

USO DE MODELOS ESTADÍSTICOS PARA LA ESTIMACIÓN DE TASAS DE DESOCUPACIÓN EN DOMINIOS PEQUEÑOS*

Leticia Susana Hachuel**

Gabriela Susana Boggio**

Cristina Beatriz Cuesta**

Facultad de Ciencias Económicas y Estadística
Universidad Nacional de Rosario

Resumen. La estimación de parámetros en áreas o dominios particulares puede ser problemática cuando el tamaño muestral correspondiente a dichos dominios es reducido ya que las “estimaciones directas” suelen ser inestables. Los modelos de efectos aleatorios constituyen una alternativa metodológica apropiada para mejorar estas estimaciones. En este trabajo se utilizan estos modelos para estimar la tasa de desocupación en dominios pequeños con información proporcionada por la Encuesta Permanente de Hogares (EPH) correspondiente al Aglomerado Gran Rosario. Se realiza, además, un estudio por simulación a fin de ampliar los resultados a diferentes escenarios de análisis. Los resultados hallados muestran la conveniencia del uso de la estimación basada en modelos toda vez que el tamaño de la muestra de los subgrupos sea reducido y la variabilidad entre ellos no sea demasiado importante.

Palabras clave: Áreas pequeñas; Modelos de efectos aleatorios; Suavizado de estimaciones

* Este trabajo es producto del Proyecto de Investigación Acreditado por la Universidad Nacional de Rosario denominado Modelos de niveles múltiples para respuesta categórica, dirigido por Leticia Susana Hachuel.

** Docentes-Investigadoras. Directora e integrantes del Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística.

Contacto: lhachuel@fcecon.unr.edu.ar.

Abstract. The parameter estimates in some domains or areas could have some problems when the sample size corresponding to those domains is reduced as the “direct estimates” could be unstable. The random effects models are a methodological alternative to improve these estimates. In this article the random effects models are used to estimate the rate of unemployment in small domains with the data provided by the Permanent Household Survey (EPH) corresponding to the Gran Rosario region. A simulation study is performed to extend the results to other scenarios of analysis. The results founded show the convenience of the use of the estimates based on models when the sample size of a group and their variability are small.

Key words: Small areas; Random effects models; Smoothing estimates

1. Introducción

En los últimos años la estimación en áreas o dominios pequeños surgió como un tema importante dentro de la estadística aplicada. El aumento en el costo de recolectar una muestra y la creciente disponibilidad de computadoras potentes se combinaron para mejorar los métodos de análisis en este tema, de modo que se puedan usar técnicas sofisticadas para estimar estadísticas relativas a áreas geográficas o dominios de interés para los cuales se dispone de escaso número de observaciones.

Se sabe que la estimación de probabilidades en áreas, dominios o subgrupos mediante la proporción simple de casos en cada subgrupo suele ser inestable cuando el tamaño muestral correspondiente a dichos subgrupos es pequeño o moderado. Parece apropiado entonces, reforzar dichas estimaciones, denominadas “estimaciones directas”, utilizando la información disponible para otros subgrupos de forma de incrementar el tamaño de muestra efectivo y de esta manera construir “estimaciones indirectas” que aumenten la precisión. Los modelos de efectos aleatorios sirven como mecanismo para mejorar las estimaciones muestrales bajo el supuesto de que las verdaderas proporciones se distribuyen de acuerdo a alguna distribución de probabilidad (Ghosh y Rao, 1994). De esta manera, las estimaciones basadas en modelos son el resultado de “enlazar” las áreas relacionadas (Rao, 1999, 2003).

En este trabajo, se utilizan modelos de efectos aleatorios para estimar la probabilidad de desocupación en dominios pequeños con información proporcionada por la Encuesta Permanente de Hogares (EPH) correspondiente al Aglomerado Gran Rosario. Los dominios de estudio están determinados por las diferentes alternativas que resultan de considerar conjuntamente el nivel de educación y el sexo de los respondentes. El Instituto Nacional de Estadísticas y Censos (INDEC) recomienda que al analizar datos provenientes de la EPH que surgen del cruce entre variables se evite el uso de información con menos de 80 casos muestrales por celda (INDEC, 2002).

