บทที่ 2

ทฤษฎีที่ใช้ในการวิจัยและผลงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่ใช้ในการวิจัย

ในการศึกษา เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณในสถานการณ์ที่ไม่ทราบค่าสัดส่วนของชั้นภูมิทั้งในขั้นตอนการวางแผนและขั้นตอนการประมาณค่ากับสถานการณ์ที่ไม่ทราบค่าสัดส่วนของชั้นภูมิในขั้นตอนการวางแผน แต่ทราบในขั้นตอนการประมาณค่า Shambhu Dayal ได้เสนอตัวประมาณสำหรับค่าเฉลี่ยของประชากร ตลอดจนความแปรปรวนของตัวประมาณดังกล่าว ตามทฤษฎีซึ่งจำแนกตามแต่ละสถานการณ์ที่ศึกษาได้ดังต่อไปนี้

2.1.1 <u>ไม่ทราบค่าสัดส่วนของชั้นภูมิทั้งในขั้นตอนการวางแผนและขั้นตอนการประมาณค่า</u> ในการศึกษาได้แบ่งออกเป็น 2 กรณี ดังนี้

2.1.1.1 <u>ทราบค่าประมาณของสัดส่วนของชั้นภูมิและความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิแล้ว</u>

กรณีนี้เป็นกรณีที่ไม่ทราบค่าสัดส่วนของชั้นภูมิ ไม่ว่าขั้นตอนใดของการวิจัย แต่ได้ทราบค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิ แล้วจากการสำรวจเบื้องต้น หรือการสำรวจอื่น ๆ ซึ่งก็หมายถึงการใช้วิธีการของการเลือกตัวอย่างสองครั้ง ขนาด n' เพื่อประมาณค่าของ W_h , S_h นั่นเอง

ดังนั้นจึงใช้ค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิที่ได้จากการสำรวจเบื้องต้น หรือการสำรวจอื่น ๆ ขนาด n' ดังกล่าว ทั้งในช่วงของการกำหนดขนาดตัวอย่างในแต่ละชั้นภูมิ (n_h) และในช่วงของการหาค่าประมาณของค่าเฉลี่ยของประชากร ($\bar{y}_{st}$)

ตัวประมาณที่ได้จากกรณีนี้ คือ

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h \qquad \text{...........(1.1.1)}$$

โดยที่ w_h คือ ค่าประมาณของ W_h ซึ่งได้จากการสำรวจเบื้องต้น ขนาด n' และจะพบว่าตัวประมาณนี้มีคุณสมบัติของความไม่เอนเอียง ซึ่งสามารถแสดงได้ดังนี้

$$
\begin{aligned}
E(\bar{y}_{st}) &= E\,\{E(\Sigma w_h \bar{y}_h \mid w_h)\} && \ldots\ldots\ldots\ldots(1.1.2) \\
&= E\,\{\Sigma w_h \bar{Y}_h\} \\
&= \Sigma W_h \bar{Y}_h \\
&= \bar{Y} && \ldots\ldots\ldots\ldots(1.1.3)
\end{aligned}
$$

จากสมการ (1.1.1) ,(1.1.2) จะพบว่าตัวประมาณอยู่ในรูปฟังก์ชันของ w_h เนื่องจากในขั้นตอนการประมาณค่าก็ยังคงไม่ทราบค่าสัดส่วนของชั้นภูมิ โดยตัวประมาณจะมีความแปรปรวนเมื่อไม่คิดค่า finite population correction หรือ f.p.c. ดังนี้

$$
\begin{aligned}
V(\bar{y}_{st}) &= \frac{(\Sigma W_h S_h)^2}{n} + \frac{1}{nn'}\Big[\,\Sigma W_h (S_h - \bar{S})^2 \\
&\quad + \frac{B_2 - 1}{4}\,\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\}\Big] \\
&\quad - \frac{B_2 - 1}{4nn'^2}\,\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} \\
&\quad + \frac{\Sigma W_h (\bar{Y}_h - \bar{Y})^2}{n'} && \ldots\ldots\ldots\ldots(1.1.4)
\end{aligned}
$$

โดยที่ $\bar{S} = \Sigma W_h S_h$ และ $B_2 = \mu_4/S^4$

<u>พิสูจน์</u>

$$
V(\bar{y}_{st}) = E\,\Big[\,E\,\{(\bar{y}_{st} - \bar{Y})^2 \mid w_1, w_2, \ldots, w_L\}\Big] \quad \ldots\ldots(1.1.5)
$$

พิจารณา $E\{(\bar{y}_{st} - \bar{Y})^2 \mid w_1, w_2, \ldots, w_L\} = \Sigma \dfrac{w_h^2 S_h^2}{n_h} + \{\Sigma(w_h - W_h)\, \bar{Y}_h\}^2$ ...(1.1.6)

ดังนั้นจะได้ว่า $V(\bar{y}_{st}) = E\{\Sigma \dfrac{w_h^2 S_h^2}{n_h}\} + E\{\Sigma(w_h - W_h)\, \bar{Y}_h\}^2$ ... (1.1.7)

จากสมการ (1.1.6) พิจารณาเทอมแรกของสมการด้านขวา แทนค่า $n_h = nw_h s_h / \Sigma w_h s_h$ ซึ่ง w_h , s_h เป็นค่าประมาณของ W_h , S_h จากการสำรวจเบื้องต้น หรือการสำรวจอื่น ๆ ขนาด n' จะได้

$$\Sigma \frac{w_h^2 S_h^2}{n_h} = \frac{1}{n} \Sigma \frac{w_h^2 S_h^2}{w_h s_h} (\Sigma w_h s_h)$$

$$= \frac{1}{n} \{\Sigma w_h^2 S_h^2 + \sum_{h \neq j} w_h w_j S_j^2 \frac{s_h}{s_j}\}$$

สำหรับการทำการสำรวจเบื้องต้น หรือการสำรวจอื่น ๆ ขนาด n' จะได้ว่า $n_1', n_2', \ldots, n_L'$ ($n_h' \geq 2$, $h = 1,2,\ldots,L$) มีการแจกแจงแบบ Truncated Hypergeometric Distribution ดังนั้นจึงทำให้

$$E(n_h') = n'W_h$$

$$V(n_h') = n'W_h(1 - W_h)$$

$$Cov(n_h', n_j') = -n'W_h W_j$$

ในกรณีที่ n' มีขนาดใหญ่ ค่าต่าง ๆ เหล่านี้จะมีผลทำให้

$$E(w_h^2) = W_h^2 + (W_h(1 - W_h))/n'$$

$$E(w_h w_j) = W_h W_j - W_h W_j / n'$$

$$= ((n' - 1)/n')\, W_h W_j$$

โดยที่ $w_h = n'_h/n'$ $(h = 1,2,\ldots,L)$ และนอกจากนี้เมื่อ n' มีขนาดใหญ่ ประกอบกับการใช้ Taylor's Expansion (Sukhatme, P.V. and Sukhatme,B.V. 1970) สามารถพิสูจน์ได้ว่า

$$
\begin{aligned}
E(w_h s_h^2) &= E\{E(w_h s_h^2 \mid n'_h \geq 2)\} \\
&= E(w_h S_h^2) \\
&= W_h S_h^2
\end{aligned}
$$

และได้
$$
\begin{aligned}
E(w_h w_j S_j^2 (s_h/s_j)) &= E\{E(w_h w_j S_j^2 (s_h/s_j) \mid n'_h, n'_j \geq 2)\} \\
&= E\{w_h w_j S_j S_h \left(1 + \frac{C^2}{4}\right)\} \\
&= ((n'-1)/n')\, W_h W_j S_h S_j \left(1 + \frac{C^2}{4}\right)
\end{aligned}
$$

โดยที่ C คือ ค่าสัมประสิทธิ์ของความแปรปรวนของ s^2 ซึ่งมีข้อสมมติให้ค่าสัมประสิทธิ์ของความแปรปรวนมีค่าคงที่ในทุก ๆ ชั้นภูมิ และจะได้ผลต่อไปว่า

$$
\begin{aligned}
E\{\Sigma \frac{w_h^2 S_h^2}{n_h}\} &= E\left[\frac{1}{n}\{\Sigma w_h^2 S_h^2 + \Sigma w_h w_j S_j^2 \frac{s_h}{s_j}\}\right] \\
&= \frac{1}{n}\left[E\{\Sigma w_h^2 S_h^2\} + E\{\Sigma w_h w_j S_j^2 \frac{s_h}{s_j}\}\right] \\
&= \frac{1}{n}\left[\Sigma\{W_h^2 + \frac{W_h(1 - W_h)}{n'}\} S_h^2\right. \\
&\quad \left. + \frac{n'-1}{n'}\left(1 + \frac{C^2}{4}\right) \Sigma W_h W_j S_h S_j\right]
\end{aligned}
$$

$$= \frac{1}{n}\left[(\Sigma W_h S_h)^2 - \frac{(\Sigma W_h S_h)^2}{n'} + \frac{\Sigma W_h S_h^2}{n'} \right.$$

$$+ \frac{C^2}{4}\{ (\Sigma W_h S_h)^2 - (\Sigma W_h^2 S_h^2)$$

$$\left. - \frac{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2}{n'} \} \right] \qquad ....(1.1.8)$$

แทนค่า $C^2 = (B_2 - 1)/n'$ (Sukhatme, P.V. and Sukhatme, B.V. 1970)

ลงในสมการ (1.1.8) จะได้ว่า

$$E\{ \Sigma \frac{w_h^2 s_h^2}{n_h} \} = \frac{(\Sigma W_h S_h)^2}{n} + \frac{1}{nn'}\left[\Sigma W_h (S_h - \bar{S})^2 \right.$$

$$\left. + \frac{B_2 - 1}{4}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} \right]$$

$$- \frac{B_2 - 1}{4nn'^2}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} \qquad ....(1.1.9)$$

สำหรับเทอมที่ 2 ของสมการด้านขวาของ (1.1.7) เมื่อกำหนดให้ $(N-n')/(N-1) = 1$

จะมีค่าดังนี้

$$E\{ \Sigma (w_h - W_h) \bar{y}_h\}^2 = \frac{\Sigma W_h (\bar{Y}_h - \bar{Y})^2}{n'} \qquad ......(1.1.10)$$

จากสมการ (1.1.7) (1.1.8) และ (1.1.10) สามารถสรุปได้ว่า

$$V(\bar{y}_{st}) = \frac{(\Sigma W_h S_h)^2}{n} + \frac{1}{nn'}\left[\Sigma W_h (S_h - \bar{S})^2 \right.$$

$$\left. + \frac{B_2 - 1}{4}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} \right]$$

$$- \frac{B_2 - 1}{4nn'^2}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} + \frac{\Sigma W_h (\bar{Y}_h - \bar{Y})^2}{n'} \qquad ...(1.1.11)$$

สำหรับค่าประมาณของความแปรปรวนของตัวประมาณตามสมการที่ (1.1.11) คือ

$$\hat{V}(\bar{y}_{st}) = \frac{(\Sigma w_h s_h)^2}{n} + \frac{1}{nn'}\left[\Sigma w_h (s_h - \bar{s})^2 + \frac{B_2' - 1}{4}\{(\Sigma w_h s_h)^2 - \Sigma w_h^2 s_h^2\}\right] - \frac{B_2' - 1}{4nn'^2}\{(\Sigma w_h s_h)^2 - \Sigma w_h^2 s_h^2\} + \frac{\Sigma w_h (\bar{y}_h - \bar{y})^2}{n'} \qquad \text{........(1.1.12)}$$

โดยที่ B_2' , w_h, s_h, $\bar{y}_h$ และ $\bar{y}$ คือค่าประมาณของ B_2, W_h, S_h, $\bar{Y}_h$ และ $\bar{Y}$ ซึ่งได้จากการทำการสำรวจเบื้องต้น หรือการสำรวจอื่น ๆ ขนาด n' ตามที่กล่าวแล้ว

2.1.1.2 <u>ไม่ทราบค่าประมาณของสัดส่วนของชั้นภูมิและความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิมาก่อน</u>

สำหรับกรณีนี้เป็นกรณีที่ไม่ทราบค่าสัดส่วนของชั้นภูมิไม่ว่าขั้นตอนใดของการวิจัย รวมทั้งไม่ทราบค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิมาก่อน นั่นก็คือ ไม่เคยมีการสำรวจข้อมูลที่จะใช้ในการวิจัยมาเลยในอดีตไม่ว่าจะเป็น การสำรวจเบื้องต้น หรือการสำรวจอื่นใดก็ตาม

ดังนั้นในการวิจัย จะทำการสำรวจข้อมูลโดยแบ่งขอบข่ายออกเป็น 2 ส่วน ในส่วนที่ 1 ขนาด n' จะใช้ในการหาค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิ หลังจากนั้นจะนำค่าประมาณเหล่านี้ไปใช้หาค่าต่าง ๆ ในส่วนที่ 2 ขนาด n'' ต่อไป $(n = n' + n'')$ ทั้งนี้จะเห็นว่ายังคงใช้ค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิทั้งในช่วงของการกำหนดขนาดตัวอย่างในแต่ละชั้นภูมิ (n_h) และในช่วงของการหาค่าประมาณของค่าเฉลี่ยของประชากร $(\bar{y}_{st})$

