บทที่ 3

ตัวประมาณค่ารวมประชากรที่เสนอแนะ

ในวิธีการประมาณค่ารวมประชากร ดังที่ได้กล่าวมาแล้วในบทที่ 2 จะเห็นว่า ตัวประมาณที่ใช้ในการประมาณค่ารวมประชากร หรือค่าเฉลี่ยประชากรจะอยู่ในรูปค่าเฉลี่ยตัวอย่าง ($\bar{y}$) หรือฟังก์ชันของค่าเฉลี่ยตัวอย่าง และเนื่องจากการทำวิจัยโดยทั่วไปมักจะมีการเก็บรวบรวมข้อมูลของตัวแปรต่าง ๆ อีกหลายตัวแปร เพื่อใช้ในการวิเคราะห์และหาข้อสรุปสำหรับตอบสนองต่อวัตถุประสงค์ที่ได้กำหนดไว้ ซึ่งโดยทั่วไปมักจะพบว่าในบรรดาตัวแปรเหล่านี้มีตัวแปรบางตัวที่มีความสัมพันธ์เชิงเส้นต่อกัน เพราะฉะนั้นผู้วิจัยน่าจะพิจารณาถึงวิธีการประมาณค่ารวมประชากรที่ใช้ตัวแปรอีกตัวแปรหนึ่งนอกเหนือจากตัวแปรที่เราสนใจศึกษามาเพิ่มประสิทธิภาพของการประมาณค่ารวมประชากร

วิทยานิพนธ์นี้จึงได้นำเอาคุณสมบัติเกี่ยวกับตัวประมาณความถดถอยและการแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้วมาใช้เพื่อที่จะหารูปแบบตัวประมาณค่ารวมประชากรใหม่ ทั้งนี้จะได้ลดอิทธิพลของการที่มีตัวอย่างสุ่มอย่างง่ายบางหน่วยที่มีค่าสูงมาก เมื่อพบว่า ตัวแปร Y ที่เราสนใจศึกษามีความสัมพันธ์เชิงเส้นต่อกันกับตัวแปร X โดยที่ตัวแปร X ไม่มีค่าสังเกตค่าใดเป็นค่าสูงมาก (ตัวแปร X อาจเป็นค่าของตัวแปร Y ในเวลาที่ต่างกัน ตัวอย่างเช่น จากการทำสำมะโนครั้งที่แล้ว เป็นต้น) ทั้งนี้ ต้องทราบค่าเฉลี่ยประชากรของตัวแปร X ในแต่ละกลุ่มด้วย กล่าวคือ ทราบค่าเฉลี่ยประชากรของตัวแปร X ในกลุ่มที่มีค่าสังเกตของตัวแปร Y เป็นค่าสูงมาก และทราบค่าเฉลี่ยประชากรของตัวแปร X ในกลุ่มที่ไม่มีค่าสังเกตของตัวแปร Y เป็นค่าสูงมาก

โดยสมมติให้เลือกตัวอย่างแบบสุ่มอย่างง่ายชนิดไม่ใส่คืนมาขนาด n จากประชากรขนาด N คือ $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)\}$ ซึ่งมีจำนวนค่าสังเกตที่เป็นค่าสูงมากเท่ากับ N_1 คือ $\{(X_1,Y_1),(X_2,Y_2),\ldots,(X_{N_1},Y_{N_1})\}$ อยู่ในประชากร แล้วแบ่งตัวอย่างที่สุ่มมาได้ออกเป็น 2 กลุ่มคือ กลุ่มที่ประกอบด้วยหน่วยตัวอย่าง Y ที่มีค่าสูงมาก ซึ่งมีจำนวนเท่ากับ n_1 ตัวอย่างคือ $\{(x_1, y_1), (x_2, y_2), \ldots, (x_{n1}, y_{n1})\}$ กับกลุ่มที่ประกอบด้วยหน่วยตัวอย่าง y ที่มีค่าปกติ โดยที่แต่ละหน่วยตัวอย่างจะตกอยู่ในประชากรย่อยหรือกลุ่มได้เพียงกลุ่มเดียวเท่านั้น และกำหนดให้

$\bar{Y}_1$ = ค่าเฉลี่ยประชากรของกลุ่มตัวแปร Y ที่มีค่าสังเกตเป็นค่าสูงมาก

$\bar{Y}_2$ = ค่าเฉลี่ยประชากรของกลุ่มตัวแปร Y ที่ไม่มีค่าสังเกตใดเป็นค่าสูงมาก หรืออีกนัย-หนึ่งก็คือ ค่าเฉลี่ยประชากรของกลุ่มตัวแปร Y ที่มีค่าสังเกตเป็นค่าปกติ

$\bar{X}_1$ = ค่าเฉลี่ยประชากรของตัวแปร X ของกลุ่มที่มีค่าสังเกต Y เป็นค่าสูงมาก

$\bar{X}_2$ = ค่าเฉลี่ยประชากรของตัวแปร X ของกลุ่มที่มีค่าสังเกต Y เป็นค่าปกติ

$\bar{y}_1$ = ค่าเฉลี่ยตัวอย่างของตัวแปร Y ของกลุ่มที่มีค่าสังเกต เป็นค่าปกติ

$\bar{y}_2$ = ค่าเฉลี่ยตัวอย่างของกลุ่มตัวแปร Y ที่หน่วยตัวอย่างมีค่าปกติ

$\bar{x}_1$ = ค่าเฉลี่ยตัวอย่างของตัวแปร X จากกลุ่มตัวแปร Y ที่มีหน่วยตัวอย่างมีค่าสูงมาก

$\bar{x}_2$ = ค่าเฉลี่ยตัวอย่างของตัวแปร X จากกลุ่มตัวแปร Y ที่มีหน่วยตัวอย่างเป็นค่าปกติ

N_1 = จำนวนค่าสังเกตของตัวแปร Y ในประชากรที่เป็นค่าสูงมาก

n_1 = จำนวนหน่วยตัวอย่างของตัวแปร Y ที่มีค่าสูงมากที่พบได้ในตัวอย่าง

ตัวประมาณค่ารวมประชากรที่เสนอแนะคือ

กรณีที่ทราบค่า N_1

$$\hat{Y}_1 = N_1 \bar{y}_{\ell r1} + N_2 \bar{y}_{\ell r2}$$

กรณีที่ไม่ทราบค่า N_1

$$\hat{Y}_2 = \frac{N}{n} (n_1 \bar{y}_{\ell r1} + n_2 \bar{y}_{\ell r2})$$

$$\hat{Y}_3 = n_1\bar{y}_{\ell r1} + (N-n_1)\bar{y}_{\ell r2}$$

เมื่อ $\bar{y}_{\ell rj} = \bar{y}_j - b_j(\bar{x}_j - \bar{X}_j)$; $j = 1, 2$ และ b_j เป็นค่าสัมประสิทธิ์ความถดถอยของแต่ละกลุ่ม

3.1 <u>การพิจารณาค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean square error) ของแต่ละตัวประมาณ</u> (MSE $(\hat{Y}k)$; $k = 1, 2, 3$)

การศึกษาคุณสมบัติและการเปรียบเทียบตัวประมาณที่เสนอแนะกับตัวประมาณ $\hat{Y}_o$ ที่มีค่าเท่ากับ $N\bar{y}$ และตัวประมาณที่เสนอโดยไม่เคิล และคาตาบา ในปี ค.ศ. 1981 จะพิจารณาจากค่าประสิทธิภาพสัมพัทธ์เป็นหลัก ซึ่งค่าประสิทธิภาพสัมพัทธ์เป็นอัตราส่วนระหว่างค่าความคลาดเคลื่อนกำลังสองเฉลี่ย 2 ค่า โดยในการคิดค่า MSE $(\hat{Y}k)$ จะคิดอย่างมีเงื่อนไขเมื่อกำหนดค่า n_1 (conditional mean square error on n_1) หรืออย่างไม่มีเงื่อนไขเกี่ยวกับค่า n_1 (unconditional mean square error on n_1) ถ้า

