บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีที่เกี่ยวข้อง

ดังได้กล่าวแล้วว่าการสำรวจตัวอย่างเป็นการศึกษาคุณลักษณะต่าง ๆ ที่มีอยู่ในตัวอย่างแล้วนำผลที่ได้ไปอ้างอิงในระดับประชากร ในหัวข้อนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้องกับการสำรวจตัวอย่าง ดังนี้

1. ประชากร (Population)

ประชากร หมายถึง หน่วยตัวอย่างทั้งหมดที่อยู่ภายใต้ขอบเขตการศึกษา ถ้าประชากรนั้นประกอบด้วยหน่วยตัวอย่างเป็นจำนวนหน่วยที่นับได้แน่นอนจะเรียกว่า ประชากรจำกัด (Finite Population) แต่ถ้าประชากรนั้นประกอบไปด้วยหน่วยตัวอย่างเป็นจำนวนหน่วยที่นับไม่ได้แน่นอนจะเรียกว่า ประชากรอนันต์ (Infinite Population) สำหรับการวิจัยครั้งนี้จะศึกษาเฉพาะในส่วนของประชากรจำกัดเท่านั้น

2. ตัวอย่าง (Sample)

ตัวอย่าง หมายถึง ส่วนหนึ่งของประชากรที่สุ่มมาเพื่อใช้เป็นตัวแทนในการศึกษาหารายละเอียดข้อเท็จจริงที่ต้องการเกี่ยวกับประชากร

3. พารามิเตอร์ (Parameter)

พารามิเตอร์ คือ ค่าคงตัวที่แสดงคุณลักษณะของประชากร ดังนั้นพารามิเตอร์จึงเป็นฟังก์ชันของค่าของหน่วยตัวอย่างต่าง ๆ ทั้งหมดในประชากร

ถ้าประชากรประกอบหน่วยตัวอย่าง N หน่วย มีค่าสังเกตคือ $Y_1, Y_2, ..., Y_N$ จากหน่วยที่ $1, 2, ..., N$ แล้ว พารามิเตอร์ที่สำคัญในการสำรวจตัวอย่าง คือ

ยอดรวมประชากร

$$Y = \sum_{i=1}^{N} Y_i$$

ค่าเฉลี่ยประชากร

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

ความแปรปรวนของประชากร

$$S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2$$

โดยที่ N คือขนาดประชากร (Population Size) Y_i คือค่าสังเกตของหน่วยตัวอย่างที่ i เมื่อ $i = 1, 2, 3, ..., N$

4. การชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ (Stratified Sampling)

การชักตัวอย่างแบบแบ่งเป็นชั้นภูมิ หมายถึง การชักตัวอย่างที่มีการแบ่งประชากรออกเป็นส่วนย่อย แต่ละส่วนย่อยจะเรียกว่า ชั้นภูมิ (Stratum) โดยที่หน่วยตัวอย่างแต่ละหน่วยจะถูกจัดให้อยู่ในชั้นภูมิใดชั้นภูมิหนึ่งเพียงชั้นภูมิเดียวเท่านั้น แล้วจึงชักหน่วยตัวอย่างมาจากชั้นภูมิต่าง ๆ โดยวิธีการที่เหมาะสมอย่างเป็นอิสระกัน ถ้าตัวอย่างในแต่ละชั้นภูมิได้มาโดยวิธีการชักตัวอย่างอย่างง่าย จะเรียกว่า การชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิ (Stratified Random Sampling)

ถ้าประชากรประกอบด้วยหน่วยตัวอย่าง N หน่วย ซึ่งถูกแบ่งออกเป็น L ชั้นภูมิ โดยชั้นภูมิต่าง ๆ มีขนาดเป็น $N_1,\ N_2, ...,\ N_L$ ตามลำดับ และชักตัวอย่างด้วยวิธีการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิมาขนาด n โดยแต่ละชั้นภูมิชักตัวอย่างขนาด $n_1,\ n_2, ...,\ n_L$ ตามลำดับ จะได้ว่า $\sum_{h=1}^{L} N_h = N$ และ $\sum_{h=1}^{L} n_h = n$

กำหนดให้ y_{hi} และ x_{hi} เป็นค่าของตัวแปรที่สนใจศึกษาและตัวแปรช่วยจากหน่วยตัวอย่างที่ i ในชั้นภูมิที่ h ของตัวอย่าง และหาค่าเฉลี่ยตัวอย่างของชั้นภูมิที่ h ได้ดังนี้

$$\overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi} \quad \text{และ} \quad \overline{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$$

