

LOS MÉTODOS FACTORIALES Y LA CLASIFICACIÓN DE DATOS

LOS MÉTODOS FACTORIALES Y LA CLASIFICACIÓN DE DATOS

Las técnicas factoriales están diseñadas para el análisis de grandes cantidades de variables que, por su volumen, son poco manejables con otras técnicas estadísticas. Estos métodos factoriales trabajan reduciendo la variedad de las observaciones a unas pocas variables-resumen o macrovariables que presentan, de manera sintética, las asociaciones existentes entre las variables de origen que fueron utilizadas para el cálculo inicial. Estas macrovariables –que realmente se llaman factores– se expresan como vectores sobre los cuales tanto las variables como los individuos tienen unas coordenadas concretas. Es la posición de éstas sobre los factores lo que nos ayuda a interpretar su sentido. Por ejemplo, si estamos analizando a los jóvenes, y vemos que éstos se ordenan en torno al factor según el tipo de actividades que ejecutan en su tiempo libre, el factor se refiere entonces a actividades realizadas durante el ocio, tales como actividades intelectuales, manuales o deportivas.

El punto de partida del análisis factorial son las tablas cruzadas, donde en una matriz se cruzan las variables y las modalidades de éstas¹ contra ellas mismas. Cuando el volumen de las variables que se deben considerar en el análisis es alto, las

tablas cruzadas resultan difíciles de manejar, pues la cantidad de casillas que hay que tener en cuenta es inmanejable para el cálculo y la interpretación. Si éste fuera el caso, y se hiciera el análisis por la tabla de datos, se buscarían aquellas celdas de las tablas donde se concentran los datos en una proporción mayor o menor a la distribución del universo en cuestión. Es la misma lógica que explicamos con las tablas de contingencia para la descripción de los tipos culturales en el análisis de la Encuesta del Proyecto Pobreza. Estas últimas son funcionales para búsquedas bien específicas –tal como lo hicimos en ese caso– no obstante resultan limitadas si buscamos relaciones entre muchas variables.

El análisis factorial tiene entonces por objetivo resumir esta gran cantidad de variables en unos pocos valores que hagan legible el fenómeno en estudio, para así poder comprender más fácilmente la variedad de relaciones existentes entre las variables. Por esto, el análisis factorial es principalmente una técnica exploratoria, busca las posibles asociaciones existentes dentro del conjunto de variables. Más que trabajar con las variables, trabaja con las modalidades de las variables; por ejemplo, no trabaja directamente con la variable “nivel educativo” sino con los valores que ésta puede tomar: sin nivel educativo, primaria, secundaria, superior.

El factor, siendo entonces un vector, presenta valores positivos y negativos a ambos lados, y es la oposición entre los polos de estos factores lo que ayuda a describir el fenómeno sintetizado en el factor. En el ejemplo que mencionábamos, sobre los jóvenes y el tiempo libre, pudiéramos encontrar en un primer factor, por un lado, los casos de los jóvenes con mucho tiempo disponible, y por el otro los casos de los jóvenes con poco o ningún tiempo disponible para el ocio. Probablemente en un segundo factor, pudiéramos encontrar los casos de aquellos que disponen de tiempo libre, poniendo en un lado los jóvenes dedicados al deporte y, en el otro extremo, los jóvenes dedicados a actividades intelectuales.

Cada uno de estos factores tiene un indicador que señala cuánto aporta dicho factor al conjunto total de factores, lo que ayuda a la exploración de los datos. No todos los factores tienen la misma capacidad de explicación sobre el fenómeno en estudio. El primer factor calculado es el que tiene la mayor capacidad de explicación, seguido de manera jerárquica en el mismo orden por los otros factores calculados. También se calcula

1

La variable es la característica que se busca medir en los casos investigados, y las dimensiones son los valores que toma esta variable. Por ejemplo: una variable es la creencia religiosa, y sus dimensiones son catolicismo, judaísmo, islamismo, protestantismo, y otras que pueden ser agrupadas según el interés del investigador.

2

El análisis de correspondencias múltiples es el tipo de análisis factorial que aquí nos ocupa. Sin embargo, el resto de las técnicas de análisis factorial operan en general bajo la misma lógica.

3

Aunque realmente usaríamos la mitad de las 784 casillas, pues habría un doble conteo. Por ejemplo, al cruzar el tipo de vivienda con el servicio de agua tendríamos la misma información cuando se llegara al cruce inverso.

cuánto contribuye cada variable a la construcción de los factores. De esta manera podemos observar cuál es la variable que ayuda a definir el factor. En el caso del análisis de correspondencias múltiples², todo el conjunto de factores (cuyo número responde a la regla siguiente: número de modalidades – número de variables) explican el 100% de las variaciones existentes entre las variables, esto a pesar de que, por lo general, más allá del cuarto o quinto factor es difícil el análisis, pues realmente se trata de los residuos “estadísticos” del análisis.