1. กรณีที่ทำการอนุมานหลังจากการสุ่มตัวอย่างแล้วและทราบว่าในตัวอย่างมีจำนวนหน่วยตัวอย่างที่เป็นค่าสูงมากอยู่เป็นจำนวนเท่าใดที่แน่นอน พร้อมทั้งรู้ว่ามีตัวประมาณตัวหนึ่งมีประสิทธิภาพสูงกว่าหรือน้อยกว่าตัวอื่น ๆ มาก ก็จะคิดค่า MSE ของ $\hat{Y}k$ อย่างมีเงื่อนไข

2. กรณีที่ทำการอนุมานหลังจากการสุ่มตัวอย่างแล้ว และไม่ทราบว่าจะใช้เกณฑ์อะไรบ่งชี้ว่าค่าสังเกตค่าใดเป็นค่าสูงมาก หรือไม่สนใจว่าในตัวอย่างที่เลือกมาจะมีหน่วยตัวอย่างที่มีค่าสูงมากเป็นจำนวนเท่าใด แต่คาดว่าจะต้องมีอย่างน้อยตั้งแต่ 1 หน่วยขึ้นไป ก็จะคิดค่า MSE ของ $\hat{Y}k$ อย่างไม่มีเงื่อนไข

รายละเอียดต่าง ๆ ในแต่ละแบบของค่า MSE เป็นดังนี้

3.1.1 การอนุมานอย่างมีเงื่อนไขหรือเมื่อกำหนดค่า n_1 (conditional inference)

ก. กรณีที่ทราบค่า b_j

<u>ทฤษฎีที่ 1</u> ในการเลือกตัวอย่างแบบสุ่มอย่างง่ายชนิดไม่ใส่คืน กรณีที่ทราบค่า b_j โดยให้ $b_j = \sum_{i=1}^{N_j}(X_{ij}-\bar{X}_j)(Y_{ij}-\bar{Y}_j)/\sum_{i=1}^{N_j}(X_{ij}-\bar{X}_j)^2$ จะได้ $\hat{Y}_1 = N_1\bar{y}_{\ell r1} + N_2\,\bar{y}_{\ell r2}$ เป็นตัวประมาณที่ไม่มีความเอนเอียงของค่ารวมประชากร $(B(\hat{Y}_1|n_1) = 0)$ และมี

ความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างมีเงื่อนไขเท่ากับ $\sum_{j=1}^{2} N_j^2 \left(\frac{1}{n_j} - \frac{1}{N_j}\right)$.

$$\frac{1}{N_j-1} \cdot \sum_{i=1}^{N_j}((Y_{ij}-\bar{Y}_j) - b_j(X_{ij}-\bar{X}_j))^2$$

<u>พิสูจน์</u>

1.1 $B(\hat{Y}_1|n_1) = 0$

จาก $\hat{Y}_1 = N_1 \bar{y}_{\ell r1} + N_2 \bar{y}_{\ell r2}$

$$\therefore E(\hat{Y}_1|n_1) = E[(N_1 \bar{y}_{\ell r1} + N_2 \bar{y}_{\ell r2})|n_1]$$

$$= N_1 E(\bar{y}_{\ell r1}) + N_2 E(\bar{y}_{\ell r2})$$

จากคุณสมบัติของตัวประมาณความถดถอยที่ว่า $E(\bar{y}_{\ell rj}) = \bar{Y}_j$ ดังนั้นจะได้

$$E(\hat{Y}_1|n_1) = N_1 \bar{Y}_1 + N_2 \bar{Y}_2$$

$$= \sum_{i=1}^{N_1} Y_i + \sum_{i=N_1+1}^{N} Y_i$$

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

$$= Y$$

ดังนั้น $B(\hat{Y}_1|n_1) = E(\hat{Y}_1|n_1)-Y = 0$ ซ.ต.พ.

1.2 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างมีเงื่อนไขของ $\hat{Y}_1$

จากค่าความคลาดเคลื่อนกำลังสองอย่างมีเงื่อนไข = ค่าความแปรปรวนอย่างมีเงื่อนไข + ความเอนเอียงอย่างมีเงื่อนไขยกกำลังสอง

หรือ $MSE(\hat{Y}k|n_1) = V(\hat{Y}k|n_1)+B(\hat{Y}k|n_1)^2$ โดยที่ $V(\hat{Y}k|n_1)$

คือค่าความแปรปรวนอย่างมีเงื่อนไข

และ $\hat{Y}_1 = N_1\,\bar{y}_{\ell r1} + N_2\,\bar{y}_{\ell r2}$ ซึ่งเขียนในอีกรูปแบบหนึ่งได้เป็น

$$\hat{Y}_1 = N\sum_{j=1}^{2} W_j\,\bar{y}_{\ell rj} \quad \text{เมื่อให้ } W_j = N_j/N$$

$$\therefore\; V(\hat{Y}_1|n_1) = V(N\sum_{j=1}^{2} W_j\,\bar{y}_{\ell rj})$$

$$= N^2\sum_{j=1}^{2} W_j^2\,V(\bar{y}_{\ell rj}|n_1)$$

จากคุณสมบัติของตัวประมาณความถดถอย เมื่อเลือกตัวอย่างแบบแบ่งชั้นภูมิจะได้

$$V(\hat{Y}_1|n_1) = N^2\sum_{j=1}^{2} W_j^2\,(1-\frac{n_j}{N_j}).\frac{1}{n_j}\;\frac{1}{N_j-1}\sum_{i=1}^{N_j}\{(Y_{ij}-\bar{Y}_j)-b_j(X_{ij}-\bar{X}_j)\}^2$$

เนื่องจาก $B(\hat{Y}_1|n_1) = 0$

และ $MSE\,(\hat{Y}_1|n_1) = V(\hat{Y}_1|n_1) + B(\hat{Y}_1|n_1)^2$

$$\therefore\; MSE\,(\hat{Y}_1|n_1) = N^2\sum_{j=1}^{2} W_j^2(\frac{1}{n_j}-\frac{1}{N_j}).\frac{1}{N_j-1}\sum_{i=1}^{N_j}\{(Y_{ij}-\bar{Y}_j)-b_j(X_{ij}-\bar{X}_j)\}^2$$

แทนค่า $W_j = N_j/N$ จะได้

$$MSE\,(\hat{Y}_1|n_1) = \sum_{j=1}^{2} N_j^2(\frac{1}{n_j}-\frac{1}{N_j}).\frac{1}{N_j-1}\sum_{i=1}^{N_j}\{(Y_{ij}-\bar{Y}_j)-b_j(X_{ij}-\bar{X}_j)\}^2$$

ทฤษฎีที่ 2 ในการเลือกตัวอย่างแบบสุ่มอย่างง่ายชนิดไม่ใส่คืน กรณีที่ทราบค่า b_j โดยให้ b_j มีค่าเหมือนในทฤษฎีที่ 1 จะได้ $\hat{Y}_2 = \frac{N}{n}(n_1 \bar{y}_{\ell r1} + n_2 \bar{y}_{\ell r2})$ เป็นตัวประมาณที่มีความเอนเอียงของค่ารวมประชากร โดยมีความเอนเอียงอย่างมีเงื่อนไข $(B(\hat{Y}_2|n_1))$ เท่ากับ $N\left[(\frac{n_1}{n} - \frac{N_1}{N})\bar{Y}_1 + (\frac{n_2}{n} - \frac{N_2}{N})\bar{Y}_2\right]$ และมีความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างมีเงื่อนไขเท่ากับ $N^2\left[\frac{1}{n^2}\sum_{j=1}^{2}(n_j - \frac{n_j^2}{N_j}) \cdot \frac{1}{N_j-1}\sum_{i=1}^{N_j}\{(Y_{ij}-\bar{Y}_j)-b_j \cdot (X_{ij}-\bar{X}_j)\}^2 + \{\sum_{j=1}^{2}(\frac{n_j}{n} - \frac{N_j}{N})\bar{Y}_j\}^2\right]$

พิสูจน์

2.1 $$B(\hat{Y}_2|n_1) = N\left[(\frac{n_1}{n} - \frac{N_1}{N})\bar{Y}_1 + (\frac{n_2}{n} - \frac{N_2}{N})\bar{Y}_2\right]$$

