

## **Modelo de Clasificación de la Población de Argentina**

*".....Esas ambigüedades, redundancias y deficiencias recuerdan las que el doctor Franz Kuhn atribuye a cierta enciclopedia china que se titula Emporio celestial de conocimientos benévolos. En sus remotas páginas está escrito que los animales se dividen en: a) pertenecientes al Emperador b) embalsamados c) amaestrados, d) lechones, e) sirenas, f) fabulosos g) perros sueltos h) incluidos en esta clasificación, i) que se agitan como locos, j) innumerables, k) dibujados con un pincel finísimo de pelo de camello, l) etcétera, m) que acaban de romper el jarrón, m) que de lejos parecen moscas....."*

El idioma analítico de John Wilkins; Ensayo de Jorge Luis Borges, publicado en la colección Otras Inquisiciones.

**Por Juan Jose Lloret**

ALEPH ZERO marketing & social research

Se describen distintos caminos para clasificar la población de Argentina para uso comercial, utilizando bases de datos públicas. Se muestra como es posible clasificar a los hogares y a los individuos con un criterio de nivel social. Se analizan diversas características del modelo y se plantean algunas miradas cuantitativas del proceso clasificatorio, a desarrollar en el futuro.

**Palabras clave:** Clasificación, ENCUESTA PERMANENTE DE HOGARES (EPH), Decil, Perfil, Incidencia, Ingresos, ESTADO DE ACTIVIDAD DEL HOGAR (EAH), MODELO DE NIVEL SOCIAL (MNS), MAPA SOCIAL DE INGRESOS (MSI)

Different methods that aim at classifying the Argentine population for commercial purposes using public databases are described. The possibility of classifying households and individuals with a social level criterion is shown. The diverse features of the model are analyzed and some quantitative-qualitative perspectives of the classifying process are introduced for future development.

### **I - Antecedentes**

Las primeras clasificaciones socio demográficas en Argentina, fueron realizadas cerca del año 1946, en donde aparecen agrupamientos por Edad, Area, Sexo y "Social Class", según figura en algunos informes de la época.

Ha pasado mucho tiempo y hoy en el 2003, en el escenario del Marketing actúan distintas fuerzas que van modificando progresivamente los criterios de agrupamiento tradicionales y nos enfrenta con nuevos desafíos si deseamos tener una clasificación útil y eficiente de la población. Algunas de las fuerzas actuantes son:

**Fenómeno "One to One"**: La necesidad de llegar a los grupos objetivos de forma mas selectiva, a través de un mensaje específico y personalizado. La búsqueda de una adecuada identificación de los hogares y los individuos aparece entonces en todos los procesos de segmentación.

**El Fenómeno de Múltiples Usuarios**: el uso de la clasificación para distintas aplicaciones (MR, MD, SR entre otras) plantea la necesidad de adecuar la clasificación a las distintas necesidades y posibilidades de los usuarios, pero basándose en criterios similares y lo mas universales posibles.

**El Fenómeno Ecológico**: La mayor dificultad para obtener los datos de los individuos, originado por cansancio, privacidad, situaciones sociales, regulaciones gubernamentales, etc., afectará las formas y metodología de clasificación que se utilizaran en el futuro próximo.

**El Impacto Tecnológico**: la disponibilidad de recursos tecnológicos para el tratamiento de los datos y construcción de la información, permite desarrollar modelos de alta complejidad con austeros recursos.

**Los Yacimientos Demográficos**: el progresivo acceso a las bases públicas de datos, con las cuales es posible construir patrones de la población, a nivel de micro modelos. En nuestro caso hemos utilizado la información provista por el INDEC (Instituto Nacional de Estadísticas y

Censos) de Argentina, quien provee los datos del Censo Nacional de Población y Viviendas del 2001, así como la Encuesta Permanente de Hogares (EPH)<sup>(1)</sup> en forma semestral.

La presencia de estas fuerzas y su interpretación nos permitió elaborar un MODELO DE CLASIFICACIÓN Y ESTIMACIÓN para la población de la Argentina. El modelo fue desarrollado para describir distintas estructuras de agrupamiento para los HOGARES y los INDIVIDUOS. Se muestra un ensayo del SOCIAL GRADE en dichos niveles. Complementariamente, nos permitió disponer de UN ESTIMADOR DE INGRESOS para las distintas categorías. El modelo entendemos, es aplicable en otros Países si se cuenta con una apropiada base de datos tal como la descripta en el punto 5.