Para ilustrar esto, utilizaremos los mismos datos de la Encuesta del Proyecto Pobreza. Bien se sabe que si tomamos un conjunto de variables como el agua, el aseo, el nivel educativo, el tipo de vivienda, el centro poblado y el ingreso, estaremos midiendo en conjunto la calidad de vida de la persona, es decir, si ésta es pobre o no. Al ingresar esas variables para un análisis factorial obtenemos 22 factores, pues hay un total de 28 modalidades menos 6 variables. Si hiciéramos uso de tablas cruzadas tendríamos una tabla de 784 casillas, 28 x 28 modalidades³. En el cuadro siguiente podemos observar cuáles son estas variables y sus modalidades, y el peso porcentual de cada una en el cálculo de los tres primeros factores. Cada modalidad de las variables pesa porcentualmente por separado en la construcción de cada uno de los factores; la suma de estas contribuciones equivale a la contribución total de la variable. Cuando se calculó el peso de las variables en la orientación actitudinal, en la parte sobre el análisis de datos, hicimos uso de estas tablas. Podemos notar cómo el tipo de vivienda es la variable que más contribuye al cálculo del primer factor, especialmente la modalidad “apartamento”, como también el nivel “A” de los ingresos: la clase alta. Igualmente podemos notar la presencia significativa de otras modalidades en el extremo opuesto del factor, como la ausencia de aseo, y el resto de las ciudades pequeñas como lugar de residencia.

VARIABLES	FACTORES		
MODALIDADES	1	2	3
Agua			
EN LA VIVIENDA	1,1	0,8	0
FUERA DE LA VIVIENDA	5,2	7,7	4,8
NO TIENE	2,8	0,1	15,2
CONTRIBUCIÓN (%) TOTAL	9,2	8,7	20
Aseo			
LLEGA A LA VIVIENDA	2,4	4,5	0,3
NO LLEGA A LA VIVIENDA	1,4	12	17,6
NO TIENE	8,4	1,6	20,2
CONTRIBUCIÓN (%) TOTAL	12,1	18,1	38,1
Nivel educativo			
ANALFABETO	1,8	0,9	3,1
SIN NIVEL	2,6	1,7	0,2
PRIMARIA	3,3	0,4	1,8
SECUNDARIA	2,1	1,5	0,5
TÉCNICO	2,7	0	0,3
UNIVERSITARIO	7,5	3,9	5,5
CONTRIBUCIÓN (%) TOTAL	20	8,5	11,3
Tipo de vivienda			
RANCHO	4,9	6,6	2,9
APARTAMENTO	11,8	12,4	0,5
CASA	1,9	4,1	0
CASA QUINTA	4	2,4	1,4
CONTRIBUCIÓN (%) TOTAL	22,5	25,5	4,8
Tamaño del centro poblado			
GRAN CARACAS	5,8	16,8	2,7
GRANDES CIUDADES	1,9	3	0,5
CIUDADES INTERMEDIAS	0	5,6	0,6
CIUDADES PEQUEÑAS	0,3	3	1,5
RESTO	9,6	1,5	5
CONTRIBUCIÓN (%) TOTAL	17,7	29,8	10,3
Ingreso per cápita			
NIVEL G	0,3	0,9	1,7
NIVEL F	3,7	0	2,9
NIVEL E	1,8	0,2	0,9
NIVEL D	0,5	0,9	3,6
NIVEL C	0	1,6	1,6
NIVEL B	1,1	0,7	0,8
NIVEL A	11,1	5,2	4
CONTRIBUCIÓN (%) TOTAL	18,5	9,5	15,6

Como el orden en que los factores son calculados obedece a su capacidad explicativa; en otras palabras, el factor que mejor resume los datos es el primero, y el que menos los resume el

4

La noción de negativo o positivo en el cálculo del análisis factorial obedece a hechos matemáticos y no valorativos.

último. Por eso la mayoría de las veces se utilizan en el análisis los primeros cuatro o cinco factores que por lo general logran un buen nivel explicativo; el resto de los factores explican más bien las relaciones estadísticamente residuales. En este caso los cinco primeros factores explican un 33,43% de la variación de los datos, tal como podemos observar en la tabla siguiente, donde se presentan los porcentajes explicativos de los 22 factores utilizados. Si quisiéramos abarcar una mayor capacidad explicativa nos veríamos en la obligación de considerar, por lo menos, los primeros ocho factores. El problema es que, tal como lo hemos mencionado, el análisis más allá del cuarto o quinto factor resultaría arduo. Por esto, la tendencia en este tipo de análisis es tratar, en la medida de lo posible, a través de algunos reprocesamientos de la información, que los primeros cinco factores tengan una explicación mayor, en este caso, al 33%.

