถิติทดสอบ

ภายใต้สมมติฐาน $H_0 : \underline{\mu}_1 = \underline{\mu}_2$ หรือ $H_0 : \underline{\mu}_1 - \underline{\mu}_2 = 0$ จะทำให้ T^2 กลายเป็น

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

โดยที่ $F_{p, n_1 + n_2 - p - 1}$ มีการแจกแจงแบบเอฟ ที่จำนวนองศาความเป็นอิสระ p และ $n_1 + n_2 - p - 1$ ตามลำดับ

การสรุปผล

จะปฏิเสธสมมติฐาน H_0 ถ้า $T^2 > \frac{\left(n_1 + n_2 - 2\right)p}{n_1 + n_2 - p - 1} F_{p, n_1 + n_2 - p - 1}$ หรือพิจารณาจาก

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

จะได้สถิติทดสอบ

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

การสรุปผล

จะปฏิเสธสมมติฐาน H_0 ถ้า $F > F_{\alpha\,;\,p, n_1 + n_2 - p - 1}$

ตัวอย่างที่ 3 สุ่มตัวอย่างบ้านที่มีเครื่องปรับอากาศมา 45 หลัง และบ้านที่ไม่มีเครื่องปรับอากาศมา 55 หลัง ทำการวัดปริมาณการใช้กระแสไฟฟ้า โดยครั้งแรกวัดปริมาณการใช้กระแสไฟฟ้าในน้อยระหว่างเดือนกรกฎาคม และครั้งที่สองวัดปริมาณการใช้กระแสไฟฟ้ามากในระหว่างเดือนเดียวกัน ข้อมูลที่เก็บรวบรวมได้เป็นดังนี้

$$\overline{X}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 13{,}825.3 & 23{,}823.4 \\ 23{,}823.4 & 73{,}107.4 \end{bmatrix}, \quad n_1 = 45$$

$$\overline{X}_2 = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 8{,}632.0 & 19{,}616.7 \\ 19{,}616.7 & 55{,}964.5 \end{bmatrix}, \quad n_2 = 55$$

จงทดสอบว่าปริมาณการใช้กระแสไฟฟ้าน้อยและมากโดยเฉลี่ยแตกต่างกันระหว่างบ้านที่มีและไม่มีเครื่องปรับอากาศหรือไม่ ที่ระดับนัยสำคัญ 0.05 โดยสมมติว่า $\Sigma_1 = \Sigma_2$

ก. ใช้การแจกแจงแบบโฮเทลลิ่งทีกำลังสอง

ข. ใช้การแจกแจงแบบเอฟ

วิธีทำ ก. ใช้การแจกแจงแบบโฮเทลลิ่งทีกำลังสอง

1. กำหนดสมมติฐาน

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

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

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

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

3. คำนวณค่าสถิติทดสอบ

$$S_p = \frac{\left(n_1 - 1\right)S_1}{n_1 + n_2 - 2} + \frac{\left(n_2 - 1\right)S_2}{n_1 + n_2 - 2}$$

$$= \frac{\left(45 - 1\right)\begin{bmatrix} 13{,}825.3 & 23{,}823.4 \\ 23{,}823.4 & 73{,}107.4 \end{bmatrix}}{45 + 55 - 2} + \frac{\left(55 - 1\right)\begin{bmatrix} 86{,}32.0 & 19{,}616.7 \\ 19{,}616.7 & 55{,}964.5 \end{bmatrix}}{45 + 55 - 2}$$

$$= \frac{44\begin{bmatrix} 13{,}825.3 & 23{,}823.4 \\ 23{,}823.4 & 73{,}107.4 \end{bmatrix}}{98} + \frac{54\begin{bmatrix} 86{,}32.0 & 19{,}616.7 \\ 19{,}616.7 & 55{,}964.5 \end{bmatrix}}{98}$$

$$= \frac{\begin{bmatrix} 608{,}313.2 & 1{,}048{,}229.6 \\ 1{,}048{,}229.6 & 3{,}216{,}725.6 \end{bmatrix}}{98} + \frac{\begin{bmatrix} 746{,}566.2 & 1{,}286{,}463.6 \\ 1{,}286{,}463.6 & 3{,}947{,}799.6 \end{bmatrix}}{98}$$