ที่มีความแปรปรวน

$$Var\left(\overline{y}_h\right) = \frac{\left(1-f_h\right)}{n_h} S_{Yh}^2 \tag{2.1}$$

และ

$$Var\left(\overline{x}_h\right) = \frac{\left(1-f_h\right)}{n_h} S_{Xh}^2 \tag{2.2}$$

ตามลำดับ

และความแปรปรวนร่วมของ $\bar{y}_h$ และ $\bar{x}_h$ เป็นดังนี้

$$Cov\left(\bar{x}_h, \bar{y}_h\right) = \frac{\left(1-f_h\right)}{n_h} S_{XYh} \tag{2.3}$$

พบว่าตัวประมาณค่า $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$ และ $\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h$ เป็นตัวประมาณค่าที่ไม่เอนเอียงของ $\bar{Y}$ และ $\bar{X}$ ตามลำดับ โดยที่มีความแปรปรวนดังนี้

$$Var\left(\bar{y}_{st}\right) = \sum_{h=1}^{L} W_h^2 \frac{\left(1-f_h\right)}{n_h} S_{Yh}^2 \tag{2.4}$$

และ

$$Var\left(\bar{x}_{st}\right) = \sum_{h=1}^{L} W_h^2 \frac{\left(1-f_h\right)}{n_h} S_{Xh}^2 \tag{2.5}$$

(Cochran, 1977)

5. ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

ให้ $\hat{\theta}$ เป็นตัวประมาณค่าของพารามิเตอร์ θ ค่าคลาดเคลื่อนกำลังสองเฉลี่ยของ $\hat{\theta}$ นิยามดังนี้

$$\begin{aligned} MSE\left(\hat{\theta}\right) &= E\left(\hat{\theta}-\theta\right)^2 \\ &= E\left[\hat{\theta}-E\left(\hat{\theta}\right)\right]^2 + \left[E\left(\hat{\theta}\right)-\theta\right]^2 \\ &= Var\left(\hat{\theta}\right) + \left[Bias\right]^2 \end{aligned}$$

ถ้า $Bias = 0$ นั่นคือ $\hat{\theta}$ เป็นตัวประมาณค่าที่ไม่เอนเอียงของ θ แล้ว $MSE(\hat{\theta}) = Var(\hat{\theta})$ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นค่าที่ใช้วัดความถูกต้อง (Accuracy) ของตัวประมาณค่า หรือค่าที่ใช้วัดความผันแปร (Variation) ของตัวประมาณค่าที่เบี่ยงเบนไปจากค่าที่แท้จริงหรือพารามิเตอร์ของประชากร

6. สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

สัมประสิทธิ์สหสัมพันธ์ คือค่าที่ใช้วัดความสัมพันธ์เชิงเส้นระหว่างตัวแปร X กับ Y เขียนแทนด้วย ρ มีค่าตั้งแต่ -1 ถึง 1

สัมประสิทธิ์สหสัมพันธ์ของตัวแปร X_h กับ Y_h เขียนแทนด้วย $\rho_h\left(Y_h, X_h\right)$ หรือ $\rho_h\left(X_h, Y_h\right)$ หรือ ρ_h นิยามดังนี้

$$\rho_h = \frac{Cov(X_h, Y_h)}{\sqrt{Var(X_h)} \cdot \sqrt{Var(Y_h)}}$$

$$= \frac{S_{XYh}}{S_{Xh} S_{Yh}} \tag{2.6}$$

โดยที่ ความแปรปรวนของ Y_h

$$S_{Yh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(Y_{hi} - \bar{Y}_h \right)^2$$

ความแปรปรวนของ X_h

$$S_{Xh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(X_{hi} - \bar{X}_h \right)^2$$

และความแปรปรวนร่วมของ Y_h กับ X_h

$$S_{XYh} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} \left(X_{hi} - \bar{X}_h \right) \left(Y_{hi} - \bar{Y}_h \right)$$

7. สัมประสิทธิ์การแปรผัน (Coefficient of Variation)