## II - Desarrollo del modelo

Nos hemos propuesto desarrollar un MODELO DESCRIPTIVO DE LOS HOGARES (INDIVIDUOS), dirigido a: 1) **CLASIFICAR** y 2) **DIMENSIONAR** distintas características de la población (principalmente cantidad e ingresos).

Asumimos una serie de supuestos que detallaremos brevemente:

a) Las poblaciones son clasificables según diferentes características básicas: la edad del jefe de familia, el tamaño del grupo familiar, el nivel de educación, el estado de actividad, la zona de residencia o el sexo, o bien a través de combinaciones de distintas variables, tal es el caso del modelo de Nivel Social o Ciclo de Vida. Dichas clasificaciones además de cumplir determinadas condiciones formales, deben reflejar la naturaleza de la clasificación, es decir deben tener sentido, en cuanto interpretación, aplicación y utilidad.<sup>2</sup>

b) Con el fin de poder clasificar o proyectar características de las muestras a la población, es necesario contar con Reglas Clasificadoras que permitan identificar a cada hogar u individuo perteneciente a alguna categoría.

c) Paralelamente debemos desarrollar Perfiles de la población que permitan representar al universo. Es en este aspecto que según el nivel de información que se tenga cada registro (hogar, individuo, etc.), se puede asignar una determinada distribución de probabilidades.

d) Para la construcción de los perfiles recurrimos a la EPH. En ella, para cada hogar se relevan múltiples variables, referidas al hogar, al jefe del hogar y a cada uno de los miembros del mismo. Varias de las variables que nos interesan son registradas con un nivel de detalle muy minucioso, tal es el caso de la renta y la ocupación. Con la información de ingresos, es posible construir la curva de ingresos totales de los hogares CITH, la curva de ingreso de los individuos CITI y en consecuencia la correspondiente clasificación de los hogares y de los individuos en deciles. Cada decil, puede caracterizarse entre otras por las siguientes medidas: promedio de ingresos en el decil, valor máximo y valor mínimo.

e) Restringiéndonos a los hogares, cada hogar de la EPH tiene la siguiente característica: pertenecer a un determinado decil de ingresos. Y si aplicamos un criterio clasificador a la población, por ejemplo  $S_x$ , donde los jefes de familia son clasificados por sexo, podemos construir una distribución por deciles de la categoría Mujer o la categoría Varón<sup>(3)</sup>. Esto nos permite calcular varios descriptores del hogar. Como ejemplo, para la Ciudad de Córdoba, tenemos los siguientes resultados:

---

<sup>1</sup> Encuesta Permanente de Hogares (EPH) es un estudio en hogares realizado por el INDEC sobre ocupación, ingresos, y perfil de los hogares, en los 32 principales centros urbanos del país, representativa del 61% de los hogares del País. Tiene periodicidad de 6 meses.

<sup>2</sup> Ver: La estructura de los conceptos científicos: Jesus Mosterin – Investigación y Ciencia - Enero 1978

<sup>3</sup> Si  $H$  es el conjunto de hogares,  $\forall i$ , el hogar  $\{h_i\} \in$  al decil  $D_k$ , para algún  $k$ . Si  $\Gamma$  es una clasificación de  $H$ , tal que  $H = \cup C_j / C_j \in \Gamma / C_i \cap C_k = \emptyset$ , es posible definir  $F: \Gamma \rightarrow [0,1]$ <sup>10</sup>; dicha función asocia a cada categoría  $C_j$  con una distribución de probabilidades. Es posible entonces hablar de una ESPERANZA DE INGRESO (**EI**) para el hogar perteneciente a la categoría  $C_j$ . Podemos definir, como **EI** el valor  $\sum f_i \times I_i$ ; donde  $f_i$  es la probabilidad de un hogar de caer en el decil  $I$  con la clasificación  $\Gamma$  y  $I_i$  el ingreso promedio en el decil  $I$ .