El enfoque que se propone para mejorar las estimaciones muestrales directas a través del uso de modelos de efectos aleatorios se evalúa a través de un estudio por simulación generando datos binomiales bajo diferentes escenarios. A partir del mismo se presentan recomendaciones acerca de las situaciones en que estas estimaciones resultan más adecuadas que las estimaciones directas.

En la sección siguiente se presenta una síntesis sobre los modelos de efectos aleatorios que se utilizan en la estimación de las tasas de desocupación para jefes de hogares en áreas pequeñas.

2. Metodología

El interés se centra en estimar la probabilidad de un evento asociado a una variable dicotómica para áreas o dominios de interés de los que se tienen relativamente pocas observaciones.

Se supone que en el área, dominio o subgrupo i -ésimo se tienen n_i observaciones dicotómicas $\{y_{ij}, j = 1, \dots, n_i\}$ independientes con $P(Y_{ij}=1) = \delta_i$ y $P(Y_{ij}=0) = 1 - \delta_i$ tal que $y_i = \sum y_{ij}$ a través de todos los elementos del subgrupo. Es decir, se tienen $\{Y_{ij}\}$ variables binomiales independientes con tamaños de muestra $\{n_i\}$ y parámetros $\{\delta_i\}$, donde δ_i es la verdadera proporción del evento en el subgrupo i , $i=1, \dots, I$.

Las proporciones muestrales $\{p_i = y_i/n_i\}$ son las estimaciones máximo-verosímiles de $\{\delta_i\}$ bajo el siguiente modelo lineal generalizado de efectos fijos para variable respuesta binomial:

$$\text{logit}(\delta_i) = \alpha + \hat{\alpha}_i, \quad i=1, \dots, I$$

donde, para resolver el problema de la identificabilidad del modelo, se fija un conjunto de restricciones en los parámetros, tal como: $\sum \hat{\alpha}_i = 0$ ó $\hat{\alpha}_1 = 0$. Para esta última restricción, α representa el logaritmo del odds de ocurrencia del evento para la categoría de referencia, $\delta_i/(1-\delta_i)$, y los $\{\hat{\alpha}_i\}$ representan los efectos diferenciadores respecto del i -ésimo subgrupo o dominio.

Este modelo es saturado por tener I parámetros no redundantes para I observaciones binomiales y por lo tanto las estimaciones que se obtienen a partir de él coinciden con las estimaciones muestrales directas $\{p_i\}$. Para muestras chicas, estas estimaciones directas tienen errores estándares grandes, es decir suelen mostrar mucha más variabilidad que los verdaderos valores $\{\delta_i\}$, especialmente cuando estas probabilidades son similares. Una forma de disminuir esta variabilidad y orientar las estimaciones hacia la media general es mediante el uso de modelos de efectos aleatorios.

$$\hat{\pi}_i = \exp(\hat{\alpha} + \hat{u}_i) / [1 + \exp(\hat{\alpha} + \hat{u}_i)]$$

Estos modelos consideran que las observaciones dentro de cada uno de ellos tienden a ser más parecidas que las observaciones de subgrupos distintos.

Agresti et al. (2000) proponen utilizar el siguiente modelo lineal generalizado de efectos aleatorios para mejorar las estimaciones directas,

$$\text{logit}(\delta_i) = \alpha + u_i, \quad i=1, \dots, I \quad (1)$$

donde los $\{u_i\}$ son los efectos aleatorios asociados a cada subgrupo, que se suponen independientes y con distribución $N(0, \sigma^2)$. Bajo esta formulación los $\{\delta_i\}$ resultan idénticos cuando $\sigma = 0$.

