บทที่ 2

ผลงานวิจัยที่เกี่ยวข้องและสถิติที่ใช้ในการวิจัย

2.1 ผลงานวิจัยที่เกี่ยวข้อง

กู๊ดแมน (Goodman (1949)) ได้เสนอให้ T' เป็นตัวประมาณค่าที่ใช้สำหรับประมาณจำนวนกลุ่มย่อย K กลุ่มของประชากรขนาด N โดยศึกษาจากตัวอย่างขนาด n ที่เลือกโดยไม่ใส่คืน และพบว่าตัวประมาณค่า T' ประมาณค่า K ได้เหมาะสมที่สุด เพราะมีค่าความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด

ซึ่ง
$$T' = S' = N - \frac{N(N-1)}{n(n-1)} x_2 \quad , \text{ ถ้า } S' \geq \sum_{i=1}^{n} x_i$$
$$= \sum_{i=1}^{n} x_i \quad , \text{ ถ้า } S' < \sum_{i=1}^{n} x_i$$

นอกจากนี้ได้นำตัวประมาณค่า T' ไปใช้กับข้อมูลที่เลือกแบบ binomial sampling ซึ่ง T' ประมาณค่า K ได้ใกล้เคียงมากเมื่อ $1-(2/q)^{\frac{1}{q}-2} < P < 1/3$, $q = \max(n_s)$

ฮอร์วิทซ์และทอมป์สัน (Horvitz and Thompson (1952)) ได้เสนอให้ $\hat{T}$ เป็นตัวประมาณค่าที่ไม่เอนเอียงของยอดรวมประชากร เมื่อเลือกตัวอย่างด้วยวิธีสุ่มอย่างง่ายโดยไม่ใส่คืน ด้วยความน่าจะเป็นในการเลือกตัวอย่างไม่เท่ากัน

ซึ่ง
$$\hat{T} = \sum_{i=1}^{N} \frac{y_i \alpha_i}{P(y_i)}$$

โดยที่ y_i เป็นค่าของสิ่งที่ต้องการศึกษา , $i = 1,2,\ldots,N$

$P(y_i)$ เป็นความน่าจะเป็นที่จะเลือกได้ y_i เป็นตัวอย่าง

กำหนดให้
$$\alpha_i = 1 \quad , y_i \text{ อยู่ในตัวอย่าง}$$
$$= 0 \quad , \text{ อื่นๆ}$$

และ
$$V(\hat{T}) = \sum_{i=1}^{N} \frac{y_i^2(1-P(y_i))}{P(y_i)} + \sum_{i=1}^{N} y_i y_j \frac{P(y_i y_j)-P(y_i)P(y_j)}{P(y_i)\,P(y_j)}$$

นอกจากนี้ฮอร์วิทซ์และทอมป์สัน (1952) ได้เสนอวิธีการเลือกตัวอย่าง ที่ทำให้ความแปรปรวนของตัวประมาณค่ายอดรวมประชากรลดลง ซึ่งมีอยู่ 2 วิธี คือ

วิธีที่หนึ่งเป็นการเลือกตัวอย่างขนาด n โดยไม่ใส่คืน จากประชากรขนาด N ด้วยความน่าจะเป็นในการเลือกตัวอย่างแต่ละครั้งไม่เท่ากัน โดยที่ $y_i (i = 1,...,N)$ จะถูกเลือกเป็นตัวอย่างครั้งแรกด้วยความน่าจะเป็น p_{i_1} และเลือกตัวอย่างครั้งต่อไปด้วยความน่าจะเป็นที่ทุกหน่วยที่เหลือจากการเลือกครั้งแรกจะถูกเลือกมาเป็นตัวอย่างเท่ากัน

กำหนดให้ α_i มีลักษณะเช่นเดียวกับ ----(1)

$$E(\alpha_i) = p_{i_1} + p_{i_2} + p_{i_3} + \dots + p_{i_n}$$

$P(y_i)$ = p_i + ความน่าจะเป็นที่ครั้งแรกไม่เลือก y_i และเลือก y_i ในครั้งอื่น ๆ ที่เหลือ n-1 ครั้ง

$$= p_i + (1-p_i)\frac{(n-1)}{N-1}$$

$$= \frac{(N-n)p_i}{N-1} + \frac{(n-1)}{N-1}$$

$E(\alpha_i\alpha_j)$ = ความน่าจะเป็นที่ y_i และ y_j อยู่ในตัวอย่างขนาด n

$P(y_iy_j)$ = ความน่าจะเป็นที่ครั้งแรกเลือก y_i และเลือก y_j ในครั้งอื่นๆที่เหลือ n-1 ครั้ง + ความน่าจะเป็นที่ครั้งแรกเลือก y_j และเลือก y_i ในครั้งอื่นๆ ที่เหลือ n-1 ครั้ง + ความน่าจะเป็นที่ครั้งแรกไม่เลือก y_i, y_j แต่เลือก y_i, y_j ในครั้งอื่นๆ ที่เหลืออีก n-1 ครั้ง

$$= p_i\frac{(n-1)}{N-1} + p_j\frac{(n-1)}{N-1} - (1-p_i-p_j)\frac{(n-1)}{N-1}\frac{(n-2)}{N-2}$$

$$= \frac{(n-1)}{N-1}\left[\frac{(N-n)}{N-2}(p_i+p_j)\right] + \frac{(n-2)}{N-2}$$

