

UNIVERSIDAD DE LOS ANDES
FACULTAD DE CIENCIAS ECONOMICAS Y SOCIALES
INSTITUTO DE ESTADÍSTICA APLICADA Y COMPUTACION

ANÁLISIS ESTADÍSTICO DE PRESUPUESTOS FAMILIARES EN VENEZUELA

Diseño e implementación de una base de datos y análisis estadístico de los datos
suministrados por la II encuesta nacional de presupuestos familiares, con base en el año

1997

Autor: Víctor Ernesto Márquez Pérez

Tutor: Giampaolo Orlandoni Merli

TRABAJO DE GRADO

Presentado ante la ilustre Universidad de los Andes
como requisito final para optar al grado Magíster en Estadística Aplicada

Mérida, Julio del 2004

ÍNDICE

Tabla de Contenido

Introducción.....	1
Capítulo I. Diseño de la base de datos.....	5
Capítulo II. Metodología Estadística.....	10
2.1. Análisis general.....	11
2.1.1. Ajuste en R^p	11
2.1.2. Ajuste en R^n	17
2.1.3. Relación entre los ajustes.....	18
2.1.4. Reconstrucción de los datos.....	20
2.1.5. Contribuciones.....	21
2.2. Análisis de componentes principales.....	25
2.2.1. Enfoque geométrica.....	26
2.2.2. Variabilidad captada por los componentes.....	28
2.2.3. Interpretación de los resultados.....	30
2.2.4. Contribución del individuo a la componente.....	31
2.2.5. Contribución de la componente al individuo (calidad de representación).....	32
2.2.6. Interpretación de la nube de variables.....	33
2.2.7. Contribución de la variable a la componente.....	34
2.2.8. Contribución de la componente a la variable.....	35
2.3. Análisis de correspondencia simple.....	35
2.3.1. Análisis general ponderado por una métrica cualquiera.....	36
2.3.2. Definiciones y notaciones.....	40
2.3.3. Nube de puntos.....	41
2.3.3.1. En R^p	41
2.3.3.2. En R^n	41
2.3.4. Inercia de la nube de puntos.....	42
2.3.4.1. En R^p	42
2.3.4.2. En R^n	43
2.3.5. Análisis en R^p	43
2.3.5.1. Coordenadas de proyección de los perfiles fila.....	44
2.3.6. Análisis en R^n	45
2.4. Modelos Loglineales.....	46
2.4.1. Modelos saturados para dos variables.....	46
2.4.2. Modelo saturado general.....	47

2.4.3.Otros tipos de modelos.....	48
2.4.4.Cálculo de los efectos.....	48
2.4.5.Bondad de ajuste prueba de los k efectos y prueba de asociación parcial.....	49
Capítulo III. Análisis descriptivo univariante.....	52
Capítulo IV. Análisis multivariante.....	65
4.1. Análisis de correspondencias.....	65
4.1.1.Análisis de correspondencias: nivel educativo vs tipo de vivienda.....	66
4.1.2.Análisis de correspondencias: apariciencia de vivienda vs tipo de vivienda.....	70
4.1.3.Análisis de correspondencias: zona geográfica vs tipo de vivienda.....	73
4.1.4.Análisis de correspondencias: seguro de previsión social vs nivel educativo.....	75
4.1.5.Análisis de correspondencia múltiple: nivel educativo, tipo de vivienda, zona geográfica, seguro de previsión social y número de empleos.....	76
4.2. Modelos loglineales.....	78
4.3 Análisis de componentes principales.....	83
4.3.1.Análisis de componentes principales para los dos grupos de variables de gastos, categorizado por entidad geográfica.....	85
4.3.1.1.Variables de gastos mensuales.....	85
4.3.1.2.Variables de gastos trimestrales o anuales.....	91
4.3.2. Análisis de componentes principales para los dos grupos de variables de gastos, categorizado por nivel educativo.....	96
4.3.2.1.Variables de gastos mensuales.....	96
4.3.2.2.Variables de gastos trimestrales o anuales.....	102
4.3.3. Análisis de componentes principales para los dos grupos de variables de gastos, categorizado por tipo de vivienda.....	106
4.3.3.1.Variables de gastos mensuales.....	106
4.3.3.2.Variables de gastos trimestrales o anuales.....	112
4.3.4. Análisis de componentes principales para el grupo de variables de gastos mensuales, categorizado por apariciencia de la vivienda.....	114
Conclusiones y recomendaciones.....	118
Bibliografía.....	124
Apéndice A.....	125
Apéndice B.....	136
Apéndice C.....	153
Apéndice D.....	154

Tabla de Figuras

Figura N° 1: Diagrama referencial de la base de datos para para la II Encuesta de Presupuestos Familiares en Venezuela.....	7
Figura N° 2: Proyección en R^3	13
Figura N° 3: Tabla de coordenadas y contribuciones de la variable al factor.....	22
Figura N° 4: Tabla de coordenadas y contribuciones del factor a la variable.....	23
Figura N° 5: Tabla de coordenadas y contribuciones del individuo al factor.....	24
Figura N° 6: Tabla de coordenadas y contribuciones del factor al individuo.....	25
Figura N° 7: Nube de puntos fila de la matriz X con traslación de origen.....	27
Figura N° 8: Representación gráfica de la calidad de representación.....	32
Figura N° 9: Representación gráfica de la correlación entre dos variables.....	34
Figura N° 10: Proyección en R^3	37

Tabla de Cuadros

Cuadro N° 1: Distribución de frecuencias de Uso de la Vivienda.....	53
Cuadro N° 2: Distribución de frecuencias de Número de hogares por vivienda.....	54
Cuadro N° 3: Estadísticas básicas de miembro por hogar.....	54
Cuadro N° 4: Estadísticas básicas de miembros por dormitorio.....	54
Cuadro N° 5: Distribución de frecuencias de miembros por hogar.....	55
Cuadro N° 6: Distribución de frecuencias de tipo de vivienda.....	56
Cuadro N° 7: Distribución de frecuencias de material de paredes.....	56
Cuadro N° 8: Distribución de frecuencias de material del piso.....	57
Cuadro N° 9: Distribución de frecuencias de material del techo.....	58
Cuadro N° 10: Distribución de frecuencias de estado de la vivienda.....	58
Cuadro N° 11: Distribución de frecuencias de apariencia de la vivienda.....	59
Cuadro N° 12: Distribución de frecuencias de	

analfabetismo	60
Cuadro N° 13: Distribución de frecuencias de nivel educativo	60
Cuadro N° 14: Distribución de frecuencias de sistema de previsión social	61
Cuadro N° 15: Distribución de frecuencias de medio de transporte	62
Cuadro N° 16: Distribución de frecuencias para la posesión de trabajo	63
Cuadro N° 17: Distribución de frecuencias para la búsqueda de trabajo	63
Cuadro N° 18: Distribución de frecuencias de trabajo adicional	64
Cuadro N° 19: Distribución de frecuencias de número de trabajo	64
Cuadro N° 20: Porcentaje de inercia captada por los factores en análisis de correspondencia nivel educativo vs tipo de vivienda	66
Cuadro N° 21: Contribución y calidad de representación para las modalidades de la variable tipo de vivienda en ACS: nivel educativo vs tipo de vivienda	67
Cuadro N° 22: Contribuciones y calidad de representación para las modalidades de la variable nivel educativo en ACS: nivel educativo vs tipo de vivienda	68
Cuadro N° 23: Test. Efectos de orden k igual a cero	79
Cuadro N° 24: Test de asociación parcial	80
Cuadro N° 25: Estimación de los efectos principales de las categorías nivel educativo	81
Cuadro N° 26: Estimación de los efectos principales de las categorías tipo de vivienda	81
Cuadro N° 27: Estimación de los efectos principales de las categorías zonas geográficas	82
Cuadro N° 28: Estimación de los efectos principales de las sistemas de previsión social	82
Cuadro N° 29: Grupo de variables de gastos mensuales	83
Cuadro N° 30: Grupo de variables de gastos trimestrales o anuales	84
Cuadro N° 31: Código de las entidades geográficas	85
Cuadro N° 32: Autovalores y % de inercia	87
Cuadro N° 33: Correlación variable factor, coordenadas de los autovectores	88
Cuadro N° 34: Contribuciones y calidad de representación de los individuos	89
Cuadro N° 35: Autovalores y % de inercia	92

Cuadro N° 36: Correlación variable factor, coordenadas de los autovectores.....	93
Cuadro N° 37: Contribuciones y calidad de representación de los individuos.....	93
Cuadro N° 38: Código del nivel educativo.....	97
Cuadro N° 39: Autovalores y % de inercia.....	98
Cuadro N° 40: Correlación variable factor, coordenadas de los autovectores.....	99
Cuadro N° 41: Contribuciones y calidad de representación de los individuos.....	100
Cuadro N° 42: Autovalores y % de inercia.....	104
Cuadro N° 43: Correlación variable factor, coordenadas de los autovectores.....	104
Cuadro N° 44: Código de tipo de vivienda.....	107
Cuadro N° 45: Autovalores y % de inercia.....	108
Cuadro N° 46: Correlación variable factor, coordenadas de los autovectores.....	108
Cuadro N° 47: Contribuciones y calidad de representación de los individuos.....	110
Cuadro N° 48: Código apariencia de vivienda.....	114
Cuadro N° 49: Autovalores y % de inercia.....	115
Cuadro N° 50: Correlación variable factor, coordenadas de los autovectores.....	115
Cuadro N° 51: Contribuciones y calidad de representación de los individuos.....	116

Tabla de gráficos

Gráfico N° 1: Plano factorial: nivel educativo vs tipo de vivienda.....	69
Gráfico N° 2: Plano factorial: apariencia de vivienda vs tipo de vivienda.....	71
Gráfico N° 3: Plano factorial: nivel educativo vs tipo de vivienda excluyendo las modalidades precaria, vivienda rural y rancho campesino.....	72
Gráfico N° 4: Plano factorial: zona geográfica vs tipo de vivienda.....	74
Gráfico N° 5: Plano factorial: tipo de vivienda vs zona geográfica excluyendo la modalidad otros.....	74
Gráfico N° 6: Plano factorial: sistema de previsión social vs nivel educativo.....	76
Gráfico N° 7: Plano factorial: nivel educativo, tipo de vivienda, zona geográfica sistema de previsión social y números de empleos.....	77

Gráfico N° 8: Proyección de los individuos en el plano factorial categorizado por entidad geográfica.....	90
Gráfico N° 9: Proyección de las variables en el plano factorial (variables de gastos trimestrales o anuales).....	95
Gráfico N° 10: Proyección de los individuos en el plano factorial categorizado por entidad geográfica.....	96
Gráfico N° 11: Proyección de las variables en el plano factorial (variables de gastos mensuales).....	100
Gráfico N° 12: Proyección de los individuos en el plano factorial categorizado por nivel educativo.....	101
Gráfico N° 13: Proyección de las variables en el plano factorial (variables de gastos trimestrales o anuales).....	105
Gráfico N° 14: Proyección de los individuos en el plano factorial categorizado por nivel educativo.....	106
Gráfico N° 15: Proyección de las variables en el plano factorial (variables de gastos mensuales).....	110
Gráfico N° 16: Proyección de los individuos en el plano factorial categorizado por tipo de vivienda.....	111
Gráfico N° 17: Proyección de los individuos en el plano factorial categorizado por tipo de vivienda.....	113
Gráfico N° 18: Proyección de los individuos en el plano factorial categorizado por apariencia de vivienda.....	117

RESUMEN

ANÁLISIS ESTADÍSTICO DE PRESUPUESTOS FAMILIARES EN VENEZUELA

Diseño e implementación de una base de datos y análisis estadístico de los datos suministrados por la II Encuesta Nacional de Presupuestos Familiares, con base en el año

1997

Por

Víctor Ernesto Márquez Pérez

Unas de las necesidades que impulsó la realización de este trabajo fue el rescatar los datos de la II Encuesta de Presupuestos Familiares y almacenarlos en un dispositivo adecuado, para luego aprovechar al máximo esta información. Por ende, se diseñó e implementó una base de datos, con tecnología referencial, que permitiera de forma eficiente realizar cualquier tipo de consulta a los datos, así como también que sirviera de materia prima para análisis estadísticos de mayor envergadura. Es decir, se trataba de retribuir en investigación el alto costo económico que significó la realización de esta encuesta y evitar a toda costa la pérdida de esta valiosa información.

Por otro lado, aprovechando las bondades de la base de datos, se realizó una serie de análisis estadísticos como: análisis de correspondencias simple y múltiples, los cuales permitieron describir de forma clara la relación existente entre distintas variables socioeconómicas de carácter cualitativo. Luego, con el objeto de explicar sucintamente las relaciones encontradas en el análisis de correspondencia, se construyó un modelo loglineal

el cual se considera es el modelo que subyace detrás de las tablas de contingencias utilizadas para la elaboración de los análisis de correspondencias.

Por último con la idea de relacionar las variables cualitativas antes mencionadas con variables cuantitativas de gastos del hogar, se realizó una serie de análisis de componentes principales sobre estas últimas variables, categorizando a los hogares con las variables socioeconómicas del tipo categóricas o cualitativas.

INTRODUCCIÓN

En el marco del programa de actualización de la estimación macroeconómica (PRACEM), se realiza la II encuesta nacional de presupuestos familiares, con base en el año 1997, a objeto de revisar la base de referencia del Índice de Precios al Consumidor del área metropolitana de Caracas (IPC) e incorporar en el mismo los cambios de los patrones de consumo de los hogares, ocurridos en el país desde 1989. En este último año se hizo la última encuesta de presupuestos familiares y por ende la actualización de la canasta de productos y su estructura de ponderaciones en el IPC. Así como, mejorar las estimaciones de la cuenta del sector hogares y estimar, por vía indirecta, los niveles de algunas actividades económicas.

La encuesta de presupuestos familiares, abarcó un período de levantamiento de datos de quince meses, comprendidos entre enero de 1997 y marzo de 1998. Desde el punto de vista geográfico, la encuesta fue diseñada para obtener estimaciones con alcance nacional. En este sentido, se investigó una muestra de aproximadamente 9.904 hogares, en el contexto de una población total del país estimada entonces, en 22.777.153 habitantes.

El levantamiento de información se desarrolló en forma conjunta entre el Banco Central de Venezuela (BCV) y la Oficina Central de Estadística e Informática (OCEI), hoy INE. En este trabajo correspondió al primero la responsabilidad de la recolección y procesamiento de información en las siguientes ciudades: Área metropolitana de Caracas, Maracaibo, Barquisimeto, Maracay y Puerto la Cruz-Barcelona, en lo que se denominó área BCV. Por su parte, la OCEI reveló información en las ciudades medianas, pequeñas y poblaciones de menos de 25.000 habitantes, denominados área OCEI.

Adicionalmente, la Corporación Venezolana de Guayana (CVG) y la Universidad de los Andes (ULA), mediante convenio con el BCV realizaron la recolección y codificación de la información en las regiones Guayana y Andes, respectivamente.

De manera más detallada puede afirmarse que la Encuesta sobre presupuestos Familiares es una investigación estadística por muestreo que se realiza a los hogares con el fin de obtener información sobre sus ingresos, egresos, características de las viviendas que habitan, composición del hogar y otras variables económicas y sociales de los miembros que lo integran a los fines de:

- Actualizar el Índice de Precios al Consumidor (IPC), en cuanto a la composición de la canasta y estructura de ponderaciones de los bienes y servicios que la componen.
- Elaborar las principales cuentas del sector hogares, según los requerimientos del sistema de cuentas nacionales.

- Estimar el nivel y estructura del ingreso de los hogares según su origen y naturaleza

Unos de los problemas y motivos de la realización de la investigación que se propone es la ausencia de una base de datos con tecnología referencial que permita aprovechar al máximo la información recogida en la encuesta. Esto trae como una subutilización de los datos recogidos, es decir, se realiza un enorme esfuerzo (costos) en la realización de una encuesta para luego utilizar los datos así obtenidos una sola vez. Además, al implementar la base de datos, el producto resultante permitirá integrar toda la información y así realizar análisis estadísticos de mayor alcance o envergadura tales como análisis multivariante, construcción de modelos, entre otros. Por otro lado también será una herramienta que facilite la consulta de la información recolectada para así diagnosticar problemas sociales, tales como: pobreza y desigualdad, nutrición, equipamiento familiar, educación, economía informal, etc. Las definiciones básicas empleadas en la Encuesta de Presupuestos familiares se presentan en el apéndice A.

OBJETIVOS

Esta investigación tiene como objetivo principal realizar un análisis estadístico sobre la información recolectada en la II encuesta nacional de presupuestos familiares, con base en el año 1997. Con este propósito se ha planteado los siguientes objetivos específicos:

- Diseñar e implementar una base de datos, a partir de la información obtenida en la encuesta nacional de presupuestos familiares.
- Realizar una completa descripción estadística de la información así organizada

- Construir un conjunto de indicadores sociales para el estudio de las condiciones de vida de los hogares
- Determinar la estructura de las variables socioeconómicas, identificando cuáles están asociadas y cuáles son aquellas que van en el mismo sentido o sentido opuesto.
- Determinar qué variables de gastos contribuyen a identificar el comportamiento o nivel de calidad de vida de los hogares.
- Construir modelos loglineales que permitan modelar la estructuras de las variables socioeconómicas del tipo categóricas.

CAPITULO I

DISEÑO DE LA BASE DE DATOS

Antes de entrar al diseño de la base de datos es oportuno describir las ventajas, sus elementos y características.

Una base de datos tiene como propósito mantener la información y hacer que esté disponible cuando se solicite. Las ventajas de una base de datos se pueden percibir de manera inmediata: se dispone en cualquier momento de información precisa y al día, y la máquina puede obtener y modificar datos con mucha mayor velocidad que un ser humano. Así es posible satisfacer con rapidez consultas de casos particulares, del momento, sin necesidad de búsquedas visuales o manuales que requieren de mucho tiempo. También es posible disminuir la redundancia, es decir, en los sistemas sin bases de datos cada aplicación tiene sus propios archivos privados provocando redundancia en los datos almacenados. En este trabajo se utilizará el concepto de base de datos relacional, el cual no es más que un conjunto de tablas relacionadas entre sí mediante campos. Cada tabla contiene un conjunto de registros que a su vez contienen un conjunto de atributos o tipos de

datos llamados campos. Otra característica importante de las tablas son las claves primarias, las cuales son conjuntos de campos que identifican de forma única a cada registro de las tablas.

Para el diseño de la base de datos para la II Encuesta de Presupuestos Familiares se realizó un estudio del conjunto de variables para poder agruparlas en entidades y luego establecer las relaciones entre ellos. En la figura número uno se presenta el diagrama referencial de la base de datos, el cual describe las relaciones entre las diferentes entidades de la base de datos. En otras palabras describe la arquitectura de la base de datos, punto esencial para una efectiva implementación.

En el diagrama referencial (figura N° 1) se muestran las distintas relaciones entre las entidades (tablas) de la base de datos. Cada tabla está representada por un rectángulo con su nombre. Las relaciones con otras tablas se indican por medio de una flecha con el nombre del campo que las relaciona. A continuación se presenta un resumen de cada una de las tablas. En el apéndice B esta en detalle el diseño de cada una de las tablas (Nombre del Campo, Descripción y tipo de dato).

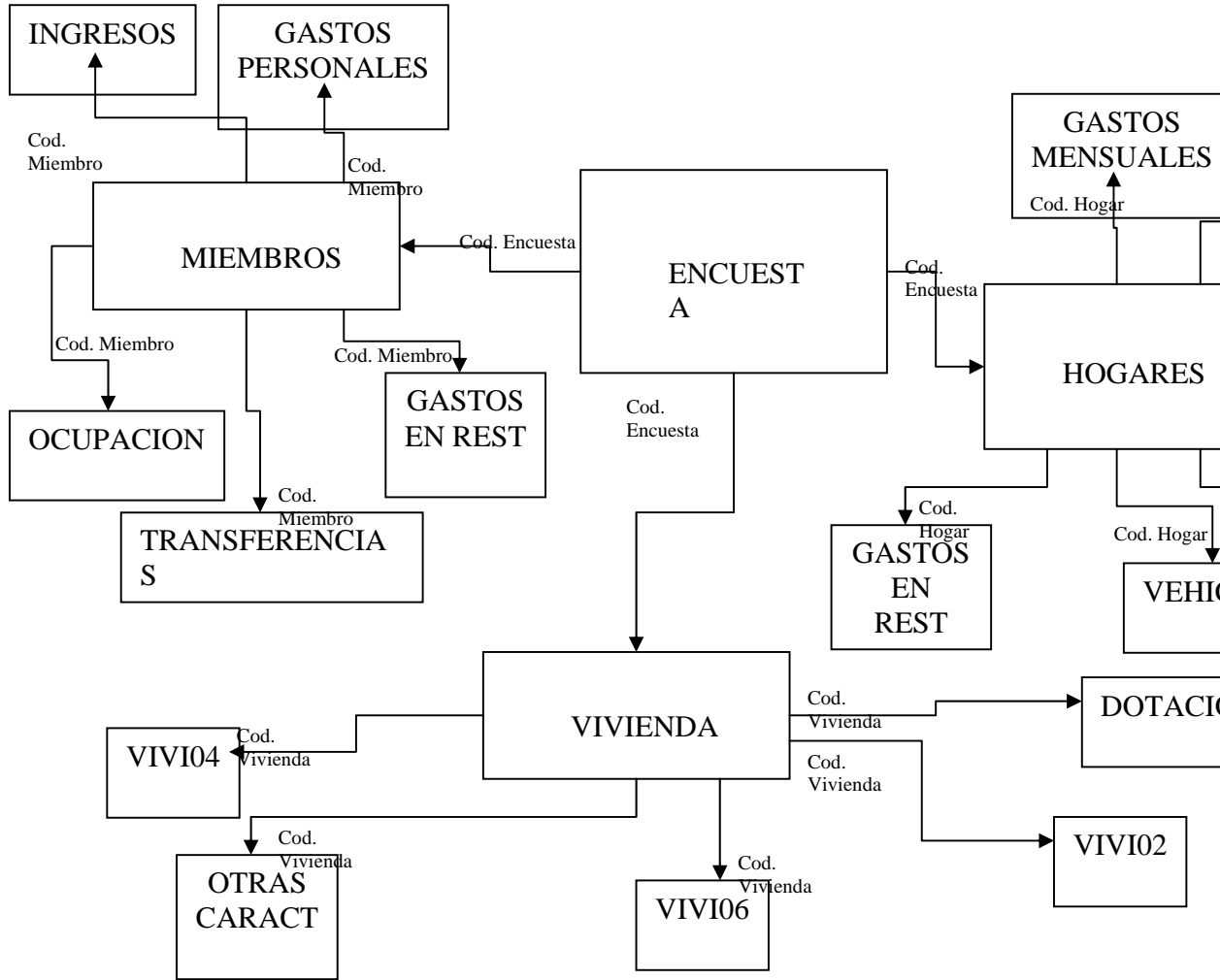
- Región: Almacena la información correspondiente a situación geográfica e identificación de los responsables en realizar el levantamiento de la información.
- Vivi02: Almacena la información referente a la identificación de los miembros que residen en la vivienda y además a qué hogar pertenecen (considerando que puede existir más de un hogar en cada vivienda).

- Vivi03: En esta tabla se recoge la información sobre aspectos generales e identificación de los hogares de cada vivienda.

Figura N° 1

Diagrama referencial de la base de datos para la II Encuesta de Presupuestos

Familiares de Venezuela



- Vivi04: Almacena la información referente a las características físicas y servicios que posee la vivienda.
- Vivi06: Almacena la información relacionada con el tipo de actividad económica que se realiza en la vivienda y qué porcentaje de ésta se destina para ese fin.
- Miembros: Almacena todas las características sociales y laborales de cada miembro del hogar entrevistado.
- Viviendas: En esta tabla se almacenan todas las características de la vivienda referentes al tipo de ocupación (alquilado, propio, traspaso) y todos los aspectos legales y mercantiles de la misma.
- Vehículo: Almacena las características principales de los vehículos usados por los hogares.
- Actividad: Almacena las características de las actividades económicas que se realizan en las viviendas y cómo se emplean sus ganancias.
- Dotación: Almacena las características de los equipos con los que está dotada la vivienda.
- Gastos diarios del hogar: Registra todos los aspectos principales referente a los gastos diarios del hogar.
- Gastos de hogar en restaurantes: Almacena todas las características más relevantes de los gastos del hogar en restaurantes
- Gastos diarios personales: Almacena todo lo referente a los gastos diarios personales de los residentes de las viviendas encuestadas
- Gastos personales en restaurante: Almacena todo lo referente a los gastos personales en restaurante de los residentes de las viviendas encuestadas.

- Gastos Mensuales: Almacenan todas las características de los gastos mensuales de los hogares.
- Gastos Trimestrales y anuales: Almacenan todas las características de los gastos Trimestrales y anuales de los hogares.
- Ocupación: Almacena las características de la ocupación de cada miembro del hogar.
- Ingresos: Almacena los ingreso de cada miembro
- Transferencia: Almacena los ingreso por transferencia de cada miembro

La implementación de la base de datos se realizo en Access 2000 ®. Por ser un manejador de base de datos bastante conocido y fácil de manejar, así como por sus bondades de compatibilidad con los programas de Microsoft y especialmente Office. (En el Cd adjunto a este trabajo se encuentra almacenado la implementación del diseño de la base de datos expuesta anteriormente)

CAPITULO II

MARCO TEÓRICO

El objetivo de este capítulo es describir o presentar los fundamentos teóricos de los diferentes métodos estadísticos utilizados en este trabajo. Así en la primera parte se presenta “el Análisis General”, el cual es la base o núcleo común de las principales técnicas del análisis factorial. En la segunda parte se describe la técnica de Análisis de Componentes Principales; luego como una introducción para explicar el Análisis de Correspondencia se esboza la generalización del Análisis General (Análisis General ponderado con una métrica cualquiera), y por último se muestra una introducción sobre teoría y diseño de base de datos con la finalidad de aclarar conceptos a la hora de explicar el diseño de la base de datos que presenta este trabajo.

Con este capítulo se pretende dar un preámbulo para el buen entendimiento de los capítulos posteriores.

2.1 Análisis General.

El análisis general comprende una serie de herramientas del álgebra matricial que constituyen el núcleo matemático común de las principales técnicas del análisis factorial de datos.

El punto de partida para el análisis lo constituye una matriz de datos $X_{n \times p}$, cuyas filas corresponden a los individuos y cuyas columnas corresponden a las variables, de modo que el término genérico x_{ij} representa el valor de la variable j en el individuo i .

La notación que se utilizara a estos efectos es la siguiente:

$$X_{n \times p} = (X^1, X^2, \dots, X^p) = \begin{pmatrix} X_1^t \\ X_2^t \\ \cdot \\ X_n^t \end{pmatrix}$$

Donde:

$(X^j)_{n \times 1}$ representa la j -ésima columna de X y

$(X_i)_{p \times 1}$ representa la i -ésima fila de X

2.1.1 Ajuste en R^p

La nube de puntos en R^p , que se denotara por $N(I)$, está definida por el conjunto de las n filas X_1, X_2, \dots, X_n , dotadas de la métrica euclidea. El objetivo del ajuste es encontrar el subespacio de R^p de dimensión q ($q < p$), con direcciones ortonormalizadas, que mejor se ajuste a la nube $N(I)$ en el sentido de los mínimos cuadrados. Esto es, se trata de la

búsqueda de un subespacio de representación de datos que garantice el mayor parecido global de las filas X_i a sus aproximaciones sobre tal subespacio.

El procedimiento a seguir para la búsqueda del subespacio que denominaremos como de “mejor ajuste”, consistirá en una sucesión de etapas, en cada una de las cuales se construye una nueva dirección del subespacio. En una primera etapa hallaremos el subespacio de dimensión 1, es decir, la recta que mejor se ajusta a la nube. El vector director de esta recta será denotado por v , del cual exigiremos además que sea unitario. Matemáticamente, se trata de un problema de mínimos cuadrados.

$$\text{Min } \sum d^2(x_i, Px_i)$$

donde Px_i es la proyección ortogonal de x_i sobre la recta.

Con el objeto de facilitar el procedimiento de minimización conviene tomar en cuenta que:

- La distancia cuadrado del origen a un punto x_i , que es un valor fijo, es igual a:

$$d^2(O, x_i) = \|x_i\|^2 = x_i^t x_i$$

- La distancia cuadrado del origen a la proyección Px_i , que depende del vector v , es igual a

$$\begin{aligned} d^2(O, Px_i) &= \|Px_i\|^2 = (x_i^t v)^2 \\ &= (x_i^t v)^t (x_i^t v) = v^t x_i x_i^t v \end{aligned}$$

- En virtud del teorema de Pitágoras se establece la relación fundamental:

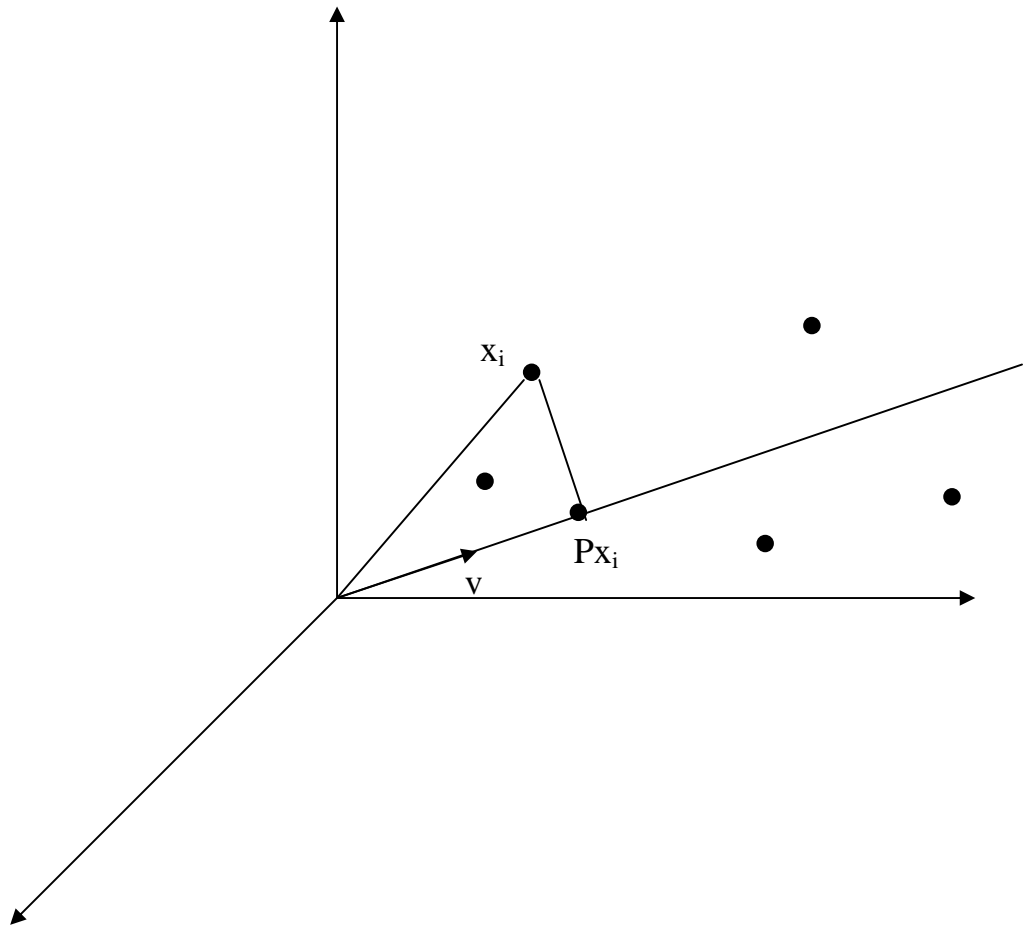
$$d^2(O, x_i) = d^2(O, Px_i) + d^2(x_i, Px_i)$$

de donde:

$$d^2(x_i, P_{x_i}) = d^2(O, x_i) - d^2(O, P_{x_i})$$

FIGURA N° 2

Proyección en \mathbb{R}^3



De acuerdo con lo anterior, minimizar la suma $\sum d^2(x_i, P_{x_i})$, resulta equivalente a maximizar la suma:

$$\sum d^2(O, P_{x_i}) = \sum v^t x_i x_i^t v = v^t \sum x_i x_i^t v = v^t X^t X v$$

Para maximizar esta forma cuadrática dado la restricción de que v sea un vector unitario, construimos el lagrangiano:

$$L = v^t X^t X v - \lambda (v^t v - 1)$$

derivamos respecto de v , e igualamos al vector nulo θ :

$$2 X^t X v - 2 \lambda v = \theta$$

obteniéndose la ecuación:

$$X^t X v = \lambda v$$

lo que determina que el vector buscado v sea un autovector de la matriz $X^t X$ asociado con el autovalor λ .

Esto quiere decir que una condición necesaria para que v defina un máximo en la forma cuadrática en referencia, es que sea un autovector de $X^t X$. Dentro del conjunto de autovectores de esa matriz nos interesamos por aquel que efectivamente maximice la forma cuadrática:

$$\text{Max} (v^t X^t X v) = \text{Max} (v^t \lambda v) = \text{Max} (\lambda v^t v) = \text{Max} (\lambda)$$

De modo que el vector director de la recta que mejor se ajusta a la nube de puntos es el autovector de $X^t X$, que denotaremos por v^1 , asociado con el mayor autovalor, que denotaremos por λ_1

En la segunda etapa del procedimiento hallaremos el subespacio de dimensión 2, es decir, el plano que mejor se ajuste a la nube. Este plano tendrá como primera dirección la definida por v^1 , mientras que la segunda quedará determinada por el vector unitario v^2 ,

ortogonal a v^1 que minimice la suma de cuadrados $\sum d^2(x_i, Px_i)$, donde ahora Px_i es la proyección ortogonal de x_i sobre el plano.

El Lagrangiano que toma en cuenta tales restricciones es ahora:

$$L = v^t X^t X v - \lambda (v^t v - 1) - \gamma (v^t v^1)$$

Si derivamos respecto de v , e igualamos al vector nulo:

$$2 X^t X v - 2 \lambda v - \gamma v^1 = 0$$

Premultiplicando por $(v^1)^t$ se obtiene que:

$$2(v^1)^t X^t X v - 2 \lambda (v^1)^t v - \gamma (v^1)^t v^1 = -\gamma = 0$$

ya que $(v^1)^t v = 0$. Entonces, al igual que en la primera etapa, queda que $X^t X v = \lambda v$. Por lo tanto, las direcciones del plano que mejor se ajustan a la nube de puntos en R^p están definidas por los vectores v^1 y v^2 , autovectores ortonormalizados de la matriz $X^t X$ asociados con sus dos mayores autovalores λ_1 y λ_2 .

Reiterando q veces el mismo procedimiento se obtiene que:

“Una base ortonormalizada de dimensión q que mejor se ajusta a la nube de puntos en R^p , está constituida por los q autovectores v^1, v^2, \dots, v^q correspondientes a los q mayores autovalores $\lambda_1, \lambda_2, \dots, \lambda_q$ de la matriz $X^t X$ ”.