ให้ $w_j = n_j/n$ และ $W_j = N_j/N$

จาก $$\hat{Y}_2 = N(\frac{n_1}{n}\bar{y}_{\ell r1} + \frac{n_2}{n}\bar{y}_{\ell r2})$$

$$\therefore \quad \hat{Y}_2 = N\sum_{j=1}^{2} w_j \bar{y}_{\ell rj}$$

$$= N\sum_{j=1}^{2} W_j \bar{y}_{\ell rj} + N\sum_{j=1}^{2} w_j \bar{y}_{\ell rj} - N\sum_{j=1}^{2} W_j \bar{y}_{\ell rj}$$

$$= N\sum_{j=1}^{2} W_j \bar{y}_{\ell rj} + N\left[\sum_{j=1}^{2}(w_j - W_j)\bar{Y}_{\ell rj}\right]$$

ดังนั้น $$E(\hat{Y}_2|n_1) = N\sum_{j=1}^{2} W_j\, E(\bar{y}_{\ell rj}) + N\left[\sum_{j=1}^{2} (w_j - W_j)\, E(\bar{y}_{\ell rj})\right]$$

$$= N\sum_{j=1}^{2} W_j\, \bar{Y}_j + N\left[\sum_{j=1}^{2} (w_j - W_j)\, \bar{Y}_j\right]$$

จากคุณสมบัติที่ว่า $$\sum_{j=1}^{2} W_j\, \bar{Y}_j = \sum_{j=1}^{2} \frac{N_j\, \bar{Y}_j}{N} = \sum_{i=1}^{N} \frac{Y_i}{N} = \bar{Y}$$

$$\therefore\; E(\hat{Y}_2|n_1) = N.\,\bar{Y} + N\left[\sum_{j=1}^{2} (w_j - W_j)\, \bar{Y}_j\right]$$

$$= Y + N\left[\sum_{j=1}^{2} (w_j - W_j)\, \bar{Y}_j\right]$$

ดังนั้น $$B(\hat{Y}_2|n_1) = E(\hat{Y}_2|n_1) - Y = N\left(\sum_{j=1}^{2} (w_j - W_j)\, \bar{Y}_j\right)$$

แทนค่า w_j, W_j จะได้

$$B(\hat{Y}_2|n_1) = N\left[\left(\frac{n_1}{n} - \frac{N_1}{N}\right)\bar{Y}_1 + \left(\frac{n_2}{n} - \frac{N_2}{N}\right)\bar{Y}_2\right]$$ ซ.ต.พ.

2.2 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างมีเงื่อนไขของ $\hat{Y}_2$

จาก $$\hat{Y}_2 = \frac{N}{n}(n_1\, \bar{y}_{\ell r1} + n_2\, \bar{y}_{\ell r2})$$

$$= N\sum_{j=1}^{2} w_j\, \bar{y}_{\ell rj} \quad ; \quad w_j = n_j/n$$

$$\therefore\; V(\hat{Y}_2|n_1) = V\left(N\sum_{j=1}^{2} w_j\, \bar{y}_{\ell rj}\right)$$

$$= N^2 \sum_{j=1}^{2} w_j^2 \, V(\bar{y}_{\ell rj})$$

$$= N^2 \sum_{j=1}^{2} w_j^2 (1-\frac{n_j}{N_j}) \cdot \frac{1}{n_j} \cdot \frac{1}{N_j-1} \sum_{i=1}^{N_j} \{(Y_{ij}-\bar{Y}_j)-b_j(X_{ij}-\bar{X}_j)\}^2$$

แทนค่า w_j จะได้

$$\therefore \; V(\hat{Y}_2|n_1) = \frac{N^2}{n^2} \sum_{j=1}^{2} (n_j-\frac{n_j^2}{N_j}) \cdot \frac{1}{N_j-1} \sum_{i=1}^{N_j} \{(Y_{ij}-\bar{Y}_j)-b_j(X_{ij}-\bar{X}_j)\}^2$$

และจาก $B(\hat{Y}_2|n_1) = N \sum_{j=1}^{2} (\frac{n_j}{n} - \frac{N_j}{N}) \bar{Y}_j$

$$\therefore \; MSE(\hat{Y}_2|n_1) = V(\hat{Y}_2|n_1) + B(\hat{Y}_2|n_1)^2$$

$$= N^2 \Big[\frac{1}{n^2} \sum_{j=1}^{2} (n_j-\frac{n_j^2}{N_j}) \cdot \frac{1}{N_j-1} \sum_{i=1}^{N_j} \{(Y_{ij}-\bar{Y}_j)-b_j(X_{ij}-\bar{X}_j)\}^2 + \{\sum_{j=1}^{2} (\frac{n_j}{n} - \frac{N_j}{N})\bar{Y}_j\}^2 \Big]$$

ช.ต.พ.

<u>ทฤษฎีที่ 3</u> ในการเลือกตัวอย่างแบบสุ่มอย่างง่ายชนิดไม่ใส่คืน กรณีที่ทราบค่า b_j โดยให้ b_j มีค่าเหมือนในทฤษฎีที่ 1 จะได้ $\hat{Y}_3 = n_1 \bar{y}_{\ell r1} + (N-n_1)\bar{y}_{\ell r2}$ เป็นตัวประมาณค่ารวมประชากรที่มีความเอนเอียง โดยมีความเอนเอียงอย่างมีเงื่อนไข $(B(\hat{Y}_3|n_1))$ เท่ากับ $(N_1-n_1)(\bar{Y}_2-\bar{Y}_1)$ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างมีเงื่อนไข

เท่ากับ $(n_1-\frac{n_1^2}{N_1}) \cdot \sum_{i=1}^{N_1} \{(Y_{i1}-\bar{Y}_1)-b_1(X_{i1}-\bar{X}_1)\}^2/(N_1-1)+(N-n_1)^2 \; (\frac{1}{n_2} - \frac{1}{N_2}) \cdot$

$$\frac{\sum_{i=N_1+1}^{N} \{(Y_{i2}-\bar{Y}_2)-b_2(X_{i2}-\bar{X}_2)\}^2 + \left[(N_1-n_1)(\bar{Y}_2-\bar{Y}_1)\right]^2}{N_2-1}$$

พิสูจน์

3.1 $B(\hat{Y}_3|n_1) = (N_1-n_1)(\bar{Y}_2-\bar{Y}_1)$

จาก $\hat{Y}_3 = n_1 \bar{y}_{\ell r1} + (N-n_1)\bar{y}_{\ell r2}$

$$= n_1 \bar{y}_{\ell r1}+(N-n_1)\bar{y}_{\ell r2} + \sum_{j=1}^{2} N_j \bar{y}_{\ell rj} - \sum_{j=1}^{2} N_j \bar{y}_{\ell rj}$$

$$= \sum_{j=1}^{2} N_j\bar{y}_{\ell rj}+n_1(\bar{y}_{\ell r1}-\bar{y}_{\ell r2})+N\bar{y}_{\ell r2}-N_1\bar{y}_{\ell r1}-N_2\bar{y}_{\ell r2}$$

จาก $\hat{Y}_1 = \sum_{j=1}^{2} N_j \bar{y}_{\ell rj}$

$\therefore \quad \hat{Y}_3 = \hat{Y}_1 + n_1(\bar{y}_{\ell r1}-\bar{y}_{\ell r2})+(N_1+N_2)\bar{y}_{\ell r2}-N_1\bar{y}_{\ell r1}-N_2\bar{y}_{\ell r2}$

$$= \hat{Y}_1 + n_1(\bar{y}_{\ell r1}-\bar{y}_{\ell r2}) + N_1\bar{y}_{\ell r2} - N_1\bar{y}_{\ell r1}$$

$$= \hat{Y}_1 + (N_1-n_1)(\bar{y}_{\ell r2}-\bar{y}_{\ell r1})$$

$\therefore \quad E(\hat{Y}_3|n_1) = E\hat{Y}_1 + (N_1-n_1)(E(\bar{y}_{\ell r2})-E(\bar{y}_{\ell r1}))$