Luego de estimar α y σ se estima el $\text{logit}(\delta_i)$ usando la estimación de α y de u_i , efecto aleatorio predicho basado en los datos observados. El efecto aleatorio predicho \hat{u}_i es la media estimada de la distribución de u_i , la cual depende de todos los datos y no sólo de los del subgrupo i . A partir de esta estimación del $\text{logit}(\delta_i)$ se obtiene la estimación de la probabilidad del evento en cada subgrupo de acuerdo a la siguiente expresión (McCulloch y Searle, 2001; Demidenko, 2004):

la cual difiere de la proporción muestral p_i .

Si la estimación de θ es nula, la estimación de cada δ_i utilizando el modelo es

$$\rho = \left(\frac{\sum_{i=1}^I \sum_{j=1}^{n_i} y_{ij}}{\sum_{i=1}^I n_i} \right)$$

es decir igual a la proporción muestral total considerando conjuntamente los I subgrupos.

En la medida que la estimación de θ sea positiva pero pequeña, las estimaciones de δ_i se alejarán un tanto de este promedio general y resultarán mejores estimaciones que las proporciones muestrales separadas $\{p_i = y_i/n_i\}$. Este “pedirle prestado al todo”, que provoca el ajuste del modelo (1), proporciona la ventaja de suavizar las estimaciones directas y efectivamente basa los resultados en tamaños de muestra más grandes que si se usaran los datos propios de cada muestra por separado. Las estimaciones de los efectos aleatorios producen un reajuste de las estimaciones separadas orientándolas hacia la proporción muestral general. Este direccionamiento o “encogimiento” decrece a medida que la variabilidad aumenta. Lo mismo sucede cuando las $\{n_i\}$ crecen ya que al aumentar el tamaño de cada muestra las proporciones muestrales separadas resultan más eficientes (Agresti et al., 2000).

Si bien las estimaciones a partir del modelo propuesto de efectos aleatorios tienden a estar más cercanas a las verdaderas probabilidades poblacionales que las proporciones muestrales, la mejoría puede ser aún mayor si se utiliza además información adicional sobre el evento en estudio, en una etapa anterior. Por ejemplo, si se conoce la proporción del evento en estudio en un período anterior, $\{q_i\}$, puede usarse esta información en el modelo de la siguiente manera:

$$\text{logit}(\delta_i) = \text{logit}(q_i) + \alpha + u_i, \quad i=1, \dots, I \quad (2)$$

Es decir, se agrega al modelo (1) el término $\text{logit}(q_i)$ donde los $\{q_i\}$ son conocidos y considerados “offsets” para el modelo (2); esto es, con coeficientes predeterminados.

Una forma alternativa de expresar el modelo (2) es la siguiente:

$$\log \left(\frac{\pi_i / (1 - \pi_i)}{q_i / (1 - q_i)} \right) = \alpha + u_i, \quad i=1, \dots, I$$

En esta formulación, $(\alpha + u_i)$ representa el logaritmo de la razón de odds, para el i -ésimo subgrupo, de presentar el evento en comparación a haberlo presentado en un momento anterior.

Bajo este modelo, cuando las estimaciones de θ y α son iguales a cero, las probabilidades estimadas obtenidas a partir del mismo coinciden con q_i . En cambio, si sólo la estimación de la variancia es nula, dichas probabilidades dependerán de la estimación de α y de los valores considerados como “offset”.

3. Estudio de la desocupación en Rosario

La estimación de la tasa de desocupación suele ser un objetivo fundamental en muchos análisis económicos. En particular la estimación para dominios específicos puede ser requerida a los efectos de instrumentar políticas de planificación.

Este trabajo se focaliza en la estimación de la tasa de desocupación correspondiente a la subpoblación de los jefes de hogar desagregada según los dominios definidos por las combinaciones de nivel de educación (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, universitario incompleto, universitario completo) y sexo (femenino, masculino) del jefe de hogar.