NÚMERO DEL FACTOR	PORCENTAJE	PORCENTAJE ACUMULADO	NÚMERO DEL FACTOR	PORCENTAJE	PORCENTAJE ACUMULADO
1	10,53	10,53	12	4,44	65,62
2	6,76	17,29	13	4,35	69,97
3	5,88	23,17	14	4,22	74,19
4	5,18	28,35	15	4,11	78,3
5	5,08	33,43	16	3,9	82,2
6	4,82	38,25	17	3,72	85,92
7	4,65	42,9	18	3,56	89,48
8	4,63	47,53	19	2,95	92,42
9	4,62	52,15	20	2,74	95,17
10	4,55	56,71	21	2,6	97,77
11	4,47	61,17	22	2,23	100

Así, a efectos de esta explicación, obtenemos los 22 factores que describen las relaciones implícitas dentro del volumen de datos analizado. Como lo mencionábamos anteriormente, cada factor es un vector sobre el cual tanto los individuos como las modalidades de las variables tienen una coordenada; y la distribución a lo largo del factor, en particular de las modalidades, es lo que nos permite desentrañar el fenómeno implícito en los factores. En el cuadro siguiente podemos observar, en los dos primeros factores, cómo se distribuyen las modalidades más significativas de las variables a ambos extremos de los factores. Para el primer factor tenemos que el extremo negativo⁴ presenta las modalidades relacionadas con la mala calidad o ausencia de servicios públicos, muy bajo nivel educativo, viviendas

marginales, bajos niveles de ingreso y localización en el resto de las ciudades pequeñas. Al otro lado del factor tenemos que se señala exactamente lo contrario: altos niveles educativos, buenas viviendas, altos niveles de ingreso, y grandes ciudades, especialmente Caracas. La existencia de los servicios públicos no aparece en este extremo –si bien apareció en el otro– pues la presencia de agua y aseo no es sólo un atributo de las clases más pudientes sino también, de manera más general, de las clases no pobres. En pocas palabras, este factor opone a los ricos contra la pobreza extrema.

El segundo factor construye una oposición complementaria. Si observamos el lado positivo, veremos que aparecen los atributos que caracterizan tanto a la pobreza extrema como a la clase más pudiente: sin servicios públicos, vivienda marginal, junto a apartamentos y nivel universitario. El lado negativo del factor nos presenta justamente esa parte del universo que se encuentra entre esos dos extremos, incluyendo hasta los casos caracterizados por la no respuesta. Este mundo “del medio” se caracteriza principalmente por su niveles de ingreso C y D y su localización en las ciudades intermedias, pequeñas y grandes pero no Caracas.

Primer factor

COORDENADAS	MODALIDAD DE LA VARIABLE	VARIABLE
-1,52	NO TIENE	ASEO
-1,25	NO TIENE	AGUA
-1,23	FUERA DE LA VIVIENDA	AGUA
-1,06	ANALFABETO	NIVEL EDUCATIVO
-0,99	RANCHO	TIPO DE VIVIENDA
-0,93	RESTO	TAMAÑO DEL CENTRO POBLADO
-0,80	SIN NIVEL	NIVEL EDUCATIVO
-0,78	NIVEL F	INGRESO PER CÁPITA
-0,67	NO CONTESTA	ASEO

Zona central del factor

0,41	SECUNDARIA	NIVEL EDUCATIVO
0,43	GRANDES CIUDADES	TAMAÑO DEL CENTRO POBLADO
0,43	NIVEL B	INGRESO PER CÁPITA
0,82	GRAN CARACAS	TAMAÑO DEL CENTRO POBLADO
0,96	TÉCNICO	NIVEL EDUCATIVO
0,96	CASA QUINTA	TIPO DE VIVIENDA
1,32	NIVEL A	INGRESO PER CÁPITA
1,46	APARTAMENTO	TIPO DE VIVIENDA
1,53	UNIVERSITARIO	NIVEL EDUCATIVO