จาก 1.1 ที่ว่า $E\hat{Y}_1 = Y$

ดังนั้น $E(\hat{Y}_3|n_1) = Y + (N_1-n_1)(\bar{Y}_2-\bar{Y}_1)$

$$\therefore \quad B(\hat{Y}_3|n_1) = E(\hat{Y}_3|n_1)-Y$$

$$= (N_1-n_1)(\bar{Y}_2-\bar{Y}_1) \qquad \text{ซ.ต.พ.}$$

3.2 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างมีเงื่อนไขของ $\hat{Y}_3$

จาก $\hat{Y}_3 = n_1\,\bar{y}_{\ell r1} + (N-n_1)\bar{y}_{\ell r2}$

$$\therefore \quad V(\hat{Y}_3|n_1) = n_1^2V(\bar{y}_{\ell r1})+(N-n_1)^2V(\bar{y}_{\ell r2}) \quad \left[\because \bar{y}_{\ell r1} \text{ เป็นอิสระกันกับ } \bar{y}_{\ell r2}\right]$$

$$= n_1^2\left(1-\frac{n_1}{N_1}\right).\frac{1}{n_1}\frac{\sum_{i=1}^{N_1}\{(Y_{i1}-\bar{Y}_1)-b_1(X_{i1}-\bar{X}_1)\}^2}{N_1-1} + (N-n_1)^2.$$

$$\left(1-\frac{n_2}{N_2}\right).\frac{1}{n_2}\frac{\sum_{i=n_1+1}^{N}\{(Y_{i2}-\bar{Y}_2)-b_2(X_{i2}-\bar{X}_2)\}^2}{N_2-1}$$

และจาก $B(\hat{Y}_3|n_1) = (N_1-n_1)(\bar{Y}_2-\bar{Y}_1)$

ดังนั้น $MSE\,(\hat{Y}_3|n_1) = V(\hat{Y}_3|n_1)+B(\hat{Y}_3|n_1)^2$

$$= (n_1-\frac{n_1^2}{N_1}) \cdot \frac{1}{N_1-1} \sum_{i=1}^{N_1} \{(Y_{i1}-\bar{Y}_1)-b_1(X_{i1}-\bar{X}_1)\}^2$$

$$+ (N-n_1)^2(\frac{1}{n_2} - \frac{1}{N_2}) \cdot \frac{1}{N_2-1} \sum_{i=N_1+1}^{N} \{(Y_{i2}-\bar{Y}_2)-b_2(X_{i2}-\bar{X}_2)\}^2$$

$$+ [(N_1-n_1)(\bar{Y}_2-\bar{Y}_1)]^2$$

ซ.ต.พ.

จากความเอนเอียงอย่างมีเงื่อนไขของ $\hat{Y}_1$, $\hat{Y}_2$ และ $\hat{Y}_3$ จะเห็นว่า

1. $\hat{Y}_2$ จะเป็นตัวประมาณที่ไม่เอนเอียง ถ้า $n_1 = \frac{n}{N} N_1$

2. ถ้า $n_1 = N_1$ แล้ว $\hat{Y}_1$ และ $\hat{Y}_3$ จะเป็นตัวประมาณที่ไม่เอนเอียง และมีค่าเท่ากัน

3. ถ้า $\hat{Y}_1 = \hat{Y}_2$ แล้ว $\hat{Y}_2$ และ $\hat{Y}_3$ จะเป็นตัวประมาณที่ไม่เอนเอียง

สำหรับการประมาณค่าความแปรปรวนของตัวประมาณ $\hat{Y}k$; $k = 1, 2, 3$ จากตัวอย่างไม่สามารถประมาณค่าความแปรปรวนของตัวประมาณ $\hat{Y}_2$ และ $\hat{Y}_3$ ได้เพราะไม่ทราบค่า N_1 ซึ่งในวิทยานิพนธ์จึงไม่ขอกล่าวถึงการประมาณค่า N_1 ด้วยตัวอย่าง และสูตรตัวประมาณค่าความแปรปรวนของ $\hat{Y}_2$ และ $\hat{Y}_3$ แต่จะเสนอเฉพาะตัวประมาณค่าความแปรปรวนอย่างมีเงื่อนไขของ $\hat{Y}_1$ เท่านั้น เนื่องจาก $\hat{Y}_1$ จะใช้ได้เมื่อทราบค่า N_1 โดยจะประมาณ $V(\hat{Y}_1|n_1)$ ด้วย $v(\hat{Y}_1|n_1)$ ดังนี้

$$v(\hat{Y}_1|n_1) = \sum_{j=1}^{2} N_j^2(\frac{1}{n_j} - \frac{1}{N_j}) \cdot \frac{1}{n_j-1} \sum_{i=1}^{n_j} \{(y_{ij}-\bar{y}_j)-b_j(x_{ij}-\bar{x}_j)\}^2$$

ซึ่งจะได้ $v(\hat{y}_1|n_1)$ เป็นตัวประมาณที่ไม่เอนเอียงของ $V(\hat{Y}_1|n_1)$ เพราะ $\frac{1}{n_j-1}$ $\cdot \sum_{i=1}^{n_j} \{(y_{ij}-\bar{y}_j)-b_j(x_{ij}-\bar{x}_j)\}^2$ เป็นตัวประมาณที่ไม่เอนเอียงของ $\frac{1}{N_j-1} \sum_{i=1}^{N_j} \{(Y_{ij}-\bar{Y}_j)$ $-b_j(X_{ij}-\bar{X}_j)\}^2$

ข. กรณีที่ไม่ทราบค่า b_j

สำหรับในกรณีที่ไม่ทราบค่า b_j นี้ จะประมาณ b_j จากตัวอย่างด้วย

$$\hat{b}_j = \frac{\sum_{i=1}^{n_j} (x_{ij}-\bar{x}_j)\cdot(y_{ij}-\bar{y}_j)}{\sum_{i=1}^{n_j} (x_{ij}-\bar{x}_j)^2}$$

ซึ่งจะทำให้ $\bar{y}_{\ell rj}$ เป็นตัวประมาณที่เอนเอียงใน order $\frac{1}{n_j}$ ของ $\bar{Y}_j$ โดยทฤษฎีที่ว่าถ้า $\hat{b}$ เป็นตัวประมาณกำลังสองน้อยที่สุด (least square estimator) ของ B ซึ่ง B มีค่าเท่ากับ $\frac{\sum_{i=1}^{N} (X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{N} (X_i-\bar{X})^2}$ และ $\hat{y}_{\ell r} = \bar{y} - \hat{b}(\bar{x}-\bar{X})$

ในการเลือกตัวอย่างแบบสุ่มอย่างง่ายขนาดตัวอย่าง n เมื่อ n มีขนาดใหญ่ จะได้ $E(\bar{y}_{\ell r}-\bar{Y})^2 \doteq$

$$V(\bar{y}_{\ell r}) = (1-\frac{n}{N})\cdot\frac{1}{N-1}\sum_{i=1}^{N}\{(y_i-\bar{Y})-B(X_i-\bar{X})\}^2 \quad [1]$$

ดังนั้น จากทฤษฎีดังกล่าว จะได้ $E[(\hat{Y}k-Y)|n_1]^2$ ในกรณีไม่ทราบค่า $b_j \doteq$

$MSE(\hat{Y}k|n_1)$ ในกรณีทราบค่า b_j ; k = 1, 2, 3 เมื่อค่า $MSE(\hat{y}_k|n_1)$ มีค่าดังที่ได้กล่าวมาแล้ว และจะประมาณ $V(\hat{Y}_1|n_1)$ ด้วย $\hat{V}(\hat{Y}_1|n_1)$ ดังนี้คือ

---

[1]Cochran, W. G. Sampling techniques, 3 d ed. (John-Wiley & Sons Inc, New York : 1977), P. 194.

$$\hat{V}(\hat{Y}_1|n_1) = \sum_{j=1}^{2} N_j^2\left(\frac{1}{n_j} - \frac{1}{N_j}\right)\cdot\frac{1}{n_j-2}\sum_{i=1}^{n_j}\{(y_{ij}-\bar{y}_j)-\hat{b}_j(x_{ij}-\bar{x}_j)\}^2$$

สำหรับตัวประมาณของ $V(\hat{Y}_2|n_1)$ และ $V(\hat{Y}_3|n_1)$ ในกรณีที่ไม่ทราบค่า b_j จะไม่ยอกล่าวถึง ซึ่งมีเหตุผลดังที่ได้กล่าวมาแล้วในกรณีที่ทราบค่า b_j

