บทที่ 2

สถิติที่ใช้ในการวิเคราะห์ตัวแปรร่วม

การทดสอบสมมติฐานเกี่ยวกับความแตกต่างระหว่างอิทธิพลของทรีตเมนต์ เมื่อค่าสังเกตที่ได้จากการศึกษาอยู่ใต้อิทธิพลของทรีตเมนต์และตัวแปรร่วม โดยการวิเคราะห์ตัวแปรร่วม (Analysis of Covariance) อาจใช้สถิติแบบพาราเมตริก (Parametric Statistics) กับสถิติทดสอบแบบนอนพาราเมตริกซ์ (Nonparametric Statistics) โดยการใช้ F-test ทั้ง 2 วิธี

ในบทนี้จะกล่าวถึงรายละเอียดเกี่ยวกับสถิติทดสอบที่ใช้ในการทดสอบสมมติฐานในการวิเคราะห์โควาเรียนซ์ ของแผนการทดลองแบบสุ่มตลอด (Analysis of Covariance of Completly Randomized Design) ตลอดจนข้อตกลงเบื้องต้นและผลกระทบในการฝ่าฝืน (Violate) ข้อตกลงเบื้องต้น

2.1 สถิติทดสอบพาราเมตริกซ์ (Parametric Statistics)

2.1.1 พาราเมตริกซ์แอนโควา (Parametric ANCOVA) เป็นสถิติทดสอบพาราเมตริกซ์ที่ผู้วิจัยส่วนมากเลือกใช้ และมีข้อจำกัดเกี่ยวกับลักษณะการแจกแจงของประชากร คือ ประชากรต้องมีการแจกแจงแบบปกติ

การวิเคราะห์ความแปรปรวนร่วม เป็นการรวมการวิเคราะห์ความแปรปรวนกับการวิเคราะห์ความถดถอยเข้าด้วยกัน เมื่อมีตัวแปรร่วมที่มากับหน่วยทดลอง ตัวแปรร่วมนี้จะเป็นแหล่งความแปรปรวนที่มิได้ควบคุมด้วยการทดลองและเชื่อว่ามีผลต่อตัวแปรตาม ในการวิเคราะห์ตัวแปรร่วมนี้จึงเป็นการปรับผลการทดลองโดยขจัดส่วนที่มิได้ควบคุมอันเป็นผลจากตัวแปรร่วมออกไป

ตัวแปรของการวิเคราะห์ความแปรปรวนของแผนการทดลองแบบสุ่มตลอด

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \text{.....(2.1)}$$

$$i = 1,2,..., t$$

$$j = 1,2,..., n$$

ในการวิเคราะห์ตัวแปรร่วม ตัวแปรร่วมมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรตาม ซึ่งมีตัวแบบเป็นเส้นถดถอยดังนี้

$$Y_{ij} = \mu + \beta (X_{ij} - \bar{X}_{,,}) + \varepsilon_{ij} \qquad \text{.....(2.2)}$$

$$i = 1,2,..., t$$

$$j = 1,2,..., n$$

ตัวแบบสำหรับการวิเคราะห์ตัวแปรร่วมของแผนการทดลองลุ่มตลอด เกิดจาก 2 โมเดลย่างต้นรวมกันดังนี้

$$Y_{ij} = \mu + \beta (X_{ij} - \bar{X}..) + \tau_i + \varepsilon_{ij} \qquad \text{.....(2.3)}$$

$$i = 1,2,..., t$$

$$j = 1,2,..., n$$

โดยที่ Y_{ij} หมายถึง ค่าสังเกตจากหน่วยทดลองที่ j ได้รับทรีตเมนต์ที่ i

μ หมายถึง ค่าเฉลี่ยของประชากร

β หมายถึง สัมประสิทธิ์ของความถดถอยรวม

$X_{ij} - \bar{X}..$ หมายถึง ค่าเบี่ยงเบนของตัวแปรร่วมที่ (i, j) จากค่าเฉลี่ยทั้งหมดของตัวแปรร่วม

ε_{ij} หมายถึง ค่าความคลาดเคลื่อนของการทดลองจากหน่วยทดลองที่ (i,j)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 2.1 แสดงข้อมูลของการวิเคราะห์ตัวแปรร่วมของแผนการทดลองแบบสุ่มตลอด

