

**LA SOCIEDAD DE LA INFORMACIÓN ANALIZADA MEDIANTE OBJETOS  
SIMBÓLICOS**

**Patricia Calvo Garrido, Yolanda Pérez Díez**



**EUSKAL ESTATISTIKA ERAKUNDEA  
INSTITUTO VASCO DE ESTADISTICA**

Donostia-San Sebastián, 1

01010 VITORIA-GASTEIZ

Tel.: 945 01 75 00

Fax.: 945 01 75 01

E-mail: [eustat@eustat.es](mailto:eustat@eustat.es)

[www.eustat.es](http://www.eustat.es)

# LA SOCIEDAD DE LA INFORMACIÓN ANALIZADA MEDIANTE OBJETOS SIMBÓLICOS <sup>1</sup>

Patricia Calvo Garrido, Yolanda Pérez Díez

## RESUMEN

A través de este trabajo y utilizando los resultados de la Encuesta de la Sociedad de la Información 2000, elaborada en Eustat, se pretende conocer los grupos que destacan en el uso de nuevos equipamientos: informáticos, televisivos y otros con ellos relacionados (fax, teléfono móvil, mensáfono...) desde una nueva perspectiva del análisis estadístico, el de los Objetos Simbólicos. Los Objetos Simbólicos son un modo de representación de datos complejos que surgen al analizar grandes ficheros de datos. Cada objeto representa un grupo de individuos con características comunes que resultan del cruce de variables sociodemográficas (sexo, edad, ...) y se tratan como nuevas unidades estadísticas. Este tipo de análisis nos permite estudiar la Sociedad de la Información mediante la creación de grupos de población que reflejen las diferencias de uso de los distintos equipamientos. Estos grupos estarían descritos por las siguientes variables:

- Los equipamientos y el uso que de ellos hace la población en su hogar, en su centro de estudio o en su centro de trabajo.
- Características de los usuarios de Internet: frecuencia, duración, antigüedad y fin principal de la conexión.

**Palabras clave:** Sociedad de la Información, Objeto Simbólico, Metodología.

---

<sup>1</sup> Artículo presentado en el V Congreso Vasco de Sociología celebrado en Bilbao en Marzo de 2001.

---

# Indice

LA SOCIEDAD DE LA INFORMACIÓN ANALIZADA MEDIANTE OBJETOS SIMBÓLICOS .....	1
INDICE .....	3
INTRODUCCIÓN .....	4
DEFINICIÓN .....	4
METODOLOGÍA.....	4
VISUALIZACIÓN DE CLASES DE INDIVIDUOS.....	6
DATOS DE LA ENCUESTA DE LA SOCIEDAD DE LA INFORMACIÓN .....	7
INFORMACIÓN RECOGIDA EN LA ENCUESTA .....	7
RESULTADOS.....	7
ANÁLISIS DE CLUSTERS .....	10
CONCLUSIONES .....	13
BIBLIOGRAFÍA.....	14

# Introducción

## Definición

Un Objeto Simbólico es un nuevo concepto estadístico que resume grandes cantidades de información almacenadas en bases de datos. Los datos van a ser complejos, es decir, ya no van a ser de individuos simples sino que van a representar grupos de individuos y las variables no van a tener un solo valor o categoría en cada individuo sino que sus valores pueden ser no atómicos (un grupo de valores, un intervalo de valores o una distribución de probabilidad). Podemos ver un ejemplo de un grupo compuesto por una variable de intervalo: Grupo1 = [20,60] o un grupo descrito por una variable con varias categorías, Grupo2 = {Ocupado (0.47), Parado (0.11), Inactivo (0.42)}.

Además, los objetos simbólicos al ser unidades resumen van a contener una variabilidad interna asociada.

## Metodología

El proceso de creación de objetos simbólicos tiene como punto de partida consultas a una base de datos relacional. Por medio de estas consultas se agregan individuos con características comunes, como por ejemplo, familias, regiones, etc. El proceso de selección de la población se hace mediante una consulta SQL, desde el software SODAS a una base de datos, que especifique qué clases se quieren formar y qué variables van a describir esos grupos. El resultado de una consulta va a ser un conjunto de objetos simbólicos que describen clases o grupos de individuos.

