EL USO DEL MÉTODO DE EXPONENCIACIÓN PARA ESTIMAR PREVALENCIAS

Antonio Arcos[*1], Juan Francisco Muñoz,[**2], Encarnación Álvarez,[**3] y Maria del Mar Rueda[*4]
[*]Departamento de Estadística e Investigación Operativa. Universidad de Granada. C.P.18071.Spain.
[**]Departamento de Métodos Cuantitativos para la Economía y la Empresa. Universidad de Granada. C.P.18071. Spain.

ABSTRACT

The problem of estimating a population proportion is an important topic that has numerous applications to many areas. For example, application can be found in the health field with the problem of estimating prevalences. Recently, many papers about the estimation of a population proportion and based upon auxiliary information have been published, and the ratio, difference and regression techniques were used. This work contributes to the literature by defining exponentiation type estimators for a population proportion. Using data extracted from the Spanish National Health Survey, the proposed estimator is numerically compared to alternative estimators. The aim in this study was to estimate prevalences for the diseases asthma and allergy. Results derived from this study reveal that the proposed exponentiation estimator can be more efficient than alternative estimators.

KEY WORDS: proportion, auxiliary information, inclusion probabilities.

RESUMEN

El problema de la estimación de una proporción poblacional es un tema de gran interés que tiene numerosas aplicaciones en distintos ámbitos, tal como en el campo de la salud a través de la estimación de prevalencias. Recientemente se han publicado numerosos trabajos donde se planteaban estimar una proporción poblacional utilizando información auxiliar en la etapa de estimación, y para ello se utilizaron estimadores de tipo razón, diferencia y regresión. Este trabajo contribuye a esta literatura mediante el desarrollo de estimadores de tipo exponenciación para una proporción poblacional. Utilizando datos de la Encuesta Nacional de Salud Española, el estimador propuesto se compara numéricamente mediante estudios de simulación con el resto de estimadores anteriormente comentados. El objetivo en estos estudios es la estimación de la proporción de personas que padecen de asma y alergia. Los resultados derivados de este estudio de simulación muestran que el estimador propuesto de tipo exponenciación puede ser más eficiente que el resto de estimadores.

PALABRAS CLAVE: proporción, información auxiliar, probabilidades de inclusión.

1. INTRODUCCIÓN

En el análisis de encuestas de salud es frecuente que las variables de respuesta que se consideran sean variables dicotómicas, y la estimación por tanto se refiere a la proporción de individuos que tienen una enfermedad, unos síntomas, que presentan mejorías ante un cierto tratamiento, etc.

La estimación de proporciones se aborda en los textos de muestreo de una forma clásica a partir del estimador de Horvitz-Thompson (ver [6]), estudiándose su aplicación para los diseños muestrales más usuales: muestreo estratificado, por conglomerados, con probabilidades proporcionales a una variable o combinaciones de estos diseños.

Un tema que ha sido objeto de muchas contribuciones en el campo del muestreo es el uso de información auxiliar. Existen muchos trabajos en los que estos métodos y modificaciones suyas son usadas para mejorar las estimaciones de parámetros lineales como medias y totales. Una revisión de estos métodos puede verse en [2] y [15]. Los estimadores de razón, diferencia o regresión que fueron introducidos para medias y totales (ver por ejemplo [14]), han sido extendidos después para otros parámetros como varianzas, coeficientes de regresión, funciones de distribución (ver [9] y [10]), medianas y otros cuantiles ([13]). Es un hecho que los estimadores directos (los que se basan solo en la información muestral proporcionada por la variable de interés) pueden ser mejorados por el estimador general de regresión, puesto que éste es capaz de incorporar la información auxiliar.

[1] arcos@ugr.es, mrueda@ugr.es
[2] jfmunoz@ugr.es
[3] encarniav@ugr.es
[4] , mrueda@ugr.es

La extensión de dichos métodos al caso de una proporción es matemáticamente simple puesto que la proporción es simplemente una media de variables indicadoras, pero la justificación de dichos métodos para el caso discreto no es inmediata, además de que puede producir problemas en la estimación (valores fuera del intervalo [0,1] por ejemplo).

