บทที่ ๒

ทฤษฎีแผนแบบการสำรวจอย่างมีชั้นภูมิ

ในบทนี้ได้ศึกษาถึงแผนแบบการสำรวจอย่างมีชั้นภูมิที่สำคัญ ๆ ๓ แผนแบบ คือ แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวโดยสุ่มตัวอย่างมาชั้นภูมิละหนึ่งหน่วย แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อย และแผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ ในการศึกษาแผนแบบการสำรวจอย่างมีชั้นภูมิทั้ง ๓ แบบนี้ จะศึกษาเกี่ยวกับวัตถุประสงค์ที่สำคัญที่เลือกใช้แผนแบบนั้น ๆ ข้อดี ข้อเสีย วิธีการเลือกตัวอย่าง การประมาณค่าพารามิเตอร์ ซึ่งได้แก่ ค่าเฉลี่ย และความแปรปรวนของค่าเฉลี่ยของแต่ละแผนแบบ ในส่วนสุดท้ายจะกล่าวถึงการเปรียบเทียบประสิทธิภาพระหว่างค่าเฉลี่ยจากตัวอย่างที่เลือกมาจากแผนแบบการสำรวจชนิดต่าง ๆ ที่กล่าวมาแล้วโดยใช้ข้อมูลจากสำมะโนประชากรและเคหะ พ.ศ.๒๕๑๓ ของสำนักงานสถิติแห่งชาติ

๒.๑ <u>แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวโดยสุ่มตัวอย่างมาชั้นภูมิละหนึ่งหน่วย</u>

(One-way stratification with one unit per stratum)

การใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวโดยสุ่มตัวอย่างมาชั้นภูมิละหนึ่งหน่วย ส่วนใหญ่มีจุดประสงค์เพื่อต้องการทราบค่าประมาณทางสถิติ เช่น ค่าเฉลี่ย อย่างคร่าว ๆ โดยใช้จำนวนตัวอย่างให้น้อยที่สุดเพื่อเป็นการประหยัดเวลา และค่าใช้จ่ายในการเก็บรวบรวมข้อมูล และมีความถูกต้องพอประมาณ จุดประสงค์เช่นนี้อาจทำได้โดยการแบ่งประเภทของหน่วยตัวอย่างออกเป็นพวก ๆ พวกที่มีความคล้ายคลึงกันจะจัดอยู่ในชั้นภูมิเดียวกัน แล้วสุ่มตัวอย่างจากแต่ละชั้นภูมิมาชั้นภูมิละ ๑ หน่วยตัวอย่าง (unit) หน่วยตัวอย่างที่สุ่มขึ้นมาจะเป็นตัวแทนของชั้นภูมิได้ดีเพียงไรขึ้นอยู่กับว่าหน่วยตัวอย่างในชั้นภูมินั้นมีความคล้ายคลึงกันมากแค่ไหน ถ้าหน่วยตัวอย่างในชั้นภูมิมีความคล้ายคลึงกันมากหรือมีลักษณะเหมือนกัน หน่วยตัวอย่างที่สุ่มได้ก็จะเป็นตัวแทนที่ดี การประมาณค่าทางสถิติย่อมมีความถูกต้องใกล้เคียงพอสมควร **ในทางตรงกันข้าม** ถ้าหน่วยตัวอย่างในชั้นภูมิมีความแตกต่างกันมาก หน่วยตัวอย่างที่สุ่มได้อาจจะเป็นตัวแทนที่ไม่ดี การประมาณค่าทางสถิติก็ย่อมมีความคลาดเคลื่อนจากค่าจริงมาก

ข้อเสียของการใช้แผนแบบการสำรวจชนิดนี้ก็คือ ไม่สามารถประมาณค่าความแปรปรวนภายในแต่ละชั้นภูมิโดยตรงได้ ดังนั้น จึงไม่สามารถประมาณค่าความแปรปรวนรวมได้ ในกรณีที่แต่ละชั้นภูมิมีขนาดตัวอย่าง (n_i) เพียง ๑ หน่วยนี้ การประมาณค่าความแปรปรวนของค่าเฉลี่ยจากตัวอย่างจะใช้วิธีจัดกลุ่มชั้นภูมิเข้าด้วยกันเป็นกลุ่ม (Collapsed strata) ซึ่งมีหลักในการจัดกลุ่มชั้นภูมิเข้าด้วยกัน ดังนี้

ก) แต่ละชั้นภูมิที่มาจัดคู่เข้าด้วยกัน จะต้องมีจำนวนหน่วยตัวอย่างทั้งหมดใกล้เคียงกัน

ข) แต่ละชั้นภูมิที่มาจัดคู่เข้าด้วยกัน จะต้องมีค่าเฉลี่ยของประชากรในชั้นภูมิไม่ต่างกันมากนัก

ถ้าให้สัญลักษณ์ที่ใช้ในการประมาณค่าพารามิเตอร์เมื่อใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวโดยสุ่มตัวอย่างมาชั้นภูมิละหนึ่งหน่วย เป็นดังนี้

N = จำนวนหน่วยตัวอย่างทั้งหมดที่ประกอบกันเป็นประชากรที่ต้องการศึกษา

N_i = จำนวนหน่วยตัวอย่างทั้งหมดในชั้นภูมิที่ i

$N_{i'}$ = จำนวนหน่วยตัวอย่างทั้งหมดในชั้นภูมิที่ i' ที่นำมาจัดคู่กับชั้นภูมิที่ i

$N_{i''}$ = จำนวนหน่วยตัวอย่างทั้งหมดในชั้นภูมิที่ i'' ที่นำมาจัดกลุ่มกับชั้นภูมิที่ i และ i'

y_{ij} = ค่าของลักษณะที่ต้องการศึกษาของหน่วยตัวอย่างที่ j ในชั้นภูมิที่ i

$\bar{y}_i$ = ค่าเฉลี่ยของตัวอย่างในชั้นภูมิที่ i

$\bar{y}_{i'}$ = ค่าเฉลี่ยของตัวอย่างในชั้นภูมิที่ i' ที่นำมาจัดคู่กับชั้นภูมิที่ i

$\bar{y}_{i''}$ = ค่าเฉลี่ยของตัวอย่างในชั้นภูมิที่ i'' ที่นำมาจัดกลุ่มกับชั้นภูมิที่ i และ i'

$\bar{\bar{Y}}_{st}$ = ค่าเฉลี่ยของประชากร (population mean)

$\bar{y}_{st}$ = ค่าเฉลี่ยของตัวอย่าง (sample mean)

T = ยอดรวมของประชากร

n_i = ขนาดตัวอย่างในชั้นภูมิที่ i ; n = ขนาดตัวอย่างทั้งหมด (sample size)

๒.๒.๑ การประมาณค่าเฉลี่ยและความแปรปรวนของค่าเฉลี่ย

ก. การประมาณค่าเฉลี่ย ($\bar{y}_{st}$)

สมมุติว่าในการวางแผนการสำรวจได้แบ่งประชากรออกเป็น L ชั้นภูมิ แต่ละชั้นภูมิประกอบด้วยจำนวนหน่วยตัวอย่าง $N_1, N_2, \ldots, N_L$ หน่วย ตามลำดับ เลือกตัวอย่าง

จากแต่ละชั้นภูมิมาชั้นภูมิละ ๑ หน่วย โดยการสุ่ม

ให้ y_{ij} เป็นค่าของลักษณะที่ต้องการศึกษาของหน่วยตัวอย่างที่ j ในชั้นภูมิที่ i ($i = 1, 2, 3, \ldots, L$; $j = 1, 2, 3, \ldots, N_i$)

ค่าเฉลี่ยของ Y คือ

$$\bar{\bar{Y}}_{st} = \frac{\sum_{i=1}^{L} \sum_{j=1}^{N_i} y_{ij}}{\sum_{i=1}^{L} N_i}$$

และค่าประมาณที่ไม่มีความเอียงเฉ (unbiased estimator) ของ $\bar{\bar{Y}}_{st}$ คือ

$$\bar{y}_{st} = \frac{\sum_{i=1}^{L} N_i \bar{y}_i}{\sum_{i=1}^{L} N_i} \; ; \; \bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

