

SEGMENTACIÓN SOCIAL DE LOS HOGARES A LA LUZ DEL CENSO 2000 ¹

Aplicación a los hogares del Marco Muestral de Viviendas

M.Sc. Cathalina García Santamaría²
Unimer Research International
Julio del 2002
cgarcia@unimerri.com

RESUMEN

El objetivo del estudio fue segmentar los hogares del Censo 2000 de acuerdo con la información demográfica contenida en el formulario. Usualmente la clasificación socioeconómica de los hogares, se refiere a elementos de la economía del hogar, específicamente. En este caso, se construyeron indicadores que proveen información sobre la composición por edad y nivel educativo del hogar, así como sobre su ciclo de vida. Se tomaron en consideración también algunos indicadores y sistemas de codificación definidos en el artículo de Johnny Madrigal “Estratificación de hogares y segmentos por niveles de ingreso en los Censos 2000”. Para ello se escogieron diecinueve variables, cuya estructura factorial reportó buenas adecuaciones muestrales. Este análisis definió nueve dimensiones a saber: a) Miembros con posibilidad de aportar al ingreso familiar; b) Desarrollo social; c) Orientación a educación; d) Fortaleza del jefe del hogar; e) Tipología del hogar y f) Movilidad del jefe. Con base en los indicadores se corrió el procedimiento Cluster con el algoritmo K-Means del SPSS versión 9.0 y se agruparon los hogares en 10 segmentos. Posteriores análisis permiten reagrupar esos segmentos en un menor número de categorías según el interés del investigador.

¹ Borrador para discusión – No citar.

² Cathalina García es Directora General Técnica de Unimer Research International desde el año 1997.

JUSTIFICACIÓN

Este documento se refiere a una segmentación de mercado demográfica basada en el Censo del 2000, la cual se pretende introducir dentro del marco muestral como un indicador para estratificación socio-demográfica de los hogares.

La segmentación de mercados describe la división de un mercado en grupos homogéneos que van a responder de forma diferente a promociones, comunicaciones, publicidad y otras variables de mercadeo. Cada grupo o segmento puede ser atendido de una forma diferente, porque los segmentos son creados para minimizar las diferencias inherentes entre los respondientes dentro del segmento y maximizar las diferencias entre los segmentos.

La segmentación de mercados es un término que aparece por primera vez en los años 1950, cuando la diferenciación de productos era la principal estrategia de mercado utilizada. En los años 1970 y 1980, la segmentación de mercados comenzó a tomar importancia como un medio para incrementar las ventas y obtener ventajas competitivas. En los años 1990, el mercadeo directo usó técnicas más sofisticadas para ofrecer a sus potenciales compradores o usuarios un producto a la medida (DSS Research, 2001).

Hay buenas razones para dividir un mercado en grupos pequeños (Peter & Olson, 1993). Las principales son:

- Facilitar el mercadeo. Es más fácil atender las necesidades de grupos de consumidores, clientes o usuarios, particularmente si tienen muchas características en común (buscan los mismos beneficios, tienen la misma edad, pertenecen al mismo género, etc.)
- Encontrar nichos de mercado. Identificar mercados que no están siendo atendidos o lo son parcialmente. Usando el mercado orientado a nichos, la segmentación puede ayudar a una empresa o un producto nuevos a orientar su

estrategia hacia esos compradores y ayuda a un producto o una empresa maduros a buscar nuevos compradores.

- Eficiencia. La posibilidad de orientar los recursos de mercadeo hacia los segmentos que mejor pueden recibir su oferta (productos, precios, promociones, locaciones) puede aumentar la eficiencia de los mismos. La segmentación puede ayudar para evitar que se envíe el mensaje equivocado o que se envíe el mensaje a un grupo equivocado.

No hay una forma simple para determinar las bases más relevantes para proceder a segmentar un mercado. Algunas dimensiones que inicialmente pueden ser de interés, se determinan a partir de estudios previos, tendencias de ventas y compras y juicio gerencial. Cuatro estilos clásicos de segmentación son: (Peter & Olson, 1993)

- Beneficios. Este estilo trata de medirle sistema de valores del consumidor y las percepciones acerca de las marcas en una categoría de productos. Los individuos son distribuidos en grupos según los beneficios que buscan al consumir un determinado producto.
- Psicográfica. En este caso, el mercado se divide en grupos según los estilos de vida del consumidor, comprador o usuario. Los estilos de vida se miden a través de las actividades, intereses y opiniones que tienen los individuos.
- Persona / Situación. Aquí se usan las diferencias individuales de los consumidores en conjunto con la situación de uso del producto, es decir, se tienen personas particulares en situaciones de uso particular.
- Demográfica. Aquella que se basa en información de las personas o de sus hogares.

El Censo de Población del 2000 constituye una oportunidad para efectuar un ejercicio de segmentación socio-demográfica de las familias, el cual podría facilitar la labor de investigación de mercados en nuestro país.

Ejercicios similares han sido efectuados por otras empresas. En enero del 2002, Maxwell Statistics Corporation publicó un reporte de una segmentación demográfica de personas de 15 años o más, basada en la muestra representativa de estadounidenses que usó el U.S. Census Bureau para su estudio *Current Population Survey (CPS)* en octubre del 2001. Estos segmentos se incluyen en sus estudios de mercado para clientes, para efectos de realizar cruces de variables o para seleccionar aquellos que tengan mayor potencial para el cliente (Maxwell, 2002). Maxwell utilizó variables como:

- Edad
- Grupo étnico
- Número de niños
- Tamaño del hogar
- Servicio militar
- Nivel educativo
- Ingreso del hogar
- Participación de la fuerza laboral
- Número semanal de horas de trabajo

Y con base en ellas lograron identificar seis segmentos: padres, hispanos, empleados, minorías, pensionados y veteranos.