En [7] abordan el problema de la estimación de proporciones cuando existe información auxiliar de tipo continuo o discreto. Los autores en vez de usar el estimador de regresión generalizado, justifican el uso de un estimador de regresión logística basado en un modelo logístico que describe la distribución conjunta de la función indicadora de la clase. Sin embargo para construir este modelo se necesita el conocimiento de todos los valores de la variable auxiliar para cada individuo de la población.

Con frecuencia no se dispone de dicha información, y lo único que se dispone es de la proporción de individuos que presentan una característica que está relacionada con la variable objeto de estudio. Por ejemplo, se quiere estudiar la proporción de mujeres que tiene problemas de infertilidad, y se conoce de censos y otros estudios la proporción de mujeres que no tienen hijos; o se quiere estimar la proporción de individuos que sufren alergia y se conoce la proporción de individuos a los que se les ha recetado un tipo de medicamentos en el último año. En estas circunstancias, en [8] y [11] se definen estimadores de razón para la proporción. El uso de la regresión para la construcción de estimadores de la proporción fue abordado en [12]. En esta última referencia se hace una recopilación de estos métodos y se aplican para estimar la proporción de lagos con riesgo de acidificación.

Otros trabajos también relacionados con la estimación de proporciones, y en particular, relacionados con la estimación mediante intervalos de confianza son [1], [3], [4], [5] y [16].

En este trabajo se propone un nuevo estimador de la proporción basado en el método de exponenciación, se estudian sus propiedades y se comprueba su funcionamiento en la práctica mediante un estudio de simulación. Basándonos en datos reales de la Encuesta Nacional de Salud Española, se calcula dicho estimador junto con estimadores alternativos con el objetivo de estimar la proporción de personas que padecen de asma y alergia. Los resultados derivados del estudio de simulación muestran que el estimador de tipo exponenciación propuesto puede tener un mejor comportamiento en términos de sesgo relativo y error cuadrático relativo que otros estimadores para una proporción poblacional.

2. EL ESTIMADOR DE EXPONENCIACIÓN DE UNA PROPORCIÓN

Se considera la población finita $U = \{1, \cdots, N\}\}$ que contiene *N* unidades distintas e identificadas. $A_1, \cdots, A_N$ denota los valores del atributo de interés *A* donde A_i =1 si la *i*-ésima unidad presenta el atributo *A* y $A_i = 0$ en caso contrario. Sea *B* un atributo auxiliar asociado con el atributo de interés *A* y cuyos valores vienen dados por $B_1, \cdots, B_N$. Para ello, asumimos que se ha extraído una muestra *s*, de tamaño *n*, la cual ha sido seleccionada de *U* bajo muestreo aleatorio simple.

El objetivo será estimar la proporción poblacional de individuos que poseen el atributo de interés *A*, es decir

$$P_A = \frac{1}{N}\sum_{i=1}^{N} A_i$$

Si asumimos una población finita, el estimador estándar de P_A viene dado por

$$\hat{p}_A = \frac{1}{n}\sum_{i=1}^{n} A_i$$

En [5] se define un estimador de tipo razón para P_A mediante $\hat{p}_r = \hat{R}P_B$, donde $\hat{R} = \hat{p}_A/\hat{p}_B$ es un estimador de la razón poblacional $R = P_A/P_B$, siendo

$$\hat{p}_B = \frac{1}{n}\sum_{i=1}^{n} B_i$$

la proporción muestral de individuos que presentan el atributo auxiliar *B* y

$$P_B = \frac{1}{N}\sum_{i=1}^{N} B_i$$

la proporción poblacional del atributo *B* que asumimos es conocida a partir de un censo o se ha estimado sin error. Por su parte, el estimador de tipo diferencia viene dado por

$$\hat{p}_d = \hat{p}_A + (P_B - \hat{p}_B)$$

En [12] también se propone un estimador de tipo regresión de la forma

$$\hat{p}_{reg} = \hat{p}_A + \hat{b}_{opt}(P_B - \hat{p}_B)$$

donde

$$\hat{b}_{opt} = \hat{\phi}\sqrt{\frac{\hat{p}_A\hat{q}_A}{\hat{p}_B\hat{q}_B}}$$