ข. <u>การประมาณความแปรปรวนของค่าเฉลี่ย</u> $\left\{\hat{V}_c(\bar{y}_{st})\right\}$

การประมาณค่าความแปรปรวนของค่าเฉลี่ยทำได้โดยวิธีจัดกลุ่มชั้นภูมิเข้าด้วยกัน (Collapsed strata) ในที่นี้จะจัดกลุ่มทีละสองชั้นภูมิ กลุ่มสุดท้ายอาจจะมีสองหรือสามชั้นภูมิ

ถ้าจัดชั้นภูมิได้ k คู่ หรือ จำนวนชั้นภูมิทั้งหมดที่นำมาจัดมี 2k ชั้นภูมิ ชั้นภูมิที่นำมาจัดคู่กันคือ ชั้นภูมิ i และ i'

เนื่องจากความแปรปรวนของค่าเฉลี่ยในชั้นภูมิที่ i และ i' คือ

$$V(\bar{y}_i) = \frac{N_i - n_i}{N_i} \cdot \frac{1}{n_i} S_i^2$$

และ $$V(\bar{y}_{i'}) = \frac{N_{i'} - n_{i'}}{N_{i'}} \cdot \frac{1}{n_{i'}} S_{i'}^2$$

โดยที่ S_i^2 และ $S_{i'}^2$ คือความแปรปรวนของประชากรในชั้นภูมิที่ i และ i'

ดังนั้น ความแปรปรวนรวมของค่าเฉลี่ยของชั้นภูมิทั้งหมดที่จับคู่กัน

$$V_c(\bar{y}_{st}) = \frac{1}{N^2}\left\{\sum_{(i,i')\leq 2k}\left(N_i^2\frac{N_i-n_i}{N_i}\frac{1}{n_i}S_i^2 + N_{i'}^2\frac{N_{i'}-n_{i'}}{N_{i'}}\frac{1}{n_{i'}}S_{i'}^2\right) + \sum_{i>2k} N_i^2 V(\bar{y}_i)\right\} \qquad \text{.............(1)}$$

ซึ่งค่าประมาณความแปรปรวนรวมของค่าเฉลี่ยของชั้นภูมิทั้งหมดที่จับคู่กัน คือ

$$\hat{V}_c(\bar{y}_{st}) = \frac{1}{N^2}\left\{\sum_{(i,i')\leq 2k}\left(N_i^2\frac{N_i-n_i}{N_i}\frac{1}{n_i}s_i^2 + N_{i'}^2\frac{N_{i'}-n_{i'}}{N_{i'}}\frac{1}{n_{i'}}s_{i'}^2\right) + \sum_{i>2k} N_i^2 v(\bar{Y}_i)\right\}$$

เมื่อ s_i^2 และ $s_{i'}^2$ เป็นค่าประมาณความแปรปรวนของชั้นภูมิที่ i และ i'

จากสูตรเราไม่สามารถคำนวณค่า s_i^2 และ $s_{i'}^2$ ได้ เนื่องจากแต่ละชั้นภูมิมีขนาดตัวอย่างเพียง ๑ หน่วย จึงต้องเปลี่ยนสูตรข้างต้นให้เป็นรูปใหม่ ดังนี้

จากนิยาม $V(X_i) = E(X_i-\bar{X})^2$

$$= E(X_i^2) - \{E(X_i)\}^2$$

หรือ $E(X_i^2) = V(X_i) + \{E(X_i)\}^2$ ....(2)

ถ้าให้ $X_i = N_i\bar{y}_i - N_{i'}\bar{y}_{i'}$

แทนค่า $X_i = N_i\bar{y}_i - N_{i'}\bar{y}_{i'}$ จะได้

$$E(N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2 = V(N_i\bar{y}_i - N_{i'}\bar{y}_{i'}) + \{E(N_i\bar{y}_i - N_{i'}\bar{y}_{i'})\}^2$$

$$= V(N_i\bar{y}_i - N_{i'}\bar{y}_{i'}) + \{N_iE(\bar{y}_i) - N_{i'}E(\bar{y}_{i'})\}^2$$

$$= V(N_i\bar{y}_i - N_{i'}\bar{y}_{i'}) + \{N_i\bar{\bar{Y}}_i - N_{i'}\bar{\bar{Y}}_{i'}\}^2$$

เมื่อ $\bar{\bar{Y}}_i$ และ $\bar{\bar{Y}}_{i'}$ คือ ค่าเฉลี่ยของประชากรในชั้นภูมิที่ i และ i'

$$= V(N_i\bar{y}_i - N_{i'}\bar{y}_{i'}) + (T_i - T_{i'})^2$$

เมื่อ $T_i = N_i\bar{\bar{Y}}_i$ และ $T_{i'} = N_{i'}\bar{\bar{Y}}_{i'}$

$$= N_i^2\frac{N_i-n_i}{N_i}\frac{1}{n_i}S_i^2 + N_{i'}^2\frac{N_{i'}-n_{i'}}{N_{i'}}\frac{1}{n_{i'}}S_{i'}^2 + (T_i - T_{i'})^2$$

เนื่องจาก $E(N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2 \cong (N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2$ และ $(T_i - T_{i'})^2$ มีค่าน้อยมาก[๑]

ดังนั้น $$(N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2 \cong N_i^2 \frac{N_i - n_i}{N_i} \frac{1}{n_i} s_i^2 + N_{i'}^2 \frac{N_{i'} - n_{i'}}{N_{i'}} \frac{1}{n_{i'}} s_{i'}^2$$

นั่นคือ จาก (1) จะได้

$$\hat{V}_c(\bar{y}_{st}) \cong \frac{1}{N^2} \sum_{(i,i') \leq 2k} (N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2 + \frac{1}{N^2} \sum_{i > 2k} N_i^2 \, v(\bar{y}_i)$$

เนื่องจากเทอมแรกทางขวามือ ; $\frac{1}{N^2} \sum_{(i,i') \leq 2k} (N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2$ ประกอบด้วยชั้นภูมิที่ (i,i')k คู่

และเทอมที่สองทางขวามือ ; $\frac{1}{N^2} \sum_{i > 2k} N_i^2 v(\bar{y}_i)$ ประกอบด้วยชั้นภูมิที่จัดกลุ่ม ๓ ชั้นภูมิเข้าด้วยกัน

แต่ในแต่ละชั้นภูมิของเทอมที่สองนี้ไม่สามารถคำนวณค่า $v(\bar{y}_i)$ ได้ ดังนั้น จะประมาณความแปรปรวนของชั้นภูมิสุดท้ายที่ประกอบด้วย ๓ ชั้นภูมิด้วยสูตรที่คล้ายกับเทอมแรก

$$\frac{1}{2N^2}\left\{(N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2 + (N_i\bar{y}_i - N_{i''}\bar{y}_{i''})^2 + (N_{i'}\bar{y}_{i'} - N_{i''}\bar{y}_{i''})^2\right\}$$

ดังนั้น สูตรสำหรับการประมาณความแปรปรวนของค่าเฉลี่ย เมื่อใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวโดยสุ่มตัวอย่างมาชั้นภูมิละหนึ่งหน่วย คือ

$$\hat{V}_c(\bar{y}_{st}) \cong \frac{1}{N^2} \sum_{(i,i') \leq 2k} (N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2 + \frac{1}{2N^2}\left\{(N_i\bar{y}_i - N_{i'}\bar{y}_{i'})^2 + (N_i\bar{y}_i - N_{i''}\bar{y}_{i''})^2 + (N_{i'}\bar{y}_{i'} - N_{i''}\bar{y}_{i''})^2\right\}$$

---

[๑] Konijn H.S, Statistical Theory of Sample Survey Design and Analysis, North-Holland publishing company London 1973, p. 144

๒.๑.๒ ความเอียงเฉในทางบวกของความแปรปรวนในการสุ่มตัวอย่างมาชั้นภูมิละหนึ่งหน่วย (Positive Bias of Variance with one unit per stratum)

