Efectos del redondeo de los factores de expansión en los estimadores de media y varianza en el muestreo aleatorio simple

Alberto M. Padilla Terán

Consultor independiente

1 Antecedentes

- En el muestreo probabilístico se emplean los **ponderadores muestrales** o factores de expansión para construir las estimaciones de promedios y varianzas, entre otras cantidades.
  - Ejemplos: ENOE, ENIGH, ENVIPE, entre otras.
- Dichos **ponderadores** pueden incluir.
  - Ajustes por no respuesta.
  - Ajustes a totales conocidos de población.
  - Calibración con información de variables auxiliares correlacionadas positivamente con las variables de interés de la encuesta-
  - Redondeo de los factores de expansión a números enteros.
- El **uso del redondeo** en los ponderadores se efectúa en encuestas relevantes para la toma de decisiones como :
  - ENOE (Encuesta Nacional de Ocupación y Empleo) levantada por el INEGI –encuesta trimestral.
  - La encuesta mensual del IMEF (Instituto Mexicano de Ejecutivos de Finanzas) con la cual se construyen los Indicadores IMEF manufactureros y no manufactureros.
    - Estos indiicadores sirven para anticipar la dirección de la actividad económica, véase Heath y Domínguez
- En los documentos metodológicos de estas dos encuestas no se describe el método de redondeo empleado ni las razones para su uso.

2 Objetivo

En esta presentación se mostrará que el estimador del promedio muestral bajo muestreo aleatorio simple es sesgado al emplear el redondeo de los ponderadores y se ejemplificará para algunas poblaciones.

- También se comentará el efecto sobre el muestreo sistemático con iguales probabilidades de selección.

3 Diseños muestrales empleados

Se analizarán dos diseños muestrales ampliamente usados en la práctica, véase Cochran o Särndal et al.:

- Muestreo aleatorio simple sin reemplazo, ***mas***.
- Muestreo sistemático con igual probabilidad de selección de las unidades, ***ms***.
  - Se ejemplificará con el muestreo sistemático circular.

Notación

- **N** = total de elementos en población
- **n** = total de elementos en muestra
- **ω** = N/n = factor de expansión **sin redondear**
- $\boldsymbol{\omega_{r,inf}}$ = [ω] - 1 = factor de expansión **redondeado hacia abajo,** [·] es la función entero mayor
- $\boldsymbol{\omega_{r,sup}}$ = [ω] = factor de expansión **redondeado hacia arriba**
- **r** = residuo de la división de N por n
- **c** = cociente de la división de N por n

5 Fórmula para evaluar los efectos del redondeo en factores de expansión o ponderadores

Debido a que el manejo de la función entero [ ] no es útil para el desarrollo deseado, obsérvese que N puede expresarse como:

N = n c + r

De esta expresión y la notación se puede ver que c = $\omega_{r,inf}$, por lo cual:

N = n $\omega_{r,inf}$ + r

A partir de esto, los valores de los factores de expansión redondeados se escriben como:

$\boldsymbol{\omega_{r,inf}}$ = (N-r)/n y $\boldsymbol{\omega_{r,sup}}$ = (N-r)/n+1

Con esta definiciones, sólo falta **determinar el número de elementos en muestra** que se redondean hacia arriba y los que se redondean hacia abajo, de tal forma que su suma sea **n**.

- $\mathbf{n_{inf}}$, denotará el número de elementos en muestra redondeados hacia abajo
- $\mathbf{n_{sup}}$, el número de elementos redondeados hacia arriba.

La suma de los dos debe ser **n**: $n = n_{inf} + n_{sup}$

7 Determinación de tamaños de muestra redondeados

- Los valores de $n_{inf} + n_{sup}$ pueden obtenerse solucionando el siguiente sistema de ecuaciones lineal:

$$n_{inf} + n_{sup} = n$$
$$\omega_{r,inf}\ n_{inf} + \omega_{r,sup}\ n_{sup} = N$$

solución:

$$\mathbf{n_{inf} = (N - n\ \omega_{r,sup}) / (\omega_{r,inf} - \omega_{r,sup})}$$
$$\mathbf{n_{sup} = n - n_{inf}}$$

