



Hachuel, Leticia
Boggio, Gabriela
Mendez, Fernanda

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística

MODELOS DE REGRESIÓN LOGÍSTICA PARA EL ESTUDIO DE LA DESOCUPACIÓN: ENFOQUES PROMEDIO POBLACIONAL Y GRUPO ESPECÍFICO*

1. INTRODUCCIÓN

En la Argentina, la Encuesta Permanente de Hogares (EPH) ha sido históricamente una herramienta de suma relevancia para el análisis del problema de la desocupación. Surgió como un intento de dar respuesta a las necesidades de información sobre distintas problemáticas sociales, principalmente las relacionadas con el mercado de trabajo. La información se releva a través de una muestra de tipo estratificado, con etapas múltiples de selección, según la cual se eligen en una etapa final hogares encuestándose a todos sus integrantes.

Una consecuencia de tal diseño muestral es que mientras la información de diferentes hogares se puede suponer estadísticamente independiente, las respuestas de los distintos integrantes de un mismo hogar dependen en algún grado unas de otras. Más aun, resulta claro reconocer características medidas en diferentes niveles, a nivel individuo y a nivel hogar. Si los análisis se realizan de acuerdo a métodos estándares, que suponen independencia entre las observaciones tanto entre como dentro de los hogares, los resultados pueden presentar distorsiones.

En el caso particular del estudio de la desocupación a través de una respuesta binaria en su relación con características personales y laborales de los individuos, resulta pertinente la aplicación de modelos de regresión logística. Con el fin de tener en cuenta la correlación inducida por el agrupamiento de los individuos en hogares es útil la aplicación del método "Alternating Logistic Regressions" (ALR) para el ajuste de modelos logísticos marginales. Otra forma alternativa de considerar la posible correlación entre las respuestas es ajustar los denominados modelos logísticos mixtos o con efectos aleatorios, teniendo en cuenta el carácter jerárquico de la información, consecuencia del diseño muestral. Esta modelización, denominada también en niveles múltiples, es una técnica estadística que en términos generales, incluye en el modelo efectos aleatorios relacionados con los diferentes niveles de información (Goldstein, 2003).

Si bien el uso de efectos aleatorios en modelos lineales para variables respuestas con distribución normal está bien estudiado, su consideración para el caso de variables respuestas categóricas es de desarrollo reciente. Los modelos que abarcan variables respuesta no

* Este trabajo fue realizado en el marco del proyecto ECO29 "Modelos de niveles múltiples para respuesta categórica".



normales y efectos fijos y aleatorios en el predictor, constituyen la clase denominada modelos lineales generalizados mixtos –MLGM- (Agresti, 2002; Agresti et al, 2000).

La elección de uno u otro enfoque depende fundamentalmente de los objetivos que se pretenden alcanzar. Si resulta de interés medir el efecto global de las covariables sobre la respuesta promedio, es útil la aplicación de los modelos marginales también llamados promedio poblacionales. Si, en cambio, es de interés medir el efecto particular de las covariables para un hogar específico, es más apropiado el enfoque provisto por los modelos con efectos aleatorios.

Estos dos enfoques se ponen a prueba para modelar la desocupación en el Gran Rosario a partir de la información de la EPH correspondiente a mayo de 2003 y se comparan los resultados. A continuación se presenta la metodología.

2. METODOLOGÍA

Los objetivos principales en la modelización estadística con datos binarios correlacionados son, por un lado, adaptar a este contexto las herramientas de regresión convencionales para relacionar la variable respuesta con variables explicativas y por otro lado, tener en cuenta la correlación entre los sujetos.

Al igual que para datos binarios independientes, los modelos de regresión logística (Hosmer, Lemeshow, 2000) son los más populares para datos binarios correlacionados. Dos importantes metodologías para el análisis de este tipo de datos son el enfoque grupo-específico, o con efectos aleatorios, (GE) y el enfoque promedio poblacional (PP) o marginal (Aerts et al., 2002).

Notación

Los datos agrupados se obtienen midiendo la respuesta de interés, que es una variable aleatoria binaria Y_{ij} , en la j -ésima unidad $j = 1, \dots, n_i$, del i -ésimo grupo, $i = 1, \dots, m$.

