บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การศึกษาวิจัยในปัจจุบันได้นำวิธีการทางสถิติไปประยุกต์ใช้ในหลายขั้นตอน ไม่ว่าจะเป็นการสำรวจตัวอย่าง การวางแผนในการเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูลและการสรุปผล เป็นต้น ซึ่งการสำรวจตัวอย่างเพื่อนำเอาข้อสรุปที่ได้ไปอ้างอิงในระดับประชากรนั้นเป็นการศึกษาคุณลักษณะต่าง ๆ ที่มีอยู่ในตัวอย่างและเชื่อว่าคุณลักษณะต่าง ๆ นั้นถูกถ่ายทอดมาจากประชากร นั่นคือ หากประชากรมีคุณลักษณะเช่นไรตัวอย่างที่ชักมานั้นก็ควรจะมีคุณลักษณะดังกล่าวเช่นเดียวกัน การที่จะนำข้อสรุปที่ได้จากตัวอย่างไปสรุปผลในระดับประชากรนั้นเราต้องทำการอนุมานทางสถิติซึ่งได้แก่ การประมาณค่าพารามิเตอร์และการทดสอบสมมติฐานทางสถิติและในการสำรวจตัวอย่างโดยส่วนใหญ่จะศึกษาในส่วนของการประมาณค่าพารามิเตอร์

เนื่องจากพารามิเตอร์คือค่าคงตัวซึ่งไม่ทราบค่า และในการหาข้อสรุปเกี่ยวกับพารามิเตอร์จะต้องประมาณค่าพารามิเตอร์ดังกล่าวว่ามีค่าเป็นเท่าใดหรือควรมีค่าอยู่ในช่วงใด ในการประมาณค่าพารามิเตอร์เราจะสร้างฟังก์ชันที่เกี่ยวข้องกับตัวแปรสุ่ม ซึ่งเรียกฟังก์ชันดังกล่าวว่า **ตัวประมาณค่า (Estimator)** พารามิเตอร์ที่สำคัญและนิยมประมาณค่า ได้แก่ ค่าเฉลี่ยประชากร ยอดรวมประชากร และสัดส่วนประชากร

ถ้าประชากรประกอบด้วยหน่วยตัวอย่าง N หน่วย ซึ่งถูกแบ่งออกเป็น L ชั้นภูมิ และชักตัวอย่างด้วยวิธีการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิ (Stratified Random Sampling) ให้ n_h เป็นขนาดของตัวอย่างที่ชักตัวอย่างจากประชากรในชั้นภูมิที่ h ที่มีขนาด N_h โดยที่ $\sum_{h=1}^{L} n_h = n$ และ $\sum_{h=1}^{L} N_h = N$

ให้ y เป็นตัวแปรที่สนใจศึกษา (Study variable) และ $y_{h1}, y_{h2}, \ldots, y_{hn_h}$ เป็นค่าของตัวแปร y จากหน่วยที่ $1, 2, \ldots, n_h$ ในชั้นภูมิที่ h

ในการประมาณค่าเฉลี่ยประชากร $\bar{Y}$ เราพบว่าตัวประมาณค่า $\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$ เป็นตัวประมาณค่าตัวหนึ่งที่ใช้ประมาณค่า $\bar{Y}$ ซึ่งตัวประมาณค่านี้มีคุณสมบัติเป็นตัวประมาณค่าที่ไม่เอนเอียง (Unbiased estimator) และมีความแปรปรวนเท่ากับ $Var(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 \frac{(1-f_h)}{n_h} S_h^2$ เมื่อ W_h คือน้ำหนักของชั้นภูมิที่ h มีค่า $W_h = N_h / N$

$\bar{y}_h$ คือค่าเฉลี่ยตัวอย่างของตัวแปรที่สนใจศึกษาในชั้นภูมิที่ h เท่ากับ $(1/n_h)\sum_{i=1}^{n_h} y_{hi}$

f_h คือสัดส่วนตัวอย่างของชั้นภูมิที่ h มีค่า $f_h = n_h/N_h$

S_h^2 คือความแปรปรวนของประชากรของตัวแปรที่สนใจศึกษาในชั้นภูมิที่ h

(Cochran,1977)