สำหรับตัวประมาณที่ได้จากกรณีนี้ นอกจากจะเป็นตัวประมาณที่เกิดจากการนำค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิ มาใช้ในส่วนที่ 2 แล้ว Shambhu Dayal ได้เสนอให้นำตัวประมาณในส่วนที่ 1

มาใช้ด้วย ทั้งนี้เพื่อเป็นการใช้ข่าวสาร (information) ที่ได้จากการสำรวจอย่างครบถ้วน โดยกำหนดตัวประมาณเป็น 2 แบบ คือ

2.1.1.2.1 นำตัวประมาณจากส่วนที่ 1 ($\bar{y}'$) และส่วนที่ 2 ($\bar{y}''_{st}$) มาเฉลี่ยโดยให้น้ำหนักเท่ากัน

ตัวประมาณที่ได้จากกรณีนี้ คือ

$$\bar{y}_{st} = \frac{\bar{y}' + \bar{y}''_{st}}{2} \qquad ....(1.2.1.1)$$

โดยที่ $\bar{y}'$ คือ ตัวประมาณจากส่วนที่ 1 ขนาด n ซึ่งมีลักษณะเป็นการเลือกตัวอย่างสุ่มอย่างง่าย เนื่องจากใช้ข้อมูลจากหน่วยตัวอย่างทุกหน่วยของส่วนที่ 1 นั่นคือ

$$\bar{y}' = \sum_{i=1}^{n'} y_i / n'$$

และ $\bar{y}''_{st}$ คือ ตัวประมาณจากส่วนที่ 2 ขนาด n ซึ่งมีลักษณะเป็นการเลือกตัวอย่างมีชั้นภูมิแบบสุ่มอย่างง่าย โดยนำค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบน-มาตรฐานของแต่ละชั้นภูมิที่หาได้จากส่วนที่ 1 มาใช้ในส่วนที่ 2 มีสูตรดังนี้

$$\bar{y}''_{st} = \sum_{h=1}^{L} w_h \bar{y}''_h$$

โดยที่ w_h คือ ค่าประมาณของ W_h ซึ่งได้จากการประมาณค่าในส่วนที่ 1 และเราจะพบว่าตัวประมาณตามสมการ (1.2.1,1) มีคุณสมบัติของความไม่เอนเอียง สำหรับการพิสูจน์แสดงได้ดังนี้

$$E(\bar{y}_{st}) = \frac{E(\bar{y}') + E(\bar{y}''_{st})}{2} \qquad .....(1.2.1.2)$$

พิจารณาเทอมแรกของสมการด้านขวาของ (1.2.1.2) คือ $E(\bar{y}')$ จะเห็นว่าเนื่องจากเราใช้ทุกหน่วยตัวอย่างจากการสำรวจในส่วนที่ 1 จึงมีลักษณะเป็นการเลือกตัวอย่างสุ่มอย่างง่าย ซึ่งจากคุณสมบัติของการเลือกตัวอย่างสุ่มอย่างง่าย จะได้ว่าตัวประมาณมีคุณสมบัติของความไม่

เอนเอียง ดังนี้

$$E(\bar{y}') = E\left\{ \frac{1}{n'} \sum_{i=1}^{n'} y_i \right\} \qquad \text{......(1.2.1.3)}$$

$$= \frac{1}{n'} \sum_{i=1}^{n'} \{ E(y_i) \}$$

$$= \frac{1}{n'} \sum_{i=1}^{n'} \bar{Y}$$

$$= \bar{Y} \qquad \text{......(1.2.1.4)}$$

สำหรับเทอมที่ 2 ของสมการด้านขวาของ (1.2.1.2) คือ $E(\bar{y}_{st})$ จะเห็นว่ามีลักษณะเป็นการเลือกตัวอย่างมีชั้นภูมิแบบสุ่มอย่างง่าย โดยนำค่า w_h , s_h ที่ประมาณจากส่วนที่ 1 ขนาด n' มาใช้ในส่วนที่ 2 ขนาด n'' ซึ่งสามารถพิสูจน์ได้ว่าตัวประมาณมีคุณสมบัติของความไม่เอนเอียง ดังนี้

$$E(\bar{y}''_{st}) = E\{E(\Sigma w_h \bar{y}''_h \mid w_h)\} \qquad \text{......(1.2.1.5)}$$

$$= E\{\Sigma w_h \bar{Y}_h\}$$

$$= \Sigma W_h \bar{Y}_h$$

$$= \bar{Y} \qquad \text{......(1.2.1.6)}$$

จากสมการ (1.2.1.2) (1.2.1.4) และ (1.2.1.6) สามารถสรุปได้ว่า

$$E(\bar{y}_{st}) = \frac{\bar{Y} + \bar{Y}}{2}$$

$$= \bar{Y} \qquad \text{......(1.2.1.7)}$$

สำหรับความแปรปรวนของตัวประมาณตามสมการ (1.2.1.1) มีค่าดังนี้ (ไม่คิดค่า f.p.c.)

$$V(\bar{y}_{st}) = \frac{1}{4}\Big[\frac{\Sigma W_h S_h^2}{n'} + \frac{\Sigma\, W_h(\bar{Y}_h-\bar{Y})^2}{n'} + \frac{(\Sigma W_h S_h)^2}{n''}$$

$$+ \frac{1}{n'n''}\{\Sigma W_h(S_h-\bar{S})^2 + \frac{B_2-1}{4}\left[(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\right]\}$$

$$- \frac{B_2-1}{4n'^2 n''}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} + \frac{\Sigma W_h(\bar{Y}_h - \bar{Y})^2}{n'}\Big] \quad \ldots(1.2.1.8)$$

<u>พิสูจน์</u>

$$V(\bar{y}_{st}) = V\{\frac{\bar{y}' + \bar{y}''_{st}}{2}\} \quad \ldots\ldots\ldots(1.2.1.9)$$

$$= \frac{1}{4}\left[V(\bar{y}') + V(\bar{y}''_{st}) + 2\,Cov(\bar{y}', \bar{y}''_{st})\right]$$

เนื่องจากการประมาณ $\bar{y}'$ กับ $\bar{y}''_{st}$ กระทำในแต่ละส่วน โดย $\bar{y}'$ ได้จากการสำรวจในส่วนที่ 1 และ $\bar{y}''_{st}$ ได้จากการสำรวจในส่วนที่ 2 จึงทำให้ $Cov(\bar{y}', \bar{y}''_{st}) = 0$ ดังนั้นจะได้ว่า

$$V(\bar{y}_{st}) = \frac{1}{4}\left[V(\bar{y}') + V(\bar{y}''_{st})\right] \quad \ldots\ldots\ldots(1.2.1.10)$$

พิจารณาเทอมแรกของสมการด้านขวาของ (1.2.1.10) คือ $V(\bar{y}')$ จะเห็นว่ามีลักษณะเป็นการเลือกตัวอย่างสุ่มอย่างง่าย ซึ่งให้ค่าความแปรปรวนดังนี้

$$V(\bar{y}') = \frac{\Sigma W_h S_h^2}{n'} + \frac{\Sigma W_h(\bar{Y}_h - \bar{Y})^2}{n'} \quad \ldots\ldots\ldots(1.2.1.11)$$

ส่วนเทอมที่ 2 ของสมการด้านขวาของ (1.2.1.10) คือ $V(\bar{y}''_{st})$ ซึ่งมีลักษณะเป็นการเลือกตัวอย่างมีชั้นภูมิแบบสุ่มอย่างง่าย โดยจะให้ค่าความแปรปรวนดังนี้

$$V(\bar{y}''_{st}) \quad = \quad E\left[E\{(\bar{y}''_{st} - \bar{Y})^2 \mid w_1, w_2, \ldots, w_L\}\right] \quad \ldots\ldots(1.2.1.12)$$

พิจารณา $E\{(\bar{y}''_{st} - \bar{Y})^2 \mid w_1, w_2, \ldots, w_L\} = \Sigma \frac{w_h^2 S_h^2}{n''_h} + \{\Sigma(w_h - W_h)\bar{Y}_h\}^2 \ldots(1.2.1.13)$

ดังนั้น $V(\bar{y}''_{st}) = E\{\Sigma \frac{w_h^2 S_h^2}{n''_h}\} + E\{\Sigma(w_h - W_h)\bar{Y}_h\}^2 \quad \ldots\ldots(1.2.1.14)$

จะเห็นว่าสมการ (1.2.1.14) อยู่ในรูปแบบเดียวกับสมการ (1.1.7) การพิสูจน์ต่อไปจึงมีรูปแบบและวิธีการทำนองเดียวกัน ตั้งแต่ขั้นตอนของสมการ (1.1.7) ถึงขั้นตอนของสมการ (1.1.11) ทั้งนี้โดยการแทนค่า $n''_h = n'' w_h s_h / \Sigma w_h s_h$ ของการพิสูจน์ในกรณีดังกล่าวนี้แทนการใช้ค่า $n_h = n w_h s_h / \Sigma w_h s_h$ ของการพิสูจน์สมการ (1.1.7) ถึงสมการ (1.1.11) ดังนั้นจะได้ว่า

$$\begin{aligned} V(\bar{y}''_{st}) &= \frac{(\Sigma W_h S_h)^2}{n''} + \frac{1}{n'n''}\Big[\Sigma W_h(S_h - \bar{S})^2 \\ &\quad + \frac{B_2 - 1}{4}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\}\Big] \\ &\quad - \frac{B_2 - 1}{4n'^2 n''}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} + \frac{\Sigma W_h(\bar{Y}_h - \bar{Y})^2}{n'} \quad \ldots(1.2.1.15) \end{aligned}$$

จากสมการ (1.2.1.10) (1.2.1.11) และ (1.2.1.15) สามารถสรุปได้ว่า

$$\begin{aligned} V(\bar{y}_{st}) &= \frac{1}{4}\Big[\frac{\Sigma W_h S_h^2}{n'} + \frac{\Sigma W_h(\bar{Y}_h - \bar{Y})^2}{n'} + \frac{(\Sigma W_h S_h)^2}{n''} \\ &\quad + \frac{1}{n'n''}\{\Sigma W_h(S_h - \bar{S})^2 + \frac{B_2 - 1}{4}[(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2]\} \\ &\quad - \frac{B_2 - 1}{4n'^2 n''}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} + \frac{\Sigma W_h(\bar{Y}_h - \bar{Y})^2}{n'}\Big] \ldots\ldots(1.2.1.16) \end{aligned}$$

สำหรับค่าประมาณของความแปรปรวนของตัวประมาณตามสมการที่ (1.2.1.16) คือ

$$\hat{V}(\bar{y}_{st}) = \frac{1}{4}\Big[\frac{\Sigma w_h s_h^2}{n'} + \frac{\Sigma w_h(\bar{y}_h - \bar{y})^2}{n'} + \frac{(\Sigma w_h s_h)^2}{n''}$$

$$+ \frac{1}{n' n''}\{ \Sigma w_h (s_h - \bar{s})^2 + \frac{B_2' - 1}{4}\left[(\Sigma w_h s_h)^2 - \Sigma w_h^2 s_h^2\right] \}$$

$$- \frac{B_2' - 1}{4n'^2 n''}\{(\Sigma w_h s_h)^2 - \Sigma w_h^2 s_h^2\} + \frac{\Sigma w_h(\bar{y}_h - \bar{y})^2}{n'} \Big] \quad ..(1.2.1.17)$$

จากสมการ (1.2.1.16) อาจเขียนในอีกรูปแบบหนึ่งได้ว่า

$$V(\bar{y}_{st}) = \frac{v}{4n'} + \frac{m}{4n''} + \frac{t}{4n' n''} + \frac{f}{4n'^2 n''} \quad ......(1.2.1.18)$$

โดยที่ v, m, t และ f มีค่าดังนี้

$$v = \Sigma W_h S_h^2 + 2\Sigma W_h(\bar{Y}_h - \bar{Y})^2$$

$$m = (\Sigma W_h S_h)^2$$

$$t = \Sigma W_h(S_h - \bar{S})^2 + \frac{B_2 - 1}{4}\left[(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\right]$$

$$f = -\frac{B_2 - 1}{4}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\}$$

สำหรับการกำหนดขนาดตัวอย่างในส่วนที่ 1 และขนาดตัวอย่างในส่วนที่ 2 ซึ่งให้ค่าความแปรปรวนของตัวประมาณมีค่าต่ำสุด จะไม่คำนึงถึงเทอมที่มี $4n'^2 n''$ เป็นตัวหาร เนื่องจากมีค่าน้อย จึงไม่มีผลกระทบต่อการกำหนดขนาดตัวอย่างในแต่ละส่วนมากนัก นั่นคือ สมการด้านขวาของ (1.2.1.18) จะลดลงเหลือเพียง 3 เทอม ดังนี้

$$V(\bar{y}_{st}) = \frac{v}{4n'} + \frac{m}{4n''} + \frac{t}{4n'n''} \quad \text{..........(1.2.1.19)}$$