$\hat{q}_A = 1 - \hat{p}_A$, $\hat{q}_B = 1 - \hat{p}_B$ y

$$\hat{\phi} = \frac{n_{11}n_{22} - n_{12}n_{21}}{\sqrt{n_{1.}n_{2.}n_{.1}n_{.2}}}$$

es el coeficiente *V* de Cramer basado en la clasificación dada en la tabla de doble entrada a nivel muestral siguiente

	B	B^c	
A	n_{11}	n_{12}	$n_{1.}$
A^c	n_{21}	n_{22}	$n_{2.}$
	$n_{.1}$	$n_{.2}$	n

Supongamos ahora una situación más general en la que la muestra *s* es seleccionada mediante un diseño muestral general *d* con probabilidades de primer y segundo orden dadas por π_i y π_{ij} respectivamente y que asumimos son estrictamente positivas. Los pesos del diseño asociados a la unidad *i* vienen dados pues por $d_i = \pi_i^{-1}$.

La proporción poblacional P_A puede estimarse usando el estimador de Horvitz-Thompson dado por

$$\hat{p}_A = \frac{1}{N}\sum_{i=1}^{N} d_i A_i$$

Siguiendo la idea dada por [16], proponemos el siguiente estimador, al que llameremos estimador de exponenciación:

$$\hat{p}_{exp.\alpha} = \hat{p}_A\left(\frac{P_B}{\hat{p}_B}\right)^{\alpha}$$

Este estimador es consistente (en el sentido de consistencia en poblaciones finitas) y sesgado. A partir de este estimador observamos

a) Si $\alpha = 1$ el estimador coincide con el estimador de razón propuesto por [11].

b) Si $\alpha = 0$ el estimador coincide con el estimador de Horvitz-Thompson.

El parámetro α se puede determinar de forma que el estimador $\hat{p}_{exp.\alpha}$ sea óptimo en sentido de mínimo error. Para ello lo primero que hacemos es determinar una expresión aproximada para el error cuadrático medio.

Consideremos las variables

$$e_0 = \frac{\hat{p}_A - P_A}{P_A}\,;\, e_1 = \frac{\hat{p}_B - P_B}{P_B}$$

Aproximando la expresión $\hat{p}_{exp.\alpha} - P_A$ por su desarrollo en series de Taylor de segundo orden (si $|e_i| < 1$, $|\alpha e_1| < 1$), se obtiene la siguiente aproximación de orden n^{-1} para el error cuadrático medio:

$$ECM(\hat{p}_{exp.\alpha}) = V(\hat{p}_A) + \alpha^2 V(\hat{p}_B)R^2 + 2\alpha R cov(\hat{p}_A, \hat{p}_B) \quad (1)$$

Derivando esta expresión con respecto a α se obtiene el valor óptimo

$$\alpha^{opt} = -\frac{cov(\hat{p}_A, \hat{p}_B)}{V(\hat{p}_B)}\frac{P_B}{P_A} \quad (2)$$

el cual da lugar al estimador óptimo

$$\hat{p}_{exp}^{opt} = \hat{p}_A\left(\frac{P_B}{\hat{p}_B}\right)^{\alpha^{opt}}$$

Sustituyendo el valor óptimo de α en la expresión (1) se obtiene

$$ECM_{min}(\hat{p}_{exp.\alpha}) = ECM(\hat{p}_{exp}^{opt}) = V(\hat{p}_A) - \frac{cov(\hat{p}_A, \hat{p}_B)}{V(\hat{p}_B)}$$

A partir de esta última expresión resulta inmediato concluir que el estimador óptimo de tipo exponenciación es asintóticamente más eficiente que el estimador directo y que el estimador de razón propuesto en [11].

2.1 Aplicación a muestreo aleatorio simple

Si el diseño muestral considerado es el muestreo aleatorio simple se tienen las siguientes expresiones para varianzas y covarianzas:

$$V(\hat{p}_A) = \frac{N-n}{(N-1)n} P_A Q_A$$

$$V(\hat{p}_B) = \frac{N-n}{(N-1)n} P_B Q_B \tag{3}$$

$$cov(\hat{p}_A, \hat{p}_B) = \frac{N-n}{(N-1)n} \phi \sqrt{P_A Q_A P_B Q_B} \tag{4}$$

donde $Q_A = 1 - P_A$ y $Q_B = 1 - P_B$ son, respectivamente, las proporciones complementarias de las proporciones P_A y P_B, y

$$\phi = \frac{N_{11}N_{22} - N_{12}N_{21}}{\sqrt{N_{1.}N_{2.}N_{.1}N_{.2}}}$$

es el coeficiente *V* de Cramer basado en la clasificación dada en la tabla de doble entrada