3.1.2 การอนุมานอย่างไม่มีเงื่อนไข (Unconditional inference)

ในกรณีที่ไม่สนใจว่าในตัวอย่างที่เลือกมาจะมีจำนวนตัวอย่างที่มีค่าสังเกตของตัวแปร Y เป็นค่าสูงมากอยู่จำนวนเท่าใด เพียงแต่คาดว่าจะต้องพบอย่างน้อยตั้งแต่ 1 ตัวอย่างขึ้นไปที่ค่าสังเกต Y_i เป็นค่าสูงมาก ดังนั้น ค่าความเอนเอียงและค่าความคลาดเคลื่อนกำลังสองจึงเกี่ยวข้องกับค่าโมเมนต์ที่ k ของ n_1 แทนค่า n_1

เนื่องจากเป็นการสุ่มตัวอย่างชนิดไม่ใส่คืนจากประชากรที่ประกอบด้วยคู่ลำดับของค่าสังเกต (X_i, Y_i) ; $i = 1, 2, \ldots, N$ 2 กลุ่มคือ กลุ่มค่าสังเกต Y_i ที่มีค่าสูงมาก และกลุ่มค่าสังเกต Y_i ที่เป็นค่าปกติ ดังนั้น การแจกแจงของ n_1 จึงมีลักษณะแบบ positive Hypergeometric ที่มีฟังก์ชันความหนาแน่นคือ

$$P(n_1|n, N, N_1) = \frac{1}{1-d}\,\frac{\binom{N_1}{n_1}\binom{N_2}{n_2}}{\binom{N}{n}}$$

เมื่อ $n_1 = 1, 2, \ldots, N_1$

$N_2 = N-N_1 \geq n$

$n > N_1 \geq 1$

และ $d = \prod_{k=0}^{N_1-1}\left(1 - \frac{n}{N-k}\right)$

โดยมี $E(n_1) = \frac{n}{(1-d)}\frac{N_1}{N}$ และ $V(n_1) = (\frac{1}{1-d})^2 \frac{n N_1}{N(N-1)N} \cdot [(1-d)N((N-n)+$ $N_1(n-1))-nN_1(N-1)]$

ก. <u>กรณีที่ทราบค่า b_j</u>

<u>ทฤษฎีที่ 4</u> ในการเลือกตัวอย่างแบบสุ่มอย่างชนิดไม่ใส่คืน สำหรับกรณีที่ทราบค่า b_j ซึ่งมีค่าเหมือนในทฤษฎีที่ 1 จะได้ตัวประมาณ $\hat{Y}_1$ เป็นตัวประมาณค่ารวมประชากรที่ไม่เอนเอียง $(B(\hat{Y}_1) = 0)$ และมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างไม่มีเงื่อนไขเท่ากับ $\sum_{j=1}^{2} N_j^2 \left(E(\frac{1}{n_j})-\frac{1}{N_j}\right) \cdot \frac{1}{N_j-1} \sum_{i=1}^{N_j} \{(Y_{ij}-\bar{Y}_j)-b_j(X_{ij}-\bar{X}_j)\}^2$

<u>พิสูจน์</u> 4.1 $B(\hat{Y}_1) = 0$

จาก $B(\hat{Y}_1|n_1) = 0$

$\therefore \; E(B(\hat{Y}_1|n_1)) = E(0) = 0$

ดังนั้น $B(\hat{Y}_1) = E(B(\hat{Y}_1|n_1)) = 0$ ซ.ต.พ.

4.2 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างไม่มีเงื่อนไขของ $\hat{Y}_1$

จาก $MSE(\hat{Y}_1|n_1) = \sum_{j=1}^{2} N_j^2 (\frac{1}{n_j}-\frac{1}{N_j}) \cdot \frac{1}{N_j-1} \sum_{i=1}^{N_j} (Y_{ij}-\bar{Y}_j)- b_j(X_{ij}-\bar{X}_j)\}^2$

$\therefore \; MSE(\hat{Y}_1) = E(MSE(\hat{Y}_1|n_1))$

$$= \sum_{j=1}^{2} N_j^2 \left(E\left(\frac{1}{n_j}\right) - \frac{1}{N_j}\right) . \frac{1}{N_j-1} \sum_{i=1}^{N_j} \{(Y_{ij}-\bar{Y}_j) - b_j(X_{ij}-\bar{X}_j)\}^2 \qquad \text{ซ.ต.พ.}$$

<u>ทฤษฎีที่ 5</u> ในการเลือกตัวอย่างแบบสุ่มอย่างง่ายชนิดไม่ใส่คืนสำหรับกรณีที่ทราบค่า b_j ซึ่งมีค่า b_j เหมือนในทฤษฎีที่ 1 จะได้ตัวประมาณ $\hat{Y}_2$ เป็นตัวประมาณค่ารวมประชากรที่มีความเอนเอียง โดยมีค่าความเอนเอียงอย่างไม่มีเงื่อนไข $(B(\hat{Y}_2))$ เท่ากับ $N_1 . \frac{d}{1-d}(\bar{Y}_1-\bar{Y}_2)$ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างไม่มีเงื่อนไข เท่ากับ

$$N^2 \Big[\sum_{j=1}^{2} . \frac{1}{n^2}\left(E(n_j) - \frac{En_j^2}{N_j}\right) . \frac{1}{N_j-1} \sum_{i=1}^{N_j} \{(Y_{ij}-\bar{Y}_j) - b_j(X_{ij}-\bar{X}_j)\}^2 + \left(\frac{En_1^2}{n^2} - \frac{2N_1\, En_1}{nN} + \frac{N_1^2}{N^2}\right)\bar{Y}_1^2 + 2\left\{\frac{(nEn_1 - En_1^2)}{n^2} + \frac{N_1\, N_2}{N^2} - \frac{(N_1\, n - N_1 En_1)}{nN} - \frac{N_2\, En_1}{Nn}\right\} . \bar{Y}_1\, \bar{Y}_2 + \left\{\frac{(n^2 - 2nEn_1 + En_1^2)}{n^2} - \frac{2(n - En_1)N_2}{nN} + \frac{N_2^2}{N^2}\right\}\bar{Y}_2^2$$

<u>พิสูจน์</u>

5.1 $B(\hat{Y}_2) = N_1 . \frac{d}{1-d}(\bar{Y}_1-\bar{Y}_2)$

จาก $B(\hat{Y}_2|n_1) = N\left[\left(\frac{n_1}{n} - \frac{N_1}{N}\right)\bar{Y}_1 + \left(\frac{n_2}{n} - \frac{N_2}{N}\right)\bar{Y}_2\right]$

$\therefore \quad B(\hat{Y}_2) = E(B(\hat{Y}_2|n_1))$

$$= E\left(N\left[\left(\frac{n_1}{n}-\frac{N_1}{N}\right)\bar{Y}_1 + \left(\frac{n_2}{n}-\frac{N_2}{N}\right)\bar{Y}_2\right]\right)$$

$$= N\left[\left(\frac{En_1}{n} - \frac{N_1}{N}\right)\bar{Y}_1 + \left(\frac{E(n-n_1)}{n} - \frac{N_2}{N}\right)\bar{Y}_2\right] \left[\because n_2 = n-n_1\right]$$

$$= N\left(\frac{n.\,N_1}{(1-d)N.n} - \frac{N_1}{N}\right)\bar{Y}_1 + \left(\frac{n-\frac{nN_1}{N(1-d)}}{n} - \frac{N_2}{N}\right)\bar{Y}_2$$

$$= \left(\frac{N_1}{1-d} - N_1\right)\bar{Y}_1 + \left(\left(N - \frac{N_1}{1-d}\right) - N_2\right)\bar{Y}_2$$

$$= \frac{N_1}{1-d}\bar{Y}_1(1 - (1 - d)) + \frac{\bar{Y}_2}{1-d}\,(N(1-d)-N_1-N_2(1-d))$$

$$= \frac{N_1\bar{Y}_1}{1-d}\cdot d + \frac{\bar{Y}_2}{1-d}\,(N-Nd-N_1-N_2+N_2d)$$

$$= N_1\bar{Y}_1\cdot\frac{d}{1-d} + \frac{\bar{Y}_2}{1-d}\,((-N_1-N_2)d+N_2d)$$

$$= N_1\bar{Y}_1\frac{d}{1-d} + -N_1\frac{d}{1-d}\bar{Y}_2$$

$$= N_1\frac{d}{1-d}(\bar{Y}_1-\bar{Y}_2)$$ ซ.ต.พ.

