

ESCUELA POLITÉCNICA NACIONAL

**FACULTAS CIENCIAS
CARRERA DE INGENIERÍA MATEMÁTICA**

**EL ESCALAMIENTO ÓPTIMO CON BASE EN EL ANÁLISIS DE
COMPONENTES PRINCIPALES NO LINEALES PARA LA
CONSTRUCCIÓN DE ÍNDICES DE CONDICIONES DE VIDA Y
SOCIOECONÓMICOS.
APLICACIÓN EN EL ÁMBITO NACIONAL.**

**PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO
MATEMÁTICO**

JESÚS ELOY TAPIA LÓPEZ

DIRECTOR: MSc FERNANDO CARRASCO

Quito, agosto 2007

DECLARACIÓN

Yo, Jesús Eloy Tapia López, declaro que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional, puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

Jesús Eloy Tapia López

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Jesús Eloy Tapia López, bajo mi supervisión.

MSc Fernando Carrasco
DIRECTOR DE PROYECTO

AGRADECIMIENTOS

Deseo expresar mi más profundo agradecimiento al Msc. Fernando Carrasco no solo por la gran colaboración, dedicación y apoyo en este trabajo; sino también por la oportunidad brindada para desarrollar las habilidades adquiridas en el transcurso de mi carrera, ayudándome a descubrir el verdadero potencial que poseemos.

Quiero agradecer además a todos los profesores que contribuyeron con mi formación y a la empresa Habitus Investigaciones S.A. que me brindó todas las facilidades para el desarrollo de este proyecto.

DEDICATORIA

Primero a Dios por ser esa fuerza invisible que me ayudo a culminar mis estudios, a mi madre por su amor y apoyo incondicional, a mi padre por su amistad sin reparos, a Carlos mi camarada, a Edison por enseñarme a ver más allá de lo que los ojos ven y a Isabel por su amor y paciencia infinita.

A mis amigos y amigas por ser parte de lo bueno y lo malo que hay en mí, sin ustedes el camino hubiera sido bastante gris.

Y finalmente al resto de mi familia por haber creído en mi.

CONTENIDO

Declaración.....	II
Certificación.....	III
Agradecimiento.....	IV
Dedicatoria.....	V
Contenido.....	VI
Resumen.....	IX
Presentación.....	X
Capítulo 1: Escalamiento Óptimo.....	1
1.1 Introducción.....	1
1.1.1 Reunión y Encuentro (Join and Meet).....	2
1.1.2 Dimensionalidad y Manejo de Datos.....	5
1.2 Técnicas Principales de Aplicación.....	6
1.2.1 Regresión Categórica.....	6
1.2.2 Análisis de Homogeneidad o de Correspondencias Múltiples.....	6
1.2.3 Escalamiento Multidimensional.....	8
1.2.4 Análisis de Correlación Canónica.....	8
Capítulo 2: Análisis de Componentes Principales no Lineal.....	10
2.1 Introducción.....	10
2.2 Algoritmo de Optimización.....	12
2.2.1 Algoritmo de Mínimos Cuadrados Alternantes.....	12
2.2.2 Algoritmo del Análisis de Componentes Principales no lineal.....	17
2.2.3 Algoritmo Resolutivo para el Método PRINCALS.....	22
2.2.4 Manejo de casos simples.....	25
2.2.5 Manejo de casos perdidos.....	26
2.2.6 Manejo de variables con nivel de medidas mixtas.....	26

2.2.6.1 Primer Método.....	27
2.2.6.2 Segundo Método.....	28
2.2.7 Generalización a Múltiples Conjuntos de variables.....	30
2.3 Resultados.....	32
2.4 Procedimiento de Construcción de Índices.....	34
2.4.1 Ejemplo.....	37
2.4.1.1 Construcción inversa del Índice.....	44
Capítulo 3: Investigación sobre aplicaciones en la construcción de índices con base en las técnicas de Escalamiento Óptimo.....	50
3.1. El caso SELBEN para distribución del bono de desarrollo humano.....	51
3.2 Otras Aplicaciones.....	57
3.2.1 Índice de Participación Ciudadana.....	57
3.2.2 Tipología de la intervención social en el Ecuador.....	59
3.2.3 Escala de inhibición para preescolares.....	60
CAPÍTULO 4: Aplicación del Análisis de Componentes Principales no lineal para la construcción de Índices de condiciones de vida.....	62
4.1. Aplicación con el Censo de Población y Vivienda del 2001.....	62
4.1.1 Selección de la muestra.....	63
4.1.2 Construcción del Modelo.....	74
4.1.3 Implementación del modelo construido a partir del Censo de Población y Vivienda 2001.....	90
4.2 Aplicación con la Encuesta de Condiciones de vida 2005 – 2006.....	92

4.2.1 Construcción del Modelo.....	95
Capítulo 5: Construcción de un Índice con base en el Análisis de Componentes Principales Lineal.....	101
Capítulo 6: Conclusiones y Recomendaciones.....	107
6.1 Conclusiones.....	107
6.1 Recomendaciones.....	111
Capítulo 7: Bibliografía.....	112
Anexo 1: Sintaxis usada en el paquete SPSS para el método PRINCALS.....	113
Anexo 2: Diccionario de variables.....	116
Anexo 3: Cuadros Descriptivos de las variables del Censo de Población y Vivienda 2001.....	122
Anexo 4: Cuadros Descriptivos de las variables de la Encuesta de Condiciones de vida 2005 – 2006.....	138

RESUMEN

En este trabajo se realiza un análisis de las variables que describen las condiciones de vida, tenencia de bienes y acceso a servicios del Censo de Población y Vivienda 2001; y de las de la Encuesta de Condiciones de Vida 2005-2006, a través de la construcción de indicadores; usando el Análisis de Componentes Principales no lineal, los índices sintetizan la información de un conjunto de variables en una sola medida.

Se determinó que un indicador que mide necesidades o carencias es el complemento de otro indicador que mide satisfacción o acceso a bienes o servicios.

Las cuantificaciones óptimas calculadas sirven como línea de base para realizar comparaciones con distintas fuentes de información, en este caso se comparan los valores del Censo de Población y Vivienda 2001, con los de la Encuesta de Vida 2005 - 2006.

Además los resultados del Análisis de Componentes Principales no lineal se comparan con los obtenidos usando el Análisis de Componentes Principales lineal para el caso del Censo del 2001 y se encontró que los indicadores construidos con el Análisis de Componentes Principales no lineal discriminan mejor que un índice elaborado en base al análisis clásico.

INTRODUCCIÓN

El presente trabajo describe el Escalamiento Óptimo, más específicamente el Análisis de Componentes Principales no lineal y realiza una aplicación del método en cuestión con los datos del Censo de Población y Vivienda del 2001 y con los de la Encuesta de Condiciones de vida 2005-2006.

El proyecto surge con la necesidad de crear mecanismos (índices) que discriminen a los hogares y a los individuos de la población por sus características de condición de vida, o por la tenencia de bienes y acceso a servicios, dado que realizar una segmentación por la variable ingreso puede llevar a errores sobre todo por su poca confiabilidad, debido que en general las personas no son muy sinceras al momento de proporcionar dicha información.

Los índices creados serán una fuerte herramienta en el análisis de la población ecuatoriana en campos como la economía, sociología y la investigación de mercados ya que en general la gente tiene ciertos parámetros de comportamiento con respecto a cada segmento socioeconómico en los campos anteriormente mencionados.

Las variables cualitativas de condiciones de vida en general son de más fácil acceso debido a que, por ejemplo, en el Censo de Población y Vivienda del 2001, el registro de bienes y servicios al que tiene acceso un hogar y consecuentemente los miembros del mismo podría hacerse inclusive por simple observación del investigador; otro tipo de variables como el nivel educativo de los miembros del hogar y otras variables que miden las cualidades de un hogar son más confiables que las variables ingreso y gasto, lo que para una segmentación de la población sería importante contar esta información.

Una solución para el manejo de esta información de tipo cualitativa es asignar valores a cada una de las categorías de las variables, y observar que hogares o individuos tienen cualidades similares con respecto a un conjunto de variables, y esto

se puede realizar con un consenso de expertos y realizar las asignaciones mencionadas. Pues bien esto puede funcionar en un número reducido de individuos, pero ya dentro de un grupo de hogares o personas mucho más amplio y con la heterogeneidad que existe en un número grande de individuos este tipo de procedimientos pueden llevar a calificaciones subjetivas o erróneas de la población.

Es por esto que se utiliza la técnica de Análisis de Componentes Principales no lineal que trabaja con variables cualitativas asignando cuantificaciones numéricas a las categorías, realizando una transformación de dichas variables con la propiedad de tener la máxima correlación lineal entre las variables transformadas.

La metodología del Análisis de Componentes para la estimación de parámetros utiliza el método de Mínimos Cuadrados Alternantes; primero, cuantificando las categorías de las variables de modo de maximizar la correlación entre las variables y segundo, utilizando estas cuantificaciones para estimar los parámetros del análisis de componentes principales lineal.

Se presenta entonces una descripción del Escalamiento Óptimo y del método de los Mínimos Cuadrados Alternantes y se profundiza en el método de Análisis de Componentes Principales no lineal en su parte teórica, para luego pasar a una aplicación práctica del modelo con la construcción de índices, utilizando las bases de datos mencionadas anteriormente.

Cabe señalar que en la construcción de índices se consideran fundamentalmente características del jefe de hogar y variables de acceso a servicios del hogar. Además se incluye la ubicación geográfica del hogar, dada por su ubicación y sea en el área urbana o rural.

Adicionalmente se presenta un análisis que determinará si un índice que mide las necesidades básicas insatisfechas es o no el complemento de otro indicador que mide en cambio las necesidades básicas satisfechas. Se resalta este punto debido a

la controversia que se ha dado en esto, puesto que utilizando otros procedimientos se han dado resultados contradictorios.

Con respecto a la implementación del modelo, se utiliza el programa estadístico SPSS, el cual usa el algoritmo llamado PRINCALS desarrollado por el departamento de Teoría de Datos de la Universidad de Leiden en Holanda, el cuál nos entrega las cuantificaciones o ponderaciones óptimas de cada categoría dentro de cada variable y teniendo además la posibilidad de transformar dichas cuantificaciones a la escala que sea más útil, por ejemplo de cero a cierto valor dentro de cada variable, o de cero a cien utilizando en conjunto las ponderaciones de todas las variables usadas en el análisis.

CAPÍTULO 1.

ESCALAMIENTO ÓPTIMO

1.1 INTRODUCCIÓN

Con la necesidad de trabajar con variables cuantitativas se han desarrollado un conjunto de técnicas basadas en lo que se ha dado por llamar “Escalaiento Óptimo”, que se basa en la asignación de cuantificaciones numéricas a cada una de las categorías de las variables.

Las cuantificaciones se obtienen por el procedimiento de los mínimos cuadrados alternantes, con base en un proceso iterativo, utilizando las asignaciones previas para estimar las siguientes, hasta converger a la solución.

El proceso iterativo de los mínimos cuadrados alternantes, se realiza en dos pasos, en el primer paso se calcula una base óptima para los valores dados de la transformación, en el segundo paso se calculan nuevos valores para las transformaciones óptimas para la base calculada en el primer paso.

Este proceso que en cada iteración consta de los dos pasos mencionados anteriormente se conoce como Escalamiento Óptimo, porque las transformaciones son escogidas de tal manera que estas minimizan la función de pérdida.

En el sentido del análisis no lineal si contamos con un modelo basado en un conjunto finito de desigualdades o sea el caso de un número finito de ecuaciones que deben cumplirse para el modelo, entonces expresamos la pérdida midiendo el incumplimiento de un determinado número de inecuaciones o ecuaciones y el objeto de minimizar la pérdida es reducir el número de violaciones o no cumplimientos en determinado procedimiento.

Supongamos que contamos con una matriz de datos $n \times m$ y que las m variables son divididas en k subconjuntos, llamamos a la intersección de los k -subconjuntos

como el encuentro (**meet**) y a la suma lineal que pertenece al espacio de todas las combinaciones lineales de k-vectores reunión (**join**); entonces la función de pérdida mide la falta de ajuste tanto de la reunión y del encuentro.

1.1.1 REUNIÓN Y ENCUESTRO (JOIN AND MEET)

Supongamos que \mathcal{W} es el conjunto de todos los subespacios de \mathbb{R}^n nosotros podemos ordenar parcialmente \mathcal{W} diciendo que $W_1 \leq W_2$, si W_1 es un subespacio de W_2 . Nosotros podemos decir que \mathcal{W} es completo a través de la definición de **MEET**(W_1, W_2) que es la intersección de W_1 y W_2 y de **JOIN**(W_1, W_2) que es la suma lineal de W_1 y W_2 (es decir que para todo $x \in \mathbb{R}^n$ de la forma $x=y+z$ con $y \in W_1$ y con $z \in W_2$).

El encuentro (meet) de W_1 y W_2 es el subespacio más grande que contiene a ambos conjuntos y la reunión (join) de W_1 y W_2 es el subespacio más pequeño que contiene a los dos conjuntos mencionados.

Ahora bien si W_1, \dots, W_m son subespacios de \mathbb{R}^n podemos también definir **MEET**(W_1, \dots, W_m) y **JOIN**(W_1, \dots, W_m), se puede también definir el rango de encuentro (**MRK**) de W_1, \dots, W_m a través de la dimensión (DIM) del subespacio conformado por la intersección de W_1, \dots, W_m es decir:

$$MRK(W_1, \dots, W_m) \equiv DIM(MEET(W_1, \dots, W_m)) \quad 1.1.1.1$$

y consecuentemente podemos notar como el rango de reunión (**JRK**) a través de:

$$JRK(W_1, \dots, W_m) \equiv DIM(JOIN(W_1, \dots, W_m)) \quad 1.1.1.2$$

Diremos que W_1, \dots, W_m tiene **p-encuentros (p-meet)** si $MRK(W_1, \dots, W_m) \geq p$ y que W_1, \dots, W_m tiene **p-reuniones (p-join)** si $JRK(W_1, \dots, W_m) \leq p$; una característica muy útil de estas definiciones es que no dependen de la base que se tome en \mathbb{R}^n .

El **problema de encuentro (meet problem)** para el caso cualitativo es encontrar un subespacio W de \mathbb{R}^n tal que $DIM(W)=p$ y $W=MEET(W_1, \dots, W_m)$; así dados

W_1, \dots, W_m y dado un entero p con $1 \leq p \leq n$ se trata de encontrar un subespacio de dimensión p que aproximadamente contiene a todos los W_j .

Antes de introducir el término de función de pérdida, primeramente reformularemos nuestro problema en forma matricial, así supongamos que W_j es el subespacio de todas las combinaciones lineales de una matriz A_j la cual no es necesariamente la matriz indicador, entonces W_1, \dots, W_m tienen p -encuentros si y solo si existe una matriz $X_{n \times p}$ de rango p y matrices Y_j de dimensión $(k_j \times p)$ tal que:

$$X = A_1 Y_1 = \dots = A_m Y_m \quad 1.1.1.3$$

Más aun, W_1, \dots, W_m tiene p -reuniones si y solo si existe una matriz $X_{n \times p}$ y matrices Y_j de dimensión $(k_j \times p)$ tal que:

$$A_j = X Y_j' \quad \text{para todo } j = 1, \dots, m \quad 1.1.1.4$$

Si acoplamos las A_j en una matriz $(n \times \sum k_j)$ particionada A y los Y_j en una matriz Y de dimensión $\sum k_j \times p$; así podemos escribir (1.1.1.4) como:

$$A = X Y' \quad 1.1.1.5$$

Donde A es la matriz formulación, ahora bien nótese que no se requiere que X tenga rango p , ya que como se vio anteriormente basta que $\text{RANK}(X) = q < p$, entonces W_1, \dots, W_m tiene una p -reunión.

A través de la formulación matricial se nos es posible definir las funciones de pérdida de mínimos cuadrados, donde la función de pérdida de encuentro está definida por:

$$\sigma_M(X; Y_1, \dots, Y_{m1}) = m^{-1} \sum_j SSQ(X - A_j Y_j) \quad 1.1.1.6$$

Los subespacios W_1, \dots, W_m tienen una p -reunión si y solo si el mínimo de la función de pérdida σ_M sobre cada Y_j y sobre todo X de rango p existe y es igual a cero; desde el punto de vista computacional esta condición no es útil, pero la forma más conveniente de cumplir con la condición mencionada, parte del hecho que W_1, \dots, W_m tiene un p -encuentro si y solo si el mínimo de σ_M sobre cada Y_j y sobre todo X satisface que $X'X = I$ existe, y dicho mínimo es igual a cero.

De la misma manera podemos definir la función de pérdida de reunión así:

$$\sigma_J(X; Y, \dots, Y_{m1}) = m^{-1} \sum_j SSQ(A_j - XY_j') \quad 1.1.1.7$$

Nótese que existen importantes diferencias con σ_M , en primer lugar el particionamiento de la matriz A en submatrices A_j es innecesario para σ_J , claramente no para σ_M .

De hecho podemos escribir:

$$\sigma_J(X; Y) = m^{-1} \sum_j SSQ(A - XY') \quad 1.1.1.8$$

La segunda diferencia está en que no se necesitan restricciones con respecto al rango de X, los subespacios tienen una p-reunión si y solo si el mínimo de σ_J sobre cada Y_j y sobre todo X es igual a cero; más aún este mínimo siempre se alcanzará y no siempre es igual a cero, otra forma de escribir la función de pérdida de reunión de la siguiente manera:

$$\sigma_J(X; Y) = m^{-1} \sum_l SSQ(a_l - Xy_l) \quad 1.1.1.9$$

Donde la sumatoria se realiza sobre todas las columnas de la matriz A, de esta forma reducimos el problema de reunión a un subespacio de una sola dimensión.

Las funciones de pérdida son todas del tipo de mínimos cuadrados y tienen la propiedad de que estas son igual a cero para un valor particular de los parámetros si y solo si este valor es una solución perfecta para la correspondiente función de pérdida.

Como se dijo anteriormente el procedimiento de mínimos cuadrados alternantes se realiza en dos pasos, en el primero se encuentra una base óptima para los valores dados de las transformaciones y en el segundo paso calculamos nuevos valores para las transformaciones óptimas para la base dada en el primer paso; al alternar estos pasos en cada iteración se produce un decrecimiento secuencial de

los valores de la función de pérdida, la cual siempre converge porque la función de pérdida está acotada inferiormente por cero.

Bajo algunas condiciones de regularidad, podemos además probar que las bases y transformaciones convergen a valores correspondientes con un valor estacionario de la función de pérdida.

Y esta forma particular de calcular las transformaciones es llamada Escalamiento Óptimo, dado que las transformaciones son escogidas de tal manera que estas minimicen la función de pérdida.

Claro esta, el término óptimo no debe ser interpretado en sentido amplio, es decir no queremos decir que este procedimiento sea mejor que otros, sino que este es el mejor procedimiento en términos de la función de pérdida que nosotros escogimos.

1.1.2 DIMENSIONALIDAD Y MANEJO DE DATOS

Obsérvese que el método tiene dos dimensiones de trabajo, nosotros hemos escogido el rango de reunión (join) y el de encuentro (meet) y claro esta, el número de diferentes soluciones que puede tener, esta selección depende del tipo de procedimiento que deseamos usar, por ejemplo, si usamos el método HOMALS en un problema de reunión e implica que se escoja un rango de reunión igual a uno y tenemos que decidir cuántas soluciones deseamos calcular; ahora si deseamos utilizar el método PRINCALS nosotros tenemos que buscar solo una solución, pero tenemos que decidir que rango de reunión queremos usar.

En el presente trabajo describiremos el método PRINCALS (“Análisis de Componentes Principales no Lineal”) y utilizaremos rangos y/o dimensionalidades iguales a uno o dos, dado que principalmente en el estudio de los gráficos es más conveniente trabajar con máximo dos dimensiones, además que generalmente las primeras dos dimensiones nos brindan una clara idea de los efectos más importantes en los datos.

1.2 TÉCNICAS PRINCIPALES DE APLICACIÓN

Dentro de las principales técnicas de aplicación del Escalamiento Óptimo, tenemos:

- Regresión categórica
- Análisis de homogeneidad o de correspondencias múltiples
- Escalamiento multidimensional
- Análisis de correlación canónica no lineal, y
- Análisis de componentes principales no lineales.

1.2.1 REGRESIÓN CATEGÓRICA

Predice los valores de una variable de respuesta categórica o nominal ordenada o no a partir de una combinación de variables predictoras cualitativas. Regresiona las variables predictoras categóricas cuantificadas usando métodos de escalamiento óptimo de tal manera que estén máximamente relacionadas con variables respuestas numéricas, nominales y ordinales.

1.2.2 ANÁLISIS DE HOMOGENEIDAD O DE CORRESPONDENCIAS MÚLTIPLES

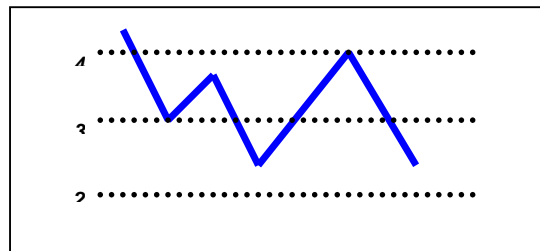
El análisis de homogeneidad en sentido más amplio se refiere a ciertos criterios para el análisis de datos multivariantes partiendo del hecho característico de optimización de homogeneidad de variables bajo varias formas de manipulación y simplificación.

Históricamente el término homogeneidad está estrictamente relacionada con la idea que varias variables pueden medir la misma cosa; suponiendo que esto es cierto la matriz de datos (suponiendo que las variables se encuentran normalizadas) pueden dar valores idénticos en cada columna o si dibujamos las observaciones como perfiles, cada uno de estos resultara en una línea horizontal.

Las variables miden la misma cosa pero con un error distinto, las columnas de la matriz de datos pueden tener elementos que varían de alguna manera.

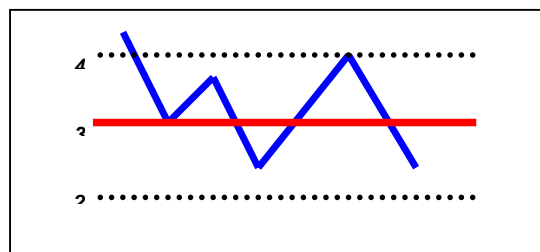
Un gráfico de los perfiles podría resultar en curvas en forma de zigzag en diferentes niveles (figura 1.2.2.1).

Figura 1.2.2.1



Reemplazando tales perfiles por una línea recta, que claro esta resulta en pérdida de información, la homogeneidad se da cuando al remplazar por una línea recta dichos perfiles la pérdida de información es mínima (figura 1.2.2.2):

Figura 1.2.2.2



1.2.3 ESCALAMIENTO MULTIDIMENSIONAL

Esta técnica es una generalización el Análisis de Componentes Principales, con la diferencia que en vez de contar de una matriz de individuos y de un grupo de variables se cuenta con una matriz de distancias o disimilitudes entre los elementos de un conjunto, es decir se cuenta con una matriz que mide similitudes o inversamente disimilitudes.

Estas distancias pueden obtenerse a partir de ciertas variables o como resultado de una estimación directa, por ejemplo consultando a un grupo de expertos.

Es decir a partir de la matriz de distancias se pretende obtener una matriz de un número determinado de variables para los individuos, básicamente el Escalamiento Multidimensional estudia las distancias en un grupo de individuos.

1.2.4 ANÁLISIS DE CORRELACIÓN CANÓNICA

El Análisis de Correlación Canónico estudia las relaciones entre dos grupos homogéneos, donde el primer grupo incluye p variables y el segundo q variables, con $p+q=k$.

Para esto consideremos que se tiene una matriz X de dimensión $(n \times p)$ que contiene las p variables, además consideremos una matriz Y de dimensión $(n \times q)$ que claro está contiene las q variables del otro grupo.

Para explicar en conjunto la relación entre los dos grupos de variables se estudia la matriz de covarianzas V con el total de las variables centradas; donde la matriz V puede ser simétrica donde no se da preferencia a ninguno de los grupos de variables para explicar el otro; o la matriz V puede ser asimétrica donde un conjunto de variables es utilizado para explicar el comportamiento del otro.

Para explicar la relación se considera que la matriz V es simétrica, para esto se busca un par de variables que resuman cada grupo y que estas tengan la máxima correlación, es decir se requiere encontrar:

$$x^* = X\alpha = \sum_{i=1}^p \alpha_i x_i \quad 1.2.4.1$$

$$y^* = Y\beta = \sum_{j=1}^q \beta_j y_j \quad 1.2.4.2$$

Se desea entonces encontrar los vectores α y β de tal manera que las nuevas variables x^* y y^* tengan como se dijo máxima correlación.

CAPÍTULO 2.

ANÁLISIS DE COMPONENTES PRINCIPALES NO LINEAL

2.1 INTRODUCCIÓN

El Análisis de Componentes Principales clásico reduce el conjunto original de variables a un conjunto más pequeño de componentes no correlacionados que capture la mayor parte de la información contenida en las variables originales.

Pero, que sucede cuando se cuenta con variables cualitativas o categóricas, por ejemplo, el nivel de aceptación de un producto, o si se necesita resumir un grupo de variables que califican (no cuantifican) para construir un índice que resuma las condiciones de acceso a bienes y servicios de una familia.

El análisis de las componentes principales no lineales, fue diseñado precisamente para dar una respuesta a esta última situación. La idea de este análisis es transformar las variables originales asignando valores a las categorías de cada una de las variables y luego correlacionarlas para caracterizar o analizar la estructura de los datos.

