QÜESTIIÓ, Vol 12, nº 2, pp. 145-148, 1988

SESGO DE NO RESPUESTA REITERADA EN LA ESTIMACIÓN DE LA VARIANZA POBLACIONAL

M. RUIZ ESPEJO y J. SANTOS PEÑAS

Universidad Complutense de Madrid

Obtenemos la expresión del sesgo de respuesta en el n-ésimo intento, dentro de la subpoblación que no respondería en 1,2,... y n-1 intentos sucesivos, en estudios por muestreo con reiteración en la subpoblación de no respuesta, para estimar la varianza poblacional.

Clasificación AMS: 62 D 05

Non-response bias in the n-trial for to estimate the population variance.

Keywords: Reitered sampling, Non-response, Bias, Population variance.

1. INTRODUCCIÓN

Recientemente ha sido tratado el problema del cálculo del sesgo de no respuesta en el intento n (Ruiz, 1986) para estimar la media poblacional en el muestreo de poblaciones finitas, generalizando el resultado clásico dado por diferentes autores del sesgo del estimador media muestral de las respuestas para inferir sobre la media poblacional en muestreo sin reiteración.

Siguiendo una lógica similar a la presentada en dicho artículo es posible calcular el sesgo de no respuesta en el intento n para estimar la varianza poblacional, a partir de una aproximación dada por Sukhatme y Sukhatme (1970). Sin embargo, la formulación que aparece es mucho más compleja según veremos.

La fórmula conocida de partida, para su generalización a n intentos de captar la respuesta de la población es

–M. Ruiz Espejo i J. Santos Peñas - Dept. d'Organització d'Empreses. Facultat de Ciències Econòmiques i Empresarials - Universitat Complutense - 28023 Madrid

–Article rebut el juliol de 1988.

(1) $$E\left(s_1^2\right) - S^2 \cong P_2\left(S_1^2 - S_2^2\right) - P_1 P_2 \left(\bar{y}_1 - \bar{y}_2\right)^2$$

siendo S^2 la varianza poblacional, P_1 y P_2 las proporciones de unidades que responden ("1") y que no responden ("2") respectivamente, que pueden ser estratificadas en dichos tamaños relativos. Con tal postestratificación, S_h^2 e $\bar{y}_h$ son la varianza y la media del estrato $h(= 1, 2)$. La fórmula (1) es exacta cuando el diseño muestral empleado es el muestreo aleatorio simple, y

$$s_1^2 = \frac{(N_1 - 1) \sum_{i=1}^{n_1} (y_{1i} - \bar{y}_{1s})^2}{N_1(n_1 - 1)}$$

siendo N_1 el tamaño del estrato de respuesta, $N_1 = NP_1$ con N el tamaño de la población finita; N_1 es desconocido, salvo que el diseño empleado sea censal. Además, si N es suficientemente grande entonces $(N_1 - 1)/N_1$ es aproximadamente 1 con lo que S_1^2 es prácticamente la cuasivarianza muestral de respuestas.

De cualquier modo, si N es grande y s_1^2 es la varianza muestral, (1) es una buena aproximación del sesgo en el primer intento según indicaron Sukhatme y Sukhatme (1970, cap. X, (77)).

Suponiendo, por tanto, que $E\left(s_1^2\right) = S_1^2$ y dando entonces por válida la aproximación (1), podemos generalizar tal relación al caso en que se intenta obtener respuesta por segunda vez en el estrato de no respuesta en el primer intento. Así obtenemos

(2) $$S_{12}^2 - S_{21}^1 = P_{22}\left(S_{12}^2 - S_{22}^2\right) - P_{12}\, P_{22}\left(\bar{y}_{12} - \bar{y}_{22}\right)^2$$

conservando la misma notación en los subíndices que en el artículo pionero de Ruiz (1986).

El tamaño relativo $P_2 P_{22}$ de la población, que no contesta en el primer ni en el segundo intento, es de nuevo interrogado por tercera vez, obteniendo la relación siguiente

(3) $$S_{13}^2 - S_{22}^2 = P_{23}\left(S_{13}^2 - S_{23}^2\right) - P_{13}\, P_{23}\left(\bar{y}_{13} - \bar{y}_{23}\right)^2$$

con la notación ya indicada.