สัมประสิทธิ์การแปรผัน คือ อัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานกับค่าเฉลี่ยเลขคณิตของข้อมูลชุดนั้น ๆ แสดงอยู่ในรูป (สุรินทร์ นิยมางกูร, 2546)

$$C_{Xh} = \frac{S_{Xh}}{\bar{X}_h}$$

และ

$$C_{Yh} = \frac{S_{Yh}}{\bar{Y}_h}$$

8. ความเอนเอียงของตัวประมาณค่าอัตราส่วน

โดยปกติการพิจารณาความเอนเอียงของตัวประมาณค่า มักจะพิจารณาอัตราส่วนของความเอนเอียงต่อค่าคลาดเคลื่อนมาตรฐาน (Standard Error หรือ S.E.) ซึ่งจะสังเกตได้ว่าค่าอัตราส่วนระหว่างความเอนเอียงและค่าคลาดเคลื่อนมาตรฐานของตัวประมาณค่า $\hat{R}, \bar{y}_R$ และ $\hat{Y}_R$ จะมีค่าเหมือนกันทั้งหมด ดังนั้นความเอนเอียงของ $\hat{R}$ หาได้ดังนี้ (สุชาดา กีระนันทน์, 2538)

$$\hat{R} - R = \frac{\bar{y} - R\bar{x}}{\bar{x}}$$

9. ตัวประมาณค่าอัตราส่วนแบบแยก

ในการประมาณค่าโดยใช้ตัวประมาณค่าอัตราส่วนแบบแยกนี้ จะประมาณค่า R_h ในชั้นภูมิที่ h เมื่อ $h = 1, 2, 3, \ldots, L$ จาก

$$R_h = \frac{Y_h}{X_h} = \frac{\overline{Y}_h}{\overline{X}_h}$$

เมื่อ R_h คือ อัตราส่วนของยอดรวมประชากรในชั้นภูมิที่ h หรืออัตราส่วนของค่าเฉลี่ยประชากรในแต่ละชั้นภูมิซึ่งเป็นพารามิเตอร์ที่ไม่ทราบค่าตัวหนึ่งโดยทั่วไปแล้วเราประมาณค่า R_h ด้วยตัวประมาณค่า $\hat{R}_h$ (สุรินทร์ นิยมางกูร, 2546) ดังนี้

$$\hat{R}_h = \frac{y_h}{x_h} = \frac{\overline{y}_h}{\overline{x}_h}$$

โดยที่ $y_h = \sum_{i=1}^{n_h} y_{hi}$ คือ ยอดรวมตัวอย่างในชั้นภูมิที่ h ของตัวแปรที่สนใจศึกษา y

$x_h = \sum_{i=1}^{n_h} x_{hi}$ คือ ยอดรวมตัวอย่างในชั้นภูมิที่ h ของตัวแปรช่วย x

$\overline{y}_h = \frac{y_h}{n_h}$ คือ ค่าเฉลี่ยตัวอย่างในชั้นภูมิที่ h ของตัวแปรที่สนใจศึกษา y

$\overline{x}_h = \frac{x_h}{n_h}$ คือ ค่าเฉลี่ยตัวอย่างในชั้นภูมิที่ h ของตัวแปรช่วย x

ตัวประมาณค่าอัตราส่วนแบบแยกดั้งเดิม (Classical Separate Ratio Estimator) ของค่าเฉลี่ยประชากรกำหนด ดังนี้

$$\overline{y}_{RS} = \sum_{h=1}^{L} W_h \hat{R}_h \overline{X}_h \tag{2.7}$$

โดยมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของ $\overline{y}_{RS}$

$$MSE(\overline{y}_{RS}) \cong \sum_{h=1}^{L} W_h^2 \frac{(1-f_h)}{n_h}\left(R_h^2 S_{Xh}^2 - 2R_h S_{XYh} + S_{Yh}^2\right) \tag{2.8}$$

และความเอนเอียง (bias) ของตัวประมาณค่า $\overline{y}_{RS}$ (Singh & Mangat, 1996) คือ

$$b(\overline{y}_{RS}) \cong \sum_{h=1}^{L} W_h \frac{(1-f_h)}{n_h \overline{X}_h}\left(R_h S_{Xh}^2 - S_{XYh}\right) \tag{2.9}$$

10. วิธีอนุกรมเทย์เลอร์ (Taylor's Series Method)

วิธีอนุกรมเทย์เลอร์ คือ วิธีการประยุกต์อนุกรมเทย์เลอร์มาใช้ประมาณค่าฟังก์ชันที่ต่อเนื่องและสามารถหาอนุพันธ์ได้ (Continuously Differentiable Function) โดยงานวิจัยนี้จะใช้อนุกรมเทย์เลอร์ประมาณค่า ดังนี้ (Lohr, 1999)

$$h(\bar{x}_h,\bar{y}_h)\cong h(\bar{X}_h,\bar{Y}_h)+\left.\frac{\partial h(c,d)}{\partial c}\right|_{\bar{X}_h,\bar{Y}_h}(\bar{x}_h-\bar{X}_h)+\left.\frac{\partial h(c,d)}{\partial d}\right|_{\bar{X}_h,\bar{Y}_h}(\bar{y}_h-\bar{Y}_h) \quad (2.10)$$

