



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

EL GASTO ALIMENTARIO DE LOS
HOGARES CON ADULTOS MAYORES

(Un análisis con el Método de Componentes Principales)

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

PRESENTA:

MARTÍN GALICIA VELÁZQUEZ

TUTORA:

M. EN C. MARÍA ISABEL MONTERRUBIO GÓMEZ



2009



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de datos del jurado

1. Datos del alumno

Galicia
Velázquez
Martín
58424413
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
082039774

2. Datos del tutor

M. en C.
María Isabel
Monterrubio
Gómez

3. Datos del sinodal 1

Dra.
María Edith
Pacheco
Gómez Muñoz

4. Datos del sinodal 2

Mat.
Margarita Elvira
Chávez
Cano

5. Datos del sinodal 3

M. en D.
María Teresa
Velázquez
Uribe

6. Datos del sinodal 4

M. en C.
María Teresa Leonor
Muciño
Porras

7. Datos del trabajo escrito

El gasto alimentario de los hogares con adultos mayores
(Un análisis con el Método de Componentes Principales)
154 p
2009

DEDICATORIA

Al **creador** (Dios).

Por permitirme nacer, llenarme de salud, comprensión, sabiduría, bondad, amor y dejarme llegar a este punto con el fin de lograr un objetivo más.

A mis **padres** (P. Margarita y F. Martín).

Quienes me dieron vida, cuidaron con su esfuerzo, su amor, comprensión y que motivaron con sus enseñanzas durante todo este tiempo, para llegar a esta meta.

A mi **esposa** e **hijos** (Patricia, Jonathan, Margarita y Diana).

Por su colaboración, consejos, motivación, comprensión y amor para poder realizar este trabajo.

A mis **hermanos** (Isabel, Sonia, Antonio y Arturo).

Quienes influyeron con su apoyo incondicional y comprensión en todas aquellas situaciones surgidas durante el desarrollo de mis estudios.

A mis **abuelos** (Juan e Isabel).

Las personas más nobles y bondadosas que inculcaron parte de su sabiduría y conocimientos de mi vida y que marcaron parte de mi camino.

A mis **amigos** [Guillermina González, Elvira Fragoso, Virginia Abrin, Rosa Hernández, Beatriz Vargas, Alicia, Eduardo, María, Alejandra, Liliana, Inés, José y M. Elena, compañeros de trabajo del INEGI, de la DGASNIEG, DEN (hoy DCNSNIEG), del CIMAT], quienes fueron en forma indirecta o directa un gran apoyo dentro de mis estudios, y las personas importantes para la realización de este trabajo, por sus consejos, apoyo y otras motivaciones.

A mis **asesores** (María Edith Pacheco Gómez Muñoz, Margarita Chávez Cano, María Teresa Muciño Porras, María Teresa Velázquez Uribe) por sus valiosas apreciaciones, sugerencias, comentarios y detalles encontrados para la mejora de este trabajo.

Al maestro **José Vences Rivera** por sus valiosos comentarios y observaciones identificadas en este trabajo.

Y en especial, a **mi directora** de tesis **María Isabel Monterrubio Gómez**, por su confianza, disposición, su paciencia, su tiempo, su voluntad, la transmisión de experiencias, los consejos, el apoyo motivacional, los valiosos señalamientos y las observaciones detectadas para la estructura y contenido en la elaboración del trabajo.

A la **Universidad Nacional Autónoma de México**, como alma mater profesional y en particular a la **Facultad de Ciencias** por permitirme formar parte de una más, de las generaciones de profesionistas productivos.

AGRADECIMIENTOS

A todos (Dios, Papá, Mamá, Esposa, Hijos, Hermanos, Tíos, Directora de Tesis, Asesores, Amigas(os), Compañeras(os) de trabajo, Profesores y más) los que en algún momento estuvieron y están cerca de mí, que me brindaron(dan) sus experiencias, enseñanzas, consejos, motivaciones, tiempo, dedicación, paciencia y sabiduría... que tendieron las manos... que me proporcionaron todo tipo de herramientas... o que me facilitaron información (en especial al INEGI), para y en la realización de este trabajo... gracias... desde lo más profundo de mi corazón... muchas gracias a todos.

Martín

Recordar que:

*“La mayor sabiduría que existe es conocerse a sí mismo” (uno sabe hasta dónde llegar...uno decide cuanto y en qué tiempo se alcanza) y por otro lado tener presente que... “las matemáticas son el alfabeto con el cual Dios ha escrito el universo” (y donde las estadísticas son oraciones creadas dentro de ese entorno)...frases alusivas de.... **Galileo Galilei (1564-1642)** y pensamientos propios.*

ÍNDICE

| | Pág. |
|--|-----------|
| Introducción | 1 |
| Capítulo 1. Antecedentes | 7 |
| Capítulo 2. Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH) | 13 |
| Capítulo 3. Metodología de Generación de Información | 21 |
| Capítulo 4. Antecedentes de los Componentes Principales | 29 |
| Capítulo 5. El gasto de los hogares con adultos mayores | 45 |
| 5.1 Hogares con al menos un adulto de 65 y más años de edad | 47 |
| 5.2 Hogares con exclusivamente población de 65 y más años de edad | 80 |

| | |
|---|------------|
| Capítulo 6. Conclusiones y consideraciones finales | 105 |
| Anexos | 113 |
| Glosario | 133 |
| Bibliografía | 137 |

Índice de cuadros y gráficas

Cuadros (Parte 1)

| Número | Título | Página |
|---------------------|--|--------|
| Cuadro 1.1 | Indicadores seleccionados de la población 1930 a 2006 | 11 |
| Cuadro 2.1 | Bases de datos de la ENIGH | 18 |
| Cuadro 3.1 | Estructura de la base de datos Gasto65.dbf | 26 |
| Cuadro 3.2 | Estructura de la base de datos Gasto65e.dbf | 28 |
| Cuadro 5.1 | Total del gasto por tipo de alimento según entidad federativa, 2006 (Pesos) | 48 |
| Cuadro 5.2 | Total del gasto por tipo de alimento según entidad federativa, 2006 (Logaritmos) | 49 |
| Cuadro 5.3 | Principales estadísticas de las trece variables de alimentos | 52 |
| Cuadro 5.4 | Prueba de normalidad con la estadística de Shapiro -Wilks | 60 |
| Cuadro 5.5 | Prueba de normalidad con la estadística de Shapiro -Wilks (valores en logaritmos) | 62 |
| Cuadro 5.6 | Matriz de correlaciones de las variables del gasto en alimentos | 65 |
| Cuadro 5.7 | Matriz de varianzas y covarianzas de las variables del gasto en alimentos | 67 |
| Cuadro 5.8 | Vector de valores propios de las variables del gasto en alimentos | 68 |
| Cuadro 5.9 | Matriz de vectores propios de las variables del gasto en alimentos | 69 |
| Cuadro 5.10 | Desviación estándar por componente principal | 70 |
| Cuadro 5.11 | Acumulación de la varianza por componente principal (Acumulación de componentes) | 70 |
| Cuadro 5.12 | Correlaciones entre las componentes principales(C) y las variables originales(X) | 71 |
| Cuadro 5.13 Parte 1 | Puntuaciones (scores) de las componentes principales 1 y 2 según entidades federativas | 73 |
| Cuadro 5.13 Parte 2 | Puntuaciones (scores) de las componentes principales 1 y 2 según entidades federativas | 74 |

Cuadros (Parte 2)

| Número | Título | Página |
|-------------------------|---|--------|
| Cuadro 5. 14 Parte 1 | Puntuaciones (scores) de las entidades federativas en los 13 componentes principales | 77 |
| Cuadro 5. 14 Parte 2 | Puntuaciones de las entidades federativas en los 13 componentes principales | 78 |
| Cuadro 5. 15 | Gasto de los hogares con exclusivamente residentes de 65 y más años de edad por tipo de alimento | 81 |
| Cuadro 5. 16 | Gasto de los hogares con exclusivamente residentes de 65 y más años de edad por tipo de alimento (valores en logaritmo) | 82 |
| Cuadro 5. 17 | Principales estadísticas de las trece variables de alimentos | 83 |
| Cuadro 5. 18 | Prueba de Normalidad de Shapiro – Wilks | 87 |
| Cuadro 5. 19 | Principales estadísticas (valores en logaritmos) | 90 |
| Cuadro 5. 20 | Prueba de Normalidad de Shapiro-Wilks(valores en logaritmos) | 93 |
| Cuadro 5. 21 Parte 1 | Matriz de varianzas y covarianzas | 96 |
| Cuadro 5. 21 Parte 2 | Matriz de varianzas y covarianzas | 97 |
| Cuadro 5. 22 Parte 1 | Matriz de correlaciones | 97 |
| Cuadro 5. 22 Parte 2 | Matriz de correlaciones | 98 |
| Cuadro 5. 23 | Valores Propios | 98 |
| Cuadro 5. 24 | Vectores Propios | 99 |
| Cuadro 5. 25 | Desviaciones estándar por componente | 100 |
| Cuadro 5. 26 | Acumulación de varianza por componente | 100 |
| Cuadro 5. 27 | Correlaciones entre las componentes principales(C) y las variables originales(X) | 101 |
| Cuadro 5. 28 | Puntuaciones de los componentes | 103 |

Gráficas (Parte 1)

| Número | Título | Página |
|--------------|--|--------|
| Gráfica 5.1 | Concentración de Carnes, Cereales, Verduras y Leche | 54 |
| Gráfica 5.2 | Concentración de Otros alimentos, Frutas, Pescado y Huevo | 55 |
| Gráfica 5.3 | Concentración de Azúcar, Aceite, Tubérculos, Café y Especias | 55 |
| Gráfica 5.4 | Gráfica multivariada de las variables del gasto | 58 |
| Gráfica 5.5 | Gráfica de cuartiles de las variables del gasto | 59 |
| Gráfica 5.6 | Gráfica de probabilidad Ji cuadrada de cuartiles | 59 |
| Gráfica 5.7 | Gráfica multivariada de las variables del gasto (valores en logaritmos) | 60 |
| Gráfica 5.8 | Gráficas de cuartiles de las variables del gasto (valores en logaritmos) | 61 |
| Gráfica 5.9 | Gráfica de probabilidad Ji cuadrada de cuartiles (valores en logaritmos) | 61 |
| Gráfica 5.10 | Tendencia de componentes (Gráfica de rodillo o codo) | 73 |
| Gráfica 5.11 | Concentración del gasto de las entidades según componentes 1 y 2 | 75 |
| Gráfica 5.12 | Distribución del gasto según variables del gasto | 84 |
| Gráfica 5.13 | Distribución del gasto de los hogares según variables en forma multivariada | 85 |
| Gráfica 5.14 | Distribución del gasto de los hogares (caras de Chernoff) | 86 |
| Gráfica 5.15 | Distribución del gasto según variables de alimentos (gráfica de cuartiles) | 88 |
| Gráfica 5.16 | Gráfica de probabilidad Ji cuadrada | 89 |
| Gráfica 5.17 | Distribución del gasto por variable (valores en logaritmos) | 91 |
| Gráfica 5.18 | Distribución de del gasto de los hogares según variables en forma multivariada (valores en logaritmos) | 91 |

Gráficas (Parte 2)

| Número | Título | Página |
|--------------|---|--------|
| Gráfica 5.19 | Distribución del gasto de los hogares en logaritmos (caras de Chernoff) | 92 |
| Gráfica 5.20 | Distribución del gasto según variables de alimentos (valores en logaritmos) | 94 |
| Gráfica 5.21 | Gráfica de probabilidad Ji cuadrada (valores en logaritmos) | 95 |
| Gráfica 5.22 | Tendencia de componentes | 102 |
| Gráfica 5.23 | Concentración de las entidades según componentes 1 y 2 | 104 |

INTRODUCCIÓN

Introducción

En los últimos años, efectuar diversos estudios con técnicas estadísticas es relevante en el contexto nacional e internacional. Para este trabajo la intención es mostrar un estudio del tipo multivariado, donde los componentes son fijos, a partir de información de la Encuesta de Ingreso y Gasto de los Hogares 2006, obtenida durante el tercer trimestre de dicho año (agosto a noviembre).

El objetivo de este trabajo de tesis es analizar el gasto de aquellos hogares que tienen al menos un integrante con población de 65 y más años de edad en 13 variables definidas por los tipos de alimentos (carne, cereales, leche, pescado, huevo, aceites, tubérculos, verduras, frutas, azúcar, especias, café, otros alimentos), en los que se invierte el mayor gasto, por medio del método estadístico de componentes principales; y comparar los resultados con el gasto ejercido en aquellos hogares que tienen exclusivamente residentes de 65 y más años de edad para las mismas componentes de alimentos, de tal forma que se identifiquen algunas diferencias en cuanto a la concentración del gasto en los alimentos, así como identificar el comportamiento de los componentes obtenidos (máxima variabilidad, correlaciones o dependencias) y que están asociados a estos resultados los gastos ejercidos de los alimentos de análisis.

Para facilitar el cumplimiento del objetivo general se ha estructurado este trabajo en seis capítulos.

Capítulo 1: Antecedentes

En este primer apartado se describe de manera general la conceptualización de la población de 65 y más años de edad en el contexto internacional y nacional, así como su importancia social con la mención de algunos indicadores estadísticos vistos desde la parte demográfica con respecto a este estrato de población adulta, señalando algunos factores asociados como son la mortalidad y la migración.

Capítulo 2: Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH)

Se describen las características más importantes de dicha encuesta dado que es el instrumento de levantamiento de información que registra y obtiene resultados de los ingresos y gastos de los hogares con particularidades de su obtención, y para este caso, el gasto en alimentos, instrumento de donde se tomó la información para poder llevar a cabo este estudio.

Capítulo 3: Metodología de Generación de Información

Aquí se explica la forma de extraer la información referente al gasto de los hogares que tienen como residentes a algún adulto mayor (población de 65 y más años de edad), así como de hogares donde exclusivamente residen habitantes que satisfacen dicho rango de edad, utilizando las bases de datos que proporciona la ENIGH y los filtros realizados para obtener las bases finales de trabajo.

Capítulo 4: Antecedentes de los Componentes Principales

Aquí se muestra la parte teórica de los componentes principales, así como los puntos relevantes de esta técnica estadística.

Capítulo 5: El Gasto de los Hogares con Adultos Mayores

En este capítulo se describe cada elemento estadístico, desde un análisis exploratorio de la información, hasta la utilización del método de componentes principales, con el fin de cumplir el objetivo planteado.

Capítulo 6: Conclusiones y consideraciones finales

Por último, se presentan las conclusiones donde se hace un análisis del comparativo entre ambos ejercicios analizados, referente al gasto de los hogares que tienen como residentes a algún adulto mayor (población de 65 y más años de edad), así como de hogares donde exclusivamente residen habitantes que satisfacen dicho rango de edad, y se indica la conclusión final de este trabajo.

CAPÍTULO 1. ANTECEDENTES

Capítulo 1. Antecedentes

El envejecimiento de la población es un producto del avance social y económico que se hace notorio no sólo como cuantificación demográfica sino también por sus consecuencias. Una mirada de la vejez es conocer los efectos de su desarrollo y sus posibilidades, como es la pérdida de independencia (económica, emocional y salud) y el regreso a la manutención familiar. Por un lado, la población envejecida requiere de apoyos económicos y alimenticios por parte de la población ya sea de los integrantes de su hogar (en edad productiva) o de un programa público de beneficencia del tipo alimentario, que pueden reflejarse en pensiones económicas o en especie como las ayudas alimentarias en despensas de víveres y las cuales consideran los productos de alimentos básicos como son leche, huevo, sopas, frijoles, arroz, aceite, azúcar, vegetales, y carnes. Por otra parte, se presentan las limitaciones de acceso a los servicios de salud, o bien por beneficiarse por el derecho de un descendiente, o quizás, inscribirse a algún programa gubernamental que se ha implementado y reforzado a inicios del nuevo siglo. Y por último, se plantea la posible interrelación en el hogar familiar y que puede estar fundamentada en la parte emocional con los hijos u otros parientes en las actividades internas sobre todo del tipo doméstico destacando la limpieza, alimentos, cuidado de niños, entre otros (Ver. *Paradigmas y nuevos contratos sociales*, Roberto Ham Chande).

La Organización de las Naciones Unidas (ONU) considera a la población de 65 y más años de edad a las personas en proceso de envejecimiento que dependen económicamente de la población en edad productiva (15 a 64 años de edad). Sin embargo, se debe señalar que para el caso de México debido a que se han elaborado leyes en beneficio de este grupo de personas, a nivel federal se identifica al adulto mayor como aquella persona que tienen 60 y más años de edad y que tiene la posibilidad de recibir prestaciones económicas en la adquisición de bienes o servicios, en cambio, en el Distrito Federal, el adulto mayor que recibe ayuda económica alimentaria es aquella que tiene 70 y más años de edad.

El paulatino envejecimiento de la población mexicana en el contexto económico hace relevante conocer cuál es la situación social y en particular la alimentaria de la población de 65 y más años de edad. Dado que es importante redefinir el papel institucional, social y familiar que se tiene con este grupo de individuos cada vez más vulnerable conforme avanza su edad.

Es conocido, y estudiado por algunos demógrafos, que conforme la población envejece cada vez hay más mujeres que hombres por lo que la supervivencia de ellas es mayor que la del sexo masculino. En consecuencia, la población femenina podría verse cada vez en desventaja donde se encuentra en riesgo su seguridad física, emocional y económica, y donde nuevamente la atención alimentaria se hace participe en la cuestión de accesibilidad, independientemente del tipo y cantidad de alimento al que tienen derecho.

Así, por recomendación internacional y por disponibilidad la información de acuerdo con estadísticas obtenidas de los Censos de Población y Vivienda (1930 y 1970) y la ENIGH 2006, permiten confirmar que la población de 65 y más años de edad se ha multiplicado poco más de diez veces en los últimos setenta y seis años; el número de hombres y mujeres cada vez se aprecia con mayor diferencia, a favor de ellas. Por otra parte, la esperanza de vida al nacer se duplicó en el mismo periodo al pasar en las mujeres de 37 años (1930) a 77 años para 2006, en el caso de los hombres de 35 a 72 años (Cuadro 1.1).

En el caso de la población adulta mayor, las estimaciones obtenidas del Consejo Nacional de Población, señalan que este estrato de grupo de edad incrementó su lapso de vida al pasar de 13 años más (ambos sexos) en 1930 a 21 años (hombres) y 23 años en mujeres para 2006. Asimismo, al revisar el indicador sobre dependencia económica de la población, adulta mayor con respecto a la población económicamente activa, se tiene que ésta casi se duplicó durante los 76 años transcurridos; en el caso de las mujeres, el indicador se redujo drásticamente

en el tiempo debido a la cada vez mayor introducción de ellas en el mercado laboral.

Indicadores seleccionados de la población 1930 a 2006

Cuadro 1.1

| Indicador | 1930 | 1970 | 2006 |
|--|----------------|------------------|------------------|
| Población de 65 y más años | 488,837 | 1,791,385 | 6,742,016 |
| - Hombres | 241,877 | 859,166 | 3,081,692 |
| - Mujeres | 246,960 | 932,219 | 3,660,324 |
| Esperanza de vida al nacer (en años) | 36 | 61 | 75 |
| - Hombres | 35 | 59 | 72 |
| - Mujeres | 37 | 63 | 77 |
| Esperanza de vida de la población de 60 y más años | 13 | 18 | 22 |
| - Hombres | 13 | 18 | 21 |
| - Mujeres | 13 | 19 | 23 |
| Índice de dependencia de población de 65 y más años | 9 | 14 | 15 |
| - Hombres | 5 | 8 | 11 |
| - Mujeres | 103 | 35 | 22 |

Nota: Las cifras de la esperanza de vida se redondearon los decimales al entero próximo. El índice de dependencia económica corresponde a la población de 65 y más años entre la población económicamente activa, por cada 100.

Fuente: INEGI. Censos de población y vivienda, 1930 y 1970.

Encuesta nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Los adultos mayores en México. Perfil Sociodemográfico al inicio del siglo XXI.

Aguascalientes, 2005.

CONAPO. Tablas de mortalidad por sexo para la República Mexicana, 1930-2050

También hay que mencionar que a partir del año 2000, se comenzaron a aplicar diversos programas gubernamentales reglamentados de ayuda alimentaria y de seguridad social tanto estatales como federales, con el fin de proporcionar un mejor bienestar a esta población adulta mayor.

En particular, esta ayuda alimentaria ha visto incrementada su aportación económica en el transcurso del tiempo con el fin de cubrir la alimentación básica

diaria y que a la postre ha reducido aquellos efectos de poca atención (alimentaria) dentro del seno familiar o individual por la escasa disponibilidad de recursos económicos para estos programas hasta antes de dicho año de referencia.

Por último, señalar que el impacto social y económico del envejecimiento de esta población está asociado al estado de salud que lleva y que implícitamente es reflejo de las condiciones de vida que llevó en su fase productiva, principalmente de la alimentación, lo que se refleja en la mortalidad próxima que tiene por las enfermedades causadas y que se han venido modificando a lo largo del tiempo, por señalar que hasta antes de los años setenta predominaban las enfermedades transmisibles (infecciosas y parasitarias) al tener mayor afectación en este estrato de población adulta, y que se ha modificado en los últimos años dado que las causas que actualmente provocan el deceso son las denominadas crónico degenerativas (diabetes mellitus) y problemas del corazón, y que están asociadas a la alimentación que llevaron o llevan actualmente.

CAPÍTULO 2. ENCUESTA NACIONAL DE INGRESO Y GASTO DE LOS HOGARES (ENIGH)

Capítulo 2. Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH)

Para llevar a cabo el análisis estadístico por componentes principales del gasto realizado en alimentación de los hogares con población de 65 y más años de edad, se procedió a obtener la información de la ENIGH 2006, debido a que es el único instrumento que recopila información del país en los rubros de ingreso y gasto de los hogares. Este estudio solamente se enfocó en el concepto del gasto referido a los alimentos (la encuesta agrega en 13 conceptos la mayoría de los alimentos), y por otra parte el interés es analizar en un momento dado del tiempo, para esto se toma como referencia la información levantada en el año de 2006, realizada por el ahora Instituto Nacional de Estadística y Geografía (INEGI).

Para ello se describen algunas características importantes. La Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), tiene sus antecedentes en varias encuestas realizadas por diferentes dependencias públicas.

En 1956 y 1958 la Dirección General de Estadística (DGE), entonces dependiente de la Secretaría de Industria y Comercio (SIC), levantó la encuesta Ingresos y Egresos de la Población en México, y en 1960 en las 16 Ciudades de la República Mexicana, se obtuvieron Ingresos y Egresos Familiares. Por su parte, para los años 1963 y 1968, el Banco de México llevó a cabo el levantamiento de la encuesta Ingresos y Gastos Familiares.

Durante el periodo 1969-1970, nuevamente la DGE-SIC realizó la encuesta Ingresos y Egresos de la República Mexicana. Posteriormente, en 1975, la Secretaría del Trabajo y Previsión Social (STPS), efectuó Ingresos y Gastos Familiares. En 1977 la DGE, como parte de la Secretaría de Programación y Presupuesto (SPP), desarrolló la Encuesta Nacional de Ingreso-Gasto de los Hogares, trabajo que constituyó el antecedente inmediato de las encuestas que ha

llevado a cabo la DGE-INEGI para los periodos: 1983-1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005 y 2006.

Es a partir de 1984 cuando se inició una tendencia a ampliar los objetivos y a homogeneizar la metodología, tomando en cuenta las recomendaciones internacionales y los requerimientos de información de los diferentes usuarios, cuidando la comparabilidad histórica.

También se incorporaron nuevos productos con el objetivo de adecuarse a los cambios económicos del país y obtener resultados que reflejan la realidad. A partir de 1992 se realiza con una periodicidad bienal con excepción del 2005, ya que fue un levantamiento extraordinario ^{2.1}.

La ENIGH proporciona información sobre la distribución, monto y estructura del ingreso y gasto de los hogares.

Permite generar información de la organización del ingreso corriente de los hogares, según la fuente de donde provenga; la disposición del gasto corriente en la adquisición de bienes de consumo final, (duradero y no duradero) entre ellas los alimentos, la estructura de las erogaciones y percepciones financieras y de capital.

También es el campo de estudio para conocer las características sociodemográficas, la condición de actividad y las características ocupacionales de los integrantes del hogar de 12 y más años, a la vez que se estudian las características de infraestructura de la vivienda y equipamiento del hogar.

2.1 Ver. Encuesta Nacional de Ingreso y Gasto de los Hogares. Caracterización General. (www.inegi.org.mx)

La población objetivo la constituyen los hogares de nacionales o extranjeros, que residen habitualmente en viviendas particulares dentro del territorio nacional.

El tamaño de la muestra efectiva es de 25 443 viviendas en el año 2006, la unidad de observación es el hogar, la unidad de muestreo es la vivienda, la unidad de análisis son el hogar, la vivienda y los integrantes del hogar.

Así, se consultan los microdatos que contienen información detallada del monto y composición de los ingresos de los hogares y del destino de sus gastos; además, permiten asociar esta información con las características de las viviendas y con los rasgos sociodemográficos de sus integrantes.

Los resultados de esta encuesta son representativos tanto a nivel nacional como para los ámbitos rural y urbano.

Solo en casos específicos y para ciertos años se obtuvo información a nivel estatal. Por señalar en 1996 para las entidades de Campeche, Coahuila, Distrito Federal, Guanajuato, Hidalgo, Jalisco, Estado de México, Oaxaca y Tabasco; en el año 2000, la obtención específica del estado de Veracruz; y en 2004 para Nuevo León y Distrito Federal.

La ENIGH dispone de las siguientes siete bases de datos:

Bases de datos de la ENIGH

Cuadro 2.1

| Descripción de base de datos | Nombre |
|---|-----------------|
| 1. Características de los hogares, de las viviendas que habitan | hogares.dbf |
| 2. Características sociodemográficas y ocupacionales de los miembros de los hogares | población.dbf |
| 3. Ingresos y percepciones de capital de cada uno de los miembros de los hogares | ingresos.dbf |
| 4. Gastos realizados por el hogar | gastos.dbf |
| 5. Erogaciones de capital por hogar | erogaciones.dbf |
| 6. Gastos o ingresos realizados por hogar y algunos por persona | nomon.dbf |
| 7. Principales variables por hogar | concen.dbf |

Para cumplir con el objetivo de este trabajo, sólo se consideraron las bases de hogares, población y concentrado, debido a que son las que contienen las variables de interés para desarrollarlo. En el anexo se incluyen las estructuras correspondientes.

Asimismo, es importante señalar que esta encuesta no es representativa a nivel entidad federativa en 2006; sin embargo, para este estudio se utiliza la información a este nivel para poder obtener resultados que proporciona la técnica de componentes principales. Por otra parte cabe mencionar la importancia del muestreo complejo aplicado para la encuesta, con la finalidad de obtener la muestra aún cuando en este estudio no se profundiza sobre este tema.

El marco muestral utilizado, fue una submuestra del marco de la Encuesta Nacional de la Dinámica Demográfica (ENADID) y del marco muestral de la Encuesta Nacional de Empleo Urbano (ENEU) para áreas metropolitanas, otro marco muestral utilizado es el de propósitos múltiples del INEGI, constituido por la información demográfica y cartográfica obtenida a partir del levantamiento del Censo de Población y Vivienda del 2000 y que el esquema de muestreo fue probabilístico, su diseño es polietápico, estratificado y por conglomerados, donde la unidad última de selección es la vivienda y la unidad de observación es el hogar; en consecuencia, los resultados obtenidos de la encuesta se generalizan a toda la población.

CAPÍTULO 3. METODOLOGÍA DE GENERACIÓN DE INFORMACIÓN

Capítulo 3. Metodología de Generación de Información

A continuación se describe el procedimiento de obtención de las dos bases de datos que sirvieron como insumo para realizar el análisis estadístico correspondiente y cumplir con los objetivos planteados, estas son:

1. Gasto65.dbf Considera registros de hogares con al menos un integrante de 65 y más años de edad y las trece variables de gasto por tipo de alimento.
2. Gasto65e.dbf Integra los registros de hogares con residentes exclusivamente que tienen 65 y más años de edad y las trece variables de gasto por tipo de alimento.

Solo se señalará en forma detallada la obtención de la primera base dado que el procedimiento es similar para la obtención de la segunda.

Obtención de la base Gasto65.dbf

Debido a que las bases de datos de la ENIGH consideran los registros de toda la población de las viviendas muestreadas, se procedió a seleccionar sólo aquellos que cumplieran las siguientes condiciones de acuerdo al orden.