	B	B^c	
A	N_{11}	N_{12}	$N_{1.}$
A^c	N_{21}	N_{22}	$n_{2.}$
	$N_{.1}$	$N_{.2}$	N

donde $N_{1.} = \sum_{i=1}^{N} A_i$ es el número de individuos en la población que poseen el atributo *A*, $N_{2.} = \sum_{i=1}^{N} A_i^c$ es el número de individuos en la población que poseen el atributo A^c, $N_{.1} = \sum_{i=1}^{N} B_i$ es el número de individuos en la población que poseen el atributo *B* y $N_{.2} = \sum_{i=1}^{N} B_i^c$ es el número de individuos en la población que poseen el atributo B^c. De forma análoga, N_{11} es el número de individuos en la población que simultáneamente poseen los atributos *A* y *B*, N_{12} es el número de individuos en la población que simultáneamente poseen los atributos *A* y B^c, etc.

Utilizando las expresiones (3) y (4) en la ecuación (2), se puede obtener fácilmente que el peso óptimo para α bajo muestreo aleatorio simple viene dado por

$$\alpha^{opt} = \phi \sqrt{\frac{Q_A P_B}{P_A Q_B}}$$

3. ESTIMADORES CON ÓPTIMOS ESTIMADOS

El valor óptimo de α dado depende de parámetros desconocidos, por lo que en la práctica no es operativo. Se puede dar un estimador alternativo sin más que sustituir el valor óptimo α por un estimador consistente. En el caso de un diseño muestral general, el estimador del valor óptimo de α que podría considerarse es

$$\hat{\alpha} = -\frac{\widehat{cov}(\hat{p}_A, \hat{p}_B)}{\hat{V}(\hat{p}_B)} \frac{\hat{p}_B}{\hat{p}_A}$$

y bajo muestreo aleatorio simple dicho estimador sería

$$\hat{\alpha} = \hat{\phi} \sqrt{\frac{\hat{q}_A \hat{p}_B}{\hat{p}_A \hat{q}_B}}$$

En ambos casos, el estimador de tipo exponenciación que resulta es

$$\hat{p}_{exp} = \hat{p}_A \left(\frac{P_B}{\hat{p}_B}\right)^{\hat{\alpha}}$$

4. ESTUDIO DE SIMULACIÓN Y APLICACIÓN A LA ESTIMACIÓN DE PREVALENCIAS

En esta sección se compara empíricamente el estimador de tipo exponenciación propuesto con otros estimadores para la proporción propuestos en la literatura. Para este estudio de simulación utilizaremos N=9063 datos extraídos de la Encuesta Nacional de Salud (ENS) Española del año 2006 con el fin de estimar prevalencias para las enfermedades crónicas de asma y alergia. Este estudio de simulación consiste en considerar estos datos muestrales como una población en la cual distintas muestras serán extraídas. Las simulaciones realizadas nos conducen a una comparación del comportamiento de varios estimadores en una situación que podría presentarse en la práctica. Los detalles más reseñables sobre la información contenida en la población ENS se describen a continuación.

La población ENS asume un diseño muestral complejo donde las unidades de muestreo de la primera etapa son secciones censales (2236 secciones censales se han seleccionado en la muestra). Las unidades de la primera etapa se agrupan en estratos de acuerdo con el tamaño del municipio. Las unidades de la segunda etapa son las viviendas familiares principales. Dentro de las unidades de la segunda etapa no se lleva a cabo un submuestreo, y las viviendas con residentes se encuestan. Dentro de cada hogar, un adulto (de una edad igual o superior a 16 años) se selecciona y se completa el cuestionario de adultos, mientras que si hay menores (de 0 a 15 años) en el hogar, uno de ellos se selecciona y se completa el cuestionario de los menores. Los datos usados en el estudio de simulación se refieren a los cuestionarios de menores. Se seleccionaron un total de 31300 hogares donde se rellenaron en total N=9063 cuestionarios sobre menores.

