

Estratificación Socioeconómica: Aproximación Conceptual y Diseño Metodológico*

*Dario Cuervo Villafañe**
Manuel Ramírez Gómez****

I. Introducción

Hace aproximadamente tres décadas que en Colombia viene siendo utilizada la estratificación socioeconómica como instrumento para adelantar diferentes tareas, tanto con carácter investigativo como para establecer medidas de política. Sin embargo hasta el momento muy poco se ha conocido acerca del desarrollo conceptual que la soporta, y mucho menos existe un consenso sobre el o los métodos más adecuados para abordarla.

La marcada importancia que para el país ha venido tomando esta temática, hace necesario explorar un poco las bases teóricas sobre las cuales se puede apoyar y, sobre todo, tratar de establecer a la luz de ese marco de referencia una posible metodología que permita acercarse a mejores sistemas de estimación, que logren aceptables resultados en términos de adecuadas clasificaciones.

* Las reflexiones de este artículo tuvieron como base el estudio sobre el diseño y aplicación metodológica elaborado por el CEDE para las Empresas Públicas municipales de Cali. Fueron múltiples los aportes y colaboraciones de más de un centenar de personas que de una u otra forma participaron en él, resaltando sobre todo a los doctores Gilberto Aristizabal, Alvaro Montenegro, Ligia Melo, Ricardo Gómez, Marisol Suárez y a la firma Napoleón Franco y Cia. encargada de tomar la información de campo.

** Economista, Universidad de San Buenaventura. Magister en Economía, Universidad de los Andes. Profesor-Investigador, Facultad de Economía, Universidad de los Andes.

*** Ingeniero Civil, Universidad Javeriana. Magister y Ph. D. en Economía, Yale University. Profesor-Investigador, Facultad de Economía, Universidad de los Andes.

Si bien es cierto pueden existir diferentes formas de realizar estratificaciones de esta índole, el método científico enseña que dada una definición y delimitación del problema que se desea analizar, así como el establecimiento preciso de los supuestos sobre los cuales se debe trabajar, pocas pueden ser las alternativas metodológicas sobre las cuales se pueden desarrollar las temáticas, de tal forma que los resultados obtenidos se constituyan en un producto lógico que corresponda adecuadamente a esos elementos planteados.

Sobre esta base surgen entonces los propósitos de este trabajo, donde se pretende plantear algún esquema metodológico que permita adelantar un proceso aceptable de estratificación socioeconómica que responda a lo que se podría entender como tal, permitiendo simultáneamente determinar los mejores criterios para lograr las discriminaciones adecuadas.

La generalidad esperada en esta clase de metodologías se constituye en un condicionamiento apenas obvio, donde de alcanzarse el objetivo de su diseño, debe estar en capacidad de ser implementada y reproducida en cualquier escenario, siempre y cuando se presenten las condiciones propias de su planteamiento.

Sin embargo, la necesidad de realizar alguna prueba empírica que valide el diseño que se logre construir, hace que para este caso se hayan tomado los datos de la ciudad de Cali, para con ellos adelantar la evidencia requerida, y constatar así su funcionalidad y operatividad.

II. Aproximación teórica

Fue en la década de los cincuenta cuando la escuela *estructural funcionalista* decidió, bajo una visión sociológica, adentrarse en la temática de la estratificación socioeconómica, y sobre la base de sus postulados desarrollar un marco teórico y conceptual que permitiera la exploración y análisis de las comunidades dentro del rigor exigido por el método científico.

Para ellos era impreciso hablar de este tema sin haber definido previamente los elementos de relación existentes, ya que en muchos aspectos la estratificación desde el punto de vista social podría estar en contravía con la que pudiera resultar desde una óptica económica, caso en el que, por ejemplo, las categorías podían establecerse según el monto de la riqueza, lo que no necesariamente guardaba una relación con la clasificación social, pues quizás era necesario valorar su origen y el tiempo de adquisición.

Lo anterior se podría explicar, debido a que en algunas sociedades se le otorga mayor importancia a la riqueza heredada que a la recién adquirida, permitiendo con esto que las familias tradicionales sean valoradas positivamente, frente al prestigio social negativo que darían a aquellas cuyas riquezas fueran recientes o de origen dudoso.

Por este y otros muchos motivos, la estratificación debe entonces tomarse como una estructura de desigualdades regularizadas, en la que los individuos son situados más arriba o más abajo de acuerdo con un sistema de valoración que puede presuponer

privilegios y poderes, susceptibles de ser representados por medio de mejores ingresos, mejores posiciones de trabajo, de vivienda, mayor respeto, admiración, etc.

Obviamente, la metodología y los criterios de valoración que se usen para adelantar cualquier proceso de estratificación, deberán mantener una relación muy directa con la manera como se conceptualice la idea de lo que puede significar una estratificación socioeconómica.

Partiendo entonces de lo que la sociología entiende que podría considerarse como estratificación socioeconómica, se define como "la forma en que una comunidad determinada, en un momento dado, reconoce que se encuentra dividida, así como los criterios que en su concepto son los que permiten ubicar los individuos en los estratos que ella misma reconoce".

Dicho en otras palabras, se puede entender que los individuos mantienen unas condiciones o características que los *identifican* como pertenecientes a una clase (estrato) determinada, por lo que es su estilo de vida, y la forma como lo reflejan, lo que permite que se les reconozca como perteneciente a ella.

De acuerdo con esto, la tarea del investigador debe estar circunscrita a tratar de "identificar" los estratos en que se encuentra dividida la comunidad que se observa, así como también a reconocer los criterios mediante los cuales la sociedad permite ubicar los individuos en dichos estratos. Así, la metodología apuntará hacia la definición de indicadores de pertenencia a los estratos, que faciliten el reconocimiento de las diferentes manifestaciones de la conducta de los miembros de esa comunidad.

Tomando en consideración todo lo anteriormente expuesto, se tiene que el proceso de estratificar socioeconómicamente una comunidad en particular, estará orientado a tratar de obtener de ella una radiografía de las diferentes clases en que se encuentra dividida, en un momento del tiempo, y de los elementos que igualmente ella considera como los más adecuados para caracterizar la diferenciación entre ellos.

Bajo estas circunstancias la tarea de estratificar una agrupación social queda circunscrita a la realización de varias exploraciones, donde sobre la base informativa que se obtenga de la comunidad, se adelanten los ejercicios necesarios que permitan aproximarse a los criterios que reflejen más adecuadamente los diferentes perfiles en que se encuentra dividida.

Una vez establecida la definición sobre la cual desarrollar toda la exploración, surgen entonces una serie de dificultades provenientes del sistema de valoración que se utilice. Efectivamente, la decisión acerca del método a ser usado conlleva la necesidad de enfrentar, entre otras, las siguientes limitaciones:

- La primera, y quizás la más importante, se conoce como el efecto de la *subjetividad*. Ateniéndose a la forma como se definió el concepto de estratificación, cualquier elemento que de manera exógena sea determinado, ya sea por parte del investigador, individuo o entidad diferente a la información suministrada por la comunidad a través del trabajo de campo, implica la violación del principio de identificación.

Si bien este elemento puede aparecer durante cualquiera de las etapas del estudio, y a la luz de la experiencia epistemológica resulta inevitable su existencia, se debe por todos los medios posibles tratar de minimizar su influencia, de tal manera que con ello pueda garantizarse un poco más la pureza de los resultados que otorga la comunidad que se analiza.

Esta búsqueda por minimizar su influencia tiene mucho que ver con acciones que quizás no resultan muy concientes para aquel que desarrolla el trabajo, ya que por ejemplo la sola escogencia de la unidad de análisis para la recolección de la información, o la decisión previa acerca del número de estratos en los cuales se va a dividir la comunidad, así como otros muchos actos del proceso de estudio, están inmediatamente involucrando aspectos que en ningún momento han sido producto de lo que se está obteniendo del terreno, ocasionando con ello sesgos en las estimaciones que alejan los resultados obtenidos de la realidad imperante.

- Otra de las limitaciones propias en esta clase de trabajos tiene que ver con la *escogencia de los criterios* para estratificar, y por consiguiente de la *importancia relativa* de cada uno de ellos como elementos discriminadores.

Como bien se sabe, una de las formas de adelantar estudios sobre evidencia empírica, es aquella que toma una metodología o una serie de criterios previos, para con ellos tratar de estimar lo que sucede en el universo que se analiza.

Sin embargo cuando se parte del principio, como sucede en este caso, que las comunidades establecen en su interior formas o mecanismos mediante los cuales los individuos se reconocen como pertenecientes a un estrato, y que la tarea consiste en tratar de identificar esas condiciones o características, mal podría entonces *a priori* el investigador, es decir sin tomar en consideración lo que dice esa comunidad, tomarse la atribución de decidir cuáles serán esas características, y menos aún, establecer unilateralmente la importancia que cada una de ellas tiene como elemento diferenciador de cada uno de esos estratos.

La decisión de establecer estos cursos de acción está evidenciando además que el investigador le está imprimiendo un elemento subjetivo al análisis, que no es producto de lo que la información le provee, sino de su intuición, visión, esperanza, o cualquier otra razón personal.

Por estos motivos, se hace necesario que sólo mediante el análisis de las observaciones se pueda llegar a la determinación de cuáles deben ser los elementos que mejor caracterizan a cada uno de los estratos de esa agrupación social, así como también a establecer las ponderaciones que cada uno de ellos tiene en cada estrato.

Cumplir el principio de la no escogencia previa de criterios (variables) y ponderaciones, implica necesariamente que sólo hasta el final del estudio ellos serán claramente identificados.

- Aunque bajo una óptica de investigación básica no sería necesario tomar en consideración situaciones de carácter legal, la estratificación socioeconómica como instru-

mento base para aplicar medidas de política, especialmente en lo que concierne al establecimiento de tarifas de servicios públicos residenciales, obligó al Estado a tener que establecer ciertos parámetros sobre los cuales se deberían adelantar todos estos análisis.

Entre los condicionamientos más importantes aparece aquel que establece que toda comunidad que se analice estará compuesta, como máximo, por seis estratos socioeconómicos. De igual manera, ha determinado cuáles pueden ser los posibles criterios o variables con los cuales se debe adelantar el proceso de estratificación, donde en forma adicional expresa que el DANE se encargará de establecer las ponderaciones que esas variables deberán tener.

La pregunta que surge tiene que ver con el deseo de conocer cuáles fueron los argumentos que las autoridades tuvieron para llegar al establecimiento de esa normatividad. ¿Están ellos seguros de que las comunidades establecidas dentro del territorio Colombiano están divididas como máximo en seis grupos? ¿No podría darse el caso que una agrupación de esas, por sus condiciones y relaciones pudiera estar conformada por siete, nueve, o cualquier otro número de estratos? ¿Poseen ellos la evidencia acerca de las características socioeconómicas mediante las cuales están conformados *todos* los municipios que forman parte de esta nación?

De igual manera, cómo es posible establecer, sin observar qué dice la comunidad, los pesos relativos que esas variables con las que se adelantará el proceso de estratificación tendrán, de tal forma que se pueda llegar al establecimiento de las diferenciaciones entre cada uno de esos estratos?

Parece ser que son las condiciones de carácter operativo, más que las del principio conceptual, las que inspiraron estas reglamentaciones, buscando con ellas dar una orientación para que cada una de las municipalidades pudiera adelantar su proceso dentro de un marco de referencia más o menos similar, de tal forma que se pudieran aplicar medidas de política de carácter general.

Sin entrar en la consideración acerca de la conveniencia de esta reglamentación, lo cierto es que para efectos del proceso de identificación de las condiciones sobre las cuales se ha planteado este trabajo, la *legislación* se constituye en una limitante más del estudio, que contiene altos ingredientes de subjetividad.

- Finalmente, existe otro inconveniente que surge de la decisión acerca de cuáles deben ser los criterios que se deben buscar en la agrupación social, mediante el trabajo de campo, para con ellos adelantar todo el proceso de clasificación.

Si no se desea violar el principio de la identificación, e imprimir subjetividad al escoger esos criterios, se debe tratar de buscar la mayor cantidad posible de variables, de tal forma que con ellas se pueda abarcar todo el universo.

Nuevamente, bajo una visión pura de la investigación, se puede intuir que deberían contemplarse una gran cantidad de ellos, situación que resulta muy poco práctica y onerosa de lograr. Además, ateniéndose estrictamente a la definición establecida, sería

necesario tomar en consideración aquellos que tienen en su interior fuertes componentes de valoración subjetiva, tales como el prestigio, el poder, el abolengo, etc., que implican dificultades al momento de desarrollar las estimaciones pertinentes.

Como es imposible evitar la subjetividad en este aspecto, se debe intentar tomar una buena gama de variables que permitan abarcar la mayor cantidad posible de características relevantes de esa comunidad, de tal manera que se logre con ellas identificar las más comunes a cada estrato, siendo simultáneamente capaces de diferenciar a todos esos estratos.