Por su parte, Claritas Inc. en conjunto con Geographic Data Technology Inc. ofrece la última base de datos disponible del Censo 2000, en Estados Unidos, con una variedad de reportes además de su adopción de la metodología de mapas digitales MapInfo MapXtreme. Esta base está segmentada según diversos criterios: grupos étnicos, tenencia de vivienda, géneros, grupos de edad, etc. El concepto es que los mejores grupos de consumidores o usuarios potenciales, se definen en primer lugar, de acuerdo con sus características demográficas: edad, ingreso, ocupación, educación, tamaño del hogar y otros. Estos criterios pueden ser ligados a requerimientos específicos de los clientes usando encuestas propias sobre estilos de vida, uso de tiempo libre, consumo de categorías, etc.

De acuerdo con esas experiencias, segmentaciones básicas pueden construirse a partir de los indicadores individuales, pero también se pueden lograr resultados más completos mediante la combinación de variables.

OBJETIVOS Y ESTRATEGIA

El objetivo de este estudio es construir una segmentación demográfica de los hogares costarricenses.

Para ello se trabajó con una muestra de 20,000 hogares del Censo 2000, distribuidos en todo el país.

A partir de ahí, el ejercicio debe ampliarse para ligarse al marco muestral de viviendas 2000 y hacer una estratificación demográfica de segmentos en el tamaño actual o ampliados, que sean de utilidad tanto en investigación de mercados como en mercadeo. Esa estratificación debe llevarse a una referencia digital en mapas, no solamente para facilitar su representación como también su análisis.

Para cumplir con estos objetivos, primero hay que enfrentar la ausencia de información sobre ingreso en el Censo 2000, tal y como lo indica Madrigal en su artículo. Por ello utilizaremos sus resultados, específicamente en la construcción del indicador de ocupación del jefe del hogar.

La estrategia general para realizar este trabajo se define en los siguientes pasos:

- Construcción de indicadores referentes a:
 - Hogar
 - Jefe del hogar
 - Vivienda

- Resumen de dimensiones medidas por los indicadores, mediante el uso de análisis factorial
- Construcción de los segmentos de hogares, usando el algoritmo Cluster K-Means del SPSS.
- Definición del perfil de los segmentos.

DETALLES METODOLOGICOS

El Instituto Nacional de Estadística y Censos (INEC) suministró una muestra de 20000 hogares, seleccionados sistemáticamente del marco muestral, ordenado según la división territorial administrativa. El subarchivo contenía variables correspondientes a la vivienda, al jefe del hogar y variables resumen del hogar. El INEC excluyó los hogares 2 de las 20000 viviendas, lo que corresponde a 400 hogares.

El uso de una muestra permite un manejo más expedito de los procesamientos en SPSS, pero tiene la consecuencia de que los resultados aquí obtenidos deben ser trasladados al Marco Muestral completo. Además, esto orientó hacia un diseño simple de los indicadores, de manera que el cálculo a nivel del marco completo no sea engorroso.

Como se indica arriba, para construir los índices se utilizaron procedimientos simples, en general de suma de puntos o de cálculo de razones. En el caso de la ocupación del jefe se utilizó el resultado del procedimiento aplicado por Johnny Madrigal en su artículo “Estratificación de hogares y segmentos por niveles de ingreso en los Censos 2000”, por cuanto la variable está correlacionada positivamente con el ingreso per cápita del hogar (Madrigal, 2002).

Mediante el procedimiento FACTOR del SPSS se evaluaron las relaciones entre los indicadores. Este tipo de análisis multivariado tiene el objetivo de descubrir las dimensiones subyacentes de un conjunto de variables. Reduce el espacio de atributos, constituido por los indicadores propuestos, a un número más pequeño de factores. Uno

de los objetivos para aplicar esta técnica estadística es la identificación de grupos o clasificación de casos.

No requiere de una teoría a priori y se utilizan las cargas factoriales para intuir la estructura factorial de los datos.

Hay varios métodos para identificar o extraer los factores. En este caso se utilizó el Análisis de componentes principales y la rotación usada fue la de Normalización Varimax con Kaiser. La rotación sirve para hacer que los resultados obtenidos del análisis sean más comprensibles y sea más fácil interpretar los factores. La rotación VARIMAX rota los ejes de los factores, que tiene el efecto de diferenciar el aporte de las variables originales a los factores extraídos. Es decir, minimiza el número de variables que “cargan alto” en un determinado factor.

El análisis de componentes principales busca una combinación lineal de variables que maximice la varianza extraída de las variables. Una vez que construye la primera combinación lineal, remueve esa varianza y procede a buscar una segunda combinación lineal que maximice la proporción que resta de varianza y así sucesivamente. Esto es conocido con el nombre de ejes principales y los factores resultantes no están correlacionados.