$$= \begin{bmatrix} 6{,}207.278 & 10{,}696.220 \\ 10{,}696.220 & 32{,}823.731 \end{bmatrix} + \begin{bmatrix} 7{,}618.022 & 13{,}127.180 \\ 13{,}127.180 & 40{,}283.669 \end{bmatrix}$$

$$= \begin{bmatrix} 13{,}825.3 & 23{,}823.4 \\ 23{,}823.4 & 73{,}107.4 \end{bmatrix}$$

$$S_p^{-1} = \frac{1}{(13{,}825.3)(73{,}107.4) - (23{,}823.4)(23{,}823.4)} \begin{bmatrix} 73{,}107.4 & -23{,}823.4 \\ -23823.4 & 13{,}825.3 \end{bmatrix}$$

$$= \frac{1}{1{,}010{,}731{,}737.22 - 567{,}554{,}387.56} \begin{bmatrix} 73{,}107.4 & -23{,}823.4 \\ -23823.4 & 13{,}825.3 \end{bmatrix}$$

$$= \frac{1}{443{,}177{,}349.66}\begin{bmatrix} 73{,}107.4 & -23{,}823.4 \\ -23823.4 & 13{,}825.3 \end{bmatrix}$$

$$S_p^{-1} = \begin{bmatrix} 0.00016 & -0.000054 \\ -0.000054 & 0.0000312 \end{bmatrix}$$

$$(\overline{X}_1 - \overline{X}_2) = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix} - \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}$$

$$= \begin{bmatrix} 74.4 \\ 201.6 \end{bmatrix}$$

$$(\overline{X}_1 - \overline{X}_2)' = \begin{bmatrix} 74.4 & 201.6 \end{bmatrix}$$

$$\therefore T^2 = \frac{(45)(55)}{45+55}\begin{bmatrix}74.4 & 201.6\end{bmatrix}\begin{bmatrix}0.00016 & -0.000054 \\ -0.000054 & 0.0000312\end{bmatrix}\begin{bmatrix}74.4 \\ 201.6\end{bmatrix}$$

$$= \frac{2{,}475}{100}\begin{bmatrix}74.4 & 201.6\end{bmatrix}\begin{bmatrix}0.00016 & -0.000054 \\ -0.000054 & 0.0000312\end{bmatrix}\begin{bmatrix}74.4 \\ 201.6\end{bmatrix}$$

$$= 24.75\begin{bmatrix}74.4 & 201.6\end{bmatrix}\begin{bmatrix}0.00016 & -0.000054 \\ -0.000054 & 0.0000312\end{bmatrix}\begin{bmatrix}74.4 \\ 201.6\end{bmatrix}$$

$$= \begin{bmatrix}1841.4 & 4989.6\end{bmatrix}\begin{bmatrix}0.00016 & -0.000054 \\ -0.000054 & 0.0000312\end{bmatrix}\begin{bmatrix}74.4 \\ 201.6\end{bmatrix}$$

$$= \begin{bmatrix}1841.4 & 4989.6\end{bmatrix}\begin{bmatrix}0.0010176 \\ 0.00227232\end{bmatrix}$$

$$= 13.213$$

4. หาบริเวณวิกฤต

เปิดตาราง $T^2_{\alpha;\, p, n_1+n_2-2} = T^2_{0.05\,;\,2,45+55-2} = T^2_{0.05\,;\,2,98} = 6.245$

5. สรุปผล

เนื่องจาก $T^2 = 13.213$ มีค่ามากกว่า 6.245 จึงปฏิเสธสมมติฐาน H_0 นั่นคือ $\underline{\mu}_1 \neq \underline{\mu}_2$

ข. ใช้การแจกแจงแบบเอฟ

จากข้อ ก. $T^2 = 13.213$ จะได้ว่า

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

$$= \frac{45 + 55 - 2 - 1}{(45 + 55 - 2)2}(13.213)$$

$$= \frac{97}{196}(13.213)$$

$$= (0.4949)(13.213)$$

$$= 6.539$$

2. หาบริเวณวิกฤต

เปิดตารางหาบริเวณวิกฤตที่ $F_{\alpha ; p, n_1 + n_2 - p - 1} = F_{0.05 ; 2, 45 + 55 - 2 - 1} = F_{0.05 ; 2, 97} = 3.1007$

3. สรุปผล เนื่องจากสถิติทดสอบ F = 6.539 มีค่ามากกว่า 3.1007 จึงปฏิเสธสมมติฐาน H_0
นั่นคือ $\underline{\mu}_1 \neq \underline{\mu}_2$