1. Se identificaron las siguientes tres bases de datos con las variables relacionadas para este estudio, éstas son: Hogares, Población y Concentrado con extensión dbf.
2. Se depuraron las bases y se dejaron los campos de utilidad, en el caso de la base de población sólo se dejaron Folio, Num_ren y edad; en el caso de hogares se consideran al folio y ubica_geo, y del archivo de concentrado a

los campos folio, ubica_geo, p65mas, cereales, carnes, pescado, leche, huevo, aceites, tubérculo, verduras, frutas, azúcar, café, especias, otro_ali (otros alimentos) y hog (hogares, factor de expansión). Esto con el fin de optimizar recursos de procesamiento.

3. De la base de datos de población se seleccionaron aquellos registros que cumplieron la condición de 65 y más años de edad, dichos registros se guardaron en una base auxiliar llamada pob65.dbf con la intención de verificar los universos con lo señalado en el campo P65mas de la base denominada concentrado.
4. La nueva base obtenida se relacionó con la base de hogares mediante el campo Folio_hog para obtener aquellos hogares que registraron dicho estrato de población. Con ello se obtuvo una nueva base denominada hog_pob65.dbf en la que se guardaron los registros que cumplieron dicha condición.
5. Posteriormente se relacionó esa nueva base con la de gastos para así identificar todos los gastos que realizaban dichos hogares en las variables seleccionadas de alimentación, estos registros se guardaron en la base Gasto65.dbf
6. Por último, a partir del campo de referencia geográfica, se filtraron los registros de tal forma que se seleccionaron y ordenaron de manera que se acumularon (sumaron) los registros por nivel entidad federativa, con el fin de obtener la base de datos final que sirve como insumo, dado que el interés del análisis es identificar de acuerdo con las entidades federativas, cuál es la situación del gasto alimentario de la población de 65 y más años de edad en los hogares.

Para cumplir con los puntos anteriores, se utilizó un manejador de bases de datos Foxpro, con el fin de realizar los procesos de filtrado de información; en la base de población se seleccionaron todos aquellos registros que cumplían la condición de personas de 65 y más años de edad.

Posteriormente, se elaboró un programa (ver anexo) para hacer una relación entre las bases de población filtrada con la de hogares con el fin de que se obtuvieran aquellos hogares donde se registra al menos un residente en dicho rango de edad (65 y más años de edad).

En una tercera etapa, esta base filtrada de hogares se relacionó con la base de gastos para obtener los registros de hogares con gastos en las distintas variables de alimentos seleccionadas.

Y por último, se aplicó el factor de expansión señalado en dicha base de gastos conforme a cada uno de los registros para acumular (sumar) posteriormente por entidad federativa, de tal forma que se obtuviera la base final con solamente 32 registros y los campos que se utilizan para el estudio. Esta es la estructura de dicha base:

| # | Descripción | Clave | Tamaño (numérico) |
|----|--|-----------|----------------------|
| 1 | Nombre de la entidad federativa (01..32) | entidad | 99 |
| 2 | Gasto en cereales | cereales | 9,999,999 |
| 3 | Gasto en carnes | carnes | 9,999,999 |
| 4 | Gasto en pescado | pescado | 9,999,999 |
| 5 | Gasto en leche | leche | 9,999,999 |
| 6 | Gasto en huevo | huevo | 9,999,999 |
| 7 | Gasto en aceites | aceites | 9,999,999 |
| 8 | Gasto en tubérculo | tubérculo | 9,999,999 |
| 9 | Gasto en verduras | verduras | 9,999,999 |
| 10 | Gasto en frutas | frutas | 9,999,999 |
| 11 | Gasto en azúcar | azúcar | 9,999,999 |
| 12 | Gasto en café | café | 9,999,999 |
| 13 | Gasto en especias | especias, | 9,999,999 |
| 14 | Gasto en otros alimentos | otro_ali | 9,999,999 |

Se debe señalar que en cada proceso de filtrado de bases, se cuidó la congruencia de los universos con la intención de garantizar la calidad de los resultados. Por ejemplo, se identificó que la población de 65 y más años de edad filtrada coincidiera con los resultados publicados, así como que los universos de hogares y de gastos coincidieran.

Obtención de la Base de datos Gasto65e.dbf

El objetivo de esta base es considerar todos aquellos registros que consideran exclusivamente residentes de 65 y más años de edad. Para ello, se realizó el mismo procedimiento señalado en el punto anterior con la condición de que el filtrado solamente identificara a los hogares con residentes de 0 a 64 años de edad igual a cero unidades y que la población de edad de 65 y más años fuera distinto de cero y se acumuló en una base de datos a nivel entidad federativa con 32 registros, uno para cada entidad federativa del país.

| # | Descripción | Clave | Tamaño (numérico) |
|----|--|-----------|----------------------|
| 1 | Nombre de la entidad federativa (01..32) | entidad | 99 |
| 2 | Gasto en cereales | cereales | 9,999,999 |
| 3 | Gasto en carnes | carnes | 9,999,999 |
| 4 | Gasto en pescado | pescado | 9,999,999 |
| 5 | Gasto en leche | leche | 9,999,999 |
| 6 | Gasto en huevo | huevo | 9,999,999 |
| 7 | Gasto en aceites | aceites | 9,999,999 |
| 8 | Gasto en tubérculo | tubérculo | 9,999,999 |
| 9 | Gasto en verduras | verduras | 9,999,999 |
| 10 | Gasto en frutas | frutas | 9,999,999 |
| 11 | Gasto en azúcar | azúcar | 9,999,999 |
| 12 | Gasto en café | café | 9,999,999 |
| 13 | Gasto en especias | especias, | 9,999,999 |
| 14 | Gasto en otros alimentos | otro_ali | 9,999,999 |

CAPÍTULO 4. ANTECEDENTES DE LOS COMPONENTES PRINCIPALES

Capítulo 4. Antecedentes de los Componentes Principales^{4.1/}

El Análisis de Componentes Principales es una técnica estadística multivariada que sirve para reducir la dimensión del conjunto original de variables y tiene sentido cuando éstas están correlacionadas.

Algebraicamente, las componentes principales son combinaciones lineales de las p variables aleatorias originales X_1, \dots, X_p . Geométricamente, estas combinaciones lineales representan la selección de un nuevo sistema de coordenadas obtenido por la rotación del sistema original con X_1, \dots, X_p como los ejes coordenados.

Los nuevos ejes representan las direcciones con máxima variabilidad y proporcionan una descripción más simple de la estructura de covarianza.

Sea el vector aleatorio $X' = [X_1, \dots, X_p]$, con matriz de covarianzas Σ , y con valores propios $\lambda_1 \geq \dots \geq \lambda_p \geq 0$.

Considérense las combinaciones lineales

$$\begin{aligned} Y_1 &= \mathbf{a}'_1 X = a_{11}X_1 + \dots + a_{1p}X_p \\ &\dots \\ Y_p &= \mathbf{a}'_p X = a_{p1}X_1 + \dots + a_{pp}X_p \end{aligned} \tag{1}$$

4.1 Ver. Análisis Multivariante. Montanero Fernández Jesús. (www.kolmogorov.unex.es/)

Por lo tanto,

$$\text{Var}(Y_i) = a'_i \Sigma a_i, \quad i = 1, \dots, p \quad (2)$$

$$\text{Cov}(Y_i, Y_k) = a'_i \Sigma a_k, \quad i, k = 1, \dots, p \quad (3)$$

Las componentes principales son aquellas combinaciones lineales no correlacionadas Y_1, \dots, Y_p cuyas varianzas en (2) son tan grandes como sea posible.

Se define entonces:

Primera componente principal = combinación lineal $a'_1 X$ que maximiza

$$\text{Var}(a'_1 X) \quad \text{sujeto a} \quad a'_1 a_1 = 1$$

Segunda componente principal = combinación lineal $a'_2 X$ que maximiza

$$\text{Var}(a'_2 X) \quad \text{sujeto a} \quad a'_2 a_2 = 1 \text{ y}$$

$$\text{Cov}(a'_1 X, a'_2 X) = 0$$

En el i -ésimo paso, la i -ésima componente principal = combinación lineal $a'_i X$ que maximiza

$$\text{Var}(a'_i X) \quad \text{sujeto a} \quad a'_i a_i = 1 \text{ y}$$

$$\text{Cov}(a'_i X, a'_k X) = 0 \quad \text{para } k < i$$

Lo anterior, está sustentado por los siguientes teoremas sobre componentes principales, en cuanto a que las componentes no están correlacionadas, tienen varianzas iguales a los valores propios, de la acumulación de varianza de las componentes que explican la varianza total.

Teorema 1. Sea el vector aleatorio $X' = [X_1, \dots, X_p]$ con matriz de covarianza Σ con valores propios $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ y vectores propios asociados e_1, \dots, e_p , respectivamente. Entonces la i -ésima componente principal está dada por:

$$Y_i = e_i'X = e_{i1}X_1 + \dots + e_{ip}X_p \quad i = 1, \dots, p \quad (4)$$

$$\begin{aligned} \text{Con } \text{Var}(Y_i) &= e_i'\Sigma e_i = \lambda_i & i &= 1, \dots, p \\ \text{Cov}(Y_i, Y_k) &= e_i'\Sigma e_k = 0 & i &\neq k \end{aligned} \quad (5)$$

Si algunas λ_i son iguales, entonces las respectivas Y_i no son únicas.

Demostración. De maximización de formas cuadráticas para puntos en la esfera unitaria, con $\beta = \Sigma$, se tiene que

$$\text{Max}_{a \neq 0} \quad a'\Sigma a / a'a = \lambda_1 \quad (\text{alcanzado cuando } a = e_1)$$

$$\text{Pero } e_1'e_1 = 1. \quad \text{Así que } \text{Max}_{a \neq 0} \quad a'\Sigma a / a'a = \lambda_1 = e_1'\Sigma e_1 / e_1'e_1 = \text{Var}(Y_1)$$

Similarmente, por los mismos resultados de maximización tenemos:

$$\text{Max } a' \mathbf{e}_1, \dots, \mathbf{e}_k \quad a' \Sigma a / a' a = \lambda_{k+1} \quad k = 1, \dots, p-1$$

Tomando $a = \mathbf{e}_{k+1}$ resulta: $\mathbf{e}'_{k+1} \Sigma \mathbf{e}_{k+1} / \mathbf{e}'_{k+1} \mathbf{e}_{k+1} = \text{Var}(Y_{k+1}) = \text{Var}(Y_{k+1}) = \lambda_{k+1}$

Basta probar que $\mathbf{e}'_i \mathbf{e}_k = 0$, $i \neq k$ implica $\text{Cov}(Y_i, Y_k) = 0$. Ahora, los vectores propios de Σ son ortogonales si todos los valores propios λ_i son diferentes. Si no todos los valores propios son distintos, los correspondientes vectores propios a los valores propios comunes pueden escogerse ortogonales. Por lo tanto, para cualquier par de vectores propios \mathbf{e}_i y \mathbf{e}_k con $\mathbf{e}'_i \mathbf{e}_k = 0$, $i \neq k$. Dado que $\Sigma \mathbf{e}_k = \lambda_k \mathbf{e}_k$, premultiplicando por \mathbf{e}'_i resulta:

$$\text{Cov}(Y_i, Y_k) = \mathbf{e}'_i \Sigma \mathbf{e}_k = \mathbf{e}'_i \lambda_k \mathbf{e}_k = \lambda_k \mathbf{e}'_i \mathbf{e}_k = 0$$

para todo $i \neq k$, y **la prueba está completa.**

El resultado anterior indica que las componentes principales no están correlacionadas y tienen varianzas iguales a los valores propios de Σ (matriz de covarianzas).

Teorema 2. Sea el vector aleatorio $X' = [X_1, \dots, X_p]$ con matriz de covarianza Σ , y valores propios $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ y vectores propios asociados $\mathbf{e}_1, \dots, \mathbf{e}_p$, respectivamente. Sean $Y_1 = \mathbf{e}'_1 X, \dots, Y_p = \mathbf{e}'_p X$ las componentes principales.

Entonces

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) \quad \text{con } i=1 \dots p$$

Demostración. Recordéese que $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \text{tr}(\Sigma)$. Además Σ puede escribirse como $\Sigma = P\Delta P'$, con Δ diagonal de valores propios y $P = [e_1, \dots, e_p]$ y, $PP' = P'P = I$.

Entonces

$$\text{tr}(\Sigma) = \text{tr}(P\Delta P') = \text{tr}(\Delta P P') = \text{tr}(\Delta) = \sum_{i=1}^p \text{Var}(Y_i) \quad \text{con } i=1 \dots p$$

El resultado anterior dice que

$$\text{Varianza total de población} = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp}$$

$$= \lambda_1 + \dots + \lambda_p \quad (6)$$

y consecuentemente, la proporción de la varianza total debida a la k -ésima componente principal es

$$\left[\begin{array}{l} \text{proporción de la varianza} \\ \text{poblacional total} \\ \text{debida a la } k\text{-ésima} \\ \text{componente principal} \end{array} \right] = \frac{\lambda_k}{(\lambda_1 + \dots + \lambda_p)} \quad (7)$$

para $k = 1, \dots, p$.

Si más (por ejemplo, del 70% al 90%) de la varianza poblacional total, para p grande, puede atribuirse a unas “cuantas” componentes, entonces éstas pueden “reemplazar” a las p variables originales, sin mucha pérdida de información.

Teorema 3. Si $Y_1 = e_1'X, \dots, Y_p = e_p'X$ son las componentes principales obtenidas de la matriz de covarianza Σ , entonces:

$$\rho(Y_i, X_k) = \frac{e_{ik} \lambda_i^{1/2}}{\sigma_{kk}^{1/2}} \quad (8)$$

$i, k = 1, \dots, p$, son los coeficientes de correlación entre las componentes Y_i y las variables X_k .

Demostración. Sea $a'_k = [0, \dots, 0, 1, 0, \dots, 0]$, luego $X_k = a'_k X$ y $\text{Cov}(X_k, Y_i) = \text{Cov}(a'_k X, e_i' X) = a'_k \Sigma e_i$. Dado que $\Sigma e_i = \lambda e_i$, $\text{Cov}(X_k, Y_i) = a'_k \lambda e_i = \lambda e_{ik}$.

Entonces $\text{Var}(Y_i) = \lambda_i$ y $\text{Var}(X_k) = \sigma_{kk}$ conducen a que

$$\rho(Y_i, X_k) = \frac{\text{Cov}(Y_i, X_k)}{(\text{Var}(Y_i))^{1/2} (\text{Var}(X_k))^{1/2}}$$

$$= \frac{\lambda_i e_{ik}}{\lambda_i^{1/2} \sigma_{kk}^{1/2}}$$

$$= \frac{e_{ik} \lambda_i^{1/2}}{\sigma_{kk}^{1/2}} \quad i, k = 1, \dots, p$$

Así, se puede considerar el análisis de componentes principales como la técnica estadística multivariada de síntesis de la información, o reducción de la dimensión (número de variables). Esto es, para grandes bases de datos con muchas variables, se trata de reducir a un menor número de ellas sin mucha pérdida de información. Donde los nuevos componentes o factores principales serán combinación lineal de las variables originales y además serán independientes entre sí. Se creará una estructura de interdependencia entre variables cuantitativas para crear nuevas variables que estarán en función lineal de las originales y de las que se podrán realizar representaciones gráficas para tener una mejor interpretación de los datos.

Las nuevas componentes determinan lo esencial de las variables originales al ser combinación lineal de ellas con las siguientes propiedades:

- a) Son ortogonales
- b) Cada componente no está correlacionada con la anterior
- c) La primera componente presenta la mayor variabilidad y la j -ésima tiene más varianza que la $j + 1$ ésima.
- d) Las primeras j componentes dan solución de mínimos cuadrados del modelo $Y=XB+E$, con Y dependiente, X independiente, B matriz de coeficientes constantes y E los errores del modelo.

La interpretación de las componentes, se facilitará tras observar la relación entre los componentes y las variables originales, esto es, se debe revisar el signo de los resultados como magnitud de las correlaciones; por otra parte, la experiencia dentro del tema ayuda a proporcionar mejor interpretación.

Para poder reducir la dimensión del problema estadístico, se utiliza el paquete estadístico gratuito R (The R Foundation for Statistical Computing, www.r-project.org/). Además, se describen las siguientes tres actividades a realizar dentro de componentes principales y donde es importante identificarlos en su forma de conceptualización.

Análisis de la matriz de correlaciones

Un análisis de componentes principales tiene sentido si existen altas correlaciones entre las variables, ya que esto es señal de que existe información redundante o dependiente y que en consecuencia solamente habrán algunos factores que explicarán la mayoría de la variabilidad total.

Selección de componentes

La elección de los componentes principales se lleva a cabo a partir de identificar la primera componente que contenga la mayor proporción posible de la variación total, la segunda componente debe de agrupar la segunda mayor variabilidad posible que no fue considerada por la primera componente y así sucesivamente. Para que el total de componentes que se elijan, se considerarán sólo aquellos que concentren el porcentaje de variabilidad suficiente, por señalar 70%, 80% o 90% de la varianza original y que serán llamados componentes principales.

Análisis de factores

Ya seleccionadas las componentes principales, éstas se representan en forma de matriz, cada elemento representa los coeficientes factoriales de las variables (las

correlaciones entre las variables y las componentes principales, en tantas filas como variables se consideren.

Uso de matriz de covarianzas y matriz de correlaciones

Para el caso de utilizar la **matriz de varianzas y covarianzas** los valores de X se dan en las *unidades originales*, mientras que si se utiliza la **matriz de correlaciones** se toman los *valores estandarizados*; desde otro punto de vista esta última se realiza sobre todo para determinar si hay asociación o posibles asociaciones, entre variables, y por tanto el de aplicar las componentes principales.

Descomposición espectral.

El teorema de descomposición espectral, muestra las condiciones bajo las cuales un operador o una matriz pueden ser diagonalizadas (representadas como una matriz diagonal en alguna base). Se identifica un tipo de operadores lineales que pueden representarse como una multiplicación de operadores.

Una vez obtenida la matriz de correlaciones se **obtuvieron** los componentes, **ésto** es obtener las combinaciones lineales de las variables que hacen la máxima varianza por lo cual se procede a obtener la descomposición espectral, esto es, determinar los eigenvalores y los eigenvectores, con la intención de determinar cuál (cuáles) de los componentes son más importantes.

Desviaciones estándar por componentes.

Se muestra para cada componente la desviación que puede tener el valor medio para que sea considerado correcto conforme a un orden de importancia (de los componentes), esto es, el primer componente reflejará la mayor desviación y

posteriormente los siguientes componentes registrarán cada vez menor dispersión entre ellas.

Acumulación de la varianza por componente principal.

Es una matriz donde se muestra la importancia de la variabilidad de cada componente principal, el primer rubro es la desviación estándar (desviación con respecto a la media), en seguida la proporción de la variabilidad explicada por cada componente y por último, la proporción de la variabilidad acumulada entre cada componente, a partir de esta acumulación se identifica el mínimo número de componentes principales que explican la mayoría de la variabilidad del conjunto de variables.

Gráfica de rodilla o codo.

Es una gráfica donde se muestra el comportamiento de los eigenvalores de los componentes, la regla es que se toman los primeros componentes antes de que se doble la tendencia, aunque eso es conservador.

Gráfica de variables o cargas.

Representa las variables originales por medio de flechas, éstas tienen las siguientes características: son oblicuas, pueden tener longitudes cortas o largas (de acuerdo a la importancia) de cada variable en los dos primeros componentes principales.

La distancia menos corta es aquella que tiene menos influencia en el análisis. Asimismo se muestran las claves de identificación de cada registro, que para este caso es el número asignado a la entidad federativa. El cuadrante se conforma de partes positivas y las correspondientes negativas y puede variar su tamaño.

Gráficas de caras de Chernoff.

Permite identificar la tendencia de datos donde se comparan varias variables al mismo tiempo. Los datos se representan en una cara.

A cada variable se asocia un rasgo o característica de una cara por ejemplo:

- 1.- altura de la cara,
 - 2.- anchura de la cara,
 - 3.- forma de la cara,
 - 4.- altura de la boca,
 - 5.- anchura de la boca,
 - 6.- curva de la sonrisa longitud de nariz,
 - 7.- altura de los ojos,
 - 8.- anchura de los ojos,
 - 9.- altura de cabellos,
 - 10.- anchura de cabellos,
 - 11.- estilo de cabellos,
 - 12.- altura de nariz,
 - 13.- anchura de nariz,
 14. anchura de orejas,
- entre otros.

Entre más significativa la variable, más visible o amplia es el rasgo.

Descripción de la forma de la cara de Chernoff (versión original).

1-height of face,
2-width of face
3-shape of face
4-height of mouth
5-width of mouth
6-curve of smile
7-height of eyes
8-width of eyes
9-height of hair
10-width of hair
11-styling of hair
12-height of nose
13-width of nose
14-width of ears
15-height of ears

Fuente: Package 'TeachingDemos'. <http://www.r-project.org/>

Transformación de valores absolutos a logaritmos.

Históricamente, la utilidad de transformar valores reales a logaritmos es el de facilitar la solución de diferentes problemas tanto aritméticos como geométricos ^{4.2}

En la estadística es recomendable hacerlo cuando se tienen cifras considerables, por ejemplo en millones, cientos de millones, miles de millones de unidades, etcétera.

4.2 Ver. John Napier (1614) (<http://es.wikipedia.org>) y Henry Briggs (1556-1631) (<http://sapimates.blogspot.com/>)

Gastos en alimentos de los hogares per cápita.

Es importante señalar que, otra forma de realizar el estudio es identificar como se distribuye el gasto de los alimentos de los hogares con respecto al total de la población (per cápita), y que para este estudio sería contra la población de los hogares que tuvieran al menos un adulto mayor, así como el correspondiente a los hogares con exclusivamente uno o más adultos de 65 y más años de edad, con el fin de contrastar resultado y señalar la mejor aportación de esta técnica de componentes principales, sin embargo se deja abierta la posibilidad de realizar en un futuro para contrastar los resultados e identificar las bondades del uso de este método de componentes principales y en su caso encontrar otra alternativa de análisis.

CAPÍTULO 5. EL GASTO DE LOS HOGARES CON ADULTOS MAYORES

Capítulo 5. El gasto de los hogares con adultos mayores

En diversas investigaciones, la aplicación del análisis exploratorio de la información es importante con el fin de identificar el comportamiento real de las observaciones. En este trabajo se revisó dicho método, antes de identificar por medio de la técnica de componentes principales, a partir de los resultados de trece tipos de alimentos (cereales, carnes, pescado, leche, huevo, aceites, tubérculo, verduras, frutas, azúcar, café, especias y otros alimentos) obtenidos de la Encuesta de Ingreso y Gasto de los Hogares 2006, con el grupo de hogares con población mayor o igual a 65 años las preguntas propuestas son: ¿cuál sería el número de componentes que explicarían en mayor medida la variabilidad de las variables en el gasto alimentario de los hogares en las 32 entidades federativas del país, que tienen al menos un residente de 65 y más años de edad?; qué variables son las que tienen más influencia en el gasto de dichos hogares y en que entidades son más importantes de acuerdo con su distribución.

Finalmente, se revisaron las mismas situaciones en los hogares que tienen exclusivamente adultos de 65 y más años de edad con la finalidad de encontrar diferencias entre ambos modelos.

5.1 Hogares con al menos un adulto de 65 y más años de edad

En el siguiente Cuadro 5.1 se presentan los datos del gasto realizado en los hogares donde hay al menos un habitante de 65 y más años de edad por entidad federativa en el 2006.

**Total del gasto por tipo de alimento según entidad federativa, 2006
(Pesos)**

Cuadro 5. 1

| CLAVE | NOMBRE ENTIDAD | CEREALES | CARNES | PESCADO | LECHE | HUEVO | ACEITES | TUBERCULO | VERDURAS | FRUTAS | AZUCAR | CAFE | ESPECIAS | OTRO_ALI |
|-------|---------------------|-------------|-------------|-------------|-------------|-------------|------------|------------|-------------|-------------|------------|------------|------------|-------------|
| 01 | Aguascalientes | 51 131 819 | 34 849 819 | 1 388 102 | 37 289 760 | 4 424 615 | 3 947 715 | 3 767 681 | 34 356 389 | 9 539 182 | 3 423 138 | 1 942 340 | 2 351 154 | 28 163 952 |
| 02 | Baja California | 68 022 306 | 93 226 790 | 13 897 509 | 68 522 558 | 15 302 067 | 3 487 809 | 6 841 395 | 57 690 505 | 32 655 040 | 4 496 167 | 3 097 486 | 2 033 522 | 19 480 761 |
| 03 | Baja California Sur | 7 227 229 | 5 871 421 | 2 005 602 | 8 402 246 | 711 148 | 410 986 | 543 366 | 4 244 716 | 2 674 364 | 644 784 | 180 487 | 154 034 | 1 065 003 |
| 04 | Campeche | 29 624 444 | 49 788 009 | 3 124 410 | 9 829 907 | 3 521 767 | 1 449 035 | 2 890 264 | 16 789 646 | 8 678 488 | 3 061 040 | 4 300 230 | 1 414 086 | 20 224 935 |
| 05 | Chiapas | 77 690 058 | 83 730 066 | 7 850 295 | 45 896 670 | 14 993 913 | 7 497 758 | 13 295 215 | 51 373 449 | 18 424 564 | 8 243 821 | 10 360 110 | 3 744 283 | 51 459 839 |
| 06 | Chihuahua | 38 936 011 | 38 897 445 | 4 122 105 | 26 685 288 | 4 083 896 | 3 586 696 | 1 679 193 | 18 221 631 | 4 988 596 | 1 314 755 | 1 412 722 | 933 768 | 17 914 786 |
| 07 | Coahuila | 168 586 754 | 243 771 413 | 42 382 698 | 49 376 929 | 33 001 965 | 15 832 623 | 5 743 671 | 114 671 121 | 30 516 107 | 25 396 124 | 9 354 112 | 8 540 972 | 94 335 406 |
| 08 | Colima | 160 727 757 | 150 080 758 | 7 012 318 | 147 128 832 | 22 057 571 | 13 324 666 | 17 222 829 | 93 161 671 | 30 050 619 | 10 536 887 | 12 715 587 | 5 412 073 | 49 322 176 |
| 09 | Distrito Federal | 422 354 123 | 879 769 414 | 122 824 426 | 401 166 400 | 69 182 213 | 23 742 902 | 49 234 720 | 438 549 328 | 289 296 243 | 34 594 035 | 28 177 532 | 28 499 128 | 203 274 068 |
| 10 | Durango | 65 093 108 | 47 177 930 | 5 287 323 | 47 226 063 | 11 318 030 | 5 676 972 | 7 478 620 | 37 776 646 | 13 745 015 | 3 871 931 | 5 825 952 | 2 853 992 | 23 948 572 |
| 11 | Guerrero | 201 049 600 | 218 810 483 | 11 578 730 | 123 403 356 | 24 319 095 | 12 704 210 | 15 036 040 | 148 155 472 | 42 088 974 | 13 026 557 | 4 402 556 | 6 396 058 | 88 653 053 |
| 12 | Guanajuato | 165 559 873 | 199 402 142 | 53 834 034 | 113 580 734 | 26 796 859 | 18 471 214 | 6 652 587 | 121 693 283 | 18 537 617 | 11 125 789 | 3 845 189 | 4 784 559 | 32 488 905 |
| 13 | Hidalgo | 84 785 673 | 100 974 368 | 8 907 507 | 52 210 750 | 14 893 181 | 7 086 156 | 5 513 721 | 69 270 393 | 25 920 412 | 9 602 306 | 5 040 398 | 2 824 631 | 21 405 385 |
| 14 | Jalisco | 208 241 772 | 241 042 045 | 36 358 515 | 171 675 508 | 31 058 601 | 9 746 895 | 15 799 185 | 159 578 785 | 48 321 134 | 23 273 184 | 14 997 026 | 10 322 863 | 86 095 546 |
| 15 | México | 591 987 599 | 847 380 065 | 124 378 385 | 436 842 866 | 109 500 994 | 47 175 151 | 47 202 499 | 467 172 011 | 191 921 029 | 47 560 770 | 27 022 333 | 37 617 450 | 204 043 796 |
| 16 | Michoacán | 214 422 104 | 193 403 655 | 21 499 386 | 157 118 874 | 21 449 547 | 21 703 647 | 27 020 691 | 188 283 062 | 65 790 803 | 13 970 523 | 8 200 904 | 6 497 547 | 43 586 270 |
| 17 | Morelos | 84 770 649 | 82 881 231 | 5 764 841 | 48 757 127 | 8 985 180 | 6 992 229 | 3 723 279 | 56 519 227 | 29 912 645 | 4 281 683 | 6 852 019 | 2 165 320 | 18 984 987 |
| 18 | Nayarit | 33 088 847 | 28 252 403 | 8 471 212 | 26 298 521 | 3 519 925 | 1 757 521 | 2 601 069 | 15 764 317 | 6 336 697 | 1 454 314 | 2 133 623 | 1 647 353 | 30 027 638 |
| 19 | Nuevo León | 174 912 084 | 272 679 406 | 15 683 728 | 154 298 763 | 32 977 460 | 14 623 323 | 13 471 183 | 100 197 802 | 42 678 799 | 7 325 460 | 10 927 550 | 4 438 825 | 119 605 962 |
| 20 | Oaxaca | 185 002 474 | 201 640 930 | 25 120 055 | 73 550 116 | 41 418 608 | 19 544 945 | 12 697 034 | 178 309 311 | 73 597 729 | 25 510 293 | 9 390 829 | 13 322 049 | 90 625 112 |
| 21 | Puebla | 251 920 367 | 240 318 351 | 15 539 000 | 100 633 263 | 34 151 435 | 20 712 348 | 12 926 225 | 208 782 952 | 46 642 994 | 14 445 608 | 12 590 743 | 11 159 693 | 46 968 879 |
| 22 | Querétaro | 58 470 087 | 45 331 492 | 3 987 885 | 46 035 677 | 5 709 662 | 2 362 601 | 6 241 545 | 35 889 827 | 14 698 794 | 1 220 631 | 1 485 201 | 1 980 885 | 23 473 363 |
| 23 | Quintana Roo | 21 668 976 | 42 810 113 | 4 467 189 | 14 809 294 | 4 207 131 | 2 126 905 | 1 977 180 | 29 188 387 | 4 966 247 | 2 102 981 | 1 209 026 | 1 783 673 | 13 544 203 |
| 24 | San Luis | 115 808 927 | 125 235 409 | 6 276 133 | 67 998 490 | 17 676 897 | 8 925 236 | 13 060 968 | 82 795 269 | 30 139 454 | 8 558 676 | 8 880 267 | 3 046 617 | 71 636 113 |
| 25 | Sinaloa | 81 567 417 | 83 853 491 | 18 271 813 | 63 426 318 | 14 111 047 | 5 469 320 | 7 828 364 | 44 589 212 | 10 527 165 | 6 885 843 | 5 713 518 | 3 004 424 | 36 576 777 |
| 26 | Sonora | 77 304 886 | 92 894 614 | 22 381 275 | 57 029 054 | 15 308 612 | 12 905 117 | 10 836 544 | 54 982 764 | 21 397 121 | 6 468 476 | 8 366 755 | 5 525 752 | 22 316 472 |
| 27 | Tabasco | 41 664 769 | 109 771 091 | 10 432 057 | 20 220 508 | 5 479 933 | 4 855 052 | 3 052 983 | 33 035 063 | 8 339 012 | 5 746 338 | 5 006 982 | 1 799 072 | 17 447 304 |
| 28 | Tampico | 141 724 185 | 150 599 796 | 13 990 634 | 76 781 485 | 20 374 332 | 11 722 084 | 10 312 671 | 58 421 161 | 20 286 787 | 7 339 461 | 13 100 042 | 4 520 935 | 70 316 181 |
| 29 | Tlaxcala | 68 613 833 | 65 766 866 | 4 506 846 | 21 215 152 | 20 355 038 | 2 276 862 | 3 934 730 | 48 881 618 | 14 537 599 | 16 429 837 | 1 421 728 | 3 275 916 | 26 496 271 |
| 30 | Veracruz | 320 237 153 | 414 017 796 | 65 575 315 | 194 905 221 | 40 318 019 | 39 166 977 | 15 442 653 | 201 915 403 | 57 411 200 | 44 310 836 | 20 218 107 | 14 009 326 | 125 825 977 |
| 31 | Yucatán | 83 227 815 | 112 327 667 | 1 496 186 | 39 030 179 | 7 806 115 | 4 334 202 | 3 449 631 | 41 856 652 | 17 514 392 | 8 543 381 | 6 846 790 | 6 883 392 | 89 723 188 |
| 32 | Zacatecas | 48 359 593 | 39 877 781 | 581 856 | 32 461 532 | 7 778 325 | 3 013 596 | 4 739 235 | 24 463 044 | 10 438 170 | 2 960 288 | 2 408 238 | 1 377 228 | 12 847 344 |