<b>S<sub>x</sub> :SEXO DEL JEFE DE FAMILIA</b>	<b>INCIDENCIA <sup>4</sup></b>	<b>PARTICIPACIÓN EN INGRESOS<sup>5</sup></b>	<b>INGRESOS MENSUALES PROMEDIO<sup>6</sup></b>
▪ <b>MUJER</b>	29%	23%	\$ 576
▪ <b>HOMBRE</b>	71%	77%	\$ 792
<b>TOTAL</b>	360.530 <sup>7</sup>	262.956.161 <sup>8</sup>	\$ 729

Fuente: EPH Córdoba - Mayo 2003 – Hogares

Es entonces que con diversos criterios Clasificadores, siguiendo los lineamientos expuestos en los puntos d) y e) y tomando como datos de entrenamiento los disponibles en la Encuesta Permanente en Hogares, que podemos elaborar para cada conglomerado urbano un conjunto de descriptores del conglomerado. A continuación se muestran algunas aplicaciones del Modelo

### **III - Aplicaciones del MODELO**

#### **III-1 Construcción de un modelo de ESTADO DE ACTIVIDAD DEL HOGAR**

Si definimos la clasificación **ESTADO DE ACTIVIDAD DEL HOGAR (EAH)** como la situación en que se halla su jefe de familia. Podemos identificar tres categorías: a) **HOGAR ACTIVO** – donde el jefe de familia tiene actualmente un trabajo, b) **HOGAR DESOCUPADO** – en donde el jefe de familia no tiene actualmente un trabajo y c) **HOGAR No ACTIVO** – con jefe de familia sin ocupación. Cada una de estas categorías muestra características muy diferenciadas, lo cual hace interesante el estudio de cada una de ellas, a los fines de comprender posibles mecanismos de hábitos e ingresos.

El conocimiento de estas estructuras en el caso de los **HOGAR ACTIVOS** permite la elaboración de un **NIVEL SOCIAL**, el cual será desarrollado en detalles en **III-2**. Por otra parte para los **HOGARES No ACTIVOS**, la tipología encontrada señala hogares con jefes jóvenes (estudiantes) o de edad avanzada (jubilados, pensionados, etc.) y a la predominancia de mujeres como jefe de familia. Para el caso de **HOGAR DESOCUPADO**, el análisis a través de las variables, permite inferir cómo se distribuye **LA DESOCUPACIÓN** por las diferentes categorías etarias, sexo, educación, etc.

A continuación se describe cada categoría para el área del Gran Buenos Aires.

<b>EAH: ESTADO DE ACTIVIDAD DEL HOGAR</b>	<b>INCIDENCIA</b>	<b>PARTICIPACIÓN EN INGRESOS</b>	<b>INGRESOS MENSUALES PROMEDIO</b>
▪ <b>HOGAR ACTIVO</b>	63%	77%	\$ 993
▪ <b>HOGAR No ACTIVO</b>	26%	17%	\$ 518
▪ <b>HOGAR DESOCUPADO</b>	11%	6%	\$ 442
<b>TOTAL</b>	3.418.389	\$ 2.768.904.000	\$ 810

Fuente: EPH Buenos Aires - Mayo 2003 – Hogares

<sup>4</sup> INCIDENCIA: la definimos como la proporción de unidades en cada categoría. Muestra la importancia física de cada categoría

<sup>5</sup> PARTICIPACIÓN EN INGRESOS: es el porcentaje ponderado por ingresos, muestra la importancia de la categoría en términos económicos.

<sup>6</sup> INGRESOS MENSUALES PROMEDIO: ingresos totales mensuales del hogar

<sup>7</sup> Cantidad de hogares en Ciudad de Córdoba, según Censo 2001

<sup>8</sup> Estimación de ingresos totales mensuales disponibles en los hogares en la Ciudad de Córdoba

### **III-2 Construcción de un MODELO de NIVEL SOCIAL (MNS)**

Nuestro modelo sigue las ideas planteadas en el documento: Standard Demographic Classification (1997)<sup>9</sup> por ESOMAR, donde se construye para los hogares activos el "Social Grade" a través del cruce de las variables: Ocupación y Educación. Recurriendo a los datos del mes de Mayo del 2003 de la Encuesta Permanente de Hogares, hemos agrupado a los hogares de los 31<sup>10</sup> conglomerados, dando lugar a la clasificación MNS con categorías Alto, Medio y Bajo.

Como variables predictoras, utilizamos: 1) la variable Ocupación con valores: Dirección, Jerarquía Media, Obreros/Empleados Calificados y Obreros/Empleados no calificados. y 2) la variable Educación, con valores: Alta, Media y Baja.

