



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ciencias Matemáticas

Escuela Profesional de Estadística

**Determinación de la estructura socioeconómica de los
hogares de Lima Metropolitana para el año 2007**

MONOGRAFÍA

Para optar el Título Profesional de Licenciado en Estadística

AUTOR

David Ismael ZÚÑIGA VELA

ASESOR

Grabiela Yolanda MONTES QUINTANA

Lima, Perú

2008



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Zúñiga, D. (2008). *Determinación de la estructura socioeconómica de los hogares de Lima Metropolitana para el año 2007*. Tesis para optar el título profesional de Licenciado en Estadística. Escuela Profesional de Estadística, Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima, Perú.

HOJA DE METADATOS COMPLEMENTARIOS

Código Orcid del autor (dato opcional):

No cuenta con código

Código Orcid del asesor o asesores (dato obligatorio):

0000-0002-9234-7049

DNI del autor:

40127096

Grupo de investigación:

No tiene grupo de investigación

Institución que financia parcial o totalmente la investigación:

Autofinanciado

Ubicación geográfica donde se desarrolló la investigación. Debe incluir localidades y coordenadas geográficas:

Lima Metropolitana, Longitud: -77.0282400, Latitud: -12.0431800

Año o rango de años que la investigación abarcó:

2007



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú, DECANA DE AMERICA
FACULTAD DE CIENCIAS MATEMATICAS

PROGRAMA DE TITULACIÓN POR EXAMEN DE
SUFICIENCIA PROFESIONAL 2007
ESCUELA ACADEMICO PROFESIONAL DE ESTADISTICA

ACTA DE SUSTENTACION DE MONOGRAFIA PARA OBTENER EL TITULO
PROFESIONAL DE LICENCIADO EN ESTADISTICA

En la Ciudad Universitaria, Facultad de Ciencias Matemáticas, siendo las *17:54* hrs. del día 27 del mes de **junio** del año 2008 se reunieron las docentes designadas como miembros del Jurado de Monografía:

Lic. Blanca Martínez Portuguez (Presidenta)

Lic. Grabiela Yolanda Montes Quintana (Asesora)

Para la sustentación de la Monografía intitulada **“DETERMINACIÓN DE LA ESTRUCTURA SOCIOECONÓMICA DE LOS HOGARES DE LIMA METROPOLITANA PARA EL AÑO 2007”**, presentada por el participante Bachiller **DAVID ISMAEL ZUÑIGA VELA**, para obtener el Título Profesional de Licenciado en Estadística.

Luego de la exposición de la Monografía, la Presidenta invitó al expositor a dar respuesta a las preguntas formuladas.

Hecha la evaluación correspondiente por los miembros del jurado, el expositor mereció la aprobación... **SOBRESALIENTE** con un calificativo promedio de **DIECISIETE (17)**.....(letra y números).

A continuación los miembros del Jurado, dan manifiesto que el participante Bachiller **DAVID ISMAEL ZUÑIGA VELA**, en virtud de haber aprobado la sustentación de su Monografía, será propuesto para que se le otorgue el Título Profesional de Licenciado en Estadística.

Siendo las *18:50* horas, se levantó la Sesión, firmando para constancia la presente Acta en tres (3) copias originales.

Lic. Grabiela Montes Quintana
Asesora

Lic. Blanca Martínez Portuguez
Presidenta

RESUMEN

DETERMINACIÓN DE LA ESTRUCTURA SOCIOECONÓMICA DE LOS HOGARES DE LIMA METROPOLITANA PARA EL AÑO 2007

DAVID ISMAEL ZÚÑIGA VELA

JUNIO – 2008

Orientadora: Gabriela Montes Quintana

En la presente monografía se determinará, a través de técnicas estadísticas de análisis multivariante, una estructura socioeconómica para los hogares de Lima Metropolitana. Se utilizará el Análisis de Componentes Principales Categórico (CATPCA) para elaborar esta clasificación socioeconómica.

Para esto, se cuenta con una muestra de 1200 hogares de Lima Metropolitana, seleccionada dentro de 39 distritos.

Al comparar los resultados obtenidos mediante la aplicación del método CATPCA, y otros resultados obtenidos por algunas empresas de investigación de mercado en el Perú, podemos apreciar que la segmentación calculada en esta monografía es válida y eficiente.

PALABRAS CLAVES: CATPCA
 COMPONENTES PRINCIPALES CATEGÓRICOS
 NIVELES SOCIOECONÓMICOS
 HOGARES
 VARIABLES

ABSTRACT

DETERMINATION OF THE SOCIOECONOMIC STRUCTURE OF THE
HOUSEHOLDS OF LIMA METROPOLITANA FOR THE YEAR 2007

DAVID ISMAEL ZÚÑIGA VELA

JUNE - 2008

Orientation: Grabiela Montes Quintana

In this monograph, a socioeconomic structure for households in Metropolitan Lima will be determined through multivariate statistical techniques. Categorical Principal Components Analysis (CATPCA) will be used to develop this socioeconomic classification.

For this, there is a sample of 1200 households in Metropolitan Lima, selected within 39 districts.

When comparing the results obtained by applying the CATPCA method, and other results obtained by some market research companies in Peru, we can see that the segmentation calculated in this monograph is valid and efficient.

KEY WORDS:

CATPCA

MAIN CATEGORICAL COMPONENTS

SOCIOECONOMIC LEVELS

HOUSEHOLDS

VARIABLES

ÍNDICE

1.	Introducción.....	6
1.1.	Objetivos de la investigación.....	6
1.2.	Antecedentes y justificación de la investigación.....	6
1.3.	Descripción resumida de los contenidos de cada capítulo	9
2.	Definición de términos técnicos utilizados en la investigación.....	10
3.	Marco Teórico	12
3.1.	Universo y Ámbito de estudio	12
3.2.	Marco Muestral.....	12
3.3.	Tamaño Muestral	12
3.4.	Método y Selección de la muestra	14
3.5.	Error estimado	14
3.6.	Instrumento de medición	14
3.7.	Técnica de Entrevista.....	14
3.8.	Consistencia y Validación	15
3.9.	Factor de expansión	15
4.	Análisis de Componentes Principales Categórico (CATPCA).....	16
4.1.	Introducción.....	16
4.2.	Datos.....	17
4.3.	Supuestos	17
4.4.	Notación.....	18

4.5.	Discretización	21
4.6.	Imputación de valores perdidos	22
4.7.	Configuración	23
4.8.	Optimización de la Función Objetivo.....	24
4.9.	Objetos Suplementarios	30
4.10.	Variables Suplementarias	30
4.11.	Diagnóstico.....	31
5.	Aplicación.....	41
5.1.	Introducción.....	41
5.2.	Variables seleccionadas	41
5.3.	Análisis exploratorio de las variables analizadas	42
5.4.	Metodología de la clasificación socioeconómica	52
5.5.	Resultados.....	53
5.6.	Perfiles de los hogares por nivel socioeconómico.....	57
5.7.	Comparación con resultados de empresas de investigación de mercados... 57	
6.	Bibliografía.....	65

1. Introducción

1.1. Objetivos de la investigación

1.1.1. Objetivo General

El objetivo del presente trabajo es determinar, a través de técnicas estadísticas de análisis multivariante, una estructura socioeconómica de los hogares de Lima Metropolitana.

1.1.2. Objetivos Específicos

- Determinar las variables que influyen en la determinación de los niveles socioeconómicos en Lima Metropolitana.
- Conocer los perfiles de los hogares en cada una de los niveles socioeconómicos de Lima Metropolitana.

1.2. Antecedentes y justificación de la investigación

El tema de segmentar las poblaciones en estratos o niveles socioeconómicos data de hace mucho tiempo. Aproximadamente en los años 30, en países como Inglaterra o Estados Unidos se ensayan las primeras clasificaciones, a raíz de esto se generaron muchas críticas y discusiones, desencadenando es un debate que siempre está vigente.

Conceptos como “clases”, “niveles” o “estratos sociales” traen consigo muchos temas o matices ideológicos. Por este motivo, en casi todos los países, si no es en todos, estas clasificaciones de la población no la realiza o la trabaja la estadística oficial o el gobierno, en la mayoría de los casos esto se lleva a cabo a través de empresas de investigación de mercados y opinión pública.

Existen muchas opiniones sobre si tiene o no sentido hablar de “clases” o “niveles sociales”, debido a que se considera que ya en estos tiempos no existen clases o niveles cerrados o claramente identificables como si había antes. Actualmente, tanto la movilidad social y la permeabilidad entre niveles son actualmente muy altos, con esto los límites entre los grupos ya no son tan marcados como lo eran antes.

Está claro que resultaría casi imposible ser parte de una comunidad donde no existan desigualdades en el acceso a educación y cultura, poder adquisitivo, entre muchas otras cosas, por lo que se necesita clasificar para poder esquemas y se pueda explicar la realidad que se está viviendo.

La idea es que la población se distribuya en segmentos o niveles que se diferencien en base a variables como nivel de cultura, nivel de educación, estilo de vida, condición material de vida y capacidad adquisitiva. Se debería dejar de utilizar un concepto estricto como “nivel” o “clase social”, y en su lugar, llegar a utilizar conceptos como “status” o “nivel socioeconómico”, para conocer por ejemplo patrones de consumo, y con esto poder evaluar la necesidad de algunos servicios y servicios.

El objetivo sería construir una escala, considerando tanto a los que tienen un mayor status socioeconómico, como los que menos tienen, estando dentro de estos extremos el resto de la población. Cuando ya se tenga esta gradación, se deberían establecer algunos cortes que permitan dividir la población en un determinado número de grupos, y luego se le debería asignar un nombre a cada grupo para que estén identificados. Obviamente se pueden hacer la cantidad de grupos que se desee, sólo se debe tener en cuenta siempre que lo importante es la gradación continua de los individuos, y esto nos permitirá distinguirlos.

Para armar este orden por niveles socioeconómicos, normalmente se les asignan puntuaciones a los integrantes de la población. El puntaje de cada hogar resulta de acumular puntos de acuerdo a todas las variables que se tiene, y también de la combinación de las mismas de acuerdo a la importancia que puedan tener. Estas variables que son tomadas para establecer el sistema de puntajes, serán las mismas que ayudarán a entender las diferencias que pueden existir entre los grupos, los patrones de consumo y estilos de vida.

Normalmente se utilizan variables como el ingreso, grado de instrucción, condición laboral, calidad de la vivienda y bienes que posee. Las variables a considerar deberían ser de estos dos tipos:

-
- Indicadores Económicos, se refiere a los atributos o bienes donde su valor es expresado en términos monetarios de forma directa, por ejemplo podemos considerar los ingresos, bienes domésticos, propiedades.
 - Indicadores Sociales, se refiere a los atributos o bienes donde su valor es expresado mayormente en su función como símbolos de status, por ejemplo se podría considerar la ocupación, nivel de educación.

Algunos atributos o bienes pueden ser considerados en ambas funciones, podríamos considerar por ejemplo el automóvil y la vivienda.

Del mismo modo, se debería considerar en las variables estos dos tipos de indicadores:

- Indicadores de Flujo, simbolizan un flujo de valor y ayudan a conocer la situación actual. Se puede considerar en este rubro la ocupación, ingreso y si se tiene algunos bienes en particular.
- Indicadores de Stock, se refiere a propiedades, bienes o fortuna que puede haber obtenido en todo lo que lleva de su vida, sin importar la situación que se tenga actualmente (podría contar con bienes y ser un desempleado o cesante). Por ejemplo podemos considerar el nivel educativo, y la vivienda como un bien permanente.

Se debe tener en cuenta que muchos indicadores pueden estar relacionados entre sí, por ejemplo, el nivel educativo puede determinar en la ocupación que se tenga, la ocupación puede determinar también los ingresos, y este último puede determinar en gran parte medida que tan posible o no es comprar algunos bienes.

El Nivel Socioeconómico es una variable no definida a nivel oficial, teóricamente controvertida, también se podría decir que es una variable no directa, no observable y es indudable la influencia que tiene en las condiciones de comportamiento de la población.

Una pregunta que se podría surgir sería: ¿Cuántas variables como mínimo se podrían necesitar para poder construir un buen indicador socioeconómico?.

Ante esto, se podría considerar teóricamente lo siguiente:

- Considerar variables observables que representen de algún modo varios indicadores contenidos en el concepto analizado.
- Analizar su congruencia para estructurar la variable que se busca, desde el punto de vista de la utilidad y valor.

-
- Estimar el valor de la variable Nivel Socioeconómico de acuerdo con los criterios que se observaron, buscando mantener tendencias históricas.

Entonces, el reto es construir un Índice de Clasificación Socioeconómica que cumpla con lo siguiente, debe ser:

- Universal, aplicable a diferentes tipos de encuestas.
- Simple, basado en pocas variables y que sea fácil de calcular.
- Objetivo, basado en variables objetivas y reproducibles.
- Coherente, que cuando se aplique lleve a un resultado único.
- Consistente con la historia, que se base en similares principios metodológicos.

1.3. Descripción resumida de los contenidos de cada capítulo

En el primer capítulo se presentan los objetivos, antecedentes y justificación del trabajo realizado.

En el segundo capítulo se muestran definiciones de algunos términos utilizados durante el desarrollo del presente trabajo.

En el tercer capítulo veremos detalles relacionados al marco teórico del trabajo, donde incluye universos, marco muestral, tamaño muestral, entre otros.

En el cuarto capítulo se encuentra la teoría relacionada con el Análisis de Componentes Principales Categórico (CATPCA).

En el quinto capítulo se desarrolla la aplicación en sí, que consiste en la determinación de la estructura socioeconómica en Lima Metropolitana.

2. Definición de términos técnicos utilizados en la investigación

➤ **Análisis Multivariante:**

Es un método estadístico que tiene como objetivo analizar un grupo de datos que son multivariantes a la vez o de manera simultánea, teniendo en cuenta que existen para cada objeto estudiado más de una variable medida. Existe y es importante porque permite tener una mejor comprensión de lo que se está estudiando, y se puede conseguir información donde otros métodos que son univariantes y también bivariantes no consiguen.

➤ **Variable:**

Variable es una característica que se está midiendo en el estudio. Estadísticamente se puede decir que una variable es la que representa las mediciones en el análisis. Datos que están medidos en una escala ordinal o continua se expresan a través de una variable, tal como sucede con variables nominales que tienen sólo dos categorías. No obstante, para el caso de los datos nominales que presentan más de dos categorías se deben expresar con más de una variable.

➤ **Variables Socioeconómicas:**

Estas variables informan una situación o estado de un individuo u hogar, podemos considerar por ejemplo: nivel educativo, nivel de ingresos y ocupación. Para dividir o clasificar mercados estos son criterios objetivos. Para determinar o definir una clase social a menudo se combinan variables.