La información disponible para ello es la proveniente de la EPH para el Aglomerado Gran Rosario en el primer trimestre del año 2007. Esta encuesta no está programada para la consideración de los dominios o subgrupos recién definidos, por lo cual es probable que el número de observaciones en cada uno de ellos sea reducido (INDEC, 2003).

La Tabla 1 muestra las proporciones muestrales de desocupación de los jefes de hogar, es decir las estimaciones directas de las probabilidades de desocupación provistas por el INDEC para cada uno de los dominios definidos.

Tabla 1
Proporción de desocupados según dominio

Dominio	Proporción de desocupados	Tamaño
Masculino, primaria incompleta	0.068	74
Masculino, primaria completa	0.061	281
Masculino, secundaria incompleta	0.068	133
Masculino, secundaria completa	0.014	219
Masculino, universitario incompleto	0.041	97
Masculino, universitario completo	0.019	106
Femenino, primaria incompleta	0.039	26
Femenino, primaria completa	0.138	80
Femenino, secundaria incompleta	0.097	31
Femenino, secundaria completa	0.102	49
Femenino, universitario incompleto	0.171	41
Femenino, universitario completo	0.057	53

Se puede señalar que en 6 de los dominios de interés propuestos no se cumple la recomendación del INDEC acerca del tamaño de muestra requerido. Ello conduce a la aplicación de los modelos presentados en la sección anterior para la obtención de estimaciones más estables.

En la Tabla 2 se presentan, junto con las estimaciones directas p_i , las estimaciones indirectas halladas en base a los modelos (1) y (2), $p_i^{(1)}$ y $p_i^{(2)}$ respectivamente. El modelo (2) utiliza como información adicional la estimación directa de la tasa de desocupación de cada dominio para el último trimestre del año 2006 a modo de “offset” (Tabla A1, Anexo).

Tabla 2
Estimaciones de las tasas de desocupación según dominio

Dominio	p_i	$p_i^{(1)}$	$p_i^{(2)}$
Masculino, primaria incompleta	0.068	0.064	0.070
Masculino, primaria completa	0.061	0.060	0.075
Masculino, secundaria incompleta	0.068	0.065	0.062
Masculino, secundaria completa	0.014	0.025	0.025
Masculino, universitario incompleto	0.041	0.048	0.022
Masculino, universitario completo	0.019	0.035	0.007
Femenino, primaria incompleta	0.039	0.053	0.036
Femenino, primaria completa	0.138	0.112	0.119
Femenino, secundaria incompleta	0.097	0.074	0.101
Femenino, secundaria completa	0.102	0.081	0.100
Femenino, universitario incompleto	0.171	0.117	0.135
Femenino, universitario completo	0.057	0.058	0.065

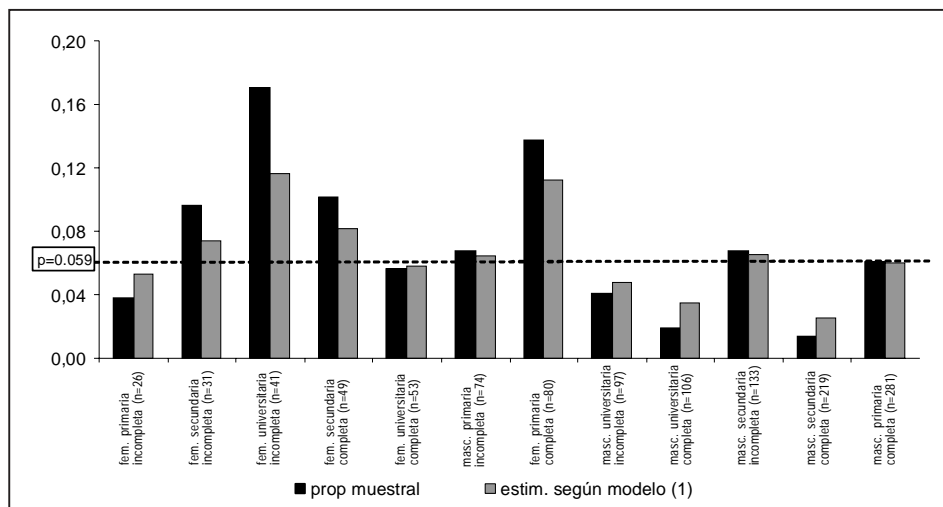
p_i : estimación directa; $p_i^{(1)}$: estimación bajo el modelo aleatorio sin “offset”; $p_i^{(2)}$: estimación bajo el modelo aleatorio con “offset”.

