บทที่ 2
แนวคิดและทฤษฎี

ในการศึกษาครั้งนี้จะทำการศึกษาการแก้ไขปัญหาข้อมูลตอบสนองที่ได้จากการทดลองนั้นไม่ได้มีการแจกแจงแบบปกติด้วยการแปลงข้อมูล เพื่อหารูปแบบการแปลงข้อมูลตอบสนองที่เหมาะสมสำหรับข้อมูลที่เศษเหลือมีค่า ความเบ้ ความโด่งและความแปรปรวนที่แตกต่างกันให้มีการแจกแจงแบบปกติและยังเป็นไปตามข้อสมมติของการวิเคราะห์ความแปรปรวน

2.1 **แผนแบบการทดลองสุ่มตลอด** (Completely Randomized Design)

แผนแบบการทดลองสุ่มตลอดเป็นแผนแบบที่เหมาะสมกับการทดลองที่มีปัจจัยที่ต้องการศึกษาเพียงอย่างเดียวและหน่วยทดลองที่นำมาศึกษานั้นมีลักษณะไม่แตกต่างกันก่อนที่จะได้รับวิธีการทดลอง ซึ่งแต่ละหน่วยทดลองนั้นจะได้รับวิธีการทดลองที่เป็นไปอย่างสุ่ม

สำหรับแผนแบบการทดลองแบบสุ่มตลอด มีตัวแบบเชิงสถิติ (Statistical Model) ดังนี้

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad ; i = 1,2,\ldots,k \; ; j = 1,2,\ldots,n$$

เมื่อ Y_{ij} = ข้อมูลตอบสนองที่ได้จากการทดลอง จากวิธีทดลองที่ i หน่วยทดลองที่ j

μ = พารามิเตอร์ค่าเฉลี่ยรวม

τ_i = พารามิเตอร์ผลกระทบจากวิธีทดลองที่ i โดยที่ $\sum_{i=1}^{k} \tau_i = 0$

ε_{ij} = ความคลาดเคลื่อนสุ่มของการทดลองจากวิธีทดลองที่ i หน่วยทดลองที่ j

k = จำนวนวิธีทดลอง

n = จำนวนหน่วยทดลองในแต่ละวิธีทดลอง

2.2 **การแจกแจงแลมดาของตูกีร์** (Tukey's Lambda Distribution)[1]

Ramberg และ Schmeiser ได้เสนอวิธีการสร้างตัวแปรสุ่มที่ขึ้นอยู่กับความเบ้ (Skewness: α_3) และความโด่ง (Kurtosis: α_4) โดยตัวแปรสุ่มนี้มีการแจกแจงที่เรียกว่า "การแจกแจงแลมดาของตูกีร์" โดยที่ตัวแปรสุ่มนั้นจะถูกกำหนดจากค่าพารามิเตอร์ 4 ค่า ซึ่งสัมพันธ์กับค่าความเบ้และค่าความโด่ง ดังนี้

$$X = R(p) = \lambda_1 + \left(\left[p^{\lambda_3} - (1-p)^{\lambda_4}\right] / \lambda_2\right) \qquad ; 0 \le p \le 1 \qquad (1)$$

โดยที่ p เป็นเลขสุ่มที่มีค่าระหว่าง 0 และ 1

λ_1 เป็นพารามิเตอร์กำหนดตำแหน่ง (Location Parameter)

---

[1] Ramberg,J,R. *A Probability Distribution and It's Uses Fitting Data.*Technometrics.21(May1979):201

λ_2 เป็นพารามิเตอร์แสดงขนาด (Scale Parameter)

λ_3, λ_4 เป็นพารามิเตอร์แสดงรูปร่าง (Shape Parameter) ซึ่งขึ้นกับค่าความเบ้และค่าความโด่งที่กำหนด ถ้าการแจกแจงเป็นแบบสมมาตร จะได้ว่า $\lambda_3 = \lambda_4$