จะเห็นว่าขนาดตัวอย่างในส่วนที่ 1 ซึ่งทำให้ความแปรปรวนของตัวประมาณตามสมการ (1.2.1.19) มีค่าต่ำสุดก็ต่อเมื่อเป็นไปตามเงื่อนไขต่อไปนี้

$$\frac{d}{dn'} V(\bar{y}_{st}) = 0 \quad \text{............(1.2.1.20)}$$

แทนค่า $V(\bar{y}_{st})$ จากสมการ (1.2.1.19) ลงในสมการ (1.2.1.20) จะได้ว่า

$$\frac{d}{dn'} \left[\frac{v}{4n'} + \frac{m}{4n''} + \frac{t}{4n'n''} \right] = 0 \quad \text{.......(1.2.1.21)}$$

หลังจากนั้นแทนค่า $n'' = n - n'$ และใช้หลักการของการหาอนุพันธ์ (Differentiation Methods) เพื่อหาค่า n' ซึ่งทำให้ $V(\bar{y}_{st})$ มีค่าต่ำสุด

$$\frac{d}{dn'} \left[\frac{v}{4n'} + \frac{m}{4(n-n')} + \frac{t}{4n'(n-n')} \right] = 0 \quad \text{.....(1.2.1.22)}$$

$$\frac{v}{4}\left[- \frac{1}{n'^2} \right] + \frac{m}{4}\left[\frac{1}{(n-n')^2} \right] + \frac{t}{4}\left[\frac{2n' - n}{n'^2 (n-n')^2} \right] = 0$$

$$\frac{- v(n-n')^2 + mn'^2 + 2n't - nt}{4n'^2 (n-n')^2} = 0 \quad \text{......(1.2.1.23)}$$

จากสมการ (1.2.1.23) จะเห็นว่า $4n'^2(n-n')^2 \neq 0$ และเป็นเทอมที่หาค่าได้ (finite term) ดังนั้นจะได้ว่า

$$-v(n-n')^2 + mn'^2 + 2n't - nt = 0 \quad ......(1.2.1.24)$$

$$-vn^2 + 2vnn' - vn'^2 + mn'^2 + 2n't - nt = 0$$

$$(m-v)\,n'^2 + 2(vn+t)n' - (vn+t)n = 0 \quad ......(1.2.1.25)$$

เนื่องจากสมการ (1.2.1.25) อยู่ในรูปแบบของสมการกำลังสอง (Quadratic Equation) โดยมี n' เป็นตัวแปร ดังนั้นจึงสามารถใช้สูตรทั่วไปในการหาราก (Root) ของสมการดังกล่าวได้ ซึ่งจะได้ค่าของ n' ดังนี้

$$n' = \frac{-2(vn+t) \pm \sqrt{4(vn+t)^2 + 4(m-v)(vn+t)n}}{2(m-v)}$$

$$= \frac{-(vn+t) \pm \sqrt{(vn+t)^2 + (m-v)(vn+t)n}}{m-v} \quad .....(1.2.1.26)$$

ส่วน n'' จะมีค่าดังนี้

$$n'' = n - n'$$

$$= n - \left[\frac{-(vn+t) \pm \sqrt{(vn+t)^2 + (m-v)(vn+t)n}}{m-v}\right] \quad ..(1.2.1.27)$$

2.1.1.2.2 <u>นำตัวประมาณจากส่วนที่ 1 ($\bar{y}'$) และส่วนที่ 2 ($\bar{y}''_{st}$) มาเฉลี่ยโดยให้น้ำหนักตามขนาดตัวอย่างในแต่ละส่วน</u>

ตัวประมาณที่ได้จากกรณีนี้ คือ

$$\bar{y}_{st} = \frac{n'\bar{y}' + \Sigma n''_h \bar{y}''_h}{n' + n''}$$

$$= \frac{(n'/n'')\bar{y}' + \Sigma w''_h \bar{y}''_h}{1 + (n'/n'')} \quad ....(1.2.2.1)$$

โดยที่ $w_h'' = n_h''/n''$ ซึ่งหาได้จากส่วนที่ 2 แต่จะพบว่าการใช้ w_h ดังกล่าวนี้ในสมการ (1.2.2.1) จะทำให้ตัวประมาณที่ได้รับมีความเอนเอียง (biased) สำหรับการแก้ปัญหา Shambhu Dayal ได้เสนอให้ใช้ w_h ซึ่งเป็นค่าประมาณของ W_h แทนการใช้ w_h'' ดังนั้นตัวประมาณที่จะนำไปใช้จริงในกรณีนี้ คือ

$$\bar{y}_{st} = \frac{(n'/n'')\bar{y}' + \Sigma w_h \bar{y}_h''}{1 + (n'/n'')}$$

$$= \frac{(n'/n'')\bar{y}' + \bar{y}_{st}''}{1 + (n'/n'')} \qquad \text{........(1.2.2.2)}$$

โดยที่ w_h คือ ค่าประมาณของ W_h ซึ่งได้จากการประมาณค่าในส่วนที่ 1 และเราจะพบว่าตัวประมาณตามสมการ (1.2.2.2) มีคุณสมบัติของความไม่เอนเอียง สำหรับการพิสูจน์แสดงได้ดังนี้

$$E(\bar{y}_{st}) = \frac{E(n'/n'')\bar{y}' + E(\bar{y}_{st}'')}{1 + (n'/n'')} \qquad \text{........(1.2.2.3)}$$

พิจารณาเทอมแรกของสมการด้านขวาของ (1.2.2.3) คือ $E(n'/n'')\bar{y}'$ จะเห็นว่าเนื่องจากเราใช้ทุกหน่วยตัวอย่างจากการสำรวจในส่วนที่ 1 จึงมีลักษณะเป็นการเลือกตัวอย่างสุ่มอย่างง่าย ซึ่งจากคุณสมบัติของการเลือกตัวอย่างสุ่มอย่างง่ายจะได้ว่าตัวประมาณมีคุณสมบัติของความไม่เอนเอียง ดังนั้นจะได้ว่า

$$E(n'/n'')\bar{y}' = (n'/n'')E(\bar{y}')$$

$$= (n'/n'')\bar{Y} \qquad \text{........(1.2.2.4)}$$

สำหรับเทอมที่ 2 ของสมการด้านขวาของ (1.2.2.3) คือ $E(\bar{y}_{st}'')$ จะเห็นว่ามีลักษณะเป็นการเลือกตัวอย่างมีชั้นภูมิแบบสุ่มอย่างง่าย ซึ่งสามารถพิสูจน์ได้ว่าตัวประมาณมีคุณสมบัติของความไม่เอนเอียงดังนี้

$$E(\bar{y}''_{st}) = E\{E(\Sigma w_h \bar{y}''_h | w_h)\} \qquad .....(1.2.2.5)$$

$$= E\{\Sigma w_h \bar{Y}_h\}$$

$$= \Sigma W_h \bar{Y}_h$$

$$= \bar{Y} \qquad .....(1.2.2.6)$$

จากสมการ (1.2.2.3), (1.2.2.4) และ (1.2.2.6) สามารถสรุปได้ว่า

$$E(\bar{y}_{st}) = \frac{(n'/n'')\bar{Y} + \bar{Y}}{1+(n'/n'')}$$

$$= \frac{((n'/n'') + 1)\bar{Y}}{1+(n'/n'')}$$

$$= \bar{Y} \qquad .....(1.2.2.7)$$

สำหรับความแปรปรวนของตัวประมาณตามสมการ (1.2.2.2) มีค่าดังนี้ (ไม่คิดค่า f.p.c)

$$V(\bar{y}_{st}) = \frac{1}{(1+(n'/n''))^2}\Big[\{\frac{n'}{n''}\}^2 \{\frac{\Sigma W_h S_h^2}{n'} + \frac{\Sigma W_h(\bar{Y}_h-\bar{Y})^2}{n'}\} + \frac{(\Sigma W_h S_h)^2}{n''}$$

$$+ \frac{1}{n' n''}\{\Sigma W_h(S_h-\bar{S})^2 + \frac{B_2 - 1}{4}[(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2]\}$$

$$- \frac{B_2-1}{4n'^2 n''}\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} + \frac{\Sigma W_h(\bar{Y}_h-\bar{Y})^2}{n'}\Big] \qquad ....(1.2.2.8)$$

<u>พิสูจน์</u>

$$V(\bar{y}_{st}) = V\{\frac{(n'/n'')\bar{y}' + \bar{y}''_{st}}{1+(n'/n'')}\} \qquad .....(1.2.2.9)$$

เนื่องจากการประมาณ $\bar{y}'$ กับ $\bar{y}''_{st}$ กระทำในแต่ละส่วน จึงทำให้ Cov $(\bar{y}', \bar{y}''_{st}) = 0$ ดังนั้นจากสมการ (1.2.2.9) จะได้ว่า

$$V(\bar{y}_{st}) = \frac{1}{(1+(n'/n''))^2}\left[\left\{\frac{n'}{n''}\right\}^2 V(\bar{y}') + V(\bar{y}''_{st})\right] \quad ....(1.2.2.10)$$

พิจารณาเทอม $V(\bar{y}')$ จะเห็นว่ามีลักษณะเป็นการเลือกตัวอย่างสุ่มอย่างง่าย ซึ่งให้ค่าความแปรปรวนดังนี้

$$V(\bar{y}') = \frac{\Sigma W_h S_h^2}{n'} + \frac{\Sigma W_h(\bar{Y}_h - \bar{Y})^2}{n'} \quad ....(1.2.2.11)$$

สำหรับเทอม $V(\bar{y}''_{st})$ เนื่องจากมีลักษณะเป็นการเลือกตัวอย่างมีชั้นภูมิแบบสุ่มอย่างง่าย จึงให้ค่าความแปรปรวนดังนี้

$$V(\bar{y}''_{st}) = E\left[E\{(\bar{y}''_{st} - \bar{Y})^2 | w_1, w_2, \ldots, w_L\}\right] \quad ......(1.2.2.12)$$

พิจารณา $E\{(\bar{y}''_{st} - \bar{Y})^2 | w_1, w_2, \ldots, w_L\} = \Sigma \frac{w_h^2 S_h^2}{n''_h} + \{\Sigma(w_h - W_h)\bar{Y}_h\}^2$ ..(1.2.2.13)

ดังนั้น $V(\bar{y}''_{st}) = E\{\Sigma \frac{w_h^2 S_h^2}{n''_h}\} + E\{\Sigma(w_h - W_h)\bar{Y}_h\}^2$ ....(1.2.2.14)

จะเห็นว่าสมการ (1.2.2.14) อยู่ในรูปแบบเดียวกับสมการ (1.1.7) การพิสูจน์ต่อไปจึงมีรูปแบบและวิธีการทำนองเดียวกัน ตั้งแต่ขั้นตอนของสมการ (1.1.7) ถึงขั้นตอนของสมการ (1.1.11) ทั้งนี้โดยการแทนค่า $n''_h = n'' w_h s_h / \Sigma w_h s_h$ ของการพิสูจน์ในกรณีดังกล่าวนี้แทนการใช้ค่า $n_h = n w_h s_h / \Sigma w_h s_h$ ของการพิสูจน์สมการ (1.1.7) ถึงสมการ (1.1.11) ดังนั้นจะได้ว่า

$$V(\bar{y}''_{st}) = \frac{(\Sigma W_h S_h)^2}{n''} + \frac{1}{n' n''}\Big[\Sigma W_h (S_h - \bar{S})^2$$

$$+ \frac{B_2 - 1}{4} \{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \} \Big]$$

$$- \frac{B_2 - 1}{4 n'^2 n''} \{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \} + \frac{\Sigma W_h (\bar{Y}_h - \bar{Y})^2}{n'} \quad ..(1.2.2.15)$$

จากสมการ (1.2.2.10) (1.2.2.11) และ (1.2.2.15) สามารถสรุปได้ว่า

$$V(\bar{y}_{st}) = \frac{1}{(1+(n'/n''))^2} \Big[\{ \frac{n'}{n''}\}^2 \{ \frac{\Sigma W_h S_h^2}{n'}$$

$$+ \frac{\Sigma W_h (\bar{Y}_h - \bar{Y})^2}{n'} \} + \frac{(\Sigma W_h S_h)^2}{n''}$$

$$+ \frac{1}{n' n''}\{ \Sigma W_h (S_h - \bar{S})^2 + \frac{B_2 - 1}{4} [(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2] \}$$

$$- \frac{B_2 - 1}{4 n'^2 n''}\{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \} + \frac{\Sigma W_h (\bar{Y}_h - \bar{Y})^2}{n'} \Big] \quad ...(1.2.2.16)$$

สำหรับค่าประมาณของความแปรปรวนของตัวประมาณตามสมการที่ (1.2.2.16) คือ

$$\hat{V}(\bar{y}_{st}) = \frac{1}{(1+(n'/n''))^2} \left[\left\{ \frac{n'}{n''} \right\}^2 \left\{ \frac{\Sigma w_h s_h^2}{n'} \right. \right.$$

$$\left. + \frac{\Sigma w_h (\bar{y}_h - \bar{y})^2}{n'} \right\} + \frac{(\Sigma w_h s_h)^2}{n''}$$

$$+ \frac{1}{n' n''} \left\{ \Sigma w_h (s_h - \bar{s})^2 + \frac{B_2' - 1}{4} \left[(\Sigma w_h s_h)^2 - \Sigma w_h^2 s_h^2 \right] \right\}$$

$$\left. - \frac{B_2' - 1}{4n'^2 n''} \left\{ (\Sigma w_h s_h)^2 - \Sigma w_h^2 s_h^2 \right\} + \frac{\Sigma w_h (\bar{y}_h - \bar{y})^2}{n'} \right] \quad \ldots(1.2.2.17)$$