Como parte del procedimiento del SPSS, se chequea la adecuación muestral usando la medida de Kaiser-Meyer-Olkin (KMO) que predice si los datos van a factorizar bien. KMO varía entre 0 y 1 y tiene que ser superior a 0.6 para proceder con el análisis. Otra prueba que se aplica es la de esfericidad de Bartlett. Ya que una de las metas del análisis factorial es descubrir los factores que ayuden a explicar las correlaciones entre las variables, es necesario que las variables estén correlacionadas. Si las correlaciones entre las variables son pequeñas, es poco probable que compartan dimensiones comunes. La prueba de Bartlett, tiene la hipótesis nula de que la matriz de correlaciones es igual a una matriz identidad (unos en la diagonal y ceros fuera de ella).

Siguiendo el criterio de Kaiser, se seleccionan los factores más importantes, que corresponden a aquellos componentes con valores característicos (“eigenvalues”) de 1.0 o superiores. Con base en esas combinaciones lineales seleccionadas, se construyen nuevas variables que forman parte de la base (SPSS, 2000).

Para realizar la estratificación de los hogares se usó el algoritmo Cluster K-Means del SPSS. Desde el punto de vista del cálculo, este método es similar a un análisis de variancia (ANOVA) “en reversa”. El programa comienza con k conglomerados aleatorios y luego va moviendo individuos entre aquellos grupos que (1) minimizan la variabilidad dentro conglomerados y (2) maximizan la variabilidad entre conglomerados. Se dice que es un ANOVA en reversa, pues el programa trata de mover los casos de tal forma que obtenga los resultados más significativos en el ANOVA (StatSoft, 2002).

VARIABLES DEL ESTUDIO

Se definieron indicadores con base en las variables censales, tomando en cuenta la experiencia y tratando de cubrir toda la temática del Censo. Las variables pueden clasificarse inicialmente en tres grandes categorías:

Categoría	Variable
Sobre la vivienda	Zona de residencia (urbana o rural)
	Calidad del techo
	Material del techo
	Calidad del piso
	Material del piso
	Calidad de las paredes exteriores
	Material de las paredes exteriores
	Pertenenencias del hogar (TV Color, Teléfono, Refrigeradora, Horno de Microondas, Ducha para agua caliente, Tanque para agua caliente, Lavadora, Microcomputadora, Carro – no de trabajo -)
	Acceso a servicios (abastecimiento de agua, servicio sanitario, abastecimiento de electricidad)

Categoría	Variable
Sobre el hogar	Tipo de hogar (Propuesta de Tipología de Hogar, INEC Censo 2000)
	Número de residentes habituales
	Número de residentes ocupados
	Número de dormitorios
	Número de residentes menores de 5 años
	Número de residentes de 5 a 12 años
	Número de residentes de 13 a 17 años
	Número de residentes mayores de 18 años
	Número de residentes que estudian actualmente
	Número de residentes con educación primaria
	Número de residentes con educación secundaria
	Número de residentes con educación universitaria o superior
	Sobre el jefe del hogar
Edad	
Estado conyugal	
Lugar de nacimiento	
Tipo de seguro social	
Condición de discapacidad	
Asistencia a entidad educativa	
Nivel de educación	
Ocupación principal	
Categoría ocupacional	
Sector institucional	
Hijos nacidos vivos	
Hijos vivos actualmente	

La categoría ocupacional fue reconstruida con base en el sector institucional para diferenciar los empleados del Estado y de las Instituciones Autónomas, siguiendo el procedimiento de Madrigal, 2002.

CONSTRUCCIÓN DE INDICES

Tal y como se indicó arriba, se construyeron índices tratando de que el procedimiento fuera muy simple. En algunos casos, se siguió el procedimiento definido por Madrigal, para efectos de comparabilidad.

1. Calidad de la vivienda. Se resumió en un índice a partir de una suma simple de puntajes de las tres áreas consideradas: piso, techo y paredes exteriores. Luego se recodificó en 5 categorías como sigue:

Categoría	Puntajes incluidos	Código asignado
Muy buena	9	5
Buena	7,8	4
Regular	6	3
Mala	4,5	2
Muy mala	3	1

2. Pertenencias del hogar. En un primer paso, se efectuó un análisis de conglomerados para analizar categorías de hogares según tenencia de artículos, lo cual resultó en una distribución interesante. El análisis discriminante posterior, mostró una proporción de clasificación correcta del 95%. Se definieron tres grupos, con los siguientes porcentajes de tenencia:

Artículo	Grupo 1	Grupo 2	Grupo 3
Porcentaje de hogares	<u>30,3%</u>	<u>20,7%</u>	<u>49,1%</u>
TV Color	99,2%	95,7%	70,9%
Teléfono	91,9%	73,3%	24,0%
Refrigeradora	99,6%	97,0%	69,1%
Horno microondas	100,0%	0,0%	2,9%
Ducha para agua caliente	76,5%	85,5%	0,8%
Tanque para agua caliente	13,3%	3,1%	0,3%
Lavadora	97,9%	94,2%	64,7%
Microcomputadora	36,3%	14,1%	0,7%
Carro (no de trabajo)	55,5%	36,7%	4,9%

En comparación con otros artículos, el tanque de agua caliente y la microcomputadora muestran valores de tenencia más bajos y, por lo tanto, las probabilidades de que un hogar seleccionado al azar los posea son inferiores a 0,5. El primer segmento identifica un grupo en el que la probabilidad de que un hogar tenga todos los artículos, se incrementa significativamente. El segundo grupo, muestra probabilidades menores de tener microondas o vehículo que no sea de trabajo. Y, finalmente, el tercer grupo, muestra hogares con alta probabilidad de tener TV a color, Refrigerador y Lavadora, únicamente. Si este ejercicio se repite para construir seis segmentos, el comportamiento de las variables tanque para agua caliente y microcomputadora es muy similar. El grupo que prácticamente tiene de todos los artículos, representa un 23.3%. En ese ejercicio, un 14% de los hogares no tiene alguno de los artículos incluidos en la boleta. Los grupos restantes, muestran combinaciones diferentes de posesiones en esa escala. Entonces, al abrir más los grupos, los artículos disponibles sirven para discriminar hacia abajo, es decir entre hogares en diferentes grados de posesión hasta el que no tiene alguno, pero no logra romper la estructura del grupo con un nivel de posesión más alto. Para ello, sería necesario afinar el tipo de vehículo (modelo-año) que poseen e incluir artículos que discriminen más entre hogares de un nivel alto o muy alto.