Fuente: INEGI. Encuesta nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Debido al formato expresado de la información (en millones de pesos) se procede a transformar a logaritmos con el fin de presentar ambos análisis exploratorios y tratar de obtener un mejor comportamiento de la información. Por lo que se explicará más adelante los resultados obtenidos (Cuadro 5.2).

**Total del gasto por tipo de alimento según entidad federativa, 2006
(Logaritmos)**

Cuadro 5. 2

| CLAVE | NOMBRE ENTIDAD | CEREALES | CARNES | PESCADO | LECHE | HUEVO | ACEITES | TUBERCULO | VERDURAS | FRUTAS | AZUCAR | CAFE | ESPECIAS | OTRO_ALI |
|-------|---------------------|----------|--------|---------|-------|-------|---------|-----------|----------|--------|--------|-------|----------|----------|
| 01 | Agascalientes | 17.75 | 17.37 | 14.14 | 17.43 | 15.30 | 15.19 | 15.14 | 17.35 | 16.07 | 15.05 | 14.48 | 14.67 | 17.15 |
| 02 | Baja California | 18.04 | 18.35 | 16.45 | 18.04 | 16.54 | 15.06 | 15.74 | 17.87 | 17.30 | 15.32 | 14.95 | 14.53 | 16.78 |
| 03 | Baja California Sur | 15.79 | 15.59 | 14.51 | 15.94 | 13.47 | 12.93 | 13.21 | 15.26 | 14.80 | 13.38 | 12.10 | 11.94 | 13.88 |
| 04 | Campeche | 17.20 | 17.72 | 14.95 | 16.10 | 15.07 | 14.19 | 14.88 | 16.64 | 15.98 | 14.93 | 15.27 | 14.16 | 16.82 |
| 05 | Chiapas | 18.17 | 18.24 | 15.88 | 17.64 | 16.52 | 15.83 | 16.40 | 17.75 | 16.73 | 15.92 | 16.15 | 15.14 | 17.76 |
| 06 | Chihuahua | 17.48 | 17.48 | 15.23 | 17.10 | 15.22 | 15.09 | 14.33 | 16.72 | 15.42 | 14.09 | 14.16 | 13.75 | 16.70 |
| 07 | Coahuila | 18.94 | 19.31 | 17.56 | 17.71 | 17.31 | 16.58 | 15.56 | 18.56 | 17.23 | 17.05 | 16.05 | 15.96 | 18.36 |
| 08 | Colima | 18.90 | 18.83 | 15.76 | 18.81 | 16.91 | 16.41 | 16.66 | 18.35 | 17.22 | 16.17 | 16.36 | 15.50 | 17.71 |
| 09 | Distrito Federal | 19.86 | 20.60 | 18.63 | 19.81 | 18.05 | 16.98 | 17.71 | 19.90 | 19.48 | 17.36 | 17.15 | 17.17 | 19.13 |
| 10 | Durango | 17.99 | 17.67 | 15.48 | 17.67 | 16.24 | 15.55 | 15.83 | 17.45 | 16.44 | 15.17 | 15.58 | 14.86 | 16.99 |
| 11 | Guerrero | 19.12 | 19.20 | 16.26 | 18.63 | 17.01 | 16.36 | 16.53 | 18.81 | 17.56 | 16.38 | 15.30 | 15.67 | 18.30 |
| 12 | Guanajuato | 18.92 | 19.11 | 17.80 | 18.55 | 17.10 | 16.73 | 15.71 | 18.62 | 16.74 | 16.22 | 15.16 | 15.38 | 17.30 |
| 13 | Hidalgo | 18.26 | 18.43 | 16.00 | 17.77 | 16.52 | 15.77 | 15.52 | 18.05 | 17.07 | 16.08 | 15.43 | 14.85 | 16.88 |
| 14 | Jalisco | 19.15 | 19.30 | 17.41 | 18.96 | 17.25 | 16.09 | 16.58 | 18.89 | 17.69 | 16.96 | 16.52 | 16.15 | 18.27 |
| 15 | México | 20.20 | 20.56 | 18.64 | 19.90 | 18.51 | 17.67 | 17.67 | 19.96 | 19.07 | 17.68 | 17.11 | 17.44 | 19.13 |
| 16 | Michoacán | 19.18 | 19.08 | 16.88 | 18.87 | 16.88 | 16.89 | 17.11 | 19.05 | 18.00 | 16.45 | 15.92 | 15.69 | 17.59 |
| 17 | Morelos | 18.26 | 18.23 | 15.57 | 17.70 | 16.01 | 15.76 | 15.13 | 17.85 | 17.21 | 15.27 | 15.74 | 14.59 | 16.76 |
| 18 | Nayarit | 17.31 | 17.16 | 15.95 | 17.09 | 15.07 | 14.38 | 14.77 | 16.57 | 15.66 | 14.19 | 14.57 | 14.31 | 17.22 |
| 19 | Nuevo León | 18.98 | 19.42 | 16.57 | 18.85 | 17.31 | 16.50 | 16.42 | 18.42 | 17.57 | 15.81 | 16.21 | 15.31 | 18.60 |
| 20 | Oaxaca | 19.04 | 19.12 | 17.04 | 18.11 | 17.54 | 16.79 | 16.36 | 19.00 | 18.11 | 17.05 | 16.06 | 16.40 | 18.32 |
| 21 | Puebla | 19.34 | 19.30 | 16.56 | 18.43 | 17.35 | 16.85 | 16.37 | 19.16 | 17.66 | 16.49 | 16.35 | 16.23 | 17.66 |
| 22 | Querétaro | 17.88 | 17.63 | 15.20 | 17.64 | 15.56 | 14.68 | 15.65 | 17.40 | 16.50 | 14.01 | 14.21 | 14.50 | 16.97 |
| 23 | Quintana Roo | 16.89 | 17.57 | 15.31 | 16.51 | 15.25 | 14.57 | 14.50 | 17.19 | 15.42 | 14.56 | 14.01 | 14.39 | 16.42 |
| 24 | San Luis | 18.57 | 18.65 | 15.65 | 18.03 | 16.69 | 16.00 | 16.39 | 18.23 | 17.22 | 15.96 | 16.00 | 14.93 | 18.09 |
| 25 | Sinaloa | 18.22 | 18.24 | 16.72 | 17.97 | 16.46 | 15.51 | 15.87 | 17.61 | 16.17 | 15.74 | 15.56 | 14.92 | 17.41 |
| 26 | Sonora | 18.16 | 18.35 | 16.92 | 17.86 | 16.54 | 16.37 | 16.20 | 17.82 | 16.88 | 15.68 | 15.94 | 15.52 | 16.92 |
| 27 | Tabasco | 17.55 | 18.51 | 16.16 | 16.82 | 15.52 | 15.40 | 14.93 | 17.31 | 15.94 | 15.56 | 15.43 | 14.40 | 16.67 |
| 28 | Tampico | 18.77 | 18.83 | 16.45 | 18.16 | 16.83 | 16.28 | 16.15 | 17.88 | 16.83 | 15.81 | 16.39 | 15.32 | 18.07 |
| 29 | Tlaxcala | 18.04 | 18.00 | 15.32 | 16.87 | 16.83 | 14.64 | 15.19 | 17.70 | 16.49 | 16.61 | 14.17 | 15.00 | 17.09 |
| 30 | Veracruz | 19.58 | 19.84 | 18.00 | 19.09 | 17.51 | 17.48 | 16.55 | 19.12 | 17.87 | 17.61 | 16.82 | 16.46 | 18.65 |
| 31 | Yucatán | 18.24 | 18.54 | 14.22 | 17.48 | 15.87 | 15.28 | 15.05 | 17.55 | 16.68 | 15.96 | 15.74 | 15.74 | 18.31 |
| 32 | Zacatecas | 17.69 | 17.50 | 13.27 | 17.30 | 15.87 | 14.92 | 15.37 | 17.01 | 16.16 | 14.90 | 14.69 | 14.14 | 16.37 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Lo interesante de aquí es ver si hay ciertos tipos de alimentos que son más necesarios en el consumo de los hogares que consideran por lo menos un integrante de este estrato de población adulta mayor; si estos son influenciados por la producción de cada entidad o bien tratar de ver si existe una relación entre todas las variables (combinaciones de ellas) sabiendo que no son independientes entre sí; por ejemplo si el consumo de leche tiene relación con el consumo de cereales, frutas y/o azúcar, o el consumo de carnes o pescados tiene relación con las verduras, tubérculos, el aceite, las especias y otros alimentos (Cuadro 5.2).

Es importante señalar que conforme la cifra sea más alta, mayor será el gasto realizado de acuerdo con cada uno de los 13 conceptos de alimentos, y conforme sea menor el dato menor será la cantidad invertida.

Análisis exploratorio de la información

Para iniciar se responde la siguiente pregunta: ¿Qué es un análisis exploratorio de datos? En estadística el análisis exploratorio es analizar el comportamiento, la estructura y las relaciones de los datos antes de aplicarle cualquier método estadístico. Con la intención de entender y captar el cómo se presentan los datos del fenómeno de estudio en forma individual o en las interrelaciones de las variables. Asimismo, efectuar este sencillo paso, permite organizar y preparar los datos y detectar errores en el diseño y captación de los mismos; en el tratamiento y evaluación de los datos aberrantes o atípicos y, en consecuencia probar algún supuesto adherido y señalado en la teoría estadística, y para este caso multivariado, como son: normalidad, linealidad y homocedasticidad.

Es importante señalar que esto último se revisó en forma breve y no se profundizó tanto en el tema, debido a que el fin del estudio es el uso del método de componentes principales que son afectados por los puntos extremos.

Primero se inició con un *Análisis individual* tanto gráfico como numérico de las variables consideradas de este trabajo; posteriormente se probó por medio de un *Análisis multivariado* la posible existencia de normalidad y la identificación de datos extraños (aberrantes o atípicos).

Análisis individual

En el siguiente Cuadro 5.3 se presentan las principales estadísticas de las variables de análisis, como son: el mínimo y el máximo valor de gasto observado, la mediana, la media o promedio aritmético, y los cuartiles primero y tercero, con la idea de mostrar como es el comportamiento de los datos para cada una de las variables involucradas.

Se puede apreciar que las *carnes*, los *cereales*, las *verduras* y la *leche* son las cuatro variables más importantes en cuanto a concentración del gasto de los hogares donde al menos uno de sus residentes tiene 65 y más años de edad. El concepto más relevante en cuanto al gasto son las *carnes*; se identifica una entidad que tiene el máximo valor con 879 millones 769 mil 414 pesos, que corresponde a lo invertido en el Distrito Federal (ver Cuadro 5.1). En contraste el mínimo valor se identifica en Baja California Sur con 5 millones 871 mil 42. Por otra parte, más de la mitad de las 32 entidades no rebasan los 106 millones de pesos en el consumo de este alimento (mediana) y que se encuentra por debajo de los 173 millones 13 mil 571 de pesos que registra la media aritmética de todo el conjunto. De ahí que, al revisar los cuartiles, se aprecia una diferencia amplia entre el máximo valor (antes señalado) y el límite del tercer cuartil (205 millones 933 mil 318) de poco más de 674 millones. En contra parte, con una diferencia menos significativa, entre el límite del primer cuartil (49 millones 135 mil 489) y el mínimo valor (enunciado con anterioridad) es de poco más de 44 millones. De ahí se puede deducir la existencia de entidades que registran cifras en el gasto para esta variable fuera de lo normal de lo registrado por la mayoría del conjunto.

Los *cereales* se ubican en segundo orden de importancia, la situación es semejante, la diferencia entre la mediana de las observaciones (83 millones 999 mil 232) y el promedio (135 millones 743 mil 197) es cercana a los 52 millones. Entre el tercer cuartil (177 millones 434 mil 681) y el máximo valor (591 millones

987 mil 599) que se registra en el Estado de México hay un margen cercano a los 324 millones. En tanto que entre el primer cuartil (poco más de 56 millones) y el mínimo valor (alrededor de 7 millones) que se registra en Baja California Sur, la distancia es cercana a los 49 millones. Lo que significaría que es más corta la variabilidad entre las entidades.

Principales estadísticas de las trece variables de alimentos

Cuadro 5. 3

| CEREALES | CARNES | PESCADO | LECHE |
|---------------------|----------------------|---------------------|----------------------|
| Mín. : 7 227 229 | Mín. : 5 871 421 | Mín. : 581 856 | Mín. : 8 402 246 |
| 1st Qu.: 56 635 520 | 1st Qu.: 49 135 489 | 1st Qu.: 4 496 932 | 1st Qu.: 36 082 703 |
| Mediana: 83 999 232 | Mediana :105 372 729 | Mediana : 9 669 782 | Mediana : 54 619 902 |
| Media :135 743 197 | Mean : 173 013 571 | Mean : 21 531 168 | Mean : 91 681 482 |
| 3rd Qu.:177 434 681 | 3rd Qu.: 205 933 318 | 3rd Qu.: 21 719 859 | 3rd Qu.: 116 036 390 |
| Máx. : 591 987 599 | Máx. : 879 769 414 | Máx. : 124 378 385 | Máx. : 436 842 866 |

| HUEVO | ACEITES | TUBERCULO | VERDURAS |
|----------------------|---------------------|---------------------|----------------------|
| Mín. : 711 148 | Mín. : 410 986 | Mín. : 543 366 | Mín. : 4 244 716 |
| 1st Qu.: 7 261 160 | 1st Qu.: 356 1974 | 1st Qu.: 3 756 580 | 1st Qu.: 35 506 467 |
| Mediana : 15 305 339 | Mediana : 7 291 957 | Mediana : 7 160 008 | Mediana : 57 104 866 |
| Mean : 21 462 287 | Mean : 11 144 774 | Mean : 11 006 780 | Mean : 101 143 129 |
| 3rd Qu.: 27 862 295 | 3rd Qu.: 14 925 648 | 3rd Qu.: 13 339 207 | 3rd Qu.: 128 308 830 |
| Máx. : 109 500 994 | Máx. : 47 175 151 | Máx. : 49 234 720 | Máx. : 467 172 011 |

| FRUTAS | AZÚCAR | CAFÉ |
|----------------------|---------------------|---------------------|
| Mín. : 2 674 364 | Mín. : 644 784 | Mín. : 180 487 |
| 1st Qu.: 10 504 917 | 1st Qu.: 3 759 733 | 1st Qu.: 2 925 174 |
| Mediana : 20 841 954 | Mediana : 7 791 641 | Mediana : 6 336 371 |
| Mean : 38 830 406 | Mean : 11 803 935 | Mean : 8 044 574 |
| 3rd Qu.: 42 236 430 | 3rd Qu.: 14 089 295 | 3rd Qu.: 10 501 970 |
| Máx. : 289 296 243 | Máx. : 47 560 770 | Máx. : 28 177 532 |

| ESPECIAS | OTRO_ALI |
|---------------------|----------------------|
| Mín. : 154 034 | Mín. : 1 065 003 |
| 1st Qu.: 2 020 363 | 1st Qu.: 21 110 272 |
| Mediana : 3 510 099 | Mediana : 34 532 841 |
| Mean : 6 385 018 | Mean : 56 308 694 |
| 3rd Qu.: 6 594 008 | 3rd Qu.: 86 734 923 |
| Máx. : 37 617 450 | Máx. : 204 043 796 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En tercer orden de importancia del gasto, se encuentra las *verduras*, al registrar un promedio de gasto entre las entidades del país cercano a los 101 millones de pesos y donde la mitad de ellos tienen gastos superiores a los 57 millones. La cifra más alta corresponde al Estado de México y la menos importante a Baja California Sur. Situación semejante se presenta en la Leche, donde ambas entidades registran la misma importancia de las cifras de gasto, y donde la mitad de las entidades tienen gastos en este alimento básico de poco menos de 54 millones de pesos.

El resto de las variables presenta cifras menos importantes, y donde las diferencias entre la media y la mediana cada vez son más cercanas; sin embargo, al señalar algunas distinciones propias, en el caso del pescado, Zacatecas tiene una cifra menos importante al invertir en este producto menos de 600 mil pesos, y el Estado de México la más alta con más de 124 millones.

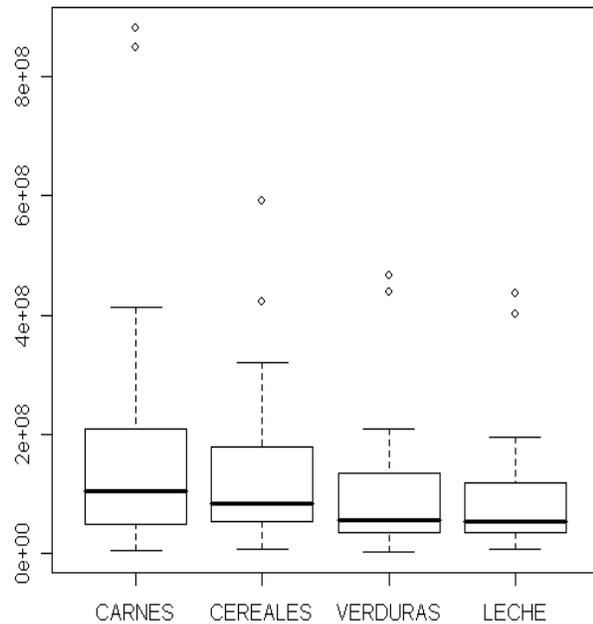
En huevo, aceites, azúcar, especias y otros alimentos, Baja California Sur es la que menos gasta en el consumo de este producto y el Estado de México la que más gasta. En tubérculo, café y en frutas, las entidades que destacan son Baja California Sur con el mínimo valor y el Distrito Federal con la cifra más importante, precisamente en Café, la diferencia entre la media y la mediana es la menos significativa con cerca de 1.7 millones, su promedio es de 8 millones y la mediana es de 6.3 millones.

Al revisar el comportamiento de la información para cada variable según su distribución (en diagramas de caja o boxplot) se distinguen los siguientes resultados generales:

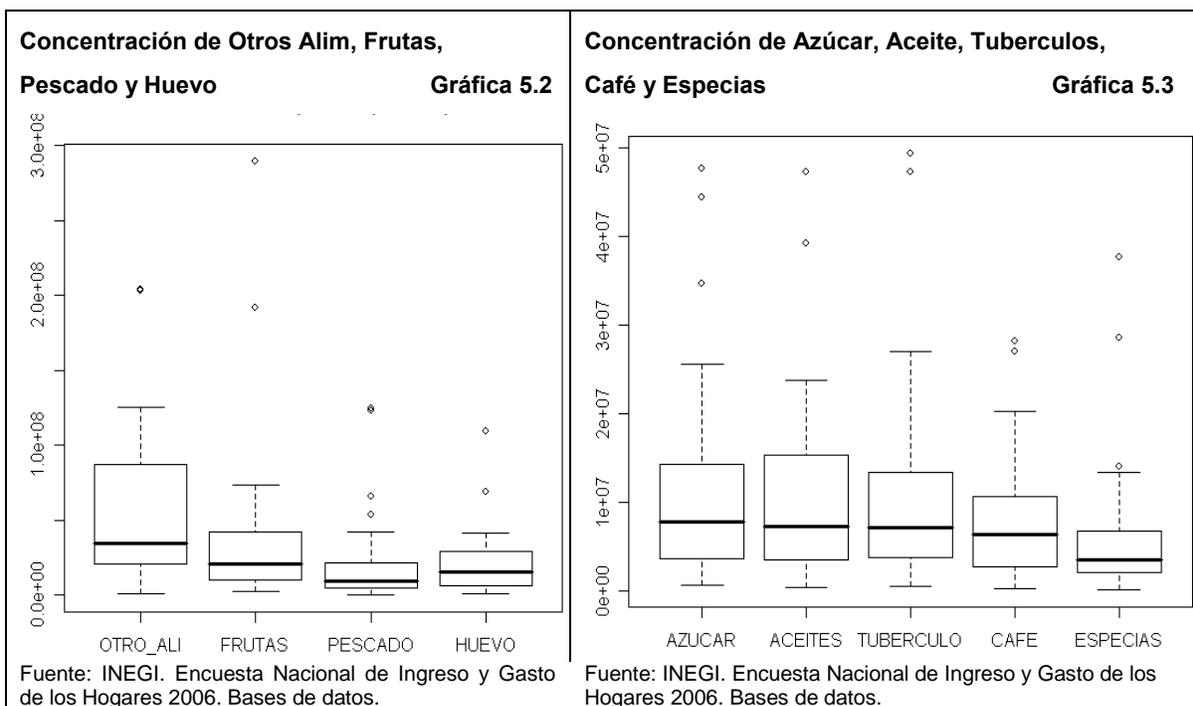
1. La mayoría de las distribuciones de las variables, reflejan medianas (líneas horizontales) que están por debajo de la media (mitad del rectángulo).
2. Las colas superiores (o bigotes, líneas punteadas) son más largas que las inferiores, asimismo, las distancias son amplias entre el tercer cuartil (límite superior del rectángulo) y el valor máximo.
3. Hay valores extraños, atípicos o aberrantes (puntos superiores alejados del rectángulo), fuera del conjunto de datos (por arriba), unos más lejanos que otros.
4. Algunas concentraciones son pequeñas debido a gastos menos representativos que en las otras variables (tamaño del rectángulo).

Concentración de Carnes, Cereales, Verduras y Leche

Gráfica 5.1



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.



En general se tiene el siguiente resultado: los hogares con al menos un residente de 65 y más años de edad gastan más en el consumo de carnes, cereales, verduras y leche y destina pocos recursos económicos para la adquisición de tubérculos, café y especias. Hay entidades /Estado de México y Distrito Federal) que registran cifras superiores a lo registrado por la mayoría de ellas (casos atípicos o fuera de la concentración), debido posiblemente a la alta concentración de población y por tanto es mayor el consumo de dicho alimento.

Es posible que al comparar con los gastos de los hogares totales, el comportamiento sea semejante, quizás el orden de las variables sea distintas por la cantidad de gasto asignado, lo que se analizaría en un estudio de comparabilidad en otro momento.

Análisis multivariado (puntos extremos, detección y tratamiento)

Para complementar el análisis exploratorio de la información habría que responder a lo siguiente: las variables seleccionadas del gasto de los hogares ¿Pueden modelarse como una distribución normal multivariada aún cuando el Análisis de Componentes Principales no lo exige? En caso negativo, se trató de encontrar un subconjunto de variables para las cuales se pueda suponer normalidad multivariada. Y posteriormente se identificó si existen datos aberrantes (datos que están alejados de la concentración de la mayoría de registros).

Para dar respuesta a lo anterior, se procede a revisar el comportamiento y las relaciones entre variables. Para ello se presentan las dos siguientes gráficas múltiples, una donde los datos están en valores absolutos y otra con datos transformados en logaritmos; esto debido a que posiblemente se pueda presentar el caso de reflejar distribuciones unimodales y asimétricas y en consecuencia la no existencia de normalidad.

En el primer ejercicio, se muestran las gráficas de las relaciones bivariadas (en Gráfica 5.4) entre las trece variables (con sus datos originales) las cuales muestran alta dispersión de la información y donde gran parte de los datos se concentran al principio, claramente se nota en los histogramas que son asimétricas, por lo que tratar de modelar a las variables como distribución multivariada es poco factible ya que vaticina la no existencia de normalidad lo que se confirma con la prueba de Shapiro-Wilks (es una prueba estadística que se aplica a lo más a 50 unidades muestrales, ver su metodología en glosario, y los resultados (Cuadro 5.4) se obtuvieron mediante el paquete estadístico R 2.5.1) al registrar estadísticas calculadas (p valor) muy pequeños en comparación con los datos reales de la estadística de Shapiro-Wilks (W), y por otra parte las gráficas de cuartiles (Gráfica 5.5) donde se observa que todas las variables muestran problemas sobre todo en las puntas superiores ya que hay puntos muy alejados

con respecto al grueso de observaciones, y que pueden ser determinadas como puntos aberrantes o atípicos, y que influyen en que la distribución de la variable no sea normal.

La gráfica de probabilidad Ji cuadrada (Gráfica 5.6), muestra como una gran parte de las observaciones se alejan de la recta, y en consecuencia su estructura no se comporta como una normal.