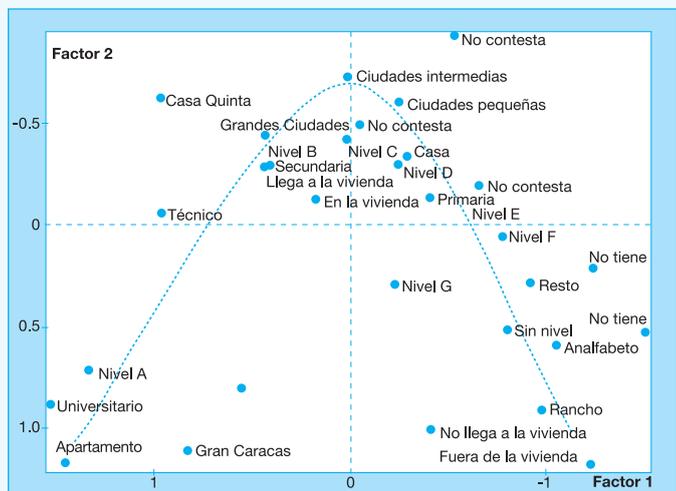
Segundo factor

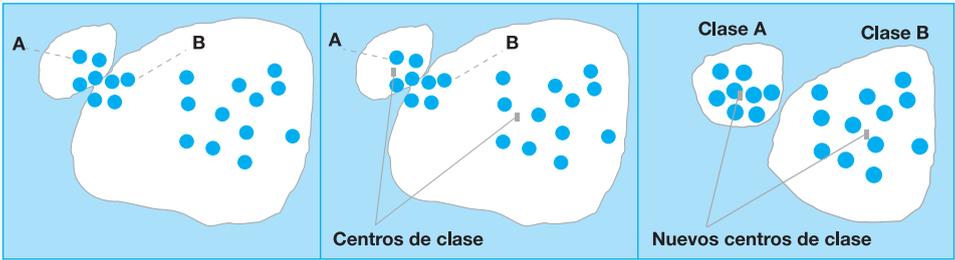
COORDENADAS	MODALIDAD DE LA VARIABLE	VARIABLE
-0,92	NO CONTESTA	NIVEL EDUCATIVO
-0,71	CIUDADES INTERMEDIAS	TAMAÑO DEL CENTRO POBLADO
-0,6	CASA QUINTA	TIPO DE VIVIENDA
-0,59	CIUDADES PEQUEÑAS	TAMAÑO DEL CENTRO POBLADO
-0,48	NO CONTESTA	AGUA
-0,42	GRANDES CIUDADES	TAMAÑO DEL CENTRO POBLADO
-0,41	NIVEL C	INGRESO PER CÁPITA
-0,31	NIVEL D	INGRESO PER CÁPITA
-0,31	CASA	TIPO DE VIVIENDA
Zona central del factor		
0,54	NO TIENE	ASEO
0,6	ANALFABETO	NIVEL EDUCATIVO
0,72	NIVEL A	INGRESO PER CÁPITA
0,89	UNIVERSITARIO	NIVEL EDUCATIVO
0,91	RANCHO	TIPO DE VIVIENDA
1,01	NO LLEGA A LA VIVIENDA	ASEO
1,12	GRAN CARACAS	TAMAÑO DEL CENTRO POBLADO
1,2	APARTAMENTO	TIPO DE VIVIENDA
1,2	FUERA DE LA VIVIENDA	AGUA

Siendo que ambos factores son vectores, podemos perfectamente cruzarlos para ilustrar mejor lo que se contempla en estos dos primeros factores. Al cruzar los factores y ubicar las modalidades según sus coordenadas en cada factor, obtenemos un gráfico como el que se muestra a continuación. El factor 1, el cual representaba los extremos (ricos y pobreza extrema) se encuentra en el eje horizontal; el factor 2, el de las clases intermedias y las clases extremas, se encuentra en el eje vertical. El cruce produce entonces una curva donde desde la parte inferior izquierda se encuentran las clases más pudientes (nivel A de ingresos, universitarios, apartamento y Caracas). Siguiendo la curva observamos que en la parte superior izquierda –exactamente antes de llegar a la cima de la curva–, se encuentran los grupos medios (técnicos, niveles B y C, ciudades grandes e intermedias, casas quintas). Al empezar a bajar la curva notamos que las modalidades describen a los grupos más vulnerables: nivel D, E y F, nivel educativo de primaria, casas, mala calidad de servicios públicos) hasta llegar a la pobreza extrema: servicios públicos ausentes o fuera del hogar, vivienda marginal, analfabeto o sin nivel educativo.

Obviamente la representación de la curva –ajustada “a mano” y no calculada por el programa– es relativa. Algunas modalidades como “Casa-quinta” o “No contesta” tienden a salirse de la curva; pero sin lugar a dudas se conservan dentro de la tendencia; si observamos como se ordenan las modalidades de cada variable, podremos ver que cada una sigue la tendencia de la curva. Los cuadrantes inferiores contienen a las modalidades de las clases extremas, los cuadrantes superiores a las de las clases medias; a su vez, los cuadrantes del lado izquierdo contienen a las clases no pobres, los cuadrantes del lado derecho a lo que pudiéramos llamar las clases populares y pobres. Finalmente, cada cuadrante contiene, si quisiéramos verlo de esta manera, una clase social. Pero en ningún momento significa que cada cuadrante contenga un 25% de la población; cada grupo es suficientemente representativo, pero el volumen de casos y una clasificación más fina la obtendríamos, valga la redundancia, a través de la clasificación automática.