Este modelo multivariante se lo conoce también por los nombres de:

- (i) Análisis de componentes principales categóricos por escalamiento óptimo, o
- (ii) Análisis de componentes principales no métrico, o
- (iii) Análisis de componentes principales mediante mínimos cuadrados alternantes

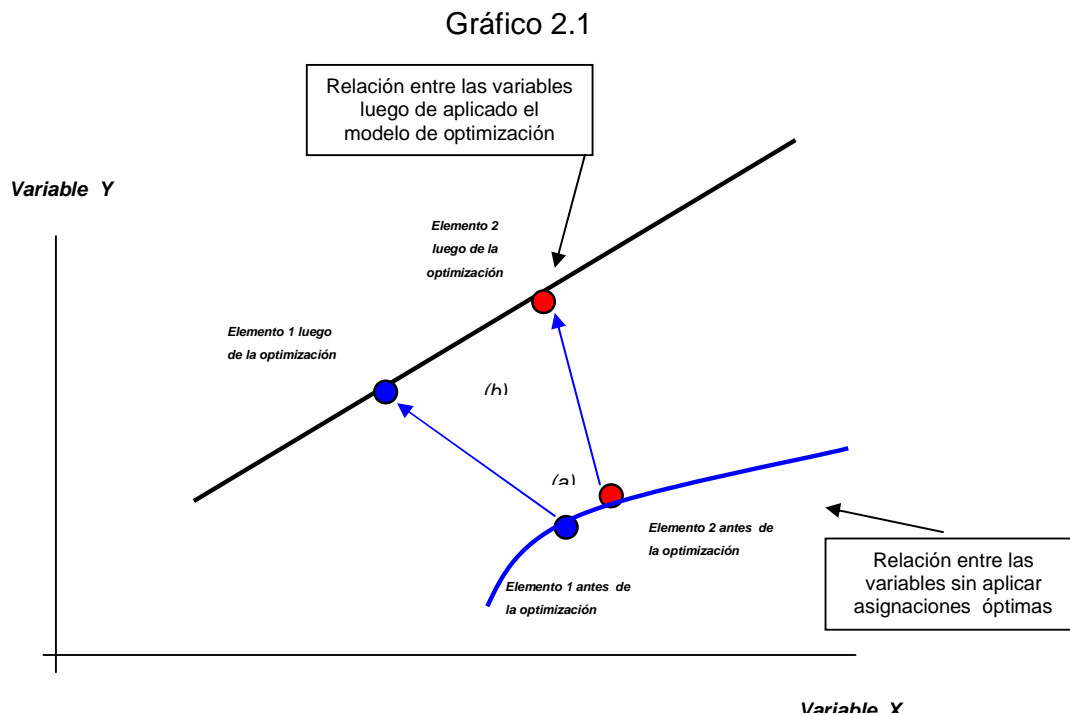
y, los investigadores sociales en nuestro medio, le han dado por llamar como el (iv) “método de asignaciones óptimas” o “método de ponderaciones óptimas”.

Con la valoración de las variables cualitativas por este procedimiento, se logra también maximizar la correlación lineal entre las variables dadas, consiguiendo

con esto transformar las variables cualitativas en variables cuantitativas, disponiendo así, de la mejor combinación lineal posible de las variables tratadas, sirviendo esto, por ejemplo, para la construcción de modelos de regresión (en donde interesa precisamente la linealidad), o poder discriminar o caracterizar mejor casos ambiguos, entre muchos otros análisis que se podrían efectuar.

En el gráfico 2.1 puede apreciarse la ventaja de valorar las variables, ya que generalmente al establecer un sistema de valoraciones o ponderaciones, sin ningún criterio técnico, se mantiene el problema de la subjetividad y también puede prestarse a manipulaciones.

Así, en la parte (a) del gráfico se muestra una relación común entre dos variables, que no puede discriminar adecuadamente a los elementos; mientras que aplicando el procedimiento de Análisis de Componentes Principales no lineal, se logra diferenciar mejor las condiciones entre los elementos, lo que se observa en la parte (b) del gráfico, logrando así la máxima correlación lineal y por ende todas las ventajas que ello implica en términos de sus usos y aplicaciones.



2.2 ALGORITMO DE OPTIMIZACIÓN

Sea una matriz H de dimensión (nxm), esta puede ser formulada en términos de su función de pérdida, la misma que se escribe de la siguiente manera:

$$\sigma_j(X, A) \equiv SSQ(H - XA') \quad 2.2.1$$

donde X es una matriz (nxp) de rango p, A es una matriz de dimensión (m xp) y SSQ(.) representa la suma de cuadrados de los elementos de la matriz (.).

Las matrices X y A que minimizan la SSQ pueden ser encontradas calculando la descomposición en valores singulares de H y reteniendo el **valor singular** p más grande con su correspondiente **vector singular**, o por el método de los mínimos cuadrados alternantes que se presenta a continuación.

2.2.1 ALGORITMO DE MÍNIMOS CUADRADOS ALTERNANTES

Supongamos que contamos con la siguiente función de pérdida:

$$\sigma(x, a) \equiv m^{-1} \sum_j SSQ(x - a_j h_j) = m^{-1} \sum_j tr(x - a_j h_j)' (x - a_j h_j) \quad 2.2.1.1$$

Entonces este algoritmo tiene dos pasos para determinar los valores óptimos x^* y pesos a^* , en el primero se minimiza la función de pérdida con respecto a x para un a fijo y en el otro paso se minimiza a para x cualquiera.

Se muestran dos variaciones de estos algoritmos que corresponden a las dos formas de normalización mencionadas en conexión con las soluciones triviales de la función (2.2.1.1), es decir para $x=0$ y $a=0$.

Supongamos que las columnas de la matriz de datos H han sido centradas y normalizada de tal manera que $H'H=R$, donde R es la matriz de correlaciones.

El algoritmo de normalización tiene que satisfacer $x'x=1$, además el algoritmo requiere de un valor inicial \tilde{a} (\tilde{a} distinto de cero), lo que garantiza que se descarte la solución trivial y se procede con los siguientes pasos:

- (1) *Se busca un valor de x :* $\tilde{x} \leftarrow H\tilde{a} / m$
- (2) *Normalización:* $x^* \leftarrow \tilde{x}(\tilde{x}'\tilde{x})^{-1/2}$
- (3) *Actualizamos los valores de a :* $a^* \leftarrow H'x^*$
- (4) *Test de convergencia:* *Se regresa al paso uno fijando $\tilde{a} \leftarrow a^*$, con tal que los valores de x^* y a^* no estén suficientemente estabilizados (de acuerdo a cierto criterio de exactitud).*

Pues bien el paso uno corresponde al cálculo no restringido de \tilde{x} , de la función (2.2.1.1) para un \tilde{a} cualquiera, donde $H\tilde{a}/m$ es el vector de filas de la matriz de columnas $\tilde{a}_j h_j$.

Los valores $\tilde{x} = H\tilde{a}/m$ por consiguiente minimizan la pérdida relativa W/T para la matriz recalculada H con pesos \tilde{a} .

En el segundo paso se calcula la proyección de \tilde{x} en la esfera de todos los vectores normalizados unitarios x ; esto transfiere la minimización que se realizó sin ninguna restricción a la región factible, que es la contiene todas las soluciones que satisface la restricción.

En el paso tres se calcula el peso a^* para cualquier x^* , dado que x^* y las columnas de H están centradas y normalizada a la unidad y a^* es un vector de correlaciones.

El algoritmo converge monótonamente desde el paso uno al dos¹, el paso tres siempre produce un pequeño valor de la función de pérdida el cuál está acotado inferiormente por el cero.

¹ La convergencia puede ser medida con el criterio dado por el método HOMALS.

Para minimizar la función de pérdida se utilizan los pasos mencionados en el algoritmo como se presenta a continuación²:

Primer paso: se encuentran los valores de x . Para esto minimizamos la expresión (2.2.1.1) con respecto a x manteniendo fijo el valor de a , es decir se tendría la siguiente ecuación normal:

$$mx = \sum_j h_j \tilde{a}_j \quad 2.2.1.2$$

de donde :

$$\tilde{x} = m^{-1} \sum_j h_j \tilde{a}_j \quad 2.2.1.3$$

Segundo paso: se normaliza el valor de \tilde{x} obteniendo la proyección x^* de dicho valor en la región factible a través de la expresión:

$$x^* = \tilde{x}(\tilde{x}'\tilde{x})^{-1/2} \quad 2.2.1.4$$

Tercer paso: se encuentra el valor de a minimizando con respecto a dicho parámetro manteniendo fijo x , con lo que se obtendría el conjunto de ecuaciones normales dadas por la expresión:

$$V_j a_j = h_j' x \quad 2.2.1.5$$

Donde $V_j = h_j' h_j$ es la matriz diagonal que contiene las frecuencias marginales de cada variable j . Por lo que se tendría que:

$$a^*_{j} = V_j^{-1} h_j' x \quad 2.2.1.6$$

Una vez que el algoritmo converge la función de pérdida (2.2.1.1) se puede escribir de la siguiente manera:

² Albert Gifi, Nonlinear Multivariate Analysis, John Wiley and Sons, 1990.

$$\sigma(x^*, a^*) \equiv m^{-1} \sum_j SSQ(x^* - a^*_j h_j) = m^{-1} \sum_j tr(x^* - a^*_j h_j)' (x^* - a^*_j h_j) \quad 2.2.1.1$$

donde:

$$\begin{aligned} (x^* - a^*_j h_j)' (x^* - a^*_j h_j) &= ((x^*)' - h'_j (a^*_j)') (x^* - a^*_j h_j) \\ (x^* - a^*_j h_j)' (x^* - a^*_j h_j) &= (x^*)' x^* - (x^*)' a^*_j h_j - h'_j (a^*_j)' x^* + h'_j (a^*_j)' a^*_j h_j \end{aligned}$$

Y como $(x^*)' a^*_j h_j = ((x^*)' a^*_j h_j)' = h'_j (a^*_j)' x^*$ se tiene que:

$$\begin{aligned} (x^* - a^*_j h_j)' (x^* - a^*_j h_j) &= (x^*)' x^* - 2h'_j (a^*_j)' x^* + h'_j (a^*_j)' a^*_j h_j \\ (x^* - a^*_j h_j)' (x^* - a^*_j h_j) &= (x^*)' x^* - 2h'_j (a^*_j)' x^* + (a^*_j)' V_j a^*_j \\ (x^* - a^*_j h_j)' (x^* - a^*_j h_j) &= (x^*)' x^* - (a^*_j)' V_j a^*_j \\ (x^* - a^*_j h_j)' (x^* - a^*_j h_j) &= nI_p - (a^*_j)' V_j a^*_j \\ m^{-1} \sum_j tr(x^* - a^*_j h_j)' (x^* - a^*_j h_j) &= m^{-1} \sum_j tr(nI_p - (a^*_j)' V_j a^*_j) \\ \sigma(x, a) &= np - m^{-1} \sum_j tr((a^*_j)' V_j a^*_j) \end{aligned} \quad 2.2.1.7$$

El término $\sum_j tr((a^*_j)' V_j a^*_j)$ se denomina el “ajuste” de la solución y la medida de discriminación de la variable j en una dimensión dada (d) (con $d=1, \dots, p$) es:

$$\mu_{jd}^2 = (a^*_j)' (d, \cdot) V_j a^*_j (\cdot, d) / n \quad 2.2.1.8$$

Donde

(d, \cdot) o (\cdot, d) se refieren a la d -ésima fila o columna de la matriz $(a^*_j)'$ o a^*_j respectivamente.

La medida de discriminación nos da el promedio de los cuadrados de las distancias de las cuantificaciones de las categorías al origen del espacio p -dimensional; cabe señalar que dichas distancias se encuentran ponderadas por las frecuencias marginales y de este modo la función de pérdida se puede escribir de la siguiente manera:

$$n\left(p - m^{-1} \sum_j \sum_d \mu_{jd}^2\right) = n\left(p - \sum_d \lambda_d\right) \quad 2.2.1.9$$

Donde $\lambda_d = m^{-1} \sum_j \mu_{jd}^2$, con $d=1, \dots, p$, son los valores propios y corresponden al promedio de las medidas de discriminación.

2.2.2 ALGORITMO DEL ANÁLISIS DE COMPONENTES PRINCIPALES NO LINEAL

El teorema de Eckart-Young, dice que el Análisis de Componentes Principales de una matriz de H de dimensión $(n \times m)$ puede ser formulado en términos de su función de pérdida.

$$\sigma_j(X, A) \equiv SSQ(H - XA')$$

Existen variaciones de esta función como las propuestas por Roskam, Tenenhaus, Young, Takane y De Leeuw³; en estos casos usan el método de los mínimos cuadrados alternantes para minimizar la función de pérdida.

La función que ha sido mayormente desarrollada y programada, que se procede a minimizar es:

$$\sigma_j(Q, X, A) = \sum_j SSQ(q_j - Xa_j) m^{-1} \quad 2.2.2.1$$

la minimización se realiza sobre las matrices \mathbf{X} , \mathbf{A} y \mathbf{q}_j , perteneciendo este último a un cono convexo⁴ definido en \mathbf{R}^n , el que satisface que la media vale 0, y $SSQ(\mathbf{q}_j)=1$, donde \mathbf{q}_j es una transformación de los datos originales, dada por $\mathbf{q}_j = \mathbf{G}_j \mathbf{y}_j$.

donde:

- \mathbf{G}_j , es la matriz indicatriz de las categorías de la variable \mathbf{h}_j , siendo ésta la variable j de la matriz de datos \mathbf{H} .
- \mathbf{y}_j son las valoraciones buscadas para la variable \mathbf{h}_j de la matriz de datos \mathbf{H} .
- \mathbf{a}_j es un vector columna igual a la fila j de la matriz \mathbf{A} .

³ Albert Gifi, Nonlinear Multivariate Analysis, John Wiley and Sons, 1990, página 151-153.

⁴ Un conjunto A en \mathbf{R}^n se dice que es un cono convexo si satisface:

(i) Si $x \in A$, $kx \in A$, $\forall k > 0$, y

(ii) Si $x, y \in A$, $(x+y) \in A$

Es importante aclarar que esto implica que si $x \in A$, $-x \notin A$.

Así, la función de pérdidas (2.2.2.1), puede ser expresada como:

$$\sigma(Y, X, A) = \sum_j SSQ(G_j y_j - X a_j) m^{-1} \quad 2.2.2.2$$

La función (2.2.2.2) se minimiza bajo las siguientes restricciones:

- $u^t G_j y_j = 0$,
- $y_j^t D_j y_j = 1$,
- $y_j \in C_j$, cono convexo.

donde

- u^t es un vector de unos,
- D_j matriz diagonal conteniendo las sumas de las columnas de G_j

De donde el algoritmo de mínimos cuadrados alternantes para minimizar la función (2.2.2.2) estaría dado de la siguiente manera:

Se inicializa el procedimiento con una matriz $A_0 \equiv A$ de datos cualquiera.

Primer paso:	$X = HA(A'A)^{-1}$
Segundo paso:	$A = H'X(X'X)^{-1}$
Tercer paso, se calcula:	$y_j = D_j^{-1} G_j' X a_j$, y $q_j = G_j y_j$,
Finalmente se minimiza	$SSQ(Q - XA')$

Dichos pasos se realizan de manera iterativa, hasta lograr encontrar las valoraciones o cuantificaciones de las categorías de cada una de las variables.

Utilizando esta metodología se han desarrollado tres procedimientos para minimizar la función en cuestión, denominados:

- MNNFAEX
- PRINQUAL
- PRINCALS

El procedimiento MNNFAEX, trabaja con variables ordinales y numéricas, mientras que los otros dos métodos pueden manejar y combinar variables cualitativas (ordinales y nominales) y cuantitativas.

Los algoritmos mencionados minimizan como se dijo anteriormente la función de pérdidas, con base en los mínimos cuadrados alternantes, es decir la función se minimiza con respecto a \mathbf{X} , permaneciendo \mathbf{a}_j fijo, luego se minimiza \mathbf{a}_j , con \mathbf{X} fijo; es decir, alternadamente.

La alternación en la transformación de los datos se da con el modelo estadístico $\mathbf{h}_j = \sum \mathbf{x}_s \mathbf{a}_{js}$, el cual no es sino el análisis de componentes principales clásico, que no se lo debe interpretar en términos de ajuste de un modelo estadístico. Este modelo es parte de la formulación del teorema de Eckart-Young.

El procedimiento PRINQUAL esta programado en el paquete estadístico SAS, mientras que el método PRINCALS se lo encuentra en el paquete estadístico SPSS (Statistical Package for the Social Sciences) el mismo que será usado para el desarrollo práctico de este trabajo.

Realizadas las transformaciones de las variables cualitativas por este procedimiento, se consigue la máxima correlación lineal entre las variables dadas, ya que el valor mínimo⁵ de la función de pérdida, está dado por $\sigma^* = 1 - r_{..}$; donde $r_{..}$ es el promedio de las correlaciones entre todas las variables transformadas por las cuantificaciones óptimas, consecuentemente se logra tener la máxima correlación entre las variables transformadas.

Cabe señalar que si el procedimiento se aplica con variables cuantitativas, este es equivalente al análisis de componentes principales clásico; mientras que, si se aplica con variables cualitativas solamente, el análisis es equivalente al análisis de homogeneidad o de correspondencias múltiples.

⁵ Albert Gifi, Nonlinear Multivariate Analysis, John Wiley and Sons, 1990, páginas 81 a 93

El Análisis de Componentes Principales basados en la función de pérdida (2.2.2.2) tiene ciertos inconvenientes, por ejemplo, las cuantificaciones múltiples deben ser calculadas sucesivamente y entre tanto en el análisis de Homogeneidad estas pueden realizarse simultáneamente, adicionalmente que los valores perdidos incorporan otros problemas; por lo que usamos la siguiente función de pérdida que usa también en el análisis de homogeneidad:

$$\sigma(X, Y) = m^{-1} \sum_j SSQ(X - G_j Y_j) \quad 2.2.2.3$$

La que puede ser escrita como:

$$\sigma(X, Y) = \sum_j tr(X - G_j Y_j)'(X - G_j Y_j) \quad 2.2.2.4$$

Y reemplazando los Y_j por $(\hat{Y}_j + (Y_j - \hat{Y}_j))$ se tendría lo siguiente:

$$\sigma(X, Y) = \sum_j tr(X - G_j(\hat{Y}_j + (Y_j - \hat{Y}_j)))'(X - G_j(\hat{Y}_j + (Y_j - \hat{Y}_j))) \quad 2.2.2.5$$

$$= \sum_j tr(X - G_j \hat{Y}_j)'(X - G_j \hat{Y}_j) + \sum_j tr(Y_j - \hat{Y}_j)' D_j (Y_j - \hat{Y}_j) \quad 2.2.2.6$$

Introduciendo las restricciones de rango uno sobre los Y_j se tiene que minimizar el segundo término de 2.2.2.6 escribiendo $Y_j = y_j a_j'$ con respecto a y_j y a_j .

$$\sigma(X, Y) = \sum_j tr(X - G_j \hat{Y}_j)'(X - G_j \hat{Y}_j) + \sum_j tr(y_j a_j' - \hat{Y}_j)' D_j (y_j a_j' - \hat{Y}_j) \quad 2.2.2.7$$

Entonces minimizando con respecto a a_j usando el algoritmo de mínimos cuadrados alternantes manteniendo fijo y_j se obtienen las siguientes ecuaciones normales:

$$w_j a_j' = \hat{Y}_j D_j y_j \quad 2.2.2.8$$

Con: $w_j = y_j' D_j y_j$ escalar, es decir podemos escribir 2.2.2.8 como:

$$\hat{a}_j = \hat{Y}_j D_j y_j / w_j \quad 2.2.2.9$$

Ahora bien si escribimos la segunda parte de 2.2.2.7 restando y sumando \hat{a}_j de tal manera que:

$$\begin{aligned} \sum_j tr(y_j a_j' - \hat{Y}_j)' D_j (y_j a_j' - \hat{Y}_j) &= \sum_j tr(y_j a_j' - \hat{Y}_j)' D_j (y_j a_j' - \hat{Y}_j) + \\ &\sum_j tr(a_j' - \hat{a}_j)' D_j (a_j' - \hat{a}_j') \end{aligned} \quad 2.2.2.10$$

No existen mayores restricciones sobre los coeficientes \hat{a}_j , de modo que podemos escribir $\hat{a}_j = a_j$ y el paso inicial de la minimización del término en discusión esta finalizado; luego se minimiza con respecto a y_j para un a_j cualesquiera con lo se tendría el siguiente conjunto de ecuaciones normales:

$$v_j D_j y_j = D_j \hat{Y}_j a_j \quad 2.2.2.11$$

con $v_j = a_j' a_j$ escalar, con lo que:

$$\hat{y}_j = D_j \hat{Y}_j a_j / v_j \quad 2.2.2.12$$

Debemos ahora sumar y restar \hat{y}_j en la segunda parte de 2.2.2.7, es decir:

$$\begin{aligned} \sum_j tr(y_j a_j' - \hat{Y}_j)' D_j (y_j a_j' - \hat{Y}_j) &= \sum_j tr(\hat{y}_j \hat{a}_j' - \hat{Y}_j)' D_j (\hat{y}_j \hat{a}_j' - \hat{Y}_j) D_j + \\ &\sum_j tr(y_j - \hat{y}_j)' D_j (y_j - \hat{y}_j) v_j \end{aligned} \quad 2.2.2.13$$

Calcular los y_j implica proyectar \hat{y}_j sobre algún cono C_j y para el caso de variables nominales el cono apropiado es el espacio \mathfrak{R}^{k_j} obteniendo la proyección $y_j = \hat{y}_j$.

Escribimos $\hat{Y}_j = \hat{y}_j \hat{a}_j'$ con j que pertenece a las J coordenadas de cada fila de la matriz de datos y se procede a calcular las puntuaciones de los objetos, este procedimiento es el llamado PRINCALS.

2.2.3 ALGORITMO RESOLUTIVO PARA EL MÉTODO PRINCALS

A continuación presentamos el algoritmo para el método PRINCALS:

Primer paso: Inicializamos con una matriz X de tal manera que se cumpla $u'X=0$ y $X'X=nl_p$

Segundo paso: Calculamos las cuantificaciones categóricas múltiples con:

$$\hat{Y}_j = D_j^{-1} G_j X, \text{ con } j \in J \quad 2.2.3.1$$

Tercer paso: Se estiman los pesos:

$$\hat{a}_j = \hat{Y}_j D_j y_j / w_j, \text{ con } j \in J \quad 2.2.3.2$$

Cuarto paso: Estimamos las cuantificaciones categóricas simples a través de:

$$\hat{y}_j = \hat{Y}_j a_j / v_j, \text{ con } j \in J \quad 2.2.3.3$$

Quinto paso: Se relaciona el nivel de medida de la ***j-ésima*** variable realizando una regresión lineal o monótona según el caso.

Sexto paso: Se actualizan las cuantificaciones categóricas múltiples con:

$$\hat{Y}_j = \hat{y}_j \hat{a}_j' \text{ con } j \in J \quad 2.2.3.4$$

Séptimo paso: Se calculan las puntuaciones de los objetos:

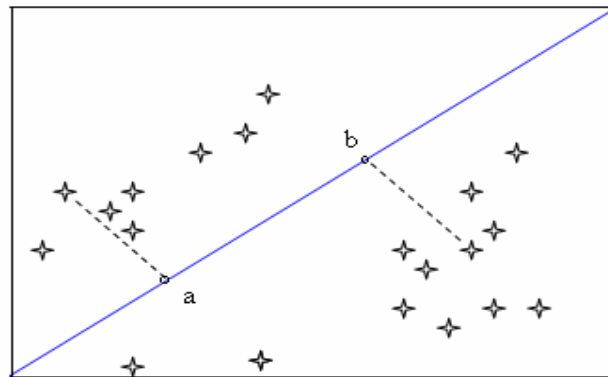
$$\hat{X} = J^{-1} \sum_{j=1}^J G_j Y_j \quad 2.2.3.5$$

Octavo paso: Se centran las columnas y ortonormalizan la matriz de puntuaciones de los objetos.

Noveno paso: Se utiliza un criterio de convergencia el mismo que puede ser dado por el Análisis de Homogeneidad.

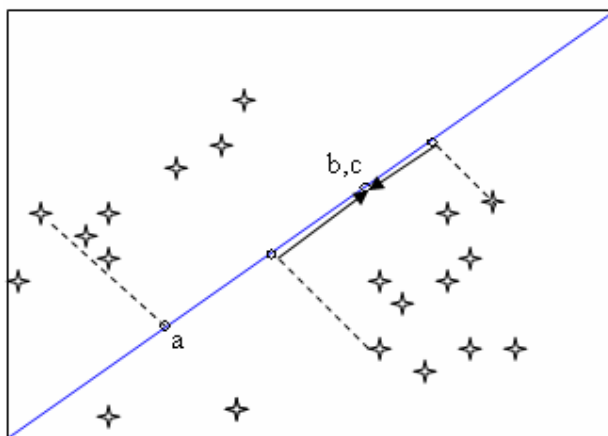
Al trabajar con variables nominales simples, cada una de estas puede ser proyectada en una línea, además las categorías se posicionan en la línea en cualquier orden, es decir se puede considerar para este caso al conjunto de categorías como una variable, gráficamente se tendría lo siguiente:

Gráfico 2.2.3.1



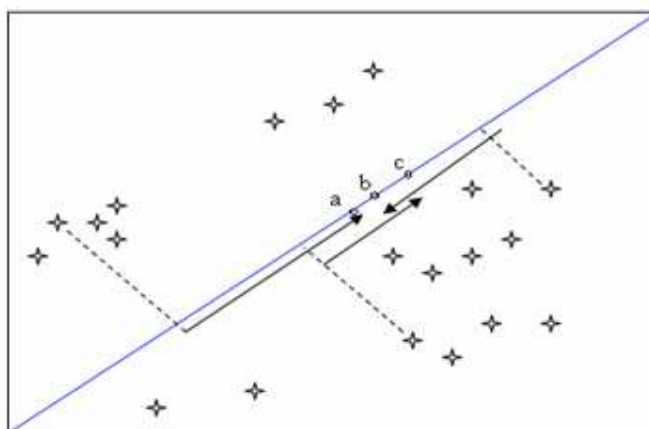
Para el caso de variables ordinales simples también las cuantificaciones se encuentran en una línea recta, pero además estas se encontraran ordenadas así:

Gráfico 2.2.3.2



Ahora si la variable se trata como una numérica simple las puntuaciones de las categorías deben seguir un orden y deben estar separadas por la misma distancia ya que estas variables constan orden y medida; más aún las soluciones por este procedimiento son encajadas, es decir no importa el espacio de la p solución contienen la misma información para cualquier p como se presenta en el siguiente gráfico:

Gráfico 2.2.3.3



2.2.4 MANEJO DE CASOS SIMPLES

Para estudiar variables ordinales simples escribamos la función de pérdida separándola en dos partes

$$\sigma(X, Y) = \sum_j tr(X - G_j \hat{Y}_j)'(X - G_j \hat{Y}_j) + \sum_j tr(\hat{y}_j \hat{a}_j' - \hat{Y}_j)' D_j (\hat{y}_j \hat{a}_j' - \hat{Y}_j) \quad 2.2.4.1$$

Usando 2.2.1.9 podemos escribir el primer término de 2.2.4.1 de la siguiente manera:

$$\sum_j tr(X - G_j \hat{Y}_j)'(X - G_j \hat{Y}_j) = N(p - \sum_j \sum_s \lambda_{js}^2) \quad 2.2.4.2$$

Que se denomina pérdida múltiple, donde λ_{js}^2 es el ajuste múltiple de la variable j en la dimensión s .