En el segundo caso, donde la gráfica multivariada (Gráfica 5.7) muestra los datos transformados, aún cuando se muestra dispersión de los datos, los histogramas presentan simetrías más visibles tales son los casos de pescado, leche, cereales y carnes. Por lo que aún cuando no se pueda modelar a todas las variables como una distribución multivariada, algunas de ellas si se pueden modelar dado que confirman normalidad, tales son los casos de leche, pescado, frutas, azúcar, verduras, carnes y aceites al registrar estadísticas calculadas cercanas (Cuadro 5.5) a las reales de Shapiro-Wilks.

En tanto que en las gráficas de cuartiles (Gráfica 5.8), la mayoría de las gráficas muestran tendencias casi uniformes, aunque en las puntas hay problemas ligeros atribuibles por la presencia de puntos aberrantes. La gráfica de probabilidad Ji cuadrada (Gráfica 5.9), muestra como en la parte central y en el extremo superior las observaciones se alejan de la recta, por lo que su estructura no se comporta como una normal.

Caso de datos originales sin transformación (valores absolutos)

Gráfica multivariada de las variables del gasto

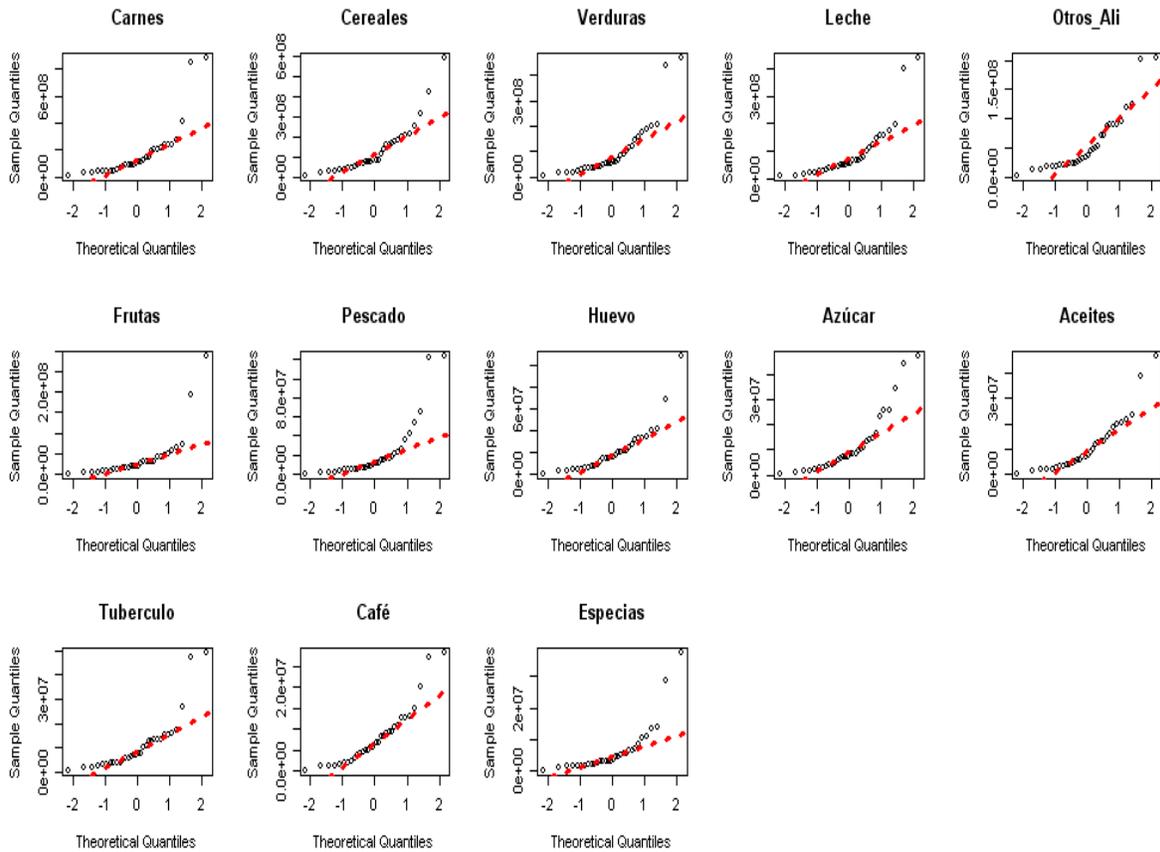
Gráfica 5.4



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Gráfica de cuartiles de las variables del gasto

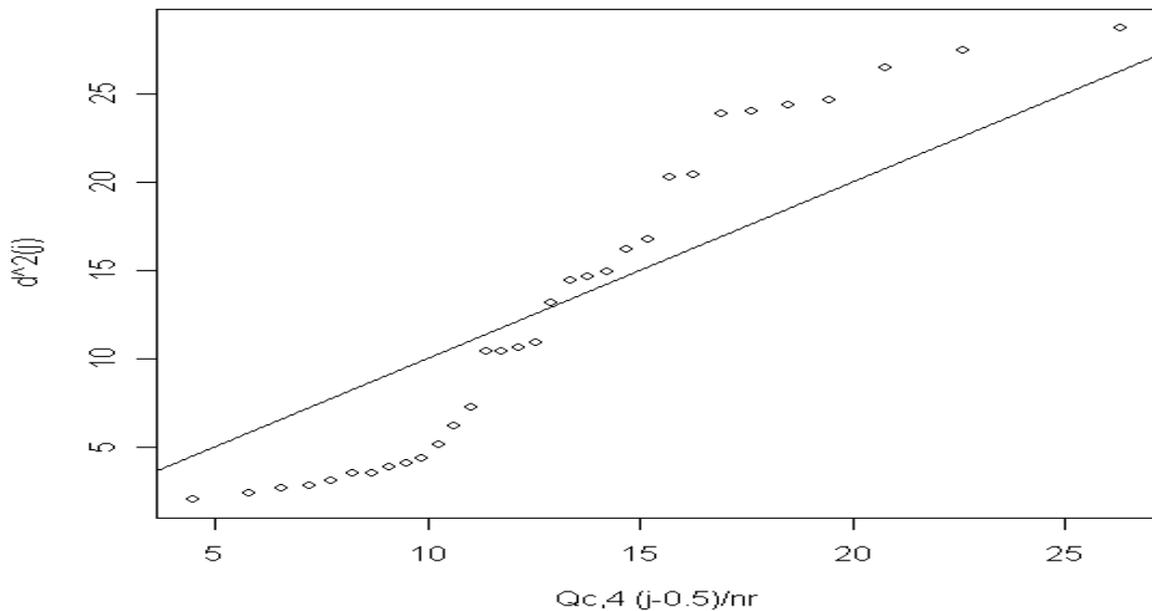
Gráfica 5.5



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Gráfica de probabilidad Ji cuadrada de cuartiles

Gráfica 5.6



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Prueba de normalidad con la estadística de Shapiro-Wilks

Cuadro 5. 4

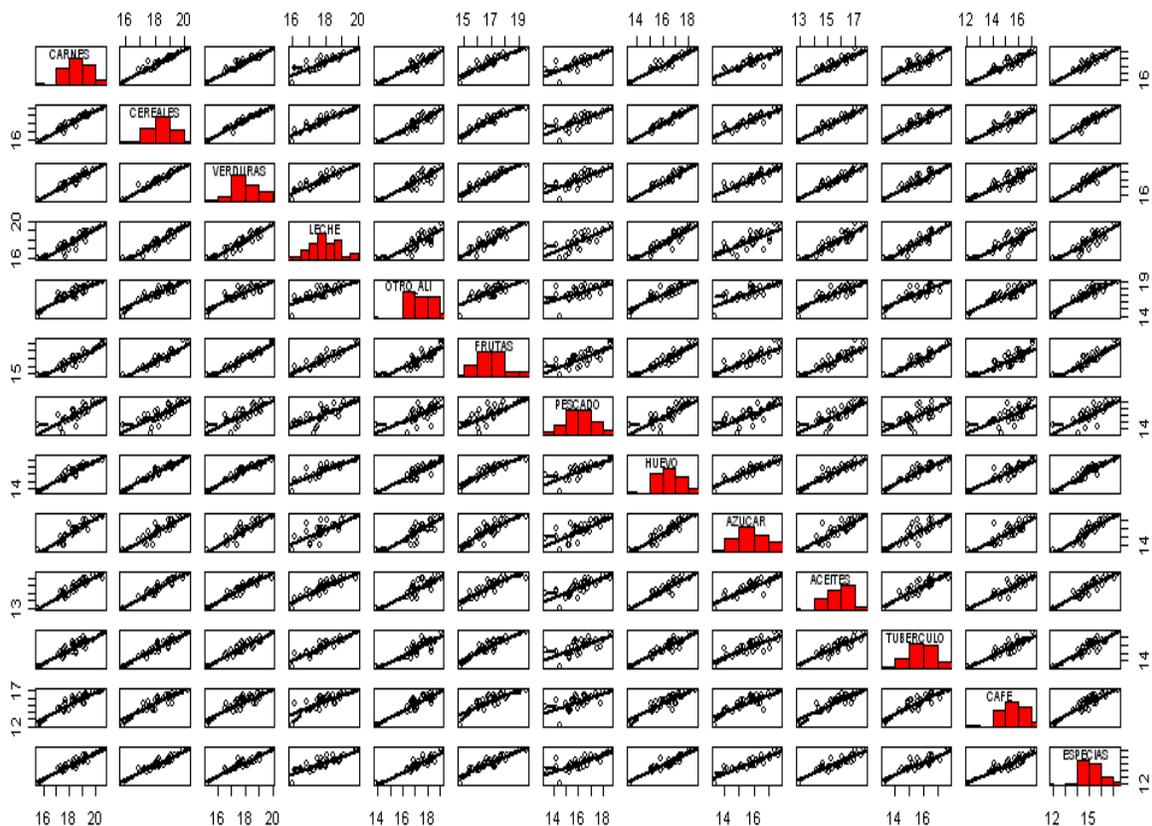
| | | | | |
|------|-------------|-----------------------------|-----------------------------|---------------------|
| "1" | "CARNES" | Shapiro-Wilk normality test | data: datos[, i] W = 0.6556 | p-value = 2.008e-07 |
| "2" | "CEREALES" | Shapiro-Wilk normality test | data: datos[, i] W = 0.7919 | p-value = 2.884e-05 |
| "3" | "VERDURAS" | Shapiro-Wilk normality test | data: datos[, i] W = 0.7244 | p-value = 2.060e-06 |
| "4" | "LECHE" | Shapiro-Wilk normality test | data: datos[, i] W = 0.7045 | p-value = 1.015e-06 |
| "5" | "OTRO_ALI" | Shapiro-Wilk normality test | data: datos[, i] W = 0.8037 | p-value = 4.783e-05 |
| "6" | "FRUTAS" | Shapiro-Wilk normality test | data: datos[, i] W = 0.5489 | p-value = 9.102e-09 |
| "7" | "PESCADO" | Shapiro-Wilk normality test | data: datos[, i] W = 0.635 | p-value = 1.057e-07 |
| "8" | "HUEVO" | Shapiro-Wilk normality test | data: datos[, i] W = 0.7474 | p-value = 4.828e-06 |
| "9" | "AZUCAR" | Shapiro-Wilk normality test | data: datos[, i] W = 0.7872 | p-value = 2.361e-05 |
| "10" | "ACEITES" | Shapiro-Wilk normality test | data: datos[, i] W = 0.8116 | p-value = 6.793e-05 |
| "11" | "TUBERCULO" | Shapiro-Wilk normality test | data: datos[, i] W = 0.7162 | p-value = 1.535e-06 |
| "12" | "CAFE" | Shapiro-Wilk normality test | data: datos[, i] W = 0.8502 | p-value = 0.00042 |
| "13" | "ESPECIAS" | Shapiro-Wilk normality test | data: datos[, i] W = 0.6479 | p-value = 1.575e-07 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

b) Caso de datos transformados a logaritmos

Gráfica multivariada de las variables del gasto (valores en logaritmos)

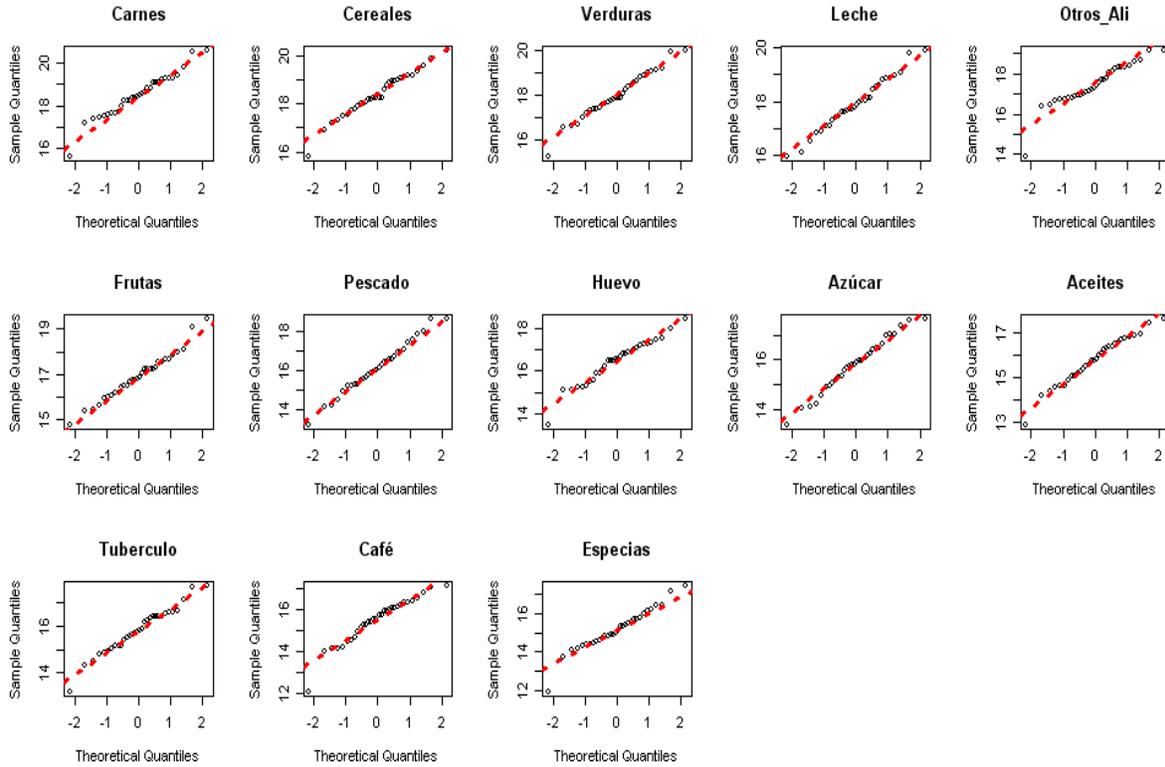
Gráfica 5.7



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Gráficas de cuartiles de las variables del gasto (valores en logaritmos)

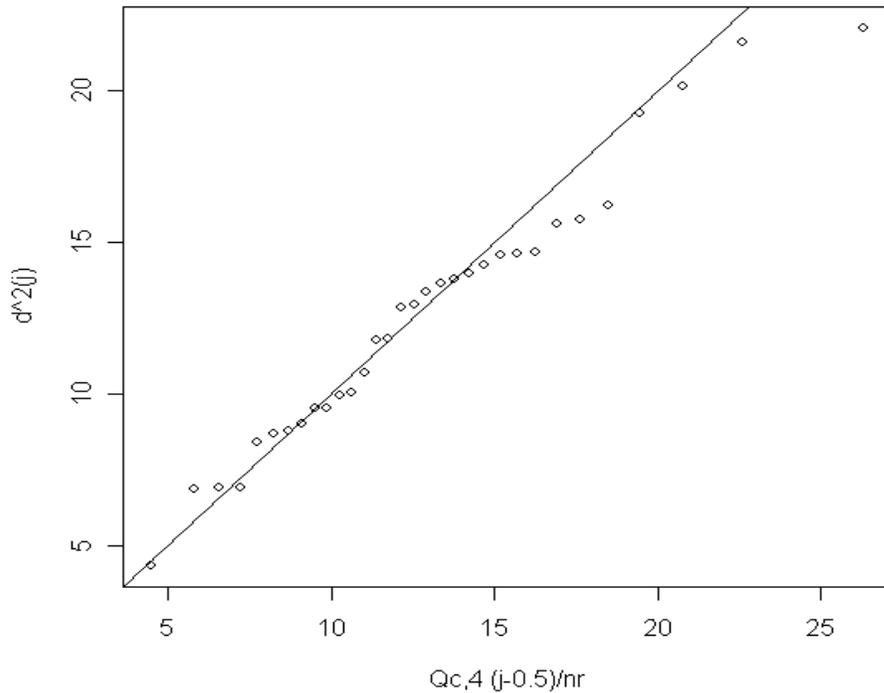
Gráfica 5.8



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Gráfica de probabilidad Ji cuadrada de cuartiles (valores en logaritmos)

Gráfica 5.9



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

**Prueba de normalidad con la estadística de Shapiro-Wilks
(valores en logaritmos)**

Cuadro 5.5

| | | | | |
|------|-------------|-----------------------------|------------------------------|-------------------|
| "1" | "CARNES" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9662, | p-value = 0.4005 |
| "2" | "CEREALES" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9778, | p-value = 0.732 |
| "3" | "VERDURAS" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9793, | p-value = 0.7796 |
| "4" | "LECHE" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9841, | p-value = 0.9057 |
| "5" | "OTRO_ALI" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9045, | p-value = 0.00809 |
| "6" | "FRUTAS" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9817, | p-value = 0.8474 |
| "7" | "PESCADO" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9885, | p-value = 0.977 |
| "8" | "HUEVO" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9646, | p-value = 0.3642 |
| "9" | "AZUCAR" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9817, | p-value = 0.8461 |
| "10" | "ACEITES" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9778, | p-value = 0.7322 |
| "11" | "TUBERCULO" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9762, | p-value = 0.6853 |
| "12" | "CAFE" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9367, | p-value = 0.06026 |
| "13" | "ESPECIAS" | Shapiro-Wilk normality test | data: datos[, i] W = 0.9599, | p-value = 0.2735 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Por último, la prueba de normalidad multivariada la cual es determinada al comparar las estadísticas de normalidad con la de la Ji cuadrada, se tiene los siguientes resultados obtenidos al comparar la estadística de prueba Ji cuadrada, y la de normalidad.

Primer caso. El valor de la Ji cuadrada es menor que el de la normalidad, entonces, no hay normalidad multivariada.

$$1 - pchisq(nr*Ap/6, nc*(nc+1)*(nc+2)/6) \quad [1] \quad 0$$

$$1 - pnorm(Kp,nc*(nc+2), 8*nc*(nc+2)/nr) \quad [1] \quad 0.2141498$$

Segundo caso. El valor de la Chi cuadrada es mayor que el de la normalidad, entonces, hay normalidad multivariada:

| | | |
|---|-----|-----------|
| $1 - \text{pchisq}(nr \cdot Ap/6, nc \cdot (nc+1) \cdot (nc+2)/6)$ | [1] | 0.9236461 |
| $1 - \text{pnorm}(Kp, nc \cdot (nc+2), 8 \cdot nc \cdot (nc+2)/nr)$ | [1] | 0.644744 |

De ahí que si el modelo considerara los valores transformados, entonces sí se puede modelar en forma multivariada para un subconjunto de variables por la existencia de normalidad en ellas. Para nuestro estudio, con base en estos resultados se procedió a modelar todas las variables sin tomar en cuenta el hecho del supuesto de normalidad debido a que queremos utilizar el método de componentes principales dado que hasta ahora no se ha realizado algún ejercicio semejante. La prueba de normalidad multivariada se realizó bajo el supuesto de identificar la existencia de modelar la información con supuesto de normalidad sin aplicar la técnica de componentes principales con el fin de observar el comportamiento natural de la información bajo el enfoque de análisis exploratorio.

Aplicación del método de componentes principales

Así, una vez que se identificó por el análisis exploratorio de los datos el mejor comportamiento de éstos, se aplicó el método de componentes principales a la información de las trece variables del gasto en alimentos transformados a logaritmos.

Análisis de la matriz de correlaciones

Al revisar los resultados de esta matriz, se identificó si hay asociación entre variables y por tanto el de aplicar los Componentes Principales, con la idea de reducir la dimensionalidad de la información.

Asimismo al revisar la matriz de correlaciones (Cuadro 5.6) se distinguió lo siguiente:

1. Las carnes están más correlacionados con verduras, cereales, huevo y especias, y es menor con el pescado.
2. Las cereales tienen alta correlación con las verduras, carnes, huevo, aceites, y es menor con otros alimentos y café.
3. Las verduras está más correlacionado con cereales, carnes, huevo, frutas y especias, y con menor importancia con café y pescado.
4. La leche está muy correlacionado con las cereales, y con menor relación con pescado.
5. Otros alimentos tiene una relación alta con especias y menor relación con pescado.

Matriz de correlaciones de las variables del gasto en alimentos

Cuadro 5. 6

| | CARNES | CEREALES | VERDURAS | LECHE | OTRO_ALI | FRUTAS | PESCADO |
|-----------|---------|----------|----------|----------|-----------|----------|-----------|
| CARNES | 1.00000 | 0.954241 | 0.956222 | 0.875338 | 0.8872274 | 0.909772 | 0.8062680 |
| CEREALES | 0.95424 | 1.000000 | 0.970669 | 0.935565 | 0.8868921 | 0.926422 | 0.7526193 |
| VERDURAS | 0.95622 | 0.970669 | 1.000000 | 0.908863 | 0.8389127 | 0.944633 | 0.7834222 |
| LECHE | 0.87534 | 0.935565 | 0.908863 | 1.000000 | 0.7879698 | 0.896810 | 0.7471075 |
| OTRO_ALI | 0.88723 | 0.886892 | 0.838913 | 0.787970 | 1.0000000 | 0.798861 | 0.6292036 |
| FRUTAS | 0.90977 | 0.926422 | 0.944633 | 0.896810 | 0.7988610 | 1.000000 | 0.7312368 |
| PESCADO | 0.80627 | 0.752620 | 0.783422 | 0.747108 | 0.6292036 | 0.731237 | 1.0000000 |
| HUEVO | 0.93770 | 0.958898 | 0.953840 | 0.877992 | 0.8514679 | 0.903573 | 0.7739270 |
| AZUCAR | 0.90482 | 0.893250 | 0.905529 | 0.757403 | 0.8082845 | 0.846610 | 0.7339590 |
| ACEITES | 0.91743 | 0.944457 | 0.928321 | 0.883514 | 0.8114282 | 0.847381 | 0.7700048 |
| TUBERCULO | 0.86484 | 0.912770 | 0.903060 | 0.916003 | 0.8168634 | 0.905523 | 0.6899570 |
| CAFE | 0.88612 | 0.871252 | 0.828151 | 0.790783 | 0.8602758 | 0.820261 | 0.6697259 |
| ESPECIAS | 0.93215 | 0.937288 | 0.940415 | 0.839593 | 0.9077894 | 0.889497 | 0.7397449 |

| | HUEVO | AZUCAR | ACEITES | TUBERCULO | CAFE | ESPECIAS |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| CARNES | 0.9376984 | 0.9048227 | 0.9174273 | 0.8648373 | 0.8861161 | 0.9321540 |
| CEREALES | 0.9588976 | 0.8932497 | 0.9444569 | 0.9127700 | 0.8712518 | 0.9372883 |
| VERDURAS | 0.9538401 | 0.9055288 | 0.9283210 | 0.9030599 | 0.8281505 | 0.9404150 |
| LECHE | 0.8779921 | 0.7574033 | 0.8835143 | 0.9160027 | 0.7907830 | 0.8395930 |
| OTRO_ALI | 0.8514679 | 0.8082845 | 0.8114282 | 0.8168634 | 0.8602758 | 0.9077894 |
| FRUTAS | 0.9035729 | 0.8466104 | 0.8473807 | 0.9055231 | 0.8202610 | 0.8894968 |
| PESCADO | 0.7739270 | 0.7339590 | 0.7700048 | 0.6899570 | 0.6697259 | 0.7397449 |
| HUEVO | 1.0000000 | 0.9175255 | 0.9097497 | 0.8966697 | 0.8442026 | 0.9224054 |
| AZUCAR | 0.9175255 | 1.0000000 | 0.8521060 | 0.7892255 | 0.8141048 | 0.9093857 |
| ACEITES | 0.9097497 | 0.8521060 | 1.0000000 | 0.8615819 | 0.8727770 | 0.8893500 |
| TUBERCULO | 0.8966697 | 0.7892255 | 0.8615819 | 1.0000000 | 0.8583265 | 0.8682404 |
| CAFE | 0.8442026 | 0.8141048 | 0.8727770 | 0.8583265 | 1.0000000 | 0.8711922 |
| ESPECIAS | 0.9224054 | 0.9093857 | 0.8893500 | 0.8682404 | 0.8711922 | 1.0000000 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

6. Las frutas se relacionan mucho con las verduras y en menor medida con otros alimentos.
7. El pescado tiene la relación más alta en pescado y la más baja en otros alimentos.

8. El huevo tiene altas correlaciones con cereales y verduras, y con menor importancia en pescado.
9. El azúcar tienen alta relación con el huevo, y tienen menor relación con el pescado.
10. El aceite sobresale con su relación con los cereales y su menor cantidad con los pescados.
11. El tubérculo distingue su relación con los cereales y la leche y en menor cantidad con el pescado.
12. En tanto que el café se distingue por tener más relación con los cereales y las especias y menos relación con el pescado.
13. Las especias destacan por su alta relación con las especias y su menor relación con la leche.

Análisis de la matriz de varianzas y covarianzas

Al revisar la matriz de varianzas y covarianzas, en primer lugar se observó que es una matriz de coeficientes positivos y altos.

Sobresalieron (por columnas) las correspondientes de pescado, huevo y azúcar que registran mayor cantidad de registros altos (más de 1.0), le siguen las especias, aceite y carnes, en tanto que la leche y los tubérculos mostraron más registros con menos significancia (Cuadro 5.7).

Matriz de varianzas y covarianzas de las variables del gasto en alimentos

Cuadro 5. 7

| | CARNES | CEREALES | VERDURAS | LECHE | OTRO_ALI | FRUTAS | PESCADO |
|----------|----------|-----------|-----------|----------|-----------|-----------|-----------|
| CARNES | 1.034881 | 0.8910492 | 0.9798767 | 0.843030 | 0.9165788 | 0.9435135 | 1.0357823 |
| CEREALES | 0.891049 | 0.8425518 | 0.8975051 | 0.813006 | 0.8267207 | 0.8669171 | 0.8724038 |
| VERDURAS | 0.979877 | 0.8975051 | 1.0146936 | 0.866738 | 0.8581712 | 0.9700662 | 0.9965687 |
| LECHE | 0.843030 | 0.8130059 | 0.8667378 | 0.896279 | 0.7575669 | 0.8655510 | 0.8931999 |
| OTRO_ALI | 0.916579 | 0.8267207 | 0.8581712 | 0.757567 | 1.0312868 | 0.8270497 | 0.8069094 |
| FRUTAS | 0.943514 | 0.8669171 | 0.9700662 | 0.865551 | 0.8270497 | 1.0393008 | 0.9413965 |
| PESCADO | 1.035782 | 0.8724038 | 0.9965687 | 0.893200 | 0.8069094 | 0.9413965 | 1.5947324 |
| HUEVO | 0.984820 | 0.9086973 | 0.9919545 | 0.858146 | 0.8927024 | 0.9510044 | 1.0090034 |
| AZUCAR | 0.982210 | 0.8749171 | 0.9733420 | 0.765147 | 0.8758903 | 0.9209795 | 0.9890344 |
| ACEITES | 0.978352 | 0.9087805 | 0.9802666 | 0.876827 | 0.8638103 | 0.9055820 | 1.0193325 |
| TUBERCUL | 0.847034 | 0.8066427 | 0.8758013 | 0.834911 | 0.7986577 | 0.8887747 | 0.8388561 |
| CAFE | 0.954063 | 0.8464148 | 0.8829131 | 0.792356 | 0.9246314 | 0.8850421 | 0.8951220 |
| ESPECIAS | 0.999359 | 0.9066927 | 0.9983335 | 0.837682 | 0.9715462 | 0.9556605 | 0.9844973 |

| | HUEVO | AZUCAR | ACEITES | TUBERCUL | CAFE | ESPECIAS |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| CARNES | 0.9848203 | 0.9822099 | 0.9783524 | 0.8470344 | 0.9540629 | 0.9993587 |
| CEREALES | 0.9086973 | 0.8749171 | 0.9087805 | 0.8066427 | 0.8464148 | 0.9066927 |
| VERDURAS | 0.9919545 | 0.9733420 | 0.9802666 | 0.8758013 | 0.8829131 | 0.9983335 |
| LECHE | 0.8581456 | 0.7651465 | 0.8768268 | 0.8349105 | 0.7923557 | 0.8376819 |
| OTRO_ALI | 0.8927024 | 0.8758903 | 0.8638103 | 0.7986577 | 0.9246314 | 0.9715462 |
| FRUTAS | 0.9510044 | 0.9209795 | 0.9055820 | 0.8887747 | 0.8850421 | 0.9556605 |
| PESCADO | 1.0090034 | 0.9890344 | 1.0193325 | 0.8388561 | 0.8951220 | 0.9844973 |
| HUEVO | 1.0658533 | 1.0107938 | 0.9845759 | 0.8912565 | 0.9224369 | 1.0035966 |
| AZUCAR | 1.0107938 | 1.1386531 | 0.9531646 | 0.8108086 | 0.9194271 | 1.0226629 |
| ACEITES | 0.9845759 | 0.9531646 | 1.0988973 | 0.8695542 | 0.9683293 | 0.9825166 |
| TUBERCUL | 0.8912565 | 0.8108086 | 0.8695542 | 0.9269218 | 0.8746117 | 0.8809478 |
| CAFE | 0.9224369 | 0.9194271 | 0.9683293 | 0.8746117 | 1.1201662 | 0.9717261 |
| ESPECIAS | 1.0035966 | 1.0226629 | 0.9825166 | 0.8809478 | 0.9717261 | 1.1106501 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Descomposición espectral de la matriz (valores propios y vectores propios)

Se procedió a determinar los valores propios y los vectores propios, con la intención de identificar las componentes que son más importantes.