เนื่องจาก $n = n' + n''$ ดังนั้นสมการ (1.2.2.16) อาจเขียนในอีกรูปแบบหนึ่งได้ว่า

$$V(\bar{y}_{st}) = \frac{1}{n^2} \left[n' \{ \Sigma W_h S_h^2 + \Sigma W_h (\bar{Y}_h - \bar{Y})^2 \} + n'' (\Sigma W_h S_h)^2 \right.$$

$$+ \frac{n''}{n'} \left\{ \Sigma W_h (S_h - \bar{S})^2 + \frac{B_2 - 1}{4} \left[(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \right] \right\}$$

$$\left. - \frac{n''(B_2 - 1)}{4n'^2} \left\{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \right\} + \frac{n''^2}{n'} \Sigma W_h (\bar{Y}_h - \bar{Y})^2 \right]$$

$$= \frac{1}{n^2} \left[n'(a+b) + n'' m + \frac{n''}{n'} t + \frac{n''}{n'^2} f + \frac{n''^2}{n'} b \right] \quad \ldots\ldots(1.2.2.18)$$

โดยที่ a, b, m, t และ f มีค่าดังนี้

$$a = \Sigma W_h S_h^2$$

$$b = \Sigma W_h (\bar{Y}_h - \bar{Y})^2$$

$$m = (\Sigma W_h S_h)^2$$

$$t = \Sigma W_h (S_h - \bar{S})^2 + \frac{B_2 - 1}{4} \left[(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \right]$$

$$f = -\frac{B_2 - 1}{4} \{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \}$$

สำหรับการกำหนดขนาดตัวอย่างในส่วนที่ 1 และขนาดตัวอย่างในส่วนที่ 2 ซึ่งให้ค่าความแปรปรวนของตัวประมาณมีค่าต่ำสุด จะไม่คำนึงถึงเทอมที่มี n^2 n'^2 เป็นตัวหารเนื่องจากมีค่าน้อย จึงไม่มีผลกระทบต่อการกำหนดขนาดตัวอย่างในแต่ละส่วนมากนัก นั่นคือ สมการด้านขวาของ (1.2.2.18) จะลดลงเหลือเพียง 4 เทอม ดังนี้

$$V(\bar{y}_{st}) = \frac{1}{n^2} \left[n'(a+b) + n''m + \frac{n''}{n'}t + \frac{n''^2}{n'} b \right] \quad .....(1.2.2.19)$$

จะเห็นว่าขนาดตัวอย่างในส่วนที่ 1 ซึ่งทำให้ความแปรปรวนของตัวประมาณตามสมการ (1.2.2.19) มีค่าต่ำสุดก็ต่อเมื่อเป็นไปตามเงื่อนไข ต่อไปนี้

$$\frac{d}{dn'} V(\bar{y}_{st}) = 0 \quad ....(1.2.2.20)$$

แทนค่า $V(\bar{y}_{st})$ จากสมการ (1.2.2.19) ลงในสมการ (1.2.2.20) จะได้ว่า

$$\frac{d}{dn'} \frac{1}{n^2} \left[n'(a+b) + n''m + \frac{n''}{n'}t + \frac{n''^2}{n'} b \right] = 0 \quad ....(1.2.2.21)$$

หลังจากนั้นแทนค่า $n'' = n - n'$ และใช้หลักการของการหาอนุพันธ์ เพื่อหาค่า n' ซึ่งทำให้ $V(\bar{y}_{st})$ มีค่าต่ำสุด

$$\frac{d}{dn'}\;\frac{1}{n^2}\left[\, n'(a+b) + (n-n')m + \frac{(n-n')}{n'}t + \left\{\frac{n^2}{n'} + n' - 2n\right\} b \,\right] = 0$$

$$\frac{1}{n^2}\left[\, a+b-m - \frac{nt}{n'^2} - \frac{n^2 b}{n'^2} + b \,\right] = 0 \qquad \text{....(1.2.2.22)}$$

จากสมการ (1.2.2.22) จะเห็นว่า $n^2 \neq 0$ และเป็นเทอมที่หาค่าได้ ดังนั้นจะได้ว่า

$$a + b - m - \frac{nt}{n'^2} - \frac{n^2 b}{n'^2} + b = 0 \qquad \text{.......(1.2.2.23)}$$

$$a + 2b - m = \frac{nt + n^2 b}{n'^2}$$

$$n'^2 = \frac{nt + n^2 b}{a + 2b - m} \qquad \text{......(1.2.2.24)}$$

จากสมการ (1.2.2.24) จะได้ค่าของ n' ซึ่งเป็นขนาดตัวอย่างในส่วนที่ 1 ดังนี้

$$n' = \sqrt{\frac{nt + n^2 b}{a + 2b - m}} \qquad \text{......(1.2.2.25)}$$

ส่วน n'' ซึ่งเป็นขนาดตัวอย่างในส่วนที่ 2 จะมีค่าดังนี้

$$n'' = n - n'$$

$$= n - \sqrt{\frac{nt + n^2 b}{a+2b-m}} \qquad \text{......(1.2.2.26)}$$

2.1.2 ไม่ทราบค่าสัดส่วนของชั้นภูมิในขั้นตอนการวางแผนแต่ทราบในขั้นตอนการประมาณค่า ในการศึกษาได้แบ่งออกเป็น 2 กรณี ดังนี้

2.1.2.1 ทราบค่าประมาณของสัดส่วนของชั้นภูมิและความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิแล้ว

กรณีนี้เป็นกรณีที่ไม่ทราบค่าสัดส่วนของชั้นภูมิ ในขั้นตอนการวางแผนของการวิจัย แต่สามารถทราบค่านี้ในขั้นตอนการประมาณค่า รวมทั้งได้ทราบค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิแล้วจากการทำการสำรวจเบื้องต้น หรือการสำรวจอื่น ๆ ขนาด n'

ดังนั้นจึงใช้ค่าประมาณของสัดส่วนของชั้นภูมิและค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิที่ได้จากการทำการสำรวจเบื้องต้น หรือการสำรวจอื่น ๆ ขนาด n' ดังกล่าวในช่วงของการกำหนดขนาดตัวอย่างในแต่ละชั้นภูมิ และใช้ค่าสัดส่วนของชั้นภูมิในช่วงของการหาค่าประมาณของค่าเฉลี่ยของประชากร

ตัวประมาณที่ได้จากกรณีนี้ คือ

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \qquad \text{..........(2.1.1)}$$

ตัวประมาณนี้มีคุณสมบัติของความไม่เอนเอียงซึ่งสามารถแสดงได้ดังนี้

$$E(\bar{y}_{st}) = E\,\{\Sigma W_h \bar{y}_h\} \qquad \text{..........(2.1.2)}$$

$$= \Sigma W_h\, E(\bar{y}_h)$$

$$= \Sigma W_h \bar{Y}_h$$

$$= \bar{Y} \qquad \text{..........(2.1.3)}$$

จากสมการ (2.1.1) (2.1.2) จะเห็นว่าตัวประมาณอยู่ในรูปฟังก์ชันของ W_h เนื่องจากในขั้นตอนการประมาณค่าสามารถที่จะทราบค่าสัดส่วนของชั้นภูมิได้ ซึ่งตัวประมาณจะมีความแปรปรวนเมื่อไม่คิดค่า f.p.c. ดังนี้

$$V(\bar{y}_{st}) = \frac{1}{n}\Big[(\Sigma W_h S_h)^2 + \frac{B_2-5}{4n'} \{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\}$$

$$+ \frac{1}{n'} \{\Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2\} + \frac{B_2-1}{4n'^2} \big[\{\Sigma S_h \Sigma W_h S_h$$

$$- \Sigma W_h S_h^2\} - \{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} \big] \Big] \qquad .....(2.1.4)$$

<u>พิสูจน์</u>

ในกรณีที่ไม่คิดค่า f.p.c. จะได้ว่า

$$V(\bar{y}_{st}) = \Sigma \frac{W_h^2 S_h^2}{n_h} \qquad ....(2.1.5)$$

แทนค่า $n_h = nw_h s_h / \Sigma w_h s_h$ ลงในสมการ (2.1.5) จะได้

$$V\{\bar{y}_{st} | w_h, s_h \ (h = 1,2,...,L)\} = \Sigma \frac{W_h^2 S_h^2}{nw_h s_h} \Sigma w_h s_h$$

$$= \frac{1}{n}\Big[\Sigma W_h^2 S_h^2 + \sum_{h \neq j} W_j^2 S_j^2 \frac{w_h}{w_j} \frac{s_h}{s_j}\Big] \quad ...(2.1.6)$$

จากสมการ (2.1.5) และ (2.1.6) ดังนั้นจะได้ว่า

$$V(\bar{y}_{st}) = E\left[V\{\bar{y}_{st} | w_h, s_h \ (h = 1,2,...,L)\}\right]$$

$$= \frac{1}{n}\Big[\Sigma W_h^2 S_h^2 + \sum_{h \neq j} W_j^2 S_j^2 \ E\{\frac{w_h}{w_j}\} \ E\{\frac{s_h}{s_j}\}\Big] \qquad ......(2.1.7)$$

เนื่องจาก w_h และ s_h เป็นอิสระต่อกัน ดังนั้น

$$V(\bar{y}_{st}) = \frac{1}{n}\left[\Sigma W_h^2 S_h^2 + \sum_{h\neq j} W_j^2 S_j^2 \frac{W_h}{W_j}\left\{1 + \frac{1-W_j}{n'W_j}\right\}\frac{S_h}{S_j}\left\{1+\frac{C^2}{4}\right\}\right]$$

$$= \frac{1}{n}\left[\Sigma W_h^2 S_h^2 + \left\{1 + \frac{C^2}{4}\right\}\left\{1 - \frac{1}{n'}\right\}\sum_{h\neq j} W_h W_j S_h S_j\right.$$

$$\left. + \frac{4 + C^2}{4n'} \sum_{h \neq j} W_h S_h S_j\right]$$

$$= \frac{1}{n}\left[(\Sigma W_h S_h)^2 + \left\{\frac{C^2}{4} - \frac{4 + C^2}{4n'}\right\}\left\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\right\}\right.$$

$$\left. + \frac{4 + C^2}{4n'}\left\{\Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2\right\}\right] \qquad \ldots\ldots(2.1.8)$$

แทนค่า $C^2 = (B_2 - 1)/n'$ ลงในสมการ (2.1.8)

$$V(\bar{y}_{st}) = \frac{1}{n}\left[(\Sigma W_h S_h)^2 + \left\{\frac{B_2 - 1}{4n'}\right.\right.$$

$$\left. - \frac{4}{4n'} - \frac{B_2 - 1}{4n'^2}\right\}\left\{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\right\}$$

$$\left. + \left\{\frac{4}{4n'} + \frac{B_2 - 1}{4n'^2}\right\}\left\{\Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2\right\}\right]$$

$$= \frac{1}{n} \Big[(\Sigma W_h S_h)^2 + \Big\{ \frac{B_2 - 5}{4n'} - \frac{B_2 - 1}{4n'^2} \Big\} \{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \} + \Big\{ \frac{1}{n'} + \frac{B_2 - 1}{4n'^2} \Big\} \{ \Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2 \} \Big] \quad \ldots\ldots(2.1.9)$$

จากสมการ (2.1.9) สามารถสรุปได้ว่า

$$V(\bar{y}_{st}) = \frac{1}{n} \Big[(\Sigma W_h S_h)^2 + \frac{B_2 - 5}{4n'} \{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \} + \frac{1}{n'} \{ \Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2 \} + \frac{B_2 - 1}{4n'^2} \Big[\{ \Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2 \} - \{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \} \Big] \Big] \quad \ldots\ldots(2.1.10)$$

ค่าประมาณของความแปรปรวนของตัวประมาณตามสมการที่ (2.1.10) คือ

$$\hat{V}(\bar{y}_{st}) = \frac{1}{n} \Big[(\Sigma W_h s_h)^2 + \frac{B_2' - 5}{4n'} \{ (\Sigma W_h s_h)^2 - \Sigma W_h^2 s_h^2 \} + \frac{1}{n'} \{ \Sigma s_h \Sigma W_h s_h - \Sigma W_h s_h^2 \} + \frac{B_2' - 1}{4n'^2} \Big[\{ \Sigma s_h \Sigma W_h s_h - \Sigma W_h s_h^2 \} - \{ (\Sigma W_h s_h)^2 - \Sigma W_h^2 s_h^2 \} \Big] \Big] \quad \ldots\ldots(2.1.11)$$

2.1.2.2 <u>ไม่ทราบค่าประมาณของสัดส่วนของชั้นภูมิและความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิมาก่อน</u>

สำหรับกรณีนี้เป็นกรณีที่ไม่ทราบค่าสัดส่วนของชั้นภูมิ ในขั้นตอนการวางแผน แต่สามารถทราบค่านี้ในขั้นตอนการประมาณค่า นอกจากนี้ยังไม่ทราบค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิมาก่อน นั่นคือไม่เคยมีการสำรวจข้อมูลที่จะใช้ในการวิจัยมาเลยในอดีต ไม่ว่าจะเป็นการสำรวจเบื้องต้นหรือการสำรวจอื่นใดก็ตาม