อย่างไรก็ตามถ้าทราบสารสนเทศ (Information) ของตัวแปรช่วย (Auxiliary Variable) X ในแต่ละชั้นภูมิ เช่น ทราบค่าเฉลี่ยประชากร $\bar{X}_h$ หรือยอดรวมประชากร X_h เราสามารถใช้สารสนเทศเหล่านี้มาช่วยในการประมาณค่า $\bar{Y}$ หรือ Y เพื่อให้ตัวประมาณค่ามีประสิทธิภาพมากยิ่งขึ้น ถ้าตัวแปรที่สนใจศึกษา Y มีความสัมพันธ์ในรูปเชิงเส้นกับตัวแปรช่วย X แล้วตัวประมาณค่าที่นำตัวแปรช่วยเข้ามาใช้มีอยู่ 2 ลักษณะ คือ ตัวประมาณค่าอัตราส่วนและ ตัวประมาณค่าการถดถอย

สำหรับตัวประมาณค่าอัตราส่วนของค่าเฉลี่ยประชากรแบบดั้งเดิม (Classical Ratio Estimator) ของการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิมี 2 รูปแบบ (Cochran, 1977) คือ ตัวประมาณค่าอัตราส่วนแบบรวม (Combined Ratio Estimator) และตัวประมาณค่าอัตราส่วนแบบแยก (Separate Ratio Estimator) ตัวประมาณค่าอัตราส่วนแบบรวม กำหนดได้ดังนี้

$$\bar{y}_{RC} = \frac{\bar{y}_{st}}{\bar{x}_{st}}\bar{X} \tag{1.1}$$

เมื่อ $\hat{R}_{st} = \frac{\bar{y}_{st}}{\bar{x}_{st}}$

โดยมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE)

$$MSE(\bar{y}_{RC}) \cong \sum_{h=1}^{L} W_h^2 \frac{(1-f_h)}{n_h}\left(R^2 S_{Xh}^2 - 2RS_{XYh} + S_{Yh}^2\right) \tag{1.2}$$

และตัวประมาณค่าอัตราส่วนแบบแยก กำหนดได้ดังนี้

$$\bar{y}_{RS} = \sum_{h=1}^{L} W_h \hat{R}_h \bar{X}_h \tag{1.3}$$

เมื่อ $\hat{R}_h = \frac{\bar{y}_h}{\bar{x}_h}$

โดยมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE(\bar{y}_{RS}) \cong \sum_{h=1}^{L} W_h^2 \frac{(1-f_h)}{n_h}\left(R_h^2 S_{Xh}^2 - 2R_h S_{XYh} + S_{Yh}^2\right) \tag{1.4}$$

สำหรับตัวประมาณค่าการถดถอยของค่าเฉลี่ยประชากรแบบดั้งเดิม (Classical Regression Estimator) กำหนดได้ดังนี้

$$\bar{y}_{lrs} = \sum_{h=1}^{L} W_h \bar{y}_{lrh} \tag{1.5}$$

เมื่อ $\bar{y}_{lrh} = \bar{y}_h - b_h(\bar{X}_h - \bar{x}_h)$

โดยมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE\left(\bar{y}_{lrs}\right) \cong \sum_{h=1}^{L} W_h^2 \frac{\left(1-f_h\right)}{n_h}\left(R_h^2 S_{Xh}^2 - 2R_h S_{XYh} + S_{Yh}^2\right) \tag{1.6}$$

(Cochran, 1977)

โดยที่ $\bar{y}_{RS}$ คือตัวประมาณค่าอัตราส่วนแบบแยกของค่าเฉลี่ยประชากรแบบดั้งเดิม

$\bar{y}_{RC}$ คือตัวประมาณค่าอัตราส่วนแบบรวมของค่าเฉลี่ยประชากรแบบดั้งเดิม

$\bar{y}_{lrs}$ คือตัวประมาณค่าการถดถอยของค่าเฉลี่ยประชากรแบบดั้งเดิม

b_h คือค่าสัมประสิทธิ์การถดถอยภายในชั้นภูมิที่ h

S_X^2 คือความแปรปรวนของตัวแปร X

S_Y^2 คือความแปรปรวนของตัวแปร Y

S_{XY} คือความแปรปรวนร่วมของตัวแปร X และ Y

R_h คืออัตราส่วนระหว่าง $\bar{Y}_h$ และ $\bar{X}_h$ ของชั้นภูมิที่ h

y_{hi} คือค่าของหน่วยตัวอย่างที่ i ในชั้นภูมิที่ h ของตัวแปรที่สนใจศึกษา

x_{hi} คือค่าของหน่วยตัวอย่างที่ i ในชั้นภูมิที่ h ของตัวแปรช่วย

$\bar{x}_{st}$ คือตัวประมาณค่าที่ไม่เอนเอียงของ $\bar{X}$ มีค่าเท่ากับ $\sum_{h=1}^{L} W_h \bar{x}_h$