De la misma manera, la necesidad de que esos criterios estén exentos de valoraciones subjetivas, exige que aquellas variables consideradas como cualitativas tengan que ser obviadas de este proceso de estimación. Adicionalmente, las variables cuantitativas deben ser homogéneas en el tiempo de modo que permitan adelantar comparaciones entre los diferentes individuos, así como a lo largo del tiempo.

III. Método de trabajo

Esta sección presenta los aspectos fundamentales que fueron tenidos en cuenta en el proceso de estimación, es decir los supuestos con los cuales se trabajó, el contenido del formulario de encuesta, y la forma como se le dio el tratamiento al diseño de la muestra.

A. Supuestos fundamentales

1. Teniendo en cuenta la conceptualización de estratificación económica adoptada para este trabajo, se desprende como necesidad sustancial la condición de *identificar* las características que podrían existir al interior de una agrupación social, que hacía que ellos se reconocieran como pertenecientes a un estrato dado, y al mismo tiempo diferentes del de los demás.

El establecimiento de esta noción sociológica de identificación hace que se rompa en forma tajante con los métodos comúnmente utilizados, al menos en Colombia desde la década de los sesenta, para adelantar procesos de estratificación socioeconómica, ya que su principio fundamental se encontraba enmarcado en la necesidad de establecer cuáles eran los criterios que en forma más adecuada permitirían al investigador realizar la labor de *asignar* el estrato.

Así, bajo esta concepción se pueden adelantar tantas estratificaciones como propósitos se tengan, y de acuerdo con esto, la tarea del investigador estará enmarcada en la búsqueda de los elementos que le permitieran asignar los individuos a determinado estrato.

Si bien puede suceder que esta forma de abordar el problema tenga toda la consistencia y solidez necesaria, a la luz de lo que se conoce como estratificación socioeconómica, parece poco probable que haciendo uso de estos métodos se logren resultados que puedan ser considerados como aceptables.

Este supuesto obliga entonces a que sea la comunidad, por medio de la información que se tome en el terreno, la que suministre todos los elementos sustanciales y relevantes para el análisis.

2. Partir del supuesto de la identificación, implica la necesidad de mantener la mayor *objetividad* posible frente a cada uno de los aspectos que se abordan, ya que como ampliamente se ha demostrado en la sección anterior, los elementos de subjetividad que no puedan ser evitados, se constituyen en aspectos nocivos que atentan en forma directa contra la capacidad de explicación de la estimación.

Sin embargo, la imposibilidad de lograr una situación óptima no es óbice para que se intente minimizar las acciones y los efectos que pueden lograr los elementos subjetivos durante todo el proceso.

3. La acción conjunta de los supuestos de identificación y objetividad, hace que surga el tercero de ellos, que tiene que ver con el intento de evitar el establecimiento previo de los criterios que puedan ser considerados como relevantes para lograr el reconocimiento de los estratos, así como las ponderaciones que ellos deben tener en el proceso de discriminación.

Efectivamente, si se parte del principio del desconocimiento acerca de las características que hacen que los miembros de una agrupación social se reconozcan como pertenecientes a un estrato, mal podría entonces, establecerse aquellas condiciones que permiten la diferenciación entre cada una de ellas, sin verificar qué es lo que dice esa agrupación al respecto, y mucho más la importancia relativa que cada una tiene en cada uno de esos estratos.

Esta situación necesariamente implica que al adelantar procesos de estratificación socioeconómica, a diferencia de una buena mayoría de estudios, no se podrán utilizar variables ni modelos previamente establecidos, pues ellos se constituirán en productos del análisis y no en insumos, como comúnmente sucede.

4. La legislación vigente impide que los procesos de estratificación que se adelanten tengan la posibilidad de arrojar la existencia de más de seis estratos, motivo por el cual esta restricción, completamente exógena a las relaciones sociales, exige que se establezca como supuesto básico la determinación de lograr clasificaciones en ese número de estratos como máximo.

5. De igual manera, la misma norma legal establece que como unidad de análisis se podrá tomar como máximo a la manzana, pudiendo ser la cuadra-lado, o inclusive la vivienda individual. Cómo entrar a tomar información de cada una de las viviendas existentes dentro de una agrupación urbana o rural puede ser una tarea bastante dispendiosa y onerosa, y partiendo del supuesto (subjetivo) que puede existir cierto grado de homogeneidad entre las viviendas que conforman la manzana, se decidió utilizar entonces la manzana como unidad básica de análisis, debiendo para ello escoger una vivienda representativa de esa manzana para adelantar en el hogar(es) que la habitaba, la toma de información requerida.

6. La imposibilidad de tomar en consideración todos los criterios que la agrupación puede mostrar como reconocedores de los estratos, implica establecer cuál o cuáles de ellos pueden acercarse a una adecuada explicación de esas condiciones diferenciadoras.

Puede pensarse que existe una buena relación directa entre el ingreso que perciben las personas y el estrato al que ellas pertenecen, es decir, que parece ser que en la medida en que los individuos tienen la posibilidad de obtener mayores ingresos, la sociedad los reconoce, y ellos mismos lo reflejan, como pertenecientes a un estrato más alto, donde no sólo tienen la posibilidad de lograr mejores condiciones de vida material, sino además de lograr valoraciones sociales más acordes con su condición.

Sin embargo, es ampliamente reconocida la dificultad que existe al tratar de establecer estimaciones que se acerquen a una medida adecuada de los niveles de ingreso y riqueza que poseen las personas. Existe suficiente evidencia empírica que demuestra la tendencia por parte de los que tienen mayores ingresos a subvalorar lo que perciben, dándose el caso contrario por parte de aquellos cuyos valores son menores.

Ante esto surge la necesidad de buscar una medida que permita aproximarse a la determinación del nivel de ingreso que poseen las personas. Parece ser que una adecuada *proxi* puede lograrse cuando se observan algunos otros elementos, donde el reflejo y reconocimiento se percibe por las cosas o bienes que ellos muestran ser poseedores. Así, percibir ingresos más altos les permite tener acceso a una mayor y mejor cantidad de bienes y servicios, que aquellos cuyos ingresos no son suficientes para lograr adquirirlos.

La libre autodeterminación que permite el ejercicio de la *democracia*, otorga la posibilidad de que las personas en forma autónoma decidan los bienes que desean adquirir así como el sitio donde quieren vivir, aspectos que bajo un sistema de economía capitalista permite que sean los precios los que en principio se encarguen de asignar los recursos.

Si adicionalmente el libre albedrío, junto con el instinto gregario que induce a las personas a que intenten agruparse entre semejantes, están presentes en la toma de decisiones frente a situaciones como la escogencia de los sitios de vivienda, se puede llegar a suponer que la conformación espacial de las agrupaciones urbanas, también guarda relación directa con los ingresos y los estratos.

Por lo tanto, estas viviendas y su entorno, junto con todos aquellos bienes y servicios que las personas hacen uso por la posibilidad de ser adquiridos dependiendo del ingreso que perciben, pueden tomarse como criterios que permiten, por la vía del ingreso, acercarse al establecimiento de las características que logran identificar la pertenencia a un estrato y no a otro.

De igual forma, la configuración física de estos elementos, permite además que sean fácilmente reconocidos, identificados, y además con capacidad de homogeneidad e inmutabilidad en el tiempo.

Por lo tanto, se toma entonces como supuesto, que las condiciones en que viven las personas, representadas por las viviendas que habitan y su entorno, así como algunos bienes y servicios a que tienen acceso, pueden ser considerados como una aceptable *proxi* del ingreso, que a su vez se constituye en una adecuada forma que permite la identificación de pertenencia al estrato.

B. Formulario de encuesta

El principio de identificación exige que se tomen del terreno las condiciones que, según la comunidad, son las que ayudan a caracterizar y determinar la posibilidad de pertenencia a cada uno de los estratos.

Sobre esta base, y siendo completamente estrictos, podría darse el caso en que el número total de criterios que reflejaran los conceptos de la comunidad llegaran a ser de una magnitud bastante considerable. Por lo tanto, ante las limitaciones de carácter logístico y temporal, se hizo necesario que sólo se escogieran (en un ejercicio subjetivo) cuáles podrían ser los criterios más relevantes, que además cumplieran con los requisitos de homogeneidad e inmutabilidad.

Una vez realizado un detallado examen acerca de aquellos que podrían ser los más relevantes, se llegó a la determinación de tomar 55 criterios, pudiendo diseñar con ellos las diferentes preguntas para el formulario de encuesta de acuerdo a una agrupación en términos de sus posibles afinidades.

El cuestionario se dividió en ocho partes, conformadas de la siguiente manera:

a. Identificación

Dónde se establecieron todos los aspectos relativos a la ubicación de la manzana, vivienda, etc.

b. Datos de control

Para establecer algunas formas de chequeo del encuestador.

c. Datos del entorno

Con el propósito de conocer las condiciones que rodeaban las viviendas escogidas para el diligenciamiento del formulario.

d. Datos de la vivienda

Para conocer las condiciones externas e internas físicas de las viviendas, considerando además aspectos tales como las áreas del lote y de construcción, la disponibilidad de servicios públicos y los últimos tres datos sobre el consumo de los mismos.

e. Datos del hogar

Se buscó conocer aspectos de los miembros del hogar, su formación, ocupación, relación con el jefe del hogar, hacinamiento, y disponibilidad de electrodomésticos.

f. Datos sobre gastos del hogar

La idea fundamental estaba orientada a que esta sección sirviera como elemento de control a las preguntas sobre los ingresos, que eran más detalladas. Sin embargo, se tomaron datos sobre los gastos en arrendamiento, cuotas de corporación, hipotecas, etc, así como lo destinado mensualmente a educación, y lo pagado por concepto de impuestos municipales.

g. Datos de ingreso

El diligenciamiento de esta sección sólo correspondería a aquellos miembros del hogar que tuvieran una actividad laboral, para lo cual las preguntas fueron bastante similares a las realizadas por el DANE en su encuesta de hogares.

h. Preguntas para el encuestador

Esta sección estaba orientada a conocer el grado de confiabilidad que se podría dar a la información recogida por el encuestador, y a la expresada por el encuestado.

Como se puede apreciar, si bien era poco probable alcanzar a cubrir todas y cada una de las opciones posibles, la escogencia de estos 55 criterios sí parecía recoger la mayor cantidad de los generalizados, de forma tal que se pudiera adelantar la tarea de caracterización y discriminación.

C. Diseño de la muestra

Según el supuesto que se estableció acerca de la limitante legal de tomar la manzana como la unidad de análisis máxima, se decidió adelantar el diseño de la muestra haciendo uso de dos elementos diferentes.

Para el primero de ellos se tomó la totalidad de las manzanas residenciales de la ciudad y tomando éstas como el universo, se decidió tomar una submuestra de aproximadamente el 33% de ese total, haciendo uso del método sistemático, de forma tal que no se quedara ninguna parte de la ciudad por fuera de esta submuestra, esto con el fin de realizar inicialmente los análisis con ella, para luego ser comparados con las estimaciones realizadas con el total de la muestra.

En el segundo momento, considerado como el más relevante, se diseñó una muestra que tomara una vivienda por cada una de las manzanas residenciales de la ciudad. Para conservar el principio de la aleatoriedad necesaria en esta clase de trabajos, se escogieron esas viviendas mediante un proceso estocástico, realizado directamente por el computador.

Sin embargo podría presentarse un problema, ya que en la mayoría de las organizaciones urbanísticas se da la posibilidad de que al interior de las manzanas puedan existir viviendas que se encuentren por fuera del promedio de las demás. Cuando esta situación se presenta, se considera que la manzana en vez ser homogénea, es heterogénea. Dicho en otras palabras, es normal esperar que todas las viviendas que conforman una manzana mantengan una gran similitud en sus características externas, por lo que a esa manzana se la puede considerar como *homogénea*.

Sin embargo, en ocasiones se da el caso que dentro de una de esas manzanas se encuentra una diversidad de condiciones externas de las viviendas que permiten pensar que sus moradores no poseen las mismas características del resto de los miembros de la manzana. Por lo tanto, a esta clase de unidades de análisis se las considera como *heterogéneas*.

La razón de la diferenciación entre estas dos clases de manzanas se apoya tanto en la vivienda a la cual tomarle la información, como en el número de encuestas a diligenciar por manzana.