De los valores propios, el número de éstos que sean diferentes de cero proporcionan la dimensión del espacio en el que se encuentran las observaciones; un valor propio nulo revelaría la dependencia lineal entre las variables originales. Así se tiene que el valor propio correspondiente a la primera componente es superior a la unidad, por lo que es susceptible de explicar la máxima variabilidad del modelo al representar 11.355 (Cuadro 5.8).

Vector de valores propios de las variables del gasto en alimentos

Cuadro 5. 8

| | | | | | | | |
|-----|--------|-------|-------|-------|-------|-------|------|
| [1] | 11.355 | 0.44 | 0.345 | 0.24 | 0.167 | 0.145 | 0.09 |
| [8] | 0.067 | 0.052 | 0.036 | 0.032 | 0.014 | 0.011 | |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En tanto que de la matriz de vectores propios, al revisar a cada componente se tiene que el primero claramente mostró datos negativos, esto es, conforme se gastó en carnes (1) se estuvo gastando en el resto de los productos alimenticios (2 al 13). Para el resto de las componentes (columnas) se tuvieron relaciones inversas (algunos datos son positivos y otros negativos), esto es, conforme se gastó en ciertos alimentos (datos positivos) también se estuvo reduciendo el gasto por otros conceptos de alimentos distintos, datos negativos (Cuadro 5.9).

Matriz de vectores propios de las variables del gasto en alimentos

Cuadro 5. 9

| | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] |
|-------|------------|----------|-------------|-------------|-------------|-------------|----------|
| [1,] | -0.2892994 | 0.033635 | -0.16008579 | -0.02629716 | -0.068117 | 0.03029107 | -0.37822 |
| [2,] | -0.2922776 | -0.05629 | 0.09783196 | 0.09050933 | -0.0070648 | 0.27589872 | -0.05411 |
| [3,] | -0.2902399 | 0.08041 | 0.06814628 | 0.28607325 | -0.003241 | 0.09908494 | -0.15240 |
| [4,] | -0.2743329 | 0.10412 | 0.54423362 | -0.09187560 | -0.1473527 | 0.25876461 | -0.02570 |
| [5,] | -0.2662602 | -0.43135 | -0.24680063 | -0.25857670 | -0.625329 | 0.15619273 | 0.07138 |
| [6,] | -0.2794435 | 0.02036 | 0.27220216 | 0.29797069 | -0.0881317 | -0.47887788 | -0.50290 |
| [7,] | -0.2392107 | 0.80935 | -0.21578206 | -0.36493277 | -0.2046650 | -0.16771812 | 0.10605 |
| [8,] | -0.2873922 | 0.03046 | -0.04622550 | 0.22783648 | 0.061606975 | 0.06600716 | 0.42267 |
| [9,] | -0.2722740 | 0.01005 | -0.45897990 | 0.45448138 | 0.249387903 | -0.12097345 | 0.20676 |
| [10,] | -0.2809547 | 0.06264 | 0.01693636 | -0.16724094 | 0.479193670 | 0.54908442 | -0.09397 |
| [11,] | -0.2760404 | -0.12920 | 0.44850168 | -0.09026398 | 0.058865021 | -0.34338568 | 0.55376 |
| [12,] | -0.2685944 | -0.30889 | -0.14021126 | -0.55482875 | 0.446172877 | -0.35396613 | -0.15150 |
| [13,] | -0.2849085 | -0.14069 | -0.22023101 | 0.09797082 | -0.186395 | -0.02539437 | 0.03207 |

| | [,8] | [,9] | [,10] | [,11] | [,12] | [,13] |
|-------|--------------|--------------|-------------|-------------|-------------|-------------|
| [1,] | 0.297315032 | 0.240150017 | -0.43697200 | 0.50257348 | 0.27137750 | -0.27750353 |
| [2,] | 0.132293694 | -0.063584409 | 0.15213781 | -0.05165680 | -0.67046944 | -0.56360913 |
| [3,] | -0.207822327 | 0.232274031 | -0.24513857 | 0.13851854 | -0.44792023 | 0.64242019 |
| [4,] | 0.187666165 | -0.574651648 | 0.12226598 | 0.22273602 | 0.21065313 | 0.21344009 |
| [5,] | 0.133264809 | 0.037056872 | -0.09983006 | -0.37098120 | 0.03819858 | 0.16295672 |
| [6,] | -0.001364215 | 0.093574875 | 0.20717450 | -0.42345548 | 0.17858991 | -0.07564526 |
| [7,] | -0.043634089 | -0.008994677 | 0.02900557 | -0.11842473 | -0.09797524 | -0.01656229 |
| [8,] | 0.356670901 | 0.450797707 | 0.52751979 | 0.14633825 | 0.18286492 | 0.10982676 |
| [9,] | 0.185986795 | -0.519670939 | -0.23013681 | -0.16238502 | 0.06365000 | 0.03063698 |
| [10,] | -0.251132130 | 0.159550225 | -0.08617857 | -0.39763382 | 0.30543084 | -0.03977805 |
| [11,] | -0.155530382 | 0.133824755 | -0.43933255 | -0.02814076 | 0.01132040 | -0.18367121 |
| [12,] | 0.105680226 | -0.101374244 | 0.21311454 | 0.13814260 | -0.18498968 | 0.19355590 |
| [13,] | -0.731981940 | -0.130308646 | 0.28291573 | 0.34527530 | 0.15406762 | -0.16953737 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Por lo tanto, cuando se obtienen los valores de las componentes, claramente se distingue que la primera tiene la máxima variabilidad al mostrar la desviación estándar más significativa 3.41242, la segunda componente con 0.78, en tanto que la componente trece solamente registra 0.098 como el valor menos importante (Cuadro 5.10).

Desviación estándar para cada componente principal

Cuadro 5. 10

| <i>Comp.1</i> | <i>Comp.2</i> | <i>Comp.3</i> | <i>Comp.4</i> | <i>Comp.5</i> | <i>Comp.6</i> | <i>Comp.7</i> |
|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 3.41242 | 0.78135 | 0.57994 | 0.51183 | 0.41936 | 0.38187 | 0.30166 |

| <i>Comp.8</i> | <i>Comp.9</i> | <i>Comp.10</i> | <i>Comp.11</i> | <i>Comp.12</i> | <i>Comp.13</i> |
|---------------|---------------|----------------|----------------|----------------|----------------|
| 0.26520 | 0.22756 | 0.18990 | 0.18153 | 0.11456 | 0.09789 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En el cuadro que se presenta a continuación, cuando se analiza la proporción acumulada de la varianza, claramente se ve que la primera componente explica el 86.38% de la variabilidad total, que es una cantidad suficiente para explicar correctamente al conjunto de 13 variables de alimentos que integran al modelo. Al acumular la primera y la segunda componente ambas ya explican el 90.91% de la variabilidad total; se puede decir que las demás componentes tienen poca aportación por lo que se desecharían (Cuadro 5.11).

Acumulación de la varianza por componente principal

Cuadro 5. 11

Importancia de componentes

| | <i>Comp.1</i> | <i>Comp.2</i> | <i>Comp.3</i> | <i>Comp.4</i> | <i>Comp.5</i> |
|--------------------|----------------|----------------|----------------|---------------|----------------|
| Standard deviation | 3.4124273 | 0.78135134 | 0.57994007 | 0.51183133 | 0.41936756 |
| Proportion of Var. | 0.8638454 | 0.04528996 | 0.02495028 | 0.01943403 | 0.01304665 |
| Cumulative Prop. | 0.8638454 | 0.90913537 | 0.93408565 | 0.95351969 | 0.96656633 |
| | <i>Comp.6</i> | <i>Comp.7</i> | <i>Comp.8</i> | <i>Comp.9</i> | <i>Comp.10</i> |
| Standard deviation | 0.38187875 | 0.301667964 | 0.265209734 | 0.227564759 | 0.189909610 |
| Proportion of Var. | 0.01081833 | 0.006750992 | 0.005217809 | 0.003841663 | 0.002675489 |
| Cumulative Prop. | 0.97738466 | 0.984135655 | 0.989353464 | 0.993195127 | 0.995870615 |
| | <i>Comp.11</i> | <i>Comp.12</i> | <i>Comp.13</i> | | |
| Standard deviation | 0.181538937 | 0.1145600034 | 0.097897095 | | |
| Proportion of Var. | 0.002444831 | 0.0009735881 | 0.000710966 | | |
| Cumulative Prop. | 0.998315446 | 0.9992890340 | 1.000000000 | | |

Fuente: INEGI. Encuesta nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Al analizar las correlaciones existentes entre las componentes principales (C) encontradas con las variables originales (X) (Cuadro 5.12), se distingue que en el caso de la primera componente (dado que es la que concentra la mayor variabilidad del modelo, presenta dos clases de correlación con las variables alimenticias originales, por un lado está altamente correlacionada positivamente con las variables originales X11 (Tubérculos) con poco más de 1, con X8 (Huevo) en 0.98, X12 (Café) con 0.89 y X9 (Azúcar) con 0.79 y en segundo lugar por la correlación en forma negativa con X10 (Aceites) con -1.44, X7 (pescado) con -1.25, X1 (Carnes) con -0.95 y con X13 (Especias) en -0.91. El resto de las componentes casi no tienen correlación significativa con las variables originales.

**Correlaciones entre las componentes principales(C)
y las variables originales(X)**

Cuadro 5. 12

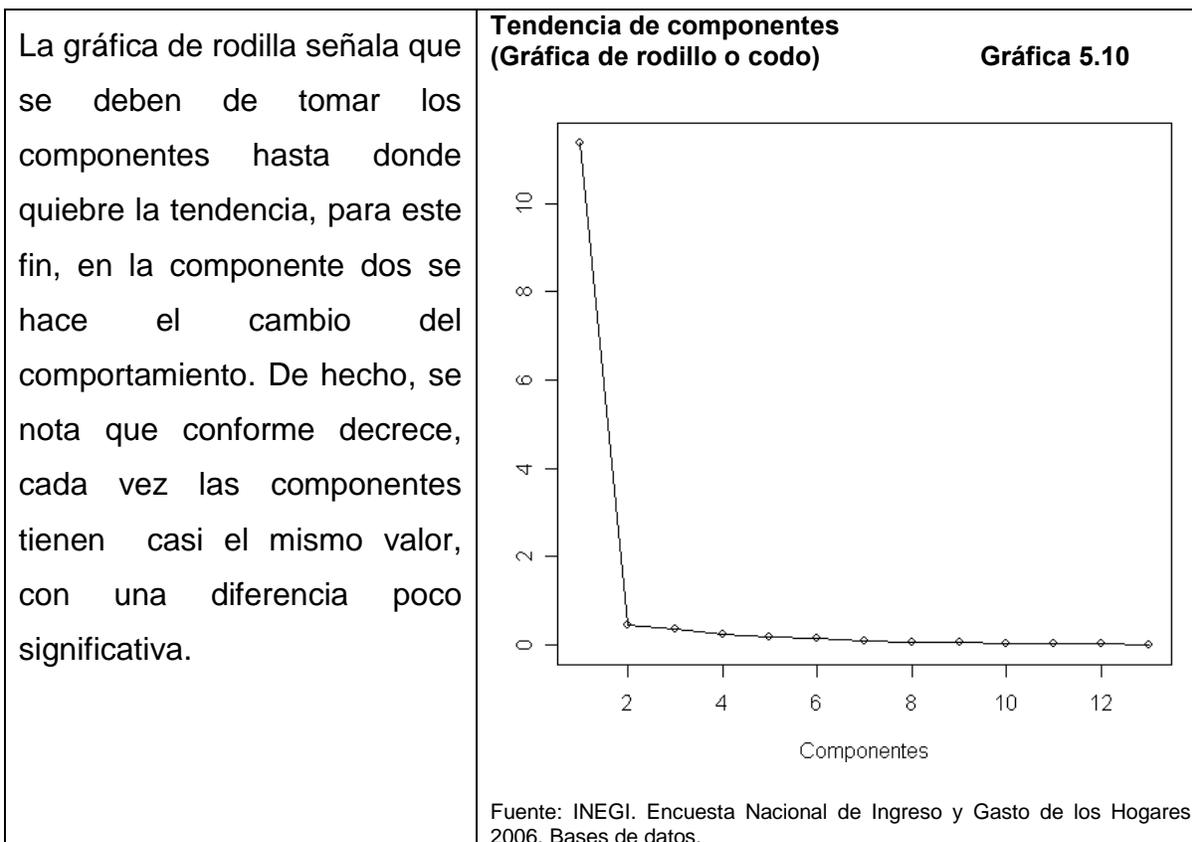
| Comp\Var.Orig. | X1 | X2 | X3 | X4 | X5 | X6 | X7 |
|----------------|----------|----------|----------|----------|----------|----------|----------|
| [C1.] | -0.95831 | 0.11142 | -0.53029 | -0.08711 | -0.22564 | 0.10034 | -1.25286 |
| [C2.] | -0.21162 | -0.04076 | 0.07083 | 0.06553 | -0.00512 | 0.19976 | -0.03918 |
| [C3.] | -0.16925 | 0.04689 | 0.03974 | 0.16682 | -0.00189 | 0.05778 | -0.08887 |
| [C4.] | -0.14199 | 0.05389 | 0.28168 | -0.04755 | -0.07626 | 0.13393 | -0.01330 |
| [C5.] | -0.10717 | -0.17361 | -0.09933 | -0.10407 | -0.25168 | 0.06286 | 0.02873 |
| [C6.] | -0.10427 | 0.00760 | 0.10157 | 0.11118 | -0.03288 | -0.17868 | -0.18765 |
| [C7.] | -0.05761 | 0.19493 | -0.05197 | -0.08789 | -0.04929 | -0.04039 | 0.02554 |
| [C8.] | -0.07233 | 0.00767 | -0.01163 | 0.05734 | 0.01551 | 0.01661 | 0.10638 |
| [C9.] | -0.05854 | 0.00216 | -0.09868 | 0.09772 | 0.05362 | -0.02601 | 0.04446 |
| [C10.] | -0.05114 | 0.01140 | 0.00308 | -0.03044 | 0.08722 | 0.09995 | -0.01710 |
| [C11.] | -0.05155 | -0.02413 | 0.08375 | -0.01686 | 0.01099 | -0.06412 | 0.10341 |
| [C12.] | -0.03003 | -0.03453 | -0.01567 | -0.06203 | 0.04988 | -0.03957 | -0.01694 |
| [C13.] | -0.02784 | -0.01375 | -0.02152 | 0.00957 | -0.01822 | -0.00248 | 0.00314 |

| Comp\Var.Orig. | X8 | X9 | X10 | X11 | X12 | X13 |
|----------------|----------|----------|----------|----------|----------|----------|
| [C1.] | 0.98486 | 0.79550 | -1.44747 | 1.66478 | 0.89894 | -0.91923 |
| [C2.] | 0.09578 | -0.04604 | 0.11015 | -0.03740 | -0.48544 | -0.40807 |
| [C3.] | -0.12119 | 0.13545 | -0.14295 | 0.08078 | -0.26120 | 0.37462 |
| [C4.] | 0.09713 | -0.29742 | 0.06328 | 0.11528 | 0.10903 | 0.11047 |
| [C5.] | 0.05364 | 0.01491 | -0.04018 | -0.14931 | 0.01537 | 0.06559 |
| [C6.] | -0.00051 | 0.03492 | 0.07730 | -0.15800 | 0.06664 | -0.02823 |
| [C7.] | -0.01051 | -0.00217 | 0.00699 | -0.02852 | -0.02360 | -0.00399 |
| [C8.] | 0.08977 | 0.11346 | 0.13276 | 0.03683 | 0.04602 | 0.02764 |
| [C9.] | 0.03999 | -0.11173 | -0.04948 | -0.03491 | 0.01369 | 0.00659 |
| [C10.] | -0.04571 | 0.02904 | -0.01569 | -0.07238 | 0.05560 | -0.00724 |
| [C11.] | -0.02904 | 0.02499 | -0.08204 | -0.00525 | 0.00211 | -0.03430 |
| [C12.] | 0.01181 | -0.01133 | 0.02383 | 0.01544 | -0.02068 | 0.02164 |
| [C13.] | -0.07153 | -0.01273 | 0.02765 | 0.03374 | 0.01506 | -0.01657 |

Fuente: INEGI. Encuesta nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En resumen, la primera componente se retendría debido a lo siguiente: 1) la primera componente presenta un valor propio superior a la unidad al registrar casi 11.4 (Cuadro 5.8). 2) el primer vector propio, registra todos los valores negativos y no superan el 0.29 (Cuadro 5.9); 3) la desviación estándar de la primera componente es la más significativa 3.14 superior a la unidad. (Cuadro 5.10); 4) la proporción de varianza acumulada de la primera componente es de casi 86.4% (Cuadro 5.11), con la segunda se acumula 90.9% y explican casi la totalidad de la variabilidad (ya sea la primera o las dos primeras). Es suficiente trabajar con una componente principal. Asimismo, señalar que hay una correlación alta en forma positiva entre la primera componente con las variables originales (Cuadro 5.12) Tubérculos, el Huevo y el Café, y una correlación negativa con aceites, pescado y carnes, esto porque posiblemente los primeros son alimentos blandos y de bajo costo, y el segundo grupo son ricos en grasas y de mayor costo de adquisición.

Gráfica de rodilla o codo.



Puntuaciones

Al revisar las puntuaciones (scores) de las entidades, en el caso de la componente 1 se distingue que el Estado de México fue la entidad que registró el valor más alto negativamente con -7.19, le siguió el Distrito Federal con -6.7.

En la parte positiva, Baja California Sur aportó la cifra más alta con 9.66 y Quintana Roo con 4.15.

Puntuaciones (scores) de las componentes principales 1 y 2 según entidades federativas **Cuadro 5. 13**
Parte 1

| Entidad | Comp.1 | Comp.2 |
|---------------------|------------|-------------|
| Aguascalientes | 2.9754354 | -1.16514224 |
| Baja California | 0.7486711 | 0.71402148 |
| Baja California Sur | 9.6570193 | 1.66711025 |
| Campeche | 3.8054302 | -0.39819779 |
| Chiapas | -0.1141333 | -0.56973657 |
| Chihuahua | 4.1376485 | 0.33720891 |
| Coahuila | -2.4265165 | 0.66655728 |
| Colima | -1.5981978 | -0.86041686 |
| Distrito Federal | -6.7049051 | 0.45190169 |
| Durango | 1.3361562 | -0.37149891 |
| Guerrero | -2.0439057 | -0.38879370 |

En el componente 2, el dato más significativo en forma negativa correspondió a Yucatán con -1.99 siguiéndole Zacatecas con -1.71; en tanto que positivamente, destacaron Baja California Sur con 1.66 y Guanajuato con 1.55.

Puntuaciones (scores) de las componentes principales 1 y 2 según entidades federativas Cuadro 5. 13 Parte 2

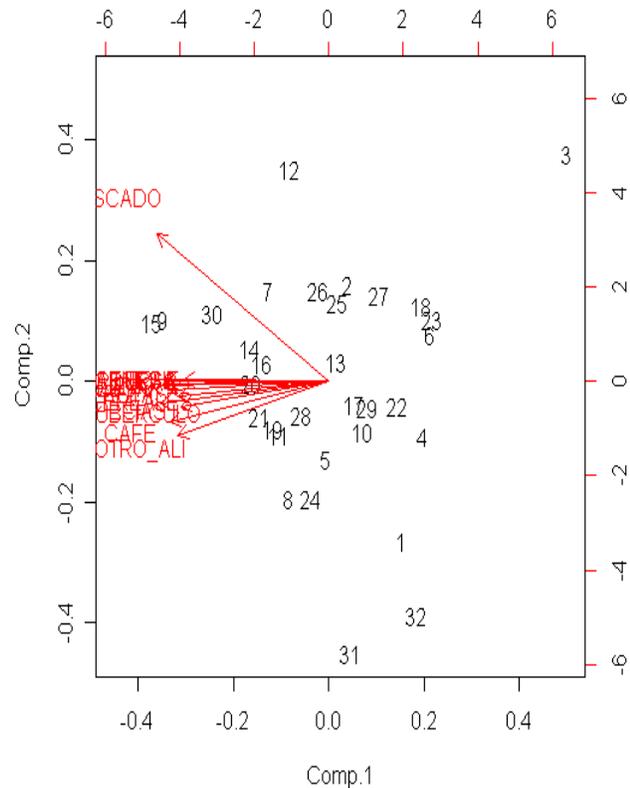
| Entidad | Comp.1 | Comp.2 |
|--------------|------------|-------------|
| Guanajuato | -1.6188800 | 1.55345416 |
| Hidalgo | 0.2720756 | 0.14489090 |
| Jalisco | -3.2067920 | 0.24268572 |
| México | -7.1904389 | 0.42862720 |
| Michoacán | -2.7401868 | 0.12871969 |
| Morelos | 0.9973397 | -0.16874887 |
| Nayarit | 0.7264955 | 0.55714314 |
| Nuevo León | -2.2699764 | -0.34895725 |
| Oaxaca | -3.1597381 | -0.01372230 |
| Puebla | -2.8061479 | -0.25737230 |
| Querétaro | 2.7728173 | -0.17731834 |
| Quintana Roo | 4.1474800 | 0.45571179 |
| San Luis | -0.7238511 | -0.85705458 |
| Sinaloa | 0.3465817 | 0.57268577 |
| Sonora | -0.4467965 | 0.66811393 |
| Tabasco | 2.0281191 | 0.63122213 |
| Tampico | -1.1232815 | -0.24161268 |
| Tlaxcala | 1.5582327 | -0.19181926 |
| Veracruz | -4.7169177 | 0.49627499 |
| Yucatán | 0.8645859 | -1.99337750 |
| Zacatecas | 3.5165770 | -1.71255990 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En la gráfica de cargas de los componentes principales se observó que hay un elemento, pescado, que se distinguió al salir del grupo, al referenciarse hacia la parte negativa del componente 1 y hacia la parte positiva del componente 2. Guanajuato (12) estuvo más afectada por la parte positiva del componente 2 y poco por la parte negativa del componente 1, mientras que en el sentido opuesto del componente 1 se encuentra Baja California Sur (3) debido a que es una característica de producción y consumo.

En tanto que, el estado de México (15) y el Distrito Federal (9), estuvieron muy cercanos al conjunto de conceptos alimenticios, debido a que se caracterizan por registrar los más altos volúmenes de gasto, y en especial de hogares con al menos un residente de 65 y más años de edad. Por su parte, Yucatán (31) y Zacatecas (32) se encontraron ubicados en la parte inferior debido al efecto (negativo) más importante producido por el componente 2.

Concentración del gasto de las entidades según componentes 1 y 2
Gráfica 5.11



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Modelo con la primera componente

Así la componente principal (dado que todos tienen el mismo **signo** negativo, se puede escribir en forma positiva) y quedaría la siguiente combinación lineal:

En término transformado (logaritmos):

$$Z_1 = 0.2893 \text{ carnes} + 0.2923 \text{ cereales} + 0.2902 \text{ verduras} + 0.2743 \text{ leche} + 0.2663 \text{ otro_alim} + 0.2794 \text{ frutas} + 0.2392 \text{ pescado} + 0.2874 \text{ huevo} + 0.2723 \text{ aceites} + 0.2810 \text{ azúcar} + 0.2760 \text{ tubérculo} + 0.2686 \text{ café} + 0.12849 \text{ especias}$$

En término real o de origen:

$$Z_1' = 0.7488 \text{ carnes} + 0.7466 \text{ cereales} + 0.7481 \text{ verduras} + 0.7601 \text{ leche} + 0.7662 \text{ otro_alim} + 0.7562 \text{ frutas} + 0.7872 \text{ pescado} + 0.7502 \text{ huevo} + 0.7616 \text{ aceites} + 0.7551 \text{ azúcar} + 0.7588 \text{ tubérculo} + 0.7645 \text{ café} + 0.7521 \text{ especias}$$

La interpretación del modelo es el siguiente: dado que casi todos los factores son parecidos en forma positiva y casi con el mismo valor, conforme tenga más valor la componente, la influencia de la mayoría de los integrantes será en casi la misma proporción. La primera Componente Principal es una medida del tamaño del fenómeno que se está estudiando.