Los modelos tipo grupo- específico ó modelos de regresión logística mixtos se explicitan en su forma general:

$$\text{logit}\{P(Y_{ij} = 1 / X_{ij}, \alpha_i, \beta_{GE})\} = \alpha_i + X'_{ij}\beta_{GE} \quad (1)$$

donde α_i es el término intercepto aleatorio para el i -ésimo grupo, X_{ij} es un vector de covariables de interés para el j -ésimo individuo en el i -ésimo grupo y β_{GE} es el correspondiente vector de parámetros.

El vector X_{ij} puede contener tanto componentes a nivel de grupo, $X_{(f)i}$ como también componentes que varían dentro del grupo $X_{(v)ij}$. En correspondencia con estas componentes, β_{GE} se clasifica en $\beta_{GE}^{(f)}$ y $\beta_{GE}^{(v)}$. Con esta notación el modelo (1) se escribe:

$$\text{logit}\{P(Y_{ij} = 1 / X_{ij}, \alpha_i, \beta_{GE})\} = \alpha_i + X'_{(f)i}\beta_{GE}^{(f)} + X'_{(v)ij}\beta_{GE}^{(v)}. \quad (2)$$

La verosimilitud bajo el modelo de regresión logística de efectos mixtos se obtiene integrando la ecuación (2) bajo el supuesto que α_i se distribuye según G . La distribución G más usada es la que supone $\alpha_i \rightarrow N(0, \sigma^2)$ (Stiratelli et al., 1984).

El efecto del término α_i en la ecuación (2) es aumentar la correlación entre respuestas dentro de un grupo en relación a la correlación entre grupos. La idea básica es que las probabilidades logísticas para observaciones de la respuesta en un grupo tengan un valor común α_i . Por lo tanto, si los α_i son muy diferentes, las respuestas de un mismo grupo están más altamente correlacionadas que el conjunto total de respuestas. Es decir, cuanto mayor



es la diferencia entre los α_i , mayor es la correlación intra que entre grupos. La heterogeneidad en los α_i es simplemente una función de su variancia, por lo que la correlación intra-grupo aumenta cuando aumenta σ^2 .

Por otro lado, el modelo general promedio poblacional (PP) establece que una función conocida de la esperanza marginal de la variable dependiente es una función lineal de una o más variables explicativas. El modelo de regresión logística marginal se explicita como:

$$\text{logit}\{P(Y_{ij} = 1 / X_{ij}, \beta_{PP})\} = \alpha_{PP} + X'_{ij} \beta_{PP}, \tag{3}$$

y su análogo a la ecuación (2):

$$\text{logit}\{P(Y_{ij} = 1 / X_{ij}, \beta_{PP})\} = \alpha_{PP} + X'_{(f)ij} \beta_{PP}^{(f)} + X'_{(v)ij} \beta_{PP}^{(v)}. \tag{4}$$

La probabilidad $P(Y_{ij} = 1 / X_{ij}, \beta_{PP})$ representa ahora la distribución marginal de Y_{ij} promediada a través de todos los grupos.

El enfoque de estimación utilizado para este tipo de modelos es el método de ecuaciones generalizadas de estimación o GEE, introducidas por Liang y Zeger (1986). En él se usa un conjunto de ecuaciones que resultan versiones ponderadas de las ecuaciones de verosimilitud utilizadas para datos independientes. Las ponderaciones involucran una aproximación de la matriz de covariancias de las observaciones correlacionadas intra-grupo. Esto requiere realizar supuestos acerca de la naturaleza de la correlación (Hardin, Hilbe, 2003; Diggle et al., 2002; Zeger, Liang, 1992).

Cuando la variable respuesta es binaria, la mejor forma de expresar la asociación entre las respuestas de individuos de un mismo grupo es a través de razones de odds. Así, la razón de odds entre las respuestas de dos individuos distintos, j y j' , en un mismo grupo i resulta:

$$\Psi_{ijj'} = \frac{P(Y_{ij} = Y_{ij'} = 1)P(Y_{ij} = Y_{ij'} = 0)}{P(Y_{ij} = 1; Y_{ij'} = 0)P(Y_{ij} = 0; Y_{ij'} = 1)} \tag{5}$$

Carey et al. (1993) sugirió la alternativa de utilizar la especificación de la asociación a través de estas razones de odds en el método denominado de "Alternating Logistic Regressions" (ALR). El mismo ajusta, en forma alternada, el modelo (4) y un modelo para las razones de odds entre pares de respuestas:

$$\log \Psi_{ijj'} = z'_{ijj'} \alpha, \tag{6}$$

donde $z_{ijj'}$ es un vector de dimensión $qx1$ de covariables que especifican la forma de la asociación entre Y_{ij} e $Y_{ij'}$.