Cuando se encontró que la manzana era homogénea, la vivienda escogida por el método aleatorio podía ser considerada como *típica* de la manzana, por lo que se procedía a diligenciar el formulario con la certeza de que la información obtenida en ella representaba el promedio del resto de la manzana. En cambio, cuando se consideraba que la manzana tenía las características para ser tomada como homogénea (subjetivo), era necesario en primer lugar determinar la clase de *atipicidad* que contenía, ya que podría darse el caso de que las viviendas atípicas estuvieran por debajo del promedio, o por encima del mismo. Cuando existía sólo una clase de atipicidad, además del diligenciamiento a la vivienda considerada como típica de la mayoría de la manzana, era necesario escoger una vivienda atípica y diligenciar también en ella la encuesta. Si se daba el caso de existir las dos clases de atipicidad, se adelantaban entonces dos encuestas adicionales.

Con la realización de esta clase de actividades, la muestra que se tomaba como representativa de las condiciones socioeconómicas de la ciudad recogía dentro de cada una de las unidades de análisis todas las características que al interior de ella se encontrarán.

D. Técnicas de estimación

Dado el supuesto de identificación, que permite el reconocimiento de las características comunes de cada uno de los estratos establecidos, era necesario encontrar la importancia relativa de cada una de esas características para cada estrato. Para ello se decidió utilizar tres técnicas de estimación.

La primera de ellas, conocida como análisis cluster, permite mediante un proceso iterativo y automático agrupar (para este caso en seis categorías) de acuerdo con los atributos que encuentre como comunes entre las observaciones (encuestas).

Como el análisis cluster sólo agrupa, y es necesario establecer la ponderación de cada variable dentro de cada estrato, se optó por usar el análisis multinomial discriminante, que permite, además de determinar las ecuaciones representativas para cada estrato, comparar la estratificación estimada con esta técnica con la realizada por el cluster. Así, es posible confirmar por dos métodos diferentes la ubicación de cada observación en cada estrato.

Como dentro de los supuestos establecidos se consideró la relación directa existente entre el ingreso y el estrato, una vez realizadas las estimaciones anteriores, se procedió a contrastar la estratificación obtenida con los ingresos que reportaron las familias encuestadas, con el fin de verificar la validez del supuesto. Para ello se hizo uso de una técnica conocida como Logit multinomial, que en este caso específico permite establecer la probabilidad de pertenencia a un estrato, dado el nivel de ingreso que reportó la familia encuestada.

1. Análisis cluster

La prueba con cluster permite por medio de un proceso iterativo crear una variable que clasifica la muestra en un número de grupos (cluster) especificado previamente, teniendo en cuenta un conjunto de características comunes a cada grupo. Dentro de las funciones principales del cluster se pueden encontrar las de nombrar, mostrar, resumir, predecir y requerir explicación. Además, a todos los objetos en el mismo cluster les es dado el mismo nombre.

Un aspecto importante en esta técnica es que los objetos de la observación están expuestos con el propósito de que las diferencias sutiles puedan llegar a ser más evidentes, ya que se encuentran físicamente muy cercanos dentro del mismo cluster. Esta información está resumida más en términos de las propiedades de los cluster, que en las propiedades individuales de los objetos, de tal forma que si algunos de los elementos de un cluster tienen cierta propiedad, sería de esperar que los demás tuvieran esa misma propiedad.

Esta técnica se ha empleado en diferentes disciplinas. Inicialmente se utilizó en astronomía con el fin de clasificar los planetas menores que se anunciaba eran descubiertos; también en taxonomía para la clasificación de animales y plantas de acuerdo con sus características comunes. En medicina ha sido usado fundamentalmente para la clasificación de enfermedades, y en arqueología y antropología para la clasificación de diferentes hallazgos. En economía, en la consideración de matrices insumo-producto, donde las filas y las columnas tienen los mismos nombres, los cluster de las filas y las columnas deben ocurrir simultáneamente. En investigación de mercados, su empleo ha estado centrado en la clasificación de industrias, de acuerdo a diferentes características financieras y operacionales.

Existen varios métodos para agrupar cada uno de los casos dentro de un cluster. Quizás el más común es el de ir agrupando en cada una de las etapas del proceso los cluster con la distancia promedio más corta (distancia Euclidiana). Este proceso concluye cuando el conjunto de casos de la muestra se agrupa en el número de cluster especificados.

Vale la pena resaltar que la creación de la variable resultante no establece en ningún momento las ponderaciones necesarias, por ejemplo para un sistema de ecuaciones, sino que sólo determina las agrupaciones teniendo en cuenta apenas las características comunes.

2. Análisis discriminante multinomial

Por estos motivos, y con el propósito de además de lograr la estimación que permitiera la determinación de los pesos relativos de cada variable en cada estrato, se necesitaba de otra técnica que facilitara la posibilidad de constatar los resultados de los cluster. Para ello se decidió entonces escoger el uso del análisis Discriminante.

Tomando X_i observaciones, de $i = 1, \dots, n$ individuos que pertenecen a una población $j = 0, 1, \dots, m$ donde:

$$X_i | y_{i=j} \sim N(\mu_j, \Sigma) \quad (1)$$

con probabilidades marginales:

$$P(y_{i=j}) = q_j \quad (2)$$

¿Cuál será la probabilidad de que $P(g_i = j | x_i)$?

La respuesta se logra mediante el análisis discriminante.

Si se dice que:

$$P(y_{i=j} | x_i) P(x_i) = P(x_i | y_{i=j}) P(y_{i=j})$$

Entonces:

$$P(y_{i=j} | x_i) = \frac{P(x_i | y_{i=j}) P(y_{i=j})}{P(x_i)}$$

Por lo tanto:

$$= \frac{P(x_i | y_{i=j}) P(y_{i=j})}{P(x_i | y_{i=0}) P(y_{i=0}) + P(x_i | y_{i=1}) P(y_{i=1}) + \dots + P(x_i | y_{i=m})} \quad (3)$$

Dejando $g_i(x_i)$ como la distribución condicional de x_i , y y_i como la obtenida en la ecuación (1), y reescribiendo la ecuación (3) como la regla de Bayes, se tiene que:

$$P(y_{i=j} | x_i) = \frac{g_j(x_i) q_j}{\sum_{k=0}^m g_k(x_i) q_k} \quad (4)$$

A las observaciones x_i , se le asignan los individuos j , para obtener la ecuación (4) mucho más grande. Luego g_j se define como una distribución normal, (para todos importa el numerador, dado que todos los denominadores son iguales), siendo lo importante dentro del numerador el exponente de la distribución normal (aún en el exponente algunos términos pueden ser cancelados).

En general se tiene que:

$$q_j g_j(x_i) = (2\pi)^{-T/2} \Sigma^{-1/2} e^{-1/2(X_i - \mu_j)' \Sigma^{-1}(x_i - \mu_j)} + \log q_j$$

El término $(2\pi)^{-T/2} \pi^{-1/2}$ ha sido cancelado en la ecuación (4).

Las variables del lado izquierdo están dadas por la siguiente expresión exponencial:

$$\begin{aligned} & -1/2(x_i - \mu_j)' \Sigma^{-1}(x_i - \mu_j) + \log q_j \\ & = -1/2(x_i' \Sigma^{-1} - \mu_j' \Sigma^{-1})(x_i - \mu_j) + \log q_j \\ & = -1/2(x_i' \Sigma^{-1} x_i - x_i' \Sigma^{-1} \mu_j - \mu_j' \Sigma^{-1} x_i + \mu_j' \Sigma^{-1} \mu_j) + \log q_j \end{aligned}$$

El término $-1/2(x_i' \Sigma^{-1} x_i)$ también fue cancelado de la ecuación (4). Nótese que todos los términos en la expresión de arriba dada, tienen igual su transpuesta así:

$$\begin{aligned} & -1/2(-2 x_i' \Sigma^{-1} \mu_j + \mu_j' \Sigma^{-1} \mu_j) + \log q_j \\ & = x_i' \Sigma^{-1} \mu_j - 1/2 \mu_j' \Sigma^{-1} \mu_j + \log q_j \\ & = \alpha_j + x_i' \beta_j \end{aligned} \quad (5)$$

donde:

$$= -1/2 \mu_j' \Sigma^{-1} \mu_j + \log q_j$$

$\beta_j = \Sigma^{-1} \mu_j$ es un coeficiente que Maddala (1983) llama *Linear Discriminant Scores*, siendo una ecuación de la forma (5) para todos los individuos $j = 0, 1, \dots, m$.

Si se piensa en la clasificación de coeficientes, dado x_i , se puede calcular (5) para todo j y asignar x_i a los individuos j , para obtener la ecuación (5) más grande.

Ahora, se supone que se requiere de la ecuación (4).

$$\frac{P(y_{i=j} | x_i)}{P(y_{i=0} | x_i)} = \frac{q_j g_j(x_i)}{q_0 g_0(x_i)}$$

$$j = 1, 2, \dots, m.$$

Donde se está usando $j = 0$ como referencia.

$$\begin{aligned} & -1/2(x_i - \mu_j)' \Sigma^{-1} (x_i - \mu_j) + \log q_j - [-1/2(x_i - \mu_0)' \Sigma^{-1} (x_i - \mu_0) + \log q_0] \\ & -1/2(x_i' \Sigma^{-1} - \mu_j' \Sigma^{-1})(x_i - \mu_j) + \log q_j + 1/2(x_i' \Sigma^{-1} - \mu_0' \Sigma^{-1})(x_i - \mu_0) - \log q_0 \\ & = -1/2 x_i' \Sigma^{-1} x_i + 1/2 x_i' \Sigma^{-1} \mu_j + 1/2 \mu_j' \Sigma^{-1} x_i - 1/2 \mu_j' \Sigma^{-1} \mu_j + \log q_j \end{aligned}$$

Por razones similares, sólo el exponente puede importar, por lo tanto se tiene que:

$$\alpha_j = -1/2 \mu_0' \Sigma^{-1} \mu_0 - 1/2 \mu_j' \Sigma^{-1} \mu_j + \log q_j - \log q_0 \quad y$$

$$\beta_j = \Sigma^{-1} (\mu_j - \mu_0)$$

Debe anotarse que esta función discriminante tiene una ecuación menos que el número de categorías, donde m es el número de ecuaciones y $m+1$ el número de categorías.

3. Análisis modelos logit

Estos modelos determinan la probabilidad de que ocurra un evento J , dado X_i (vector de características). Cuando la decisión es entre 1 (que el evento ocurra) y 0 (que el evento no ocurra) el modelo es binomial; cuando se realiza entre más de dos eventos el modelo es multinomial.

Logit Binomial

Formalmente, si P_i es la probabilidad de que el evento que interesa ocurra, sobre la i -ésima vez, la variable aleatoria y_i será 1 si el evento ocurre y cero si no lo hace teniendo una función de probabilidad:

$$\text{donde } Y_i = \begin{cases} 1 & \text{con probabilidad } P_i \\ 0 & \text{con probabilidad } 1 - P_i \end{cases}$$

Esta función, definida como de máxima verosimilitud.

La función de máxima verosimilitud se puede representar como:

$$L = \prod_{i=1}^n P_i^{y_i} (1-P_i)^{1-y_i} \quad (6)$$

Cuando el logit se incrementa, P_i se especifica como la función logística (CDF) evaluada en $X'_i \beta$, donde $F(X'_i \beta)$ la denota, especificada por:

$$L = \prod_{i=1}^n [F(X'_i \beta)]^{y_i} [1 - F(X'_i \beta)]^{1-y_i} \quad (7)$$

donde el Logaritmo de la función de máxima verosimilitud será:

$$\text{Ln } L = \sum_{i=1}^n [y_i \ln F(X'_i \beta)] + (1-y_i) \ln [1 - F(X'_i \beta)] \quad (8)$$

La Función logística será entonces:

$$F(X'_i \beta) = \frac{e^{(X'_i \beta)}}{1 + e^{(X'_i \beta)}} \quad (9)$$

El valor máximo de la función de máxima verosimilitud será uno, sólo si es un producto de probabilidades, por lo que el máximo valor de $\text{Ln } L$ será cero.

Para maximizar el $\text{Ln } L$ con respecto a β se puede tentativamente encontrar el $\hat{\beta}$, considerado como el mejor estimador de β .

Derivando $\text{Ln } L$ e igualando a cero:

$$\frac{\delta \text{Ln } L}{\delta \beta} = \sum_{i=1}^n \left[y_i \frac{f(X'_i \beta)}{F(X'_i \beta)} - (1-y_i) \frac{f(X'_i \beta)}{F(X'_i \beta)} \right] X_i \quad (10)$$

Se obtiene que esta ecuación es no-lineal en β , maximizando entonces numéricamente $\text{Ln } L$.

Un método popular para realizar esta estimación, se logra con el procedimiento iterativo de Newton - Raphson:

$$\beta_{t+1} = \beta_t - \left[\frac{\delta^2 \text{Ln } L}{\delta \beta \delta \beta} \Big|_{\beta = \beta_t} \right]^{-1} \left[\frac{\delta \text{Ln } L}{\delta \beta} \Big|_{\beta = \beta_t} \right] \quad (11)$$

Donde $\hat{\beta}_t$ es el t-ésimo valor estimado, y las derivadas son evaluadas en β_t .