Las estimaciones de las tasas de desocupación mediante el modelo (1), $p_i^{(1)}$, varían entre 0.025 y 0.117, es decir con un rango de variación igual a 0.092, mientras que las proporciones muestrales, p_i , varían entre 0.014 y 0.171 (rango igual a 0.157). Esta disminución en la variación de las estimaciones basadas en el modelo, comúnmente denominado encogimiento, es producto del valor moderado (0.58) obtenido para la variabilidad estimada de los efectos aleatorios del modelo.

A los efectos de una mejor visualización de los resultados se grafican las proporciones directas y estimadas bajo el modelo para cada dominio. Se ordenan

los mismos de acuerdo a su tamaño con el objeto de evaluar el efecto del suavizado de la estimación a medida que el tamaño de la muestra aumenta (Figura 1).

Figura 1
Estimaciones de las tasas de desocupación bajo el modelo (1) según dominio



Se puede apreciar un acercamiento de las tasas de desocupación estimadas bajo el modelo hacia la proporción muestral total obtenida considerando conjuntamente los 12 dominios o subgrupos, la cual resulta igual a 0.059. En particular, en el dominio de menor tamaño, n=26, conformado por los jefes de hogar de sexo femenino con estudios primarios incompletos, se produce un acercamiento importante a esta media global ya que la estimación directa es de 0.039 y la basada en el modelo resulta igual a 0.053 (Figura 1).

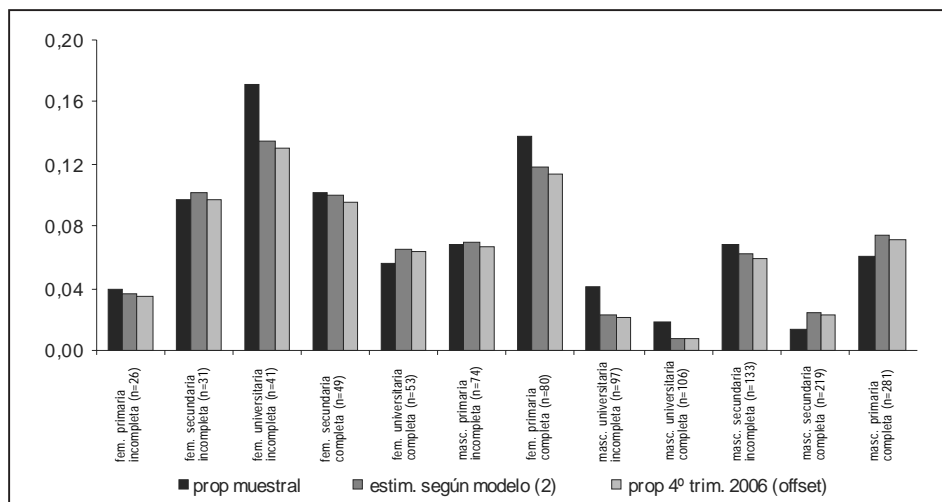
El ajuste del modelo (2), que incluye información adicional, también produce estimaciones suavizadas de las tasas de desocupación que varían entre 0.007 y 0.135. Las estimaciones halladas para los parámetros del modelo (0.05 y 0.00 para α y δ respectivamente) provocan que, en esta oportunidad, las tasas de desocupación estimadas se acerquen considerablemente a las tasas de desocupación del último trimestre del 2006, utilizadas como “offset” tal como se observa en la Figura 2.