Las coordenadas de proyección de los n puntos fila sobre el eje α conforman el

vector:

$$\psi^\alpha = \begin{pmatrix} X_1^t v^\alpha \\ X_2^t v^\alpha \\ \cdot \\ \cdot \\ X_n^t v^\alpha \end{pmatrix} = \begin{pmatrix} X_1^t \\ X_2^t \\ \cdot \\ \cdot \\ X_n^t \end{pmatrix} v^\alpha = X v^\alpha = \sum_{j=1}^p v_{j\alpha} X^j$$

que constituye una combinación lineal de las columnas (variables) de la matriz X , y además es tal que:

$$\sum_{i=1}^n \psi_{i\alpha}^2 = (\psi^\alpha)^t \psi^\alpha = (X v^\alpha)^t X v^\alpha = (v^\alpha)^t X^t X v^\alpha$$

$$\sum_{i=1}^n \psi_{i\alpha}^2 = (v^\alpha)^t \lambda_\alpha v^\alpha = \lambda_\alpha (v^\alpha)^t v^\alpha = \lambda_\alpha$$

y

$$(\psi^\alpha)^t \psi^\beta = \lambda_\beta (v^\alpha)^t v^\beta = 0$$

Las coordenadas de los n puntos filas sobre q ejes pueden organizarse en la matriz

$$\Psi(q) = (\psi^1, \psi^2, \dots, \psi^q) = X V(q)$$

siendo $V(q)$ la matriz cuyas columnas son los q autovectores ortonormalizados asociados con los q mayores autovalores de la matriz $X^t X$.

En resumen:

El vector ψ^α de coordenadas de proyección de los puntos fila sobre el eje α es tal que:

- Es una combinación lineal de las columnas (variables) de la matriz X
- Su norma cuadrado es igual a λ_α
- Es ortogonal a ψ^β , para cualquier $\beta \neq \alpha$

2.1.2 Ajuste en R^n

La nube de puntos en R^n , que se denotara por $N(J)$, está definida por el conjunto de las p columnas X^1, X^2, \dots, X^p , dotadas de la métrica euclidea. El objetivo del ajuste es encontrar el subespacio de R^n de dimensión q ($q < p$), con direcciones ortonormalizadas, que mejor se ajuste a la nube $N(J)$ en el sentido de los mínimos cuadrados.

Procediendo en forma similar al caso anterior se encuentra que:

“Una base ortonormalizada de dimensión q que mejor se ajusta a la nube de puntos en R^n , está constituida por los q autovectores u^1, u^2, \dots, u^q correspondientes a los q mayores autovalores $\lambda_1, \lambda_2, \dots, \lambda_q$ de la matriz XX^t ”.

Las coordenadas de proyección de los p puntos columna sobre el eje α conforman el vector:

$$\varphi^\alpha = \begin{pmatrix} (X^1)^t u^\alpha \\ (X^2)^t u^\alpha \\ \cdot \\ (X^p)^t u^\alpha \end{pmatrix} = \begin{pmatrix} (X^1)^t \\ (X^2)^t \\ \cdot \\ (X^p)^t \end{pmatrix} u^\alpha = X u^\alpha = \sum_{i=1}^n u_{i\alpha} X_i$$

que constituye una combinación lineal de las filas (individuos) de la matriz X, y además es tal que:

$$\sum_{j=1}^p \varphi_{j\alpha}^2 = (\varphi^\alpha)^t \varphi^\alpha = (X u^\alpha)^t X u^\alpha = (u^\alpha)^t X X^t u^\alpha$$

$$\sum_{j=1}^p \varphi_{j\alpha}^2 = (u^\alpha)^t \lambda_\alpha u^\alpha = \lambda_\alpha (u^\alpha)^t u^\alpha = \lambda_\alpha$$

y $(\varphi^\alpha)^t \varphi^\beta = \lambda_\beta (u^\alpha)^t u^\beta = 0$

Las coordenadas de los p puntos columna sobre q ejes pueden organizarse en la matriz

$$\varphi(q) = (\varphi^1, \varphi^2, \dots, \varphi^q) = X^t U(q)$$

siendo U(q) la matriz cuyas columnas son los q autovectores ortonormalizados asociados con los q mayores autovalores de la matriz XX^t .

En resumen:

El vector φ^α de coordenadas de proyección de los puntos fila sobre el eje α es tal que:

- Es una combinación lineal de las filas (individuos) de la matriz X
- Su norma cuadrado es igual a λ_α
- Es ortogonal a φ^β , para cualquier $\beta \neq \alpha$

2.1.3 Relación entre los ajustes

Como existen unas relaciones entre los autovectores v de la matriz $X^t X$ y los autovectores u de la matriz XX^t , a partir de ellos se definen las siguientes ecuaciones matriciales, conocidas como “relaciones de doble transición”:

$$V = X^t U D^{-1}$$

$$U = X V D^{-1}$$

donde D es la matriz diagonal cuyos elementos son los r valores singulares de la matriz X . Por consiguiente, basta solamente con diagonalizar una de las dos matrices $(X^t X)$ o (XX^t) , para hallar las coordenadas de las filas y de las columnas sobre el subespacio de mejor ajuste. Por lo general se diagonaliza $X^t X$, ya que usualmente el número de variables es menor que el número de individuos.

A continuación estudiaremos las relaciones entre las coordenadas de los dos análisis, en R^p y en R^n .

- Si se ha diagonalizado la matriz $X^t X$ y se han obtenido los autovectores v^α y los autovalores λ_α , se pueden calcular las coordenadas de los puntos columna por medio de:

$$\varphi = X^t U = V D$$

de modo que:

$$\varphi^\alpha = \sqrt{\lambda_\alpha} v^\alpha$$

- Si se ha diagonalizado la matriz XX^t y se han obtenido los autovectores u^α y los autovalores λ_α , se pueden calcular las coordenadas de los puntos fila por medio de:

$$\psi = X V = U D$$

de modo que:

$$\psi^\alpha = \sqrt{\lambda_\alpha} u^\alpha$$

- Por otro lado, como:

$$V = (X^t U) D^{-1} = \phi D^{-1} \text{ y } U = (XV) D^{-1} = \psi D^{-1}$$

entonces:

$$\phi = X^t U = X^t \psi D^{-1}$$

luego:

$$\phi^\alpha = X^t \psi^\alpha \frac{1}{\sqrt{\lambda_\alpha}}$$

y por tanto:

$$\phi_{j\alpha} = (X^j)^t \psi^\alpha \frac{1}{\sqrt{\lambda_\alpha}} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n x_{ij} \psi_{i\alpha} \quad j = 1, 2, 3 \dots p$$

- En forma similar, de la relación:

$$\psi = XV = X\phi D^{-1}$$

se obtiene:

$$\psi^\alpha = X\phi^\alpha \frac{1}{\sqrt{\lambda_\alpha}}$$

$$\text{de donde: } \psi_{i\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^p x_{ij} \phi_{j\alpha} \quad i = 1, 2, 3 \dots n$$

2.1.4. Reconstrucción de los datos

Por la descomposición en valores singulares de una matriz se tiene que:

$$X = UDV^t = \sum_{\alpha=1}^q \delta_\alpha u^\alpha (v^\alpha)^t$$

siendo δ_α los valores singulares y r el rango de la matriz X . Tal relación puede considerarse como una vía para reconstruir la matriz original a partir de los autovalores y autovectores de las matrices $X^t X$ y XX^t . Al considerar solamente los q ($q < r$) primeros autovalores y autovectores, se obtiene una reconstrucción aproximada de la matriz X , la cual constituye la mejor aproximación mínimo-cuadrática de rango q :

$$X \cong X(q) = \sum_{\alpha=1}^q \sqrt{\lambda_\alpha} u^\alpha (v^\alpha)^t$$

En particular, el término (i,j) :

$$x_{ij} \cong \sum_{\alpha=1}^q \sqrt{\lambda_\alpha} u_{i\alpha} v_{j\alpha} \quad i = 1, 2, \dots, n ; j = 1, 2, \dots, p$$

2.1.5 Contribuciones:

En las diferentes técnicas factoriales se utilizan ciertos coeficientes denominados contribuciones, que constituyen elementos importantes para la interpretación de resultados.

Estos coeficientes se basan en los siguientes resultados:

Para las variables:

- $\sum_{j=1}^p \phi_{j\alpha}^2 = \lambda_\alpha$
- $\sum_{\alpha=1}^p \phi_{j\alpha}^2 = T_j = \sum_{i=1}^n x_{ij}^2$

y para los individuos:

- $\sum_{i=1}^n \psi_{i\alpha}^2 = \lambda_\alpha$
- $\sum_{\alpha=1}^p \psi_{i\alpha}^2 = T_i = \sum_{j=1}^p x_{ij}^2$

Se define como “Contribución de la variable j al factor α ” a la expresión:

$$CV_j F_\alpha = \frac{\varphi_{j\alpha}^2}{\lambda_\alpha} \quad (j=1, 2, \dots, p)$$

Este valor usualmente se expresa en términos de porcentaje y se interpreta como la contribución de cada variable j a la variabilidad del factor α .

Aquellas variable con mayor contribución a la variabilidad de un determinado factor serán las que se utilizarán para interpretarlo. Por otro lado, como ocurre que $\varphi_{j\alpha} = \sqrt{\lambda_\alpha} v_{j\alpha}$,

se tiene que: $CV_j F_\alpha = v_{j\alpha}^2 \quad (j=1, 2, \dots, p)$

Esquemáticamente:

FIGURA N° 3

**TABLA DE COORDENADAS Y CONTRIBUCIONES
DE LA VARIABLE AL FACTOR**

Coordenadas Factor						Contribuciones $CV_j F_\alpha$				
Variable	1	...	α	...	p	1	...	α	...	P
1								$\varphi_{1\alpha}^2 / \lambda_\alpha$		
⋮								⋮		
J	φ_{j1}	...	$\varphi_{j\alpha}$...	φ_{jp}	$\varphi_{j1}^2 / \lambda_1$...	$\varphi_{j\alpha}^2 / \lambda_\alpha$...	$\varphi_{jp}^2 / \lambda_p$
⋮								⋮		
P			$\varphi_{p\alpha}$					$\varphi_{p\alpha}^2 / \lambda_\alpha$		
$\sum_{j=1}^p \varphi_{j\alpha}^2$	λ_1	...	λ_α	...	λ_p	$\sum \lambda_\alpha$	Σ	1	...	1

Se define como “Contribución del factor α a la variable j ” a la expresión:

$$CF_{\alpha} V_j = \frac{\varphi_{j\alpha}^2}{T_j} \quad (\alpha = 1, 2, \dots, p)$$

Este valor usualmente se expresa en términos de porcentaje y se interpreta como la contribución de cada factor α a la variabilidad de la variable j . Aquellos factores con mayor contribución a la variable serán los ejes en los cuales la variable estará mejor representada.

Esquemáticamente:

FIGURA N° 4

**TABLA DE COORDENADAS Y CONTRIBUCIONES
DEL FACTOR A LA VARIABLE**

Coordenadas							Contribuciones				
Factor							$CF_{\alpha} V_j$				
Variable	1	...	α	...	p	$\sum_{\alpha=1}^p \varphi_{j\alpha}^2$	1	...	α	...	P
1			$\varphi_{1\alpha}$			T_1			$\varphi_{1\alpha}^2 / T_1$		
.						.			.		
.						.			.		
J	φ_{j1}	...	$\varphi_{j\alpha}$...	φ_{jp}	T_j	φ_{j1}^2 / T_j	...	$\varphi_{j\alpha}^2 / T_j$...	φ_{jp}^2 / T_j
.						.			.		
.						.			.		
P			$\varphi_{p\alpha}$			T_p			$\varphi_{p\alpha}^2 / T_p$		
						$\sum \lambda_{\alpha}$					

Se define como “Contribución del individuo i al factor α ” a la expresión:

$$CI_i F_{\alpha} = \frac{\psi_{i\alpha}^2}{\lambda_{\alpha}} \quad (i = 1, 2, \dots, n)$$

Este valor usualmente se expresa en términos de porcentaje y se interpreta como la contribución de cada individuo i a la variabilidad del factor α . Aquellos individuos con mayor contribución a la variabilidad de un determinado factor serán los que se utilizarán para interpretarlo. Por otro lado, como ocurre que $\psi_{i\alpha} = \sqrt{\lambda_\alpha} u_{i\alpha}$, se tiene que:

$$CI_{iF_\alpha} = u_{i\alpha}^2$$

Esquemáticamente:

FIGURA N° 5
TABLA DE COORDENADAS Y CONTRIBUCIONES
DEL INDIVIDUO AL FACTOR

Coordenadas Factor						Contribuciones CI_{iF_α}				
Variable	1	...	α	...	p	1	...	α	...	P
1								$\psi_{1\alpha}^2 / \lambda_\alpha$		
⋮								⋮		
i	ψ_{i1}	...	$\psi_{i\alpha}$...	ψ_{ip}	ψ_{i1}^2 / λ_1	...	$\psi_{i\alpha}^2 / \lambda_\alpha$...	ψ_{ip}^2 / λ_p
⋮								⋮		
n			$\psi_{n\alpha}$					$\psi_{n\alpha}^2 / \lambda_\alpha$		
$\sum_{i=1}^n \psi_{i\alpha}^2$	λ_1	...	λ_α	...	λ_p	$\text{tr}(X^t X)$		1	...	1

Se define como “Contribución del factor α al individuo i ” a la expresión:

$$CF_{\alpha} I_i = \frac{\psi_{i\alpha}^2}{T_i} \quad (\alpha=1,2,\dots,p)$$

Este valor usualmente se expresa en términos de porcentaje y se interpreta como la contribución de cada factor α a la variabilidad del individuo i en el espacio R^p . Aquellos

factores con mayor contribución al individuo serán los ejes en los cuales el individuo estará mejor representado.

Esquemáticamente:

FIGURA N 6

**TABLA DE COORDENADAS Y CONTRIBUCIONES
DEL FACTOR AL INDIVIDUO**

Variable	Coordenadas Factor					$\sum_{\alpha=1}^p \Psi_{i\alpha}^2$	Contribuciones $CF_{\alpha I_i}$				
	1	...	α	...	p		1	...	α	...	P
1			$\Psi_{1\alpha}$			T_1			$\Psi_{1\alpha}^2 / T_1$		
·						·			·		
·						·			·		
·						·			·		
i	Ψ_{i1}	...	$\Psi_{i\alpha}$...	Ψ_{ip}	T_i	Ψ_{i1}^2 / T_i	...	$\Psi_{i\alpha}^2 / T_i$...	Ψ_{ip}^2 / T_i
·						·			·		
·						·			·		
·						·			·		
n			$\Psi_{n\alpha}$			T_n			$\Psi_{n\alpha}^2 / T_n$		
											$\sum \lambda_{\alpha}$

2.2. Análisis de Componentes Principales

La técnica multivariante conocida como análisis de componentes principales (ACP), puede ser presentada según el enfoque geométrico de K. PEARSON (1901) o según el enfoque clásico de H. HOTTELING (1933). De acuerdo con el primero, el ACP tiene por objeto la búsqueda de planos de representación óptima de una nube de puntos, según el criterio de mínimos cuadrados. De acuerdo con el segundo, el ACP tiene por objeto obtener combinaciones lineales de las variables originales, con propiedades especiales en términos

de varianza y correlación. En ambos casos se obtienen nuevas variables, denominadas componentes principales, las cuales deben retener lo esencial de la información contenida en las variables originales en cuanto a variabilidad e intercorrelación.

2.2.1 Enfoque Geométrico.

El problema planteado en el enfoque geométrico, es la búsqueda del subespacio q -dimensional de mejor ajuste a la nube de puntos fila (o columna) de una matriz de datos, utilizando la técnica de los mínimos-cuadrados. Con este fin vamos a adaptar los resultados del análisis general (AG) pero con la idea de que nos interesa no sólo reducir, sino también describir la matriz original de datos. En este sentido, dado que la descripción estadística enfatiza los aspectos relacionados con la variabilidad, lo que se debe maximizar no es la suma de cuadrados de las distancias en proyección con respecto al origen, sino la suma de cuadrados de estas distancias con respecto al centro de gravedad de la nube proyectada:

$$\begin{aligned} \sum d^2(P_{x_i}, \bar{P}_{x_i}) &= \sum d^2(P_{x_i}, P_{\bar{x}_i}) = \sum (x_i^t v - \bar{x}^t v)^2 \\ &= \sum ((x_i - \bar{x})^t v)^2 = v^t \hat{X}^t \hat{X} v \end{aligned}$$

A diferencia de los resultados obtenidos en el análisis general, la matriz a considerar ahora no es $X_{n \times p}$. Por lo tanto, la solución al problema planteado puede resumirse en la forma siguiente:

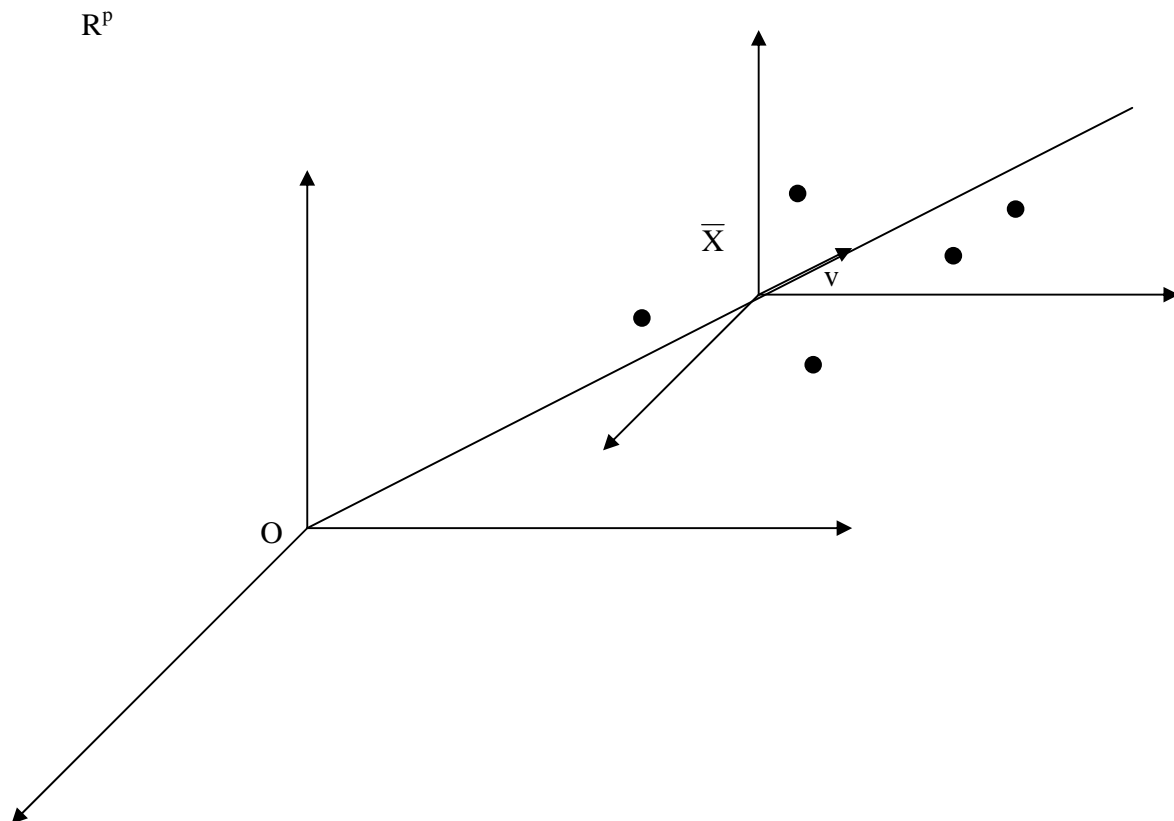
“Las direcciones ortonormalizadas del subespacio q -dimensional de R^p , que mejor se ajusta a la nube de puntos fila centrados, coincide con los q autovectores v^1, v^2, \dots, v^q correspondientes a los q mayores autovalores $\lambda_1, \lambda_2, \dots, \lambda_q$ de la matriz $\hat{X}^t \hat{X}$ ”.

Por razones de interpretación de los resultados, algunos autores introducen el factor $(1/\sqrt{n-1})$ en la matriz \hat{X} de modo que la matriz a diagonalizar sea la matriz de varianza y covarianza S .

Haciendo abstracción de la dimensión R^p , representamos en el siguiente figura la nube de puntos fila de la matriz X con traslación de origen, lo que da lugar a una nube de puntos cuyo centro de gravedad coincide con el centro de coordenadas.

FIGURA N° 7

NUBE DE PUNTOS FILA DE LA MATRIZ X CON TRASLACIÓN DE ORIGEN



Las coordenadas de proyección de los n puntos fila sobre el eje α , se organizan sobre el vector:

$$\psi^\alpha = \hat{X} v^\alpha = \sum_{j=1}^p v_{j\alpha} \hat{x}^j$$

el cual define una nueva variable, combinación lineal de las p variables originales centradas, la cual denominaremos α -ésima componente principal.

Estas coordenadas tienen importante propiedades estadísticas siguientes:

$$\text{Med}(\psi^\alpha) = (1/n) \sum_{i=1}^n \psi_{i\alpha} = \mathbf{1}^t \psi^\alpha = (1/n) \mathbf{1}^t \hat{X} v^\alpha = 0$$

$$\begin{aligned} \text{Var}(\psi^\alpha) &= (1/(n-1)) \sum_{i=1}^n \psi_{i\alpha}^2 = (1/(n-1)) (v^\alpha)^t \hat{X}^t \hat{X} v^\alpha \\ &= (1/(n-1)) \lambda_\alpha (v^\alpha)^t v^\alpha = (1/(n-1)) \lambda_\alpha = \lambda_\alpha^* \quad (\text{autovalor de S}) \end{aligned}$$

$$\text{Cov}(\psi^\alpha, \psi^\beta) = (1/(n-1)) \sum_{i=1}^n \psi_{i\alpha} \psi_{i\beta} = (1/(n-1)) (v^\alpha)^t \hat{X}^t \hat{X} v^\beta = \lambda_\beta (v^\alpha)^t v^\beta = 0$$

2.2.2 Variabilidad explicada por las componentes

Hemos señalado que uno de los aspectos importantes del ACP es que permite resumir la variabilidad contenida en los datos originales, en unas pocas variables. A este respecto es conveniente tomar en cuenta los siguientes elementos:

- La traza de la matriz de varianza y covarianza es una medida de variabilidad total.

Por su parte:

$$\text{tr}(\mathbf{S}) = \sum_{j=1}^p \mathbf{S}_j^2 = \frac{1}{n-1} \text{tr}(\hat{\mathbf{X}}^t \hat{\mathbf{X}}) = \frac{1}{n-1} \sum_{j=1}^p \lambda_j = \sum_{j=1}^p \lambda_j^*$$

y además :

$$\begin{aligned} \frac{1}{n-1} \sum_{i=1}^n d^2(x_i, \bar{x}) &= \frac{1}{n-1} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_{.j})^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(x_{ij} - \bar{x}_{.j})^2}{n-1} \\ &= \sum_{j=1}^p \mathbf{S}_j^2 = \text{tr}(\mathbf{S}) \end{aligned}$$

Es decir $\text{tr}(\mathbf{S})$ es un indicador de la variabilidad total de la nube de puntos, que mide globalmente el grado de similitud entre cada fila y su centro de gravedad, al cual se denomina Inercia total.

Esta expresión puede escribirse también como:

$$\text{Tr}(\mathbf{S}) = \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{s=1}^n d^2(x_i, x_s)$$

Es decir, se trata también de una medida de la desemejanzas entre filas de una matriz de datos, tomados dos a dos.

- La variabilidad explicada por las q primeras componentes es la mejor aproximación a $\text{tr}(\mathbf{S})$. Si denotamos por $\psi_{i(q)}$ al vector que contiene las coordenadas de proyección del i -ésimo individuo sobre el subespacio q -dimensional de mejor ajuste, entonces la variabilidad de la nube proyectada sobre el subespacio viene dada por:

$$\frac{1}{n-1} \sum_{i=1}^n d^2(\psi_{i(q)}, \bar{\psi}_{(q)}) = \frac{1}{n-1} \sum_{i=1}^n \sum_{\alpha=1}^q \psi_{i\alpha}^2 = \frac{1}{n-1} \sum_{\alpha=1}^q \lambda_\alpha = \sum_{\alpha=1}^q \lambda_\alpha^*$$

expresión que toma un valor máximo (RAO, 1964)

- La proporción de variabilidad total captada por las q primeras componentes principales viene dada por la:

$$\text{Tasa de inercia} = \frac{\sum_{\alpha=1}^q \lambda_{\alpha}}{\sum_{\alpha=1}^p \lambda_{\alpha}} \cdot 100$$

- Un indicador de la importancia de la α -ésima componente sería:

$$\frac{\lambda_{\alpha}}{\sum \lambda_{\alpha}} \cdot 100 = \% \text{ de la variabilidad total explicada por la } \alpha\text{-ésima}$$

componente

2.2.3. Interpretación de los resultados

Tomando en cuenta que la α -ésima componente principal es una combinación lineal de las variables originales cuyos coeficientes son las coordenadas del vector propio v^{α} :

$$\psi^{\alpha} = \sum_{j=1}^p v_{j\alpha} \hat{x}^j$$

entonces la importancia de la j-ésima variable en la formación de la componente viene dada por $v_{j\alpha}$.

Para darle una interpretación a las componentes principales deben tomarse en cuenta sus correlaciones con las variables originales. En una situación extrema en la cual exista una relación lineal directa (inversa) perfecta entre la α -ésima componente y la j-ésima variable, la componente define un criterio que permite ordenar los individuos de acuerdo con valores crecientes (o decrecientes) de la variable j.

Para evaluar esta relación obtenemos el coeficiente de correlación entre la variable j y el factor α , para lo cual hacemos uso de que:

$$\text{Cov}(X^j, \psi^\alpha) = (1/(n-1))(\hat{X}^j)^t \psi^\alpha = (1/(n-1))(\hat{X}^j)^t \sqrt{\lambda_\alpha} u^\alpha = (1/(n-1))\sqrt{\lambda_\alpha} \varphi_{j\alpha}$$

$$\text{Var}(X^j) = S_j^2$$

y

$$\text{Var}(\psi^\alpha) = (1/(n-1))\lambda_\alpha$$

entonces la correlación entre la j -ésima variable y la α -ésima componente es, salvo el termino $1/(\sqrt{n-1}S_j)$, la componente de proyección de la j -ésima variable sobre el eje α :

$$\begin{aligned} \text{Corr}(X^j, \psi^\alpha) &= \frac{(1/(n-1))\sqrt{\lambda_\alpha} \varphi_{j\alpha}}{S_j(\sqrt{(1/(n-1))\lambda_\alpha})} \\ &= \frac{\varphi_{j\alpha}}{\sqrt{n-1} \cdot S_j} \end{aligned}$$

2.2.4. Contribución del individuo a la componente

Partiendo de la relación: $\sum_{i=1}^n \psi_{i\alpha}^2 = \lambda_\alpha = (n-1)\lambda_\alpha^*$, se define como contribución del

individuo i a la variabilidad de la componente α , a la expresión:

$$\text{CI}_i F_\alpha = \frac{\psi_{i\alpha}^2}{\lambda_\alpha} \cdot 100\%$$

para la interpretación de los resultados deben tomarse en cuenta solamente aquellos individuos con mayor contribución a la variabilidad de la componente.

2.2.5 Contribución de la componente al individuo (Calidad de representación)

Partiendo de la relación:

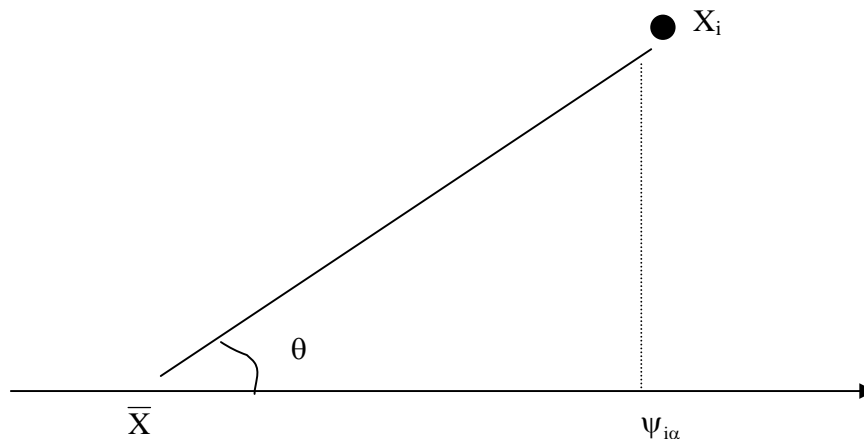
$$\begin{aligned} \sum_{\alpha=1}^p \Psi_{i\alpha}^2 &= \sum_{\alpha=1}^p (\hat{x}_i^t v^\alpha) (\hat{x}_i^t v^\alpha)^t = \hat{x}_i^t \left(\sum_{\alpha=1}^p (v^\alpha) (v^\alpha)^t \right) \hat{x}_i \\ &= \hat{x}_i^t \hat{x}_i = \sum_{j=1}^p (x_{ij} - \bar{x}_{.j})^2 = d^2(x_i, \bar{x}) \end{aligned}$$

se define como contribución del eje α a la ubicación del individuo i , a la expresión:

$$CF_{\alpha} I_i = \frac{\Psi_{i\alpha}^2}{\sum_{\alpha=1}^p \Psi_{i\alpha}^2} \cdot 100\%$$

FIGURA N° 8

Representación grafica de la calidad de representación



$$\cos^2 \theta = \frac{d^2(\Psi_{i\alpha}, \bar{x})}{d^2(x_i, \bar{x})} = \frac{\Psi_{i\alpha}^2}{\sum_{\alpha=1}^p \Psi_{i\alpha}^2}$$

Si $\theta \rightarrow 0^\circ \Rightarrow \cos^2 \theta \rightarrow 1$
 (Punto bien representado)
 Si $\theta \rightarrow 90^\circ \Rightarrow \cos^2 \theta \rightarrow 0$
 (punto mal representado)

2.2.6. Interpretación de la nube de variables

Las coordenadas de los p puntos columna sobre el eje α , vienen dadas por:

$$\varphi^\alpha = \hat{X}^t u^\alpha = \sum_{i=1}^n u_{i\alpha} \hat{x}_i = \sqrt{\lambda_\alpha} v^\alpha$$

y por tanto definen un nuevo individuo, combinación lineal de los n individuos.

Si asumimos que la matriz \hat{X} es de rango r entonces:

$$(n-1) \cdot S = \hat{X}^t \hat{X} = V(r) D^2(r) V^t(r) = \varphi(r) \varphi^t(r)$$

donde $\varphi(r) = (\varphi^1, \varphi^2, \dots, \varphi^r)$

Por consiguiente, es posible reconstruir las varianzas y covarianzas entre las variables originales a partir de las coordenadas de proyección de las variables sobre los ejes factoriales:

$$(n-1) s_j^2 = \sum_{\alpha=1}^r \varphi_{j\alpha}^2 \quad \text{y} \quad (n-1) s_{jk} = \sum_{\alpha=1}^r \varphi_{j\alpha} \varphi_{k\alpha}$$

De esta manera tenemos que el coeficiente de correlación entre las variables j y k:

$$r_{jk} = \frac{\varphi_j^t(r) \varphi_k(r)}{\|\varphi_j(r)\| \|\varphi_k(r)\|} = \text{Cos} \theta_{jk}(r)$$

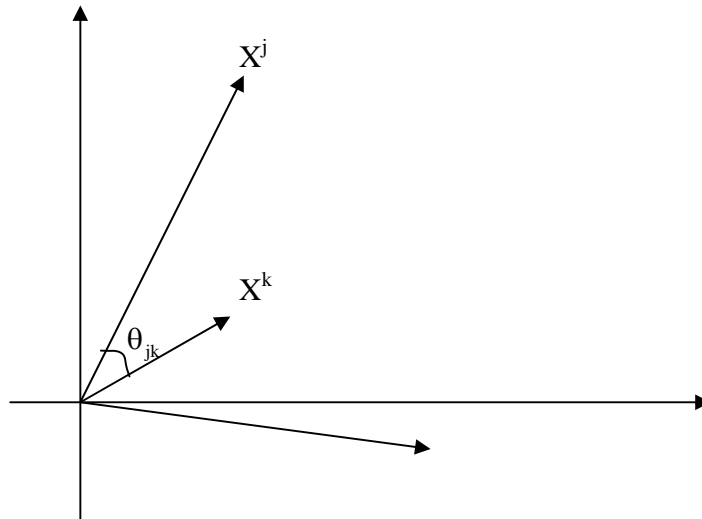
siendo $\theta_{jk}(r)$ el ángulo entre $\varphi_j(r)$ y $\varphi_k(r)$

Como el coeficiente de correlación r_{jk} entre las variables j y k puede aproximarse a través de:

$$r_{jk} = \text{Cos} \theta_{jk}(r) \approx \text{Cos} \theta_{jk}(q)$$

es entonces posible construir un gráfico sobre el cual se represente aproximadamente la estructura de las varianzas y las covarianzas:

FIGURA N° 9
Representación grafica de la correlación entre dos variables



2.2.7. Contribución de la variable a la componente

Partiendo de la relación: $\sum_{j=1}^p \varphi_{j\alpha}^2 = \lambda_{\alpha}$, se define como contribución de la variable j a

la variabilidad de la componente α , a la expresión:

$$CV_{j\alpha} = \frac{\varphi_{j\alpha}^2}{\lambda_{\alpha}} \cdot 100\% = v_{j\alpha} \cdot 100\%$$

Para la interpretación de los resultados deben tomarse en cuenta solamente aquellas variables con mayor contribución a la variabilidad de la componente.

2.2.8 Contribución de la componente a la variable.

Partiendo de la relación $\sum_{\alpha=1}^p \varphi_{j\alpha}^2 = (n-1)s_j^2$, se define como contribución de la componente α a la variabilidad de la variable j , a la expresión:

$$CF_{\alpha} V_j = \frac{\varphi_{j\alpha}^2}{(n-1)s_j^2} \cdot 100\%$$

la cual, como vimos anteriormente, es igual a $(\text{Corr}(X^j, \psi^\alpha))^2$, es decir, el coeficiente de determinación entre X^j y ψ^α .

2.3 Análisis de Correspondencia Simple

El Análisis de correspondencia simple, es un método del análisis multivariante cuyo objetivo consiste fundamentalmente en explicar la asociación existente entre dos variables cualitativas a través de dispositivos gráficos constituidos a manera de diagrama de dispersión, denominados planos factoriales. Sobre estos gráficos se representan simultáneamente los perfiles de las modalidades de las filas y de las columnas de una tabla de contingencia. Y son construidos como planos de mejor ajuste a los perfiles de las modalidades filas o columnas, en el sentido de los mínimos cuadrados, obteniéndose así que el primer plano factorial garantiza una representación óptima, debido a que por construcción sus direcciones captan la máxima proporción posible de la asociación existente entre las variables cualitativas consideradas; ello queda reflejado en las representaciones en las que aparecen agrupados perfiles filas similares (columnas), separándose de aquéllos respecto de los cuales difieren.