Bajo condiciones regulares, la función tendrá una distribución:

$$\sqrt{n} (\beta - \beta) \rightarrow N(0, \text{Lim} [-n^{-1} E \frac{\delta^2 \text{LnL}}{d\beta * d\beta}]^{-1}) \quad (12)$$

Para muestras finitas, la distribución asintótica de β puede ser aproximada por:

$$N(\beta, -[\frac{\delta^2 \text{LnL}}{\delta\beta * \delta\beta} \beta = \beta]^{-1}) \quad (13)$$

El hallazgo del mejor β , exige optimizar los valores de la primera y segunda derivada (no las ecuaciones). Estos podrían ser obtenidos numéricamente para cada β_i por aproximación. Sin embargo para el logit, este proceso de optimización se puede lograr a partir de las siguientes ecuaciones:

$$\frac{\delta \text{LnL}}{\delta\beta} = \sum_{i=1}^n y_i \frac{X_i'}{1 + e^{X_i \beta}} - \sum_{i=1}^n (1-y_i) \frac{X_i'}{1 + e^{-X_i \beta}} \quad (14)$$

$$\frac{\delta^2 \text{LnL}}{\delta\beta * \delta\beta} = \sum_{i=1}^n \frac{e^{-X_i \beta}}{(1 + e^{-X_i \beta})^2} X_i X_i' \quad (15)$$

Estas ecuaciones pueden ser evaluadas en cualquier β_i y usadas en la ecuación (11) con cualquier valor estimado. La maximización global presentada en (15) tendrá una matriz negativa en β . Este proceso iterativo continuará hasta que el criterio sea encontrado.

Logit Multinomial

Si se supone que existen $J > 2$ alternativas, donde y_{ij} es una variable binomial que será uno, si la i -ésima alternativa es elegida; de lo contrario será cero. Dejando que $P_{ij} = \text{Prob} [y_{ij} = 1]$.

Entonces:

$$\sum_{j=1}^j y_{ij} = \sum_{j=1}^j P_{ij} = 1 \quad (16)$$

$$L = \sum_{i=1}^n P_{i1} y_{i1}, P_{i2} y_{i2}, \dots, P_{ij} y_{ij} \quad (17)$$

$$\text{Ln}L = \sum_{i=1}^n (\text{Log } P_{i1}y_{i1} + \text{Log } P_{i2}y_{i2} + \dots + \text{Log } P_{ij}y_{ij}) \quad (18)$$

Para este caso, la probabilidad de que ocurra el evento J estará dada por:

$$P_{ij} = \frac{e^{X_i \beta_j}}{J} \quad \text{donde: } P_{ij} \leq 1 \quad (19)$$

$$\sum_{j=1} e^{X_i \beta_j} \quad \text{y:} \quad \sum P_{ij} = 1$$

Para encontrar los parámetros del modelo se realiza la siguiente normalización $B_1 = 0$, entonces:

$$P_{ij} = \frac{1}{J \left(1 + \sum_{j=1} e^{X_i \beta_j} \right)} \quad (20)$$

$$P_{ij} = \frac{e^{X_i \beta_j}}{J \left(1 + \sum_{j=1} e^{X_i \beta_j} \right)} \quad j = 2, \dots, J \quad (21)$$

La función de máxima verosimilitud más apropiada se obtendrá entonces sustituyendo los P_{ij} en la ecuación (18).

IV. Análisis de resultados

Una vez adelantados los procesos de organización y depuración de la información, se iniciaron las tareas del análisis, tomando en primer lugar todas las variables encuestadas y realizando con ellas un análisis de frecuencias.

Con esta tarea se buscaba descartar en primera instancia aquellas variables que, por tener una alta participación en las observaciones, no tendrían la capacidad de ser discriminatorias. El ejemplo más característico puede ser el de la disponibilidad del servicio de energía eléctrica, que para el caso de la ciudad de Cali tuvo una cobertura del 98%, lo que permite establecer que no puede ser un criterio que ayude a diferenciar estratos.

Con el resto de variables que si mostraron una aceptable distribución de frecuencias se iniciaron entonces las diferentes pruebas que pudieran separar las observaciones en función de grupos determinados de variables, empezando para ello con el uso de análisis de cluster.

Cabe resaltar que dada la discusión acerca de la suficiencia por parte de cierta clase de criterios para lograr aceptables niveles de explicación, se decidió dividir las variables en grupos, y con ellos realizar estimaciones.

Los principales grupos conformados fueron:

- a. Variables de observación externa directa de las viviendas y su entorno.
- b. Las anteriores más las de observación directa interna de las viviendas.
- c. Disponibilidad de electrodomésticos.
- d. Todas las anteriores.
- e. Disponibilidad de vehículo, teléfono más todas las anteriores.
- f. Las del aparte "e" más el consumo de agua.
- g. Las del aparte "e" más algunas variables del hogar.

Todos los ejercicios exploratorios fueron inicialmente ejecutados con las 2785 encuestas de la sub-muestra proporcional sistemática de la ciudad, para posteriormente ser realizados con la totalidad de la muestra escogida, es decir, la muestra de cada una de las manzanas residenciales de la ciudad.

Como el método de trabajo determinaba que cuando se encontrara una manzana, que por sus condiciones externas pudiera ser considerada como heterogénea debería realizarse en ella encuestas adicionales dependiendo de la posición con respecto a la tipicidad, de los 9.111 del total de unidades dedicadas a vivienda residencial se encontró que un 14% de ellas tenían estas características, ascendiendo entonces el número de encuestas diligenciadas a un total de 10.805.

Con la técnica del cluster y los grupos de variables antes mencionados, se realizaron los ejercicios de exploración, con el propósito de encontrar el grupo de variables que contribuyera a explicar en forma significativa una estratificación, por medio de la diferenciación con grupos de viviendas cuyas características fueran similares entre sí. Como por definición se había determinado que no existiera ponderación alguna, era necesario que como dato de entrada todas las variables tuvieran el mismo peso. Por lo tanto se decidió utilizar las categorías de cada variable como dummies, para todas menos una, y así evitar la posible autocorrelación entre ellas.

De acuerdo con las características comunes que comparten cada uno de los grupos formados por el procedimiento cluster, se asoció cada uno a un estrato socioeconómico particular. La diferenciación de las viviendas fue determinada entonces, por el grupo de variables incluídas en el proceso, a través de la agrupación de observaciones con características comunes. Vale resaltar que de las diferentes pruebas realizadas con el cluster, cuya clasificación inicial fue en seis grupos, no permitieron identificar directamente los 6 estratos socioeconómicos, obligando así a realizar diferentes pruebas en forma iterativa con subgrupos de viviendas.

Del primer análisis cluster generalmente fue fácil la identificación de los grupos que reúnen las viviendas de los estratos socioeconómicos más bajo y más alto; las demás viviendas, en la mayoría de los casos, se reúnen en grupos muy grandes de los cuales no fue posible diferenciar los demás estratos. En consecuencia, con los casos reunidos en estos grupos, se ejecutó un nuevo proceso cluster con el propósito de identificar más claramente las viviendas que por sus características pertenecían a los restantes estratos. En algunas pruebas fue necesario repetirlo más de una vez.

Después de realizar diversas pruebas, siguiendo las etapas descritas que permitían detectar las variables más significativas en la contribución a la estratificación socioeconómica de la ciudad, se definió un grupo de variables que incluía, además de las características físicas (exteriores e interiores y estado de las vías adyacentes), la disponibilidad de ciertos elementos en las viviendas, tales como la disponibilidad de teléfono, de electrodomésticos, y de vehículo particular, que aportaban significativamente en la clasificación en un estrato particular (Cuadro 1).

Para realizar el proceso de estratificación se dividió el número total de encuestas en 6 grupos usando el análisis cluster, en función de las variables descritas. La distribución porcentual de los grupos encontrados se muestra en el Cuadro 1.

Cuadro 1

**CLUSTER EN FUNCION DE LAS CARACTERISTICAS EXTERIORES,
INTERIORES Y DISPONIBILIDAD DE CIERTOS ELEMENTOS**

Cluster	Casos	Porcentaje
I	2730	25.2
II	461	4.3
III	628	5.8
IV	471	4.4
V	3957	36.6
VI	2558	23.7

Como se expresó, con la agrupación total en 6 grupos (es decir con las 10.805 encuestas) igual que en los ejercicios desarrollados con la muestra, no se logró diferenciar directamente los 6 estratos socioeconómicos, siendo posible sólo diferenciar el estrato I y el estrato VI, y observando la distribución de frecuencias de las variables empleadas, para cada uno de los cluster formados.

Por sus características, las observaciones clasificadas en el cluster 3 (5.8% del total de las encuestas realizadas), se pueden asociar claramente al estrato I, donde aparecie-

ron las condiciones físicas menos favorables, siendo a su vez aquellas que carecen de la mayoría de los electrodomésticos empleados en el procedimiento; tampoco disponen de servicio telefónico, ni de vehículo particular.

Las clasificadas en el cluster 4 parecen estar asociadas al estrato socioeconómico VI. Este grupo se caracteriza por que las viviendas presentan las mejores características físicas tanto internas como externas, y por poseer la mayor parte de los electrodomésticos, presentando además la mayor frecuencia en disponibilidad de vehículo particular y de servicio telefónico.

Con el primer proceso de análisis cluster los estratos II, III, IV, y V no se identificaron claramente; en consecuencia se procedió a realizar un nuevo proceso de análisis en función del mismo grupo de variables utilizadas en el primer ejercicio, esta vez con los casos clasificados en los grupos 1, 2, 5 y 6. Esta submuestra se dividió en 12 cluster, con el propósito de identificar los casos de los estratos restantes (Cuadro 2).

Con este segundo ejercicio, se lograron extraer las que por sus características pertenecen al estrato socioeconómico II, (cluster 1, 5, 7 y 12). Estas viviendas corresponden a aquellas que tras el nuevo ejercicio presentaron las características exteriores e interiores menos favorables y registraron la menor cantidad de elementos en el hogar.

Con los cluster restantes no se identificaron claramente los estratos III, IV y V. Así que se procedió una vez más a clasificar los casos de los cluster restantes.

Cuadro 2

**CLUSTER EN FUNCION DE VARIABLES CARACTERISTICAS
EXTERIORES, INTERIORES Y DISPONIBILIDAD DE
CIERTOS ELEMENTOS**

Cluster	Casos	Porcentaje sobre el total de casos
I	1288	11.9
II	1252	11.6
III	532	4.9
IV	387	3.6
V	712	6.6
VI	65	0.6
VII	483	4.5
VIII	2323	21.5
IX	1331	12.3
X	677	6.2
XI	226	2.1
XII	430	4.0

Se realizó así con la submuestra un nuevo cluster para dividir los casos restantes en 12 grupos, en función del grupo de variables consideradas a lo largo del ejercicio, con el propósito de detectar los casos de los estratos faltantes (Cuadro 3).

Este nuevo ejercicio contribuyó a diferenciar más claramente, por las características de las viviendas y por la disponibilidad de los elementos considerados, las observaciones asociadas a los estratos III (cluster 2, 4, 6, 8, 10), IV (cluster 9, 11, 12) y V (cluster 1, 3, 5, 7).

Cuadro 3

**CLUSTER EN FUNCION DE VARIABLES CARACTERISTICAS
EXTERIORES, INTERIORES Y DISPONIBILIDAD DE
CIERTOS ELEMENTOS**

Cluster	Casos	Porcentaje sobre el total de casos
I	535	4.9
II	114	1.1
III	205	1.9
IV	1613	15.0
V	448	4.1
VI	22	0.2
VII	107	1.0
VIII	1520	14.1
IX	1630	15.1
X	55	0.5
XI	103	0.9
XII	441	4.1

Una vez detectados los 6 grupos con características afines a cada uno de los estratos socioeconómicos: bajo-bajo, bajo, medio-bajo, medio, medio-alto, alto, haciendo uso del cluster, se adelantaron pruebas con el uso del análisis discriminante de la variable obtenida con el uso del análisis cluster. Como se expresó, con la aplicación de este nuevo modelo se garantiza que cada una de las observaciones sea clasificada en el estrato correcto, de acuerdo a las variables tenidas en cuenta a lo largo del proceso de estratificación.

Vale anotar que esta técnica de clasificación encuentra una combinación lineal de atributos que separa los grupos lo máximo posible, asignando cada uno de los casos a un grupo, utilizando la comparación de los resultados estimados por las funciones con los valores estimados en el cluster.