8 Estimador del total *mas* con efectos del redondeo

- Se emplea el estimador de Horvitz-Thompson del total, véase Särndal et al. (1992) para una variable y_k.
- El **factor de expansión** para el ***mas*** es N/n.
- El estimador del total es: $\hat{y} = \sum_k \frac{N}{n} y_k$
- El índice corre sobre todos los elementos de la muestra.
- El estimador del total, usando los factores de expansión redondeados, se construye con los factores redondeados hacia arriba (índices en *r1*) y abajo (índices en *r2*) como:
- $\hat{y} = \sum_{k\in r1} \omega_{k,r,inf} y_k + \sum_{k\in r2} \omega_{k,r,sup} y_k$
- Para obtener el **total estimado redondeado** bajo ***mas***, $\hat{y}_{red}$, se sustituyen los valores de $\boldsymbol{\omega_{r,inf}}$ y $\boldsymbol{\omega_{r,sup}}$ en $\hat{y}$

$$\hat{y}_{red} = \frac{\boldsymbol{N-r}}{\boldsymbol{n}} \sum_k \boldsymbol{y_k} + \sum_{k\in r2} \boldsymbol{y_k}$$

- El índice *k* corre sobre todos los elementos de la muestra.
- Nótese que la segunda parte del estimador sólo depende de los valores de la variable y_k en la parte de la muestra que se redondeó hacia arriba. La primera parte tiene un efecto del residuo de la división N/n.
- El valor esperado bajo ***mas*** es: $E(\hat{y}_{red}) = \left(1 - \frac{r}{N}\right) y_U + E(\sum_{\boldsymbol{k\in r2}} \boldsymbol{y_k})$
- Es sesgado y la segunda parte del valor esperado depende de los valores que se redondean hacia arriba.

8 Ejemplos

Ejemplo 1: Población: en una población con ***N=8*** elementos y valores y_k igual a {3,34,29,36,43,31,20,17} y se extraen todas las posibles muestras bajo ***mas*** de tamaño ***n=3*** que son 56.

En este ejemplo: *f=N/n=2.67*, *r=2*, $\omega_{r,inf}$ = 2, $\omega_{r,inf}$ = 3, n_{inf} = 1, n_{sup} = 2. Se calculan, para todas las posibles muestras los posibles redondeos; es decir; el redondeo hacia abajo en el primer elemento de la muestra (denotado por red_1); el redondeo en el segundo elemento de muestra (red_2) y en e tercero (red_3) y se comparan con el estimador sin redondeo (red_0), que emplea *f=N/n*.

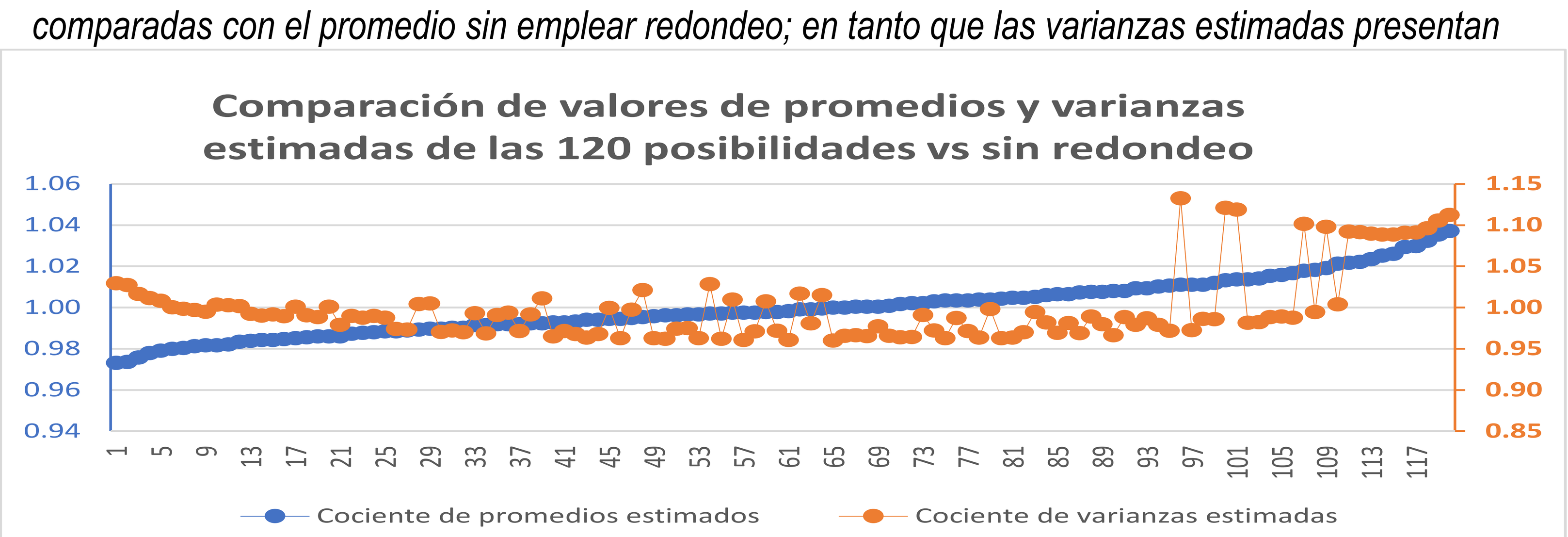
- *Los valores poblacionales son:*
  - ***Promedio** = 26.625*
  - ***S²** = 161.411*
  - ***Varianza** del estimador del **promedio** = 33.627*
- *En la tabla a la derecha se aprecia que el redondeo tiene efectos leves en el promedio y grandes en la varianza.*