ข. ใช้การแจกแจงแบบเอฟ

จากข้อ ก. $T^2 = 13.213$ จะได้ว่า

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

$$= \frac{45 + 55 - 2 - 1}{(45 + 55 - 2)2}(13.213)$$

$$= \frac{97}{196}(13.213)$$

$$= (0.4949)(13.213)$$

$$= 6.539$$

2. หาบริเวณวิกฤต

เปิดตารางหาบริเวณวิกฤตที่ $F_{\alpha ; p, n_1 + n_2 - p - 1} = F_{0.05 ; 2, 45 + 55 - 2 - 1} = F_{0.05 ; 2, 97} = 3.1007$

3. สรุปผล เนื่องจากสถิติทดสอบ F = 6.539 มีค่ามากกว่า 3.1007 จึงปฏิเสธสมมติฐาน H_0 นั่นคือ $\underline{\mu}_1 \neq \underline{\mu}_2$

1.2 กรณีเมทริกซ์ค่าความแปรปรวนร่วมเท่ากัน ($\sum_1 \neq \sum_2$)

ในกรณีที่ไม่ทราบค่าความแปรปรวนของประชากร $\sum_1 , \sum_2$ แต่ทราบว่าความแปรปรวนประชากรไม่เท่ากัน ($\sum_1 \neq \sum_2$) จะต้องทำการประมาณเมทริกซ์ความแปรปรวนร่วมของประชากร ดังนี้

จาก $\mathrm{Cov}\left(\overline{X}_1 - \overline{X}_2\right) = \frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}$

ทำการประมาณโดย $\hat{\mathrm{C}}\mathrm{ov}\left(\overline{X}_1 - \overline{X}_2\right) = \frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}$

ซึ่งถ้าขนาดตัวอย่าง n_1 และ n_2 มีขนาดใหญ่ จากทฤษฎีลิมิตเข้าสู่ส่วนกลาง (Central Limit Theorem : CLT) จะได้ว่า $\overline{X}_1 - \overline{X}_2$ มีการแจกแจงเข้าสู่การแจกแจงแบบปกติหลายตัวแปร

$N_p\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 , \frac{\Sigma_1}{n_1}+\frac{\Sigma_2}{n_2}\right)$ หรือก็คือ $\overline{X}_1 - \overline{X}_2 \sim N_p\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 , \frac{\Sigma_1}{n_1}+\frac{\Sigma_2}{n_2}\right)$

โดยมีขั้นตอนการทดสอบสมมติฐาน ดังนี้

1. กำหนดสมมติฐาน

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

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

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

$$T^2 = \left[\left(\overline{X}_1 - \overline{X}_2\right) - \left(\Sigma_1 - \Sigma_2\right)\right]\left(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}\right)^{-1}\left[\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)\right]$$

ภายใต้สมมติฐาน $H_0 : \underline{\mu}_1 = \underline{\mu}_2$ หรือ $H_0 : \underline{\mu}_1 - \underline{\mu}_2 = 0$ จะทำให้ T^2 กลายเป็น

$$T^2 = \left(\overline{X}_1 - \overline{X}_2\right)'\left(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}\right)^{-1}\left(\overline{X}_1 - \overline{X}_2\right)$$

ซึ่งถ้าไม่ทราบค่า Σ_1 และ Σ_2 ถ้าตัวอย่าง n_1 และ n_2 มีขนาดใหญ่จะประมาณค่า Σ_1 ด้วย S_1 และประมาณค่า Σ_2 ด้วย S_2 ทำให้ T^2 กลายเป็น

$$T^2 = \left(\overline{X}_1 - \overline{X}_2\right)'\left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1}\left(\overline{X}_1 - \overline{X}_2\right)$$