La variable resultante de la aplicación de las funciones del análisis discriminante para cada uno de los 6 grupos, es la variable que se emplea en la estratificación, garantizando que cada uno de los nuevos estratos tenga características similares, y que exista una diferenciación precisa entre los grupos.

Siguiendo el procedimiento descrito, y tomando como referencia los resultados de los análisis de cluster y de los análisis discriminante, se realizaron diferentes pruebas, inicialmente con la sub-muestra de aproximadamente el 33% del total de encuestas elaboradas para el estudio, y posteriormente con todo el conjunto de encuestas. En primer lugar se corrió el modelo en función de las variables de observación directa externa de las viviendas y su entorno, con el 49.1% del total de casos correctamente clasificados. Como se puede apreciar (Cuadro 4), para el estrato I aparece el 78% de los casos clasificados correctamente, entendiéndose como tal la concordancia entre las dos técnicas de estimación. Así mismo, el 16.2% clasificados inicialmente (cluster) en ese estrato, al ser utilizadas sólo en este grupo de variables quedarían ubicadas en el estrato II.

Observando la diagonal principal de esa matriz, se puede corroborar la concordancia entre los dos métodos de estimación, cuando se toman sólo las variables de observación directa externa de las viviendas. En términos generales se puede pensar que cuando se usan sólo estas variables, existe la probabilidad de que el 50.9% de los casos queden incorrectamente clasificados.

Cuadro 4

**DISCRIMINANTE DE LA ESTRATIFICACION PROPUESTA EN
FUNCION DE LAS VARIABLES OBSERVABLES**

Estrato	Casos	Porcentaje de casos clasificados por estrato Estratificación propuesta					
		I	II	III	IV	V	VI
I	628	78.0	16.2	2.1	1.1	1.1	1.4
II	2912	11.5	56.9	12.8	12.8	4.5	1.4
III	3323	1.0	11.2	39.6	25.5	19.8	2.8
IV	2174	0.5	12.2	15.8	49.2	19.6	2.8
V	1294	0.4	6.6	15.3	25.1	37.4	15.1
VI	466	0.0	1.5	4.7	8.8	24.7	60.3
Total	10797	873	2489	2266	2664	1822	683

Porcentaje de casos por estrato correctamente clasificados: 49.1%

Al realizar el ejercicio en función de las características exteriores, interiores y del entorno de la vivienda, el porcentaje de casos correctamente clasificados se incrementó a 68.5% (Cuadro 5). Nótese que todos los datos de la diagonal principal aumentaron el porcentaje de casos correctamente clasificados, por lo que se puede pensar que en la medida en que se incrementa el número de las variables, donde ya se tomaron en consideración aquellas que se encuentran al interior de las viviendas aumenta la posibilidad de lograr mayores estimaciones.

Cuadro 5

**DISCRIMINANTE DE LA ESTRATIFICACION PROPUESTA EN
FUNCION DE LAS CARACTERISTICAS EXTERIORES
E INTERIORES DE LAS VIVIENDAS**

Estrato	Casos	Porcentaje de casos clasificados por estrato Estratificación propuesta					
		I	II	III	IV	V	VI
I	628	80.6	15.1	1.9	0.0	1.3	1.1
II	2912	6.4	74.7	11.9	3.8	2.2	1.0
III	3323	1.1	6.3	73.4	10.3	6.6	2.3
IV	2174	0.5	10.6	13.1	57.2	16.2	2.5
V	1294	0.3	4.9	11.4	10.4	57.2	15.9
VI	466	0.0	0.4	13.1	4.3	19.1	63.1
Total	10797	742	2775	3289	1852	1472	667

Porcentaje de casos por estrato correctamente clasificados: 68.5%

Finalmente, al realizar el ejercicio en función de todo el conjunto de variables escogidas, el porcentaje total de casos bien clasificados se incrementó a 74.3%, y el porcentaje de casos de cada estrato considerado individualmente mejoró también notablemente, donde para el estrato I alcanzó 81.5%; para el estrato II, 75.7%; para el estrato III, 78.6%; para el estrato IV, 67.0%; para el estrato V, 66.7%; y para el estrato VI, 81.8% (Cuadro 6).

Como se mencionó desde un principio, el objetivo del diseño de la estratificación debía pretender además reflejar la relación existente con el nivel de ingresos percibidos por los habitantes, donde, de acuerdo con los avances en las estimaciones, se hacía necesario incluir variables adicionales a las solas características físicas de

las viviendas. Con esto se buscaba que, además de contribuir a separar más claramente los diferentes grupos, ellos reflejaran esta relación de las personas que habitan las viviendas.

Con el fin de verificar esta relación, se realizó la constatación usando un logit multinomial con la variable obtenida a partir del discriminante en función del ingreso total de los hogares, para así observar la probabilidad de pertenencia a un estrato socioeconómico determinado, dado el nivel de ingresos (Cuadro 7).

Cuadro 6

DISCRIMINANTE DE LA ESTRATIFICACION PROPUESTA EN FUNCION DE LAS CARACTERISTICAS EXTERIORES E INTERIORES DE LAS VIVIENDAS Y DISPONIBILIDAD DE CIERTOS ELEMENTOS

Estrato	Casos	Porcentaje de casos clasificados por estrato Estratificación propuesta					
		I	II	III	IV	V	VI
I	628	81.5	14.8	1.4	0.0	2.2	0.0
II	2912	6.1	75.7	12.3	3.5	2.3	0.1
III	3323	1.1	6.0	78.6	7.9	6.1	0.4
IV	2174	0.4	10.5	9.0	67.0	12.0	1.2
V	1294	0.2	5.3	7.9	9.6	66.7	10.4
VI	466	0.0	1.3	6.0	3.2	7.7	81.8
Total	10797	736	2796	3306	1957	1445	557

Porcentaje de casos por estrato correctamente clasificados: 74.3%

Cuadro 7

LOGIT MULTINOMIAL DE LA ESTRATIFICACION PROPUESTA EN FUNCION DEL INGRESO TOTAL DE LOS HOGARES

Coefic.	Func. 1	Func. 2	Func. 3	Func. 4	Func. 5
Const.	1.14421	1.46892	0.537970	-0.683624	-2.5808
Ingtot	0.529E-05	0.137E-04	0.169E-04	0.196E-04	0.218E-04

Graficando esta función se puede apreciar que existe una relación bastante definida entre el estrato estimado y el nivel de ingresos total de los hogares. La probabilidad de que una observación pertenezca al estrato I disminuye a medida que aumenta el nivel del ingreso; una situación similar se presenta para las clasificadas en el estrato II, aunque la disminución de este último se presenta menos rápidamente que en el primer caso (Gráfico 1).

La probabilidad de pertenecer al estrato III aumenta hasta ingresos de \$100.000.00; para ingresos superiores, la probabilidad de pertenecer a este estrato disminuye rápidamente a medida que aumentan los niveles de ingreso. Sin embargo para ingresos menores de \$280.000.00, predomina la probabilidad de pertenecer al estrato socioeconómico III.

La probabilidad de pertenecer al estrato IV, aumenta suavemente hasta niveles de ingreso de \$375.000.00, y a partir de este nivel de ingresos, la probabilidad de pertenecer a este estrato disminuye lentamente hasta ingresos de \$600.000.00, donde a partir de éste dicha probabilidad disminuye rápidamente. Para rangos de ingresos comprendidos entre \$280.000.00 y \$400.000.00, predomina la probabilidad de pertenecer al estrato socioeconómico IV.

La probabilidad de que una vivienda esté clasificada en el estrato V o VI, aumenta con el nivel de ingreso total de los hogares. La probabilidad de pertenecer al estrato V, predomina entre ingresos de \$450.000.00 y \$875.000.00, y a partir de este nivel, la probabilidad de que una observación se ubique en él comienza a descender, predominando para ingresos mayores la probabilidad de que las viviendas pertenezcan al estrato VI.

Estos resultados permiten por tanto inferir que la estratificación obtenida guarda una estrecha relación con el nivel de ingreso de los hogares, ya que la probabilidad de que una vivienda esté clasificada en los estratos bajos disminuye a medida que aumenta el ingreso total de los hogares, mientras que la probabilidad de que la vivienda esté estratificada en niveles altos aumenta con el nivel de ingreso de los hogares.

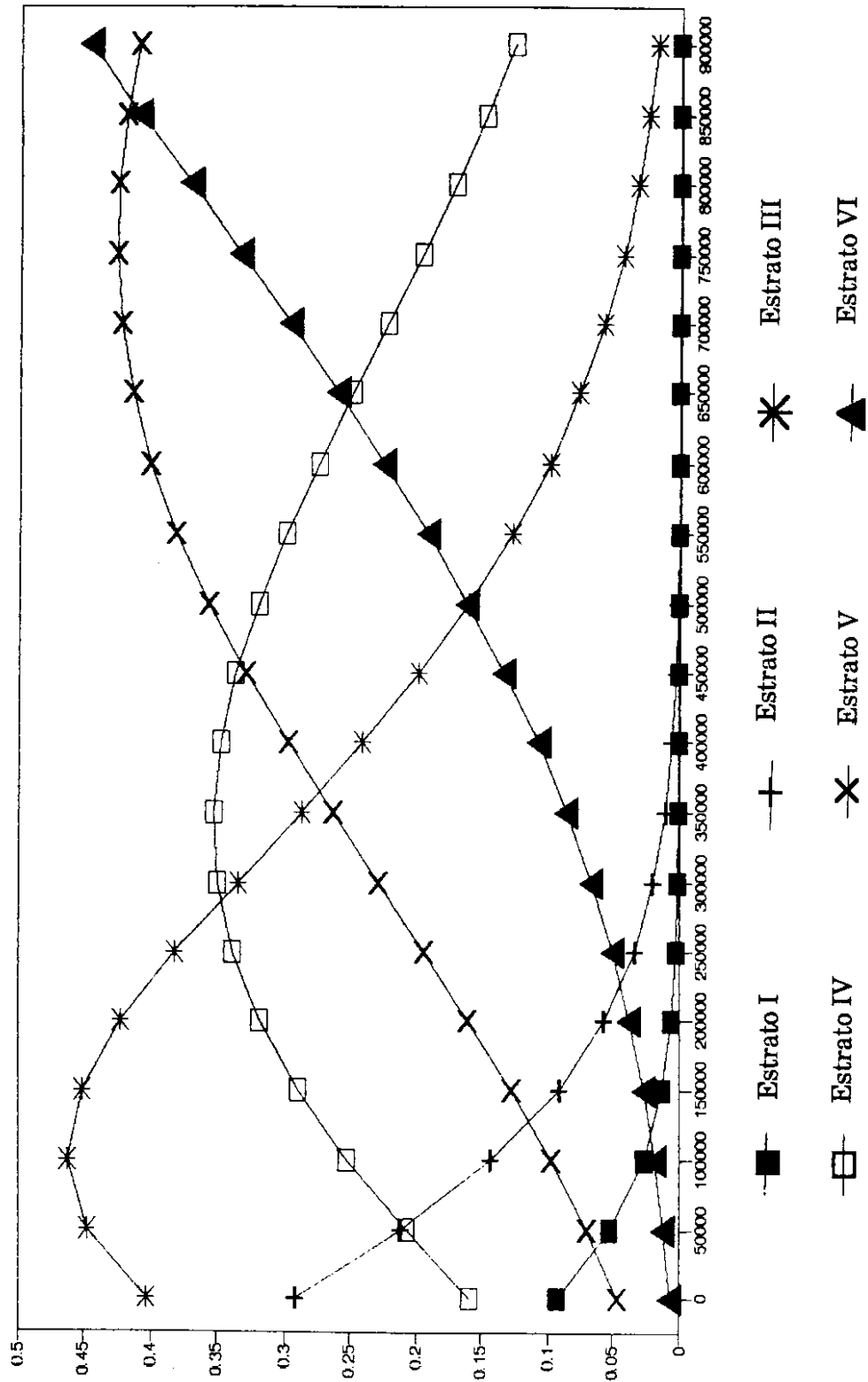
Así mismo, resulta relevante observar cómo hasta los niveles de ingreso de \$275.000.00 la mayor probabilidad de pertenencia ocurre para el estrato III, siendo para el estrato IV en los rangos de ingreso entre \$275.000,00 y \$450.000.00. De igual forma, para aquellas familias cuyos ingresos totales se encuentran entre \$450.000.00, y \$850.000.00 la mayor posibilidad se presenta para pertenecer al estrato IV.

La evidencia de todo esto permite aseverar que si bien la estratificación diseñada con este grupo de variables apenas logra explicar correctamente el 74.3%, logra una adecuada relación con la distribución de ingresos reportada, por lo que al ser utilizada como instrumento de política que necesite alguna referencia a los ingresos, tendrá una aceptable probabilidad de acertar en el diseño de esos programas.