5.2 ค่าความคลาดเคลื่อนอย่างไม่มีเงื่อนไขของ $\hat{Y}_2$

จาก $$MSE(\hat{Y}_2|n_1) = N^2\left[\frac{1}{n^2}\sum_{j=1}^{2}\frac{(n_j-n_j^2)}{N_j}\cdot\frac{1}{N_j-1}\sum_{i=1}^{N_j}\{(Y_{ij}-\bar{Y}_j)-b_j\cdot(X_{ij}-\bar{X}_j)\}^2 + \left\{\sum_{j=1}^{2}\left(\frac{n_j}{n}-\frac{N_j}{N}\right)\bar{Y}_j\right\}^2\right]$$

$$\therefore \quad MSE(\hat{Y}_2) = E(MSE(\hat{Y}_2|n_1))$$

$$= N^2\Big[\frac{1}{n^2}\sum_{j=1}^{2}\Big(En_j - \frac{En_j^2}{N_j}\Big)\cdot\frac{1}{N_j-1}\sum_{i=1}^{N_j}\{(Y_{ij}-\bar{Y}_j)-b_j\cdot(X_{ij}-\bar{X}_j)\}^2 + E\{(\frac{n_1}{n}-\frac{N_1}{N})\bar{Y}_1 + (\frac{n_2}{n}-\frac{N_2}{N})\bar{Y}_2\}^2\Big]$$

พิจารณาเทอม $E\{(\frac{n_1}{n} - \frac{N_1}{N})\bar{Y}_1 + (\frac{n_2}{n} - \frac{N_2}{N})\bar{Y}_2\}^2$ จะเห็นว่า

$$E\{(\frac{n_1}{n}-\frac{N_1}{N})\bar{Y}_1 + (\frac{n_2}{n}-\frac{N_2}{N})\bar{Y}_2\}^2 = E\{(\frac{n_1}{n}-\frac{N_1}{N})^2\bar{Y}_1^2 + 2(\frac{n_1}{n}-\frac{N_1}{N})(\frac{n_2}{n}-\frac{N_2}{N})\cdot\bar{Y}_1\bar{Y}_2 + (\frac{n_2}{n}-\frac{N_2}{N})^2\,\bar{Y}_2^2\}$$

$$= E\{(\frac{n_1^2}{n^2} - \frac{2n_1N_1}{nN} + \frac{N_1^2}{N^2})\cdot\bar{Y}_1^2 + 2(\frac{n_1n_2}{n^2} + \frac{N_1N_2}{N^2} - \frac{N_1n_2}{Nn} - \frac{N_2n_1}{nN})\cdot\bar{Y}_1\bar{Y}_2 + (\frac{n_2^2}{n^2} - \frac{2n_2N_2}{nN} + \frac{N_2^2}{N^2})\cdot\bar{y}_2^2\}$$

แทนค่า $n_2 = n - n_1$ จะได้

$$= E\Big\{\Big(\frac{n_1^2}{n^2} - \frac{2n_1N_1}{nN} + \frac{N_1^2}{N^2}\Big)\bar{Y}_1^2 + 2\Big[\frac{n_1(n-n_1)}{n^2} + \frac{N_1N_2}{N^2} - N_1\frac{(n-n_1)}{Nn} -$$

$$\frac{N_2n_1}{nN}\Big] \cdot \bar{Y}_1\bar{Y}_2 + \Big[(n^2-2n_1n+n_1^2) - 2(n-n_1)\frac{N_2}{nN} + \frac{N_2^2}{N^2}\Big] \cdot \bar{Y}_2^2\Big\}$$

$$= E\Big\{\Big(\frac{n_1^2}{n^2} - \frac{2n_1N_1}{nN} + \frac{N_1^2}{N^2}\Big)\bar{Y}_1^2 + 2\Big[\frac{n_1n-n_1^2}{n^2} + \frac{N_1N_2}{N^2} - N_1\frac{(n - n_1)}{Nn} - \frac{N_2n_1}{nN}\Big] \cdot$$

$$\bar{Y}_1\bar{Y}_2 + \Big[\frac{(n^2 - 2nn_1 + n_1^2)}{n^2} - 2(n-n_1)\frac{N_2}{nN} + \frac{N_2^2}{N^2}\Big]\bar{Y}_2^2\Big\}$$

$$= \Big(\frac{En_1^2}{n^2} - \frac{2N_1En_1}{nN} + \frac{N_1^2}{N^2}\Big)\bar{Y}_1^2 + 2\Big[\frac{nEn_1 - En_1^2}{n^2} + \frac{N_1N_2}{N^2} - \frac{N_1(n - En_1)}{Nn}$$

$$-\frac{N_2En_1}{nN}\Big]\bar{Y}_2\bar{Y}_1 + \Big[\frac{(n^2 - 2nEn_1 + En_1^2)}{n^2} - 2(n - En_1)\frac{N_2}{nN} + \frac{N_2^2}{N^2}\Big]\bar{Y}_2^2$$

$$\therefore \quad MSE(\hat{Y}_2) = N^2\Big\{\frac{1}{n^2}\sum_{j=1}^{2}\Big(En_j - \frac{En_j^2}{N_j}\Big) \cdot \frac{1}{N_j-1}\sum_{i=1}^{N_j}\{(Y_{ij}-\bar{Y}_j)-b_j(X_{ij}-\bar{X}_j)\}^2$$

$$+ \Big(\frac{En_1^2}{n^2} - \frac{2N_1En_1}{nN} + \frac{N_1^2}{N^2}\Big) \cdot \bar{Y}_1^2 + 2\Big[\frac{nEn_1 - En_1^2}{n^2} + \frac{N_1N_2}{N^2}$$

$$- \frac{N_1(n-En_1)}{Nn} - \frac{N_2En_1}{nN}\Big] \cdot \bar{Y}_2\bar{Y}_1 + \Big\{\frac{(n^2 - 2nEn_1 + En_1^2)}{n^2} -$$

$$2(n - En_1)\frac{N_2}{nN} + \frac{N_2^2}{N^2}\Big\}\bar{Y}_2^2\Big\}$$

ทฤษฎีที่ 6 ในการเลือกตัวอย่างแบบสุ่มอย่างง่ายชนิดไม่ใส่คืน สำหรับกรณีที่ทราบค่า b_j ซึ่งมีค่า b_j เหมือนในทฤษฎีที่ 1 จะได้ตัวประมาณ $\hat{Y}_3$ เป็นตัวประมาณค่ารวมประชากรที่ความเอนเอียง โดยมีความเอนเอียงอย่างไม่มีเงื่อนไข ($B(\hat{Y}_3)$) เท่ากับ $N_1(1 - \frac{n}{N(1-d)})(\bar{Y}_2 - \bar{Y}_1)$ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเท่ากับ

$$(En_1 - \frac{En_1^2}{N_1}) \cdot \frac{1}{N_1-1} \sum_{i=1}^{N_1} \{(Y_{i1} - \bar{Y}_1) - b_1(X_{i1} - \bar{X}_1)\}^2 + \Big[N^2 . E(\frac{1}{n_2}) -$$

$$E\frac{(n_1(2N - n_1))}{n_2} - \frac{(N^2 - 2NEn_1 + En_1^2)}{N_2}\Big] \cdot \frac{1}{N_2-1} \sum_{i=N_1+1}^{N} \{(Y_{i2} - \bar{Y}_2) - b_2 \cdot$$

$$(X_{i2} - \bar{X}_2)\}^2 + (N_1^2 - 2N_1 En_1 + En_1^2)(\bar{Y}_2 - \bar{Y}_1)^2$$

เมื่อ $E(\frac{1}{n_2}) = E(\frac{1}{n-n_1})$, $E(n_2) = n - En_1$ และ $E(n_2^2) = n^2 - 2nEn_1 + En_1^2$