3. คำนวณค่าสถิติทดสอบ

4. หาบริเวณอาณาเขตวิกฤต $\chi^2_{\alpha\,;\,p}$

5. สรุปผล จะปฏิเสธสมมติฐาน H_0 ถ้า $T^2 > \chi^2_{\alpha\,;\,p}$

ตัวอย่างที่ 4 สุ่มตัวอย่างบ้านที่มีเครื่องปรับอากาศมา 45 หลัง และบ้านที่ไม่มีเครื่องปรับอากาศมา 55 หลัง ทำการวัดปริมาณการใช้กระแสไฟฟ้า โดยครั้งแรกวัดปริมาณการใช้กระแสไฟฟ้าในน้อยระหว่างเดือนกรกฎาคม และครั้งที่สองวัดปริมาณการใช้กระแสไฟฟ้ามากในระหว่างเดือนเดียวกัน ข้อมูลที่เก็บรวบรวมได้เป็นดังนี้

$$\overline{X}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix},\quad S_1 = \begin{bmatrix} 13{,}825.3 & 23{,}823.4 \\ 23{,}823.4 & 73{,}107.4 \end{bmatrix},\quad n_1 = 45$$

$$\bar{X}_2 = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, \ S_2 = \begin{bmatrix} 8,632.0 & 19,616.7 \\ 19,616.7 & 55,964.5 \end{bmatrix}, \ n_2 = 55$$

จงทดสอบว่าปริมาณการใช้กระแสไฟฟ้าน้อยและมากโดยเฉลี่ยแตกต่างกันระหว่างบ้านที่มีและไม่มีเครื่องปรับอากาศหรือไม่ ที่ระดับนัยสำคัญ 0.05 โดยสมมติว่า $\Sigma_1 \neq \Sigma_2$

วิธีทำ 1. กำหนดสมมติฐาน

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

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

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

$$T^2 = \left(\bar{X}_1 - \bar{X}_2\right)' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} \left(\bar{X}_1 - \bar{X}_2\right)$$

3. คำนวณค่าสถิติทดสอบ

จาก

$$T^2 = \left(\overline{X}_1 - \overline{X}_2\right)' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} \left(\overline{X}_1 - \overline{X}_2\right)$$

หาค่า

$$\left(\overline{X}_1 - \overline{X}_2\right) = \begin{bmatrix} 204.4 - 130.0 \\ 556.6 - 355.0 \end{bmatrix} = \begin{bmatrix} 74.4 \\ 201.6 \end{bmatrix}$$

$$\left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right) = \frac{\begin{bmatrix} 13{,}825.3 & 23{,}823.4 \\ 23{,}823.4 & 73{,}107.4 \end{bmatrix}}{45} + \frac{\begin{bmatrix} 8{,}632.0 & 19{,}616.7 \\ 19{,}616.7 & 55{,}964.7 \end{bmatrix}}{55}$$

$$= \begin{bmatrix} 307.229 & 529.409 \\ 529.409 & 1{,}624.609 \end{bmatrix} + \begin{bmatrix} 156.945 & 356.667 \\ 356.667 & 1{,}017.54 \end{bmatrix}$$

$$= \begin{bmatrix} 464.174 & 886.076 \\ 886.076 & 2{,}642.149 \end{bmatrix}$$

$$\left(\frac{S_1}{n_1}+\frac{S_2}{n_2}\right)^{-1} = \frac{1}{(464.174)(2642.149)\ -\ (886.076)(886.076)}\begin{bmatrix} 2642.149 & -886.076 \\ -886.076 & 464.174 \end{bmatrix}$$

$$= \frac{1}{1{,}226{,}416.87 - 785{,}130.678}\begin{bmatrix} 2642.149 & -886.076 \\ -886.076 & 464.174 \end{bmatrix}$$

$$= \frac{1}{442{,}286.192}\begin{bmatrix} 2642.149 & -886.076 \\ -886.076 & 464.174 \end{bmatrix}$$

$$= \begin{bmatrix} 0.006 & -0.002 \\ -0.002 & 0.001 \end{bmatrix}$$

$$\therefore\quad T^2 = \begin{bmatrix} 74.4 & 201.6 \end{bmatrix}\begin{bmatrix} 464.174 & 886.076 \\ 886.076 & 2642.149 \end{bmatrix}^{-1}\begin{bmatrix} 74.4 \\ 201.6 \end{bmatrix}$$

$$= \begin{bmatrix} 74.4 & 201.6 \end{bmatrix}\begin{bmatrix} 0.006 & -0.002 \\ -0.002 & 0.001 \end{bmatrix}\begin{bmatrix} 74.4 \\ 201.6 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0432 & 0.0528 \end{bmatrix} \begin{bmatrix} 74.4 \\ 201.6 \end{bmatrix}$$

