บทที่ 2

ทฤษฎีและตัวสถิติที่ใช้ในการวิจัย

2.1 การประมาณค่ายอดรวมประชากรเมื่อกลุ่มถูกเลือกด้วยความน่าจะเป็นไม่เท่ากันแบบไม่ใส่คืน

เมื่อกำหนดให้ y_i คือค่ายอดรวมของกลุ่ม i p_i เป็นความน่าจะเป็นที่กลุ่ม i จะถูกเลือก และ π_i คือความน่าจะเป็นที่กลุ่ม i อยู่ในตัวอย่างขนาด n กลุ่มด้วยความน่าจะเป็นแบบไม่เท่ากันแบบไม่ใส่คืนแล้ว ตัวประมาณค่ายอดรวมประชากรของ ของฮอร์วิทซ-ทอมป์สัน คือ

$$\hat{Y}_{HTWOR} = \sum_{i=1}^{n} y_i / \pi_i$$

ตัวประมาณ $\hat{Y}_{HTWOR}$ จะเป็นตัวประมาณไม่เอนเอียงของ Y และมีค่าความแปรปรวน

$$V(\hat{Y}_{HTWOR}) = \sum_{i=1}^{N} \sum_{j>i} (\pi_i \pi_j - \pi_{ij})(y_i/\pi_i - y_j/\pi_j)^2$$

เมื่อ π_{ij} เป็นความน่าจะเป็นที่กลุ่ม i และกลุ่ม j อยู่ในตัวอย่างขนาด n

จากตัวประมาณค่ายอดรวมประชากรของฮอร์วิทซ-ทอมป์สัน รวมทั้งสูตรความแปรปรวนของตัวประมาณ มีความยุ่งยากในการหาค่า π_i และ π_{ij} จากการเลือกหน่วยประชากรด้วยความน่าจะเป็นไม่เท่ากันแบบไม่ใส่คืน ทำให้ π_i และ π_{ij} ไม่เป็นสัดส่วนโดยตรงกับ p_i ซึ่งความน่าจะเป็นที่กลุ่ม i จะถูกเลือกในการเลือกครั้งที่ j จะขึ้นอยู่กับผลการเลือกครั้งที่ผ่านมาว่ากลุ่มใดถูกเลือกไปแล้วบ้าง เพื่อให้การประมาณค่าเป็นไปได้ในทางปฏิบัติ ได้มีผู้คิดค้นวิธีการต่างๆหลายวิธี โดยส่วนใหญ่จะดำเนินการโดยการเปลี่ยนรูปตัวประมาณ แต่คงวิธีการเลือกตัวอย่างไว้ หรือปรับวิธีการเลือกตัวอย่าง แต่คงรูปตัวประมาณไว้เช่น แผนการเลือกตัวอย่างของ Vasantha kumar ,E. , Srivenkataramana ,T. และ Srinath ,K.P. และ แผนการเลือกตัวอย่างของ Tommy Wright ดังต่อไปนี้

2.1.1 แผนการเลือกตัวอย่างของ Vasantha kumar ,E. , Srivenkatamana ,T. และ Srinath ,K.P. (1996)

แผนการเลือกตัวอย่างด้วยความน่าจะเป็นไม่เท่ากันแบบไม่ใส่คืน กับขนาดตัวอย่างที่มากกว่า 2 ค่อนข้างซับซ้อนสำหรับการหาความน่าจะเป็นของเลือกตัวอย่างแรกและการเลือกครั้งที่ 2 ดังนั้น ภายใต้ข้อสมมติซึ่งกลุ่ม U_i มีอันดับที่ i จากการเรียงลำดับจากน้อยไปหามากของขนาดของตัวแปร X จะทำการสุ่ม 2 ครั้งแรกด้วยความน่าจะเป็นที่เป็นสัดส่วนกับขนาด โดย ความน่าจะเป็นของการเลือกกลุ่ม U_i ในการสุ่มครั้งแรก คือ p_i โดย

$$p_i = \frac{2i}{(N+1)N} \quad \text{,กรณีใช้ค่าอันดับเป็นค่าวัดขนาด}$$

$$p_i = \frac{x_i}{\sum_{i=1}^{N} x_i} \quad \text{,กรณีใช้ค่าของตัวแปร x เป็นค่าวัดขนาด} \qquad (2.1)$$

