บทที่ 1
ความรู้เบื้องต้นเกี่ยวกับตัวแปรพหุ

การวิเคราะห์ตัวแปรพหุมีพื้นฐานแนวคิดมาจากการวิเคราะห์แบบการวิเคราะห์ปัจจัยเดี่ยว (Univariate) และ การวิเคราะห์เชิงสัมพันธ์ (Bivariate) ซึ่งในการทำความเข้าใจเบื้องต้นนั้นควรต้องมีการทำความเข้าใจเกี่ยวกับระบบความสัมพันธ์ และ ความเกี่ยวโยงกันของความแปรปรวนเสียก่อน โดยทั้งสองเรื่องนี้จัดว่าเป็นเรื่องที่มีความเฉพาะเจาะจงแต่มีแนวความคิดที่คล้ายคลึงกัน โดยในการศึกษานั้นผู้ที่ทำการศึกษาจะต้องมีความรู้ทั้งในเรื่องของความแปรปรวน ระดับการวัดหรือมาตราการวัด และ ระดับนัยสำคัญทางสถิติหรือระดับความเชื่อมั่น ซึ่งแต่ละอย่างล้วนเป็นพื้นฐานที่ประยุกต์ใช้กับการวิเคราะห์ตัวแปรพหุทั้งสิ้น

ความแปรปรวน (The Variate)

เป็นการกระทำภายใต้เงื่อนไขที่กำหนด และมีส่วนกระทำร่วมจากหลายสิ่งเร้า หรือหลายตัวแปรอิสระ เพื่อร่วมกันแปรปรวนในตัวแปรตาม โดยจะมีการให้ความสำคัญของค่าน้ำหนัก (Weights) ที่แฝงอยู่รวมกับตัวแปรที่ทำการศึกษา ดังสมการต่อไปนี้

$$\text{ค่าความแปรปรวน} = w_1X_1 + w_2X_2 + w_3X_3 + \ldots\ldots + w_nX_n$$

โดยกำหนดให้ w_1 ถึง w_n คือ ค่าน้ำหนักที่ใช้อธิบายตัวแปรนั้น ๆ ที่ร่วมแปรปรวน และ X_1 ถึง X_n คือ ตัวแปรที่ใช้ในการวิจัย

ผลลัพธ์ที่ได้จากการร่วมกันพยากรณ์ หรือร่วมกันแปรปรวนนั้นจะมีค่าเพียงค่าเดียว และค่าน้ำหนักที่กระทำต่อตัวแปรที่ใช้ในการศึกษารายตัวแปรนั้น โดยส่วนมากจะเรียงลำดับตามความสำคัญ โดย การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) จะให้ความสำคัญของค่าน้ำหนัก (w) ในเรื่องของความสัมพันธ์ ของตัวแปรที่ทำการศึกษากับตัวแปรตาม การวิเคราะห์การจำแนกกลุ่ม (Discriminant Analysis) จะให้ความสำคัญของค่าน้ำหนัก (w) ในเรื่องของคะแนนที่สามารถจำแนกความแตกต่างระหว่างกลุ่ม และในการวิเคราะห์องค์ประกอบ (Factor Analysis) จะให้ความสำคัญของค่าน้ำหนัก (w) ในเรื่องของการพยากรณ์ภายใต้โครงสร้างที่มีความสัมพันธ์กัน เป็นต้น

มาตราการวัด (Measurement Scale)

การแบ่งประเภทของมาตรการวัดของตัวแปรนั้น สามารถแบ่งได้เป็น 2 ประเภท คือ

ประเภทที่ 1 แบ่งโดยการเรียงจากความหยาบไปหาละเอียดสุด สามารถแบ่งได้เป็น 4 มาตราคือ

1. นามบัญญัติ (Nominal Scale)

เป็นระดับที่ใช้จำแนกความแตกต่างของสิ่งที่ต้องการวัดออกเป็นกลุ่ม ๆ โดยใช้ตัวเลข เช่น ตัวแปรเพศ แบ่งออกเป็นกลุ่มเพศชายและกลุ่มเพศหญิง ในการกำหนดตัวเลขอาจจะใช้เลข 1 แทนเพศชาย และเลข 2 แทนเพศหญิง เป็นต้น ตัวเลข 1 หรือ 2 หรือ 3 ที่ใช้แทนกลุ่มต่าง ๆ นั้น ถือเป็นตัวเลขในระดับนามบัญญัติไม่สามารถนำมาบวก ลบ คูณ หาร หรือหาสัดส่วนได้

2. เรียงอันดับ (Ordinal Scale)

เป็นระดับที่ใช้สำหรับจัดอันดับที่หรือตำแหน่งของสิ่งที่ต้องการวัด ตัวเลขในมาตรการวัดระดับนี้เป็นตัวเลขที่บอกความหมายในลักษณะมาก-น้อย สูง-ต่ำ เก่ง-อ่อน กว่ากัน เช่น การประกวดร้องเพลง นางสาวเขียวได้รางวัลที่ 1 นางสาวชมพูได้รางวัลที่ 2 นางสาวเหลืองได้รางวัลที่ 3 เป็นต้น ตัวเลขอันดับที่แตกต่างกันไม่สามารถบ่งบอกถึงปริมาณความแตกต่างได้ เช่น ไม่สามารถบอกได้ว่าผู้ที่ประกวดร้องเพลงได้รางวัลที่ 1 มีความเก่งมากกว่าผู้ที่ได้รางวัลที่ 2 ในปริมาณเท่าใด ตัวเลขในระดับนี้สามารถนำมาบวกหรือลบ กันได้

3. อันตรภาคชั้น (Interval Scale)