2.3.1. Análisis general ponderado con una métrica cualquiera.

El análisis general ponderado es una generalización del análisis general, en la que se consideran los puntos provistos de masas, y se considera además una métrica cualquiera. Esta generalización es de utilidad en el estudio del análisis de correspondencia simple y múltiple.

El punto de partida lo constituye la matriz de datos X_{np}

$$X_{np} = \begin{pmatrix} X_1^t \\ X_2^t \\ \cdot \\ \cdot \\ X_n^t \end{pmatrix}$$

considerada la nube de puntos en R^p , provistos de la masa $(w_1, w_2, w_3, \dots, w_n)$ y de una métrica $M_{p \times p}$ (si $w_i = 1 \forall i$ y $M=I$, obtenemos como caso particular el análisis general).

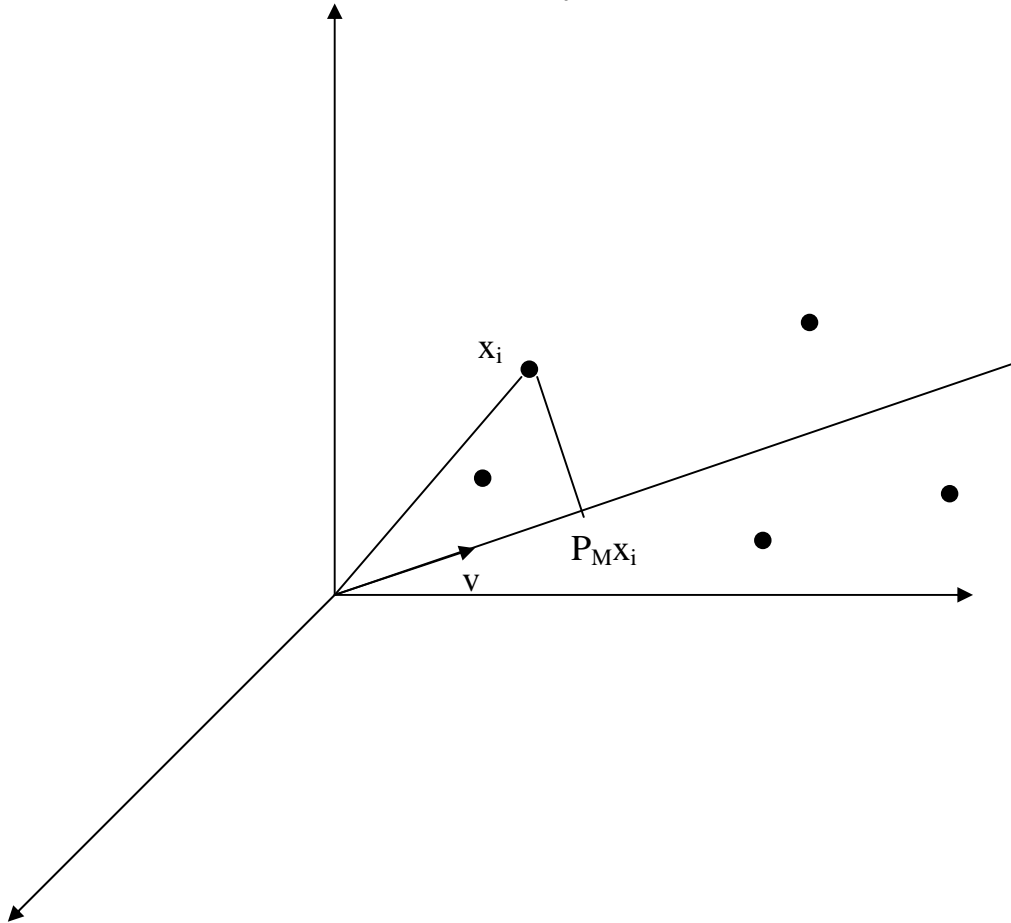
Nuestro objetivo es encontrar el subespacio de R^p de dimensión q ($q < p$), con direcciones M -ortonormalizadas, que mejor se ajuste a la nube $N(I)$. En primer lugar hallaremos el subespacio de dimensión 1, es decir, la recta que mejor se ajuste a la nube. Denotaremos por v al vector M -unitario director de la recta de mejor ajuste, y para su determinación utilizaremos el criterio clásico de los mínimos cuadrados, es decir, se trata de hallar el vector v tal que:

$$\min \sum w d_M^2(x_i, P_M x_i)$$

donde $P_M x_i$ es la proyección M -ortogonal de x_i sobre la recta cuyo vector director es v

FIGURA N° 10

Proyección en \mathbb{R}^3



La distancia cuadrado del origen al punto x_i

$$d_M^2(O, x_i) = x_i^t M x_i$$

es un valor fijo. La distancia cuadrado del origen a la proyección $P_M x_i$,

$$d_M^2(O, P_M x_i) = (x_i^t M v)^2 = (x_i^t M v)^t (x_i^t M v) = v^t M x_i x_i^t M v$$

y por el teorema de Pitágoras

$$d_M^2(O, x_i) = d_M^2(O, P_M x_i) + d_M^2(x_i, P_M x_i)$$

de donde:

$d_M^2(x_i, P_{Mx_i}) = d_M^2(O, x_i) - d_M^2(O, P_{Mx_i})$ así que minimizar la suma $\sum w_i d_M^2(x_i, P_{Mx_i})$, resulta equivalente a maximizar la suma:

$$\begin{aligned} \sum w d_M^2(O, P_{Mx_i}) &= \sum w (v^t M x_i x_i^t M v) = v^t M \sum w (x_i x_i^t) M v \\ &= v^t M X^t W X M v \end{aligned}$$

siendo $W_{n \times n}$ la matriz diagonal de pesos. De tal manera que nuestro problema consiste en maximizar la forma cuadrática $v^t A v$, siendo $A = M X^t W X M$, con la condición $v^t M v = 1$. (nuevamente, si $w_i = 1 \forall i$ y $M=I$, se obtiene $v^t X^t X v$, que es la forma cuadrática a maximizar en el análisis general)

Para maximizar esta forma cuadrática dado la restricción de que v sea un vector M -unitario, construimos el lagrangiano:

$$L = v^t A v - \lambda (v^t M v - 1)$$

derivamos respecto de v , e igualamos al vector nulo:

$$2 A v - 2 \lambda M v = 0$$

obteniéndose la ecuación:

$$A v = \lambda M v$$

Así que el vector buscado v es un autovector de la matriz $S = M^{-1} A = X^t W X M$ asociado con el autovalor λ . Esto quiere decir que una condición necesaria para que v defina un máximo es que sea un autovector de S . Dentro del conjunto de autovectores de esa matriz nos interesamos por aquel que efectivamente maximice la forma cuadrática:

$$\underset{v}{\text{Max}} (v^t A v) = \underset{v}{\text{Max}} (v^t \lambda M v) = \underset{v}{\text{Max}} (\lambda v^t M v) = \underset{v}{\text{Max}} (\lambda)$$

De modo que el vector director de la recta que mejor se ajusta a la nube de puntos es el autovector de S , que denotaremos por v^1 , asociado con el mayor autovalor, que denotaremos por λ_1

En la segunda etapa del procedimiento hallaremos el subespacio de dimensión 2, es decir, el plano que mejor se ajuste a la nube. Este plano tendrá como primera dirección la definida por v^1 , mientras que la segunda quedará determinada por el vector M -unitario v^2 , M -ortogonal a v^1 que minimice la suma de cuadrados $\sum w_i d_M^2(x_i, P_{Mx_i})$, donde ahora P_{Mx_i} es la proyección M -ortogonal de x_i sobre el plano.

El Lagrangiano que toma en cuenta tales restricciones es ahora:

$$L = v^t A v - \lambda (v^t M v - 1) - \gamma (v^t M v^1)$$

Si derivamos respecto de v , e igualamos al vector nulo:

$$2 A v - 2 \lambda M v - \gamma M v^1 = 0$$

Premultiplicando por $(v^1)^t$ se obtiene que:

$$2(v^1)^t A v - 2 \lambda (v^1)^t M v - \gamma (v^1)^t M v^1 = -\gamma = 0$$

ya que $(v^1)^t M v = 0$. Entonces, al igual que en la primera etapa, queda que $A v = \lambda M v$; por tanto, $Sv = \lambda v$. En consecuencia las direcciones del plano que mejor se ajustan a la nube de puntos en R^p están definidas por los vectores v^1 y v^2 , autovectores M -ortonormalizados de la matriz S asociados con sus dos mayores autovalores λ_1 y λ_2 .

Reiterando el mismo procedimiento se obtiene que:

“Una base M-ortonormalizada de dimensión q que mejor se ajusta a la nube de puntos en R^p , está constituida por los q autovectores v^1, v^2, \dots, v^q correspondientes a los q mayores autovalores $\lambda_1, \lambda_2, \dots, \lambda_q$ de la matriz $S=M^{-1}A=X^tWXM$ ”.

2.3.2 Definiciones y Notaciones

- **Tabla de contingencia:** Es una tabla de frecuencias que describe el comportamiento de n objetos caracterizados simultáneamente mediante dos variables cualitativas. Las categorías o modalidades de las dos variables consideradas se describen a través de dos conjuntos I y J de cardinales respectivos n y p. El elemento genérico en la posición (i,j) de la tabla de contingencia, que se denota k_{ij} , describe el número de elementos que poseen simultáneamente la modalidad i del conjunto I y la j del conjunto J, la tabla es entonces:

$$K_{n \times p} = (k_{ij})$$

Algunas características de interés obtenidas después de hacer operaciones matriciales son:

- Totales por fila de la matriz K: $(K_{1.}, K_{2.}, \dots, K_{n.}) = 1_p^t K_{p \times n}^t$
- Totales por columna de la matriz K: $(K_{.1}, K_{.2}, \dots, K_{.p}) = 1_n^t K_{n \times p}$
- Matriz de Frecuencias Relativas F: $F_{n \times p} = (f_{ij} = k_{ij} / N)$
- Totales por fila de la matriz F: $r_{i \times n}^t (f_{1.}, f_{2.}, \dots, K_{n.}) = 1_p^t F_{p \times n}^t$
- Totales por columna de la matriz F: $c_{1 \times p}^t (f_{.1}, f_{.2}, \dots, f_{.p}) = 1_n^t F_{n \times p}$
- Perfil de la i-ésima fila: $R_{i \times p}^t = (f_{i1}/f_{i.}, f_{i2}/f_{i.}, \dots, f_{ip}/f_{i.}) \quad i=1,2,\dots,n$
- Perfil de la j-ésima columna: $C_{i \times p}^t = (f_{1j}/f_{.j}, f_{2j}/f_{.j}, \dots, f_{pj}/f_{.j}) \quad j=1,2,\dots,p$

- Matriz diagonal de pesos fila: $D_n = (\text{diag}(r.))$
- Matriz diagonal de pesos columna: $D_p = (\text{diag}(c.))$
- Matriz de perfiles fila: $R_{n \times p} = (f_{ij}/f_i) = D_n^{-1} F$
- Matriz de perfiles columna: $C_{p \times n} = (f_{ij}/f_j) = D_p^{-1} F^t$

2.3.3 Nube de puntos

2.3.3.1 En R^p

La nube de puntos en R^p , $N(I)$, es el conjunto de perfiles fila $R_1^t, R_2^t, \dots, R_n^t$, dotados de las masas $r^t = (f_{1.}, f_{2.}, \dots, f_{n.})$ y la distancia chi-cuadrado:

$$d^2(i, s) = \sum_{j=1}^p \frac{1}{f_j} \left(\frac{f_{ij}}{f_i} - \frac{f_{sj}}{f_s} \right)^2$$

$$(R_i - R_s)^t D_p^{-1} (R_i - R_s)$$

El centro de gravedad de la nube $N(I)$:

$$G_{1 \times p}^t = r^t R = r^t D_n^{-1} F = 1_n^t F = c^t = (f_{.1}, f_{.2}, \dots, f_{.p})$$

2.3.3.2 En R^n

La nube de puntos en R^n , $N(J)$, es el conjunto de perfiles columna $C_1^t, C_2^t, \dots, C_n^t$, dotados de las masas $c^t = (f_{.1}, f_{.2}, \dots, f_{.p})$ y la distancia chi-cuadrado:

$$d^2(i, s) = \sum_{j=1}^n \frac{1}{f_i} \left(\frac{f_{ij}}{f_j} - \frac{f_{ik}}{f_k} \right)^2$$

$$(C_j - C_k)^t D_n^{-1} (C_j - C_k)$$

El centro de gravedad de la nube N(J):

$$H_{1 \times n}^t = c^t C = c^t D_p^{-1} F = 1_p^t F^t = r^t = (f_{i.}, f_{2.}, \dots, f_{n.})$$

2.3.4. Inercia de la nube de puntos:

2.3.4.1 En R^p

$$\begin{aligned} \text{Inercia (N(I))} &= \sum_{i=1}^n f_i d^2(R_i, G) \\ &= \sum_{i=1}^n f_i (R_i - G)^t D_p^{-1} (R_i - G) \\ &= \sum_{i=1}^n \sum_{j=1}^p \frac{f_i}{f_{.j}} \left(\frac{f_{ij}}{f_i} - f_{.j} \right)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^p \frac{1}{f_{.j} f_i} (f_{ij} - f_{.j} f_i)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^p \frac{f_{ij}^2}{f_{.j} f_i} - 1 \end{aligned}$$

Que es una medida global del parecido de la nube de perfiles fila y su centro de gravedad

2.3.4.2 En \mathbb{R}^n

$$\begin{aligned}
 \text{Inercia } (N(J)) &= \sum_{j=1}^p f_{.j} d^2(C_j, H) \\
 &= \sum_{j=1}^p f_{.j} (C_j - H)^t D_n^{-1} (C_j - H) \\
 &= \sum_{i=1}^n \sum_{j=1}^p \frac{f_{.j}}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - f_{i.} \right)^2 \\
 &= \sum_{i=1}^n \sum_{j=1}^p \frac{1}{f_{.j} f_{i.}} (f_{ij} - f_{.j} f_{i.})^2 \\
 &= \sum_{i=1}^n \sum_{j=1}^p \frac{f_{ij}^2}{f_{.j} f_{i.}} - 1
 \end{aligned}$$

Que es una medida global del parecido de la nube de perfiles columna y su centro de gravedad

2.3.5 Análisis en \mathbb{R}^p

Se trata de hallar el subespacio que mejor se ajusta a la nube $N(I)$ constituida por las filas de la matriz:

$$R_{n \times p} = D_n^{-1} F$$

provistas de los pesos $r^t = (f_{1.}, f_{2.}, \dots, f_{n.})$ y la distancia chi-cuadrado:

$$\begin{aligned}
 d^2(i, s) &= \sum_{j=1}^p \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{sj}}{f_{s.}} \right)^2 \\
 &= (R_i - R_s)^t D_p^{-1} (R_i - R_s)
 \end{aligned}$$

por consiguiente la matriz diagonal de pesos será $D_n = \text{diag}(f_{1.}, f_{2.}, \dots, f_{n.})$ y la métrica será

$$D_p^{-1} = \text{diag}(f_{1.}^{-1}, f_{2.}^{-1}, \dots, f_{n.}^{-1}).$$

De acuerdo con los resultados del análisis general ponderado, la matriz a diagonalizar será entonces:

$$S = R^t D_n R D_p^{-1} = F^t D_n^{-1} F D_p^{-1}$$

En consecuencia:

“Las direcciones del subespacio de dimensión q ($q < p$) que mejor ajusta a la nube de perfiles fila $R = D_n^{-1} F$, están definidas por los autovectores v^1, v^2, \dots, v^q , vectores D_p^{-1} ortonormalizados asociados a los q mayores autovalores $\lambda_1, \lambda_2, \dots, \lambda_q$ de la matriz S ”

2.3.5.1 Coordenadas de proyección de los perfiles fila

$$\psi^\alpha = R D_p^{-1} V^\alpha \quad \alpha = 1, 2, 3, \dots, \min(n, p)$$

para los efectos de interpretación consideremos las siguientes propiedades:

- La traza de la matriz a diagonalizar es tal que:

$$\sum_{\alpha=1}^p \lambda_\alpha = \text{Traza}(S) = \text{Traza}(F^t D_n^{-1} F D_p^{-1}) = \sum_{i=1}^n \sum_{j=1}^p \frac{f_{ij}^2}{f_{.j} f_{i.}}$$

- Los autovalores de la matriz S son tales $0 \leq \lambda \leq 1$
- El máximo autovalor de la matriz S es $\lambda_1 = 1$; se dice que λ_1 es un autovalor trivial y se descarta del análisis puesto que las coordenadas de proyección sobre el eje asociado son todas iguales a uno.
- Las coordenadas de proyección ψ^α sobre el eje α son tales que:

$$\sum_{i=1}^n f_i \psi_{i\alpha}^2 = \lambda_\alpha$$

por consiguiente, el aporte de la i -ésima modalidad fila a la asociación captada por el factor α . se define mediante:

$$CF_i F_\alpha = \frac{f_i \cdot \psi_{i\alpha}^2}{\lambda^\alpha} \quad i=1,2,\dots,n$$

- Las coordenadas de proyección de la i-ésima modalidad fila sobre los p factores es tal que:

$$\sum_{\alpha=1}^p \psi_{i\alpha}^2 = d^2(R_i, G)$$

por consiguiente, el aporte del factor α al posicionamiento de la i-ésima modalidad fila se define mediante:

$$CF_\alpha F_i = \frac{\psi_{i\alpha}^2}{\sum_{\alpha=1}^p \psi_{i\alpha}^2} \quad \alpha=1,2,\dots,nf$$

2.3.6 Análisis en R^n

De forma similar al análisis anterior se obtiene que:

“Las direcciones del subespacio de dimensión q ($q < p$) que mejor ajusta a la nube de perfiles columna $C = D^{-1} F^t$, están definidas por los autovectores u^1, u^2, \dots, u^q , vectores D^{-1}_n ortonormalizados asociados a los q mayores autovalores $\lambda_1, \lambda_2, \dots, \lambda_q$ de la matriz

$$T = F D^{-1}_p F^t D^{-1}_n$$

Por consiguiente las coordenadas de proyección de los perfiles columna:

$$\varphi^\alpha = C D_n^{-1} U^\alpha$$

y las contribuciones se expresan se definen como sigue:

- Contribución de la j-ésima modalidad columna al factor α

$$CC_j F_\alpha = \frac{f_{.j} \phi_{j\alpha}^2}{\lambda^\alpha}$$

- Contribución del factor α a la representación de la j-ésima modalidad columna

$$CF_\alpha C_j = \frac{\phi_{j\alpha}^2}{\sum_{\alpha=1}^p \phi_{i\alpha}^2}$$

2.4 Modelos Loglineales

2.4.1 Modelo saturado para dos variables

Sea una muestra de N individuos sobre los que se dispone de observaciones correspondientes a dos variables cualitativas, X e Y, con I y J categorías, respectivamente. El modelo que se postula para explicar la frecuencia en cada celda de la tabla de contingencia entre las dos variables es:

$$\ln(f_{ij}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \quad \forall i = 1, \dots, I \quad \forall j = 1, \dots, J$$

con las restricciones:

$$\sum_{i=1}^I \lambda_i^X = 0 \quad \sum_{j=1}^J \lambda_j^Y = 0 \quad \text{y} \quad \sum_{i=1}^I \lambda_{ij}^{XY} = \sum_{j=1}^J \lambda_{ij}^{XY} = 0$$

donde:

- f_{ij} es la frecuencia en la celda (i,j). Es decir, el numero de individuos de la muestra que presentan, simultáneamente, las cualidades correspondientes a las categorías i-ésima y j-ésima de X e Y, respectivamente.
- μ es el efecto debido al promedio de las frecuencias en todas las celdas de la tabla.

- λ_i^X es el efecto debido a la i-ésima categoría de la variable X
- λ_j^Y es el efecto debido a la j-ésima categoría de la variable Y
- λ_{ij}^{XY} es el efecto debido a la interacción de las categorías i y j de X e Y, respectivamente.

Este modelo, denominado modelado saturado, incluye todos los posibles efectos y, por tanto, reproducirá exactamente las frecuencias observadas en la tabla de contingencia.

2.4.2 Modelo saturado general

La generalización del modelo saturado al caso de k variables X_1, \dots, X_k , con I_1, \dots, I_k categorías, respectivamente, es:

$$\ln(f_{i_1 i_2 \dots i_k}) = \mu + \lambda_{i_1}^{X_1} + \dots + \lambda_{i_k}^{X_k} + \lambda_{i_1 i_2}^{X_1 X_2} + \dots + \lambda_{i_1 i_k}^{X_1 X_k} + \dots + \lambda_{i_{k-1} i_k}^{X_{k-1} X_k} \\ + \dots + \lambda_{i_1 i_2 i_k}^{X_1 X_2 X_k} + \dots + \lambda_{i_1 i_2 \dots i_k}^{X_1 X_2 \dots X_k}$$

con la restricción:

$$\forall \{i_{r_1}, \dots, i_{r_s}\} \subset \{i_1, \dots, i_k\} \quad \sum_{i_{r_j}} \lambda_{i_{r_1} \dots i_{r_s}}^{X_{r_1} \dots X_{r_s}} = 0 \quad j = 1, \dots, s$$

Es decir, para explicar la frecuencia observada en la celda (i_1, \dots, i_k) , determinada por el cruce de la categoría i_1, \dots, i_k de X_1, \dots, X_k , respectivamente, el modelo incluye todos los efectos principales (los debidos a la categoría correspondiente de cada una de las k variables), los efectos de la interacción de cualquier par de variables, de cualquier terna, y así sucesivamente, hasta el efecto de la interacción de las k variables.

2.4.3 Otros tipos de modelos

Cualquier subconjunto de los efectos considerados en el modelo saturado determina en si mismo un modelo. Un caso particular es el modelo de independencia, en el que se supone que los efectos debidos a las interacciones de las variables son nulos. Es decir, el modelo de independencia únicamente incluye los efectos principales:

$$\ln(f_{i_1 i_2 \dots i_k}) = \mu + \lambda_{i_1}^{X_1} + \dots + \lambda_{i_k}^{X_k}$$

De entre todos los posibles modelos, una clase de especial relevancia es la de los modelos jerárquicos. Un modelo jerárquico esta caracterizado porque, si el efecto de la interacción de un conjunto de variables está presente, entonces todos los efectos de orden inferior correspondientes a interacciones entre las variables del conjunto también lo están. En particular, el modelo saturado y el de independencia se encuentran dentro de esta clase de modelos.

2.4.4 Calculo de los efectos.

En el modelo saturado cada efecto se obtiene a partir de los datos observados. Por ejemplo, en el modelo con dos variables:

$$\ln(f_{ij}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

μ es el promedio de los logaritmos de las frecuencias en todas las celdas:

$$\mu = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \ln(f_{ij})$$

λ_i^X es el promedio de los logaritmos de las frecuencias en la fila i eliminando el efecto

$$\text{promedio: } \lambda_i^X = \frac{1}{J} \sum_{j=1}^J \ln(f_{ij}) - \mu$$

λ_j^Y es el promedio de los logaritmos de las frecuencias en la columna j eliminando el efecto

$$\text{promedio: } \lambda_j^Y = \frac{1}{I} \sum_{i=1}^I \ln(f_{ij}) - \mu$$

λ_{ij}^{XY} es el logaritmo de la frecuencia en la celda (i,j) eliminando los efectos de orden

inferior: $\lambda_{ij}^{XY} = \ln(f_{ij}) - (\mu + \lambda_i^X + \lambda_j^Y)$. Puesto que, para cada una de las variables, la suma

de los efectos principales es nula: $\sum_{i=1}^I \lambda_i^X = 0$ y $\sum_{j=1}^J \lambda_j^Y = 0$, para obtener todos los efectos

bastará con calcular los I – 1 primeros, en el caso de la variable X, y los J – 1 primeros, en

el de la variable Y. Por otro lado $\sum_{i=1}^I \lambda_{ij}^{XY} = \sum_{j=1}^J \lambda_{ij}^{XY} = 0$. Luego para obtener todos los efectos

debidos a la interacción de las variables X e Y bastará con calcular los (I – 1)(J – 1) primeros.

En la práctica, dado que es posible que, en una o mas celdas de la tabla, la frecuencia observada sea igual a cero, para evitar la imposibilidad de construir el modelo en dicha situación (al tener que aplicar la función logaritmo sobre el valor cero), a todas las frecuencias de la tabla se les sumara una determinada cantidad que distorsione poco la frecuencia observada en la tabla. (algunos Software usan por defecto el valor 0,5).

2.4.5 Bondad de ajuste, prueba de los k efectos y prueba de asociación parcial

Al considerar modelos no saturados, el modelo ajustado será adecuado cuando las frecuencias observadas y las estimadas mediante el modelo sean muy parecidas. Los estadísticos chi-cuadrado de Pearson y razón de verosimilitud chi-cuadrado comparan la frecuencia observada y la esperada mediante diferencia entre ellas, en el primer caso, y

mediante el cociente, en el segundo. Si el p-valor asociado al estadístico es menor que α se rechazará la hipótesis nula de que el modelo es adecuado al nivel de significación α

El interés de los modelos loglineales reside en determinar los efectos significativos, dado que, a partir de ellos, será posible interpretar las relaciones entre las variables. La prueba de los k efectos permite contrastar tanto la hipótesis nula de que todos los efectos debidos a las interacciones de orden k o superior son iguales a cero, como la hipótesis nula de que todos los efectos debidos a las interacciones de orden k son iguales a cero. En ambos casos, los estadísticos de contraste son la chi-cuadrado de Pearson y la razón de verosimilitud chi-cuadrado. En el primer caso ambos estadísticos se calculan para el modelo obtenido al eliminar los efectos de orden k o superior. Si dicho modelo es adecuado, se podrá concluir que los efectos de orden k o superior no son significativamente distintos de cero. En el segundo caso, los estadísticos chi-cuadrado de Pearson y razón de verosimilitud chi-cuadrado se calculan como la diferencia entre el estadístico asociado al modelo sin efectos de orden k o superior y el asociado al modelo sin efectos de orden $k + 1$ o superior. Si la probabilidad asociada a la diferencia es menor que α se rechazará la hipótesis nula al nivel de significación α .

La prueba de los k efectos contrasta la hipótesis general de que los efectos de orden k son nulos. Rechazar dicha hipótesis no implica que, en particular, cada uno de ellos sea necesariamente distinto de cero. Para contrastar la hipótesis nula de que un efecto en particular es nulo, la prueba adecuada es la de asociación parcial. Si el p-valor asociado al

estadístico de contraste (en este caso, la chi-cuadrado parcial) es menor que α se rechazará la hipótesis nula al nivel de significación α .

CAPITULO III

ANÁLISIS DESCRIPTIVO UNIVARIANTE

Sobre una muestra de 9600 viviendas que comprenden 10259 hogares y 40852 miembros, se realizó un análisis estadístico a las principales variables, con el fin de obtener una primera descripción de las observaciones más relevantes.

Empezaremos describiendo la composición de las viviendas con respecto al número de hogares y miembros que habitan en la misma. En cuadros número uno y dos podemos apreciar que el 94% de las viviendas son principales, esto quiere decir, que la mayoría de viviendas encuestadas fueron aquellas donde los hogares la utilizan para su desenvolvimiento diario. Además en un 91.6% de viviendas existe sólo un hogar; este resultado lo que indica que en la mayoría de viviendas hay un solo fondo para cubrir las necesidades de la misma. Por último observamos que el total de miembros promedio por hogar y vivienda es de: 4.22 y 4.65 personas respectivamente; estos resultados son muy parecidos y esto se debe que en casi todas las viviendas hay solo un hogar. (ver cuadro N° 3)

CUADRO N° 1

Distribución de frecuencias de Uso de la Vivienda

Uso de la Vivienda	Frecuencia	Porcentaje	Porcentaje Acumulado
Principal	9668	94.2	94.2
Auxiliar	99	1.0	95.2
Descanso	54	0.5	95.7
Otros	438	4.3	100.0
Total	10259	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

CUADRO N° 2

Distribución de frecuencias de Número de hogares por vivienda

Hogares por Viv.	Frecuencia	Porcentaje	Porcentaje Acumulado
1	8588	91.6	91.6
2	681	7.3	98.9
3	103	1.1	100.0
Total	9372	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

Para tener una idea de la relación entre número de miembros en la vivienda y el tamaño de ésta se calculó para cada vivienda la razón: número de miembros por dormitorios, esto con la finalidad de obtener una medida del grado de comodidad o

hacinamiento promedio en las viviendas. Observemos entonces (Cuadro N° 4) que en general hay cierto grado de comodidad ya que en promedio hay aproximadamente 2 miembros por habitación. Por otro lado observando el cuadro N° 5 notamos que en el 78 % de las viviendas la relación miembros por dormitorio es menor a 2 pers/dorm; sólo observamos un hacinamiento del 5% (si asumimos como criterio viviendas hacinadas si existe miembros por dormitorio mayor que 4)

CUADRO N° 3

Estadísticas básicas de Miembros por hogar

	N	Mínimo	Máximo	Suma	Media	Desv. Est.
Miembros por Hogar	10259	1	21	43390	4.22	2.2513
Miembros por Vivienda	9372	1	21	43390	4.62	2.25

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

CUADRO N° 4

Estadísticas básicas de Miembros por dormitorio

	N	Mínimo	Máximo	Media	Desv. Est.
Miembros por Dormitorio	9195	0.13	14	1.72	1.11

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

CUADRO N° 5

Distribución de frecuencias de los miembros por dormitorio

Miembros por Dorm.	Frecuencia	Porcentaje	Porcentaje Acumulado
0 - 2	7190	78.2	78.2
2 - 4	1722	18.7	96.9
4 - 6	214	2.3	99.2
6 - 8	53	0.6	99.8
8 - 10	9	0.1	99.9
10 - 12	4	0	100.0
12 - 14	3	0	100.0
Total	9195	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

En los cuadros números 6, 7, 8, 9 se describe las características físicas de las viviendas, mas relevantes, por ejemplo, podemos observar: que un 7% de las viviendas son piezas en casa de vecindad o ranchos, y un 97% son viviendas mas aptas para vivir. Por otro lado los materiales predominante para paredes es el bloque (90%), para el piso existe una división entre materiales de lujo 48,7% y cemento rustico 49,4% y en materiales de techo también existe dos grupos, las vivienda con techo de platabanda con un 51% y los techo con laminas metálicas 38%

CUADRO N° 6**Distribución de frecuencias de Tipo de Vivienda**

Tipo de Vivienda	Frecuencia	Porcentaje	Porcentaje Acumulado
Quinta y/o casa	6314	67.4	67.4
Apartamento en edificio con ascensor	1790	19.1	86.5
Apartamento en edificio sin ascensor	599	6.4	92.9
Local no construido para vivienda	12	0.1	93.0
Apartamento en casa o quinta	65	0.7	93.7
Pieza en casa de vecindad	4	0.0	93.7
Vivienda rustica (rancho)	394	4.2	97.9
Rancho campesino	186	2.0	99.9
Otros	8	0.1	100.0
Total	9372	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

CUADRO N° 7**Distribución de frecuencias de Material en las paredes**

Material.	Frecuencia	Porcentaje	Porcentaje Acumulado
Bloque o ladrillo frizado	7782	83.0	83.0
Bloque o ladrillo sin frizar	696	7.4	90.5
Concreto (prefabricado)	244	2.6	93.1
Madera aserrada	60	0.6	93.7

CUADRO N° 7 (CONTINUACIÓN)

Material.	Frecuencia	Porcentaje	Porcentaje Acumulado
Adobe, tapia o bahareque frizado	240	2.6	96.3
Adobe, tapia o bahareque sin frizar	63	0.7	96.9
Otros (caña, palma, tabla y similares)	287	3.1	100.0
Total	9372	100	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

CUADRO N° 8**Distribución de frecuencias de Material del piso**

Material.	Frecuencia	Porcentaje	Porcentaje Acumulado
Mosaico, granito, vinil, cerámica, ladrillo, terracota y similares	4566	48.7	48.7
Cemento	4629	49.4	98.1
Tierra	155	1.7	99.8
Otros (tablas, tablones y similares)	22	0.2	100.0
Total	9372	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

CUADRO N° 9

Distribución de frecuencias de Material del techo

Material.	Frecuencia	Porcentaje	Porcentaje Acumulado
Platabanda	4837	51.6	51.6
Teja	312	3.3	54.9
Asbesto	583	6.2	61.2
Laminas metálicas (zinc y similares)	3575	38.1	99.3
Otros (caña, tabla y similares)	65	0.7	100.0
Total	9372	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

También observamos que la mayoría de las viviendas, aproximadamente el 60% necesitan por lo menos reparaciones pequeñas y tienen una apariencia modesta (48.4%).
Estos últimos resultados los podemos apreciar en los dos siguientes cuadros

CUADRO N° 10

Distribución de frecuencias de Estado de la vivienda

Estado de la vivienda	Frecuencia	Porcentaje	Porcentaje Acumulado
No necesita reparaciones	3918	41.8	41.8
Necesita reparaciones pequeñas	3167	33.8	75.6
Necesita reparaciones importantes	2287	24.4	100.0
Total	9372	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

CUADRO N° 11

Distribución de frecuencias de Apariencia de la vivienda

Apariencia de la vivienda	Frecuencia	Porcentaje	Porcentaje Acumulado
Lujosa	55	0.6	0.6
Muy Buena	649	6.9	7.5
Buena	3425	36.5	44.1
Modesta	4539	48.4	92.5
Precaria	704	7.5	100.0
Total	9372	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

Las actividades económicas más comunes en las viviendas son el comercio al por menor (bodegas) con un 37% Elaboración de productos alimenticios y bebidas (comida rápida) con un 17% y la fabricación de prendas de vestir con un 10%. Las demás actividades económicas tienen un porcentaje menor al 10%. (ver Apéndice C)

Ahora se analizarán características educativas, sociales y laborales del jefe del hogar, ya que, se asume que sus características socioeconómicas son una buena representación de las características de los demás miembros del hogar.

En el cuadro número diez podemos observar que el porcentaje de Analfabetismo en los miembros jefes del hogar es aproximadamente 5 %.

CUADRO N° 12

Distribución de frecuencias de Analfabetismo

	Frecuencia	Porcentaje	Porcentaje Acumulado
Sabe leer y Escribir	8774	95.1	95.1
No sabe leer ni escribir	450	4.9	400.0

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

El cuadro anterior se puede complementar con el cuadro numero 13, que refleja el nivel educativo de los jefes del hogar, en esta podemos apreciar que se mantiene un 5% de miembros jefes del hogar que no tienen nivel educativo (analfabetas); por otro lado llama mucho la atención que solo el 20% de los jefes de hogares tienen un nivel educativo universitario o postgrado, y la mayoría de las cabezas del hogar tienen un nivel educativo Básico, Hasta 9no grado, con un 47 %.