Otro modelo alternativo

Si bien el diseño logrado hasta este momento cumplía con gran parte de los requisitos exigidos, las condiciones del estudio, cuyo enfoque estaba orientado a la

Gráfico 1
 PROBABILIDAD DE PERTENECER A CADA ESTRATO DE ACUERDO
 CON EL INGRESO



“exploración” sobre la base de diferentes combinaciones de variables, obligaba a seguir intentando probar en busca de un modelo mejor.

Por tal motivo se decidió agregar variables adicionales, tales como el número de miembros del hogar, el número de hogares por vivienda, y la condición de vivienda de los miembros que la habitan. Las tres primeras variables se trabajaron en forma continua, mientras que con la última cada una de las categorías se trabajó como variable dummy. Dado que el ejercicio combina variables dummy y variables continuas, fue necesario estandarizar una vez más todas las variables para darles igual ponderación a todo el conjunto de estratos.

Una vez realizadas las diferentes agrupaciones de las viviendas siguiendo las diferentes etapas iterativas, a través del análisis cluster, se obtuvo una variable a la cual se le aplicaron, como en el caso de todas las anteriores, las pruebas del análisis discriminante para los diferentes grupos de variables.

El ejercicio de estratificación en función de las variables externas y del entorno observables, (Cuadro 8), arrojó que el porcentaje de casos correctamente clasificados fue de 64.9%. Sin embargo, al observar los casos bien clasificados de cada estrato en particular, fue el estrato III el que agrupó la mayor parte de las observaciones (35.3%), registrando el mayor porcentaje de casos bien clasificados (80.0%). De ahí que el promedio sea alto con relación al obtenido en la estratificación anterior.

Los porcentajes de los estratos restantes fueron comparativamente más bajos, siendo respectivamente, para el estrato I de 46.6%, para el estrato II de 69.1%, para el estrato IV de 53.7%, para el estrato V de 48.3% y para el estrato VI de 40.3%.

Cuadro 8

**DISCRIMINANTE DE LA ESTRATIFICACION ALTERNATIVA EN
FUNCION DE LAS VARIABLES OBSERVABLES**

Estrato	Casos	Porcentaje de casos clasificados por estrato Estratificación propuesta					
		I	II	III	IV	V	VI
I	597	46.6	28.1	16.9	6.2	1.8	0.3
II	2324	7.5	69.1	17.5	3.1	2.2	0.6
III	3696	0.5	7.8	80.0	9.1	1.1	1.5
IV	2270	0.3	3.5	36.5	53.7	3.8	2.2
V	1022	0.2	1.1	19.0	11.4	48.3	20.0
VI	556	0.4	0.7	30.4	14.9	13.3	40.3
Total	10465	482	2158	4654	1865	757	549

Porcentaje de casos por estrato correctamente clasificados: 64.9%

El nuevo ejercicio de discriminante, en función de las características exteriores, interiores y del entorno de las viviendas, mostró que el porcentaje de observaciones bien clasificadas se eleva a 72.6% con respecto al obtenido con la estratificación analizada anteriormente (Cuadro 9). Individualmente mejora el porcentaje de los estratos II, III, IV, V, que pasan respectivamente de 74.7% a 80.1%, de 73.4% a 79.8%, de 57.2% a 62.4% y de 57.2% a 60.3%; así mismo, disminuye el porcentaje de explicación de los estratos I y VI que respectivamente pasan de 80.6% a 80.4%, y de 50.2% a 63.1%.

Cuadro 9

**DISCRIMINANTE DE LA ESTRATIFICACION ALTERNATIVA
FUNCION DE LAS CARACTERISTICAS EXTERIORES E
INTERIORES DE LAS VIVIENDAS**

Estrato	Casos	Porcentaje de casos clasificados por estrato Estratificación propuesta					
		I	II	III	IV	V	VI
I	597	80.4	14.1	4.0	0.7	0.8	0.0
II	2324	5.0	80.1	10.5	2.3	1.8	0.2
III	3696	0.2	5.3	79.8	11.1	1.7	1.9
IV	2270	3.3	1.9	26.2	62.4	3.4	2.9
V	1022	0.4	1.1	7.3	10.9	60.3	20.1
VI	556	1.8	0.4	13.3	20.1	14.2	50.2
Total	10465	690	2198	3961	2106	884	626

Porcentaje de casos por estrato correctamente clasificados: 72.6%

Cuando se consideró todo el conjunto de variables usadas en la elaboración del ejercicio, es decir las variables de características de las viviendas, disponibilidad de elementos, el número de miembros del hogar, el número de hogares por vivienda y las condiciones de vivienda de las personas que la habitan, el porcentaje de casos correctamente clasificados se incrementó considerablemente, llegando a tener una explicación del 83.3% (Cuadro 10).

Cuadro 10

**DISCRIMINANTE DE LA ESTRATIFICACION ALTERNATIVA
FUNCION DE LAS VARIABLES EMPLEADAS EN EL MODELO**

Estrato	Casos	Porcentaje de casos clasificados por estrato Estratificación propuesta					
		I	II	III	IV	V	VI
I	597	84.9	11.6	2.7	0.3	0.5	0.0
II	2324	3.7	85.2	7.0	2.2	1.6	0.1
III	3696	0.1	3.8	88.2	6.1	1.8	0.0
IV	2270	3.0	1.5	16.9	74.5	3.1	1.0
V	1022	0.5	1.3	7.8	7.3	80.4	2.6
VI	556	1.6	0.4	1.6	5.6	9.0	81.8
Total	10465	680	2240	3911	2077	1049	508

Porcentaje de casos por estrato correctamente clasificados: 83.3%

A pesar de que el ejercicio resultó bastante aceptable al aplicar las funciones obtenidas con el análisis discriminante para cada uno de los estratos socioeconómicos al total de las encuestas, al usar el modelo logit multinomial a la variable obtenida en función del total de ingresos de los hogares no se encontró una relación bien definida entre las dos variables como la obtenida en el ejercicio anterior (Gráfico 2).

Se puede apreciar que la probabilidad de que una observación haya quedado clasificada en los estratos I y II disminuye claramente a medida que aumenta el nivel de ingresos de los hogares; la probabilidad de pertenecer al estrato I no predomina para ningún rango de ingresos, mientras que la probabilidad de pertenecer al estrato II, predomina hasta ingresos aproximados de \$ 80.000.00.

La probabilidad de que una observación esté clasificada en el estrato III aumenta hasta ingresos aproximados de \$180.000.00, y a partir de este nivel de ingresos dicha probabilidad comienza a descender rápidamente. La probabilidad de pertenecer al estrato III predomina con respecto a la probabilidad de pertenecer a los demás estratos socioeconómicos para niveles de ingreso comprendidos entre \$ 80.000.00 y \$ 225.000.00. Así mismo la probabilidad de pertenecer al estrato IV predomina sobre la probabilidad de pertenecer a los demás estratos para ingresos comprendidos entre \$ 225.000.00 y \$550.000.00, donde para este nivel de ingresos es igual la probabilidad de que la vivienda esté estratificada en el estrato IV o en el estrato VI.

De otro lado, la probabilidad de pertenecer al estrato V no predomina para ningún rango de ingresos específico, y para niveles de ingresos de \$ 450.000.00 existe la misma

probabilidad de que la vivienda sea estratificada en el estrato III, V o VI. Como se puede apreciar, este modelo que contempla otro grupo adicional de variables, si bien logró una mejor estimación en términos de mayor nivel de explicación que el ejercicio anterior, no alcanza a establecer una adecuada relación con la distribución del ingreso, siendo por lo tanto un instrumento que no puede garantizar que aquellas medidas de política que puedan utilizarse sobre la base, tengan las respuestas esperadas.

Sin embargo parece ser que de seguir realizando otros ejercicios de exploración, con otras combinaciones de variables (parece ser de tipo social), quizás se puedan lograr mejores niveles de explicación, no sólo en términos de la identificación del estrato, sino además en su relación con la distribución del ingreso que reporta la agrupación social que se analiza.

Es importante anotar también que los casos que no están correctamente clasificados en el estrato respectivo, están localizados en los estratos más cercanos, a pesar de que para el estrato IV un 10.2% de las viviendas presentan características del estrato II, para el estrato V un 7.9% presenta características del estrato III y para el estrato VI, un 6.0% presenta también características del estrato III, evidenciando con ello el problema que se conoce como la "dificultad en los límites", donde resulta bastante complejo determinar la pertenencia a uno u otro estrato.

Una vez escogidos los modelos de estratificación, que en una buena parte cumplían con los requisitos escogidos, era necesario conocer las características de las funciones halladas para cada estrato.

Como se puede observar (Cuadro 11), las variables que se lograron determinar como representativas fueron aquellas que el terreno consideró que mejor lograban diferenciar las características para cada uno de los estratos. Así, tomando en consideración cada una de las categorías asignadas, se mostraron las ponderaciones relativas para cada una de ellas en cada estrato (Cuadro 12).

Las funciones resultantes del análisis discriminante aplicadas como base para la estratificación, mostraron por ejemplo para la variable ANDEN1 (sin andén ni antejardín), que al ser dummy tiene las opciones de cero o uno, que tiene como coeficiente para el estrato I 3.245044, para el estrato II 1.458627, para el estrato III 0.89696, para el estrato IV 0.913604, para el estrato V 1.219971 y para el estrato VI 1.007776. Una vez que se tiene el valor para cada una de las variables consideradas en el proceso (cero o uno), se aplican las seis ecuaciones resultantes para cada uno de los seis estratos socioeconómicos, y la vivienda se asigna al estrato para el cual el resultado sea más alto.

En el evento de aplicar las funciones para estratificar una nueva manzana es necesario reunir los valores de las diferentes variables consignadas en el Cuadro 12.

Es importante tener en cuenta que, de las categorías consideradas en cada una de las variables, sólo una de ellas puede ser uno, mientras las restantes categorías tendrán valor cero. Por ejemplo, para la variable andenes y antejardín (con relación a la vivienda), la vivienda tendrá el valor 1, sólo en una de las categorías: sin andén ni

antejardín, sin andén y con antejardín, con andén y sin antejardín, con andén y con antejardín. Es necesario tener en cuenta este proceso en cada una de las variables. En el caso de la disponibilidad de los diferentes elementos considerados, se asignará uno si la vivienda dispone del elemento respectivo; de lo contrario tendrá un valor cero. Una vez registrados los valores (cero o uno) para todas las variables, se aplicarían las seis ecuaciones y la estratificación de la manzana respectiva se realizaría comparando los resultados obtenidos y tomando el mayor valor.

Cuadro 11

VARIABLES EMPLEADAS EN LA ESTRATIFICACION

1. Andenes y antejardín	
Sin andén ni antejardín	(ANDEN1)
Sin andén y con antejardín	(ANDEN2)
Con andén y sin antejardín	(ANDEN3)
Con andén y con antejardín	
2. Zonas verdes y recreativas	
No existen	(ZONAS1)
Existen no acondicionadas	(ZONAS2)
Comunitarias abiertas	(ZONAS3)
Recreativas de conjunto	(ZONAS4)
Privadas	
3. Vías locales	
Sendero o camino sin pavimento	(VIAS1)
Vehicular sin pavimento	(VIAS2)
Vehicular pavimentada y en mal estado	(VIAS3)
Vehicular pavimentada en buen estado	
4. Condiciones de localización	
Zona de alto riesgo ambiental	(CON-LO1)
Zona de deterioro social	(CON-LO2)
Zona residencial mixta	(CON-LO3)
Zona residencial integrada a la industria	(CON-LO4)
Zona residencial integrada al comercio	(CON-LO5)
Zona residencial exclusiva	
5. Tipo de vivienda	
Casa independiente	(TIPO-VI1)
Casa en condominio	(TIPO-VI2)
Apartamento en edificio	(TIPO-VI3)
Apartamento en condominio	

(Sigue)

(Continuación Cuadro 11)

6. Condiciones de la vivienda	
Totalmente construida	(CON-VI1)
Vivienda en construcción	(CON-VI2)
Casa - lote	(CON-VI3)
Rancho, choza o cabaña	(CON-VI4)
Otro	
7. Situación de la calle que cruza por el frente con la vivienda	
Destapada	(SIT-CAL1)
Asentada	(SIT-CAL2)
Adoquinada	(SIT-CAL3)
Pavimentada a la mitad	(SIT-CAL4)
Completamente pavimentada	
8. Ubicación de la vivienda con relación a las vías	
Frente a calle secundaria	(UB-VIA1)
Frente a calle principal	(UB-VIA2)
Frente a autopista	
9. Ubicación de la vivienda con relación a la avenida	
Menos de cinco cuadras de avenida	(UB-AV1)
Más de cinco cuadras de avenida	
Nota: La categoría de vivienda en edificación que no es de habitación se reunió con la categoría de "otro", debido a que individualmente no resultaba significativa.	
10. Material predominante en las paredes exteriores de la vivienda	
Esterilla, cartón, láta, guadua, tabla	(MEX1)
Adobe, bahareque	(MEX2)
Ladrillo rúst. a la vista o rep.sin pintar	(MEX3)
Bloque prefabricado a la vista	(MEX4)
Repello pintado o ladrillo pulido	(MEX5)
Revestimiento plástico	(MEX6)
Acabados de lujo	
11. Material predominante en la puerta de acceso y en las ventanas exteriores de la vivienda	
Lámina de hierro prensada	(MPU1)
Aluminio corriente	(MPU2)
Aluminio anodizado	(MPU3)
Madera fina	(MPU4)
Madera corriente	(MPU5)
Tabla o lata	(MPU6)
Otro	