ในการสุ่มตัวอย่างมาชั้นภูมิละหนึ่งหน่วย เนื่องจากในแต่ละชั้นภูมิที่นำมาจัดคู่กันต้องมีจำนวนหน่วยตัวอย่างใกล้เคียงกัน และค่าเฉลี่ยของประชากรของชั้นภูมิที่นำมาจัดคู่กันต้องไม่แตกต่างกันมากนัก ถ้าค่าเฉลี่ยของประชากรของชั้นภูมิที่นำมาจัดคู่กันต่างกันมากจะก่อให้เกิดความเอียงเฉในทางบวกของค่าความแปรปรวนของค่าประมาณ ดังนี้

ถ้าผลการสุ่มตัวอย่างในแต่ละคู่เป็น y_{i1}, y_{i2} โดยที่ i มีค่าตั้งแต่ 1 ถึง $L/2$ ในทำนองเดียวกับ (2) ได้ว่า

$$E(y_{i1}-y_{i2})^2 = (\bar{\bar{Y}}_{i1}-\bar{\bar{Y}}_{i2})^2 + \frac{N_i-1}{N_i}(S_{i1}^2+S_{i2}^2) \quad \ldots\ldots\ldots(3)$$

โดยที่ $N_i = N_{i'} = N_h$ ซึ่งคือจำนวนประชากรในแต่ละชั้นภูมิในคู่นี้

พิจารณาค่าประมาณ

$$\hat{V}(\bar{y}_{st}) = \frac{1}{N^2}\sum_{i=1}^{L/2} N_i^2(y_{i1}-y_{i2})^2 \quad \ldots\ldots\ldots(4)$$

ค่าเฉลี่ยของจำนวนนี้ก็คือ

$$E\,\hat{V}(\bar{y}_{st}) = E\left[\frac{1}{N^2}\sum_{i=1}^{L/2} N_i^2(y_{i1}-y_{i2})^2\right] \quad \ldots\ldots\ldots(5)$$

แทนค่า (3) ลงใน (5)

$$\therefore E\,\hat{V}(\bar{y}_{st}) = \frac{1}{N^2}\left[\sum_{h=1}^{L} N_h(N_h-1)S_h^2 + \sum_{i=1}^{L/2} N_i^2(\bar{\bar{Y}}_{i1}-\bar{\bar{Y}}_{i2})\right] \ldots(6)$$

เทอมแรกทางขวามือเป็นค่าความแปรปรวนที่ถูกต้อง (Correct variance) ($n_h = 1$) เทอมที่สองคือ ความเอียงเฉของความแปรปรวนในทางบวก ความเอียงเฉดังกล่าวขึ้นอยู่กับความแตกต่างของค่าเฉลี่ยที่แท้จริง (true mean) ของชั้นภูมิที่นำมาจับคู่กัน และจาก (5) ให้ข้อสังเกตว่าถ้าจัดคู่ชั้นภูมิที่มีค่าจากตัวอย่าง (sample value) ต่างกันเพียงเล็กน้อย จะทำให้ค่าความแปรปรวนที่ได้เป็นค่าประมาณที่ต่ำกว่าความเป็นจริง (Underestimate)

๒.๒ แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อย (Two-way stratification with small sample)

ในการวางแผนการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อยนี้ จะแบ่งหน่วยตัวอย่างออกเป็นพวก ๆ ตามลักษณะ ๒ ลักษณะ คือ ทางแถว (row) และทางสดมภ์ (column) แต่ละพวกที่แบ่งแล้วเรียกว่า ชั้นภูมิ (stratum) หรือเซลล์ (Cell) แล้วเลือกตัวอย่างจากแต่ละชั้นภูมิ โดยที่ตัวอย่างที่เลือกในแต่ละชั้นภูมินี้เป็นอิสระต่อกัน หลักในการจัดชั้นภูมิแบบสองทางนี้เป็นเช่นเดียวกันกับในแผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียว กล่าวคือ ภายในชั้นภูมิเดียวกันตัวอย่างควรประกอบด้วยหน่วยตัวอย่างที่มีความคล้ายคลึงกัน (homogeneous) และชั้นภูมิที่แตกต่างกันควรประกอบด้วยหน่วยตัวอย่างที่มีความแตกต่างกัน (heterogeneous)

จุดประสงค์ที่สำคัญของการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทาง ก็เพื่อลดความคลาดเคลื่อนภายในชั้นภูมิให้น้อยลง ซึ่งจะทำให้ความคลาดเคลื่อนในระดับประชากรรวมมีค่าต่ำไปด้วย การใช้แผนแบบนี้มีประสิทธิภาพสูงกว่าการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียว หรือกล่าวอีกอย่างหนึ่ง คือ การใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทาง โดยใช้ตัวอย่างจำนวนน้อยนี้จะให้ค่าประมาณที่มีความแปรปรวนหรือความคลาดเคลื่อนน้อยกว่าการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวเสมอ

ส่วนข้อเสียก็คือ การใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทาง จะต้องเสียเวลาและค่าใช้จ่ายมากขึ้นในการประมาณพารามิเตอร์ เนื่องจากมีงานเพิ่มมากขึ้นทั้งในขั้นการเลือกตัวอย่างและขั้นการประมาณผล และในการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อยนี้ ขนาดตัวอย่างที่ใช้จะต้องกำหนดให้มากพอที่จะกระจายตกในชั้นภูมิต่าง ๆ ถ้าขนาดตัวอย่างที่ใช้น้อยเกินไป จะมีปัญหาในการประมาณค่าความแปรปรวนภายในชั้นภูมิได้

ถ้าให้ R แทนจำนวนแถว C แทนจำนวนสดมภ์ และ n แทนขนาดตัวอย่าง การแบ่งประเภทของหน่วยตัวอย่างออกเป็น R แถว และC สดมภ์ จะได้ชั้นภูมิทั้งหมด RC ชั้นภูมิ จำนวนหน่วยตัวอย่างในแต่ละชั้นภูมิ (n_{ij}) ที่ใช้ประมาณความแปรปรวนภายในชั้นภูมิควรจะมีอย่างน้อยเท่ากับ ๒ หน่วยตัวอย่าง นั่นคือ ขนาดตัวอย่างที่ต้องใช้ในการประมาณความแปรปรวนอย่างน้อยต้องมากกว่าหรือเท่ากับ 2 RC ($n \geq 2$ RC)

สำหรับปัญหาที่เกิดขึ้น เมื่อขนาดตัวอย่างที่ใช้ไม่เพียงพอที่จะกระจายตกในแต่ละชั้นภูมิได้อย่างน้อย ๒ หน่วยตัวอย่าง ไบรอัน (Bryant) ฮาร์ทลี่ (Hartley) และ เจสเสน (Jessen) ได้คิดวิธีประมาณค่าเฉลี่ยและความแปรปรวนรวม ซึ่งมีวิธีการดังต่อไปนี้

สมมุติให้หน่วยตัวอย่างทั้งหมดถูกแบ่งออกเป็นชั้นภูมิทั้งหมด ij ชั้นภูมิ คือมี i แถว และ j สดมภ์ ถ้าให้