TREATMENT(i = 1,2,..., t)							
1		2		3		. . .	t	
Y	X	Y	X	Y	X		Y	X
Y_{11}	X_{11}	Y_{21}	X_{21}	Y_{31}	X_{31}		Y_{t1}	X_{t1}
Y_{12}	X_{12}	Y_{22}	X_{22}	Y_{32}	X_{32}		Y_{t2}	X_{t2}
Y_{13}	X_{13}	Y_{23}	X_{23}	Y_{33}	X_{33}		Y_{t3}	X_{t3}
.		.		.			.	
.		.		.			.	
.		.		.			.	
Y_{1n}	X_{1n}	Y_{2n}	X_{2n}	Y_{3n}	X_{3n}		Y_{tn}	X_{tn}

โดยที่ ε_{ij} ในตัวแบบการวิเคราะห์ตัวแปรร่วมจะน้อยกว่าตัวแบบของการวิเคราะห์ความแปรปรวน เพราะได้มีการขจัดความผันแปรของ Y ที่เกิดจากความแตกต่างในค่าที่เป็น Covariate ซึ่งแสดงได้ดังนี้

จากตัวแบบ (2.2) อาจเขียนเป็นตัวแบบจากตัวอย่างได้

$$\hat{Y}_{ij} = \bar{Y}.. + b(X_{ij} - \bar{X}..)$$

$$Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}.. - b(X_{ij} - \bar{X}..)$$

หรือ $y = bx + e$ .....(2.4)

โดยวิธี Least-Square Method

$$b = \frac{\Sigma xy}{\Sigma x^2}$$

จาก (2.4) $e = y - bx$

$$\Sigma e^2 = \Sigma (y - bx)^2$$

$$= \Sigma y^2 - 2b \Sigma xy + b^2 \Sigma x^2$$

แทนค่า b ได้

$$\Sigma e^2 = \Sigma y^2 - \frac{2\Sigma xy}{\Sigma x^2} \cdot \Sigma xy + \left(\frac{\Sigma xy}{\Sigma x^2}\right)^2 \Sigma x^2$$

$$= \Sigma y^2 - \frac{(\Sigma xy)^2}{\Sigma x^2}$$

เพราะฉะนั้นจะเห็นได้ว่า ปริมาณความแปรปรวนลุ่มคือเป็นผลรวมกำลังสองลดลงเท่ากับ $\frac{(\Sigma xy)^2}{\Sigma x^2}$ ผลรวมกำลังสองที่หักอิทธิพลความถดถอยของ X ออกแล้ว เรียกว่า ผลรวมกำลังสองปรับแล้ว

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 2.2 ตารางการวิเคราะห์ตัวแปรร่วม

S.V.	d.f.	SS and S.C.P			Adjust SS(Y)			F
		XX	XY	YY	d.f.	SS	MS	
1. Bet. tretment	t - 1	B_{XX}	B_{XY}	B_{YY}	t - 1			
2. with tretment	t(r-1)	S_{XX}	S_{XY}	S_{YY}	t(r-1)-1	Sadj	Sadj /t(r-1)-1 = A	$F = \frac{B}{A}$
3. total	rt=N-1	T_{XX}	T_{XY}	T_{YY}	N - 2	Tadj		
Adj Bet. Tret					t-1	Badj	Badj/(t-1) = B	

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สัญลักษณ์ที่ใช้ในการคำนวณ

$$T_{YY} = \Sigma Y^2 - \frac{(\Sigma Y)^2}{N}$$

$$T_{XX} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$T_{XY} = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{N}$$

$$B_{YY} = \Sigma \frac{Y_{i.}^2}{n} - \frac{(\Sigma Y)^2}{N}$$

$$B_{XX} = \Sigma \frac{X_{i.}^2}{n} - \frac{(\Sigma X)^2}{N}$$

$$B_{XY} = \Sigma \frac{X_{i.} Y_{i.}}{n} = \frac{(\Sigma X)(\Sigma Y)}{N}$$

$$S_{YY} = T_{YY} - B_{YY}$$

$$S_{XX} = T_{XY} - B_{XY}$$

$$S_{XY} = T_{XY} - B_{XY}$$

$$Tadj = T_{YY} - \frac{(T_{XY})^2}{T_{XX}}$$

$$Sadj = S_{YY} - \frac{(S_{XY})^2}{S_{XX}}$$

$$Badj = Tadj - Sadj$$

2.1.2 <u>ข้อตกลงเบื้องต้นในการวิเคราะห์ตัวแปรร่วม</u>

เนื่องจากการวิเคราะห์ตัวแปรร่วมเป็นวิธีผสมระหว่างการวิเคราะห์ความแปรปรวนกับการวิเคราะห์ความถดถอย ดังนั้นข้อสมมติของการวิเคราะห์ตัวแปรร่วมร่วมจึงเป็นข้อตกลงเบื้องต้นร่วมกันระหว่างการวิเคราะห์ความแปรปรวน กับการวิเคราะห์ความถดถอย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1. ข้อกำหนดตามปกติของการวิเคราะห์ความแปรปรวนคือ