➤ **Universo:**

Como su nombre lo indica, es todo lo que existe, todo lo que hay, se toma como una totalidad. Número total de hogares o personas que se consideran dentro de un área geográfica específica, para fines de un estudio. Ejemplos: Población de la ciudad de Lima, practicantes

de un deporte, etc.

➤ **Muestra:**

Es un subconjunto o grupo de datos que se extrae de la población, y se toma con la finalidad de hacer inferencias sobre la población a partir del análisis del comportamiento de los elementos que la conforman. Obviamente, si la muestra es una parte de la población entonces la muestra tendrá un menor número o cantidad de elementos. Cuando el procedimiento utilizado para selección permite calcular la probabilidad de seleccionar cada una de las unidades o individuos de la población, se dice que la muestra es probabilística

➤ **Tamaño muestral:**

Se refiere a la cantidad de datos que se extraerán de la población, y que pasan a formar parte de la muestra. Cuando se recogen datos relativos a las características de un grupo de objetos o individuos suele ser imposible observar a todo el grupo, sobre todo si este grupo es muy grande. En lugar de examinar a todo el grupo (población o universo), se estudia sólo una parte pequeña (muestra).

➤ **Métodos de Muestreo:**

La selección de la muestra se puede realizar a través de procedimientos aleatorios o no aleatorios. En el caso de procedimientos aleatorios se tratará de muestreo probabilístico, mientras que en el caso de procedimientos no aleatorios se trata de un muestreo no probabilístico.

➤ **Error muestral:**

Este error también es conocido como error de estimación y es originado por la desviación entre un valor estimado que pertenece a una muestra y un valor real que pertenece a una población.

3. Marco Teórico

3.1. Universo y Ámbito de estudio

Para este estudio, se considera a 39 distritos en Lima Metropolitana, excluyendo los distritos considerados como balnearios (los cuales cuentan con un número de viviendas bajo, además de implicar mayores costos de desplazamientos).

Para el año 2007, según proyecciones realizadas en CPI, Lima Metropolitana cuenta con 1 869 514 hogares.

Entendamos que se considera un hogar cuando existe un grupo de personas que habitan una vivienda, pero adicionalmente preparan y consumen sus alimentos entre ellos mismos.

3.2. Marco Muestral

El marco muestral empleado para este estudio fue el Marco de unidades primarias de muestreo (UPM's) y viviendas CPI 2007, el cual se construyó en base a los resultados del Censo 2005. Se contó además con cartografía digital del INEI del año 2005.

3.3. Tamaño Muestral

Se consideró un tamaño de muestra de 1 200 hogares, asignada proporcionalmente a la distribución de viviendas en cada uno de los distritos de Lima Metropolitana. Esta muestra asignada fue distribuida en 310 unidades primarias de muestreo (UPM's). Las UPM's son conglomerados de 80 viviendas en promedio.

DISTRIBUCIÓN DE LA MUESTRA		
DISTRITO	MUESTRA HOGARES	MUESTRA UPM's
LIMA	1 070	276
LIMA	48	12
ATE	64	16
BARRANCO	6	2
BREÑA	14	4
CARABAYLLO	28	7
CHACLACAYO	6	2
CHORRILLOS	40	10
COMAS	66	17
EL AGUSTINO	24	6
INDEPENDENCIA	28	7
JESUS MARIA	10	3
LA MOLINA	20	5
LA VICTORIA	32	8
LINCE	10	3
LOS OLIVOS	44	11
LURIGANCHO	14	4
MAGDALENA DEL MAR	10	3
PUEBLO LIBRE	14	4
MIRAFLORES	16	4
PUENTE PIEDRA	32	8
RIMAC	28	7
SAN BORJA	18	5
SAN ISIDRO	12	3
SAN JUAN DE LURIGANCHO	122	31
SAN JUAN DE MIRAFLORES	50	13
SAN LUIS	8	2
SAN MARTIN DE PORRES	90	23
SAN MIGUEL	22	6
SANTA ANITA	24	6
SANTIAGO DE SURCO	46	12
SURQUILLO	16	4
VILLA EL SALVADOR	54	14
VILLA MARIA DEL TRIUNFO	54	14
CALLAO	130	34
CALLAO	60	15
BELLAVISTA	12	3
CARMEN DE LA LEGUA REYNOSO	6	2
LA PERLA	10	3
LA PUNTA	2	1
VENTANILLA	40	10
TOTAL LIMA METROPOLITANA	1 200	310

3.4. Método y Selección de la muestra

Se realizó la selección de esta muestra en cuatro etapas empleando métodos estrictamente probabilísticos:

- En primera etapa se seleccionaron las unidades primarias de muestreo (UPM's).
- En segunda etapa se seleccionaron sistemáticamente 4 viviendas dentro de cada UPM seleccionada.
- En tercera etapa, en caso de existir en la vivienda más de un hogar, se seleccionó uno de ellos de forma aleatoria.
- En una última etapa, se entrevistó a las persona que MÁS APORTA al sostenimiento del hogar (jefe del hogar).

3.5. Error estimado

Los tamaños de muestra consideran un error aproximado de +/- 3%.

3.6. Instrumento de medición

Se aplicó un cuestionario estructurado y pre-codificado en el hogar del entrevistado, el cual recoge variables para realizar una evaluación socioeconómica.

3.7. Técnica de Entrevista

Entrevista personal en cada hogar seleccionado.

3.8. Consistencia y Validación

La base de datos obtenida del trabajo de campo fue consistenciada y validada a través de la construcción de tablas de contingencia y la aplicación de rangos de validez.

3.9. Factor de expansión

El factor de acuerdo con el diseño, corresponde al aplicado en la muestra en la cantidad de etapas mencionadas anteriormente, asignándole una probabilidad de selección a cada hogar elegido, dependiendo de cuantos hogares se tenga en cada distrito (conglomerado), por tanto el factor de expansión será la inversa de esta probabilidad asignada.

Se interpreta este factor como la cantidad de hogares que representa en una población, un hogar seleccionado en una muestra.

4. Análisis de Componentes Principales Categórico (CATPCA)

4.1. Introducción

A este procedimiento se le conoce más por la abreviatura CATPCA, que proviene de su mismo nombre en inglés (CATegorical Principal Components Analysis).

Tiene como objetivo reducir un conjunto original de variables (sin importar que tipos de variables sean) en un conjunto reducido de componentes que no se encuentren relacionadas entre ellas, y que representen la mayor cantidad de la información que se encuentra en este conjunto original de variables.

Se podría decir que la principal diferencia entre el análisis clásico de componentes principales (PCA) y el análisis de componentes principales categórico (CATPCA) se refiere al tipo de variables que usa. El PCA requiere variables que se encuentren en una escala métrica, y por otro lado el CATPCA usa cualquier tipo de variables. Debido a que las variables que se van a considerar en el análisis pueden tener un nivel de media ordinal, nominal y numérico, será necesario usar CATPCA.

Es de mayor utilidad la técnica cuando se tiene una gran cantidad de variables y estas no permiten hacer una correcta interpretación de las relaciones existentes que hay entre objetos. Cuando se reducen las dimensiones, se va a interpretar una cantidad reducida de componentes, y ya no un gran número de variables.

Para el análisis clásico de componentes principales se considera que entre las variables numéricas existen relaciones lineales. No obstante, la aproximación o estimación por escalamiento óptimo CATPCA facilita hacer un escalamiento a diferentes categorías o niveles a las variables. Las variables con categorías son cuantificadas en las dimensiones especificadas de manera óptima y gracias a esto, se modelan entre las variables relaciones no lineales.

Para poder estimar los parámetros, el análisis de componentes principales categórico (CATPCA) usa como algoritmo computacional el método de los mínimos cuadrados

alternativos.

Las variables con categorías son cuantificadas por CATPCA a través del óptimo escalamiento, y con esto resultan óptimos componentes principales para las variables nuevas o transformadas. Estas variables pueden generar niveles óptimos de escalamiento mixtos y es innecesario hacer alguna suposición sobre la distribución de variables.

En este análisis categórico, las dimensiones corresponden a componentes (esto quiere decir que un análisis con dos dimensiones va a dar como resultado dos componentes), y las puntuaciones o puntajes de objetos corresponden a las puntuaciones de los componentes.

4.2. Datos

Sobre los datos se podría decir lo siguiente, se convierten los valores que tienen las variables de cadena en enteros positivos considerando orden alfabético de manera ascendente. Se va a considerar como valores perdidos a todos los valores definidos por el usuario con esta condición (como perdidos), los que fueron definidos por el sistema, y todos los valores que sean menores a 1 también. Para evitar que estos valores perdidos se pierdan, a las variables que presentan valores menores a 1 se les puede añadir valor o se les puede hacer una recodificación.

4.3. Supuestos

Sobre este tema se debe considerar que los datos deben tener como mínimo tres casos que se puedan considerar como válidos. Se enfoca el análisis en datos de tipo entero y con valores positivos. Con alternativa de discretización automáticamente se va a categorizar una variable que presenta o tiene valores continuos o fraccionados, haciendo una agrupación de sus valores en categorías aproximando su distribución a una “normal”, y automáticamente convierte en enteros positivos los valores que presentan las variables con el tipo cadena. Es posible especificar otros esquemas diferentes de discretización.

4.4. Notación

Se define la siguiente notación:

- n Número de objetos o casos analizados.
- n_w Número ponderado de casos analizados: $\sum_{i=1}^n w_i$
- n_{tot} Número total de casos (analizados + suplementario)
- w_i Ponderación del objeto i ; $w_i = 1$ si los casos no están ponderados; $w_i = 0$ si el objeto i es suplementario.
- W** Matriz diagonal $n_{tot} \times n_{tot}$, con w_i en la diagonal.
- m Número de variables analizadas.
- m_w Número ponderado de variables analizadas ($m_w = \sum_{j=1}^m v_j$)
- m_{tot} Número total de variables (análisis y suplementario)
- m_1 Número de variables analizadas con nivel múltiple de escalamiento nominal.
- m_2 Número de variables analizadas con nivel no-múltiple de escalamiento.
- m_{w1} Número ponderado de variables analizadas con nivel múltiple de escalamiento nominal.
- m_{w2} Número ponderado de variables analizadas con nivel no-múltiple de escalamiento.
- J** Índice que registra que variables tienen nivel de escalamiento nominal múltiple.
- H** Matriz de datos (indicadores categóricos), con orden $n_{tot} \times m_{tot}$, después de discretización, imputación de valores perdidos, y eliminación mediante

lista, de ser aplicable.

p Número de dimensiones.

Considerar en variable j , que $j = 1, \dots, m_{tot}$.

v_j Peso de la variable. $v_j = 1$ si el peso de la variable j no es especificado o si la variable j es suplementaria.

k_j Número de categorías de la variable j (número de valores distintos en h_j , por lo tanto, incluyendo objetos suplementarios)

\mathbf{G}_j Matriz del indicador para la variable j , de orden $n_{tot} \times k_j$

Los elementos de \mathbf{G}_j se definen como $i = 1, \dots, n_{tot}$; $r = 1, \dots, k_j$

$g_{(j)ir} = \begin{cases} 1 & \text{cuando el } i\text{-ésimo objeto está en la } r\text{-ésima categoría de la variable } j \\ 0 & \text{cuando el } i\text{-ésimo objeto está en la } r\text{-ésima categoría de la variable } j \end{cases}$

\mathbf{D}_j Matriz diagonal $k_j \times k_j$, contiene los marginales univariados ponderados, es decir, las sumas de las columnas ponderadas de \mathbf{G}_j ($\mathbf{D}_j = \mathbf{G}'_j \mathbf{W} \mathbf{G}_j$).

\mathbf{M}_j Matriz Diagonal de orden $n_{tot} \times n_{tot}$, con elementos de la diagonal definidos como

$m_{(j)ii} = \begin{cases} 0 & \text{cuando la } i\text{-ésima observación está perdida y la variable estratégica} \\ & \text{perdida } j \text{ es pasiva} \\ 0 & \text{cuando el } i\text{-ésimo objeto está en la } r\text{-ésima categoría de la variable} \\ & j \text{ y la } r\text{-ésima categoría es usada sólo por objetos suplementarios (es} \\ & \text{decir, cuando } d_{(j)rr} = 0) \\ v_j & \text{otro caso} \end{cases}$

$\mathbf{M}_* = \sum_j \mathbf{M}_j$

\mathbf{S}_j I-líneaSP base de la variable j , de orden Matriz diagonal $k_j \times (s_j + t_j)$
(ver Ramsay (1988) para detalles)

\mathbf{b}_j Vector líneaSP de coeficientes, de orden $s_j + t_j$

d_j Intercepto de líneaSP.

s_j Grado de polinomio.

t_j El número de nudos al interior

La cuantificación de matrices y parámetros de vectores es:

\mathbf{X} Puntuaciones de los objetos, de orden $n_{tot} \times p$

\mathbf{X}_w Puntuaciones ponderadas de los objetos ($\mathbf{X}_w = \mathbf{W}\mathbf{X}$)

\mathbf{X}^n \mathbf{X} normalizado según la opción de normalización requerida

\mathbf{Y}_j Coordenadas del centroide, de orden $k_j \times p$. Para variables con nivel múltiple de escalamiento nominal, esto es cuantificaciones categóricas.

\mathbf{y}_j Cuantificaciones categóricas para las variables con nivel no-múltiple de escalamiento, de orden k_j

\mathbf{a}_j Saturaciones que presentan componentes con nivel no-múltiple de escalamiento, orden p

\mathbf{a}_{n_j} \mathbf{a}_j normalizado según la opción de normalización requerida

\mathbf{Y} Colección de cuantificaciones categóricas (coordenadas del baricentro) para variables con nivel múltiple de escalamiento nominal (\mathbf{Y}_j), y coordenadas del vector para nivel no-múltiple de escalamiento ($\mathbf{Y}_j \mathbf{a}'_j$).

Nota: Las matrices \mathbf{W} , \mathbf{G}_j , \mathbf{M}_j , \mathbf{M}_* , y \mathbf{D}_j , son dispositivos exclusivos de notación, y son almacenados en forma reducida, y el programa se beneficia completamente de sus escasas multiplicaciones por reemplazo de la matriz con acumulación selectiva.

4.5. Discretización

Se agrupan en siete categorías (o en un número menor, dependiendo de los diferentes valores de la variables) las variables que tienen valores discontinuos o fraccionados considerando una distribución casi normal, de no especificarse algo distinto. Siempre se convierten a valores enteros positivos las variables de tipo cadena, asignando indicadores de categoría considerando un orden alfanumérico de manera ascendente. Todas las variables que restan no sufren ninguna alteración. Y luego de esto, se usan estas variables que se han discretizado en el análisis.

Esta discretización se efectua en los datos no ponderados.