วิธีที่สองเป็นการเลือกตัวอย่างขนาด 2 โดยไม่ใส่คืน จากประชากรขนาด N ซึ่ง y_i (i = 1,2,...,N) จะถูกเลือกเป็นตัวอย่างครั้งแรกด้วยความน่าจะเป็น p_{i_1} และเลือกตัวอย่างครั้งที่สองด้วยความน่าจะเป็นที่เป็นปฏิภาคกับความน่าจะเป็นที่หน่วยที่ i ถูกเลือกมาเป็นตัวอย่างในครั้งแรก

กำหนดให้ α_i มีลักษณะเช่นเดียวกับ ---- (1)

$E(\alpha_i)$ = ความน่าจะเป็นที่ y_i อยู่ในตัวอย่างขนาด 2

$$P(y_i) = p_{i_1} + p_{i_2}$$

$$= p_i + p_j \sum_{i \neq j}^{N} \frac{p_i}{1-p_j}$$

$E(\alpha_i \alpha_j)$ = ความน่าจะเป็นที่ y_i และ y_j อยู่ในตัวอย่างขนาด 2

$P(y_i y_j)$ = ความน่าจะเป็นที่ครั้งแรกเลือกได้ y_i และครั้งที่สองเลือกได้ y_j + ความน่าจะเป็นที่ครั้งแรกเลือกได้ y_j และครั้งที่สองเลือกได้ y_i

$$= p_{i_1} p_{j_2/i} + p_{j_1} p_{i_2/j}$$

$$= p_i p_j \left[\frac{1}{1-p_i} + \frac{1}{1-p_j} \right]$$

เยสท์และกรันดี้ (Yates and Grundy (1953)) ได้แสดงว่าในการเลือกตัวอย่างแบบแบ่งพวกโดยไม่ใส่คืน ด้วยความน่าจะเป็นที่แต่ละหน่วยในพวกเดียวกันถูกเลือกมาเป็นตัวอย่างเป็นปฏิภาคกับขนาดของข้อมูล จะมีผลทำให้ตัวประมาณค่ายอดรวมประชากรมีคุณสมบัติเอนเอียง ดังนั้นจึงเสนอวิธีการเลือกตัวอย่างโดยใช้วิธีที่สองของฮอร์วิทซ์และทอมป์สัน ซึ่งพบว่าตัวประมาณค่ายอดรวมประชากรมีคุณสมบัติไม่เอนเอียง พร้อมกันนี้ได้เสนอแนะว่าการหาความแปรปรวนจากตัวอย่างขนาด 2 ที่เลือกแบบแบ่งพวก ควรใช้วิธีการวิเคราะห์ความแปรปรวน เพราะถ้าแต่ละคู่ซึ่งเลือกมาเป็นตัวอย่างมีสหสัมพันธ์กัน จะทำให้ความแปรปรวนมีค่าเพิ่มขึ้น

นอกจากนี้เยสท์และกรันดี้ (1953) ได้ปรับปรุงวิธีหาความแปรปรวนของตัวประมาณค่ายอดรวมประชากรที่ฮอร์วิทซ์และทอมป์สันเสนอไว้ เพื่อให้มีประสิทธิภาพเพิ่มขึ้น คือ

$$V(T) = \sum_{i<j}^{N}\sum^{N} \pi_i \pi_j - \pi_{ij} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2$$

โดยที่ π_i เป็นผลรวมทั้งหมดของความน่าจะเป็นที่ y_i อยู่ในตัวอย่าง

π_{ij} เป็นผลรวมทั้งหมดของความน่าจะเป็นที่ y_i, y_j อยู่ในตัวอย่าง

และ $\pi_{ij} < \pi_i \pi_j$

โกแดมเบ (Godambe(1955)) ได้แสดงว่าตัวประมาณค่าที่ไม่เอนเอียง และมีความคลาดเคลื่อนกำลังสองเฉลี่ยน้อยที่สุด ของยอดรวมประชากรไม่ได้มีเพียงตัวเดียว

ซิงค์และนเรนทร์ (Singh and Narain (1989)) เสนอให้ T_2 เป็นตัวประมาณค่าที่ไม่เอนเอียงของยอดรวมประชากร เมื่อตัวอย่างที่นำไปศึกษามีขนาดเป็นตัวแปรสุ่ม นอกจากนี้ได้ทำการศึกษาประสิทธิภาพระหว่างตัวประมาณค่ายอดรวมประชากร T_1 เทียบกับตัวประมาณค่ายอดรวมประชากร T_2 ซึ่งใช้ประสิทธิภาพสัมพัทธ์เป็นเกณฑ์ โดยทดสอบกับประชากรที่เป็น bivariate normal มี Y เป็นตัวแปรสุ่มแทนลักษณะของสิ่งที่ต้องการศึกษา ความน่าจะเป็นที่จะได้ข้อมูลโดยเฉลี่ยมีค่าระหว่าง 0.5 - 0.8 สัมประสิทธิสหสัมพันธ์ระหว่างลักษณะของสิ่งที่ต้องการศึกษากับความน่าจะเป็นที่จะได้ข้อมูลมีค่าอยู่ระหว่าง(-0.9) - 0.9 และขนาดตัวอย่างเท่ากับขนาดประชากรซึ่งมีขนาดต่างๆกันคือ 4 , 6 , 8 , 10 โดยทำซ้ำ 5 ครั้งในแต่ละสถานการณ์ พบว่าตัวประมาณค่ายอดรวมประชากร T_2 มีประสิทธิภาพสูงกว่าตัวประมาณค่ายอดรวมประชากร T_1 เพราะความแปรปรวนของตัวประมาณค่ายอดรวมประชากร T_1 มีค่ามากกว่าความแปรปรวนของตัวประมาณค่ายอดรวมประชากร T_2 ดังตาราง 2.1