พิสูจน์

6.1 $B(\hat{Y}_3) = N_1(1 - \frac{n}{N(1-d)})(\bar{Y}_2 - \bar{Y}_1)$

จาก $B(\hat{Y}_3|n_1) = (N_1 - n_1)(\bar{Y}_2-\bar{Y}_1)$

$$\therefore B(\hat{Y}_3) = E(B(\hat{Y}_3|n_1))$$

$$= E((N_1 - n_1)(\bar{Y}_2 - \bar{Y}_1)$$

$$= (N_1 - E(n_1))(\bar{Y}_2 - \bar{Y}_1)$$

$$= (N_1 - \frac{nN_1}{N(1-d)})(\bar{Y}_2 - \bar{Y}_1) \quad \left[\because E(n_1) = \frac{nN_1}{N(1-d)}\right]$$

$$= N_1(1 - \frac{n}{N(1-d)})(\bar{Y}_2 - \bar{Y}_1) \qquad \text{ซ.ต.พ.}$$

6.2 ค่าความคลาดเคลื่อนกำลังสองอย่างไม่มีเงื่อนไขของ $\hat{Y}_3$

จาก $$MSE(\hat{Y}_3|n_1) = (n_1 - \frac{n_1^2}{N_1}) \cdot \sum_{i=1}^{N_1} \{(Y_{i1} - \bar{Y}_1) - b_1(X_{i1} - \bar{X}_1)\}^2 + (N - n_1)^2 (\frac{1}{n_2} - \frac{1}{N_2}) \cdot \sum_{i=N_1+1}^{N} \frac{\{(Y_{i2} - \bar{Y}_2) - b_2 (X_{i2} - \bar{X}_2)\}^2}{N_2 - 1} + \left[(N_1 - n_1)(\bar{Y}_2 - \bar{Y}_1)\right]^2$$

$$\therefore MSE(\hat{Y}_3) = E(MSE(\hat{Y}_3|n_1))$$

$$= (En_1 - \frac{En_1^2}{N_1}) \cdot \sum_{i=1}^{N_1} \frac{\{(Y_{i1} - \bar{Y}_1) - b_1 (X_{i1} - \bar{X}_1)\}^2}{N_1 - 1} + E\left[(N - n_1)^2(\frac{1}{n_2} - \frac{1}{N_2}) \sum_{i=N_1+1}^{N} \frac{\{(Y_{i2}-\bar{Y}_2)-b_2(X_{i2} - \bar{X}_2)\}^2}{N_2 - 1}\right] + E\left[(N_1 - n_1)(\bar{Y}_2 - \bar{Y}_1)\right]^2$$

$$= (En_1 - \frac{En_1^2}{N_1}) \cdot \frac{1}{N_1-1} \sum_{i=1}^{N_1} \{(Y_{i1} - \bar{Y}_1) - b_1(X_{i1} - \bar{X}_1)\}^2 +$$

$$E\left[(N^2 - 2n_1N + n_1^2)(\frac{1}{n_2} - \frac{1}{N_2}) \cdot \frac{1}{N_2-1} \sum_{i=N_1+1}^{N} \{(Y_{i2} - \bar{Y}_2) -\right.$$

$$\left. b_2(X_{i2} - \bar{X}_2)\}^2\right] + E\left[(N_1^2 - 2n_1N_1 + n_1^2)(\bar{Y}_2 - \bar{Y}_1)^2\right]$$

$$= (En_1 - \frac{En_1^2}{N_1}) \cdot \frac{1}{N_1-1} \sum_{i=1}^{N_1} \{(Y_{i1} - \bar{Y}_1) - b_1(X_{i1} - \bar{X}_1)\}^2 +$$

$$E\left[\{(\frac{N^2}{n_2} - \frac{n_1}{n_2}(2N - n_1) - \frac{(N^2 - 2n_1N + n_1^2)}{N_2}\} \cdot \frac{1}{N_2-1} \cdot\right.$$

$$\left.\sum_{i=N_1+1}^{N} \{(Y_{i2} - \bar{Y}_2) - b_2(X_{i2} - \bar{X}_2)\}^2\right] + \left[N_1^2 - 2N_1En_1 +\right.$$

$$\left. En_1^2\right] \cdot (\bar{Y}_2 - \bar{Y}_1)^2$$

$$\therefore \text{MSE}(\hat{Y}_3) = (En_1 - \frac{En_1^2}{N_1}) \cdot \frac{1}{N_1-1} \sum_{i=1}^{N_1} \{(Y_{i1} - \bar{Y}_1) - b_i(X_{i1} - \bar{X}_1)\}^2$$

$$\{N^2 \cdot E(\frac{1}{n_2}) - E\left[\frac{n_1(2N - n_1)}{n_2}\right] - \frac{(N^2 - 2NEn_1 + En_1^2)}{N_2}\} \cdot$$

$$\frac{1}{N_2-1} \cdot \sum_{i=N_1+1}^{N} \{(Y_{i2} - \bar{Y}_2) - b_2(X_{i2} - \bar{X}_2)\}^2 + \left[N_1^2 - 2N_1En_1 +\right.$$

$$\left. En_1^2\right] \cdot (\bar{Y}_2 - \bar{Y}_1)^2$$

ช.ต.พ.

จะเห็นได้ว่าในกรณีการอนุมานแบบไม่มีเงื่อนไข ไม่สามารถประมาณค่าความแปรปรวนของตัวประมาณ $\hat{Y}_2$ และ $\hat{Y}_3$ ได้เพราะไม่ทราบค่า N_1 เพราะฉะนั้น จึงไม่เสนอถึงประมาณค่าความแปรปรวนของ $\hat{Y}_2$ และ $\hat{Y}_3$ ในการทำวิจัยครั้งนี้ แต่จะเสนอเฉพาะตัวประมาณค่าความแปรปรวนของ $\hat{Y}_1$ เท่านั้น เนื่องจาก $\hat{Y}_1$ จะใช้ได้เมื่อทราบค่า N_1 โดยประมาณ $V(\hat{Y}_1)$ ด้วย $v(\hat{Y}_1)$ ดังนี้

$$v(\hat{Y}_1) = \sum_{i=1}^{2} N_j^2 \left(E\left(\frac{1}{n_j}\right) - \frac{1}{N_j}\right) \cdot \frac{1}{n_j-1} \sum_{i=1}^{n_j} \{(y_{ij}-\bar{Y}_j) - b_j(x_{ij} - \bar{x}_j)\}^2$$

โดย $v(\hat{y}_1)$ จะเป็นตัวประมาณที่ไม่เอนเอียงสำหรับ $V(\hat{Y}_1)$ ด้วย (เหตุผลเช่นเดียวกับตัวประมาณ $v(\hat{Y}_1|n_1)$ ดังที่ได้กล่าวในกรณีการอนุมานอย่างมีเงื่อนไข)

ข. กรณีที่ไม่ทราบค่า b_j

สำหรับกรณีที่ไม่ทราบค่า b_j จะประมาณ b_j จากตัวอย่างด้วย $\hat{b}_j$ ซึ่งมีค่าเหมือนกันกับในแบบการอนุมานอย่างมีเงื่อนไข และอาศัยคุณสมบัติเช่นเดียวกันจะได้ว่าถ้าไม่อาศัยทฤษฎีความถดถอยและ n มีขนาดใหญ่แล้ว $E(\hat{Y}k - Y)^2$ ในกรณีไม่ทราบค่า $b_j \doteq MSE\ (\hat{Y}k)$ ในกรณีทราบค่า b_j ; k = 1, 2, 3 และประมาณ $V(\hat{Y}_1)$ ด้วย

$$\hat{V}(\hat{Y}_1) = \sum_{j=1}^{2} N_j^2 \left(E\left(\frac{1}{n_j}\right) - \frac{1}{N_j}\right) \cdot \frac{1}{n_j-2} \sum_{i=1}^{n_j} \{(y_{ij} - \bar{y}_j) - \hat{b}_j(x_{ij} - \bar{x}_j)\}^2$$

3.2 <u>การประมาณค่าความน่าจะเป็นแบบ positive Hypergeometric ในกรณีที่ประชากรมีขนาดใหญ่</u>