Entre los métodos de discretización se tiene:

4.5.1. Agrupación

Tenemos que redefinir en una cantidad de específica de categorías, o redefinir por intervalos. Cuando se hace la discretización por agrupación tenemos estas opciones disponibles:

➤ **En una cantidad establecida de categorías con distribución normal**

Inicialmente, se estandariza la variable original. Entonces, usando intervalos definidos en Max (1960), los casos son asignados a categorías.

➤ **En una cantidad establecida de categorías con distribución uniforme**

Se calcula inicialmente la frecuencia objetivo dividiendo n entre el número de categorías especificadas, y el resultado se redondea. Luego, las categorías originales se asignarán a categorías agrupadas de tal forma que las frecuencias de las categorías agrupadas estén tan cerca de la frecuencia objetivo como esto sea posible.

➤ **Agrupando los intervalos iguales de tamaño específico**

En este caso, se definen los intervalos como *el menor valor + el tamaño del intervalo*, *el menor valor + 2*el tamaño del intervalo*, etc. Entonces los casos con valores en el k -ésimo intervalo son asignados a la categoría k .

4.5.2. Asignación de rangos o Ranking

Se clasifica en orden ascendente la variable original, de acuerdo al valor alfanumérico.

4.5.3. Multiplicación

Inicialmente, se estandariza la variable original. Luego de esto, se multiplican los valores estandarizados por 10 y se redondea el resultado, después se le agrega un valor de modo que el menor valor sea 1.

4.6. Imputación de los valores perdidos

Entre las estrategias de imputación se maneja lo siguiente: se puede optar por la exclusión de valores perdidos (se denomina tratamiento pasivo a esto), imputación de valores perdidos (se llama tratamiento activo a esto) o exclusión de objetos que presenten valores perdidos (se conoce como eliminación por lista a este proceso).

4.6.1. Exclusión de valores perdidos.

Si en una variable seleccionada se tienen objetos que presentan valores perdidos, estos objetos no aportan en el análisis. De aplicarse tratamiento pasivo a las variables, serán tratados como suplementarios los objetos con que presentan valores perdidos. Si en el cuadro

de diálogo *Resultados* se especifican correlaciones, luego del análisis, se van a imputar los valores perdidos con la moda para las correlaciones de las variables iniciales. Se puede seleccionar el método de imputación para correlaciones de una variable óptimamente escalada. Si se desea sustituir perdidos por la cuantificación de una categoría adicional seleccionar *Categoría adicional*. Con esto, a los objetos que tengan en esta variable un valor perdido se les considerará como parte de la categoría adicional.

4.6.2. Imputación de valores perdidos.

En la variable que se seleccionó, se imputarán los valores de los objetos que presentan valores perdidos. El método para la imputación puede ser escogido. Se debe seleccionar *Moda* si los valores perdidos se reemplazarán por la categoría con mayor frecuencia. Si se diera el caso de tener varias modas, hay que elegir la moda que tiene el indicador de categoría menor. Si se desea sustituir perdidos por la cuantificación de una categoría adicional seleccionar *Categoría adicional*.

4.6.3. Exclusión de objetos con valores perdidos.

Se excluyen del análisis los objetos que en la variable que se seleccionó presentaran valores perdidos. Para las variables que son suplementarias no está disponible esta opción.

4.7. Configuración

El análisis de componentes principales categórico (CATPCA) puede leer desde un archivo una configuración, para ser utilizada como la configuración inicial, o también como una configuración fija en la que se ajusten las variables.

Para una configuración inicial revisar en la sección Optimización, el paso 1.

Una configuración fija \mathbf{X} es centrada y ortonormalizada como se indica en la sección

Optimización en el paso 3 (con \mathbf{X} en lugar de \mathbf{Z}) y paso 4 (excepto por el factor $n_w^{1/2}$), y el resultado posteriormente es multiplicado por $\Lambda^{1/2}$ (con esto la configuración queda sin cambios, esta ya es centrada y ortogonal). En el análisis, las variables se fijan como suplementarias y se fijan a uno los pesos de la variable. Siendo así, CATPCA procede como se indica en la sección Variables Suplementarias.

4.8. Optimización de Función Objetivo

4.8.1. Función Objetivo

CATPCA tiene como objetivo encontrar puntuaciones de objetos \mathbf{X} y un conjunto de $\underline{\mathbf{Y}}_j$ (para $j=1,\dots,m$) – el subrayado indica que pueden ser restringidos en diferentes maneras – de modo que la función

$$\sigma(\mathbf{X}; \underline{\mathbf{Y}}) = n_w^{-1} \sum_j c^{-1} \text{tr}((\mathbf{X} - \mathbf{G}_j \underline{\mathbf{Y}}_j) \mathbf{M}_j \mathbf{W} (\mathbf{X} - \mathbf{G}_j \underline{\mathbf{Y}}_j)), \text{ con}$$

igual a p si $j \in J$ y c es igual a 1 si $j \notin J$,

sea mínima, bajo la restricción de normalización $\mathbf{X} \mathbf{M}_* \mathbf{W} \mathbf{X} = n_w m_w \mathbf{I}$ (\mathbf{I} es la matriz identidad $p \times p$). La inclusión de \mathbf{M}_j en $\sigma(\mathbf{X}; \underline{\mathbf{Y}})$ garantiza que no exista una influencia de los valores perdidos pasivos (valores perdidos en variables con la opción de perdidos pasivos, o la opción de perdidos sin especificar). \mathbf{M}_* comprende para cada objeto el número de valores de datos activos. También son centradas las puntuaciones de objeto; esto significa que satisfacen $\mathbf{u}' \mathbf{M}_* \mathbf{W} \mathbf{X} = 0$ con \mathbf{u} que denota un n -vector con unos.

4.8.2. Niveles de Escalamiento Óptimo

Estos niveles de escalamiento óptimo son distinguidos en CATPCA:

-
- **LíneaSP Ordinal:** $\underline{Y}_j = \mathbf{y}_j \mathbf{a}'_j$, y , $\mathbf{y}_j = d_j + \mathbf{S}_j \mathbf{b}_j$ (restricciones de igualdad, rango-uno, y líneaSP monotónica), con \mathbf{b}_j limitado a contener elementos que no sean negativos (para asegurar I-líneasSP monotónicas). Se conserva el orden de las categorías de la variable observada en la variable óptimamente escalada. Los puntos de categoría se encontrarán sobre una recta o vector que pasará por el origen. La transformación que resulta sería un polinomio monotónico por tramos suave del orden seleccionado. Se precisan las partes por el número de nudos interiores que el usuario ha definido y el procedimiento determina su posición en función del número de nudos interiores.

 - **LíneaSP Nominal:** $\underline{Y}_j = \mathbf{y}_j \mathbf{a}'_j$, y , $\mathbf{y}_j = d_j + \mathbf{S}_j \mathbf{b}_j$ (restricción de igualdad, rango-uno, y líneaSP). De la variable observada únicamente se conserva, en la variable óptimamente escalada, la agrupación de los objetos en categorías. El orden de las categorías de la variable observada se pierde, no se conserva. Los puntos de categoría se encontrarán sobre una recta o vector que pasará por el origen. La transformación que resulta es un polinomio, podría ser monotónico, por tramos suave del orden que se seleccionó. Se especifican las partes en función del número de nudos interiores que ha definido el usuario y se determina su posición por el procedimiento, en función del número de nudos interiores.

 - **Nominal Múltiple:** $\underline{Y}_j = \mathbf{Y}_j$ (sólo restricción de igualdad). De la variable observada únicamente se conserva, en la variable óptimamente escalada, la agrupación de los objetos en categorías. El orden de las categorías de la variable observada se pierde, no se conserva. Para las categorías particulares, estarán en el centroide de los objetos, los puntos de categoría. La palabra Múltiple nos avisa que para cada dimensión se pueden obtener distintos grupos de cuantificaciones.

 - **Ordinal:** $\underline{Y}_j = \mathbf{y}_j \mathbf{a}'_j$, y , $\mathbf{y}_j \in \mathbf{C}_j$ (restricción de igualdad, monotonicidad y rango-uno). La restricción de monotonía $\mathbf{y}_j \in \mathbf{C}_j$ indica que \mathbf{y}_j debe localizarse en el cono

convexo de los k_j -vectores con elementos no decrecientes.

- **Nominal:** $\underline{\mathbf{Y}}_j = \mathbf{y}_j \mathbf{a}'_j$ (restricción de igualdad y rango-uno). En la variable óptimamente escalada sólo se guarda la segmentación en categorías de los objetos de la variable en observación. El ordenamiento de categorías de nuestra variable en observación no se guarda. Los puntajes de categoría encontrarán sobre una recta (vector) que pasará por el origen. Resulta una transformación que se ajusta mejor que la de líneaSp nominal pero tiene una menor suavidad.

- **Numérico:** $\underline{\mathbf{Y}}_j = \mathbf{y}_j \mathbf{a}'_j$, $\mathbf{y}_j \in \mathbf{L}_j$ (restricción de igualdad, linealidad y rango-uno). La restricción de linealidad $\mathbf{y}_j \in \mathbf{L}_j$ se refiere a que \mathbf{y}_j se debe localizar en todos los subespacios de los k_j -vectores siendo una transformación lineal del vector que está conformada de k_j enteros sucesivos. Se manejan las categorías como si estuvieran ordenadas y espaciadas de manera uniforme (a nivel de intervalo). Se guardan en nuestra variable óptimamente escalada tanto el ordenamiento de las categorías, como la equidistancia entre los números de estas categorías. Los puntajes de categoría encontrarán sobre una recta o vector), cruza el origen.

Los niveles se pueden escoger de manera independiente para cada una de las variables. Los indicadores de categorías iguales deben recibir cuantificaciones iguales. Para las opciones no múltiples $\underline{\mathbf{Y}}_j = \mathbf{y}_j \mathbf{a}'_j$; es decir, $\underline{\mathbf{Y}}_j$ es de rango-uno; para una mejor identificación, \mathbf{y}_j siempre se normaliza para que $\mathbf{y}'_j \mathbf{D}_j \mathbf{y}_j = n_w$.

4.8.3. Optimización

El sistema ejecuta lo siguiente para la optimización:

1. Inicialización I o II

-
2. Actualización de las cuantificaciones de categoría
 3. Actualización de las puntuaciones de objeto
 4. Ortonormalización
 5. Prueba de convergencia: Repetir (2)-(4) o seguir
 6. Rotación y reflexión

Se utiliza la inicialización I en la configuración inicial, además las variables sin nivel de escalamiento óptimo Nominal Múltiple o Numérico son tratadas como numéricas de manera temporal. La inicialización II es utilizada para la configuración final. Ahora se hace una descripción detallada de todos los pasos (del 1 al 6).

1. Inicialización

I. De no indicar una configuración inicial, el sistema inicializa las puntuaciones del objeto \mathbf{X} con números aleatorios. Luego \mathbf{X} se ortonormaliza de tal forma que $\mathbf{u}'\mathbf{M}_* \mathbf{W}\mathbf{X} = 0$ y $\mathbf{u}'\mathbf{M}_* \mathbf{W}\mathbf{X} = n_w m_w \mathbf{I}$, originando \mathbf{X}_w^+ . En las componentes iniciales, las saturaciones son calculadas como productos cruzados de \mathbf{X}_w^+ con las variables centradas originales, $(\mathbf{I} - \mathbf{M}_j \mathbf{u}\mathbf{u}'\mathbf{W}) / (\mathbf{u}'\mathbf{M}_j \mathbf{W}\mathbf{u}) \mathbf{h}_j$, ajustando nuevamente a la longitud de la unidad.

II. En su totalidad, las cantidades que relevantes se copian de los resultados obtenidos del primer ciclo.

2. Actualización de las cuantificaciones de categoría; ciclo de repetición a través de las variables $j = 1, \dots, m$:

Con los valores actuales \mathbf{X}_w^+ la actualización de \mathbf{Y}_j sería

$$\mathbf{Y}_j = \mathbf{D}_j^{-1} \mathbf{G}'_j \mathbf{X}_w^+$$

Nominal Múltiple: $\mathbf{Y}_j^+ = \mathbf{Y}_j$.

Inicialmente, para niveles de escalamiento no-múltiple, se calcula una actualización sin restricciones de este modo:

$$\mathbf{Y}_j = \mathbf{D}_j^{-1} \mathbf{G}'_j \mathbf{X}_w^+$$

luego se ejecuta un ciclo de un algoritmo ALS para calcular una descomposición rango-uno de \mathbf{Y}_j , resultando

$$\mathbf{y}_j = \tilde{\mathbf{Y}}_j \mathbf{a}_j$$

Nominal: $\mathbf{y}_j^* = \tilde{\mathbf{y}}_j$.

Para los cuatro niveles de escalamiento óptimo que siguen, si hubo una imputación con una categoría adicional en la variable j , en la fase inicial \mathbf{y}_j^* es categoría inclusiva k_j , y en la fase final es categoría exclusiva k_j .

LíneaSP nominal y líneaSP ordinal: $\mathbf{y}_j^* = d_j + \mathbf{S}_j \mathbf{b}_j$.

Se hace el cálculo de la transformación líneaSP como una regresión ponderada de $\tilde{\mathbf{y}}_j$ (con pesos de la diagonal de \mathbf{D}_j) en la I-líneaSP base \mathbf{S}_j . Se restringen los elementos de \mathbf{b}_j para el nivel de escalamiento líneaSP para que no resulten negativos, con esto \mathbf{y}_j^* sería creciente monotónicamente.

Ordinal: $\mathbf{y}_j^* \leftarrow WMON(\tilde{\mathbf{y}}_j)$.

$WMON()$ se usa para representar el proceso de regresión monotónico ponderado, que hace monotónica creciente a \mathbf{y}_j^* . Se usan los elementos de la diagonal de \mathbf{D}_j como pesos y se usa el algoritmo de mínimos infractores up-and-down-blocks

Numérico: $\mathbf{y}_j^* \leftarrow WLIN(\mathbf{y}_j)$.

La notación $WLIN(\)$ se usa para representar el proceso de regresión lineal ponderado. Se usan los elementos de la diagonal de \mathbf{D}_j como pesos.

Se normaliza \mathbf{y}_j^* (si se imputó con una categoría adicional la variable j , a partir de este momento \mathbf{y}_j^* es categoría inclusiva k_j):

$$\mathbf{y}_j^+ = n_w^{1/2} \mathbf{y}_j^* (\mathbf{y}_j^{*\prime} \mathbf{D}_j \mathbf{y}_j^*)^{-1/2}$$

Se actualiza las saturaciones de componentes:

$$\mathbf{a}_j^+ = n_w^{-1} \mathbf{Y}_j' \mathbf{D}_j \mathbf{y}_j^+$$

Resultando $\underline{\mathbf{Y}}_j^+ = \mathbf{y}_j^+ \mathbf{a}_j^{+\prime}$.