เป็นระดับที่สามารถกำหนดค่าตัวเลขโดยมีช่วงห่างระหว่างตัวเลขเท่า ๆ กัน สามารถนำตัวเลขมาเปรียบเทียบกันได้ว่ามีปริมาณมากน้อยเท่าใด แต่ไม่สามารถบอกได้ว่าเป็นกี่เท่าของกันและกัน เพราะมาตราการวัดระดับนี้ไม่มี 0 (ศูนย์) แท้ มีแต่ 0 (ศูนย์) สมมติ เช่น นายวันดีสอบได้ 0 คะแนน ไม่ได้หมายความว่าเขาไม่มีความรู้ เพียงแต่เขาไม่สามารถทำข้อสอบซึ่งเป็นตัวแทนของความรู้ทั้งหมดได้ เป็นต้น ตัวเลขในระดับนี้สามารถนำมาบวก ลบ คูณ หรือหารกันได้

4. มาตราส่วน (Ratio Scale)

เป็นระดับที่สามารถกำหนดค่าตัวเลขให้กับสิ่งที่ต้องการวัด มี 0 (ศูนย์) แท้ เช่น น้ำหนัก ความสูง อายุ เป็นต้น ระดับนี้สามารถนำตัวเลขมาบวก ลบ คูณ หาร หรือหาอัตราส่วนกันได้ คือสามารถบอกได้ว่า ถนนสายหนึ่งยาว 50 กิโลเมตร ยาวเป็น 2 เท่าของถนนอีกสายหนึ่งที่ยาวเพียง 25 กิโลเมตร

ประเภทที่ 2 แบ่งมาตราการวัดของตัวแปรโดยใช้คุณลักษณะเป็นเกณฑ์จะแบ่งได้ 2 มาตรา คือ

1. กลุ่มตัวแปรเชิงคุณภาพจะประกอบด้วย นามบัญญัติ (Nominal Scale) เรียงอันดับ (Ordinal Scale)

2. กลุ่มตัวแปรเชิงปริมาณ จะประกอบด้วย อันตรภาคชั้น (Interval Scale) และมาตราส่วน (Ratio Scale)

โดยมาตราการวัดนี้ไม่ใช่มีไว้เพื่อแบ่งหรือจำแนกประเภทของตัวแปรเท่านั้น แต่จะเป็นการบ่งบอกถึงการเลือกใช้สถิติในการแก้ปัญญาได้อย่างถูกต้องและเหมาะสมกับชุดข้อมูล รวมไปถึงการรายงานผลการวิจัยซึ่งจะทำให้ผลการวิจัยมีความน่าเชื่อถือมากขึ้น โดยการเลือกใช้หรือการให้ความสำคัญของมาตราการวัด มีความสำคัญดังนี้

1. เป็นการกระตุ้นให้นักวิจัยเข้าใจธรรมชาติของตัวแปร และเข้าใจความหมายและธรรมชาติของข้อมูลที่ทำการเก็บรวบรวมมาได้ เช่น เพศ จะพบกันบ่อยที่กำหนด ให้ เพศชาย เป็น 1 และ เพศหญิงเป็น 2 เมื่อพิจารณาจะพบว่า เพศเป็นตัวแปรในมาตราการวัด นามบัญญัติ (Nominal Scale) หากมีการรายงานเป็นค่าเฉลี่ย (Mean) จะทำให้ค่าที่ได้นั้นไม่มีความหมาย เพราะธรรมชาติบ่งชี้ให้เห็นถึงแค่เพศชายหรือหญิง ส่วนการกำหนดตัวเลขนั้นเป็นการแปรข้อมูลในมาตรการวัดเชิงคุณภาพให้เป็นเชิงประมาณ (Dummy Variable)

2. เป็นการใช้เพื่อตระหนักถึงเทคนิคที่จะใช้วิเคราะห์ข้อมูล ที่ทำการเก็บรวมรวมมาว่าจะมีการวิเคราะห์ข้อมูลโดยใช้สถิติอะไร เพื่อความเหมาะสมและความถูกต้องตรงตามคุณลักษณะของตัวแปรที่ทำการศึกษา

ลักษณะของข้อมูลที่ใช้ในการวิเคราะห์หลายตัวแปร

ในการวิเคราะห์หลายตัวแปร (ตัวแปรพหุ) จะเป็นการเก็บรวบรวมข้อมูลจำนวน p ตัวแปร คือ $X_1 , X_2 , \dots , X_2$ โดยที่ $p > 1$ และกำหนดให้ X_{ij} = เป็นตัวแปรที่ i ของหน่วยที่ j ; i = 1, 2, ... , n ; j = 1, 2, ..., p ดังตาราง

ตารางแสดงข้อมูล n หน่วย p ตัวแปร

หน่วยที่	ตัวแปรที่			
	1	2	...	P
1	X_{11}	X_{12}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2p}
.	.	.		.
.	.	.		.
.	.	.		.
i	X_{i1}	X_{i2}	...	X_{ip}
.	.	.		.
.	.	.		.
.	.	.		.
n	X_{n1}	X_{n2}	...	X_{np}

จากตารางสามารถแสดงในรูปเมตริก X ได้ดังนี้

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & & & \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

กรณีที่เก็บรวบรวมข้อมูลจากตัวอย่าง n หน่วย โดยที่แต่ละหน่วยมี p ลักษณะ หรือ p ตัวแปร จะได้เมตริก X เป็นเมตริกของข้อมูลตัวอย่าง ดังนี้