ดังนั้นในการวิจัย จะทำการสำรวจข้อมูลโดยแบ่งขอบข่ายออกเป็น 2 ส่วน ในส่วนที่ 1 ขนาด n' จะใช้ในการหาค่าประมาณของสัดส่วนของชั้นภูมิ และค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิ หลังจากนั้นจะนำค่าประมาณเหล่านี้ไปใช้หาค่าต่าง ๆ ในส่วนที่ 2 ขนาด n'' ต่อไป

ตัวประมาณที่ได้จากกรณีนี้ นอกจากจะเป็นตัวประมาณที่เกิดจากการนำค่าประมาณของสัดส่วนของชั้นภูมิ ค่าประมาณของความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิตลอดจนค่าสัดส่วนของชั้นภูมิมาใช้ในส่วนที่ 2 แล้ว Shambhu Dayal ได้เสนอให้นำตัวประมาณในส่วนที่ 1 มาใช้ด้วยเช่นกัน ทั้งนี้เพื่อเป็นการใช้ข่าวสารที่ได้จากการสำรวจอย่างครบถ้วน

ตัวประมาณที่ได้จากกรณีนี้ คือ

$$\bar{y}_{st} = \frac{\bar{y}' + \bar{y}''_{st}}{2} \qquad \text{........(2.2.1)}$$

โดยที่ $\bar{y}'$ คือ ตัวประมาณจากส่วนที่ 1 ขนาด n' ซึ่งมีลักษณะเป็นการเลือกตัวอย่างสุ่มอย่างง่าย เนื่องจากใช้ข้อมูลจากหน่วยตัวอย่างทุกหน่วยของส่วนที่ 1 นั่นคือ

$$\bar{y}' = \sum_{i=1}^{n'} y_i/n'$$

และ $\bar{y}''_{st}$ คือ ตัวประมาณจากส่วนที่ 2 ขนาด n'' ซึ่งมีลักษณะเป็นการเลือกตัวอย่างมีชั้นภูมิแบบสุ่มอย่างง่าย โดยนำค่าสัดส่วนของชั้นภูมิที่สามารถทราบได้ในขั้นตอนการประมาณค่า มาใช้ในส่วนที่ 2 มีสูตรดังนี้

$$\bar{y}''_{st} = \sum_{h=1}^{L} W_h \bar{y}''_h$$

เราจะพบว่าตัวประมาณตามสมการ (2.2.1) มีคุณสมบัติของความไม่เอนเอียง ซึ่งการพิสูจน์แสดงได้ดังนี้

$$E(\bar{y}_{st}) = \frac{E(\bar{y}') + E(\bar{y}''_{st})}{2} \quad \text{.........(2.2.2)}$$

พิจารณาเทอมแรกของสมการด้านขวาของ (2.2.2) คือ $E(\bar{y}')$ จะเห็นว่าเนื่องจากได้ใช้ทุกหน่วยตัวอย่างจากการสำรวจในส่วนที่ 1 จึงมีลักษณะเป็นการเลือกตัวอย่างสุ่มอย่างง่าย ซึ่งจากคุณสมบัติของการเลือกตัวอย่างสุ่มอย่างง่าย จะได้ว่าตัวประมาณมีคุณสมบัติของความไม่เอนเอียง ดังนี้

$$E(\bar{y}') = E\left\{ \frac{1}{n'} \sum_{i=1}^{n'} y_i \right\} \quad \text{.......(2.2.3)}$$

$$= \frac{1}{n'} \sum_{i=1}^{n'} \{ E(y_i) \}$$

$$= \frac{1}{n'} \sum_{i=1}^{n'} \bar{Y}$$

$$= \bar{Y} \quad \text{.......(2.2.4)}$$

สำหรับเทอมที่ 2 ของสมการด้านขวาของ (2.2.2) คือ $E(\bar{y}''_{st})$ จะเห็นว่ามีลักษณะเป็นการเลือกตัวอย่างมีชั้นภูมิแบบสุ่มอย่างง่าย โดยนำค่าสัดส่วนของชั้นภูมิที่ทราบในขั้นตอนการประมาณค่ามาใช้ในส่วนที่ 2 ซึ่งสามารถพิสูจน์ได้ว่าตัวประมาณมีคุณสมบัติของความไม่เอนเอียงดังนี้

$$E(\bar{y}''_{st}) = E\{\Sigma W_h \bar{y}''_h\} \quad \text{.......(2.2.5)}$$

$$= \Sigma W_h \, E(\bar{y}''_h)$$

$$= \quad \Sigma W_h \bar{Y}_h$$

$$= \quad \bar{Y} \qquad ....(2.2.6)$$

จากสมการ (2.2.2) (2.2.4) และ (2.2.6) สามารถสรุปได้ว่า

$$E(\bar{y}_{st}) \quad = \quad \frac{\bar{Y} + \bar{Y}}{2}$$

$$= \quad \bar{Y} \qquad ....(2.2.7)$$

สำหรับความแปรปรวนของตัวประมาณตามสมการ (2.2.1) มีค่าดังนี้ (ไม่คิดค่า f.p.c.)

$$V(\bar{y}_{st}) \quad = \quad \frac{1}{4}\Big[\frac{\Sigma W_h S_h^2}{n'} + \frac{\Sigma W_h (\bar{Y}_h - \bar{Y})^2}{n'} + \frac{(\Sigma W_h S_h)^2}{n''}$$

$$+ \frac{B_2 - 5}{4n' n''} \{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \} + \frac{1}{n' n''}\{ \Sigma S_h \Sigma W_h S_h$$

$$- \Sigma W_h S_h^2 \} + \frac{B_2 - 1}{4n'^2 n''} \Big[\{\Sigma S_h \Sigma W_h S_h$$

$$- \Sigma W_h S_h^2 \} - \{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} \Big] \Big] \qquad ......(2.2.8)$$

<u>พิสูจน์</u>

$$V(\bar{y}_{st}) \quad = \quad V \{ \frac{\bar{y}' + \bar{y}''_{st}}{2} \} \qquad ......(2.2.9)$$

$$= \quad \frac{1}{4} [V(\bar{y}') + V(\bar{y}''_{st}) + 2Cov(\bar{y}', \bar{y}''_{st})]$$

เนื่องจากการประมาณ $\bar{y}'$ กับ $\bar{y}''_{st}$ กระทำในแต่ละส่วน โดย $\bar{y}'$ ได้จากการสำรวจในส่วนที่ 1 และ $\bar{y}''_{st}$ ได้จากการสำรวจในส่วนที่ 2 จึงทำให้ $Cov(\bar{y}', \bar{y}''_{st}) = 0$ ดังนั้นจะได้ว่า

$$V(\bar{y}_{st}) \quad = \quad \frac{1}{4}\left[\, V(\bar{y}') + V(\bar{y}''_{st}) \,\right] \qquad \ldots\ldots(2.2.10)$$

พิจารณาเทอมแรกของสมการด้านขวาของ (2.2.10) คือ $V(\bar{y}')$ จะเห็นว่ามีลักษณะเป็นการเลือกตัวอย่างสุ่มอย่างง่าย ซึ่งให้ค่าความแปรปรวนดังนี้

$$V(\bar{y}') \quad = \quad \frac{\Sigma W_h S_h^2}{n'} + \frac{\Sigma W_h (\bar{Y}_h - \bar{Y})^2}{n'} \qquad \ldots\ldots\ldots\ldots(2.2.11)$$

ส่วนเทอมที่ 2 ของสมการด้านขวาของ (2.2.10) คือ $V(\bar{y}''_{st})$ ซึ่งมีลักษณะเป็นการเลือกตัวอย่างมีชั้นภูมิแบบสุ่มอย่างง่าย โดยจะให้ค่าความแปรปรวนดังนี้ (ไม่คิดค่า f.p.c.)

$$V(\bar{y}''_{st}) \quad = \Sigma \frac{W_h^2 S_h^2}{n''_h} \qquad \ldots\ldots\ldots\ldots(2.2.12)$$

จะเห็นว่าสมการ (2.2.12) อยู่ในรูปแบบเดียวกับสมการ (2.1.5) การพิสูจน์ต่อไปจึงมีรูปแบบและวิธีการทำนองเดียวกัน ตั้งแต่ขั้นตอนของสมการ (2.1.5) ถึงขั้นตอนของสมการ (2.1.10) ทั้งนี้โดยการแทนค่า $n''_h = n'' w_h s_h / \Sigma w_h s_h$ ของการพิสูจน์ในกรณีดังกล่าวนี้ แทนการใช้ค่า $n_h = n w_h s_h / \Sigma w_h s_h$ ของการพิสูจน์สมการ (2.1.5) ถึงสมการ (2.1.10) ดังนั้นจะได้ว่า

$$\begin{aligned} V(\bar{y}''_{st}) \quad = \quad & \frac{1}{n''}\Big[\, (\Sigma W_h S_h)^2 + \frac{B_2 - 5}{4n'} \left\{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \right\} \\ & + \frac{1}{n'} \left\{ \Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2 \right\} + \frac{B_2 - 1}{4n'^2}\Big[\left\{ \Sigma S_h \Sigma W_h S_h \right. \\ & \left. - \Sigma W_h S_h^2 \right\} - \left\{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \right\} \Big]\Big] \qquad \ldots\ldots(2.2.13) \end{aligned}$$

จากสมการ (2.2.10) (2.2.11) และ (2.2.13) สามารถสรุปได้ว่า

$$V(\bar{y}_{st}) = \frac{1}{4}\Big[\frac{\Sigma W_h S_h^2}{n'} + \frac{\Sigma W_h(\bar{Y}_h - \bar{Y})^2}{n'} + \frac{(\Sigma W_h S_h)^2}{n''}$$

$$+ \frac{B_2 - 5}{4n'n''}\{ (\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} + \frac{1}{n'n''}\{ \Sigma S_h \Sigma W_h S_h$$

$$- \Sigma W_h S_h^2 \} + \frac{B_2 - 1}{4n'^2 n''}\Big[\{ \Sigma S_h \Sigma W_h S_h$$

$$- \Sigma W_h S_h^2 \} - \{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2 \} \Big] \Big] \qquad \ldots\ldots(2.2.14)$$

สำหรับค่าประมาณของความแปรปรวนของตัวประมาณตามสมการที่ (2.2.14) คือ

$$\hat{V}(\bar{y}_{st}) = \frac{1}{4}\Big[\frac{\Sigma W_h s_h^2}{n'} + \frac{\Sigma W_h(\bar{y}_h - \bar{y})^2}{n'} + \frac{(\Sigma W_h s_h)^2}{n''}$$

$$+ \frac{B_2' - 5}{4n'n''}\{ (\Sigma W_h s_h)^2 - \Sigma W_h^2 s_h^2\} + \frac{1}{n'n''}\{ \Sigma s_h \Sigma W_h s_h$$

$$- \Sigma W_h s_h^2 \} + \frac{B_2' - 1}{4n'^2 n''}\Big[\{\Sigma s_h \Sigma W_h s_h$$

$$- \Sigma W_h s_h^2 \} - \{ (\Sigma W_h s_h)^2 - \Sigma W_h^2 s_h^2 \} \Big] \Big] \qquad \ldots(2.2.15)$$

จากสมการ (2.2.14) อาจเขียนในอีกรูปแบบหนึ่งได้ว่า

$$V(\bar{y}_{st}) = \frac{S^2}{4n'} + \frac{m}{4n''} + \frac{u}{4n'n''} + \frac{g}{4n'^2 n''} \qquad \ldots(2.2.16)$$

โดยที่ S^2, m, u และ g มีค่าดังนี้

$$S^2 = \Sigma W_h S_h^2 + \Sigma W_h (\bar{Y}_h - \bar{Y})^2$$

$$m = (\Sigma W_h S_h)^2$$

$$u = \frac{B_2 - 5}{4} \{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\} + \{\Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2\}$$

$$g = \frac{B_2 - 1}{4} [\{\Sigma S_h \Sigma W_h S_h - \Sigma W_h S_h^2\} - \{(\Sigma W_h S_h)^2 - \Sigma W_h^2 S_h^2\}]$$

สำหรับการกำหนดขนาดตัวอย่างในส่วนที่ 1 และขนาดตัวอย่างในส่วนที่ 2 ซึ่งให้ค่าความแปร-ปรวนของตัวประมาณมีค่าต่ำสุดจะไม่คำนึงถึงเทอมที่มี $4n'^2n''$ เป็นตัวหาร เนื่องจากมีค่าน้อยจึงไม่มีผลกระทบต่อการกำหนดขนาดตัวอย่างในแต่ละส่วนมากนัก นั่นคือ สมการด้านขวาของ (2.2.16) จะลดลงเหลือเพียง 3 เทอม ดังนี้

$$V(\bar{y}_{st}) = \frac{S^2}{4n'} + \frac{m}{4n''} + \frac{u}{4n'n''} \qquad ....(2.2.17)$$

จะเห็นว่าขนาดตัวอย่างในส่วนที่ 1 ซึ่งทำให้ความแปรปรวนของตัวประมาณตามสมการ (2.2.17) มีค่าต่ำสุดก็ต่อเมื่อเป็นไปตามเงื่อนไข ต่อไปนี้

$$\frac{d}{dn'} V(\bar{y}_{st}) = 0 \qquad ....(2.2.18)$$

แทนค่า $V(\bar{y}_{st})$ จากสมการ (2.2.17) ลงในสมการ (2.2.18) จะได้ว่า

$$\frac{d}{dn'} \left[\frac{S^2}{4n'} + \frac{m}{4n''} + \frac{u}{4n'n''} \right] = 0 \qquad ....(2.2.19)$$