N_{ij} = จำนวนหน่วยตัวอย่างทั้งหมดในแถวที่ i สดมภ์ที่ j
หรือ จำนวนหน่วยตัวอย่างทั้งหมดในชั้นภูมิหรือเซลล์ที่ ij

n = ขนาดตัวอย่าง

n_{ij} = ขนาดตัวอย่างในแถวที่ i สดมภ์ที่ j หรือขนาดตัวอย่างในชั้นภูมิที่ ij

$n_{i.}$ = ขนาดตัวอย่างในแถวที่ i
$= nP_{i.}$

$n_{.j}$ = ขนาดตัวอย่างในสดมภ์ที่ j
$= nP_{.j}$

P_{ij} = โอกาสที่หน่วยตัวอย่างในชั้นภูมิที่ ij จะถูกเลือก

$$= \frac{\text{จำนวนหน่วยตัวอย่างทั้งหมดในแถวที่ i, สดมภ์ที่ j}}{\text{จำนวนหน่วยตัวอย่างทั้งหมด}}$$

$$= \frac{N_{ij}}{\sum_{ij} N_{ij}}$$

$N_{i.}$ = จำนวนหน่วยตัวอย่างทั้งหมดในแถวที่ i

$$= \sum_{j} N_{ij}$$

$$P_{i.} = \frac{\text{จำนวนหน่วยตัวอย่างทั้งหมดในแถวที่ i}}{\text{จำนวนหน่วยตัวอย่างทั้งหมด}}$$

$$= \frac{N_{i.}}{\sum_{ij} N_{ij}}$$

$N_{.j}$ = จำนวนหน่วยตัวอย่างทั้งหมดในสดมภ์ที่ j

$$= \sum_{i} N_{ij}$$

$$P_{.j} = \frac{\text{จำนวนหน่วยตัวอย่างทั้งหมดในสดมภ์ที่ j}}{\text{จำนวนหน่วยตัวอย่างทั้งหมด}}$$

$$= \frac{N_{.j}}{\sum_{ij} N_{ij}}$$

ตารางที่ ๑ แสดงแผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทาง

แถว	สดมภ์ 1	สดมภ์ 2	สดมภ์ 3	. . .	สดมภ์ j	. . .	สดมภ์ C	ยอดรวม	ขนาดตัวอย่าง
1	N_{11} P_{11}	N_{12} P_{12}	N_{13} P_{13}	. . .	N_{1j} P_{1j}	. . .	N_{1C} P_{1C}	$N_{1.}$ $P_{1.}$	$n_{1.}$
2	N_{21} P_{21}	N_{22} P_{22}	N_{23} P_{23}	. . .	N_{2j} P_{2j}	. . .	N_{2C} P_{2C}	$N_{2.}$ $P_{2.}$	$n_{2.}$
3	N_{31} P_{31}	N_{32} P_{32}	N_{33} P_{33}	. . .	N_{3j} P_{3j}	. . .	N_{3C} P_{3C}	$N_{3.}$ $P_{3.}$	$n_{3.}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮	⋮
i	N_{i1} P_{i1}	N_{i2} P_{i2}	N_{i3} P_{i3}	. . .	N_{ij} P_{ij}	. . .	N_{iC} P_{iC}	$N_{i.}$ $P_{i.}$	$n_{i.}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮	⋮
R	N_{R1} P_{R1}	N_{R2} P_{R2}	N_{R3} P_{R3}	. . .	N_{Rj} P_{Rj}	. . .	N_{RC} P_{RC}	$N_{R.}$ $P_{R.}$	$n_{R.}$
ยอดรวม	$N_{.1}$ $P_{.1}$	$N_{.2}$ $P_{.2}$	$N_{.3}$ $P_{.3}$	. . .	$N_{.j}$ $P_{.j}$	. . .	$N_{.C}$ $P_{.C}$	$\sum_{ij} P_{ij} = 1$	
ขนาดตัวอย่าง	$n_{.1}$	$n_{.2}$	$n_{.3}$	. . .	$n_{.j}$	. . .	$n_{.C}$		

การกำหนดขนาดตัวอย่างให้แก่แต่ละชั้นภูมิกระทำได้โดยคำนวณขนาดตัวอย่างในแต่ละแถวด้วย $n_{i.} = nP_{i.}$ และคำนวณขนาดตัวอย่างในแต่ละสดมภ์ด้วย $n_{.j} = nP_{.j}$ เมื่อปัดเศษตัวเลขทศนิยมที่คำนวณได้ให้เป็นเลขจำนวนเต็ม (integer) จะได้ $\sum_{i} n_{i.} = n$ และ $\sum_{j} n_{.j} = n$ ตารางที่ใช้ในการสุ่มตัวอย่างในแต่ละชั้นภูมิเป็นดังนี้

ตารางที่ ๒ ตารางที่ใช้ในการสุ่มตัวอย่างในแต่ละชั้นภูมิ

สดมภ์ / แถว		1				. . .	j				. . .	C				จำนวนตัวอย่าง
		1	2	. . .	$nP_{.1}$	. . .	1	2	. . .	$nP_{.j}$	. . . .	1	2	. . .	$nP_{.c}$	
1	1															
	2															
	:															$nP_{1.}$
	$nP_{1.}$															
:	:															
i	1															
	2															
	:															$nP_{i.}$
	$nP_{i.}$															
R	1															
	2															
	:															$nP_{R.}$
	$nP_{R.}$															
จำนวนตัวอย่าง		$nP_{.1}$					$nP_{.j}$					$nP_{.c}$				n

จากตารางที่ใช้ในการสุ่มตัวอย่างในแต่ละชั้นภูมิข้างต้น สุ่มตัวอย่างในแต่ละแถวเป็นอิสระต่อกัน โดยใช้วิธีการสุ่มแบบธรรมดาโดยไม่มีการแทนที่ (random sampling without replacement)

๑. จากแถวที่ ๑ สุ่มตัวอย่างมา ๑ สดมภ์ นำสดมภ์ที่ถูกสุ่มแล้วออก

๒. จากแถวที่ ๒ สุ่มตัวอย่างสดมภ์ที่เหลือจากการสุ่มในแถวที่ ๑ มาอีก ๑ สดมภ์ แล้วนำสดมภ์ที่ถูกสุ่มแล้วออก

ทำเช่นนี้จนกระทั่งถึงแถวที่ n จะได้ตัวอย่างขนาด n ที่กระจายในแต่ละชั้นภูมิ ด้วยความน่าจะเป็น

$P_{ij} \cong \frac{n_{i.}\, n_{.j}}{n^2}$ โดยที่ $\sum_{ij} P_{ij} = 1$

การกำหนดขนาดตัวอย่างให้แก่แต่ละแถวหรือแต่ละสดมภ์ ($n_{i.}$ หรือ $n_{.j}$) แถวใดหรือสดมภ์ใดมีจำนวนหน่วยตัวอย่างมาก จำนวนตัวอย่างที่เลือกมาจากแถวนั้นหรือสดมภ์นั้นก็จะมีจำนวนมาก และแถวใดหรือสดมภ์ใดมีจำนวนหน่วยตัวอย่างน้อย จำนวนตัวอย่างที่เลือกมาจากแถวนั้นหรือสดมภ์นั้นก็จะมีจำนวนน้อย

๒.๒.๑ การประมาณค่าเฉลี่ยและความแปรปรวนของค่าเฉลี่ย

ก. การประมาณค่าเฉลี่ย

สมมุติให้ y_{ij} เป็นค่าของลักษณะที่ต้องการศึกษา หรือข้อมูลของแถวที่ i สดมภ์ที่ j ($i = 1, 2, 3, \ldots, R$; $j = 1, 2, 3, \ldots, C$)

สูตรสำหรับการประมาณค่าเฉลี่ยของ Y คือ

$$\bar{y}_u = \frac{1}{n} \sum \frac{n^2 P_{ij}}{n_{i.}\, n_{.j}} y_{ij}$$

ค่าเฉลี่ย $\bar{y}_u$ ที่ได้เป็นค่าประมาณที่ไม่มีความเอียงเฉของ $\bar{Y}_{..}$

หรือ $E(\bar{y}_u) = \bar{Y}_{..}$ โดยที่ $\bar{Y}_{..} = \sum_{ij} P_{ij} \bar{Y}_{ij}$

$\bar{Y}_{..}$ = ค่าเฉลี่ยของประชากร

$\bar{Y}_{ij}$ = ค่าเฉลี่ยของประชากรในแต่ละเซลล์

ข. การประมาณค่าความแปรปรวนของค่าเฉลี่ย $\left(\hat{V}(\bar{y}_u)\right)$

สูตรที่ใช้สำหรับหาค่าความแปรปรวนของค่าเฉลี่ยเมื่อใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อย คือ

$$\hat{V}(\bar{y}_u) \cong \frac{1}{n}\sum_{ij}P_{ij}s_{ij}^2 + \frac{1}{n-1}\sum_{ij}P_{ij}(\bar{y}_{ij}-\bar{y}_{i.})^2 - \frac{1}{n-1}\sum_{j}P_{.j}(\bar{y}_{.j}-\bar{y}_u)^2$$