3. Actualización de las puntuaciones del objeto

Inicialmente se calcula la matriz auxiliar de puntuaciones \mathbf{Z} de la siguiente manera:

$$\mathbf{Z} \leftarrow \sum_j \mathbf{M}_j \mathbf{G}_j \underline{\mathbf{Y}}_j^+$$

centrada con respecto a \mathbf{W} y \mathbf{M}_* :

$$\mathbf{X}^* (\mathbf{I} - \mathbf{M}_* \mathbf{u} \mathbf{u}' \mathbf{W} / (\mathbf{u}' \mathbf{M}_* \mathbf{W} \mathbf{u})) \mathbf{Z}$$

4. Ortonormalización

Para conseguir una \mathbf{M}_* -ortonormal \mathbf{X}^+ cercano a \mathbf{X}^* , se utiliza para la rotación Procrustes (Cliff, 1966) la descomposición de valor singular $m_w^{1/2} \mathbf{M}_*^{-1/2} \mathbf{W}^{1/2} \mathbf{X}^* = \mathbf{K} \mathbf{\Lambda}^{1/2} \mathbf{L}'$, entonces $n_w^{1/2} m_w^{1/2} \mathbf{M}_*^{-1/2} \mathbf{W}^{1/2} \mathbf{K} \mathbf{L}'$ producen puntuaciones de objeto ponderadas \mathbf{M}_* -ortonormal:

$$\mathbf{X}_w^+ \leftarrow n_w^{1/2} m_w \mathbf{M}_*^{-1} \mathbf{W} \mathbf{X}^* \mathbf{L} \mathbf{\Lambda}^{-1/2} \mathbf{L}', \text{ y } \mathbf{X}^+ = \mathbf{W}^{-1} \mathbf{X}_w^+.$$

5. Prueba de convergencia

La diferencia de los valores consecutivos de

$$\text{TFIT} = (pn_w)^{-1} \sum_{j \in J} v_j \text{tr}(\mathbf{Y}'_j \mathbf{D}_j \mathbf{Y}_j) + \sum_{j \notin J} v_j \mathbf{a}'_j \mathbf{a}_j,$$

se compara con el criterio de convergencia usuario-especifico ε - un número positivo pequeño. Podría mostrar esto que $\text{TFIT} = m_{w1} + pm_{w2} - \sigma(\mathbf{X}; \mathbf{Y})$. Siempre y cuando la diferencia de pérdida exceda a ε , los pasos 2 hasta 4 se repiten.

Después la convergencia TFIT es también igual a $\text{tr}(\mathbf{\Lambda}^{1/2})$, con $\mathbf{\Lambda}$ como se calculó durante la última iteración en el paso 4.

6. Rotación y reflexión

\mathbf{X}^+ se rota con la matriz \mathbf{L} , para conseguir la orientación de los ejes principales. Si para la dimensión s , el promedio de las saturaciones al cuadrado con signo negativo resulta mayor que el promedio de las saturaciones al cuadrado con signo positivo, la s -ésima columna de \mathbf{X}^+ se refleja. Se produce la rotación y posiblemente el reflejo de las cuantificaciones y las saturaciones ejecutando el paso 2.

4.9. Objetos Suplementarios

Para realizar el cálculo de las puntuaciones de objeto para los objetos suplementarios, después de que los pasos 2 y 3 se repitan, con los ceros en \mathbf{W} establecidos de manera temporal en el cálculo de \mathbf{Z} y \mathbf{X}^+ . Se aplicará un tratamiento pasivo si un objeto suplementario presenta valores perdidos.

4.10. Variables Suplementarias

Luego de la convergencia, se calculan las cuantificaciones para las variables

suplementarias. Se va a ejecutar el paso 2 sólo una vez para variables suplementarias que tengan un nivel de óptimo escalamiento del tipo nominal múltiple. Por otro lado, para las variables suplementarias no-múltiples, se calcula un a_j inicial como en el paso 1. Mientras la diferencia entre los valores consecutivos de $\mathbf{a}'_j \mathbf{a}_j$ exceda a .00001, con un máximo de 100 iteraciones, se repiten el rango-uno y los subpasos de restricción del paso 2.

4.11. Diagnóstico

4.11.1. Rango máximo

La notación $p_{m\acute{a}x}$ indica el máximo número de dimensiones que se pueden calcular para un conjunto de datos. Tenemos que

$$p_{m\acute{a}x} = \text{mín} \left(n-1, \left(\sum_{j \in J} k_j \right) - m_1 + m_2 \right)$$

si hay variables que tienen un nivel de óptimo escalamiento del tipo nominal múltiple que no presentan valores perdidos que serán tratados como pasivos. Si se presentan variables que tienen un nivel de óptimo escalamiento del tipo nominal múltiple y hay valores perdidos con tratamiento pasivo, el rango máximo sería

$$p_{m\acute{a}x} = \text{mín} \left(n-1, \left(\sum_{j \in J} k_j \right) - \text{máx}(m_3, 1) + m_2 \right)$$

donde m_3 sería el número de variables que tienen un nivel de óptimo escalamiento del tipo nominal múltiple sin valores perdidos en tratamiento pasivo.

k_j es un objeto suplementario exclusivo (no se cuenta en el cálculo del máximo rango una categoría que solamente es utilizada por objetos suplementarios). A pesar de que el número de dimensiones no triviales podría ser menor que $p_{m\acute{a}x}$ cuando $m=2$, CATPCA admite todas las dimensionalidades hasta el $p_{m\acute{a}x}$.

El programa se detiene cuando, a raíz de la presencia de categorías vacías en los datos, ya que el rango se deteriora por debajo de la dimensionalidad que se especifica.

4.11.2. Descriptivos

Tablas descriptivas ofrecen para cada variable unas frecuencias marginales univariadas ponderadas y también unos números ponderados de valores perdidos (definidos por sistema, perdidos por definición del usuario, y valores menores o iguales a 0).

Se va a considerar como valores perdidos a todos los valores definidos por el usuario con esta condición (como perdidos), los que fueron definidos por sistema, y todos los valores que sean menores a 1 también

4.11.3. Medidas de Ajuste y de Pérdida

Con la opción HISTORIAL, se tiene estas medidas de ajuste y de pérdida:

- a. Ajuste total o Factor de Ajuste (VAF). Es la cantidad TFIT definida en el paso 5.
- b. Pérdida total. Es $\sigma(\mathbf{X}; \mathbf{Y})$, y se calcula sumando las pérdidas múltiples y pérdidas simples que se definirán luego.
- c. Pérdidas múltiples. El cálculo se realiza de la siguiente manera

$$TMLOSS = (m_{w1} + pm_{w2}) - \left(n_w p^{-1} \sum_{j \in J} v_j tr(\mathbf{Y}'_j \mathbf{D}_j \mathbf{Y}_j) + n_w^{-1} \sum_{j \notin J} v_j tr(\mathbf{Y}'_j \mathbf{D}_j \mathbf{Y}_j) \right)$$

- d. Pérdidas simples. Solamente en caso existan variables que sean simples se hace el cálculo de esto:

$$SLOSS = n_w^{-1} \sum_{j \notin J} v_j tr(\mathbf{Y}'_j \mathbf{D}_j \mathbf{Y}_j) - \sum_{j \notin J} v_j \mathbf{a}'_j \mathbf{a}_j$$

4.11.4. Resumen del modelo

➤ **Alfa de Cronbach**

Alfa de Cronbach por dimensión ($s = 1, \dots, p$):

$$\alpha_s = m_w (\lambda_s^{1/2} - 1) / (\lambda_s^{1/2} (m_w - 1)),$$

Alfa de Cronbach total es

$$\alpha = m_w (\sum_s \lambda_s^{1/2} - 1) / \sum_s \lambda_s^{1/2} (m_w - 1)$$

donde λ_s es el s -ésimo elemento de la diagonal de Λ calculado en la última iteración en el paso 4.

➤ **Varianza Considerada Para**

Varianza Considerada Para, por dimensión ($s = 1, \dots, p$):

- **Variables Nominales Múltiples**

$$\text{VAF1}_s = n_w^{-1} \sum_{j \in J} v_j \text{tr}(\mathbf{y}'_{(j)s} D_j \mathbf{y}_{(j)s}), \text{ (\% de varianza es } \text{VAF1}_s \times 100 / m_{w1} \text{)},$$

- **Variables No-Múltiples**

$$\text{VAF2}_s = \sum_{j \in J} v_j a_{js}^2, \text{ (\% de varianza es } \text{VAF2}_s \times 100 / m_{w2} \text{)}.$$

- **Autovalor**

Autovalor por dimensión

$$\lambda_s^{1/2} = \text{VAF1}_s + \text{VAF2}_s,$$

donde λ_s es el s -ésimo elemento de la diagonal de Λ que se calculó en la última

iteración del paso 4.

La Varianza Total Considerada para variables nominales múltiples es el promedio sobre dimensiones, y para variables no-múltiples sería la suma sobre dimensiones. De esta manera, el autovalor total es

$$tr(\Lambda^{1/2}) = p^{-1} \sum_s VAF1_s + \sum_s VAF2_s .$$

Si existen valores perdidos pasivos, entonces los autovalores $\Lambda^{1/2}$ son los de la matriz de correlación (ver secciones Correlaciones y Autovalores) ponderados con pesos de variable:

$$r_{jj}^w = v_j r_{jj}, \text{ y } r_{jl}^w = r_{lj}^w = v_j^{1/2} r_{jl}$$

Si existen valores perdidos pasivos, en consecuencia los autovalores son los de la matriz $m_w \mathbf{Q}'_c \mathbf{M}_*^{-1} \mathbf{Q}_c$, con

$$\mathbf{Q}_c = n_w^{-1/2} (\mathbf{I} - \mathbf{M}_* \mathbf{u} \mathbf{u}' \mathbf{W} / (\mathbf{u}' \mathbf{M}_* \mathbf{W} \mathbf{u})) \mathbf{Q},$$

(para \mathbf{Q} ver secciones Correlaciones y Autovalores) que no es necesariamente una matriz de correlación, aunque es semi-definida positiva. Esta matriz sería ponderada con pesos de la variable en la misma forma que \mathbf{R} .

4.11.5. Varianza Considerada Para

La tabla de Varianza Considerada Para, brinda por dimensión y por variable el VAF para coordenadas de centroides, y para variables no-múltiples también para coordenadas del vector.

➤ **Coordenadas de centroide**

$$VAF_{js} = v_j tr(\mathbf{Y}'_{js} \mathbf{D}_j \mathbf{Y}_{js}).$$

➤ **Coordenadas del vector**

$$\text{VAF}_{js} = v_j a_{js}^2, \text{ para } j \notin J.$$

4.11.6. Correlaciones y Autovalores

➤ **Antes de la transformación**

$$\mathbf{R} = n_w^{-1} \mathbf{H}'_c \mathbf{W} \mathbf{H}_c, \text{ con } \mathbf{H}_c \text{ ponderado centrado y normalizado } \mathbf{H}.$$

Para la descomposición de autovalores de \mathbf{R} (para el cálculo de los autovalores), primero, si j es una variable suplementaria son retiradas de \mathbf{R} la fila j y la columna j , y luego r_{ij} se multiplica por $(v_i v_j)^{1/2}$.

Si para una variable, a sus valores perdidos se les aplicase tratamiento pasivo, se van a imputar estos valores perdidos utilizando su moda, esto es independiente de la especificación de imputación pasiva.

➤ **Después de la transformación**

Si son no-múltiples todas las variables en el análisis, y no existen valores perdidos, indicados con tratamiento pasivo, la matriz de correlación sería:

$$\mathbf{R} = n_w^{-1} \mathbf{Q}' \mathbf{W} \mathbf{Q}, \text{ con } \mathbf{q}_j = \mathbf{G}_j \mathbf{y}_j.$$

Los p autovalores primeros de \mathbf{R} igual a $\mathbf{\Lambda}^{1/2}$.

Si en el análisis existen variables nominales múltiples, se calculan p matrices de correlación ($s = 1, \dots, p$):

$$\mathbf{R}_{(s)} = n_w^{-1} \mathbf{Q}'_{(s)} \mathbf{W} \mathbf{Q}_{(s)},$$

con $\mathbf{q}_{(s)j} = \mathbf{G}_j \mathbf{y}_j$ para el caso de variables no-múltiples

y $\mathbf{q}_{(s)j} = n_w^{1/2} \mathbf{G}_j \mathbf{y}_{(j)s} (\mathbf{Y}'_{(j)s} \mathbf{D}_j \mathbf{Y}_{(j)s})^{-1/2}$ para el caso de variables nominales múltiples.

Para los más altos autovalores, mayormente, el primer autovalor de $\mathbf{R}_{(s)}$ es igual a $\lambda_s^{1/2}$. Mayormente los valores más pequeños o bajos de $\Lambda^{1/2}$ son el segundo autovalor o los siguientes autovalores de $\mathbf{R}_{(s)}$.

De existir valores perdidos, que se especificaron con tratamiento pasivo, se imputa el valor de la moda de la variable cuantificada o se imputa la cuantificación de la categoría adicional (se utiliza la moda por defecto de no especificarse) antes de hacer los cálculos de las correlaciones.

No igualan a $\Lambda^{1/2}$ los autovalores de la matriz de correlación. Se calcula la cuantificación de una categoría adicional para el caso de variables nominales múltiples de la siguiente manera

$$Y_{(j)(k_{j+1})} = \left(\sum_{i \in I} w_i \right)^{-1} \sum_{i \in I} w_i x_{is} ,$$

Donde I es índice que va a registrar que objetos son los que tienen valores perdidos.

Para el caso de variables no-múltiples, para la cuantificación de una categoría adicional, previamente se hace el cálculo de $Y_{(j)(k_{j+1})_s}$, y con esto

$$y_{(k_{j+1})j} = n_x^{1/2} \left(\sum_s a_{js}^2 \right)^{-1} \sum_s a_{js} Y_{(j)(k_{j+1})_s} .$$

Para realizar el cálculo de los autovalores de \mathbf{R} , si se encuentra una variable suplementaria j , inicialmente la fila j y la columna j se retirarán \mathbf{R} , y con esto r_{ij} se multiplicará por $(v_i v_j)^{1/2}$.

4.11.7. Las puntuaciones y las saturaciones de Objeto

➤ Normalización

De contar las variables en su totalidad con un nivel de escalamiento del tipo no-múltiple, la normalización separa los p primeros valores singulares de $n_w^{-1/2} \mathbf{W}^{1/2} \mathbf{QV}^{1/2}$

dividido entre m_w sobre puntuaciones de objeto \mathbf{X} y saturaciones \mathbf{A} , donde \mathbf{Q} la matriz de variables cuantificadas, y \mathbf{V} una matriz diagonal con elementos v_j .