El logaritmo de la razón de odds de dos observaciones cualesquiera de diferentes grupos se supone igual a cero, lo que implica independencia entre ellos. El algoritmo admite diferentes estructuras para los logaritmos de las razones de odds, siendo este enfoque útil cuando la estructura de asociación es en sí misma también de interés.

En síntesis, el enfoque de Carey et al. (1993) caracteriza la dependencia entre respuestas en términos de razones de odds y usa las GEE para estimar los parámetros de regresión alternando con regresiones logísticas de cada respuesta sobre otras en el mismo grupo para actualizar las estimaciones de los parámetros razones de odds.



Interpretación de los parámetros

Para considerar la interpretación de los parámetros en los modelos GE y PP, resulta conveniente tener en cuenta efectos de covariables de nivel grupo y de las que varían dentro del grupo; se supone por simplicidad una única covariable a nivel grupo y una única covariable a nivel sujeto.

Bajo estos supuestos, el parámetro α_i en el modelo GE (2), refleja el log-odds de la respuesta para cualquiera de las j unidades del grupo i , dado que $X_{(f)i}=0$ y $X_{(v)ij}=0$.

En el modelo PP, el parámetro α_{PP} representa el log-odds poblacional de la respuesta en el grupo base, es decir en la población con $X_{(f)i}=0$ y $X_{(v)ij}=0$:

$$\alpha_{PP} = \log \left\{ \frac{P(Y = 1 / X_{(f)} = 0, X_{(v)} = 0)}{1 - P(Y = 1 / X_{(f)} = 0, X_{(v)} = 0)} \right\}. \quad (7)$$

Es decir, la interpretación de los coeficientes resulta diferente según el tipo de modelo que se haya ajustado. Por tal razón, si bien ambos modelos, el específico por grupo y el promedio poblacional, se pueden ajustar a datos que contienen covariables a nivel grupo y a nivel sujeto, la elección del modelo debería considerar qué tipo de inferencia se intenta lograr. El modelo grupo-específico es más útil cuando el objetivo es proveer inferencias acerca de los efectos de covariables medidas a nivel individual. Alternativamente, el modelo promedio poblacional es más útil cuando se intenta inferir acerca de efectos de covariables a nivel grupal.

3. APLICACIÓN

Se utilizan los datos proporcionados por la EPH correspondiente a la onda de mayo de 2003 para el estudio de la desocupación en el Gran Rosario. El esquema muestral de dicha encuesta impone que se registren datos respecto a todos los individuos que conforman cada hogar seleccionado.

Después de realizada las depuraciones pertinentes, la información disponible referida a la población económicamente activa (PEA), es decir personas que trabajan o buscan trabajo con edades comprendidas entre 15 y 65 años de edad, queda conformada por 298 hogares cuya composición se detalla en la siguiente tabla (Tabla 1).

Tabla 1: Número de hogares según sea la cantidad de integrantes pertenecientes a la PEA

Cantidad de individuos de la PEA por hogar	Cantidad de hogares (%)
1	176 (59.06)
2	96 (32.21)
3	19 (6.38)
4	4 (1.34)
5	2 (0.67)
6	1 (0.34)
Total	298 (100.00)



Se observa en la tabla que más del 50% de los hogares tienen sólo un integrante económicamente activo, el 32% tiene dos integrantes pertenecientes a la PEA y es notoria la baja de este porcentaje para tres o más integrantes.

En cuanto a la distribución de los desocupados, el porcentaje global de los mismos en el grupo bajo estudio es de 18%, coincidiendo con el porcentaje de desocupados correspondiente al grupo de hogares con un sólo integrante perteneciente a la PEA. Dicho porcentaje se reduce al 12% si se considera el grupo de hogares con dos integrantes pertenecientes a la PEA.

En base a la información descripta se pretende construir modelos capaces de explicar la probabilidad de desocupación a partir de un conjunto de variables explicativas relacionadas con características personales y laborales. La elección de dichos factores se basa en argumentos económicos, los cuales han sido descriptos en trabajos previos (Servy et al., 2000). Ellos se traducen en las siguientes variables, a saber:

Sexo: femenino, masculino.

Edad: variable continua medida en años.

Escolaridad: primaria incompleta, primaria completa, secundaria incompleta, secundaria completa, nivel superior o universitario incompleto y nivel superior o universitario completo.