หลังจากนั้นแทนค่า $n'' = n - n'$ และใช้หลักการของการหาอนุพันธ์ เพื่อหาค่า n' ซึ่งทำให้ $V(\bar{y}_{st})$ มีค่าต่ำสุด

$$\frac{d}{dn'}\left[\frac{S^2}{4n'} + \frac{m}{4(n-n')} + \frac{u}{4n'(n-n')}\right] = 0 \qquad ....(2.2.20)$$

$$\frac{S^2}{4}\left[-\frac{1}{n'^2}\right] + \frac{m}{4}\left[\frac{1}{(n-n')^2}\right] + \frac{u}{4}\left[\frac{2n'-n}{n'^2(n-n')^2}\right] = 0$$

$$\frac{-S^2(n-n')^2 + mn'^2 + 2n'u - nu}{4n'^2(n-n')^2} = 0 \qquad .....(2.2.21)$$

จากสมการ (2.2.21) จะเห็นว่า $4n'^2(n-n')^2 \neq 0$ และเป็นเทอมที่หาค่าได้ ดังนั้นจะได้ว่า

$$-S^2(n-n')^2 + mn'^2 + 2n'u - nu = 0 \qquad ....(2.2.22)$$

$$-S^2n^2 + 2S^2nn' - S^2n'^2 + mn'^2 + 2n'u - nu = 0$$

$$(m - S^2)n'^2 + 2(S^2n + u)n' - (S^2n + u)n = 0 \qquad ...... (2.2.23)$$

เนื่องจากสมการ (2.2.23) อยู่ในรูปแบบของสมการกำลังสอง โดยมี n' เป็นตัวแปร ดังนั้นจึงสามารถใช้สูตรทั่วไปในการหารากของสมการดังกล่าวได้ ซึ่งจะได้ค่าของ n' ดังนี้

$$n' = \frac{-2(S^2n + u) \pm \sqrt{4(S^2n + u)^2 + 4(m - S^2)(S^2n + u)n}}{2(m - S^2)}$$

$$= \frac{-(S^2n + u) \pm \sqrt{(S^2n + u)^2 + (m - S^2)(S^2n + u)n}}{m - S^2} \qquad .....(2.2.24)$$

ส่วน n'' จะมีค่าดังนี้

$$n'' = n - n'$$

$$= n - \left[\frac{-(S^2n + u) \pm \sqrt{(S^2n + u)^2 + (m - S^2)(S^2n + u)n}}{m - S^2}\right] \quad ..(2.2.25)$$

2.2 ผลงานวิจัยที่เกี่ยวข้อง

จะเห็นว่าในการวิจัยครั้งนี้ ในส่วนของขอบเขตของการวิจัยนอกจากจะได้ระบุถึงขอบเขตทั่วไป นั่นคือ ขนาดของประชากร รูปแบบการแจกแจงของประชากร และขนาดตัวอย่างที่ใช้ในการเลือกตัวอย่าง ดังเช่นงานวิจัยทั่ว ๆ ไปแล้ว วิธีการแบ่งช่วงของชั้นภูมิและจำนวนชั้นภูมิที่ใช้ก็นับได้ว่าเป็นสิ่งสำคัญที่จะต้องระบุถึงสำหรับงานวิจัยนี้ด้วย เนื่องจากเป็นงานวิจัยที่ใช้วิธีการของการเลือกตัวอย่างแบบมีชั้นภูมิในการเลือกตัวอย่าง ซึ่งในการวิจัยครั้งนี้ได้กำหนดวิธีการแบ่งช่วงของชั้นภูมิเป็น 2 วิธี คือ วิธีการแบ่งโดยให้มีช่วงของชั้นภูมิเท่ากันทุกชั้นภูมิ และวิธี Cumulative $\sqrt{f}$ ส่วนจำนวนชั้นภูมิที่ใช้ได้กำหนดเป็น 6 ชั้นภูมิทุกรูปแบบที่ศึกษา

สำหรับวิธีการแบ่งช่วงของชั้นภูมิซึ่งได้กำหนดเป็น 2 วิธี ตามที่ได้กล่าวนั้น ทั้งนี้ก็เพื่อผลการวิจัยที่ต้องการให้ครอบคลุมในรูปแบบของการแบ่งช่วงที่พบกันทั่ว ๆ ไปในทางปฏิบัติ โดยจะเห็นว่า วิธีการแบ่งโดยให้มีช่วงของชั้นภูมิเท่ากันทุกชั้นภูมิแม้จะไม่มีผลในแง่ของการเพิ่มคุณภาพของตัวประมาณ แต่ก็พบว่าวิธีนี้ยังคงใช้กันมาก อันอาจเนื่องมาจากนักวิจัยไม่มีข่าวสารเพียงพอที่จะเลือกใช้วิธีการอื่น หรืออาจเนื่องมาจากสาเหตุอื่นใดก็แล้วแต่ จึงจำเป็นที่จะต้องเลือกใช้วิธีการนี้ ส่วนวิธีการแบ่งช่วงของชั้นภูมิโดยใช้วิธี Cumulative $\sqrt{f}$ มีทฤษฎีที่กล่าวถึงในแง่ของการเพิ่มคุณภาพของตัวประมาณเนื่องจากอยู่ภายใต้เงื่อนไขที่ต้องการทำให้ความแปรปรวนของตัวประมาณมีค่าต่ำที่สุด ซึ่งทฤษฎีดังกล่าวนี้จะได้กล่าวถึงต่อไปในหัวข้อ 2.2.1

ขอบเขตของการวิจัยสุดท้ายที่ได้กล่าวถึงก็คือ จำนวนชั้นภูมิที่ใช้ ซึ่งได้กำหนดเป็น 6 ชั้นภูมิ ทุกรูปแบบที่ศึกษา ทั้งนี้แม้จะพบว่าจำนวนชั้นภูมิที่กำหนดยิ่งมากเท่าใดยิ่งเป็นการดีต่อการเพิ่มคุณภาพของตัวประมาณ แต่ในทางปฏิบัติจริงมักจะไม่นิยมกำหนดจำนวนชั้นภูมิมากเกินความจำเป็น เนื่องจากการกำหนดจำนวนชั้นภูมิที่เพิ่มมากขึ้นเป็นสิ่งที่บ่งบอกถึงค่าใช้จ่ายที่จะต้องมากขึ้นตามไปด้วย และนอกจากนี้ยังได้มีผลการวิจัยที่แสดงให้เห็นว่าจำนวนชั้นภูมิ 6 ชั้นภูมิที่นักวิจัยเลือกใช้นับได้ว่าเพียงพอแล้วสำหรับโครงการหรืองานวิจัยทั่ว ๆ ไป ซึ่งจะได้กล่าวถึงผลงานวิจัยดังกล่าวในหัวข้อ 2.2.2

2.2.1 การแบ่งช่วงของชั้นภูมิโดยวิธี Cumulative $\sqrt{f}$

การใช้วิธี Cumulative $\sqrt{f}$ ในการแบ่งช่วงของชั้นภูมิ ของ Dalenius และ Hodges (1959) มีหลักเกณฑ์ในการพิจารณาหาตำแหน่งของจุด y_h ซึ่งเป็นจุดแบ่งช่วง

ของชั้นภูมิเช่นเดียวกับวิธีที่ Dalenius (1957) เคยเสนอไว้ นั่นก็คือ จะหาจุด y_h เพื่อแบ่งช่วงระหว่างชั้นภูมิที่ h กับชั้นภูมิที่ h + 1 ภายใต้เงื่อนไขที่ต้องการทำให้ค่าของ $\Sigma W_h S_h$ มีค่าต่ำที่สุดเช่นเดียวกัน ซึ่งในที่นี้จะได้กล่าวถึงวิธีการที่ Dalenius เคยเสนอไว้ด้วย เพื่อให้เห็นถึงเหตุผลในการปรับปรุงวิธีการนั้นมาใช้วิธี Cumulative $\sqrt{f}$ ในการแบ่งช่วงของชั้นภูมิ

พิจารณาความแปรปรวนของตัวประมาณที่ใช้วิธีการของเนย์แมนในการกำหนดขนาดตัวอย่าง ในกรณีที่ไม่คิดค่า f.p.c. จะได้ว่า

$$V(\bar{y}_{st}) = \frac{(\Sigma W_h S_h)^2}{n} \qquad \ldots\ldots\ldots\ldots (3.1)$$

กำหนดให้ y_o และ y_L คือ ค่าต่ำสุดและค่าสูงสุดของตัวแปร y สำหรับประชากรที่ทำการศึกษา จะหาตำแหน่งของจุด $y_1, y_2, \ldots, y_{L-1}$ ซึ่งทำให้ความแปรปรวนของตัวประมาณตามสมการ (3.1) มีค่าต่ำสุด

จะเห็นว่าความแปรปรวนของตัวประมาณตามสมการ (3.1) จะมีค่าต่ำที่สุดก็ต่อเมื่อ $\Sigma W_h S_h$ มีค่าต่ำที่สุดนั่นเอง ซึ่งในการจะหาตำแหน่งของจุด y_h $(h = 1,2,\ldots,L-1)$ จะพบว่า y_h จะมีผลกระทบต่อชั้นภูมิที่ h และชั้นภูมิที่ h + 1 เท่านั้น ดังนั้นจะได้ว่า

$$\frac{\partial}{\partial y_h}(\Sigma W_h S_h) = \frac{\partial}{\partial y_h}(W_h S_h) + \frac{\partial}{\partial y_h}(W_{h+1} S_{h+1}) \quad \ldots\ldots \quad (3.2)$$

กำหนดให้ f(y) เป็นฟังก์ชันความถี่ของตัวแปร y จะได้สัดส่วนของชั้นภูมิในรูปของฟังก์ชันความถี่ดังนี้

$$W_h = \int_{y_{h-1}}^{y_h} f(t)\,dt \qquad \ldots\ldots \quad (3.3)$$

และได้ค่าอนุพันธ์ของสัดส่วนของชั้นภูมิดังนี้

$$\frac{\partial W_h}{\partial y_h} = f(y_h) \qquad \ldots\ldots \quad (3.4)$$

สำหรับเทอม $W_h S_h^2$ ซึ่งเป็นเทอมที่ Dalenius กำหนดขึ้นเพื่อใช้เป็นเทอมประกอบในการไปสู่คำตอบที่ต้องการ ให้ค่าในรูปของฟังก์ชันความถี่เท่ากับ

$$W_h S_h^2 = \int_{y_{h-1}}^{y_h} t^2 f(t)dt - \frac{\left[\int_{y_{h-1}}^{y_h} t f(t)dt\right]^2}{\int_{y_{h-1}}^{y_h} f(t)dt} \quad \text{.....(3.5)}$$

หาอนุพันธ์ของสมการ (3.5) จะได้ว่า

$$S_h^2 \frac{\partial W_h}{\partial y_h} + 2W_h S_h \frac{\partial S_h}{\partial y_h} = y_h^2 f(y_h) - 2y_h \mu_h f(y_h) + \mu_h^2 f(y_h) \quad \text{....(3.6)}$$

โดยที่ μ_h คือ ค่าเฉลี่ยของตัวแปร y ของประชากรในชั้นภูมิที่ h และจากสมการ (3.6) นี้ จะบวกค่า $S_h^2\ \partial W_h/\partial y_h$ ทางสมการด้านซ้าย และบวกค่า $S_h^2\ f(y_h)$ ซึ่งมีค่าเท่ากันทางสมการด้านขวา หลังจากนั้นจะหารสมการที่ได้ด้วยค่า $2S_h$ ดังนั้นจะให้ค่าผลลัพธ์เท่ากับ

$$\frac{\partial(W_h S_h)}{\partial y_h} = S_h \frac{\partial W_h}{\partial y_h} + W_h \frac{\partial S_h}{\partial y_h}$$

$$= \frac{1}{2} f(y_h) \frac{(y_h - \mu_h)^2 + S_h^2}{S_h} \quad \text{......(3.7)}$$

ในทำนองเดียวกันจะได้ว่า

$$\frac{\partial(W_{h+1} S_{h+1})}{\partial y_h} = -\frac{1}{2} f(y_h) \frac{(y_h - \mu_{h+1})^2 + S_{h+1}^2}{S_{h+1}} \quad \text{....(3.8)}$$

จะเห็นว่าค่าของ y_h ซึ่งทำให้ $\Sigma W_h S_h$ มีค่าต่ำที่สุด ก็ต่อเมื่อเป็นไปตามเงื่อนไขต่อไปนี้

$$\frac{\partial(\Sigma W_h S_h)}{\partial y_h} = 0 \quad \text{.......(3.9)}$$

แทนค่าอนุพันธ์ของสมการ (3.9) ด้วยสมการ (3.2) ดังนั้นจะได้ว่า

$$\frac{\partial}{\partial y_h}(W_h S_h) + \frac{\partial}{\partial y_h}(W_{h+1} S_{h+1}) = 0 \quad ....(3.10)$$