El cálculo o descomposición de valores singulares de $n_w^{-1/2} \mathbf{W}^{1/2} \mathbf{Q} \mathbf{V}^{1/2}$ sería

$$SVD(n_w^{-1/2} \mathbf{W}^{1/2} \mathbf{Q} \mathbf{V}^{1/2}) = \mathbf{K} \mathbf{\Phi}^{1/2} \mathbf{L}'.$$

Donde $\mathbf{X} = \mathbf{K}_p$ (el subíndice identificado con p denota las p primeras columnas de \mathbf{K}) y $\mathbf{A} = (\mathbf{L} \mathbf{\Phi}^{1/2})_p$, $\mathbf{X} \mathbf{A}'$ ofrece la mejor estimación o aproximación p -dimensional de $n_w^{-1/2} \mathbf{W}^{1/2} \mathbf{Q} \mathbf{V}^{1/2}$.

Los p primeros valores singulares de $\mathbf{\Phi}_p^{1/2}$ iguales a $\mathbf{\Lambda}^{1/4}$, con $\mathbf{\Lambda}$ como se calcularon durante la última iteración en el paso 4.

Para separar los p primeros valores singulares escribimos

$$(\mathbf{K} \mathbf{\Phi}^{1/2} \mathbf{L}')_p = \mathbf{K}_p \mathbf{\Phi}_p^{a/2} \mathbf{\Phi}_p^{b/2} \mathbf{L}'_p = \mathbf{K}_p \mathbf{\Lambda}^{a/4} \mathbf{\Lambda}^{b/4} \mathbf{L}'_p, (a + b = 1, \text{ ver debajo}).$$

Estando en la etapa de optimización, la variable principal de normalización se utilizará. Resultando después de la convergencia $\mathbf{X} = n_w^{1/2} \mathbf{W}^{-1/2} \mathbf{K}_p$ y $\mathbf{A} = \mathbf{V}^{-1/2} \mathbf{L}_p \mathbf{\Lambda}^{1/4}$.

De ser requerida o necesaria la variable principal de normalización, $\mathbf{X}^n = \mathbf{X}$ y $\mathbf{A}^n = \mathbf{A}$, además

$$\mathbf{X}^n = \mathbf{X} \mathbf{\Lambda}^{a/4}$$

$$\mathbf{A}^n = \mathbf{A} \mathbf{\Lambda}^{1/4(b-1)}$$

donde $a = (1+q)/2$, $b = (1-q)/2$, y q toma algún valor dentro del intervalo cerrado $[-1,1]$, haciendo la excepción en la normalización independiente: no existe q y $a = b = 1$. Con $q = -1$ resulta igual a la variable principal de normalización, $q = 1$ resulta igual al objeto principal de normalización, $q = 0$ resulta igual a la normalización simétrica.

En el caso de que en el análisis existan variables nominales múltiples, existen p matrices $\mathbf{Q}_{(s)}$, $s = 1, \dots, p$. A consecuencia de esto existe uno de los valores singulares de $n_w^{-1/2} \mathbf{W}^{1/2} \mathbf{Q} \mathbf{V}^{1/2}$ igual a $\mathbf{\Lambda}_s^{1/4}$.

De existir una variable que tiene un nivel de escalamiento del tipo nominal múltiple, entonces estará reflejado en los centroides: $\mathbf{Y}_j^n = \mathbf{Y}_j \mathbf{\Lambda}^{1/4(b-1)}$ el factor de normalización.

Una de estas 5 alternativas se debe precisar o detallar para la normalización de los puntajes de objeto y también de las variables. Tener en cuenta que se debe especificar solamente un método para normalizar en cierto análisis.

- **Método Principal por variable:** La alternativa indicada se caracteriza por optimizar la asociación entre las variables. En el espacio de objetos, las coordenadas que tienen las variables serán las saturaciones en los componentes. Se utiliza esto cuando si el interés primordial que se tiene está enfocado en la correlación que puede existir en las variables que se estudian.
- **Método Principal por objeto:** En el caso de esta alternativa de normalización, lo que se busca es la optimización de distancias que existen entre objetos. Se utiliza esto cuando se está interesando principalmente en buscar las semejanzas o disimilitudes que existen entre todos los objetos estudiados.
- **Método Simétrico:** La alternativa es utilizada cuando el principal interés es conocer el nexo que puede existir entre las variables y los objetos usados.
- **Método Independiente:** Es empleado cuando se quiere observar de manera independiente las distancias de objetos y de correlaciones de variables.
- **Método Personalizado.-** Para esta alternativa hay que precisar cualquier valor dentro del intervalo [-1, 1]. Si se toma el valor de 1 resultaría lo mismo que el método Principal por variable, en caso se tome el valor 0 sería lo mismo que el método Simétrico, o en caso de que se tome un valor de -1 sería lo mismo que el método Principal por objeto. La utilidad principal de esta opción es la generación de generar

diagramas a medida de dispersión biespaciales y triespaciales.

4.11.8. Cuantificaciones

Para el caso de variables que presentan un nivel de escalamiento del tipo no-múltiple se observa lo siguiente: cuantificaciones \mathbf{y}_j , vector de coordenadas $\mathbf{y}_j(\mathbf{a}_j^n)$, y coordenadas de centroide: \mathbf{Y}_j con variable principal de normalización, $\mathbf{D}_j^{-1}\mathbf{G}'_j \mathbf{W}\mathbf{X}^n$ con uno de las otras alternativas de normalización.

En el caso de las variables nominales múltiples las cuantificaciones serán las coordenadas de centroide \mathbf{Y}_j^n .

Si sólo es usada una categoría por objetos suplementarios, se van a mostrar solamente coordenadas de centroide para esta categoría, y el cálculo sería el siguiente:

$$y_{(j)r} = n_w^{1/2} n_{jr}^{-1} \sum_{i \in I} x_i^n, \text{ cuando hay variables que tienen nivel de escalamiento del tipo}$$

no-múltiple y

$$y_{(j)r} = n_w^{1/2} n_{jr}^{-1} \sum_{i \in I} x_i \Lambda^{1/4(b-1)}, \text{ cuando hay variables que presentan un nivel de}$$

escalamiento del tipo nominal múltiple.

Aquí tenemos $y_{(j)r}$ es la r -ésima fila de \mathbf{Y}_j , n_{jr} es el número o cantidad de objetos que tienen la categoría r , y el índice I es donde se registra que objetos se encuentran en la categoría r .

➤ Residuos

En el caso de las variables no-múltiples, los Residuos producen un gráfico de la variable cuantificada j ($\mathbf{G}_j \mathbf{y}_j$) vs. la aproximación, $\mathbf{X}\mathbf{a}_j$. Para el caso de variables nominales múltiples, los Residuos muestran un gráfico de $\mathbf{G}_j \mathbf{y}_{(j)s}^n$ versus la aproximación x_s^n .

➤ **Centroides proyectados**

Los centroides proyectados de la variable I sobre la otra variable j , $j \notin J$, serán

$$Y_I a_j (a'_j a_j)^{-1/2}$$

➤ **Factor de escalamiento Biespacial, triespacial, y grafico de saturaciones**

En gráficos que incluyen tanto las puntuaciones de objeto o los centroides, y las saturaciones las puntuaciones de objeto y centroides se van a reajustar usando el factor de escalamiento siguiente:

$$\text{FactordeEscala} = \frac{2 \sum_{s=1}^p \max(a_{1s}^n, \dots, a_{ms}^n)}{\sum_{s=1}^p \left[\min(x_{1s}^n, \dots, x_{ns}^n) + \max(x_{1s}^n, \dots, x_{ns}^n) \right]}$$

5. Aplicación

5.1. Introducción

Podemos mencionar que el nivel o estrato socioeconómico es un tipo de estratificación social que agrupa individuos que comparten ciertas características, como estatus social y recursos económicos, y estas características influyen de manera directa en el estilo de vida que los individuos puedan llevar.

El CATPCA no permitirá hacer la clasificación socioeconómica de los hogares de Lima Metropolitana, de acuerdo a toda la metodología que se ha descrito anteriormente.

5.2. Variables seleccionadas

Siempre cuando se intenta realizar una clasificación socioeconómica o construir una estructura socioeconómica, prácticamente se utilizan las mismas variables que de manera comprobada ayudan a discriminar a los hogares. Entonces se tienen en cuenta por ejemplo las variables que tienen relación con el hogar (cantidad de personas en hogar, hacinamiento, etc.), y jefe del hogar (grado de instrucción, ocupación, seguro médico, etc.), también variables que tienen relación con la vivienda (cantidad de dormitorios, cantidad de baños, material en vivienda, etc.)

A continuación se presentan 22 variables, seleccionadas de un conjunto de variables, que fueron consideradas para realizar el análisis.

Estas variables se dividieron inicialmente en 3 grupos de acuerdo a su naturaleza:

- Jefe del hogar.
- Bienes y servicios.
- Vivienda.

➤ **Jefe del Hogar:**

Variable	Nombre Variable	Tipo de variable
INSTRUC	Grado de instrucción	Cualitativa ordinal
OCUP_AP	Ocupación	Cualitativa ordinal
HOSPITA	Hospitalización	Cualitativa nominal
SEGMEDI	Seguro Médico	Cualitativa nominal

➤ **Bienes y servicios:**

Variable	Nombre Variable	Tipo de variable
TELEFON	Teléfono fijo	Cualitativa nominal
COMPUTA	Computadora o laptop	Cualitativa nominal
INTERNE	Conexión a internet	Cualitativa nominal
TVCABLE	Conexión de TV por cable	Cualitativa nominal
SERVDOM	Servicio doméstico	Cualitativa nominal
AUTOMOV	Automóvil o camioneta de uso particular	Cualitativa nominal
REFRIGE	Refrigeradora en uso	Cualitativa nominal
LAVADOR	Lavadora en uso	Cualitativa nominal
THERMA	Calentador o therma	Cualitativa nominal

➤ **Vivienda:**

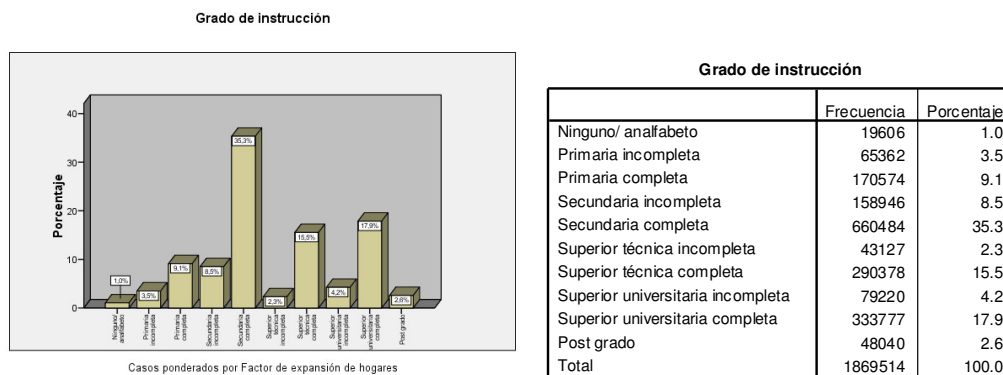
Variable	Nombre Variable	Tipo de variable
NUMPERH	Nro. de personas del hogar	Cuantitativa discreta
DORMITO	Nro. de dormitorios en el hogar	Cuantitativa discreta
HACINAM	Hacinamiento	Cualitativa nominal
BAÑOS	Nro. de baños en el hogar	Cuantitativa discreta
PAREDVI	Material que predomina en paredes de vivienda	Cualitativa nominal
TECHOVI	Material que predomina en techos de vivienda	Cualitativa nominal
PISOVIV	Material que predomina en pisos de vivienda	Cualitativa nominal
AGUAVIV	Medio para abastecerse de agua en vivienda	Cualitativa nominal
SSHVIV	Tipo de servicio higiénico en vivienda	Cualitativa nominal

5.3. Análisis exploratorio de las variables analizadas

Realizaremos, a continuación un análisis exploratorio de las variables mencionadas

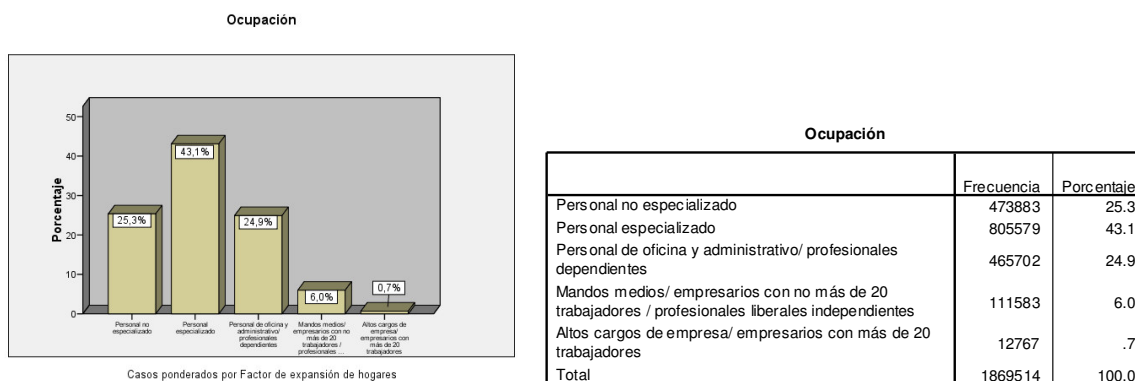
mediante el uso de tablas de frecuencias y gráficos de barras.

➤ Grado de instrucción (INSTRUC)



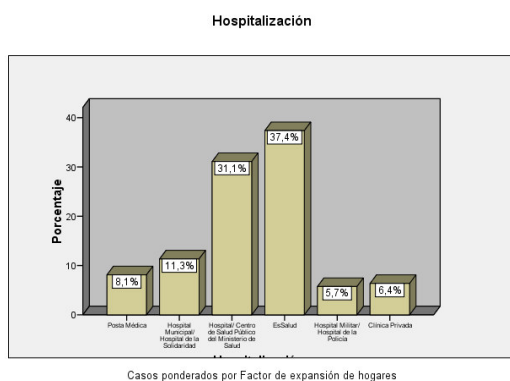
El 35.3% de los jefes de hogar de Lima Metropolitana cuenta con secundaria completa como grado de instrucción (35.3%), y sólo el 1.0% es considerado como analfabeto.

➤ Ocupación (OCUP_AP)



El 43.1% de los jefes de hogar es personal especializado, por otro lado, sólo el 0.7% ocupa altos cargos de empresas o son empresarios con más de 20 trabajadores.