เมื่อ s_{ij}^2 = ความแปรปรวนของหน่วยตัวอย่างในชั้นภูมิที่ ij

$$= \frac{\sum_{ij}(y_{ij}-\bar{y}_{ij})^2}{n_{ij}-1}$$

$\bar{y}_{ij}$ = ค่าเฉลี่ยของหน่วยตัวอย่างในชั้นภูมิที่ ij

$\bar{y}_{i.}$ = ค่าเฉลี่ยของหน่วยตัวอย่างในแถวที่ i

$\bar{y}_{.j}$ ค่าเฉลี่ยของหน่วยตัวอย่างในสดมภ์ที่ j

พิจารณาเทอม $\frac{1}{n}\sum_{ij}P_{ij}s_{ij}^2$ เนื่องจากชั้นภูมิที่ ij บางชั้นภูมิมีตัวอย่างตกเพียง ๑ หน่วยเท่านั้น การประมาณค่าความแปรปรวนภายในชั้นภูมินั้นจึงไม่สามารถกระทำได้ แต่มีวิธีการที่สามารถประมาณค่าความแปรปรวนรวมได้โดยการจัดกลุ่มชั้นภูมิเข้าด้วยกัน (Collapsed strata) ซึ่งวิธีการนี้ได้กล่าวโดยละเอียดแล้วในแผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียว โดยสุ่มตัวอย่างมาชั้นภูมิละหนึ่งหน่วย ในที่นี้จะกล่าวเฉพาะการประยุกต์สูตรมาใช้

จาก $\frac{1}{n}\sum_{ij}P_{ij}s_{ij}^2$ = ความแปรปรวนของชั้นภูมิที่นำมาจัดกลุ่ม + ความแปรปรวนของชั้นภูมิที่ไม่จัดกลุ่ม

$= \sum_{k}P_{C_k}\hat{V}_{C_k}(\bar{y}_{st})$ + ความแปรปรวนของชั้นภูมิที่ไม่จัดกลุ่ม

$$= \frac{1}{N_C^2}\sum_{(h,h')\leq 2k} P_{C_{(h,h')}}(N_h\bar{y}_h - N_{h'}\bar{y}_{h'})^2$$

$$+ \frac{1}{2N_C^2}P_{C_{(h,h',h'')}}\left[(N_h\bar{y}_h - N_{h'}\bar{y}_{h'})^2 + (N_h\bar{y}_h - N_{h''}\bar{y}_{h''})^2 + (N_{h'}\bar{y}_{h'} - N_{h''}\bar{y}_{h''})^2\right]$$

+ ความแปรปรวนของชั้นภูมิที่ไม่จัดกลุ่ม

ซึ่ง

$P_{C_{(h,h')}}$ = สัดส่วนของจำนวนหน่วยตัวอย่างในชั้นภูมิที่ h+h' ต่อจำนวนหน่วยตัวอย่างทั้งหมด

$P_{C_{(h,h',h'')}}$ = สัดส่วนของจำนวนหน่วยตัวอย่างในชั้นภูมิที่ h+h'+h" ต่อจำนวนหน่วยตัวอย่างทั้งหมด

N_h = จำนวนหน่วยตัวอย่างทั้งหมดในชั้นภูมิที่ h

$N_{h'}$ = จำนวนหน่วยตัวอย่างทั้งหมดในชั้นภูมิที่ h' ที่นำมาจัดคู่กับชั้นภูมิที่ h

$N_{h''}$ = จำนวนหน่วยตัวอย่างทั้งหมดในชั้นภูมิที่ h" ที่นำมาจัดกลุ่มกับชั้นภูมิที่ h และ h'

$\bar{y}_h$ = ค่าเฉลี่ยของชั้นภูมิที่ h

$\bar{y}_{h'}$ = ค่าเฉลี่ยของชั้นภูมิที่ h' ที่นำมาจัดคู่กับชั้นภูมิที่ h

$\bar{y}_{h''}$ = ค่าเฉลี่ยของชั้นภูมิที่ h" ที่นำมาจัดกลุ่มกับชั้นภูมิที่ h และ h'

N_C = จำนวนหน่วยตัวอย่างทั้งหมดในชั้นภูมิที่นำมาจัดกลุ่มกัน

ในทำนองเดียวกันพิจารณาเทอม $\frac{1}{n-1}\sum_{ij} P_{ij}(\bar{y}_{ij}-\bar{y}_{i.})^2$ เนื่องจากบางแถวมีตัวอย่างตกเพียง ๑ หน่วย การประมาณค่าความแปรปรวนในแถวนั้นจึงไม่สามารถกระทำได้ การประมาณค่าความแปรปรวนรวมจะใช้วิธีการจับคู่แถวที่มีข้อมูลตัวอย่างเพียง ๑ หน่วยตัวอย่างเข้าด้วยกัน

จาก $\frac{1}{n-1}\sum_{ij} P_{ij}(\bar{y}_{ij}-\bar{y}_{i.})^2$ = ความแปรปรวนของแถวที่นำมาจัดกลุ่ม + ความแปรปรวนของแถวที่ไม่จัดกลุ่ม

= $n_c \sum_k P_{C_k} \hat{V}_{C_k}(\bar{y}_{st})$ + ความแปรปรวนของแถวที่ไม่จัดกลุ่ม

$$= \frac{n_c}{N_c^2} \sum_{(h,h')\leq 2k} P_{C_{(h,h')}}(N_h\bar{y}_h - N_{h'}\bar{y}_{h'})^2$$

+ ความแปรปรวนของถวที่ไม่จัดกลุ่ม

n_c = จำนวนตัวอย่างที่มาจัดกลุ่มกัน

$P_{C_{(h,h')}}$ = สัดส่วนของจำนวนหน่วยตัวอย่างในแถวที่ h+h' ต่อจำนวนหน่วยตัวอย่างทั้งหมด

N_h = จำนวนของค่าเฉลี่ยของชั้นภูมิในแถวที่ h

$N_{h'}$ = จำนวนของค่าเฉลี่ยของชั้นภูมิในแถวที่ h'

$\bar{y}_h$ = ค่าเฉลี่ย (sample mean) ของ $\bar{y}_{ij}$ ในแถวที่ h

$\bar{y}_{h'}$ = ค่าเฉลี่ย (sample mean) ของ $\bar{y}_{ij}$ ในแถวที่ h'

N_C = จำนวนค่าเฉลี่ย ($\bar{y}_{ij}$) ทั้งหมดทุกแถวที่นำมาจัดกลุ่มกัน

๒.๒.๒ การเปรียบเทียบประสิทธิภาพระหว่างแผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อย กับแผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียว

ในการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อยนั้น วัตถุประสงค์สำคัญก็เพื่อที่จะทำให้สามารถประมาณค่าพารามิเตอร์ที่สนใจ เช่น ค่าเฉลี่ย ให้มีความแปรปรวนหรือความคลาดเคลื่อนน้อยกว่าการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียว

จากสูตรความแปรปรวนของค่าเฉลี่ยเมื่อใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อย

$$V(\bar{y}_u) = \frac{1}{n}\sum P_{ij}S_{ij}^2 + \frac{1}{n-1}\sum P_{ij}(\bar{y}_{ij}-\bar{y}_{i.})^2 - \frac{1}{n-1}\sum P_{.j}(\bar{y}_{.j}-\bar{Y}_{..})^2 \cdots\cdots(1)$$

ถ้าใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียว โดยใช้แถวเป็นชั้นภูมิ ค่าเฉลี่ยของ Y คือ

$$\bar{y}_{st} = \sum_i P_{i.}\bar{y}_{n_{i.}}$$

เมื่อ $P_{i.}$ = สัดส่วนของจำนวนหน่วยตัวอย่างในชั้นภูมิที่ i ต่อจำนวนหน่วยตัวอย่างทั้งหมด