จากนั้นจะแทนค่าอนุพันธ์ของทั้ง 2 เทอมตามสมการ (3.10) ด้วยสมการ (3.7) และ (3.8) ซึ่งจะพบว่าผลลัพธ์สุดท้ายที่ต้องการ นั่นก็คือ y_h ซึ่งเป็นจุดแบ่งช่วงระหว่างชั้นภูมิที่ h กับชั้นภูมิที่ h + 1 ซึ่งทำให้ค่าของ $\Sigma W_h S_h$ มีค่าต่ำที่สุดก็ต่อเมื่อค่า y_h ที่ได้นั้นจะต้องทำให้

$$\frac{(y_h - \mu_h)^2 + S_h^2}{S_h} = \frac{(y_h - \mu_{h+1})^2 + S_{h+1}^2}{S_{h+1}} \; ; \; h = 1,2,...,L-1 \quad ..(3.11)$$

จะเห็นได้ว่า ในการหาค่า y_h ที่สอดคล้องกับสมการ (3.11) ตามที่ Dalenius เสนอนั้น จะต้องทราบค่าของ μ_h และ S_h^2 ของทุก ๆ ชั้นภูมิเสียก่อน ดังนั้นการใช้วิธีนี้จึงเป็นเรื่องยากในทางปฏิบัติ ทั้งนี้เนื่องจากทั้ง μ_h และ S_h^2 ซึ่งเป็นค่าที่จะต้องทราบและนำมาใช้นั้นต่างก็ขึ้นอยู่กับค่าของ y_h ที่ต้องการหานั่นเอง

จากความยุ่งยากดังกล่าว Dalenius และ Hodges(1959) จึงได้พยายามหาวิธีการใหม่ซึ่งอยู่ในรูปแบบที่ง่ายกว่า อีกทั้งในทางปฏิบัติก็มิได้ยุ่งยากมากนัก แต่ก็ยังคงไว้ซึ่งเงื่อนไขที่ต้องการทำให้ค่าของ $\Sigma W_h S_h$ มีค่าต่ำที่สุดเช่นเดิม นั่นก็คือ วิธี Cumulative $\sqrt{f}$ นั่นเอง ทั้งนี้โดยมีที่มาของทฤษฎีดังนี้

กำหนดให้

$$Z(y) = \int_{y_o}^{y} \sqrt{f(t)}\, dt \quad ....(4.1)$$

เมื่อพิจารณาจากฟังก์ชันความถี่ จะเห็นว่า ในกรณีที่มีจำนวนชั้นภูมิมาก ๆ จะทำให้ช่วงของชั้นภูมิแคบลง จึงมีผลทำให้ f(y) มีลักษณะใกล้เคียงกับการแจกแจงแบบรูปสี่เหลี่ยมผืนผ้า (Rectangular Distribution) ดังนั้นจึงให้ค่าของสัดส่วนของชั้นภูมิ ความแปรปรวนของแต่ละชั้นภูมิ ตลอดจนความเบี่ยงเบนมาตรฐานของแต่ละชั้นภูมิ ดังนี้

$$W_h = \int_{y_{h-1}}^{y_h} f(t)dt$$

$$\simeq f_h(y_h - y_{h-1}) \qquad \text{.........(4.2)}$$

$$S_h^2 \simeq \frac{1}{12}(y_h - y_{h-1})^2 \qquad \text{.........(4.3)}$$

$$S_h \simeq \frac{1}{\sqrt{12}}(y_h - y_{h-1}) \qquad \text{.........(4.4)}$$

สำหรับเทอม $Z_h - Z_{h-1}$ ซึ่งเป็นเทอมที่ Dalenius และ Hodges กำหนดขึ้นเพื่อใช้เป็นเทอมประกอบในการไปสู่คำตอบที่ต้องการ ให้ค่าในรูปของฟังก์ชันความถี่เท่ากับ

$$Z_h - Z_{h-1} = \int_{y_{h-1}}^{y_h} \sqrt{f(t)}\, dt$$

$$\simeq \sqrt{f_h}\,(y_h - y_{h-1}) \qquad \text{..........(4.5)}$$

โดยที่ f_h เป็นค่าความถี่ $f(y)$ ของชั้นภูมิที่ h

พิจารณาจากสมการ (4.2) และ (4.4) จะได้ว่า

$$\Sigma W_h S_h = \frac{1}{\sqrt{12}} \Sigma f_h (y_h - y_{h-1})^2 \qquad \text{...... (4.6)}$$

แทนค่าเทอมของตัวแปรที่มีค่าเท่ากันตามสมการ (4.5) ลงในสมการ (4,6)

$$\Sigma W_h S_h = \frac{1}{\sqrt{12}} \Sigma (Z_h - Z_{h-1})^2 \qquad \text{...... (4.7)}$$

จะเห็นว่าสมการด้านขวาของ (4.6) และ (4,7) ต่างก็มีค่าเท่ากับ $\Sigma W_h S_h$ ซึ่งผลจากการมีค่าที่เท่ากันนี้ จึงอาจกล่าวต่อไปได้ว่า หากสามารถหาค่า y_h ซึ่งเป็นจุดแบ่งช่วงของชั้นภูมิจากการ minimize เทอม $\Sigma(Z_h - Z_{h-1})^2$ เทียบกับ y_h แล้ว y_h นั้น จะเป็น

ค่าเดียวกับ y_h ที่ทำให้เทอม $\Sigma f_h(y_h - y_{h-1})^2$ มีค่าต่ำที่สุดนั่นเอง

ในการหาค่า y_h ที่มีผลทำให้ $\Sigma(Z_h - Z_{h-1})^2$ มีค่าต่ำสุด จะอาศัยทฤษฎีทางคณิตศาสตร์ช่วยในการหาคำตอบ ซึ่งจะพบว่า $\Sigma(Z_h - Z_{h-1})^2$ จะให้ค่าต่ำที่สุดก็ต่อเมื่อได้จัดแบ่งให้ $Z_h - Z_{h-1}$ มีค่าคงที่ นั่นก็คือ เท่ากันในทุกชั้นภูมิ และเมื่อพิจารณาสมการ (4.5) และ (4.6) สามารถอธิบายในทำนองเดียวกันได้ว่า หากสามารถทำให้ $\sqrt{f(y)}$ มีค่าเท่ากันในทุกชั้นภูมิ ค่าของ y_h ซึ่งใช้เป็นจุดแบ่งช่วงของชั้นภูมินั้น ๆ จะมีผลทำให้ $\Sigma f_h(y_h - y_{h-1})^2$ มีค่าต่ำที่สุด ซึ่งก็หมายถึง $\Sigma W_h S_h$ มีค่าต่ำที่สุดนั่นเอง

สำหรับวิธีการที่จะจัดทำเพื่อให้สอดคล้องกับทฤษฎีดังกล่าว จึงใช้วิธีสะสมข้อมูลในรูปของ $\sqrt{f(y)}$ จนครบทุกช่วงตามที่แบ่ง จากนั้นจึงนำผลลัพธ์สุดท้ายหารด้วยจำนวนชั้นภูมิที่ต้องการ ซึ่งจะเห็นว่าวิธีการที่ Dalenius และ Hodges ได้เสนอดังที่กล่าวนี้มีความยุ่งยากลดน้อยลงกว่าวิธีการที่ Dalenius ได้เคยเสนอไว้

2.2.2 <u>การกำหนดจำนวนชั้นภูมิ</u>

ในการกำหนดจำนวนชั้นภูมิสำหรับโครงการต่าง ๆ มีผลการวิจัยซึ่งแสดงให้เห็นถึงจำนวนชั้นภูมิที่เหมาะสมและนักวิจัยสามารถนำไปใช้เป็นหลักในทางปฏิบัติได้ โดยผลงานดังกล่าวเป็นผลการศึกษาร่วมกันระหว่าง Cochran(1961) กับ Dalenius และ Gurney (1951) ซึ่งก่อนที่จะกล่าวถึงผลงานนั้นจะได้กล่าวถึงทฤษฎีที่เกี่ยวข้องกับการกำหนดจำนวนชั้นภูมิเสียก่อน เพื่อให้เห็นถึงที่มาตลอดจนเหตุผลในการค้นคว้าหาหลักเกณฑ์ในการกำหนดจำนวนชั้นภูมิดังกล่าว

พิจารณาโดยเริ่มจากการแจกแจงในรูปแบบง่าย ๆ นั่นคือ ให้ตัวแปร y มีการแจกแจงแบบรูปสี่เหลี่ยมผืนผ้าในช่วง (a, a+d) ดังนั้นจะได้ความแปรปรวนของ y เมื่อยังมิได้แบ่งชั้นภูมิดังนี้

$$S_y^2 = \frac{d^2}{12} \qquad \text{...........(5.1)}$$

และสำหรับการเลือกตัวอย่างสุ่มอย่างง่าย ขนาดเท่ากับ n จะได้ว่า

$$V(\bar{y}) = \frac{d^2}{12n} \qquad \text{...........(5.2)}$$

ถ้าแบ่งประชากรของการแจกแจงดังกล่าวออกเป็น L ชั้นภูมิ โดยให้สัดส่วนของแต่ละชั้นภูมิเท่ากัน ตลอดจนกำหนดขนาดตัวอย่างในแต่ละชั้นภูมิเท่ากัน นั่นคือ $W_h = 1/L$ และ $n_h = n/L$ จะได้ความแปรปรวนของตัวแปร y ในแต่ละชั้นภูมิ ดังนี้

$$S_{yh}^2 = \frac{d^2}{12L^2} \qquad \text{...........(5.3)}$$

และได้

$$V(\bar{y}_h) = \frac{d^2}{12L^2 n_h}$$

$$= \frac{d^2}{12L^2 n_h} \qquad \text{...........(5.4)}$$

พิจารณาความแปรปรวนของตัวประมาณที่ใช้วิธีการของเนย์แมนในการกำหนดขนาดตัวอย่าง ในกรณีที่ไม่คิดค่า f.p.c. จะได้ว่า

$$V(\bar{y}_{st}) = \frac{1}{n}\{\sum_{h=1}^{L} W_h S_{yh}\}^2 \qquad \text{...........(5.5)}$$

แทนค่า W_h และ S_{yh} ลงในสมการ (5.5)

$$V(\bar{y}_{st}) = \frac{1}{n}\{\sum_{h=1}^{L} \frac{1}{L} \cdot \frac{d}{\sqrt{12}\,L}\}^2$$

$$= \frac{d^2}{12\,L^2 n}$$

$$= \frac{V(\bar{y})}{L^2} \qquad \text{.........(5.6)}$$

จากสมการ (5.6) จะเห็นว่า L เป็นสัดส่วนผกผันกับ $V(\bar{y}_{st})$ ดังนั้นจึงสามารถอธิบายได้ว่าจำนวนชั้นภูมิยิ่งมากเท่าใด ยิ่งทำให้ความแปรปรวนของตัวประมาณลดลงเท่านั้น

จากที่ได้กล่าวมาจะพบว่าเป็นผลลัพธ์จากข้อสมมติที่กำหนดให้ตัวแปร y มีการแจกแจงแบบรูปสี่เหลี่ยมผืนผ้าในทางปฏิบัติ การแจกแจงของข้อมูลอาจไม่เป็นเช่นนี้ ดังนั้น Cochran (1961) จึงได้ศึกษากับข้อมูลจริง โดยใช้ข้อมูลจากประชากรต่าง ๆ กัน 8 ชุด เพื่อหาค่า $V(\bar{y}_{st})/V(\bar{y})$ แล้วนำมาเปรียบเทียบกับกรณีของการแจกแจงแบบรูปสี่เหลี่ยมผืนผ้า โดยใช้จำนวนชั้นภูมิเท่ากับ 2,3และ 4 ชั้นภูมิ ผลจากการศึกษาพบว่า $V(\bar{y}_{st})/V(\bar{y})$ ให้ค่าเท่ากับ 0.232 0.098 0.053 สำหรับข้อมูลจริงที่ศึกษาและเท่ากับ 0.250 0.111 0.062 สำหรับข้อมูลจากการแจกแจงแบบรูปสี่เหลี่ยมผืนผ้า เมื่อ $L = 2$ 3 และ 4 ตามลำดับ ซึ่งจะเห็นว่าค่าที่ได้ไม่แตกต่างกันมากนัก

จะเห็นว่า ผลจากการศึกษาตามที่กล่าวนั้น Cochran ใช้เป็นเพียงพื้นฐานที่จะทำการศึกษาต่อไป ทั้งนี้เนื่องจากในทางปฏิบัติ ตัวแปร y ที่จะทำการศึกษานั้นเป็นตัวแปรที่ยังไม่ทราบค่า จึงไม่สามารถนำมาใช้ในการกำหนดชั้นภูมิได้ ดังนั้นจะใช้ตัวแปร x ซึ่งสัมพันธ์กับ y และทราบค่าแล้ว เป็นตัวแปรกำหนดชั้นภูมิ ซึ่งถ้า $\phi(x) = E(y|x)$ สามารถเขียนได้ว่า

$$y = \phi(x) + \ell \qquad \text{........(5.7)}$$

โดยที่ ϕ และ ℓ ไม่มีความสัมพันธ์กัน ดังนั้นจะได้ว่า

$$S_y^2 = S_\phi^2 + S_\ell^2 \qquad \text{........(5.8)}$$

พิจารณาเทอมแรกของสมการด้านขวาของ (5.8) นั่นคือ S_ϕ^2 จะพบว่าเมื่อมีการแบ่งชั้นภูมิเกิดขึ้น S_ϕ^2 จะมีค่าลดลง ทั้งนี้เนื่องมาจากการใช้ x เป็นตัวแปรกำหนดชั้นภูมินั่นเอง โดย S_ϕ^2 จะลดลงเรื่อย ๆ ตามจำนวนชั้นภูมิที่เพิ่มมากขึ้น ซึ่งถ้าตัวแปร y กับ $\phi(x)$ มีรูปแบบของความสัมพันธ์ในรูปเชิงเส้น (Linear Regression Model) และกำหนดจำนวนชั้นภูมิเท่ากับ L ชั้นภูมิ จะมีผลทำให้ S_ϕ^2 ลดลงเป็น S_ϕ^2/L^2