➤ **Hospitalización (HOSPITA)**

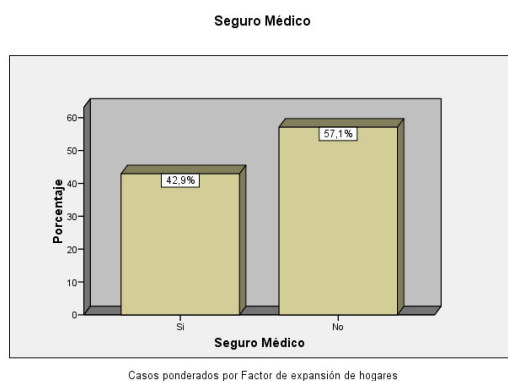


Hospitalización		
	Frecuencia	Porcentaje
Posta Médica	151614	8.1
Hospital Municipal/ Hospital de la Solidaridad	211785	11.3
Hospital/ Centro de Salud Público del Ministerio de Salud	581469	31.1
EsSalud	698501	37.4
Hospital Militar/ Hospital de la Policía	107254	5.7
Clínica Privada	118891	6.4
Total	1869514	100.0

El 37.4% de los hogares en Lima Metropolitana se atienden en EsSalud, el 31.1% en un Hospital o Centro de Salud Público del Ministerio de Salud.

Por otro lado, sólo el 5.7% de los hogares en el Hospital Militar o en el Hospital de la Policía.

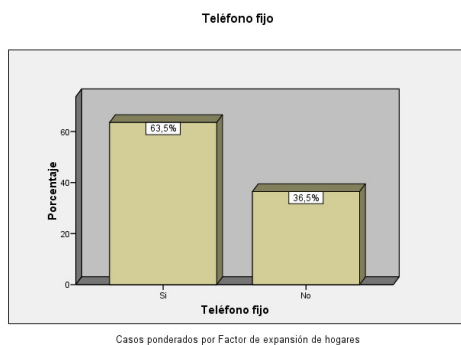
➤ **Seguro Médico (SEGMEDI)**



Seguro Médico		
	Frecuencia	Porcentaje
Si	802302	42.9
No	1067212	57.1
Total	1869514	100.0

Un 42.9% de hogares de Lima Metropolitana cuenta con algún tipo de seguro médico.

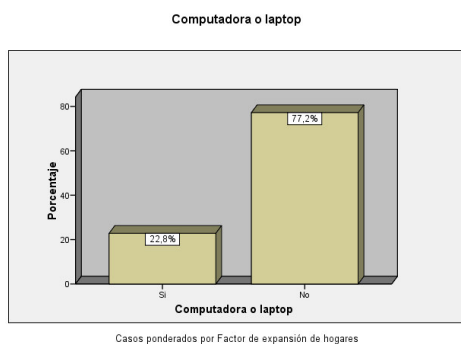
➤ **Teléfono fijo (TELEFON)**



	Frecuencia	Porcentaje
Si	1187968	63.5
No	681546	36.5
Total	1869514	100.0

El 63.5% de hogares cuentan con servicio de telefonía fija.

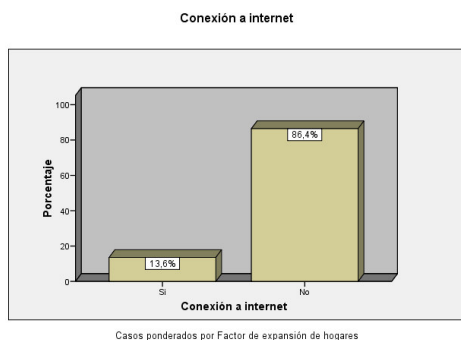
➤ **Computadora o Laptop (COMPUTA)**



	Frecuencia	Porcentaje
Si	426631	22.8
No	1442883	77.2
Total	1869514	100.0

El 22.8% de los hogares en Lima Metropolitana cuentan con computadora o laptop.

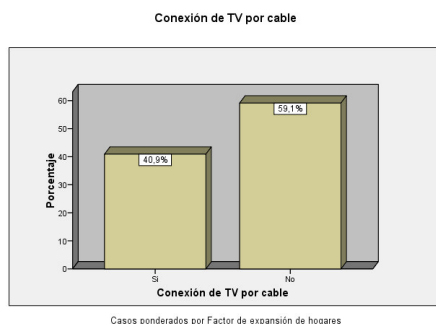
➤ **Conexión a internet (INTERNE)**



	Frecuencia	Porcentaje
Si	253330	13.6
No	1616184	86.4
Total	1869514	100.0

El 13.6% de los hogares cuentan con conexión a internet.

➤ **Conexión de TV por cable (TVCABLE)**

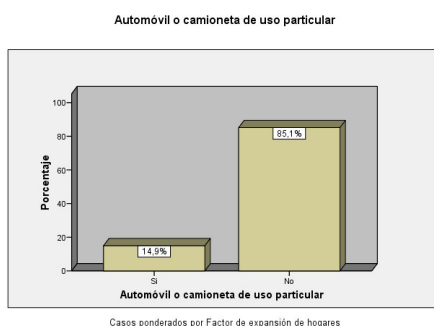


Conexión de TV por cable

	Frecuencia	Porcentaje
Si	764352	40.9
No	1105162	59.1
Total	1869514	100.0

EL 40.9% de los hogares de Lima Metropolitana tiene conexión de TV por cable.

➤ **Automóvil o camioneta de uso particular (AUTOMOV)**

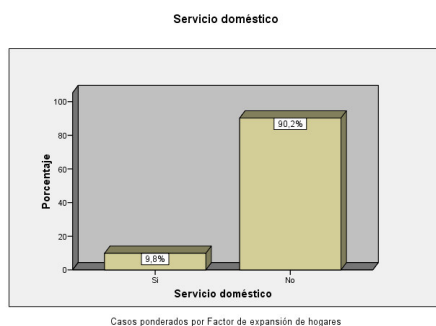


Automóvil o camioneta de uso particular

	Frecuencia	Porcentaje
Si	278880	14.9
No	1590634	85.1
Total	1869514	100.0

EL 14.9% de los hogares tiene automóvil o camioneta de uso particular.

➤ **Servicio doméstico (SERVDOM)**

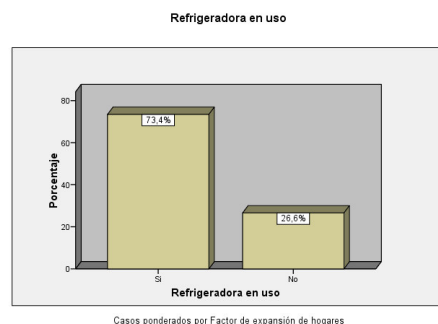


Servicio doméstico

	Frecuencia	Porcentaje
Si	183315	9.8
No	1686199	90.2
Total	1869514	100.0

Solamente el 9.8% de los hogares de Lima Metropolitana tiene servicio doméstico.

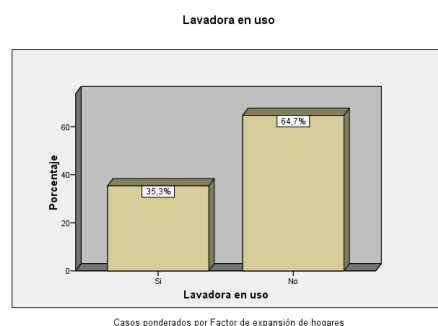
➤ **Refrigeradora en uso (REFRIGE)**



	Frecuencia	Porcentaje
Si	1372558	73.4
No	496956	26.6
Total	1869514	100.0

En Lima Metropolitana, el 73.4% de los hogares tiene su refrigeradora en uso.

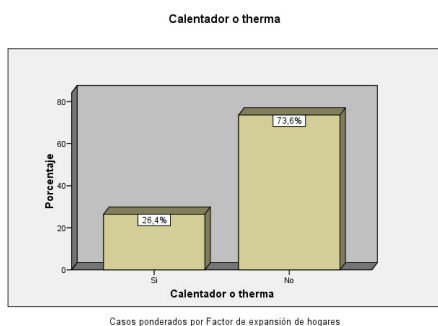
➤ **Lavadora en uso (LAVADOR)**



	Frecuencia	Porcentaje
Si	660867	35.3
No	1208647	64.7
Total	1869514	100.0

El 35.3% de los hogares en Lima Metropolitana cuenta con lavadora en uso.

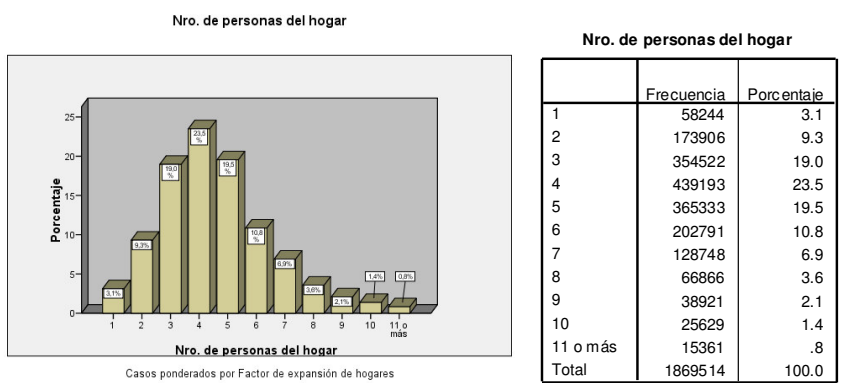
➤ **Calentador o therma (THERMA)**



	Frecuencia	Porcentaje
Si	493937	26.4
No	1375577	73.6
Total	1869514	100.0

El 26.4% de hogares de Lima Metropolitana cuentan con calentador o therma.

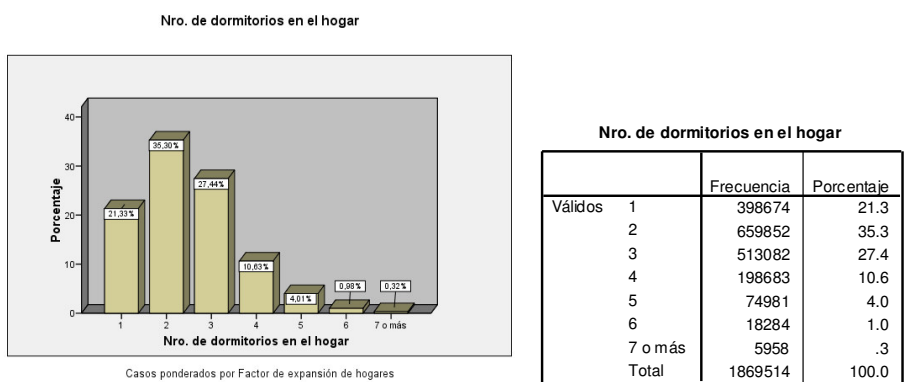
➤ **Nro. de personas del hogar (NUMPERH)**



En su mayoría, los hogares se encuentran conformados por cuatro personas (23.5%); y existe un pequeño porcentaje de hogares conformados por 11 o más personas (0.8%).

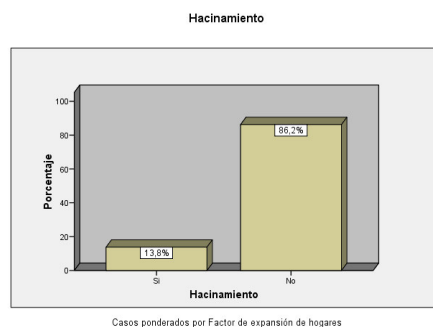
Gráficamente se puede observar que la distribución de esta variable es asimétrica a la derecha.

➤ **Nro. de dormitorios en el hogar (DORMITO)**



El 35.3% de los hogares de Lima Metropolitana cuenta con dos dormitorios. Además se aprecia que el 84.1% de los hogares tiene a lo más tres dormitorios.

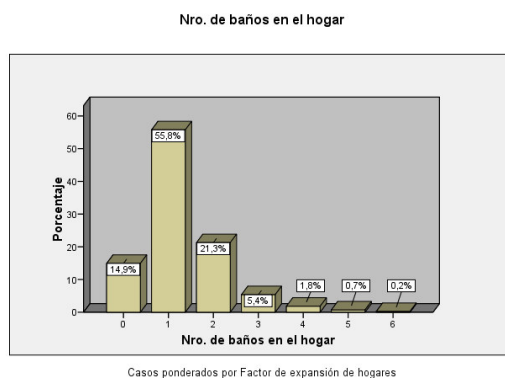
➤ **Hacinamiento (HACINAM)**



Hacinamiento		
	Frecuencia	Porcentaje
Si	257548	13.8
No	1611966	86.2
Total	1869514	100.0

Se consideran como viviendas hacinadas a aquellas en las que habitan más de 3 personas por cuarto, excluyendo la cocina, baño y garaje, es decir, por dormitorio. Entonces, vemos que en Lima Metropolitana, el 13.8% son viviendas hacinadas.

➤ **Nro. de baños en el hogar (BAÑOS)**



Nro. de baños en el hogar		
	Frecuencia	Porcentaje
0	278979	14.9
1	1042393	55.8
2	397595	21.3
3	100020	5.4
4	33800	1.8
5	12683	0.7
6	4044	0.2
Total	1869514	100.0

El 55% de los hogares de Lima Metropolitana cuenta con un baño (55.8%), un 14.9% no cuentan con baño.

➤ **Material que predomina en paredes de vivienda (PAREDVI)**

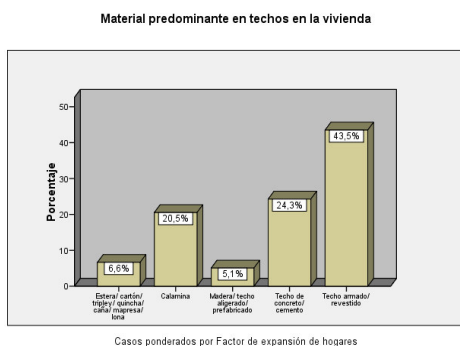


Material predominante en paredes en la vivienda

	Frecuencia	Porcentaje
Estera/ cartón/ triplay/ tablonos	113987	6.1
Adobe/ quincha	77679	4.2
Madera prensada (fibrablock)/ prefabricado	66030	3.5
Ladrillo sin revestir/ bloque de cemento	392079	21.0
Ladrillo revestido/ enchapado	1219739	65.2
Total	1869514	100.0

El 65.2% de los hogares de Lima Metropolitana cuenta con paredes de ladrillo revestido/enchapado.