Si tenemos en cuenta que $\hat{Y}_j' D_j y_j a_j' = N a_j a_j'$ y además que las restricciones de normalización $y_j' a_j' y_j = N$, de 2.2.2.9 la segunda parte de 2.2.4.1 se escribe de la siguiente manera:

$$\sum_j tr(\hat{Y}_j' D_j \hat{Y}_j - N a_j a_j') = N(\sum_j \sum_s (\lambda_{js}^2 - a_{js}^2)) \quad 2.2.4.3$$

Al segundo término de la ecuación 2.2.4.3 se la llama pérdida simple y al término a_{js}^2 se lo denomina ajuste simple; además de la ecuación 2.2.4.1 se tiene que si la variable es tratada como nominal múltiple esta no contribuye en nada a la pérdida simple.

Nótese que la pérdida simple está formada de dos partes, la primera es la restricción de rango uno que indica que las cuantificaciones simples de las categorías deben estar sobre una línea en el espacio conjunto y la segunda, la restricción del nivel de medida que indica que las cuantificaciones simples deben

estar en el orden adecuado para el caso de variables ordinales o igualmente espaciados para el caso numérico.

2.2.5 MANEJO DE CASOS PERDIDOS

Cuando se cuenta con valores perdidos (missing values) se utiliza la siguiente transformación de la ecuación 2.2.2.6

$$\begin{aligned} \sigma(X, Y) &= \sum_j \text{tr}(X - G_j Y_j)'(X - G_j Y_j) \\ &= \sum_j \text{tr}(X - G_j \hat{Y}_j)' M_j (X - G_j \hat{Y}_j) + \sum_j \text{tr}(Y_j - \hat{Y}_j)' D_j (Y_j - \hat{Y}_j) \end{aligned} \quad 2.2.5.1$$

Donde la matriz M_j es una diagonal binaria que indica cuales observaciones son valores perdidos es decir si i está perdido para algún j el elemento diagonal i de M_j es cero caso contrario es uno, de este modo los valores perdidos no afectan al algoritmo.

2.2.6 MANEJO DE VARIABLES CON NIVEL DE MEDIDAS MIXTAS

El algoritmo de Mínimos cuadrados alternantes nos permite trabajar con variables de medidas mixtas, es decir tratar a las categorías de una variable como nominales y otras como ordinales, para esto se presentan dos métodos que nos ayudarán a tratar estos casos.

2.2.6.1 Primer Método

Para este caso tratamos las variables como simples, inicialmente se particiona el vector de cuantificaciones de categorías simples en dos; la que contiene restricciones que es la correspondiente a datos ordinales y numéricos, la segunda que no cuenta con restricciones⁶ que corresponde a los datos nominales.

$$y_j = \begin{pmatrix} y_j^K \\ y_j^L \end{pmatrix} \quad 2.2.6.1$$

Usando 2.2.2.9 y 2.2.2.12 se estiman los valores \hat{a}_j y \hat{y}_j , quedando por calcular $tr(y_j - \hat{y}_j)' D_j (y_j - \hat{y}_j)$, la dificultad radica en que el cono C_j , ya que no se requiere la linealidad y la monotonía para la parte nominal, para esto particionamos la matriz D_j así:

$$D_j = \begin{pmatrix} D_j^K & 0 \\ 0 & D_j^L \end{pmatrix} \quad 2.2.6.2$$

De donde se tiene:

$$tr(y_j - \hat{y}_j)' D_j (y_j - \hat{y}_j) = tr(y_j^K - \hat{y}_j^K)' D_j^K (y_j^K - \hat{y}_j^K) + tr(y_j^L - \hat{y}_j^L)' D_j^L (y_j^L - \hat{y}_j^L) \quad 2.2.6.3$$

La segunda parte de 2.2.6.3 minimiza simplemente escribiendo $y_j^L = \hat{y}_j^L$; entre tanto la primera parte presenta dos casos, para las variables ordinales se realiza una regresión sobre la parte con restricción y_j^K .

La diferencia radica en que la métrica que cambia a D_j^K entonces juntamos las dos partes y normalizamos con lo que se obtiene:

$$y_j' y_j^K' D_j^K Q_j^K + y_j^L' D_j^L Q_j^L = N \quad 2.2.6.4$$

⁶ Salazar Bermeo, Franklin Eduardo, Análisis de Homogeneidad en el Análisis Multivariante no lineal, Estudio en centros de rehabilitación social, Septiembre 2000.

Para el caso de variables numéricas simples se debe cumplir que:

$$y_j^K = \eta_j + \zeta_j s_j \quad 2.2.6.5$$

Con s_j normalizada a través de:

$$u' D_j^K s_j = 0 \quad y \quad 2.2.6.6$$

$$s_j' D_j^K s_j = 1 \quad 2.2.6.7$$

Se calculan η_j y ζ_j usando regresiones y se estiman los y_j^K ; luego se incorpora la sección con restricciones y sin restricciones de y_j y normalizamos con $y_j' D_j y_j = N$.

2.2.6.2 Segundo Método

Una de las condiciones que exige el primer método⁷ es que la parte sin restricción sea nominal simple; en este procedimiento la sección nominal es tratada como nominal múltiple.

En es método se particiona la matriz indicador de la siguiente manera:

$$G_j = [G_j^K \parallel G_j^L] \quad 2.2.6.8$$

Y de igual manera se tendría que:

$$Y_j = [Y_j^K \parallel Y_j^L]' \quad 2.2.6.9$$

$$D_j = \oplus_{r=K,L} D_j^r \quad 2.2.6.10$$

Minimizando la función de pérdida se obtiene:

⁷ Salazar Bermeo, Franklin Eduardo, Análisis de Homogeneidad en el Análisis Multivariante no lineal, Estudio en centros de rehabilitación social, Septiembre 2000.

$$\hat{Y}_j^K = (D_j^K)^{-1} G_j^K' X \quad 2.2.6.11$$

$$\hat{Y}_j^L = (D_j^L)^{-1} G_j^L' X \quad 2.2.6.12$$

Planteamos restricciones de rango uno en la segunda parte para los datos ordinales o numéricos, es decir: $Y_j^K = z_j \alpha_j'$ con lo que la ecuación 2.2.2.6 sería:

$$\sigma(X, Y) = \sum_j \text{tr}(X - G_j \hat{Y}_j)' (X - G_j \hat{Y}_j) + \sum \text{tr}(z_j \alpha_j' - \hat{Y}_j)' D_j^K (z_j \alpha_j' - \hat{Y}_j) \quad 2.2.6.13$$

Luego se deben estimar los z_j y α_j como en (2.2.2.8) y (2.2.2.9); y usando la normalización $z_j' D_j^K z_j = N_K$ con $N_K = u_j' D_j^K u_j$ se obtiene que la componente de pérdida simple está dada por:

$$N_K \left(\sum_j^J \sum_s^p (\phi_{js}^2(K) - a_{js}^2) \right) \quad 2.2.6.14$$

En donde $\phi_{js}^2(K) = y_j^K'(\cdot, s) D_j^K y_j^K(\cdot, s) / N_K$ la pérdida simple esta dada por la sección restrictiva de la función de pérdida.

2.2.7 GENERALIZACIÓN A MÚLTIPLES CONJUNTOS DE VARIABLES

Para dar una generalización para múltiples conjuntos de variables al grupo de índices J de las J variables se lo particiona en un subconjunto B , tal que $J(1), \dots, J(b), \dots, J(B)$; con lo que la función de pérdida estaría dada por:

$$\sigma(X; Y_1, \dots, Y_j) = B^{-1} \sum_b SSQ(X - \sum_{j \in J(b)} G_j Y_j) \quad 2.2.7.1$$

Sujeta a las restricciones $X'X = NI_p$ y $u'X = 0$, pero en este caso las transformaciones óptimas de una variable j dentro un conjunto $J(b)$ depende de las transformaciones de las restantes variables del conjunto $J(b)$ con lo que es necesario corregir la contribución de las otras variables, lo que se realiza en el siguiente algoritmo:

Primer paso: dada X la transformación óptima de Y_j es:

$$\hat{Y}_j = D_j^{-1} G_j^{-1} (X - V_{kj}) \quad \text{con } j \in J \quad 2.2.7.2$$

Con $V_{kj} = \sum_{j \in J(b)} G_j Y_j - G_j Y_j$.

Segundo paso: Dados Y_j el óptimo X esta dado por:

$$\hat{X} = B^{-1} \sum_b \sum_{j \in J(b)} G_j Y_j \quad 2.2.7.3$$

Tercer paso: Se centran y ortonormalizan las puntuaciones de los objetos para satisfacer las restricciones de normalización.

Las ecuaciones (2.2.7.2) y (2.2.7.3) ilustran el principio de centroide⁸ que no son más que las cuantificaciones de las categorías de los objetos corregidos por la influencia del resto de variables del conjunto y las puntuaciones de los objetos son los promedios de variables cuantificadas.

⁸ Salazar Bermeo, Franklin Eduardo, Análisis de Homogeneidad en el Análisis Multivariante no lineal, Estudio en centros de rehabilitación social, Septiembre 2000.

2.3 RESULTADOS

A continuación se presentan las principales consecuencias o resultados que son de utilidad para el análisis de datos usando la técnica en discusión.

Uno de los resultados más importantes es el que el Análisis de Componentes Principales no lineal, garantiza la máxima correlación cuadrada media entre las variables recodificadas por las cuantificaciones categóricas y los componentes.

El procedimiento PRINCALS puede utilizar variables de tipo nominales múltiples, nominales simples, ordinales y numéricas; mientras que otros trabajan solo con variables ordinales discretas o continuas simples como es el caso del procedimiento MNNFAEX desarrollado por Roskam E. y la diferencia radica en la función de pérdida que se utiliza.

A diferencia del Análisis de Homogeneidad en el que las soluciones sobre las cuantificaciones categóricas no tienen restricciones, en el Análisis de Componentes Principales no lineal se incorporan restricciones sobre el orden y la medida de las variables.

Debemos anotar que el algoritmo de mínimos cuadrados alternantes calcula proyecciones sobre conos, si este cono es cerrado siempre existirá solución en un espacio de dimensión finita y si adicionalmente es un convexo la proyección es única.

Con la valoración de las categorías con base en este método, se logra maximizar la correlación lineal entre las variables dadas, consecuentemente se logra transformar las variables cualitativas en variables cuantitativas, es decir se cuenta con la mejor combinación lineal posible de las variables tratadas, sirviendo esto, por ejemplo, para la construcción de modelos de regresión (en donde interesa precisamente la linealidad), o poder discriminar o caracterizar mejor casos ambiguos, entre muchos otros análisis que se podrían efectuar.

Este procedimiento no garantiza que se produzcan mejores transformaciones que otros métodos en el sentido de la optimización como tal; sino que es mejor con respecto a la función de pérdida elegida para este mecanismo.

Y, finalmente es importante mencionar que a través de las cuantificaciones obtenidas pueden construirse índices o indicadores que recojan la información de un conjunto de variables de interés.

2.4 PROCEDIMIENTO DE CONSTRUCCIÓN DE ÍNDICES

Como se mencionó en la parte final de la sección anterior las cuantificaciones obtenidas nos permiten construir indicadores con base en un conjunto de variables nominales u ordinales; para esto sean \hat{Y}_{jk} las cuantificaciones categóricas óptimas de las J variables con $j=1, \dots, m$ y de la k -ésima categoría dentro de cada variable, es decir que por ejemplo la cuantificación \hat{Y}_{14} es la cuantificación de la variable uno y de la cuarta categoría u orden de la misma.

Entonces se obtiene un vector \hat{Y} que está formado por las cuantificaciones $\hat{Y}_{1k}, \dots, \hat{Y}_{jk}, \dots, \hat{Y}_{mk}$ más claramente se tendría:

$$\hat{Y} = (\hat{Y}_{1k}, \dots, \hat{Y}_{jk}, \dots, \hat{Y}_{mk})' \quad 2.4.1$$

Ahora bien analicemos el caso de la primera variable \hat{Y}_{1k} y para esto supongamos que la variable en cuestión tiene s_1 categorías es decir nuestro vector de cuantificaciones para la primera variable estaría dado por:

$$\hat{Y}_{1k} = (\hat{y}_{11}, \dots, \hat{y}_{1k}, \dots, \hat{y}_{1s_1})' \quad 2.4.2$$

Dentro de cada variable nosotros podemos definir un orden de las ponderaciones ya que no necesariamente la cuantificación de la primera categoría será inferior a las demás; para esto podemos también definir un mínimo dentro de las cuantificaciones óptimas así:

$$\min(\hat{Y}_{jk}) = \mu_j \text{ con } j \in J \quad 2.4.3$$

Y el valor de la siguiente cuantificación al mínimo estará dado por:

$$\min(\hat{Y}_{jk}) + \delta_i = \mu_j + \delta_i = \mu_{j+1} \text{ con } j \in J \quad 2.4.4$$

Donde δ_i es un escalar que mide la diferencia entre una cuantificación menor y otra más grande con $i=1, \dots, k-1$ es decir por ejemplo δ_1 mide la diferencia que existe entre el mínimo y la segunda cuantificación con respecto a los valores dados por el algoritmo.

Es decir nuestro vector de cuantificaciones estaría dado por:

$$\hat{Y} = \begin{pmatrix} \hat{Y}_{1k} \\ \cdot \\ \cdot \\ \cdot \\ \hat{Y}_{jk} \\ \cdot \\ \cdot \\ \cdot \\ \hat{Y}_{mk} \end{pmatrix} = \begin{pmatrix} (\mu_1, \dots, \mu_{1+(s_1-1)})' \\ \cdot \\ \cdot \\ \cdot \\ (\mu_j, \dots, \mu_{j+(s_j-1)})' \\ \cdot \\ \cdot \\ \cdot \\ (\mu_m, \dots, \mu_{m+(s_m-1)})' \end{pmatrix} \quad 2.4.5$$

Se puede entonces realizar una transformación restando el mínimo de cada variable a las cuantificaciones de las mismas y así se consigue cambiar las puntuaciones a una escala que va desde cero que correspondería al valor mínimo hasta un valor dado por transformación ya que $\mu_j < \mu_{j+1}, \dots, < \mu_{j+(s_j-2)} < \mu_{j+(s_j-1)}$; por lo que se tendría $0 < \mu_{j+1} - \mu_j, \dots, < \mu_{j+(s_j-2)} - \mu_j < \mu_{j+(s_j-1)} - \mu_j$ y de este modo se logra obtener una escala que va desde cero hasta cierto valor que sería el máximo menos el mínimo; así nuestras cuantificaciones estarían dadas por:

$$\hat{Y} = \begin{pmatrix} (0, \dots, \mu_{1+(s_1-1)} - \mu_1)' \\ \cdot \\ \cdot \\ \cdot \\ (0, \dots, \mu_{j+(s_j-1)} - \mu_j)' \\ \cdot \\ \cdot \\ \cdot \\ (0, \dots, \mu_{m+(s_m-1)} - \mu_m)' \end{pmatrix} \quad 2.4.6$$

Dadas las cuantificaciones en la escala de cero a un valor máximo dentro de la variable debemos ahora realizar la suma de dichos máximos, para esto tenemos:

$$\max(\hat{Y}_{jk}) = \mu_{j+(s_j-1)} - \mu_j = \rho_j \text{ con } j \in J \quad 2.4.7$$

Adicionalmente nosotros podemos calcular la suma de los valores máximos de la siguiente manera:

$$\sum_j \rho_j = \gamma \text{ con } j \in J \quad 2.4.8$$

obteniendo el mayor valor posible que puede tener un individuo con respecto a las variables de interés, de este modo se puede obtener un **factor de escalamiento** que no es sino el valor de una escala **z** deseada (por ejemplo diez o cien) dividido por el valor γ , es decir tendríamos que nuestro factor de escalamiento estaría dado por:

$$fe = \frac{z}{\gamma} \quad 2.4.9$$

Así nosotros transformamos nuestras cuantificaciones óptimas a la escala deseada con lo que se tendría:

$$\hat{Y} = \begin{pmatrix} (0, \dots, (fe(\mu_{1+(s_1-1)} - \mu_1))' \\ \cdot \\ \cdot \\ \cdot \\ (0, \dots, (fe(\mu_{j+(s_j-1)} - \mu_j))' \\ \cdot \\ \cdot \\ \cdot \\ (0, \dots, (fe(\mu_{m+(s_m-1)} - \mu_m))' \end{pmatrix} \quad 2.4.10$$

Finalmente se obtiene el indicador imputando a las categorías la puntuación correspondiente y para cada individuo se suman las cuantificaciones obtenidas en (2.4.10) es decir nuestro indicador T estará dado por:

$$T = \sum_{j=1}^m fe(\mu_{j+(s_j-1)} - \mu_j)$$

$$= fe \sum_{j=1}^m (\mu_{j+(s_j-1)} - \mu_j) \quad 2.4.11$$

2.4.1 EJEMPLO

Supongamos que contamos con 20 individuos y siete variables (Tabla 2.4.1.1), las cuales describen el nivel de educación del entrevistado y las otras seis acerca de la tenencia de bienes; las variables son las siguientes: Educación del entrevistado, número de televisores a color, número de computadoras, líneas telefónicas, tenencia de autos, tenencia de lavadora de ropa y si cuentan con servicio doméstico.

Tabla 2.4.1.1

n	Educación del entrevistado	TV a color	Computadoras	Líneas Telefónicas	Autos	Lavadora de ropa	Servicio doméstico
1	Superior no universitaria	Tres o más	Una	Una	Ninguno	No	No
2	Superior / universitario	Tres o más	Una	Una	Uno	No	No
3	Ninguna / primaria	Una o dos	Ninguna	Una	Ninguno	No	Si
4	Superior no universitaria	Una o dos	Una	Una	Uno	No	No
5	Secundaria	Una o dos	Una	Una	Uno	No	No
6	Superior / universitario	Una o dos	Una	Dos o más	Uno	No	No
7	Superior / universitario	Tres o más	Una	Dos o más	Dos	No	No
8	Superior / universitario	Tres o más	Dos o tres	Una	Ninguno	Si	Si
9	Superior / universitario	Una o dos	Una	Una	Uno	Si	Si
10	Secundaria	Tres o más	Una	Una	Dos	Si	Si
11	Superior / universitario	Una o dos	Una	Una	Ninguno	No	No
12	Superior / universitario	Tres o más	Dos o tres	Dos o más	Uno	Si	Si
13	Superior no universitaria	Una o dos	Una	Ninguna	Ninguno	No	No
14	Secundaria	Una o dos	Una	Ninguna	Ninguno	Si	Si
15	Superior / universitario	Tres o más	Dos o tres	Una	Dos	Si	Si
16	Superior no universitaria	Una o dos	Una	Una	Uno	No	No
17	Postgrado	Tres o más	Una	Una	Tres o más	Si	Si
18	Secundaria	Tres o más	Una	Una	Uno	Si	Si
19	Superior / universitario	Tres o más	Una	Dos o más	Uno	Si	Si
20	Superior / universitario	Tres o más	Dos o tres	Dos o más	Uno	Si	Si

Debemos definir que valor es mayor o menor, según lo que queramos medir es decir si cuantificamos positivamente o negativamente, en nuestro caso mediremos de menor a mayor las categorías dentro de cada variable, para esto se tienen las cuantificaciones categóricas para cada una de las variables:

Educación del entrevistado	
Ninguna / Primaria	1
Secundaria	2
Superior no universitaria	3
Superior / Universitaria	4
Postgrado	5

Televisores a color	
Ninguno	1
Uno o dos	2
Tres o más	3

Computadoras	
Ninguna	1
Una	2
Dos o tres	3
Cuatro o más	4

Líneas Telefónicas	
Ninguna	1
Una	2
Dos o más	3

Automóviles	
Ninguno	1
Uno	2
Dos	3
Tres o mas	4

Lavadora de ropa	
No	1
Si	2

Servicio doméstico	
No	1
Si	2

Dadas las cuantificaciones categóricas usamos el algoritmo PRINCALS implementado en el paquete estadístico SPSS que usa la sintaxis que se anota en el anexo 1 y se obtienen los siguientes resultados:

Variable: Nivel educativo del entrevistado		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
1 Ninguna / primaria	1	-1,81
2 Secundaria	4	-1,71
3 Superior no universitaria	4	0,46
4 Superior / universitaria	10	0,62
5 Postgrado	1	0,62

Variable: Televisores a color		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
1 Ninguna	0	0
2 Una o dos	9	-1,11
3 Tres o más	11	0,9

Variable: Computadora		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
1 Ninguna	1	-2,91
2 Una	15	-0,24
3 Dos o tres	4	1,63
4 Cuatro o más	0	0

Variable: Líneas telefónicas		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
1 Ninguna	2	-1,96
2 Una	13	-0,28
3 Dos o más	5	1,51

Variable: Automóviles		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
1 Ninguno	6	-1,31
2 Uno	10	0,17
3 Dos	3	1,54
4 Tres o más	1	1,54

Variable: Lavadora de ropa		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
1 No	5	-1,73
2 Si	15	0,58

Variable: Servicio doméstico		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
1 No	10	-1
2 Si	10	1

Así en el gráfico (2.4.1.1) se ve como las categorías del nivel educativo del entrevistado más bajas se encuentran alrededor de la categoría de no tenencia de lavadora, adicionalmente se observa que los individuos que no cuentan con ninguna línea telefónica tampoco cuentan con computadora.

Para la construcción del índice se utiliza la metodología presentada a través de la siguiente de la tabla 2.4.1.2:

Tabla 2.4.1.2

Variables	Proyecciones Óptimas		Mínimos	Transformación de cero a mayor	Máximos	Cuantificaciones Ponderadas por el Índice de Escalamiento
Educación de la entrevistada						
Ninguna / Primaria	1	-1,81	-1,81	0	2,43	0
Secundaria	2	-1,71		0,1		0,509943906
Superior no universitaria	3	0,46		2,27		11,57572667
Superior / Universitaria	4	0,62		2,43		12,39163692
Postgrado	5	0,62		2,43		12,39163692
Televisores a color						
Ninguno	1	0	-1,11	0	2,01	0
Uno o dos	2	-1,11		0		0
Tres o más	3	0,9		2,01		10,24987251
Computadoras						
Ninguna	1	-2,91	-2,91	0	4,54	0
Una	2	-0,24		2,67		13,61550229
Dos o tres	3	1,63		4,54		23,15145334
Cuatro o más	4	0				0
Líneas Telefónicas						
Ninguna	1	-1,96	-1,96	0	3,47	0
Una	2	-0,28		1,68		8,567057624
Dos o más	3	1,51		3,47		17,69505354
Automóviles						
Ninguno	1	-1,31	-1,31	0	2,85	0
Uno	2	0,17		1,48		7,547169811
Dos	3	1,54		2,85		14,53340133
Tres o mas	4	1,54		2,85		14,53340133
Lavadora de ropa						
No	1	-1,73	-1,73	0	2,31	0
Si	2	0,58		2,31		11,77970423
Servicio doméstico						
No	1	-1	-1	0	2	0
Si	2	1		2		10,19887812
				Suma de los valores máximos	19,61	100
				Índice de Escalamiento	5,1	

Cabe señalar que cuando no existen individuos dentro de una categoría en una variable, la cuantificación es cero.

De este modo, asignando las cuantificaciones a cada categoría de cada variable de los individuos se tendrían los siguientes valores del índice construido para cada elemento como se ve en la tabla 2.4.1.3:

Tabla 2.4.1.3

n	Educación del entrevistado	TV a color	Computadoras	Líneas Telefónicas	Autos	Lavadora de ropa	Servicio doméstico	I _i Indicador
1	Superior no universitaria	Tres o más	Una	Una	Ninguno	No	No	55,79
2	Superior / universitario	Tres o más	Una	Una	Uno	No	No	64,15
3	Ninguna / primaria	Una o dos	Ninguna	Una	Ninguno	No	Si	8,57
4	Superior no universitaria	Una o dos	Una	Una	Uno	No	No	53,09
5	Secundaria	Una o dos	Una	Una	Uno	No	No	30,24
6	Superior / universitario	Una o dos	Una	Dos o más	Uno	No	No	63,03
7	Superior / universitario	Tres o más	Una	Dos o más	Dos	No	No	80,27
8	Superior / universitario	Tres o más	Dos o tres	Una	Ninguno	Si	Si	76,34
9	Superior / universitario	Una o dos	Una	Una	Uno	Si	Si	64,1
10	Secundaria	Tres o más	Una	Una	Dos	Si	Si	57,67
11	Superior / universitario	Una o dos	Una	Una	Ninguno	No	No	46,35
12	Superior / universitario	Tres o más	Dos o tres	Dos o más	Uno	Si	Si	93,01
13	Superior no universitaria	Una o dos	Una	Ninguna	Ninguno	No	No	36,97
14	Secundaria	Una o dos	Una	Ninguna	Ninguno	Si	Si	24,32
15	Superior / universitario	Tres o más	Dos o tres	Una	Dos	Si	Si	90,87
16	Superior no universitaria	Una o dos	Una	Una	Uno	No	No	53,09
17	Postgrado	Tres o más	Una	Una	Tres o más	Si	Si	81,34
18	Secundaria	Tres o más	Una	Una	Uno	Si	Si	50,69
19	Superior / universitario	Tres o más	Una	Dos o más	Uno	Si	Si	83,48
20	Superior / universitario	Tres o más	Dos o tres	Dos o más	Uno	Si	Si	93,01

En la tabla anterior puede observarse como funciona el índice construido para estos veinte casos, es claro que el individuo tres es el que se encuentra en peores condiciones con respecto a las variables presentadas ya que el jefe del hogar no tiene sino tal vez un nivel máximo de educación primaria, no poseen computadora, ningún auto y no poseen lavadora de ropa ni servicio doméstico; por el contrario los elementos que se encuentran mejor (individuos 12 y 20) ya

que el nivel educativo de los entrevistados es superior universitario, tienen vehículo y poseen servicio doméstico y lavadora de ropa.

El caso 18 no sobresale de los demás pese a que tiene mejores condiciones que el elemento 12, por que la variable de líneas telefónicas discrimina mucho más que el otro grupo de variables dentro de este conjunto de datos.