Los objetos simbólicos creados son también almacenados en tablas, llamadas tablas simbólicas. Cada celda de estas tablas, con objetos simbólicos por filas y variables por columnas, puede contener datos de diferentes tipos, tales como:

1. *Un valor:*
  - *cuantitativo:* edad (w) = 23

- *cualitativo*: sexo ( $w$ ) = mujer
2. *Varios valores*:
- *cuantitativos*: peso ( $w$ ) = {48, 52, 56}
  - *cualitativos*: estado civil ( $w$ ) = {soltero, casado}
3. *Intervalo*: edad ( $w$ ) = [20, 25]
4. *Varios valores con pesos*: edad ( $w$ ) = [20 (0.65), 25 (0.35)], histograma o función de pertenencia;

Siendo *edad*, *sexo*, *estado civil* y *peso* variables y  $w$  unidades.

**Tabla 1.** *Tabla Simbólica*

	<b>Sexo</b>	<b>Edad</b>	<b>Profesión</b>
OS 1	{mujer (0.33), varón (0.67)}	{{25:57}}	{Técnico y profesionales superiores (0.35), Personal Directivo (0.25), Jefes Administrativos (0.4)}
OS 2	{mujer (0.5), varón (0.5)}	{{18:42}}	{Comerciantes y Vendedores (0.55), Administrativos (0.45)}

En la tabla 1, cada objeto simbólico (por filas) representa un grupo de individuos con características comunes descrito por 3 variables. La tabla entera corresponde a una consulta a una base de datos.

Las clases de individuos (objetos agregados) pueden crearse, aparte de con conocimiento experto o análisis previos para formar grupos, mediante una sola variable o combinación de varias. Un ejemplo en el que el objeto simbólico es una sola modalidad de una variable, en este caso, del estado civil:

os "**Casado**" (1684) =

[sexo = {"Mujer"(0.498812), "Varón"(0.501188)}] ^ [nivel\_instrucción = {"Estudios Primarios o menos"(0.675178), "Estudios Secundarios"(0.224466), "Estudios Universitarios"(0.100356)}] ^ [relación\_actividad = {"Parados que buscan empleo"(0.00296912), "Parados que han trabajado"(0.0647268), "Ocupados"(0.465558), "Inactivos"(0.466746)}] ^ [edad = [16:90]]

Si el atributo de grupo se compone del cruce de dos o más variables nominales, se obtendrán tantos objetos simbólicos como producto de modalidades de las variables, sexo por estado civil:

os "**Casado / Varón**" (844) =

```
[nivel_instrucción = {"Estudios Primarios o menos"(0.622043), "Estudios Secundarios"(0.25979), "Estudios Universitarios"(0.118167)}] ^ [relación_actividad = {"Parados que buscan empleo"(0.00130055), "Parados que han trabajado"(0.034125), "Ocupados"(0.677499), "Inactivos"(0.287075)}] ^ [edad = [16:90]
```

## Visualización de Clases de Individuos

La visualización de un objeto simbólico se hace mediante un Zoom Star, que es un gráfico radial donde cada eje representa una variable. En el mismo gráfico pueden representarse variables categóricas, intervalos, con pesos,...

Hay dos tipos de representación, en 2D y 3D, que nos proveen diferentes niveles de detalle. La representación en 2D permite una impresión global del objeto, mientras que la representación en 3D nos da información más detallada.

En 2D los ejes están unidos por una línea que conecta los valores más frecuentes de cada variable. Cuando existe una variable intervalo la línea se une a los límites inferior y superior. En la representación 3D se ve la distribución correspondiente a cada variable con pesos. Las variables numéricas se representan por rectángulos que van del valor mínimo al máximo.

En el siguiente apartado se muestra un ejemplo de visualización en 3D.

## Datos de la Encuesta de la Sociedad de la Información

La E.S.I.(Encuesta de la Sociedad de la Información) es una encuesta por muestreo sobre la población de la CAE de 15 y más años. Se toma como base del muestreo el panel de 3.750 viviendas familiares seleccionadas para la P.R.A. (Población en Relación con la Actividad) en el mismo período. Se selecciona una primera persona por vivienda y, de forma especial, determinadas cuotas de estudiantes y de ocupados, hasta completar una muestra cercana a los 5.000 individuos (4.627 en IV/2000).

### Información recogida en la Encuesta

**Equipamiento en Tecnologías de la Información:** Equipamiento Informático (ordenador personal, módem-fax, lector de CD-Rom, correo electrónico, conexión a internet), Equipamiento Televisivo (aparato de vídeo, TV con teletexto, antena parabólica, descodificador para TV de pago, TV por cable) y Otro Equipamiento Tecnológico (fax, teléfono móvil, mensáfono, busca).