Nivel de ingreso: bajo (1º, 2º y 3º decil de ingreso), medio (4º a 7º decil) y alto (8º a 10º decil).

Rama de actividad: construcción, manufacturera, servicios comerciales y de transporte, intermediación financiera, administración pública o defensa e instrucción pública o servicios de salud, otros servicios.

Tamaño de la empresa: unipersonal, de 2 a 5 personas, de 6 a 25 personas, 26 a 100 personas y más de 100 personas.

El análisis estadístico que se presenta a continuación tiene en cuenta la posible correlación de la información introducida por el agrupamiento de los individuos en hogares, de dos maneras diferentes: enfoque marginal y cluster específico. Se emplean para ello los procedimientos GENMOD y NLMIXED de SAS (SAS Institute, 1999) respectivamente.

Modelo marginal

El ajuste del modelo marginal se realiza mediante la metodología ALR descripta en la sección 2. Para ello es necesario determinar una estructura de asociación factible entre las observaciones. Se elige especificar una razón de odds constante entre las respuestas de pares de individuos de un mismo hogar (Preisser et al., 2003). El modelo marginal postulado, con dicha estructura de asociación, incluye las covariables categóricas antes definidas y edad como efectos principales. Respecto de esta última covariable, medida en escala continua, se modela a través de una componente lineal y otra cuadrática a fin de captar los diferentes cambios en la probabilidad de desocupación según sea la edad.

En la Tabla 2 se presentan las estimaciones de los parámetros de regresión y de asociación. Con respecto a los parámetros de regresión, resultan significativos en forma global los relacionados con la escolaridad y rama de actividad de los individuos. Si bien esto no ocurre con el test global para edad, los tests univariados de Wald correspondientes a sus dos componentes resultan levemente significativos. Por tal razón, y al ser la edad una variable medida en escala continua se resuelve graficar la probabilidad de desocupación estimada según la edad para algún perfil de individuos particular. Así por ejemplo para los hom-

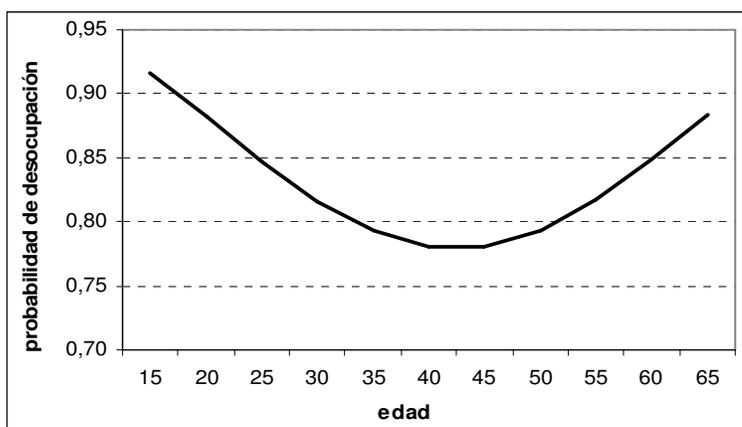


bres con primaria incompleta que trabajan o trabajaron en la rama manufacturera en empresas unipersonales, el gráfico resulta según la Figura 1.

Tabla 2: Estimaciones de los parámetros del modelo logit marginal con asociación intra-hogar constante

Parámetro	Estimación	Error Estándar	P
Constante	2.7449	1.3767	0.0462
SEXO			
Masculino	0		
Femenino	0.4715	0.3292	0.1521
EDAD			0.1249
Edad	-0.1272	0.0624	0.0415
Edad al cuadrado	0.0015	0.0008	0.0439
ESCOLARIDAD			0.0030
Primaria incompleta	0	-	-
Primaria completa	0.3068	0.4698	0.1537
Secundaria incompleta	-0.0937	0.5357	0.8611
Secundaria completa	-0.4686	0.5492	0.3935
Superior o univ. incompleta	1.4195	0.6032	0.0186
Superior o univ. completa	-0.4318	0.8093	0.5937
RAMA DE ACTIVIDAD			0.0000
Manufactura	0	-	-
Construcción	-1.4156	0.5949	0.0173
Servicios comerciales	-2.2626	0.6357	0.0004
Intermediación financiera	-3.2905	0.9619	0.0006
Adm.Pública y Def. eIns. Pub.	-3.8855	0.8159	0.0000
Otras activ. de servicio	-2.9819	0.6745	0.0000
TAMAÑO DE LA EMPRESA			0.1605
Unipersonal	0	-	-
2-5 personas	0.3309	0.3644	0.3639
6-25 personas	-0.4969	0.5156	0.3351
26-100 personas	0.8187	0.5562	0.1410
101 o más personas	-1.0559	0.8578	0.2183
Log asoc.intra-hogar	1.2128	0.6425	0.0591