$= \frac{N_{i.}}{N}$

$\bar{y}_{i.}$ = ค่าเฉลี่ยที่ได้จากตัวอย่างในชั้นภูมิที่ i

และความแปรปรวนของ Y คือ

$$V(\bar{y}_{st}) \cong \sum_i \frac{P_{i.}^2 S_i^2}{n_{i.}} \qquad \text{..........(2)}$$

เราทราบว่า

$$S_i^2 = \frac{1}{NP_{i.}-1}\sum_{jk}(y_{ijk}-\bar{y}_{i.})^2$$

$$= \frac{1}{NP_{i.}-1}\sum_{jk}(y_{ijk}-\bar{y}_{ij}+\bar{y}_{ij}-\bar{y}_{i.})^2$$

$$= \frac{1}{NP_{i.}-1}\left[\sum_j (N_{ij}-1)S_{ij}^2+\sum_j N_{ij}(\bar{y}_{ij}-\bar{y}_{i.})^2\right]$$

$$= \sum_i \frac{P_{ij}}{P_{i.}} S_{ij}^2 + \sum_j \frac{P_{ij}}{P_{i.}}(\bar{y}_{ij}-\bar{y}_{i.})^2$$

ดังนั้น จาก (2)

$$V(\bar{y}_{st}) \cong \frac{1}{n}\sum_{ij} P_{ij}S_{ij}^2 + \frac{1}{n}\sum_{ij}P_{ij}(\bar{y}_{ij}-\bar{y}_{i.})^2 \text{..........(3)}$$

$$(3)-(1);\ V(\bar{y}_{st})-V(\bar{y}_u) = \frac{1}{n-1}\sum_j P_{.j}(\bar{y}_{.j}-\bar{Y}_{..})^2 - \frac{1}{n(n-1)}\sum_{ij}P_{ij}(\bar{y}_{ij}-\bar{y}_{i.})^2$$

ตามหลักการจัดชั้นภูมิ ภายในชั้นภูมิเดียวกันจะประกอบด้วยหน่วยตัวอย่างที่มีความคล้ายคลึงกัน และต่างชั้นภูมิกันจะประกอบด้วยหน่วยตัวอย่างที่มีความแตกต่างกันในลักษณะที่สนใจจะศึกษา ดังนั้น เทอมแรกทางขวามือ $\frac{1}{n-1}\sum_j P_{.j}(\bar{y}_{.j}-\bar{Y}_{..})^2$ ซึ่งเป็นความแตกต่างระหว่างชั้นภูมิ ย่อมต้องมีค่ามากกว่า เทอมที่สอง $\frac{1}{n(n-1)}\sum_{ij} P_{ij}(\bar{y}_{ij}-\bar{y}_{i.})^2$ ซึ่งเป็นความแตกต่างภายในชั้นภูมิเสมอ ผลที่ได้มีค่าเป็นบวก หรือ $V(\bar{y}_{st}) - V(\bar{y}_u)$ มีค่ามากกว่า ๐ นั่นคือ การใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อยจะมีประสิทธิภาพสูงกว่าการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวเสมอ

๒.๓ <u>แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ</u>

(Two-way stratification with systematic pps - sampling)

ในการวางแผนการสำรวจแบบนี้ จะแบ่งประเภทของหน่วยตัวอย่างออกตามลักษณะ ๒ อย่าง เช่นเดียวกับในแผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อย คือ

โดยใช้ตัวอย่างจำนวนน้อย คือ แบ่งหน่วยตัวอย่างออกเป็นทางแถวและทางสดมภ์ ซึ่งทำให้เกิดชั้นภูมิ หรือเซลล์ ภายในชั้นภูมิเดียวกันจะประกอบด้วยหน่วยตัวอย่างที่มีความคล้ายคลึงกัน และต่างชั้นภูมิจะประกอบด้วยหน่วยตัวอย่างที่มีความแตกต่างกัน

การสุ่มตัวอย่างแบบมีระบบ (systematic sampling) ต่างกับการสุ่มตัวอย่างแบบธรรมดา (simple random sampling) คือ การสุ่มตัวอย่างแบบมีระบบจะได้ตัวอย่างที่กระจายเป็นระบบในประชากรแน่นอนกว่าการสุ่มตัวอย่างแบบธรรมดา และโอกาสในการเลือก (selection probability) เมื่อใช้การสุ่มตัวอย่างแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจขึ้นอยู่กับจำนวนและขนาดของหน่วยตัวอย่าง สำหรับจำนวนของหน่วยตัวอย่างในประชากรจะทราบได้ไม่ยากนัก แต่สิ่งที่เป็นปัญหาคือ จะใช้อะไรเป็นค่าที่ใช้วัดขนาดของลักษณะที่ต้องการศึกษา ซึ่งจะได้กล่าวในตอนต่อไป

จุดประสงค์สำคัญของการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ คือ เพื่อลดความคลาดเคลื่อนในระดับประชากรรวมให้มีค่าต่ำ ความคลาดเคลื่อนในระดับประชากรรวมจะมีค่ามากหรือน้อยขึ้นอยู่กับโอกาสในการลือก ผู้วางแผนการสำรวจจะต้องพยายามกำหนดให้โอกาสในการเลือกเป็นปฏิภาคโดยตรงกับค่าที่ใช้วัดขนาดของลักษณะที่ต้องการศึกษา ซึ่งค่าที่ใช้วัดขนาดของลักษณะที่ต้องการศึกษานี้จะต้องมีส่วนสัมพันธ์ใกล้ชิดกับลักษณะที่ต้องการศึกษา ถ้าผู้วางแผนการสำรวจสามารถเลือกใช้ค่าที่ใช้วัดขนาดให้มีส่วนสัมพันธ์ใกล้ชิดกับลักษณะที่ต้องการศึกษาได้มากเท่าไร ค่าประมาณที่ได้ย่อมมีความแปรปรวนหรือความคลาดเคลื่อนน้อยลงเท่านั้น ในทางปฏิบัติจริง ๆ ไม่ทราบค่าของลักษณะที่ต้องการศึกษา จึงต้องพยายามหาค่าที่จะใช้เป็นตัวแทนของลักษณะที่ต้องการศึกษา

หลักสำคัญในการเลือกใช้ค่าที่ใช้วัดขนาดของลักษณะที่ต้องการศึกษา คือ ค่าที่ใช้วัดขนาดต้องมีส่วนสัมพันธ์ใกล้ชิดกับลักษณะที่ต้องการศึกษา หรือมีความเป็นตัวแทนลักษณะที่ต้องการศึกษา และสามารถให้ข้อเท็จจริงเกี่ยวกับลักษณะที่ต้องการศึกษาได้มากที่สุด

ในการสุ่มตัวอย่างจากชั้นภูมิใช้วิธีการสุ่มตัวอย่างแบบมีระบบ การสุ่มตัวอย่างวิธีนี้เป็นวิธีการสุ่มตัวอย่างที่ให้ความน่าจะเป็นที่หน่วยตัวอย่างหนึ่ง ๆ จะถูกเลือกเข้าไปอยู่ในตัวอย่าง

เท่ากันหมด (equal probability sampling หรือ random sampling) เทคนิคการสุ่มตัวอย่างแบบนี้จะคำนวณช่วงสุ่ม (sampling interval) สมมุติเท่ากับ I เลือกจุดเริ่มสุ่ม (random start) r ตัวอย่างที่ได้จะประกอบด้วยหน่วยตัวอย่างที่ r, r+I, r+2I,... ซึ่งเป็นการเลือกหน่วยตัวอย่างมา ๑ หน่วยทุก ๆ I หน่วย คล้ายกับประชากรถูกแบ่งออกเป็น I ชุด แบบมีระบบก่อนที่จะถูกเลือกมาใช้อย่างสุ่ม ๆ ๑ ชุด