Las reglas de clasificación para la construcción de las categorías fueron:

<b>(Ocupación <math>\wedge</math> Educación) <math>\Rightarrow</math> NIVEL</b>
---

- $\{Dirección \wedge (i) Alta\} \Rightarrow ALTA$
  - $\{Dirección \wedge Media\} \Rightarrow ALTA$
  - $\{Jerarquía Media \wedge Alta\} \Rightarrow ALTA$
  - $\{Dirección \wedge Baja\} \Rightarrow MEDIA$
  - $\{Jerarquía Media \wedge Media\} \Rightarrow MEDIA$
  - $\{Jerarquía Media \wedge Baja\} \Rightarrow MEDIA$
  - $\{Obreros/Empleados Calificados \wedge Alta\} \Rightarrow MEDIA$
  - $\{Obreros/Empleados Calificados \wedge Media\} \Rightarrow MEDIA$
  - $\{Obreros/Empleados Calificados \wedge Baja\} \Rightarrow MEDIA$
  - $\{Obreros/Empleados No Calificados \wedge Alta\} \Rightarrow MEDIA$
  - $\{Obreros/Empleados No Calificados \wedge Media\} \Rightarrow BAJA$
  - $\{Obreros/Empleados No Calificados \wedge Baja\} \Rightarrow BAJA$
- (i)es el operador conjuntivo

Para el caso del conglomerado de Buenos Aires. y sus suburbios obtuvimos la siguiente descripción de la clasificación MNS:

<b>NS : NIVEL SOCIAL</b>	<b>INCIDENCIA</b>	<b>PARTICIPACIÓN EN INGRESOS</b>	<b>INGRESOS MENSUALES PROMEDIO</b>
▪ <b>ALTO</b>	6%	15%	\$ 2.471
▪ <b>MEDIO</b>	41%	54%	\$ 1.310
▪ <b>BAJO</b>	53%	31%	\$ 581
<b>TOTAL</b>	2.153.592	\$ 2.149.392.000	\$ 993

Fuente: EPH Buenos Aires - Mayo 2003 – Muestra 2300 Hogares

Vale la pena mencionar algunas líneas de investigación que estamos realizando con el **Modelo de Nivel Social**

Los antecedentes sobre este tipo de clasificación en Argentina, están focalizados a una clasificación cuantitativa de la población, basada en un índice lineal de variables, con pesos asignados con variados criterios. Nuestra tesis es que el vector [actividad, ocupación, educación] genera variables categóricas y ordinales, cada una de las cuales encierra propiedades básicas bien diferentes y en donde el estudio a través de la lente cuanti-cualitativa

<sup>9</sup> En dicho documento, se plantea la construcción de la matriz MIES, basada en el producto cartesiano de la ocupación y educación, en el caso de hogares con principal sostén del hogar activo. Para clasificar solo, los hogares con principal sostén del hogar no activo, se recomienda utilizar el cruce de educación por una lista de bienes.

<sup>10</sup> La Clasificación fue realizada sobre 31 conglomerados, al no estar disponible la información de la Ciudad de Santa Fe, en esta medición.

















puede ayudar a una mejor comprensión de la población de Argentina en los grandes conglomerados.

Este modelo de Nivel Social, permite empalmar las series de datos clasificados con criterios del NSE 1996<sup>11</sup>, ello posibilita la realización de análisis longitudinales de los datos de mercado. Creo conveniente mencionar sintéticamente cómo se define esta clasificación.

### **III-3 Mapa Social de Ingresos (MSI) para los individuos de 18 a 65 años**

Un tercer desarrollo, se basa en tomar como unidad de análisis a los individuos de 18-65 años, con ocupación declarada y elaborar una clasificación a nivel de personas con el criterio **EDUCACIÓN x OCUPACIÓN**. Creemos que esta forma de identificar a los individuos, permite una mayor precisión en aquellos mercados donde lo individual toma trascendencia como pueden ser los mercados de belleza, telefonía móvil, medicamentos OTC, productos financieros, vestimenta, bebidas, tabaco, etc.

Hemos construido con dicho criterio un mapa descriptivo del País. En el siguiente gráfico de iconos, se muestran dos descriptores para cada región: a) el ingreso mensual per capita de los individuos entre 18 y 65 años y b) el número de individuos de nivel social bajo por cada individuo de nivel social alto. Los resultados hallados nos indican marcados contrastes en la posibilidad de consumo y en la composición social entre las distintas regiones.