➤ **Material que predomina en techos de vivienda (TECHOVI)**

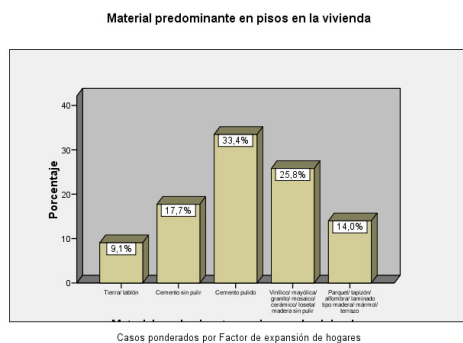


Material predominante en techos en la vivienda

	Frecuencia	Porcentaje
Estera/ cartón/ triplay/ quincha/ caña/ mapresa/ lona	124103	6.6
Calamina	383696	20.5
Madera/ techo aligerado/ prefabricado	94710	5.1
Techo de concreto/ cemento	453826	24.3
Techo armado/ revestido	813179	43.5
Total	1869514	100.0

El 43.5% de los hogares de Lima Metropolitana cuenta con techo armado/ revestido.

➤ **Material que predomina en pisos de vivienda (PISOVIV)**

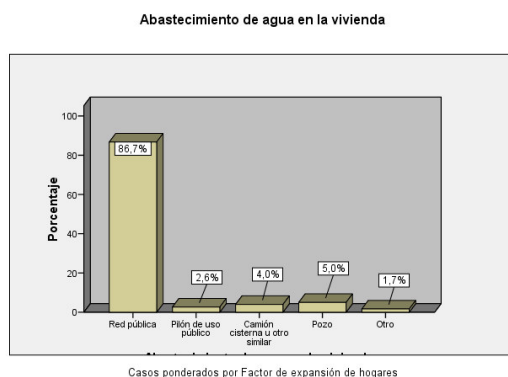


Material predominante en pisos en la vivienda

	Frecuencia	Porcentaje
Tierra/ tablón	169357	9.1
Cemento sin pulir	331307	17.7
Cemento pulido	625336	33.4
Vinílico/ mayólica/ granito/ mosaico/ cerámico/ loseta/ madera sin pulir	481966	25.8
Parquet/ tapicería/ alfombra/ laminado tipo madera/ mármol/ terrazo	261548	14.0
Total	1869514	100.0

El 33.4% de los hogares de Lima Metropolitana cuenta con piso de cemento pulido.

➤ **Medio para abastecerse de agua en vivienda (AGUAVIV)**

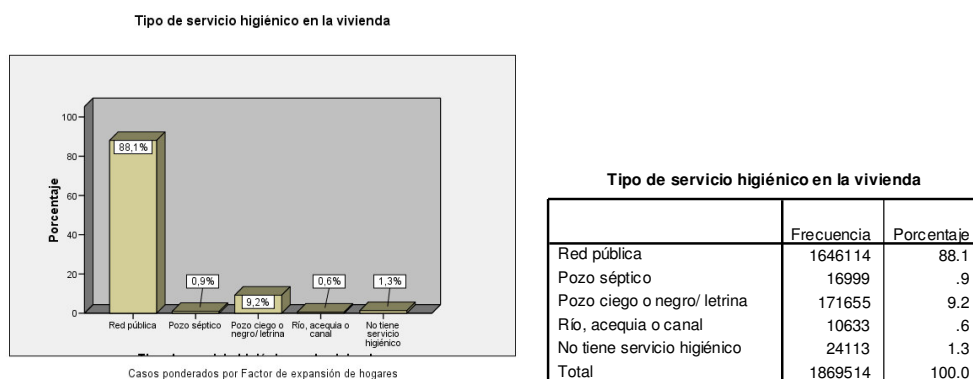


Abastecimiento de agua en la vivienda

	Frecuencia	Porcentaje
Red pública	1621644	86.7
Piñón de uso público	49131	2.6
Camión cisterna u otro similar	74437	4.0
Pozo	93355	5.0
Otro	30947	1.7
Total	1869514	100.0

El 86.7% de los hogares en Lima Metropolitana se abastece de agua mediante la red pública.

➤ **Tipo de servicio higiénico en vivienda (SSHHVIV)**



El 88.1% de los hogares En Lima Metropolitana cuenta con servicio higiénico con conexión a la red pública.

5.4. Metodología de la clasificación socioeconómica

El Análisis de Componentes Principales es el procedimiento adecuado cuando las características observadas son de tipo cuantitativo, pero, para la clasificación de hogares, muchas de las características que se han observado son variables de tipo cualitativo (variables en escala nominal u ordinal), esto no permite la aplicación directa de este procedimiento.

Para dar solución a este tema, se debe transformar variables cualitativas a variables cuantitativas, es decir dar valor a las categorías dentro de cada una de las variables.

Este proceso de asignación de asignar valores puede presentar algunas dificultades, se han identificado dos: la primera, esto es un tema subjetivo, y la valoración podría cambiar de acuerdo a quien la haga; y la segunda dificultad es que la persona que hace la valoración mayormente valora las categorías de una variable sin contemplar la relación que existe con las categorías de otras variables de interés, esto quiere decir que para cada variable cualitativa solo se realiza una valoración unidimensional, perdiéndose información valiosa sobre la relación multivariante de todas las variables en conjunto.

Existe un procedimiento llamado PRINCIPALS (elaborado por Young, Takane y de Leeuw en 1978), que desarrolla para todas las variables el análisis de los componentes

principales, a través de combinaciones entre variables cualitativas y las variables cuantitativas.

Mediante combinaciones en las variables iniciales se generan los componentes principales.

La suma ponderada de todo el conjunto original de variables resultaría ser la primera componente principal, y esta componente se puede usar en vez de variables originales. Es la componente que presenta mayor varianza.

Si se llega a tener el mismo número de componentes principales y cantidad de variables originales, se puede decir que se obtiene un modelo de componentes completo, de no ser así se obtendría un modelo de componentes truncado.

Se elaboró esta clasificación socioeconómica de hogares en función de varias variables como: características de la vivienda, bienes del hogar, ocupación, grado de instrucción y edad del jefe de hogar.

Componentes Principales no lineal, PRINCALS, fue el método utilizado para crear la clasificación o ranking de los hogares. Este método genera una puntuación para cada uno de los hogares en función de las variables que se han incluido, logrando resumir la posición relativa del hogar en la estratificación que se ha realizado.

En SPSS, encontramos el método de Análisis de Componentes Principales no Lineal, dentro del módulo "Reducción de datos", y ahí se encuentra una opción que dice "Escalamiento óptimo".

Para el análisis se han utilizado todas las variables que se han detallado anteriormente, se ha seleccionado la alternativa de normalización simétrica.

Se está considerando la imputación con una categoría adicional si se presentaran valores perdidos. Finalmente con todas estas consideraciones se han guardado las puntuaciones óptimas de los objetos (hogares) generadas para una única dimensión.

5.5. Resultados

La puntuación mínima obtenida es -1.678 mientras que la mayor es 2.139. Teniendo

en cuenta estos números, los vamos a considerar como los límites de un intervalo, y a este intervalo lo dividimos en otros 5 de igual amplitud.

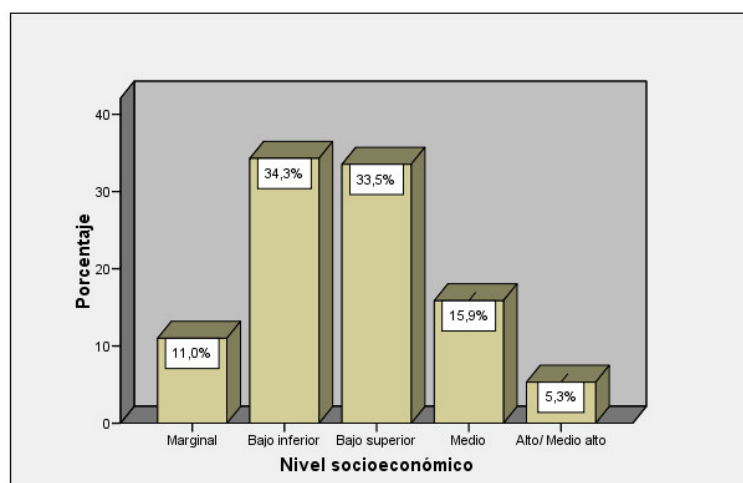
Usaremos los siguientes niveles para la categorización que se está construyendo. Estos niveles son independientes.

CARACTERIZACIÓN DE NIVELES SOCIOECONÓMICOS		
Nivel socioeconómico	NSE	Categoría
Marginal	E	1
Bajo inferior	D	2
Bajo superior	C	3
Medio	B	4
Alto/ Medio alto	A	5

Con estos parámetros, se obtiene una clasificación socioeconómica de 5 niveles:

DISTRIBUCIÓN DE HOGARES SEGÚN PUNTAJE PRINCALS						
Intervalo	Límite inferior	Límite superior	Amplitud	Nº de hogares	%	% acumulado
1	-1.678	-0.915	0.7636	205768	11.0	11.0
2	-0.914	-0.152	0.7636	641121	34.3	45.3
3	-0.151	0.612	0.7636	626913	33.5	78.8
4	0.613	1.375	0.7636	296556	15.9	94.7
5	1.376	2.139	0.7636	99156	5.3	100.0
				1969514	100.0	

Nivel socioeconómico



Podemos decir que las puntuaciones que se han obtenido son consistentes. El índice de Cronbach que se obtiene con estos datos es 0.905. Si el valor del alfa de Cronbach en la tabla es mínimo 0,5, dentro de un intervalo de 0 hasta 1, podemos decir que consideramos como escala, puntaje o índice fuerte o consistente. Cuando el valor está más cerca de 1, es más fiable.

Resumen del modelo

Dimensión	Alfa de Cronbach
1	.905
Total	.905

En el siguiente cuadro vemos el porcentaje del aporte de cada una de las variables consideradas, a la variabilidad total del modelo:

APORTE PORCENTUAL DE CADA VARIABLE A VARIANZA TOTAL DEL MODELO							
Nº	Variables	Descripción	% Aporte	Nº	Variables	Descripción	% Aporte
1	PISOVIV	Material que predomina en pisos de vivienda	8.422	12	TVCABLE	Conexión de TV por cable	4.407
2	BAÑOS	Nro. de baños en el hogar	7.654	13	REFRIGE	Refrigeradora en uso	4.132
3	THERMA	Calentador o therma	6.283	14	OCUP_AP	Ocupación	3.882
4	INSTRUC	Grado de instrucción	6.258	15	AUTOMOV	Automóvil o camioneta de uso particular	3.876
5	LAVADOR	Lavadora en uso	6.054	16	SERVDOM	Servicio doméstico	3.763
6	HOSPITA	Hospitalización	5.652	17	AGUAVIV	Medio para abastecerse de agua en vivienda	3.685
7	COMPUTA	Computadora o laptop	5.531	18	SSHHVIV	Tipo de servicio higiénico en vivienda	3.372
8	TECHOVI	Material que predomina en techos de vivienda	5.286	19	SEGMEDI	Seguro Médico	3.010
9	TELEFON	Teléfono fijo	4.980	20	DORMITO	Nro. de dormitorios en el hogar	2.618
10	PAREDAVI	Material que predomina en paredes de vivienda	4.956	21	HACINAM	Hacinamiento	1.275
11	INTERNE	Conexión a internet	4.805	22	NUMPERH	Nro. de personas del hogar	0.102

Por otro lado, para poder establecer grupos y clasificar en categorías a los hogares, se hizo un análisis de conglomerados. Pero como este método de conglomerados no se utiliza

para los casos con variables dicotómicas o cualitativas, resultado que no era lo correcto.

Otra opción para llevar a cabo este análisis de conglomerados es considerar a las variables transformadas como discriminantes (variables que derivan de las variables originales mediante un escalamiento óptimo) que nos brinda el método CATPCA.

Veamos los resultados de este análisis:

ANÁLISIS DE CONGLOMERADOS		
Conglomerado	Nº de hogares	%
1	172293	9.2
2	244229	13.1
3	216914	11.6
4	655274	35.1
5	580804	31.1
Total	1869514	100.0

A continuación se tiene una tabla cruzada, para que se pueda hacer una comparación entre la clasificación calculada con PRINCALS v/s los grupos obtenidos con el método de Conglomerados.

NSE	CONGLOMERADOS					Total
	1	2	3	4	5	
1	0	0	169573	36195	0	205768
2	0	0	47341	579158	14622	641121
3	1376	97086	0	39921	488530	626913
4	73024	145880	0	0	77652	296556
5	97893	1263	0	0	0	99156
Total	172293	244229	216914	655274	580804	1869514

Se puede ver que hay diferencias en la conformación de los grupos para cada método, esto se debe en mayor proporción a la normalización que ejecuta PRINCALS. Como se ha descrito anteriormente esta normalización en PRINCALS busca que se optimice la distancia entre hogares y la asociación de variables, y por otro lado, en conglomerados se busca optimizar sólo la distancia entre hogares. Adicionalmente para el caso de conglomerados, no existe un orden, sólo se agrupa hogares optimizando sus distancias.

Como se indica, podemos apreciar claras diferencias entre los grupos formados por cada método, eso ocurre principalmente por la normalización de las variables ejecutada por el método PRINCALS, ya que esta normalización se hace para optimizar la distancia que existe entre los hogares al mismo tiempo que también se optimiza la asociación entre variables, mientras que en el otro caso, en el análisis de conglomerados de hogares se optimiza solamente la distancia entre los hogares. Además, el análisis de conglomerados la clasificación no es ordenada en su clasificación, lo que hace simplemente es agrupar hogares maximizando (optimizando) las distancias entre ellos.

5.6. Perfiles de los hogares por nivel socioeconómico

Aquí se muestra la distribución de los hogares utilizando PRINCALS. Se hace una descripción separando las variables en tres grupos, tal como se hizo en el análisis exploratorio de las variables, según:

- Jefe del hogar.
- Bienes y servicios.
- Vivienda.