El análisis factorial, si bien otorga coordenadas a los casos, ayuda a explorar más bien las relaciones entre las variables. Para conocer cómo se ordenarían los casos según lo expresado en los factores, se hace uso de los métodos de clasificación automática. Así como los métodos factoriales exploran las relaciones entre las variables, en el caso de los métodos de clasificación automática se trabaja sobre las relaciones entre los individuos. Estas relaciones entre los individuos toman como punto de referencia los factores construidos en el análisis factorial, pues estos, como lo hemos expresado, ya resumen la información de las variables consideradas en el análisis. Además, es de recordar que los individuos poseen unas coordenadas sobre estos factores. Como todo sistema de coordenadas, éstas ayudan





a ubicar un punto en el espacio y a calcular la distancia entre los puntos. En otras palabras, permiten saber qué puntos se encuentran más cercanos entre sí, y cuáles están más lejanos.

Existen diferentes modelos de clasificación automática: unos trabajan de manera ascendente (de par en par, van agregando primero individuos y luego grupos) y otros de manera descendente (se parte de la división mínima de un conjunto, en dos grupos y luego cada grupo, en momentos diferentes, se divide en dos y así sucesivamente). En el fondo son técnicas parecidas pero que operan en sentido contrario, una divide grupos y la otra los suma. La intención de estos procesos es crear grupos *muy parecidos dentro de sí pero muy diferentes entre ellos*. Para el procesamiento realizado se usó una clasificación descendente.

Explicábamos entonces que cada individuo tiene una posición y, por ende, una distancia con el resto de los individuos que es fijada por las coordenadas sobre los factores. De esta manera, si pudiéramos graficar todos los individuos, por ejemplo, en un gráfico bidimensional de dos factores, observaríamos que se forman especies de nubes de puntos donde cada punto es un individuo. Ahora bien, dentro de esta nube de puntos se pueden elegir al azar dos individuos, digamos A y B. Para el resto de los individuos, se calcularía la distancia que tienen con A y B para identificar dos nubes: los individuos que estén más cerca de A y los que estén más cerca de B. Siendo que el proceso elige los puntos al azar, este mismo proceso aplica mecanismos para corregir y ajustar la identificación de las dos nubes. Tal como lo reflejamos en la gráfica anexa, es probable que la elección de los individuos al azar no refleje, en este primer proceso, dos grupos bien homogéneos. En efecto, algunos individuos están realmente más cerca de A que, por ejemplo, gran parte de los individuos en B.

Para corregir este efecto causado por la elección al azar, el proceso estadístico calcula los centros de cada nube identificada,

es decir, identifica un punto medio dentro de las nubes –que no necesariamente es un individuo– a partir del cual se equiparan las distancias de cada individuo a su centro del grupo o de clase. A partir de este momento se recalculan las distancias de los individuos, pero esta vez con relación a los centros de clase, y los individuos son reagrupados según el centro que les queda más cercano. Este proceso es repetido (se calculan los nuevos centros de clase, se recalculan las distancias y se reagrupan los individuos) hasta el punto en que estos recálculos no llevan a cambiar a los individuos de grupo y, en consecuencia, los centros de clase quedan siempre igual. Este proceso de ajuste se llama iteración. Nótese que en ningún momento se afectan o se alteran las coordenadas originales de los individuos; lo que continuamente se recalcula y se reajusta son las distancias y los centros de clase.

Ahora bien, este proceso que se ha realizado para conformar dos clases –una vez que éstas fueron estabilizadas por las continuas iteraciones– se aplica de igual manera para originar nuevas clases, sólo que en este caso cada grupo es tratado de manera independiente: para cada uno de nuestros dos grupos originados se toman dos individuos, se calculan las distancias, luego los centros de clase, y se recalculan las distancias reafectando a los individuos a los centros de clase más cercano. Así tantas veces como el proceso de iteración continúe. Sin embargo existe una pregunta: ¿Cuál de nuestros grupos dividiremos primero? Si observamos la clase A podemos notar que es más compacta que la clase B, es decir es más homogénea. Esto es así porque la distancia de los puntos al centro de clase es menor en la clase A que en la B. En consecuencia, se divide primero al grupo B pues éste es menos homogéneo que el grupo A. Ese va a ser el criterio del proceso estadístico para dividir subsecuentemente las nuevas clases: dividir primero a los grupos menos homogéneos.

¿Hasta qué punto se crean las nuevas clases? ¿Cuándo se detiene el proceso? Recordemos que la intención es crear grupos bien homogéneos hacia el interior, pero bien heterogéneos en relación con los otros grupos. Si lo decimos en términos de las distancias calculadas, se trata de minimizar las distancias de los individuos hacia el centro de la clase, y de maximizar las distancias “entre” los centros de clase, a través de los procesos de iteración y creación de nuevas clases. El proceso se detiene en el momento en el cual, a través de un cálculo de ponderación, se llega a un óptimo donde las distancias dentro de las

clases son pequeñas en comparación con las distancias que existen entre las clases. Si lo ponemos en un ejemplo extremo, donde en un universo de 100 individuos se crean 50 grupos de dos individuos cada uno, tenemos que al calcular las distancias hacia el interior de estos 50 grupos, y al compararlas con las distancias entre los 50 centros de clases, observaremos que las distancias de los 50 grupos, ponderada en un índice, es casi igual que las de los 100 individuos por separado. Simplemente se agruparon individuos, pero no se clasificaron.