Del mismo modo, para el intento n-1, generalizando (2) y (3) tenemos

$$
\begin{aligned}
S^2_{1(n-1)} - S^2_{2(n-2)} &\doteq P_{2(n-1)}\left(S^2_{1(n-1)} - S^2_{2(n-1)}\right) \\
&-P_{1(n-1)}P_{2(n-1)}\left(\bar{y}_{1(n-1)} - \bar{y}_{2(n-1)}\right)^2
\end{aligned} \tag{4}
$$

y en el intento n,

$$
S^2_{1n} - S^2_{2(n-1)} = P_{2n}\left(S^2_{1n} - S^2_{2n}\right) - P_{1n}\, P_{2n}\left(\bar{y}_{1n} - \bar{y}_{2n}\right)^2 \tag{5}
$$

aplicando el principio de inducción de n.

2. SESGO DE NO RESPUESTA EN LA APLICACIÓN n

En el intento de respuesta n-ésimo, tenemos el estimador s^2_{1n} de S^2_{1n}, cuyo sesgo puede calcularse de esta manera

$$
\begin{aligned}
S^2_{1n} - S^2 &= \left(S^2_{1n} - 2_{1(n-1)}\right) + \left(S^2_{1(n-1)} - S^2_{1(n-2)}\right) + \ldots \\
&+ \left(S^2_{12} - S^2_1\right) + \left(S^2_1 - S^2\right) = \sum_{k=1}^{n}\left(S^2_{1k} - S^2_{1(k-1)}\right)
\end{aligned} \tag{6}
$$

donde hemos denotado $S^2_{11} = S^2_1$ y $S^2_{10} = S^2$.

Ahora, de (4) tenemos que

$$
\begin{aligned}
S^2_{2(n-1)} = S^2_{1(n-1)} &- \frac{1}{P_{2(n-1)}}\left(S^2_{1(n-1)} - S^2_{2(n-2)}\right) \\
&-P_{1(n-1)}\left(\bar{y}_{1(n-1)} - \bar{y}_{2(n-1)}\right)^2
\end{aligned}
$$

y de (5) finalmente tenemos

$$
\begin{aligned}
S^2_{1n} - S^2_{1(n-1)} &= P_{2n}\left(S^2_{1n} - S^2_{2n}\right) - P_{1n}\, P_{2n}\left(\bar{y}_{1n} - \bar{y}_{2n}\right)^2 \\
&-\frac{1}{P_{2(n-1)}}\left(S^2_{1(n-1)} - S^2_{2(n-2)}\right) \\
&-P_{1(n-1)}\left(\bar{y}_{1(n-1)} - \bar{y}_{2(n-1)}\right)^2
\end{aligned} \tag{7}
$$

donde $n = 1, 2, ...$ y por ello quedan concretados los sumandos elementales de (6).

3. DISCUSIÓN

Aunque la relación (6) es válida siempre y puede ser explicada más explícitamente sustituyendo en cada sumando la expresión (7), dicha relación es especialmente práctica cuando el muestreo reiterado se comienza con un censo y posteriormente nos interesamos en las sucesivas subpoblaciones de no respuesta, como hemos explicado. También se pueden extraer conclusiones de la tendencia de S_{1k}^2 cuando $k = 1, 2, ..., n, ...$ (recorre los sucesivos intentos), y esto hace posible estudiar el efecto que tiene la no respuesta persistente sobre la variable de interés y.

Partiendo de un censo en el primer intento y siguiendo como hemos indicado, disponemos de la relación aproximada

$$E\left(s_{1n}^2\right) - S^2 \cong S_{1n}^2 - S^2$$

expresión obtenida en (6) y más explicitada con (7). Conviene observar que debido a la definición de varianza muestral s_{1n}^2, su esperanza es sólo aproximadamente igual a S_{1n}^2. Permanecería S_{2n}^2 como único parámetro desconocido tras los n intentos de obtener respuesta. En este caso podría obtenerse una estimación de S_{2n}^2 en el sentido propuesto por Hansen y Hurwitz (1946), lo cual nos permitiría completar estimativamente la tendencia de S_{1k}^2, con $S_{1(n+1)}^2 = S_{2n}^2$ estimado por el método de los autores americanos del modo $s_{1(n+1)}^2 = \hat{S}_{2n}^2$.

BIBLIOGRAFÍA

[1] **Hansen, M.H.** y **Hurwitz, W.N.** (1946) "The problem of nonresponse in sample surveys." J. Amer Statist. Assoc. 41, 517-529.

[2] **Ruiz, M.** (1986). "Sesgo de no respuesta en el intento n". Estadíst. Española 112-113, 75-78.

[3] **Sukhatme, P.V.** y **Sukhatme, B.V.** (1970). "Sampling Theory of Surveys with Applications". Ames: Iowa State University Press.