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

กำหนดให้ x เป็นเวคเตอร์ตัวแปร p ตัว โดยที่

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

สำหรับสัญลักษณ์ที่ใช้ในเรื่องการวิเคราะห์หลายตัวแปร จะเป็นดังนี้

ตัวอักษรใหญ่และหนา หมายถึง เมทริกซ์

ตัวอักษรเล็กและหนา หมายถึง เวคเตอร์

ตัวอักษรใหญ่และไม่หนา หมายถึง ตัวแปร

ตัวอักษรเล็กและไม่หนา หมายถึง ค่าของตัวแปร

ลักษณะและคุณสมบัติของตัวแปร

1. ลักษณะและคุณสมบัติของตัวแปร 1 ตัว

ค่าเฉลี่ยของตัวอย่าง

$E(X) = \mu \quad V(X) = \sigma^2$

$E(\overline{X}) = \mu \quad V(\overline{X}) = \frac{\sigma^2}{n}$

2. ลักษณะและคุณสมบัติของตัวแปร 2 ตัว

ความแปรปรวนร่วมของตัวแปร 2 ตัว

กำหนดให้ cov (X,Y) = σ_{XY} เป็นความแปรปรวนร่วมของตัวแปร X และ Y โดยที่

$$\sigma_{XY} = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

$$= \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}$$ กรณีสุ่มมาจากประชากร (Population)

$$S_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{n-1}$$ กรณีสุ่มมาจากตัวอย่าง (Sample)

หมายเหตุ ความแปรปรวนร่วมใช้วัดความสัมพันธ์ในรูปเชิงเส้นเท่านั้น และในกรณีที่ X Y เป็นตัวแปรที่มีการแจกแจงแบบปกติหลายตัวแปร ค่าเฉลี่ยของ Y โดยกำหนด X จะอยู่ในรูปของเส้นตรง

ถ้าตัวแปรทั้ง 2 เป็นอิสระกัน ค่าความแปรปรวนร่วมจะมีค่าเป็น 0 แต่ถ้าความแปรปรวนร่วมเป็น 0 ไม่จำเป็นที่ตัวแปรทั้ง 2 ต้องเป็นอิสระกันเสมอไป

3. ความหมายของความแปรปรวนร่วม

1. ค่าความแปรปรวนร่วมขึ้นอยู่กับหน่วยของข้อมูล

2. ค่าความแปรปรวนร่วมมีค่าเป็นไปได้ทั้งบวกหรือลบ $-\infty < S_{XY}, \sigma_{XY} < \infty$

4. ลักษณะและคุณสมบัติของตัวแปรหลายตัว

กำหนดให้ x เป็นเวคเตอร์ของตัวแปรสุ่ม p ตัว ซึ่งวัดจากหน่วยเดียวกัน โดยที่ N เป็นขนาดของประชากร n เป็นขนาดของตัวอย่าง ถ้าสุ่มตัวอย่างขนาด n หน่วย จะได้เวคเตอร์ของค่าที่ได้จากตัวอย่าง $\mathbf{x_1, x_2, \ldots, x_n}$

$$\mathbf{x}_i = \begin{bmatrix} \mathbf{X}_{i1} \\ \mathbf{X}_{i2} \\ \vdots \\ \mathbf{X}_{ip} \end{bmatrix} ; i=1, 2, \ldots, n$$

เวคเตอร์ค่าเฉลี่ย (Mean Vectors)

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{\mathbf{X}}_1 \\ \bar{\mathbf{X}}_2 \\ \vdots \\ \bar{\mathbf{X}}_p \end{bmatrix}$$

กำหนดให้ X เป็นเมทริกซ์ของข้อมูลตัวอย่างขนาด n ซึ่งมีตัวแปร p ตัว ในที่นี้จะให้แถวนอนแสดงหน่วยของตัวอย่าง และแถวตั้งแทนตัวแปร

$$\mathbf{X} = \begin{bmatrix} \mathbf{X_1'} \\ \mathbf{X_2'} \\ \vdots \\ \mathbf{X_i'} \\ \vdots \\ \mathbf{X_n'} \end{bmatrix}$$

โดยที่ $n>p$; $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, p$

ค่าคาดหวังของ X

$$E(X) = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \mu$$

ในทำนองเดียวกันค่าคาดหวังของ $\overline{X}$

$$E(\overline{X}) = \mu$$

เมทริกซ์ของความแปรปรวนร่วม (Covariance Matrices)

สัญลักษณ์ที่ใช้แทนความแปรปรวนร่วม ได้แก่ $\sum$

$$\sum = Cov(X) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{11} & \sigma_{11} & \cdots & \sigma_{11} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{11} & \sigma_{11} & \cdots & \sigma_{11} \end{bmatrix}_{p \times p}$$

หมายเหตุ ค่าบนเส้นทแยงมุม คือ ค่าความแปรปรวนของประชากรของตัวแปรที่ i และค่านอกเส้นทแยงมุมเป็นค่าความแปรปรวนร่วมประชากรของตัวแปรที่ i และ j ; $I \neq j$

โดยที่ $\sum=[(X-\mu)(X-\mu)']$

เมทริกซ์ของความสัมพันธ์ (Correlation Matrices)

$$\rho=(\rho_{ij})=\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}$$

$$\rho_{ij}=\frac{\sigma_{ij}}{\sigma_i\sigma_j}$$

ระดับนัยสำคัญทางสถิติและอำนาจการทดสอบ (Significant Level and Power of the Test)

ในการทดสอบสมมติฐานทางสถิติจะเกิดความคลาดเคลื่อน 2 ชนิด คือ

1. ความคลาดเคลื่อนชนิดที่ 1 (Type I error) เป็นความคลาดเคลื่อนที่เกิดจากการปฏิเสธสมมติฐาน H_0 เมื่อ H_0 เป็นจริง