Para efectos de ser simples, se siguió la opción de Madrigal como una sumatoria de ceros y unos para todos los artículos. En este caso el valor mínimo del índice es cero y el máximo 9.

Artículo	Códigos en boleta censal		Puntaje si posee el artículo
	Sí	No	
TV Color	1	2	1
Teléfono	1	2	1
Refrigeradora	1	2	1
Ducha para agua caliente	1	2	1
Horno Microondas	1	2	1
Microcomputadora	1	2	1
Tanque para agua caliente	1	2	1
Carro (no de trabajo)	1	2	1

3. Acceso a servicios. De igual forma, mediante una sumatoria simple de los códigos en preguntas sobre abastecimiento de agua, servicio sanitario y acceso a electricidad y su división por el número de variables incluidas, se construyó un índice de acceso a servicios directamente en el hogar. El valor mínimo del índice es 1 y el máximo 3. Un hogar con valor 1 significa que tiene los servicios a mano directamente en el hogar.
4. Tipo de hogar. Según la propuesta de tipología de hogar del INEC, se definen cuatro grandes categorías, las cuales se subdividen en un total de 13 grupos. Debido a las bajas frecuencias de algunos grupos, que pueden considerarse extraños, se tomó la decisión de abrir la primera gran categoría de hogar nuclear en sus tres grupos y para el resto se mantuvo la definición de la categoría: hogar extenso, hogar compuesto y otros tipos.

Categoría de hogar	Subcategoría de hogar	Código del Tipo
Hogar nuclear	Conyugal sin hijos	1
	Conyugal con hijos	2
	Monoparental con hijos	3
Hogar extenso	Conyugal y otros familiares	4
	Monoparental y otros familiares	4
	Jefe (a) y otros familiares	4
Hogar compuesto	Nuclear y no familiares	5
	Extenso y no familiares	5
	Jefe (a), familiares y no familiares	5
Otros hogares	Unipersonal	6
	No familiares	6

5. Hacinamiento. Se calculó la razón de personas por dormitorio de la vivienda. Los dormitorios hacen referencia a los cuartos que se usan exclusivamente para dormir. El resultado fue codificado en 4 niveles, donde el 1 corresponde a un hacinamiento bajo, lo que indica mayor comodidad y el 4 corresponde a un hacinamiento alto, lo que se espera que represente ingresos insuficientes.
6. Dentro de los aspectos demográficos del hogar, se calcularon las proporciones de personas que:
- Trabajan
 - Estudian
 - Menores de 5 años
 - De 5 a 12 años
 - De 13 a 17 años
 - Mayores de 18 años
 - Con educación primaria
 - Con educación secundaria
 - Con educación universitaria y más

Donde el denominador de las proporciones es el número total de residentes habituales de la vivienda. Estas proporciones fueron recodificadas de la siguiente forma:

Indicador	Rango	Código
Proporción que trabaja	No hay personas que trabajan	0
	Menor o igual a 0,20	1
	Más de 0,20 a 0,30	2
	Más de 0,30 a 0,40	3
	Más de 0,40 a 0,50	4
	Más de 0,50 a 0,60	5
	Más de 0,60 a 0,70	6
	Más de 0,70	7
Proporción de niños menores de 5 años	No tiene niños menores de 5 años	0
	Menor o igual a 0,10	1
	Más de 0,10 a 0,20	2
	Más de 0,20 a 0,30	3
	Más de 0,30 a 0,40	4
	Más de 0,40 a 0,50	5
	Más de 0,50 a 0,60	6
	Más de 0,60	7
Proporción de niños de 5 a 12 años	No tiene niños de 5 a 12 años	0
	Menor o igual a 0,20	1
	Más de 0,20 a 0,30	2
	Más de 0,30 a 0,40	3
	Más de 0,40 a 0,50	4
	Más de 0,50	5

Indicador	Rango	Código
Proporción de jóvenes de 13 a 17 años	No tiene jóvenes de 13 a 17 años	0
	Menor o igual a 0,20	1
	Más de 0,20 a 0,30	2
	Más de 0,30 a 0,40	3
	Más de 0,40 a 0,50	4
	Más de 0,50	5
Proporción de personas de 18 años o más	No tiene personas de 18 años o más	0
	Menor o igual a 0,20	1
	Más de 0,20 a 0,30	2
	Más de 0,30 a 0,40	3
	Más de 0,40 a 0,50	4
	Más de 0,50 a 0,60	5
	Más de 0,60 a 0,70	6
	Más de 0,70 a 0,80	7
	Más de 0,80	8
Proporción de personas que estudian	No hay personas que estudian	0
	Menor o igual a 0,20	1
	Más de 0,20 a 0,30	2
	Más de 0,30 a 0,40	3
	Más de 0,40 a 0,50	4
	Más de 0,50 a 0,60	5
	Más de 0,60 a 0,70	6
	Más de 0,70	7