**Puntuaciones (scores) de las entidades federativas
en los 13 componentes principales**

Cuadro 5. 14

Parte 1

| Clave | Nombre Entidad | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 |
|-------|---------------------|---------|---------|---------|---------|---------|---------|---------|
| 01 | Aguascalientes | 2.9754 | -1.1651 | 0.1755 | -0.2411 | -0.1513 | -0.5025 | -0.0471 |
| 02 | Baja California | 0.7487 | 0.7140 | 0.7358 | -0.3128 | -0.2730 | 0.4965 | -0.0521 |
| 03 | Baja California Sur | 9.6570 | 1.6671 | 0.6460 | -0.6485 | 0.1484 | 0.3245 | -0.1676 |
| 04 | Campeche | 3.8054 | -0.3982 | -0.8955 | 0.4659 | -0.0808 | 0.8960 | -0.1815 |
| 05 | Chiapas | -0.1141 | -0.5697 | -0.2001 | 0.5238 | 0.2533 | 0.3924 | 0.5549 |
| 06 | Chihuahua | 4.1376 | 0.3372 | 0.0530 | 0.4947 | -0.1332 | -0.8641 | -0.2333 |
| 07 | Coahuila | -2.4265 | 0.6666 | -1.3436 | -0.1849 | -0.0071 | -0.2408 | -0.0856 |
| 08 | Colima | -1.5982 | -0.8604 | 0.5700 | 0.2230 | 0.4704 | -0.0432 | 0.2246 |
| 09 | Distrito Federal | -6.7049 | 0.4519 | 0.4852 | -0.0004 | -0.7291 | 0.7409 | -0.3708 |
| 10 | Durango | 1.3362 | -0.3715 | 0.2863 | 0.3943 | 0.3345 | 0.1187 | 0.3949 |
| 11 | Guerro | -2.0439 | -0.3888 | 0.3851 | -0.5241 | -0.4513 | -0.4600 | 0.0223 |
| 12 | Guanajuato | -1.6189 | 1.5535 | 0.0215 | -0.2387 | 0.2425 | -0.9188 | 0.0726 |
| 13 | Hidalgo | 0.2721 | 0.1449 | 0.0120 | -0.4688 | 0.4842 | 0.1833 | -0.1636 |
| 14 | Jalisco | -3.2068 | 0.2427 | -0.1271 | 0.0154 | 0.2599 | 0.3551 | 0.1561 |
| 15 | México | -7.1904 | 0.4286 | 0.2620 | -0.1583 | -0.2864 | 0.0545 | -0.0237 |
| 16 | Michoacán | -2.7402 | 0.1287 | 0.9654 | -0.2504 | 0.3571 | 0.0528 | -0.0282 |
| 17 | Morelos | 0.9973 | -0.1687 | 0.3185 | 0.0828 | 0.5753 | 0.2044 | -0.7970 |
| 18 | Nayarit | 0.7265 | 0.5571 | -0.2234 | 0.9490 | -0.9007 | -0.0228 | 0.1419 |
| 19 | Nuev León | -2.2700 | -0.3490 | 0.4197 | 0.5644 | -0.3196 | -0.3314 | -0.0993 |
| 20 | Oaxaca | -3.1597 | -0.0137 | -0.4624 | -0.6106 | -0.0267 | 0.0996 | -0.0419 |
| 21 | Puebla | -2.8061 | -0.2574 | 0.1003 | -0.3153 | 0.5164 | -0.1730 | -0.2406 |
| 22 | Querétaro | 2.7728 | -0.1773 | 1.0777 | 0.1480 | -0.8783 | -0.1087 | -0.0227 |
| 23 | Quintana Roo | 4.1475 | 0.4557 | -0.4969 | -0.1485 | -0.2843 | -0.2350 | -0.0700 |
| 24 | San Luis | -0.7239 | -0.8571 | 0.1947 | 0.2050 | 0.0456 | 0.1160 | 0.1350 |
| 25 | Sinaloa | 0.3466 | 0.5727 | 0.1989 | 0.4937 | -0.0285 | -0.0478 | 0.6340 |
| 26 | Sonora | -0.4468 | 0.6681 | 0.0940 | 0.4857 | 0.6276 | 0.2137 | 0.2545 |
| 27 | Tabasco | 2.0281 | 0.6312 | -0.8938 | 0.3831 | 0.5508 | 0.1733 | -0.3288 |
| 28 | Tampico | -1.1233 | -0.2416 | -0.1763 | 0.8307 | 0.1611 | -0.1414 | 0.1898 |
| 29 | Tlaxcala | 1.5582 | -0.1918 | -0.8790 | -1.6792 | -0.3518 | 0.1586 | 0.6328 |
| 30 | Veracruz | -4.7169 | 0.4963 | -0.5436 | 0.0869 | 0.3232 | -0.3171 | -0.0980 |
| 31 | Yucatán | 0.8646 | -1.9934 | -0.9889 | -0.0545 | -0.4439 | -0.1290 | -0.5269 |
| 32 | Zacatecas | 3.5166 | -1.7126 | 0.6267 | -0.5103 | 0.5154 | -0.0446 | 0.1652 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Puntuaciones de las entidades federativas en los 13 componentes principales Cuadro 5.14

Parte 2

| Clave | Nombre Entidad | Comp.8 | Comp.9 | Comp.10 | Comp.11 | Comp.12 | Comp.13 |
|-------|---------------------|---------|---------|---------|---------|---------|---------|
| 01 | Aguascalientes | 0.4017 | 0.3294 | -0.1185 | -0.1610 | 0.0475 | 0.1103 |
| 02 | Baja California | -0.3926 | -0.1895 | 0.1679 | 0.0801 | -0.0440 | 0.1138 |
| 03 | Baja California Sur | -0.0614 | 0.4625 | 0.0185 | -0.1665 | -0.1049 | -0.0551 |
| 04 | Campeche | -0.0739 | -0.1763 | -0.2241 | 0.0334 | 0.0733 | -0.1856 |
| 05 | Chiapas | 0.0598 | 0.0604 | -0.1632 | -0.1816 | -0.0152 | 0.0771 |
| 06 | Chihuahua | -0.1893 | -0.1194 | 0.0624 | -0.0903 | -0.0356 | -0.0416 |
| 07 | Coahuila | -0.1240 | -0.1649 | 0.0999 | -0.2404 | 0.0418 | -0.0424 |
| 08 | Colima | -0.0700 | 0.2668 | 0.0185 | 0.1967 | -0.0144 | 0.0287 |
| 09 | Distrito Federal | 0.0495 | 0.0551 | -0.0630 | 0.1101 | -0.1078 | 0.0256 |
| 10 | Durango | 0.2309 | -0.0637 | 0.2782 | -0.0538 | 0.0335 | 0.0523 |
| 11 | Guerro | -0.0993 | 0.0086 | -0.3685 | -0.0679 | -0.0028 | -0.0731 |
| 12 | Guanajuato | -0.1006 | 0.0317 | 0.0014 | 0.2019 | 0.0378 | -0.0331 |
| 13 | Hidalgo | -0.1819 | 0.0085 | 0.0761 | -0.0280 | 0.0328 | 0.1005 |
| 14 | Jalisco | -0.0779 | 0.4590 | 0.1214 | 0.2551 | 0.2234 | 0.1087 |
| 15 | México | 0.1701 | 0.0254 | 0.1138 | 0.0983 | -0.1493 | -0.0849 |
| 16 | Michoacán | 0.2777 | 0.0971 | -0.4425 | -0.1844 | 0.0853 | -0.0801 |
| 17 | Morelos | -0.1431 | -0.0784 | 0.3090 | -0.1493 | 0.1832 | 0.0445 |
| 18 | Nayarit | 0.2172 | 0.1508 | 0.2032 | -0.1690 | 0.0842 | -0.0071 |
| 19 | Nuev León | -0.6536 | -0.2350 | 0.0300 | -0.0615 | -0.1719 | 0.0919 |
| 20 | Oaxaca | 0.3444 | -0.2679 | 0.1256 | -0.4381 | -0.0612 | 0.0788 |
| 21 | Puebla | 0.3249 | -0.2839 | 0.1563 | 0.2480 | 0.2612 | -0.0892 |
| 22 | Querétaro | 0.2599 | -0.2950 | -0.0339 | 0.0735 | 0.1077 | -0.1598 |
| 23 | Quintana Roo | 0.4140 | -0.3595 | 0.2126 | 0.3401 | -0.0337 | 0.2937 |
| 24 | San Luis | -0.3606 | -0.1137 | -0.2854 | -0.2748 | 0.1197 | 0.1206 |
| 25 | Sinaloa | -0.1508 | 0.2087 | 0.0263 | 0.1704 | 0.0869 | 0.0096 |
| 26 | Sonora | 0.5929 | -0.1656 | 0.1238 | -0.0126 | -0.2609 | -0.0406 |
| 27 | Tabasco | 0.1336 | -0.0413 | -0.4337 | 0.2629 | -0.0683 | -0.0178 |
| 28 | Tampico | -0.2275 | -0.0684 | 0.1195 | -0.0115 | -0.0064 | -0.1490 |
| 29 | Tlaxcala | -0.2368 | -0.1154 | 0.0987 | 0.0213 | 0.0208 | -0.0895 |
| 30 | Veracruz | -0.0337 | 0.4670 | -0.0683 | -0.1473 | -0.0521 | -0.0162 |
| 31 | Yucatán | 0.1507 | 0.3753 | 0.1534 | 0.1547 | -0.1452 | -0.0267 |
| 32 | Zacatecas | -0.1833 | -0.0842 | 0.1094 | 0.1915 | -0.1654 | -0.0643 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En la Componente 1 se observaron las cifras más notorias por señalar, la entidad menos poblada Baja California Sur registró el dato más importante positivamente 9.66, en tanto que la más habitada, el Estado de México, con la cifra negativa más alta -7.2. En las siguientes componentes, los valores que presentaron las

entidades están entre 2 y -2. Por señalar la componente 2 donde sobresalieron (3) Baja California Sur y (31) Yucatán; y para el componente 3 destacaron (7) Coahuila y (22) Querétaro.

Se observó que aquellas entidades que asignan más recursos (15) Estado de México y (9) Distrito Federal o que de acuerdo a su producción (2) Baja California, (31) Yucatán en los gastos de alimentos como carnes, cereales, verduras y leche son las que registraron componentes más altos (influenciadas por su alta variabilidad), con esto podemos afirmar que no hay necesidad de analizar todas las variables dado que los alimentos más adecuados y donde más gasto dirigen los hogares que tienen al menos un adulto mayores son los alimentos ricos en vitaminas y proteínas.

5.2 Hogares con exclusivamente población de 65 y más años de edad

En este apartado se analiza cómo es el comportamiento del gasto alimentario de los hogares que únicamente registran personas de 65 y más años de edad, bajo la misma técnica de componentes principales. Para ello resumiremos todos los puntos antes señalados:

Análisis exploratorio

Al igual que en el ejercicio anterior, se procedió a hacer un análisis breve sobre el comportamiento de la información, para ello se describió la parte univariada y posteriormente la multivariada.

A continuación se presentan los cuadros con los datos correspondientes, en primera instancia a los valores reales sin transformación y en seguida los datos transformados en logaritmos (Cuadro 5.15 y Cuadro 5.16).

Es necesario aclarar que aquellos registros que presentaron un dato nulo o cero unidades, en el momento de transformar a logaritmos se le asignó el valor de uno para continuar con la consistencia del dato ya que un logaritmo de un valor nulo se identifica como valor no determinado

**Gasto de los hogares con exclusivamente residentes de 65 y más años de edad
por tipo de alimento**

Cuadro 5. 15

| ent nom | CEREALES | CARNES | PESCADO | LECHE | HUEVO | ACEITES | TUBERCULO | VERDURAS | FRUTAS | AZUCAR | CAFE | ESPECIAS | OTRO_ALI |
|------------------------|----------|-----------|----------|----------|----------|---------|-----------|----------|----------|----------|---------|----------|----------|
| 01 Aguascalientes | 3016167 | 723749 | 14680 | 2539936 | 268573 | 140791 | 149043 | 1060515 | 504088 | 83623 | 233456 | 12543 | 2599060 |
| 02 Baja California | 7961279 | 26006063 | 5881011 | 20972570 | 6273511 | 106685 | 1952144 | 15448235 | 3947794 | 916640 | 750958 | 0 | 4957178 |
| 03 Baja California Sur | 1437564 | 1452569 | 1166073 | 841977 | 130795 | 0 | 31652 | 776175 | 99558 | 101585 | 32152 | 0 | 32505 |
| 04 Campeche | 6871989 | 8107863 | 1062757 | 2033016 | 630549 | 139356 | 509539 | 3520167 | 1976298 | 1202655 | 1086977 | 211355 | 2919459 |
| 05 Chiapas | 6065004 | 5004853 | 0 | 4525694 | 362093 | 300415 | 539508 | 4243394 | 4699575 | 276590 | 157040 | 104326 | 3477197 |
| 06 Chihuahua | 5017148 | 6375590 | 746809 | 3487048 | 589310 | 436595 | 113874 | 2182852 | 1259064 | 99225 | 258460 | 93207 | 5148447 |
| 07 Coahuila | 14009383 | 27429541 | 1623207 | 4369187 | 2361519 | 1328735 | 1345794 | 10322335 | 3570383 | 1841718 | 2902021 | 195992 | 1476385 |
| 08 Colima | 39681413 | 24216281 | 505726 | 37553798 | 5883973 | 1633778 | 3618825 | 25605713 | 16438568 | 3758420 | 5369405 | 63856 | 4251263 |
| 09 Distrito Federal | 64722453 | 128330208 | 30702393 | 69662863 | 10640532 | 2010120 | 7773247 | 77245394 | 63554038 | 3701588 | 5498332 | 3346440 | 21420467 |
| 10 Durango | 12884505 | 13280259 | 1560092 | 16730419 | 1925192 | 794723 | 878817 | 11592825 | 3963873 | 689400 | 1638935 | 848340 | 4674586 |
| 11 Guerrero | 25793900 | 30792319 | 2843353 | 18697138 | 2895927 | 3224704 | 2239883 | 25233999 | 11170695 | 1674349 | 746085 | 1065638 | 8754012 |
| 12 Guanajuato | 21323188 | 20390689 | 7466702 | 18025214 | 5359606 | 1541977 | 530960 | 21654293 | 1647878 | 2706447 | 819311 | 744156 | 5002327 |
| 13 Hidalgo | 3843833 | 7446253 | 1423128 | 8135973 | 61633 | 546793 | 134269 | 4517500 | 3838265 | 379798 | 1072604 | 164621 | 2824526 |
| 14 Jalisco | 41805711 | 86790996 | 10526060 | 41589804 | 6457578 | 4006934 | 3442944 | 42378730 | 19134563 | 4711139 | 1040441 | 880026 | 15347219 |
| 15 México | 73711355 | 98421677 | 29097796 | 58066405 | 13745279 | 8304506 | 7139308 | 90541703 | 30523359 | 11284546 | 5001592 | 2928636 | 23195718 |
| 16 Michoacán | 47915101 | 43666220 | 2231237 | 27772414 | 1373970 | 2259938 | 6401908 | 43541125 | 16005756 | 4458754 | 1947263 | 728084 | 9269162 |
| 17 Morelos | 16462450 | 16310334 | 450348 | 12566519 | 1979924 | 4180998 | 1039986 | 14175464 | 6916918 | 1002494 | 1287902 | 552256 | 3786648 |
| 18 Nayarit | 10773207 | 3221015 | 1081643 | 7039725 | 557953 | 137075 | 732149 | 3854228 | 2631998 | 405746 | 312573 | 801563 | 4867830 |
| 19 Nuev León | 25422682 | 21990451 | 222791 | 19369849 | 5836483 | 2639307 | 2247425 | 14456498 | 4483077 | 317637 | 2084031 | 87410 | 20317043 |
| 20 Oaxaca | 31210781 | 15079835 | 1184138 | 9608591 | 5268576 | 4432950 | 1517504 | 40383334 | 11721079 | 3984461 | 483797 | 2065408 | 5221373 |
| 21 Puebla | 33384273 | 22653975 | 1155062 | 13797888 | 2618901 | 1831983 | 1340750 | 18355621 | 3098297 | 437035 | 1264000 | 2882956 | 5342720 |
| 22 Querétaro | 2912282 | 1383478 | 9907 | 5090676 | 137865 | 45611 | 34796 | 2596118 | 331656 | 13071 | 11160 | 1134 | 165427 |
| 23 Quintana Roo | 4544116 | 5122208 | 4428838 | 5130326 | 325546 | 921485 | 13623 | 1358530 | 1487194 | 449398 | 626871 | 900003 | 1808764 |
| 24 San Luis | 16020110 | 12942660 | 4192611 | 14156538 | 1346579 | 1564584 | 720302 | 13258200 | 7411452 | 110906 | 58503 | 50621 | 8053786 |
| 25 Sinaloa | 21934810 | 20219741 | 8713257 | 19571125 | 3092112 | 1622798 | 1839042 | 15203901 | 3218295 | 1958935 | 2119487 | 836928 | 4118838 |
| 26 Sonora | 12540614 | 13345231 | 1306677 | 9936756 | 2679260 | 1614348 | 1509327 | 8423867 | 3569290 | 2130097 | 1587538 | 450232 | 4446597 |
| 27 Tabasco | 4798640 | 11106968 | 1692805 | 3529749 | 489142 | 128262 | 156339 | 2793798 | 869496 | 769820 | 596582 | 0 | 1779977 |
| 28 Tampico | 21872476 | 15555315 | 795639 | 17773214 | 2307087 | 2770564 | 2110499 | 11763163 | 6970243 | 2201502 | 2365328 | 1058723 | 15271861 |
| 29 Tlaxcala | 2484291 | 2538726 | 1299934 | 1430530 | 347364 | 154512 | 534235 | 2484869 | 1335338 | 128565 | 96334 | 31936 | 2005776 |
| 30 Veracruz | 47233756 | 54090790 | 3891247 | 24220115 | 3829836 | 4811579 | 2028466 | 28939887 | 7978765 | 8419344 | 2445340 | 923239 | 20243497 |
| 31 Yucatán | 11362371 | 11577721 | 937792 | 7948845 | 1066615 | 206338 | 319977 | 5583052 | 4309251 | 1538115 | 170259 | 720641 | 15826718 |
| 32 Zacatecas | 13086236 | 2748808 | 0 | 5500257 | 1445401 | 599661 | 1402748 | 5920946 | 1846590 | 1687499 | 1057451 | 465791 | 2658057 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Gasto de los hogares con exclusivamente residentes de 65 y más años de edad por tipo de alimento (valores en logaritmos)

Cuadro 5. 16

| ent nom | CEREALES | CARNES | PESCADO | LECHE | HUEVO | ACEITES | TUBERCULO | VERDURAS | FRUTAS | AZUCAR | CAFE | ESPECIAS | OTRO_ALI |
|------------------------|----------|--------|---------|-------|-------|---------|-----------|----------|--------|--------|-------|----------|----------|
| 01 Aguascalientes | 14.92 | 13.49 | 9.59 | 14.75 | 12.50 | 11.86 | 11.91 | 13.87 | 13.13 | 11.33 | 12.36 | 9.44 | 14.77 |
| 02 Baja California | 15.89 | 17.07 | 15.59 | 16.86 | 15.65 | 11.58 | 14.48 | 16.55 | 15.19 | 13.73 | 13.53 | 0.00 | 15.42 |
| 03 Baja California Sur | 14.18 | 14.19 | 13.97 | 13.64 | 11.78 | 0.00 | 10.36 | 13.56 | 11.51 | 11.53 | 10.38 | 0.00 | 10.39 |
| 04 Campeche | 15.74 | 15.91 | 13.88 | 14.53 | 13.35 | 11.84 | 13.14 | 15.07 | 14.50 | 14.00 | 13.90 | 12.26 | 14.89 |
| 05 Chiapas | 15.62 | 15.43 | 0.00 | 15.33 | 12.80 | 12.61 | 13.20 | 15.26 | 15.36 | 12.53 | 11.96 | 11.56 | 15.06 |
| 06 Chihuahua | 15.43 | 15.67 | 13.52 | 15.06 | 13.29 | 12.99 | 11.64 | 14.60 | 14.05 | 11.51 | 12.46 | 11.44 | 15.45 |
| 07 Coahuila | 16.46 | 17.13 | 14.30 | 15.29 | 14.67 | 14.10 | 14.11 | 16.15 | 15.09 | 14.43 | 14.88 | 12.19 | 14.21 |
| 08 Colima | 17.50 | 17.00 | 13.13 | 17.44 | 15.59 | 14.31 | 15.10 | 17.06 | 16.62 | 15.14 | 15.50 | 11.06 | 15.26 |
| 09 Distrito Federal | 17.99 | 18.67 | 17.24 | 18.06 | 16.18 | 14.51 | 15.87 | 18.16 | 17.97 | 15.12 | 15.52 | 15.02 | 16.88 |
| 10 Durango | 16.37 | 16.40 | 14.26 | 16.63 | 14.47 | 13.59 | 13.69 | 16.27 | 15.19 | 13.44 | 14.31 | 13.65 | 15.36 |
| 11 Guerrero | 17.07 | 17.24 | 14.86 | 16.74 | 14.88 | 14.99 | 14.62 | 17.04 | 16.23 | 14.33 | 13.52 | 13.88 | 15.99 |
| 12 Guanajuato | 16.88 | 16.83 | 15.83 | 16.71 | 15.49 | 14.25 | 13.18 | 16.89 | 14.31 | 14.81 | 13.62 | 13.52 | 15.43 |
| 13 Hidalgo | 15.16 | 15.82 | 14.17 | 15.91 | 11.03 | 13.21 | 11.81 | 15.32 | 15.16 | 12.85 | 13.89 | 12.01 | 14.85 |
| 14 Jalisco | 17.55 | 18.28 | 16.17 | 17.54 | 15.68 | 15.20 | 15.05 | 17.56 | 16.77 | 15.37 | 13.86 | 13.69 | 16.55 |
| 15 México | 18.12 | 18.40 | 17.19 | 17.88 | 16.44 | 15.93 | 15.78 | 18.32 | 17.23 | 16.24 | 15.43 | 14.89 | 16.96 |
| 16 Michoacán | 17.68 | 17.59 | 14.62 | 17.14 | 14.13 | 14.63 | 15.67 | 17.59 | 16.59 | 15.31 | 14.48 | 13.50 | 16.04 |
| 17 Morelos | 16.62 | 16.61 | 13.02 | 16.35 | 14.50 | 15.25 | 13.85 | 16.47 | 15.75 | 13.82 | 14.07 | 13.22 | 15.15 |
| 18 Nayarit | 16.19 | 14.99 | 13.89 | 15.77 | 13.23 | 11.83 | 13.50 | 15.16 | 14.78 | 12.91 | 12.65 | 13.59 | 15.40 |
| 19 Nuev León | 17.05 | 16.91 | 12.31 | 16.78 | 15.58 | 14.79 | 14.63 | 16.49 | 15.32 | 12.67 | 14.55 | 11.38 | 16.83 |
| 20 Oaxaca | 17.26 | 16.53 | 13.98 | 16.08 | 15.48 | 15.30 | 14.23 | 17.51 | 16.28 | 15.20 | 13.09 | 14.54 | 15.47 |
| 21 Puebla | 17.32 | 16.94 | 13.96 | 16.44 | 14.78 | 14.42 | 14.11 | 16.73 | 14.95 | 12.99 | 14.05 | 14.87 | 15.49 |
| 22 Querétaro | 14.88 | 14.14 | 9.20 | 15.44 | 11.83 | 10.73 | 10.46 | 14.77 | 12.71 | 9.48 | 9.32 | 7.03 | 12.02 |
| 23 Quintana Roo | 15.33 | 15.45 | 15.30 | 15.45 | 12.69 | 13.73 | 9.52 | 14.12 | 14.21 | 13.02 | 13.35 | 13.71 | 14.41 |
| 24 San Luis | 16.59 | 16.38 | 15.25 | 16.47 | 14.11 | 14.26 | 13.49 | 16.40 | 15.82 | 11.62 | 10.98 | 10.83 | 15.90 |
| 25 Sinaloa | 16.90 | 16.82 | 15.98 | 16.79 | 14.94 | 14.30 | 14.42 | 16.54 | 14.98 | 14.49 | 14.57 | 13.64 | 15.23 |
| 26 Sonora | 16.34 | 16.41 | 14.08 | 16.11 | 14.80 | 14.29 | 14.23 | 15.95 | 15.09 | 14.57 | 14.28 | 13.02 | 15.31 |
| 27 Tabasco | 15.38 | 16.22 | 14.34 | 15.08 | 13.10 | 11.76 | 11.96 | 14.84 | 13.68 | 13.55 | 13.30 | 0.00 | 14.39 |
| 28 Tampico | 16.90 | 16.56 | 13.59 | 16.69 | 14.65 | 14.83 | 14.56 | 16.28 | 15.76 | 14.60 | 14.68 | 13.87 | 16.54 |
| 29 Tlaxcala | 14.73 | 14.75 | 14.08 | 14.17 | 12.76 | 11.95 | 13.19 | 14.73 | 14.10 | 11.76 | 11.48 | 10.37 | 14.51 |
| 30 Veracruz | 17.67 | 17.81 | 15.17 | 17.00 | 15.16 | 15.39 | 14.52 | 17.18 | 15.89 | 15.95 | 14.71 | 13.74 | 16.82 |
| 31 Yucatán | 16.25 | 16.26 | 13.75 | 15.89 | 13.88 | 12.24 | 12.68 | 15.54 | 15.28 | 14.25 | 12.05 | 13.49 | 16.58 |
| 32 Zacatecas | 16.39 | 14.83 | 0.00 | 15.52 | 14.18 | 13.30 | 14.15 | 15.59 | 14.43 | 14.34 | 13.87 | 13.05 | 14.79 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Análisis individual (numérico y gráfico)

Al analizar las correspondientes estadísticas de cada una de las variables del gasto alimentario de aquellos hogares con exclusivamente habitantes adultos mayores, se identificó lo siguiente: las carnes, las verduras, los cereales, la leche y

las frutas son los conceptos donde se invirtió más para su consumo al registrar las cifras más significativas, tan solo en carne, el mayor gasto 128 millones registrado en el Distrito Federal son cerca de 177 veces a lo que invierte el estado de Aguascalientes, y que dista mucho del promedio(cerca de 24 millones) que registran las entidades del país (Cuadro 5.15).

Para el caso de las verduras, el estado de México registró el dato más alto con poco más de 90 millones de pesos, en cambio en Baja California Sur, la inversión es inferior al millón de pesos. Más de la mitad de las entidades gastaron en su consumo por debajo de los 12 millones de pesos (Cuadro 5.15).

Principales estadísticas de las trece variables de alimentos **Cuadro 5. 17**

| CEREALES | CARNES | PESCADO | LECHE | HUEVO |
|--------------------------|---------------------------|--------------------|--------------------------|-------------------|
| Min. : 1 437 564 | Min. : 723 749 | Min. : 0 | Min. : 841 977 | Min. : 61633 |
| 1st Qu.: 5 803 040 | 1st Qu.: 6 062 245 | 1st Qu.: 783 432 | 1st Qu.: 4 949 431 | 1st Qu.: 540750 |
| Medn : 13 547 809 | Medn : 14 212 533 | Medn : 1 303 305 | Medn : 11 251 638 | Medn : 1952558 |
| Mean : 20 378 221 | Mean : 23 697 575 | Mean : 4 006 678 | Mean : 15 989 817 | Mean : 2 884 021 |
| 3rd Qu.: 27 148 120 | 3rd Qu.: 24 663 726 | 3rd Qu.: 3 966 588 | 3rd Qu.: 19 420 168 | 3rd Qu.: 4189521 |
| Max. : 73 711 355 | Max. : 128 330 208 | Max. : 30 702 393 | Max. : 69 662 863 | Max. : 13 745 279 |

| ACEITES | TUBERCULO | VERDURAS | FRUTAS | AZUCAR |
|------------------|------------------|--------------------------|--------------------------|-------------------|
| Min. : 0 | Min. : 13623 | Min. : 776175 | Min. : 99558 | Min. : 13071 |
| 1st Qu.: 193382 | 1st Qu.: 462148 | 1st Qu.: 3770713 | 1st Qu.: 1796912 | 1st Qu.: 364257 |
| Median : 1435356 | Median : 1190368 | Median : 11677994 | Median : 3893029 | Median : 1102575 |
| Mean : 1701191 | Mean : 1698403 | Mean : 17794263 | Mean : 7828522 | Mean : 1982534 |
| 3rd Qu.: 2354780 | 3rd Qu.: 2048974 | 3rd Qu.: 22549219 | 3rd Qu.: 7553280 | 3rd Qu.: 2327738 |
| Max. : 8 304 506 | Max. : 7 773 247 | Max. : 90 541 703 | Max. : 63 554 038 | Max. : 11 284 546 |

| CAFE | ESPECIAS | OTRO_ALI |
|------------------|------------------|-------------------|
| Min. : 11160 | Min. : 0 | Min. : 32505 |
| 1st Qu.: 299045 | 1st Qu.: 81521 | 1st Qu.: 2782908 |
| Median : 1048946 | Median : 509023 | Median : 4771208 |
| Mean : 1410068 | Mean : 725502 | Mean : 7227013 |
| 3rd Qu.: 1981455 | 3rd Qu.: 885020 | 3rd Qu.: 8882799 |
| Max. : 5 498 332 | Max. : 3 346 440 | Max. : 23 195 718 |

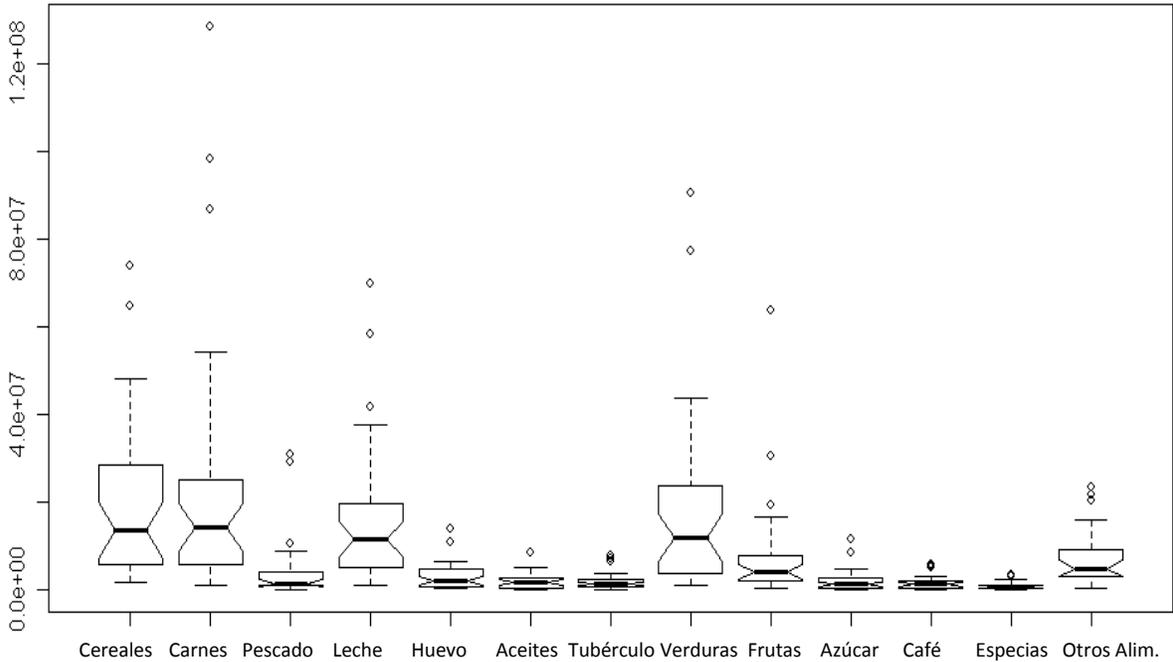
Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En los cereales, la situación es semejante donde el estado de México ubicó el valor máximo y Baja California Sur con el mínimo, sin embargo aquí, el 50% de las entidades presentaron gastos por debajo de los 13 millones de pesos para su consumo (Cuadro 5.15 y Cuadro 5.17).