1.1 อิทธิพลของทรีตเมนต์และสิ่งแวดล้อมอื่น ๆ เป็นแบบบวก (Additive)

1.2 ความคลาดเคลื่อนของการทดลองเกิดขึ้นโดยสุ่มเป็นอิสระต่อกัน และมีลักษณะการแจกแจงแบบปกติมีค่าเฉลี่ยเป็น 0 และมีความแปรปรวนเป็น σ^2

2. ในการวิเคราะห์ความแปรปรวนร่วมมีข้อกำหนดต่อไปอีกคือ

2.1 การถดถอยของตัวแปรตาม Y ที่ขึ้นต่อตัวแปรร่วม X เป็นเส้นตรง

2.2 สัมประสิทธิ์ความถดถอย ในกลุ่มประชากรต้องเป็นเอกพันธ์ (Homogeneous) สำหรับ ทรีตเมนต์ t ระดับ

2.3 ตัวแปรร่วม X เป็นตัวแปรอิสระ ถือว่าไม่ถูกอิทธิพลของทรีตเมนต์ และการวัดไม่มีความคลาดเคลื่อน

เมื่อข้อมูลที่ศึกษาเป็นไปตามข้อตกลงเบื้องต้น การใช้สถิติทดสอบพาราเมตริกซ์ก็จะมีประสิทธิภาพ ในการสรุปผลการวิจัยเป็นไปอย่างถูกต้อง แต่สิ่งที่น่าสนใจก็คือ เมื่อข้อมูลไม่สอดคล้องหรือฝ่าฝืนกับข้อตกลงเบื้องต้น ผลกระทบที่ตามมาดังจะได้กล่าวต่อไป

2.1.3 <u>การฝ่าฝืนข้อตกลงเบื้องต้น</u>

จากหัวข้อ 2.1.2 ได้เน้นถึงความสำคัญของข้อตกลงเบื้องต้น ว่ามีความจำเป็นในการทดสอบสมมติฐาน Cocharan และ Cox (1957) ได้กล่าวว่าการขาดข้อตกลงเบื้องต้นจะมีผลต่อระดับความมีนัยสำคัญของการทดสอบและความไวของตัวทดสอบ เช่น ถ้าทดสอบที่ระดับนัยสำคัญ 0.05 ซึ่งที่จริงอาจจะอยู่ที่ระดับ 0.04 หรือ 0.07 ก็ได้ การขาดความไวก็จะทำให้มีผลต่ออำนาจในการทดสอบลดลง แต่การแจกแจง F นั้น ถ้าขาดข้อตกลงเบื้องต้นบางส่วนก็ทำให้เกิดผลลำเอียงในการทดสอบ (อุทุมพร ทองอุไทย, 2523)

การฝ่าฝืนอิทธิพลของทรีตเมนต์และสิ่งแวดล้อมอื่น ๆ เป็นแบบบวก (Additive) มีผู้ทำการศึกษาหลายท่านพบว่ามีผลกระทบน้อย อาจทำการแก้ไขโดยการแปลงคะแนนเพื่อให้ได้ผลรวมตามที่กำหนด

การฝ่าฝืนความคลาดเคลื่อน (error) เป็นอิสระต่อกัน มีการแจกแจงแบบปกติ และมีความแปรปรวนเท่ากัน ในข้อนี้ ผู้วิจัยหลายท่านชี้ว่า การวิเคราะห์ความแปรปรวนร่วมจะมีความแกร่ง (Robust) เพียงใดขึ้นอยู่กับผลกระทบจากการฝ่าฝืนข้อตกลงเบื้องต้นนี้

การฝ่าฝืน ตัวแปรตามและตัวแปรร่วมมีความสัมพันธ์เชิงเส้นตรง จะทำให้การเปรียบเทียบค่าเฉลี่ยลำเอียง (bias) ได้