Siguiendo el procedimiento definido por Madrigal, se usó el nivel educativo del jefe del hogar, sin considerar el detalle de la variable de años de educación. En este caso se tienen las siguientes categorías:

Categoría	Códigos en la boleta	Puntaje
Sin estudios	00	0
Primaria incompleta	21, 22, 23, 24, 25	1
Primaria completa	26	2
Secundaria incompleta	31, 32, 33, 34,41,42, 43, 44, 45	3
Secundaria completa	35, 46	4
Universitaria incompleta	51, 52, 53, 61, 62, 63	5
Universitaria completa	64, 65, 66, 67,68	6

Se siguió también el procedimiento de Madrigal para la construcción del índice de ocupación. Únicamente se varió el punto de que si el jefe del hogar no está ocupado se asignó el código 0. En esta categoría se clasificó el 28.3% de los casos. Para los jefes ocupados, se definió el mismo sistema de puntajes que Madrigal.

RELACIONES ENTRE LOS INDICADORES

Mediante el procedimiento FACTOR del SPSS se evaluaron las relaciones entre los indicadores.

En el primer ejercicio realizado con todos los indicadores, como parte del procedimiento del SPSS, se chequeó la adecuación muestral usando la medida de Kaiser-Meyer-Olkin (KMO) que predice si los datos van a factorizar bien. El valor fue inferior a 0,6, lo que limitó proseguir con el análisis. Otra prueba que se aplicó es la de esfericidad de Bartlett. La prueba de Bartlett, tiene la hipótesis nula de que la matriz de correlaciones es igual a una matriz identidad (unos en la diagonal y ceros fuera de ella). En este caso la hipótesis fue rechazada.

Aunque los factores obtenidos estaban limitados por el resultado de la prueba de adecuación muestral, era interesante continuar revisando los resultados de las cargas

factoriales y las correlaciones entre las variables, con el propósito de discriminar aquellas variables o indicadores que no estaban aportando significativamente al modelo.

Con base en esto, se eliminaron los siguientes indicadores:

- Asistencia a educación del jefe del hogar
- Proporción de niños menores de 5 años
- Proporción de jóvenes de 13 a 17 años
- Proporción de personas con educación secundaria

Una vez realizado este proceso, el siguiente ejercicio de análisis factorial, nos ofreció mejores resultados en las pruebas de adecuación:

KMO = 0,75 (superando el límite de 0,60 requerido)

Bartlett = 105447.657 y Sig. = .000

Con lo que se puede proseguir con el análisis de agrupación. Usando el método de extracción de Análisis de Componentes Principales, se seleccionaron 6 factores básicos que explican el 61.9% de la variabilidad total.

Se aplicó el método de rotación Varimax con Kaiser, que convergió en 9 iteraciones. Las dimensiones identificadas son:

1. Miembros con posibilidad de aportar al ingreso familiar:
 - a. Proporción de personas de 18 años o más
 - b. Proporción de personas que trabajanE inversamente con
 - c. Proporción de personas de 5 a 12 años
 - d. Proporción de personas que estudian
2. Desarrollo social:

- a. Servicios públicos (agua, electricidad y cloacas) – la carga en este caso es negativa por la dirección del indicador.
 - b. Zona de residencia (urbana-rural)
 - c. Pertenencias del hogar
 - d. Estado de la vivienda
3. Orientación a educación
 - a. Proporción de miembros con educación universitaria
 - b. Nivel de educación del jefe del hogarE inversamente con
 - c. Proporción de miembros con educación primaria
4. Fortaleza del jefe del hogar
 - a. Índice de ocupación del jefeE inversamente con
 - b. Edad del jefe
 - c. Discapacidad del jefe
5. Tipología del hogar
 - a. Sexo del jefe del hogar
 - b. Tipo de hogar
6. Movilidad del jefe
 - a. Lugar de nacimiento – la carga factorial es negativa, por cuanto los jefes de hogares extranjeros tendrán menos posibilidad de acceder al seguro social.
 - b. Tenencia de seguro social

ESTRATIFICACIÓN DE LOS HOGARES

Para clasificar los hogares se usó el algoritmo K-medias (SPSS versión 9) y se le solicitó que construyera diversos números de agrupaciones. Con un número mayor a diez, se tiene que hay funciones discriminantes que no se asocian fuertemente con alguno de los indicadores. La solución de diez agrupaciones produjo un resultado

interesante. Se procedió a clasificar los hogares asumiendo una estructura única en todo el país, sin diferenciar por zona urbana o rural. Esto con el propósito de hacer comparable la clasificación, a nivel nacional. Una vez efectuado este análisis se procedió a evaluar su bondad mediante un análisis discriminante.

Si se corre el análisis discriminante para obtener ciertos estadísticos de clasificación, se obtiene que el 90% de los casos son clasificados correctamente. En esta solución, la probabilidad de clasificar un caso correctamente, por azar, es del 10% y por tanto el resultado global del modelo es muy bueno.

La matriz de estructura del análisis discriminante nos indica el aporte a la explicación de la varianza y los indicadores más altamente correlacionados con los puntajes discriminantes. A partir de los resultados de este ejercicio, se eliminaron del análisis de conglomerados, las variables: lugar de nacimiento del jefe, sexo del jefe, discapacidad del jefe, tenencia de seguro social y zona de residencia, que por su nivel nominal y, en algunos casos, bajo nivel de incidencia de alguna categoría, no varían significativamente de un segmento a otro. Esas variables serán usadas posteriormente para definir el perfil del segmento.