Figura 1: Probabilidad estimada de desocupación para hombres con primaria incompleta de la rama manufacturera en empresas unipersonales, según edad



Como era de esperar la forma de la curva muestra una disminución en la probabilidad estimada de desocupación hasta alrededor de los 40 años, momento a partir del cual comienza nuevamente a ascender.

Con respecto al nivel de instrucción, su influencia sólo se muestra significativa cuando se compara uno de los niveles más altos de escolaridad respecto al nivel más bajo. Así la chance de desocupación para los individuos con nivel superior o universitario incompleto es cuatro veces mayor ($\exp(1.4195) = 4.13$) que para los individuos sin primaria completa. Este resultado hace pensar que las personas con alto nivel de instrucción pero que no lograron formalizarlo en un título habilitante tienen dificultades para el acceso laboral aún mayores que las personas con escasa instrucción.

En relación a la rama de actividad, el riesgo de desocupación resulta diferente en promedio según cuál sea dicha rama, siendo la situación más desfavorable la correspondiente a manufactura, a diferencia de lo comprobado en estudios similares para el período 2001-2002 (Hachuel et al, 2003) donde la construcción era la rama de actividad con mayor probabilidad de desocupación. Además en ambos trabajos se pudo verificar que las ramas administración pública y defensa, instrucción pública e intermediación financiera presentan las menores tasas de desocupación. Puntualmente los resultados encontrados en este caso permiten afirmar que las chances de desocupación para las personas que trabajan o trabajaron en la construcción son ($\exp(-1.4156) = 0.24$) alrededor del 75% menor que para las personas dedicadas a la rama manufacturera.

Resta mencionar que ni el sexo ni el tamaño de la empresa donde trabaja o trabajó el individuo tuvieron un efecto significativo sobre la probabilidad de desocupación.

En relación a la estructura de dependencia, la falta de contundencia en la significación de la asociación intra-hogar ($p = 0.05981$), indicaría que los factores considerados en el modelo postulado no resultan del todo suficientes para explicar la asociación entre individuos de un mismo hogar. En este sentido, análisis adicionales permitieron comprobar que la inclusión en el modelo de una covariable medida a nivel hogar, como lo es el nivel de ingreso



familiar, elimina completamente la significación de la asociación intra-hogar. Este resultado muestra que esta variable medida a nivel del hogar e imputada con el mismo valor a cada integrante del mismo, alcanza para mitigar la asociación entre las respuestas de los integrantes del hogar.

Modelo con efectos aleatorios

Bajo este enfoque grupo específico la posible asociación entre las respuestas de los individuos de un mismo hogar se tiene en cuenta mediante la incorporación de un efecto aleatorio específico para cada uno de ellos. Los resultados del ajuste del modelo con el mismo conjunto de covariables considerado en el enfoque marginal se presentan en la siguiente tabla (Tabla 3).

Tabla 3: Estimaciones de los parámetros del modelo logit con efecto aleatorio para cada hogar

Parámetro	Estimación	Error Estándar	P
Constante	3.3982	1.8053	0.0608
SEXO			
Masculino	0	-	-
Femenino	0.5579	0.4053	0.1697
EDAD			
Edad	-0.1625	0.0853	0.0578
Edad al cuadrado	0.0020	0.0011	0.0671
ESCOLARIDAD			
Primaria incompleta	0	-	-
Primaria completa	0.4654	0.5819	0.4235
Secundaria incompleta	0.2022	0.6535	0.7572
Secundaria completa	-0.4585	0.6679	0.4930
Superior o univ. incompleta	1.8562	0.8051	0.0218
Superior o univ. completa	-0.3456	0.8969	0.7003
RAMA DE ACTIVIDAD			
Manufactura	0	-	-
Construcción	-1.7427	0.8206	0.0345
Servicios comerciales	-2.7703	0.8365	0.0010
Intermediación financiera	-3.8991	1.2047	0.0013
Adm. Pública y Def.e Ins. Pub.	-4.7301	1.2233	0.0001
Otras activ. de servicio	-3.6389	0.9652	0.0002
TAMAÑO DE LA EMPRESA			
Unipersonal	0	-	-
2-5 personas	0.3824	0.4207	0.3641
6-25 personas	-0.5753	0.5952	0.3346