การฝ่าฝืนสัมประสิทธิ์ความถดถอยในกลุ่มประชากรต้องเป็นเอกพันธ์ (Homogeneous) มีผู้ศึกษาน้อยเกี่ยวกับการฝ่าฝืนข้อตกลงเบื้องต้นนี้ (Pechman 1968, อ้างโดย Albert R. Wildt, 1978) ได้ศึกษาจากมอนติคาร์โลซิมูเลชั่น พบว่าการแจกแจงของตัวอย่าง (Empirical Sampling distribuation) ของสถิติเอฟ จะแตกต่างกันเพียงเล็กน้อยจากการแจกแจงของทฤษฎี (Theoretical distribution) นอกจากว่าสัมประสิทธิ์ความถดถอยต้องเป็นเอกพันธ์จะถูกฝ่าฝืนอย่างมาก

การฝ่าฝืน ตัวแปรร่วมเป็นตัวแปรอิสระ ผลทดลองตามกำหนด (Fix effect) ไม่ถูกอิทธิพลของทรีตเมนต์ และการวัดไม่มีความคลาดเคลื่อน พบว่าถ้าตัวแปรร่วมไม่เป็นอิสระจากทรีตเมนต์การปรับด้วยเส้นถดถอยคลุมเครือทำให้การเปรียบเทียบทรีตเมนต์มีความคลาดเคลื่อน และถ้าการวัดมีการคลาดเคลื่อน (error in measurement) ก็จะทำให้การวิเคราะห์มีความแม่นยำ (Precision) ต่ำ (Albert R. Wildt, 1978)

2.1.4 <u>การทดสอบสมมติฐาน</u>

สมมติฐานที่ทดสอบ

$H_o : \tau_1 = \tau_2 = \ldots = \tau_t$

H_1 : มีทรีตเมนต์อย่างน้อยที่สุด 2 ทรีตเมนต์ มีความแตกต่างกัน

$$\text{ตัวสถิติ} \quad F = \frac{Badj/t-1}{Sadj/(t(r-1)-1)} \quad ; \quad F(t-1 , tr-t-1)$$

เกณฑ์การตัดสิน

จะปฏิเสธสมมติฐาน H_o เมื่อค่าสถิติเอฟที่คำนวณได้ มีค่ามากกว่าค่าที่ได้จากการเปิดตารางเอฟ ที่องศาอิสระเท่ากับ (t-1) และ (tr-t-1)

2.2 สถิติทดสอบนอนพาราเมตริกซ์ (Nonparametric Statistics)

2.2.1 นอนพาราเมตริกซ์แอนโควา (Nonparametric ANCOVA) เป็นสถิติทดสอบนอนพาราเมตริกซ์ ซึ่งไม่มีข้อจำกัดเกี่ยวกับลักษณะการแจกจงของประชากร Quade ได้ศึกษาและสร้างสถิติทดสอบ ในการเปรียบเทียบทรีตเมนต์ เมื่อค่าสังเกตอยู่ใต้อิทธิพลของทรีต-เมนต์ จึงเรียกสถิติทดสอบนี้ว่า Quade's nonparametric ANCOVA หรือ Rank Analysis of Covariance

หลักการ

จากค่าสังเกตที่ได้จากการศึกษาหรือตัวแปรตาม (y) อยู่ใต้อิทธิพลของทรีต-เมนต์และตัวแปรร่วม (x) จากนั้นจะสร้างตัวแปรใหม่ (z) ได้จากการจัดลำดับ(Rank) ตัวแปรตามและตัวแปรร่วม ซึ่งตัวแปร z นี้จะวิเคราะห์ได้เช่นเดียวกับการวิเคราะห์ความแปร-ปรวนแบบทางเดียว(One-way Analysis of Variance)

วิธีการ

1. ให้ S_{ij} เป็นค่าที่เรียงลำดับของค่าสังเกตหรือตัวแปรตาม Y (รวมทุกกลุ่ม = N) จาก 1 ถึง N จากน้อยไปหามาก ถ้าลำดับเท่ากันจะใช้ลำดับเฉลี่ย

ให้ R_{ij} เป็นค่าที่เรียงลำดับ S_{ij} โดยที่ $\sum_{j=1}^{r}\sum_{i=1}^{t} R_{ij} = 0$

นั่นคือ $R_{ij} = \left[S_{ij} - (N+1)/2\right]$

R เป็นเมตริกซ์ Nx1

013923

2. ให้ T_{ij} เป็นค่าที่เรียงลำดับของตัวแปรร่วม X (รวมทุกกลุ่ม =N) จาก 1 ถึง tr จากน้อยไปหามาก ถ้าลำดับเท่ากันจะใช้ลำดับเฉลี่ย