$$= 13.858$$

4. หาบริเวณอาณาเขตวิกฤต $\chi^2_{\alpha\,;\,p} = \chi^2_{0.05\,;\,2} = 5.99$

5. สรุปผล

เนื่องจากสถิติทดสอบ $T^2 = 13.858$ มีค่ามากกว่า 5.99

จึงปฏิเสธสมมติฐาน H_0 นั่นคือ $\underline{\mu}_1 \neq \underline{\mu}_2$

2. การทดสอบสมมติฐานผลต่างของเวกเตอร์ค่าเฉลี่ยประชากร 2 กลุ่ม เมื่อสุ่มตัวอย่างไม่เป็นอิสระกัน (มีความสัมพันธ์กันหรือเป็นลักษณะแบบจับคู่)

ในกรณีที่สุ่มตัวอย่างแต่ละประชากรอย่างเป็นอิสระกันในบางครั้งหากมีอิทธิพลจากปัจจัยอื่นๆเข้ามาเกี่ยวข้องก็อาจส่งผลให้การสรุปผลการทดลองไม่ถูกต้องได้ จึงต้องมีการควบคุมหรือกำจัดอิทธิพลของปัจจัยอื่น ๆ ก่อนที่จะทำการวัดปัจจัยที่ต้องการเปรียบเทียบ ซึ่งการกำจัดอิทธิพลของปัจจัยอื่นๆอาจทำได้โดยควบคุมลักษณะหรือปัจจัยอื่น ๆ ให้เหมือนกับหรืออาจใช้หน่วยตัวอย่างเดียวกันแต่วัดผลซ้ำ 2 ครั้งและพิจารณาตัวแปรหลาย ๆ ตัวในเวลาเดียวกัน สำหรับขั้นตอนในการทดสอบสมมติฐานเป็นดังนี้

จากรูปแบบของตาราง

คู่ที่	ทรีตเมนต์ที่ 1 X_{1j}	ทรีตเมนต์ที่ 2 X_{2j}	ผลต่าง $d_j = X_{1j} - X_{2j}$
1	X_{11}	X_{21}	$d_1 = X_{11} - X_{21}$
2	X_{12}	X_{22}	$d_2 = X_{12} - X_{22}$
.	.	.	.
.	.	.	.
.	.	.	.
n	X_{1n}	X_{2n}	$d_n = X_{1n} - X_{2n}$

ให้ $\underline{X}_{1j}$ แทน เวกเตอร์ของค่าตัวอย่างในกลุ่มที่ 1 หรือ ทรีตเมนต์ที่ 1 ของคู่ที่ j ; $j = 1, 2, \ldots, n$

$\underline{X}_{2j}$ แทน เวกเตอร์ของค่าตัวอย่างในกลุ่มที่ 2 หรือ ทรีตเมนต์ที่ 2 ของคู่ที่ j ; $j = 1, 2, \ldots, n$

$\underline{d}_j$ แทน เวกเตอร์ผลต่างของทรีตเมนต์ที่ 1 และ ทรีตเมนต์ที่ 2 ของคู่ที่ j

1. กำหนดสมมติฐาน

$H_0 : \mu_d = 0$

$H_1 : \mu_d \neq 0$

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

$$T^2 = \overline{d}'\left(\frac{S_d}{n}\right)^{-1}\overline{d}$$

$$= \frac{\overline{d}' Sd^{-1} \overline{d}}{n^{-1}}$$

$$= n\overline{d}' Sd^{-1} \overline{d}$$

โดยที่ $D^2 = \overline{d}' Sd^{-1} \overline{d}$

$$\therefore T^2 = nD^2$$

3. คำนวณค่าสถิติทดสอบ

4. หาบริเวณอาณาเขตวิกฤตที่ $T^2_{\alpha;\, p,n-1}$

5. สรุปผล

ภายใต้สมมติฐาน H_0 จะพบว่า T^2 มีการแจกแจงแบบโฮเทลลิ่งทีกำลังสองที่องศาความเป็นอิสระ p และ $n-1$ ตามลำดับ โดยจะปฏิเสธสมมติฐานถ้า $T^2 > T^2_{\alpha;\, p,n-1}$