Las bases de datos de estadística nacional de salud contienen datos de indicadores de salud, incluyendo indicadores básicos demográficos y socioeconómicos, algunos indicadores están relacionados con los estilos de vida, el entorno y los cuidados de la salud, utilización y gastos en medicación. La dolencia del asma crónica y la alergia se investigaron en la ENS. La Organización Mundial de la Salud reconoce que el asma es una de las mayores preocupaciones de la salud pública. La Organización Mundial de la Salud juega el rol de coordinadora internacional de todos los esfuerzos contra esta dolencia. El estudio internacional del asma y las alergias en niños y adolescentes reveló que las mismas siguen en crecimiento en los niños europeos. En estos momentos el objetivo básico es el uso alternativo de estimadores puntuales más precisos para las comentadas prevalencias y la construcción de intervalos de confianza con cobertura deseable y la mínima anchura. Los datos usados en las simulaciones se refieren a tres variables. La variable de interés (atributo A) indica si el niño ha sufrido de la dolencia (asma o alergia). Por otro lado la información auxiliar procede de la variable "ha recetado el médico medicamentos al niño para el asma o para la alergia?" (atributo auxiliar B_1) y " ha consumido el niño los medicamentos para la alergia o para el asma en las \últimas dos semanas?" (atributo B_2). Para el caso de asma, las proporciones poblacionales de los atributos son $P_A = 0.07$, $P_{B1} = P_{B2} = 0.04$, mientras que los coeficientes V de Cramer entre el atributo A y los atributos B_1 y B_2 son, respectivamente, $\phi_1 = 0.583$ y $\phi_2 = 0.57$. Para el caso de alergia, estos parámetros son $P_A = 0.12$, $\mathrm{P}_{\mathrm{B}1} = \mathrm{P}_{\mathrm{B}2} = 0.03$, $\phi_1 = 0.51$ y $\phi_2 = 0.495$.

Destacamos que el Sistema de Información de la Salud del Sistema Nacional de Salud Española posee toda la información referida a los medicamentos prescritos, los cuales son clasificados por grupos terapéuticos, edad, sexo, dolencias, etc. La prescripción electrónica se implantará en un futuro muy cercano y su información será mayor. Por otro lado, los estudios llevados a cabo por agencias públicas o privadas tales como la del Colegio Oficial de Farmacéuticos o la Federación de Farmacéuticos Españoles también tienen información muy relevante en lo que respecta a las recetas médicas, especialmente en el área de salud infantil. Estos argumentos indican que el estimador propuesto se puede aplicar en estas situaciones, siempre y cuando la proporción poblacional de recetas médicas sea conocida. Por otra parte, la proporción de medicamentos consumidos se puede obtener del Sistema de Información de la Salud. Sin embargo, nuestro interés reside en el estudio del comportamiento empírico del estimador propuesto cuando es utilizado con datos reales.

El estudio de simulación consiste en la selección de D=10000 muestras con el fin de comparar empíricamente los distintos estimadores en términos de sesgo relativo (SR) y error cuadrático medio relativo (ECMR), donde

$$SR = \frac{E[\hat{p}] - P_A}{P_A}$$

$$ECMR = \frac{\sqrt{ECM[\hat{p}]}}{P_A}$$

$\hat{p}$ es un determinado estimador y la esperanza empírica ($E[\cdot]$) y el error cuadrático medio empírico ($ECM[\cdot]$) vienen dados por

$$E[\hat{p}] = \frac{1}{D}\sum_{d=1}^{D} \hat{p}_d$$

$$ECM[\hat{p}] = \frac{1}{D}\sum_{d=1}^{D} (\hat{p}_d - P_A)^2$$

donde $\hat{p}_d$ denota el estimador $\hat{p}$ obtenido en la d-ésima simulación. El estimador propuesto $\hat{p}_{exp}$ se comparará empíricamente con los estimadores $\hat{p}_A$ (estimador estándar), $\hat{p}_r$ (estimador de tipo razón), $\hat{p}_d$ (estimador de tipo diferencia) y $\hat{p}_{reg}$ (estimador de tipo regresión) definidos en la Sección 2. Por último destacamos que el estudio de simulación se ha llevado a cabo con muestras de distinto tamaño y que dichas muestras se han seleccionado utilizando, por un lado, muestreo aleatorio simple y, por otro lado, utilizando como diseño muestral con probabilidades desiguales, muestreo estratificado aleatorio con afijación uniforme.

