

Clasificación por niveles socioeconómicos de los hogares entrevistados para la Encuesta Nacional de Salud y Nutrición 2005/06: nota metodológica

Juan Pablo Gutiérrez¹

Introducción

Como parte del análisis de la Encuesta Nacional de Salud y Nutrición 2005/06 (Ensanut), se ha considerado la agrupación de hogares de acuerdo a su nivel socioeconómico, para lo cual es necesario contar con un mecanismo de clasificación que identifique el estatus de cada hogar con base en la información recolectada por la encuesta. Si bien el cuestionario incluyó información sobre el gasto del hogar, el cual se ha utilizado en diversos estudios como una medida de nivel socioeconómico, dado que por la naturaleza de la encuesta esta información no se obtuvo de forma exhaustiva, se consideró la necesidad de desarrollar un indicador socioeconómico que considerara en mayor amplitud la información recolectada, para a partir de la misma, construir un índice de nivel socioeconómico.

El punto de partida para la generación de un sistema de clasificación es que la misma no parte de un supuesto teórico a probar, pero de una intención de empíricamente maximizar la capacidad predictiva para identificar correctamente una característica desconocida a partir de una base de datos. Es decir, la validación esencial del sistema es su capacidad predictiva, medida con datos conocidos.

La metodología que se plantea en esta nota se basa en un trabajo previo realizado con el fin de clasificar por deciles de ingreso a las familias, la cual fue modificada tanto en función de la información disponible como de las necesidades específicas del análisis de la Ensanut.

¹ Centro de Investigación en Evaluación y Encuestas, Instituto Nacional de Salud Pública.
Contacto: jpgutier@insp.mx

Antecedentes

En términos generales, el estatus socioeconómico es típicamente usado como una medida que agrupa variables que caracterizan la condición de las personas, las familias y los hogares respecto a su capacidad de crear o consumir bienes de valor en la sociedad.^{i,ii} En sentido estricto, si se pretende categorizar a individuos o familias de acuerdo a su nivel socioeconómico es necesario seleccionar el indicador relevante y a partir de éste estimar los puntos de corte que permitan generar una clasificación con el número de grupos deseado. El indicador seleccionado se considera como una medida de la riqueza de los hogares, la que puede entenderse a su vez como un indicador del bienestar del hogar.

Existen diferentes medidas que pueden funcionar como indicadores de riqueza, entre ellas se encuentran el ingreso, el gasto y los bienes acumulados en el hogar.

Además, al calcular el nivel socioeconómico de un hogar, es necesario considerar aspectos vinculados a la composición del mismo y el nivel de ingreso total, es decir el ingreso que perciben en conjunto todos sus miembros. Sin embargo, muchas de las principales mediciones del nivel socioeconómico (NSE) no están disponibles o no son confiables, en especial para las economías en desarrollo, donde una parte importante de las transacciones se desarrollan fuera del mercado económico²

Es importante señalar que, para ejercicios de este tipo, la estimación exacta del monto del ingreso es menos relevante que la asignación en el decil apropiado, ya que aun con un margen de error importante al estimar el ingreso es posible clasificar correctamente a una familia o individuo.

El método más utilizado para predecir la distribución de las familias en los deciles ha sido asignar medidas o pesos a cada una de las variables observadas de acuerdo a ciertas consideraciones económicas (asignar un valor monetario a bienes durables) así como herramientas estadísticas (por ejemplo, análisis de

² Fuera del Mercado se refiere a transacciones en el mercado informal o mercado negro, tales como trueques, intercambios o actividades ilícitas

componentes principales), con lo cual se obtiene una medida univariada de bienestar.

Idealmente, la identificación de las familias requeriría examinar diversos aspectos: el bienestar de cada integrante, una consideración minuciosa de sus ingresos, la estructura familiar, los precios locales, el valor de los bienes personales, etcétera, y posteriormente comprobar la información obtenidaⁱⁱⁱ.

Debido a las limitaciones para la recolección de información detallada en encuestas temáticas (por los costos y tiempo necesario para obtenerla), resulta deseable contar con una metodología que permita estimar el nivel SE de forma parsimoniosa.

Este documento describe un abordaje metodológico para imputar el nivel SE (definido como nivel de ingreso por adulto equivalente de los hogares) a la Encuesta Nacional de Salud y Nutrición 2006.