**Usuarios de Internet:** Comercio electrónico, días de mayor acceso, lugar de acceso, posibilidad de acceso a internet, servicios tecnológicos (consultas a páginas Web, correo electrónico, transferencia de ficheros (FTP), chats y conversaciones).

### Resultados

Los objetos simbólicos que se han creado resultan de la combinación de 4 variables: sexo (2 modalidades), edad (7), nivel de instrucción (3) y relación con la actividad (4). Tendríamos 168 objetos simbólicos, pero al extraer solo los usuarios de internet en casa, que es el colectivo que vamos a describir a continuación, nos encontramos que se reduce a 53 objetos.

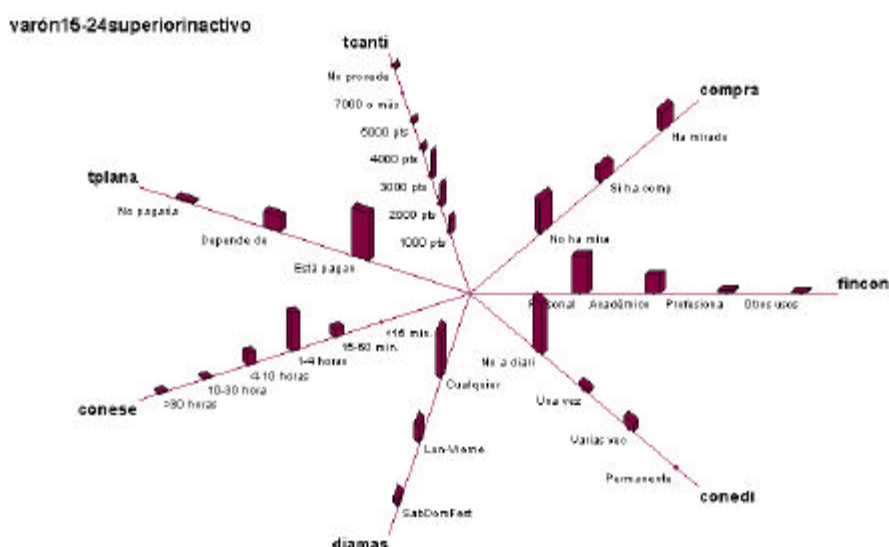
Las variables que describen los grupos de usuarios de Internet en casa son: Fin principal de la conexión, *fincon*; Conexión diaria, *conedi*; Conexión semanal, *conese*; Días de mayor acceso, *díamas*; Tarifa plana, *tplana*; Importe de la tarifa, *tcanti*; Compra a través de internet, *compra*.

Se han seleccionado dos objetos que tienen cierta relevancia en el uso de internet, jóvenes comprendidos entre los 15 y 24 años, de ambos sexos, realizando estudios superiores, y que no están trabajando ni buscando empleo.

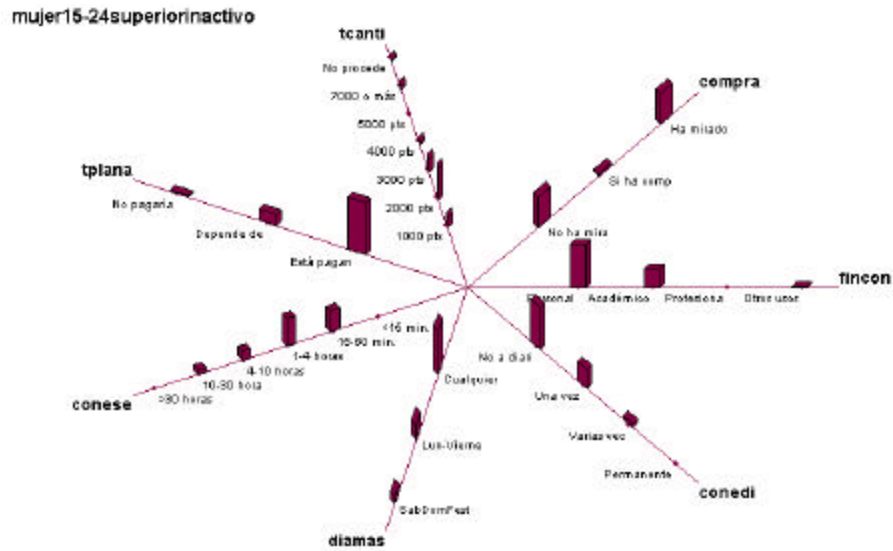
En los siguientes gráficos podemos ver que en ambos sexos el fin principal de la conexión es personal, no realizan diariamente una conexión pero la duración semanal de la/s conexiones se contabiliza entre 1 y 4 horas, con un predominio entre los varones, ya que entre ellos el 60% está conectado en ese intervalo de tiempo, mientras que entre las mujeres se reduce ese porcentaje a un 44 %; lo que sí comparten es que pagan tarifa plana las tres cuartas partes de los usuarios jóvenes de internet.