2.4.1.1 Construcción inversa del Índice

Pues bien en este punto es pertinente ver que es lo que sucede si tomamos el orden inverso de las categorías, es decir si asignamos el valor más pequeño a la categoría que en el ejemplo tiene el valor más grande, más claramente que relación existe con un nuevo índice construido de manera inversa, para esto asignamos las siguientes cuantificaciones categóricas:

Educación del entrevistado	
Ninguna / Primaria	5
Secundaria	4
Superior no universitaria	3
Superior / Universitaria	2
Postgrado	1

Televisores a color	
Ninguno	3
Uno o dos	2
Tres o más	1

Computadoras	
Ninguna	4
Una	3
Dos o tres	2
Cuatro o más	1

Líneas Telefónicas	
Ninguna	3
Una	2
Dos o más	1

Automóviles	
Ninguno	4
Uno	3
Dos	2
Tres o mas	1

Lavadora de ropa	
No	2
Si	1

Servicio doméstico	
No	2
Si	1

Entonces corriendo el algoritmo usando estas categorías se obtiene lo siguiente:

Variable: Nivel educativo del entrevistado		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
5 Ninguna / primaria	1	1,81
4 Secundaria	4	1,71
3 Superior no universitaria	4	-0,46
2 Superior / universitaria	10	-0,62
1 Postgrado	1	-0,62

Variable: Televisores a color		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
3 Ninguna	0	0
2 Una o dos	9	1,11
1 Tres o más	11	-0,9

Variable: Computadora		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
4 Ninguna	1	2,91
3 Una	15	0,24
2 Dos o tres	4	-1,63
1 Cuatro o más	0	0

Variable: Líneas telefónicas		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
3 Ninguna	2	1,96
2 Una	13	0,28
1 Dos o más	5	-1,51

Variable: Automóviles		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
4 Ninguno	6	1,31
3 Uno	10	-0,17
2 Dos	3	-1,54
1 Tres o más	1	-1,54

Variable: Lavadora de ropa		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
2 No	5	1,73
1 Si	15	-0,58

Variable: Servicio doméstico		
Tipo: Ordinal		
Valores perdidos: 0		
Categoría	Frecuencia Marginal	Cuantificación
2 No	10	1
1 Si	10	-1

Como puede verse, las cuantificaciones en el caso inverso tienen los mismos valores pero con signo distinto es decir que el algoritmo reproduce los mismos valores para las ponderaciones pero medidos en forma inversa que era lo que se esperaba.

Utilizando el mismo procedimiento de construcción de índices veamos que es lo que sucede utilizando este orden inverso de las categorías dentro de cada variable, lo que se presenta en la tabla 2.4.1.4:

Tabla 2.4.1.4

Variables		Proyecciones Óptimas	Mínimos	Transformación de cero a mayor	Máximos	Cuantificaciones Ponderadas por el Índice de Escalamiento
Educación de la entrevistada						
Ninguna / Primaria	5	1,81	-0,62	2,43	2,43	12,39163692
Secundaria	4	1,71		2,33		11,88169301
Superior no universitaria	3	-0,46		0,16		0,81591025
Superior / Universitaria	2	-0,62		0		0
Postgrado	1	-0,62		0		0
Televisores a color						
Ninguno	3	0	-0,9	0	2,01	0
Uno o dos	2	1,11		2,01		10,24987251
Tres o más	1	-0,9		0		0
Computadoras						
Ninguna	4	2,91	-1,63	4,54	4,54	23,15145334
Una	3	0,24		1,87		9,535951045
Dos o tres	2	-1,63		0		0
Cuatro o más	1	0				0
Líneas Telefónicas						
Ninguna	3	1,96	-1,51	3,47	3,47	17,69505354
Una	2	0,28		1,79		9,12799592
Dos o más	1	-1,51		0		0
Automóviles						
Ninguno	4	1,31	-1,54	2,85	2,85	14,53340133
Uno	3	-0,17		1,37		6,986231515
Dos	2	-1,54		0		0
Tres o mas	1	-1,54		0		0
Lavadora de ropa						
No	2	1,73	-0,58	2,31	2,31	11,77970423
Si	1	-0,58		0		0
Servicio doméstico						
No	2	1	-1	2	2	10,19887812
Si	1	-1		0		0
				Suma de los valores máximos	19,61	100
				Índice de Escalamiento	5,1	

Como se ve el índice de escalamiento va a ser el mismo dadas las condiciones de construcción del indicador.

Ahora bien asignemos los valores de las cuantificaciones ponderadas por el índice de escalamiento y observemos que es lo que sucede con nuestro nuevo índice.

Tabla 2.4.1.5

n	Educación del entrevistado	TV a color	Computadoras	Líneas Telefónicas	Autos	Lavadora de ropa	Servicio doméstico	I ₂ Indicador Dos
1	Superior no universitaria	Tres o más	Una	Una	Ninguno	No	No	44,21
2	Superior / universitario	Tres o más	Una	Una	Uno	No	No	35,85
3	Ninguna / primaria	Una o dos	Ninguna	Una	Ninguno	No	Si	91,43
4	Superior no universitaria	Una o dos	Una	Una	Uno	No	No	46,91
5	Secundaria	Una o dos	Una	Una	Uno	No	No	69,76
6	Superior / universitario	Una o dos	Una	Dos o más	Uno	No	No	36,97
7	Superior / universitario	Tres o más	Una	Dos o más	Dos	No	No	19,73
8	Superior / universitario	Tres o más	Dos o tres	Una	Ninguno	Si	Si	23,66
9	Superior / universitario	Una o dos	Una	Una	Uno	Si	Si	35,9
10	Secundaria	Tres o más	Una	Una	Dos	Si	Si	42,33
11	Superior / universitario	Una o dos	Una	Una	Ninguno	No	No	53,65
12	Superior / universitario	Tres o más	Dos o tres	Dos o más	Uno	Si	Si	6,99
13	Superior no universitaria	Una o dos	Una	Ninguna	Ninguno	No	No	63,03
14	Secundaria	Una o dos	Una	Ninguna	Ninguno	Si	Si	75,68
15	Superior / universitario	Tres o más	Dos o tres	Una	Dos	Si	Si	9,13
16	Superior no universitaria	Una o dos	Una	Una	Uno	No	No	46,91
17	Postgrado	Tres o más	Una	Una	Tres o más	Si	Si	18,66
18	Secundaria	Tres o más	Una	Una	Uno	Si	Si	49,31
19	Superior / universitario	Tres o más	Una	Dos o más	Uno	Si	Si	16,52
20	Superior / universitario	Tres o más	Dos o tres	Dos o más	Uno	Si	Si	6,99

Dados estos dos indicadores puede observarse que si nosotros sumamos los dos se obtiene el valor máximo del indicador que en este caso es cien puntos, es decir los dos indicadores son **complementarios** como se observa en la tabla 2.4.1.6:

Tabla 2.4.1.6

n	I ₁ Indicador Uno	I ₂ Indicador Dos	I ₁ +I ₂
1	55,79	44,21	100
2	64,15	35,85	100
3	8,57	91,43	100
4	53,09	46,91	100
5	30,24	69,76	100
6	63,03	36,97	100
7	80,27	19,73	100
8	76,34	23,66	100
9	64,1	35,9	100
10	57,67	42,33	100
11	46,35	53,65	100
12	93,01	6,99	100
13	36,97	63,03	100
14	24,32	75,68	100
15	90,87	9,13	100
16	53,09	46,91	100
17	81,34	18,66	100
18	50,69	49,31	100
19	83,48	16,52	100
20	93,01	6,99	100

Y si calculamos la correlación entre estos dos índices se tendrá que esta será igual a menos uno, lo que significa que son complementarios, consecuentemente puede calcularse un indicador como de necesidades o de satisfacción.

CAPÍTULO 3.

INVESTIGACIÓN SOBRE APLICACIONES EN LA CONSTRUCCIÓN DE ÍNDICES CON BASE EN LAS TÉCNICAS DE ESCALAMIENTO ÓPTIMO.

En el presente capítulo se exponen unas aplicaciones del procedimiento del Análisis de Componentes Principales no lineal.

En primer lugar se presenta el caso SELBEN (Sistema de Selección de Beneficiarios) para la distribución del bono de desarrollo humano; el cual tiene como antecedentes el modelo colombiano y chileno, siendo una poderosa herramienta que permite asignar dicho beneficio de manera más justa.

Adicionalmente se presenta una aplicación desarrollada por el CEPLAES (Centro de Planificación de Estudios Sociales) para medir la participación ciudadana, otra en el campo de la intervención social en el Ecuador desarrollado por el CIMACYT y finalmente se presenta una aplicación en el campo de la psicología desarrollado por la Universidad de Oviedo que mide la inhibición conductal en preescolares en el año 2003.

3.1. EL CASO SELBEN PARA DISTRIBUCION DEL BONO DE DESARROLLO HUMANO

El SELBEN registra a las potenciales familias beneficiarias de subsidios y servicios sociales, identifica su nivel socioeconómico y califica su elegibilidad para recibir dichos beneficios.

Este sistema permitió focalizar de mejor manera beneficios y servicios sociales para familias que se encuentran en condiciones de indigencia o pobreza.

El sistema parte de un formulario de caracterización socioeconómica, dicho instrumento recaba información con el propósito de determinar su posible participación en ciertos beneficios o programas sociales; el instrumento contempla la recopilación de información para siete grupos de variables que ayudaron a determinar si una persona debe o no participar de los beneficios mencionados.

El tipo de variables utilizadas fueron cualitativas ya que son más fáciles de medir que variables como el ingreso o gasto; el conjunto de variables utilizadas fueron:

- Características demográficas.
- Educación.
- Salud.
- Vivienda.
- Servicios.
- Bienes y.
- Empleo.

La información del formulario consecuentemente sirvió a su vez, para construir un "índice de situación socioeconómica" de los hogares, cuyo objetivo es diferenciar a las familias según su nivel de bienestar.

En la construcción de este indicador se observa que variables tienen mayor poder de discriminación o segregación para dividir a los individuos de acuerdo

a su nivel de vida. A partir de este índice se determinó un punto de corte que permite identificar a quienes eran los posibles beneficiarios de subsidios y acciones sociales.

El sistema se implementó en todas las provincias del país, pero priorizando las parroquias rurales y las ciudades fueron incluidas en cada etapa del programa según la incidencia de pobreza, en las parroquias rurales según la información que se disponía en el mapa de la pobreza, y la disponibilidad de servicios básicos, en las parroquias urbanas.

La primera etapa duró seis meses durante los cuales se realizaron 300.000 encuestas en todo el país: 180.000 encuestas en áreas rurales y 120.000 en las ciudades.

El levantamiento de información fue distinto en áreas urbanas y rurales, definidas según la división político administrativa; en el sector urbano, el levantamiento de la información se realizó mediante visitas a todos los hogares de las parroquias o barrios seleccionados; con respecto al área rural la información se recopiló en las parroquias priorizadas mediante un proceso de convocatoria de las familias a un sitio previamente identificado que facilite el acceso de la población (por ejemplo, escuelas, colegios o centros comunitarios).

Se levantaron 4.000 encuestas en sectores urbanos y rurales amanzanados de la sierra y la costa; en donde se seleccionaron zonas con niveles medios de pobreza, para evaluar la capacidad de discriminación del índice.

Dicho índice tiene como objetivo diferenciar a las familias según su nivel de bienestar y se lo construye mediante una combinación lineal de variables de condiciones de vida que tienen el mayor poder de discriminación para ordenar a los hogares de acuerdo a su nivel de bienestar, esta metodología se basó en

una propuesta elaborada inicialmente por el CONAM y que fue afinada finalmente por el SIISE.

Para seleccionar los pesos se usó el método expuesto en el capítulo dos del presente trabajo maximizando así la correlación lineal entre dos variables, es decir se aumenta la distancia entre elementos o individuos (beneficiarios), diferenciando de mejor manera entre casos ambiguos.

Se examinó una serie de variables que permiten describir las condiciones de vida de los hogares de los sectores educación, salud, vivienda, servicios, bienes y empleo. Las variables fueron escogidas a base de pruebas de consistencia, estabilidad y por la facilidad de ser medidas. Se optó por las siguientes variables:

- **Ubicación geográfica (diferencia de los hogares según la región y área)**
- **Características de la vivienda y acceso a servicios básicos**
 - Material del piso de la vivienda.
 - Tipo de servicio eléctrico.
 - Disponibilidad de ducha.
 - Tipo de servicio de eliminación de excretas.
 - Combustible que utiliza para cocinar.
 - Número de personas del hogar por dormitorio.
- **Tenencia de bienes durables**
 - Disponibilidad de aparato de cocina/cocineta para preparar alimentos.
 - Disponibilidad de televisión a color.
 - Disponibilidad de refrigeradora.
 - Disponibilidad de líneas telefónicas.
 - Disponibilidad de carros/automóviles.

- Disponibilidad de equipo de sonido.
- Disponibilidad de VHS.

- **Recursos productivos**
 - Tenencia de tierra para cultivos.
 - Tenencia de crédito bancario.
- **Aspectos demográfico-culturales**
 - Número de niños menores de 6 años.
 - Número de hijos nacidos vivos que han muerto.
 - Idioma que habla el jefe del hogar.
 - Miembros del hogar en edad de trabajar que no perciben ingresos.

- **Educación**
 - Nivel de instrucción del jefe/a del hogar.
 - Nivel de instrucción del/la cónyuge del jefe del hogar.

- **Seguridad social**
 - Afiliación a algún seguro por parte del jefe/a de hogar.

- **Salud**
 - Número de miembros del hogar con alguna discapacidad.

Así se determinó que las variables que más alta relación tienen con el índice son la disponibilidad de televisión a color, la tenencia de ducha, el tipo de servicio de

eliminación de excretas, el material del piso de la vivienda y el acceso a líneas telefónicas

Para medir la confiabilidad del índice se calculó su correlación con el nivel de consumo per cápita; donde el índice de bienestar mide pobreza estructural, y no la capacidad de consumo de los hogares, por lo que estas medidas no necesariamente coinciden, pero se supone que debe existir una correlación positiva entre ellas y se encontró una correlación entre el índice y el consumo per cápita de 0.626, valor que nos indica una alta asociación.

Además se calcularon proporciones de los quintiles de consumo y quintiles del índice de bienestar (Tabla 3.1.1):

Tabla 3.1.1

		Quintiles del índice de bienestar				
		1	2	3	4	5
Quintiles de consumo	1 (Más pobres)	65,5%	23,0%	8,8%	2,6%	0,1%
	2	24,5%	37,6%	24,0%	12,5%	1,3%
	3	8,1%	27,4%	33,9%	25,0%	5,7%
	4	1,6%	9,4%	28,2%	37,6%	23,2%
	5 (Menos pobres)	0,3%	2,6%	5,1%	22,3%	69,6%
		100%	100%	100%	100%	100%

Se observa que el 90% de quienes tienen la mayor opción de ser identificados como beneficiarios pertenecen a los 2 quintiles más pobres, habiendo apenas un 0.3% de hogares del quintil más rico que podría ser identificados como beneficiarios, lo que con restricciones a priori para ser calificados al sistema de bonos, permitiría filtrar casi totalmente los casos atípicos.

Dado que por practicidad se debe contar con un número adecuado de variables que sirvan para discriminar a los individuos e identificar a la población pobre; así para seleccionar las variables necesarias se consideraron aquellas que más aportaban a la varianza del primer eje factorial; entonces para analizar la estabilidad se calculan las correlaciones entre los índices que cuentan con un número distinto de variables incluidas, así:

- Se calcula el índice uno (I_1) con n_1 variables
- Calcular el índice (I_2) con n_2 variables (siendo $n_1 < n_2$)
- Calcular el índice (I_3) con n_3 variables (siendo $n_1 < n_2 < n_3$)
- Calcular la correlación lineal entre los índices I_1, I_2, I_3 .

Si la correlación entre los índices supera el valor de 0.9, se puede considerar que el índice es estable.

Luego de seguir el procedimiento señalado, se encontraron correlaciones superiores a 0.98, lo que significa que agregar o quitar una variable no incide en el resultado ni la utilidad final del índice.

Comprobado el buen funcionamiento del índice, es conveniente determinar los puntos de corte, los mismos que son arbitrarios y dependen del criterio de comparación que se desee; como un ejemplo se analizó un punto de corte equivalente al de la línea de pobreza utilizado en el mapa de pobreza⁹

Para establecer un valor equivalente a la línea de pobreza e indigencia, se construyeron modelos de regresión lineal tomando como variable dependiente el INDICE y como variable independiente el logaritmo natural del consumo per cápita mensual familiar, es decir:

$$\text{Índice} = a + b \cdot \ln(\text{Consumo})$$

donde a, y b son los parámetros del modelo a estimar.

⁹ La línea de pobreza utilizada corresponde a la que se utilizó para calcular la pobreza a partir de la ECV95 obtenida por el Banco Mundial y proyectado posteriormente al Censo de población y vivienda de 1990. Los valores fueron 90701 y 45900 sucres quincenales per cápita respectivamente para las líneas de pobreza e indigencia.

Este procedimiento se aplicó a nivel nacional y por cada una de las regiones del país, obteniendo los siguientes modelos (y coeficientes de determinación):

$$\text{Índice} = -115.8558352 + 12.523502914 \ln(\text{Consumo})$$

Es posible obtener los niveles del consumo para diferentes valores del índice, tomando el exponencial de la función inversa.

La descripción aquí presentada es un resumen de la construcción del índice utilizado por el SELBEN, y que demuestra la utilidad del Análisis de Componentes Principales no lineal especialmente cuando no se cuenta con variables numéricas que diferencien de mejor manera un conjunto de individuos, hogares o sector geográfico, a continuación se presentan algunas aplicaciones adicionales.

3.2 OTRAS APLICACIONES

Dentro de las distintas ramas en las cuales es aplicable el método expuesto presentamos una descripción general de una aplicación para medir la participación ciudadana en los distintos municipios del país desarrollado por el CEPLAES y otra en el campo de la intervención social en el Ecuador desarrollado por el CIMACYT.

3.2.1 ÍNDICE DE PARTICIPACIÓN CIUDADANA

La utilización del método PRINCALS surge por la necesidad de medir el grado de participación ciudadana en 32 municipios sin caer en criterios subjetivos de personas expertas en el tema. En dichos municipios se realizaron entrevistas a dos grupos objetivo, el uno era el conjunto conformado por dirigentes y el otro por representantes de elección popular por lo que se contaban con dos formularios.

Utilizando estos dos formularios se necesitaba observar que variables aportaban más a un índice de este tipo, donde se evaluaron variables cualitativas como:

- Días a la semana que dedica al trabajo en la organización.
- Personas que conforman la organización (mayoría).
- Actividades de la organización donde hay mayor participación.
- Actividades de la organización donde participan mujeres.
- Actividades de la organización donde participan hombres.
- Tipo de información que utilizan para la definición de actividades.
- Miembro que dispone la información para los asuntos de la organización.
- Calificación de la gestión de los líderes sobre administración de recursos.
- Calificación de la gestión de los líderes sobre ejecución de obras.
- Calificación de la participación de los miembros de la organización.
- La instrucción índice en mayor participación de las personas.
- Afecta la participación de los miembros la falta de reconocimiento legal de la organización.
- La falta de reconocimiento legal de la organización impide influir en las decisiones.
- Participación que despierta mayor entusiasmo en la comunidad.
- Las formas de participación establecidas en la constitución cubren las expectativas.

Evaluando dichas variables con sus respectivas categorías se obtuvo el índice que mide la participación ciudadana; siendo el cantón Lago Agrio de la provincia de Sucumbíos el que consta de una mayor participación ciudadana y el cantón Rumiñahui de Pichincha el que tiene una menor participación de la comunidad.

3.2.2 TIPOLOGÍA DE LA INTERVENCIÓN SOCIAL EN EL ECUADOR

El trabajo realizado por el CIMACYT () se centraba en observar la distribución de los beneficios que ofrecen algunos programas sociales como el ORI, Programa de Alimentación Escolar (PAE) y el acceso al Bono de Desarrollo Humano.

Y adicionalmente para observar cuál es la oferta de educación y salud; se utilizaron variables como el número de alumnos por profesor, número de personal médico.

Se construyó el índice de oferta social (IOS), mediante el método (PRINCALS), las poblaciones de referencia son proyecciones con base en el censo de población y vivienda 2001; así se desarrolló una segmentación a nivel cantonal, para todo el país en la cual se mostró la existencia de tres grupos consistentes de cantones, así como la deficiencia en la cobertura de los programas sociales en determinados sectores.

Con la técnica mencionada se obtuvo que los tres cantones con peores coberturas por los programas sociales fueran el cantón Palanda de la provincia Zamora Chinchipe, Montalvo en la provincia de Los Ríos y el peor cantón es el de Las Golondrinas que se encuentra en una zona no delimitada.

Y los tres cantones con mejor cobertura fueron Pablo VI en Morona Santiago y Santa Clara Mera y Pastaza.

De esta manera se logró determinar que es importante que los programas sociales como el ORI, PAE, Bono de desarrollo humano y demás, es tomar en cuenta a los cantones pertenecientes al grupo de alta pobreza y Zonas no delimitadas, para prestar sus beneficios, ya que éstos presentan altas incidencias de la pobreza y bajísimas coberturas.

3.2.3 ESCALA DE INHIBICIÓN PARA PREESCOLARES

Dada la necesidad de medir el grado de inhibición de los niños puesto que esta se encuentra relacionada con ciertos problemas que se presentan en la adolescencia la Universidad de Oviedo presenta un mecanismo para medir dicha característica en preescolares usando el método expuesto en este trabajo.

Esta herramienta surge como respuesta a la falta de instrumentos que permitan medir dicho comportamiento puesto que la forma más utilizada de evaluar dicha característica era a través del laboratorio y la observación en un medio ambiente determinado o ya sea la utilización de escalas de temperamento general; de este modo se desarrollo un instrumento de medición luego de recoger algunos criterios de profesionales en la materia y se aplicó a los padres y profesores de niños en preescolar en la localidad de Montcada y Reixach durante el año 2000, dicho formulario constaba de las siguientes preguntas en las cuales las posibles respuestas eran nunca, alguna vez, casi siempre o siempre.

- ¿Cómo reacciona el niño ante otro niño de aproximadamente la misma edad que no le es familiar:?
 - ¿Presenta una expresión facial de miedo o preocupación?
 - ¿Evita entrar en contacto con él otro niño?
 - ¿Se pone a llorar?
 - ¿Se aferra al maestro o maestra?
 - ¿Deja de jugar o hacer lo que estaba haciendo?
 - ¿Deja de hablar o cantar?
 - ¿Queda paralizado?
 - ¿Hay que animarlo para que se acerque?
 - ¿Enseguida le coge confianza?
- ¿Cómo se comporta en el patio o en una situación de juego en el aula?
 - ¿Pasa más tiempo jugando solo que en grupo?
 - ¿Le cuesta relacionarse con los otros niños de su edad?
 - ¿Se pasa la mayor parte del tiempo llorando?
 - ¿Evita los lugares y juegos de riesgo (ej. Lugares altos)?

- ¿Juega enseguida con la mayoría de niños, aunque no los conozca?

Así gracias al método de Análisis de Componentes Principales no lineales se pudo determinar que una de las características más importantes en un niño inhibido es la necesidad de que el padre le anime al niño a que se acerque a otro que no conoce (**Hay que animarlo para que se acerque**), presenta una actitud facial de miedo o preocupación ante el desconocido y adicionalmente evita entrar en contacto con él.

Evaluando las características de inhibición del niño en su entorno educativo se pudo encontrar que los niños que presentan un grado inhibición les caracteriza el hecho de que les cuesta relacionarse con la mayoría de los niños de su edad y también le cuesta jugar enseguida con los otros niños.

De este modo se presenta la utilidad del método en distintas áreas en donde la cuantificación no es posible y es necesario crear herramientas de tipo cualitativo que nos den la idea de un problema o situación sin entrar en juicios subjetivos de cierto grupo de expertos o entendidos.

En el siguiente capítulo se presenta la aplicación de este método en el Censo de Población y Vivienda del 2001 en Ecuador y en la Encuesta de Condiciones de Vida 2005 y 2006.

CAPÍTULO 4.

APLICACIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES NO LINEAL PARA LA CONSTRUCCIÓN DE ÍNDICES DE CONDICIONES DE VIDA.

En el presente capítulo se presentan las aplicaciones del algoritmo PRINCALS en el Censo de Población y Vivienda del 2001 y la Encuesta de Condiciones de Vida 2005 – 2006, con la idea de comparar los índices calculados con cada una de las bases.

De este modo para que los indicadores sean comparables debemos utilizar el conjunto de variables comunes en las dos bases de datos y así reproducir el modelo a nivel nacional para el caso del Censo y por los dominios correspondientes en la Encuesta de condiciones de vida.

4.1. APLICACIÓN CON EL CENSO DE POBLACIÓN Y VIVIENDA DEL 2001

Para la construcción del indicador usaremos una muestra del Censo que sea representativa con desglose provincial y así luego de construir el modelo con la muestra mencionada se reproducirá a nivel de toda la base de datos.

El hecho de que se tomen en cuenta inicialmente un conjunto de variables comunes no implica que estas sean las que formen parte del modelo definitivo, sino que se estudiará que variables son las que más aportan al modelo en construcción, de este modo las variables que se utilizarán inicialmente son:

- Tipo de Vivienda.
- Material predominante del techo de la vivienda.
- Material predominante de las paredes exteriores de la vivienda.
- Material predominante de piso de la vivienda.

- Cómo se obtiene el agua de la vivienda.
- De donde proviene el agua que recibe para la vivienda.
- Forma de eliminación de aguas negras o servidas.
- Posee servicio de luz eléctrica en la vivienda.
- Posee servicio telefónico para la vivienda.
- Método de eliminación de la basura.
- Hacinamiento.
- Principal combustible que se usa para cocinar.
- Tipo de servicio higiénico con el que cuenta la vivienda.
- Acceso a ducha en la vivienda.
- La vivienda es propia, arrendada, etc.
- Presencia de personas de 18 años y más analfabetas.
- Educación del jefe del hogar.

En el Anexo 2 se presenta la descripción de todas las variables mencionadas anteriormente.

4.1.1 SELECCIÓN DE LA MUESTRA

Como se mencionó inicialmente se tomará una muestra de los hogares del Censo de Población y Vivienda 2001 para construir el modelo y luego replicarlo en toda la base y obteniendo un índice, que puede ser agregado a cualquier nivel geográfico.

Para esto usaremos el método de muestreo aleatorio simple pero tomando en cuenta la representación de cada provincia en la muestra obtenida, el número de hogares y los pesos correspondientes se presentan en la tabla 4.1.1.1.

Se cuentan con 2887087 hogares de tal manera que usando muestreo aleatorio simple con más del 99% de confiabilidad, más precisamente con un error del 0,5% y con la máxima heterogeneidad $p=0.5$ se tiene una muestra de 37912 hogares, pero para efectos de practicidad utilizamos una muestra de 32000

hogares, así se asegura que en esta muestra se encuentre un comportamiento similar al del universo de hogares con respecto al conjunto de variables mencionadas.

Así en la tabla 4.1.1.1 se muestra la repartición de los hogares según el peso que tiene cada provincia.