Retomando el ejemplo que veníamos trabajando con el análisis factorial, observamos que los factores calculados nos resuman las variables en torno al fenómeno de las condiciones de vida. Si aplicamos la clasificación automática a estos casos se parte de las coordenadas de los individuos calculadas para los factores, es decir, de la agrupación de los individuos según sus condiciones de vida, pues de esto tratan los factores. No sería sorpresa observar que las dos primeras clases resultantes, luego de sus iteraciones, son justamente los pobres y los no pobres. Bien sabemos –y no hace falta la clasificación– que estas dos clases son insuficientes, pues existe un mundo de individuos que se ubica entre estos dos polos descritos por aquel primer factor. Necesariamente serían afectados a una clase pero, analíticamente hablando, sabríamos que sería inconsistente hablar de dos clases, en cuanto a condiciones de vida se refiere.

En efecto, al proceder con una clasificación automática con esta información, el proceso estadístico “nos sugirió” –pues siempre existe la opción de escoger, por nuestra propia cuenta, las clases deseadas– la conformación de seis clases. Esta clasificación en seis grupos responde al equilibrio entre las diferencias de las distancias dentro de las clases, con las distancias entre los centros de clases.

Una vez que estas clases están conformadas, el procesamiento estadístico las caracteriza. Para realizar esto, se vale del uso de las tablas de contingencia para cada clase, pero presentadas de una manera diferente. Existen dos criterios: las modalidades que describen a una clase y las modalidades que la discriminan. La modalidad que describe a una clase consiste en las modalidades que se presentan con mayor proporción dentro de la clase. Por ejemplo, la primera clase obtenida para este ejemplo, que podemos observar en los cuadros siguientes, se describe porque el 90% de esta clase vive en casas. Igualmente se describe porque el 98% de la clase tiene

servicio de agua dentro de la vivienda. Sin embargo, el dato de “vivir en casas” es más significativo que el dato de “agua en la vivienda”. Esto se debe a que el 88% de las personas entrevistadas tiene el servicio de agua en la vivienda, y el porcentaje presentado para la primera clase, 98%, sólo representa un 10% más que las proporciones globales. En el caso de “vivir en casa”, el 65% del global vive en viviendas, lo que hace mayor diferencia con el 90% de la primera clase. En otras palabras, si los porcentajes de una modalidad dentro de una clase reproducen más o menos los mismos porcentajes del global, ésta no sería muy significativa. Lo que realmente sería significativo es que los porcentajes dentro de las clases se diferenciara de los porcentajes globales.

La segunda manera de caracterizar una clase es con las modalidades que discriminan a una clase. Si la pregunta en el caso de la descripción de una clase fuese ¿cuánta gente de la clase 1 tiene la modalidad A?, la pregunta en el caso de la discriminación sería ¿cuánta de la gente con la modalidad A está en la clase 1?. Siguiendo con nuestro ejemplo, podemos observar que el 57% de las personas que viven en casa está en la clase 1. Si en el caso anterior de la descripción el parámetro de la comparación era el porcentaje de la modalidad en la distribución global, en este caso lo es el porcentaje de la clase dentro del global. Que el 57% de las personas que viven en casa está en la clase A resulta bastante revelador, tomando en cuenta que esta clase concentra el 42% de la población. Pero resulta más revelador en el caso de la clase 3 donde está el 74% de las personas que tienen el servicio de agua fuera de la vivienda, siendo que esta clase sólo concentra el 8% del global.

De esta manera, las seis clases obtenidas son caracterizadas, como puede observarse en las tablas siguientes que responden al mismo formato usado para el anexo B, ordenando de arriba hacia abajo las modalidades que mejor caracterizan a las clases. La columna *CLA/MOD* corresponde al porcentaje que tiene esa clase dentro de la distribución global de la modalidad –la discriminación– y la columna *MOD/CLA* corresponde a la distribución porcentual de la modalidad dentro de la clase: la descripción. La combinación de ambos datos ayuda a conformar la caracterización de la clase. La columna global indica cómo se distribuye esa modalidad dentro de la población global; es un dato que se repite para cada clase. El análisis no ordena o jerarquiza las clases –eso es trabajo del investigador, si es de

su interés—. En este caso, las ordenamos para hacer más clara la explicación.

De las seis clases, podemos decir en resumen que las tres primeras corresponden a niveles o situaciones diferentes de pobreza, la cuarta corresponde a una especie de clase popular o media-baja y las dos últimas a las clases acomodadas.