En contra parte, el café y las especias mostraron los datos más bajos de gasto, y donde destaca el Distrito Federal en ambos conceptos al registrar el mayor consumo, en tanto que Querétaro tuvo el dato más bajo en café, y las Bajas Californias no invirtieron en especias. Así mismo la distancia entre su media y su mediana son cercanas (Cuadro 5.15 y Cuadro 5.17).

Al revisar la distribución del gasto por variable de acuerdo a la siguiente gráfica se observa que: las carnes, los cereales, la leche y las verduras tienen las concentraciones más altas del gasto.

Distribución del gasto según variables del gasto Gráfica 5.12



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

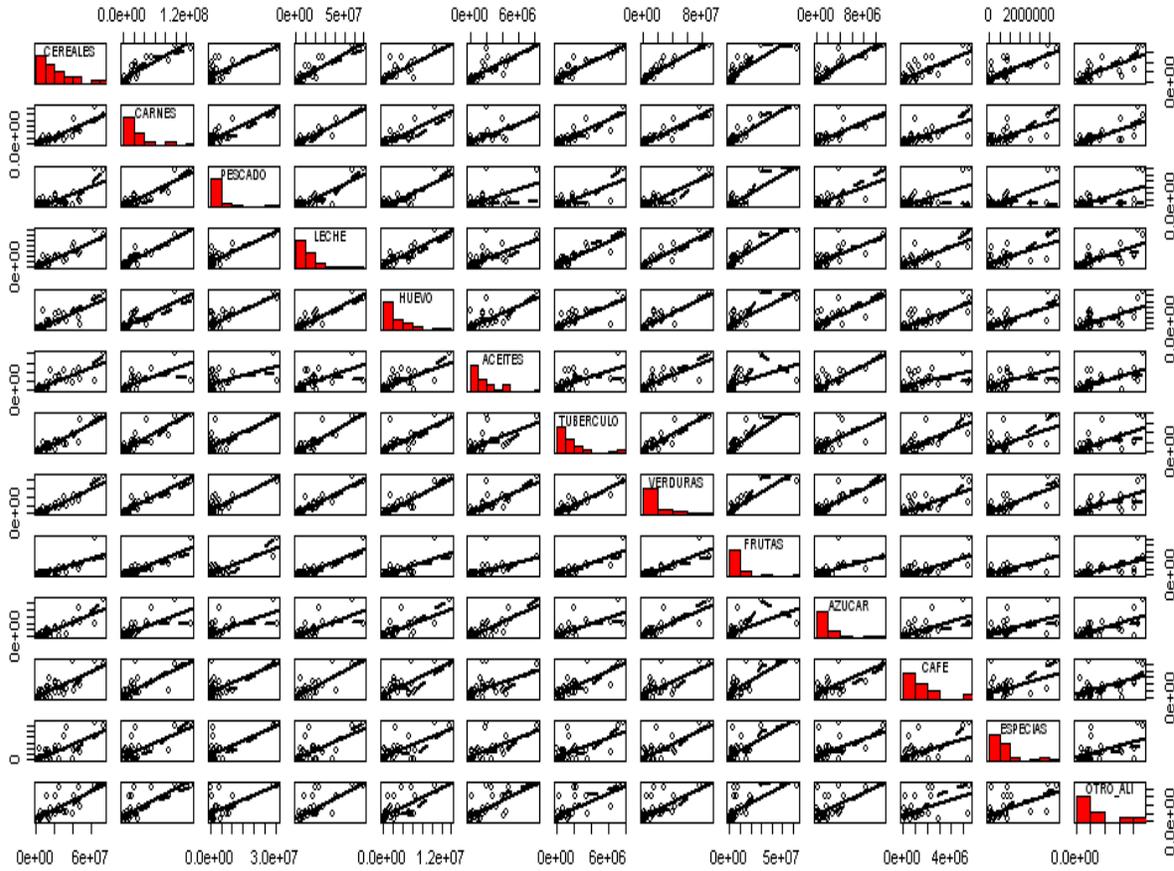
En la parte superior se distinguen las entidades que registran gastos superiores al conjunto de ellas tales son los casos del Distrito Federal y el estado de México y que distan de la mediana en forma considerable. Las colas superiores son más relevantes que las inferiores (Gráfica 5.12 y Cuadro 15).

Análisis multivariado (puntos extremos, detección y tratamiento)

Al realizar el análisis multivariado, e identificar posibles relaciones, se observó en la gráfica que en la mayoría de esas, hay dispersión ya que la mayoría de las observaciones se encuentran fuera de la tendencia lineal, y que se corrobora con la gráfica de histogramas ya que en la parte izquierda hay una gran proporción de observaciones agrupadas en las primeras barras, una menor cantidad de entidades pero con cifras de gasto alto (Gráfica 5.13).

Distribución del gasto de los hogares según variables en forma multivariada

Gráfica 5.13



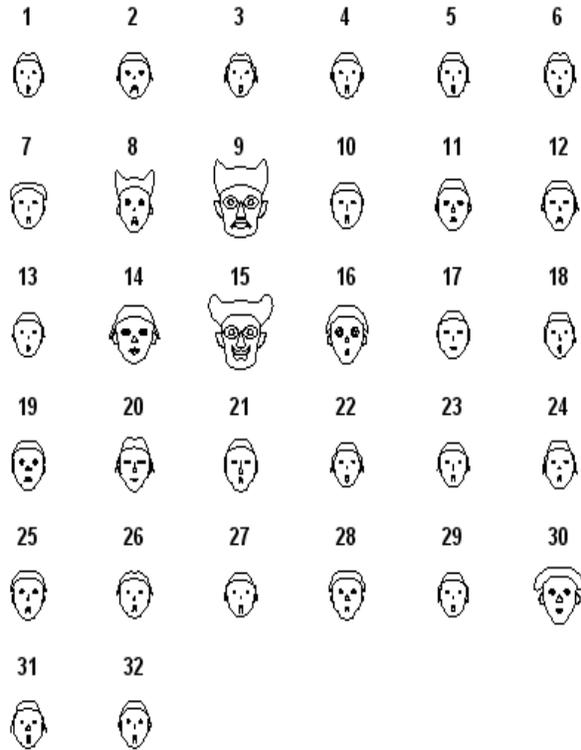
Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En la gráfica de caras, se identifica claramente que el Distrito Federal (9) y el Estado de México (15), el gasto en los trece conceptos de alimentos es fuerte ya que las imágenes son claramente definidas y robustas.

En la descripción, el Distrito Federal y el Estado de México presentan imágenes notorias: caras amplias (influido por cereales), orejas grandes (carnes), altura de la cara grande (pescado), mentón grande (leche), gorro grande (huevo), nariz grande (aceites), posición de boca notoria (tubérculo), curvatura grande de la boca o sonrisa (verduras), longitud amplia de boca (frutas), altura de ojos grandes (azúcar), separación de ojos notoria (café), inclinación de ojos amplia (especias) y forma amplia de los ojos (otros alimentos).

En cambio entidades como Aguascalientes (1), y las dos Baja Californias (2 y 3), las imágenes muestran poco gasto ya que las caras son poco robustas o de poca claridad, por señalar la cara, nariz y los ojos que son pequeñas.

Distribución del gasto de los hogares (caras de Chernoff) Gráfica 5.14



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Al revisar la prueba de normalidad univariada de Shapiro, se identifica (Cuadro 5.18) que todos los componentes muestran indicadores superiores a lo estimado (ver Prueba de Shapiro, glosario) por lo que no son susceptibles de distribuirse en forma normal.

Prueba de Normalidad de Shapiro – Wilks

Cuadro 5. 18

| | | | | |
|----------|-------------|--|-------------|---------------------|
| [1] "1" | "CEREALES" | Shapiro-Wilk normality test data: datos[, i] | W = 0.8549, | p-value = 0.0005304 |
| [1] "2" | "CARNES" | Shapiro-Wilk normality test data: datos[, i] | W = 0.6948, | p-value = 7.269e-07 |
| [1] "3" | "PESCADO" | Shapiro-Wilk normality test data: datos[, i] | W = 0.5354, | p-value = 6.357e-09 |
| [1] "4" | "LECHE" | Shapiro-Wilk normality test data: datos[, i] | W = 0.7905, | p-value = 2.712e-05 |
| [1] "5" | "HUEVO" | Shapiro-Wilk normality test data: datos[, i] | W = 0.7973, | p-value = 3.627e-05 |
| [1] "6" | "ACEITES" | Shapiro-Wilk normality test data: datos[, i] | W = 0.8174, | p-value = 8.81e-05 |
| [1] "7" | "TUBERCULO" | Shapiro-Wilk normality test data: datos[, i] | W = 0.7462, | p-value = 4.611e-06 |
| [1] "8" | "VERDURAS" | Shapiro-Wilk normality test data: datos[, i] | W = 0.7387, | p-value = 3.477e-06 |
| [1] "9" | "FRUTAS" | Shapiro-Wilk normality test data: datos[, i] | W = 0.5758, | p-value = 1.892e-08 |
| [1] "10" | "AZUCAR" | Shapiro-Wilk normality test data: datos[, i] | W = 0.7333, | p-value = 2.85e-06 |
| [1] "11" | "CAFE" | Shapiro-Wilk normality test data: datos[, i] | W = 0.7961, | p-value = 3.446e-05 |
| [1] "12" | "ESPECIAS" | Shapiro-Wilk normality test data: datos[, i] | W = 0.7511, | p-value = 5.558e-06 |
| [1] "13" | "OTRO_ALI" | Shapiro-Wilk normality test data: datos[, i] | W = 0.8028, | p-value = 4.598e-05 |

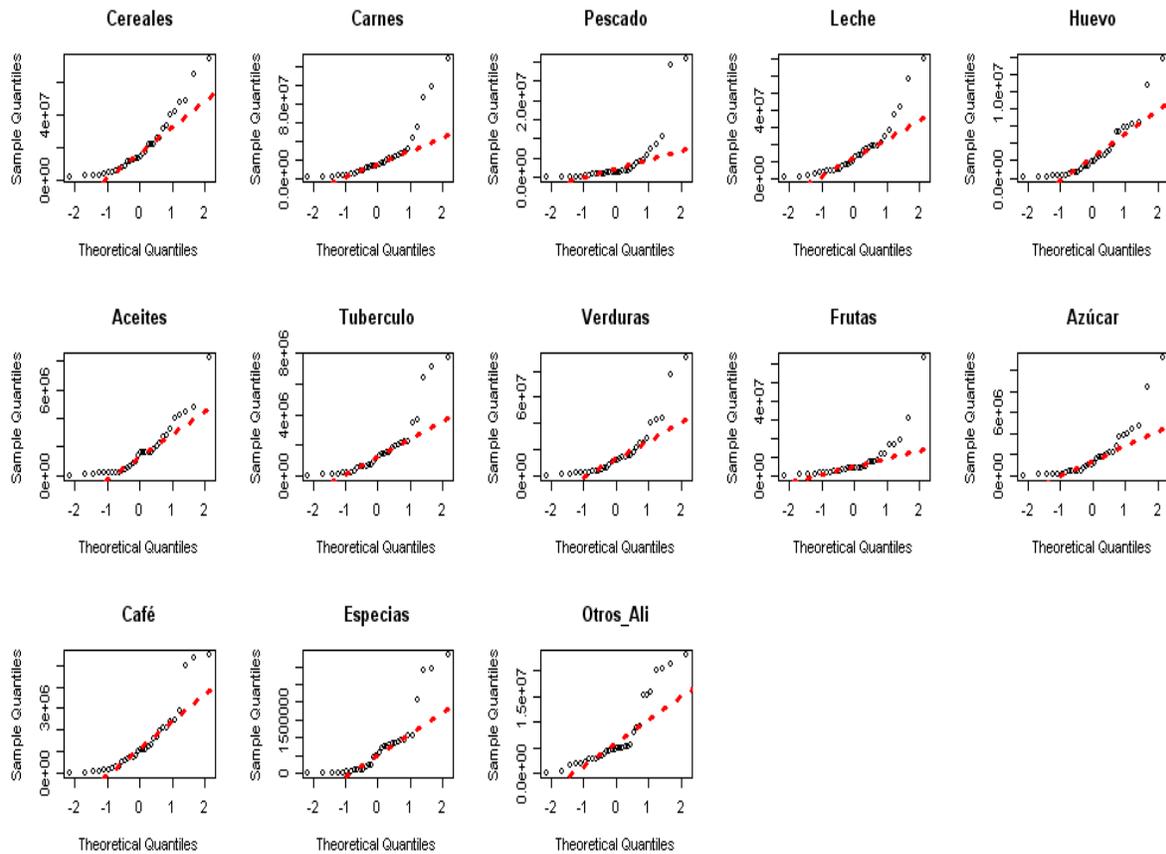
Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Gráfica de cuartiles

Aquí se distinguió como la mayoría de las observaciones en las trece componentes están sobre la tendencia, con excepción de aquellas entidades que registran las cifras más altas, poco menos de 6 observaciones. Lo que se traduce en una posible no existencia de normalidad de las observaciones (Gráfica 5.15).

Distribución del gasto según variables de alimentos (gráfica de cuartiles)

Gráfica 5.15



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

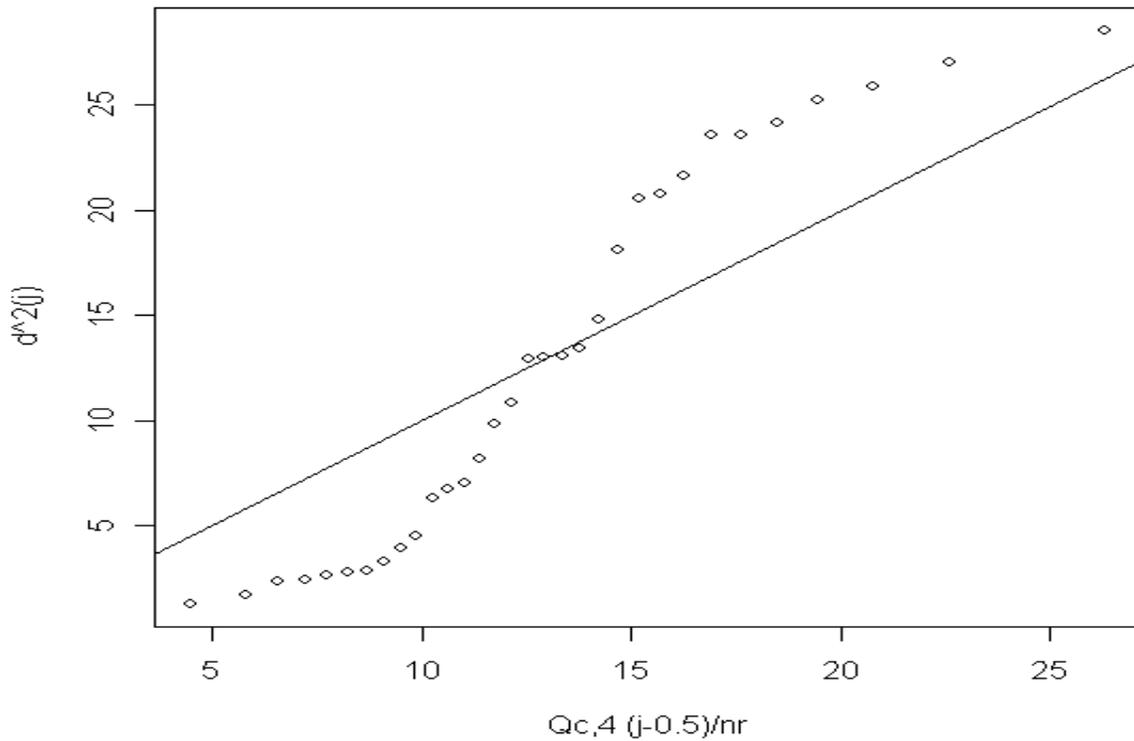
Prueba de normalidad multivariada

Al revisar el contraste de normalidad (entre la Ji cuadrada y la normal) se identificó que desde la parte estricta no existe normalidad ya que son distintos los valores, sin embargo si comparamos los datos, estos son muy cercanos por lo que la existencia de normalidad es latente.

$$1 - pchisq(nr \cdot Ap/6, nc \cdot (nc+1) \cdot (nc+2)/6) \quad [1] \quad 0$$

$$1 - pnorm(Kp, nc \cdot (nc+2), 8 \cdot nc \cdot (nc+2)/nr) \quad [1] \quad 0.1896237$$

Al revisar la gráfica de probabilidad Chi cuadrada, claramente se distinguió que las observaciones no caen en la línea de tendencia por lo que no se tiene normalidad (Gráfica 5.16).



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Datos transformados a logaritmos

Nuevamente se realizó el análisis de la información en datos transformados a logaritmos para ver si existe normalidad de la información y poder así determinar con que información (reales o en logaritmos) utilizar la técnica de componentes principales.

Se observa que carnes, verduras, cereales y leche registraron las estadísticas más importantes, con los registros más altos, así como las medias y las medianas más importantes. De hecho es la misma situación de la versión en absolutos (Cuadro 5.19).

Principales estadísticas (valores en logaritmos)

Cuadro 5. 19

| CEREALES | CARNES | PESCADO | LECHE | HUEVO |
|----------------|----------------|----------------|----------------|----------------|
| Min. : 14.18 | Min. : 13.49 | Min. : 0 | Min. : 13.64 | Min. : 11.03 |
| 1st Qu.: 15.57 | 1st Qu.: 15.61 | 1st Qu.: 13.57 | 1st Qu.: 15.41 | 1st Qu.: 13.20 |
| Median : 16.42 | Median : 16.47 | Median : 14.08 | Median : 16.23 | Median : 14.48 |
| Mean : 16.39 | Mean : 16.33 | Mean : 13.32 | Mean : 16.11 | Mean : 14.18 |
| 3rd Qu.: 17.11 | 3rd Qu.: 17.02 | 3rd Qu.: 15.19 | 3rd Qu.: 16.78 | 3rd Qu.: 15.24 |
| Max. : 18.12 | Max. : 18.67 | Max. : 17.24 | Max. : 18.06 | Max. : 16.44 |

| ACEITES | TUBERCULO | VERDURAS | FRUTAS | AZUCAR |
|----------------|----------------|----------------|----------------|-----------------|
| Min. : 0 | Min. : 9.52 | Min. : 13.56 | Min. : 11.51 | Min. : 9.478 |
| 1st Qu.: 12.16 | 1st Qu.: 13.02 | 1st Qu.: 15.14 | 1st Qu.: 14.40 | 1st Qu.: 12.803 |
| Median : 14.17 | Median : 13.98 | Median : 16.27 | Median : 15.17 | Median : 13.909 |
| Mean : 13.25 | Mean : 13.54 | Mean : 16.05 | Mean : 15.12 | Mean : 13.652 |
| 3rd Qu.: 14.67 | 3rd Qu.: 14.53 | 3rd Qu.: 16.93 | 3rd Qu.: 15.84 | 3rd Qu.: 14.656 |
| Max. : 15.93 | Max. : 15.87 | Max. : 18.32 | Max. : 17.97 | Max. : 16.239 |

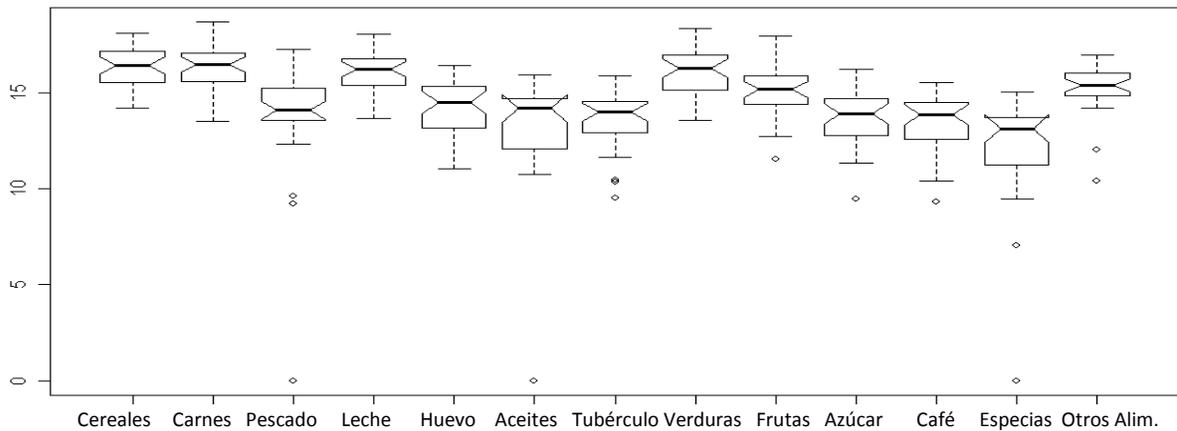
| CAFE | ESPECIAS | OTRO_ALI |
|----------------|----------------|----------------|
| Min. : 9.32 | Min. : 0 | Min. : 10.39 |
| 1st Qu.: 12.61 | 1st Qu.: 11.30 | 1st Qu.: 14.84 |
| Median : 13.86 | Median : 13.14 | Median : 15.38 |
| Mean : 13.46 | Mean : 11.51 | Mean : 15.26 |
| 3rd Qu.: 14.50 | 3rd Qu.: 13.69 | 3rd Qu.: 16.00 |
| Max. : 15.52 | Max. : 15.02 | Max. : 16.96 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Al revisar la distribución de la información mediante los diagramas de caja, se observó que tienen comportamientos similares, y donde destaca que la mediana está por arriba de la mitad de los conjuntos, asimismo se distinguieron colas notorias en ambos extremos, y datos atípicos en variables como pescado, aceites, tubérculo y frutas, entre otras (Gráfica 5.17).

Distribución del gasto por variable (valores en logaritmos)

Gráfica 5.17

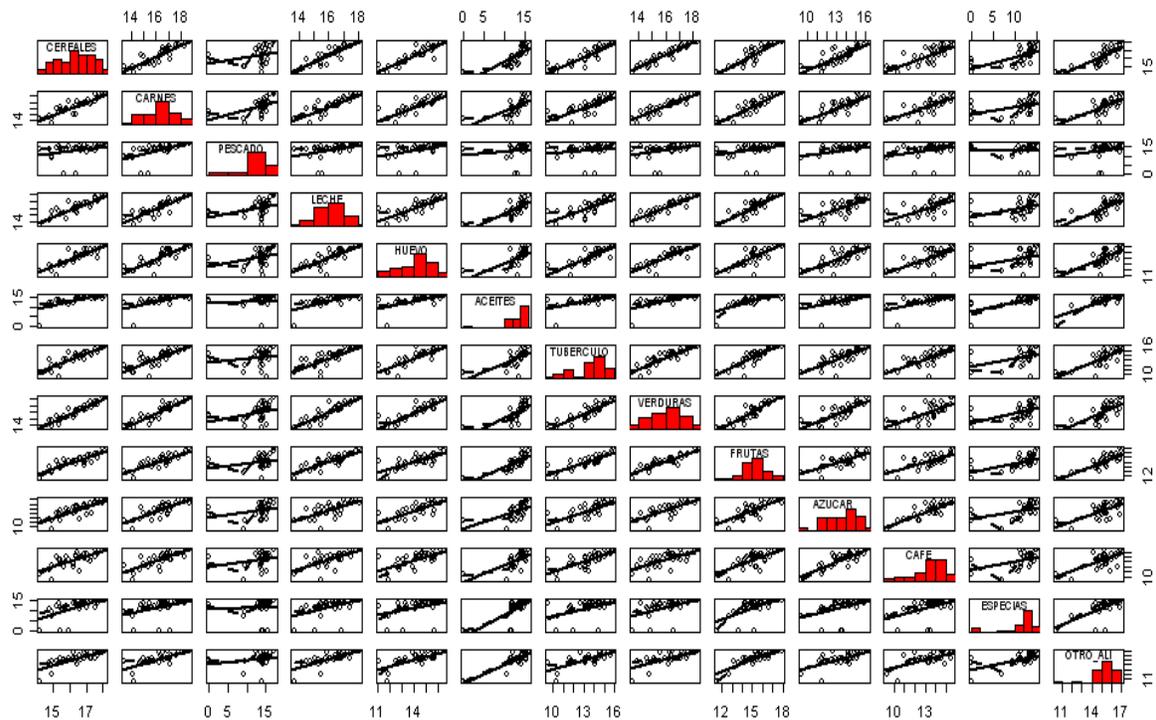


Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En la revisión gráfica del comportamiento multivariado, o las relaciones bivariadas entre las trece componentes, se presentó un mejor comportamiento de la información a diferencia de los valores reales, sin embargo sigue predominando la no linealidad de la generalidad de las observaciones, lo que se confirma con los histogramas. (Gráfica 5.18)

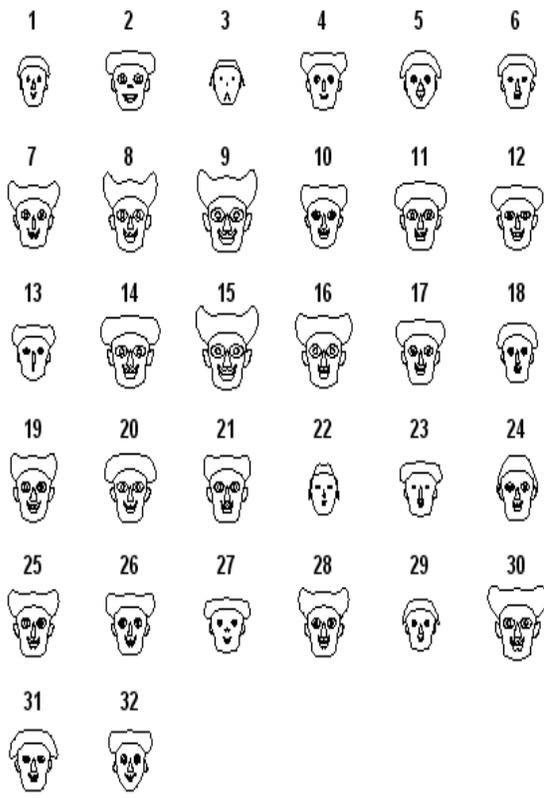
Distribución de del gasto de los hogares según variables en forma multivariada (valores en logaritmos)

Gráfica 5.18



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Distribución del gasto de los hogares en logaritmos (caras de Chernoff) Gráfica 5.19



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Al revisar la grafica de caras o imágenes se distinguen mejores formas, de hecho hay más entidades que aportan más en la cuestión del gasto por variable, dado que sus características son definidas. En tanto que hay pocas entidades donde el gasto es poco significativo lo que se traduce en imágenes más pequeñas y no definidas.

Nuevamente el Distrito Federal, el Estado de México, ahora Jalisco, Michoacán y Veracruz presentan imágenes notorias: caras amplias (influido por cereales), orejas grandes (carnes), altura de la cara grande(pescado), mentón grande de la cara (leche), gorro grande (huevo), nariz grande (aceites), posición de boca notoria (tubérculo), curvatura grande de la boca o sonrisa (verduras), longitud amplia de boca (frutas), altura de ojos grandes (azúcar), separación de ojos notoria (café), inclinación de ojos amplia (especias) y forma amplia de los ojos (otros alimentos).

Al revisar la estadística de Shapiro - Wilks por variable en comparación con la estadística obtenida (p valor), se distinguió (Cuadro 5.20) que en algunos casos las cifras son cercanas por lo que existe la posibilidad de normalidad en leche, verduras y frutas (ver prueba de Shapiro en el glosario).