Con el nuevo conjunto de 14 indicadores, se corrió un análisis de conglomerados. En este nuevo escenario, la probabilidad de clasificación correcta es de 91.6%. La matriz de estructura obtenida en el análisis discriminante señala las correlaciones más altas con indicadores, que se muestran en la tabla siguiente.

Función discriminante	% de varianza	Indicadores significativamente correlacionados
1	43,1%	Proporción de miembros con universidad Nivel de escolaridad del jefe Pertenenencia de artículos
2	24,3%	Proporción de miembros de 18 años o más (inversa) Proporción de miembros de 5 a 12 años Edad del jefe (inversa) Proporción de miembros que estudian
3	13,7%	Estado de la vivienda
4	8,5%	Proporción de miembros que trabajan
5	5,4%	Proporción de miembros con educación primaria (inversa)
6	2,5%	Acceso a servicios públicos
7	1,2%	Tipo de hogar (inversa)
8	0,7%	Índice de ocupación del jefe (inversa)
9	0,6%	Hacinamiento

Las primeras dos funciones canónicas discriminantes explican el 67,4% de la variabilidad. Correlaciones negativas o coeficientes estandarizados negativos implican que altos valores del indicador, producen un valor menor en la función discriminante. Si se evalúan las ecuaciones discriminantes canónicas en las medias de los grupos, se encuentran los siguientes perfiles, basados en las variables o indicadores que tienen mayor nivel de asociación con las funciones discriminantes.

Conglomerado	Principales características (que discriminan más)	Proporción de la muestra	Tamaño promedio del hogar	Tipo de hogar más frecuente
1	Mayor proporción de miembros con educación universitaria Mayor nivel educativo del jefe Mayor nivel de tenencia de artículos Mayor proporción de mayores de 18 años Menor proporción de niños de 5 a 12 años Mayor edad del jefe Menor proporción que estudian Menor índice de ocupación del jefe	3,7%	3,3	Nuclear con hijos
2	Mayor nivel educativo del jefe Mayor nivel de tenencia de artículos Mayor índice de ocupación del jefe Menor proporción de mayores de 18 años Mayor proporción de niños de 5 a 12 años Menor edad del jefe Mayor proporción que estudian Mejor estado de la vivienda	11,1%	4,4	Nuclear conyugal con hijos
3	Menor nivel de tenencia de artículos Menor nivel de escolaridad del jefe Mayor proporción de mayores de 18 años Menor proporción de niños de 5 a 12 años Menor proporción que estudian Mal estado de la vivienda Menor nivel de acceso a servicios Menor índice de ocupación del jefe	5,1%	3,2	Hogar extenso o de otros tipos (unipersonal – no familiares)
4	Nivel medio de escolaridad del jefe Mayor proporción de mayores de 18 años Menor proporción de 5 a 12 años Mayor proporción que trabaja Mayor proporción con educación primaria	16,7%	3,4	Nuclear conyugal
5	Nivel medio de escolaridad del jefe Nivel medio de tenencia de artículos Menor proporción de mayores de 18 años Mayor proporción de 5 a 12 años Mayor proporción que estudia Menor proporción que trabaja	19,1%	4,7	Nuclear conyugal con hijos
6	Nivel bajo de escolaridad del jefe Menor índice de ocupación del jefe Mayor edad del jefe Mayor proporción de 18 años o más Menor proporción de 5 a 12 años Menor proporción que estudian Menor proporción que trabaja	15,6%	3,0	Hogar extenso y de otros tipos Nuclear conyugal sin hijos
7	Mayor proporción con educación universitaria Mayor nivel de escolaridad del jefe Mayor nivel de tenencia de artículos Alto índice de ocupación del jefe Buen estado de la vivienda Buen acceso a servicios Mayor proporción que trabaja Menor proporción con educación primaria	5,7%	3,6	Nuclear conyugal con hijos

Conglomerado	Principales características (que discriminan más)	Proporción de la muestra	Tamaño promedio del hogar	Tipo de hogar más frecuente
8	Bajo nivel de escolaridad del jefe Bajo nivel de tenencia de artículos Alto nivel de hacinamiento Mayor proporción de niños de 5 a 12 años Menor proporción de mayores de 18 años Baja proporción que trabaja Mal estado de la vivienda Bajo acceso a servicios	8,5%	5,8	Nuclear conyugal con hijos
9	Bajo nivel de edad del jefe Bajo nivel de tenencia de artículos Menor proporción que estudia Regular estado de la vivienda Baja proporción de niños de 5 a 12 años	12,0%	4,2	Nuclear conyugal con hijos
10	Alto nivel de escolaridad del jefe Alto índice de ocupación del jefe Bajo nivel de hacinamiento Mayor proporción de 18 años o más Mayor proporción que trabaja Menor proporción con educación primaria Mayor proporción con educación universitaria	2,6%	1,8	Otros tipos (unipersonal – no familiares)

DESCRIPCIÓN DE LOS CONGLOMERADOS

En la siguiente tabla se describen los segmentos ordenados según nivel de escolaridad del jefe del hogar.

