APLICACIONES DE MODELOS DE SUPERPOBLACIÓN A LA ASIGNACIÓN ÓPTIMA

Rubén Belmonte Coloma

Un método muy conocido en el muestreo de poblaciones finitas estratificado para asignar tamaños de muestra a los estratos es el método de Neyman. Este método esta destinado a poblaciones fijas, en la actualidad existe muchos estudios sobre muestreo en superpoblaciones, en este artículo se encara este método.

1.- MODELO SUPERPOBLACIONAL

El uso de modelos superpoblacionales se remonta a trabajos desarrollados por los fundadores de la Teoría del Muestreo. Cochran (1939) la utilizó, Godambe (1955) recomendó su uso para hacer inferencias al considerar que Y es generado por un mecanismo aleatorio denominado superpoblación. Este enfoque es considerado como parcialmente Bayesiano pues a pesar de utilizar como principio el inferir a partir de una distribución a priori no busca la minimización del riesgo aposteriori. Un modelo popular es el dado por asumir que una variable conocida X permite establecer la familia a la que pertenece la distribución de Y que a su vez es caracterizada por un modelo superpoblacional. El modelo para la estratificación es:

$$Y_{hj} = \alpha_{h0} + \alpha_{h1} X_{hj} + \varepsilon_{hj}, j = 1, \ldots, N_h, h = 1, \ldots, H$$

Donde $E_{\mathfrak{P}}(\varepsilon_{hj}) = 0$

$$Cov_{\mathfrak{P}}(\varepsilon_{hj}, \varepsilon_{h'j}{}') = \begin{cases} \sigma_h^2 g(X_{hj}) & si \quad h \neq h' \quad y \quad j \neq j' \\ \rho_{\prime h} \sigma_h^2 \sqrt{g(X_{hj}) g(X_{hj'})} & si \quad h = h' \quad y \quad j \neq j' \\ 0 \; en \; otro \; caso & \end{cases}$$

α_{ho} α_h y g (X_{hj})son estrictamente positivos y_ρ es el coeficiente de correlación entre X y Y.
Simplificaremos este modelo haciendo $\rho_h = \alpha_{ho}=0$ y $g(X_{hj}) = 1$. Esto permite modelar problemas en los que se espera una expansión de Y como función de X.
En estudios del crecimiento en que X mide los resultados 'antes' y la Y 'después' por ejemplo.
Se espera suavizar la restricción de insesgamiento de mu por la de que esta se cumpla respecto al modelo. Entonces tendremos el problema de optimización

$$Min\left\{\sum_{h=1}^{H} \beta_h^2 \frac{\sigma_h^2}{n_h} \middle| \sum_{h=1}^{H} \beta_h \mu_{X(h)} - \mu_X = 0,\ \beta_h > 0, h = 1, \ldots, H\right\}$$

Los pesos óptimos se definen convenientemente a partir de los resultados de X y

$$\beta_h^* = \frac{Q_h^*}{\sum_{h=1}^{H} Q_h^*}$$

donde

$$Q_h^* = \frac{n_h \mu_{X(h)} / \sigma_h^2}{\sum_{h=1}^{H} n_h \mu_{X(h)} / \sigma_h^2}$$

y el error está dado por

$$V(m_\beta) = \sum \frac{\beta_h^{*2}\sigma_h^2}{n_h}$$

Si se acepta que E(A/B)≅E(A)/E(B) (linearización de Taylor) tendremos que $E_{\mathfrak{P}}(Q_h) \cong Q_h^*$ y $E_{\mathfrak{P}}(\beta_h) \cong \beta_h^*$

De ahí que

$$E\big(E(m_\beta)\big) \cong E\left(\sum_{h=1}^{H} \beta_h \mu_h\right) \cong \sum_{h=1}^{H} \beta_h^* \mu_h$$

por lo que

$$\sum_{h=1}^{H} \beta_h^* m_h$$

es aproximadamente insesgado

2.- COMPARACIÓN DE LOS PESOS OPTIMOS

En muchas aplicaciones los estratos están definidos a partir de intervalos de una variable continua. Tal es el caso cuando se usan variables como el área de parcelas, los ingresos de las familias, etc. En ellos es muy común que la media y la varianza crezcan en forma no proporcional y que una medida relativa sea más adecuada para medir la precisión de los estimadores.

Población fija

Analizamos algunos ejemplos de los libros de texto para fijar el comportamiento de este método. En algunso casos lo datos considerados como meustrales son de una población artificial.

Se analiza la ganancia en precisión debida al uso del método propuesto respecto al uso de la Asignación de Neymann, y la proporcional $n_h^p = nW_h$. Para cualcular las ponderaciones óptimas son usadas n_h^0 y n_h^P

Ejemplo

Tabla 1 Parámetros del Ejemplo

h	w_h	s_h^2	c_h	n_h^0	μ_h	Q_h^0	β_h^0
1	0,5	3,2	1	2	5	3,1	0,6
2	0,5	14,8	4	2	15	2,39	0,4

Tabla 2 Eficiencia de las Ponderaciones Óptimas en el Ejemplo

Criterio	Varianza	V(mβ)/Varianza
Asignación de Neymann	1,50	1,50
Asignación Proporcional	1,50	1,50
Ponderación Optima	1,76	1,00

En este caso el suso del método prouesto de las ponderaciones óptimas (PO) tiene un comportamiento peor que la asignación de Neymann (AN) y que el de la proporcional (AP). Si tuviéramos que μ1=15 y μ2=5 los resultados cambian y por tanto son otros los valores de los parámetros.

Tabla 3 . Parámetros al variar los valores de la media en el Ejemplo.

H	W_h	s_h^2	μ_h	Q_h^0	β_h^0
1	0,5	3,2	15	9,40	0,92
2	0,5	14,8	5	0,78	0,08

Ahora hay una mayor eficiencia al tener una relación diferente entre varianza y media: el método es mejor que la AP.

Tabla 1.4. Eficiencia del método de las proporciones óptimas con los parámetros variados para el Ejemplo:

Criterio	Varianza	$V(m_\beta)$/Varianza
Asignación de Neymann	1,50	1,17
Asignación Proporcional	2,75	0,64
Ponderación Optima	1,40	1,00

BIBLIOGRAFIA

ALLENDE,S. and C. BOUZA (1993): "Stochastic Programming approaches for the estiamtion of the mean in stratified populations", Inv. Operacional, **13,109-118.**

COCHRAN, W. (1977) : **Sampling Techniques,** Willey, New York.