Tabla 4.1.1.1

n	Provincia	Número de hogares	Porcentaje de Participación	Muestra
1	Azuay	145394	5,0%	1592
2	Bolívar	39647	1,4%	432
3	Cañar	49410	1,7%	500
4	Carchi	36845	1,3%	411
5	Cotopaxi	80998	2,8%	890
6	Chimborazo	99933	3,5%	1089
7	El Oro	126405	4,4%	1333
8	Esmeraldas	84534	2,9%	929
9	Guayas	784630	27,2%	8763
10	Imbabura	82493	2,9%	994
11	Loja	94468	3,3%	1088
12	Los Ríos	147716	5,1%	1617
13	Manabí	256018	8,9%	2772
14	Morona santiago	24062	0,8%	253
15	Napo	15096	0,5%	163
16	Pastaza	13340	0,5%	140
17	Pichincha	613345	21,2%	6858
18	Tungurahua	110621	3,8%	1214
19	Zamora Chinchipe	16364	0,6%	186
20	Galápagos	4918	0,2%	45
21	Sucumbíos	27810	1,0%	344
22	Orellana	17234	0,6%	196
23	Zonas No Delimitadas	15806	0,5%	191
	TOTAL	2887087	100%	32000

Ahora examinemos las características de algunas de las variables con respecto a la muestra dada

Por ejemplo, observemos como se encuentran los hogares con respecto a las variables de interés para el modelo.

Inicialmente puede observarse que las provincias de Bolívar (84%) y Cañar (84%) tienen el porcentaje más alto de hogares que viven en una casa o villa y que las provincias en las cuales la proporción de hogares que tienen casa o villa es menor son Pichincha, 55% y Pastaza, 58%, pero así mismo Pichincha es la provincia en la que los hogares poseen el mayor porcentaje de departamentos, 23% como puede observarse en el Anexo 3.

A nivel global se tiene que el 72% de los hogares del país poseen casa o villa, el 10% viven en departamentos, es decir que más del 80% de los hogares ecuatorianos viven en casa o departamento, siendo esto una posible indicatriz de que esta variable no aporte como se espera en la construcción del modelo ya que no ayudaría a discriminar a los hogares.

Con respecto al material predominante del techo de la vivienda puede observarse que a nivel de país cerca del 45% de los hogares tiene una vivienda cuyo techo es de zinc; adicionalmente el 56% de los hogares en Pichincha tienen una vivienda en la que el techo es de loza u hormigón, así como un 38% en Tungurahua.

La provincia que cuenta con mayor porcentaje de viviendas con techo de asbesto o eternit es la provincia de Galápagos, 46% y la provincia en la que las viviendas tienen un mayor porcentaje de techo de paja es la de Morona Santiago (17%).

Para ver si la muestra tiene un comportamiento similar a todo el conjunto de hogares, se presenta en la tabla 4.1.1.2 la distribución de la variable tipo de vivienda para todos los hogares y para la muestra.

Tabla 4.1.1.2

	Tipo de Vivienda	Universo Censo 2001	Muestra
	Casa o Villa	71,90%	71,80%
	Departamento	9,90%	9,80%
	Cuarto en casa de Inquilinato	6,70%	6,80%
	Mediagua	6,00%	6,10%
	Rancho	3,10%	3,20%
	Covacha	1,20%	1,20%
	Choza	0,70%	0,70%
	Otras particulares	0,30%	0,40%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

De igual manera, para la variable que describe el material del techo de la vivienda se presenta en la tabla 4.1.1.3, la distribución para ambos casos.

Tabla 4.1.1.3

	Material del techo de la vivienda	Universo Censo 2001	Muestra
	Loza de hormigón	26,4%	26,3%
	Asbesto/Eternit	11,8%	11,8%
	Zinc	44,5%	44,3%
	Teja	13,4%	13,8%
	Paja	1,6%	1,5%
	Otros	2,4%	2,2%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

Y como puede verse en las dos tablas anteriores, no existen diferencias significativas entre las distribuciones del Censo del 2001 y la muestra, siendo esta representativa y por lo tanto útil para construir el indicador.

Continuando con el análisis de algunas de las variables veamos como se encuentra la distribución de la variable que describe el material del piso de la vivienda en los hogares ecuatorianos.

Como puede verse en el Anexo 3.3 el 36% de los hogares ecuatorianos tienen en su vivienda piso entablado o parquet y un 33% tienen un piso de cemento o ladrillo, las provincias que tienen el porcentaje más alto de viviendas con piso entablado o parquet son las de Zamora Chinchipe, 63% y la de Orellana (59%); entre tanto las provincias que presentan un porcentaje más alto de viviendas con

piso de baldosa o vinyl son Galápagos, 27% y Guayas, 22%; finalmente las provincias con peores condiciones con respecto a la variable en cuestión son Manabí, 18% de viviendas con piso de caña, Los Ríos 12% de viviendas con piso de caña, Chimborazo, 37% de viviendas con piso de tierra y Bolívar, 28 % de viviendas con piso de tierra.

En el Anexo 3.4 puede verse la distribución de la variable que describe la forma de obtener el agua para la vivienda, observándose que a nivel global el 52% de los hogares ecuatorianos obtienen el agua por tubería dentro de la vivienda, que es un buen indicador de la calidad de vida de los hogares ecuatorianos, pero en cambio se tiene también que tres provincias y las zonas no delimitadas poseen un alto porcentaje de hogares que en su vivienda no reciben agua por tubería, así Zonas no Delimitadas, 60%, Orellana, 60%, Sucumbíos, 59% y Esmeraldas, 47% con lo que puede concluirse que esta variable probablemente aportará al modelo en construcción.

Ahora con respecto a la procedencia del agua para la vivienda (ver Anexo 3.5) se observa que el 67% de los hogares ecuatorianos tienen acceso a agua a través de red pública, pero Carchi, 86% y Pichincha, 84% son las que mejor se encuentran con respecto al acceso a agua y mientras que Orellana, 33%, Sucumbíos, 28% y las Zonas no Delimitadas, 25% son las provincias y sectores que peor se encuentran con respecto a esta variable.

En términos globales el 48% de los hogares en el Ecuador cuentan con un sistema de eliminación de aguas negras o servidas a través de una red pública de alcantarillado (ver Anexo 3.6), el 19% a través de pozo séptico y un 16% usan pozo ciego; a nivel provincial nuevamente Pichincha, 77% y Carchi, 67% se destacan con respecto a la variable en mención, entre tanto en Galápagos el 59% de los hogares usan como sistema de eliminación de excretas un pozo séptico y un 30% de las Zonas no Delimitadas usan pozo ciego.

A continuación se presentan las comparaciones entre la muestra y todo el conjunto de hogares, de las variables descritas.

Tabla 4.1.1.4

	Material Predominante del Piso de la Vivienda	Universo Censo 2001	Muestra
	Enablado o parquet	36,20%	35,90%
	Baldosa o vinyl	12,90%	13,20%
	Ladrillo o cemento	33,20%	33,20%
	Caña	4,00%	3,90%
	Tierra	11,10%	11,10%
	Otros	2,70%	2,60%

Fuente: Censo de Población y Vivienda del 2001 - INEC

Tabla 4.1.1.5

	Forma de obtener el agua para la vivienda	Universo Censo 2001	Muestra
	Por tubería dentro de la vivienda	52,10%	51,90%
	Por tubería fuera de la vivienda, pero dentro del edificio,	20,90%	21,30%
	Por tubería fuera del edificio, lote o terreno	5,10%	5,00%
	No recibe agua por tubería	21,90%	21,80%

Fuente: Censo de Población y Vivienda del 2001 - INEC

Tabla 4.1.1.6

	Procedencia del agua para la vivienda	Universo Censo 2001	Muestra
	De red publica	67,50%	67,40%
	De pozo	12,50%	12,60%
	De río, vertiente, acequia o canal	12,20%	12,20%
	De carro repartidor	6,00%	5,90%
	Otro	1,90%	1,90%

Fuente: Censo de Población y Vivienda del 2001 - INEC

Tabla 4.1.1.7

	Forma de eliminación de aguas negras o servidas	Universo Censo 2001	Muestra
	Por red publica de alcantarillado	48,00%	47,90%
	Pozo ciego	15,60%	16,00%
	Pozo séptico	19,10%	18,70%
	Otras	17,30%	17,30%

Fuente: Censo de Población y Vivienda del 2001 - INEC

Como se ve en las cuatro tablas anteriores no existen diferencias significativas entre la muestra y el universo de hogares ecuatorianos, lo que sigue corroborando la buena representación de la muestra. A continuación se observa como se encuentra la situación de los hogares con respecto al acceso a luz eléctrica (Anexo 3.7), donde se observa que el 90% de los hogares ecuatorianos cuentan

con el servicio de electricidad en su vivienda, con lo que esta variable podrá ayudar a identificar que provincias se encuentran en peores condiciones, particularmente Galápagos es la provincia que tiene mayor cobertura de luz eléctrica, 98% y las provincias que cuentan con menor cobertura son Orellana, 47%, Morona Santiago, 42% y Napo, 39% es decir las provincias del Oriente se encuentran desatendidas con respecto al servicio de luz eléctrica.

Así mismo en el Anexo 3.7 se observa que más de la tercera parte de los hogares en el Ecuador no cuentan con servicio telefónico, 68% denotando una gran deficiencia de la cobertura de las empresas telefónicas en el país especialmente en las Zonas No Delimitadas, 91% y en provincias como Sucumbíos, 89%, Orellana, 89% y Bolívar en la sierra central, 88% dando un indicio de que será una variable que colaborará en la construcción del indicador.

Con respecto a las comparaciones entre el universo y la muestra, se observa en las tablas 4.1.1.8 y 4.1.1.9 que las distribuciones son prácticamente las mismas.

Tabla 4.1.1.8

	Disponibilidad de Luz Eléctrica	Universo Censo 2001	Muestra
	Si dispone	89,70%	89,70%
	No dispone	10,30%	10,30%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

Tabla 4.1.1.9

	Disponibilidad de Servicio Telefónico	Universo Censo 2001	Muestra
	Si dispone	32,20%	32,10%
	No dispone	67,80%	67,90%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

Ahora siguiendo con la revisión de las variables puede observarse en el Anexo 3.8 que más del 60% de los hogares tienen acceso a un sistema de recolección de basura a través de un carro recolector, lo que es un indicio de que este sector de la población tiene una mejor condición de vida, especialmente en provincias como Galápagos, 93%, Pichincha, 85% y Guayas, 73%.

Así mismo, existen provincias en las cuales la forma de eliminación de la basura se realiza lanzando a terrenos baldíos o quebradas como por ejemplo en Napo, 43%, Bolívar, 41% y Zamora Chinchipe, 38% lo que muestra una menor calidad de vida de estas provincias.

La comparación que se tiene en la tabla 4.1.1.10 muestra que al igual que las variables anteriormente descritas, la muestra calculada tiene un comportamiento prácticamente similar lo que sigue demostrando la fiabilidad de la muestra que será utilizada para implementar el modelo.

Tabla 4.1.1.10

	Método de Eliminación de Basura	Universo Censo 2001	Muestra
	Por carro recolector	62,70%	62,80%
	En terreno baldío o quebrada	14,70%	14,50%
	Por incineración o entierro	17,90%	17,70%
	Otra forma	4,80%	4,90%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

Siguiendo con el hacinamiento¹⁰, puede observarse en el Anexo 3.9 que el 85% de los hogares ecuatorianos son no hacinados; pero en provincias como Orellana, Morona Santiago y Pastaza más de la cuarta parte de los hogares se encuentran hacinados, lo que muestra nuevamente la menor calidad de vida de las provincias del oriente del Ecuador, sin dejar de lado que en la provincia de Esmeraldas, más del 20% de los hogares tienen problemas de hacinamiento; es decir, esto indica que esta variable ayudará a discriminar dentro del modelo.

En forma global puede verse en la tabla 4.1.1.11 las distribuciones globales y de la muestra con respecto al hacinamiento.

Tabla 4.1.1.11

	Hacinamiento	Universo Censo 2001	Muestra
	No	84,70%	84,50%
	Si	15,30%	15,50%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

¹⁰ Se define que un hogar se encuentra hacinado si el número de personas por dormitorio es mayor a tres

Con respecto al combustible para cocinar puede mirarse en el Anexo 3.10 que el 84% de los hogares cuentan con gas para cocinar, pero un 13% de los hogares en el país cocina con leña o carbón lo que muestra una menor calidad de vida, más aun en provincias como Bolívar, 54%, Chimborazo, 41%, Cotopaxi y Morona Santiago, 39%; con respecto a la comparación entre la muestra y el universo de hogares, se tiene en la tabla 4.1.1.12 que se sigue manteniendo las similitudes.

Tabla 4.1.1.12

	Principal combustible para cocinar	Universo Censo 2001	Muestra
	Gas	84,50%	84,30%
	Electricidad	0,80%	0,80%
	Gasolina	0,10%	0,10%
	Kérex o diesel	0,30%	0,30%
	Leña o carbón	13,00%	13,20%
	Otros	0,10%	0,10%
	Ninguno/no cocina	1,10%	1,20%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

En el Anexo 3.11 puede apreciarse que el 64% de los hogares ecuatorianos cuentan con excusado de uso exclusivo especialmente en Galápagos, 80%, Pichincha, 75% y El Oro que al igual que Manabí cuentan con un 68% de hogares que cuentan con excusado de uso exclusivo, pero en cambio un 17% del país no cuenta con ningún tipo de excusado lo que denota una mala calidad de vida y que se presenta más aun en las provincias de Morona Santiago, 48%, Orellana, 46%, Zamora Chinchipe, 44%, Napo, 40% y Sucumbíos 31%, lo que afirma que el oriente ecuatoriano se encuentra en las peores condiciones con respecto a condiciones de vida; en la tabla 4.1.1.13 se ven las mínimas diferencias entre la población total de hogares y la muestra con respecto a la variable en cuestión.

Tabla 4.1.1.13

	Servicio Higiénico con que cuenta	Universo Censo 2001	Muestra
	Excusado de uso exclusivo	64,50%	64,20%
	Excusado de uso común	11,10%	11,30%
	Letrina	7,80%	8,00%
	No tiene	16,60%	16,60%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

El Anexo 3.12 muestra que el 55% de los hogares en el Ecuador cuentan con ducha de uso exclusivo, pero en cambio es notorio que un 34% no cuentan con ducha en su hogar lo que implica que este sector de la población es vulnerable a ciertos problemas de salud lo que se traduce en una mala calidad de vida, que se resalta en la provincia de Orellana, 66% y en Cotopaxi, 61%, la tabla 4.1.1.14 muestra la respectiva comparación entre el censo y la muestra.

Tabla 4.1.1.14

	Tipo de Ducha con que cuenta	Universo Censo 2001	Muestra
	De uso exclusivo	55,20%	54,80%
	De uso común	11,80%	11,70%
	No tiene	33,00%	33,50%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

En el Anexo 3.13 puede observarse que el 23% de los hogares arriendan un lugar para vivir, es decir este es un grupo que debe destinar una parte importante de sus ingresos para pagar el alquiler de la vivienda en donde residen, siendo Pichincha, 38% la provincia que cuenta con el porcentaje más alto de hogares que rentan el lugar donde viven, quizás tal vez a la gran cantidad de migrantes de otras provincias hacia la ciudad de Quito.

La tabla 4.1.1.15 muestra las comparaciones entre el censo y la muestra de esta variable.

Tabla 4.1.1.15

	La vivienda que habita es:	Universo Censo 2001	Muestra
	Propia	67,30%	67,30%
	Arrendada	23,40%	23,20%
	Anticresis	0,40%	0,40%
	Gratuita	5,60%	5,70%
	Por servicios	2,60%	2,70%
	Otros	0,70%	0,80%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

El anexo 3.14 muestra el porcentaje de hogares con la presencia de una o más personas de más de 18 años analfabetas, 30%, siendo esta proporción más notable en las provincias de Bolívar, 45%, Cotopaxi, 44% y Chimborazo, 43% lo que muestra que las provincias de la sierra central son las que se encuentran en peores condiciones con respecto a la variable en estudio.

La tabla 4.1.1.16 muestra las comparaciones entre la muestra y todos los hogares del país observándose un comportamiento similar.

Tabla 4.1.1.16

	Presencia de personas de 18 años y más Analfabetas en el hogar	Universo Censo 2001	Muestra
	No	69,60%	69,60%
	Si	30,40%	30,40%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

Finalmente en el Anexo 3.15 tenemos la educación del jefe del hogar, esta variable será usada en el modelo ya que se presume que ésta ayudará a discriminar de mejor manera a los hogares con respecto a su situación socioeconómica, así, puede verse que el 48% de los hogares en el país cuentan con un jefe de hogar que alcanzado a terminar la primaria y una cuarta parte de los hogares cuentan con un jefe de hogar que tiene educación secundaria, es notable ver que en el 14% de los hogares cuentan con un jefe de hogar que tiene instrucción superior; en la tabla 4.1.1.17 se muestran las comparaciones correspondientes.

Tabla 4.1.1.17

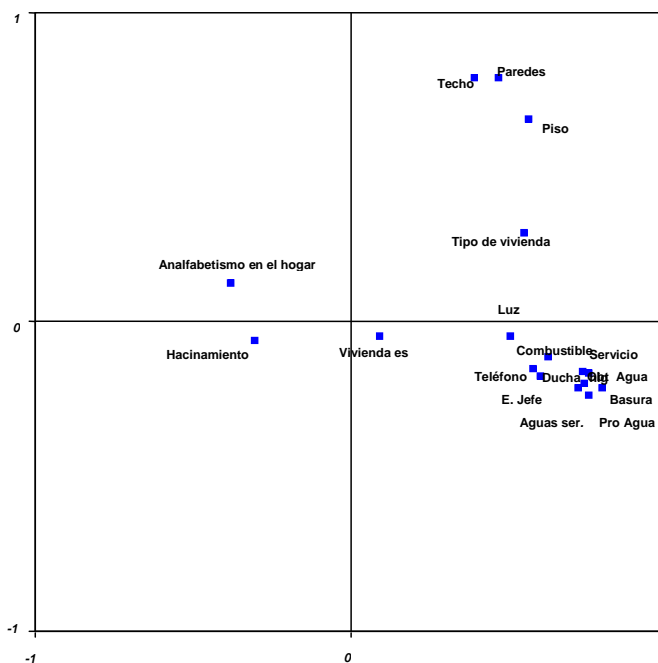
	Educación del Jefe del Hogar	Universo Censo 2001	Muestra
	Ninguno	10,30%	10,50%
	Centro de alfabetización	1,00%	1,00%
	Primaria	47,00%	47,00%
	Secundaria	24,70%	24,60%
	Ciclo Post - bachillerato	2,00%	2,00%
	Superior	13,90%	13,80%
	Postgrado	1,00%	1,10%
Fuente: Censo de Población y Vivienda del 2001 - INEC			

4.1.2 CONSTRUCCIÓN DEL MODELO.

Con las variables descritas anteriormente se procede a construir el modelo, así puede verse en el gráfico 4.1.2.1 que las variables techo de la vivienda, paredes de la vivienda, tipo de vivienda y el piso de la misma se encuentran relacionadas (lo que se esperaría); además puede observarse como las variables de acceso a servicios básicos forman otro grupo y adicionalmente éstas se encuentran relacionadas con la educación del jefe del hogar.

La presencia de analfabetismo en el hogar se encuentra relacionada con el hacinamiento en el hogar y el tipo de tenencia de vivienda con la que cuenta el hogar.

Gráfico 4.1.2.1



Para obtener las cuantificaciones óptimas asignamos a las categorías de cada variable un orden, las mismas que se presentan en la tabla 4.1.2.1 y entre tanto en la tabla 4.1.2.2 se muestran las cuantificaciones óptimas para cada variable obtenidos con el algoritmo.

Tabla 4.1.2.1

Categ	VARIABLES	Categ	VARIABLES
Tipo de vivienda		Servicio telefónico	
1	Choza	1	No
2	Covacha	2	Si
3	Rancho	Forma de eliminación de basura de la vivienda	
4	Mediagua	1	De otra forma
5	Cuarto en casa de inquilinato	2	Por incineración o entierro
6	Departamento	3	En terreno baldío o quebrada
7	Casa o villa / Otras particulares	4	De carro recolector
		Hacinamiento	
Techo de la vivienda		1	No
1	Otros materiales	2	Si
2	Paja o similares	Combustible para cocinar	
3	Teja	1	Ninguno (no cocina)
3	Zinc	2	Otro
5	Eternit	3	Leña o carbón
6	Losa de hormigón	4	Kérex o diesel
Paredes exteriores		5	Gasolina
1	Otros materiales	6	Gas
2	Caña no revestida	7	Electricidad
3	Caña revestida o bahareque	Servicio higiénico del hogar	
4	Madera	1	No tiene
5	Adobe o tapia	2	Letrina
6	Hormigón, ladrillo o bloque	3	Excusado de uso común en varios hogares
Piso de la Vivienda		4	Excusado de uso exclusivo del hogar
1	Otros materiales	Ducha en el hogar	
2	Tierra	1	No tiene
3	Caña	2	De uso común a varios hogares
4	Ladrillo	3	De uso exclusivo en el hogar
5	Baldosa o vynil	La vivienda que ocupa es	
6	Entablado o parquet	1	Otra
Cómo obtiene el agua para la vivienda		2	Por servicios
1	No recibe, agua por tubería, sino por otros medios	3	Gratuita
2	Por tubería fuera del edificio, lote o terreno	4	En anticresis
3	Por tubería fuera de la vivienda pero dentro del edificio	5	Arrendada
4	Por tubería dentro de la vivienda	6	Propia
De dónde proviene el agua que recibe		Personas de 18 años y más Analfabetas en el hogar	
1	Otro	1	No
2	De carro repartidor	2	Si
3	De Río, vertiente, acequia o canal	Educación del Jefe del Hogar	
4	De Pozo	1	Ninguno
5	De red pública	2	Centro de alfabetización
Forma de eliminación de aguas negras o servidas		3	Primaria / educación básica
1	Otra forma	4	Secundaria / educación Media
2	Por Pozo séptico	5	Ciclo Post - bachillerato
3	Por pozo ciego	6	Superior
4	Por red pública de alcantarillado	7	Postgrado
Luz Eléctrica			
1	No		
2	Si		

Fuente: Censo de Población y Vivienda del 2001 - INEC

Tabla 4.1.2.2

Categ	Variables	Cuantificaciones Óptimas	Categ	Variables	Cuantificaciones Óptimas
Tipo de vivienda			Servicio telefónico		
1	Choza	-6,33	1	No	-0,7
2	Covacha	-3,86	2	Si	1,44
3	Rancho	-3,86	Forma de eliminación de basura de la vivienda		
4	Mediagua	-0,49	1	De otra forma	-1,46
5	Cuarto en casa de inquilinato	0,27	2	Por incineración o entierro	-1,27
6	Departamento	0,27	3	En terreno baldío o quebrada	-1,27
7	Casa o villa / Otras particulares	0,27	4	De carro recolector	0,77
Techo de la vivienda			Hacinamiento		
1	Otros materiales	-5,01	1	No	-0,43
2	Paja o similares	-4,64	2	Si	2,33
3	Teja	-0,05	Combustible para cocinar		
3	Zinc	-0,02	1	Ninguno (no cocina)	-2,41
5	Eternit	0,47	2	Otro	-2,41
6	Losa de hormigón	0,54	3	Leña o carbón	-2,41
Paredes exteriores			4	Kérex o diesel	-1,95
1	Otros materiales	-5,61	5	Gasolina	-1,39
2	Caña no revestida	-2,78	6	Gas	0,41
3	Caña revestida o bahareque	-1,32	7	Electricidad	1,25
4	Madera	-0,57	Servicio higiénico del hogar		
5	Adobe o tapia	-0,57	1	No tiene	-1,84
6	Hormigón, ladrillo o bloque	0,48	2	Letrina	-1,59
Piso de la Vivienda			3	Excusado de uso común en varios hogares	0,34
1	Otros materiales	-4,02	4	Excusado de uso exclusivo del hogar	0,6
2	Tierra	-1,69	Ducha en el hogar		
3	Caña	-1,69	1	No tiene	-1,42
4	Ladrillo	0,43	2	De uso común a varios hogares	0,64
5	Baldosa o vynil	0,43	3	De uso exclusivo en el hogar	0,72
6	Entablado o parquet	0,43	La vivienda que ocupa es		
Cómo obtiene el agua para la vivienda			1	Otra	-1,45
1	No recibe, agua por tubería, sino por otros medios	-1,53	2	Por servicios	-1,45
2	Por tubería fuera del edificio, lote o terreno	-0,98	3	Gratuita	-1,45
3	Por tubería fuera de la vivienda pero dentro del edificio	-0,38	4	En anticresis	-1,45
4	Por tubería dentro de la vivienda	0,89	5	Arrendada	-1,45
De dónde proviene el agua que recibe			6	Propia	0,69
1	Otro	-1,92	Personas de 18 años y más Analfabetas en el hogar		
2	De carro repartidor	-1,5	1	No	-0,67
3	De Río, vertiente, acequia o canal	-1,5	2	Si	1,5
4	De Pozo	-1,29	Educación del Jefe del Hogar		
5	De red pública	0,69	1	Ninguno	-1,86
Forma de eliminación de aguas negras o servidas			2	Centro de alfabetización	-1,59
1	Otra forma	-1,66	3	Primaria / educación básica	-0,47
2	Por Pozo séptico	-0,51	4	Secundaria / educación Media	0,72
3	Por pozo ciego	-0,51	5	Ciclo Post – bachillerato	0,94
4	Por red pública de alcantarillado	0,95	6	Superior	1,54
Luz Eléctrica			7	Postgrado	1,72
1	No	-2,97			
2	Si	0,34			

Fuente: Censo de Población y Vivienda del 2001 - INEC

En la tabla anterior se puede ver que es necesario juntar ciertas categorías dentro de cada variable ya que el procedimiento proporciona las mismas cuantificaciones

para algunas de ellas; es así como se unen en la variable “Tipo de vivienda”, (Covacha con Rancho), además Cuarto en casa de inquilinato, Departamento y Casa o villa/ Otras particulares); en la variable “Paredes exteriores” se juntan Madera y adobe o tapia.

En “Piso de la vivienda” debemos juntar, Tierra con caña y Ladrillo con Baldosa y Entablado; en la variable “De donde proviene el agua que recibe” se unirán las categorías, De carro repartidor con de río o vertiente.

Con respecto a “La forma de eliminación de excretas” se juntarán, Pozo séptico con pozo ciego; en “Forma de eliminación de la basura”, Por incineración con terreno baldío, en la variable “Combustible para cocinar” se unen Ninguna no cocina, otro y leña con carbón); finalmente en “la vivienda que ocupa es se juntarán, Otra, por servicios, gratuita, en anticresis y arrendada.

Realizando estos cambios puede verse en la tabla 4.1.2.3 como al correr nuevamente el algoritmo se obtiene el mismo valor de la cuantificación para las categorías juntadas y produciendo variaciones muy pequeñas en el resto de cuantificaciones en las variables donde se realizaron los cambios, además en el resto de variables se mantienen los valores dados en la tabla anterior.