Si bien ambos modelos consiguen suavizar las estimaciones superando el problema de la baja frecuencia de algunos dominios, es posible compararlos desde un punto de vista estadístico a través de dos medidas descriptivas de bondad de ajuste: el criterio de información de Akaike, AIC (Akaike, 1973, 1974) y el criterio de información bayesiano, BIC (Schwarz, 1978). Los valores encontrados para estas estadísticas son $AIC^{(1)} = 66.8$ y $BIC^{(1)} = 67.8$ para el modelo (1) y $AIC^{(2)} = 49.5$ y $BIC^{(2)} = 50.5$ para

el modelo (2). La regla establece que los valores más bajos indican un modelo más apropiado, por lo que, como era de esperar, las estimaciones proporcionadas por el modelo que incorpora información suplementaria resultan más satisfactorias.

De todas maneras, con el fin de evaluar el comportamiento de las probabilidades estimadas por el modelo de efectos aleatorios en un espectro más amplio que el de la situación real particular presentada se realiza un estudio por simulación.

Figura 2
Estimaciones de las tasas de desocupación bajo el modelo (2) según dominio



4. Estudio de simulación

Los estudios por simulación constituyen una herramienta apropiada para generalizar los resultados más allá de un caso particular analizado. Consisten en construir poblaciones hipotéticas con valores de los parámetros conocidos y generar muestras de dichas poblaciones observando el comportamiento de las estimaciones halladas para dichos parámetros conocidos.

En este trabajo se evalúa el comportamiento del modelo de efectos aleatorios (1) en cuatro escenarios de análisis compuestos por dominios con diferentes características¹. El primer escenario fija las probabilidades de ocurrencia para cada uno de 12 dominios, $\{\delta_i, i=1, \dots, 12\}$, todas iguales a 0.2; el segundo asigna valores a dichas probabilidades entre 0.10 y 0.32 y el tercero, entre 0.10 y 0.65. El

¹ En esta oportunidad se desestima la evaluación del modelo (2) debido a que el requerimiento de información adicional ("offset") hace más compleja la definición de los escenarios.

último escenario se define con características semejantes al caso real analizado (Tabla 3).

Tabla 3
Descripción de los escenarios bajo estudio según los valores de $\{\delta_i\}$ asignados en cada dominio

Escenarios	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	δ_8	δ_9	δ_{10}	δ_{11}	δ_{12}
1	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
2	0.10	0.12	0.14	0.16	0.18	0.20	0.22	0.24	0.26	0.28	0.30	0.32
3	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65
4	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13

En cada uno de los 12 dominios o subgrupos se genera una observación proveniente de una variable binomial con probabilidad de ocurrencia δ_i y el mismo tamaño de muestra n para cada uno de ellos. Luego se ajusta el modelo de efectos aleatorios (1) a los datos generados obteniéndose las estimaciones de $\hat{\delta}_i$ y $\hat{\delta}_i^{(1)}$ registrándose para cada dominio:

1. la diferencia en valor absoluto entre la estimación directa y la verdadera probabilidad: $|p_i - \delta_i|$,
2. la diferencia en valor absoluto entre la estimación basada en el modelo de efectos aleatorios y la verdadera probabilidad: $|p_i^{(1)} - \delta_i|$.

Luego se promedian estas diferencias a través de todos los dominios obteniéndose: $\bar{O}|p_i - \delta_i|/12$, y $\bar{O}|p_i^{(1)} - \delta_i|/12$. Se comparan estos dos promedios y se registra cuál de ellos es menor.

Esta secuencia se repite 1000 veces y se calcula:

- el promedio de las estimaciones de $\hat{\delta}_i$,
- el promedio de las diferencias promedio halladas en cada repetición,
- el porcentaje de veces en que el promedio de las diferencias en valor absoluto de los verdaderos valores con respecto a las estimaciones basadas en el modelo es menor que con respecto a las estimaciones directas.