โดยทั่วไปการแจกแจงแบบไฮเปอร์จีออเมตริก (Hypergeometric) เป็นลักษณะของการเลือกตัวอย่างที่จะพบความสำเร็จตามจำนวนที่ต้องการแบบไม่ใส่คืน ซึ่งถ้าขนาดของประชากร (N) มีขนาดใหญ่เมื่อเทียบกับขนาดตัวอย่างแล้วกรณีการเลือกตัวอย่างแบบไม่ใส่คืนและการเลือกตัวอย่างแบบใส่คืนจะให้ค่าความน่าจะเป็นไม่แตกต่างกันมากนัก จากคุณสมบัติดังกล่าวเพื่อเป็นการแก้ปัญหาในการคำนวณหาค่าความน่าจะเป็นสำหรับแต่ละค่าที่เป็นไปได้ของจำนวนหน่วยตัวอย่าง

ที่มีค่าสูงมากสำหรับกรณีที่ประชากรมีขนาดใหญ่เมื่อเทียบกับขนาดตัวอย่าง เพื่อใช้ในการอนุมานอย่างไม่มีเงื่อนไข จึงได้ใช้การแจกแจงที่มีฟังกชันความหนาแน่น $P(n_1)$ (ซึ่งจะได้กล่าวถึงต่อไป) มาประมาณฟังกชันความหนาแน่น $P(n_1|n, N, N_1)$ โดยที่ฟังกชันความหนาแน่น $P(n_1)$ ได้มาจากคุณสมบัติของการใช้การแจกแจงแบบทวินาม (binomial) มาประมาณการแจกแจงแบบไฮเปอร์จีออเมตริก โดยกำหนดให้ฟังกชันความหนาแน่น $P(n_1)$ มีลักษณะเป็นดังนี้คือ

$$P(n_1) = \frac{1}{1-D} \binom{n}{n_1} P^{n_1} q^{n-n_1} \; ; \; n_1 = 1, 2, \ldots, N_1 \text{ และ } n \geq N_1$$

เมื่อ $P = \frac{N_1}{N}$, $q = 1-P$ และ $D = \sum_{n_1=N_1+1}^{n} \binom{n}{n_1} P^{n_1} q^{n-n_1} + q^n$

สำหรับวิธีการหาฟังกชันความหนาแน่น $P(n_1)$ มีดังนี้คือ

$$\text{จาก} \quad P(n_1|n, N, N_1) = \frac{1}{1-d} \frac{\binom{N_1}{n_1}\binom{N-N_1}{n-n_1}}{\binom{N}{n}}$$

$$= \frac{1}{1-d} \frac{N_1!}{(N_1-n_1)!\; n_1!} \; \frac{x(N-N_1)!}{(n-n_1)!\; (N-N_1-n+n_1)!}$$

$$x \frac{n!\,(N-n)!}{N!}$$

$$= \frac{1}{1-d} \cdot \frac{N_1 \cdot (N_1-1) \ldots (N_1-n_1+1)}{n_1!} \; x$$

$$\frac{(N-N_1)(N-N_1-1) \ldots (N-N_1-n+n_1+1)}{(n-n_1)!} \; x$$

$$\frac{n!}{N(N-1) \ldots (N-n+1)}$$

$$= \frac{1}{1-d} \frac{n!}{n_1! \, .(n-n_1)!} \times \frac{N_1(N_1-1)\ldots(N_1-n_1+1)\times(N-N_1)}{N(N-1)\ldots(N-n+1)}$$

$$\frac{(N-N_1-1)\ldots(N-N_1-n+n_1+1)}{\quad}$$

$$= \frac{1}{1-d} \frac{n!}{n_1!\,(n-n_1)} \frac{N_1}{N} \overbrace{\left(\frac{N_1}{N-1} - \frac{1}{N-1}\right)\ldots\left(\frac{N_1}{N-n_1+1} - \frac{(n_1-1)}{N-n_1+1}\right)}^{\text{มีอยู่ } n_1 \text{ เทอม}}$$

$$\overbrace{\frac{(N-N_1)}{N-n_1} \cdot \left(\frac{N-N_1}{N-n_1-1} - \frac{1}{N-n_1-1}\right) \ldots \left(\frac{N-N_1}{N-(n-1)} - \frac{(n-n_1-1)}{N-(n-1)}\right)}^{\text{มีอยู่ } n-n_1 \text{ เทอม}}$$

กำหนดให้ $P = \frac{N_1}{N}$, $q = 1 - p$ และ $N - 1 \simeq N , \ldots , N - n + 1 \simeq N$

เมื่อ $\frac{n}{N}$ มีค่าน้อย ๆ

$$\therefore \lim_{N\to\infty} P(n_1|n, N, N_1) = \frac{1}{1-d} \frac{n!}{n_1!\,(n-n_1)!} \cdot \overbrace{P.P\ldots P}^{\text{มี } n_1 \text{ ตัว}} \times \overbrace{q.\,q\ldots q}^{\text{มี } n-n_1 \text{ ตัว}}$$

$$= \frac{1}{1-d} \binom{n}{n_1} P^{n_1} \cdot q^{n-n_1} \; ; n_1 = 1, 2, \ldots, N_1$$

แต่เนื่องจาก $\frac{1}{1-d} \binom{n}{n_1} P^{n_1} q^{n-n_1}$; $n_1 = 1, 2, \ldots, N_1$ ยังไม่เป็นฟังกชัน

ความหนาแน่น ดังนั้น ต้องพยายามจัดรูปใหม่

จาก $$\sum_{n_1=o}^{n} \binom{n}{n_1} P^{n_1} q^{n-n_1} = \sum_{n_1=1}^{N_1} \binom{n}{n_1} P^{n_1} q^{n-n_1} + \sum_{n_1=N_1+1}^{n} \binom{n}{n_1} P^{n_1} \cdot q^{n-n_1} + \binom{n}{o} P^o q^{n-o}$$

$$= 1$$

$$\therefore \sum_{n_1=1}^{N_1} \binom{n}{n_1} p^{n_1} q^{n-n_1} = 1 - \sum_{n_1=N_1+1}^{n} \binom{n}{n_1} p^{n_1} q^{n-n_1} - q^n$$

และจะได้

$$\frac{1}{1-d} \sum_{n_1=1}^{N} \binom{n}{n_1} p^{n_1} q^{n-n_1} = \frac{1}{1-d} \times \left(1 - \sum_{n_1=N_1+1}^{n} \binom{n}{n_1} p^{n_1} q^{n-n_1} - q^n\right)$$

$$\therefore \frac{\frac{1}{1-d} \sum_{n_1=1}^{N} \binom{n}{n_1} p^{n_1} q^{n-n_1}}{\frac{1}{1-d} \left(1 - \sum_{n_1=N_1+1}^{n} \binom{n}{n_1} p^{n_1} q^{n-n_1} - q^n\right)} = 1$$

ดังนั้น จึงกำหนดให้ $P(n_1) = \frac{1}{1-D} \binom{n}{n_1} p^{n_1} q^{(n-n_1)}$; $n_1 = 1, 2, \ldots, N_1$

$n \geq N_1$

และเมื่อ $D = \sum_{n_1=N_1+1}^{n} \binom{n}{n_1} p^{n_1} q^{n-n_1} + q^n$

ซึ่งจะเห็นได้ว่า $P(n_1)$ เป็นฟังกชันความหนาแน่น เพราะ

1. $P(n_1) > 0$ สำหรับทุก ๆ ค่าของ n_1 เมื่อ $n_1 = 1, 2, \ldots, N_1$
2. $\sum_{n_1=1}^{N_1} P(n_1) = \sum_{n_1=1}^{N_1} \cdot \frac{1}{1-D} \binom{n}{n_1} p^{n_1} q^{n-n_1}$

$$= \frac{1}{1-D} \sum_{n_1=1}^{N_1} \binom{n}{n_1} p^{n_1} q^{n-n_1}$$

$$= \frac{1 - D}{1 - D}$$

$$= 1$$

$$\therefore \lim_{N\to\infty} P(n_1|n, N, N_1) \simeq P(n_1)$$