CUADRO N° 13

Distribución de frecuencias de Nivel educativo

Nivel Educativo	Frecuencia	Porcentaje	Porcentaje Acumulado
Preescolar	9	0.1	0.1
Básica	4351	47.2	47.3
Media Diversificada	1871	20.3	67.6
Técnico Superior	526	5.7	73.3

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

CUADRO N° 13 (CONTINUACIÓN)

Nivel Educativo	Frecuencia	Porcentaje	Porcentaje Acumulado
Universitario	1631	17.7	90.9
Postgrado	307	3.3	94.3
INCE y similares	33	0.4	94.6
Sin nivel	496	5.4	100.0
Total	9224	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Prepuesto Familiares

El 60 % de los miembros de los hogares encuestados no poseen un seguro de previsión social, solo el 30 % goza del seguro social obligatorio y un 6% usa los servicios de un seguro privado, ya sea pagado por cuenta propia o por la empresa donde trabaja (ver cuadro numero 14)

CUADRO N° 14

Distribución de frecuencias de Sistema de previsión Social

Sistema de previsión social	Frecuencia	Porcentaje	Porcentaje acumulado
Seguro Social obligatorio	12284	30,06	30.06
Seguro privado de asistencia médica	1161	2,84	32.9
Seguro colectivo de asistencia médica	1218	2,98	35.88
Otros (IPASME, IPSFA, IMPRES, etc)	1444	3,53	39.41
Ninguno	24745	60,57	100.00
Total	40852	100	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Prepuesto Familiares

En el cuadro número quince observamos que el transporte de pasajeros es el medio de traslado mas común (48 %), le sigue el vehículo automotor propio con un (36 %). Podemos concluir que la mayoría no posee un carro propio para su traslado (64 %).

CUADRO N° 15

Distribución de frecuencias de Medio de transporte

Nivel Educativo	Frecuencia	Porcentaje	Porcentaje Acumulado
Vehículo automotor propio	3338	36.2	36.2
Vehículo automotor de terceros	245	2.7	38.9
Transporte de pasajeros	4431	48.0	86.9
Otros medios de transporte propio no automotor	213	2.3	89.2
Transporte de la empresa	155	1.7	90.9
Transporte publico subsidiado	17	0.2	91.1
Metro, metrobus	115	1.2	92.3
No usa habitualmente transporte	711	7.7	100.0
Total	9225	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

Ahora analizaremos una serie de tablas que describen la situación laboral del miembro jefe del hogar. Primero observamos que el 15 % de los jefes del hogar no tienen trabajo, es decir existe un 15 % de desempleo a nivel de cabezas del hogar. (ver cuadro numero 16)

CUADRO N° 16

Distribución de frecuencias para la Posesión de Trabajo

Posee Trabajo	Frecuencia	Porcentaje	Porcentaje Acumulado
Si	7825	84.8	84.8
No	1399	15.2	100.0
Total	9225	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

Del 15 % de desempleados, el 12 % ha trabajado antes y sólo el 3% está buscando trabajo. El número grande de desempleados que no busca trabajo parece contradictorio, pero esto se debe que la mayoría de estas personas esta jubilado o pensionado.

CUADRO N° 17

Distribución de frecuencias para la Búsqueda de Trabajo

Posee Trabajo	Frecuencia	Porcentaje	Porcentaje Acumulado
No buscan Porque ya tienen	7825	84.8	84.8
Busca trabajo porque no tiene	304	3.3	88.1
No busca trabajo	1096	11.9	100.0
Total	9225	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Presupuesto Familiares

El 67 % de los miembros jefes del hogar están empleados y no buscan un trabajo adicional, es decir, que sus trabajos son tal que satisfacen sus necesidades. Además podemos observar en el cuadro numero 18, que el 75 % tiene solo un trabajo y aproximadamente el 10 % tiene mas de un trabajo.

CUADRO N° 18

Distribución de frecuencias de Trabajo Adicional

Trabajo Adicional	Frecuencia	Porcentaje	Porcentaje Acumulado
No Posee Trabajo	1389	15.1	15.1
Busca trabajo adicional al que tiene	1639	17.8	32.8
No busca trabajo	6197	67.2	100.0
Total	9225	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Prepuesto Familiares

CUADRO N° 19

Distribución de frecuencias de Número de trabajos

Numero de trabajos	Frecuencia	Porcentaje	Porcentaje Acumulado
No Posee Trabajo	1346	14.6	14.6
Uno	6963	75.5	90.1
Dos	811	8.8	98.9
Tres	98	1.1	99.9
Mas de Tres	7	0.1	100.0
Total	9225	100.0	

Fuente: Elaboración propia, con base en los datos de la II Encuesta de Prepuesto Familiares

CAPITULO IV

ANÁLISIS MULTIVARIANTE

4.1. Análisis de correspondencia

En esta sección se presentaran una serie de análisis de correspondencia simple y múltiple, en los cuales, se describirá las relaciones de variables cualitativas relacionadas con aspectos socioeconómicos como:

- Nivel educativo (del jefe del hogar)
- Tipo de Vivienda
- Apariencia de la vivienda
- Empleo
- Zona geográfica
- Tipo de Seguro

Primero se realizarán varios análisis con distintos pares de variables (Análisis de correspondencia simple), que nos permitirá ir describiendo las relaciones entre pares de variables para luego, con una idea preliminar, realizar un análisis de correspondencia múltiple.

4.1.1 Análisis de correspondencia: Nivel Educativo vs. Tipo de Vivienda

En el cuadro N° 20, observamos el porcentaje de inercia o variabilidad captada por los factores, en este, se aprecia que el primer plano factorial (plano compuesto por los dos primeros factores) capta un 97 % de esta inercia, por lo tanto, podemos afirmar que el primer plano capta de forma satisfactoria la asociación entre las dos variables de estudio (nivel educativo y tipo de vivienda)

CUADRO N° 20

Porcentaje de inercia captada por los factores en análisis de correspondencia simple:

Nivel educativo vs tipo de vivienda

NUMERO	Valor Propio	Porcentaje	Porcentaje acumulado
1	0.2127	84.92	84.92
2	0.0303	12.10	97.02
3	0.0063	2.53	99.55
4	0.0009	0.35	99.90
5	0.0002	0.10	99.99
6	0.0000	0.01	100.00
7	0.0000	0.00	100.00

Otro aspecto importante en el análisis de correspondencia, es la contribución de las modalidades de las diferentes variables a la construcción de los factores, ya que con este índice, podemos sugerir un significado real a los factores. Dicha medida e índices de calidad de representación son mostrados en el siguiente cuadro (N° 21); cabe señalar que este último índice (calidad de representación) es de útil importancia, ya que, nos indica qué tan bien están representadas las modalidades en el plano factorial, y por ende, nos proporciona una idea de la confianza a la hora de interpretar las proyecciones en el plano.

CUADRO N° 21

Contribuciones y calidad de representación para las modalidades de la variable tipo de vivienda en el ACS nivel educativo vs tipo de vivienda

Tipo de Vivienda	Contribuciones					Calidad de Representación				
	1	2	3	4	5	1	2	3	4	5
Casa o Quinta	11.0	13.6	1.6	0.0	0.1	0.85	0.15	0.00	0.00	0.00
Edif. Con Ascensor	55.8	14.6	9.6	0.0	0.0	0.96	0.04	0.00	0.00	0.00
Edif.. sin Ascensor	6.4	1.7	70.8	8.6	0.2	0.73	0.03	0.24	0.00	0.00
Local no apto para vivir	0.3	0.2	0.4	0.8	3.3	0.85	0.09	0.03	0.01	0.01
Anexo	0.2	0.5	3.3	82.2	3.8	0.26	0.11	0.14	0.48	0.01
Pieza en vecindad	0.1	0.4	0.2	0.0	0.8	0.51	0.43	0.04	0.00	0.01
Vivienda Rural (rancho)	14.6	9.1	2.0	2.4	6.3	0.91	0.08	0.00	0.00	0.00
Rancho campesino	11.1	59.9	9.1	0.9	3.0	0.56	0.43	0.01	0.00	0.00
Otros	0.5	0.0	3.1	5.1	82.5	0.70	0.01	0.13	0.03	0.13

Si todas las modalidades de tipo de vivienda tuviesen una contribución homogénea, es decir igual, entonces su contribuciones serian alrededor de 11 %. Por lo tanto siguiendo este criterio podemos afirmar que las modalidades que contribuyen de forma significativa en la formación del primer factor es Edificio con ascensor (55,8 %), Vivienda rural (14,6%), casa o quinta y Rancho campesino con 11 % ambos. Con respecto al segundo factor, se observa que las modalidades con mayor contribución son: Rancho campesino, Edif. con ascensor y casa o quinta con 60, 14,56 y 13. 6 % respectivamente.

Por otro lado observamos que la calidad del plano factorial es bastante buena, ya que la modalidad peor representada por el plano es Anexo con un 37% (26% + 11%). Las demás modalidades tienen una calidad de representación por el plano entre 70 y 90%. Esto

nos garantiza una relativa confianza a la hora de analizar las proyecciones en el plano factorial.

A continuación se presenta el mismo análisis pero para las modalidades de la variable nivel educativo.

CUADRO N° 22

Contribuciones y calidad de representación para las modalidades de la variable Nivel educativo en el ACS nivel educativo vs tipo de vivienda

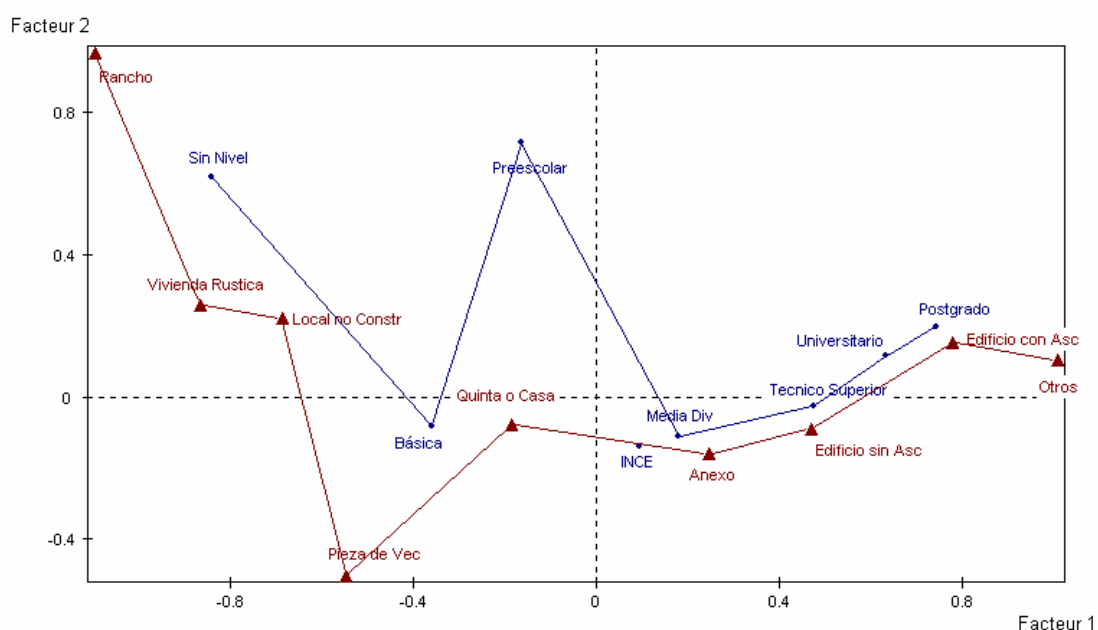
Nivel Educativo	Contribución					Calidad de Representación				
	1	2	3	4	5	1	2	3	4	5
Preescolar	0	1.5	1.4	2	0.8	0.04	0.75	0.15	0.03	0
Básica	28	10.2	12.2	0.9	2.1	0.94	0.05	0.01	0	0
Media Diversificado	3	8.3	31.3	9.7	27.4	0.58	0.23	0.18	0.01	0.01
Técnico Superior	6.3	0.1	25.3	32.3	27.1	0.87	0	0.1	0.02	0
Universitario	35.5	8	9.4	16.6	10	0.96	0.03	0.01	0	0
Postgrado	9.5	4.7	13.1	32.5	32	0.89	0.06	0.04	0.01	0
INCE	0	0.2	0	5.5	0.3	0.2	0.44	0.02	0.31	0
Sin Nivel	17.7	66.9	7.2	0.4	0.3	0.64	0.35	0.01	0	0

Las modalidades de Nivel educativo que contribuyen de forma significativa a la construcción del primer factor son Universitario, Básica y Sin Nivel con 36, 28 y 18 % respectivamente. Hay que tomar en cuenta también a la hora de estudiar el plano factorial la importante contribución de la modalidad Sin nivel, a la construcción del segundo factor (67%), que nos ayudará a interpretar este factor. Con respecto a la calidad de representación se observa en el cuadro, que la modalidad con la representación más baja es Ince con un 64%; además es importante señalar la buena representación del segundo factor hacia las

modalidades preescolar e Ince. A continuación se presentará el primer plano factorial con las proyecciones de las distintas modalidades de nivel educativo y tipo de vivienda.

GRAFICO N° 1

Plano factorial: Nivel Educativo vs Tipo de Vivienda



En el plano observamos como el primer factor es un eje de ordenamiento natural, ya que acomoda de forma ordenada a las distintas modalidades de las dos variables, es decir, a medida que se recorre el primer eje de derecha a izquierda se van encontrando las modalidades que describen un nivel educativo alto y tipo de vivienda de mayor calidad de vida. Un aspecto resaltante es, mientras el primer eje describe un orden natural, el segundo eje indica una separación o discrimina entre los niveles educativos mas bajos (Sin nivel y Preescolar) del resto de los niveles, es decir, sugiere la existencia de una diferencia, mas allá, del nivel educativo en si. En otras palabras el segundo eje nos indica que jefes de

hogar con nivel educativo preescolar o sin nivel tienen diferencias con los jefes de niveles educativos superiores, incluso los de básica; al menos en lo que respecta al tipo de Vivienda.

Como las proyecciones de las modalidades de una variables son promedios ponderados por la frecuencias de las modalidades de la otra variable, entonces una proyección de cierta modalidades de nivel educativo cercana a modalidades de tipo de vivienda sugieren una asociación entre las modalidades. Siguiendo este criterio, observamos en el gráfico numero uno, que los hogares con jefes sin nivel educativo se asocia con hogares con viviendas de tipo rancho campesino, rustica y local no apto para vivienda; en contraste a esto, los hogares con jefes cuyo nivel educativo es alto (universitario, Postgrado) se asocian con viviendas de tipo Edificio con o sin ascensor y aquellos con nivel educativo medio son asociados con anexos casas o quintas.

4.1.2 Análisis de correspondencia: Apariencia de la vivienda vs. Tipo de Vivienda

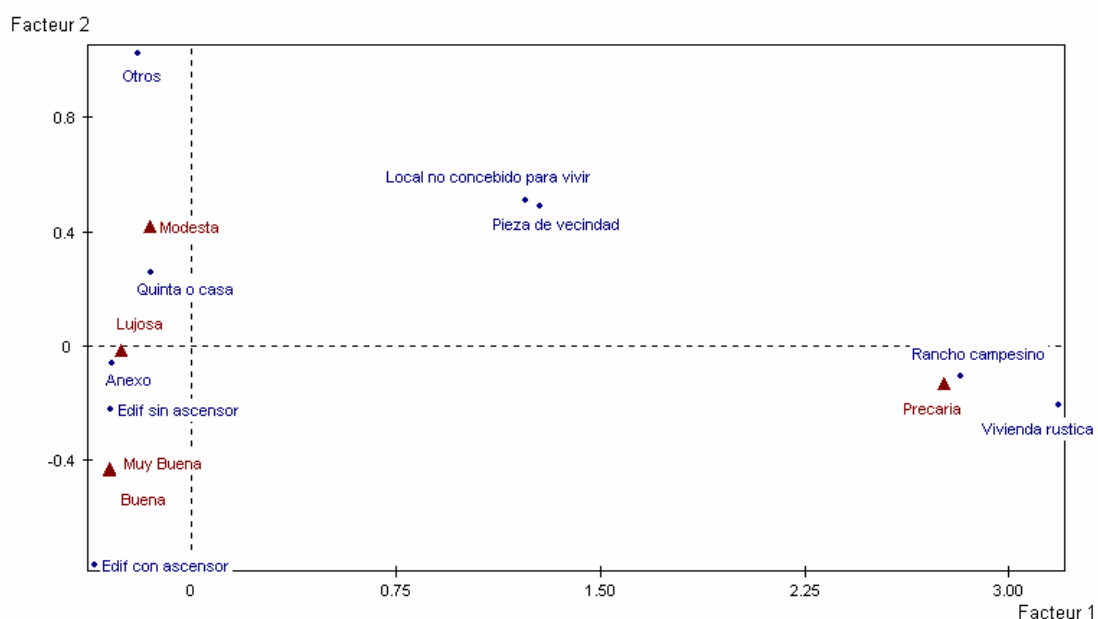
En este análisis se desea describir la asociación entre tipo de vivienda y su apariencia (lujosa, muy buena, buena, moderada y precaria), para tener una idea mas clara de la condición o calidad de vida a la cual se le puede asociar el tipo de vivienda, ya que, no necesariamente un rancho campesino debe tener una apariencia precaria. En el apéndice D se muestran los resultados referentes a contribuciones y calidad de representación de las modalidades. En este se aprecia que la modalidad precaria de la variable apariencia de la vivienda define prácticamente el primer factor con un 91 %. El segundo eje esta definido por las modalidades buena y moderada con un 40 y 50% respectivamente. Con respecto a las modalidades del tipo de vivienda, ocurre un patrón similar, las modalidades bajas

(vivienda rustica, rancho campesino) definen prácticamente el primer factor con 67 y 25 % respectivamente. Mientras que el segundo esta definido por las modalidades altas (Quinta o casa, Edificio con ascensor) definen el segundo factor (26 y 69%).

La calidad de representación para todas las modalidades es bastante aceptable. Esta alrededor de un 75 a 80% para el primer plano factorial. La representaciones mas bajas, pero significativa, son Lujosa con 41% y Edificio sin ascensor con 55%. (ver apéndice D).

GRAFICO N° 2

Plano factorial: Apariencia de la vivienda vs Tipo de Vivienda

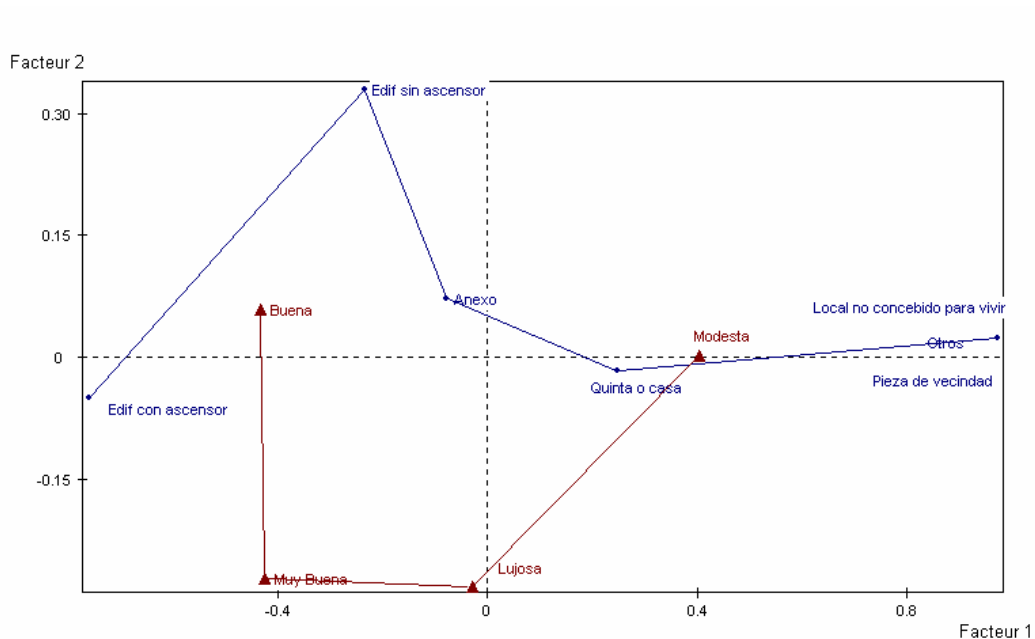


En este plano observamos rápidamente que rancho campesino y vivienda rustica esta asociado con una apariencia precaria, Sin embargo estas tres modalidades distorsionan el plano factorial impidiendo observar las relaciones entre las otras modalidades. Por esta

razón. Se hizo otro análisis excluyendo las modalidades: precaria, Vivienda rustica y rancho campesino. Sin embargo antes de realizar este segundo análisis es importante notar como el premier eje identifica tres grupos de tipo de vivienda. El primero conformado por: Edificio con ascensor, Edificio sin ascensor, Anexo, caso o quinta y otros. El segundo por: Local no apto para vivir y pieza de vecindad y por ultimo el tercer grupo: Vivienda rustica y rancho campesino. Otro aspecto de relevancia es que el segundo factor sirve como un eje establece un orden en el primer grupo. Dicho orden se repite, pero ahora en el primer factor, cuando hacemos el análisis excluyendo las modalidades que mencionamos anteriormente.

GRAFICO N° 3

Plano factorial: Apariencia de la vivienda vs Tipo de Vivienda excluyendo a las modalidades Precaria, Viv. Rust y Rancho campesino



Del grafico anterior se observa que las modalidades local no concebido para vivir, otros y pieza de vecindad son muy similares, lo cual sugiere la idea de unir estas modalidades en una sola. Por otro lado el primer factor realiza un orden (idéntico al que hizo el segundo factor en el primer análisis) el cual sugiere que la modalidad con mayor nivel o calidad de vida es edificio con ascensor, el cual esta asociada con apariencias muy buena y buena. Por otro lado también se evidencia una relación entre quinta o casa y apariencia modesta, esta ultima afirmación puede parecer contradictoria, pero cabe señalar que esta modalidad (casa o quinta) es muy general, el cual puede abarcar cualquier tipo de casa.

4.1.3 Análisis de correspondencia: Zona geográfica vs. Tipo de Vivienda

Ahora describiremos la relación entre el tipo de vivienda y la zona geográfica, los resultados referente a las contribuciones y calidad de representación de las modalidades se presentan en el apéndice D. Observando el plano factorial entre estas dos variables (grafico cuatro) se percata dos grupos de zonas geográficas, el primero conformado por la zona del centro y el segundo por el resto de las zonas (occidente, sur, oriente y llano). Este resultado es bastante razonable si consideramos que el centro del país se caracteriza por ubicar las ciudades principales y por ser una zona netamente urbana, a diferencia del resto del país, en el cual predomina la zona rural. Otro resultado razonable es la asociación del tipo de vivienda edificio con ascensor y anexo con la zona del centro, ya que, por ser áreas urbanas predominan los edificios y abundan los jóvenes profesionales que viven en anexos de vivienda. Sin embargo la modalidad otros distorsiona la asociación entre las otras modalidades, por lo que es conveniente sacarla del análisis (grafico numero 5)

GRAFICO N° 4

Plano factorial: Zona geográfica vs Tipo de Vivienda

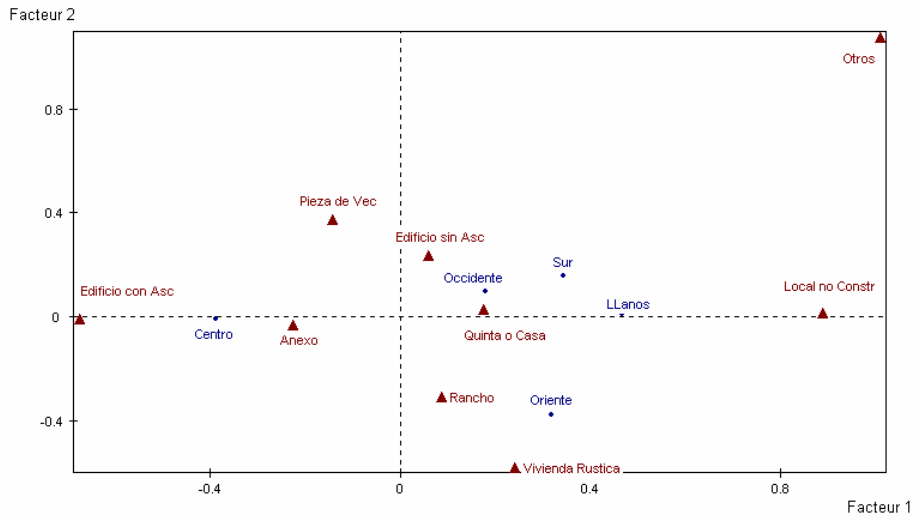
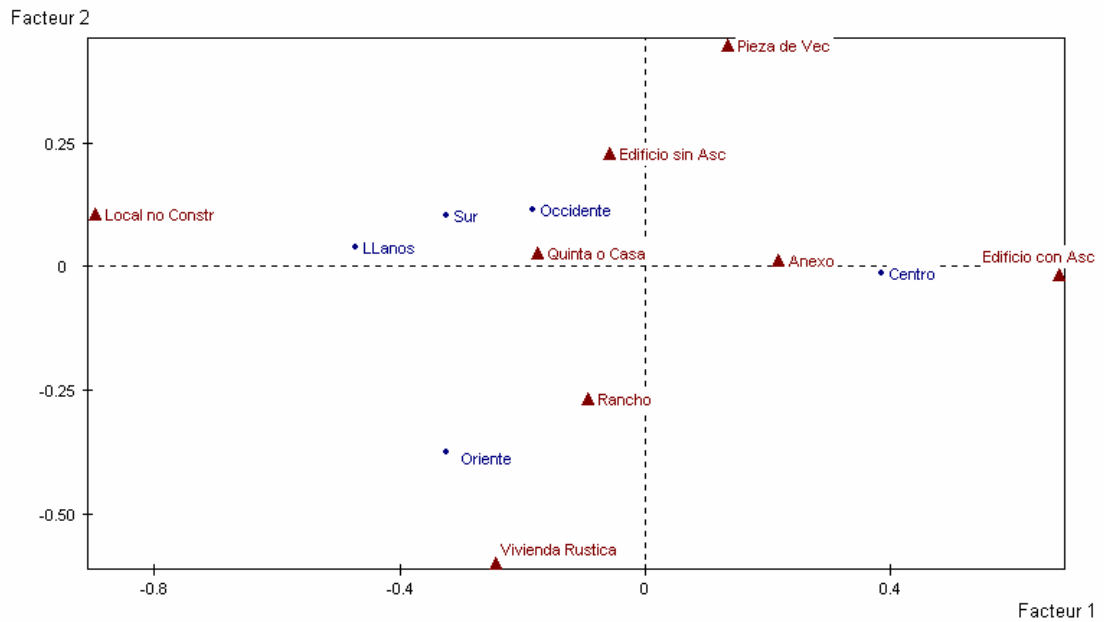


GRAFICO N° 5

Plano factorial: Zona geográfica vs Tipo de Vivienda excluyendo a la modalidad

Otros



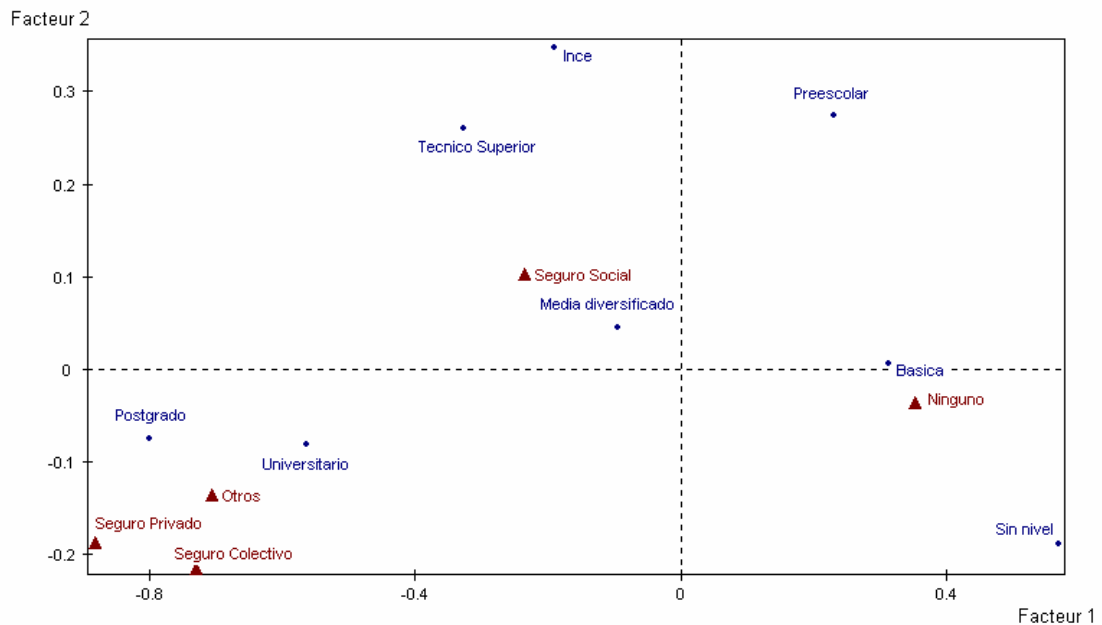
En este último gráfico se sigue observando la diferencia entre los dos grupos de zonas geográficas mencionado anteriormente, Además también sigue prevaleciendo la asociación entre edificio con ascensor y anexo con la zona del centro. Pero sale a relucir la notoria relación que existe entre la zona oriental con el tipo de vivienda rústica y rancho campesino. También el segundo eje sugiere una diferencia entre Oriente y el resto de las zonas. Por último también podemos afirmar que las zonas Occidente, Llanos y Sur están asociados a viviendas de tipo casa o quinta.

4.1.4 Análisis de correspondencia: Sistema de seguro vs. Nivel Educativo

En el siguiente gráfico se observa que el primer factor del plano factorial, describe las siguientes asociaciones entre las modalidades de nivel educativo y sistema de seguro: los jefes de hogar con nivel educativo de postgrado y universitario están relacionados con sistemas de seguro privado y colectivo, mientras que los que poseen nivel educativo intermedio (Técnico superior, Ince y medio diversificado) están asociados con el seguro social obligatorio y por último los que poseen un nivel educativo bajo (Sin nivel, Preescolar y básica) no tienen sistema de previsión social.

GRAFICO N° 6

Plano factorial: Sistema de seguro vs Tipo de Nivel Educativo

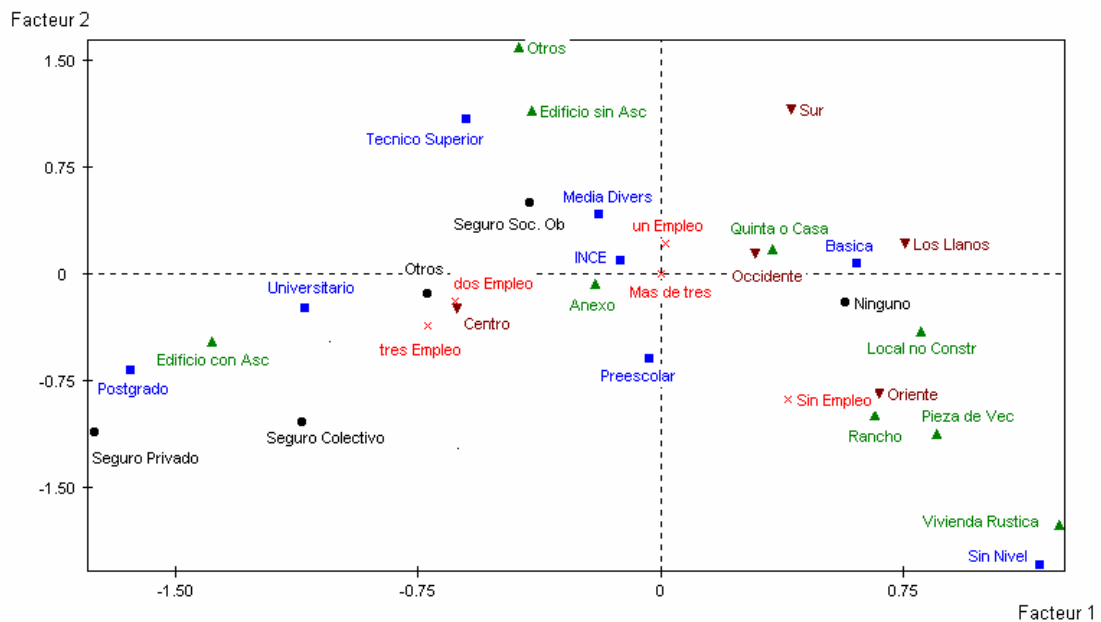


4.1.5 Análisis de correspondencia múltiple: Nivel Educativo, Tipo de Vivienda, Zona, Sistema de seguro y Numero de empleos del jefe del hogar.

En esta sección se realizó un análisis de correspondencia múltiple, el cual pretende describir la relación simultánea de las variables categóricas que han intervenido a lo largo de los distintos análisis de correspondencia simple. Los resultados referentes a las contribuciones y calidad de proyección de las distintas modalidades, como en los anteriores análisis, se presentan en el apéndice D. En el gráfico número siete se presenta el primer plano factorial del análisis de correspondencia múltiple.

GRAFICO N° 7

Plano factorial: Nivel educativo, Tipo de vivienda, Zona geográfica, Sistema de seguro y numero de empleos del jefe del hogar.



En el gráfico anterior se observa dos grupos de hogares discriminados por el primer eje: la zona del centro donde los jefes del hogar tienden a tener entre dos y tres empleos, las viviendas que predominan son los edificios y el nivel educativo es alto. La descripción anterior se contrapone con la zona de Oriente, los Llanos, Occidente y Sur, donde predominan las casas o quinta, vivienda rústica y ranchos campesinos. Las personas tienen un nivel educativo más bajo y por lo general tienen un solo empleo. Esta afirmación se puede sustentar, por el hecho de que las zonas geográficas, excluyendo el centro (Aragua, Carabobo y Distrito Federal) son prácticamente en su totalidad rurales, predomina los niveles educativos bajos, las viviendas rurales y la escasez de sistemas de previsión social. Por lo tanto este primer eje puede ser útil a la hora de crear un índice de calidad de vida, es

decir, individuos cuyas proyecciones estén a la izquierda del primer factor indicaría una buena calidad de vida, en términos de un Nivel educativo alto, apariencia de la vivienda buena, posee un sistema de previsión social, posee una buena remuneración alta, entre otros; y a medida que las proyecciones de los individuos estén a la derecha del factor, indicará un descenso en lo que respecta a: empleo, nivel educativo, tipo de vivienda, etc.

4.2. Modelos Loglineales

Hasta ahora hemos descrito mediante análisis de correspondencia las diferentes asociaciones o relaciones existentes entre las varias variables socioeconómicas de tipo categóricas. En esta sección se pretende explicar las relaciones antes descritas mediante modelos loglineales, en otras palabras lo que se quiere es dar a conocer el modelo que subyace o que está por detrás de los resultados arrojados por los análisis de correspondencia.

Empezaremos planteando un modelo saturado (incluyendo todas las interacciones de orden superior) de las siguientes variables: Nivel Educativo, Tipo de Vivienda, Zona y Sistema de previsión social.