การสุ่มครั้งที่ 2 จะจัดอันดับประชากร N-1 กลุ่มที่เหลือ จาก 1 ถึง N-1 จากน้อยไปหามากตามค่าของ X สังเกตุว่า อันดับของกลุ่ม U_{i+1} , U_{i+2} , . . . , U_N จะกลายเป็น i , i +1, . . . , (N-1) ในขณะที่อันดับก่อนหน้ากลุ่ม i ที่ถูกเลือกจะมีอันดับเป็น 1,2,. . .,(i – 1) แล้วเลือก 1กลุ่มจากกลุ่มเหล่านั้นด้วยความน่าจะเป็นที่เป็นสัดส่วนกับอันดับที่จัดใหม่

ความน่าจะเป็นแบบมีเงื่อนไขของกลุ่ม U_j , $j \neq i$ ที่ถูกเลือกในการสุ่มครั้งที่ 2 คือ

$$p_j^* = \frac{2j}{N(N-1)} \quad , j < i$$

$$p_j^* = \frac{2(j-0.5)}{N(N-1)} \quad , \text{j อยู่ในกลุ่มที่มีค่าซ้ำกัน}$$

$$p_j^* = \frac{2(j-1)}{N(N-1)} \quad , j > i$$

$j = 1, 2, \ldots, N; \quad j \neq i = 1, 2, \ldots, N$

และ

$$p_j^* = \frac{x_j}{\sum_{j=1}^{N-1} x_j} \quad \text{,กรณีใช้ค่าของตัวแปร x เป็นค่าวัดขนาด} \qquad (2.2)$$

และการสุ่ม n-2 ครั้งต่อๆไป จะกระทำแบบใส่คืน ด้วยความน่าจะเป็นที่เป็นสัดส่วนกับขนาดของ N-2 กลุ่มที่เหลือหลังจากการสุ่ม 2 ครั้งแรก

ความน่าจะเป็นแบบมีเงื่อนไขที่กลุ่ม r จะถูกเลือกในการสุ่มครั้งที่สามหรือครั้งต่อๆไป เป็น

$$p_r^{\bullet} = \frac{2r}{(N-2)(N-1)}$$, r < i และ r < j

$$p_r^{\bullet} = \frac{2(r-0.5)}{(N-2)(N-1)}$$, r , j อยู่ในกลุ่มที่มีค่าซ้ำกัน และ j < i หรือ

r , i อยู่ในกลุ่มที่มีค่าซ้ำกัน และ j > i

$$p_r^{\bullet} = \frac{2(r-1)}{(N-2)(N-1)}$$, i < r < j หรือ j < r < i หรือ

I, j, r อยู่ในกลุ่มที่มีค่าซ้ำกัน

$$p_r^{\bullet} = \frac{2(r-1.5)}{(N-2)(N-1)}$$, r , j อยู่ในกลุ่มที่มีค่าซ้ำกัน และ j > i หรือ

r , i อยู่ในกลุ่มที่มีค่าซ้ำกัน และ j < i

$$p_r^{\bullet} = \frac{2(r-2)}{(N-2)(N-1)}$$, r > j และ r > I (2.3)

และ $$p_r^{\bullet} = \frac{x_r}{\sum_{r=1}^{N-2} x_r}$$,กรณีใช้ค่าของตัวแปร x เป็นค่าวัดขนาด

เมื่อ r = 1 2, . . . , N ; r $\neq$ i $\neq$ j = 1, 2, . . . ,N

ตัวประมาณไม่เอนเอียงของ Y สำหรับการสุ่มแบบไม่คืนที่และคืนที่ ตามลำดับดังนี้

$$t_3 = \frac{y_i}{\pi_i} + \frac{y_j}{\pi_j}$$

$$t_4 = (y_i + y_j) + \frac{1}{n-2} \sum_{r=1}^{n-2} \left(\frac{y_r}{p_r^{\bullet}} \right)$$

(2.4)