(sigue)

(Continuación Cuadro 11)

12. Material predominante en el piso de la vivienda	
Tierra	(MPIS1)
Cemento o adoquín	(MPIS2)
Madera	(MPIS3)
Baldosa	(MPIS4)
Alfombra	
13. Area del lote	
Hasta 75 m ²	(ARLOT1)
75 a 110 m ²	(ARLOT2)
100 a 150 m ²	(ARLOT3)
150 a 200 m ²	(ARLOT4)
200 a 300 m ²	(ARLOT5)
Más de 300 m ²	
14. Area construída	
30 a 75 m ²	(ARCOS1)
60 a 90 m ²	(ARCOS2)
90 a 120 m ²	(ARCOS3)
120 a 150 m ²	(ARCOS4)
150 a 200 m ²	(ARCOS5)
Más de 200 m ²	
15. Garaje	
Sin garaje	(GARAJ1)
Garaje descubierto o semi-descubierto	(GARAJ2)
Garaje cubierto sencillo	(GARAJ3)
Garaje cubierto múltiple	
16. Material predominante en el baño principal	
Sin enchape	(MTBAÑ1)
Con enchape	
17. Disponibilidad de piscina	
Posee piscina	(PISCINA)
No posee piscina	
18. Material de la cocina	
Sin mobiliario	(MCOC1)
Madera	(MCOC2)
Loza de cemento sin enchape	(MCOC3)
Con enchape en baldosín	(MCOC4)
Enchape en cerámica o metal inoxidable	

(Sigue)

(Continuación Cuadro 11)

-
- 19. Combustible utilizado en la vivienda para cocinar**
- | | |
|---------------|---------|
| Electricidad | (COMB1) |
| Gas | (COMB2) |
| Gasolina | (COMB3) |
| Petróleo | (COMB4) |
| Carbón y leña | |
- 20. Disponibilidad de teléfono**
- | | |
|------------------|--------|
| Sin teléfono | (TEL1) |
| Servicio público | (TEL2) |
| Servicio privado | |
- 21. Disponibilidad en el hogar**
- Estufa de gas
 - Estufa eléctrica
 - Estufa a gasolina o petróleo
 - Estufa de carbón o leña
 - Licuada

Nota: Las categorías carbón y leña se reunieron en una variable, dado que individualmente no resultaban significativas.

- Auxiliar de Cocina
- Aspiradora
- Brilladora
- Nevera
- Máquina lavadora de ropa
- Máquina secadora de ropa
- Máquina lavadora de platos
- Horno eléctrico
- Horno micro-ondas
- Televisor
- Betamax
- Equipo de sonido o radiola
- Vehículo para negocio
- Motocicleta
- Bicicleta
- Aire acondicionado
- Ventilador
- Calentador eléctrico (o a gas)
- Computador y/o juegos electrónicos
- Vehículo Particular.

Cuadro 12

**FUNCIONES DEL DISCRIMINANTE RESULTANTES DE LA
ESTRATIFICACION PROPUESTA**

Variable	Estr. I	Estr. II	Estr. III	Estr. IV	Estr. V	Estr. VI
ANDEN1	3.245044	1.458627	0.890696	0.913604	1.219971	1.007776
ANDEN2	3.426651	3.051181	3.164702	3.729491	3.120589	3.303284
ANDEN3	0.648474	0.943216	1.149679	1.298675	1.227878	1.526374
ZONAS1	47.72011	47.26615	47.23255	47.69757	47.81358	47.78891
ZONAS2	49.65723	49.16942	49.16788	49.61490	49.45543	48.66502
ZONAS3	48.39024	47.94236	48.14069	47.59789	47.84071	46.68748
ZONAS4	54.07482	53.77012	53.50413	53.66077	53.47733	53.02897
VIAS1	1.110721	-0.73842	-1.43673	-1.88395	-1.57643	-1.68015
VIAS2	-2.69424	-1.43671	-0.64234	-0.69109	-0.31181	-0.58148
VIAS3	1.145416	1.097191	1.068556	1.012060	1.241368	1.100141
CON-LOC1	1.475478	0.660221	0.099048	0.466531	-0.07898	0.099766
CON-LOC2	1.239349	0.942504	0.118652	-0.14005	-0.29872	-0.16177
CON-LOC3	-0.17744	0.196695	0.320014	1.646180	1.183310	1.138505
CON-LOC4	-1.85384	-1.70558	-1.93285	-2.52521	-2.78828	-2.74490
CON-LOC5	3.843307	3.468581	4.063884	3.894873	3.830240	4.152998
TIPO-VI1	68.61286	68.96016	68.48638	68.86047	68.12622	65.10627
TIPO-VI2	59.38986	59.85838	58.91694	58.88612	60.14755	57.96224
TIPO-VI3	70.67670	71.52959	70.53381	70.00426	70.47124	72.84147
COND-VI1	116.4679	117.2374	118.9926	117.6091	117.9398	117.4238
COND-VI2	111.1922	113.3071	113.8669	114.0312	113.0387	112.7831
COND-VI3	114.3845	112.2961	111.4837	110.7658	110.5570	110.3541
COND-VI5	107.5361	110.5024	110.9135	110.3550	109.8263	109.6373
SIT-CAL1	1.328870	0.037443	-0.83505	-0.73058	-0.58193	-0.32985
SIT-CAL2	0.854112	2.549294	3.357495	4.099011	3.524752	3.474290
SIT-CAL3	2.506731	2.840923	3.375645	2.595294	2.696407	2.461859
SIT-CAL4	3.406765	4.695174	4.828263	3.968985	3.971513	4.52645
UB-VIA1	113.3502	112.4832	114.0511	113.5289	113.0523	113.1228
UB-VIA2	114.5520	114.9450	113.9488	114.2222	113.6386	117.3674
UB-AV1	7.066696	8.691653	7.401518	7.555918	7.259099	7.279057
MEX1	21.78740	11.16078	10.94463	11.64716	11.65161	9.108227
MEX2	6.984954	6.950674	7.401609	7.274454	7.334389	4.986939
MEX3	6.596462	7.605370	7.317086	7.814048	8.305212	6.002397
MEX4	7.372140	7.883665	9.798465	9.428449	8.879973	6.829791
MEX5	7.206491	7.599201	7.150947	8.479209	7.720277	5.292077
MEX6	8.299216	8.534851	8.712615	8.488545	8.879234	7.069317
MPU1	67.76694	68.69832	68.08568	69.27485	67.83225	68.43523

(Sigue)

(Continuación Cuadro 12)

Variable	Estr. I	Estr. II	Estr. III	Estr. IV	Estr. V	Estr. VI
MPU2	68.59973	68.67541	68.36165	68.87508	68.32649	70.73353
MPU3	82.52716	83.45727	82.59287	83.16709	82.43399	91.34156
MPU4	93.38709	94.34455	93.76054	94.20809	95.12827	103.1269
MPU5	69.83960	70.37543	68.94392	69.43297	68.37200	70.06911
MPU6	74.83563	67.39174	66.22036	66.56818	65.85282	66.97389
MPIS1	243.8350	231.7046	230.2254	230.3914	231.0962	229.8703
MPIS2	232.1495	234.9888	230.6840	231.3775	232.5847	231.4897
MPIS3	229.9251	232.9489	232.9231	232.5196	232.5020	232.6401
MPIS4	230.1097	231.1639	231.8839	230.8393	231.8520	230.4921
ARLOT1	10.48669	10.19407	11.29456	9.993333	10.01210	9.337528
ARLOT2	9.053684	10.15939	8.933105	8.541558	8.743333	8.925647
ARLOT3	11.99034	12.42479	12.30621	13.24364	12.29038	11.65764
ARLOT4	10.76775	10.82661	11.57718	11.39103	12.03762	10.64800
ARLOT5	13.13527	12.44731	12.61154	12.30703	14.18144	12.95084
ARCOS1	10.87324	10.17698	10.01021	9.208464	8.886326	9.093728
ARCOS2	10.69262	10.73428	10.07539	9.447989	9.086562	9.481991
ARCOS3	8.512607	8.942627	8.720964	10.24187	8.516184	8.875139
ARCOS4	6.496788	6.537460	7.098567	7.111192	7.287000	6.465355
ARCOS5	3.64010	3.141875	3.111071	3.149019	4.794638	2.691210
GARAJ1	27.25205	28.13674	29.56491	27.29226	26.49677	26.95042
GARAJ2	32.46420	32.91147	33.30860	33.17294	35.44780	31.41177
GARAJ3	26.82986	27.50838	27.65337	28.77467	29.00117	26.34723
MTBAN1	0.779701	1.245343	-0.09044	-0.20047	-0.45898	0.192747
PISCINA	24.33598	22.98612	23.38114	23.09727	23.29506	26.50085
MCOC1	3.127336	1.270142	-1.13838	-0.74866	-1.90602	-1.49096
MCOC2	-1.82954	2.707245	0.612977	0.817242	0.248779	0.228939
MCOC3	-0.65457	-0.87926	-1.24724	-1.74901	-2.64183	-2.38175
MCOC4	1.082244	0.655312	1.406724	1.772436	-0.06792	-0.31061
COMB1	23.28462	20.79546	22.52609	21.92930	19.86195	21.66996
COMB2	19.79350	17.71960	18.98795	20.94802	15.21565	15.91855
COMB3	21.83117	18.37730	19.41233	18.86780	15.91795	18.12873
COMB4	24.42107	16.58751	17.42357	17.00507	14.83775	16.29448
ESTGAS	1.514477	1.848396	2.178277	2.939934	2.889745	2.062589
ESTELE	-1.39176	-0.03797	0.220121	-0.11984	0.035245	-0.68382
ESTGSO	3.058980	1.804242	2.095701	1.553076	2.155341	2.15690
ESTCAR	1.625926	2.231139	3.442630	3.299749	2.936695	2.696009
LICUADO	1.628223	1.104305	1.653428	1.314016	1.149950	1.069086
AUXCOC	2.356015	1.906636	1.957360	1.732726	3.973328	2.139938
ASPIRAD	0.618909	0.733825	0.343822	0.064491	1.024083	9.389677
BRILLAD	-3.41894	-3.08909	-3.29537	-3.52472	-2.90744	-0.23223

(Sigue)

(Continuación Cuadro 12)

Variable	Estr. I	Estr. II	Estr. III	Estr. IV	Estr. V	Estr. VI
NEVERA	-1.00039	0.725836	1.209354	1.026194	0.252054	0.950362
LAVADORA	1.071214	0.640348	0.537443	0.570925	2.663672	3.548205
SECADORA	-2.77922	-2.77479	-2.98045	-2.75428	-2.45885	3.082963
LAVPLA	1.760149	1.521293	1.284187	0.922202	2.037239	1.912087
HORNELC	0.718068	0.475682	0.788186	0.813161	1.378197	4.343300
HORNOMI	-14.0375	-13.4415	-13.7559	-14.2912	-14.9405	-4.09021
TELEVIS	-2.78558	-0.21638	0.743112	0.758750	-0.24365	0.08939
BETAMAX	1.021015	0.724383	0.368521	2.117104	1.287062	1.800570
EQ-SON	-0.20772	0.611777	0.045534	1.182579	0.223564	0.168359
VEH-NEG	2.065022	1.946497	1.939223	2.068768	1.940404	1.048659
MOTO	0.291207	0.460578	-0.43455	-0.13375	-0.12545	0.984305
BICICLE	-0.50341	-0.03751	-0.17310	-0.56093	-0.33125	-1.40493
AIRE-ACO	-0.310187	-0.11314	0.071428	-0.07320	-0.16945	3.051312
VENTILA	-1.20192	-0.94160	-1.48438	-1.30670	-0.10771	0.325032
CALENT	6.148994	6.300736	5.921254	5.390627	7.588184	13.15212
COMPUT	1.298084	0.976994	0.983849	0.932475	1.222893	7.602974
VEH-PAR1	0.879279	1.126576	0.722575	2.253477	1.523251	3.682776
CONST	-373.428	-367.062	-368.261	-368.089	-361.866	-372.473

La distribución porcentual de la nueva estratificación es sustancialmente diferente de la que actualmente existe para la ciudad de Cali. Actualmente un 61.5% de la población se concentra en los estratos socioeconómicos I y II (el estrato I con 27.9%, y el II con 33.6%) en contraste los estratos IV y VI (que agrupan en conjunto tan sólo un 7.3% de la población total), el estrato IV (con 3.7%), el estrato VI (con 3.6%), el estrato III (20.8%) y el V (10.4%) (Gráfico 3).