ดังนั้นฟังก์ชันความหนาแน่นของตัวแปรสุ่ม X ที่ได้จากสมการ (1) คือ

$$\begin{aligned} f(x) &= f[R(p)] \\ &= 1/R'(p) \qquad ;0 \le p \le 1 \qquad (2) \\ &= \lambda_2\left[\lambda_3 p^{\lambda_3-1} + \lambda_4(1-p)^{\lambda_4-1}\right]^{-1} \end{aligned}$$

และสามารถหาค่า $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ เมื่อกำหนดค่าความเบ้และค่าความโด่งต่างๆ ได้จากตาราง Ramberg โดยที่ค่า λ_1, λ_2 เป็นค่าที่ค่าเฉลี่ยเป็นศูนย์และความแปรปรวนเท่ากับ 1 แต่ถ้าค่าเฉลี่ยเป็น μ และความแปรปรวนเป็น σ^2 จะต้องแปลงค่า λ_1, λ_2 จากตารางดังนี้

$$\lambda_1(\mu,\sigma^2) = \lambda_1(0,1)\sigma + \mu$$
$$\lambda_2(\mu,\sigma^2) = \lambda_2(0,1)/\sigma$$

ความเบ้ (Skewness)

เป็นค่าที่ใช้วัดว่าการแจกแจงนั้นมีการแจกแจงแบบสมมาตรหรือไม่ โดยวิธีที่ใช้ในการวัดค่าความเบ้ คือ วิธีโมเมนต์ (Moment) ซึ่งมีสูตรดังนี้

$$\alpha_3 = \frac{\pi_3}{\sqrt{\pi_2^3}}$$

เมื่อ

π_2 คือ โมเมนต์ของค่าเฉลี่ยอันดับที่ 2

π_3 คือ โมเมนต์ของค่าเฉลี่ยอันดับที่ 3

สามารถประมาณค่าความเบ้จากข้อมูลตัวอย่างได้โดยที่

$$\pi_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

$$\pi_3 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n}$$

จะได้ว่า

$$\text{ความเบ้} = \frac{\sqrt{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{3/2}}$$

ถ้าค่าความเบ้จะเข้าใกล้ 0 เมื่อการแจกแจงของข้อมูลนั้นเข้าใกล้สมมาตร ถ้าค่าความเบ้มีค่าเป็นบวก จะมีการแจกแจงแบบเบ้ขวาและ ถ้ามีค่าเป็นลบ จะมีการแจกแจงแบบเบ้ซ้าย

ความโด่ง(Kurtosis)

เป็นค่าที่ใช้แสดงว่ามีค่าจำนวนมากหรือน้อยที่อยู่ใกล้กึ่งกลางของการแจกแจงของข้อมูลนั้นๆ โดยวิธีที่ใช้ในการวัดค่าความโด่ง คือ วิธีโมเมนต์ (Moment) ซึ่งมีสูตรดังนี้

$$\alpha_4 = \frac{\pi_4}{\pi_2^2}$$

เมื่อ

π_2 คือ โมเมนต์ของค่าเฉลี่ยอันดับที่ 2

π_4 คือ โมเมนต์ของค่าเฉลี่ยอันดับที่ 4

สามารถประมาณค่าความโด่งจากข้อมูลตัวอย่างได้โดยที่

$$\pi_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

$$\pi_4 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}$$

จะได้ว่า

$$\text{ความโด่ง} = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{2}}$$

ถ้าค่าความโด่ง น้อยกว่า 3 แสดงว่าการแจงแจงนั้นมีลักษณะแบนราบกว่าการแจกแจงแบบปกติ ถ้าค่าความโด่งมีค่ามากกว่า 3 แสดงว่าการแจกแจงนั้นมีลักษณะที่โด่งกว่าการแจกแจงแบบปกติ และจะมีค่าเข้าใกล้ 3 เมื่อการแจกแจงของข้อมูลนั้นเข้าใกล้การแจกแจงแบบปกติ

ในการศึกษาครั้งนี้ได้ใช้สูตรการประมาณค่าความเบ้ ความโด่ง ดังนี้

$$\text{ความเบ้} = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 / n-1\right)^{3/2}}$$

$$\text{ความโด่ง} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4 n(n+1)/(n-1) - 3\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2}{\left(\sqrt{\sum(x_i - \bar{x})^2 / (n-1)}\right)^4 (n-2)(n-3)}$$