ตาราง 2.1 แสดงค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณค่ายอดรวมประชากร T_1 เทียบกับ ตัวประมาณค่ายอดรวมประชากร T_2 เมื่อทุกหน่วยของประชากรมีความน่าจะเป็นที่จะให้ข้อมูลไม่เท่ากัน

N	ρ^*	μ_p	$V(T_1)$	$V_3(T_2)$	E
4	0.93	0.75	240.66	12.40	20.00
	0.40	0.61	712.62	291.48	3.25
	-0.06	0.72	378.26	86.76	4.39
	-0.34	0.75	306.64	51.11	6.00
	-0.92	0.68	583.23	287.41	2.03
6	0.92	0.67	623.53	27.47	23.07
	0.51	0.64	581.25	83.76	7.00
	0.00	0.69	577.25	45.69	12.82
	-0.51	0.70	822.82	131.76	6.27
	-0.90	0.65	931.35	349.08	2.66
8	0.90	0.52	1708.45	168.89	10.16
	0.50	0.63	1271.04	63.94	20.17
	-0.01	0.58	1729.43	343.29	5.04
	-0.52	0.70	993.08	39.40	25.46
	-0.86	0.69	970.80	123.59	7.88
10	0.87	0.54	1544.54	131.06	11.78
	0.49	0.67	951.47	34.30	27.97
	0.00	0.71	1086.62	45.53	24.13
	-0.51	0.67	1354.28	206.59	6.50
	-0.92	0.63	1922.38	385.90	4.99

ρ^* เป็นค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าของสิ่งที่ต้องการศึกษากับความน่าจะเป็นที่แต่ละหน่วยจะให้ข้อมูล

μ_p เป็นค่าเฉลี่ยของความน่าจะเป็นที่จะให้ข้อมูล

2.2 วิธีการทางสถิติ

2.2.1 วิธีประมาณค่ายอดรวมประชากรจากตัวอย่างที่มีขนาดเป็นตัวแปรสุ่มตามแนวคิดของฮอร์วิทซ์และทอมป์สัน

ในการประมาณค่ายอดรวมประชากรที่มีขนาด N โดยศึกษาจากตัวอย่างขนาด n ซึ่งเลือกมาด้วยวิธีสุ่มอย่างง่ายโดยไม่ใส่คืน และมี p_i เป็นค่าความน่าจะเป็นที่แต่ละหน่วยในประชากรให้ข้อมูล ภายหลังการสำรวจตัวอย่างพบว่ามีข้อมูลที่ศึกษาได้เพียง m หน่วย จากตัวอย่างขนาด n ดังนั้น m = 1 , 2 ,..., n จะเห็นว่าตัวอย่างที่จะนำไปศึกษามีขนาดเป็นตัวแปรสุ่ม

ให้ T_1 เป็นตัวประมาณค่ายอดรวมประชากรที่มีขนาดเป็นตัวแปรสุ่ม

$$T_1 = 0 \qquad , m=0$$

$$= \frac{N}{n}\sum_{i=1}^{N}\frac{y_i\alpha_i\varepsilon_i}{P_i} \qquad , m>0$$

โดยที่

$$\alpha_i = 0 \quad , \text{ อื่นๆ} \qquad \text{----(1)}$$

$$= 1 \quad , y_i \text{ อยู่ในตัวอย่าง } , i=1,2,\ldots,N$$

$$\varepsilon_i = 0 \quad , \text{ อื่นๆ} \qquad \text{----(2)}$$

$$= 1 \quad , y_i \text{ ให้ข้อมูลที่ต้องการศึกษา } , i=1,2,\ldots,N$$

<u>และมีข้อสมมุติว่า</u>

1. $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_N$ อิสระกัน และ $E(\varepsilon_i) = P_i$
2. เวคเตอร์ $(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_N)$ อิสระกับเวคเตอร์ $(\alpha_1,\alpha_2,\ldots,\alpha_N)$
3. $E(\alpha_i) = \frac{n}{N}$ และ $E(\alpha_i\alpha_j) = \frac{n(n-1)}{N(N-1)}$

<u>ต้องการแสดงว่า</u>

1. T_1 เป็นตัวประมาณค่าที่ไม่เอนเอียงของยอดรวมประชากร
2. $V(T_1) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)S^2 + \frac{N}{n}\sum_{i=1}^{N}\frac{y_i^2(1-P_i)}{P_i}$ ----(3)