➤ **Jefe del Hogar:**

	NIVEL SOCIOECONÓMICO				
	NSE A ALTO/ ALTO MEDIO	NSE B MEDIO	NSE C BAJO SUPERIOR	NSE D BAJO INFERIOR	NSE E MARGINAL
GRADO DE INSTRUCCIÓN	<ul style="list-style-type: none"> ➤ Universitaria completa (54.0%) ➤ Postgrado (33.3%) 	<ul style="list-style-type: none"> ➤ Sup. Técnica completa (21.1%) ➤ Universitaria completa (51.2%) 	<ul style="list-style-type: none"> ➤ Secundaria completa (31.6%) ➤ Sup. técnica completa (24.3%) 	<ul style="list-style-type: none"> ➤ Secundaria incompleta (12.2%) ➤ Secundaria completa (51.5%) 	<ul style="list-style-type: none"> ➤ Secundaria incompleta (22.0%) ➤ Secundaria completa (45.6%)
OCUPACIÓN	<ul style="list-style-type: none"> ➤ Personal de oficina y administrativo/ profesionales dependientes (29.6%) ➤ Mandos medios/ empresarios con no más de 20 trabajadores / profesionales liberales independientes (36.5%) 	<ul style="list-style-type: none"> ➤ Personal especializado (43.4%) ➤ Personal de oficina y administrativo/ profesionales dependientes (15.8%) 	<ul style="list-style-type: none"> ➤ Personal especializado (51.7%) ➤ Personal de oficina y administrativo/ profesionales dependientes (26.0%) 	<ul style="list-style-type: none"> ➤ Personal no especializado (40.0%) ➤ Personal especializado (39.9%) 	<ul style="list-style-type: none"> ➤ Personal no especializado (48.6%) ➤ Personal especializado (35.5%)
HOSPITALIZACIÓN	<ul style="list-style-type: none"> ➤ EsSalud (49.2%) ➤ Clínica Privada (41.9%) 	<ul style="list-style-type: none"> ➤ EsSalud (58.3%) ➤ Clínica Privada (19.6%) 	<ul style="list-style-type: none"> ➤ Hospital/ Centro de Salud Público del Ministerio de Salud (25.0%) ➤ EsSalud (47.7%) 	<ul style="list-style-type: none"> ➤ Hospital/ Centro de Salud Público del Ministerio de Salud (46.6%) ➤ EsSalud (25.0%) 	<ul style="list-style-type: none"> ➤ Posta Médica (29.7%) ➤ Hospital/ Centro de Salud Público del Ministerio de Salud (49.5%)
SEGURO MÉDICO	90.9%	72.2%	51.8%	23.7%	10.4%

➤ **Bienes y servicios:**

	NIVEL SOCIOECONÓMICO				
	NSE A ALTO/ ALTO MEDIO	NSE B MEDIO	NSE C BAJO SUPERIOR	NSE D BAJO INFERIOR	NSE E MARGINAL
TELÉFONO FIJO	100.0%	97.7%	82.6%	40.6%	10.2%
COMPUTADORA O LAPTOP	96.6%	63.4%	20.6%	2.2%	0.0%
CONEXIÓN A INTERNET	89.4%	44.3%	5.3%	0.0%	0.0%
CONEXIÓN DE TV POR CABLE	96.4%	81.3%	47.8%	17.8%	6.6%
AUTOMÓVIL O CAMIONETA	95.0%	33.8%	10.6%	2.8%	0.0%
SERVICIO DOMÉSTICO	80.5%	25.6%	3.9%	0.5%	0.0%
COCINA	100.0%	100.0%	99.8%	98.8%	99.3%
ENERGÍA O COMBUSTIBLE PARA COCINAR	<ul style="list-style-type: none"> ➤ Gas (87.2%) ➤ Electricidad (12.8%) 	<ul style="list-style-type: none"> ➤ Gas (92.9%) ➤ Electricidad (7.1%) 	<ul style="list-style-type: none"> ➤ Gas (96.7%) 	<ul style="list-style-type: none"> ➤ Gas (83.8%) 	<ul style="list-style-type: none"> ➤ Gas (75.6%)
REFRIGERADORA EN USO	100.0%	99.0%	91.8%	54.9%	25.5%
LAVADORA EN USO	95.3%	83.4%	41.6%	8.8%	0.7%
CALENTADOR O THERMA	98.6%	75.9%	25.6%	1.7%	0.0%

➤ **Vivienda:**

	NIVEL SOCIOECONÓMICO				
	NSE A ALTO/ ALTO MEDIO	NSE B MEDIO	NSE C BAJO SUPERIOR	NSE D BAJO INFERIOR	NSE E MARGINAL
NRO. PROMEDIO DE PERSONAS EN HOGAR	4.5	4.3	4.5	4.6	5.0
NRO. PROMEDIO DE DORMITORIOS	4	3	3	2	2
HACINAMIENTO	0.0%	1.4%	6.6%	22.9%	31.9%
NRO. PROMEDIO DE BAÑOS	3	2	1	1	➤ 0 (85.3%) ➤ 1 (14.7%)
TIPO DE VIVIENDA	➤ Casa independiente (80.9%)	➤ Casa independiente (77.8%)	➤ Casa independiente (84.7%)	➤ Casa independiente (88.3%)	➤ Casa independiente (82.8%)
MATERIAL QUE PREDOMINA EN PAREDES	➤ Ladrillo revestido (92.8%)	➤ Ladrillo revestido (89.9%)	➤ Ladrillo revestido (83.9%)	➤ Ladrillo sin revestir (34.7%) ➤ Ladrillo revestido (47.0%)	➤ Tripley (26.4%) ➤ Ladrillo sin revestir (33.8%)
MATERIAL QUE PREDOMINA EN TECHOS	➤ Techo de concreto/ cemento (37.1%) ➤ Techo armado/ revestido (62.9%)	➤ Techo de concreto/ cemento (26.8%) ➤ Techo armado/ revestido (69.7%)	➤ Techo de concreto/ cemento (24.8%) ➤ Techo armado/ revestido (62.1%)	➤ Calamina (35.9%) ➤ Techo de concreto/ cemento (25.4%) ➤ Techo armado revestido (23.9%)	➤ Estera (22.3%) ➤ Calamina (55.8%)
MATERIAL QUE PREDOMINA EN PISOS	➤ Parquet (77.6%)	➤ Mayólica (19.8%) ➤ Loseta (15.4%) ➤ Parquet (40.2%)	➤ Cemento pulido (43.6%) ➤ Mayólica (12.5%) ➤ Loseta (19.9%)	➤ Cemento sin pulir (34.2%) ➤ Cemento pulido (44.0%)	➤ Tierra (46.0%) ➤ Cemento sin pulir (31.7%)
TIPO DE ALUMBRADO	➤ Electricidad (100.0%)	➤ Electricidad (100.0%)	➤ Electricidad (99.8%)	➤ Electricidad (99.3%)	➤ Electricidad (95.2%)
MEDIO DE ABASTECIMIENTO DE AGUA	➤ Red pública (100.0%)	➤ Red pública (98.4%)	➤ Red pública (98.5%)	➤ Red pública (89.4%)	➤ Red pública (19.3%) ➤ Camión cisterna (26.7%) ➤ Pozo (37.1%)
TIPO DE SERVICIO HIGIÉNICO	➤ Red pública (100.0%)	➤ Red pública (100.0%)	➤ Red pública (99.3%)	➤ Red pública (90.5%)	➤ Red pública (23.1%) ➤ Pozo ciego o negro/ letrina (56.1%)

Ahora haremos una descripción de las principales características según nivel socioeconómico:

➤ **NSE A: NIVEL ALTO/ MEDIO ALTO**

Características principales NSE A:

- Representan al 5.3% de los hogares en Lima Metropolitana.
- Se tiene 4 personas en promedio por hogar (4.5), adicionalmente se tiene 4 dormitorios

y 3 baños en una vivienda.

- En este nivel, la vivienda está hecha principalmente de material noble. Se abastecen de agua a través de la red pública (100.0%). Los pisos en su mayoría tienen parquet (77.6%). Los techos y las paredes son de concreto.

Bienes y servicios NSE A:

- La mayoría de los hogares cuenta con teléfono fijo (100.0%), computadora (96.6%), automóvil (95.0%), refrigeradora (100.0%), lavadora (95.3%), calentador o therma (98.6%).
- Un alto porcentaje cuenta con servicio doméstico (80.5%). La mayoría cuenta con conexión a internet (89.4%), TV por cable (96.4%).

Jefe de hogar NSE A:

- Un poco más de la mitad tiene sólo educación universitaria completa (54.0%) y un poco más del 30% ha realizado estudios de postgrado (33.3%).
- La mayoría ocupa cargos de mandos medios.
- El 90% cuenta con seguro médico en caso de hospitalización acuden a EsSalud o clínica privada.

➤ **NSE B: NIVEL MEDIO**

Características principales NSE B:

- Representan el 15.9% de los hogares de Lima Metropolitana.
- Se tiene 4 personas en promedio por hogar (4.3), adicionalmente se tiene 3 dormitorios y 2 baños en una vivienda.
- En este nivel, la vivienda está hecha principalmente de material noble. Se abastecen mayormente de agua a través de la red pública (98.4%). Existe mayor diversidad en el material usado en los pisos: se tiene parquet (40.2%), mayólica (19.8%) y loseta

(15.4%). Los techos y las paredes son de concreto.

Bienes y servicios NSE B:

- Aproximadamente una tercera parte cuenta con automóvil (33.8%), y la cuarta parte tiene servicio doméstico remunerado (25.6%). El 63.4% cuenta con computadora y el 44.3% tiene conexión a internet.
- Todos cuentan con cocina (100.0%), y la mayoría tiene refrigeradora (99.0%) y también teléfono fijo (82.6%). Un gran porcentaje cuenta con TV por cable (96.4%), lavadora (83.4%) y calentador o therma (75.9%).

Jefe de hogar NSE B:

- La mayoría es personal de oficina y administrativo, o personal especializado.
- Poco más de la mitad tiene educación universitaria completa.
- Un 72.2% tiene seguro médico. En caso de hospitalización, el 58.3% se atiende en EsSalud.

➤ **NSE C: NIVEL BAJO SUPERIOR**

Características principales NSE C:

- Son el 33.5% de los hogares en Lima Metropolitana.
- Se tiene 4 personas en promedio por hogar (4.5), adicionalmente se tiene 3 dormitorios y 1 baño en una vivienda.
- En este nivel, la vivienda está hecha principalmente de material noble. Se abastecen mayormente de agua a través de la red pública (98.5%). En los pisos hay presencia de cemento pulido (43.6%), loseta (19.9%) y mayólica (12.5%). Los techos y las paredes son de concreto.

Bienes y servicios NSE C:

- La mayoría tiene: cocina (99.8%), refrigeradora (91.8%) y teléfono fijo (82.6%).
- El 47.8% cuenta con conexión de televisión por cable, el 41.6% tiene lavadora de ropa (menos de la mitad).
- El 20.6% posee computadora, el 10.6% posee automóvil, el 5.3% tiene internet y el 3.9% servicio doméstico.

Jefe de hogar NSE C:

- Cuenta con educación secundaria completa el 31.6%, y un 24.3% tiene estudios técnicos completos.
- El 51.7% es personal calificado, un 26.0% es personal de oficina y administrativo, profesionales dependientes.
- Un poco más de la mitad tiene seguro (51.8%). Si requieren hospitalización irían a hospitales, centros de salud públicos (46.6%) o EsSalud (47.7%).

➤ **NSE D: NIVEL BAJO INFERIOR**

Características principales NSE D:

- Son el 34.3% de los hogares de Lima Metropolitana.
- Se tiene 5 personas en promedio por hogar (4.6), adicionalmente se tiene 2 dormitorios y 1 baño en una vivienda.
- Se abastecen mayormente de agua a través de la red pública (89.4%). En los pisos hay presencia de cemento pulido (44.0%) y cemento sin pulir (34.2%), loseta (19.9%) y mayólica (12.5%). En cuanto a las paredes, presentan con ladrillo revestido un 47.0%, y un 34.7% son de ladrillo sin revestir. En los techos hay presencia de calamina (35.9%), techo de concreto (25.4%) y techo armado revestido (23.9%).

Bienes y servicios NSE D:

- La mayoría tiene cocina (98.8%).
- El 54.9% posee refrigeradora, y solo el 8.8% tiene lavadora.
- El 40.6% tiene teléfono fijo y sólo el 2.8% tiene automóvil.

Jefe de hogar NSE D:

- El 51.5% cuenta con secundaria completa.
- El 40.0% es personal no calificado y el 39.9% es personal calificado.
- Sólo la cuarta parte cuenta con seguro (25.0%). Para hospitalización acudirían a un hospital o centro de salud público (46.6%), y en menor proporción a EsSalud (25.0%)

➤ **NSE E: NIVEL MARGINAL**

Características principales NSE E:

- Son el 11.0% de los hogares de Lima Metropolitana.
- Se tiene 5 personas en promedio por hogar (5.0), adicionalmente se tiene 2 dormitorios, y en cuanto a los baños en su mayoría no tienen (85.3%) o tienen a lo más 1 (14.7%).
- Sobre el abastecimiento de agua se tiene la presencia de fuentes diversas como: Pozo (37.1%), Camión cisterna (26.7%), y una minoría lo hace a través de la red pública (19.3%). En los pisos predomina la tierra (46.0%) y cemento sin pulir (31.7%). En las paredes se tiene mayor presencia de ladrillo son revestir (33.8%) y tripley (26.4%). En el caso de los techos de las viviendas mayormente son de calamina (55.8%) y esteras (22.3%).

Bienes y servicios NSE E:

- EL 99.3% posee cocina.
- La cuarta parte tiene refrigeradora (25.5%) y la lavadora es inexistente (sólo 0.7%).

- Sólo un 10.2% posee teléfono fijo, y no se registra presencia de automóvil (0.0%)

Jefe de hogar NSE E:

- El 45.6% tiene secundaria completa, y el 22.0% secundaria incompleta.
- El 48.6% es personal no especializado, y un 35.5% es personal especializado.
- Sólo el 10.4% tiene seguro médico. Para hospitalización acudirían a un hospital o centro de salud público (49.7%), o también irían a una posta (29.7%).

5.7. Comparación con resultados de empresas de investigación de mercados

Ahora mostramos una comparación con los resultados obtenidos como estructura socioeconómica para los hogares de Lima Metropolitana para el año 2007 que han obtenido unas empresas en el Perú.

NIVEL SOCIOECONÓMICO	NSE	NSE CATPCA	NSE APEIM	NSE APOYO
Alto/ Medio alto	A	5.3	5.3	5.4
Medio	B	15.9	15.7	18.1
Bajo superior	C	33.5	28.7	35.1
Bajo inferior	D	34.3	33.9	28.7
Marginal	E	11.0	16.4	12.7

Donde:

- APEIM: Asociación Peruana de Empresas de Investigación de Mercados
- APOYO: Ipsos APOYO Opinión y Mercado S. A.

6. Bibliografía

- B. Visauta Vinacua: Análisis estadístico con SPSS para Windows. Análisis Multivariante vol.2 (1998). Editorial McGraw-Hill / Interamericana de España.
- De Leeuw, J., Young, F. W., and Takane, Y. 1976. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika*, 31: 33–42.
- Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester: John Wiley and Sons Ltd.
- Kruskal, J. B. 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29: 115–129.
- Pérez López, César (2005). *Métodos Estadísticos Avanzados con SPSS*
- Meulman, Jacqueline J. & Heiser, Willem J. (2005). *SPSS Categories 14.0*.
- Ministerio de Salud de Chile (2007). *Estudio Nacional sobre satisfacción y gasto en Salud*, Santiago.

- CATPCA. SPSS 14.0 Statistical Algorithms.