	n	$f(\%)$	$\hat{p}_A$	$\hat{p}_r$	$\hat{p}_d$	$\hat{p}_{reg}$	$\hat{p}_{exp}$
Asma	50	0.6	10.7	-0.1	-0.9	0.3	-2.3
	100	1.1	1.8	18.0	-0.5	-0.2	-3.1
	250	2.8	0.2	10.8	0.0	0.0	-1.5
	500	5.5	0.0	4.2	-0.1	-0.1	-0.8
	750	8.3	0.2	3.2	0.2	0.2	-0.2
	1000	11.0	0.1	1.9	0.0	0.0	-0.3
Alergia	50	0.6	3.0	13.8	-1.3	-0.3	-2.9
	100	1.1	0.2	21.7	-0.1	0.0	-2.2
	250	2.8	-0.1	7.5	-0.1	-0.1	-1.0
	500	5.5	-0.1	3.1	-0.1	-0.1	-0.6
	750	8.3	0.1	1.8	0.0	0.0	-0.3
	1000	11.0	0.0	1.4	0.0	0.0	-0.2

Tabla 1. Valores estimados de SR (%) para distintos estimadores de P_A en la población ENS. Muestras seleccionadas mediante muestreo aleatorio simple con distintas fracciones de muestreo f=n/N.

Las Tablas 1 y 2 muestran los resultados obtenidos en el estudio de simulación en término de SR y ECMR para el caso de muestreo aleatorio simple. Destacamos que para el caso de muestreo estratificado aleatorio los resultados derivados de dicho estudio de simulación han resultado ser muy similares a los obtenidos en el caso de muestreo aleatorio simple y mostrados en las Tablas 1 y 2, y de aquí que esta información esté omitida. Del mismo modo, los resultados obtenidos con el segundo atributo auxiliar son muy similares a los resultados obtenidos con el primer atributo auxiliar, por lo que las simulaciones asociadas al segundo atributo auxiliar también se han omitido.

	n	$f(\%)$	$\hat{p}_A$	$\hat{p}_r$	$\hat{p}_d$	$\hat{p}_{reg}$	$\hat{p}_{exp}$
Asma	50	0.6	49.6	52.6	41.6	41.6	40.7
	100	1.1	34.8	69.6	29.0	29.1	28.9
	250	2.8	22.4	48.7	18.2	18.3	18.3
	500	5.5	15.8	24.5	12.8	12.8	12.8
	750	8.3	12.6	18.6	10.2	10.2	10.2
	1000	11.0	10.6	15.2	8.8	8.7	8.7
Alergia	50	0.6	38.2	66.1	34.0	33.9	33.4
	100	1.1	27.4	76.0	23.9	23.9	23.8
	250	2.8	17.1	36.8	15.1	14.9	14.9
	500	5.5	11.9	20.1	10.4	10.3	10.3
	750	8.3	9.7	15.1	8.4	8.3	8.3
	1000	11.0	8.3	12.8	7.1	7.1	7.1

Tabla 2. Valores estimados de ECMR (%) para distintos estimadores de P_A en la población ENS. Muestras seleccionadas mediante muestreo aleatorio simple con distintas fracciones de muestreo f=n/N

La Tabla 1 muestra los valores de SR obtenidos en el problema de estimación de prevalencias (asma y alergia) y para los distintos estimadores. Podemos observar que todos los estimadores obtienen, como parece razonable, sesgos más insignificantes a medida que aumenta el tamaño de muestra. Para el caso de asma, el estimador estándar obtiene un sesgo relativamente elevado para tamaños muestrales pequeños. Respecto a estimadores basados en información auxiliar, el estimador de tipo razón también obtiene sesgos elevados para tamaños muestrales pequeños. En el resto de los casos, los estimadores obtienen valores de SR inferiores al 5% en términos absolutos. Para el caso de alergia, el estimador estándar obtiene valores más pequeños, mientras que el estimador de tipo razón sigue teniendo asociados valores elevados de SR. En todos los casos, los sesgos relativos del estimador de tipo exponenciación propuesto están dentro de un rango razonable.