ซึ่งจะได้ผลในการประมาณที่ใกล้เคียงกับการใช้โปรแกรมการวิเคราะห์ข้อมูลทางสถิติ เช่น โปรแกรม MINITAB SPSS SAS และ EXCEL

2.3 การทดสอบการแจกแจง

การวิเคราะห์ความแปรปรวนนั้นมีข้อสมมติ คือ ความคลาดเคลื่อนสุ่มต้องมีการแจกแจงแบบปกติ ดังนั้นจึงต้องทำการทดสอบข้อมูลตอบสนองที่ได้ก่อนที่จะนำมาใช้ในการวิเคราะห์ ซึ่งในที่นี้ทดสอบการแจกแจงของความคลาดเคลื่อนของข้อมูลใช้วิธีของวีคและแชปปีโร (Shapiro-Wilk) ซึ่งมีสมมติฐาน ดังนี้

H_0 : ความคลาดเคลื่อนมีการแจกแจงแบบปกติ

H_1 : ความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ

<u>สถิติทดสอบ</u> คือ

$$W = \frac{\left\{ \sum_{i=1}^{k} a_{n-i+1}\left(x_{(n-i+1)} - x_{(i)}\right) \right\}^2}{\sum_{i=1}^{n} \left(x_{(i)} - \bar{x}\right)^2}$$

เมื่อ n = ขนาดตัวอย่าง

k = จำนวนเต็มที่เล็กที่สุดที่มากกว่าหรือเท่ากับ $\frac{n}{2}$

a_i = ค่าสัมประสิทธิ์ที่ได้จากการเปิดตารางในภาคผนวก เมื่อ $n \leq 50$

$x_{(i)}$ = order sample

<u>เขตปฏิเสธ</u> จะปฏิเสธสมมติฐานว่าง (H_0 : ความคลาดเคลื่อนมีการแจกแจงแบบปกติ) เมื่อค่าสถิติที่คำนวณได้มีค่าน้อยกว่าค่า W ที่ได้จากตารางของ Pearson and Hartley (1972) ที่ขนาดตัวอย่าง n และระดับนัยสำคัญที่ต้องการ

2.4 วิธีตรวจสอบของเลอวีน (Modify Levene's Test)[2]

ใช้ในการทดสอบความเท่ากันของความแปรปรวนของข้อมูลตอบสนองที่ได้จากการทดลองทุกวิธีทดลอง ซึ่งมีสมมติฐาน ดังนี้

H_0 : $\sigma_i^2 = \sigma^2$ เมื่อ i = 1,2,...,k

H_1 : เมื่อมี $\sigma_i^2 \neq \sigma^2$ อย่างน้อย 1 ค่า

<u>สถิติทดสอบ</u> คือ

$$F_L = \frac{MSTR}{MSE}$$

เมื่อ

$$MSTR = \frac{\sum_{i=1}^{k} n_i \left(\bar{d}_{i.} - \bar{d}_{..}\right)^2}{k-1}$$

---

[2] Neter,Kutner,Nachtsheim and Wasserman. <u>Applied Linear Statistical Models</u>. The McGraw-Hill Companies.1996

$$MSE = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(d_{ij} - \bar{d}_{i.}\right)^2}{n-k}$$

โดยที่

$\tilde{y}_i$ คือ ค่ามัธยฐานของค่า y_{ij} ในวิธีทดลองที่ i

$$d_{ij} = \left|y_{ij} - \tilde{y}_i\right|$$

$$\bar{d}_{i.} = \frac{\sum_{j=1}^{n_i} d_{ij}}{n_i}$$

$$\bar{d}_{..} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} d_{ij}}{n}$$

เขตปฏิเสธ ถ้า $F_L > F_{1-\alpha}$ ที่องศาอิสระ $k-1, n-k$ จะสรุปได้ว่ามีอย่างน้อย 1 วิธีการทดลองที่มีความแปรปรวนต่างจากความแปรปรวนของวิธีการทดลองอื่น