พิสูจน์ $E(T_1) = \frac{N}{n} \sum_{i=1}^{N} \frac{y_i E(\alpha_i) E(\varepsilon_i)}{P_i}$

$$= \frac{N}{n} \sum_{i=1}^{N} \frac{y_i}{P_i} \frac{n}{N} P_i$$

$$= \sum_{i=1}^{N} y_i \qquad \text{----(4)}$$

จาก----(1)และข้อสมมุติที่ 3 จะได้ว่า $E(\alpha_i)^2 = \frac{n}{N}$

----(2)และข้อสมมุติที่ 1 จะได้ว่า $E(\varepsilon_i)^2 = P_i$

$$V(T_1) = E(T_1)^2 - [E(T_1)]^2 \qquad \text{----(5)}$$

พิจารณา $E(T_1)^2 = E\left[\frac{N}{n} \sum_{i=1}^{N} \frac{y_i \alpha_i \varepsilon_i}{P_i}\right]^2$

$$= E\left[\frac{N^2}{n^2} \sum_{i=1}^{N} \frac{y_i^2 \alpha_i^2 \varepsilon_i^2}{P_i^2} + \sum_{i \neq j=1}^{N} \frac{y_i y_j \alpha_i \alpha_j \varepsilon_i \varepsilon_j}{P_i P_j}\right]$$

$$= \frac{N^2}{n^2}\left[\sum_{i=1}^{N} \frac{y_i^2 E(\alpha_i)^2 E(\varepsilon_i)^2}{P_i^2} + \sum_{i \neq j=1}^{N} \frac{y_i y_j E(\alpha_i \alpha_j) E(\varepsilon_i) E(\varepsilon_j)}{P_i P_j}\right]$$

$$= \frac{N^2}{n^2}\left[\sum_{i=1}^{N} \frac{y_i^2}{P_i^2} \frac{n}{N} P_i + \sum_{i \neq j=1}^{N} \frac{y_i y_j}{P_i P_j} \frac{n(n-1)}{N(N-1)} P_i P_j\right]$$

$$= \frac{N}{n} \sum_{i=1}^{N} \frac{y_i^2}{P_i} + \frac{N(n-1)}{n(N-1)} \sum_{i \neq j=1}^{N} y_i y_j \qquad \text{----(6)}$$

จาก----(5) จะได้ว่า;

$$[E(T_1)]^2 = \left(\sum_{i=1}^{N} y_i\right)^2 \qquad \text{----(7)}$$

นำ----(6)และ----(7)ไปแทนใน----(5);

$$V(T_1) = \frac{N}{n} \sum_{i=1}^{N} y_i^2 + \frac{N(n-1)}{n(N-1)} \sum_{i \neq j=1}^{N} y_i y_j - \left(\sum_{i=1}^{N} y_i\right)^2 \qquad \text{----(8)}$$

จาก----(6) พิจารณา $\frac{N(n-1)}{n(N-1)}\sum_{i\neq j}^{N} y_i y_j - \left(\sum_{i=1}^{N} y_i\right)^2$

$$= \frac{N(n-1)}{n(N-1)}\left[\left(\sum_{i=1}^{N} y_i\right)^2 - \sum_{i=1}^{N} y_i^2\right] - \left(\sum_{i=1}^{N} y_i\right)^2$$

$$= \left[\frac{N(n-1)}{n(N-1)}\left(\sum_{i=1}^{N} y_i\right)^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right] - \frac{N(n-1)}{n(N-1)}\sum_{i=1}^{N} y_i^2$$

$$= \left[\frac{N(n-1)}{n(N-1)} - 1\right]\left(\sum_{i=1}^{N} y_i\right)^2 - \frac{(nN-N)}{n(N-1)}\sum_{i=1}^{N} y_i^2$$

$$= \frac{nN-N-nN+n}{n(N-1)}\left(\sum_{i=1}^{N} y_i\right)^2 - \frac{N}{N-1}\sum_{i=1}^{N} y_i^2 + \frac{N}{n(N-1)}\sum_{i=1}^{N} y_i^2$$

$$= \frac{n-N}{n(N-1)}\left(\sum_{i=1}^{N} y_i\right)^2 - \frac{N}{N-1}\sum_{i=1}^{N} y_i^2 + \frac{N}{n(N-1)}\sum_{i=1}^{N} y_i^2$$

$$= \frac{\left(\sum_{i=1}^{N} y_i\right)^2}{N-1} - \frac{N}{n(N-1)}\left(\sum_{i=1}^{N} y_i\right)^2 - \frac{N}{N-1}\sum_{i=1}^{N} y_i^2 + \frac{N}{n(N-1)}\sum_{i=1}^{N} y_i^2$$

$$= -\frac{N}{N-1}\left[\sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right] + \frac{N}{n(N-1)}\left[\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right]$$

$$= -NS^2 + \frac{N}{n(N-1)}\left[(N-1)S^2 + N\bar{y}_N^2 - N^2\bar{y}_N^2\right]$$