La Tabla 2 nos permite comparar empíricamente la eficiencia de los distintos estimadores en términos de ECMR. A partir de estos resultados podemos comprobar cómo los estimadores de tipo diferencia, regresión y exponenciación son los más eficientes, siendo, en general, el estimador propuesto de tipo exponenciación ligeramente más eficiente que los estimadores de tipo diferencia y tipo regresión. El estimador de tipo razón no tiene un buen comportamiento en términos de ECMR, y es incluso menos eficiente que el estimador estándar que no utiliza información auxiliar en la etapa de estimación.

Agradecimientos: Este trabajo ha sido subvencionado por el proyecto de Excelencia SEJ-7039 de la Consejería de Economía, Innovación y Ciencia de la Junta de Andalucía, y por el proyecto nacional MTM2009-10055 del Ministerio de Educacion y Ciencia.

REFERENCIAS

[1] BLYTH, C.R. & STILL, H. A. (1983): Binomial confidence intervals. **Journal of the American Statistical Association**, 78, 108 - 116.

[2] CHAUDHURI, A. & VOS, J.W.E. (1988): **Unified Theory and Strategies of Survey Sampling**. North-Holland, Amsterdam.

[3] CHEN, C., LI, J. & ZHOU, Z. (2004): Confidence interval of the difference between two proportions with overdispersion. **Journal of Biopharmaceutical Statistics**, 14, 469 - 482.

[4] CLOPPER, C.J. & PEARSON, E. S. (1934): The use of confidence or fiducial limits illustrated in the case of binomial. **Biometrika**, 26, 404 - 413.

[5] FLEISS, J. L., LEVIN, B. & PAIK, M. C. (2003): **Statistical methods for rates and proportions (3rd~edn).** Wiley, New Yersey.

[6] HORVITZ, D.G. & THOMPSON D.J. (1952): A generalization of sampling without replacement from a finite universe. **Journal of the American Statistical Association**, 47, 663 - 685.

[6] LEHTONEN, R. & VEIJANEN, A. (1998): Logistic generalized regression estimators. **Survey Methodology**, 24, 51-55.

[7] MUÑOZ, J.F., ARCOS, A., ÁLVAREZ, E., RUEDA, M., GONZALEZ, S. & SANTIAGO, A. (2012): Optimum ratio estimators for the population proportion. **International Journal of Computers Mathematics**, 89, 357 – 365.

[8] NASCIMENTO-SILVA, P.L.D. & SKINNER, C.J. (1995): Estimating distribution functions with auxliary information using Poststratification. **Journal of Official Statistics**, 11, 277 - 294.

[9] RAO, J.N.K., KOVAR, J.G. & MANTEL, H.J. (1990): On estimating distribution function and quantiles from survey data using auxiliary information. **Biometrika**, 77, 365 -375.

[10] RUEDA M.M., MUÑOZ, J.F., ARCOS, A., ÁLVAREZ-VERDEJO, E. & MARTÍNEZ, S. (2011): Estimators and confidence intervals for the proportion using binary auxiliary information with applications to pharmaceutical studies. **Journal of Biopharmaceutical Statistics,** 21, 526 - 554.

[11] RUEDA M.M., MUÑOZ J.F., ARCOS A. & ÁVAREZ-VERDEJO, E. (2011): Indirect estimation of proportions in natural resource surveys. **Mathematicss and Computers in Simulation**, 81, 2317 – 2325.

[12] RUEDA, M., MUÑOZ, J. F., GONZÁLEZ, S. & ARCOS, A. (2007): Estimating quantiles under sampling on two occassions with arbitrary sampling designs. **Computational Statistics and Data Analysis**, 51, 6596 - 6613.

[13] SÄRNDAL, C. E., SWENSSON, B. & WRETMAN, J. H. (1992): **Model Assisted Survey Sampling. 3rd ed.** Springer-Verlag, New York.

[14] SINGH, S. (2003): **Advanced Sampling Theory With Applications: How Michael Selected Amy**. Kluwer Academic Pub., Berlin.

[15] SRIVASTAVA, S. K. (1967): An estimator using auxiliary information in sample surveys. **Calcutta Statistical Association Bulletin**, 16 121-132.

[16] VOLLSET, S. E. (1993): Confidence interval for a binomial proportion. **Statistics in Medicine**, 12, 809 - 824.