Metodología

La clasificación se realizó a partir de una estimación por fases, utilizando modelos tipo logit y logit ordenado. En un primer momento, se estimó la probabilidad de los hogares de encontrarse en pobreza alimentaria; del conjunto de hogares no en pobreza alimentaria, se estimó la probabilidad de encontrarse en pobreza patrimonial. Finalmente, para los hogares no en pobreza patrimonial, se estimó la probabilidad de ubicarse en los 6 deciles restantes.

Las variables incluidas en los modelos para identificar estas probabilidades fueron pre-identificadas del ejercicio realizado previamente para el Seguro Popular, incluyendo todas aquellas que se encontraban disponibles en la Ensanut. Con el conjunto de variables seleccionadas, se revisaron los pesos específicos de cada una, replicando la metodología utilizada previamente, y que consistió en estimar los modelos con los hogares incluidos en la Encuesta Nacional de Ingreso y Gasto de los Hogares 2006.

En un primer momento, los modelos se estimaron utilizando el 50% de la muestra de la Enigh, y utilizando el restante 50% para clasificar. El sistema se llevó a cabo en las dos direcciones (modelando con un 50% y luego con el otro

50%), con resultados similares, por lo que se optó para los pesos finales utilizar la muestra completa de la Enigh.

Variables generadas

Como parte del análisis, se generaron las siguientes variables a partir de la información recolectada: número de adultos equivalentes, número de menores y adultos trabajando, índice de dependencia demográfica, índice de dependencia económica.

Variables de localidad

Para considerar el peso del lugar de residencia, y como ya se mencionó, se incluyó en el modelo el índice de marginación elaborada por el Consejo Nacional de Población (Conapo) a partir de los resultados del Censo de Población y Vivienda. Asimismo, se incluyeron características meteorológicas de las localidades, que buscaban incluir una medida relativa para balancear la información de ciertos bienes y características del hogar.

Principales limitaciones

El modelo parece funcionar adecuadamente para los deciles en los extremos, y con menor precisión en el centro de la distribución. Una explicación posible es que si bien hay una clara diferenciación entre las características de los hogares en los extremos, esta es mucho más difusa en los deciles intermedios. De esta forma, la asignación de los hogares en los deciles intermedios tiene un mayor grado de error.

Abordajes alternativos

Considerando la principal limitación del modelo, se exploraron alternativas metodológicas discutidas con expertos para buscar mejorar la clasificación. El abordaje principal fue buscar saturar el modelo, buscando incrementar la capacidad predictiva del mismo. No obstante, las estimaciones con interacciones resultaron en modelos no válidos. Modelos con menos interacciones resultaron en estimaciones con capacidad predictiva similar a la original, por lo que se optó por el sistema propuesto.

Implementación

En las tablas siguientes se presentan las variables incluidas en las 3 fases del sistema, con sus pesos respectivos y puntos de corte para la clasificación de los hogares en los 8 grupos: pobreza alimentaria, pobreza patrimonial y deciles 5 a 10.

Fase 1: Selección de pobres alimentarios (punto de corte para ser pobre alimentario ≥ 0.25)

Variables	Coeficiente
Características del jefe de familia	
Años de educación del jefe	-0.07669
Sexo del jefe	0.07321
El jefe trabajo el mes pasado	0.58610
Características de estructura sociodemográfica	
Total de personas que trabajan	-0.60183
Número de adultos equivalentes	0.69075
Número de niños que trabajan	0.58102
Número de adultos que trabajan	0.12957
Índice de dependencia económica	-0.97643
Características de la vivienda	
No se duerme en la cocina	-0.10457
Baño con conexión de agua	-0.35034
No usa leña ni carbón	-0.68568
No tienen piso de tierra	-0.26618
Bienes del hogar	
Tiene TV a color	-0.39853
Tienen micro-ondas	-0.58174
Tiene lavadora	-0.42176
Tienen aire acondicionado	-0.20951
Tienen vehículo de motor	-0.66498
Tiene boiler	-0.39581
Tienen VCR	-0.25123
Tienen licuadora	-0.03162
Tienen refrigerador	-0.31544
Tienen ventilador	-0.23103
Tienen línea telefónica	-0.20166
Patrones familiares de consumo	
Realizaron alguna comida fuera del hogar	-0.54784
Usaron transporte	-0.56138
Pagaron por cine	-1.77126
Pagan renta por la vivienda	0.45735
Características de la localidad de residencia	
Temperatura máxima de la localidad	0.00065
Temperatura mínima en la localidad	0.04186
Altitud de la localidad	0.00007
Marginación alta	-0.40205
Marginación media	-0.34283
Marginación baja	-0.77325
Marginación muy baja	-1.19753
Constante	0.58882