ตัวประมาณ t_3 และ t_4 ไม่สัมพันธ์กัน ในขณะที่ t_4 ไม่เอนเอียงอย่างมีเงื่อนไขสำหรับการสุ่มครั้งแรกในการสุ่ม 2 ครั้ง จะได้ว่าตัวประมาณของ Y เป็นผลบวกเชิงเส้นของ t_3 และ t_4 โดย

$$t_5 = \delta t_3 + (1-\delta) t_4 \quad , 0 < \delta < 1 , \tag{2.5}$$

ด้วย $$V(t_5) = \delta^2 V(t_3) + (1-\delta)^2 V(t_4) . \tag{2.6}$$

เมื่อ $$V(t_3) = \sum_{i=1}^{N} \sum_{j>i} (\pi_i \pi_j - \pi_{ij}) (y_i/\pi_i - y_j/\pi_j)^2 \tag{2.7}$$

$$\begin{aligned} V(t_4) &= E_1 V_2(t_4) + V_1 E_2(t_4) \\ &= E_1 V_2(t_4) \end{aligned} \tag{2.8}$$

เมื่อ $V_1E_2(t_4)$ เป็น 0 เนื่องจาก $E_2(t_4) = Y$

ให้ i และ j เป็นกลุ่มที่ถูกเลือกในการสุ่ม 2 ครั้งแรก และ j > i

$$E_1V_2(t_4) = E_1[\sum_{r=1}^{N-2} y_r^2 / p_r - Y^2]/(n-2)$$

$$= \frac{(N-1)(N-2)}{2(n-2)} E_1[(\sum_{r=1}^{i-1} y_r^2 / r) + (\sum_{r=i+1}^{j-1} y_r^2 /(r-1)) + (\sum_{r=j+1}^{N} y_r^2 /(r-2))$$

$$- E_1(Y^2)/(n-2) \tag{2.9}$$

$$E_1(Y^2) = \sum_{i=1}^{N} \sum_{j>i} (Y - y_i - y_j)^2 \pi_{ij}$$

$$= Y^2 - 2Y \sum_{i=1}^{N} y_i \pi_i + \sum_{i=1}^{N} y_i^2 \pi_i + \sum_{i}^{N} \sum_{\neq j}^{N} y_i y_j \pi_{ij}$$

$$= Y^2 - 2Y \sum_{i=1}^{N} y_i \pi_i + (\sum_{i=1}^{N} y_i \pi_i)^2 + \sum_{i=1}^{N} \sum_{j>i} (y_i - y_j)^2 (\pi_i \pi_j - \pi_{ij})$$

$$= (Y - \sum_{i=1}^{N} y_i \pi_i)^2 + \sum_{i=1}^{N} \sum_{j>i} (y_i - y_j)^2 (\pi_i \pi_j - \pi_{ij}) \tag{2.10}$$

เมื่อ

$$\sum_{i=1}^{N} \sum_{j>i} (y_i - y_j)^2 (\pi_i \pi_j - \pi_{ij}) = \sum_{i=1}^{N} y_i^2 \pi_i + \sum_{i=1}^{N} y_i y_j \pi_{ij} - (\sum_{i=1}^{N} y_i \pi_i)^2$$

$$E_1(\sum_{r=1}^{i-1} y_r^2 / r) = \sum_{i=1}^{N} (\sum_{r=1}^{i-1} y_r^2 / r) \pi_i$$

$$= \sum_{i=1}^{N} y_i^2 (\sum_{r=i+1}^{N} \pi_r)/i \tag{2.11}$$

$$E_1(\sum_{r=j+1}^{N} y_r^2 /(r-2)) = \sum_{j=3}^{N} y_j^2 (\sum_{r=2}^{j-1} \pi_r)/(j-2) \tag{2.12}$$

$$E_1(\sum_{r=i+1}^{j-1} y_r^2 /(r-1)) = \sum_{i=1}^{N} \sum_{j>i} (\sum_{r=i+1}^{j-1} y_r^2 /(r-1)) \pi_{ij}$$

$$= \sum_{i=2}^{N-1} y_i^2 (\sum_{r=1}^{i-1} \sum_{\alpha=i+1}^{N} \pi_{\alpha r})/(i-1) \tag{2.13}$$