$$= -NS^2 - \frac{N}{n(N-1)}\left[(N-1)S^2 - (N-1)N\bar{y}_N^2\right]$$

$$= -NS^2 + \frac{N}{n}\left[S^2 - N\bar{y}_N^2\right]$$

$$= -NS^2 + \frac{NS^2}{n} - \frac{N}{n}\left[\sum_{i=1}^{N} y_i^2 - (N-1)S^2\right]$$

$$= -NS^2 + \frac{NS^2}{n} - \frac{N}{n}\sum_{i=1}^{N} y_i^2 + \frac{N^2S^2}{n} - \frac{NS^2}{n}$$

$$= \frac{N^2S^2}{n} - NS^2 - \frac{N}{n}\sum_{i=1}^{N} y_i^2 \,.$$

$$= N^2S^2\left(\frac{1}{n}-\frac{1}{N}\right) - \frac{N}{n}\sum_{i=1}^{N} y_i^2$$

$$= N^2\left(\frac{1}{n}-\frac{1}{N}\right)S^2 - \frac{N}{n}\sum_{i=1}^{N} y_i^2 \qquad \text{----(9)}$$

นำ----(9)ไปแทนใน----(8);

$$V(T_1) = \frac{N}{n}\sum_{i=1}^{N}\frac{y_i^2}{P_i} + N^2\left(\frac{1}{n}-\frac{1}{N}\right)S^2 - \frac{N}{n}\sum_{i=1}^{N} y_i^2$$

$$= N^2\left(\frac{1}{n}-\frac{1}{N}\right)S^2 + \frac{N}{n}\sum_{i=1}^{N}\frac{y_i^2(1-P_i)}{P_i}$$

$$\text{Est}V(T_1) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)\text{Est}S^2 \;{}^{1} + \text{Est}\;\frac{N}{n}\sum_{i=1}^{N}\frac{y_i^2(1-P_i)}{P_i} \;{}^{2}$$

$$\hat{V}(T_1) = \frac{N(N-n)}{n(n-1)}\left[\sum_{i=1}^{m}\frac{y_i^2}{p_i} - \frac{1}{n}\left(\sum_{i=1}^{m}\frac{y_i^2}{p_i} + \sum_{i\neq j}^{m}\frac{y_i y_j}{p_i p_j}\right)\right] + \frac{N^2}{n^2}\sum_{i=1}^{m}\frac{y_i^2(1-p_i)}{p_i^2}$$

----

[1]ศึกษารายละเอียดเพิ่มเติมได้ในภาคผนวก ข.หน้า 68

[2]ศึกษารายละเอียดเพิ่มเติมได้ในภาคผนวก ก.หน้า 63

2.2.2 <u>วิธีประมาณค่ายอดรวมประชากรจากตัวอย่างที่มีขนาดเป็นตัวแปรสุ่ม ตามวิธีของซิงค์และนเรนทร์</u>

ในการประมาณค่ายอดรวมประชากรตามวิธีของซิงค์และนเรนทร์ จะมีการหาค่าใหม่เพิ่มขึ้น คือ $p(m)$ เป็นความน่าจะเป็นที่จะได้ข้อมูล m หน่วยจากตัวอย่างขนาด n $p(0)$ เป็นความน่าจะเป็นที่จะไม่ได้ข้อมูลจากตัวอย่างขนาด n มีค่าเท่ากับ $q_1q_2q_3\ldots q_n$ $\pi_i(m)$ เป็นความน่าจะเป็นที่จะได้ข้อมูลจากหน่วยที่ i ในตัวอย่างขนาด n ซึ่งมีอยู่ m หน่วยที่ให้ข้อมูล $\pi_{ij}(m)$ เป็นความน่าจะเป็นที่จะได้ข้อมูลจากหน่วยที่ i,j ในตัวอย่างขนาด n ซึ่งมีอยู่ m หน่วยที่ให้ข้อมูล , $i\neq j=1,2,\ldots,N$

$$T_2 = 0 \quad ,m=0$$

$$= \frac{N}{n(1-p_o)}\sum_{i=1}^{m}\frac{y_i}{\pi_i(m)} \quad ,m>0$$

<u>ต้องการแสดงว่า</u>

1. T_2 เป็นตัวประมาณค่าที่ไม่เอนเอียงของยอดรวมประชากร
2. $V(T_2) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)S^2 +$

$$\frac{N(n-1)}{n(N-1)(1-p_o)}\sum_{i<j}^{N}\sum^{N}\left(\pi_i(m)\pi_i(m)-\pi_{ij}(m)\right)\left(\frac{y_i}{\pi_i(m)}-\frac{y_j}{\pi_j(m)}\right)^2 \text{---(10)}$$