	red_0	*red_1*	*red_2*	*red_3*
cv(promedio)	22%	19%	24%	23%
promedio estimado	26.625	27.188	25.775	26.913
varianza entre promedios	33.627	27.730	38.762	38.311
Comparación de promedios y varianzas del redondeo vs sin redondeo				
promedio red_i/red_0		2%	-3%	1%
varianza red_i/red_1		-18%	15%	14%

Nota: este ejercicio se realizó para el muestreo sistemático circular en el que hay ocho muestras y con los posibles redondeos las desviaciones de varianzas poblacionales fueron: -1%, 28% y 1%. Los promedios no cambian por el balance de las unidades en muestra en los subconjuntos que se redondean hacia arriba y abajo..

Ejemplo 2, *Ejercicio para estimar el total de legisladores conservadores en los concejos municipales de Suecia con una muestra pequeña del libro de* Särndal et al. *(1992), ejemplo 4.2.1, página 129. Es una muestra aleatoria simple de conglomerados de una etapa, de tamaño n=16 con N=50. El valor estimado del total es 2,347 y el estimador de varianza es 62,312, véase la página 130 de de* Särndal et al. *(1992).*

- *En este ejercicio, se tiene que f=N/n= 3.125, r=2,* $\omega_{r,inf}$ = 3, $\omega_{r,inf}$ = 4, n_{inf} = 14, n_{sup} = 2. *Se compararán los esimadores del promedio y varianza con la muestra fija, pero calculando todas las combinaciones de redondeo, las cuales son los subconjuntos de tamaño 2 de 16, lo cual da 120 posibles tipos de redondeo.*
- *Como se tienen 120 valores, se muestran dos gráficas de los estimadores del promedio y varianza comparados con los obtenidos por* Särndal et al. *(1992). Los valores se ordenaron por promedio estimado de manera creciente.*
- *En la gráfica se aprecia que los valores estimados del promedio presentan desviaciones de hasta 4% comparadas con el promedio sin emplear redondeo; en tanto que las varianzas estimadas presentan*



Conclusiones

- Se construyó una expresión mostrando que el estimador del total con redondeo bajo ***mas*** es sesgado y se ejemplificó en dos poblaciones pequeñas (podrían formar parte de un estrato en una población más grande).
- En los ejemplos se aprecia el efecto de desviación en la varianza estimada con redondeo comparada con la del ***mas***. Está en desarrollo el desarrollo analítico de la expresión para la varianza con efectos de redondeo.
- En el caso de muestreo sistemático se mencionó que hay efectos en la estimación de la varianza del total; empero, no hay afectación en el estimador del promedio porque los elementos en muestra aparecen el mismo número de veces en los subconjuntos que se redondean hacia arriba y abajo.

Bibliografía

[1] Cochran, W.G. (1977), Sampling Techniques, 3rd edn. New York: Wiley.
[2] Heath, J. y Domínguez, L., Marco Conceptual y Metodológico del Indicador del Entorno Empresarial Mexicano IMEF, Instituto Mexicano de Ejecutivos de Finanzas, México, (sin fecha en la nota).
[3I INEGI, Encuesta Nacional de Ocupación y Empleo, ENOE. (2007), Cómo se hace la ENOE. Métodos y Procedimientos.
[4] Murthy, M.N. & Rao, T.J. (1988) Systematic Sampling, Chapter 7 in Handbook of Statistics 6: Sampling, ed. by C.R. Rao, Amsterdam, North Holland.
[5] Padilla, A. (2009) "An Unbiased Estimator of the Variance of Simple Random Sampling using Mixed Random-Systematic Sampling", Banco de México, Documento de Investigación No. 2009-13.
[6] Särndal, C.E., Swensson, B. & Wretman, J.H., Model Assisted Survey Sampling, Springer-Verlag, New York, 1992.

Contacto

betoampt@gmail.com