แทนค่า (2.10) - (2.13) ลงใน (2.9) จะได้ว่า

$$V(t_4)=\frac{(N-1)(N-2)}{2(n-2)}[\sum_{i=1}^{N-1}(\sum_{r=i+1}^{N}\pi_r)y_i^2/i+\sum_{i=2}^{N-1}(\sum_{r=1}^{i-1}\sum_{\alpha=i+1}^{N}\pi_{r\alpha})y_i^2/(i-1)+$$

$$\sum_{i=3}^{N}(\sum_{r=2}^{i-1}\pi_r)y_i^2/(i-2)]-$$

$$[\Sigma'(\pi_i\pi_j-\pi_{ij})(y_i-y_j)^2+(Y-\sum_{i=1}^{N}y_i\pi_i)^2]/(n-2) \qquad (2.14)$$

ค่าของ $V(t_4)$ นี้เป็นกรณีที่ค่าอันดับไม่มีค่าซ้ำกัน แต่ถ้าค่าของตัวแปร x มีค่าซ้ำกันค่าอันดับจะเท่ากับค่าเฉลี่ยของกลุ่มค่าอันดับเหล่านั้น ตามเงื่อนไขใน (2.3)

ค่า δ ที่ทำให้ความแปรปรวนมีค่าต่ำสุด คือ

$$\delta_{opt} = V(t_4)\,/[\,V(t_3)+V(t_4)\,] \qquad (2.15)$$

ด้วยค่าที่เหมาะสมของ δ สามารถประมาณ $V(t_\delta)$ ด้วยการประมาณ $V(t_3)$ และ $V(t_4)$ แยกจากกัน ตัวอย่างเช่นการใช้ตัวประมาณของ Y-G สำหรับ $V(t_3)$

$$v(t_3)=\sum_{i=1}^{n}\sum_{j>i}(\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}})(y_i/\pi_i-y_j/\pi_j)^2$$

ส่วน ตัวประมาณไม่เอนเอียงอย่างง่ายของ $V(t_4)$ หาได้โดย ตัวประมาณของ $V_2(t_4)$ เมื่อ $V(t_4) = E_1(V_2(t_4))$ กำหนดให้ตัวแปรจากการสุ่ม 2 ครั้งแรกคือ $u_r = (y_r/p_r^*)$ มีความสัมพันธ์กับการสุ่ม (n-2) ครั้งสุดท้ายซึ่ง u_r แจกแจงเหมือนกันและเป็นอิสระต่อกัน เมื่อสุ่มแบบคืนที่ ตัวประมาณไม่เอนเอียงของ $V_2(t_4)$ เท่ากับ $\sum_{r=1}^{n-2}(u_r-\bar{u})^2/(n-2)(n-3)$

เมื่อ $\bar{u}=\sum_{r=1}^{n-2}u_r/(n-2)$ (Raj 1968 , p.38)

2.1.2 แผนการเลือกตัวอย่างของ Wright (1990)

เป็นแผนการเลือกตัวอย่างแบบ pps แบบไม่ใส่คืน โดยแบ่งประชากรขนาด N ออกเป็น n ชั้นภูมิเช่นชั้นภูมิที่ 1 ประกอบด้วย k หน่วยแรก(หน่วยในที่นี้คือกลุ่ม) และ k หน่วยสุดท้ายในอันดับของประชากร ชั้นภูมิที่ 2 ถูกจัดเช่นเดียวกันหลังจากตัดหน่วยที่ถูกจัดในชั้นภูมิที่ 1 แล้ว วิธีนี้จะทำให้แต่ละชั้นภูมิมี 2k หน่วย มีผลรวมของอันดับเท่ากัน คือ N(N+1) / 2n แล้วเลือกตัวอย่างชั้นภูมิละ 1 หน่วยด้วยความน่าจะเป็นที่เป็นสัดส่วนกับขนาด เมื่อใช้ค่าอันดับเป็นค่าวัดขนาด หน่วย i ในชั้นภูมิที่ h จะถูกเลือกด้วยความน่าจะเป็น np_{ih} คือ

$$p_{ih} = \frac{2i_h}{(N+1)N} \quad \text{สำหรับ } h = 1, 2, \ldots, n \qquad (2.16)$$

ผลรวมของความน่าจะเป็นที่หน่วยในชั้นภูมิที่ h เป็น

$$\sum_{i=1}^{N_h} np_{ih} = 1$$

เนื่องจากความน่าจะเป็นที่หน่วย i จะถูกเลือก คือ $\pi_i = np_i$

ถ้า π_{ij} เป็นความน่าจะเป็นที่หน่วย i และ j ถูกเลือกเป็นตัวอย่างอย่างเป็นอิสระกัน , $i \neq j$ และถูกเลือกจากต่างชั้นภูมิกัน แล้ว $\pi_{ij} = \pi_i\pi_j$, ส่วนในกรณีอื่นๆ $\pi_{ij} = 0$