$\bar{X}_h$ คือค่าเฉลี่ยของประชากรในชั้นภูมิที่ h ของตัวแปรช่วย

$\bar{x}_h$ คือค่าเฉลี่ยตัวอย่างของตัวแปรช่วยในชั้นภูมิที่ h มีค่าเท่ากับ $\left(1/n_h\right)\sum_{i=1}^{n_h} x_{hi}$

สำหรับตัวประมาณค่าอัตราส่วนของค่าเฉลี่ยประชากรในการชักตัวอย่างอย่างง่ายได้มีผู้วิจัยหลายท่านศึกษาไว้ เช่น Kadilar and Cingi (2004) ได้เสนอตัวประมาณค่าอัตราส่วน ซึ่งมีรูปแบบดังนี้

$$\bar{y}_{pr} = \frac{\bar{y} + b\left(\bar{X} - \bar{x}\right)}{\bar{x}} \bar{X} \tag{1.7}$$

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

และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE\left(\bar{y}_{pr}\right) \cong \frac{(1-f)}{n}\left(R^2 S_X^2 + S_Y^2\left(1-\rho^2\right)\right) \tag{1.8}$$

พบว่า $\bar{y}_{pr}$ มีประสิทธิภาพสูงกว่า $\bar{y}_r$ เมื่อ $\rho > \dfrac{2C_X}{C_Y}$, $(\rho > 0)$ และ $\rho < \dfrac{2C_X}{C_Y}$, $(\rho < 0)$

โดยที่ b คือค่าสัมประสิทธิ์การถดถอยของ Y บน X

ρ คือสัมประสิทธิ์สหสัมพันธ์ของตัวแปร Y กับ X

สำหรับการศึกษาเกี่ยวกับตัวประมาณค่าอัตราส่วนแบบรวมในการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมินั้น Kadilar and Cingi (2005) ได้เสนอตัวประมาณค่าอัตราส่วน ดังนี้

$$\bar{y}_{stp} = \kappa^* \bar{y}_{RC} \tag{1.9}$$

เมื่อ

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

และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$$MSE\left(\bar{y}_{stp}\right) = \left(\kappa^*\right)^2 \sum_{h=1}^{L} W_h^2 \frac{(1-f_h)}{n_h}\left(R^2 S_{Xh}^2 - 2RS_{XYh} + S_{Yh}^2\right) + \bar{Y}^2\left(\kappa^* - 1\right)^2 \tag{1.10}$$

เราพบว่า $\bar{y}_{stp}$ มีประสิทธิภาพสูงกว่า $\bar{y}_{RC}$ เมื่อ $\kappa^* > \dfrac{\bar{Y}^2 - v}{\bar{Y}^2 + v}$ โดยที่

$$v = \sum_{h=1}^{L} W_h^2 \frac{(1-f_h)}{n_h}\left(R_h^2 S_{Xh}^2 - 2R_h S_{XYh} + S_{Yh}^2\right)$$

จากการศึกษาดังกล่าวยังไม่พบงานวิจัยใดที่ศึกษาเกี่ยวกับตัวประมาณค่าอัตราส่วนแบบแยกในการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิเพื่อประมาณค่าเฉลี่ยประชากร ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาตัวประมาณค่าอัตราส่วนแบบแยกของค่าเฉลี่ยประชากรในการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิ โดยจะเสนอตัวประมาณค่าซึ่งประยุกต์จากตัวประมาณค่าของ Kadilar and Cingi (2004, 2005)

วัตถุประสงค์ของการวิจัย

1. เพื่อเสนอตัวประมาณค่าอัตราส่วนแบบแยกของค่าเฉลี่ยประชากรในการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิ

2. เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าที่นำเสนอกับตัวประมาณค่าอัตราส่วนแบบแยกของค่าเฉลี่ยประชากรแบบดั้งเดิม

ขอบเขตของการวิจัย

1. การศึกษาครั้งนี้ศึกษาภายใต้วิธีการชักตัวอย่างอย่างง่ายแบบแบ่งเป็นชั้นภูมิเท่านั้น

2. การเปรียบเทียบประสิทธิภาพของตัวประมาณค่าเปรียบเทียบ โดยใช้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของตัวประมาณค่า

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ตัวประมาณค่าสำหรับค่าเฉลี่ยประชากร

2. นำตัวประมาณค่าที่นำเสนอนี้ไปประยุกต์กับข้อมูลจริง เช่น การประมาณค่าจำนวนอาจารย์ต่อนักศึกษา (FTES) และอื่น ๆ เป็นต้น