Tabla 4.1.2.3

Categ	Variables	Cuantificaciones Óptimas	Categ	Variables	Cuantificaciones Óptimas
Tipo de vivienda			Servicio telefónico		
1	Chozas	-6,31	1	No	-0,7
2	Covacha/Rancho	-3,86	2	Si	1,44
3	Mediagua	-0,49	Forma de eliminación de basura de la vivienda		
4	Cuarto en casa de inquilinato/Departamento /Casa o villa / Otras particulares	0,27	1	De otra forma	-1,46
Techo de la vivienda			2	Por incineración o entierro/En terreno baldío o quebrada	-1,27
1	Otros materiales	-5,04	3	De carro recolector	0,77
2	Paja o similares	-4,61	Hacinamiento		
3	Teja	-0,05	1	No	-0,43
3	Zinc	-0,02	2	Si	2,33
5	Eternit	0,47	Combustible para cocinar		
6	Losa de hormigón	0,54	1	Ninguno (no cocina)/Otro/Leña o carbón	-2,41
Paredes exteriores			2	Kérex o diesel	-1,95
1	Otros materiales	-5,64	3	Gasolina	-1,39
2	Caña no revestida	-2,77	4	Gas	0,41
3	Caña revestida o bahareque	-1,32	5	Electricidad	1,25
4	Madera/Adobe o tapia	-0,57	Servicio higiénico del hogar		
5	Hormigón, ladrillo o bloque	0,48	1	No tiene	-1,84
Piso de la Vivienda			2	Letrina	-1,59
1	Otros materiales	-4,04	3	Excusado de uso común en varios hogares	0,34
2	Tierra/Caña	-1,69	4	Excusado de uso exclusivo del hogar	0,6
3	Ladrillo/Baldosa o vynil/Entablado o parquet	0,43	Ducha en el hogar		
Cómo obtiene el agua para la vivienda			1	No tiene	-1,42
1	No recibe, agua por tubería, sino por otros medios	-1,53	2	De uso común a varios hogares	0,64
2	Por tubería fuera del edificio, lote o terreno	-0,98	3	De uso exclusivo en el hogar	0,72
3	Por tubería fuera de la vivienda pero dentro del edificio	-0,38	La vivienda que ocupa es		
4	Por tubería dentro de la vivienda	0,89	1	No Propia	-1,45
De dónde proviene el agua que recibe			2	Propia	0,69
1	Otro	-1,92	Personas de 18 años y más Analfabetas en el hogar		
2	De carro repartidor/De Río, vertiente, acequia o canal	-1,5	1	No	-0,67
3	De Pozo	-1,29	2	Si	1,5
4	De red pública	0,69	Educación del Jefe del Hogar		
Forma de eliminación de aguas negras o servidas			1	Ninguno	-1,86
1	Otra forma	-1,66	2	Centro de alfabetización	-1,59
2	Por Pozo séptico/Por pozo ciego	-0,51	3	Primaria / educación básica	-0,47
3	Por red pública de alcantarillado	0,95	4	Secundaria / educación Media	0,72
Luz Eléctrica			5	Ciclo Post - bachillerato	0,94
1	No	-2,97	6	Superior	1,54
2	Si	0,34	7	Postgrado	1,72

Fuente: Censo de Población y Vivienda del 2001 - INEC

Usando dichas cuantificaciones construimos el indicador mediante el mecanismo presentado en el capítulo dos, luego de 16 iteraciones con un ajuste total de 0,4468 y con una pérdida total de 1,5532, los valores asignados a cada categoría dentro de cada variable se muestran en la tabla 4.1.2.4.

Tabla 4.1.2.4

Categoría	Variables	Cuantificaciones Óptimas	Mínimos	Cambio de Escala	Máximos	Cuantificaciones Transformadas
Tipo de vivienda						
1	Choza	-6,31	-6,31	0	6,58	0
2	Covacha/Rancho	-3,86		2,45		4,301264045
3	Mediagua	-0,49		5,82		10,21769663
4	Cuarto en casa de inquilinato/Departamento /Casa o villa / Otras particulares	0,27		6,58		11,55196629
Techo de la vivienda						
1	Otros materiales	-5,04	-5,04	0	5,58	0
2	Paja o similares	-4,61		0,43		0,75491573
3	Teja	-0,05		4,99		8,760533708
3	Zinc	-0,02		5,02		8,813202247
5	Eternit	0,47		5,51		9,673455056
6	Losa de hormigón	0,54		5,58		9,796348315
Paredes exteriores						
1	Otros materiales	-5,64	-5,64	0	6,12	0
2	Caña no revestida	-2,77		2,87		5,038623596
3	Caña revestida o bahareque	-1,32		4,32		7,584269663
4	Madera/Adobe o tapia	-0,57		5,07		8,900983146
5	Hormigón, ladrillo o bloque	0,48		6,12		10,74438202
Piso de la Vivienda						
1	Otros materiales	-4,04	-4,04	0	4,47	0
2	Tierra/Caña	-1,69		2,35		4,125702247
3	Ladrillo/Baldosa o vynil/Entablado o parquet	0,43		4,47		7,84761236
Cómo obtiene el agua para la vivienda						
1	No recibe, agua por tubería, sino por otros medios	-1,53	-1,53	0	2,42	0
2	Por tubería fuera del edificio, lote o terreno	-0,98		0,55		0,965589888
3	Por tubería fuera de la vivienda pero dentro del edificio	-0,38		1,15		2,018960674
4	Por tubería dentro de la vivienda	0,89		2,42		4,248595506
De dónde proviene el agua que recibe						
1	Otro	-1,92	-1,92	0	2,61	0
2	De carro repartidor/De Río, vertiente, acequia o canal	-1,5		0,42		0,737359551
3	De Pozo	-1,29		0,63		1,106039326
4	De red pública	0,69		2,61		4,582162921
Forma de eliminación de aguas negras o servidas						
1	Otra forma	-1,66	-1,66	0	2,61	0
2	Por Pozo séptico/Por pozo ciego	-0,51		1,15		2,018960674
3	Por red pública de alcantarillado	0,95		2,61		4,582162921
Luz Eléctrica						
1	No	-2,97	-2,97	0	3,31	0
2	Si	0,34		3,31		5,811095506

Servicio telefónico						
1	No	-0,7		0	2,14	0
2	Si	1,44	-0,7	2,14		3,757022472
Forma de eliminación de basura de la vivienda						
1	De otra forma	-1,46		0		0
2	Por incineración o entierro/En terreno baldío o quebrada	-1,27	-1,46	0,19	2,23	0,333567416
3	De carro recolector	0,77		2,23		3,91502809
Hacinamiento						
1	Si	-0,43		0		0
2	No	2,33	-0,43	2,76	2,76	4,845505618
Combustible para cocinar						
1	Ninguno (no cocina)/Otro/Leña o carbón	-2,41		0		0
2	Kérex o diesel	-1,95	-2,41	0,46	3,66	0,80758427
3	Gasolina	-1,39		1,02		1,790730337
4	Gas	0,41		2,82		4,950842697
5	Electricidad	1,25		3,66		6,425561798
Servicio higiénico del hogar						
1	No tiene	-1,84		0		0
2	Letrina	-1,59		0,25		0,438904494
3	Excusado de uso común en varios hogares	0,34	-1,84	2,18	2,44	3,827247191
4	Excusado de uso exclusivo del hogar	0,6		2,44		4,283707865
Ducha en el hogar						
1	No tiene	-1,42		0		0
2	De uso común a varios hogares	0,64	-1,42	2,06	2,14	3,616573034
3	De uso exclusivo en el hogar	0,72		2,14		3,757022472
La vivienda que ocupa es						
1	No Propia	-1,45		0		0
2	Propia	0,69	-1,45	2,14	2,14	3,757022472
Personas de 18 años y más Analfabetas en el hogar						
1	Si	-0,67		0		0
2	No	1,5	-0,67	2,17	2,17	3,809691011
Educación del Jefe del Hogar						
1	Ninguno	-1,86		0		0
2	Centro de alfabetización	-1,59		0,27		0,474016854
3	Primaria / educación básica	-0,47		1,39		2,440308989
4	Secundaria / educación Media	0,72	-1,86	2,58	3,58	4,529494382
5	Ciclo Post - bachillerato	0,94		2,8		4,915730337
6	Superior	1,54		3,4		5,969101124
7	Postgrado	1,72		3,58		6,28511236
Fuente: Censo de Población y Vivienda del 2001 - INEC				Suma de los valores máximos	56,96	100
				Factor de Escalamiento	1,755618	

En el gráfico 4.1.2.1 puede observarse el comportamiento del indicador construido, notándose una asimetría negativa y adicionalmente el conjunto de valores tienden hacia una distribución Beta, así mismo en la tabla 4.1.2.5 se muestran las correlaciones entre el indicador y las variables utilizadas en el mismo.

Gráfico 4.1.2.1

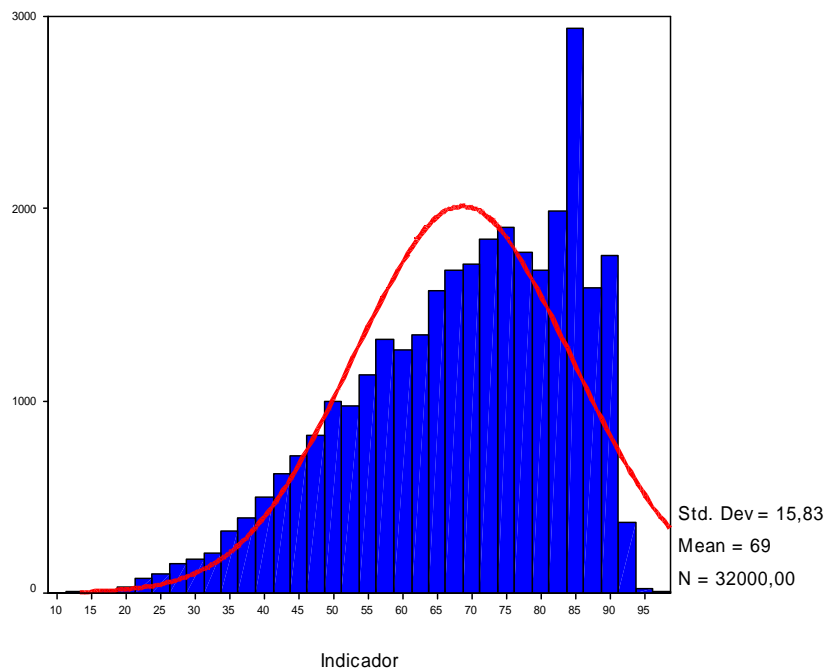


Tabla 4.1.2.5 Correlaciones

	INDICADOR
INDICADOR	1,0000000000
Forma de eliminación de aguas negras o servidas	0,7197058636
Cómo obtiene el agua para la vivienda	0,7125375205
Ducha en el hogar	0,6522921717
Servicio higiénico del hogar	0,6416903462
Forma de eliminación de basura de la vivienda	0,6328635079
De dónde proviene el agua que recibe	0,6233487097
Material del techo de la vivienda	0,5762974166
Servicio telefónico	0,5747419708
Paredes exteriores	0,5729718451
Piso	0,5489249173
Combustible para cocinar	0,5353576805
Educación del Jefe del Hogar	0,5161760595
Luz Eléctrica	0,4860826354
Tipo de vivienda	0,4574541444
Personas de 18 años y más Analfabetas en el hogar	0,1710333033
Hacinamiento	0,1439441182
La vivienda que ocupa es	-0,0868278965

Fuente: Censo de Población y Vivienda del 2001 - INEC

Como se observa en la tabla anterior, todas las variables se relacionan positivamente con nuestro indicador a excepción de la variable que describe si la vivienda es propia o no, adicionalmente las variables que describen si en el hogar existen personas analfabetas o no y la que describe si en el hogar existe hacinamiento o no, son las que presentan correlaciones bajas.

Debemos notar que nuestro modelo es una medida positiva de las características del hogar, consecuentemente los valores más altos son los que se encuentran en mejores condiciones.

Así corriendo nuevamente el algoritmo sin la variable que se encuentra relacionada inversamente se obtuvieron las siguientes cuantificaciones óptimas las mismas que se calcularon en dieciséis iteraciones con un ajuste total de 0,4711 aumentando este valor con respecto al modelo inicial; y con relación a la pérdida total se determino que esta se redujo a 1,5289 mejorando el modelo en construcción.

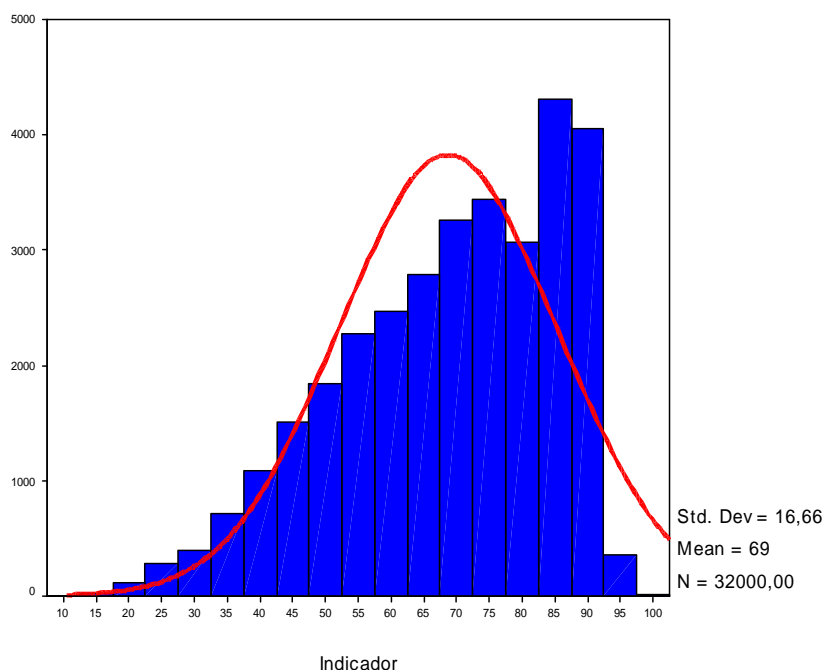
Tabla 4.1.2.6

Categoría	Variables	Cuantificaciones Óptimas	Mínimos	Cambio de Escala	Máximos	Cuantificaciones Transformadas
Tipo de vivienda						
1	Choza	-6,32	-6,32	0	6,59	0
2	Covacha/Rancho	-3,86		2,46		4,473540644
3	Mediagua	-0,48		5,84		10,62011275
4	Cuarto en casa de inquilinato/Departamento /Casa o villa / Otras particulares	0,27		6,59		11,98399709
Techo de la vivienda						
1	Otros materiales	-5,06	-5,06	0	5,59	0
2	Paja o similares	-4,6		0,46		0,83651573
3	Teja	-0,04		5,02		9,128932533
4	Zinc	-0,01		5,05		9,183487907
5	Eternit	0,45		5,51		10,02000364
6	Losa de hormigón	0,53		5,59		10,16548463
Paredes exteriores						
1	Otros materiales	-5,67	-5,67	0	6,14	0
2	Caña no revestida	-2,77		2,9		5,273686125
3	Caña revestida o bahareque	-1,32		4,35		7,910529187
4	Madera/Adobe o tapia	-0,56		5,11		9,292598654
5	Hormigón, ladrillo o bloque	0,47		6,14		11,16566648
Piso de la Vivienda						
1	Otros materiales	-4,06	-4,06	0	4,49	0
2	Tierra/Caña	-1,68		2,38		4,328059647
3	Ladrillo/Baldosa o vynil/Entablado o parquet	0,43		4,49		8,165120931

Cómo obtiene el agua para la vivienda						
1	No recibe, agua por tubería, sino por otros medios	-1,53	-1,53	0	2,42	0
2	Por tubería fuera del edificio	-0,98		0,55		1,000181851
3	Por tubería fuera de la vivienda pero dentro del edificio	-0,39		1,14		2,073104201
4	Por tubería dentro de la vivienda	0,89		2,42		4,400800145
De dónde proviene el agua que recibe						
1	Otro	-1,94	-1,94	0	2,63	0
2	De carro repartidor/De Río, vertiente, acequia o canal	-1,49		0,45		0,818330606
3	De Pozo	-1,29		0,65		1,182033097
4	De red pública	0,69		2,63		4,782687761
Forma de eliminación de aguas negras o servidas						
1	Otra forma	-1,67	-1,67	0	2,62	0
2	Por Pozo séptico/Por pozo ciego	-0,5		1,17		2,127659574
3	Por red pública de alcantarillado	0,95		2,62		4,764502637
Luz Eléctrica						
1	No	-2,97	-2,97	0	3,31	0
2	Si	0,34		3,31		6,019276232
Servicio telefónico						
1	No	-0,7	-0,7	0	2,14	0
2	Si	1,44		2,14		3,891616658
Forma de eliminación de basura de la vivienda						
1	De otra forma	-1,47	-1,47	0	2,24	0
2	Por incineración o entierro/En terreno baldío o quebrada	-1,27		0,2		0,363702491
3	De carro recolector	0,77		2,24		4,073467903
Hacinamiento						
1	Si	-2,33	-2,33	0	2,76	0
2	No	0,43		2,76		5,019094381
Combustible para cocinar						
1	Ninguno (no cocina)/Otro/Leña o carbón	-2,41	-2,41	0	3,67	0
2	Kérex o diesel	-1,95		0,46		0,83651573
3	Gasolina	-1,38		1,03		1,873067831
4	Gas	0,41		2,82		5,128205128
5	Electricidad	1,26		3,67		6,673940716
Servicio higiénico del hogar						
1	No tiene	-1,84	-1,84	0	2,45	0
2	Letrina	-1,58		0,26		0,472813239
3	Excusado de uso común en varios hogares	0,28		2,12		3,855246408
4	Excusado de uso exclusivo del hogar	0,61		2,45		4,455355519
Ducha en el hogar						
1	No tiene	-1,42	-1,42	0	2,16	0
2	De uso común a varios hogares	0,53		1,95		3,546099291
3	De uso exclusivo en el hogar	0,74		2,16		3,927986907
Personas de 18 años y más Analfabetas en el hogar						
1	Si	-1,5	-1,5	0	2,17	0
2	No	0,67		2,17		3,946172031
Educación del Jefe del Hogar						
1	Ninguno	-1,86	-1,86	0	3,61	0
2	Centro de alfabetización	-1,58		0,28		0,509183488
3	Primaria / educación básica	-0,47		1,39		2,527732315
4	Secundaria / educación Media	0,71		2,57		4,673577014
5	Ciclo Post - bachillerato	0,92		2,78		5,05546463
6	Superior	1,55		3,41		6,201127478
7	Postgrado	1,75		3,61		6,564829969
Fuente: Censo de Población y Vivienda del 2001 - INEC				Suma de los valores máximos	54,99	100
				Factor de Escalamiento	1,818512	

En el gráfico 4.1.2.2 se muestra el comportamiento de nuestro indicador que indica una asimetría negativa y tiende hacia una distribución Beta; se puede observar la conformación de grupos más grandes para cada valor de nuestro índice, el siguiente paso es asignar los valores correspondientes a todos los hogares de la base y realizar una segmentación con el indicador construido en todo el conjunto de hogares.

Gráfico 4.1.2.2



La replicación en toda la base se la realizó a nivel cantonal, así en la tabla 4.1.2.7 se pueden ver los valores de nuestro indicador siendo los cantones como Rumiñahui (81,7), Quito (81,1) San Cristóbal en Galápagos (78) e Ibarra (77) los cantones que se encuentran en mejores condiciones.

Los cantones que se encuentran en peores condiciones son Taisha en Morona Santiago (31), Arajuno en Pastaza (40), Aguarico en Orellana (41) y Huamboya en Morona Santiago (42).

Tabla 4.1.2.7

PROVINCIA	CANTÓN	INDICADOR
AZUAY	CUENCA	76,24
	GIRON	67,32
	GUALACEO	65,74
	NABON	55,85
	PAUTE	65,37
	PUCARA	57,1
	SAN FERNANDO	69,91
	SANTA ISABEL	60,56
	SIGSIG	60,27
	ONA	59,39
	CHORDELEG	65,7
	EL PAN	65,77
	SEVILLA DE ORO	66,83
	GUACHAPALA	66,24
BOLIVAR	GUARANDA	57,98
	CHILLANES	51,24
	CHIMBO	60,6
	ECHEANDIA	59,87
	SAN MIGUEL	59,37
	CALUMA	60,05
	LAS NAVES	54,44
CAÑAR	AZOGUES	70,62
	BIBLIAN	67,64
	CANAR	58,13
	LA TRONCAL	67,21
	EL TAMBO	61,92
	DELEG	61,83
CARCHI	SUSCAL	54,04
	TULCAN	75,26
	BOLIVAR (DE CARCHI)	65,2
	ESPEJO	70,32
	MIRA	67,38
	MONTUFAR	71,11
COTOPAXI	SAN PEDRO DE HUACA	70,75
	LATACUNGA	70
	LA MANA	62,41
	PANGUA	48,71
	PUJILI	55,68
	SALCEDO	64,62
	SAQUISILI	60,28
CHIMBORAZO	SIGCHOS	49,98
	RIOBAMBA	73,33
	ALAUSI	50,52
	COLTA	50,53
	CHAMBO	64,36
	CHUNCHI	56,99
	GUAMOTE	44,79
	GUANO	59,33
	PALLATANGA	56,27
	PENIPE	61,89
EL ORO	CUMANDA	63,59
	MACHALA	72,77
	ARENILLAS	65,82
	ATAHUALPA	68,86
	BALSAS	71,69
	CHILLA	54,22
	EL GUABO	64,3
	HUAQUILLAS	64,48
	MARCABELI	72,26
	PASAJE	69,58
	PIÑAS	72,05
	PORTOVELO	69,88
	SANTA ROSA	68,7
	ZARUMA	66,53
LAS LAJAS	61,86	

ESMERALDAS	ESMERALDAS	67,86
	ELOY ALFARO	45,97
	MUISNE	49,76
	QUININDE	54,97
	SAN LORENZO	55,78
	ATACAMES	61,26
	RIOVERDE	47,88
GUAYAS	GUAYAQUIL	72,71
	ALFREDO BAQUERIZO MORENO	53,68
	BALAO	60,98
	BALZAR	55,15
	COLIMES	48,29
	DAULE	56,19
	DURAN	72,76
	EL EMPALME	55,09
	EL TRIUNFO	61,85
	MILAGRO	65,13
	NARANJAL	60,44
	NARANJITO	65,44
	PALESTINA	57,5
	PEDRO CARBO	54,42
	SALINAS	70,99
	SAMBORONDON	64,67
	SANTA ELENA	63,84
	SANTA LUCIA	53,72
	URBINA JADO	50,49
	SAN JACINTO DE YAGUACHI	57,47
	PLAYAS	67,2
	SIMON BOLIVAR	54,17
	CORONEL MARCELINO MARIDUENA	67,18
	LOMAS DE SARGENTILLO	60,46
	NOBOL (PIEDRAHITA)	61,18
LA LIBERTAD	67,01	
GENERAL ANTONIO ELIZALDE (BUCAY)	65,52	
IMBABURA	ISIDRO AYORA	57,53
	IBARRA	76,77
	ANTONIO ANTE	73,57
	COTACACHI	62,55
	OTAVALO	69,87
	PIMAMPIRO	65,96
	SAN MIGUEL DE URCUQUI	68,04
LOJA	LOJA	75,94
	CALVAS	63,91
	CATAMAYO	71,04
	CELICA	61,37
	CHAGUARPAMBA	59,48
	ESPINDOLA	56,7
	GONZANAMA	57,66
	MACARA	69,14
	PALTAS	59,7
	PUYANGO	59,71
	SARAGURO	58,48
	SOZORANGA	57,07
	ZAPOTILLO	55,97
	PINDAL	56,42
	QUILANGA	59,28
LOS RIOS	OLMEDO	57,54
	BABAHOYO	62,84
	BABA	50,61
	MONTALVO	57,97
	PUEBLOVIEJO	58,44
	QUEVEDO	66,64
	URDANETA	56,05
	VENTANAS	58,14
	VINCES	56,3
	PALENQUE	48,09
BUENA FE	59,22	
VALENCIA	55,21	
MOCACHE	51,31	

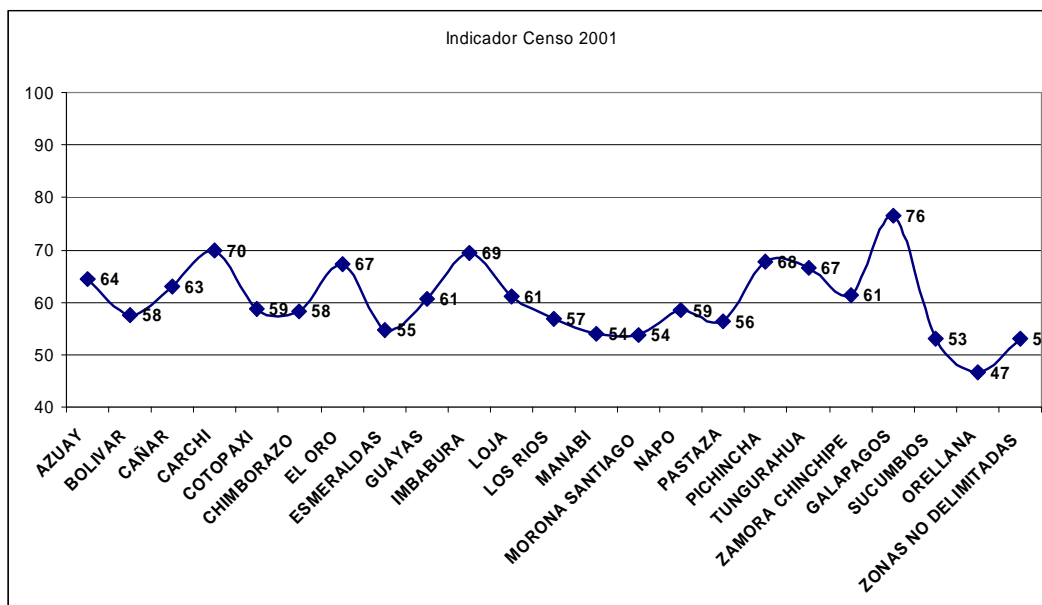
MANABI	PORTOVIEJO	65,19
	BOLIVAR (DE MANABI)	51,52
	CHONE	55,71
	EL CARMEN	56,22
	FLAVIO ALFARO	45,33
	JIPIJAPA	56,95
	JUNIN	48,19
	MANTA	72,09
	MONTECRISTI	60,91
	PAJAN	46,17
	PICHINCHA	43,96
	ROCAFUERTE	53,53
	SANTA ANA	49,58
	SUCRE	59,85
	TOSAGUA	52,72
	24 DE MAYO	45,55
	PEDERNALES	51,33
	OLMEDO	44,46
	PUERTO LOPEZ	59,4
	JAMA	50,86
JARAMIJO	62,47	
SAN VICENTE	57,61	
MORONA SANTIAGO	MORONA	59,36
	GUALAQUIZA	60,65
	LIMON INDANZA	56,19
	PALORA	60,51
	SANTIAGO	61,3
	SUCUA	64,25
	HUAMBOYA	41,89
	SAN JUAN BOSCO	55,65
	TAISHA	30,45
	LOGROÑO	47,22
PABLO VI	54,25	
NAPO	TENA	57,08
	ARCHIDONA	52,6
	EL CHACO	64,06
	QUIJOS	66,11
	CARLOS JULIO AROSEMENA TOLA	53,07
PASTAZA	PASTAZA	61,51
	MERA	69,47
	SANTA CLARA	54,38
	ARAJUNO	39,95
PICHINCHA	QUITO	81,05
	CAYAMBE	68,96
	MEJIA	73,4
	PEDRO MONCAYO	69,81
	RUMIÑAHUI	81,69
	SANTO DOMINGO	66,29
	SAN MIGUEL DE LOS BANCOS	58,34
	PEDRO VICENTE MALDONADO	58,16
PUERTO QUITO	52,23	
TUNGURAHUA	AMBATO	72,7
	BAÑOS DE AGUA SANTA	74,2
	CEVALLOS	68
	MOCHA	64,58
	PATATE	64,81
	QUERO	61,41
	SAN PEDRO DE PELILEO	65,19
	SANTIAGO DE PILLARO	65,48
TISALEO	63,15	
ZAMORA CHINCHIPE	ZAMORA	69,6
	CHINCHIPE	58,26
	NANGARITZA	56,79
	YACUAMBI	55,86
	YANTZAZA	67,35
	EL PANGUI	62,49
	CENTINELA DEL CONDOR	64,54
PALANDA	55,88	

GALAPAGOS	SAN CRISTOBAL	77,99
	ISABELA	76,74
	SANTA CRUZ	74,6
SUCUMBIOS	LAGO AGRIO	56,88
	GONZALO PIZARRO	53,88
	PUTUMAYO	44,54
	SHUSHUFINDI	55,67
	SUCUMBIOS	59,71
	CASCALES	53,29
	CUYABENO	47,13
ORELLANA	ORELLANA	53,71
	AGUARICO	40,52
	LA JOYA DE LOS SACHAS	50,07
	LORETO	42,22
ZONAS NO DELIMITADAS	LAS GOLONDRINAS	53,39
	LA CONCORDIA	58,26
	MANGA DEL CURA	45,79
	EL PIEDRERO	55,05

Fuente: Censo de Población y Vivienda del 2001 - INEC

A nivel provincial se observa en el gráfico 4.1.2.3 que las provincias que están en mejores condiciones son Galápagos (76), Carchi (70), Imbabura (69) y Pichincha (68); las que se encuentran en peores condiciones son Orellana (47), Sucumbíos (53), Zonas No Delimitadas (53) y Morona Santiago (54).