การใช้การสุ่มตัวอย่างแบบมีระบบ มีข้อดีคือ สามารถเลือกตัวอย่างได้ง่าย รวดเร็ว ประหยัดเวลาและค่าใช้จ่าย อาจสามารถลดความคลาดเคลื่อนของค่าประมาณ และค่าประมาณที่ได้จะมีความแม่นยำ (precise) กว่าวิธีการสุ่มแบบธรรมดา เนื่องจากในการสุ่มตัวอย่างแบบมีระบบตัวอย่างที่ได้จะกระจายทั่วไปในประชากร ทำให้ได้ตัวอย่างที่เป็นตัวแทนของประชากรกระจายอยู่ทั่วไป ไม่มารวมอยู่ช่วงใดช่วงหนึ่ง แต่มีข้อควรระมัดระวังในการใช้การสุ่มตัวอย่างแบบนี้คือ การเรียงลำดับของหน่วยตัวอย่างในประชากรจะต้องเรียงให้หน่วยตัวอย่างที่มีค่าใกล้กันไว้ใกล้ ๆ กัน หรือไว้ในกลุ่มย่อย (subgroup) เดียวกัน เมื่อสุ่มตัวอย่างจะได้ตัวอย่างที่มีความแตกต่างกันกระจายอยู่ทั่วไป

๒.๓.๑ <u>วิธีการเลือกตัวอย่าง</u>

สมมุติว่าในการวางแผนการสำรวจได้แบ่งประเภทของหน่วยตัวอย่างออกเป็น ๒ ทาง คือ ทางแถวและทางสดมภ์ สมมุติให้มี R แถว และ C สดมภ์

ต้องการเลือกตัวอย่างขนาด n หน่วย จากประชากรที่ประกอบด้วยหน่วยตัวอย่างทั้งหมด N หน่วย โดยให้โอกาสในการเลือกหน่วยตัวอย่างที่ k ในชั้นภูมิที่ ij เป็น

$$P_{ijk} = \frac{M_{ijk}}{\sum_{j=1}^{c} \sum_{k=1}^{N_{ij}} M_{ijk}}$$

เมื่อ M_{ijk} เป็นค่าที่ใช้วัดขนาดของหน่วยตัวอย่างที่ k ในชั้นภูมิที่ ij ซึ่งผู้วางแผนการสำรวจต้องกำหนดขึ้นมา โดยพยายามให้ M_{ijk} มีลักษณะสัมพันธ์ใกล้เคียงกับลักษณะที่ต้องการศึกษาให้มากที่สุด วิธีการต่อไป คือ

หาผลบวกสะสมของ M_{ijk}

คำนวณช่วงสุ่ม ; $I = \frac{1}{n} \sum_{ijk} M_{ijk}$; $i = 1, 2, 3, \ldots, R$

$j = 1, 2, 3, \ldots, c$

$k = 1, 2, 3, \ldots, N_{ij}$

เลือกจุดเริ่มสุ่ม r

ตัวอย่างที่ได้จะประกอบด้วยหน่วยตัวอย่างที่ $r, r+I, r+2I, \ldots, r+(n-1)I$ ก็จะได้ตัวอย่างขนาด n ตามต้องการ

๒.๓.๒ การประมาณค่าเฉลี่ยและความแปรปรวนของค่าเฉลี่ย

ในการศึกษาขั้นนี้จะประยุกต์สูตรสำหรับการประมาณค่าเฉลี่ยและความแปรปรวนของค่าเฉลี่ยจากแผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ (One - way stratification with systematic pps - sampling) มาใช้กับแผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทาง

ก. การประมาณค่าเฉลี่ย

ในการจัดข้อมูลอย่างมีชั้นภูมิแบบสองทาง ได้แบ่งประเภทของหน่วยตัวอย่างออกเป็นแถวและสดมภ์ การกำหนดค่า M_{ijk} หรือค่าที่ใช้วัดขนาดของลักษณะที่ต้องการศึกษาของหน่วยตัวอย่างที่ k ในชั้นภูมิที่ ij จะทำได้ง่ายและดีกว่าในการจัดข้อมูลอย่างมีชั้นภูมิแบบทางเดียว เนื่องจากการจัดข้อมูลอย่างมีชั้นภูมิแบบสองทางมีการจัดประเภทของข้อมูลออกโดยละเอียด หน่วยตัวอย่างที่สุ่มได้จะให้ค่า M_{ijk} ที่มีความสัมพันธ์กับลักษณะที่ต้องการศึกษามากกว่าในการจัดชั้นภูมิแบบทางเดียว ซึ่งในการจัดชั้นภูมิแบบทางเดียวนี้หน่วยตัวอย่างจะถูกจัดประเภทออกอย่างหยาบ ๆ ค่า M_{ijk} ก็ถูกกำหนดขึ้นอย่างกว้าง ๆ M_{ijk} จึงมีความสัมพันธ์ใกล้ชิดกับลักษณะที่ต้องการศึกษาน้อย ฉะนั้น การแบ่งชั้นภูมิไม่ละเอียดเท่าไร การกำหนดค่า M_{ijk} ให้มีความสัมพันธ์ใกล้ชิดกับลักษณะที่ต้องการศึกษาย่อมทำได้ยากขึ้นเท่านั้น และการประมาณค่าเฉลี่ยเมื่อใช้การจัดชั้นภูมิแบบสองทางย่อมจะมีความคลาดเคลื่อนหรือความแปรปรวนน้อยกว่าการจัดชั้นภูมิแบบทางเดียว ดังเหตุผลที่ได้กล่าวมาแล้ว

ดังที่ได้กล่าวมาแล้วว่า ในการศึกษาขั้นนี้ได้อาศัยพื้นฐานของแผ่นแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจในการเลือกตัวอย่าง และการประมาณค่าพารามิเตอร์ที่สนใจ สูตรที่นำมาประยุกต์ใช้ คือ

$$\bar{y} = \frac{1}{\sum_{i=1}^{R} N_i}\left[\sum_{i=1}^{R} \frac{1}{n_i} \sum_{k=1}^{n_i} \frac{y_{ik}}{P_{ik}}\right]$$

เมื่อ $\bar{y}$ = ค่าเฉลี่ยที่ได้จากตัวอย่างขนาด

y_{ik} = ลักษณะที่ต้องการศึกษาของหน่วยตัวอย่างที่ k ในแถวที่ i

M_{ik} = ค่าที่ใช้วัดขนาดของลักษณะที่ต้องการศึกษาของหน่วยตัวอย่างที่ k ในแถวที่ i

P_{ik} = โอกาสในการเลือกหน่วยตัวอย่างที่ k ในแถวที่ i

$= \dfrac{M_{ik}}{\sum_{k=1}^{N_i} M_{ik}}$ โดยที่ $M_{ik} \propto y_{ik}$

N_i = จำนวนหน่วยตัวอย่างทั้งหมดในแถวที่ i

$\sum_{i=1}^{R} N_i$ = จำนวนหน่วยตัวอย่างทั้งหมดของทุกแถวที่เป็นตัวอย่าง

n_i = จำนวนตัวอย่างในแถวที่ i

ข. <u>การประมาณค่าความแปรปรวนของค่าเฉลี่ย</u> $\left[\hat{V}(\bar{y})\right]$

โดยอาศัยพื้นฐานของแผนแบบการสำรวจอย่างมีชั้นภูมิแบบทางเดียวแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ สูตรที่นำมาประยุกต์ใช้ คือ

$$\hat{V}(\bar{y}) = \frac{1}{\left(\sum_{i=1}^{R} N_i\right)^2} \sum_{i=1}^{R} \frac{1}{n_i(n_i-1)} \sum_{k=1}^{n_i}\left(\frac{y_{ik}}{P_{ik}} - \hat{T}_i(y)\right)^2$$