ระดับนัยสำคัญหรืออัลฟา (α) เป็นความน่าจะเป็นที่เกิดจากความคลาดเคลื่อนชนิดที่ 1 หรือความน่าจะเป็นในการปฏิเสธสมมติฐาน H_0 เมื่อ H_0 เป็นจริง นั่นคือเป็นการปฏิเสธสิ่งที่ถูกต้อง

$$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 | H_0 \text{ true})$$

ถ้ากำหนดระดับนัยสำคัญให้เท่ากับ 0.05 หรือ 5% แล้วปรากฏว่าข้อสมมติเบื้องต้นบางข้อข้างต้นไม่จริง โดยเฉพาะอย่างยิ่งถ้าข้อสมมติเบื้องต้นข้อที่ 1 ไม่จริง คือข้อมูลไม่ได้มีการแจกแจงแบบปกติตัวแปรหลายตัวแล้ว จะทำให้ระดับนัยสำคัญที่เกิดขึ้นจริงจะมากกว่า 0.05 ทั้งๆที่กำหนดไว้เพียง 0.05 เท่านั้น

2. ความคลาดเคลื่อนชนิดที่ 2 (Type II error) เป็นความคลาดเคลื่อนที่เกิดจากการที่ไม่สามารถปฏิเสธสมมติฐาน H_0 เมื่อ H_0 ไม่เป็นจริง

เบต้า (β) เป็นความน่าจะเป็นที่เกิดจากความคลาดเคลื่อนชนิดที่ 2 หรือความน่าจะเป็นที่ไม่สามารถปฏิเสธสมมติฐาน H_0 เมื่อ H_0 ไม่เป็นจริง นั่นคือยอมรับสิ่งที่ไม่ถูกต้อง

$$\beta = P(\text{Type II error}) = P(\text{accept } H_0 | H_0 \text{ false})$$

อำนาจการทดสอบหรือกำลังการทดสอบ $(1 - \beta)$ เป็นความน่าจะเป็นในการปฏิเสธสมมติฐาน H_0 เมื่อ H_0 ไม่เป็นจริง

$$1 - \beta = P(\text{reject } H_0 | H_0 \text{ false})$$

ถ้าข้อสมมติเบื้องต้นเกี่ยวกับการแจกแจงข้อมูลไม่จริง จะทำให้อำนาจการทดสอบ $(1 - \beta)$ ต่ำ (สายชล สินสมบูรณ์ทอง, 2559, หน้า 59 – 60)

การแจกแจงแบบปกติหลายตัวแปร

การแจกแจงแบบปกติ หรือการแจกแจงแบบเกาส (Normal or Gaussian distribution) สามารถใชไดกับขอมูลสถานการณ และปรากฏการณตาง ๆ ได หลากหลาย เชน ปริมาณผลผลิตการเกษตร น้ำหนัก สวนสูง ค่าใช้จ่าย คะแนนสอบ และยังเป็นข้อตกลงเบื้องต้นของการวิเคราะห์สถิติอนุมานอีกด้วย โดยในการวิเคราะห์ตัวแปรพหุประยุกต์ก็เช่นเดียวกันที่ในการวิเคราะห์จะมีข้อตกลงเบื้องต้นเกี่ยวกับการแจกแจงแบบปกติโดยจะเรียกว่าการแจกแจงแบบปกติหลายตัวแปร ซึ่งการแจกแจงแบบปกติของตัวแปรหลายตัวแปร ก็คือรูปแบบทั่ว ๆ ของโค้งการแจกแจงตัวแปรเดียวที่มีหลายมิตินั่นเอง

ข้อตกลงเบื้องต้นของสถิติที่ใช้พารามิเตอร์สำหรับการวิเคราะห์ตัวแปรพหุประยุกต์

1. สุ่มข้อมูลจากประชากรที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate normal distribution)
2. เมทริกซ์ความแปรปรวนร่วมของทุกกลุ่มต้องเท่ากัน
3. สุ่มข้อมูลอย่างเป็นอิสระกัน

การแจกแจงแบบปกติของหลายตัวแปร (Multivariate normal distribution)

การแจกแจงแบบปกติของหลายตัวแปร โดยส่วนใหญ่จะใช้อธิบายเซตของตัวแปรสุ่มหลาย ๆ ตัวที่มีความสัมพันธ์กัน โดยที่แต่ละค่าของตัวแปรจะมีค่าเกาะกลุ่มอยู่ใกล้ ๆ กับค่าเฉลี่ย ซึ่งถ้าเวคเตอร์ของตัวแปร $\mathbf{x}$ หรือ $x' = \left[X_1, X_2, \ldots, X_p \right]$

มีการแจกแจงแบบปกติที่มีเวคเตอร์ของค่าเฉลี่ยเป็น μ เมตริกความแปรปรวนร่วม (Covariance matrix) เป็น Σ และ $\overline{X}$ เป็นเวคเตอร์ของค่าเฉลี่ยตัวอย่าง และ S เป็นเมทริกซ์ค่าความแปรปรวนร่วมของตัวอย่างจะได้ว่า

1. ฟังก์ชันความน่าจะเป็นของเวคเตอร์ตัวแปร $\mathbf{x}$ คือ

$$f(x) = \frac{1}{\left(\sqrt{2\pi}\right)^{p} |\Sigma|^{1/2}} e^{-\frac{1}{2}(X-\mu)' \Sigma^{-1} (X-\mu)}$$

โดยที่ p เป็นจำนวนตัวแปร หรือเขียนย่อเป็น $X \sim N_P(\mu, \Sigma)$

จากสมการเลขยกกำลังของค่า e คือ ระยะห่างของ X และ μ ในรูปมาตรฐานยกกำลังสอง โดยจะเรียกระยะห่างหรือเรียกว่าระยะทางมาฮาลาโนบิส (Mahalanobis distance)

$$\text{Mahalanobis distance} = \Delta^2 = (X-\mu)' \Sigma^{-1} (X-\mu)$$

2. (n – 1)S มีการแจกแจงแบบวิชาร์ท ที่มี df = n – 1

3. $\overline{X}$ และ S เป็นอิสระกัน

คุณสมบัติของเวคเตอร์ตัวแปรที่มีการแจกแจงแบบปกติ

1. ถ้าเวคเตอร์ตัวแปร **x** มีการแจกแจงแบบปกติหลายตัวแปร $X \sim N_P(\mu, \Sigma)$ จะได้ว่าเซตย่อยของเวคเตอร์ **x** จะมีการแจกแจงแบบปกติหลายตัวแปรด้วย หรือตัวแปร $\mathbf{X}_i$ แต่ละตัวจะมีการแจกแจงแบบปกติที่มีค่าเฉลี่ย μ_i และค่าความแปรปรวน σ_i^2