Se espera que el cálculo de este último porcentaje permita evaluar bajo qué condiciones resulta más conveniente la estimación basada en el modelo de efectos aleatorios que la simple proporción muestral.

Este procedimiento se lleva a cabo bajo los escenarios definidos y para tamaños de muestra en cada dominio iguales a $n=20, 50$ y 100 .

En la Tabla 4 se muestran los resultados obtenidos. Se puede observar que en el escenario 1, donde el promedio de las estimaciones de δ es pequeño, el porcentaje de veces en que el promedio de las diferencias en valor absoluto del verdadero valor con respecto a las estimaciones basadas en el modelo es menor que con respecto a las estimaciones directas, resulta muy alto cualquiera sea el tamaño de las muestras en cada dominio (columna 6, Tabla 4).

Tabla 4
Resultados de las simulaciones.

n	Escenarios	Variabilidad promedio estimada	Promedio de $ \hat{O}p_i - \delta_i /12$	Promedio de $ \hat{O}p_i^{(1)} - \delta_i /12$	% de veces que $\frac{ \hat{O}p_i^{(1)} - \delta_i /12 < \hat{O}p_i - \delta_i /12$
20	1	0.105	0.070	0.025	99.7
	2	0.349	0.071	0.054	84.4
	3	0.819	0.078	0.075	64.9
	4	0.369	0.046	0.031	93.5
50	1	0.046	0.045	0.015	99.8
	2	0.401	0.044	0.040	70.5
	3	0.829	0.051	0.050	60.4
	4	0.288	0.028	0.024	73.9
100	1	0.020	0.032	0.011	100.0
	2	0.420	0.032	0.030	67.0
	3	0.834	0.036	0.035	56.6
	4	0.260	0.0201	0.0188	68.2

p_i : estimación directa; $p_i^{(1)}$: estimación con el modelo aleatorio sin "offset".

En los escenarios 2 y 3, donde el rango de las probabilidades $\{\delta_i\}$ es 0.22 y 0.55 respectivamente, y el promedio de las estimaciones de δ es mayor alcanzando valores cercanos a 0.8 en el escenario 3, el porcentaje de veces en que el promedio de las diferencias en valor absoluto es menor con las estimaciones basadas en el modelo que con las estimaciones directas resulta mucho menor y más notoriamente cuando el tamaño de las muestras, n , es grande.

En general, se observa que a medida que aumenta la variabilidad promedio estimada disminuye este porcentaje. Si bien esta tendencia se aprecia para todos los tamaños de muestra, los porcentajes son más elevados para los tamaños muestrales más reducidos.

En síntesis, los resultados hallados muestran que las estimaciones obtenidas bajo el modelo con efectos aleatorios son realmente mejores en casos de poca variabilidad entre las probabilidades y/o tamaños de muestra pequeños.

5. Discusión

En este trabajo se ha evaluado el uso de modelos de efectos aleatorios para estimar probabilidades referidas a áreas o dominios pequeños. Si bien el término área pequeña se usa comúnmente para denotar un área geográfica con un número reducido de observaciones, también puede describir una pequeña subpoblación, como por ejemplo, la correspondiente al grupo de personas de una edad y sexo específicos (Ghosh y Rao, 1994).

En estas condiciones las estimaciones directas o proporciones muestrales no resultan adecuadas ni informativas, por lo que se han propuesto diferentes alternativas metodológicas para enfrentar este inconveniente. Rao (1999) presenta estimaciones indirectas basadas en modelos que utilizan datos suplementarios tales como datos censales o registros administrativos de carácter continuo, complementando el trabajo de Ghosh y Rao (1994) sobre el tema.

Noble et al. (2002) proponen el uso de los denominados modelos lineales generalizados para la obtención de estimaciones indirectas haciendo uso de variables adicionales. En esta línea, Agresti *et al.* (2000) incorporan efectos aleatorios a un modelo lineal generalizado para respuestas categóricas como un mecanismo de estimación de las verdaderas proporciones específicas por área a fin de mejorar las proporciones muestrales.