La primera clase podemos designarla como la población en situación de miseria. La primera modalidad que la caracteriza es el bajo nivel de ingreso, el cual incluye la ausencia de éste. Concentra casi el 60% de la población analfabeta y sin nivel educativo, siendo que esta clase sólo concentra el 14% del total. Es probable que el hecho de que la modalidad “casa” aparezca caracterizando la clase, pueda deberse al hecho de que casi la mitad de esta clase viva en zonas rurales o ciudades pequeñas, donde existen las casas rurales. Esta población en situación de miseria puede ser encontrada tanto en las zonas de los centros poblados pequeños como en las grandes ciudades.

Distribución porcentual de las modalidades para la clase 1 (13,7%)

DISTRIBUCIONES PORCENTUALES			MODALIDAD DE LA VARIABLE	VARIABLE
CLASE / MODALIDAD	MODALIDAD / CLASE	TOTAL		
60,0	63,3	14,5	ESTRATO G	INGRESO PER CÁPITA
62,3	41,7	9,2	SIN NIVEL	NIVEL EDUCATIVO
62,8	16,6	3,6	ANALFABETO	NIVEL EDUCATIVO
17,8	85,2	65,5	CASA	TIPO DE VIVIENDA
21,2	20,1	13,0	CIUDADES PEQUEÑAS	TAMAÑO DEL CENTRO POBLADO
14,7	93,8	87,8	EN LA VIVIENDA	AGUA
16,7	31,3	25,7	RESTO	TAMAÑO DEL CENTRO POBLADO
16,1	29,1	24,8	GRANDES CIUDADES	TAMAÑO DEL CENTRO POBLADO
15,6	19,8	17,5	NO LLEGA A LA VIVIENDA	ASEO
14,0	75,7	74,1	LLEGA A LA VIVIENDA	ASEO

La segunda clase la hemos identificado como la población en situación de pobreza extrema. Dos tercios de este grupo viven en ranchos y carecen de servicio de agua y de aseo. Se caracterizan principalmente por las condiciones de la vivienda pero también por el nivel educativo. Mientras la clase anterior se identificaba con la población analfabeta y sin nivel, ésta se identifica con la población sin nivel y con primaria aprobada. Es, junto a la clase siguiente, de tamaño reducido (un 8%).

Distribución porcentual de las modalidades para la clase 2 (8,4%)

DISTRIBUCIONES PORCENTUALES			MODALIDAD DE LA VARIABLE	VARIABLE
CLASE / MODALIDAD	MODALIDAD / CLASE	TOTAL		
73,4	69,8	8,0	FUERA DE LA VIVIENDA	AGUA
48,0	66,7	11,7	RANCHO	TIPO DE VIVIENDA
31,3	65,2	17,5	NO LLEGA A LA VIVIENDA	ASEO
14,6	44,5	25,7	RESTO	TAMAÑO DEL CENTRO POBLADO
11,4	59,7	44,2	PRIMARIA	NIVEL EDUCATIVO
16,3	17,8	9,2	SIN NIVEL	NIVEL EDUCATIVO
14,2	23,6	14,0	ESTRATO E	INGRESO PER CÁPITA
12,9	21,3	13,8	ESTRATO F	INGRESO PER CÁPITA
10,4	17,7	14,3	ESTRATO D	INGRESO PER CÁPITA
10,8	4,7	3,6	ANALFABETO	NIVEL EDUCATIVO

La tercera clase, la de pobreza rural, se caracteriza en mayor medida por la presencia de población radicada en los centros poblados pequeños. Presenta también condiciones más precarias que las otras dos clases, carece de servicios de agua y de aseo. Comprende sólo un 9% de la población.

Distribución porcentual de las modalidades para la clase 3 (9%)

DISTRIBUCIONES PORCENTUALES			MODALIDAD DE LA VARIABLE	VARIABLE
CLASE / MODALIDAD	MODALIDAD / CLASE	TOTAL		
81,9	76,7	8,4	NO TIENE	ASEO
81,2	38,3	4,2	NO TIENE	AGUA
25,5	72,7	25,7	RESTO	TAMAÑO DEL CENTRO POBLADO
22,0	33,7	13,8	ESTRATO F	INGRESO PER CÁPITA
10,9	79,0	65,5	CASA	TIPO DE VIVIENDA
11,9	58,6	44,2	PRIMARIA	NIVEL EDUCATIVO
22,9	9,3	3,6	ANALFABETO	NIVEL EDUCATIVO
14,0	21,8	14,0	ESTRATO E	INGRESO PER CÁPITA
14,5	18,8	11,7	RANCHO	TIPO DE VIVIENDA
12,9	13,1	9,2	SIN NIVEL	NIVEL EDUCATIVO
9,1	14,6	14,5	ESTRATO G	INGRESO PER CÁPITA

La cuarta clase corresponde a un segmento que podemos denominar la clase popular o la clase media-baja. En su mayoría vive en casas con servicio de agua y de aseo, tiene niveles de educación básica con ingresos bajos y vive en las ciudades medias. Es la clase más grande; comprende casi un 42% de la población.