ความสัมพันธ์ระหว่าง Hotelling's T^2 กับ F

จากสถิติทดสอบ $T^2 = n\bar{d}'S_d^{-1}\bar{d} \sim \frac{(n-1)p}{n-p}F_{p,n-p}$ โดยที่ $F_{\alpha;\, p,n-p}$ มีการแจกแจงแบบเอฟที่มีองศาความเป็นอิสระ p และ $n-p$ ตามลำดับ ในการสรุปผลจะปฏิเสธ H_0 ถ้า $T^2 > \frac{(n-1)p}{n-p}F_{\alpha;\, p,n-p}$

หรือพิจารณาจาก

$$T^2 = \frac{(n-1)p}{n-p}F_{p,n-p}$$

$$F = \frac{(n-1)p}{n-p}T^2$$ ในการสรุปผลจะปฏิเสธ H_0 ถ้า $F > F_{\alpha;\, p,n-p}$

ตัวอย่างที่ 5 โรงงานบำบัดน้ำเสียของเทศบาลจำเป็นต้องออกกฎหมายเพื่อเตือนการปล่อยน้ำเสียลงสู่แม่น้ำลำคลอง เพื่อให้ข้อมูลมีความน่าเชื่อถือจึงสุ่มตัวอย่างน้ำและแบ่งออกเป็น 2 ส่วน และส่งไปยังห้องทดลอง 2 แห่ง สำหรับการทดสอบตัวอย่างครึ่งหนึ่งถูกส่งไปยังห้องทดลองของรัฐบาลและตัวอย่างอีกครึ่งหนึ่งถูกส่งไปยังห้องทดลองของเอกชน ทำการวัดค่าปริมาณออกซิเจนเชิงชีวเคมี (BOD) และของแข็งที่แขวนลอยได้ (SS) จำนวน 11 ค่าจากแต่ละห้องทดลอง ได้ข้อมูลดังตาราง

ตัวอย่างที่	ห้องทดลองของเอกชน		ห้องทดลองของรัฐบาล		ผลต่าง	
	X_{1j}	Y_{1j}	X_{2j}	Y_{2j}	$d_1 = X_{1j} - X_{2j}$	$d_2 = Y_{1j} - Y_{2j}$
1	6	27	25	15	-19	12
2	6	23	28	13	-22	10
3	18	64	36	22	-18	42
4	8	44	35	29	-27	15
5	11	30	15	31	-4	-1
6	34	75	44	64	-10	11
7	28	26	42	30	-14	-4
8	71	124	54	64	17	60
9	43	54	34	56	9	-2
10	33	30	29	20	4	10
11	20	14	39	21	-19	-7

จงทดสอบว่าห้องทดลองทั้ง 2 แห่ง วัดค่าได้แตกต่างกันหรือไม่ ที่ระดับนัยสำคัญ 0.05 โดยใช้

ก. ใช้การแจกแจงแบบโฮเทลลิ่งทีกำลังสอง

ข. ใช้การแจกแจงแบบเอฟ

วิธีทำ หาเวกเตอร์ผลต่างค่าเฉลี่ย ดังนี้

$$\bar{d} = \begin{bmatrix} \bar{d}_1 \\ \bar{d}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{n_1}\sum_{j=1}^{n} d_{1j} \\ \frac{1}{n_2}\sum_{j=1}^{n} d_{2j} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{(-19)+(-22)+\ldots+(-19)}{11} \\ \frac{12+10+\ldots+(-7)}{11} \end{bmatrix}$$

$$= \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix}$$

$$S_{11} = \sum_{j=1}^{n} \frac{\left(d_{1j} - \bar{d}_1\right)^2}{n - 1}$$

$$= \frac{\left(-19 + 9.36\right)^2 + \left(-22 + 9.36\right)^2 + \ldots + \left(-19 + 9.36\right)^2}{10}$$

$$= 199.26$$

$$S_{12} = \sum_{j=1}^{n} \frac{\left(d_{1j} - \bar{d}_1\right)\left(d_{2j} - \bar{d}_2\right)}{n - 1}$$

$$= \frac{\left(-19 + 9.36\right)\left(12 + 13.27\right) + \ldots + \left(-19 + 9.36\right)\left(-7 + 13.27\right)}{10}$$