En el cuadro numero 23, se presentan los estadísticos de razón de verosimilitud chi-cuadrado y el chi-cuadrado de Pearson. Los cuales contrastan la hipótesis nula de que todos los efectos debidos a las interacciones de orden k son iguales a cero. En dicho cuadro se aprecia que los dos estadísticos coinciden en que el efecto de la interacción de orden cuarta no es significativa mientras que los efectos principales e interacciones dobles si lo son, es decir, al menos una interacción de orden dos y un efecto principal es significativo, diferente

de cero. Con respecto a los efectos de las interacciones triples existe una diferencia entre los estadísticos, ya que, usando la razón de verosimilitud no se rechaza la hipótesis nula de que los efectos de tercer orden son iguales a cero; en cambio con el estadístico chi-cuadrado de Pearson se concluye lo contrario. Por lo tanto esta discrepancia se solucionara cuando analicemos los resultados del test de asociación parcial.

CUADRO N° 23

Test. Efectos de orden k igual a cero

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob
1	23	49372.81	0	363928.236	0
2	192	6188.976	0	10203.411	0
3	688	553.891	0.9999	834.376	0.0001
4	896	183.201	1	163.267	1

Como se dijo anteriormente, el hecho de que un efecto de orden k fuese significativo en el test anterior, no implica que todas las posibles interacciones de orden k lo sean. Por lo tanto es necesario realizar el test de asociación parcial, el cual se presenta en el cuadro numero 24. En este se aprecia que todas las interacciones de primer y segundo orden son significativas, en cambio las de tercer orden solo es significativa la interacción Nivel educativo, Zona y sistema de previsión social. Es curioso observar que todas las interacciones de tercer orden donde esta involucrado Tipo de vivienda no son significativas. Esto quiere decir que en conjunto las categorías de cualquier dupla entre nivel educativo, Zona, sistema de previsión social con las categorías de tipo de vivienda son independientes.

CUADRO N° 24

Test de asociación parcial

Efectos	DF	Partial Chisq	Prob
Nivel educativo*tipo de vivienda*zona	224	198.198	0.892
Nivel educativo*tipo de vivienda*seguro	224	106.145	1
Nivel educativo*zona*seguro	112	167.559	0.0005
Tipo de Vivienda*zona*seguro	128	73.182	1
Nivel educativo*tipo de vivienda	56	1593.348	0
Nivel educativo*zona	28	147.824	0
Tipo de vivienda*zona	32	1099.754	0
Nivel educativo*seguro	28	983.853	0
Tipo de vivienda*seguro	32	205.15	0
Zona*seguro	16	182.778	0
Nivel educativo	7	12132.172	0
Tipo de vivienda	8	23191.397	0
Zona	4	3701.744	0
Seguro	4	10347.513	0

Por motivo de cantidad de parámetros sólo se mostrarán las estimaciones de los parámetros de primer orden, ya que, el número de parámetros para las interacciones de orden dos y tres es demasiado grande (el menor de orden dos es de 16 parámetros y el único de orden tres contiene 225). Sin embargo en el apéndice E. Se muestran las estimaciones e intervalos de confianza para todos los parámetros.

En el cuadro numero 25 se presenta las estimaciones de los efectos de las categorías de nivel educativo. En éste se observa que técnico superior no es significativo, es decir la frecuencia de esa categoría es la misma si todas las categorías fuesen equiprobables. Los efectos positivos y negativos debido a una categoría lo que indican es que la frecuencia

observada de dicha categoría es mayor o menor que la que cabría esperar si todas las categorías de las variables fuesen equiprobables.

CUADRO N° 25

Estimación de los efectos principales de las categorías de Nivel educativo

Parámetro	Coeficiente	Error Est.	Valor de Z	Intervalo de confianza 95%	
				Mínimo	Máximo
Preescolar	-0.5611818	0.0859	-6.53328	-0.72954	-0.39283
Básica	0.58740967	0.07241	8.1121	0.44548	0.72934
Medio divers.	0.42821738	0.07313	5.85545	0.28488	0.57156
Técnico superior	-0.01041205	0.07838	-0.13285	-0.16403	0.1432
Universitario	0.38309534	0.07447	5.14446	0.23714	0.52905
Postgrado	-0.16214831	0.08012	-2.0239	-0.31918	-0.00512
Ince	-0.48659777	0.08459	-5.75249	-0.65239	-0.3208
Sin Nivel	-0.17838246	0.08164	-2.0845	-0.3254	-0.00632

CUADRO N° 26

Estimación de los efectos principales de las categorías de Tipo de vivienda

Parámetros	Coeficiente	Error Est	Valor Valor de Z	Intervalo de confianza 95%	
				Mínimo	Máximo
Casa o Quinta	1.67261247	0.0604	27.69371	1.55423	1.79099
Edif. Con asc	0.70035849	0.07296	9.59865	0.55735	0.84337
Edif. Sin asc	0.27774982	0.0778	3.56991	0.12526	0.43024
Local no apto para vivir	-0.55145224	0.09189	-6.00131	-0.73155	-0.37135
Anexo	-0.41320962	0.08936	-4.62398	-0.58836	-0.23806
Pieza de vecindad	-0.58353885	0.09262	-6.30022	-0.76508	-0.402
Vivienda rustica	-0.18547791	0.08674	-2.13838	-0.35548	-0.01547
Rancho campesino	-0.34168209	0.08888	-3.84443	-0.51588	-0.16748
Otros	-0.57536007	0.09245	-6.01241	-0.73245	-0.37244

En esta tabla se aprecia la estimación de los efectos de las categorías tipo de vivienda. Existe en este cuadro dos aspectos relevantes. El primero es la diferencia en magnitud de la categoría casa o quinta con respecto a las demás categorías y la segunda es la diferencia en signo de las primeras categorías con las restantes, es decir, los tipos de vivienda que se consideran de un nivel alto tienen signo positivo y las demás categorías, las cuales representan niveles bajos, tienen signo negativo.

CUADRO N° 27

Estimación de los efectos principales de las categorías de Zona

Parámetros	Coeficiente	Error Est	Valor de Z	Intervalo de confianza 95%	
				Mínimo	Máximo
Centro	0.3154995	0.05791	5.44814	0.202	0.429
Occidente	0.1933681	0.05796	3.33649	0.07978	0.30696
Oriente	-0.101463	0.06002	-1.69041	-0.21911	0.01618
Llano	-0.260364	0.0615	-4.23336	-0.38091	-0.13982
Sur	-0.14704	0.07012	-1.7546	-0.26914	-0.01545

CUADRO N° 28

Estimación de los efectos principales de las categorías de Sistema de previsión social

Parámetros	Coeficiente	Error Est	Valor de Z	Intervalo de confianza 95%	
				Mínimo	Máximo
Seguro Social Oblig	0.3753019	0.05686	6.6004	0.26386	0.48675
Privado	-0.30285	0.06185	-4.89635	-0.42408	-0.18162
Colectivo	-0.321943	0.06198	-5.19408	-0.44343	-0.20046
Otro	-0.225719	0.06121	-3.68778	-0.34569	-0.10575
Ninguno	0.4752095	0.05894	6.824	0.2548	0.49345

En los últimos dos cuadros se observa la estimación de los efectos principales de las categorías de las variables zona y sistema de previsión social. En este último la categoría seguro social obligatorio tiene el mismo signo que la categoría ninguno, en contraposición a seguro privado o colectivo. Esto tiene algún sentido, ya que la posesión de un seguro social obligatorio es prácticamente lo mismo a no tener un sistema de previsión social.

4.3. Análisis de componentes principales

En esta sección, se realizará primero un conjunto de análisis de componentes principales con dos grupos específicos de variables, el primer grupo es referente a varias categorías de gastos mensuales, y el segundo grupo son categorías de gastos trimestrales o anuales, los cuales se muestran en los siguientes cuadros:

CUADRO N° 29

Grupos de variables de gastos mensuales:

Código	Descripción
VE	Vestimenta exterior
VI	Vestimenta interior
RE	Ropa escolar
VT	Vestimenta, calzado y accesorios de trabajo exclusivamente
VC	Vajilla y cristalería
UC	Utensilios de cocina
AACP	Artículos y accesorios para el cuidado y aseo personal
LTEF	Libros, textos y similares de apoyo a la educación formal y otros servicios de enseñanza
AAVS	Artículos y accesorios vinculados con la salud
AARE	Artículos y accesorios relacionados con actividades de esparcimiento
PM	Plantas flores y animales domésticos
SD	Gasto en servicio domestico
SEF	Servicio de enseñanza formal
SCP	Servicio de cuidado personal

Código	Descripción
VTM	Viaje de tres o menos días
IOC1	Ingreso por primera ocupación
IOC2	Ingreso por segunda ocupación
IOC3	Ingreso por tercera ocupación

Fuente: elaboración propia, con base en el formato de muestreo de la EPF (BCV,1997)

CUADRO N° 30

Grupos de variables de gastos trimestrales o anuales

Código	Descripción
Mu	Muebles
EEH	Equipo electrodoméstico del hogar
EC	Equipos de computación
ACP	Artículos de cuidado personal
SRF	Servicio de fabricación y reparación a todo costo
RM	Repuesto y otros materiales
Vi	Viajes
GV	Gastos en viajes
Sa	Salud y otros servicios conexos
RGV	Reparación y gastos en mant. de vivienda
RV	Remodelaciones de la vivienda
CV	Compra de vehículos
SV	Seguros de vehículos
SRV	Seguros referentes a la vivienda

Fuente: elaboración propia, con base en el formato de muestreo de la EPF (BCV,1997)

A continuación se realizarán los análisis de componentes principales para cada grupo de variables caracterizando a los individuos con las variables categóricas ya estudiadas en las secciones anteriores, lo cual nos permitirá conocer o describir mejor a los individuos o grupos de individuos frente a las variables de gastos. Además podremos

caracterizar en función de variables cualitativas patrones de comportamiento de los individuos con respecto a las variables de gastos.

4.3.1 Análisis de componentes principales para los dos grupos de variables de gastos, categorizado por entidad geográfica.

4.3.1.1 Variables de gastos mensuales

Para este análisis los individuos son los diferentes hogares, a los cuales se les midió los rubros de gastos mensuales antes mencionados. Para facilitar la interpretación de los resultados, se agruparon los individuos por entidades geográficas y se calculó el promedio de los rubros de gastos mensuales por entidad. De esta manera se reducirá el número de individuos a representar en el plano formado por las componentes principales.

Antes de empezar a mostrar los resultados se mostrará la codificación de las entidades geográficas:

CUADRO N° 31

Código de las entidades geográficas

Código	Entidad	Código	Entidad
Am	Amazonas	Ma	Aragua
An	Anzoátegui	Me	Mérida
Ap	Apure	Mi	Miranda
Ba	Barinas	Mo	Monagas
Bo	Bolívar	NE	Nueva Esparta
Ca	Carabobo	Po	Portuguesa
Co	Cojedes	Su	Sucre
DA	Delta Amacuro	Ta	Táchira
DF	Distrito Federal	Tr	Trujillo
Fa	Falcón	Ya	Yaracuy
Gu	Guarico	Zu	Zulia
La	Lara		

En los tres siguientes cuadros se presentan los primeros resultados del análisis de componentes principales. En el primero se describe el valor de los autovalores y el porcentaje de variabilidad captada por los diferentes componentes principales; observamos que los dos primeros componentes principales captan sólo un 53% de la variabilidad total, sin embargo son las componentes que de forma individual aportan más a la totalidad de la inercia 37 % y 16 % respectivamente. Por otro lado si decidimos escoger un número de componentes que cubran un 80 a 85 % de la inercia tendríamos entonces que escoger las primeras seis componentes, lo cual causaría serios trastornos a la hora de interpretación; además la ganancia de reducción de dimensión no sería del todo significativa. por esta razones, se trabajará con las dos primeras componentes. Luego se presentan dos Sub-cuadro de izquierda a derecha los cuales representan: la correlación entre las variables y las componentes y por último las coordenadas de los autovectores.

Observando las correlaciones entre los factores y las variables, existe en la mayoría de las variables una correlación negativa, esto quiere decir que individuos con valores pequeños en las componentes tendrán altos valores en las variables de estudio y viceversa. Esta reflexión es importante a la hora de analizar el grafico de los individuos sobre el plano generado por los dos primeros componentes principales. Ahora si elevamos al cuadrado las correlaciones entre las variables y el factor obtendremos la contribución del factor a la variable, es decir, el porcentaje que el factor explica a la variación de la variable. En este aspecto podemos afirmar que el plano conformado por los dos primeros factores explica en la mayoría de las variables entre 40 y 60 %. Las variables mejor representadas por el plano con un 75% aproximadamente son: Servicio de enseñanza formal y Artículos y accesorios relacionados con actividades de esparcimiento. Las variables menos representadas por el

plano, con un porcentaje entre 20 y 30 % son Ropa escolar, Vestimenta, calzado y accesorios de trabajo exclusivamente y Artículos y accesorios vinculados con la salud. Por último, observando las coordenadas de los autovectores podemos notar que las variables contribuyen aproximadamente en un mismo porcentaje en la construcción de los componentes principales.

CUADRO N° 32

Autovalores y % de Inercia

Numero	Auovalores	Porcentaje	Porcentaje acumulado
1	67.110	37.28	37.28
2	29.837	16.58	53.86
3	16.137	8.97	62.82
4	13.532	7.52	70.34
5	11.217	6.23	76.57
6	0.8696	4.83	81.40
7	0.7410	4.12	85.52
8	0.6183	3.44	88.96
9	0.5459	3.03	91.99
10	0.4155	2.31	94.30
11	0.3557	1.98	96.27
12	0.2255	1.25	97.53
13	0.1596	0.89	98.41
14	0.1328	0.74	99.15
15	0.0574	0.32	99.47
16	0.0495	0.28	99.74
17	0.0276	0.15	99.90
18	0.0183	0.10	100.00

CUADRO N° 33

Correlación variable-factor, coordenadas de autovectores

Variables	Correlación entre variable y factor					Coordenadas de Autovectores				
	1	2	3	4	5	1	2	3	4	5
VE	-0.50	0.42	-0.62	0.08	0.06	-0.19	0.24	-0.49	0.07	0.05
VI	-0.70	0.39	0.29	0.06	-0.10	-0.27	0.23	0.23	0.05	-0.10
RE	0.35	0.37	0.01	-0.57	0.56	0.13	0.21	0.00	-0.49	0.53
VT	-0.34	-0.22	0.71	0.25	0.45	-0.13	-0.13	0.56	0.21	0.42
VC	-0.43	-0.51	-0.28	0.24	-0.25	-0.17	-0.30	-0.22	0.21	-0.23
UC	-0.17	0.79	0.18	0.16	-0.21	-0.06	0.46	0.14	0.14	-0.20
AACP	-0.68	-0.05	-0.17	0.29	0.29	-0.26	-0.03	-0.13	0.25	0.27
LTEF	0.04	0.79	-0.24	0.09	0.13	0.01	0.46	-0.19	0.08	0.12
AAVS	-0.53	-0.15	-0.44	0.38	0.31	-0.21	-0.09	-0.35	0.33	0.29
AARE	-0.27	0.81	0.19	0.15	-0.24	-0.11	0.47	0.15	0.13	-0.23
PM	-0.77	0.07	-0.03	-0.09	0.09	-0.30	0.04	-0.03	-0.08	0.09
SD	-0.68	-0.39	0.16	0.03	-0.29	-0.26	-0.23	0.12	0.03	-0.27
SCP	-0.82	0.11	0.25	0.08	0.06	-0.32	0.07	0.20	0.07	0.06
SEF	-0.85	0.20	-0.05	-0.12	-0.10	-0.33	0.11	-0.04	-0.10	-0.09
VTM	-0.78	-0.25	-0.07	-0.33	-0.09	-0.30	-0.14	-0.06	-0.29	-0.08
IOC1	-0.79	-0.08	-0.14	-0.08	0.29	-0.30	-0.05	-0.11	-0.07	0.27
IOC2	-0.64	-0.02	-0.11	-0.67	-0.19	-0.25	-0.01	-0.09	-0.58	-0.18
IOC3	-0.77	-0.07	0.31	-0.13	0.06	-0.30	-0.04	0.25	-0.11	0.06

Antes de observar la representación de los individuos en el plano formado por los dos primeros componentes principales, analizaremos si están bien representados por estos observando los cósenos cuadrados (calidad de representación). (ver cuadro N° 34)

CUADRO N° 34

Contribuciones y calidad de representación de los individuos

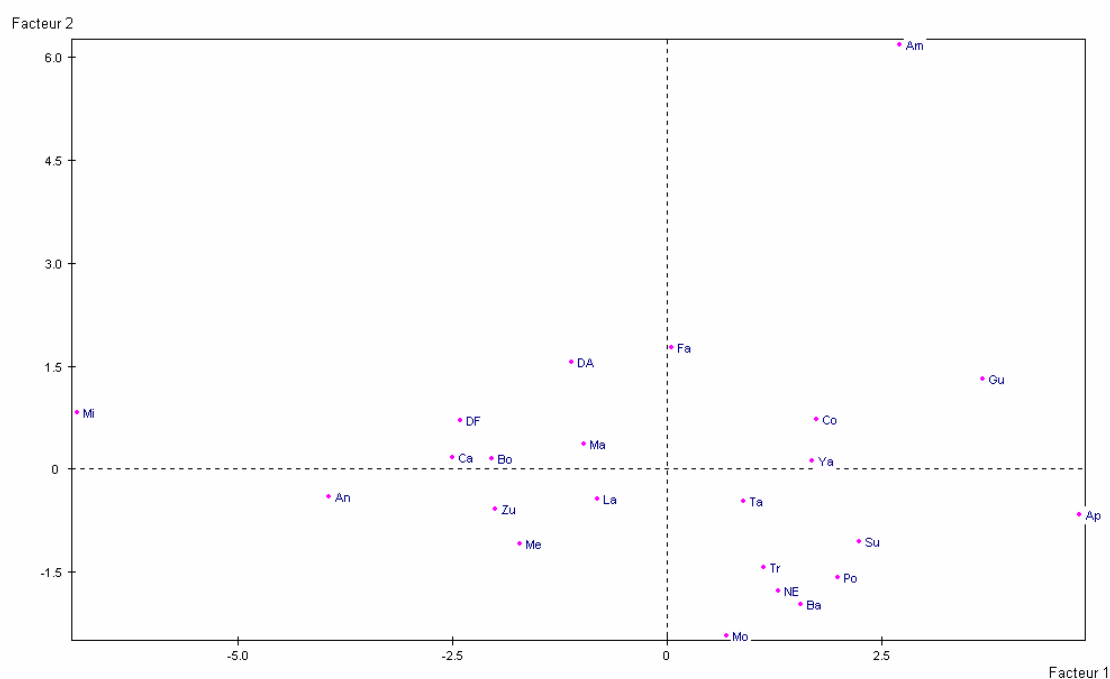
Individuos	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
Am	4.7	55.6	0.7	0.1	3.5	0.15	0.78	0.01	0.00	0.02
An	10.1	0.2	3.0	4.8	9.4	0.56	0.01	0.04	0.05	0.09
Ap	15.0	0.6	0.7	12.1	1.9	0.73	0.01	0.01	0.12	0.02
Ba	1.6	5.7	3.4	0.5	7.6	0.18	0.29	0.09	0.01	0.15
Bo	2.7	0.0	1.1	2.2	0.0	0.38	0.00	0.04	0.06	0.00
Ca	4.1	0.0	3.0	3.4	8.0	0.47	0.00	0.08	0.08	0.16
Co	1.9	0.8	1.6	1.7	0.4	0.24	0.04	0.05	0.04	0.01
DA	0.8	3.6	20.9	4.1	0.0	0.07	0.14	0.44	0.07	0.00
DF	3.8	0.7	0.4	0.0	1.6	0.64	0.06	0.02	0.00	0.05
Fa	0.0	4.6	43.4	1.1	2.9	0.00	0.13	0.65	0.01	0.03
Gu	8.8	2.5	0.6	13.8	33.9	0.44	0.06	0.01	0.14	0.29
La	0.4	0.3	0.0	0.7	3.7	0.13	0.04	0.00	0.04	0.19
Ma	0.6	0.2	1.4	1.7	0.4	0.16	0.02	0.09	0.09	0.02
Me	1.9	1.7	1.5	15.9	1.9	0.18	0.07	0.03	0.30	0.03
Mi	30.7	1.0	0.1	29.5	2.3	0.78	0.01	0.00	0.15	0.01
Mo	0.3	8.7	5.7	2.3	0.1	0.03	0.39	0.14	0.05	0.00
NE	1.1	4.6	4.3	0.6	4.2	0.11	0.21	0.10	0.01	0.07
Po	2.6	3.6	1.2	0.4	0.0	0.36	0.23	0.04	0.01	0.00
Su	3.2	1.6	0.7	2.8	0.8	0.45	0.10	0.02	0.08	0.02
Ta	0.5	0.3	0.8	1.2	5.4	0.12	0.03	0.05	0.06	0.22
Tr	0.8	3.0	3.7	0.1	4.6	0.12	0.19	0.12	0.00	0.11
Ya	1.8	0.0	1.1	0.2	2.7	0.39	0.00	0.06	0.01	0.10
Zu	2.6	0.5	0.8	0.7	4.6	0.52	0.04	0.04	0.03	0.15

Observando la matriz de los cosenos cuadrados vemos que los individuos mejor representados por el plano conformado por los dos primeros componentes son Amazonas y Miranda con 96 y 79 % respectivamente. Los demás individuos tienen una calidad de

representación entre 45 y 70 %. Sólo los individuos: Falcón, Delta Amacuro, Táchira y Lara están mal representados.

GRAFICO N° 8

Proyección de los Individuos en el plano factorial categorizados por zona geográfica



En el gráfico anterior se aprecia que las entidades llaneras (Guarico, Apure, Cojedes, Portuguesa y Barinas) son entidades que gastan poco en las mayorías de los rubros de gastos mensuales. A diferencia de las entidades centrales (Carabobo, Distrito Federal, Bolívar, Anzoátegui y Miranda) que gastan más o los gastos son mayores en los rubros de los gastos mensuales. Por otro lado el segundo componente discrimina entre Amazonas y el grupo Monagas, Barinas, Trujillo, Portuguesa y Nueva Esparta. Este segundo grupo se

diferencia de Amazonas en el sentido de que gastan poco en los siguientes rubros: Utensilios de cocina, Libros, textos y similares de apoyo a la educación formal y Artículos y accesorios relacionados con actividades de esparcimiento.

4.3.1.2 Variables de gastos Trimestrales o anuales.

Para los rubros de gastos trimestrales y anuales las componentes que más aportan de forma individual a la variabilidad de la nube de puntos son las tres primeras componentes (mayor que el 10%), acumulando un total de inercia de 63,97 %. Sin embargo para nuestro análisis tomaremos las dos primeras componentes sacrificando un porcentaje del 11%. De esta manera ganamos sencillez a la hora de interpretar los resultados. (ver cuadro N° 35).

Igual que para los rubros de gastos mensuales las correlaciones de los gastos trimestrales o anuales con el primer factor son todas negativas excepto para la variable Equipos electrodomésticos para el hogar. Esto quiere decir que entidades a la derecha del primer factor en promedio tienen pocos gastos en los rubros de gastos trimestrales o anuales. La variabilidad de las variables explicadas por el plano conformado por los dos primeros componentes esta entre 55% y 75% aproximadamente, excepto las variables Muebles, Equipos de computación y Repuestos y otros materiales, los cuales están explicados solamente por un 25% aproximadamente (ver cuadro N° 36).

Observando el cuadrado de las coordenadas de los autovectores se verifica que las variables contribuyen de manera aproximadamente igual a la construcción de los componentes principales, excepto las variables Equipos de computación, Equipos

electrodomésticos del hogar y seguros de vehículo, los que contribuyen en 1 % aproximadamente. (ver cuadro N° 36)

CUADRO N° 35

Autovalores y % de Inercia

Numero	Autovector	Porcentaje	Porcentaje acumulado
1	47.565	33.98	33.98
2	25.267	18.05	52.02
3	16.731	11.95	63.97
4	12.348	8.82	72.79
5	10.018	7.16	79.95
6	0.7660	5.47	85.42
7	0.5118	3.66	89.08
8	0.4259	3.04	92.12
9	0.3795	2.71	94.83
10	0.2478	1.77	96.60
11	0.1808	1.29	97.89
12	0.1374	0.98	98.87
13	0.1114	0.80	99.67
14	0.0466	0.33	100.00

Con respecto a los individuos, en líneas generales las entidades están bien representadas por el plano formado por las dos primeras componentes. La mayoría poseen una calidad de representación entre 50 y 75 %. La entidad mejor representada por el plano es Distrito Federal con 85%. Las entidades Maracay, Mérida, Portuguesa, Yaracuy y Táchira no deben ser objeto de análisis, ya que su calidad de representación es muy baja (menor que 10%). (ver cuadro N° 37)

CUADRO N° 36**Correlación variable-factor, coordenadas de autovectores**

Variables	Correlación entre las variables y el factor					Coordenadas de los autovectores				
	1	2	3	4	5	1	2	3	4	5
Mu	-0.46	0.23	0.17	-0.36	-0.44	-0.21	0.14	0.13	-0.33	-0.44
EC	-0.14	0.52	0.40	0.61	-0.13	-0.06	0.33	0.31	0.55	-0.13
EEH	0.11	0.63	0.13	-0.25	-0.48	0.05	0.40	0.10	-0.22	-0.48
ACP	-0.85	-0.04	-0.21	-0.01	-0.01	-0.39	-0.02	-0.16	-0.01	-0.01
SRF	-0.50	0.65	0.08	0.23	0.18	-0.23	0.41	0.06	0.21	0.18
RM	-0.46	-0.41	-0.37	0.54	-0.08	-0.21	-0.26	-0.28	0.48	-0.08
Vi	-0.67	-0.55	0.31	-0.10	-0.24	-0.31	-0.35	0.24	-0.09	-0.24
Sa	-0.72	0.03	-0.49	0.10	0.03	-0.33	0.02	-0.38	0.09	0.03
RGV	-0.64	0.01	-0.55	0.05	-0.36	-0.29	0.01	-0.43	0.04	-0.36
RV	-0.76	0.34	0.07	-0.24	0.26	-0.35	0.22	0.05	-0.22	0.26
CV	-0.59	0.44	0.36	0.25	0.04	-0.27	0.28	0.28	0.23	0.04
SV	-0.26	-0.61	0.62	0.18	0.05	-0.12	-0.38	0.48	0.16	0.05
SRV	-0.66	0.15	0.01	-0.36	0.50	-0.30	0.09	0.01	-0.33	0.50
GV	-0.72	-0.45	0.35	-0.15	-0.10	-0.33	-0.28	0.27	-0.14	-0.10

CUADRO N° 37**Contribución y calidad de representación de los individuos**

Individuos	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
Am	0.8	14.0	31.2	19.1	3.2	0.03	0.29	0.43	0.19	0.03
An	11.2	0.4	0.0	6.6	38.1	0.48	0.01	0.00	0.07	0.34
Ap	12.6	0.9	0.0	25.7	7.0	0.55	0.02	0.00	0.29	0.06
Ba	0.5	0.9	0.2	5.1	0.3	0.06	0.06	0.01	0.16	0.01
Bo	2.5	5.0	4.8	2.9	0.2	0.24	0.26	0.17	0.08	0.00
Ca	2.7	0.2	1.9	0.5	2.2	0.51	0.02	0.12	0.02	0.09
Co	2.6	0.1	0.0	0.5	1.7	0.40	0.01	0.00	0.02	0.05

CUADRO N° 34(Continuación)

Individuos	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
DA	2.2	12.4	4.4	2.6	6.5	0.15	0.44	0.10	0.04	0.09
DF	7.9	1.7	0.0	0.7	1.7	0.76	0.09	0.00	0.02	0.03
Fa	6.5	0.2	2.1	1.9	1.0	0.62	0.01	0.07	0.05	0.02
Gu	0.5	53.9	19.4	0.1	0.4	0.01	0.77	0.18	0.00	0.00
La	0.6	0.5	0.7	0.7	0.2	0.14	0.07	0.06	0.04	0.01
Ma	0.0	0.1	0.4	0.2	2.8	0.00	0.01	0.03	0.01	0.13
Me	0.3	0.4	0.6	5.5	5.5	0.04	0.03	0.03	0.19	0.15
Mi	33.3	0.2	0.0	1.4	15.2	0.86	0.00	0.00	0.01	0.08
Mo	6.4	4.3	0.0	0.9	1.4	0.55	0.20	0.00	0.02	0.02
NE	1.2	0.2	4.6	6.8	0.0	0.14	0.01	0.20	0.21	0.00
Po	0.1	0.6	3.8	13.2	0.2	0.01	0.05	0.18	0.46	0.01
Su	2.6	0.0	2.0	2.6	0.4	0.51	0.00	0.14	0.13	0.02
Ta	0.0	1.2	0.6	0.9	9.2	0.00	0.10	0.03	0.03	0.30
Tr	2.2	2.5	7.9	0.7	0.1	0.24	0.14	0.30	0.02	0.00
Ya	0.7	0.3	11.4	1.1	2.7	0.07	0.01	0.40	0.03	0.06
Zu	2.5	0.0	4.0	0.3	0.0	0.43	0.00	0.24	0.01	0.00

En los gráficos N° 9 y 10, se observa que al igual que en el primer grupo de variables (rubros de gastos mensuales) se mantiene una diferencia entre entidades llaneras y las centrales. En el sentido que las llaneras gastan menos en los rubros de gastos trimestrales y anuales que las centrales; un ejemplo de esto es la gran discriminación causada por el primer componente principal entre Apure y el grupo de entidades Miranda, Anzoátegui y Distrito federal. El segundo componente describe una notoria diferencia entre el grupo Amazonas, Delta Amacuro y la entidad Guárico. Si observamos el gráfico de las variables podremos concluir que Amazonas y Delta Amacuro gasta más en los rubros Equipos electrodomésticos, Equipos de computación y Servicio de fabricación y reparación a todo costo que Guárico. Del resto de las entidades no se puede decir mucho, ya que están muy cerca del promedio o simplemente no están bien representadas por el plano.

GRAFICO N° 9

Proyección de las variables en el plano factorial (variables de gastos trimestrales o anuales)

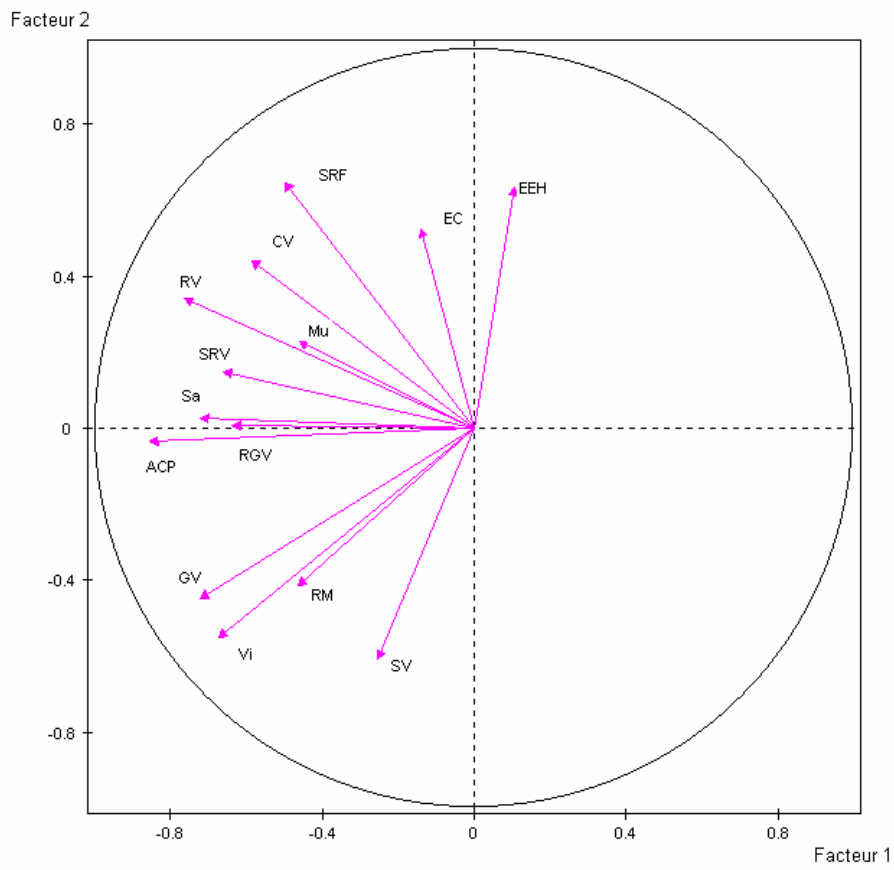
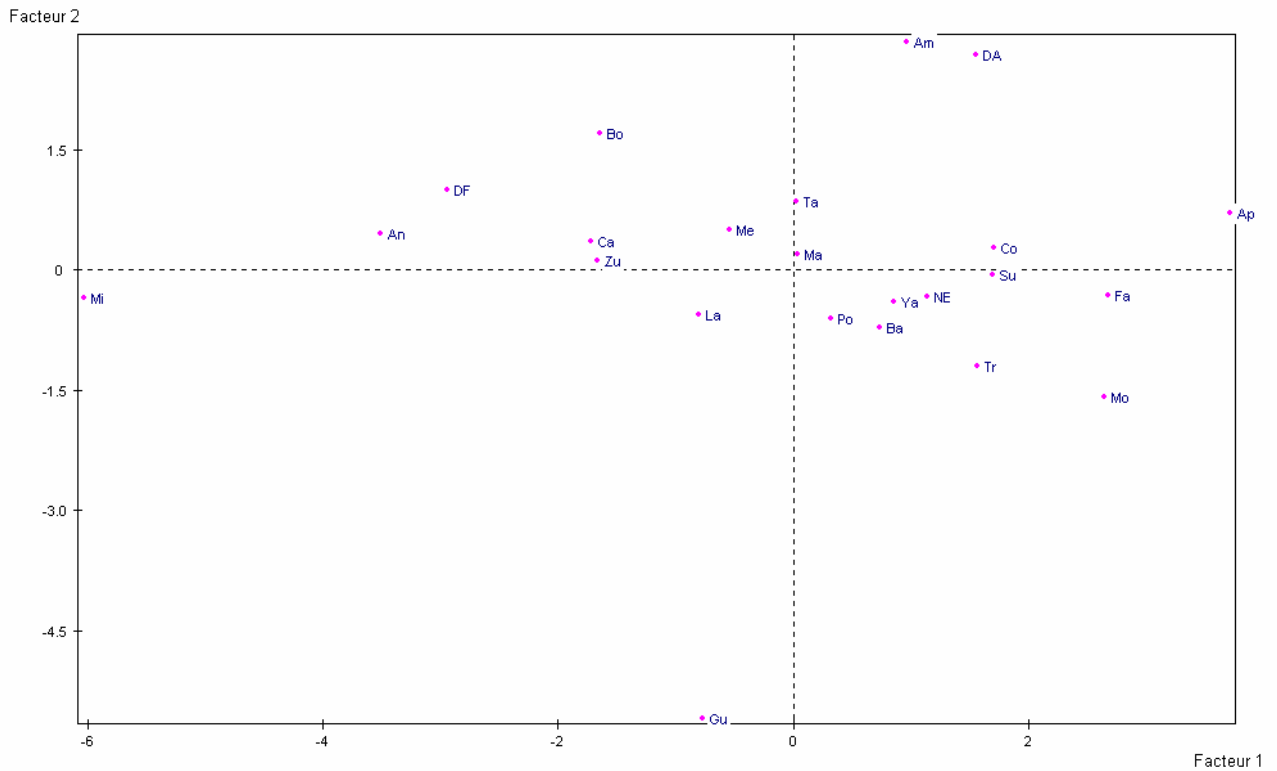


GRAFICO N° 10

Proyección de los Individuos en el plano factorial categorizados por entidades geográficas



4.3.2. Análisis de componentes principales para los dos grupos de variables de gastos, categorizado por nivel educativo del jefe del hogar

4.3.2.1 Variables de gastos mensuales

Para facilitar la interpretación de los resultados, se agruparon los jefes de los hogares por Nivel educativo y se calculó el promedio de los rubros de gastos mensuales por entidad. De esta manera se reduce el número de individuos a representar en el plano formado por las componentes principales.