26-100 personas	0.9991	0.6513	0.1261
101 o más personas	-1.1387	0.9998	0.2556
Variancia del efecto aleatorio	1.4198	1.2964	0.2743

Los resultados hallados son similares a los encontrados bajo el enfoque marginal, en el sentido de resultar significativos los efectos del nivel de escolaridad y rama de actividad y perdiendo fuerza la significación de la edad del individuo en sus dos componentes.

Con respecto al nivel de instrucción, al igual que bajo el enfoque marginal, su influencia sólo se muestra significativa cuando se compara uno de los niveles más altos de escolaridad respecto al nivel más bajo. Así en un determinado hogar, la chance de que un integrante con nivel superior o universitario incompleto esté desocupado es seis veces mayor ($\exp(1.8562) = 6.40$) que si ese individuo no hubiese completado la escuela primaria.

Con respecto a rama de actividad el signo negativo de sus coeficientes muestra que todas las ramas de actividad consideradas tienen menor probabilidad de desocupación que la manufacturera. Así por ejemplo, el coeficiente correspondiente a la rama de la construcción indica que una persona, integrante de un determinado hogar, que se declaró perteneciente a la rama de la construcción tiene una chance de estar desocupada un 82% ($\exp(-1.7427) = 0.18$) menor que si desarrollara sus actividades en la rama de la manufactura.

Por otro lado, la leve asociación intra-hogar hallada en el análisis marginal se traduce en este caso en una estimación no significativa de la variancia del efecto hogar aleatorio ($p = 0.2743$).

También bajo este enfoque se considera el efecto de una covariable a nivel hogar: nivel de ingreso. En este caso se produce una estimación nula de la variancia de la componente aleatoria del modelo, situación que puede ser provocada por una correlación intraclase muy pequeña. Este resultado ratifica el ya hallado bajo el enfoque marginal cuando la inclusión de esta variable hacía que la asociación intra-hogar perdiera su significación.

Cabría entonces preguntarse por qué recurrir a estas metodologías que tienen en cuenta la asociación intra-hogar cuando no parece ser demasiado importante, traducida en una leve significación de la razón de odds entre respuestas de pares de individuos del mismo hogar en el enfoque ALR y en una variancia del efecto aleatorio no significativa en el modelo mixto. La respuesta es posible encontrarla si se ajusta un modelo de regresión logística convencional que presupone la independencia entre las respuestas de los diferentes individuos. Sabido es el hecho que ignorar la correlación entre las respuestas provoca un sobre-dimensionamiento de los efectos. Los resultados hallados en este último ajuste así lo corroboran produciendo significación estadística en el efecto global asociado a tamaño de la empresa (Tabla A.1).

4. DISCUSIÓN

Los muestreos multietápicas, elegidos frecuentemente por cuestiones prácticas y presupuestarias, pueden introducir correlación entre las observaciones producto del agrupamiento de los individuos en unidades mayores, por ejemplo, en hogares.

El análisis de este tipo de datos, correlacionados por grupo (hogar) presenta desafíos metodológicos que son abordados satisfactoriamente mediante los modelos lineales generalizados mixtos y las ecuaciones de estimación generalizadas, que se relacionan con los enfoques denominados grupo-específico y promedio poblacional respectivamente.



Dentro de este último enfoque, el método ALR resulta apropiado en caso de modelar una respuesta de tipo binaria ya que mide la asociación intra-grupo mediante razones de odds entre las respuestas de pares de individuos de un mismo hogar.

La puesta a prueba de estas metodologías en datos sobre desocupación provenientes de la EPH resultó de interés ya que el esquema muestral, hacía pensar en una posible correlación entre las respuestas de los individuos integrantes de un mismo hogar. Estos modelos permitieron el uso eficiente y apropiado de los datos, en la medida en que la inferencia sobre las covariables se ajusta por la correlación intra-hogar, de una manera que resulta consistente con la forma en que fueron obtenidos los datos.