$$= 88.38$$

$$S_{22} = \sum_{j=1}^{n} \frac{\left(d_{2j} - \bar{d}_2\right)^2}{n-1}$$

$$= \frac{(12 - 13.27)^2 + (10 - 13.27)^2 + \ldots + (-7 - 13.27)^2}{10}$$

$$= 418.61$$

จะได้เมทริกซ์ความแปรปรวนร่วมผลต่างของตัวอย่าง คือ

$$S_d = \begin{bmatrix} 199.26 & 88.38 \\ 88.38 & 418.61 \end{bmatrix}$$

$$S_d^{-1} = \frac{1}{(199.26)(418.61)-(88.38)(88.38)} \begin{bmatrix} 418.61 & -88.38 \\ -88.38 & 199.26 \end{bmatrix}$$

$$= \begin{bmatrix} 0.005 & -0.0012 \\ -0.0012 & 0.0036 \end{bmatrix}$$

ก. ใช้การแจกแจงแบบโฮเทลลิ่งทีกำลังสอง

1. กำหนดสมมติฐาน

$H_0 : \mu_d = 0$

$H_1 : \mu_d \neq 0$

2. กำหนดตัวสถิติที่ใช้ในการทดสอบ (Test Statistics)

$$T^2 = n\overline{d}' Sd^{-1}\overline{d}$$

3. คำนวณค่าสถิติทดสอบ

$$T^2 = 11\begin{bmatrix} -9.36 & 13.27 \end{bmatrix}\begin{bmatrix} 0.0055 & -0.0012 \\ -0.0012 & 0.0036 \end{bmatrix}\begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix}$$

$$= 13.60$$

4. หาบริเวณอาณาเขตวิกฤตที่ $T^2_{\alpha; p,n-1} = T^2_{0.05; 2,11-1} = T^2_{0.05; 2,10} = 9.459$

5. สรุปผล เนื่องจากสถิติทดสอบ $T^2 = 13.60$ มีค่ามากกว่า 9.459 จึงปฏิเสธสมมติฐาน H_0 ยอมรับสมมติฐาน H_1 ที่ว่า $\mu_d \neq 0$ นั่นคือ ห้องทดลองทั้งสองแห่งวัดค่าได้แตกต่างกัน ที่ระดับนัยสำคัญ 0.05

ข. ใช้การแจกแจงแบบเอฟ

จากข้อ ก. $T^2 = 13.60$ จะได้ว่า

1. สถิติทดสอบ $F = \dfrac{(n-1)p}{n-p}T^2$

$$= \frac{(11-1)\,2}{11-2}(13.60)$$

$$= \frac{(10)2}{9}(13.60)$$

$$= \frac{20}{9}(13.60)$$

$$= (2.222)(13.60) = 30.219$$

2. หาบริเวณวิกฤต

เปิดตารางหาบริเวณวิกฤตที่

$$F_{\alpha;\, p,n-p} = F_{0.05\,;\,2,11-2} = F_{0.05\,;\,2,9} = 4.26$$

สรุปผล เนื่องจากสถิติทดสอบ F 30.219 มีค่ามากกว่า 4.26 จึงปฏิเสธสมมติฐาน H_0 ยอมรับสมมติฐาน H_1 ที่ว่า $\mu_d \neq 0$ นั่นคือห้องทดลองทั้งสองแห่งวัดค่าได้แตกต่างกัน ที่ระดับนัยสำคัญ 0.05

บทสรุป

ในการศึกษาตัวแปรหลาย ๆ ตัวจากกลุ่มตัวอย่างหรือประชากรเดียวกัน จะมีค่าข้อมูลหลาย ๆ ค่าจากตัวอย่างแต่ละหน่วย หากเราต้องการศึกษาตัวแปร p ตัว การทดสอบค่าเฉลี่ยประชากรครั้งละ 1 ตัวแปร เป็นจำนวน p ครั้ง จะทำให้ความน่าจะเป็นของการเกิดความคลาดเคลื่อนชนิดที่ 1 หรือระดับนัยสำคัญ α เพิ่มขึ้นตามจำนวนครั้งของการทดสอบ ดังนั้นในการทดสอบสมมติฐานจึงควรทำการทดสอบพร้อมกันในรูปของเวคเตอร์ค่าเฉลี่ย โดยในทดสอบสมมติฐานจะสามารถทำได้หลายวิธีขึ้นอยู่กับจำนวนประชากรว่าเป็น 1 กลุ่ม หรือ 2 กลุ่ม ซึ่งในกรณีที่ประชากรมี 2 กลุ่ม ต้องพิจารณาว่าประชากร 2 กลุ่ม เป็นอิสระกัน หรือ ไม่เป็นอิสระกันเสียก่อน