Distribución porcentual de las modalidades para la clase 4 (41.6%)

DISTRIBUCIONES PORCENTUALES			MODALIDAD DE LA VARIABLE	VARIABLE
CLASE / MODALIDAD	MODALIDAD / CLASE	TOTAL		
57,4	90,4	65,5	CASA	TIPO DE VIVIENDA
46,6	98,3	87,8	EN LA VIVIENDA	AGUA
57,8	61,4	44,2	PRIMARIA	NIVEL EDUCATIVO
61,6	24,6	16,6	CIUDADES INTERMEDIAS	TAMAÑO DEL CENTRO POBLADO
46,7	83,3	74,1	LLEGA A LA VIVIENDA	ASEO
63,3	21,7	14,3	ESTRATO D	INGRESO PER CÁPITA
61,8	21,0	14,1	ESTRATO C	INGRESO PER CÁPITA
60,6	18,9	13,0	CIUDADES PEQUEÑAS	TAMAÑO DEL CENTRO POBLADO
56,7	19,1	14,0	ESTRATO E	INGRESO PER CÁPITA
47,7	33,1	28,8	SECUNDARIA	NIVEL EDUCATIVO
47,8	15,9	13,8	ESTRATO F	INGRESO PER CÁPITA
47,3	16,5	14,5	ESTRATO B	INGRESO PER CÁPITA

La quinta clase es una de las que pudiéramos llamar clase acomodada. Tiene a un 16% de la población total. Concentra casi el 90% de quienes viven en casa-quinta. En su mayoría son personas de las grandes ciudades –pero no de Caracas– con niveles altos de ingreso y con niveles técnicos o universitarios de educación.

Distribución porcentual de las modalidades para la clase 5 (15.9%)

DISTRIBUCIONES PORCENTUALES			MODALIDAD DE LA VARIABLE	VARIABLE
CLASE / MODALIDAD	MODALIDAD / CLASE	TOTAL		
86,7	54,2	9,9	CASA QUINTA	TIPO DE VIVIENDA
37,4	58,3	24,8	GRANDES CIUDADES	TAMAÑO DEL CENTRO POBLADO
20,8	96,7	74,1	LLEGA A LA VIVIENDA	ASEO
38,0	35,3	14,8	ESTRATO A	INGRESO PER CÁPITA
50,2	21,4	6,8	TÉCNICO	NIVEL EDUCATIVO
43,8	20,3	7,4	UNIVERSITARIO	NIVEL EDUCATIVO
29,1	26,6	14,5	ESTRATO B	INGRESO PER CÁPITA
17,6	96,8	87,8	EN LA VIVIENDA	AGUA
22,2	40,2	28,8	SECUNDARIA	NIVEL EDUCATIVO
20,2	21,1	16,6	CIUDADES INTERMEDIAS	TAMAÑO DEL CENTRO POBLADO

Finalmente, la última clase corresponde a los profesionales de Caracas. Es un 11% de la población. Se encuentran en ella un 74% de quienes viven en apartamento, y concentra casi a la mitad de las personas que viven en Caracas. Poco menos de un tercio de este grupo es de nivel universitario. Junto a la clase anterior, presenta niveles altos de ingreso.

Distribución porcentual de las modalidades para la clase 6 (11.4%)

DISTRIBUCIONES PORCENTUALES			MODALIDAD DE LA VARIABLE	VARIABLE
CLASE / MODALIDAD	MODALIDAD / CLASE	TOTAL		
74,2	83,8	12,9	APARTAMENTO	TIPO DE VIVIENDA
46,8	81,9	19,9	GRAN CARACAS	TAMAÑO DEL CENTRO POBLADO
45,6	59,1	14,8	ESTRATRO A	INGRESO PER CÁPITA
48,1	31,2	7,4	UNIVERSITARIO	NIVEL EDUCATIVO
12,9	99,6	87,8	EN LA VIVIENDA	AGUA
12,6	81,9	74,1	LLEGA A LA VIVIENDA	ASEO
19,6	11,6	6,8	TÉCNICO	NIVEL EDUCATIVO
14,5	36,6	28,8	SECUNDARIA	NIVEL EDUCATIVO
11,6	17,8	17,5	NO LLEGA A LA VIVIENDA	ASEO

Un comentario final. Estas clases son tipologías y no clasificaciones exhaustivas. Por ejemplo, el hecho de que no haya salido una clase de pobreza extrema en Caracas no significa que ésta no exista o carezca de significación; este grupo se encontrará diseminado en las primeras tres o cuatro clases. Si incluyéramos otras variables, como tipo y rama de actividad económica, es probable que apareciera una clase caracterizada, no sólo por la pobreza, sino también por la actividad informal en Caracas.