En la aplicación realizada sólo se incluyen en los modelos covariables que cambian de individuo a individuo dentro del hogar, es decir no se consideran covariables que traduzcan características propias del hogar y no de los individuos en particular. Por ello parecería que es más apropiado el uso de modelos bajo el enfoque hogar-específico ya que el enfoque marginal proporciona estimaciones de razones de odds promediadas a través de los hogares.

Si bien es sabido que los modelos del tipo grupo-específico proporcionan coeficientes de mayor valor absoluto que los del tipo promedio-poblacionales, los resultados hallados son similares en los valores de los coeficientes de los modelos ajustados. Ello es consecuencia de la leve asociación intra-hogar, traducida en una razón de odds entre las respuestas de pares de individuos de un mismo hogar cercana a uno en el caso de los modelos marginales y en una estimación de la varianza del efecto aleatorio del modelo no significativa en el caso de los modelos grupo-específico.

REFERENCIAS

- Aerts, M.; Geys, H.; Molenberghs, G.; Ryan, L. 2002. *Topics in Modelling of Clustered Data*. Chapman & Hall/CRC.
- Agresti, A. 2002. *Categorical Data Analysis*. John Wiley & Sons.
- Agresti, A.; Booth, J.; Hobert, J.; Caffo, B. 2000. Random-effects modeling of categorical response data. *Sociological Methodology*, 30: 27-81.
- Carey, V.; Zeger, S. DIGGLE, P. 1993. Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, 80(3): 517-526.
- Diggle, P. J.; Heagerty, P. ; Liang K-Y.; Zeger S. L. 2002. *Analysis of Longitudinal Data*, 2nd ed. Oxford University Press.
- Hachuel, L.; Boggio, G.; Méndez, F. 2003. Modelos Marginales para el Estudio de la Desocupación en Rosario 2001-2002. [CD-ROM] XXXI Coloquio Argentino de Estadística. Sociedad Argentina de Estadística.
- Hardin, W. H.; Hilbe, J. M. 2003. *Generalized Estimating Equations*. Chapman & Hall.
- Hosmer, D. W. & Lemeshow, S. 2000. *Applied Logistic Regression*, 2nd ed. John Wiley & Sons. New York.
- Liang, K.; Zeger, S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.
- Preisser, J.; Arcury, T.; Quandt, S. 2003. Detecting patterns of occupational illness clustering with alternating logistic regressions applied to longitudinal data. *American Journal of Epidemiology*, 158: 495-501.



SAS Institute, Inc. 1999. SAS/STAT user's guide, version 8. Cary, NC :SAS Institute, Inc.

Servy, E.; Hachuel, L.; Boggio, G.; Cuesta, C; Leone, G. 2000. Modelos estadísticos para el estudio de la desocupación. Parte I: Modelos para cortes transversales. *Cuadernos del IITAE* N° 5. Escuela de Estadística. Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario.

Stiratelli, R. ; Laird, N. ; Ware, J.H. 1984. Random-effects model for serial observations with binary response. *Biometrics*, 40: 961-971.

Zeger, S. L.; Liang, K. 1992. An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, 11: 1825-1839.

ANEXO

Tabla A1: Estimaciones de los parámetros del modelo logit convencional

Parámetro	Estimación	Error estándar	P
Constante	3,2439	1,4815	0,0286
SEXO			
Masculino	0	-	-
Femenino	0,4579	0,3405	0,1787
EDAD			
Edad	-0,1514	0,0695	0,0294
Edad al cuadrado	0,0018	0,0009	0,0393
ESCOLARIDAD			
Prim incomp	0	-	-
Prim comp	0,4447	0,4892	0,3633
Sec incomp	0,1677	0,5508	0,7608
Sec omp	-0,4282	0,5658	0,4491
Sup o univ incomp	1,5020	0,6203	0,0155
Sup o univ comp	-0,4143	0,7831	0,5968
RAMA DE ACTIVIDAD			
Manufactura	0	-	-
Construccion	-1,5321	0,6548	0,0193
Ss comerciales	-2,2804	0,6181	0,0002
Interm financiera	-3,2422	0,9060	0,0003
Adm Pub y Def + Ins Pub	-3,6784	0,7782	0,0000
Otras activ de ss	-3,0796	0,7052	0,0000
TAMAÑO DE LA EMPRESA			
Unipersonal	0	-	-
2-5 pers	0,2969	0,3478	0,3933
6-25 pers	-0,5733	0,5088	0,2599
26-100 pers	0,8579	0,5248	0,1021
101 o mas pers	-1,1873	0,8783	0,1764