Antes de empezar con los resultados se mostrará la codificación de los diferentes niveles educativos:

CUADRO N° 38

Código del nivel educativo

Código	Nivel educativo
Pres	Pre-Escolar
Bas	Básica
Md	Media diversificado
TS	Técnico Superior
Un	Universitario
Post	Post-Grado
Ince	Ince o similares
SN	Sin nivel

Observando el cuadro N° 39 notamos que los dos primeros componentes acumulan un porcentaje de variabilidad del 91%. Por lo tanto, podemos escoger los dos primeros factores para construir un plano de mejor representación. Por otro lado, cabe señalar que el primer eje por si solo explica el 82.34% de la variabilidad, lo que nos hace indicar que posiblemente existe un efecto talla (variables muy correlacionadas entre si). Es decir que el primer eje es un discriminador u ordenador natural de los datos.

También se nota en el cuadro N° 40 que las variables (rubros de gastos mensuales) están muy bien representadas por el plano conformado por los dos primeros ejes factoriales. La variabilidad explicada por el plano de la mayoría de las variables están entre 90 y 95%; solo las variables: Servicio de enseñanza formal y Artículos y accesorios vinculados con la

salud tienen una representación por el plano del 85%, que sigue siendo bastante representativo.

CUADRO N° 39

Autovalores y % de Inercia

Numero	Autovalores	Porcentaje	Porcentaje acumulado
1	148.204	82.34	82.34
2	15.971	8.87	91.21
3	0.8148	4.53	95.74
4	0.3673	2.04	97.78
5	0.2724	1.51	99.29
6	0.1049	0.58	99.87
7	0.0231	0.13	100.00
8	0.0000	0.00	100.00

Por otra parte los individuos, en este caso los jefes del hogar caracterizados por el nivel educativo, también están muy bien representado por el plano factorial, ya que la mayoría tienen una calidad de representación en el plano entre 85 y 95%. Solo existe una categoría que no está bien representada, y es el nivel educativo: Técnico Superior, el cual tiene una calidad de representación del tan solo 21%. (ver cuadro N° 41)

CUADRO N° 40

Correlación variable-factor, coordenadas de autovectores

VARIAB	Correlación entre las variables y el factor					Coordenadas de los autovectores				
	1	2	3	4	5	1	2	3	4	5
VE	-0.99	-0.06	0.06	0.05	0.06	-0.26	-0.05	0.06	0.08	0.11
VI	-0.97	-0.09	-0.10	0.09	-0.13	-0.25	-0.07	-0.11	0.15	-0.26
RE	-0.98	0.15	0.08	0.12	0.03	-0.25	0.12	0.09	0.20	0.06
VT	-0.92	0.31	0.00	-0.23	0.01	-0.24	0.24	0.00	-0.38	0.02
VC	-0.53	-0.80	-0.27	0.07	-0.03	-0.14	-0.63	-0.30	0.11	-0.05
UC	-0.94	0.20	-0.06	0.27	-0.09	-0.24	0.15	-0.06	0.45	-0.18
AACP	-0.99	-0.10	0.09	0.06	0.07	-0.26	-0.08	0.10	0.10	0.13
LTEF	-0.90	0.35	-0.22	-0.10	-0.07	-0.23	0.27	-0.24	-0.16	-0.14
AAVS	-0.86	-0.34	-0.07	-0.22	0.24	-0.22	-0.27	-0.08	-0.36	0.46
AARE	-0.97	-0.18	0.09	-0.11	-0.07	-0.25	-0.14	0.10	-0.18	-0.14
PM	-0.90	-0.41	0.07	0.02	-0.10	-0.23	-0.33	0.08	0.03	-0.19
SD	-0.94	-0.12	-0.03	-0.21	0.02	-0.24	-0.10	-0.03	-0.35	0.04
SCP	-0.81	0.51	-0.27	-0.09	-0.10	-0.21	0.40	-0.30	-0.15	-0.20
SEF	-0.99	0.00	0.03	0.14	-0.05	-0.26	0.00	0.04	0.23	-0.10
VTM	-0.92	0.05	-0.39	-0.04	0.01	-0.24	0.04	-0.43	-0.06	0.01
IOC1	-0.98	0.04	0.10	0.17	0.06	-0.25	0.03	0.11	0.29	0.12
IOC2	-0.88	0.26	0.22	0.10	0.32	-0.23	0.20	0.24	0.16	0.61
IOC3	-0.75	-0.05	0.60	-0.15	-0.20	-0.20	-0.04	0.67	-0.25	-0.38

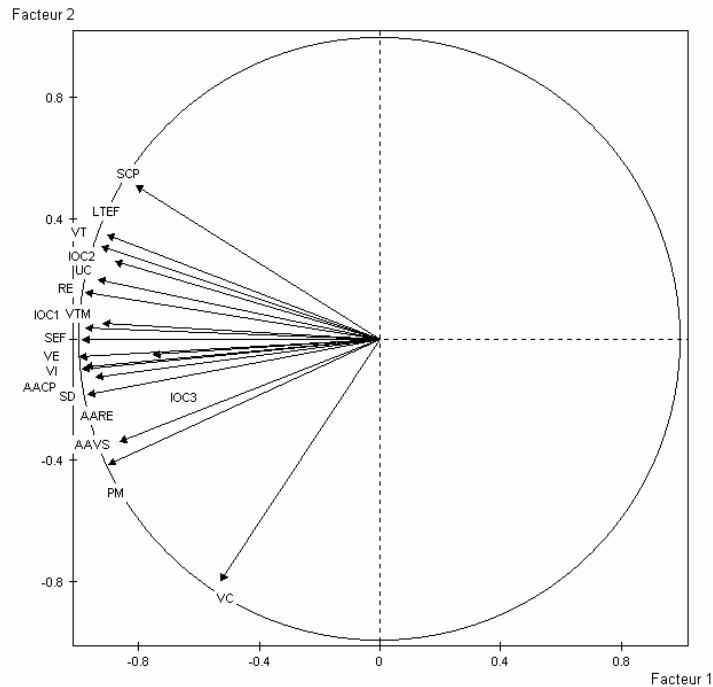
CUADRO N° 41

Contribuciones y calidad de representación de los individuos

Individuos	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
Pres	21.6	0.8	5.0	45.8	0.6	0.93	0.00	0.01	0.05	0.00
Bas	4.6	3.3	0.9	4.4	2.7	0.85	0.07	0.01	0.02	0.01
Md	0.1	14.5	3.0	18.1	11.1	0.05	0.56	0.06	0.16	0.07
TS	0.8	0.3	34.6	7.0	42.7	0.21	0.01	0.51	0.05	0.21
Un	9.5	14.7	30.9	14.1	16.4	0.71	0.12	0.13	0.03	0.02
Post	50.7	8.4	6.8	3.4	17.3	0.97	0.02	0.01	0.00	0.01
Ince	2.1	56.4	18.7	6.6	3.1	0.22	0.65	0.11	0.02	0.01
SN	10.6	1.5	0.0	0.6	5.9	0.95	0.01	0.00	0.00	0.01

GRAFICO N° 11

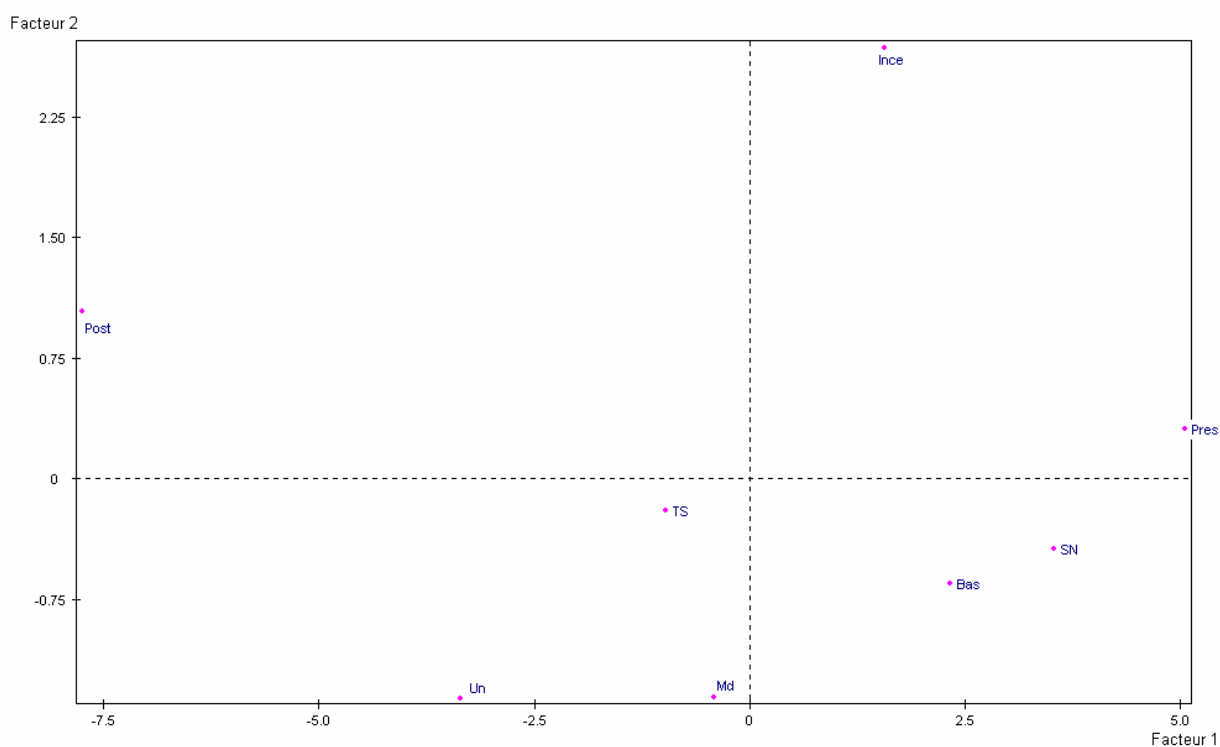
Proyección de las variables en el plano factorial (variables de gastos mensuales)



En el gráfico anterior se aprecia la alta correlación que existe entre las variables, ya que todos los vectores tienden a una sola dirección; esto también es un indicio de que existe un efecto talla.

GRAFICO N° 12

Proyección de los Individuos en el plano factorial categorizados por nivel educativo



Lo más evidente a primera vista en el gráfico anterior es el ordenamiento que hace el primer factor de izquierda a derecha, en el sentido de mayor a menor nivel educativo (Postgrado, Universitario, TSU, Medio diversificado, Ince, Básica Sin nivel y Preescolar).

Este ordenamiento sugiere que los hogares con jefes de niveles educativos altos realizan mas gastos mensuales en los rubros considerados, que los jefes con menor grado de instrucción. Además, se supone que individuos con alto nivel educativo son personas ya mayores, que ya han alcanzado un nivel de vida, en el cual le permite gastar en rubros que otra persona no gastaría.

Ahora observando el segundo componente principal, se advierte que éste discrimina entre nivel educativo Ince y Medio diversificado. Estos dos niveles educativos son considerados similares a la hora de realizar los gastos mensuales por el primer componente, y además son niveles de instrucción bastante parecidos. Por lo tanto el segundo factor explica posiblemente la forma del gasto o la diferencia implícita entre los niveles educativos medios.

4.3.2.2 Variables de gastos Trimestrales o Anuales.

Para las variables de gastos trimestrales o anuales, el análisis de componentes principales es bastante similar a hecho con las variables de gastos mensuales. También existe la presencia de un efecto talla; tanto los individuos como las variables están muy bien representadas por los planos factoriales y además el primer componente realiza un ordenamiento de los individuos según su grado de instrucción.

En el cuadro N° 42 se muestra como los dos primeros ejes captan el mayor porcentaje de variabilidad 87,15 %, a pesar de que el segundo componente aporta un 13,69%. Es evidente el efecto talla que se presenta en este análisis, es decir las variables

originales son explicadas prácticamente por el primer factor. Esto se debe a la alta correlación que existe entre las variables.

Las correlaciones entre las variables y el primer factor son todas negativas y altas en valor absoluto, esto quiere decir, que individuos con cierto nivel educativo representados con valores bajos en la componente, significa que realizan altos gastos en los rubros de gastos trimestrales o anuales y viceversa. (ver cuadro N° 43)

Cabe destacar que las variables están bien representadas por los factores. La variable que peor es explicada por el plano formado por los dos primeros componentes es Seguro referente a la vivienda con un 63%, el cual no deja de ser significativo.

Ahora haciendo el mismo análisis para los individuos, se confirma que están bien representados por el plano. La calidad de representación de la mayoría está entre 80 y 95%. Solo los individuos con nivel educativo Técnico superior y diversificado tienen una pobre representación en el plano con 29 y 16 % respectivamente.

CUADRO N° 42

Autovalores y % de Inercia

Numero	Autovalores	Porcentaje	Porcentaje acumulado
1	102.843	73.46	73.46
2	19.162	13.69	87.15
3	0.6072	4.34	91.48
4	0.5694	4.07	95.55
5	0.3517	2.51	98.06
6	0.1859	1.33	99.39
7	0.0853	0.61	100.00
8	0.0000	0.00	100.00

CUADRO N° 43

Correlación variable-factor, coordenadas de autovectores

Variables	Correlación entre las variables y el factor					Coordenadas de los autovectores				
	1	2	3	4	5	1	2	3	4	5
Mu	-0.80	0.57	-0.07	-0.09	0.12	-0.25	0.41	-0.08	-0.12	0.21
EC	-0.92	-0.24	0.26	0.01	-0.04	-0.29	-0.18	0.33	0.01	-0.07
EEH	-0.59	0.78	-0.01	-0.13	-0.19	-0.18	0.56	-0.01	-0.17	-0.32
ACP	-0.88	0.20	-0.33	-0.08	0.25	-0.27	0.14	-0.43	-0.10	0.42
SRF	-0.85	0.52	0.08	-0.05	0.05	-0.26	0.38	0.10	-0.07	0.08
RM	-0.78	0.04	-0.10	0.59	0.07	-0.24	0.03	-0.13	0.78	0.12
Vi	-0.93	-0.20	-0.21	-0.20	-0.04	-0.29	-0.15	-0.27	-0.26	-0.06
Sa	-0.92	-0.05	0.28	0.05	-0.22	-0.29	-0.03	0.36	0.07	-0.37
RGV	-0.90	-0.09	-0.05	0.30	0.04	-0.28	-0.07	-0.06	0.40	0.07
RV	-0.94	-0.20	-0.01	-0.02	-0.20	-0.29	-0.14	-0.02	-0.02	-0.34
CV	-0.99	-0.10	0.05	-0.02	-0.08	-0.31	-0.07	0.07	-0.03	-0.14
SV	-0.65	-0.68	-0.24	-0.18	0.03	-0.20	-0.49	-0.30	-0.24	0.05
SRV	-0.78	-0.17	0.45	-0.14	0.35	-0.24	-0.12	0.58	-0.19	0.59
GV	-0.96	-0.16	-0.14	-0.08	-0.09	-0.30	-0.11	-0.18	-0.11	-0.15

En la representación de los individuos en el plano factorial, observamos que se mantiene igual que en el análisis hecho con las variables de gastos mensuales. Es decir sigue apareciendo un ordenamiento de los niveles educativos según grado de instrucción, los cuales afirman que a mayor grado de instrucción del jefe del hogar, mayor es el gasto del hogar en rubros de gastos trimestrales o anuales. Por otro lado el segundo eje sigue diferenciando a los miembros con nivel educativo Ince del resto. Esto nos hace pensar que los jefes con grado de instrucción Ince tienen una estructura de gastos diferente al resto de los miembros con otros niveles educativos.

GRAFICO N° 13

Proyección de las variables en el plano factorial (variables de gastos trimestrales o anuales)

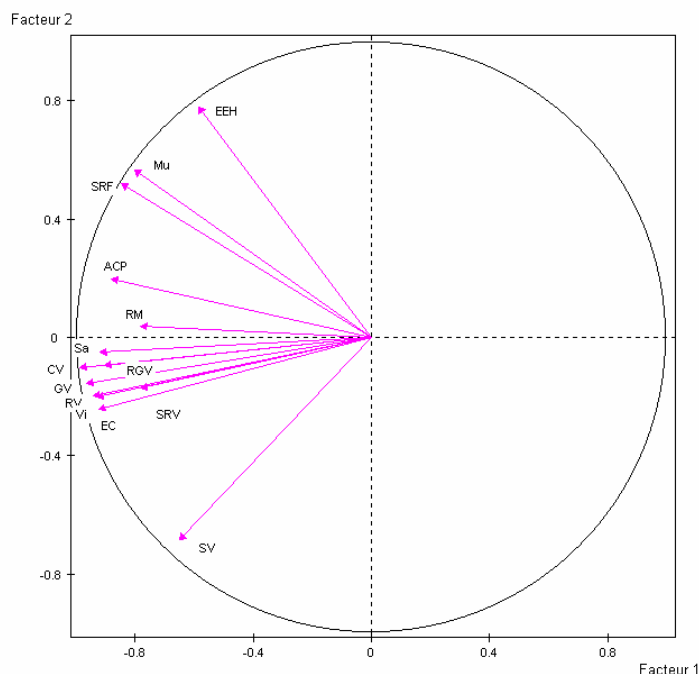
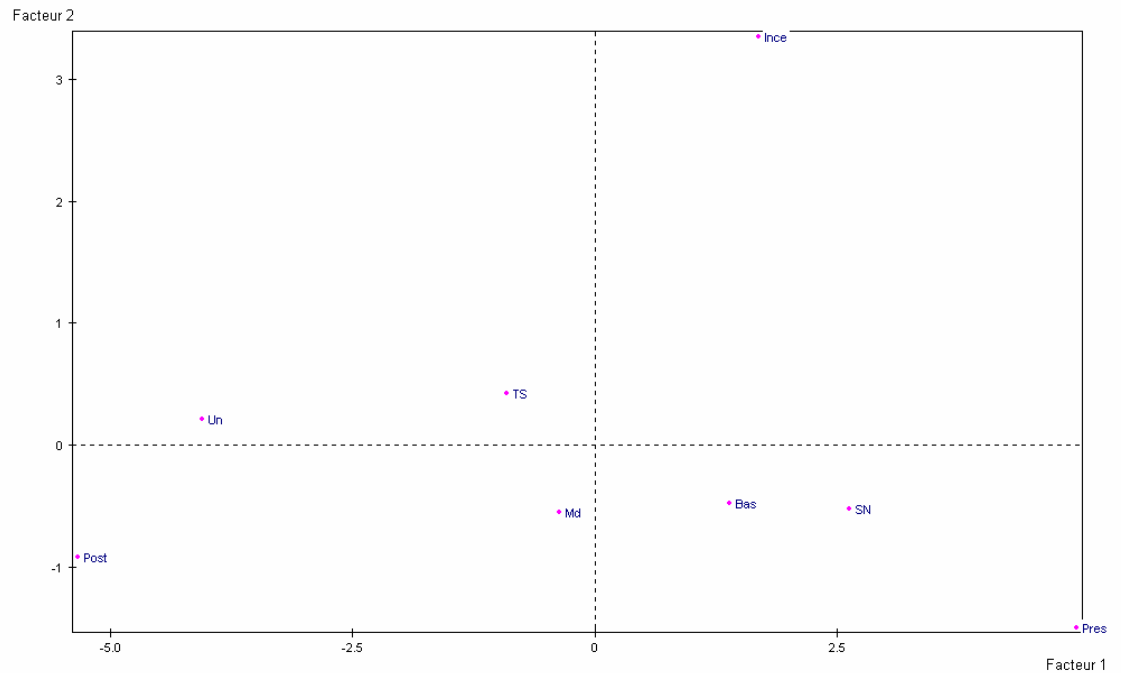


GRAFICO N° 14

Proyección de los Individuos en el plano factorial categorizado por nivel educativo



4.3.3 Análisis de componentes principales para los dos grupos de variables de gastos, categorizado por tipo de vivienda

4.3.3.1 Variables de gastos mensuales.

Se categorizó a los diferentes hogares o individuos según el tipo de vivienda como se muestra en el siguiente cuadro:

CUADRO N° 44

Códigos de tipo de vivienda

Código	Descripción
QC	Quinta o casa
Eca	Edif. Con ascensor
Esa	Edif. Sin ascensor
LncV	Local no concebido para vivienda
Acq	Apartamento en casa o quinta
Pcv	Pieza en casa de vecindad
VR	Vivienda rural
RC	Rancho campesino
Otr	Otros

Para este análisis se decidió trabajar o tomar como mejor plano de representación, el formado por los dos primeros componentes principales, ya que estos recogen el 83 % de la variabilidad de la nube de puntos; además ellos por si solos son los que más aportan a la inercia total con 71,24 y 11,77 % respectivamente (ver cuadro N° 45).

Con respecto a la correlación de las variables con el primer factor ocurre un fenómeno similar al de los anteriores análisis. La correlación es significativamente grande e inversa, esto quiere decir que los hogares con tipos de vivienda en el lado izquierdo de la componente serán los que consumen más o gastan más en los rubros de gastos mensuales.

Las variables están bien representadas por el plano; la variable con la peor representación es Vestimenta, calzado y accesorios de trabajo exclusivamente con un 64% de variabilidad explicada por el plano, el cual sin embargo sigue siendo significativo. Por

otro lado las variables contribuyen de forma uniforme en la formación del primer factor en un aproximado de 7% . (ver cuadro N° 46)

CUADRO N° 45

Autovalores y % de Inercia

Numero	Autovalores	Porcentaje	Porcentaje acumulado
1	128.226	71.24	71.24
2	21.191	11.77	83.01
3	12.698	7.05	90.06
4	0.7752	4.31	94.37
5	0.5234	2.91	97.28
6	0.3280	1.82	99.10
7	0.1296	0.72	99.82
8	0.0324	0.18	100.00
9	0.0000	0.00	100.00

CUADRO N° 46

Correlación variable-factor, coordenadas de autovectores

Variables	Correlación entre las variables y el factor					Coordenadas de los autovectores				
	1	2	3	4	5	1	2	3	4	5
VE	-0.95	-0.10	0.12	-0.26	-0.08	-0.26	-0.07	0.10	-0.30	-0.12
VI	-0.96	0.00	0.16	-0.21	-0.01	-0.27	0.00	0.15	-0.24	-0.01
RE	-0.85	-0.35	0.04	0.36	-0.07	-0.24	-0.24	0.03	0.41	-0.09
VT	-0.66	0.46	-0.46	0.12	-0.31	-0.18	0.32	-0.41	0.13	-0.43
VC	-0.73	0.40	0.48	0.23	-0.07	-0.20	0.28	0.43	0.27	-0.10
UC	-0.96	0.25	-0.02	-0.02	0.09	-0.27	0.17	-0.01	-0.02	0.13
AACP	-0.91	0.29	0.21	-0.01	0.15	-0.26	0.20	0.19	-0.02	0.21
LTEF	-0.84	0.06	-0.52	-0.09	-0.07	-0.23	0.04	-0.46	-0.10	-0.10

CUADRO N° 46 (Continuación)

Variables	Correlación entre las variables y el factor					Coordenadas de los autovectores				
	1	2	3	4	5	1	2	3	4	5
AAVS	-0.93	0.32	-0.11	0.02	0.12	-0.26	0.22	-0.10	0.02	0.17
AARE	-0.88	0.26	0.28	-0.24	0.02	-0.25	0.18	0.25	-0.28	0.03
PM	-0.75	0.14	-0.51	-0.16	0.33	-0.21	0.09	-0.46	-0.19	0.46
SD	-0.81	-0.13	-0.11	0.24	0.22	-0.23	-0.09	-0.10	0.28	0.31
SCP	-0.93	-0.32	0.08	0.16	0.03	-0.26	-0.22	0.07	0.18	0.04
SEF	-0.65	-0.64	-0.17	0.32	-0.02	-0.18	-0.44	-0.15	0.37	-0.03
VTM	-0.56	-0.69	-0.03	-0.41	-0.20	-0.16	-0.47	-0.03	-0.47	-0.27
IOC1	-0.88	-0.42	0.11	-0.08	0.00	-0.25	-0.29	0.10	-0.09	-0.01
IOC2	-0.96	-0.08	0.23	0.01	0.10	-0.27	-0.06	0.20	0.01	0.14
IOC3	-0.85	0.31	-0.02	0.07	-0.38	-0.24	0.21	-0.02	0.07	-0.53

La calidad de representación de los individuos es bastante buena, esta alrededor del 80%. solo los individuos con tipo de vivienda Quinta casa y Apartamento en quinta están representado pobremente por el plano factorial.

CUADRO N° 47

Contribuciones y calidad de representación de los individuos

Individuo	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
QC	2.9	0.2	3.7	0.0	7.0	0.62	0.01	0.08	0.00	0.06
Eca	36.0	2.0	12.0	9.4	11.2	0.93	0.01	0.03	0.01	0.01
Esa	11.1	8.6	0.5	2.6	61.7	0.72	0.09	0.00	0.01	0.16
LncV	9.5	1.2	5.1	66.9	5.3	0.66	0.01	0.03	0.28	0.01
Acq	6.5	1.6	70.2	3.5	5.6	0.46	0.02	0.49	0.01	0.02
Pcv	21.4	2.2	0.4	2.4	5.4	0.93	0.02	0.00	0.01	0.01
VR	5.4	6.9	2.1	1.5	0.3	0.76	0.16	0.03	0.01	0.00
RC	5.8	5.5	6.1	1.9	0.5	0.66	0.10	0.07	0.01	0.00
Otr	1.4	72.0	0.0	11.8	3.1	0.10	0.84	0.00	0.05	0.01

GRAFICO N° 15

Proyección de las variables en el plano factorial (variables de gastos mensuales)

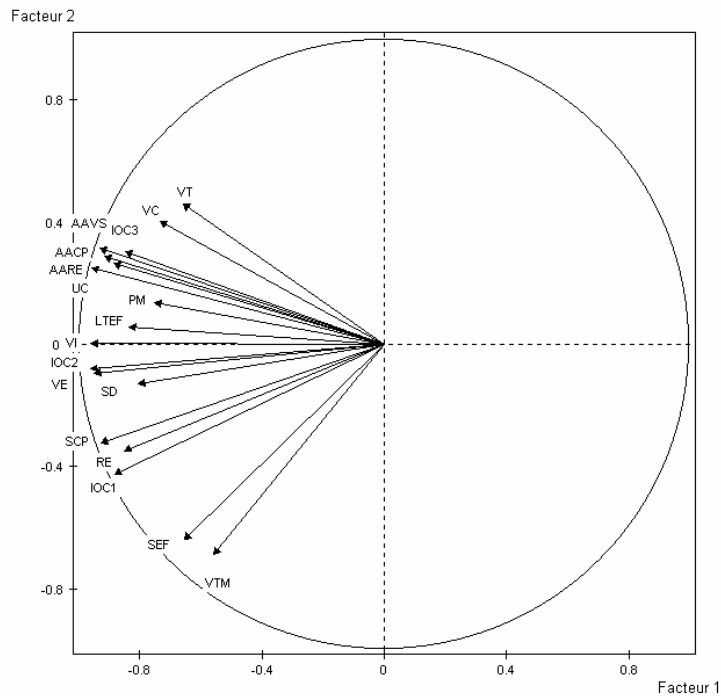
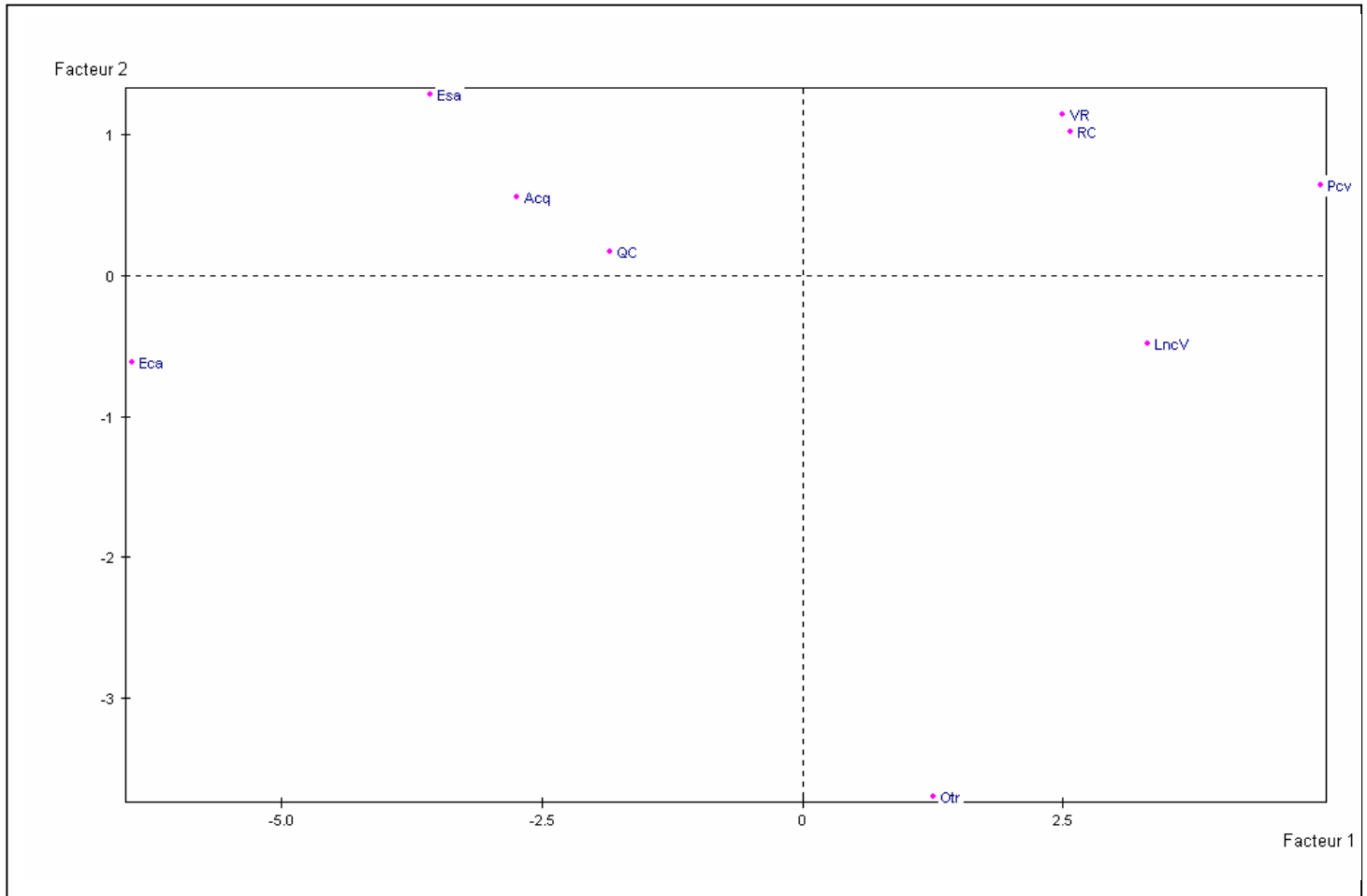


GRAFICO N° 16

Proyección de los Individuos en el plano factorial categorizados por tipo de vivienda



En el gráfico anterior se observa la formación de dos grupos de hogares, el primer grupo se ubica al lado izquierdo del segundo eje y son los hogares que poseen una vivienda, según el tipo, de mejor calidad y a su vez son los hogares que tienen gastos mensuales mas elevados. Por otra parte esta el otro grupo de hogares ubicados al lado derecho del segundo

eje, los cuales tienen como característica común una vivienda que por su tipo no cumple con algunos requerimientos mínimos y que sus gastos mensuales en los rubros considerados son bajos, esto se debe posiblemente a que sus ingresos se desvían para gastos de alta necesidad o urgencia.

4.3.3.2. Variables de gastos trimestrales o anuales.

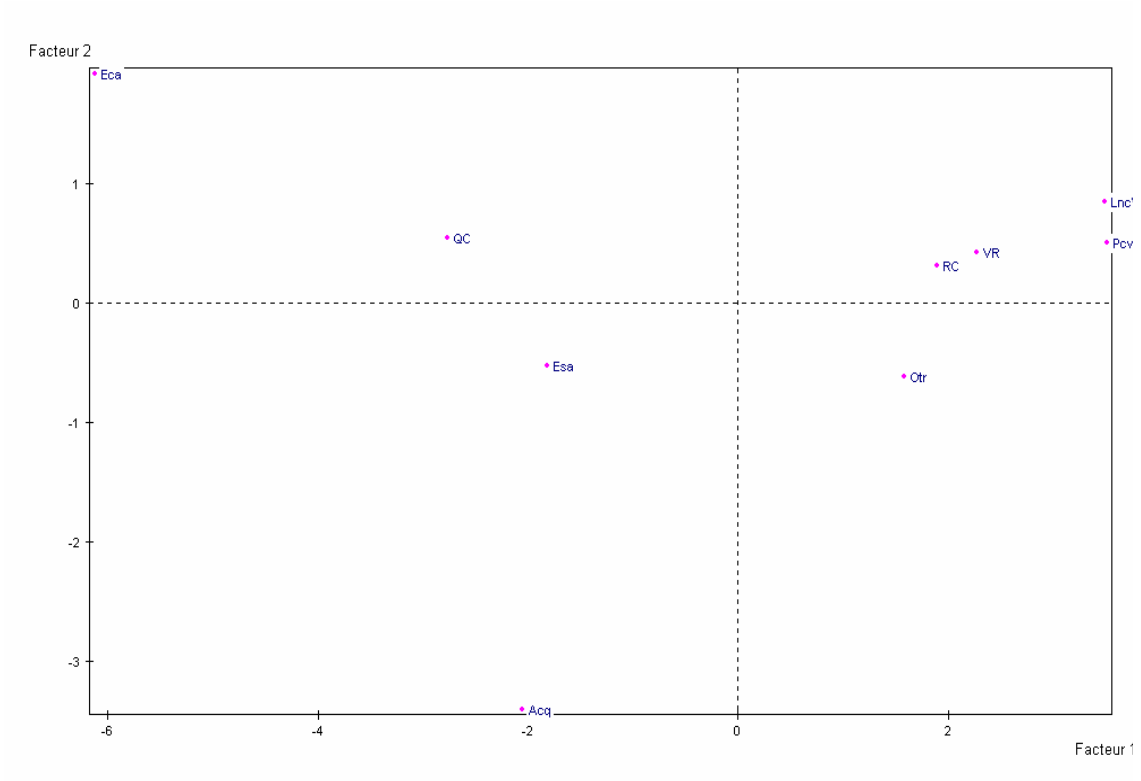
Para las variables de gastos trimestrales o anuales, el análisis de componentes principales arrojó resultados similares. El plano conformado por los dos primeros ejes ajustan de forma significativa al nube de puntos (84%), además las variables como los individuos están bien representados por el plano factorial.