En este trabajo, siguiendo el enfoque presentado por Agresti et al. (2000), se presenta la estimación de las tasas de desocupación referidas a los jefes de hogar clasificadas por sexo y nivel de instrucción. Es decir, los dominios definidos por las combinaciones de las categorías de las dos características responden al concepto de área pequeña como subpoblación a la manera de Ghosh y Rao (1994). El uso de modelos de efectos aleatorios resulta satisfactorio debido al escaso número de observaciones en alguno de los dominios considerados ya que permite reforzar la estimación particular de cada subgrupo o área pequeña utilizando la información disponible del conjunto de todas las áreas.

El primer modelo proporciona estimaciones orientadas hacia la proporción general de desocupados para el período considerado, mientras que el modelo que incluye "offset" conduce a estimaciones cercanas a la tasa de desocupación previa utilizada como información adicional. En ambos casos se logra un suavizado de las estimaciones con rangos de variación menores que el correspondiente a las estimaciones directas.

El estudio de simulación realizado muestra el comportamiento de las estimaciones obtenidas bajo el modelo de efectos aleatorios ante diferentes escenarios de análisis con la intención de generalizar los resultados. La recomendación que surge del mismo es la de recurrir a estas estimaciones indirectas toda vez que el tamaño de

la muestra de los subgrupos sea reducido y la variabilidad de la respuesta entre ellos no sea demasiado importante.

Referencias Bibliográficas

- Agresti, A. (2002). *Categorical Data Analysis*. Second Edition. New York, USA: John Wiley & Sons.
- Agresti, A.; Booth, J.; Hobert, J.; Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, 30: 27-81.
- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. En B. N. Petrov y F. Csaki, (Eds). *Proceedings of the Second International Symposium on Information Theory*, 267-281. Budapest, Hungary: Akademiai Kiado.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control AC*, 19, 716-723.
- Demidenko, E. (2004). *Mixed Models. Theory and Applications*. New York, USA: John Wiley & Sons.
- Ghosh, M.; Rao, J.N.K. (1994) Small area estimation: an appraisal. *Statistical Sciences*, 9, 55-93.
- INDEC. (2002) *Encuesta Permanente de Hogares*. Base usuaria ampliada de EPH (BUA). Rosario, Argentina. Recuperado el día 25 de abril de 2008 en <http://www.indec.mecon.ar>.
- INDEC. (2003) *La Nueva Encuesta Permanente de Hogares de Argentina*. Rosario, Argentina. Recuperado el día 25 de abril de 2008 en http://www.indec.mecon.ar/nuevaweb/cuadros/4/Metodologia_EPHContinua.pdf.
- McCulloch, C.; Searle, S. (2001). *Generalized Linear and Mixed Models*. New York, USA: John Wiley & Sons.
- Noble, A.; Haslett, S.; Arnold, G. (2002). Small area estimation via generalized linear models. *Journal of Official Statistics*, 18 (1), 45-60.
- Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25 (2), 175-186.
- Rao, J.N.K. (2003). *Small Area Estimation*. New York, USA: John Wiley & Sons.
- SAS Institute, Inc. (2004). *SAS/STAT User's guide*, version 9.1.3 Cary, MC, USA.
- Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.

ANEXO

Tabla A1

Proporciones de desocupación correspondiente al 4º trimestre del 2006 según dominio

Dominio	Proporción de desocupación	Tamaño
Masculino, primaria incompleta	0.067	90
Masculino, primaria completa	0.071	351
Masculino, secundaria incompleta	0.059	136
Masculino, secundaria completa	0.023	256
Masculino, universitario incompleto	0.021	95
Masculino, universitario completo	0.007	154
Femenino, primaria incompleta	0.035	29
Femenino, primaria completa	0.114	88
Femenino, secundaria incompleta	0.097	31
Femenino, secundaria completa	0.095	63
Femenino, universitario incompleto	0.130	54
Femenino, universitario completo	0.063	64