Se puede distinguir que los varones pagan más dinero por su conexión, alrededor de las 3000 ptas., mientras que en las mujeres rondan las 2000 ptas.

En relación a su actitud ante la compra de artículos vía internet, todavía se aprecia como incipiente, puesto que consultan o miran el 47% comprando sólo el 5%, en el caso de las mujeres, mientras que entre los varones, la proporción de los que buscan o miran para comprar a través de internet es menor, pero se deciden por la compra en un porcentaje mayor que ellas, un 20%.







**Figuras 1 y 2.** Zoom Star 3D de 2 objetos simbólicos

El comentario de estos dos objetos simbólicos no es más que un ejemplo que nos introduce en una nueva metodología para analizar los resultados de cualquier encuesta o estadística desde una perspectiva de grupo, en la que se destacan los rasgos que predominan en el grupo u objeto elegido.

Como trabajo previo a la generación de objetos se puede realizar un análisis que mida la asociación entre variables y la correlación con el fenómeno social a analizar si es que se desconoce. Siempre se puede recurrir a las variables sociodemográficas que explican la mayoría de los acontecimientos sociales, tales como sexo, edad, estado civil, nivel de instrucción, relación con la actividad, rama de actividad, tendencia política...por mencionar las más usuales en las encuestas.

## Análisis de Clusters

Uno de los muchos métodos de análisis que se puede aplicar sobre objetos simbólicos es la **clasificación divisiva**, procedimiento DIV del software SODAS, con el que se pueden encontrar grupos más o menos homogéneos de población que reflejen las diferencias en el uso de distintos equipamientos. Dicho método es un método de clasificación jerárquica “de arriba a abajo” que persigue la construcción de grupos empezando con un gran cluster y dividiendo sucesivamente hasta llegar a clusters más pequeños. Las variables que definen los objetos son las encargadas de dividir el árbol. El árbol que se crea es binario, es decir, a un lado se sitúa la modalidad de corte y en el otro el resto de modalidades que componen esa variable.

En este ejemplo se han construido 35 objetos simbólicos de población ocupada resultantes del cruce de las variables: sexo, edad y nivel de instrucción. Las variables que describen estos objetos son las que explican la disponibilidad de equipos informáticos en el lugar de trabajo. A estos objetos se les ha aplicado la clasificación divisiva con el objeto de encontrar grupos homogéneos de población teniendo en cuenta el uso que hagan de estos equipos informáticos.

A continuación se incluye la salida con los resultados de la clasificación. Se ha elegido como óptima la partición en 6 grupos y en cada grupo se detallan los objetos simbólicos que pertenecen al mismo.

```
PARTITION IN 6 CLUSTERS:
-----:
```

```
Cluster 1 (n=1):
```

```
"Varón >65 Secundario"
```

```
Cluster 2 (n=11):
```

```
"Mujer 16-24 Superior" "Mujer 25-34 Secundario" "Mujer 35-44 Secundario"
```

```
"Mujer 45-54 Secundario" "Mujer 45-54 Superior" "Mujer 55-64 Secundario"
```

```
"Varón 16-24 Superior" "Varón 25-34 Secundario" "Varón 35-44 Secundario"
```

```
"Varón 45-54 Secundario" "Varón 55-64 Secundario"
```

```
Cluster 3 (n=6):
```

"Mujer 25-34 Superior" "Mujer 35-44 Superior" "Varón 25-34 Superior" "Varón 35-44 Superior" "Varón 45-54 Superior" "Varón 55-64 Superior"

**Cluster 4** (n=1):  
"Varón >65 Primario"

**Cluster 5** (n=15):  
"Mujer >65 Primario" "Mujer >65 Secundario" "Mujer 16-24 Primario" "Mujer 16-24 Secundario" "Mujer 25-34 Primario" "Mujer 35-44 Primario" "Mujer 45-54 Primario" "Mujer 55-64 Primario" "Varón >65 Superior" "Varón 16-24 Primario" "Varón 16-24 Secundario" "Varón 25-34 Primario" "Varón 35-44 Primario" "Varón 45-54 Primario" "Varón 55-64 Primario"