2.5 การแปลงข้อมูล

ใช้วิธีการแปลงข้อมูลของ Box และ Cox

$$y' = \begin{cases} \dfrac{y^{\lambda}-1}{\lambda}, \lambda \neq 0 \\ \log y, \quad \lambda = 0 \end{cases}$$

เมื่อ $-2 \leq \lambda \leq 2$ โดยเพิ่มค่า λ ทีละ 0.5

2.6 การวิเคราะห์ความแปรปรวนหรือการทดสอบเอฟสำหรับแผนแบบการทดลองสุ่มตลอดเมื่อปัจจัยทดลองคงที่ (Completely Randomized Design with Fixed Treatment Factor)

ตารางที่ 2.1 ตารางการวิเคราะห์ความแปรปรวนสำหรับแผนแบบการทดลองสุ่มตลอด เมื่อหน่วยทดลองในแต่ละวิธีการทดลองมีจำนวนที่เท่ากัน

สาเหตุของความแปรปรวน	องศาความเป็นอิสระ	ผลรวมกำลังสอง	ผลรวมกำลังสองเฉลี่ย	ค่าเอฟ
วิธีการทดลอง	$(k-1)$	$SSTrt$	$MSTrt$	F
ความคลาดเคลื่อน	$k(n-1)$	SSE	MSE	
ผลรวม	$kn-1$	SST		

โดยที่ $SSTrt = \sum_{i=1}^{k}\sum_{j=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$$

$$MSTrt = \frac{SSTrt}{(k-1)}$$

$$MSE = \frac{SSE}{k(n-1)}$$

$$F = \frac{MSTrt}{MSE}$$

y_{ij} คือ ข้อมูลตอบสนองที่ได้จากการทดลอง จากวิธีทดลองที่ i หน่วยทดลองที่ j

$\bar{y}_{i.}$ คือ ค่าเฉลี่ยของข้อมูลตอบสนองที่ได้รับจากวิธีทดลองที่ i

$\bar{y}_{..}$ คือ ค่าเฉลี่ยของข้อมูลตอบสนองทั้งหมด

สมมติฐานในการทดสอบ

$H_0 : \mu_1 = \mu_2 = ... = \mu$

$H_1 : \mu_i \neq \mu_j$ อย่างน้อย 1 คู่ของ $i \neq j$

หรือ

$H_0 : \tau_1 = \tau_2 = ... = \tau_k = 0$

$H_1 : \tau_i$ อย่างน้อย 1 ค่าที่ไม่เท่ากับ 0

<u>เกณฑ์การตัดสินใจของการทดสอบเอฟ</u>

ในการทดสอบเอฟ จะปฏิเสธสมมติฐานว่าง เมื่อค่า F ที่คำนวณได้มีค่ามากกว่าค่า F ที่ได้จากการเปิดตาราง F ที่องศาความเป็นอิสระ v_1=(k-1) และ v_2=k(n-1) ภายใต้สมมติฐานว่างสามารถเขียนแทนด้วย F[(k-1),k(n-1)] และสำหรับภายใต้สมมติฐานแย้งซึ่งการแจกแจงของเอฟนั้นจะเป็นการแจกแจงแบบเอฟที่ห่างศูนย์กลาง(Non-central F Distribution) ที่มีองศาความเป็นอิสระ v_1=(k-1) และ v_2=k(n-1) และมีพารามิเตอร์ห่างศูนย์กลางคือ λ (Noncentral Parameter)

โดยที่ $$\lambda = \frac{n\sum_{i=1}^{k}\tau_i^2}{\sigma^2}$$

สามารถเขียนแทนด้วย F[(k-1),k(n-1), λ]

หรือ พิจารณาจากค่า p-value

ถ้าค่า p-value มีค่าน้อยกว่าระดับนัยสำคัญ(α) ที่กำหนด จะปฏิเสธสมมติฐานว่าง

ถ้าค่า p-value มีค่ามากกว่าระดับนัยสำคัญ(α) ที่กำหนด จะยอมรับสมมติฐานว่าง