2. ถ้าตัวแปร X แต่ละตัวมีการแจกแจงแบบปกติ หรือ $X_i \sim N_P(\mu_i, \sigma_i^2)$ แล้วเวคเตอร์ตัวแปร X อาจมีหรือไม่มีการแจกแจงแบบปกติหลายตัวแปรก็ได้

การตรวจสอบเงื่อนไขเกี่ยวกับการแจกแจงแบบปกติหลายตัวแปร

ในการทดสอบการแจกแจงแบบปกติหลายตัวแปรนั้น มีจำนวนวิธีการในการทดสอบไม่กี่วิธี ในที่นี้ขอยกตัวอย่าง 2 วิธี คือ การใช้พล็อตของไคกำลังสอง การใช้ความเบ้และความโด่ง

1. การใช้พล็อตของไคกำลังสอง

วิธีการนี้จะใช้ระยะห่างทั่วไปกำลังสอง (Squared generalized distance) หรือระยะห่าง Mahalanobis $\left(d_j^2\right)$ ของข้อมูลแต่ละค่า โดยที่

$$d_j^2 = \left(X_j - \overline{X}\right)' S^{-1} \left(X_j - \overline{X}\right) ; \; j=1,2,...,n$$

ถ้าประชากรมีการแจกแจงแบบปกติหลายตัวแปร และ $n \geq 25$ หรือ $n - p \geq 30$ จะได้ว่าค่า d_j^2

และค่าไคกำลังสองมาพล็อตกราฟ จึงเรียกกราฟที่ได้ว่าพล็อตของไคกำลังสอง หรือพล็อตของแกมมา

สำหรับขั้นตอนการทดสอบมีดังนี้

1. คำนวณค่า d_j^2 ทุกค่าของ j ; $j = 1, 2, \ldots, n$

2. เรียงลำดับ d_j^2 จากน้อยไปมากได้

$$d_{(1)}^2 \leq d_{(2)}^2 \leq \ldots \leq d_{(n)}^2$$

3. หาค่าไคกำลังสองจากตารางไคกำลังสองโดยหาค่า $\chi_{P,\frac{(j-0.5)}{n}}^2$ โดยที่ $\chi_{P,\frac{(j-0.5)}{n}}^2$ เป็นค่าควอไทล์ที่ $\frac{100(j-0.5)}{n}$ ของการแจกแจงไคกำลังสองที่ $df = p$ โดยที่ $\chi_{P,\frac{(j-0.5)}{n}}^2 = \chi_{P,\frac{(n-j+0.5)}{n}}^2$

4. พล็อตคู่อันดับ $\left(\chi_{P,\frac{(n-j+0.5)}{n}}^2, d_{(j)}^2\right)$

เกณฑ์การพิจารณา

ในการพิจารณาถ้ากราฟอยู่ในรูปเชิงเส้นผ่านจุดกำเนิด ที่มีความชันเป็น 1 จะสามารถสรุปได้ว่า เวคเตอร์ตัวแปร **x** มีการแจกแจงแบบปกติหลายตัวแปร แต่ถ้ากราฟที่ได้ห่างจากเส้นตรงแบบเป็นระบบแล้วจะสรุปได้ว่า **x** ไม่ได้มีการแจกแจงแบบปกติหลายตัวแปร ซึ่งในกรณีที่ข้อมูลไม่ได้มีการแจกแจงแบบปกติหลายตัวแปร หากมีความจำเป็นจะต้องแปลงข้อมูลให้อยู่ในรูปการแจกแจงแบบปกติเสียก่อน โดยอาจใช้การถอดรากที่สอง take log หรือ การ take ln ฯลฯ

วิธีการวิเคราะห์ตัวแปรหลายตัว (Analysis of multivariate)

ในกรณีที่ต้องการวิเคราะห์ตัวแปรตามพร้อมกันหลายตัว ควรใช้การวิเคราะห์แบบ multivariate จะมีความเหมาะสมกว่า การวิเคราะห์แบบ univariate หลายครั้ง ซึ่งวิธีการวิเคราะห์ตัวแปรหลายตัวมีหลายวิธี ดังนี้

1. สถิติทดสอบ Z^2 หรือ Hoteling T^2 เป็นการทดสอบค่าเฉลี่ยของตัวแปรหลายตัวกับค่าคงที่ หรือก็คือการทดสอบเวกเตอร์ค่าเฉลี่ยของตัวแปรหลายตัวกับเวกเตอร์ค่าคงที่นั่นเอง

2. การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการทางสถิติที่ใช้ศึกษาความสัมพันธ์ระหว่างตัวแปรต้นหรือตัวแปรอิสระ (Independent Variable) กับตัวแปรตาม (Dependent Variable ซึ่งเป็นตัวแปรเชิงปริมาณตั้งแต่สองตัวขึ้นไป เพื่อใช้พยากรณ์ค่าตัวแปรของตัวแปรหนึ่งจากตัวแปรอื่นๆ แบ่งเป็น 2 ลักษณะคือ