**Cluster 6** (n=1):  
"Mujer 55-64 Superior"

Explicated inertia: 80.059932

THE CLUSTERING TREE:

-----

- the number noted at each node indicates the order of the divisions
- Ng <-> yes and Nd <-> no

```

          +---- Classe 1 (Ng=1)
          !
!----4- [Móvil = Si tiene]
!      !
!      +---- Classe 5 (Nd=15)
!
!----1- [Ordenador = No tiene]
!
!      +---- Classe 2 (Ng=11)
!      !
!      !----3- [Frecuencia de uso del ordenador = Diariamente]
!      !      !
!      !      +---- Classe 4 (Nd=1)
!      !      !
!----2- [Internet = No tiene]
!
!      +---- Classe 3 (Ng=6)
!      !
!----5- [Frecuencia de uso del ordenador = Diariamente]
!
!      +---- Classe 6 (Nd=1)

```

La lectura del árbol es la siguiente: en cada corte, la rama superior es la opción afirmativa y la rama inferior es la opción negativa. Es decir, en el primer corte la variable que más discrimina la población ocupada es la disponibilidad de ordenador, por lo que a la rama de arriba pertenece la población sin ordenador en el puesto de trabajo, y a la de abajo la población con ordenador.

Los clusters 1 y 5 agrupan a los que no disponen de ordenador en el puesto de trabajo. Están compuestos por personas ocupadas, que bien su nivel de instrucción es primario o bien su edad es elevada, superando en algunos casos los 65 años. De los que no usan ordenador nos encontramos que la siguiente variable de corte es el uso del teléfono móvil para el trabajo.

Los que sí tienen ordenador (rama inferior del primer corte) se diferencian por:

- Los que tienen acceso a Internet, que además se distinguen por el uso diario o no del ordenador, clusters 3 y 6 respectivamente. En el 3 se encuentran 6 grupos, todos ellos caracterizados por un nivel de estudios superior, en el caso de las mujeres entre 25 y 44 años, para los varones la distribución de las edades abarca prácticamente toda la vida laboral, desde los 25 hasta los 64 años. El cluster 6 se define por los que sí tienen ordenador, sí tienen internet, pero el ordenador no lo utilizan a diario. Es un grupo cuya definición queda establecida por un pequeño número de mujeres ocupadas de 55-64 años con estudios superiores.
- Los que no tienen acceso a internet, clusters 2 y 4, también vuelven a distinguirse por su uso diario o no del ordenador. Así, la clase 2 está compuesta por aquellos grupos que sí utilizan el ordenador a diario, son clases compuestas sobre todo por personas con estudios secundarios y edades intermedias; a excepción de dos grupos con estudios superiores y que tienen menos de 25 años y otro, de mujeres de 45-54 años también con estudios superiores. La clase 4 está compuesta por los que no usan el ordenador a diario, y corresponde a varones de más de 65 años y con estudios primarios, lo cual representa a un grupo también especial en cuanto a sus efectivos.

---

## Conclusiones

Los objetos simbólicos son una nueva representación de conceptos estadísticos, que permiten el análisis de datos agregados a partir de la combinación de variables seleccionadas. Añade nuevas posibilidades en la visualización de la información, así como nuevos métodos de uso, diseño y análisis estadístico facilitando al investigador la interpretación de los resultados.

El interés en la utilización de este método sobre los datos de la Encuesta de la Sociedad de la Información radica en que es una encuesta todavía en periodo de implantación, y nos sirve para ir descubriendo perfiles de usuarios en las nuevas tecnologías.

## Bibliografía

[1] Bock, H. H. y Diday, E.

*Analysis of Symbolic Data*. Springer-Verlag. (2000).

[2] Calvo, P.

*Aplicaciones de los Objetos Simbólicos en la Estadística Oficial*. Cuaderno Técnico de EUSTAT. (2000).

[3] Calvo P., Prado C., Pérez Y., Ayestarán M.

*Creación de Objetos Simbólicos a partir de encuestas almacenadas en Bases de Datos Relacionales*. Metodología de Encuestas, Vol 2, Núm 2, 239-260. (2000).

[4] Diday, E.

*Analyse des données et classification automatique numérique et symbolique*. Seminario Internacional de Estadística en Euskadi. Volumen 27. EUSTAT. (1992).