<u>พิสูจน์</u>

$$T_2 = \frac{N}{n(1-p_o)}\sum_{i=1}^{m}\frac{y_i}{\pi_i(m)} \text{----(11)}$$

take E_2 บน----(11);

$$E_2(T_2) = \frac{N}{n(1-p_o)}\sum_{i=1}^{n}\frac{y_i\ \pi_i(m)}{\pi_i(m)} \text{----(12)}$$

$$E_2(T_2) = \frac{N}{n}\sum_{i=1}^{n}y_i \text{----(13)}$$

$$= N\bar{y}_n \text{----(14)}$$

take E_1 บน----(13);

$$E_1E_2(T_2) = \frac{N}{n}\sum_{i=1}^{N} y_i \quad \frac{n}{N}$$

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

$$E(T_2) = E_1E_2(T_2) = \sum_{i=1}^{N} y_i$$

ดังนั้น T_2 เป็นตัวประมาณค่าที่ไม่เอนเอียงของยอดรวมประชากร

เพราะว่า $E(T_2)$ $E_1E_2(T_2)$ และ ทฤษฎี 1.8 ของ Des Raj จะได้ว่า

$$V(T_2) = E_1V_2(T_2) + V_1E_2(T_2)$$ [3]

จาก----(14);

$$V_1E_2(T_2) = V_1[N\bar{y}_n]$$ [4]

$$= N^2\left(\frac{1}{n} - \frac{1}{N}\right) S^2 \qquad ----(15)$$

$$V_2(T_2) \doteq E_2(T_2) - [E_2(T_2)]^2$$

----

[3]ศึกษารายละเอียดเพิ่มเติมในภาคผนวก ก.หน้า 60

[4]Sukhatme,P.V. and Sukhatme,B.V.,Sampling theory of Surveys with Application.,Second Revised Edition,(Ames ,IOWA:Iowa State University,1970)P.14.

จาก----(11);

$$E_2(T_2)^2 = E_2\left[\frac{N}{n(1-p_o)}\sum_{i=1}^{m}\frac{y_i}{\pi_i(m)}\right]^2$$

$$= E_2\left[\frac{N^2}{n^2(1-p_o)^2}\left(\sum_{i=1}^{m}\frac{y_i^2}{\pi_i^2(m)} + 2\sum\sum_{i<j}^{mm}\frac{y_i y_j}{\pi_i(m)\pi_j(m)}\right)\right]$$

$$= \frac{N^2}{n^2(1-p_o)^2}\left[\sum_{i=1}^{n}\frac{y_i^2}{\pi_i(m)} + 2\sum\sum_{i<j}^{nn}\frac{y_i y_j\,\pi_{ij}(m)}{\pi_i(m)\pi_j(m)}\right] \quad \text{----(16)}$$

จาก----(14);

$$[E_2(T_2)]^2 = E_2\left[\left(\frac{N}{n(1-p_o)}\sum_{i=1}^{m}\frac{y_i}{\pi_i(m)}\right)\right]^2$$

$$= \left[\frac{N}{n(1-p_o)}\sum_{i=1}^{n} y_i\right]^2$$

$$= \frac{N^2}{n^2(1-p_o)^2}\left[\sum_{i=1}^{n} y_i^2 + 2\sum\sum_{i<j}^{nn} y_i y_j\right] \quad \text{----(17)}$$

$$V_2(T_2) = (16) - (17)$$

$$V_2(T_2) = \frac{N^2}{n^2(1-p_o)^2}\left[\sum_{i=1}^{n}\frac{y_i^2(1-\pi_i(m))}{\pi_i(m)} + 2\sum\sum_{i<j}^{nn} y_i y_j\,\frac{\pi_{ij}(m)-\pi_i(m)\pi_j(m)}{\pi_i(m)\pi_j(m)}\right]$$

$$= \frac{N^2}{n^2(1-p_o)^2}\left[\sum_{i=1}^{n}\frac{y_i^2(1-\pi_i(m))}{\pi_i(m)}^{5} - 2\sum\sum_{i<j}^{nn} y_i y_j\,\frac{\pi_i(m)\pi_j(m)-\pi_{ij}(m)}{\pi_i(m)\pi_j(m)}\right]$$

---

[5]ศึกษารายละเอียดเพิ่มเติมได้ในภาคผนวก ก.หน้า 67

$$= \frac{N^2}{n^2(1-p_o)^2}\left[\sum_{i<j}^{n}\sum^{n} \pi_i(m)\pi_j(m)-\pi_{ij}(m)\left(\frac{y_i^2}{\pi_i(m)}-\frac{y_j^2}{\pi_j(m)}\right) - 2\sum_{i<j}^{n}\sum^{n} y_i y_j \frac{\pi(m)\pi_j(m)-\pi_{ij}(m)}{\pi_i(m)\pi_j(m)}\right]$$

$$= \frac{N^2}{n^2(1-p_o)^2}\left[\sum_{i<j}^{n}\sum^{n} \pi_i(m)\pi_j(m)-\pi_{ij}(m)\left(\frac{y_i}{\pi_i(m)}-\frac{y_j}{\pi_j(m)}\right)^2\right]$$

$$E_1V_2(T_2) = \frac{N^2}{n^2(1-p_o)^2} E_1\left[\sum_{i<j}^{N}\sum^{N} \pi_i(m)\pi_j(m)-\pi_{ij}(m)\left(\frac{y_i}{\pi_i(m)}-\frac{y_j}{\pi_j(m)}\right)^2\right]$$

$$= \frac{N(n-1)}{n(N-1)(1-p_o)}\left[\sum_{i<j}^{N}\sum^{N} \pi_i(m)\pi_j(m)-\pi_{ij}(m)\left(\frac{y_i}{\pi_i(m)}-\frac{y_j}{\pi_j(m)}\right)^2\right] \text{----(18)}$$

$$V(T_2) = (15) + (18)$$

$$= N^2\left(\frac{1}{n}-\frac{1}{N}\right)S^2 +$$

$$\frac{N(n-1)}{n(N-1)(1-p_o)}\left[\sum_{i<j}^{N}\sum^{N} (\pi_i(m)\pi_j(m)-\pi_{ij}(m))\left(\frac{y_i}{\pi_i(m)}-\frac{y_j}{\pi_j(m)}\right)^2\right]$$

$$\text{Est}V(T_2) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)\text{Est}S^{2\,6} +$$

$$\text{Est}\frac{N(n-1)}{n(N-1)(1-p_o)}\sum_{i<j}^{N}\sum^{N} (\pi_i(m)\pi_j(m)-\pi_{ij}(m))\left(\frac{y_i}{\pi_i(m)}-\frac{y_j}{\pi_j(m)}\right)^{2\,7}$$

$$= \frac{N(N-n)}{n(n-1)}\left[\sum_{i=1}^{m} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{m}\frac{y_i^2}{\pi_i(m)} + 2\sum_{i<j}^{m}\sum^{m}\frac{y_i y_j}{\pi_{ij}(m)}\right)\right] +$$

$$\frac{N^2}{n^2(1-p_o)^2}\sum_{i<j}^{m}\sum^{m}\frac{(\pi_i(m)\pi_j(m)-\pi_{ij}(m))}{\pi_{ij}(m)}\left(\frac{y_i}{\pi_i(m)}-\frac{y_j}{\pi_j(m)}\right)^2$$

---

[6] ศึกษารายละเอียดเพิ่มเติมได้ในภาคผนวก ข.หน้า 68

[7] ศึกษารายละเอียดเพิ่มเติมได้ในภาคผนวก ก.หน้า 63

2.2.3 <u>การศึกษาประสิทธิภาพของตัวประมาณค่ายอดรวมประชากร T_1 เทียบกับ ตัวประมาณค่ายอดรวมประชากร T_2 ตามวิธีของซิงค์และนเรนทร์</u>

ซิงค์และนเรนทร์ใช้ประสิทธิภาพความผันแปร (Coefficient of Variation) เป็นเกณฑ์ในการศึกษาประสิทธิภาพของตัวประมาณค่ายอดรวมประชากร T_1 เทียบกับ ตัวประมาณค่ายอดรวมประชากร T_2 พร้อมกันนี้ต้องมีข้อจำกัด สำหรับใช้ในการเปรียบเทียบความแปรปรวนดังต่อไปนี้ คือ

1. $\pi_{ij}(m) < \pi_i(m)\pi_j(m)$ และ $\pi_{ij}(m) > 0$, $i \neq j = 1,2,\ldots,N$
2. $p_i = p$, $i = 1,2,\ldots,N$
3. p_o มีค่าน้อยมาก ดังนั้น $1-p_o = 1-q^n \sim 1$

เนื่องจาก

$$V(T_1) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + \frac{N}{n}\sum_{i=1}^{N}\frac{y_i^2(1-p_i)}{p_i}$$

$$V(T_2) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + \frac{N(n-1)}{n(N-1)(1-p_o)}\sum_{i<j}^{N}\sum^{N}\left(\pi_i(m)\pi_j(m)-\pi_{ij}(m)\right)\left(\frac{y_i}{\pi_i(m)}-\frac{y_j}{\pi_j(m)}\right)^2$$

จากข้อจำกัด 2 ทำให้

$$V(T_1) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + \frac{Nq}{np}\sum_{i=1}^{N} y_i^2 \qquad \text{----(19)}$$

จากข้อจำกัด1,2,3 ทำให้

$$V(T_2) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + \left[\frac{1}{1-q^n} - 1\right]\left(\sum_{k=1}^{N} y_k\right)^2$$

$$= \frac{N^2}{1-q^n}\left[E\left(\frac{1}{m}\right) - \frac{1}{N}\right]S^2 + \left[\frac{1}{1-q^n} - 1\right]\left(\sum_{k=1}^{N} y_k\right)^2 \qquad \text{--(20}$$

และจากข้อจำกัดที่ว่า $1-q^n \sim 1$ ดังนั้น

จาก----(20) จะได้

$$V(T_2) = N^2\left[E\left(\frac{1}{m}\right) - \frac{1}{N}\right]S^2$$

$$= N^2\left[E\left(\frac{1}{m}\right) - \frac{1}{n} + \frac{1}{n} - \frac{1}{N}\right]S^2$$

$$= N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + N^2\left[E\left(\frac{1}{m}\right) - \frac{1}{n}\right]S^2 \qquad \text{----(21)}$$

