บทที่ 2

ผลงานวิจัยที่เกี่ยวข้อง

ในการเลือกตัวอย่างแบบสุ่มอย่างง่ายชนิดไม่ใส่คืน เพื่อจะประมาณค่ารวมประชากร พบว่าตัวอย่างที่เลือกมามีค่า สังเกตบางหน่วยเป็นค่าสูงมาก และเป็นค่าที่มีอยู่จริงในประชากร ในกรณีนี้ถ้าขนาดตัวอย่างที่ใช้ไม่มากพอแล้ว การใช้ตัวประมาณ $\hat{Y}_o$ จะยังคงเป็นตัวประมาณที่ไม่เอนเอียง แต่ความแม่นยำในการประมาณค่าจะต่ำลงมาก ในการแก้ปัญหาดังกล่าวได้มีนักสถิติหลายท่านเสนอตัวประมาณในรูปแบบต่าง ๆ ขึ้นมาหลายรูปแบบ อาทิเช่น

ในปี ค.ศ. 1966 โดนัลด์ ที เซิรลส์ (Donald T. Searls) ได้เสนอตัวประมาณ $\bar{y}_t$ สำหรับประมาณค่าเฉลี่ยประชากร วิธีการประมาณค่าคือ ตัดหน่วยตัวอย่างที่มีค่าสูงมากทิ้งไป แล้วเปลี่ยนแปลงน้ำหนัก (weight) ของค่าเฉลี่ยตัวอย่าง ($\bar{y}$) ใหม่ โดยกำหนดให้

$$\bar{y}_t = \frac{\sum_{i=1}^{r} y_i + (n-r)t}{n} \quad ; \quad r = 0\ ,\ 1,\ 2,\ \ldots,\ n\ ;\ y_i \leq t$$

เมื่อ y_i เป็นค่าสังเกตจากกลุ่มค่าสังเกตในตัวอย่างที่มีค่าน้อยกว่าหรือเท่ากับ t

r เป็นจำนวนค่าสังเกต y_i ที่มีค่าน้อยกว่าหรือเท่ากับ t

t เป็นค่าที่ทำให้ first partial differentive ของ $MSE(\bar{y}_t)$ มีค่าต่ำสุด

โดยที่ y_i เป็นตัวแปรสุ่มที่เป็นอิสระกันที่สุ่มมาจาก original distribution ที่มีฟังก์ชั่นความหนาแน่น (probability density function = p.d.f.) คือ $f(x)$ ฟังก์ชั่นการแจกแจง (cumulative density function = c.d.f.) คือ $F(x)$ และมีจุด t เป็นจุดตัด (truncated point) ทางขวามือของการแจกแจง ให้ μ_t และ σ_t^2 เป็นค่าเฉลี่ยและความแปรปรวนของ truncated distribution จะได้

$$MSE(\bar{y}_t) = \left(\frac{\rho}{n}\right) \left[\sigma_t^2 + q\ (t-\mu_t)^2 \right] + q^2(\mu_t - t)^2$$

เมื่อ $\rho = F(t)$, $q = 1 - \rho$

และศึกษาคุณสมบัติของตัวประมาณ $\bar{y}_t$ จากค่าประสิทธิภาพสัมพัทธ์ของ $\bar{y}_t$ เทียบกับ $\bar{y}$ โดยการจำลองข้อมูลให้ประชากรมีการแจกแจงแบบเอกซ์โปเนนเชียล (exponential) ที่มีฟังก์ชันความหนาแน่น $f(y) = \frac{1}{\mu} \exp(-y/\mu)$; $0 < y$ พบว่าประสิทธิภาพของตัวประมาณจะขึ้นอยู่กับอัตราส่วนระหว่างค่า t กับ μ กล่าวคือ ประสิทธิภาพของตัวประมาณ $\bar{y}_t$ จะลดลงถ้าค่า t/μ เพิ่มขึ้น และประสิทธิภาพของตัวประมาณ $\bar{y}_t$ จะเทียบเท่ากับ $\bar{y}$ เมื่อค่า t/μ มีค่ามากกว่า 8 แต่ถ้าค่า t มีค่าเท่ากับ μ แล้วประสิทธิภาพของตัวประมาณ $\bar{y}_t$ จะต่ำกว่า $\bar{y}$ ซึ่งจะเห็นผลได้ชัดเจนเมื่อใช้ขนาดตัวอย่างใหญ่ ดังผลที่แสดงในตารางที่ 2.1

ตารางที่ 2.1 แสดงค่าประสิทธิภาพสัมพัทธ์ของตัวประมาณ $\bar{y}_t$ เทียบกับ $\bar{y}$ (%) เมื่อประชากรมีการแจกแจงแบบเอกซ์โปเนนเชียล

ค่าของ t/μ	ขนาดตัวอย่าง				
	5	10	50	100	500
1	83.4	44.1	9.2	4.6	.9
2	183.1	153.3	66.7	39.1	9.1
3	140.5	138.0	120.6	104.2	49.9
4	117.0	116.7	114.9	112.7	97.7
5	107.2	107.2	106.9	106.7	104.7
6	103.1	103.1	103.0	103.0	102.7
7	101.3	101.3	101.3	101.3	101.3
8	100.5	100.5	100.5	100.5	100.5
9	100.2	100.2	100.2	100.2	100.2
10	100.1	100.1	100.1	100.1	100.1

แหล่งที่มา : Journal of the American Statistical Association ฉบับที่ 61 ปี 1966 หน้า 1204

ต่อมาในปี ค.ศ. 1973 เจนคิน ริงเกอร์ ฮาร์ทเล่ย์ (Jenkin Ringer Hartley) ได้เสนอใช้ตัวประมาณในรูปแบบกรณีที่ล่อง (square root estimator ; $\hat{Y}$)มาประมาณค่าเฉลี่ยประชากร เมื่อ $\hat{Y} = C_1\bar{y} + C_2\bar{U}^2$ โดยที่ $U_i = \sqrt{y_i}$; $i = 1,2,\ldots, n$ ส่วน C_1 และ C_2 เป็นค่าเฉพาะสำหรับในแต่ละการแจกแจงของประชากร เพื่อที่จะทำให้ $\hat{Y}$ เป็นตัวประมาณที่เอนเอียงของค่าเฉลี่ยประชากรที่มีประสิทธิภาพสูงกว่า $\bar{y}$ พบว่าค่า C_1 และ C_2 ที่จะทำให้การใช้ตัวประมาณในรูปแบบกรณีที่ล่องมีประสิทธิภาพสูงกว่า $\bar{y}$ ในบางลักษณะของการแจกแจงของประชากรเป็นดังนี้

1. เมื่อกำหนดให้ $C_1 + C_2 = 1$ และ $C_2 = C$ ซึ่งจะแสดงค่า C ที่เหมาะสมในบางลักษณะของการแจกแจงของประชากรในตารางที่ 2.2 ดังนี้

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 2.2 แสดงค่า C ที่เหมาะสมในการแจกแจงบางลักษณะที่ควรใช้การประมาณค่าแบบ square root estimator

แฟมมิลิของการแจกแจง (family)	รูปแบบ	C
แกมม่า (Gamma)	$1/6\ y^{3}\ e^{-y}$	1
	$1/2\ y^{2}\ e^{-y}$	2
	$y\ e^{-y}$	3
	e^{-y}	4
สแควร์รูทแกมม่า (Square-Root gamma)	$(1/12)\, y e^{-\sqrt{y}}$	5
	$\frac{1}{4}\ \sqrt{y}\ e^{-\sqrt{y}}$	6
	$\frac{1}{2}\ e^{-\sqrt{y}}$	7
	$(1/2) \cdot y^{-1/2}\ e^{-\sqrt{y}}$	8
พาเรโต (Pareto)	y^{-2}	9
	$2y^{-3}$	10
	$3y^{-4}$	11
	$4y^{-5}$	12
	$5y^{-6}$	13

แหล่งที่มา : Journal of the American Statistical Association ฉบับที่ 68 ปี 1973 หน้า 416

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2. เมื่อกำหนดให้ $C_1 \neq C_2$ จะแสดงค่า C_1 , C_2 ที่เหมาะสมในบางลักษณะของการแจกแจงของประชากร ในตารางที่ 2.3 ดังนี้

ตารางที่ 2.3 แสดงค่า C_1 C_2 ที่เหมาะสมในบางลักษณะของประชากร สำหรับขนาดตัวอย่างตั้งแต่ 2 ถึง 20

	n																	
Distribution	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
(1) $\frac{1}{6} y^3 e^{-y}$	.89	.92	.94	.95	.96	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.99	.99	.99
	.96	.96	.96	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.99	.99	.99	.99	.99
(2) $\frac{1}{2} y^2 e^{-y}$	.86	.90	.92	.94	.95	.95	.96	.96	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98
	.94	.94	.95	.96	.96	.96	.97	.97	.97	.97	.98	.98	.98	.98	.98	.98	.98	.98
(3) $y e^{-y}$	.80	.86	.89	.91	.92	.93	.94	.95	.95	.96	.96	.96	.96	.97	.97	.97	.97	.97
	.92	.92	.93	.93	.94	.95	.95	.96	.96	.96	.96	.97	.97	.97	.97	.97	.98	.99
(4) e^{-y}	.67	.75	.80	.83	.86	.88	.89	.90	.91	.92	.93	.93	.94	.94	.94	.94	.95	.95
	.84	.84	.86	.87	.88	.90	.91	.91	.92	.93	.93	.94	.94	.94	.95	.95	.95	.95
(5) $\frac{1}{n} y e^{-\sqrt{y}}$	.63	.70	.74	.77	.79	.81	.82	.83	.84	.84	.85	.85	.86	.86	.86	.86	.87	.87
	.73	.73	.75	.77	.79	.81	.83	.84	.85	.86	.87	.88	.88	.89	.90	.90	.91	.91
(6) $\frac{1}{4} \sqrt{y}\, e^{-\sqrt{y}}$	.56	.64	.69	.72	.74	.76	.77	.79	.80	.80	.81	.82	.82	.82	.83	.83	.83	.84
	.67	.67	.69	.72	.74	.76	.78	.80	.81	.82	.84	.84	.85	.86	.87	.87	.88	.88
(7) $\frac{1}{2} e^{-\sqrt{y}}$	.45	.54	.59	.63	.66	.68	.70	.72	.73	.74	.75	.75	.76	.77	.77	.78	.78	.79
	.57	.57	.60	.63	.66	.68	.71	.73	.75	.76	.78	.79	.80	.81	.82	.83	.83	.84
(8) $\frac{1}{2} y^{\frac{1}{2}} e^{-\sqrt{y}}$	.28	.36	.42	.46	.50	.53	.55	.57	.59	.60	.62	.63	.64	.65	.65	.66	.67	.67
	.40	.40	.43	.47	.50	.53	.56	.58	.60	.62	.64	.66	.68	.69	.70	.71	.72	.73
(9) y^{-2} $(y \geq 1)$	.26	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.27	.26
	.28	.28	.33	.36	.39	.42	.44	.46	.48	.50	.52	.54	.54	.55	.57	.58	.59	.60
(10) $2y^{-3} (y \geq 1)$	.22	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20	.20
	.22	.22	.23	.24	.26	.27	.29	.30	.32	.33	.34	.36	.37	.38	.39	.40	.41	.42
(11) $3y^{-4} (y \geq 1)$	.35	.34	.34	.34	.34	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33	.33
	.35	.35	.36	.37	.39	.40	.41	.42	.44	.45	.46	.47	.48	.49	.50	.51	.52	.52
(12) $4y^{-5} (y \geq 1)$	.44	.42	.41	.41	.41	.41	.41	.41	.40	.40	.40	.40	.40	.40	.40	.40	.40	.40
	.45	.45	.47	.48	.50	.52	.53	.55	.56	.57	.59	.60	.61	.62	.63	.64	.65	.66
(13) $5y^{-6} (y \geq 1)$	.49	.46	.45	.44	.44	.44	.44	.44	.44	.43	.43	.43	.43	.43	.43	.43	.43	.43
	.51	.51	.52	.54	.56	.57	.59	.61	.62	.63	.65	.66	.67	.68	.69	.70	.71	.71

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

แหล่งที่มา :Journal of the American Statistical Association ฉบับที่ 68 ปี 1973 หน้า 418

และในปี ค.ศ. 1981 ไมเคิล และคาดาบา (Michael and Kadaba) ได้เสนอตัวประมาณค่ารวมประชากรขึ้นมาใหม่ 4 รูปแบบ เมื่อสมมติให้ประชากรขนาด N คือ $\{Y_1, \ldots, Y_N\}$ มีจำนวนค่าสังเกตที่เป็นค่าสูงมากในประชากรเท่ากับ N_1 เลือกตัวอย่างแบบสุ่มอย่างง่าย ชนิดไม่ใส่คืนมาขนาด n แล้วได้จำนวนหน่วยตัวอย่างที่มีค่าสูงมากเท่ากับ n_1 คือ $\{y_1, \ldots, y_{n_1}\}$ ตัวประมาณที่เสนอคือ

กรณีที่ไม่ทราบค่า N_1 เสนอ

$$\hat{Y}_{mk1} = \sum_{i=1}^{n_1} y_i + \frac{N - n_1}{n_2} \sum_{i=n_1+1}^{n} y_i$$

$$\hat{Y}_{mk2} = \frac{N}{n} \sum_{i=1}^{n} y_i - \frac{Nn_1n_2}{2n^2} \left(\sum_{i=1}^{n_1} \frac{y_i}{n_1} - \sum_{i=n_1+1}^{n} \frac{y_i}{n_2} \right)$$

$$\hat{Y}_{mk3} = r \sum_{i=1}^{n_1} y_i + \frac{N - rn_1}{n_2} \cdot \sum_{i=n_1+1}^{n} y_i \text{ ; r เป็น optimal weight}$$

กรณีที่ทราบค่า N_1 เสนอใช้

$$\hat{Y}_{mk4} = \frac{N_1}{n_1} \sum_{i=1}^{n_1} y_i + \frac{N_2}{n_2} \sum_{i=n_1+1}^{n} y_i$$

โดยที่ $N_2 = N - N_1$; $n_2 = n - n_1$ และกำหนดให้

$f = n/N$

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

$\bar{Y}_2$ = ค่าเฉลี่ยประชากร Y ที่มีค่าสังเกตเป็นค่าไม่สูงมากหรือเป็นค่าปกติ

$\delta = \bar{Y}_1/\bar{Y}_2$

$S^2_{Y_1}$ = ค่าความแปรปรวนของประชากร Y ที่มีค่าสังเกตเป็นค่าสูงมาก

$S^2_{Y_2}$ = ค่าความแปรปรวนของประชากร Y ที่มีค่าสังเกตเป็นค่าปกติหรือไม่เป็นค่าสูงมาก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

C_1 = ค่าสัมประสิทธิความแปรผันของกลุ่มที่มีค่าสังเกตสูงมาก $= \frac{\sqrt{S^2_{Y_1}}}{\bar{Y}_1}$

C_2 = ค่าสัมประสิทธิความแปรผันของกลุ่มที่มีค่าสังเกตเป็นค่าปกติ $= \frac{\sqrt{S^2_{Y_2}}}{\bar{Y}_2}$

จะได้ความเอนเอียงอย่างมีเงื่อนไขเมื่อกำหนดค่า n_1 (conditional bias on n_1) ของ $\hat{Y}mkt$; t = 1,2,3,4 สำหรับการอนุมานอย่างมีเงื่อนไข (conditional inference) เป็น

$$B(\hat{Y}mk1 \mid n_1) = -(N_1 - n_1)\ (\delta - 1)\ \bar{Y}_2$$

$$B(\hat{Y}mk2 \mid n_1) = -(N_1 - \frac{Nn_1}{n} \cdot (\frac{n + n_1}{2n})) \cdot (\delta - 1)\ \bar{Y}_2$$

$$B(\hat{Y}mk3 \mid n_1) = -(N_1 - rn_1)\ (\delta - 1)\ \bar{Y}_2$$

$$B(\hat{Y}mk4 \mid n_1) = 0$$

และมีความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างมีเงื่อนไข (conditional mean square error) เป็นดังนี้

$$MSE(\hat{Y}mk1 \mid n_1) = \left\{ n_1^2 \left(\frac{1}{n_1} - \frac{1}{N_1}\right) C_1^2\ \delta^2 + (N-n_1)^2 \left(\frac{1}{n_2} - \frac{1}{N_2}\right) C_2^2 + (N_1 - n_1)^2 (\delta - 1)^2 \right\} \bar{Y}_2^2$$

$$MSE(\hat{Y}mk2 \mid n_1) = \left\{ \left[\frac{N}{2n^2} \cdot n_1 \ (n+n_1) \right]^2 \left(\frac{1}{n_1} - \frac{1}{N_1}\right) C_1^2\ \delta^2 + \left[\frac{N}{2n^2} (2n^2 - nn_1 - n_1^2 \right. \right.$$

$$\left. \left(\frac{1}{n_2} - \frac{1}{N_2}\right) C_2^2 + \left(N_1 - \frac{Nn_1(n+n_1)}{2n^2}\right)^2 (\delta - 1)^2 \right\} \bar{Y}_2^2$$

$$MSE(\hat{Y}mk3 \mid n_1) = \left\{ r^2 n_1^2 \left(\frac{1}{n_1} - \frac{1}{N_1} \right) C_1^2 \delta^2 + (N-rn_1)^2 \left(\frac{1}{n_2} - \frac{1}{N_2} \right) C_2^2 + \right.$$

$$\left. (N_1 - rn_1)^2 \times (\delta - 1)^2 \right\} \bar{Y}_2^2$$

เมื่อ

$$r = \frac{(NN_1 (N_2 - n_2) C_2^2 + n_2 N_2 N_1^2 (\delta - 1)^2)}{n_2 N_2 \left[(N_1 - n_1) C_1^2 \delta^2 + n_1 N_1 (N_2 - n_2) C_2^2 + n_1 N_1 (\delta - 1)^2 \right]}$$

$$MSE(\hat{Y}mk4 \mid n_1) = \left\{ N_1^2 \left(\frac{1}{n_1} - \frac{1}{N_1} \right) C_1^2 \delta^2 + N_2^2 \left(\frac{1}{n_2} - \frac{1}{N_2} \right) C_2^2 \right\} \bar{Y}_2^2$$

สำหรับการอนุมานอย่างไม่มีเงื่อนไขเกี่ยวกับค่า n_1 จำเป็นต้องทราบการแจกแจงของ n_1 เนื่องจากความเอนเอียงและความคลาดเคลื่อนกำลังสองเฉลี่ยในการอนุมานอย่างไม่มีเงื่อนไขเกี่ยวข้องกับค่าคาดหมายของฟังก์ชั่นของ n_1 และในการเลือกตัวอย่างแบบสุ่มไม่ใส่คืน จะได้การแจกแจงของ n_1 เป็น positive hypergeometric distribution ที่มีฟังก์ชั่นความหนาแน่นคือ

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

เมื่อ $n_1 = 1, 2, \ldots, N_1$

$N_2 = N - N_1 \geq n$

$n > N_1$

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

011921

จะได้ความเอนเอียงแบบไม่มีเงื่อนไข (unconditional bias) สำหรับ

$\hat{Y}mkt$; t = 1,2,3,4 ดังนี้

$$B(\hat{Y}mk_1) = -N_1(\delta - 1)\left(1 - \frac{f}{1-d}\right)\bar{Y}_2$$

$$B(\hat{Y}mk2) = \frac{N_1(\delta - 1)}{1-d}\left[-(n-1)N_2 + d\right]\bar{Y}_2$$

$$B(\hat{Y}mk3) = -N_1(\delta - 1)\left(1 - \frac{rf}{1-d}\right)\bar{Y}_2$$

$$B(\hat{Y}mk4) = 0$$

และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างไม่มีเงื่อนไข (unconditional mean square error) ของ $\hat{Y}mkt$; t = 1,2,3,4 เป็น

$$MSE(\hat{Y}mk1) = \Big\{ \frac{f(N_1 - 1)(1-f)}{1-d} \cdot \frac{N}{N-1} C_1^2 \delta^2 + C_2^2 (N-n)^2 \Big[E\left(\frac{1}{n_2}\right) - \frac{1}{f\left(N - \frac{N1}{1-d}\right)} \Big] + \Big[\left(N - \frac{fN1}{1-d}\right)^2 \left(\frac{1}{N-(N_1/1-d)} - \frac{1}{N_2}\right) + \frac{df^2 N_1^2}{N_2(1-d)^2}\Big] C_2^2 +$$

$$(1-f) \cdot \frac{1}{fN_2}\Big[\frac{\left(N - \frac{fN_1}{1-d}\right)^2}{N-N_1/1-d} - \frac{f^2 N_1 N_2}{(1-d)(N-1)}\Big] C_2^2 +$$

$$(1-\delta)^2 \Big[N_1^2 \left(1 - \frac{f}{1-d}\right)^2 - df^2 N_1^2/(1-d)^2 + f\frac{N_1}{(1-d)}(1-f)\frac{N_2}{N-1}\Big]\Big\}\bar{Y}_2^2$$

$$MSE(\hat{Y}mk2) = \Bigg\{ \left[\frac{N_1(1-\delta)\ N_2(n-1)}{2n(N-1)} \right]^2 + \frac{(C_1\delta)^2}{2f} \Bigg[En_1 + \frac{2En_1^2}{n} + \frac{En_1^3}{n^2}$$

$$- \frac{1}{N_1}(En_1^2 + \frac{2En_1^3}{n} + \frac{En_1^4}{n^2}) \Bigg] + (\frac{C_2}{2f})^2 \Bigg[(4n - \frac{3En_1^2}{n} -$$

$$\frac{En_1^3}{n^2}) - \frac{1}{N_2}(4n^2 - 4nEn_1 - 3En_2^2 + \frac{2En_1^3}{n} + \frac{En_1^4}{n^2}) \Bigg]$$

$$+ (\frac{1}{2nf})^2 (1-\delta)^2 V(n\,n_1 + n_1^2) \Bigg\} \bar{Y}_2^2$$

เมื่อ $V(nn_1 + n_1^2) = n^2 V(n_1) + 2n\ cov(n_1, n_1^2) + V(n_1^2)$ และ En_1^t ; $t = 1,2$ 3,4 คือโมเมนต์ที่ t ของ n_1

$$MSE(\hat{Y}_{mk3}) = \Bigg\{ r^2 f \frac{(N_1-1)(1-f)}{1-d} \cdot \frac{N}{N-1} \cdot C_1^2 \delta^2 + C_2^2 (N-rn)^2 \Bigg[E(\frac{1}{n_2})$$

$$- \frac{1}{f[N-N_1/(1-d)]} \Bigg] + \Bigg[(N-\frac{rfN_1}{1-d})^2 (\frac{1}{N-(N_1/(1-d))} - \frac{1}{N_2})$$

$$+ \frac{r^2 df^2 N_1^2}{N_2(1-d)^2} \Bigg] C_2^2 + (\frac{1-f}{fN_2}) \Bigg[(N-\frac{rfN_1}{1-d})^2 \frac{N_2}{N-(N_1/1-d)}$$

$$- \frac{r^2 f^2 N_1 N_2}{(1-d)(N-1)} \Bigg] C_2^2 + (1-\delta)^2 \Bigg[N_1^2 (1-\frac{rf}{1-d})^2 - \frac{r^2 df^2 N_1^2}{(1-d)^2}$$

$$+ \; r^2 f \; \frac{N_1(1-f) \, . \, N_2)}{(1-d)(N-1)} \Big] \Big\} \; \bar{Y}_2^2$$

โดยที่ $r = q_1(N,f,N_1,\delta,C_2)/q_2(N,f,N_1,\delta,C_1,C_2)$

เมื่อ

$$q_1(N,f,N_1,\delta,C_2) = N^2 f C_2^2 \left[E\left(\frac{1}{n_2}\right) - \frac{1}{f\left(N - \frac{N_1}{1-d}\right)} \right] + \frac{(1-\delta)^2}{1-d} \cdot fN_1^2 \; +$$

$$\left(\frac{1-f}{N - \frac{N_1}{1-d}}\right) \cdot \frac{N_1 \, . N}{1-d} \; C_2^2 + \frac{fN_1 N}{1-d} \left[\frac{1}{N - \frac{N_1}{1-d}} - \frac{1}{N_2} \right]$$

$$\text{และ}\; q_2(N,f,N_1,\delta,C_1,C_2) = (1-\delta)^2 \; f \frac{N_1}{1-d} \left[(1-f) \frac{N_2}{N-1} + fN_1 \right] + \frac{f(1-f)(N_1-1)}{1-d} \cdot$$

$$\left[C_1^2 \; \frac{N}{N-1} \; \delta^2 + \frac{N_1}{N-(N_1/1-d)} \cdot \frac{C_2^2}{(1-d)} \right] \; +$$

$$\frac{f(1-f)}{1-d} \; N_1 \left[\frac{1}{(1-d)[N-(N_1/1-d)]} - \frac{1}{N-1} \right] C_2^2$$

$$+ \; N^2 f^2 C_2^2 \left[E\left(\frac{1}{n_2}\right) - \frac{1}{f\left(N - \frac{N_1}{1-d}\right)} \right] + \frac{f^2 N_1^2}{(1-d)^2} \; x$$

$$\left[\frac{1}{N - \frac{N_1}{1-d}} - \frac{1}{N_2} \right] C_2^2 + \frac{df^2 \, N_1^2}{N_2(1-d)^2} \; C_2^2$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

$$MSE(\hat{Y}_{mk4}) = V(\hat{Y}_{mk4}) = \Big\{ N_1^2 \left[En_1^{-1} - N_1^{-1} \right] C_1^2\, \delta^2 + N_2^2 \left[En_2^{-1} - N_2^{-1} \right] . C_2^2 \Big\} \bar{Y}_2^2$$

สำหรับตัวประมาณค่ารวมประชากร $\hat{Y}_o$ ถ้าประมาณค่าตามลักษณะวิธีการประมาณค่าของไมเคิลและคาดาบา จะได้ว่าตัวประมาณ $\hat{Y}_o$ จะเป็นตัวประมาณที่เอนเอียงของค่ารวมประชากร โดยมีความเอนเอียงเมื่อกำหนดค่า n_1 หรือความเอนเอียงอย่างมีเงื่อนไข เท่ากับ $B(\hat{Y}_o|n_1) = -(N_1 - \frac{Nn_1}{n})(\delta - 1)\bar{Y}_2$ และมีความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างมีเงื่อนไข กำหนดค่า n_1 เป็น

$$MSE(\hat{Y}_o|n_1) = \Big\{ f^{-2} \left[(n_1 - \frac{n_1^2}{N_1})\; C_1^2\, \delta^2 + \frac{n_2}{N_2} (N_2 - n_2)\; C_2^2 \right] + (N_1 - f^{-1} n_1)^2 (\delta - 1)^2 \Big\} \bar{Y}_2^2$$

สำหรับเมื่อไม่มีเงื่อนไขเกี่ยวกับจำนวนค่า n_1 จะได้ความเอนเอียงเท่ากับ $B(\hat{Y}_o) = \frac{d}{1-d} N_1 (\delta - 1)\bar{Y}_2$ และความคลาดเคลื่อนกำลังสองเฉลี่ยอย่างไม่มีเงื่อนไขเกี่ยวกับค่า n_1 เป็น

$$MSE(\hat{Y}_o) = \Big\{ (f^{-1} - 1) \frac{N}{N-1} \cdot \frac{1}{1-d} \left[\frac{N_1}{N} . N_2 (1 - \delta)^2 + (N_1 - 1) C_1^2\, \delta^2 + (N_2 - 1)\, C_2^2 \right] - \frac{d}{1-d} \left[f^{-1} \frac{N}{N_2} . (N_2 - n)\, C_2^2 + (1 - \delta)^2 N_1^2 \right] \Big\} \bar{Y}_2^2 \quad 1$$

ไมเคิลและคาดาบา พบว่าคุณสมบัติของตัวประมาณ $\hat{Y}_{mkt}$; $t = 1,2,3,4$ เมื่อเทียบกับ $\hat{Y}_o$ สำหรับในกรณีการอนุมานอย่างไม่มีเงื่อนไขเป็นดังนี้

---

[1] ศึกษาการพิสูจน์ได้ในภาคผนวก ก

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1. ประสิทธิภาพของตัวประมาณ $\hat{Y}_{mk1}$, $\hat{Y}_{mk2}$ และ $\hat{Y}_{mk4}$ จะมีค่าสูงสุด เมื่อ N_1 มีค่าน้อย ๆ แต่มีบางกรณีที่ประสิทธิภาพของตัวประมาณดังกล่าวจะเพิ่มขึ้น เมื่อ N_1 เพิ่มขึ้น แต่อย่างไรก็ตาม สามารถที่จะสรุปได้ว่า ประสิทธิภาพของตัวประมาณทั้ง 3 จะลดลง ถ้า N_1 เพิ่มขึ้น

2. เมื่อกำหนดให้ค่า C_2 , δ , f, N และ N_1 คงที่แต่ให้ค่า C_1 เปลี่ยนไปจะได้ว่าประสิทธิภาพของตัวประมาณทั้ง 4 รูปแบบจะเพิ่มขึ้น ถ้า C_1 เพิ่มขึ้น แต่ถ้าให้ค่า C_2 เปลี่ยนไป ประสิทธิภาพของตัวประมาณทั้ง 4 จะลดลง ถ้า C_2 เพิ่มขึ้น

3. ถ้าใช้ขนาดตัวอย่างใหญ่ ประสิทธิภาพของตัวประมาณจะลดลงเรื่อย ๆ

กล่าวโดยสรุป ตัวประมาณที่น่าจะนำไปใช้เมื่อศึกษาในแง่ประสิทธิภาพ คือ $\hat{Y}_{mk3}$ สำหรับ $\hat{Y}_{mk2}$ จะมีประสิทธิภาพต่ำกว่า $\hat{Y}_{mk1}$ ถ้า N_1 มีจำนวนน้อย ส่วน $\hat{Y}_{mk4}$ ประสิทธิภาพอาจต่ำกว่าประสิทธิภาพของตัวประมาณอื่น ๆ ในกรณีที่ N_1 มีจำนวนน้อย เพราะทำให้การสุ่มตัวอย่างแบ่งชั้นภูมิเมื่อเลือกตัวอย่างแล้ว ต่ำกว่าวิธีอื่น ๆ สำหรับประสิทธิภาพของตัวประมาณ เมื่อใช้การอนุมานอย่างมีเงื่อนไข ไมเคิลและคาดาบา มิได้เสนอไว้

นอกจากผลงานวิจัยที่เกี่ยวข้องดังที่ได้กล่าวมาแล้ว ยังได้มีนักสถิติท่านอื่นที่ศึกษาเกี่ยวกับเรื่องนี้อีกอาทิเช่น ทูคี และแมกลาฟลิน (Tukey และ Mclaughlin : 1963)[1] ได้ศึกษาการแจกแจงของประชากรที่มีลักษณะการแจกแจงแบบสมมาตร (Symmetric distribution) แต่มีค่าสังเกตที่เป็นค่าสูงมากร่วมอยู่ด้วยโดยใช้ตัวประมาณ Winsorized sample mean ประมาณค่าเฉลี่ยประชากร ครัว(Crow : 1964)[2] ศึกษาวิธีการถ่วงน้ำหนัก (weighting

---

[1]ศึกษาเพิ่มเติมได้ใน Tukey, John W., and Mclaughlin, D.H. (1963) "Less Vulnerable confidence and Significance Procedures for Location Based on a Single Sample : Trimming/Winsorization 1," Sankhyā, Ser.A, 25, 381-352.

[2]ศึกษาเพิ่มเติมได้ใน Grow, E.K. (1964)." The Statistical Construction of a Single Standard From Serveral Available Standards .

procedures) สำหรับกลุ่มข้อมูลที่มีค่าสูงมาก และ ฟูลเลอร์ (Fuller : 1970)[1] ได้สมมติให้ประชากรมีการแจกแจงแบบเบ้ (Skewed distribution) แล้วใช้ One sided Winsorized mean ประมาณค่าเฉลี่ยประชากร

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

---

[1] ศึกษาเพิ่มเติมได้ใน Fuller, Wayne A. (1970), Simple Estimators for the Mean of Skewed Populations, Technical Report prepared for the U.S. Bureau of the Census, Iowa State University, Dept. of Statistics.