2.1 การวิเคราะห์ถดถอยอย่างง่าย (Simple Regression Analysis) ใช้ในกรณีที่มีตัวแปรตามหนึ่งตัว และตัวแปรต้นหนึ่งตัว อยู่ในมาตราการวัดระดับช่วงและอัตราส่วน (Interval & Ratio) โดยถ้ามีตัวแปรนามบัญญัติหรือเรียงอันดับ ต้องเปลี่ยนเป็นตัวแปรดัมมี่ (ตัวแปรเทียม) เสียก่อน

2.2 การวิเคราะห์การถดถอยพหุ (Multivariate Regression Analysis) ใช้ในกรณีที่มีตัวแปรตามหนึ่งตัว และ ตัวแปรต้นมีมากกว่าหนึ่งตัว อยู่ในมาตราการวัดระดับช่วงและอัตราส่วน (Interval & Ratio) โดยถ้ามีตัวแปรนามบัญญัติหรือเรียงอันดับ ต้องเปลี่ยนเป็นตัวแปรดัมมี่ (ตัวแปรเทียม) เสียก่อน

3. การวิเคราะห์ความแปรปรวนแบบสองทาง (Two - way Analysis of Variance : Two – way ANOVA) เป็นการเปรียบเทียบความแตกต่างระหว่างตัวแปรต้นหรือตัวแปรอิสระเชิงคุณภาพสองตัว กับตัวแปรตามเชิงปริมาณหนึ่งตัว

4. การวิเคราะห์ความแปรปรวนร่วม (Analysis of Covariance : ANCOVA) เป็นการเปรียบเทียบความแตกต่างระหว่างตัวแปรต้นหรือตัวแปรอิสระเชิงคุณภาพหนึ่งตัว กับตัวแปรตามเชิงปริมาณหนึ่งตัว โดยในการเปรียบเทียบความแตกต่างจะมีการควบคุมตัวแปรเชิงปริมาณอื่น ๆ ที่มีความสัมพันธ์กับตัวแปรตาม ซึ่งตัวแปรเชิงปริมาณที่ถูกควบคุมจะเรียกว่าตัวแปรร่วม (Covariance)

5. การวิเคราะห์ความแปรปรวนตัวแปรหลายตัว (Multivariate Analysis of Variance : MANOVA) เป็นการเปรียบเทียบความแตกต่างระหว่างตัวแปรต้นหรือตัวแปรอิสระเชิงคุณภาพอย่างน้อย 1 ตัว กับตัวแปรตามเชิงปริมาณตั้งแต่ 2 ขึ้นไป

6. การวิเคราะห์ความแปรปรวนร่วมตัวแปรหลายตัว (Multivariate Analysis of Covariance : MACOVA) เป็นการเปรียบเทียบความแตกต่างระหว่างตัวแปรต้นหรือตัวแปรอิสระเชิงคุณภาพอย่างน้อย 1 ตัว กับตัวแปรตามเชิงปริมาณตั้งแต่ 2 ขึ้นไป โดยในการเปรียบเทียบความแตกต่างจะมีการควบคุมตัวแปรเชิงปริมาณอื่น ๆ ที่มีความสัมพันธ์กับตัวแปรตาม

7. เทคนิคเดลฟาย (Delphi Technique) เป็นวิธีสำรวจความเห็นจากกลุ่มผู้เชี่ยวชาญในสาขาหนึ่ง ๆ เพื่อให้ได้คำตอบที่น่าเชื่อถือมากที่สุด โดยให้ผู้เชี่ยวชาญในสาขานั้น ๆ ตอบแบบสอบถามชุดเดียวกันหลายครั้ง ซึ่งในการสำรวจรอบที่หนึ่งผู้ตอบแบบสอบถามจะตอบคำถามพร้อมข้อคิดเห็นส่วนตัวเกี่ยวกับคำถาม จากนั้นคณะวิจัยจะคำนวณหาค่าควอไทล์ (quartile) ของคำตอบและรวบรวมข้อคิดเห็นเพิ่มเติมของผู้ตอบเพิ่มลงในชุดแบบสอบถามรอบที่สอง พร้อมส่งคำตอบที่ได้ในรอบแรกคืนให้ผู้ตอบ ผู้ตอบจะเปรียบเทียบคำตอบของตนกับผู้เชี่ยวชาญท่านอื่นและข้อคิดเห็นเพิ่มเติม เพื่อตัดสินใจใหม่ว่า จะยืนยันความคิดเดิม หรือจะเปลี่ยนใจโดยไม่ต้องเผชิญหน้ากับผู้เชี่ยวชาญท่านอื่น สำหรับขั้นตอนของเทคนิคเดลฟาย จะประกอบด้วย 4 ขั้นตอน ดังนี้

1. กำหนดกรอบในการเก็บข้อมูล โดยรับผิดชอบจะสอบถามความคิดเห็นจากผู้ตัดสินใจ ว่าสนใจข้อมูลอะไร ใช้เวลาประมาณ 3-4 ชม.

2. กำหนดผู้เชี่ยวชาญ คุณสมบัติและขนาดผู้ให้ข้อมูล โดยผู้เชี่ยวชาญต้องมีส่วนร่วมในการแลกเปลี่ยนความคิดเห็น ซึ่งผู้เชี่ยวชาญจะมีขนาด 17 คนขึ้นไป เพื่อให้เกิดความคลาดเคลื่อนที่น้อยที่สุดจนคงที่

3. ขั้นเก็บข้อมูล ส่วนใหญ่ไม่เกิน 4 รอบ แต่ละรอบเตรียมข้อมูลและนำเสนอต่างกัน ดังนี้

รอบที่ 1 เก็บข้อมูลจากผู้เชี่ยวชาญ คำถามปลายเปิด อิสระในความคิดเห็น

รอบที่ 2 จากรอบแรก วิเคราะห์เนื้อหา สรุปประเด็นทั้งหมดทำเป็นแบบสอบถามปลายปิดแบบประมาณค่า นิยม 5 ระดับ คำถามมาจากความคิดเห็นรอบแรกเท่านั้น