<i>REGION</i>	$\omega_i^{12}$	<i>INGRESO MENSUAL PER CAPITA (a)</i>	<i>RELACION ENTRE NIVELES (b)</i>
<i>GBA</i>	55%	 \$ 412	 3
<i>LITORAL</i>	12%	 \$ 243	 13
<i>PAMPA</i>	8%	 \$ 367	 5
<i>CENTRO</i>	8%	 \$ 306	 4
<i>NOA</i>	7%	 \$ 271	 8
<i>CUYO</i>	6%	 \$ 294	 7
<i>SUR</i>	4%	 \$ 486	 4
<i>TOTAL</i>	100%	 \$ 364	 5

*Fuente: EPH Total conglomerados - Mayo 2003 – Hogares*

### **IV - Crítica del MODELO:**

- IV. 1 Alcance del modelo: El modelo construido permite analizar a la población de hogares e individuos de gran parte de Argentina, desde una óptica con mayor panorama que la desarrollada al presente. Es así que nos permite inferir un alcance levemente mayor al 60% de la población.
- IV. 2 Profundidad del modelo: Nuestro desarrollo incorpora una mirada de la población a nivel individual, complementado la visión "hogar-céntrica" tradicional. En algunos mercados, este parece ser un camino a explorar.
- IV. 3 Precisión del modelo: El modelo es desagregado a nivel de conglomerado, evitando extrapolar supuestos de las grandes urbes en el resto de los conglomerados. La heterogeneidad de los resultados por región parecería indicar que la inferencia de los parámetros de las grandes ciudades, en este caso, no es válida.

<sup>11</sup> Índice de Nivel Socio Económico Argentino – Asociación Argentina de Marketing - 1996

<sup>12</sup>  $\omega_i$  es la proporción de población entre 18 y 65 años en cada región

- IV. 4 Realidad del modelo: el modelo sigue una lógica, que entendemos permite por un lado comparar con criterios parecidos, lecturas de la realidad pasada. Al mismo tiempo, la incorporación de la lupa cualitativa en la categorización, intenta abrir nuevos horizontes explicativos sobre las clasificaciones naturales con las cuales trabajamos cotidianamente.
- IV. 5 Optimalidad del modelo: el modelo aprovecha la disponibilidad de datos públicos a nivel de caso y sobre él, desarrolla un sistema clasificatorio, que tiene dos propiedades interesantes a saber: bajo costo de recolección de los datos y transparencia de los mismos.
- IV. 6 Actualización y refinamiento del modelo: Vista la extensa serie histórica disponible de la EPH, es viable realizar actualizaciones periódicas de las categorías, así como refinamientos de análisis o construcciones específicas de interés.
- IV. 7 Por ultimo, pero no por ello menos importante: la construcción del modelo permitió a nuestro Grupo tratar de atisbar el complejo tejido de métodos y técnicas, tendencias y corrientes, fuerzas y modas, remolinos y fragmentos donde nuestra industria está sin dudas inmersa. Este modelo, en forma de ensayo intenta aportar una visión algo diferente sobre la trama de nuestros datos.

#### **Bibliografía:**

- ALGUNAS REFLEXIONES SOBRE LA DESCRIPCION DEMOGRAFICA DE LOS CONSUMIDORES/CLIENTES DE ARGENTINA – JUAN JOSE LLORET – SAIMOEXPRESS Nº 5 ENERO 2003
- BASE USUARIA AMPLIADA DICCIONARIO O DISEÑO DE REGISTRO; EPH ONDA MAYO 2003
- CLASSIFICATION AND REGRESSION TREES – LELAND WILKINSON – MANUAL SYSTAT
- ECONOMETRIA ESPACIAL APLICADA A LA PREDICCION-EXTRAPOLACION DE DATOS MICRO TERRITORIALES – TESIS DOCTORAL – DRA. CORO CHASCO YRIGOYEN – COMUNIDAD DE MADRID – CONSEJERIA DE ECONOMIA E INNOVACION TECNOLOGICA. ABRIL 2003-12-31
- EL IDIOMA ANALÍTICO DE JOHN WILKINS - JORGE LUIS BORGES; OTRAS INQUISICIONES; 1937-1952
- INSE; AAM 1996; INSE 2002; AAM 2003
- LA ESTRUCTURA DE LOS CONCEPTOS CIENTÍFICOS: JESUS MOSTERIN – INVESTIGACIÓN Y CIENCIA - ENERO 1978
- MACHINE LEARNING:TOM MITCHELL MC GRAW HILL BOSTON 1997
- STANDARD DEMOGRAPHIC CLASSIFICATION; ESOMAR 1997
- UPDATING THE SOCIOECONOMIC STRATIFICATION IN SPAIN – IGNACIO REDONDO BELLON; MARKETING AND RESEARCH TODAY; FEBRERO 1999
- UTILIZACIÓN DE CLASIFICADORES DE OCUPACIÓN EN LAS BASES DE DATOS DE LA ENCUESTA PERMANENTE DE HOGARES; EPH SEPTIEMBRE 2000