ทำให้สามารถหาตัวประมาณไม่เอนเอียงสำหรับ Y ของ Horvitz - Thompson ได้ดังนี้

$$\hat{Y}_{HTWOR} = \sum_{h=1}^{n} (y_h / \pi_h) \qquad (2.17)$$

และมีค่าความแปรปรวนของตัวประมาณ เท่ากับ

$$V(\hat{Y}_{HTWOR}) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} - \left[\frac{N}{n}\right]^2 \sum_{h=1}^{n} \mu_{hy}^2 \qquad (2.18)$$

เมื่อ μ_{hy}^2 เป็นค่าเฉลี่ยของประชากรในชั้นภูมิที่ h

กรณีการใช้ค่าสะสมของค่ารากที่สองของความถี่ กำหนดขอบเขตของชั้นภูมิ ทำให้แต่ละชั้นภูมิมีจำนวนหน่วยตัวอย่างไม่เท่ากันจะ สามารถหาค่าความแปรปรวนจองตัวประมาณค่ารวมประชากรได้จาก

$$V(\hat{y}_{HTWOR}) = \sum_{hi=1}^{n} \frac{1}{n_h} \sum_{i=1}^{N_h} \left(\frac{y_{hi}}{\pi_{hi}} - Y_h\right)^2 \pi_{hi}$$

(2.19)

เมื่อ

n_h เป็นจำนวนหน่วยตัวอย่างที่ถูกเลือกในชั้นภูมิที่ h ซึ่งแผนงานนี้จะเลือกชั้นภูมิละ 1 ตัวอย่าง
N_h เป็นจำนวนหน่วยตัวอย่างในชั้นภูมิที่ h
Y_h เป็นค่ารวมประชากรในชั้นภูมิที่ h

ตัวประมาณไม่เอนเอียงของค่าความแปรปรวนของ ตัวประมาณค่ายอดรวมประชากรของฮอร์วิทซ์-ทอมป์สัน (1952) , Sen (1953) และ Yates และ Grundy (1953) นิยมใช้กันอย่างกว้างขวาง แต่ทุกๆค่าของ π_{ij} ต้องไม่เท่ากับศูนย์ ในที่นี้ซึ่งมีบางกรณีที่ π_{ij} มีค่าเป็นศูนย์ จึงไม่สามารถใช้ตัวประมาณเหล่านั้นได้ แต่จะพิจารณาการประมาณค่าความแปรปรวนของตัวประมาณค่ายอดรวมประชากรภายใต้การเลือกตัวอย่าง pps แบบไม่ใส่คืน $V(\hat{Y}_{HTWOR})$ ด้วยตัวประมาณค่าความแปรปรวนของตัวประมาณค่ายอดรวมประชากรภายใต้การเลือกตัวอย่างpps แบบใส่คืน $\hat{V}(\hat{Y}_{HTWR})$

$$\hat{V}(\hat{Y}_{HTWOR}) = \hat{V}(\hat{Y}_{HTWR}) = \frac{1}{n(n-1)} \sum_{j=1}^{n} \left(\frac{y_j}{p_j} - \hat{Y}_{HTWR}\right)^2 \qquad (2.20)$$

และถ้าประมาณ $V(\hat{Y}_{HTWOR})$ ด้วย $\hat{V}(\hat{Y}_{HTWR})$ แล้ว

$$Bias\,(\hat{V}(\hat{Y}_{HTWR})) = \frac{N^2}{n(n-1)} \sum_{j=1}^{n} (\mu_{jy} - \mu_y)^2 \qquad (2.21)$$

ค่าความเอนเอียงที่ไม่เป็นลบและมีค่าต่ำสุดของตัวประมาณของ $V(\hat{Y}_{HTWOR})$ พิจารณาจาก

$\hat{V}(\hat{Y}_{HTWR}) = \alpha \hat{V}(\hat{Y}_{HTWOR})$ เมื่อ $(n-1)/(n\phi-1)<\alpha<1$

และ

$\phi = V(\hat{Y}_{HTWR}) / V(\hat{Y}_{HTWOR})$