Característica	Total	Seg. 3	Seg. 6	Seg. 9	Seg. 8	Seg. 4	Seg. 5	Seg. 2	Seg. 1	Seg. 10	Seg. 7
Orientación a educación											
Nivel educativo del jefe del hogar	59.9% con educación primaria o menos	76.6% con primaria incompleta o menos	66.7% con primaria incompleta o menos	63.9% con primaria incompleta o completa	74.2% con primaria completa o incompleta	72.8% con secundaria completa o primaria completa	75.4% con secundaria completa o primaria completa	70.4% con secundaria completa o universidad incompleta	44.4% con universidad completa o incompleta	70.4% con universidad completa o incompleta	92.5% con universidad completa o incompleta
Proporción con primaria	38.3%	24.2%	31.8%	45.3%	24.4%	67.9%	46.8%	23.2%	22.9%	4.8%	13.6%
Proporción con secundaria	13.2%	1.9%	12.0%	6.3%	1.6%	13.3%	6.1%	37.1%	18.8%	45.9%	17.6%
Proporción con universidad	6.6%	0.1%	0.5%	0.3%	0.1%	0.8%	0.5%	5.5%	47.7%	41.4%	49.5%
Capacidad de aporte al ingreso familiar											
Proporción de miembros de 18 años o más	68.9%	81.1%	92.0%	60.8%	39.3%	86.4%	47.2%	55.4%	92.2%	97.8%	74.3%
Proporción de personas que trabajan	37.1%	42.6%	19.7%	35.7%	18.6%	61.2%	26.0%	34.9%	36.5%	92.1%	55.9%
Proporción de personas que estudian	26.2%	6.1%	8.0%	11.1%	37.9%	15.9%	44.9%	46.6%	26.4%	24.1%	37.9%
Proporción de menores de 5 años	8.3%	7.5%	2.2%	22.4%	13.6%	3.8%	7.3%	10.2%	1.8%	1.3%	7.9%
Proporción de miembros de 5 a 12 años	14.0%	4.5%	2.6%	6.4%	36.2%	2.1%	31.8%	22.3%	2.0%	0.5%	10.6%
Proporción de miembros de 13 a 17 años	8.9%	6.9%	3.3%	10.4%	10.9%	7.7%	13.8%	12.2%	4.0%	0.5%	7.2%

<i>Característica</i>	<i>Total</i>	<i>Seg. 3</i>	<i>Seg. 6</i>	<i>Seg. 9</i>	<i>Seg. 8</i>	<i>Seg. 4</i>	<i>Seg. 5</i>	<i>Seg. 2</i>	<i>Seg. 1</i>	<i>Seg. 10</i>	<i>Seg. 7</i>
Tipología del hogar											
Tipo de hogar	69.2% Nuclear 18.9% Extenso	21.7% Extenso 29.7% Otros tipos	31.8% Extenso 22.3% Otros tipos 16.9% Nuclear conyugal sin hijos	58.6% Nuclear conyugal con hijos	62.6% Nuclear conyugal con hijos	59% Nuclear conyugal	71% Nuclear conyugal con hijos	71.4% Nuclear conyugal con hijos	38.9% Nuclear conyugal con hijos 21.6% Monoparental	71.2% Otros tipos (Unipersonal, no familiares)	62.1% Nuclear conyugal con hijos
Sexo del jefe del hogar	77.5% Varones	86.2% Varones	61.9% Varones	85.0% Varones	76.1% Varones	79.8% Varones	79.5% Varones	84.5% Varones	62.8% Varones	64.9% Varones	85.9% Varones
Tamaño promedio del hogar	4.0	3.2	3.0	4.2	5.8	3.4	4.7	4.4	3.3	1.8	3.6

Fortaleza del jefe del hogar

Índice de ocupación del jefe	28.4% sin ocupación Índice promedio 3.99	37.0% sin ocupación Índice promedio 2.28	84.3% sin ocupación Índice promedio 2.53	15.4% sin ocupación Índice promedio 3.03	32.7% sin ocupación Índice promedio 2.65	11.0% sin ocupación Índice promedio 3.57	19.0% sin ocupación Índice promedio 3.23	4.1% sin ocupación Índice promedio 5.69	75.5% sin ocupación Índice promedio 3.49	1.2% sin ocupación Índice promedio 6.30	0.0% sin ocupación Índice promedio 7.71
Edad del jefe	45.4 años	50.3 años	66.5 años	33.7 años	38.7 años	46.5 años	39.1 años	39.2 años	59.3 años	38.3 años	41.8 años
Discapacidad del jefe	9.2% tiene	16.9% tiene	22.6% tiene	5.9% tiene	9.5% tiene	7.7% tiene	4.7% tiene	3.6% tiene	12.2% tiene	4.1% tiene	2.7% tiene
Seguro social	82.0% tiene	63.3% tiene	89.8% tiene	76.5% tiene	66.2% tiene	84.4% tiene	80.4% tiene	89.6% tiene	88.8%	85.1%	91.2%
Lugar de nacimiento jefe	44.5% mismo cantón 45.6% otro cantón 9.9% otro país	45.9% mismo cantón 35.0% otro cantón 19.1% otro país	45.0% mismo cantón 50.0% otro cantón 5.1% otro país	45.9% mismo cantón 38.0% otro cantón 16.1% otro país	41.0% mismo cantón 39.8% otro cantón 19.3% otro país	45.5% mismo cantón 47.5% otro cantón 6.9% otro país	50.4% mismo cantón 44.2% otro cantón 5.5% otro país	42.7% mismo cantón 50.4% otro cantón 6.9% otro país	40.8% mismo cantón 47.1% otro cantón 12.2% otro país	32.0% mismo cantón 52.0% otro cantón 16.1% otro país	33.3% mismo cantón 54.2% otro cantón 12.6% otro país