ให้ C_{ij} เป็นค่าที่เรียงลำดับของ T_{ij} โดยที่ $\sum_{j=1}^{r} \sum_{i=1}^{t} T_{ij} = 0$

นั่นคือ $C_{ij} = \left[T_{ij} - (N+1)/2 \right]$

C เป็นเมตริกซ์ NxP ; P = 1 (ตัวแปรร่วม X 1 ตัวแปร)

3. หาค่า $\hat{R}$

โดย $\hat{R} = C(C\ C)^{-1} CR$

4. สร้างตัวแปรใหม่คือ Z

โดย $Z_{ij} = R_{ij} - \hat{R}_{ij}$

5. การวิเคราะห์จะใช้การวิเคราะห์ความแปรปรวนแบบทางเดียว

(One-Way Analysis of Variance)

$$\text{Variance Ratio (VR)} = \frac{(N-t) \sum_{i} \left(\sum_{j} Z_{ij} \right)^2 / n_i}{(t-1) \sum_{i} \sum_{j} Z_{ij}^2 - \sum_{i} \left(\sum_{j} Z_{ij} \right)^2 / n_i}$$

โดยที่ $\sum_{i} \sum_{j} Z_{ij} = 0$ และ d.f. = (t-1, N-t)

ตารางที่ 2.3 แสดงการวิเคราะห์ข้อมูล

ทรีตเมนต์	Y	X	S	T	R	C	R	Z
	Y_{11}	X_{11}	S_{11}	T_{11}	R_{11}	C_{11}	R_{11}	Z_{11}
	Y_{12}	X_{12}	S_{12}	T_{12}	R_{12}	C_{12}	R_{12}	Z_{12}
1	Y_{13}	X_{13}	S_{13}	T_{13}	R_{13}	C_{13}	R_{13}	Z_{13}
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	Y_{1n}	X_{1n}	S_{1n}	T_{1n}	R_{1n}	C_{1n}	R_{1n}	Z_{1n}
	Y_{21}	X_{21}	S_{21}	T_{21}	R_{21}	C_{21}	R_{21}	Z_{21}
	Y_{22}	X_{22}	S_{22}	T_{22}	R_{22}	C_{22}	R_{22}	Z_{22}
2	Y_{23}	X_{23}	S_{23}	T_{23}	R_{23}	C_{23}	R_{23}	Z_{23}
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	Y_{2n}	X_{2n}	S_{2n}	T_{2n}	R_{2n}	C_{2n}	R_{2n}	Z_{2n}
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	Y_{t1}	X_{t1}	S_{t1}	T_{t1}	R_{t1}	C_{t1}	R_{t1}	Z_{t1}
	Y_{t2}	X_{t2}	S_{t2}	T_{t2}	R_{t2}	C_{t2}	R_{t2}	Z_{t2}
t	Y_{t3}	X_{t3}	S_{t3}	T_{t3}	R_{t3}	C_{t3}	R_{t3}	Z_{t3}
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	.	.	.	.	.	.	.	.
	Y_{tn}	X_{tn}	S_{tn}	T_{tn}	R_{tn}	C_{tn}	R_{tn}	Z_{tn}
SUM	$\Sigma\Sigma Y_{ij}$	$\Sigma\Sigma X_{ij}$			0	0	0	0

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.2.2 การทดสอบสมมติฐาน

สมมติฐานที่ทดสอบ

$H_o : \tau_1 = \tau_2 = \ldots = \tau_t$

H_1 : มีทรีตเมนต์อย่างน้อยที่สุด 2 ทรีตเมนต์มีความแตกต่างกัน

ตัวสถิติ

$$F = \frac{(N-t) \sum_i \left(\sum_j Z_{ij} \right)^2 / n_i}{(t-1) \sum_i \sum_j Z_{ij}^2 - \sum_i \left(\sum_j Z_{ij} \right)^2 / n_i} , F(t-1, N-t)$$

เกณฑ์การตัดสิน

จะปฏิเสธสมมติฐาน H_o เมื่อค่าสถิติเอฟที่คำนวณได้ มีค่ามากกว่าค่า F ที่ได้จากการเปิดตาราง ที่องศาอิสระเท่ากับ (t-1) และ (N-t)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย