ESTADÍSTICA ESPAÑOLA
Vol. 50, núm. 169, 2008, págs. 605 a 623

Estimación del promedio sobre dos ocasiones cuando hay no-respuesta

por
GARCÍA LUENGO, AMELIA VICTORIA
y
OÑA CASADO, INMACULADA
Departamento de Estadística y Matemática Aplicada. Universidad de Almería.

RESUMEN

Estudiamos el problema de la estimación de la suma de medias bajo un diseño de muestreo sucesivo en dos ocasiones, cuando hay no-respuesta en ambas ocasiones, cuando hay no-respuesta sólo en la primera ocasión y cuando hay no-respuesta sólo en la segunda ocasión.

Obtenemos el porcentaje de pérdida en precisión de estos estimadores sobre el estimador de la suma cuando no hay no-respuesta. Además se calculan los tamaños muestrales y los correspondientes costes esperados para los estimadores propuestos, considerando que éstos tienen igual precisión que el estimador de la suma cuando no hay no-respuesta.

Para finalizar, los resultados obtenidos se demuestran con la ayuda de un estudio empírico, donde se observa que el porcentaje de pérdida en precisión es mayor para el estimador de la suma con no-respuesta en las dos ocasiones, mientras que es menor para el que tiene no-respuesta sólo en la primera ocasión.

Palabras clave: Muestreo sucesivo, No-respuesta, Estimador de la suma.

Clasificación AMS: 62D05.

1. INTRODUCCIÓN

Jessen (1942), Tikkiwal (1951), Yates (1949), Patterson (1950), Eckler (1955) y Raj (1979) contribuyeron a desarrollar la teoría de la estimación de la media en muestreo sucesivo. Hansen y Hurwitz (1946) sugirieron para las encuestas por correo una técnica cuando hay no-respuesta. Estas encuestas tienen la ventaja de que los datos pueden ser recogidos con bajo coste. Sin embargo, la no-respuesta es un problema común en las encuestas por correo. Cochran (1977) y más recientemente Fabian y Hyunshik (2000) extendieron la técnica de Hansen y Hurwitz al caso que estudia la característica principal junto con información auxiliar disponible. En este trabajo utilizamos la técnica de Hansen y Hurwitz para estimar la suma de medias, bajo un diseño de muestreo en ocasiones sucesivas. De esta manera, completamos el trabajo de Choudhary, Bathla y Sud (2004), que estudian la no-respuesta en el muestreo en ocasiones sucesivas.

2. ESTIMADOR DE LA SUMA DE MEDIAS CUANDO HAY NO-RESPUESTA EN LAS DOS OCASIONES

Consideremos una población finita $\Omega = \left(U_1, U_2, ..., U_N\right)$ de N unidades identificables. Sean $\left(x_i, y_i; \quad i = 1, 2, 3, ..., N\right)$ los valores de la característica en la primera y segunda ocasión respectivamente. Asumimos que la población está dividida en dos clases, unos que responderán en el primer intento y otros que no. Sean N_1 y N_2 los tamaños de esas dos clases. En la primera ocasión, las encuestas, mediante el correo, se envían a n unidades seleccionadas por muestreo aleatorio simple. En la segunda ocasión, una muestra aleatoria simple de $m = np$ unidades se retienen mientras una muestra independiente de $\mu = nq = n - m$ unidades se selecciona. Asumimos que en la parte no común de la muestra en las dos ocasiones u_1 unidades responden y u_2 unidades no lo hacen. Igualmente, en la parte común de la muestra en las dos ocasiones m_1 unidades responden y m_2 unidades no lo hacen.

Denotamos por m_{h2} el tamaño de la submuestra extraída de la clase de los que no responden en la parte común de la muestra en las dos ocasiones. Igualmente, denotamos por u_{h2} el tamaño de la submuestra extraída de la clase de los que no responden en la parte no común de la muestra en las dos ocasiones.

σ^2 y σ_2^2 , denotan la varianza poblacional y la varianza poblacional relativa a los que no responden, respectivamente. Igualmente ρ y ρ_2 denotan la correlación entre las unidades que pertenecen a la parte común y la correlación entre las unidades que no responden pertenecientes a la parte común.

Sean $\overline{x}_m^*$ y $\overline{x}_u^*$ los estimadores de la parte común y no común de la muestra en la primera ocasión y $\overline{y}_m^*$ y $\overline{y}_u^*$ los estimadores de la parte común y no común de la muestra en la segunda ocasión. El esquema es el siguiente

Primera ocasión →	$\overline{x}_u^*$	$\overline{x}_m^*$	
Segunda ocasión →		$\overline{y}_m^*$	$\overline{y}_u^*$

donde

$$\overline{y}_m^* = \frac{m_1\overline{y}_{m_1} + m_2\overline{y}_{m_{h2}}}{m}; \qquad \overline{y}_u^* = \frac{u_1\overline{y}_{u_1} + u_2\overline{y}_{u_{h2}}}{u}$$

$$\overline{x}_m^* = \frac{m_1\overline{x}_{m_1} + m_2\overline{x}_{m_{h2}}}{m}; \qquad \overline{x}_u^* = \frac{u_1\overline{x}_{u_1} + u_2\overline{x}_{u_{h2}}}{u}$$

Consideremos el siguiente estimador de la suma de medias

$$z_{12} = a\,\overline{x}_u^* + b\,\overline{x}_m^* + c\,\overline{y}_m^* + d\,\overline{y}_u^*$$

al que imponemos las condiciones de insesgadez y varianza mínima.

$$E(z_{12}) = aE(\overline{x}_u^*) + bE(\overline{x}_m^*) + cE(\overline{y}_m^*) + dE(\overline{y}_u^*) =$$

$$= a\overline{X}^* + b\overline{X}^* + c\overline{Y}^* + d\overline{Y}^* = (a+b)\overline{X}^* + (c+d)\overline{Y}^* = \overline{X}^* + \overline{Y}^*$$

que implica las siguientes condiciones

$$a + b = 1 \quad y \quad c + d = 1 \quad \text{luego} \quad b = 1 - a \quad y \quad d = 1 - c$$

El estimador será

$$z_{12} = a\,\overline{x}_u^* + (1 - a)\,\overline{x}_m^* + c\,\overline{y}_m^* + (1 - c)\,\overline{y}_u^*$$

Tomando varianzas en z_{12}

$$V(z_{12}) = a^2V(\overline{x}_u^*) + (1 - a)^2V(\overline{x}_m^*) + c^2V(\overline{y}_m^*) +$$

$$+ (1 - c)^2V(\overline{y}_u^*) + 2(1 - a)c\,\mathrm{Cov}(\overline{x}_m^*, \overline{y}_m^*)$$

Se puede comprobar que

$$\mathrm{Cov}(\overline{x}_m^*, \overline{x}_u^*) = \mathrm{Cov}(\overline{x}_m^*, \overline{y}_u^*) = \mathrm{Cov}(\overline{y}_m^*, \overline{x}_u^*) = \mathrm{Cov}(\overline{y}_m^*, \overline{y}_u^*) =$$

$$= \mathrm{Cov}(\overline{y}_u^*, \overline{x}_u^*) = 0$$

$$\mathrm{Cov}(\overline{x}_m^*, \overline{x}_m^*) = \mathrm{Var}(\overline{x}_m^*) = \left(\frac{\sigma^2}{m} + \frac{fN_2\sigma_2^2}{Nm}\right)$$

$$\mathrm{Cov}(\overline{x}_u^*, \overline{x}_u^*) = \mathrm{Var}(\overline{x}_u^*) = \left(\frac{\sigma^2}{u} + \frac{fN_2\sigma_2^2}{Nu}\right)$$

$$\mathrm{Cov}(\overline{y}_m^*, \overline{y}_m^*) = \mathrm{Var}(\overline{y}_m^*) = \left(\frac{\sigma^2}{m} + \frac{fN_2\sigma_2^2}{Nm}\right)$$

$$\mathrm{Cov}(\overline{y}_u^*, \overline{y}_u^*) = \mathrm{Var}(\overline{y}_u^*) = \left(\frac{\sigma^2}{u} + \frac{fN_2\sigma_2^2}{Nu}\right)$$

$$\mathrm{Cov}(\bar{y}_m^*, \bar{x}_m^*) = \left(\frac{\rho\sigma^2}{m} + \frac{\rho_2 f N_2 \sigma_2^2}{Nm}\right)$$

$$W_2 = \frac{N_2}{N} \qquad A = fW_2\sigma_2^2 \qquad f = \frac{m_2}{m_{h2}} = \frac{u_2}{u_{h2}}$$

Hallamos a y c para que $V(z_{12})$ sea mínima, para lo cual igualaremos a cero las respectivas derivadas parciales, respecto de a y c, obteniendo

$$a_{opt} = \frac{pq(\sigma^2 + A)(\rho\sigma^2 + \rho_2 A)}{(\sigma^2 + A)^2 - q^2(\rho\sigma^2 + \rho_2 A)^2} + \frac{q\left((\sigma^2 + A)^2 - q(\rho\sigma^2 + \rho_2 A)^2\right)}{(\sigma^2 + A)^2 - q^2(\rho\sigma^2 + \rho_2 A)^2}$$

$$c_{opt} = \frac{p(\sigma^2 + A)^2}{(\sigma^2 + A)^2 - q^2(\rho\sigma^2 + \rho_2 A)^2} - \frac{pq(\sigma^2 + A)(\rho\sigma^2 + \rho_2 A)}{(\sigma^2 + A)^2 - q^2(\rho\sigma^2 + \rho_2 A)^2}$$

Sustituyendo en z_{12}

$$z_{12} = \frac{q\left((\sigma^2 + A)^2 - q(\rho\sigma^2 + \rho_2 A)^2\right)}{(\sigma^2 + A)^2 - q^2(\rho\sigma^2 + \rho_2 A)^2}(\bar{y}_u^* + \bar{x}_u^*) +$$

$$+ \frac{p(\sigma^2 + A)^2}{(\sigma^2 + A)^2 - q^2(\rho\sigma^2 + \rho_2 A)^2}(\bar{y}_m^* + \bar{x}_m^*) +$$

$$+ \frac{pq(\sigma^2 + A)(\rho\sigma^2 + \rho_2 A)}{(\sigma^2 + A)^2 - q^2(\rho\sigma^2 + \rho_2 A)^2}\left[(\bar{x}_u^* - \bar{x}_m^*) + (\bar{y}_u^* - \bar{y}_m^*)\right]$$

Operando, nos queda

$$z_{12} = \frac{p(\sigma^2 + A)}{(\sigma^2 + A) + q(\rho\sigma^2 + \rho_2 A)}(\bar{y}_m^* + \bar{x}_m^*)$$

$$+\frac{q\left(\left(\sigma^2+A\right)+\left(\rho\sigma^2+\rho_2 A\right)\right)}{\left(\sigma^2+A\right)+q\left(\rho\sigma^2+\rho_2 A\right)}\left(\overline{y}_u^* + \overline{x}_u^*\right) \qquad [1]$$

Esta fórmula fue establecida por Yates (1960) y equivale a una media aritmética ponderada de los estimadores, $\overline{y}_m^* + \overline{x}_m^*$, $\overline{y}_u^* + \overline{x}_u^*$, con pesos

$$\frac{p\left(\sigma^2+A\right)}{\left(\sigma^2+A\right)+q\left(\rho\sigma^2+\rho_2 A\right)}, \qquad \frac{q\left(\left(\sigma^2+A\right)+\left(\rho\sigma^2+\rho_2 A\right)\right)}{\left(\sigma^2+A\right)+q\left(\rho\sigma^2+\rho_2 A\right)}$$

que evidentemente suman la unidad.

Si tomamos varianzas en la fórmula del estimador [1], tenemos

$$V(z_{12}) = \frac{\left(\sigma^2+A\right)+\left(\rho\sigma^2+\rho_2 A\right)}{\left(\sigma^2+A\right)+q\left(\rho\sigma^2+\rho_2 A\right)}$$

Se observa que, para $\left(\rho\sigma^2+\rho_2 A\right)/\left(\sigma^2+A\right)>0$, la varianza del estimador de la suma de varianzas es mínima si $q=1$, o sea, la varianza será mínima si las muestras son independientes en ambas ocasiones. En este caso

$$V(z_{12}) = \frac{2}{n}\left(\sigma^2+A\right)$$

En el caso en que $\rho=\rho_2$

$$V(z_{12}) = \frac{2}{n}\left(\sigma^2+A\right)\frac{1+\rho}{1+q\rho}$$

Si $A=0$, es decir, no hay no-respuesta, entonces $V\left(z_{12}\right)$ se reduce a

$$V(z_0) = \frac{2\sigma^2}{n}\,\frac{1+\rho}{1+q\rho}$$

donde z_0 es el estimador de la suma de las medias en el contexto del muestreo en ocasiones sucesivas cuando no hay no-respuesta.

3. ESTIMADOR DE LA SUMA DE MEDIAS CUANDO HAY NO-RESPUESTA EN LA PRIMERA OCASIÓN

Cuando hay no-respuesta solamente en la primera ocasión consideramos el siguiente estimador de la suma de medias

$$z_1 = a\,\overline{x}_u^* + b\,\overline{x}_m^* + c\,\overline{y}_m + d\,\overline{y}_u$$

donde

$$\overline{y}_m = \frac{1}{m}\sum_{i=1}^{m} y_i \quad \text{y} \quad \overline{y}_u = \frac{1}{u}\sum_{i=1}^{u} y_i$$

y al que imponemos las condiciones de insesgadez y varianza mínima.

El estimador será

$$z_1 = a\,\overline{x}_u^* + (1-a)\,\overline{x}_m^* + c\,\overline{y}_m + (1-c)\,\overline{y}_u$$

Tomando varianzas en z_1

$$V(z_1) = a^2V(\overline{x}_u^*) + (1-a)^2V(\overline{x}_m^*) + c^2V(\overline{y}_m) +$$

$$+ (1-c)^2V(\overline{y}_u) + 2(1-a)c\,\mathrm{Cov}(\overline{x}_m^*, \overline{y}_m)$$

$$V(z_1) = a^2\left(\frac{\sigma^2}{qn}+\frac{A}{qn}\right) + (1-a)^2\left(\frac{\sigma^2}{pn}+\frac{A}{pn}\right) + c^2\,\frac{\sigma^2}{pn} +$$

$$+ (1-c)^2\,\frac{\sigma^2}{qn} + 2(1-a)c\,\frac{\sigma^2\rho}{pn}$$

Hallamos a y c para que $V(z_1)$ sea mínima, para lo cual igualaremos a cero las respectivas derivadas parciales, respecto de a y c, obteniendo

$$a_{opt} = \frac{pq\sigma^2\rho}{(\sigma^2 + A) - q^2\rho^2\sigma^2} + \frac{q((\sigma^2 + A) - q\rho^2\sigma^2)}{(\sigma^2 + A) - q^2\rho^2\sigma^2}$$

$$c_{opt} = \frac{p(\sigma^2 + A)}{(\sigma^2 + A) - q^2\rho^2\sigma^2} - \frac{pq(\sigma^2 + A)\rho}{(\sigma^2 + A) - q^2\rho^2\sigma^2}$$

4. ESTIMADOR DE LA SUMA DE MEDIAS CUANDO HAY NO-RESPUESTA EN LA SEGUNDA OCASIÓN

Cuando hay no-respuesta solamente en la segunda ocasión consideramos el siguiente estimador de la suma de medias

$$z_2 = a\,\bar{x}_u + b\,\bar{x}_m + c\,\bar{y}_m^* + d\,\bar{y}_u^*$$

donde

$$\bar{x}_m = \frac{1}{m}\sum_{i=1}^{m} x_i \quad \text{y} \quad \bar{x}_u = \frac{1}{u}\sum_{i=1}^{u} x_i$$

y al que imponemos las condiciones de insesgadez y varianza mínima.

El estimador será

$$z_2 = a\,\bar{x}_u + (1 - a)\,\bar{x}_m + c\,\bar{y}_m^* + (1 - c)\,\bar{y}_u^*$$

Tomando varianzas en z_2

$$V(z_2) = a^2V(\bar{x}_u) + (1 - a)^2V(\bar{x}_m) + c^2V(\bar{y}_m^*) +$$

$$+ (1 - c)^2V(\bar{y}_u^*) + 2(1 - a)c\,\mathrm{Cov}(\bar{x}_m, \bar{y}_m^*)$$

$$V(z_2) = a^2\frac{\sigma^2}{qn} + (1-a)^2\left(\frac{\sigma^2}{pn}+\frac{A}{pn}\right) + c^2\left(\frac{\sigma^2}{pn}+\frac{A}{pn}\right) +$$

$$+(1-c)^2\left(\frac{\sigma^2}{qn}+\frac{A}{qn}\right) + 2(1-a)c\frac{\sigma^2\rho}{pn}$$

Hallamos a y c para que $V(z_2)$ sea mínima, para lo cual igualaremos a cero las respectivas derivadas parciales, respecto de a y c, obteniendo

$$a_{opt} = \frac{pq(\sigma^2+A)\rho}{(\sigma^2+A)-q^2\rho^2\sigma^2} + \frac{q((\sigma^2+A)-q\rho^2\sigma^2)}{(\sigma^2+A)-q^2\rho^2\sigma^2}$$

$$c_{opt} = \frac{p(\sigma^2+A)}{(\sigma^2+A)-q^2\rho^2\sigma^2} - \frac{pq\sigma^2\rho}{(\sigma^2+A)-q^2\rho^2\sigma^2}$$

5. COMPARACIÓN DE LA PRECISIÓN DE LOS ESTIMADORES, z_0, z_{12}, z_1 y z_2

El porcentaje de pérdida en precisión de z_{12}, z_1 y z_2 sobre z_0 vienen dados por

$$L_{12} = \left[\frac{\mathbf{V(z_{12})}}{\mathbf{V(z_0)}} - 1\right] \text{ x } 100; \qquad L_1 = \left[\frac{\mathbf{V(z_1)}}{\mathbf{V(z_0)}} - 1\right] \text{ x } 100$$

$$L_2 = \left[\frac{\mathbf{V(z_2)}}{\mathbf{V(z_0)}} - 1\right] \text{ x } 100$$

El porcentaje de pérdida en precisión de, z_{12}, z_1 y z_2 sobre z_0 para diferentes valores de $\rho, \rho_2, \sigma_2^2, \sigma^2, W_2, f$ y q se indican en los cuadros 1 y 2. Se considera que $N = 300$ y $n = 50$.

En los cuadros 1 y 2, se observa que el porcentaje de pérdida en precisión es mayor para z_{12}, mientras que es menor para z_1.

Cuadro 1

PORCENTAJE DE PÉRDIDA EN PRECISIÓN DE z_{12}, z_1 Y z_2 SOBRE z_0

PARA DIFERENTES VALORES DE ρ, ρ_2, σ_2^2 Y σ^2

ρ	ρ_2	q	f	W_2	σ_2^2	σ^2	L_{12}	L_1	L_2
				$\sigma^2 < \sigma_2^2$					
0,7	0,2	0,7	2,5	0,8	0,4	0,3	247,621	118,064	128,505
0,7	0,2	0,7	2,5	0,8	0,6	0,3	370,831	176,591	191,684
0,7	0,2	0,7	2,5	0,8	0,9	0,3	555,511	264,311	286,312
				$\sigma^2 > \sigma_2^2$					
0,6	0,2	0,3	1,5	0,6	0,2	0,3	50,748	22,866	33,715
0,6	0,2	0,3	1,5	0,6	0,2	0,7	21,819	9,889	14,591
0,6	0,2	0,3	1,5	0,6	0,2	0,9	16,982	7,707	11,373
				$\sigma^2 = \sigma_2^2$					
0,8	0,3	0,7	2,0	0,7	0,1	0,1	131,395	61,769	66,431
0,8	0,3	0,7	2,0	0,7	0,3	0,3	131,395	61,769	66,431
0,8	0,3	0,7	2,0	0,7	0,8	0,8	131,395	61,769	66,431
				$\rho < \rho_2$					
0,1	0,7	0,6	2,5	0,5	0,5	0,4	182,893	75,347	102,030
0,3	0,7	0,6	2,5	0,5	0,5	0,4	170,749	71,334	90,899
0,6	0,7	0,6	2,5	0,5	0,5	0,4	159,047	67,516	79,586
				$\rho > \rho_2$					
0,8	0,1	0,3	2,0	0,5	0,5	0,4	94,708	44,673	63,008
0,8	0,4	0,3	2,0	0,5	0,5	0,4	108,428	44,673	63,008
0,8	0,9	0,3	2,0	0,5	0,5	0,4	128,868	44,673	63,008
				$\rho = \rho_2$					
0,2	0,2	0,8	1,5	0,5	0,5	0,5	75	36,347	41,616
0,5	0,5	0,8	1,5	0,5	0,5	0,5	75	35,377	38,430
0,9	0,9	0,8	1,5	0,5	0,5	0,5	75	34,529	35,604

En concreto, se observa que para $\sigma^2 < \sigma_2^2$, el porcentaje de pérdida en precisión de los tres estimadores sobre aumenta cuando aumentan los valores de σ_2^2.

En el caso de que $\sigma^2 > \sigma_2^2$ el porcentaje de pérdida en precisión de los tres estimadores sobre z_0 disminuye cuando aumentan los valores de σ^2.

Cuando $\sigma^2 = \sigma_2^2$ el porcentaje de pérdida en precisión de los tres estimadores sobre z_0 permanece constante al aumentar los valores de σ^2 y σ_2^2.

Para $\rho < \rho_2$, el porcentaje de pérdida en precisión de los tres estimadores sobre z_0 disminuye cuando aumentan los valores de ρ.

Cuando $\rho > \rho_2$ el porcentaje de pérdida en precisión de z_1 y z_2 sobre z_0 permanece constante al aumentar los valores de ρ_2, mientras que para z_{12} el porcentaje de pérdida en precisión aumenta.

En el caso de que $\rho = \rho_2$ el porcentaje de pérdida en precisión de z_{12} sobre z_0 permanece constante al aumentar los valores de ρ y ρ_2, mientras que para z_1 y z_2 el porcentaje de pérdida en precisión disminuye.

En el cuadro 2 el porcentaje de pérdida en precisión de los tres estimadores sobre z_0 aumenta cuando aumentan los valores de W_2, f y q, exceptuando el caso en que el porcentaje de pérdida en precisión de z_2 sobre z_0, primero disminuye y después aumenta conforme los valores de q aumentan.

Cuadro 2

PORCENTAJE DE PÉRDIDA EN PRECISIÓN DE z_{12}, z_1 Y z_2 SOBRE z_0 PARA DIFERENTES VALORES DE q, f y W_2

ρ	ρ_2	q	f	W_2	σ_2^2	σ^2	L_{12}	L_1	L_2
				W_2					
0,7	0,2	0,6	2,5	0,1	0,4	0,6	15,242	7,298	8,598
0,7	0,2	0,6	2,5	0,3	0,4	0,6	45,543	21,640	25,249
0,7	0,2	0,6	2,5	0,6	0,4	0,6	90,736	42,864	49,624
				f					
0,8	0,3	0,4	1,0	0,5	0,4	0,6	28,958	13,008	16,986
0,8	0,3	0,4	3,0	0,5	0,4	0,6	86,338	38,196	49,496
0,8	0,3	0,4	3,5	0,5	0,4	0,6	100,627	44,417	57,492
				q					
0,8	0,2	0,1	1,5	0,4	0,7	0,5	60,059	25,874	45,205
0,8	0,2	0,5	1,5	0,4	0,7	0,5	72,915	34,000	41,053
0,8	0,2	0,7	1,5	0,4	0,7	0,5	77,860	37,304	40,376
0,8	0,2	0,9	1,5	0,4	0,7	0,5	82,099	40,418	41,028

Realizamos el siguiente estudio, bajo un diseño de muestreo en ocasiones sucesivas, con objeto de economizar en las encuestas realizadas por correo y asumiendo diferentes valores de $\rho, \rho_2, \sigma_2^2, \sigma^2, W_2, f$ y q.

Para ello, consideramos

$$N = 300, \qquad n = 50, \qquad c_o = 1, \qquad c_1 = 4, \qquad c_2 = 45$$

donde

c_o, es el coste por cuestionario mandado por correo.

c_1, es el coste por cuestionario con los resultados de los entrevistados en el primer intento.

c_2, es el coste por cuestionario con los datos recogidos mediante entrevistas personales.

C_{00}, es el coste total invertido en la recogida de datos mediante entrevistas personales, llevadas a cabo en la muestra completa, con lo que se asegura que no haya no-respuesta. Asumimos que el coste invertido en la recogida de datos para la parte común y no común de la muestra es el mismo y también que el coste invertido en la recogida de datos en las dos ocasiones es el mismo. De esta manera, la función coste viene dada por

$$C_{00} = 2nc_2$$

Sustituyendo los valores de n y c_2 en la expresión anterior el coste total es 4.500.

Además, n_1 denota el número de unidades que responden en el primer intento y n_2 denota el número de unidades que no responden.

1. La función coste para el caso en el que hay no-respuesta en las dos ocasiones viene dado por

$$C_{12} = 2\left[c_0 n + c_1 n_1 + \frac{c_2 n_2}{f}\right]$$

El coste esperado viene dado por

$$E(C_{12}) = 2n\left[c_0 + c_1 W_1 + \frac{c_2 W_2}{f}\right]$$

donde $W_1 = \frac{N_1}{N}$ y $W_2 = \frac{N_2}{N}$, con $W_1 + W_2 = 1$

2. La función coste para el caso en el que hay no-respuesta en la segunda ocasión viene dado por

$$C_2 = 2c_0 n + c_1 n + \left[c_1 n_1 + \frac{c_2 n_2}{f}\right]$$

y el coste esperado viene dado por

$$E(C_2) = n\left[2c_0 + c_1(W_1 + 1) + \frac{c_2 W_2}{f}\right]$$

3. La función coste para el caso en el que hay no-respuesta en la primera ocasión viene dado por

$$C_1 = \left[c_1 n_1 + \frac{c_2 n_2}{f}\right] + 2c_0 n + c_1 n$$

y el coste esperado viene dado por

$$E(C_1) = n\left[2c_0 + c_1(W_1 + 1) + \frac{c_2 W_2}{f}\right]$$

Igualando las varianzas de $V(z_{12})$, $V(z_1)$ y $V(z_2)$ respectivamente a $V(z_0)$ y utilizando valores dados de diferentes parámetros, determinamos los tamaños muestrales y el correspondiente coste esperado en la encuesta para z_{12}, z_1 y z_2. Los tamaños muestrales, considerando que los estimadores z_{12}, z_1 y z_2, tienen igual precisión que z_0, se denotan por n', n_1' y n_2'. Los resultados se muestran en los cuadros 3 y 4.

En los cuadros 3 y 4, se observa que el tamaño muestral es mayor en z_{12}, mientras que es menor para z_1.

Cuadro 3

TAMAÑO MUESTRAL Y EL CORRESPONDIENTE COSTE ESPERADO PARA DIFERENTES VALORES DE ρ, ρ_2, σ_2^2 Y σ^2, CONSIDERANDO QUE LOS ESTIMADORES z_{12}, z_1 Y z_2, TIENEN IGUAL PRECISIÓN QUE z_0

ρ	ρ_2	q	f	W_2	σ_2^2	σ^2	n'	n_1'	n_2'	$E(C_{12})$	$E(C_1)$	$E(C_2)$
				$\sigma^2 < \sigma_2^2$								
0,7	0,2	0,5	2,5	0,4	0,4	0,3	109	77	83	2.301,55	1.203,96	1.301,13
0,7	0,2	0,5	2,5	0,4	0,7	0,3	152	97	108	3.223,77	1515,92	1.680,45
0,7	0,2	0,5	2,5	0,4	0,8	0,3	167	104	116	3.530,82	1.619,63	1.806,36
				$\sigma^2 > \sigma_2^2$								
0,6	0,2	0,3	1,5	0,3	0,2	0,3	63	56	58	1.605,68	992,54	1.041,27
0,6	0,2	0,3	1,5	0,3	0,2	0,6	56	53	54	1.443,14	941,54	966,07
0,6	0,2	0,3	1,5	0,3	0,2	0,9	54	52	53	1.388,83	924,43	940,83
				$\sigma^2 = \sigma_2^2$								
0,8	0,3	0,7	2,0	0,5	0,2	0,2	97	72	74	2.764,25	1.388,96	1.423,03
0,8	0,3	0,7	2,0	0,5	0,6	0,6	97	72	74	2.764,25	1.388,96	1.423,03
0,8	0,3	0,7	2,0	0,5	0,9	0,9	97	72	74	2.764,25	1.388,96	1.423,03
				$\rho < \rho_2$								
0,1	0,7	0,6	2,5	0,5	0,4	0,6	99	70	77	2.378,30	1.191,68	1.312,78
0,5	0,7	0,6	2,5	0,5	0,4	0,6	93	68	72	2.239,37	1.162,63	1.229,16
0,8	0,7	0,6	2,5	0,5	0,4	0,6	91	68	70	2.184,70	1.151,01	1.192,21
				$\rho > \rho_2$								
0,8	0,2	0,3	2,0	0,4	0,5	0,3	103	74	83	2.554,08	1.284,02	1.453,85
0,8	0,6	0,3	2,0	0,4	0,5	0,3	112	74	83	2.786,68	1.284,02	1.453,85
0,8	0,9	0,3	2,0	0,4	0,5	0,3	119	74	83	2.944,47	1.284,02	1.453,85
				$\rho = \rho_2$								
0,3	0,3	0,8	1,5	0,3	0,6	0,4	84	66	68	2.144	1.178,07	1.213,57
0,5	0,5	0,8	1,5	0,3	0,6	0,4	84	65	67	2.144	1.173,52	1.198,19
0,8	0,8	0,8	1,5	0,3	0,6	0,4	84	65	66	2.144	1.168,54	1. 180,96

En concreto cuando $\sigma^2 < \sigma_2^2$, el coste disminuye para los tres estimadores al aumentar los valores de σ_2^2. En cuanto a los tamaños muestrales, considerando que los estimadores tienen igual precisión que z_0, aumentan al aumentar los valores de σ_2^2.

Si $\sigma^2 > \sigma_2^2$, el coste aumenta para los tres estimadores al aumentar los valores de σ^2, y los tamaños muestrales, considerando que los estimadores tienen igual precisión que z_0, disminuyen al aumentar los valores de σ^2.

Cuando $\sigma^2 = \sigma_2^2$ el coste permanece constante para los tres estimadores al aumentar los valores de σ^2 y σ_2^2. Asimismo, el tamaño muestral de los tres estimadores, considerando que tienen igual precisión que z_0, permanece constante.

Si $\rho < \rho_2$, el coste aumenta para los tres estimadores al aumentar los valores de ρ, y los tamaños muestrales, considerando que los estimadores tienen igual precisión que z_0, disminuyen al aumentar los valores de ρ.

Si $\rho > \rho_2$, el coste permanece constante para z_1 y z_2 al aumentar los valores de ρ_2, mientras que para z_{12} el coste disminuye. Asimismo, los tamaños muestrales de z_1 y z_2, considerando que tienen igual precisión que z_0, permanecen constante, mientras que el tamaño muestral de z_{12}, considerando que tiene igual precisión que z_0, aumenta.

Cuando $\rho = \rho_2$, el coste permanece constante para z_{12} al aumentar los valores de ρ y ρ_2, mientras que para z_1 y z_2 el coste aumenta. Asimismo, los tamaños muestrales de z_1 y z_2, considerando que tienen igual precisión que z_0, disminuyen, mientras que el tamaño muestral de z_{12}, considerando que tiene igual precisión que z_0, permanece constante.

Se observa en el cuadro 4 que los tamaños muestrales de los tres estimadores, considerando que tienen igual precisión que z_0, aumentan al aumentar los valores de W_2, f y q, exceptuando el caso del tamaño muestral de z_2, considerando que tiene igual precisión que z_0, que primero disminuye y después aumenta conforme aumentan los valores de q. Además, el coste de los tres estimadores aumenta cuando los valores de f aumentan y disminuye cuando los valores de W_2 y q

aumentan, exceptuando el caso del estimador z_2, cuyo coste primero aumenta y después disminuye cuando los valores de q aumentan.

Cuadro 4

TAMAÑO MUESTRAL Y EL CORRESPONDIENTE COSTE ESPERADO PARA DIFERENTES VALORES DE q, f Y W_2 CONSIDERANDO QUE LOS ESTIMADORES z_{12}, z_1 Y z_2, TIENEN IGUAL PRECISIÓN QUE z_0

ρ	ρ_2	q	f	W_2	σ_2^2	σ^2	n'	n_1'	n_2'	$E(C_{12})$	$E(C_1)$	$E(C_2)$
				W_2								
0,7	0,2	0,6	2,5	0,2	0,4	0,6	65	57	58	1.017,25	732,79	748,72
0,7	0,2	0,6	2,5	0,6	0,4	0,6	95	71	75	2555,86	1.314,36	1.376,54
0,7	0,2	0,6	2,5	0,8	0,4	0,6	110	78	83	3.576,47	1.663,42	1.756,37
				f								
0,8	0,3	0,4	1,0	0,5	0,4	0,6	64	56	58	3.288,45	1.723,37	1.784,04
0,8	0,3	0,4	1,5	0,5	0,4	0,6	72	60	63	2.580,31	1.372,77	1.440,15
0,8	0,3	0,4	3,0	0,5	0,4	0,6	93	69	75	1.956,55	1.071,02	1.158,59
				q								
0,8	0,2	0,2	1,5	0,4	0,7	0,5	82	62	72	2.521,48	1.260,25	1.467,63
0,8	0,2	0,7	1,5	0,4	0,7	0,5	89	69	70	2.739,05	1.400,50	1.431,83
0,8	0,2	0,9	1,5	0,4	0,7	0,5	91	70	71	2.804,33	1.432,26	1.438,48

REFERENCIAS

CHOUDHARY, R.K. (2003), «Use of Hansen and Hurwitz technique in successive sampling», *M. Sc. Thesis submitted to P. G. School, I.A.R.I., New Delhi*.

CHOUDHARY, R.K, BATHLA, H.V.L AND SUD, U.C. (2004), «On non-response in sampling on two occasions», *Journal of the Indian Society of Agricultural Statistics*, 58(3), 331-343.

COCHRAN, W. G. (1977), «Sampling Techniques», third edition, *John Wiley* & *Sons*, *New York.*

ECKLER, A. R. (1955), «The Annals of Mathematical Statistics», Rotation Sampling 26, 664--685.

FABIAN, C.O. AND HYUNSHIK, L. (2000), «Double sampling for ratio and regression estimation with sub-sampling the non-respondents», *Survey Methodology*, 26(2), 183-188.

GARCÍA LUENGO, A. V. (2001), «Mejora de estimadores en muestreo en ocasiones sucesivas», *Servicio de Publicaciones de la Universidad de Almería.*

GARCÍA LUENGO, A. V. Y OÑA CASADO I. (2007), «Encuestas continuas: estimación de parámetros en muestreo sucesivo», *Servicio de Publicaciones de la Universidad de Almería*.

HANSEN, M. H. AND HURWITZ, W. N. (1946), «The problem of the non-response in sample surveys», *J. Amer. Statist. Assoc.*, 41, 517-529.

JESSEN, R. J. (1942) «Statistical Investigation of a Sample Survey for Obtaining Farm Facts», *Iowa Agricultural Experiment Statistical Research Bulletin*, 304.

PATTERSON, H. D. (1950), «Sampling on Successive Occasions with Partial Replacement of Units», *Journal of the Royal Statistical Society*, B12, 241-255.

RAJ, D. (1979), «Sampling Theory», *Tata McGraw Hill*, *New Delhi*, 152-162.

RAO, C.R. (1952), «Some Theorems on Minimum-Variance Unbiased Estimation», *Sankhya* 12, 27-42.

TIKKIWAL, B. D. (1951), «Theory of Successive Sampling», *Thesis for diploma I.C.A.R.*, *New Delhi*.

YATES, F. (1949), «Sampling Methods for Censuses and Surveys», *Griffin, London*.

ESTIMATION OF SUM IN SAMPLING ON TWO OCCASIONS WHEN THERE IS NON-RESPONSE

ABSTRACT

The problem of estimation of sum in mail surveys for the current occasion in the context of sampling on two occasions is attempted when there is non-response on both the occasions, when there is non-response only on first occasion, and when there is non-response only on the second occasion.

We obtain the percentage loss in precision of all the estimators over the estimator of sum when there is no non-response. Sample sizes and saving in cost for all the estimators which give equal precision that the estimator of sum when there is no non-response are derived.

The results obtained are demonstrated with the help of an empirical study, which reveals that the percentage loss in precision is maximum for the estimator of sum when there is non-response on both the occasions, while it is least for the estimator of sum when there is non-response only on first occasion.

Key words: Successive sampling, Non-response, Estimator of sum.

AMS Classification: 62D05.