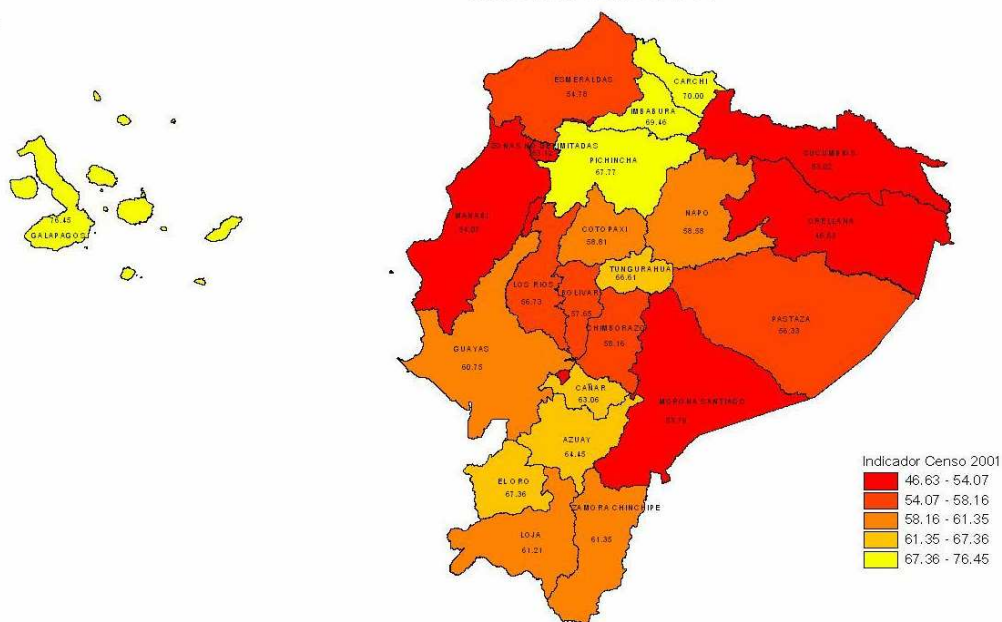
Gráfico 4.1.2.3



Así también en el gráfico 4.1.2.4 se muestran los quintiles formados a partir del indicador, notándose que la amazonía presenta los valores más bajos, por ende peores condiciones.

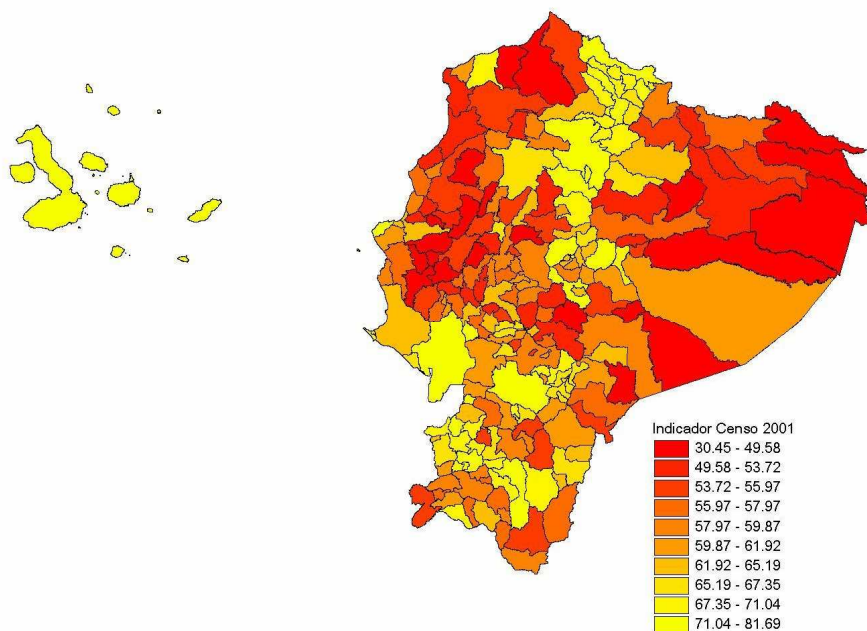
Gráfico 4.1.2.4

Indicador Censo 2001



En el gráfico 4.1.2.5 tenemos la representación a nivel cantonal de los deciles del indicador.

Gráfico 4.1.2.5



4.1.3 IMPLEMENTACIÓN DEL MODELO CONSTRUÍDO A PARTIR DEL CENSO DE POBALCIÓN Y VIVIENDA 2001

Se utilizan las cuantificaciones obtenidas en el modelo de la sección anterior para comparar el indicador obtenido con el Censo de Población y Vivienda 2001 con un nuevo índice a partir de la Encuesta de Condiciones de Vida 2005-2006; y a la vez utilizar dichas cuantificaciones óptimas como una línea de base para comparar las condiciones de vida en diferentes periodos pero con el mismo conjunto de variables.

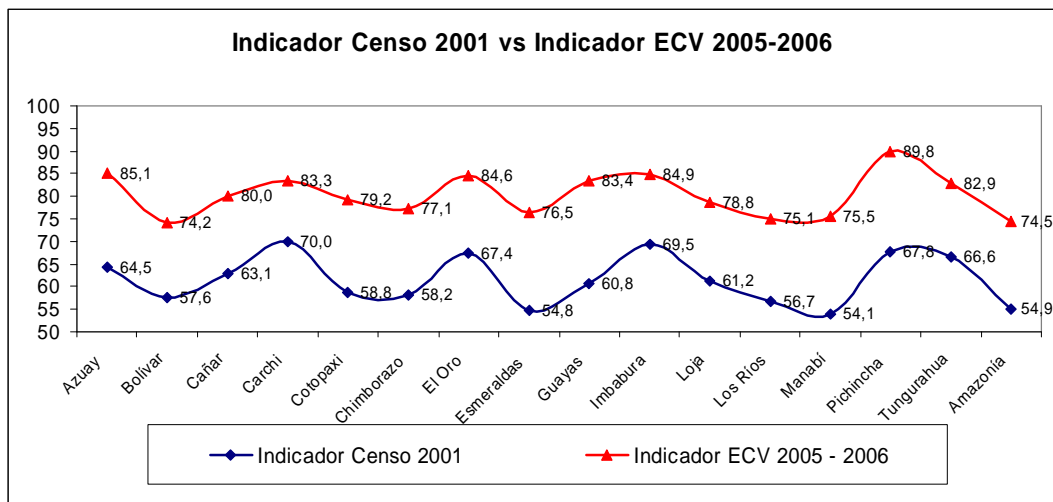
En la tabla 4.2.1.1 se muestra la comparación entre el indicador obtenido a partir del Censo de Población y Vivienda y el de la Encuesta de Condiciones y Vida 2005-2006.

Tabla 4.1.3.1

	Indicador Censo 2001	Indicador ECV 2005 - 2006
Azuay	64,5	85,1
Bolívar	57,6	74,2
Cañar	63,1	80,0
Carchi	70,0	83,3
Cotopaxi	58,8	79,2
Chimborazo	58,2	77,1
El Oro	67,4	84,6
Esmeraldas	54,8	76,5
Guayas	60,8	83,4
Imbabura	69,5	84,9
Loja	61,2	78,8
Los Ríos	56,7	75,1
Manabí	54,1	75,5
Pichincha	67,8	89,8
Tungurahua	66,6	82,9
Amazonía	54,9	74,5
Encuesta de Condiciones de Vida 2005-2006 - INEC		

Y en el gráfico 4.2.1.1 puede observarse como los indicadores tienen un comportamiento similar para cada provincia con respecto a las dos fuentes de información, observando un mejoramiento de las condiciones de vida a partir del año 2001.

Gráfico 4.1.3.1



4.2 APLICACIÓN CON LA ENCUESTA DE CONDICIONES DE VIDA 2005 - 2006

Así como en la elaboración del indicador de condiciones de vida con el Censo de población y vivienda se presentó una descripción de las distribuciones de las variables utilizadas en el índice, a continuación también se presenta una revisión de las variables análogas al Censo que se encuentran en la Encuesta de Condiciones de Vida 2005 – 2006.

En el Anexo 4.1 se aprecia que más del 90% de los hogares habitan un cuarto de inquilinato, departamento o casa y un 2% viven en choza/rancho o covacha, pero a nivel provincial se ve que en la región amazónica el 9% de los hogares viven en choza/rancho o covacha así como Esmeraldas, 7% y Bolívar, 5%.

Con respecto al material del techo de la vivienda, se puede observar en el Anexo 4.2 que en provincias como Chimborazo, 8% y Esmeraldas, 5% los hogares cuentan con techo de paja o similares, en forma global el 45% de los hogares cuentan con techo de zinc lo que se ha seguido manteniendo desde el año 2001.

Se aprecia en el Anexo 4.3 que el material predominante de las viviendas en provincias como Manabí, 32%, Los Ríos 29% y Guayas, 17% es caña no revestida, puede observarse que las tres cuartas partes de los hogares en el país cuentan con viviendas con paredes de hormigón y un 13% cuenta con paredes de madera o tapia.

En relación al piso de la vivienda se tiene en el Anexo 4.4 que el 92% de los hogares cuentan con piso de ladrillo / baldosa / parquet o mármol lo que si muestra un mejoramiento con respecto al año 2001, sin embargo un 28% de los hogares en la provincia de Loja tienen piso de tierra o caña, así como en Chimborazo y un 185 en la provincia del Carchi.

En el Anexo 4.5 podemos ver que el 52% de los hogares en el país obtienen el agua por tubería por dentro de la vivienda lo que se mantiene según el Censo del

2001, pero existen provincias en las cuales el agua se obtiene por tubería pero por fuera del edificio especialmente en la costa y es así como Manabí, 34%, Esmeraldas, 24% y Los Ríos, 22% son las que cuentan con los valores más altos con respecto a este indicador.

Se observa en el Anexo 4.6 que en el país la tercera parte de los hogares en el Ecuador cuentan con pozo séptico o ciego, lo que no varía con respecto al 2001, 34%, principalmente en provincias como Los Ríos, 54%, Guayas, 47% y Bolívar cuentan con 42% de hogares que tienen esta forma de eliminación de aguas negras o servidas; pero así mismo, en la Amazonía el 37% de los hogares tienen otra forma de eliminar las aguas negras o servidas, así como Esmeraldas, 34%, Los Ríos, 32% y Loja 30%.

Mientras tanto en el Anexo 4.7 puede observarse las distribuciones de las variables luz eléctrica y servicio telefónico.

Con respecto a la primera tenemos que en el país el 96% de los hogares cuentan con este servicio lo que sí muestra un mejoramiento con respecto al 2001, 90%, pero en cambio en provincias de la Amazonía un 23% de los hogares no cuentan con luz eléctrica aunque muestra un mejoramiento notable en relación al Censo, 66%.

El servicio telefónico en el país cuenta con una cobertura del 36% que muestra un mejoramiento de cuatro puntos en relación al 2001, en provincias como Los Ríos la cobertura de este servicio es del 85%, en Manabí y la Amazonía 82% y en Bolívar 81%.

En relación a la forma de la eliminación de basura se aprecia en el Anexo 4.8 que en el país el 27% de los hogares eliminan la basura por incineración o la botan en un terreno baldío lo que muestra una mala práctica sanitaria, pero que mejora con respecto al Censo 32%, es más notorio en Bolívar, 67%, Cotopaxi, 58% y Chimborazo e Imbabura 52% lo que sigue mostrando una falta de acceso al sistema de recolección en las provincias de la sierra central con respecto al 2001.

También en el Anexo 4.8 puede verse que el 16% de los hogares presentan problemas de hacinamiento, lo que se viene manteniendo desde el 2001, en provincias como Loja, 25%, Los Ríos, 23% y Amazonía, 20% donde dicha característica es más notoria.

Con respecto al combustible para cocinar se tiene en el Anexo 4.9 que el 11% de los hogares no cuentan con ningún servicio o cocinan con leña o carbón siendo más notorio en Bolívar, 44%, Chimborazo, 30% y la Amazonía, 28%; en forma global se muestra una reducción del 4% con respecto al 2001.

En el Anexo 4.10 se ve que el 10% de los hogares no cuentan con servicio higiénico que ha mejorado con respecto al 2001, 17%; pero las provincias en las que es más notorio la falta de servicio higiénico son las de la Amazonía, 31%, Bolívar, Cañar y Loja 23%; a nivel nacional el 7% de los hogares cuentan con letrina, manteniendo prácticamente el mismo nivel respecto al 2001, 8%.

El 47% de los hogares no cuentan con ducha exclusiva en el hogar, sino más bien de uso común como se ve el Anexo 4.11 y los valores más altos se encuentran en las provincias de Los Ríos, 75%, Bolívar, 70%, Esmeraldas y Manabí, 67% y la Amazonía, 61%.

También se aprecia que a nivel nacional el 17% de los hogares tienen personas de 18 y más años que son analfabetas, pero, muestra un mejoramiento tomando en cuenta el 30% con el que se contaba en el 2001. Las provincias en las cuales existe mayor presencia de personas de 18 y más años analfabetas son Bolívar y Chimborazo, 34%, Cotopaxi, 31% y Cañar, 28%.

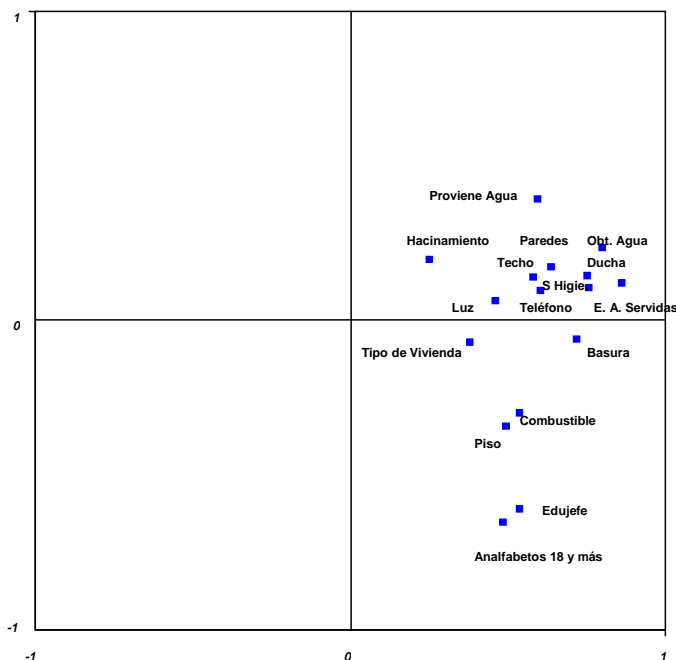
En relación a la educación del jefe del hogar se ve en el Anexo 4.12 que en el país el 7% no tienen ningún tipo de educación, centro de alfabetización 0,4% primaria 46%, las provincias en las cuales es más grande el porcentaje de hogares en los que el jefe del hogar no tienen ningún tipo de educación son Chimborazo, 19%, Bolívar, 15% y Cotopaxi, 14%.

4.2.1 CONSTRUCCIÓN DEL MODELO

Dadas las variables anteriormente descritas puede verse en el gráfico 4.2.1.1 que las variables “Educación del jefe del hogar” y “Presencia de analfabetos” se encuentran muy relacionadas; la variable “Hacinamiento” y “Tipo de vivienda” se encuentran un poco alejadas del grupo conformado por las otras variables, además las variables “Combustible para cocinar” y “Piso de la vivienda” están alejadas del resto de variables pero más relacionadas entre sí; y en general todas las variables de descripción de servicios o características de la vivienda están asociadas.

En esta sección se utilizarán todas las variables mencionadas, ya que si bien al correr el algoritmo nuevamente con las variables de la Encuesta de Condiciones de Vida 2005 – 2006 no serán comparables las cuantificaciones obtenidas con las del Censo, analizaremos el comportamiento de nuestro nuevo indicador de calidad de vida en relación al índice anterior.

Gráfico 4.2.1.1



En la tabla 4.2.1.1 se muestran las cuantificaciones óptimas resultantes para la ECV 2005 – 2006.

Tabla 4.2.1.1

Categoría	Variables	Cuantificaciones Óptimas	Mínimos	Cambio de Escala	Máximos	Cuantificaciones Transformadas
Tipo de vivienda						
2	Covacha/Rancho	-5,66	-5,66	0	5,93	0
3	Mediagua	-1,92		3,74		7,540322581
4	Cuarto en casa de inquilinato/Departamento /Casa o villa / Otras particulares	0,27		5,93		11,95564516
Techo de la vivienda						
1	Otros materiales	-2,36	-2,36	0	3,81	0
2	Paja o similares	-2,36		0		0
3	Teja	-0,66		1,7		3,427419355
4	Zinc	-0,66		1,7		3,427419355
5	Eternit	0,8		3,16		6,370967742
6	Losa de hormigón	1,45		3,81		7,681451613
Paredes exteriores						
1	Otros materiales	-2,2	-2,2	0	2,82	0
2	Caña no revestida	-2,2		0		0
3	Caña revestida o bahareque	-1,46		0,74		1,491935484
4	Madera/Adobe o tapia	-1,28		0,92		1,85483871
5	Hormigón, ladrillo o bloque	0,62		2,82		5,685483871
Piso de la Vivienda						
1	Otros materiales	-2,89	-2,89	0	3,24	0
2	Tierra/Caña	-2,89		0		0
3	Ladrillo/Baldosa o vynil/Entablado o parquet	0,35		3,24		6,532258065
Cómo obtiene el agua para la vivienda						
2	Por tubería fuera del edificio, lote o terreno	-1,89	-1,89	0	2,84	0
3	Por tubería fuera de la vivienda pero dentro del edificio	-0,56		1,33		2,681451613
4	Por tubería dentro de la vivienda	0,95		2,84		5,725806452
De dónde proviene el agua que recibe						
1	Otro	-2,13	-2,13	0	2,62	0
2	De carro repartidor/De Río, vertiente, acequia o canal	-2,13		0		0
3	De Pozo	-1,79		0,34		0,685483871
4	De red pública	0,49		2,62		5,282258065
Forma de eliminación de aguas negras o servidas						
1	Otra forma	-1,73	-1,73	0	2,61	0
2	Por Pozo séptico/Por pozo ciego	-0,22		1,51		3,044354839
3	Por red pública de alcantarillado	0,88		2,61		5,262096774
Luz Eléctrica						
1	No	-4,02	-4,02	0	4,27	0
2	Si	0,25		4,27		8,608870968
Servicio telefónico						
1	No	-0,7	-0,7	0	2,14	0
2	Si	1,44		2,14		4,314516129

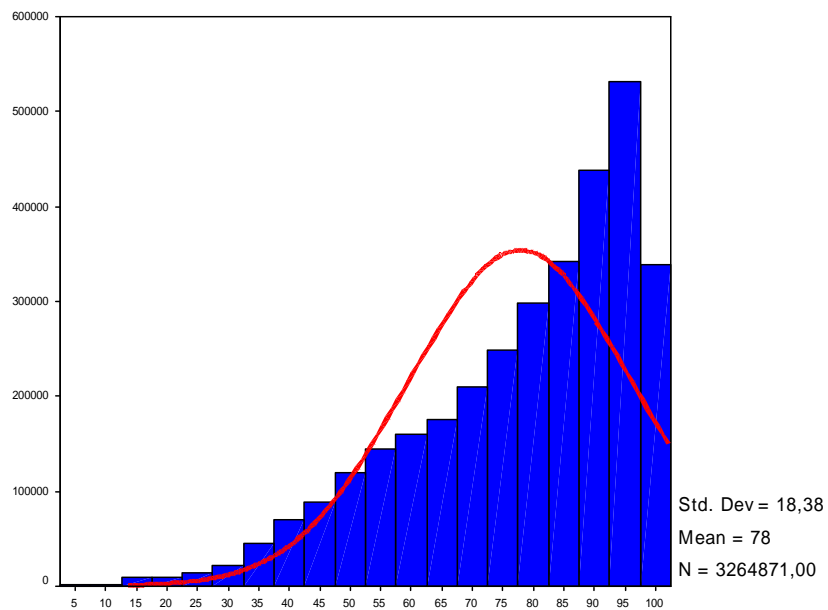
Forma de eliminación de basura de la vivienda						
1	De otra forma	-1,36	-1,36	0	2,1	0
2	Por incineración o entierro/En terreno baldío o quebrada	-1,36		0		0
3	De carro recolector	0,74		2,1		4,233870968
Hacinamiento						
1	Si	-2,25	-2,25	0	2,69	0
2	No	0,44		2,69		5,423387097
Combustible para cocinar						
1	Ninguno (no cocina)/Otro/Leña o carbón	-2,39	-2,39	0	3,57	0
4	Gas	0,42		2,81		5,665322581
5	Electricidad	1,18		3,57		7,197580645
Servicio higiénico del hogar						
1	No tiene	-2,03	-2,03	0	2,54	0
2	Letrina	-1,85		0,18		0,362903226
3	Excusado de uso común en varios hogares	0,51		2,54		5,120967742
4	Excusado de uso exclusivo del hogar	0,51		2,54		5,120967742
Ducha en el hogar						
2	De uso común a varios hogares	-1,02	-1,02	0	2	0
3	De uso exclusivo en el hogar	0,98		2		4,032258065
Personas de 18 años y más Analfabetas en el hogar						
1	Si	-2,05	-2,05	0	2,54	0
2	No	0,49		2,54		5,120967742
Educación del Jefe del Hogar						
1	Ninguno	-3	-3	0	3,88	0
2	Centro de alfabetización	-1,36		1,64		3,306451613
3	Primaria / educación básica	-0,13		2,87		5,786290323
4	Secundaria / educación Media	0,7		3,7		7,459677419
5	Ciclo Post - bachillerato	0,79		3,79		7,641129032
6	Superior	0,84		3,84		7,741935484
7	Postgrado	0,88		3,88		7,822580645
Encuesta de Condiciones de Vida 2005-2006 - INEC				Suma de los valores máximos	49,6	100
				Factor de Escalamiento	2,016129	

Asignando las cuantificaciones óptimas a cada una de las categorías dentro de cada variable se tiene en el gráfico 4.2.1.2 la distribución de nuestro indicador de condiciones de vida, que al igual que el índice construido con el Censo es un puntaje positivo con respecto a las variables usadas tiene una mayor asimetría negativa lo que muestra el mejoramiento mencionado.

Además en la tabla 4.2.1.2 se tienen las correlaciones de las variables usadas con nuestro indicador, donde al igual que con el índice construido con el Censo

del 2001, la variable “forma de eliminación de aguas negras o servidas”, es la que más se encuentra relacionada con la medida de condiciones de vida, este indicador se ajusta más a una distribución Beta como se puede ver en el gráfico 4.2.1.2.

Gráfico 4.2.1.2



Indicador ECV 05-06

Tabla 4.2.1.2 Correlaciones

	Indicador ECV 05 - 06
INDICADOR ECV 05 - 06	1
Forma de eliminación de aguas negras o servidas	0,820386488
Cómo obtiene el agua para la vivienda	0,77467956
Ducha en el hogar	0,734142856
Forma de eliminación de basura de la vivienda	0,666101338
Servicio higiénico del hogar	0,614959334
Paredes exteriores	0,610713437
Servicio telefónico	0,603453065
De dónde proviene el agua que recibe	0,573620092
Educación del Jefe del Hogar	0,566896524
Material del techo de la vivienda	0,552930635
Combustible para cocinar	0,507112011
Piso	0,487688622
Personas de 18 años y más Analfabetas en el hogar	0,483931296
Luz Eléctrica	0,452772641
Tipo de vivienda	0,414485459
Hacinamiento	0,306508675
Encuesta de Condiciones de Vida 2005-2006 - INEC	

Si creamos quintiles usando nuestro indicador, puede observarse en el gráfico 4.2.1.3 que la provincia de Bolívar es la que peor se encuentra según la medida calculada. Pichincha, Azuay, El Oro e Imbabura son las que presentan mejores condiciones de vida; si comparamos el indicador calculado con el Censo, el calculado con base en las cuantificaciones óptimas del Censo pero usando Encuesta de Condiciones de Vida y el índice calculado usando las cuantificaciones óptimas obtenidas con la Encuesta de Condiciones de Vida se puede ver en el gráfico 4.2.1.4 que las tendencias se siguen manteniendo, es decir se muestra un mejoramiento con respecto a las condiciones de vida pero las desigualdades se siguen manteniendo con el paso de los años lo que indica que en cinco años no se han tomado acciones más decisivas para ayudar a las provincias que se encuentran en peores condiciones.