Fase 2: Selección de pobres patrimoniales (punto de corte para ser pobre patrimonial ≥ 0.4)

Variables	Coefficiente
Características del jefe de familia	
Años de educación del jefe	-0.09724
Sexo del jefe	0.11782
El jefe trabajo el mes pasado	0.69698
Características de estructura sociodemográfica	
Total de personas que trabajan	-0.35952
Número de adultos equivalentes	0.65253
Número de niños que trabajan	0.40728
Número de adultos que trabajan	0.17768
Índice de dependencia económica	-1.29430
Características de la vivienda	
No se duerme en la cocina	0.00987
Baño con conexión de agua	-0.18194
No usa leña ni carbón	-0.22405
No tienen piso de tierra	-0.20649
Bienes del hogar	
Tiene TV a color	0.08966
Tienen micro-ondas	-0.47921
Tiene lavadora	-0.34565
Tienen aire acondicionado	-0.45103
Tienen vehículo de motor	-0.84546
Tiene boiler	-0.44395
Tienen VCR	-0.20246
Tienen licuadora	0.05148
Tienen refrigerador	-0.25494
Tienen ventilador	-0.07042
Tienen línea telefónica	-0.46254
Patrones familiares de consumo	
Realizaron alguna comida fuera del hogar	-0.33050
Usaron transporte	-0.27942
Pagaron por cine	-0.61463
Pagan renta por la vivienda	0.14796
Características de la localidad de residencia	
Temperatura máxima de la localidad	0.00399
Temperatura mínima en la localidad	0.00414
Altitud de la localidad	0.00006
Marginación alta	-0.36946
Marginación media	-0.55441
Marginación baja	-0.52816
Marginación muy baja	-0.80893
Constante	0.55706

Fase 3: Selección deciles 5 a 10³

Variables	Coefficiente
Características del jefe de familia	
Años de educación del jefe	0.13185
Sexo del jefe	-0.09702
El jefe trabajo el mes pasado	-0.83167
Características de estructura sociodemográfica	
Total de personas que trabajan	0.32236
Número de adultos equivalentes	-0.88575
Número de niños que trabajan	-0.92205
Índice de dependencia económica	1.25931
Características de la vivienda	
Número de cuartos	0.29565
Bienes del hogar	
Tiene estereo	0.15219
Tiene TV a color	-0.00812
Tienen micro-ondas	0.29587
Tienen aire acondicionado	0.24884
Tienen vehículo de motor	0.77125
Tienen computadora	0.68497
Tienen ventilador	0.10527
Tienen línea telefónica	0.54770
Patrones familiares de consumo	
Realizaron alguna comida fuera del hogar	0.50721
Compraron cigarros	0.27816
Usaron transporte	-0.16476
Pagaron por cine	0.65071
Pagan renta por la vivienda	0.02323
Características de la localidad de residencia	
Temperatura máxima de la localidad	-0.03727
Altitud de la localidad	-0.00014
Marginación alta	-0.95043
Marginación media	-0.77041
Marginación baja	-0.91005
Marginación muy baja	-0.63206

ⁱ. Hauser R.M., Warren J.R., **Socioeconomic Indexes for Occupations: A Review, Update, and Critique.** *Sociological Methodology*, Vol. 27 (1997), 177-298

ⁱⁱ. Duncan, O.D.(1961) **A socioeconomic index for all occupations.** In A.J. Reis, Jr. (Ed), *Occupations and social status* (pp. 109-138), New cork: Free Press.

³ resto2	Probability	Observed
5	Pr($xb+u<_{cut1}$)	0.1304
6	Pr($_{cut1}<xb+u<_{cut2}$)	0.1516
7	Pr($_{cut2}<xb+u<_{cut3}$)	0.1673
8	Pr($_{cut3}<xb+u<_{cut4}$)	0.1760
9	Pr($_{cut4}<xb+u<_{cut5}$)	0.1864
10	Pr($_{cut5}<xb+u$)	0.1883

ⁱⁱⁱ. Grosh M.E. **De los Principios a la Práctica: La Focalización de Programas Sociales en América Latina.** *América Latina y el Caribe, Programa de Estudios Regionales*, Informe No. 21, Vol. 1, septiembre de 1992.