สำหรับเทอมที่ 2 ของสมการด้านขวาของ (5.8) นั่นคือ S_{ℓ}^2 จะพบว่า เมื่อมีการแบ่งชั้นภูมิ S_{ℓ}^2 จะยังคงมีค่าเท่าเดิม เนื่องจาก x ซึ่งเป็นตัวแปรกำหนดชั้นภูมิไม่มีความสัมพันธ์กับ ℓ แต่อย่างใด ดังนั้นไม่ว่าจะมีการเพิ่มจำนวนชั้นภูมิมากขึ้นเพียงใด และตัวแปร y กับ $\phi(x)$ จะมีลักษณะของความสัมพันธ์ในรูปแบบใดก็ตาม S_{ℓ}^2 จะไม่มีการเปลี่ยน-แปลง

จากข้อเท็จจริงดังกล่าว จึงเห็นได้ว่า เมื่อมีการเพิ่มจำนวนชั้นภูมิไปถึงจุดหนึ่ง ซึ่งทำให้ S_{ϕ}^2/L^2 มีค่าน้อยมากเมื่อเทียบกับ S_{ℓ}^2 ดังนั้นจึงมีผลต่อการเปลี่ยน-แปลงของ S_y^2 ในทางลดลงน้อยมากตามไปด้วยและนั่นก็หมายถึง การมีผลต่อการลดลงของความแปรปรวนของตัวประมาณของตัวแปรที่ทำการศึกษาน้อยมากเช่นกัน

พิจารณาตัวแปร x ซึ่งเป็นตัวแปรกำหนดชั้นภูมิ จะได้ว่า ความแปรปรวนของตัวประมาณเมื่อใช้วิธีการของเนย์แมนในการกำหนดขนาดตัวอย่างในแต่ละชั้นภูมิ ในกรณีที่ไม่คิดค่า f.p.c. ดังนี้

$$V(\bar{x}_{st}) = \Sigma \frac{W_h^2 S_{xh}^2}{n_h} \qquad \text{............(5.9)}$$

เมื่อแบ่งประชากรของการแจกแจงของตัวแปร x ออกเป็น L ชั้นภูมิ โดยให้สัดส่วนของแต่ละชั้นภูมิเท่ากันตลอดจนกำหนดขนาดตัวอย่างในแต่ละชั้นภูมิเท่ากัน นั่นคือ $W_h = 1/L$ และ $n_h = n/L$ เช่นเดียวกับตัวแปร y จะได้ความแปรปรวนของตัวประมาณของตัวแปร x จากสมการ (5.9) ดังนี้

$$\begin{aligned} V(\bar{x}_{st}) &= \frac{L}{n} \Sigma W_h^2 S_{xh}^2 \\ &= \frac{L}{n} \left[\frac{1}{L^2} \Sigma S_{xh}^2 \right] \\ &= \frac{S_x^2}{nL^2} \qquad \text{............(5.10)} \end{aligned}$$

ในกรณีที่ y กับ x มีความสัมพันธ์กันในรูปเชิงเส้น นั่นคือ

$$y = \alpha + \beta x + \ell \qquad \text{............(5.11)}$$

โดยที่ S_{ℓ}^2 มีค่าคงที่ตามที่กล่าวแล้ว ดังนั้นจะได้ค่าความแปรปรวนของตัวประมาณของตัวแปรที่ศึกษา ดังนี้

$$V(\bar{y}_{st}) = \frac{L}{n} \Sigma W_h^2 S_{yh}^2$$

$$= \frac{L\beta^2}{n} \Sigma W_h^2 S_{xh}^2 + \frac{LS_{\ell}^2}{n} \Sigma W_h^2 \quad .....(5.12)$$

แทนค่าสมการ (5.10) ลงในสมการ (5.12) และจากการที่ $\Sigma W_h^2 \geqslant \frac{1}{L}$ ดังนั้นจะได้ว่า

$$V(\bar{y}_{st}) \geqslant \frac{1}{n} [\frac{\beta^2 S_x^2}{L^2} + S_{\ell}^2]$$

$$\geqslant \frac{S_y^2}{n} [\frac{\rho^2}{L^2} + (1-\rho^2)] \quad .....(5.13)$$

ซึ่ง ρ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง y กับ x เมื่อยังมิได้แบ่งชั้นภูมิ

จะเห็นว่า ความแปรปรวนของตัวประมาณจากอสมการ(5.13) จะให้ค่าต่ำที่สุด เมื่อ

$$V(\bar{y}_{st}) = \frac{S_y^2}{n}[\frac{\rho^2}{L^2} + (1-\rho^2)] \quad ....(5.14)$$

จากสมการ (5.14) ซึ่งอยู่ในรูปฟังก์ชันของ ρ กับ L Cochran ได้ทำการวิจัยต่อไป โดยใช้กับข้อมูลที่จำลองขึ้นตามค่าของสัมประสิทธิ์สหสัมพันธ์ระหว่าง y กับ x ซึ่งกำหนดค่าสัมประสิทธิ์สหสัมพันธ์เป็น 4 ระดับ คือ 0.99 0.95 0.90 และ 0.85 ในแต่ละระดับได้ทำการศึกษาเพื่อดูการเปลี่ยนแปลงของค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณจากการใช้วิธีการของการเลือกตัวอย่างสุ่มอย่างง่าย เมื่อเทียบกับตัวประมาณจากการใช้วิธีการของการเลือกตัวอย่างแบบมีชั้นภูมิ เมื่อจำนวนชั้นภูมิที่ใช้เพิ่มสูงขึ้น ซึ่งผลที่ได้จากการวิจัยแสดงไว้ในตารางที่ 2.1

ตารางที่ 2.1 แสดงค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณจากการใช้วิธีการของการเลือกตัวอย่างสุ่มอย่างง่าย เมื่อเทียบกับตัวประมาณจากการใช้วิธีการของการเลือกตัวอย่างแบบมีชั้นภูมิ เมื่อใช้โมเดลของความสัมพันธ์ระหว่างตัวแปรที่ศึกษากับตัวแปรกำหนดชั้นภูมิในรูปแบบความถดถอยเชิงเส้น จำแนกตามจำนวนชั้นภูมิที่ใช้ และค่าสัมประสิทธิ์สหสัมพันธ์จากโมเดลที่ศึกษา

จำนวนชั้นภูมิ (L)	ค่าประสิทธิภาพสัมพัทธ์			
	ρ = 0.99	ρ = 0.95	ρ = 0.90	ρ = 0.85
2	0.265	0.323	0.392	0.458
3	0.129	0.198	0.280	0.358
4	0.081	0.154	0.241	0.323
5	0.059	0.134	0.222	0.306
6	0.047	0.123	0.212	0.298
∞	0.020	0.098	0.190	0.277

จากผลการวิจัยที่ได้จะเห็นว่า ค่าประสิทธิภาพสัมพัทธ์จากการเปรียบเทียบตัวประมาณจะลดลงเรื่อย ๆ ตามจำนวนชั้นภูมิที่เพิ่มขึ้นในทุกระดับของค่าสัมประสิทธิ์สหสัมพันธ์ และเมื่อพิจารณาต่อไปจะพบว่า หลังจากที่ใช้จำนวนชั้นภูมิมากกว่า 6 ชั้นภูมิ ค่าประสิทธิภาพสัมพัทธ์ที่ได้แทบจะไม่ลดลงอีกนอกจากกรณีที่ค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 1 หรือนั่นก็คือ ความสัมพันธ์ระหว่าง y กับ x มีค่าสูงมาก นั่นเอง ซึ่งในทางปฏิบัติจริงตัวแปรที่ศึกษาและตัวแปรกำหนดชั้นภูมิที่มีความสัมพันธ์กันสูงมากในระดับดังกล่าวไม่ค่อยพบแต่อย่างใด

นอกจากนี้ยังมีผลการวิจัยที่ได้ศึกษากับข้อมูลจริง โดยใช้ข้อมูลจากประชากรต่าง ๆ กัน 3 ชุด ซึ่งเป็นผลงานของ Cochran(1961) 2 ชุด คือ ข้อมูลจำนวนนักศึกษา และข้อมูลขนาดของเมือง กับของ Dalenius และ Gurney(1951) 1 ชุด คือ ข้อมูลรายได้ของครอบครัว ซึ่งผลที่ได้จากการวิจัยแสดงไว้ในตารางที่ 2.2

<u>ตารางที่ 2.2</u> แสดงค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณจากการใช้วิธีการของการเลือกตัวอย่างสุ่มอย่างง่าย เมื่อเทียบกับตัวประมาณจากการใช้วิธีการของการเลือกตัวอย่างแบบมีชั้นภูมิ เมื่อใช้กับข้อมูลจากประชากรต่าง ๆ กัน 3 ชุด จำแนกตามจำนวนชั้นภูมิที่ใช้ และชุดของข้อมูลที่ศึกษา

จำนวนชั้นภูมิ	ค่าประสิทธิภาพสัมพัทธ์		
(L)	จำนวนนักศึกษา	ขนาดของเมือง	รายได้ของครอบครัว
2	0.197	0.295	0.500
3	0.108	0.178	0.375
4	0.075	0.142	0.244
5	0.065	0.105	0.241
6	0.050	0.104	0.212
∞	-	-	-

สำหรับตัวแปรกำหนดชั้นภูมิของประชากรทั้ง 3 ชุดนี้ พบว่า อยู่ในลักษณะอย่างเดียวกัน นั่นก็คือ ใช้ตัวแปรที่ศึกษาเป็นตัวแปรกำหนดชั้นภูมิ ทั้งนี้โดยการใช้ข้อมูลของตัวแปรที่ศึกษาซึ่งได้ผ่านการเก็บรวบรวมมาก่อนข้อมูลที่เก็บรวบรวมในปีที่ศึกษาเป็นข้อมูลของตัวแปรกำหนดชั้นภูมิ ซึ่งมีรายละเอียดดังนี้

ข้อมูลจำนวนนักศึกษา : เก็บรวบรวมข้อมูลของตัวแปรกำหนดชั้นภูมิใน ค.ศ. 1952
เก็บรวบรวมข้อมูลของตัวแปรที่ศึกษาใน ค.ศ. 1958

ข้อมูลขนาดของเมือง : เก็บรวบรวมข้อมูลของตัวแปรกำหนดชั้นภูมิใน ค.ศ. 1940
เก็บรวบรวมข้อมูลของตัวแปรที่ศึกษาใน ค.ศ. 1950

ข้อมูลรายได้ของครอบครัว : เก็บรวบรวมข้อมูลของตัวแปรกำหนดชั้นภูมิใน ค.ศ. 1929
เก็บรวบรวมข้อมูลของตัวแปรที่ศึกษาใน ค.ศ. 1933

จะเห็นว่าผลที่ได้จากการวิจัยซึ่งได้ศึกษากับข้อมูลจริง สำหรับประชากรที่ 2 และประชากรที่ 3 ก็ยังคงมีลักษณะเช่นเดียวกับผลการวิจัยที่ได้จากข้อมูลที่จำลองขึ้น นั่นก็คือ หลังจากที่ใช้จำนวนชั้นภูมิมากกว่า 6 ชั้นภูมิ ค่าประสิทธิภาพสัมพัทธ์ที่ได้จากการเปรียบเทียบตัวประมาณจะลดลงไปจากเดิมอีกน้อยมาก ทั้งนี้เมื่อพิจารณากลับไปยังโมเดลของความสัมพันธ์ระหว่างตัวแปรที่ศึกษากับตัวแปรกำหนดชั้นภูมิของทั้ง 2 ประชากร จะพบว่าต่างก็อยู่ในรูปแบบความถดถอยเชิงเส้น ส่วนประชากร 1 จะพบว่า ค่าประสิทธิภาพสัมพัทธ์ที่ได้อยู่ในลักษณะของการลดลงตามจำนวนชั้นภูมิที่เพิ่มขึ้นเช่นกัน แต่จะพบว่า สัดส่วนของการลดลงอยู่ในลักษณะที่แตกต่างไปจากกรณีของประชากรที่ 2 และประชากรที่ 3 ซึ่งเมื่อพิจารณาต่อไป จะเห็นว่า สาเหตุดังกล่าวเนื่องมาจากโมเดลของความสัมพันธ์ระหว่างตัวแปรที่ศึกษากับตัวแปรกำหนดชั้นภูมิมิได้อยู่ในรูปแบบที่เป็นไปตามทฤษฎีนั่นเอง