Por parte de las variables, sigue existiendo el mismo patrón, una alta correlación negativa con el primer eje factorial y una contribución uniforme y aceptable en la construcción de los ejes factoriales.

En la representación de los individuos se aprecia una diferencia con respecto al análisis hecho con las variables de rubros de gastos mensuales, ya que, en este análisis los hogares caracterizados con viviendas de baja calidad están más agrupadas y desplazadas a la derecha, lo que indica que los rubros de gastos trimestrales y anuales son más restringidos que los mensuales para los hogares antes mencionados.

GRAFICO N° 17

Proyección de los Individuos en el plano factorial categorizados por tipo de vivienda



4.3.4. Análisis de componentes principales para el grupo de variables de gastos mensuales, categorizado por apariencia de la vivienda

Se categorizo a los diferentes hogares o individuos según la apariencia de vivienda como se muestra en el siguiente cuadro:

CUADRO N° 48

Código de apariencia de vivienda

Código	Descripción
Lu	Lujosa
MB	Muy Buena
Bu	Buena
Mo	Moderada
Pr	Precaria

En el cuadro N° 49 presentan los autovalores y su porcentaje de inercia captada, en el cual, podemos escoger a los dos primeros componentes como el plano de mejor ajuste a la nube de puntos, ya que, ello en conjunto recogen el 96% de la inercia total.

La correlación de las variables con el primer factor son negativas y muy altas, esto quiere decir que hogares de cierta apariencia a la izquierda del componente principal son familias con altos gastos mensuales y viceversa. (ver cuadro N° 50). En cuanto a la contribución de las variables al factor es bastante homogénea para cada variable, es decir, todas las variables aportan en la misma proporción a la formación de los componentes.

Además los componentes contribuyen de forma satisfactoria a las variables, con un porcentaje de representación entre 85 y 95 %. (ver cuadro N° 50)

CUADRO N° 49

autovalores y porcentaje de inercia

Numero	Autovalores	Porcentaje	Porcentaje acumulado
1	146.967	81.65	81.65
2	27.037	15.02	96.67
3	0.5588	3.10	99.77
4	0.0409	0.23	100.00
5	0.0000	0.00	100.00

CUADRO N° 50

Correlación variable-factor, coordenadas de autovectores

Variables	Correlación entre las variables y el factor				Coordenadas de lo autovectores			
	1	2	3	4	1	2	3	4
VE	-0.99	-0.09	-0.09	-0.02	-0.26	-0.05	-0.12	-0.08
VI	-0.96	0.28	0.01	-0.04	-0.25	0.17	0.02	-0.22
RE	-0.99	-0.04	-0.10	-0.12	-0.26	-0.02	-0.13	-0.60
VT	-0.84	0.55	0.02	0.03	-0.22	0.33	0.02	0.15
VC	-0.86	-0.13	0.49	-0.04	-0.22	-0.08	0.65	-0.19
UC	-0.93	-0.13	0.33	0.05	-0.24	-0.08	0.45	0.25
AACP	-0.99	0.08	-0.14	0.00	-0.26	0.05	-0.18	0.01
LTEF	-0.92	-0.38	0.01	-0.04	-0.24	-0.23	0.01	-0.19
AAVS	-0.75	-0.64	-0.11	0.04	-0.20	-0.39	-0.15	0.20
AARE	-0.93	0.28	0.24	-0.03	-0.24	0.17	0.32	-0.16
PM	-0.95	0.28	0.05	0.11	-0.25	0.17	0.06	0.54
SD	-0.92	-0.38	-0.14	0.01	-0.24	-0.23	-0.18	0.05
SCP	-0.98	-0.18	-0.04	0.03	-0.26	-0.11	-0.06	0.14
SEF	-1.00	-0.01	-0.09	-0.01	-0.26	0.00	-0.12	-0.03
VTM	-0.80	-0.58	-0.15	0.02	-0.21	-0.35	-0.20	0.09
IOC1	-1.00	-0.07	-0.07	0.01	-0.26	-0.04	-0.10	0.05
IOC2	-0.55	0.81	-0.19	-0.03	-0.14	0.50	-0.25	-0.13
IOC3	-0.78	0.62	-0.09	0.03	-0.20	0.38	-0.12	0.17

En la representación de los individuos con el plano formado por los dos primeros componentes, observamos que el primer componente realiza un ordenamiento según la apariencia de la casa de izquierda a derecha en sentido de mejor apariencia a peor apariencia. El cual nos indica que vivienda con mejor apariencia realizan gastos mensuales mayores. por otro lado observamos que el segundo eje discrimina entre Lujosa y Muy buena; y a su vez estas del resto de las categorías. esto puede ser porque el segundo eje explica la estructura o la variabilidad de los gastos entre los rubros, es decir, una familia con una vivienda de apariencia muy buena puede tener gastos similares a otorgar con vivienda de apariencia lujosa, pero este ultimo puede tener unos gastos que el primero no tiene o una estructura de gastos mas compleja y variable. (ver cuadro N° 51)

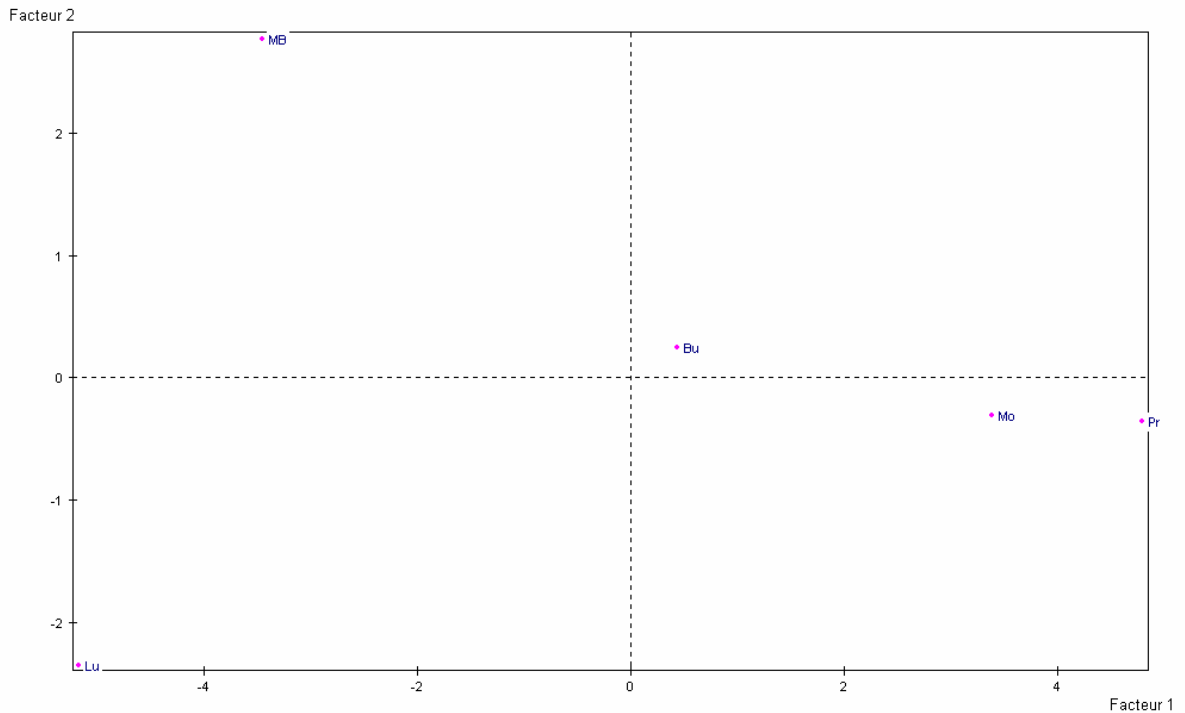
CUADRO N° 51

Contribuciones y calidad de representación de los individuos

Individuos	Contribuciones				Calidad de representación			
	1	2	3	4	1	2	3	4
Lu	36.5	41.1	2.4	0.0	0.83	0.17	0.00	0.00
MB	16.2	56.8	6.9	0.1	0.60	0.39	0.01	0.00
Bu	0.3	0.5	72.4	6.9	0.08	0.03	0.88	0.01
Mo	15.6	0.7	0.0	63.7	0.98	0.01	0.00	0.01
Pr	31.4	0.9	18.3	29.3	0.97	0.01	0.02	0.00

GRAFICO N° 18

Proyección de los Individuos en el plano factorial categorizados por apariencia de la vivienda



Hasta ahora hemos observado que los hogares a medida que alcanzan niveles ideales de las variables categóricas gastan más en los rubros de gastos tanto mensuales como trimestrales y anuales. Esto se debe, a que los hogares con niveles de vida altos tienen más recursos y pueden emplearlos con mayor comodidad que los hogares que se han caracterizados con niveles bajos.

Por otro lado hemos notado como el primer eje para la mayoría de los análisis discrimina entre hogares con calidad de vida aceptable y hogares con deficiencias en cuanto

a los estándares de calidad de vida. Además el segundo eje diferencia dentro de estos dos gran grupos, es decir, el segundo eje capta la diferencia entre hogares que tienen categorizaciones similares, por ejemplo: Nivel educativo postgrado y universitario; o apariencia de la vivienda Lujosa y muy buena; etc.

CONCLUSIONES Y RECOMENDACIONES

A partir de los resultados estadísticos se logró caracterizar a los hogares venezolanos a través de diferentes variables socioeconómicas de tipo cualitativas e incluso se logró, mediante los componentes principales relacionar estas cualidades con variables de gastos. Ello permitió construir un bosquejo sobre el patrón de gastos de un hogar dado previamente sus características cualitativas como: tipo de vivienda del hogar, nivel educativo del jefe del hogar, apariencia de la vivienda, zona geográfica, entre otros. Un ejemplo lo constituye el patrón de consumo de los hogares diferenciados por zona geográfica, es decir, los hogares del Occidente y del Centro del país son mucho más consumistas que los hogares Orientales y Llaneros. Además, también se puede concluir que los estados Guárico, Amazonas y Delta Amacuro son zonas atípicas con respecto al gasto (en relación a las demás zonas). A continuación se presentan las conclusiones y aportes más relevantes obtenidos en cada uno de los métodos estadísticos utilizados en el análisis.

1. **Análisis de Correspondencia.** Este método permitió encontrar asociaciones entre categorías de distintas variables cualitativas. Estas relaciones ayudaron a describir y confirmar las características de los diferentes hogares venezolanos. Por ejemplo,

hogares que poseen jefe de familia con nivel educativo bajo están asociados con viviendas de baja calidad de vida (Ranchos, Locales no concebidos para la vivienda, Piezas de vecindad). Este tipo de vivienda en donde la apariencia predominante es precaria están localizadas con mayor frecuencia en el interior del país, es decir, excluyendo las zonas centrales (Distrito Federal, Carabobo, Miranda), si bien es cierto en el centro del país también existen hogares con bajo nivel educativo, viviendas precarias (en fin, con un pobre nivel de vida). Sin embargo éstas se encuentran concentradas en zonas pequeñas comparadas con la gran extensión de zona rural que posee Venezuela. Por esta razón el tipo de hogares antes mencionados, a pesar de existir también en la zona central, están asociados con el interior del país. Otro aspecto importante es la información que proporciona el segundo factor del plano factorial. Mientras el primer eje en la mayoría de los análisis de correspondencia sugiere un ordenamiento natural de las categorías, el segundo describe posibles diferencias de forma entre las categorías ordenadas por el primer eje. Por ejemplo:

- a. El segundo eje discrimina a los jefes de hogar con niveles educativos: Preescolar y Sin Nivel del resto de las categorías. Esto puede ser un indicio sobre el hecho que estos hogares son totalmente diferentes a los restos de hogares, incluso de aquellos donde el jefe de hogar tiene una educación baja como por ejemplo Básica.
- b. A pesar de que las proyecciones sobre el primer eje de Rancho y vivienda rústica están relativamente cerca, indicando un parecido en el orden arrojado por este eje. En el segundo factor las proyecciones de estas categorías están distantes. Tales circunstancias dan a entender que no es lo mismo un rancho

en la zona rural que uno en la zona central del país, lo cual sugiere que a pesar de tener una vivienda y niveles educativos similares, los hogares rurales no tienen la misma calidad de vida de los hogares ubicados en barrios de la zona central del país.

Con respecto a otras relaciones encontradas entre las variables cualitativas podemos resaltar que el tipo de vivienda asociado con la mejor apariencia es edificio con ascensor, por encima de casa o quinta (la cual es la que por lo general se asocia con mejor apariencia). Esto se puede deberse a dos causas principales: una es que la categoría casa o quinta es un término muy general; y la segunda es la subjetividad que implica categorizar la apariencia de la vivienda. También, se encontró que los hogares ubicados en el centro del país, como ya se ha expuesto anteriormente, difieren de las zonas: Occidente, Llanos, Oriente y Sur, con respecto a las otras variables cualitativas como: sistema de previsión social, número de empleos del jefe del hogar, tipo de vivienda, nivel educativo, entre otros. Uno de los principales motivos de esta discriminación entre los dos grupos es que las zonas no centrales son en su mayoría rurales; ejemplo: Mérida, Zulia, Bolívar, etc.; y esto trae como consecuencia inmediata que existan más viviendas rurales, niveles educativos bajos, apariencia de viviendas no tan buenas, pocos hogares con un buen sistema de previsión social, etc.

2. **Modelos Log Lineales.** A través de estos modelos se pueden cuantificar los efectos de las distintas categorías de las variables cualitativas en las frecuencias observadas. Un resultado interesante es analizar los signos de dichos efectos para explicar las descripciones hechas en los análisis de correspondencia.

Por ejemplo:

- a. En tipo de vivienda los coeficientes de los efectos principales correspondientes a las categorías casa o quinta y edificio con o sin ascensor tienen signo positivo en contraposición de las categorías que describen viviendas de baja calidad como son Local no apto para vivir, Pieza de vecindad, Rancho campesino y vivienda rústica, los cuales tienen signo negativo. Esto pone en manifiesto una discriminación lógica entre viviendas de buena calidad con viviendas con calidad de vida baja.
- b. Los efectos de las zonas geográficas Oriente, Llano y Sur tienen signo negativo a diferencia de las zonas Centro y Occidente. Esto tiene sentido pues al observar los resultados arrojados por los distintos análisis de correspondencias, ya que en las zonas de Oriente, Sur y Llanera a diferencia del Centro y un poco del Occidente, predominan las áreas rurales. Por lo tanto, este resultado corrobora la diferencia en el estilo de vida de estos dos grupos de hogares.
- c. Ahora observando los efectos de las distintas categorías de Sistemas de Previsión Social, se nota que gozar del seguro social obligatorio va en el mismo sentido a no tener ningún sistema de previsión social. Esto indica que los hogares donde el jefe de hogar sólo cuenta con el beneficio de seguro social obligatorio tiene características similares a aquellos en los cuales no poseen ningún tipo de seguro, a diferencia de los hogares que poseen las otras categorías de sistema de previsión social (seguro privado, colectivo, otros).

3. **Análisis de Componentes Principales.** Gracias al efecto talla que originan las variables de gastos, se puede utilizar los primeros ejes factoriales para crear índices que ordenen de forma cuantitativa aspectos cualitativos de los hogares venezolanos. Así, por ejemplo, conocido los gastos mensuales de un hogar, a éste se le pueden calificar de forma numérica aspectos como: apariencia de la vivienda, que es un factor bastante subjetivo y difícil de medir. Otro resultado importante se refiere a poder identificar cuales son los estados en los cuales los hogares gastan más. Estos son: Carabobo Distrito Federal, Miranda, Anzoátegui, Zulia y Bolívar, que son estados de alto costo de vida por tradición. En contraparte se encuentran los hogares de los estados Llaneros los cuales gastan menos en comparación a los estados mencionados anteriormente. Una de las posibles razones de esta diferencia es el grado de desarrollo del comercio en ambas zonas. En las zonas Llaneras hay menos centros comerciales, industrias y menos fuentes de empleo causando una incidencia directa en el patrón de consumo de estos estados. De forma similar ocurre con los estados con patrones de consumo alto. Ellos poseen grandes centros de comercio, además de que existe en ellos mayores fuentes de empleo y por ende patrones de consumo altos.

Con respecto al consumo caracterizado por el nivel educativo del jefe del hogar, se evidencia que hogares donde se ha alcanzado un nivel educativo alto (Postgrado) presentan un alto nivel de consumo. Esto se debe a que estos hogares ya tienen cierta trayectoria (estabilidad, posesión de vivienda propia) y por ende pueden realizar gastos que no harían hogares recién formados. En general, el análisis corrobora que los hogares cuyas características cualitativas representan bajo nivel de vida (viviendas en condiciones poco favorables, bajo nivel educativo, desempleo,

zona geográfica, etc) tendrán patrones de consumo bajo, es decir desvían sus pocos ingresos a gastos de alta necesidad o urgencia.

Con base en las conclusiones anteriores, así como del trabajo y la metodología en él utilizada, se recomienda para futuros diseños de encuestas usar menos números de categorías para definir una variable cualitativa, ya que existe una falsa premisa de pensar que por tener muchas categorías en una variable se puede obtener más información. Esta práctica trae como consecuencia muchos trastornos cuando se realiza el análisis estadístico. Así ocurre, por ejemplo, con:

- Aumento exponencial de parámetros en modelos loglineales
- Poca inercia recogida en los primeros ejes factoriales
- Muchas casillas en blanco

Otro punto, no menos importante de destacar, es la implementación hecha de la base de datos en Ms-Access utilizando la información recogida por la segunda encuesta de presupuestos familiares, la cual se encuentra en el CD adjunto a este trabajo. Esta base de datos permitirá explotar de forma eficiente toda la información almacenada en éste, así como justificar mejor el elevado costo de la Encuesta de Presupuesto Familiares. Por lo tanto, se recomienda elaborar proyectos de investigación donde se use como fuente los datos almacenados en la base de datos, en cuestión, así como se sugiere aprovechar el diseño de la base de datos para implementar futuras Encuestas de Presupuestos Familiares.

Bibliografía

- ANDERSON, Theodore. 1984. **An Introduction to Multivariate Statistical Analysis**. New York: Editorial John Wiley & Sons.
- BALTER'S, Alison. 1998. **Mastering Acces 97 Development**. Editorial Sams Publishing. Segunda Edición
- BANCO CENTRAL DE VENEZUELA. 1998. **Encuesta Nacional de Presupuestos Familiares**. Caracas: (archivos electrónicos de la encuesta)
- BENZÉCRI, J. P. 1992. **Correspondence Analysis Handbook**. New York Editorial Marcel Dekker, Inc..
- CALLAHAN, Evan. 1998. **Microsoft Acces 97 Visual Basic**. Editorial Mc Graw Hill.
- CORNELL, Gary. 2000. **Visual Basic para Windows**. Editorial Mc Graw Hill.
- DATE, C.J. 1993. **Introducción a los sistemas de Base de Datos Volumen 1**. México: Editorial Wesley Iberoamericana.
- JOBSON, J.D. 1992. **Applied Multivariate Data Analysis Volume II: Categorical and Multivariate Methods** Editorial Springer-Verlag.
- LUDOVIC, Lebart. 1984. **Multivariate Descriptive Statistical Analysis. Coorespondence Analysis and Related Techniques for Large Matrices**. New York: Editorial John Wiley & Sons 1984
- MEZA, Otilia. 1998. **Notas metodológicas sobre la II Encuesta Nacional de presupuestos Familiares 1997/1998**. Caracas: Cuadernos BCV Serie Técnica
- MORRISONDonald. 1977. **Multivariate Statistical Methods**. New York: Editorial McGraw-Hill

APÉNDICE A

DEFINICIONES BÁSICAS DE LA ENCUESTA DE PREPUESTOS FAMILIARES

En esta sección se presentaran algunas definiciones básicas que nos servirán de apoyo para la interpretación de los resultados estadísticos.

Períodos de referencia: Se refiere a los períodos de tiempo, dentro de los cuales va a ser considerada la realización de determinados pagos o la percepción de ciertos ingresos.

En general los periodos de referencia se definen atendiendo principalmente a los siguientes criterios:

- Capacidad de los informantes para recordar
- Importancia del gasto, en términos de valor
- Frecuencia del gasto

Estos tres aspectos se influyen mutuamente. los gastos que son mas frecuentes y/o los que por su bajo monto son mas difíciles de recordar, tendrán periodos de referencia menores y viceversa, aquellos que por su monto y características son mas fáciles de recordar, tendrán períodos de referencia más largos. En virtud de estos criterios se han establecido para los gastos de consumo, los siguiente periodos de referencia.

- **Semanal:** Se investigan con período de referencia semanal aquellos productos con alta frecuencia de consumo o con precios bajos que pudieran ser fácilmente olvidados. Estos son: alimentos y bebidas consumidas en el hogar, artículos de cuidado personal, periódicos, pequeñas reparaciones, detergentes, etc.
- **Mensual:** Se investigan con período de referencia mensual aquellos productos, cuya frecuencia de compras o uso es mas espaciada que la semanal, por ejemplo: prendas de vestir, calzados y accesorios o aquellos servicios cuyo pago se realiza mensualmente: alquiler mensual de la vivienda, mensualidad escolar, pago de la luz, teléfono, agua, etc.

- **Trimestral:** Se investiga productos cuya frecuencia de compra es mas espaciada que la mensual, por ejemplo: muebles, equipos del hogar, servicios de salud, mantenimiento de vivienda, etc.
- **Anual:** Se investiga productos de escasa frecuencia de compra y de un valor unitario, que se considera suficiente para evitar olvidos, por ejemplo: compra de vehículos, seguros, pagos de impuestos, remodelaciones de la vivienda, etc.

Hogar: un hogar esta constituido por un conjunto de personas que comparten o dependen de una bolsa o fondo común para sus gastos o como mínimo para el gesto en alimentación y que han residido en la misma vivienda la mayor parte de los últimos seis meses o manifiestan tener intenciones de fijar allí su residencia.

Este fondo esta conformado por las aportaciones de los miembros con las cuales se atienden a los gastos que origina la satisfacción de las necesidades, tanto particulares como comunes de los mismos.

La relación de los miembros del hogar puede estar o no fundamentada en nexos consanguíneos o legales y puede depender de diferentes factores, por ejemplo, las costumbres o las estrategias de supervivencia.

Vivienda: Es toda estructura hecha o no para vivir, pero utilizada con ese fin. Esta conformada por una habitación o conjunto de habitaciones y sus dependencias, que constituyen en si misma una edificación o una parte estructuralmente separada de la misma y que, por la forma en que han sido construidas, reconstruidas, transformadas o adaptadas, están habitadas por uno o varios hogares.

Estos ambientes o habitaciones están destinadas o de algún modo se han adaptado para servir como mínimo, de área social, dormitorio, lugar de aseo personal, cocina, etc. Presentan una o mas entradas y están interconectadas entre si. Las personas que la habitan comparten en todo o en parte el uso de la misma.

Un hogar que habite una vivienda puede utilizarla como:

- **Vivienda principal:** Es la vivienda en la que el hogar reside la mayor parte del tiempo. Es la sede o residencia habitual del hogar
- **Vivienda de descanso:** Es una vivienda cuyo uso principal es el descanso y el esparcimiento. Presenta la característica que aunque no esta ocupada la mayor parte del tiempo, siempre esta a la disposición del hogar.
- **Vivienda auxiliar o secundaria:** Es una vivienda que puede ser utilizada por alguno o varios miembros del hogar con motivo de algunas actividades especiales, que no son de descanso o esparcimiento y que deben desarrollar lejos de su vivienda principal residiendo parte del tiempo en otro lugar, por ejemplo un trabajo fuera de la ciudad de su residencia habitual.

Miembros del hogar: Se consideran miembros del hogar, aquellas personas que comparten y dependen habitualmente del presupuesto de gastos comunes del hogar (colabore económicamente o no en el mismo). Se incluyen los niños y ancianos. Además han residido en la vivienda toda o la mayor parte de los últimos seis meses o han residido en la vivienda encuestada una parte del año mayor que en cualquier otra vivienda del hogar (en caso de que resida alternativamente en distintas viviendas).

Las condiciones que se establecen para determinar si una persona es o no miembro del hogar, tratan de evitar las posibilidades de que un mismo individuo pueda clasificarse en mas de un hogar o por el contrario, no pueda clasificarse en ninguno.

Jefe del hogar: Se considera jefe del hogar aquel miembro del hogar que periódicamente aporta mas al presupuesto o bolsa común que se destina a atender los gastos del hogar.

En la mayoría de los veces es el esposo o la esposa y en ocasiones el jefe del hogar puede ser el abuelo(a) u otro miembro del mismo hogar.

Servicio domestico: El tratamiento dado al personal del servicio domestico depende de las condiciones de su trabajo.

- No duerme en la vivienda: Si el servicio domestico no duerme en la vivienda, no existe ninguna duda de que es miembro de un hogar que tiene como vivienda principal una vivienda diferente a la del hogar en que trabaja. No se considera miembro del hogar.
- Duerme en la vivienda: Si el servicio domestico duerme en la vivienda, se presentan dos casos:
 - Que habite la vivienda donde presta sus servicios, la mayor parte de la semana, pero regularmente o casi todas las semanas se traslada a su hogar de origen y comparte los gastos generales de ese hogar salvo en lo que corresponde a los días que esta trabajando. Compartir los gastos generales , significa que además del aporte que pueda realizar para sufragar los demás gastos, también efectúe una parte relevante de consumo final en dicho hogar. En este caso se considera al servicio domestico miembro de su hogar de origen.
 - Que habite la vivienda donde presta sus servicios la mayor parte de la semana, pero que aun cuando se traslade regularmente a su hogar de origen y ayude en dinero o en especies a ese hogar, no comparte los gastos generales. Se considera en este caso, al servicio domestico como otro hogar separado, que tiene como vivienda principal la vivienda del hogar para el que trabaja. Este hogar estará conformado por el servicio domestico y los familiares del mismo que convivan en la vivienda, independientemente de que también presten servicios domésticos en la misma

Pensionistas: Son aquellas personas que alquilan a otro hogar alguna(s) habitación(es) de la vivienda de dicho hogar y que además de tener el derecho al uso de la misma pueden a veces recibir otros servicios que les son prestados por el hogar arrendador tales como: alimentación, lavado de ropa, etc.

VARIABLES DE ESTUDIO

GASTOS DEL HOGAR

Gastos de consumo final

Para los hogares residentes se registraron los gastos en:

- Bienes y servicios adquiridos y utilizados en el país (caso más frecuente)
- Bienes y servicios adquiridos en el país para su utilización fuera de él (paquetes turísticos, viajes de avión al exterior, etc)
- Bienes y servicios adquiridos fuera del país para su utilización en el país (compra de ropa, vehículo, perfumes, etc.)
- Bienes y servicios adquiridos y utilizados en el exterior (alquiler de vehículos, gastos en hotel y restaurantes durante viajes, etc.)

Gastos de consumo no monetario

- Consumo de la producción por cuenta propia del hogar (autoconsumo), se incluyen todos los bienes y/o servicios producidos por el hogar sin ánimo de ofrecerlos en el mercado y que están destinados para su propio consumo. Abarca los productos primarios de la agricultura, caza, silvicultura y pesca, las manufacturas de esos productos (queso, mantequilla, vinos, telas, calzados, vestidos, muebles, etc.)
- Consumo de bienes y servicios recibidos como ingreso o remuneración en especie, Se incluye el valor de los bienes y/o servicios suministrados por los empleadores a sus asalariados como parte de su remuneración por el trabajo realizado, ejemplo:
 - Alimentos y bebidas
 - Alojamiento de un tipo que puedan utilizarlo todos los miembros del hogar al que pertenece el preceptor
 - Vestimenta utilizable fuera del lugar de trabajo
 - Consumo gratuito de servicios producidos por el empleador para el mercado
 - Servicio de guarderías infantiles gratuitas para los hijos de los trabajadores

- Servicio de vehículo e de otros bienes recibidos en préstamo para uso personal.
 - Servicio de transporte, estacionamiento de vehículos, servicios médicos, etc.
- Consumo de bienes y servicios por retiro del negocio (autosuministro), se incluye el valor de los bienes y/o servicios producidos o comercializados por empresas no constituidas en sociedad, propiedad del hogar que son retirados del negocio, sin realizar ningún pago, por alguno o algunos de los miembros del hogar propietario para su consumo.
 - Consumo de bienes y servicios por donaciones en especie (regalos), se incluye el valor de los bienes y/o servicios que recibe el hogar por parte de otros hogares o de empresas constituidas en sociedad, sin ofrecer a cambio ninguna contraprestación en bienes y/o servicio y/o dinero.

Gastos no imputable al consumo

- Consumo intermedio de bienes y servicios, en su función productiva los hogares utilizan bienes y servicios y factores primarios para obtener bienes y servicios. Los bienes y servicios que los hogares adquieren para el desarrollo de sus actividades productivas constituyen consumo intermedio de las mismas, mientras que los adquiridos para actividades no productivas forman parte de su gasto de consumo.
- Gastos en reparaciones mayores, renovación o ampliación de la vivienda propiedad del hogar.
- Transferencias corrientes en efectivo y/o en especie sin contrapartida, comprende las erogaciones en dinero realizadas por los hogares sin recibir nada como contrapartida, por ejemplo los pagos de cuotas sindicales o para asociaciones políticas, etc
- Intereses por deudas de los hogares, se refiere a por ejemplo: Impuestos directos (impuesto sobre la renta), Primas netas de seguro de riesgo, contribuciones a la seguridad social y a la caja de pensiones, multas y sanciones, etc.

INGRESOS DEL HOGAR

Es el conjunto de ingresos percibidos por los miembros del hogar, preceptores de ingresos, cualesquiera sea su origen y naturaleza, en el periodo de referencia, para hacer frente a sus gastos inmediatos, a sus necesidades futuras o para incrementar su patrimonio.

Atendiendo su naturaleza, los ingresos pueden ser monetarios y no monetarios.

Ingresos monetarios

Se considera ingresos monetarios, los ingresos percibidos por los miembros del hogar en los períodos de referencia:

- Ingreso por trabajo asalariado, es la remuneración total en dinero que pagan las empresas a sus asalariados en contraprestación al trabajo realizado.
- Ingreso del trabajo por cuenta propia, Se refiere a los ingresos netos que reciben los trabajadores por el desarrollo de su actividad por cuenta propia
- Rentas del capital y la propiedad, Se consideran rentas de capital y la propiedad, los ingresos percibidos en concepto de intereses de cuentas corrientes, cuentas de ahorro y depósitos a plazo; intereses y dividendos de acciones, obligaciones, bonos, etc.
- Transferencias recibidas, son los ingresos periódicos percibidos por el hogar sin mediar contraprestación alguna.
- Transferencias extraordinarias y otros ingresos ocasionales anuales, se consideran transferencias extraordinarias, los ingresos percibidos de manera extraordinaria por el hogar,
- tanto de la administración pública como de otros hogares o instituciones, sin que exista contraprestación alguna. Se incluye los premios de loterías, kinos, terminales, etc.

VARIABLES A NIVEL DE VIVIENDA

- Nombre de las personas que habitan la vivienda, tiempo de residencia, ánimo de residencia en la vivienda y su desagregación en hogares.
- Identificación de las personas que prestan servicio domestico al hogar y que residen en la vivienda
- Características de la vivienda:
 - Tipo de vivienda
 - Apariencia de la vivienda (nivel y estado de conservación)
 - Materiales predominantes de los pisos, paredes, etc
 - Antigüedad de la construcción original
 - Número de años transcurridos desde la ultima refacción o remodelación importante
 - Tipo y numero de habitaciones y espacios, en la vivienda, con destinos especiales
 - Número de metros cuadrados construidos
 - Disponibilidad de energía eléctrica, gas directo, etc.
 - Instalaciones y forma de abastecimiento relacionadas con el agua
- Actividades económicas desarrolladas total o parcialmente en la vivienda
 - Hogar/es y miembros participantes
 - Rama de actividad económica
 - Destino de la producción: autoconsumo, transferencias en especies a terceros y de mercado
 - Características del proceso productivo

VARIABLES A NIVEL DE HOGARES Y MIEMBROS

- Características de los miembros del hogar.
 - Nombre
 - Sexo
 - Edad
 - Parentesco o relación con el jefe del hogar
 - Estado Civil
 - Nacionalidad
 - Nivel de educación alcanzado
 - Si esta estudiando, ¿dónde y que estudia?

- Relación de los miembros del hogar con la actividad económica
 - Ocupados o trabajando actualmente, desempleados, buscando trabajo por primera vez, etc.
 - Número de trabajos
 - Situación profesional en la ocupación principal y en otras ocupaciones que desarrolle (patrón o empleador, trabajador por cuenta propia, empleado remunerado, obrero remunerado, trabajador familiar, etc)
 - Ocupación o tipo de trabajo desarrollado
 - Rama de actividad de la empresa donde desarrolla su ocupación principal y otras ocupaciones que realice
 - Horas trabajadas en su ocupación principal y en otras ocupaciones que realice
 - Número de trabajadores que aproximadamente tiene la empresa donde desarrolla su ocupación principal y otras ocupaciones que realice
 - Miembros que reciben algún tipo de pensión o jubilación

- Sistema de previsión social
 - Miembros afiliados o beneficiarios de algún tipo de seguro de salud.
 - Miembros que recibieron asistencia medica u odontológica en el mes anterior a la semana de referencia

- Medios de transporte mas utilizados por los miembros del hogar
 - Vehículo que dispone el hogar
 - Tipo, marca y modelo
 - Año de producción del vehículo
 - Procedencia: Nacional o extranjera
 - Seguro del vehículo

- Hábitos de comidas
 - Tipo de comida (desayuno, almuerzo y cena) que realizan habitualmente en la semana
 - Lugares donde realizan sus comidas
 - Descripción del tipo de platos que incluye normalmente cada tipo de comida

- Combustibles utilizados por el hogar con excepción de los usados en los vehículos automotores
- Dotación y equipamiento del hogar
 - Tipo y numero de equipos de que dispone el hogar
 - Fecha de adquisición

- Vivienda de que dispone el hogar
 - Tipo (principal, secundaria o de descanso)
 - Ubicación: en el país o el extranjero
 - Tenencia: propia, alquilada, ocupada sin pago
 - En el caso de vivienda alquilada: monto del alquiler, quien lo paga

- En el caso de vivienda propia: año de adquisición e información sobre si está o no pagada
- En el caso de ocupación sin pago: información si el uso de la vivienda tiene permiso del propietario. Cual es el motivo: remuneración en especie, donación en especie por parte de instituciones privadas sin fines de lucro u otro hogares
- Valor que se estima tendría el alquiler de la vivienda si se alquilara en el periodo de referencia.