ซึ่งในที่นี้

$$h(\bar{x}_h,\bar{y}_h)=\hat{R}_{prh}$$

และ

$$h(\bar{X}_h,\bar{Y}_h)=R_h$$

งานวิจัยที่เกี่ยวข้อง

1. ตัวประมาณค่าอัตราส่วนในการชักตัวอย่างอย่างง่าย

1.1 ตัวประมาณค่าอัตราส่วนของค่าเฉลี่ยประชากรเสนอโดย Prasad

Prasad (1989) (อ้างอิงใน Kadilar, Unyazici, & Cingi, 2009)ได้เสนอตัวประมาณค่าอัตราส่วนของค่าเฉลี่ยประชากรในรูป

$$\bar{y}_p=K\frac{\bar{y}}{\bar{x}}\bar{X} \quad (2.11)$$

โดยที่สัมประสิทธิ์ $K=\dfrac{1+\gamma\rho C_Y C_X}{\gamma C_Y^2+1}$ และ $\gamma=\dfrac{1}{n}$

และมีความคลาดเคลื่อนกำลังสองเฉลี่ยเป็น

$$MSE(\bar{y}_p)\cong\gamma\left(R^2S_X^2-2RKS_{XY}+K^2S_Y^2\right)+\bar{Y}^2(K-1)^2$$

$$=\bar{Y}^2\left[\gamma\left(C_X^2-2K\rho C_Y C_X+K^2C_Y^2\right)+(K-1)^2\right] \quad (2.12)$$

เมื่อ ρ คือสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ระหว่างตัวแปร X และ Y

C_X คือสัมประสิทธิ์ความแปรผันของประชากร (Population Coefficient of Variation) ของตัวแปร X

C_Y คือสัมประสิทธิ์ความแปรผันของประชากร (Population Coefficient of Variation) ของตัวแปร Y

1.2 ตัวประมาณค่าอัตราส่วนของค่าเฉลี่ยประชากรเสนอโดย Kadilar and Cingi

Kadilar and Cingi (2004) ได้เสนอตัวประมาณค่าอัตราส่วนของค่าเฉลี่ยประชากรในรูป

$$\bar{y}_{pr} = \frac{\bar{y} + b\left(\bar{X} - \bar{x}\right)}{\bar{x}}\bar{X} = \hat{R}_{pr}\bar{X} \tag{2.13}$$

โดยที่ $b = \frac{s_{xy}}{s_x^2}$

s_{xy} คือความแปรปรวนร่วมของตัวอย่าง (Sample Covariance) ของตัวแปร X และ Y

และ s_x^2 คือความแปรปรวนของตัวอย่าง (Sample Variance) ของตัวแปร X

โดยมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็น

$$MSE\left(\bar{y}_{pr}\right) \cong \frac{\left(1-f\right)}{n}\left(R^2 S_X^2 + S_Y^2\left(1-\rho^2\right)\right) \tag{2.14}$$

2. ตัวประมาณค่าอัตราส่วนในการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิ

ตัวประมาณค่าอัตราส่วนของค่าเฉลี่ยประชากรเสนอโดย Kadilar and Cingi

Kadilar and Cingi (2005) ได้เสนอตัวประมาณค่าอัตราส่วนของค่าเฉลี่ยประชากรในรูป

$$\bar{y}_{stp} = \kappa^* \bar{y}_{RC} \tag{2.17}$$

โดยที่

$$\bar{y}_{RC} = \frac{\bar{y}_{st}}{\bar{x}_{st}}\bar{X}$$

เมื่อ

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \quad \text{และ} \quad \bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h$$

ที่มีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็น

$$MSE\left(\bar{y}_{stp}\right) = \left(\kappa^*\right)^2 \sum_{h=1}^{L} W_h^2 \frac{\left(1-f_h\right)}{n_h}\left(R^2 S_{Xh}^2 - 2RS_{XYh} + S_{Yh}^2\right) + \left(\kappa^* - 1\right)^2 \bar{Y}^2 \tag{2.18}$$

และ $MSE\left(\bar{y}_{slp}\right)$ มีค่าต่ำที่สุด ก็ต่อเมื่อ

$$\kappa^{*}=\frac{\bar{Y}^{2}}{\bar{Y}^{2}+\sum_{h=1}^{L}W_{h}^{2}\frac{\left(1-f_{h}\right)}{n_{h}}\left(R^{2}S_{Xh}^{2}-2RS_{XYh}+S_{Yh}^{2}\right)} \quad \text{โดยที่} \quad 0<\kappa^{*}<1$$