Característica	Total	Seg. 3	Seg. 6	Seg. 9	Seg. 8	Seg. 4	Seg. 5	Seg. 2	Seg. 1	Seg. 10	Seg. 7
Desarrollo social y vivienda											
Estado vivienda	56.8% Muy buena	9.9% Muy buena	52.5% Muy buena	28.9% Muy buena	7.6% Muy buena	69.6% Muy buena	68.7% Muy buena	83.5% Muy buena	81.0% Muy buena	79.0% Muy buena	89.5% Muy buena
Pertenencias	4.23	0.62	3.88	2.90	1.89	4.70	4.26	6.29	6.58	5.57	7.16
Zona	60.5% Urbano	13.0%	61.1%	50.3%	36.8%	67.6%	55.4%	81.3%	86.0%	85.5%	87.8%
Acceso a agua	92.4% tubería dentro	49.2%	95.5%	92.4%	78.5%	97.0%	96.5%	98.4%	97.7%	98.2%	98.5%
Acceso a electricidad	77.2% CNFL o ICE	37.7%	79.6%	81.3%	71.6%	79.1%	78.6%	81.3%	81.0%	84.1%	82.6%
Acceso a servicio sanitario	67.1% a Tanque séptico	23.5%	71.9%	74.7%	50.6%	72.8%	80.9%	66.2%	53.6%	53.1%	56.2%
Provincia ³		Más frecuente en Alajuela, Puntarenas, Guanacaste y Limón	Se encuentra en todas las provincias	Más frecuente en Puntarenas y Limón	Más frecuente en Limón, Puntarenas y Guanacaste	Más frecuente en las provincias centrales	Más frecuente Alajuela, Cartago	Más frecuente en San José y Heredia			

³ Se destacan diferencias respecto al total de la muestra.

CONCLUSIONES Y RECOMENDACIONES

El ejercicio de segmentación que se presenta en este artículo es un primer paso para estratificación de hogares en Costa Rica y la creación de un indicador a nivel del marco muestral, que permita orientar el diseño de muestras en investigación de mercados.

Cada nuevo indicador provee posibilidades más amplias de adaptarse a las necesidades del usuario, incluir la segmentación en sus estudios ad-hoc y sobre todo tener una referencia geográfica.

Este es el siguiente paso en este proceso. Proceder a una relación geográfica que permita el diseño de las muestras.

El análisis de las variables e indicadores creados, también orienta el diseño de las preguntas futuras. Algunos especialistas quisieran disponer de una forma de afinar la identificación de los segmentos de más alto nivel económico y social. Eso requiere que las variables que ayudan en la estimación del ingreso tengan esa oportunidad. El acceso a la tecnología es cada vez mayor y el formulario debería tener la capacidad de detectar los nuevos elementos de comunicación, que discriminen mejor en ese segmento socioeconómico.

AGRADECIMIENTOS

Quiero agradecer el apoyo de todo el personal de Unimer Research International y de la Unidad de Programación de Operaciones Censales y de Encuestas del INEC. En particular a la Directora de los Censos 2000, Lic. María Elena González y al personal a su cargo; al Bach. Gerardo Fonseca, jefe de procesamiento de Unimer RI, a M.Sc. Jimena Vargas y a la Lic. Graciela Navarro, directora y exdirectora de proyectos de Unimer RI, por sus valiosos aportes a este trabajo.

BIBLIOGRAFÍA

1. Instituto Nacional de Estadística y Censos. **Encuesta de Hogares de Propósitos Múltiples, Julio del 2001**. San José, Costa Rica. Junio del 2002.
2. Instituto Nacional de Estadística y Censos. **Clasificación de Ocupaciones de Costa Rica-2000**. San José, Costa Rica. 2000. Documento basado en la Clasificación Internacional Uniforme de Ocupaciones CIIU-88 de la Oficina Internacional del Trabajo.
3. Madrigal, Johnny. **Metodología y Construcción de un Indicador del Ingreso Familiar: Aplicación a los Censos Nacionales de Vivienda y Población 1984**. Asociación Demográfica Costarricense. San José, Costa Rica, 1986.
4. Madrigal, Johnny. **Estratificación de hogares y segmentos por niveles de ingreso en los Censos 2000**. Julio 2002
5. Maxwell, Nick. **Sample Market Segmentation Report**. Maxwell Statistics Corporation. Enero 2002.
6. Maxwell, Nick. **Working with market research clients on market segmentation**. Maxwell Statistics Corporation. Febrero 2002.
7. Peter, J.Paul and Olson, Jerry C. **Consumer Behavior and Marketing Strategy**. Richard Irwin Inc. 1993
8. SPSS/ PC + ADVANCED STATISTICS 4.0 Copyright 1990 by SPSS Inc. United States of America. Sección B1.
9. SPSS/PC + STATISTICS 4.0 Copyright 1990 by SPSS Inc. United States of America. Secciones B125, B155
10. www.claritas.com Página WEB de Claritas sobre servicios de estratificación demográfica basada en el Censo 2000 de Estados Unidos.
11. www.dssresearch.com **Understanding Market Segmentation**. DSS Research 2001.
12. www.statsoftinc.com **Cluster Analysis**. StatSoft, Inc. 1984 – 2002.