สเตฟาน(Stephan(1945))ได้ประมาณ $E\left(\frac{1}{m}\right)$ ดังนี้

$$E\left(\frac{1}{m}\right) = \frac{1}{E(m)}\left[1 + \frac{V(m)}{[E(m)]^2}\right]$$

$$= \frac{1}{np} + \frac{q}{n^2p^2}$$

นำค่า $E\left(\frac{1}{m}\right)$ ไปแทนใน----(21);

$$V(T_2) = N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + N^2\left(\frac{1}{np} + \frac{q}{n^2p^2} - \frac{1}{n}\right)S^2$$

$$= N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + \frac{N^2}{n^2p^2}(np+q-np^2)S^2$$

$$= N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + \frac{N^2}{n^2p^2}(np(1-p)+q)S^2$$

$$= N^2\left(\frac{1}{n} - \frac{1}{N}\right)S^2 + \frac{N^2}{n^2p^2}(npq+q)S^2 \qquad ----(22)$$

จะเห็นว่าเมื่อมีข้อจำกัด 3 ข้อข้างต้น แล้วสามารถปรับ $V(T_2)$คือสมการ----(10)ให้อยู่ในรูปของสมการ----(22) และในการคำนวณ $V(T_2)$ด้วยสมการ----(22) จะสะดวกกว่าใช้สมการ----(10) เพราะเทอมที่สองของสมการ----(22)จะใช้ S^2 เช่นเดียวกับเทอมที่หนึ่ง

จาก----(19)และ----(22) จะได้ว่า $V(T_1) > V(T_2)$

ถ้า $$\sum_{i=1}^{N} y_i^2 > \frac{N}{np}(np+1)S^2$$

$$> \frac{N(np+1)}{np}\ \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right]$$

$$> \frac{1}{N-1}\left[N\sum_{i=1}^{N} y_i^2 + \frac{N}{np}\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2 - \frac{1}{np}\left(\sum_{i=1}^{N} y_i\right)^2\right]$$

$$> \frac{1}{N-1}\left[\frac{(Nnp+N)}{np}\sum_{i=1}^{N} y_i^2 - \frac{(np+1)}{}(\sum_{i=1}^{N} y_i)^2\right]$$

$$\frac{Nnp+N}{np(N-1)}\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} y_i^2 < \frac{np+1}{np(N-1)}(\sum_{i=1}^{N} y_i)^2$$

$$\frac{Nnp\sum_{i=1}^{N} y_i^2 + N\sum_{i=1}^{N} y_i^2 - Nnp\sum y_i^2 + np\sum_{i=1}^{N} y_i^2}{np(N-1)} < \frac{np+1}{np+N}(\sum_{i=1}^{N} y_i)^2$$

$$(N+np)\sum_{i=1}^{N} y_i^2 < (np+1)(\sum_{i=1}^{N} y_i)^2$$

$$\frac{\sum_{i=1}^{N} y_i^2}{(\sum_{i=1}^{N} y_i)^2} < \frac{np+1}{np+N}$$

$$\frac{N\sum_{i=1}^{N} y_i^2}{(\sum_{i=1}^{N} y_i)^2} < \frac{N(np+1)}{np+N}$$

$$\frac{N\sum_{i=1}^{N} y_i^2}{(\sum_{i=1}^{N} y_i)^2} - 1 < \frac{N(np+1)}{np+N} - 1$$

$$\frac{N\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2}{(\sum_{i=1}^{N} y_i)^2} < \frac{N(np+1) - np - N}{np+N}$$

$$\frac{N\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2}{(\sum_{i=1}^{N} y_i)^2} < \frac{(N-1)np}{np+N}$$

$$\frac{N\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2}{(\sum_{i=1}^{N} y_i)^2} \cdot \frac{1}{N-1} < \frac{np}{np+N}$$

$$\frac{N\sum_{i=1}^{N} y_i^2 - (\sum_{i=1}^{N} y_i)^2}{(\sum_{i=1}^{N} y_i)^2} \frac{N}{N-1} < \frac{Nnp}{np+N} \qquad \text{----(23)}$$

พิจารณา C.V. $\quad \dfrac{S}{\bar{y}_N}$ [8]

$$= \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - \frac{1}{N}\left(\sum_{i=1}^{N} y_i\right)^2\right]} \; \frac{N}{\sum_{i=1}^{N} y_i}$$

$$= \sqrt{\frac{N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2}{N(N-1)}} \; \frac{N}{\sum_{i=1}^{N} y_i}$$

$$(C.V.)^2 = \frac{N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2}{\left(\sum_{i=1}^{N} y_i\right)^2} \; \frac{N}{N-1}$$

จาก----(23)จะได้ว่า $\quad (C.V.)^2 < \dfrac{1}{\dfrac{1}{N} + \dfrac{1}{np}}$ ----(24)

โดยทั่วไป c.v.มีค่าน้อยกว่า 1 ซึ่งจะทำให้ $(c.v.)^2$ มีค่าน้อยกว่า 1 เสมอ ดังนั้นจะพิจารณาว่า $\dfrac{1}{\dfrac{1}{N} + \dfrac{1}{np}}$ มีค่ามากกว่า 1 มีความเป็นไปได้ คือ

ถ้าขนาดตัวอย่างเท่ากับขนาดประชากรแล้ว $\dfrac{1}{N} + \dfrac{1}{np} > 1$ เมื่อ

$N = n > 4$ และ $p > 0.5$

----

[8]Sukhatme,P.V. and Sukhatme,B.V.,Sampling Theory of Surveys with Application. , Second Revised Edition.P.20