รอบที่ 3 นำข้อมูลรอบสอง สร้างแบบสอบถาม มุ่งตรวจสอบความคิดเห็นผู้ให้ข้อมูลซ้ำ เพื่อให้ได้รับข้อมูลย้อนกลับ 2 ส่วน ส่วนแรกความคิดเห็นกลุ่มด้วยค่าสถิติ ส่วนสองคำตอบผู้เชี่ยวชาญ

ในการตรวจสอบความคิดเห็นจะต้องมีการตรวจสอบอย่างน้อยร้อยละ 60 และสามารถเปลี่ยนแปลงคำตอบแต่ละรอบเพิ่ม-ลดไม่น้อยกว่าร้อยละ 15 โดยค่าสัมประสิทธิ์การกระจายหากมีค่ามากกว่าศูนย์แต่ไม่เกิน 0.5 จะได้รับฉันทามติในระดับสูง แต่หากมากกว่า 0.5 จะได้รับฉันทามติระดับต่ำ ควรดำเนินการรอบต่อไป

4. รายงานผล วิเคราะห์ข้อมูลรอบสุดท้าย เสนอผู้เชี่ยวชาญตัดสินใจ

8. การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis)

สายชล สินสมบูรณ์ทอง (2559) สรุปว่า เป็นการสร้างตัวแปรใหม่ซึ่งเป็นฟังก์ชันเชิงเส้นของตัวแปรเดิม ความผันแปรของตัวแปรเดิมจะอยู่ในตัวแปรใหม่ โดยตัวแปรใหม่หรือองค์ประกอบใหม่ที่สร้างขึ้นจะมีความแปรผันของตัวแปรเดิมมากที่สุด ซึ่งเป็นวิธีการลดจำนวนตัวแปรที่ไม่มีการแบ่งกลุ่มมาก่อน และไม่มีการแบ่งว่าตัวแปรใดเป็นตัวแปรตามหรือตัวแปรต้น

บุญอ้อม โฉมที (ม.ป.ป.) สรุปว่า เป็นเทคนิคที่ใช้สำหรับลดจำนวนตัวแปรให้น้อยลง โดยการสร้างเซตของตัวแปรใหม่ซึ่งอยู่ในรูปฟังก์ชันเชิงเส้นของตัวแปรเดิม เซตของตัวแปรใหม่ต้องสามารถอธิบายความผันแปรของข้อมูลให้ได้มากที่สุด โดยใช้จำนวนตัวแปรน้อยกว่าจำนวนตัวแปรเดิม

โดยสรุปการวิเคราะห์องค์ประกอบหลัก หมายถึง เทคนิควิธีทางสถิติที่ใช้สำหรับลดจำนวนตัวแปรให้น้อยลง โดยการสร้างเซตของตัวแปรใหม่ซึ่งเป็นฟังก์ชันเชิงเส้นของตัวแปรเดิม โดยตัวแปรใหม่หรือองค์ประกอบใหม่ที่สร้างขึ้นจะมีความแปรผันของตัวแปรเดิมมากที่สุด

9. การวิเคราะห์ปัจจัย (Factor Analysis)

เพชรน้อย สิงห์ช่างชัย (2549) ให้ความหมายคือ การวิเคราะห์องค์ประกอบเป็นเทคนิคทางสถิติ สาหรับวิเคราะห์ตัวแปรหลายตัว (Multivariate analysis techniques) ที่ออกแบบมาเพื่อช่วยให้นักวิจัยได้ใช้แสวงหาความรู้ความจริงดังกล่าว เช่น นักวิจัยสามารถใช้การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis หรือ EFA) ในการพัฒนาทฤษฎี หรือ นักวิจัยสามารถใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis หรือ CFA) ในการทดสอบหรือยืนยันทฤษฎี

กัลยา วานิชบัญชา (2551) สรุปว่า เป็นการวิเคราะห์หลายตัวแปรเทคนิคหนึ่งเพื่อการสรุปรายละเอียดของตัวแปรหลายตัว หรือเรียกว่าเป็นเทคนิคที่ใช้ในการลดจำนวนตัวแปรเทคนิคหนึ่งโดยการศึกษาถึงโครงสร้างความสัมพันธ์ของตัวแปร และสร้างตัวแปรใหม่เรียกว่า องค์ประกอบ โดยองค์ประกอบที่สร้างขึ้นจะเป็นการนาตัวแปรที่มีความสัมพันธ์กันหรือมีความร่วมกันสูงมารวมกันเป็นองค์ประกอบเดียวกัน ส่วนตัวแปรที่อยู่คนละองค์ประกอบมีความร่วมกันน้อย หรือไม่มีความสัมพันธ์กันเลย

โดยสรุปการวิเคราะห์องค์ประกอบ หมายถึง เทคนิควิธีทางสถิติที่จะจับกลุ่มหรือรวมกลุ่ม หรือรวมตัวแปรที่มีความสัมพันธ์กันไว้ในกลุ่มเดียวกัน ซึ่งความสัมพันธ์เป็นไปได้ทั้งทางบวกและทางลบ ตัวแปรภายในองค์ประกอบเดียวกัน จะมีความสัมพันธ์กันสูง ส่วนตัวแปรที่ต่างองค์ประกอบจะสัมพันธ์กันน้อยหรือไม่มี สามารถใช้ได้ทั้งการพัฒนาทฤษฎีใหม่ หรือการทดสอบหรือยืนยันทฤษฎีเดิม

10. การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นเทคนิคทางสถิติที่ใช้ในการจัดกลุ่มคน/สิ่งของ ออกเป็นกลุ่ม ๆ ที่คล้ายกันตามตัวแปรที่นำมาจัดกลุ่มออกเป็นกลุ่มย่อยอย่างน้อย 2 กลุ่ม โดยมี หลักเกณฑ์ในการจัดกลุ่มคือ “ให้หน่วยที่อยู่ในกลุ่มเดียวกันมีลักษณะที่สนใจเหมือนกันหรือ คล้ายกัน แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีลักษณะที่สนใจต่างกัน”

ซึ่งคำว่าลักษณะที่สนใจอาจจะมีหลาย ๆ ตัวแปร เช่น ถ้าสนใจความคิดเห็นทางด้าน การเมือง จะมีคำถามหลาย ๆ คำถามด้านการเมือง และจะนำคำตอบเหล่านั้นมาแบ่งกลุ่ม (กัลยา วานิชย์บัญชา. 2551 : 286)

11. การวิเคราะห์สหสัมพันธ์คาโนนิคัล (Canonical correlation Analysis) เป็นเทคนิคการวิเคราะห์ตัวแปรพหุคูณวิธีหนึ่ง ซึ่งพัฒนามาจากการวิเคราะห์แบบถดถอยพหุคูณ (Multiple Regression Analysis) ซึ่งในการวิเคราะห์จะไม่มีการแบ่งแยกตัวแปรออกเป็นตัวแปรอิสระและตัวแปรตามในลักษณะเป็นรายตัว แต่จะเป็นการแบ่งตัวแปรทั้งหมดในข้อมูลชุดเดียวกันออกเป็น 2 ชุด คือ ชุดของตัวแปรอิสระหรือตัวแปรทำนายและชุดของตัวแปรตามหรือตัวแปรเกณฑ์ โดยที่จำนวนของตัวแปรแต่ละชุดไม่จำเป็นต้องเท่ากัน ดังนั้นแต่ละชุดของตัวแปรดังกล่าวจึงมีลักษณะเป็นตัวแปรประกอบ ซึ่งหมายถึง การรวมกันของตัวแปรหลายตัว เมื่อศึกษาวิเคราะห์ความสัมพันธ์ของตัวแปรก็พิจารณาไปด้วยกันทั้งกลุ่มในลักษณะของตัวแปรหลายตัวกับตัวแปรหลายตัว ไม่ใช่ลักษณะรายคู่

12. การวิเคราะห์จำแนกประเภท (Discriminant Analysis) เป็นเทคนิคการวิเคราะห์ตัวแปรพหุคูณวิธีหนึ่งที่ใช้ในการแบ่งกลุ่มคน สัตว์ สิ่งของ ออกเป็นกลุ่มย่อยอย่างน้อย 2 กลุ่ม ซึ่งในการวิเคราะห์จะแบ่งตัวแปรออกเป็นตัวแปรต้น และตัวแปรตาม โดยตัวแปรตามจะมี 1 ตัว ซึ่งเป็นตัวแปรเชิงคุณภาพ สำหรับตัวแปรต้นจะเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพที่แปลงเป็นตัวแปร Dummy อย่างน้อย 1 ตัว ในการวิเคราะห์จำแนกประเภทจะต้องทำการแบ่งกลุ่มมาก่อนว่ามีกี่กลุ่ม และจะต้องทราบว่าหน่วยใดอยู่กลุ่มใดมาก่อน เมื่อทำการวิเคราะห์จะได้สมการที่แสดงความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามอยู่ในรูปเชิงเส้นเพื่อใช้ในการหาสาเหตุที่ทำให้คนหรือหน่วยต่างๆที่อยู่ต่างกลุ่มกันมีความแตกต่างกัน รวมทั้งใช้สมการในการพยากรณ์กลุ่มให้กับหน่วยใหม่

13. การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis) เป็นเทคนิคการวิเคราะห์ตัวแปรพหุคูณที่คล้ายกับการวิเคราะห์จำแนกประเภท คือ ตัวแปรตามจะมี 1 ตัว ซึ่งเป็นตัวแปรเชิงคุณภาพ สำหรับตัวแปรต้นจะเป็นตัวแปรเชิงปริมาณหรือตัวแปรเชิงคุณภาพที่แปลงเป็นตัวแปร Dummy อย่างน้อย 1 ตัว และในการวิเคราะห์ต้องทำการแบ่งกลุ่มมาก่อนว่ามีกี่กลุ่ม และต้องทราบว่าหน่วยใดอยู่กลุ่มใดมาก่อน แต่จะแตกต่างจากการวิเคราะห์จำแนกประเภทคือสมการที่แสดงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรต้นจะไม่อยู่ในรูปเชิงเส้น แต่จะอยู่ในของสมการเอกซ์โพเนนเชียลที่ใช้พยากรณ์โอกาสที่แต่ละหน่วยหรือหน่วยใหม่จะอยู่ในกลุ่มที่สนใจ

บทสรุป

ในการวิเคราะห์ตัวแปรพหุ ผู้ทำการวิเคราะห์ควรทำความเข้าใจเบื้องต้นเกี่ยวกับระบบความสัมพันธ์ และ ความเกี่ยวโยงกันของความแปรปรวนเสียก่อน โดยทั้งสองเรื่องนี้จัดว่าเป็นเรื่องที่มีความเฉพาะเจาะจงแต่มีแนวความคิดที่คล้ายคลึงกัน โดยในการศึกษานั้นผู้ที่ทำการศึกษาจะต้องมีความรู้ทั้งในเรื่องของความแปรปรวน ระดับการวัดหรือมาตรการวัด ระดับนัยสำคัญทางสถิติหรือระดับความเชื่อมั่น การแจกแจงแบบปกติหลายตัวแปรเสียก่อน ซึ่งแต่ละอย่างล้วนเป็นพื้นฐานที่ประยุกต์ใช้กับการวิเคราะห์ตัวแปรพหุ

The End