La estratificación obtenida concentra la mayoría de la población en el estrato socioeconómico III (30.6%), siendo seguido por el estrato II con 25.9%, el estrato IV con 18.1%, el estrato V con 13.4%; los estratos que menos viviendas agrupan son el I y el VI, con 6.8% y 5.2% respectivamente del total de las viviendas consideradas en el análisis (Gráfico 4).

Realizando un cruce entre la estratificación escogida y la estratificación actual de la ciudad se logró observar los porcentajes de viviendas que permanecieron en el mismo nivel de estrato y aquellos porcentajes de las viviendas que cambiaron de estrato (Cuadro 13).

Gráfico 3

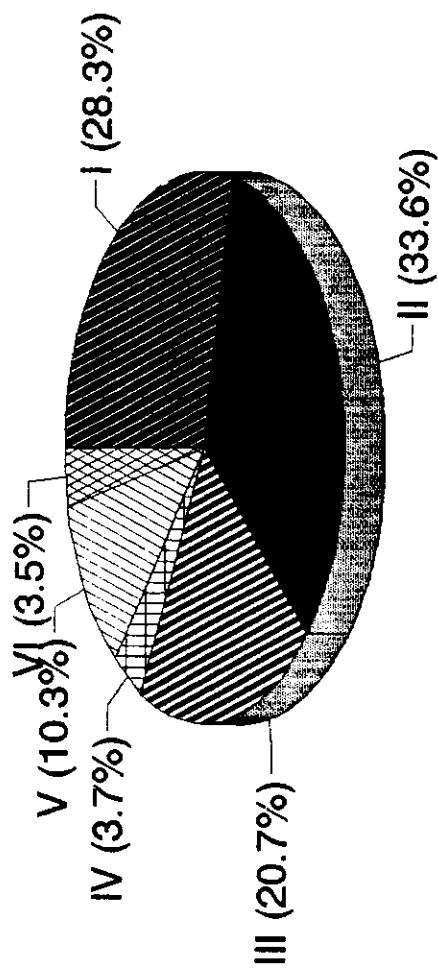
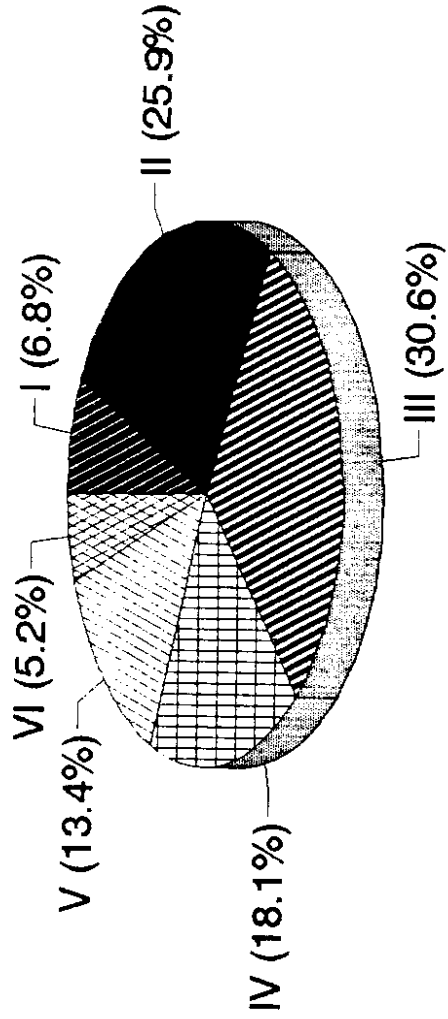
**DISTRIBUCION PORCENTUAL ESTRATIFICACION ACTUAL
DE LA CIUDAD DE CALI**

Gráfico 4

DISTRIBUCION PORCENTUAL ESTRATIFICACION
PROPUESTA PARA LA CIUDAD DE CALI



Cuadro 13

RELACION ESTRATO ACTUAL - ESTRATIFICACION PROPUESTA

Estrato	Estratificación Propuesta						Total
	I	II	III	IV	V	VI	
I	15.9	57.0	17.0	7.9	1.9	0.3	3053 28.3
II	6.4	25.1	38.9	22.0	0.7	0.6	3624 33.6
III	0.5	6.2	49.4	26.1	15.7	2.1	2239 20.7
IV	0.3	1.3	32.4	27.3	30.4	8.1	395 3.7
V	0.5	0.2	10.7	16.7	46.3	25.6	1107 10.3
VI	0.0	0.8	5.8	9.7	40.0	43.7	380 3.5
Total	737 6.8	2798 25.9	3303 30.6	1952 18.1	1449 13.4	558 5.2	100.0

De las 3052 viviendas clasificadas actualmente en el estrato I, el 15.9%, se mantuvo en este mismo nivel de estrato, el 57.0%, pasó al estrato II, el 17.0% al estrato III, el 7.9% al estrato IV, el 1.9% al estrato V y el 0.3% al estrato VI. El nivel de variaciones en este estrato se explica por el alto nivel de viviendas estratificadas actualmente en el estrato I (28.3%); sin embargo, el mayor porcentaje de las viviendas cambian al estrato II.

De las viviendas clasificadas actualmente en el estrato II, se conservan en él mismo el 25.1%, mientras que el 6.4%, se clasifica según la estratificación propuesta en el estrato I, el 38.9% en el estrato III, el 22.0% en el estrato IV, el 7.0% en el estrato V y el 0.6% en el estrato VI. Es decir el 6.4% de las viviendas clasificadas actualmente en este estrato disminuye de estrato, y el 68.5% se ubica en niveles de estrato más altos.

El 49.8% de las viviendas que actualmente están clasificadas en el estrato III se mantienen en el mismo con base en la estratificación propuesta, mientras que el 0.5% pasó al estrato I, el 6.2% al estrato II, el 26.1% al estrato IV, el 15.7% al estrato V y el 2.1% al estrato VI.

De la 395 viviendas clasificadas actualmente en el estrato IV, el 32.5% permanece en este estrato, el 0.3% pasa al estrato I, el 1.3% pasa al estrato II, el 27.3% al estrato V y el 8.1% al estrato VI. De las viviendas de este estrato un 34.0% baja de estrato socioeconómico, mientras un 38.5% de las viviendas aumenta de estrato.

Para el estrato V, el 46.3% de las viviendas no cambia de estrato socioeconómico; un 0.5% se ubica ahora en el estrato I, el 0.2% en el estrato II, el 10.7% en el estrato III, el 16.7% en el estrato IV y el 25.6% en el estrato VI. Así, de las viviendas clasificadas en este estrato, baja en el nivel de estrato el 28.1%, mientras que el 25.6% asciende.

De los 380 casos clasificados en el estrato VI, permanecen en el mismo el 43.7% de los casos, el 0.8% se clasifica en el estrato II, el 5.8% en el estrato III, el 9.7% en el estrato IV y el 40% en el estrato V. Es decir que el 57.3% de las viviendas localizadas en este estrato disminuyen de estrato socioeconómico.

Así mismo, considerando las variaciones totales de la estratificación actual respecto a la propuesta, el 30.5% de los casos no varían de estrato socioeconómico. De otro lado, el 45.9% se mueve un estrato: el 7.8% desciende mientras el 38.1% sube. El 17.5% de los casos se mueve dos estratos, de los cuales el 1.6% bajó, mientras el 15.9% subió. Finalmente, un 4.9% varió tres estratos descendiendo el 0.2% y ascendiendo el 4.7% (Cuadro 14).

Cuadro 14

**VARIACIONES DE LA ESTRATIFICACION PROPUESTA CON
RELACION A LA ESTRATIFICACION ACTUAL**

Variación	Casos	Porcentaje
-4	5	0.0
-3	22	0.2
-2	174	1.6
-1	844	7.8
0	3.314	30.5
+1	4.135	38.1
+2	1.730	15.9
+3	507	4.7
+4	63	0.6
+5	3	0.0

Las viviendas con variaciones de 4 y 5 estratos fueron bastante bajas, tanto para aquellas que bajaron de estrato como para aquellas que subieron de estrato.

V. Conclusiones

El propósito fundamental de este trabajo estaba centrado en tratar de realizar una primera aproximación al concepto de estratificación socioeconómica, buscando con ello establecer un marco de referencia sobre el cual poder desarrollar una metodología que permitiera adelantar procesos de estimación.

Las condiciones de trabajo de esta metodología debían permitir la identificación de los elementos necesarios para aportar grados de confiabilidad estadísticamente aceptables, que sirvieran además como modelos de estimación en cualquier agrupación social, de forma tal que se convirtiera en un instrumento eficaz con el cual aplicar medidas de política con altas probabilidades de éxito.

Se partió del concepto explorado por la visión sociológica del estructural-funcionalismo, donde por definición las agrupaciones sociales se encuentran conformadas en diferentes niveles o estratos socioeconómicos, y donde los individuos, por su estilo de vida, su capacidad adquisitiva, las relaciones que sostienen y otros tantos elementos, se reconocen como pertenecientes a ese nivel determinado.

De acuerdo con esto, la tarea se centraba entonces en tratar de identificar las características comunes a cada uno de esos estratos, que permitieran la diferenciación entre cada uno de ellos, logrando así simultáneamente la determinación de la importancia relativa de cada una de esas características en cada estrato.

Para tratar de acercarse a esta identificación era necesario minimizar el efecto de posibles acciones subjetivas, inevitables en cualquier proceso de análisis, con el fin de que la radiografía que se lograra obtener de la comunidad observada se aproximara lo máximo posible a lo que en la realidad se presentaba.

Aunque limitaciones de orden práctico y logístico impedían el mantenimiento total de este objetivo, se logró que de un total de 55 criterios de aceptación general, como posibles determinantes del proceso de discriminación en seis categorías, las técnicas de estimación escogieran 22 de ellos como suficientes para alcanzar parte de los propósitos buscados.

Además de lograr que fuera la comunidad observada, mediante el proceso de estimación, la que determinara cuáles eran las características que establecían la diferencia entre los estratos socioeconómicos, se pudo establecer, también en forma automática, la determinación de la importancia relativa que esas características tienen para cada estrato, logrando con ello llegar a establecer el modelo que permite reproducir el proceso de estratificación en esa comunidad analizada.

Lo que se pretende es que, partiendo de una realidad consistente en que una agrupación social posee en su interior las relaciones de clase que permiten diferenciar y reconocer la pertenencia a distintos estratos, ella no tiene por qué ser igual a cualquier otra, por lo que mal podría entonces pensarse que tanto los criterios como las ponderaciones que arroja el análisis de una comunidad pueden ser utilizados en otras.

Al observar las variables que el proceso arrojó como relevantes, cuyos niveles de explicación alcanzaron un 74.3% (aceptable desde el punto de vista estadístico), se logró establecer que incorporando grupos de variables adicionales se pueden alcanzar mayores grados de confiabilidad al momento de ser utilizada como instrumento de política, quedando aún bastante terreno por explorar en materia de los criterios que mejor ayudan a explicar las condiciones de la estratificación socioeconómica.

Dentro de los supuestos establecidos se había determinado que debería existir una relación directa entre el nivel del estrato y la distribución del ingreso, es decir que la pertenencia a un estrato socioeconómico más alto también implicaba el acceso a mayores niveles de ingreso. Así, los diferentes ejercicios de exploración en búsqueda de mejores niveles de explicación, exigían además que se realizara la prueba que permitiera la contrastación entre ellos y la distribución del ingreso, llegando al momento en que para el caso en que se obtuvo un modelo con el 74.3% de explicación, se alcanzaba también una clara diferenciación en términos de los ingresos de cada uno de los estratos en que se había dividido la comunidad analizada, no sucediendo igual para otro modelo donde se logró un nivel del 84%.

Esta evidencia permite entonces pensar que no necesariamente las variables obtenidas por el proceso de agrupamiento garantizan la relación con el ingreso, siendo entonces necesario adelantar las pruebas adicionales que permitan superar esta situación.

Bibliografía

- Amemiya, Takesh (1985). *Advanced Econometrics*. Basil Blackwell.
- Hartigan, John A. (1975). *Clustering Algorithms*. John Wiley and Sons.
- Judje, G. et.al (1985). *The Theory and Practice of Econometrics*. John Wiley and Sons.
- Maddala, G.S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, London.