เมื่อ $\hat{T}_i(y)$ = ค่าประมาณของยอดรวมของ y ในแถวที่ i

$$= \frac{1}{n_i}\sum_{k=1}^{n_i}\frac{y_{ik}}{P_{ik}}$$

เนื่องจากแถวที่ i บางแถว มีจำนวนตัวอย่างเพียง ๑ หน่วย การประมาณค่าความแปรปรวนภายในแถวนั้นจึงไม่สามารถทำได้ ซึ่งทำให้ไม่สามารถหาความแปรปรวนรวม จึงใช้เทคนิคที่ใช้ประมาณค่าความแปรปรวนเมื่อตัวอย่างตกเพียง ๑ หน่วย ซึ่งเรียกว่า วิธีจัดกลุ่มชั้นภูมิเข้าด้วยกัน ดังได้กล่าวโดยละเอียดแล้วในหัวข้อ ๒.๑ ในที่นี้จะกล่าวเฉพาะการประยุกต์สูตรมาใช้

$V(\bar{Y})$ = ความแปรปรวนของแถวที่นำมาจัดกลุ่มเข้าด้วยกัน + ความแปรปรวนของแถวที่ไม่นำมาจัดกลุ่ม

$$= \frac{1}{\left(\sum_{i=1}^{R} N_i\right)^2 N_c^2}\sum_{(i,i')\,2k}\left\{N_i\overline{\left(\frac{y_i}{P_i}\right)} - N_{i'}\overline{\left(\frac{y_{i'}}{P_{i'}}\right)}\right\}^2 + \text{ความแปรปรวนของแถว}$$

ที่ไม่นำมาจัดกลุ่ม

เมื่อ

N_i = จำนวนหน่วยตัวอย่างทั้งหมดในแถวที่ i

$N_{i'}$ = จำนวนหน่วยตัวอย่างทั้งหมดในแถวที่ i' ที่นำมาจัดกลุ่มกับแถวที่ i

$\overline{\left(\frac{y_i}{P_i}\right)}$ = ค่าเฉลี่ย (sample mean) ของ $\left(\frac{y_i}{P_i}\right)$ ของแถวที่ i

$\overline{\left(\frac{y_{i'}}{P_{i'}}\right)}$ = ค่าเฉลี่ยของ $\left(\frac{y_{i'}}{P_{i'}}\right)$ ที่นำมาจัดกลุ่มกับแถวที่ i

N_C = จำนวนหน่วยตัวอย่างทั้งหมดในแถวที่จัดกลุ่มเข้าด้วยกัน

$\sum_{i=1}^{R} N_i$ = จำนวนหน่วยตัวอย่างทั้งหมดของทุกแถวที่เป็นตัวอย่าง

๒.๓.๓ การพิสูจน์ว่าถ้าสามารถกำหนดให้โอกาสในการเลือกหน่วยตัวอย่างที่ k ในแถวที่ i เป็นสัดส่วนกับลักษณะที่ต้องการศึกษาของหน่วยตัวอย่างที่ k ในแถวที่ i แล้ว จะสามารถประมาณค่าเฉลี่ยได้โดยไม่มีความคลาดเคลื่อนเลย

จาก $$V(\bar{y}) = \frac{1}{\left(\sum_{i=1}^{R} N_i\right)^2} \sum_{i=1}^{R} \frac{1}{n_i} \sum_{k=1}^{N_i} P_{ik}\left(\frac{y_{ik}}{P_{ik}} - T_i(y)\right)^2$$

ถ้าสามารถกำหนดให้ $P_{ik} \propto y_{ik}$

หรือ $P_{ik} = ty_{ik}$ เมื่อ t เป็นค่าคงที่ใด ๆ

เนื่องจาก $$\sum_{k=1}^{N_i} P_{ik} = 1$$

$$\sum_{k=1}^{N_i} ty_{ik} = 1$$

$$t = \frac{1}{\sum_{k-1}^{N_i} y_{ik}}$$

นั่นคือ $$P_{ik} = \frac{y_{ik}}{\sum_{k=1}^{N_i} y_{ik}}$$

แทนค่า P_{ik} ใน $V(\bar{y})$

$$V(\bar{y}) = \frac{1}{\left(\sum_{i=1}^{R} N_i\right)^2} \sum_{i=1}^{R} \frac{1}{n_i} \sum_{k=1}^{N_i} \frac{y_{ik}}{\sum_{k=1}^{N_i} y_{ik}}\left(\frac{y_{ik}}{\frac{y_{ik}}{\sum_{k=1}^{N_i} y_{ik}}} - T_i(y)\right)^2$$

$$= \frac{1}{\left(\sum_{i=1}^{R} N_i\right)^2} \sum_{i=1}^{R} \frac{1}{n_i} \sum_{k=1}^{N_i} \frac{y_{ik}}{\frac{N_i}{\sum_{k=1}^{N_i} y_{ik}}} \left(\sum_{k=1}^{N_i} y_{ik} - T_i(y) \right)^2$$

$$= \quad 0 \qquad ; \quad \left(T_i(y) = \sum_{k=1}^{N_i} y_{ik} \right)$$

นั่นคือ ถ้าสามารถกำหนดให้โอกาสในการเลือกหน่วยตัวอย่างที่ k ในแถวที่ i เป็นสัดส่วนกับลักษณะที่ต้องการศึกษาของหน่วยตัวอย่างที่ k ในแถวที่ i แล้ว จะสามารถประมาณค่าเฉลี่ยได้โดยไม่มีความคลาดเคลื่อนเลย

๒.๓.๔ <u>การเปรียบเทียบประสิทธิภาพระหว่างแผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางแบบมีระบบ โดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ กับแผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางโดยใช้ตัวอย่างจำนวนน้อย</u>

ในการเปรียบเทียบประสิทธิภาพระหว่างแผนแบบทั้งสองนี้ ไม่มีรูปแบบทางคณิตศาสตร์ที่จะบอกได้แน่นอนว่าการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ มีประสิทธิภาพดีกว่า การใช้แผนแบบการสำรวจอย่างมีชั้นภูมิโดยใช้ตัวอย่างจำนวนน้อยเสมอ เนื่องจากในการใช้แผนแบบการสำรวจแบบแรกนั้น ประสิทธิภาพในการสำรวจขึ้นอยู่กับว่าผู้วางแผนการสำรวจสามารถเลือกใช้ค่าที่ใช้วัดขนาด (M_{ik}) ได้เหมาะสมเพียงใด M_{ik} ที่ใช้จะต้องมีความสัมพันธ์โดยใกล้ชิดกับลักษณะที่ต้องการศึกษา (y_{ik}) ถ้าสามารถเลือกใช้ M_{ik} ได้เหมาะสมเท่าไร ก็จะทำให้ได้ค่าประมาณที่มีความแปรปรวนหรือความคลาดเคลื่อนน้อยลงเท่านั้น แต่ถ้าสามารถกำหนดให้ $M_{ik} = y_{ik}$ แล้ว ค่าประมาณที่ได้จะไม่มีความคลาดเคลื่อนเลย ดังนั้น หลักการที่สำคัญในการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ คือ พยายามเลือกใช้โอกาสในการเลือก (P_{ik}) ให้เป็นปฏิภาคกับ M_{ik} ซึ่ง $M_{ik} \propto y_{ik}$ แต่ถ้า $M_{ik} = y_{ik}$ แล้ว แผนแบบการสำรวจนี้จะให้ค่าประมาณที่มีประสิทธิภาพสูงที่สุด

แต่ถ้าผู้วางแผนการสำรวจกำหนดค่า M_{ik} ได้ไม่เหมาะสม หรือไม่มีความสัมพันธ์กับลักษณะที่ต้องการศึกษา การใช้แผนแบบการสำรวจอย่างมีชั้นภูมิแบบสองทางแบบมีระบบโดยให้ความน่าจะเป็นของการเลือกหน่วยตัวอย่างเป็นปฏิภาคกับขนาดของหน่วยตัวอย่างที่ต้องการสำรวจ ก็อาจจะไม่ดีไปกว่าการใช้แผนแบบการสำรวจอย่างมีชั้นภูมิโดยใช้ตัวอย่างจำนวนน้อย

ดังนั้น ในการเปรียบเทียบแผนแบบทั้งสองนี้ จึงไม่สามารถสรุปได้ว่าการใช้แผนแบบใดจะให้ค่าประมาณที่มีประสิทธิภาพดีกว่า ดังเหตุผลที่กล่าวแล้ว