APÉNDICE B
TABLAS Y CAMPOS DE LA BASE DE DATOS

A continuación se presenta en detalle el diseño de cada una de las tablas. Los campos resaltados en negrita conforman la clave principal de la tabla.

Tabla: Región.

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
DOMINIO	Dominio
ENTIDAD	Entidad
Municipio	Municipio
Parroquia	Parroquia
Localidad	Localidad
SEGMENTO	Segmento
SECTOR	Sector
NUMEOCEI	Número de control OCEI
AREA1	Área
NUMESEMA	Numero de Semana
DESDEDIA	Desde el Día
DESDEMES	Desde el Mes
DESDEANO	Desde el Año
HASTADIA	Hasta el Día
HASTAMES	Hasta el Mes
HASTAANO	Hasta el Año
CODIENCU	Código del encuestador
CODISUPE	Código del supervisor
CODIGUBM	Código de la UBM
CODIVIVI	Código de la Vivienda
CONDOCUP	Condición de ocupación
TIPOUSO	Tipo de Uso
PRESENTE	Presenta en la Semana

Tabla: Vivi02.

NOMBRE DEL CAMPO	DESCRIPCIÓN
NUMERBCV	Código de la Encuesta
HOGARNUM	Numero Hogar
MIEMBRON	Numero Miembro

Tabla: Vivi03

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
RESUMENH	Número Hogar
TOTALMIE	Total Miembros
USOVIVIE	Uso de la vivienda
TIPORELA	Tipo de relación
HOGARREL	Hogar relacionado

Tabla: Vivi04

NOMBRE DE CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
TIPOVIVI	Tipo de vivienda
NUMEPISO	Numero de pisos
NUMEAPTO	Numero de Apartamentos
PAREMATE	Material en las paredes
TECHMATE	Material en el techo
PISOMATE	Material en el piso
TIENAGUA	Tiene agua
ABASAGUA	Como se abastece de agua
ELECPUBL	Eléctrico publico
RECOBASU	Recolección directa de basura
CONTBASU	Container de basura
INSTTELE	Instalación para teléfono
GASDIREC	Gas directo
ANTEPARA	Antena parabólica
PUERELEC	Puerta eléctrica
VIGISERV	Servicio de vigilancia
INSTALAR	Instalación de alarma
VIVITIEN	Excretas
ANOCONSU	Años de construcción
ANOREPAR	Años ultima reparación
METROCUA	Metros cuadrados construidos
DORMITOR	Dormitorios
RECIBOAA	Recibo

COMEDORA	Comedor
SALACOME	Sala comedor
COCINAAA	Cocina
BANOCOMP	Baño completo
BANOINCO	Baño incompleto
BANOPREC	Baño precario
BANOSERV	Baño de servicio
BIBLIOTE	Biblioteca
SALAESTA	sala estar
DEPOSITO	Depósitos
MALETERO	Maleteros
LAVANDER	Lavaderos
PATIO	Patio
GARAGE	Garaje
PUESESTA	Puesto de estacionamiento
HABIMULT	Hab. múltiple baño cocina
HCONBANO	Hab. múltiple baño no cocina
HCONCOCI	Hab. múltiple cocina no baño
HSINBACO	Hab. múltiple ni baño ni cocina
SALAFIES	Sala de fiesta
OTROSAAA	Otros
LAVATECH	Lavadero techado
LAVANTEC	Lavadero no techado
PATIOTEC	Patio techado
PATINTAA	Patio no techado
JARDFREN	Jardín al frente
JARDFOND	Jardín al fondo
OTROJARD	Otros jardines
PISCINAA	Piscina
PARQUEAA	Parque infantil
TERRTECH	Terraza techada
TERRNTEC	Terraza no techada
PUESTECH	Puesto techado
MT1A	Metros cuadrados
PUESNTEC	Puesto no techado
MT2A	Metros cuadrados
ESTAVIVA	Estado de la vivienda
APARIENC	Apariencia de la vivienda
MT0A	Metros cuadrados

Tabla: Vivi06

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
ACTIVIDA	Realiza actividad económica
CODIACT1	Código actividad1
CODIACT2	Código actividad2
CODIACT3	Código actividad3
CODIACT4	Código actividad4
PORCACTI	Porcentaje actividad económica
PORTOHOG	Porcentaje total hogares
CODIHOG1	Código hogar1
PORCHOG1	Porcentaje hogar1
CODIHOG2	Código hogar2
PORCHOG2	Porcentaje hogar2
CODIHOG3	Código hogar3
PORCHOG3	Porcentaje hogar3

Tabla: Miembros

NOMBRE DEL CAMPO	DESCRIPCIÓN
NUMERBCV	Código de la encuesta
HOGARNUM	Numero hogar
MIEMBRO	Numero miembro
TOTAHOGA	Total de hogares
PARENTES	Parentesco
FONDCOMU	Fondo común
RESIDIDO	Reside
SEXO	Sexo
EDAD	Edad
SITUCONY	Situación conyugal
NOMBREMI	Nombre y apellido del miembro
IDENNUCL	Identificación del núcleo

NEXONUCL	Nexo del núcleo
NACIONAL	Nacionalidad
LEERESCR	Sabe leer y escribir
NIVEEDUC	Nivel educativo
COMPLETO	Completo
ASISCENT	Asiste a un centro de enseñanza
NIVEASIS	A que nivel asiste
INSTASIS	Tipo de institución
SISTE01	Sistema de previsión1
SISTE02	Sistema de previsión2
SISTE03	Sistema de previsión3
SISTE04	Sistema de previsión4
SISTE05	Sistema de previsión5
REQUASIS	Requirió asistencia medica en el ult mes
LUGARM01	A que lugar asistio1
LUGARM02	A que lugar asistio2
LUGARM03	A que lugar asistio3
LUGARM04	A que lugar asistio4
LUGARM05	A que lugar asistio5
LUGARM06	A que lugar asistio6
ASISODON	Requirió asistencia odontológica
ODONTO01	A que lugar asistio1
ODONTO02	A que lugar asistio2
ODONTO03	A que lugar asistio3
ODONTO04	A que lugar asistio4
MEDIO01	Medio de transporte1
MEDIO02	Medio de transporte2
MEDIO03	Medio de transporte3
MEDIO04	Medio de transporte4
MEDIO05	Medio de transporte5
AYUDA01	Ayuda socio economica1
AYUDA02	Ayuda socio economica2
AYUDA03	Ayuda socio economica3
AYUDA04	Ayuda socio economica4
AYUDA05	Ayuda socio economica5
AYUDA06	Ayuda socio economica6
AYUDA07	Ayuda socio economica7
AYUDA08	Ayuda socio economica8

ASISESPE	Asistió a espectáculos
ESPEC01	Tipo de espectaculo1
ESPEC02	Tipo de espectaculo2
ESPEC03	Tipo de espectaculo3
ESPEC04	Tipo de espectaculo4
PAGA01	Pagado por1
PAGA02	Pagado por2
PAGA03	Pagado por3
PAGA04	Pagado por4
TRASEMAN	Trabajo en la semana anterior
NOESTTRA	Ha trabajado antes
BUSCTRAB	Busca trabajo
TRABADIC	Trabajo adicional
CUANEMPL	Cuantos empleos
CODOCU01	Código ocupación1
CODOCU02	Código ocupación2
CODOCU03	Código ocupación3
TRABA01	Tipo de trabajo1
TRABA02	Tipo de trabajo2
TRABA03	Tipo de trabajo3
EMPRE01	A que se dedica empresa1
EMPRE02	A que se dedica empresa2
EMPRE03	A que se dedica empresa3
TOTAHORA	Total de horas de trabajo semanal
HORA01	Horas semanales trabajo1
HORA02	Horas semanales trabajo2
HORA03	Horas semanales trabajo3
HORA04	Horas semanales trabajo4
PERSO01	Cuantas personas trabajan en empresa1
PERSO02	Cuantas personas trabajan en empresa2
PERSO03	Cuantas personas trabajan en empresa3
JUBILADO	Jubilado

Tabla: Viviendas

NOMBRE DEL CAMPO	DESCRIPCION
------------------	-------------

NUMERBCV	Código de la encuesta
VIVIENUM	Numero de vivienda
HOGARNUM	Numero hogar
TOTAHOGA	Total hogar
TOTAVIVI	Total viviendas
ESTVIVES	Uso de la vivienda
ESTAUBIC	Ubicación
ALOSHOGA	Años del hogar en la vivienda
ESTVIVIE	condición de posesión de la vivienda
AHORROS	% fuente de financiamiento con ahorros
DONAINST	% fuente de financiamiento con donaciones de Inst.
DONATERC	% fuente de financiamiento con donaciones de terceros a hog.
DONAEMPR	% fuente de financiamiento con donación de empresas
PRESTERC	% fuente de financiamiento con préstamo de terceros a hogares
PRESHIPO	% fuente de financiamiento con préstamo hipotecario
PRESBANC	% fuente de financiamiento con bancarios no hipotecarios
OTROPRES	% fuente de financiamiento con otros prestamos
ALOADQUI	año de adquisición
TRASPASO	pago por traspaso
ALOTRASP	año del traspaso
MONTRRAS	monto del traspaso
ELHOGAR	Contribuye el hogar en el pago de alquiler
MONTPAGA	Monto de contribución del hogar
REMUESPE	pago del alquiler por Remuneración en especies
DONAESPE	alquiler donado en especies por Inst. publicas
INSTPRIV	alquiler donado en especies por Inst. privadas
POTROHOG	alquiler donado en especies por otros hogares
OTROSHOG	Contribuye otros hogares en el pago del alquiler
MONOTRHO	Monto de contribución de otros hogares
UNIDECON	Contribuye las unidades económicas que realizan actividades
MONTUNID	Monto de contribución de otros hogares
OTROAGEN	Contribución de otros agentes
MONTTOTA	Monto total del alquiler
PAGGASCO	Paga gastos comunes
MOPROGAS	Monto promedio de gastos comunes
ALQUILAD	Se alquilo con muebles
ADMINIST	Tipo de administración de arrendamiento
MECAAJUS	posee Mecanismo de ajuste
TIEMAJUS	Tiempo de Ajuste

MECANISM	Tipo de mecanismo de ajuste
CONCOCUP	Concepto de ocupación
QUIEPERT	A quien pertenece la vivienda
ELXHOGAR	% de uso para el Hogar
OTRXHOGA	% de uso para otros hogares
PACTHOGA	% de uso para actividades económicas donde part. el hogar
PACTNHOG	% de uso para actividades económicas donde no part. el hogar
MONTMINI	Monto mínimo de posible alquiler
MONTMAXI	Monto máximo de posible alquiler
MONTPROB	Monto probable de posible alquiler

Tabla: Vehículo.

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Numero dela encuesta
HOGARNUM	Numero hogar
NUMEVEHI	Numero vehículo
CLASVEHI	Clase del vehículo
MARCVEHI	Marca
MODEVEHI	Modelo
CILIVEHI	Cilindrada
APROVEHI	Año de producción
ELVEHIES	Nacionalidad
PROPVEHI	Propietario
MESADVEH	Mes de adquisición
ALOADVEH	Año de adquisición
SEADVEHI	Se adquirió nuevo o usado
PORCVEHI	% de uso del hogar
POOTVEHI	% de uso otros hogares
POEMVEHI	% de uso por empresas del hogar
POTRVEHI	% de uso por otras empresas
SEGUVEHI	Paga seguro
MATRVEHI	Matriculado en

Tabla: Actividad.

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la Encuesta
HOGARNUM	Numero hogar
CODXACTI	Código actividad
MIEMPART	Miembros que participan
REAXHOGA	Realiza Actividad económica
AUTOCONS	% para autoconsumo
MERCADO	% para mercado
TRANSFER	% para transferencia

Tabla: Dotación

NOMBRE DEL CAMPO	DESCRIPCIÓN
NUMERBCV	Código de la encuesta
HOGARNUM	Numero Hogar
CODIEQUI	Código del equipo
CANTEQUI	Cantidad de equipo
PERTEQUI	Propiedad del equipo
MESADEQU	Mesa de adquisición
ALOADEQU	Año de adquisición

Tabla: Otros servicios

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
HOGARNUM	Numero hogar
CUANCUAR	Cuantos cuartos
EXCLBALO	Uso exclusivo de baño
CUANTBAL	cuantos baños
GASDIREC	Gas directo
GASBOMBO	Kerosén
KEROSENE	Gas por bombona
ELECTRIC	Electricida
OTROSCOM	Otros

ANTEPARA	Antena parabólica
TELECABL	Televisión por cable
TELEFONO	Teléfono
TELECELU	Teléfono celular
HOGARFAX	Fax
CORREOEL	Correo electrónico

Tabla: Gastos diarios del hogar.

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de encuesta
HOGARNUM	Numero hogar
NUMCOR01	Numero correlativo
COD_EPF	Código del gasto
MIEMBRON	Numero miembro
CANTGAST	Cantidad del gasto
CONTGAST	Contenido
CODUNME	Código de la unidad de medida
LUGORI01	Lugar y Origen del gasto
QUIPAG01	Código de quien paga
VAPAH001	Valor pagado por el hogar
FORPAG01	Forma de pago
VALTOT01	Valor total del gasto
CUADES01	Código de quien lo uso
PORDES01	Porcentaje consumido por el hogar
TIPEST01	Tipo de establecimiento
PROCOM01	Productor o Comerciante

Tabla: Gastos de hogar en restaurantes.

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
HOGARNUM	Numero hogar
NUMCOR01	Numero correlativo

COD_EPF	Código del gasto
MIEMBRON	Numero miembro
QUIPAG01	Código de quien paga
VAPAH001	Valor pagado por hogar
FORPAG01	Forma de pago
VALTOT01	Valor total
CUADES01	Código de quien lo consumió
PORDES01	Porcentaje consumido por el hogar
TIPEST01	Tipo de establecimiento
TIPOCOMI	Tipo de comida
CONSCOMP	Consumida o comprada
PARTTOTA	Total participantes
PARTHOGA	Participantes del hogar
LUGAR	Lugar

Tabla: Gastos diarios personales

NOMBER DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
HOGARNUM	Numero hogar
COD_EPF	Código del gasto
NUMCOR01	Numero correlativo
MIEMBRON	Numero miembro
CANTGAST	Cantidad del gasto
CONTGAST	Contenido del gasto
CODUNME	Código de la unidad de medida
LUGORI01	Lugar
QUIPAG01	Código quien paga
VAPAH001	Valor pagado por hogar
FORPAG01	Forma de pago
VALTOT01	Valor total
CUADES01	Código de quien lo uso
PORDES01	Porcentaje consumido por el hogar
TIPEST01	Tipo de establecimiento
PROCOM01	Productor comerciante

Tabla: Gastos personales en restaurante.

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
COD_EPF	Código del gasto
MIEMBRON	Numero miembro
NUMCOR01	Numero correlativo
HOGARNUM	Numero hogar
QUIPAG01	Código quien paga
VAPAH001	Valor pagado por hogar
FORPAG01	Forma de pago
VALTOT01	Valor total del gasto
CUADES01	Código de quien lo uso
PORDES01	Porcentaje consumido por el hogar
TIPEST01	Tipo de establecimiento
TIPOCOMI	Tipo de comida
CONSCOMP	Consumida o comprada
PARTTOTA	Total participantes
PARTHOGA	Participantes del hogar
LUGAR	Lugar

Tabla: Gastos Mensuales.

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
HOGARNUM	Numero hogar
MESFECHA	mes
ALOFecha	año
DIAENTRE	Día de entrevista
MESENTRE	Mes de entrevista
ALOENTRE	Año de entrevista
MSGAS01	Mes del gasto

ALOGAS01	Año del gasto
COD_EPF	Código del gasto
HOMUNI01	Hombre mujer niño
CANTID01	Cantidad
NUEUSA01	Nuevo o usado
LUGORI01	Lugar y origen
QUIPAG01	Código de quien paga
VAPAH001	Valor pagado por hogar
FORPAG01	Forma de pago
VALTOT01	Valor total
CRECIM01	Cuota inicial monto
CREIPM01	Inicio de los pagos mes
CREIPA01	Inicio de los pagos años
CREFPM01	Fin de los pagos mes
CREFPA01	Fin de los pagos año
CRENCU01	Numero de cuotas
CREMPR01	Monto promedio de las cuotas
CUADES01	Código cual es el destino
PORDES01	% del destino
TIPEST01	Tipo de establecimiento
PROCOM01	Productor comerciante
AUSPR101	Código ausencia presencia1
AUSPR201	Código ausencia presencia2
SEPAGO01	Se pago
NUMCOR01	Numero correlativo
CARSER01	Características del Servicio
CODUNI01	Unidad de medida
USAPAR01	Usado para
CODVIV01	
CANHOR01	Cantidad de horas semanales trabajadas
ESMIEM01	Es miembro
MIEMBR01	Numero de miembro
NUMHOR01	Numero de horas
NUMMES01	Numero de meses
SERINC01	Servicio incluye
INSTIT01	Tipo de institución
SERIO201	Serio1
SERIO301	Serio2

SERI0401	Serio3
PERTOT01	Personas cubiertas por el seguro Total
PERHOG01	Personas cubiertas por el seguro del hogar
PARTOT01	Numero de participantes total
PARHOG01	Numero de participantes del hogar
TOUR0101	Tour1
TOUR0201	Tour2
TOUR0301	Tour3
TOUR0401	Tour4
TOUR0501	Tour5
TOUR0601	Tour6
TOUR0701	Tour7
TOUR0801	Tour8

Tabla: Gastos Trimestrales y anuales

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
HOGARNUM	Numero Hogar
MESTRIDE	Mes de inicio trimestre
ALOTRIDE	Año de inicio trimestre
MESTRIHA	Mes final trimestre
ALOTRIHA	Año final trimestre
MESANUDE	Mes de inicio anual
ALOANUDE	Año de inicio anual
MESANUHA	Mes final anual
ALOANUHA	Año final Anual
MESGAS01	Mes del gasto
ALOGAS01	Año del gasto
COD_EPF	Código del gasto
CANTID01	Cantidad
NUEUSA01	Nuevo usado
LUGORI01	Lugar y origen
QUIPAG01	Código de quien paga
VAPAHO01	Valor pagado por el hogar
FORPAG01	Forma de pago

VALTOT01	Valor total
CRECIM01	Cuota inicial del monto
CREIPM01	Inicio de los pagos mes
CREIPA01	Inicio de los pagos año
CREFPM01	Fin de los pagos mes
CREFPA01	Fin de los pagos años
CRENCU01	Numero de cuotas
CREMPR01	Monto promedio de la cuota
CUADES01	Código cual es el destino
PORDES01	Porcentaje del destino
TIPEST01	Tipo de establecimiento
PROCOM01	Productor comerciante
AUSPR101	Código ausencia presencia1
AUSPR201	Código ausencia presencia2
SEPAGO01	Se pago
NUMCOR01	Numero correlativo
CARSER01	Características del servicio
CODUNI01	Unidad de medida
USAPAR01	Usado para
NUMVEH01	Numero del vehículo
NUMVIA01	Numero de viaje
MESDES01	Mes desde
ALODES01	Año desde
MESHAS01	Mes hasta
ALOHAS01	Año hasta
CODVIA01	Código del viaje
TOTPAR01	Numero de participantes total
HOGPAR01	Numero de participantes del hogar
TOUR0101	Tour1
TOUR0201	Tour2
TOUR0301	Tour3
TOUR0401	Tour4
TOUR0501	Tour5
TOUR0601	Tour6
ESMIEM01	Es miembro
MIEMBR01	Miembro
NUMVIV01	Numero de vivienda
MONSER01	Monto del servicio

MONMAT01	Monto de materiales
LUGTRA01	Lugar de transferencia
QUIVEN01	A quien se vende
GASALI01	Gasto alimentación
LUGALI01	Lugar Alimentación
GASRES01	Gasto restaurantes
LUGRES01	Lugar restaurante
GASREC01	Gasto recreación
LUGREC01	Lugar recreación
GASTRA01	Gasto Transporte
GASCOM01	Gasto comunicación
LUGCOM01	Lugar comunicación
GASOTR01	Otros gastos
LUGOTR01	lugar otros gastos

Tabla: Ocupación

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la Encuesta
HOGARNUM	Numero hogar
MIEMBRON	Numero miembro
MESINGRE	Mes ingreso
ALOINGRE	Año ingreso
EMPLEOS	Cuantos empleos
OCUPA01	Ocupación1
OCUPA02	Ocupacion2
OCUPA03	Ocupacion3
ACTIVI01	Actividad1
ACTIVI02	Actividad2
ACTIVI03	Actividad3

Tabla: Ingresos

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta

HOGARNUM	Numero hogar
MIEMBRON	Numero miembro
MESINGRE	Mes ingreso
ALOINGRE	Año ingreso
CODRUB01	Rubro
MARQUE01	Ingresos
OCUPRI01	Ocupación principal
OCUSEC01	Ocupación secundaria
OCUTER01	Ocupación Terciaria

Tabla: Transferencia

NOMBRE DEL CAMPO	DESCRIPCION
NUMERBCV	Código de la encuesta
HOGARNUM	Numero hogar
MIEMBRON	Numero miembro
MESINGRE	Mes ingreso
ALOINGRE	Año ingreso
codrub01	Código de concepto
ocupri01	Monto de transferencia

APÉNDICE C

CUADRO DE DISTRIBUCIÓN DE FRECUENCIA DE LA ACTIVIDAD ECONOMICA REALIZADA EN LA VIVIENDA

Actividad Económica	frec.	Porc
ACTIVIDADES AUXILIARES DE LA INTERMEDIACION FINANCIERA	1	.1
ACTIVIDADES DE ASOCIACIONES NCP	2	.1
ACTIVIDADES DE EDICION E IMPRESION Y DE REPRODUCCION DE GRABACIONES	5	.3
ACTIVIDADES DE ESPARCIMIENTO Y ACTIVIDADES CULTURALES Y DEPORTIVAS	18	1.0
ACTIVIDADES DE TRANSPORTE COMPLEMENTARIAS Y AUXILIARES, ACTIVIDADES DE AGENCIAS	1	.1
ACTIVIDADES INMOBILIARIAS	25	1.4
ADMINISTRACIÓN PUBLICA Y DEFENSA; PLANES DE SEGURIDAD SOCIAL DE AFILIACION OBLIG	3	.2
AGRICULTURA, GANADERIA CAZA Y ACTIVIDADES DE SERVICIOS CONEXAS	18	1.0
ALQUILER DE MAQUINARIA Y EQUIPOS SIN OPERARIOS Y DE EFECTOS PERSONALES Y ENSERES	2	.1
COMERCIO AL POR MAYOR Y EN COMISION, EXCEPTO EL COMERCIO DE VEHICULOS AUTOMOTORE	15	.9
COMERCIO AL POR MENOR, EXCEPTO EL COMERCIO DE VEHICULOS AUTOMOTORES Y MOTOCICLET	647	37.3
CONSTRUCCION	5	.3
CURTIDO Y ADOBO DE CUEROS, FABRICACION DE MALETAS, BOLSOS DE MANO, ARTICULOS DE	14	.8
ELABORACION DE PRODUCTOS ALIMENTICIOS Y BEBIDAS	287	16.5
ELABORACION DE PRODUCTOS DE TABACO	1	.1
ENSEÑANZA	48	2.8
FABRICACION DE EQUIPOS Y APARATOS DE RADIO TELEVISION Y COMUNICACIONES	1	.1
FABRICACION DE MAQUINARIA Y APARATOS ELECTRICOS	1	.1
FABRICACION DE MUEBLES; INDUSTRIA MANUFACTURERA NCP	16	.9
FABRICACION DE OTROS METALES COMUNES	2	.1
FABRICACION DE OTROS PRODUCTOS MINERALES NO METALICOS	31	1.8
FABRICACION DE PAPEL Y PRODUCTOS DE PAPEL	3	.2
FABRICACION DE PRENDAS DE VESTIR, ADOBO Y TEÑIDO DE PIELS	176	10.1
FABRICACION DE PRODUCTOS ELABORADOS DE METAL, EXCEPTO MAQUINARIA Y EQUIPO	5	.3
FABRICACION DE SUBSTANCIAS Y PRODUCTOS QUIMICOS	2	.1
FABRICACION DE TEXTILES	36	2.1
FINANCIAMIENTO DE PLANES DE SEGURO Y DE PENSIONES, EXCEPTO LOS PLANES DE SEGURID	2	.1
HOGARES PRIVADOS CON SERVICIO DOMESTICO	1	.1
HOTELES Y RESTAURANTES	77	4.4
INFORMATICA Y ACTIVIDADES CONEXAS	8	.5
INVESTIGACION Y DESARROLLO	1	.1
OTRAS ACTIVIDADES DE SERVICIOS	159	9.2
OTRAS ACTIVIDADES EMPRESARIALES	55	3.2
PESCA, EXPLOTACION DE CRIADEROS DE PECES Y GRANJAS PISCICOLAS, ACTIVIDADES DE SE	5	.3
PRODUCCION DE MADERA Y FABRICACION DE PRODUCTOS DE MADERA Y CORCHO, EXCEPTO MUEB	8	.5
SERVICIOS SOCIALES Y DE SALUD	34	2.0
SUMINISTROS DE ELECTRICIDAD, GAS VAPOR Y AGUA CALIENTE	1	.1
VENTA, MANTENIMIENTO Y REPARACION DE VEHICULOS AUTOMOTORES Y MOTOCICLETAS; VENTA	19	1.1
Total	1735	100.0

APÉNDICE D
ANÁLISIS DE CORRESPONDENCIA : APARIENCIA DE VIVIENDA VS TIPO DE VIVIENDA

Autovalores y porcentaje de inercia explicada por cada factor

Numero	Autovalores	Porcentaje	Porcentaje Acumulado
1	0.6239	78.05	78.05
2	0.1674	20.94	98.99
3	0.0080	1.00	99.99
4	0.0000	0.01	100.00

Contribuciones y calidad de representación de las modalidades de la variable tipo de vivienda

Modalidades	Contribuciones de la modalidad al factor				Calidad de representación			
	1	2	3	4	1	2	3	4
QC	2.4	26.8	2.6	0.8	0.25	0.75	0.00	0.00
EcA	4.0	69.3	6.6	0.0	0.18	0.82	0.00	0.00
EsA	0.8	1.8	90.3	0.7	0.34	0.20	0.46	0.00
LncV	0.3	0.2	0.0	6.0	0.85	0.15	0.00	0.00
Anex	0.1	0.0	0.5	75.1	0.86	0.04	0.06	0.05
Pev	0.1	0.1	0.0	2.2	0.87	0.13	0.00	0.00
VR	67.2	1.1	0.0	0.4	1.00	0.00	0.00	0.00
RC	25.0	0.1	0.0	0.7	1.00	0.00	0.00	0.00
Otr	0.0	0.6	0.0	14.2	0.03	0.96	0.00	0.01

Contribuciones y calidad de representación de las modalidades de la variable apariencia de la vivienda

Modalidades	Contribuciones				Calidad de representación			
Lujosa	0.1	0.0	6.8	92.2	0.41	0.00	0.55	0.04
Muy Buena	1.1	8.3	76.0	7.1	0.25	0.52	0.23	0.00
Buena	5.3	40.7	17.1	0.2	0.32	0.66	0.01	0.00
Modesta	1.8	50.2	0.0	0.4	0.12	0.88	0.00	0.00
Precaria	91.7	0.7	0.0	0.0	1.00	0.00	0.00	0.00

ANALISIS DE CORRESPONDENCIA : ZONA GEOGRAFICA VS TIPO DE VIVIENDA

Autovalores y porcentaje de inercia explicada por cada factor

Numero	Autovalor	Porcentaje	Porcentaje acumulado
1	0.1149	75.50	75.50
2	0.0211	13.89	89.39
3	0.0094	6.19	95.58
4	0.0067	4.42	100.00

Contribuciones y calidad de representación de las modalidades de la variable tipo de vivienda

Modalidades	Contribuciones				Calidad de representación			
	1	2	3	4	1	2	3	4
QC	17.6	2.0	2.5	8.1	0.94	0.02	0.01	0.03
EcA	77.9	0.1	0.4	0.9	1.00	0.00	0.00	0.00
EsA	0.2	16.1	7.0	62.1	0.03	0.40	0.08	0.49
LncV	0.9	0.0	5.2	0.3	0.67	0.00	0.32	0.01
Anex	0.3	0.0	7.7	0.0	0.31	0.01	0.68	0.00
Pev	0.0	0.3	0.8	3.1	0.03	0.19	0.20	0.58
VR	2.1	67.0	16.6	1.7	0.13	0.78	0.09	0.01
RC	0.1	8.9	7.6	17.6	0.04	0.48	0.18	0.30
Otr	0.9	5.5	52.1	6.2	0.14	0.16	0.65	0.06

Contribuciones y calidad de representación de las modalidades de la variable zona geográfica

Modalidades	Contribuciones				Calidad de representación			
	1	2	3	4	1	2	3	4
Centro	54.9	0.2	0.0	3.0	1.00	0.00	0.00	0.00
Occidente	7.5	12.0	11.6	42.0	0.57	0.17	0.07	0.19
Oriente	9.5	74.1	3.0	2.5	0.40	0.58	0.01	0.01
Llanos	16.0	0.0	27.8	47.6	0.76	0.00	0.11	0.13
Amazonas	12.1	13.7	57.5	4.9	0.62	0.13	0.24	0.01

**ANALISIS DE CORRESPONDENCIA : SISTEMA DE PREVISIÓN SOCIAL VS
NIVEL EDUCATIVO**

Autovalores y porcentaje de inercia explicada por cada factor

Numero	Autovalor	Porcentaje	Porcentaje acumulado
1	0.1539	94.17	94.17
2	0.0083	5.05	99.23
3	0.0011	0.68	99.90
4	0.0002	0.10	100.00

Contribuciones y calidad de representación de las modalidades de la variable nivel educativo

Modalidades	Contribuciones				Calidad de representación			
	1	2	3	4	1	2	3	4
Preescolar	0.0	0.8	0.1	0.6	0.41	0.57	0.01	0.01
Básica	29.3	0.2	0.4	0.0	1.00	0.00	0.00	0.00
Media Divers	1.2	4.9	1.0	11.7	0.81	0.18	0.00	0.01
Técnico Superior	4.1	47.8	0.0	0.0	0.61	0.39	0.00	0.00
Universitario	38.9	15.3	24.2	1.4	0.97	0.02	0.00	0.00
Postgrado	15.3	2.5	69.9	0.0	0.96	0.01	0.03	0.00
INCE	0.1	5.2	0.5	82.5	0.19	0.62	0.01	0.18
Sin Nivel	11.2	23.2	4.0	3.8	0.90	0.10	0.00	0.00

Contribuciones y calidad de representación de las modalidades de la variable sistema de previsión social

Modalidades	Contribuciones				Calidad de representación			
	1	2	3	4	1	2	3	4
Seguro social obligatorio	13.7	47.8	0.6	0.0	0.84	0.16	0.00	0.00
Seguro privado	21.3	17.9	3.3	53.4	0.95	0.04	0.00	0.00
Seguro colectivo	9.7	15.9	64.0	7.6	0.88	0.08	0.04	0.00
Otros	14.5	10.1	32.0	38.9	0.95	0.04	0.02	0.00
Ninguno	40.8	8.3	0.1	0.1	0.99	0.01	0.00	0.00

ANALISIS DE CORRESPONDENCIA MULTIPLE: NIVEL EDUCATIVO, TIPO DE VIVIENDA, ZONA, SISTEMA DE PREVISION SOCIAL Y NUMEROS DE EMPLEO

Contribuciones y calidad de representación de las modalidades de las distintas variables

Modalidades de nivel educativo	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
Basica	8.8	0.2	4.8	1.9	0.2	0.31	0.00	0.11	0.04	0.00
Media Divers	0.4	2.8	0.6	0.9	3.7	0.01	0.04	0.01	0.01	0.05
Tecnico Superior	1.1	5.4	0.6	7.7	1.1	0.02	0.07	0.01	0.09	0.01
Universitario	11.9	0.8	1.4	0.9	2.9	0.28	0.01	0.02	0.01	0.04
Postgrado	5.1	1.4	0.4	1.0	22.7	0.10	0.02	0.00	0.01	0.25
Sin Nivel	3.8	17.6	5.4	3.7	0.2	0.08	0.24	0.07	0.04	0.00
TOTAL	31.1	28.1	13.2	16.1	30.9					

Modalidades de tipo de vivienda	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
Quinta o Casa	4.3	1.5	0.0	6.8	0.2	0.26	0.06	0.00	0.24	0.01
Edificio con Asc	19.3	3.6	1.0	0.4	2.5	0.47	0.06	0.01	0.01	0.03
Edificio sin Asc	0.4	6.0	16.4	13.6	0.3	0.01	0.08	0.21	0.16	0.00
Vivienda Rustica	3.8	12.3	4.2	16.8	1.1	0.08	0.17	0.05	0.20	0.01
TOTAL	27.8	23.3	21.7	37.6	4.1					

Modalidades de zona	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
Centro	8.7	1.9	4.6	0.0	0.0	0.29	0.04	0.09	0.00	0.00
Nor Occ	1.2	0.4	16.7	0.0	1.5	0.03	0.01	0.27	0.00	0.02
Oriente	2.6	6.1	0.8	6.6	3.3	0.06	0.09	0.01	0.08	0.04
Occidente	2.5	0.3	0.2	22.3	0.0	0.05	0.00	0.00	0.27	0.00
Amazonas y Bolivar	1.0	12.4	0.9	1.9	0.1	0.02	0.18	0.01	0.02	0.00
TOTAL	16.0	21.2	23.1	30.9	5.0					

Modalidades de sistema de previsión social	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
SSO	3.2	7.3	0.8	3.4	6.2	0.10	0.15	0.02	0.06	0.11
SP	6.7	4.2	2.8	0.6	14.3	0.13	0.06	0.03	0.01	0.16
SC	1.8	2.4	2.9	0.5	1.6	0.04	0.03	0.04	0.01	0.02
Otr	1.2	0.1	21.0	4.3	0.0	0.02	0.00	0.26	0.05	0.00
Ning	8.7	1.7	0.2	0.3	2.0	0.34	0.04	0.01	0.01	0.04
TOTAL	21.7	15.6	27.8	9.2	24.2					

Modalidades de numero de empleos	Contribuciones					Calidad de representación				
	1	2	3	4	5	1	2	3	4	5
un Empleo	0.0	2.4	2.5	0.9	3.7	0.00	0.12	0.12	0.04	0.16
dos Empleo	2.2	0.3	0.0	5.1	32.0	0.05	0.00	0.00	0.06	0.38
Sin Empleo	1.1	9.1	11.8	0.1	0.1	0.03	0.14	0.17	0.00	0.00
TOTAL	3.4	11.9	14.3	6.2	35.8					