Gráfico 4.2.1.3

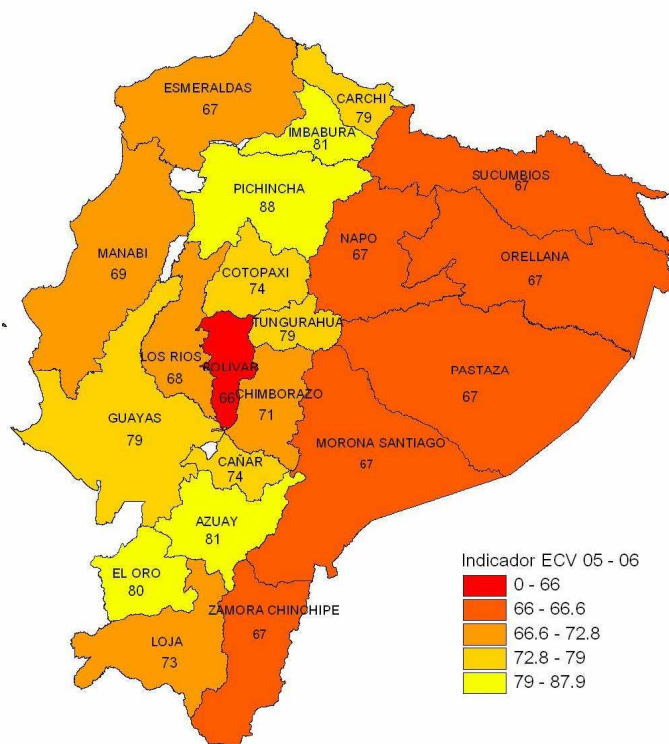
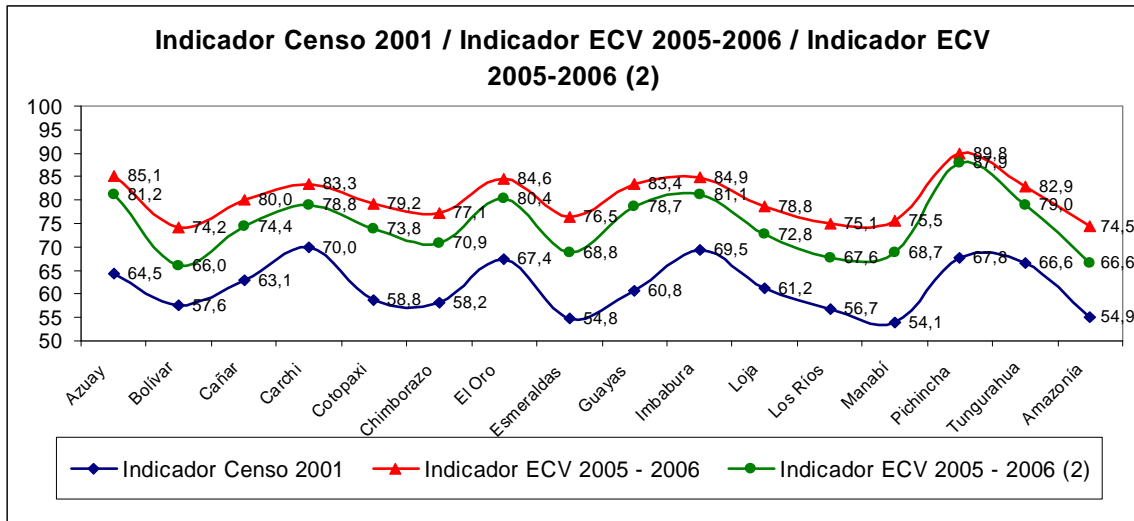


Gráfico 4.2.1.4



CAPÍTULO 5.

CONSTRUCCIÓN DE UN ÍNDICE CON BASE EN EL ANÁLISIS DE COMPONENTES PRINCIPALES LINEAL

Para la construcción de este índice se utilizan las mismas variables que se toman en la aplicación del Análisis de Componentes Principales no lineal, pero calculadas como indicador a nivel cantonal, así, las variables utilizadas son:

1. Hogares que viven en Cuarto de inquilinato / Departamento / Casa o villa.
2. Hogares con techo de losa de hormigón.
3. Hogares con paredes de hormigón, ladrillo o bloque.
4. Hogares con piso de ladrillo, baldosa, entablado o parquet.
5. Hogares que obtienen el agua por tubería dentro de la vivienda.
6. Hogares que reciben el agua de red pública.
7. Hogares que eliminan aguas negras por red pública de alcantarillado.
8. Hogares que poseen luz eléctrica.
9. Hogares que cuentan con servicio telefónico.
10. Hogares que eliminan la basura a través de carro recolector.
11. Hogares que no se encuentran hacinados.
12. Hogares que cocinan con gas.
13. Hogares que cuentan con excusado de uso exclusivo.
14. Hogares con ducha de uso exclusivo.
15. Hogares sin presencia de personas de 18 y más analfabetas.

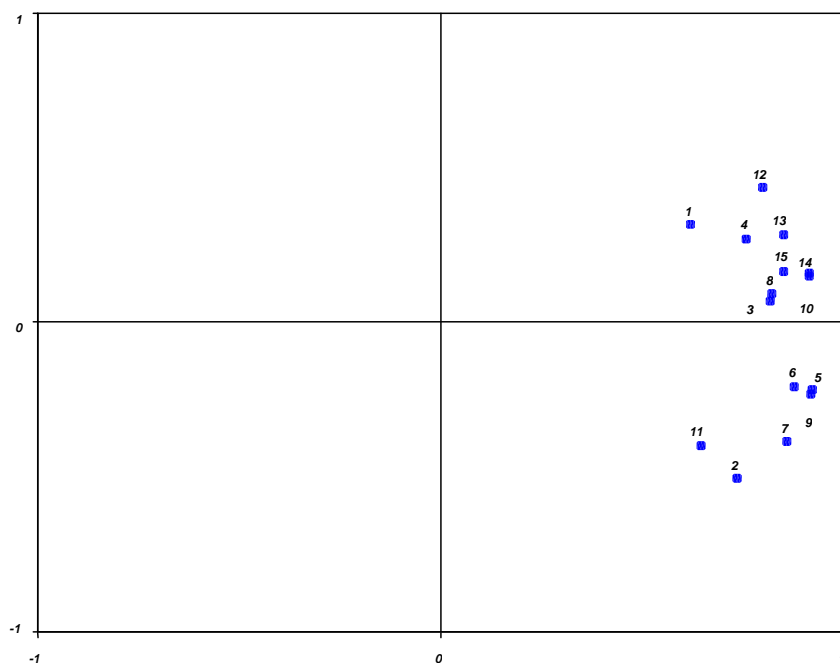
Utilizando el ACP lineal se obtuvo que la primera componente explica el 48% de la varianza, cuyos puntajes se utilizan para calcular el indicador de condiciones de vida.

De este modo en el gráfico 5.1 puede observarse como se encuentran relacionadas las variables utilizadas, así por ejemplo la variable “Hogares que eliminan la basura a través de carro recolector (10)” se encuentra muy relacionada con “Hogares con ducha de uso exclusivo (14)”, de la misma manera “Hogares con paredes de hormigón, ladrillo o bloque (3)” con “Hogares que poseen luz eléctrica (8)” y también están relacionadas “Hogares que

obtienen el agua por tubería dentro de la vivienda (5)” con “Hogares que cuentan con servicio telefónico (9)”.

Las variables “Hogares que viven en Cuarto de inquilinato / Departamento / Casa o villa (1)”, “Hogares con techo de losa de hormigón (2)” y “Hogares que no están hacinados (11)”, no están muy relacionadas con el resto de variables, lo que indica que estas variables ayudarán a discriminar de mejor manera a cada uno de los cantones.

Gráfico 5.1



Para medir la calidad del modelo utilizamos la medida de ajuste de Kaiser-Meyer-Olkin, que en este caso es de 0,923 que nos indica que el modelo es adecuado.

Para la construcción del indicador usamos la primera componente extraída que capta el 48% de la varianza, y para la elaboración del índice se utiliza la siguiente transformación lineal:

$$I = CP_1 * a + b \quad (5.1)$$

Para la construcción del indicador de condiciones de vida se probaron varias alternativas para los parámetros **a** y **b**, para la constante **b** se utiliza la media del índice creado en la sección 4.1, cuyo valor es de 68,06 para poder comparar ambos indicadores y se probaron algunos valores para la constante **a**.

Tabla 5.1

	Mínimo	Máximo	Media	Desviación Estándar	Varianza
Índice con ponderaciones óptimas	32,67	81,37	68,0624	9,7868	95,782
Índice = Cp + sumada la media	64,49	69,35	68,0624	1	1
Índice = Cp*5 + media	50,22	74,49	68,0624	5	25
Índice = Cp*10 + media	32,38	80,91	68,0624	10	100
Índice = Cp*15 + media	14,53	87,33	68,0624	15	225
Índice = Cp*20 + media	-3,31	93,76	68,0624	20	400
Índice = Cp*30 + media	-39	106,6	68,0624	30	900

En la tabla 5.1 puede observarse que por la forma de construcción del indicador, la varianza que nos servirá para determinar cuál índice discrimina mejor, siempre será igual al valor de la constante **a** elevado al cuadrado, ya que:

$$\text{Siendo } y = ax + b, \text{ entonces } \text{Var}(y) = a^2 \text{Var}(x) \quad (5.2)$$

Adicionalmente se observa que con **a=20** y **a=30** se obtienen medidas negativas para el indicador, así se utiliza **a=9,7868**; que es la desviación estándar de la medida obtenida en la sección 4.1, para que nuestro nuevo índice pueda compararse con el de la sección mencionada, así:

$$I = 9,7868 * CP_1 + 68,0624 \quad (5.3)$$

Con esta combinación lineal se obtienen quintiles con dicho indicador observándose en la tabla 5.2 los valores que describen a cada uno de los quintiles.

	Mínimo	Máximo	Media	Varianza
Quintil 1	33,1	58,7	52,8	21,6
Quintil 2	58,8	66,8	62,8	5,4
Quintil 3	67,0	73,8	70,3	5,4
Quintil 4	74,0	74,9	74,7	0,1
Quintil 5	75,5	80,6	80,6	0,2

En el gráfico 5.2 se observa el indicador calculado con el Análisis de Componentes Principales Lineal a nivel provincial y nuestro índice creado en el capítulo anterior, observando un comportamiento similar.

Gráfico 5.2

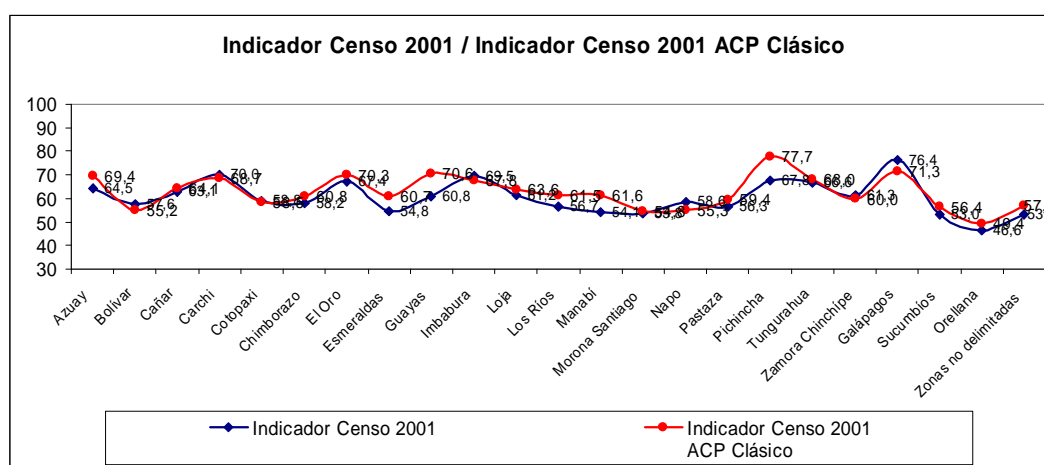


Tabla 5.3

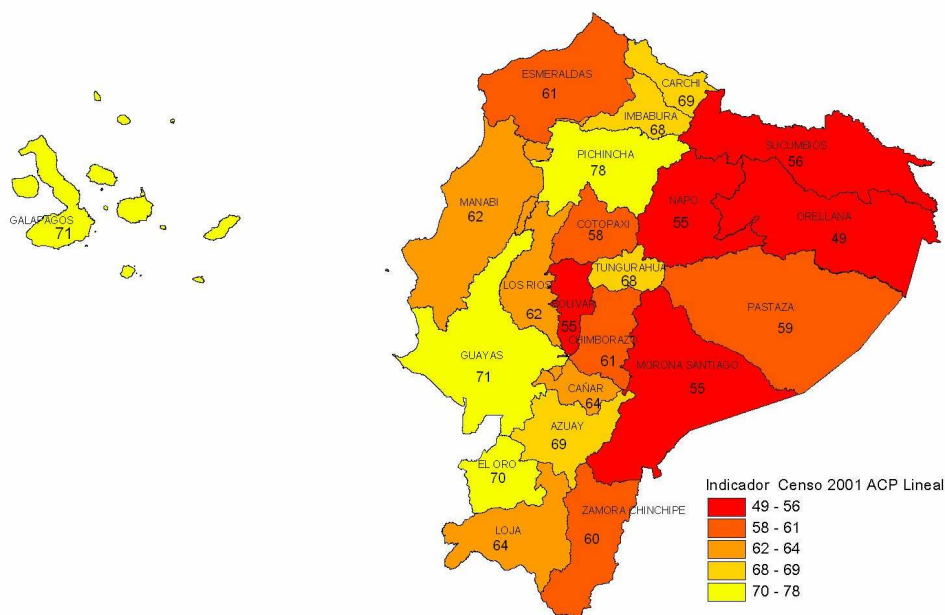
	Indicador ACP Clásico
Indicador ACP Clásico	1
Hogares que obtienen el agua por tubería dentro de la vivienda	0,92
Hogares que cuentan con servicio telefónico	0,92
Hogares que eliminan la basura a través de carro recolector	0,92
Hogares con ducha de uso exclusivo	0,91
Hogares que reciben el agua de red pública	0,88
Hogares que eliminan aguas negras por red pública de alcantarillado	0,86
Hogares sin presencia de personas de 18 y más analfabetas	0,85
Hogares que cuentan con excusado de uso exclusivo	0,85
Hogares que poseen luz eléctrica	0,82
Hogares con paredes de hormigón, ladrillo o bloque	0,82
Hogares que cocinan con gas	0,80
Hogares con piso de ladrillo, baldosa, entablado o parquet	0,76
Hogares con techo de losa de hormigón	0,73
Hogares que no se encuentran hacinados	0,64
Hogares que viven en Cuarto de inquilinato / Departamento / Casa o villa	0,62

Fuente: Censo de Población y Vivienda del 2001 - INEC

En la tabla 5.3 puede verse que las variables más relacionadas con nuestro indicador son “Hogares que obtienen el agua por tubería dentro de la vivienda”, “Hogares que cuentan con servicio telefónico”, “Hogares que eliminan la basura a través de carro recolector”, “Hogares con ducha de uso exclusivo” y “Hogares que reciben el agua de red pública”; aunque en general todas las variables aportan a la construcción del indicador.

En el gráfico 5.3 se observa como se encuentran clasificadas las provincias usando quintiles del indicador construido, donde provincias como Pichincha Galápagos, y Guayas se encuentran en el quintil más alto; pero provincias como Orellana, Morona Santiago, Bolívar, Napo, Sucumbíos y Cotopaxi son parte del grupo que se encuentra en peores condiciones con respecto a nuestro nuevo índice.

Gráfico 5.3

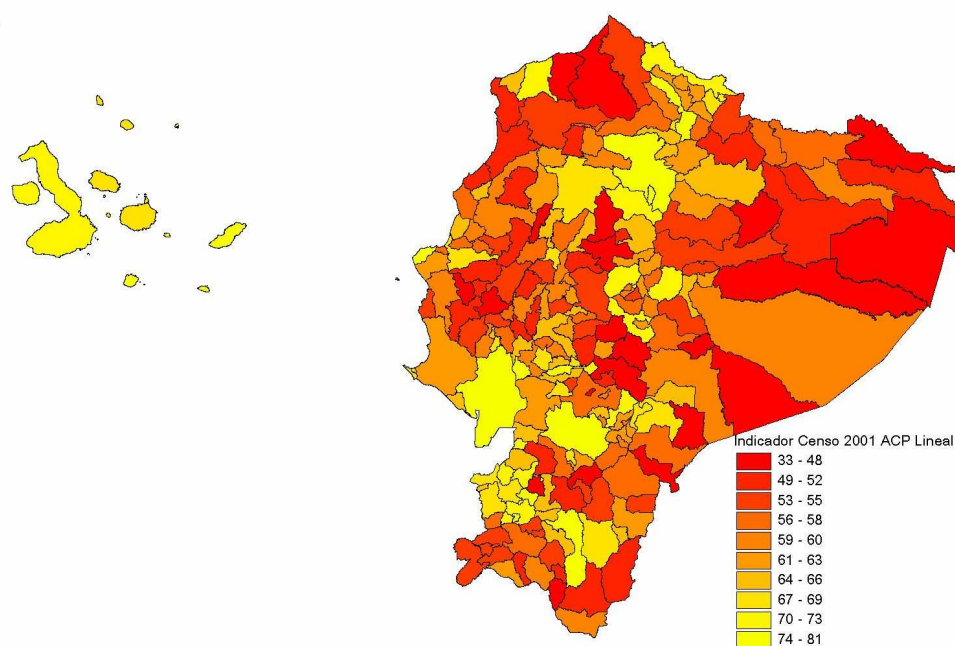


A nivel cantonal los cantones que se encuentran en mejores condiciones son Quito, 80,64, Rumiñahui 80,53, y en Guayas el cantón Coronel Marcelino Maridueña, 77,43.

En cambio los cantones que se encuentran en peores condiciones, son Taisha en Morona Santiago, 33,14, Guamote en Chimborazo, 37,93, Aguarico en Orellana, 39,61 y Huamboya en Morona Santiago, 41,1.

En el gráfico 5.4 puede apreciarse la clasificación dada a nivel de cantón, teniendo un comportamiento muy parecido al mapa cantonal (gráfico 4.1.2.5) del capítulo cuatro.

Gráfico 5.4



Para medir cuál indicador discrimina de mejor manera a nivel cantonal se plantea medir cual presenta el mayor rango, y se obtuvo que para el índice creado en el capítulo cuatro el rango es de **48,7**, mientras que para el índice presentado en este capítulo, el rango es **47,5**; es decir el indicador calculado usando el Análisis de Componentes Principales no lineal presenta una pequeña ventaja para discriminar las condiciones de vida de los cantones del país.

A nivel provincial la varianza de nuestro nuevo indicador es **45,04** y del índice construido con el procedimiento PRINCALS es **48,13**; con lo que a nivel

provincial el método utilizado en el capítulo anterior discrimina de mejor manera.

Si bien numéricamente las diferencias no son notables, en un sentido práctico puede observarse que nuestra hipótesis de que el método del Análisis de Componentes Principales no lineal, discrimina mejor a las unidades territoriales, característica muy importante para estos índices de condiciones de vida.

CAPÍTULO 6.

CONCLUSIONES Y RECOMENDACIONES

6.1 CONCLUSIONES

A continuación se presentan los resultados más importantes obtenidos en la investigación.

Usando el Análisis de Componentes Principales no lineal se hace más fácil la construcción de índices a partir de un conjunto de variables cualitativas, dado que las variables ingresan al modelo con las categorías predefinidas.

Uno de los resultados más importantes es el que, un índice construido como de necesidades es el complemento de otro que mide satisfacción (correlación igual a -1), por lo tanto da lo mismo elaborar cualquiera de los dos.

Además si se construye un índice de necesidades y luego se requiere construir un indicador de satisfacción, las cuantificaciones de cada una de las variables son el inverso negativo de las cuantificaciones iniciales, así por ejemplo para la variable “Nivel educativo del entrevistado” se tiene que:

Tabla 6.1.1

Variable: Nivel educativo del entrevistado			
Categoría (1)	Cuantificación (1)	Categoría (2)	Cuantificación (2)
5 Ninguna / primaria	1,81	1 Ninguna / primaria	-1,81
4 Secundaria	1,71	2 Secundaria	-1,71
3 Superior no universitaria	-0,46	3 Superior no universitaria	0,46
2 Superior / universitaria	-0,62	4 Superior / universitaria	0,62
1 Postgrado	-0,62	5 Postgrado	0,62

Al ser indicadores de satisfacción se determinó con el índice construido a partir del Censo de Población y Vivienda 2001, que los cantones Rumiñahui y Quito de la provincia de Pichincha son los que se encuentran en mejores condiciones con respecto al total de cantones a nivel nacional, así mismo se nota que en el 2001 los cantones que se encuentran en peores condiciones son especialmente los de la región amazónica y los de la sierra central.

A nivel provincial están en mejores condiciones Galápagos, Carchi, Imbabura y Pichincha; las que se encuentran en peores condiciones son Orellana, Sucumbíos, Las Zonas No Delimitadas y Morona Santiago.

Las cuantificaciones óptimas obtenidas a partir del Censo de Población y Vivienda 2001 constituyen una línea de base para realizar comparaciones usando las mismas variables, pero con distintas fuentes de información, así pues en el gráfico 4.2.1.1 puede observarse como las condiciones de vida de la población han mejorado, manteniéndose las diferencias entre cada una de las provincias.

Las variables que se encuentran más relacionadas con los indicadores construidos con el Censo de Población y Vivienda 2001 y con la Encuesta de Condiciones de Vida 2005 – 2006 son “Forma de eliminación de aguas negras o servidas, “Cómo obtiene el agua para la vivienda” y “Ducha en el hogar” lo que significa que las variables de acceso a estos servicios son las que más discriminan tanto el 2001 como en el 2006.

El índice construido a partir de las cuantificaciones obtenidas en la Encuesta de Condiciones de Vida 2005 – 2006, tiene un comportamiento similar al construido en base a las cuantificaciones obtenidas a partir del Censo, pero usando la información de la ECV 2005 – 2006, es decir las cuantificaciones obtenidas en el Censo del 2001 realmente constituyen una línea de base.

Con respecto al modelo construido usando la ECV 2005 – 2006, al ser una encuesta con nivel de representatividad provincial, puede observarse que Pichincha, Azuay, El Oro e Imbabura son las que se encuentran en mejores

condiciones y las que peor se encuentran son Bolívar, las provincias de la amazonía y Los Ríos.

Con el índice construido usando el Análisis de Componentes Principales lineal, se determinó que las provincias que se encuentran en mejores condiciones son Pichincha, Galápagos y Guayas; las que se encuentran peores condiciones de vida Bolívar, Morona Santiago y Orellana.

Este índice, tiene la restricción que para poder comparar diferentes fuentes de información, deberán calcularse cada uno de los indicadores para que puedan ser usados en el modelo, a diferencia del uso directo que se les da a las variables en el Análisis de Componentes Principales no lineal.

Las variables que más se encuentran relacionadas con el índice usando el análisis clásico son “Porcentaje de Hogares que obtienen el agua por tubería dentro de la vivienda”, “Porcentaje de Hogares que cuentan con servicio telefónico”, “Porcentaje de Hogares que eliminan la basura a través de carro recolector” y “Porcentaje de Hogares con ducha de uso exclusivo”.

Con respecto a que modelo discrimina mejor, puede observarse en la tabla 6.1.2, que el indicador construido usando el método PRINCALS en términos del rango discrimina mejor que el índice usando el análisis clásico, además a nivel provincial el indicador desarrollado con base en el Análisis de Componentes Principales no lineal tiene mayor variabilidad que el índice construido con el Análisis de Componentes Principales lineal.

Tabla 6.1.2

	Rango	Mínimo	Máximo	Media	Desviación Estándar	Varianza
Índice con ponderaciones óptimas	48,7	32,67	81,37	68,0624	9,7868	95,782
INDICE con Cp*varianza + media	47,5	33,14	80,64	68,0624	9,7868	95,782

Finalmente, debemos comentar que por lo expuesto en este trabajo, el Análisis de Componentes Principales no lineal es una herramienta que no se ha explotado adecuadamente en diversos campos, dado que en general es más fácil obtener información cualitativa que puede ser transformada en cuantitativa en cualquier campo de trabajo como se vio en el capítulo tres, además que es una metodología alternativa a los métodos clásicos.

6.2 RECOMENDACIONES

En primer lugar consideró que se deberían homogenizar las fuentes de información de tal manera que se puedan realizar comparaciones más eficaces para distintos periodos.

Además, se recomienda a las instituciones del Estado que desarrollen metodologías que ayuden a segmentar a la población por el acceso a bienes y servicios y por ciertas características demográficas, dado que el ingreso y gasto no son muy confiables.

Las fuentes de información con las que cuentan las instituciones del Estado no están siendo correctamente utilizadas, dado que la información recopilada en las diferentes encuestas puede ser explotada de mejor manera.

Finalmente considero que la metodología planteada podría servir especialmente a instituciones que se dedican a la investigación de mercados, en donde la diferenciación socioeconómica es de vital importancia, además en dichas instituciones puede manejarse de mejor manera la información cualitativa que es de más fácil acceso.

BIBLIOGRAFÍA

- Meulman J, Heiser W, SPSS Categorías 11.0, SPSS Inc. 2001.
- Albert Gifi, Nonlinear Multivariate Analysis, John Wiley and Sons, 1990.
- Carrillo Granda, Mario Martín, Análisis no lineal de variables categóricas de las Encuestas de Condiciones de Vida para la construcción de un indicador de calidad de vida, Junio 2000
- Salazar Bermeo, Franklin Eduardo, Análisis de Homogeneidad en el Análisis Multivariante no lineal, Estudio en centros de rehabilitación social, Septiembre 2000.
- Rodríguez Vignoli, Jorge, Segregación Residencial Socioeconómica: ¿Qué es? ¿Cómo se mide?, ¿Qué está pasando?, ¿Importa? .Serie 16, CEPAL-ECLAC. Santiago de Chile, Agosto 2001.
- INE Chile, Metodología de Clasificación Socioeconómica de los hogares chilenos, 2003.
- Luis Fernando Gamboa, José Alberto Guerra, Andrés Fernando Casas, Nohora Y. Forero, Cambios en calidad de vida en Colombia, durante 1997-2003: otra aproximación.