Prueba de Normalidad de Shapiro-Wilks (valores en logaritmos)

Cuadro 5. 20

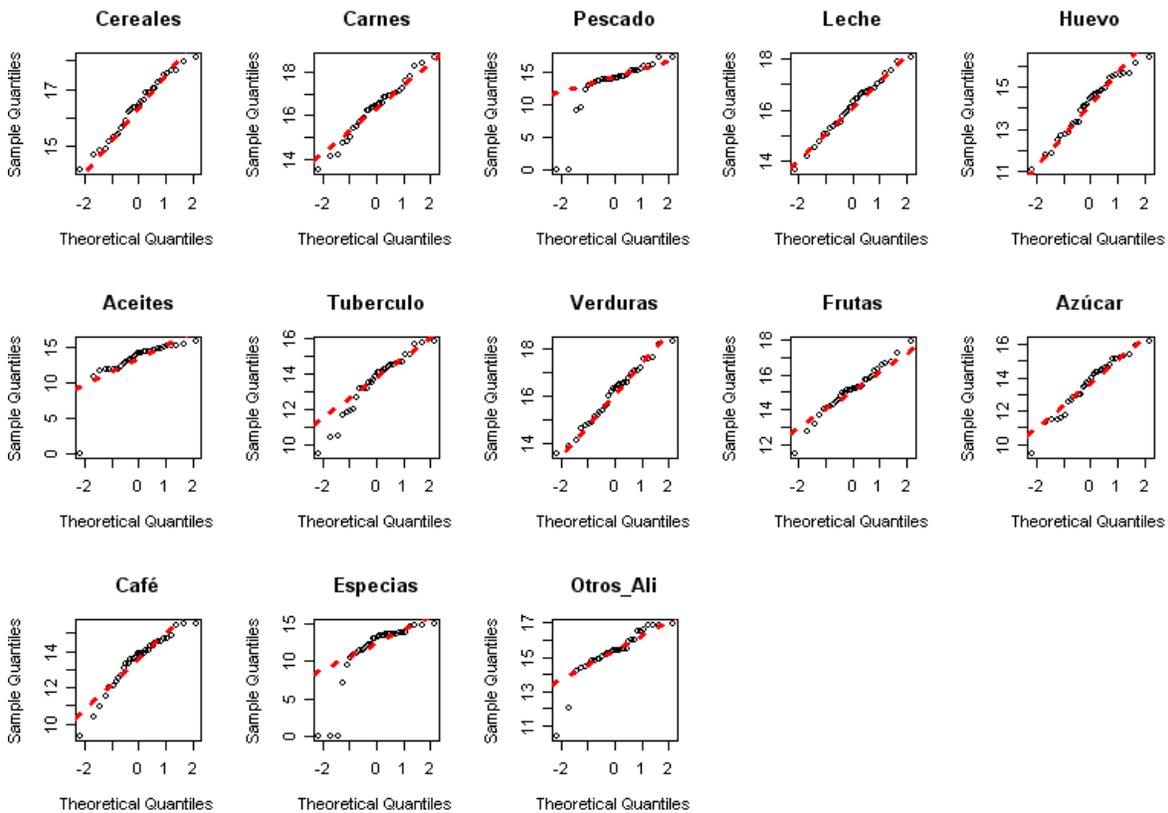
| | | | | |
|----------|-------------|--|-------------|---------------------|
| [1] "1" | "CEREALES" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9742, | p-value = 0.623 |
| [1] "2" | "CARNES" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9761, | p-value = 0.6806 |
| [1] "3" | "PESCADO" | Shapiro-Wilk normality test data: datos[, i] | W = 0.6403, | p-value = 1.243e-07 |
| [1] "4" | "LECHE" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9832, | p-value = 0.8844 |
| [1] "5" | "HUEVO" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9624, | p-value = 0.3189 |
| [1] "6" | "ACEITES" | Shapiro-Wilk normality test data: datos[, i] | W = 0.6265, | p-value = 8.177e-08 |
| [1] "7" | "TUBERCULO" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9313, | p-value = 0.04256 |
| [1] "8" | "VERDURAS" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9797, | p-value = 0.7916 |
| [1] "9" | "FRUTAS" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9782, | p-value = 0.7471 |
| [1] "10" | "AZUCAR" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9605, | p-value = 0.2837 |
| [1] "11" | "CAFE" | Shapiro-Wilk normality test data: datos[, i] | W = 0.9285, | p-value = 0.03547 |
| [1] "12" | "ESPECIAS" | Shapiro-Wilk normality test data: datos[, i] | W = 0.6767, | p-value = 3.972e-07 |
| [1] "13" | "OTRO_ALI" | Shapiro-Wilk normality test data: datos[, i] | W = 0.8255, | p-value = 0.000127 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Gráfica de cuartiles

Aquí se observó como en casi la mayoría de las observaciones en las trece componentes están sobre la tendencia, con excepción de pescado donde hay entidades que registran las cifras más altas. Lo que da la posibilidad de existencia de normalidad multivariada (Gráfica 5.20).

Distribución del gasto según variables de alimentos (valores en logaritmos) Gráfica 5.20



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Prueba de normalidad multivariada

Al contrastar las dos estadísticas, el de la Ji cuadrada y la de la normal, se identificó que la de la normal es mayor a la Ji cuadrada por lo que no existe normalidad multivariada, de las observaciones.

$$1 - \text{pchisq}(nr \cdot Ap/6, nc \cdot (nc+1) \cdot (nc+2)/6)$$

$$[1] 0.0006595728$$

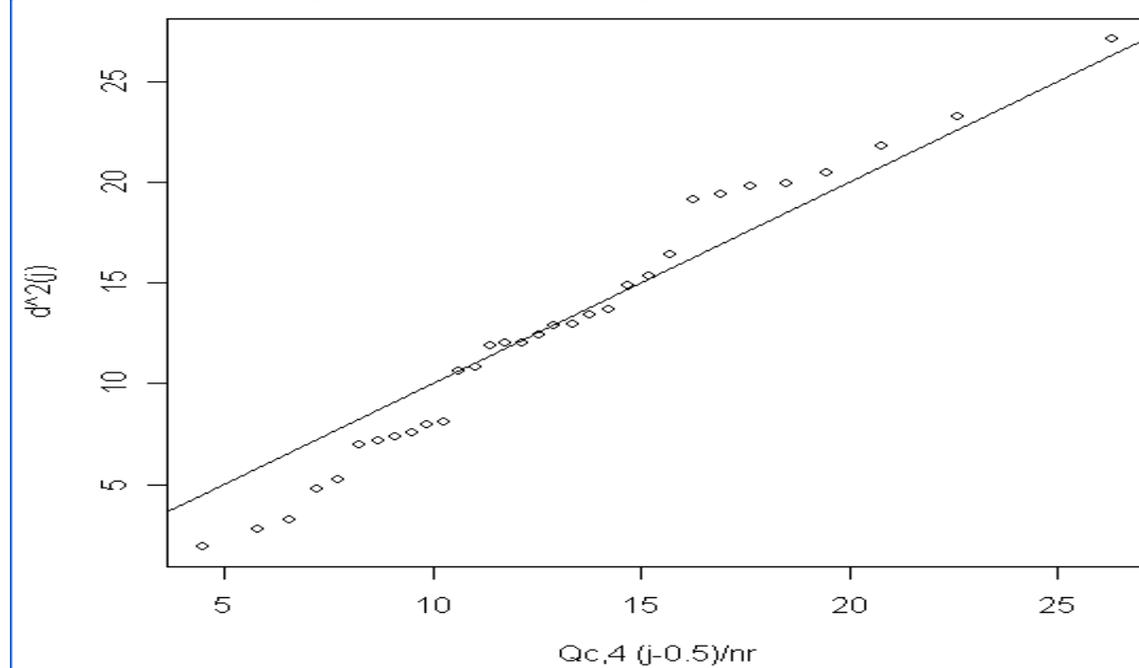
$$1 - \text{pnorm}(Kp, nc \cdot (nc+2), 8 \cdot nc \cdot (nc+2)/nr)$$

$$[1] 0.4728188$$

La gráfica de probabilidad Ji cuadrada (Gráfica 5.21) claramente distinguió la no existencia de normalidad, sin embargo la generalidad de las observaciones están más cerca de la tendencia lineal.

Gráfica de probabilidad ji-cuadrada (valores en logaritmo)

Gráfica 5.21



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Aplicación del método de componentes principales.

(Para el caso del gasto de los hogares con exclusivamente población de 65 y más años de edad).

Con los resultados anteriores, debido a que ambos ejercicios no muestran gran diferencia, se optó por presentar para este caso la aplicación de la técnica de componentes principales con los datos en valores absolutos o sin transformación logarítmica, con el fin de ejemplificar en forma distinta la utilización del método lo cual complementaria el trabajo presente.

Matriz de varianzas y covarianzas

Tiene valores positivos y significativos. Aquí se notó que las carnes presentan las cifras más altas, le siguen las verduras, los cereales y las frutas. Contrariamente, los valores menos representativos se ubican en Especies. (Cuadro 5.21).

Matriz de varianzas y covarianzas

Cuadro 5. 21
Parte 1

| | CEREALES | CARNES | PESCADO | LECHE | HUEVO | ACEITES | TUBERCULO |
|----------|-----------|----------|----------|----------|-----------|----------|-----------|
| CEREALES | 3.46e+14 | 4.83e+14 | 9.82e+13 | 2.71e+14 | 4.927e+13 | 2.74e+13 | 3.39e+13 |
| CARNES | 4.83e+14 | 8.74e+14 | 1.89e+14 | 4.46e+14 | 7.879e+13 | 3.50e+13 | 5.19e+13 |
| PESCADO | 9.82e+13 | 1.89e+14 | 5.29e+13 | 9.98e+13 | 1.912e+13 | 7.02e+12 | 1.11e+13 |
| LECHE | 2.71e+14 | 4.45e+14 | 9.98e+13 | 2.61e+14 | 4.555e+13 | 1.87e+13 | 2.96e+13 |
| HUEVO | 4.93e+13 | 7.88e+13 | 1.91e+13 | 4.56e+13 | 1.020e+13 | 4.20e+12 | 5.05e+12 |
| ACEITES | 2.740e+13 | 3.50e+13 | 7.02e+12 | 1.87e+13 | 4.203e+12 | 3.46e+12 | 2.27e+12 |
| TUBERCU | 3.39e+13 | 5.19e+13 | 1.11e+13 | 2.96e+13 | 5.051e+12 | 2.27e+12 | 4.04e+12 |
| VERDURAS | 3.72e+14 | 5.72e+14 | 1.31e+14 | 3.12e+14 | 6.017e+13 | 3.04e+13 | 3.89e+13 |
| FRUTAS | 1.80e+14 | 3.22e+14 | 7.44e+13 | 1.76e+14 | 2.896e+13 | 1.06e+13 | 2.11e+13 |
| AZUCAR | 3.95e+13 | 5.38e+13 | 1.13e+13 | 2.85e+13 | 5.872e+12 | 3.88e+12 | 3.59e+12 |
| CAFE | 2.15e+13 | 3.12e+13 | 6.91e+12 | 1.94e+13 | 3.519e+12 | 1.42e+12 | 2.38e+12 |
| ESPECIAS | 1.24e+13 | 1.81e+13 | 4.54e+12 | 9.41e+12 | 1.853e+12 | 9.82e+11 | 1.13e+12 |
| OTRO_ALI | 9.52e+13 | 1.49e+14 | 2.97e+13 | 7.96e+13 | 1.487e+13 | 8.25e+12 | 9.08e+12 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Matriz de varianzas y covarianzas**Cuadro 5. 21
Parte 2**

| | VERDURAS | FRUTAS | AZUCAR | CAFE | ESPECIAS | OTRO_ALI |
|----------|----------|----------|----------|----------|----------|----------|
| CEREALES | 3.72e+14 | 1.80e+14 | 3.95e+13 | 2.15e+13 | 1.24e+13 | 9.52e+13 |
| CARNES | 5.72e+14 | 3.22e+14 | 5.38e+13 | 3.12e+13 | 1.81e+13 | 1.49e+14 |
| PESCADO | 1.31e+14 | 7.44e+13 | 1.13e+13 | 6.91e+12 | 4.54e+12 | 2.97e+13 |
| LECHE | 3.12e+14 | 1.76e+14 | 2.85e+13 | 1.94e+13 | 9.41e+12 | 7.96e+13 |
| HUEVO | 6.02e+13 | 2.90e+13 | 5.87e+12 | 3.52e+12 | 1.85e+12 | 1.49e+13 |
| ACEITES | 3.04e+13 | 1.06e+13 | 3.88e+12 | 1.42e+12 | 9.82e+11 | 8.25e+12 |
| TUBERCU | 3.89e+13 | 2.11e+13 | 3.59e+12 | 2.38e+12 | 1.13e+12 | 9.08e+12 |
| VERDURAS | 4.50e+14 | 2.22e+14 | 4.35e+13 | 2.24e+13 | 1.46e+13 | 9.98e+13 |
| FRUTAS | 2.22e+14 | 1.47e+14 | 1.66e+13 | 1.31e+13 | 7.51e+12 | 5.18e+13 |
| AZUCAR | 4.35e+13 | 1.66e+13 | 6.31e+12 | 2.39e+12 | 1.21e+12 | 1.12e+13 |
| CAFE | 2.24e+13 | 1.31e+13 | 2.39e+12 | 2.23e+12 | 6.78e+11 | 5.40e+12 |
| ESPECIAS | 1.46e+13 | 7.52e+12 | 1.21e+12 | 6.78e+11 | 7.98e+11 | 3.25e+12 |
| OTRO_ALI | 9.98e+13 | 5.18e+13 | 1.12e+13 | 5.40e+12 | 3.25e+12 | 4.50e+13 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Matriz de correlaciones

Sobresalieron las siguientes correlaciones (Cuadro 5.22):

- Los cereales mostraron alta correlación con verduras, tubérculos y leche.
- Las carnes presentaron mucha dependencia con leche y verduras.

Matriz de correlaciones**Cuadro 5. 22
Parte 1**

| | CEREALES | CARNES | PESCADO | LECHE | HUEVO | ACEITES | TUBERCULO |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| CEREALES | 1.0000000 | 0.8795946 | 0.7262021 | 0.9023073 | 0.8298070 | 0.7906862 | 0.9068439 |
| CARNES | 0.8795946 | 1.0000000 | 0.8788228 | 0.9328952 | 0.8345598 | 0.6363300 | 0.8730955 |
| PESCADO | 0.7262021 | 0.8788228 | 1.0000000 | 0.8491243 | 0.8232995 | 0.5187814 | 0.7595325 |
| LECHE | 0.9023073 | 0.9328952 | 0.8491243 | 1.0000000 | 0.8831807 | 0.6212203 | 0.9109569 |
| HUEVO | 0.8298070 | 0.8345598 | 0.8232995 | 0.8831807 | 1.0000000 | 0.7074003 | 0.7870281 |
| ACEITES | 0.7906862 | 0.6363300 | 0.5187814 | 0.6212203 | 0.7074003 | 1.0000000 | 0.6080756 |
| TUBERCU | 0.9068439 | 0.8730955 | 0.7595325 | 0.9109569 | 0.7870281 | 0.6080756 | 1.0000000 |
| VERDURAS | 0.9420082 | 0.9123602 | 0.8518013 | 0.9112931 | 0.8880230 | 0.7692715 | 0.9123328 |
| FRUTAS | 0.7979130 | 0.8983898 | 0.8441897 | 0.8969153 | 0.7481126 | 0.4702396 | 0.8659379 |
| AZUCAR | 0.8458360 | 0.7253855 | 0.6209172 | 0.7020530 | 0.7323350 | 0.8306827 | 0.7108744 |
| CAFE | 0.7727027 | 0.7062741 | 0.6357070 | 0.8044852 | 0.7380552 | 0.5125523 | 0.7923165 |
| ESPECIAS | 0.7467544 | 0.6843034 | 0.6987433 | 0.6522727 | 0.6496291 | 0.5905791 | 0.6294949 |
| OTRO_ALI | 0.7628392 | 0.7510129 | 0.6086603 | 0.7343023 | 0.6936522 | 0.6605623 | 0.6730615 |

- c) La leche también tuvo fuerte relación con verduras y tubérculos.
- d) El Huevo mostró correlaciones importantes con casi todos los productos, resaltando las cifras de cereales, carnes y pescado.
- e) Las frutas reflejaron baja correlación con los aceites.

Matriz de correlaciones

**Cuadro 5. 22
Parte 2**

| | VERDURA | FRUTAS | AZUCAR | CAFE | ESPECIAS | OTRO_ALI |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| CEREALES | 0.9420082 | 0.7979130 | 0.8458360 | 0.7727027 | 0.7467544 | 0.7628392 |
| CARNES | 0.9123602 | 0.8983898 | 0.7253855 | 0.7062741 | 0.6843034 | 0.7510129 |
| PESCADO | 0.8518013 | 0.8441897 | 0.6209172 | 0.6357070 | 0.6987433 | 0.6086603 |
| LECHE | 0.9112931 | 0.8969153 | 0.7020530 | 0.8044852 | 0.6522727 | 0.7343023 |
| HUEVO | 0.8880230 | 0.7481126 | 0.7323350 | 0.7380552 | 0.6496291 | 0.6936522 |
| ACEITES | 0.7692715 | 0.4702396 | 0.8306827 | 0.5125523 | 0.5905791 | 0.6605623 |
| TUBERCULO | 0.9123328 | 0.8659379 | 0.7108744 | 0.7923165 | 0.6294949 | 0.6730615 |
| VERDURAS | 1.0000000 | 0.8640763 | 0.8161198 | 0.7081473 | 0.7677584 | 0.7007887 |
| FRUTAS | 0.8640763 | 1.0000000 | 0.5444097 | 0.7252854 | 0.6939442 | 0.6366264 |
| AZUCAR | 0.8161198 | 0.5444097 | 1.0000000 | 0.6385596 | 0.5399443 | 0.6627068 |
| CAFE | 0.7081473 | 0.7252854 | 0.6385596 | 1.0000000 | 0.5083889 | 0.5385190 |
| ESPECIAS | 0.7677584 | 0.6939442 | 0.5399443 | 0.5083889 | 1.0000000 | 0.5413955 |
| OTRO_ALI | 0.7007887 | 0.6366264 | 0.6627068 | 0.5385190 | 0.5413955 | 1.0000000 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Descomposición espectral de la matriz (valores propios y vectores propios)

El valor propio de la primera componente fue alto, es cercana a las 10 unidades, las restantes 12 componentes fueron inferiores a la unidad. De ahí que el modelo toma importancia en una sola componente (Cuadro 5.23).

Valores Propios

Cuadro 5. 23

| | | | | | | | | |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|
| [1] | 9.98246 | 0.91064 | 0.57561 | 0.43831 | 0.33357 | 0.26059 | 0.17357 | 0.11875 |
| [9] | 0.07855 | 0.06179 | 0.04110 | 0.01840 | 0.00666 | | | |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Los vectores propios señalan lo siguiente:

- 1) En la primera componente todos los valores tuvieron el mismo signo (negativo) y casi todos presentaron la misma representatividad al oscilar entre -0.24 (especies) y -0.30 (verduras y leche), por lo que el gasto aplicado en estos hogares tiene casi la misma proporción.
- 2) En el resto de las componentes se mostraron relaciones inversas, esto es se equilibra el gasto en el consumo de ciertos alimentos, unos en forma positiva y otros con característica negativa (Cuadro 5.24).

Vectores Propios

Cuadro 5. 24

| | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] |
|-------|---------|----------|-----------|----------|---------|----------|---------|
| [1,] | -0.3038 | 0.14171 | -0.047776 | 0.09579 | -0.2649 | -0.18108 | 0.0514 |
| [2,] | -0.2996 | -0.14456 | 0.033106 | -0.2086 | 0.13016 | -0.20247 | -0.1130 |
| [3,] | -0.2745 | -0.28595 | 0.232418 | -0.06465 | 0.5274 | 0.13457 | -0.3432 |
| [4,] | -0.3022 | -0.17941 | -0.150937 | -0.09656 | 0.0393 | 0.00558 | 0.1924 |
| [5,] | -0.2878 | 0.01771 | -0.043759 | 0.01635 | 0.3964 | 0.49145 | 0.4182 |
| [6,] | -0.2409 | 0.61094 | 0.101785 | 0.11564 | 0.0659 | 0.02411 | 0.3647 |
| [7,] | -0.2919 | -0.14138 | -0.217784 | 0.06286 | -0.1893 | -0.40951 | 0.1976 |
| [8,] | -0.3083 | 0.02409 | 0.111839 | 0.11134 | 0.1236 | -0.24633 | 0.1752 |
| [9,] | -0.2798 | -0.40480 | 0.055953 | -0.07138 | -0.1232 | -0.20904 | 0.0614 |
| [10,] | -0.2601 | 0.46496 | -0.164861 | 0.1449 | 0.2248 | -0.21931 | -0.5972 |
| [11,] | -0.2533 | -0.14964 | -0.555271 | 0.3746 | -0.2756 | 0.47262 | -0.2233 |
| [12,] | -0.2414 | -0.03795 | 0.713416 | 0.3302 | -0.4039 | 0.22962 | -0.1496 |
| [13,] | -0.2488 | 0.21692 | -0.00308 | -0.7927 | -0.3402 | 0.26400 | -0.1399 |

| | [,8] | [,9] | [,10] | [,11] | [,12] | [,13] |
|-------|---------|----------|----------|---------|---------|---------|
| [1,] | -0.2111 | 0.2532 | -0.13991 | -0.2066 | 0.6766 | 0.3765 |
| [2,] | 0.2004 | 0.5309 | -0.2916 | 0.5974 | -0.0268 | -0.0974 |
| [3,] | 0.14184 | -0.3957 | -0.2975 | -0.1836 | 0.2237 | 0.1473 |
| [4,] | -0.0708 | 0.3425 | -0.2705 | -0.6453 | -0.4166 | -0.1428 |
| [5,] | -0.3889 | 0.0962 | 0.2757 | 0.2405 | -0.0295 | 0.2143 |
| [6,] | 0.5897 | -0.1368 | -0.1555 | -0.0174 | -0.0918 | 0.0876 |
| [7,] | -0.2825 | -0.5299 | -0.24199 | 0.2576 | -0.2869 | 0.1819 |
| [8,] | -0.1164 | -0.1421 | 0.2262 | -0.0295 | 0.3021 | -0.7717 |
| [9,] | 0.4190 | 0.0130 | 0.6547 | -0.0897 | -0.0510 | 0.2681 |
| [10,] | -0.1935 | 0.0844 | 0.2935 | -0.0596 | -0.2675 | 0.0834 |
| [11,] | 0.2398 | -0.06324 | -0.0482 | 0.1021 | 0.0838 | -0.1854 |
| [12,] | -0.1518 | 0.0481 | -0.0472 | 0.0549 | -0.2306 | -0.0178 |
| [13,] | -0.0649 | -0.1941 | 0.0649 | -0.0016 | 0.0263 | -0.1042 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Desviación estándar por componente principal

Claramente se observó que la componente 1 mostró la mayor cantidad de la desviación estándar con poco más de 44 millones de unidades, en cambio, las componentes 12 y 13 registraron las cifras menos representativas (Cuadro 5.25).

Desviaciones estándar por componente

Cuadro 5. 25

| Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 |
|------------|-----------|-----------|-----------|-----------|-----------|-----------|
| 44222155.6 | 8831101.7 | 5958798.4 | 5176265.8 | 3919489.6 | 3723571.2 | 2195206.3 |

| Comp.8 | Comp.9 | Comp.10 | Comp.11 | Comp.12 | Comp.13 | |
|-----------|-----------|----------|----------|----------|----------|--|
| 1178563.4 | 1085344.7 | 840898.4 | 656061.7 | 402718.9 | 331426.2 | |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Acumulación de la varianza por componente principal.

Aquí se obtuvo la proporción de la varianza por lo que la componente uno efectivamente explicó el 91% de la variabilidad del modelo. Con la segunda componente se acumuló el 95 por ciento (Cuadro 5.26).

Acumulación de varianza por componente

Cuadro 5. 26

| | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|------------------------|------------|------------|------------|------------|------------|
| Standard deviation | 4.422e+07 | 8.831e+06 | 5.959e+06 | 5.176e+06 | 3.919e+06 |
| Proportion of Variance | 9.1644e-01 | 3.655e-02 | 1.664e-02 | 1.2556e-02 | 7.1992e-03 |
| Cumulative Proportion | 9.1644e-01 | 9.5299e-01 | 9.6963e-01 | 9.822e-01 | 9.8938e-01 |
| | Comp.6 | Comp.7 | Comp.8 | Comp.9 | Comp.10 |
| Standard deviation | .7236e+06 | 2.195e+06 | 1.177e+06 | 1.085e+06 | 8.409e+05 |
| Proportion of Variance | 6.497e-03 | 2.259e-03 | 6.509e-04 | 5.5203e-04 | 3.3137e-04 |
| Cumulative Proportion | 9.9588e-01 | 9.9814e-01 | 9.9879e-01 | 9.9934e-01 | 9.997e-01 |
| | Comp.11 | Comp.12 | Comp.13 | | |
| Standard deviation | 6.561e+05 | 4.027e+05 | 3.314e+05 | | |
| Proportion of Variance | 2.017e-04 | 7.6002e-05 | 5.1475e-05 | | |
| Cumulative Proportion | 9.9987e-01 | 9.9996e-01 | 1.000e+00 | | |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Correlaciones entre las componentes principales y las variables originales.

En esta situación se puede identificar que las variables originales juegan un papel importante de relación ya que X2(Carnes), X7 (Tubérculos), X1 (Cereales) e incluso X3 (Pescado) y X5 (Huevo), están correlacionadas en forma importante con la componente C1, tres de ellas en forma positiva y las otras dos en forma negativa.

Posteriormente se identifican otras relaciones dispersas y con menor importancia (Cuadro 5.27).

Correlaciones entre las componentes principales(C) y las variables originales(X)

Cuadro 5. 27

| | X1 | X2 | X3 | X4 | X5 | X6 | X7 |
|--------|----------|----------|----------|----------|----------|----------|----------|
| [C1,] | -0.94230 | 1.50504 | 0.78413 | 0.42795 | 0.70158 | 0.49377 | -0.98690 |
| [C2,] | -0.19586 | -0.17036 | 0.10944 | -0.09065 | 0.02621 | 0.05657 | 0.02226 |
| [C3,] | -0.11819 | -0.14285 | -0.23057 | -0.13080 | -0.33888 | -0.17601 | -0.65154 |
| [C4,] | -0.11276 | -0.01034 | -0.03186 | 0.22478 | -0.17130 | 0.06419 | 0.06368 |
| [C5,] | -0.07795 | 0.04742 | -0.04793 | 0.00116 | -0.30896 | -0.06601 | 0.09997 |
| [C6,] | -0.05924 | 0.18980 | 0.11525 | -0.16884 | -0.09526 | -0.10080 | 0.14189 |
| [C7,] | -0.04579 | 0.03095 | -0.03069 | 0.05063 | 0.03614 | 0.02867 | 0.04480 |
| [C8,] | -0.02577 | 0.02432 | -0.02632 | -0.02454 | -0.00652 | -0.00924 | 0.01952 |
| [C9,] | -0.02240 | -0.02166 | -0.04594 | 0.03327 | 0.05574 | -0.02328 | -0.00541 |
| [C10,] | -0.01480 | 0.03472 | 0.02704 | -0.03538 | -0.02479 | 0.00276 | -0.02172 |
| [C11,] | -0.01124 | 0.00820 | -0.00357 | 0.04279 | -0.00233 | 0.02577 | -0.05555 |
| [C12,] | -0.00670 | 0.00840 | -0.00958 | -0.00965 | 0.01453 | -0.01233 | -0.05014 |
| [C13,] | -0.00574 | 0.00225 | 0.02206 | 0.00806 | -0.00383 | -0.04309 | 0.00476 |

| | X8 | X9 | X10 | X11 | X12 | X13 |
|--------|----------|----------|----------|----------|----------|----------|
| [C1,] | 0.05129 | 0.30269 | 0.01878 | -0.11922 | 0.25380 | 0.06707 |
| [C2,] | -0.00505 | 0.00842 | -0.00350 | 0.00424 | -0.00518 | 0.00096 |
| [C3,] | 0.08115 | -0.00409 | 0.01683 | 0.04641 | 0.09154 | 0.01653 |
| [C4,] | 0.04166 | -0.03171 | 0.03145 | -0.00859 | -0.01691 | -0.00629 |
| [C5,] | -0.88015 | 0.64903 | -0.19744 | -0.24806 | 0.30493 | -0.21309 |
| [C6,] | -0.51706 | -0.19873 | 1.51525 | 1.08109 | 0.40529 | -0.27912 |
| [C7,] | 0.11866 | -0.13664 | -0.59481 | 0.59417 | 0.30164 | -0.62468 |
| [C8,] | 0.00650 | -0.00103 | -0.00346 | 0.00081 | -0.00619 | 0.00553 |
| [C9,] | -0.01506 | -0.00475 | 0.00731 | -0.00595 | 0.00763 | -0.00346 |
| [C10,] | -0.12780 | -0.26444 | -0.01018 | -0.13139 | -0.00269 | -0.08727 |
| [C11,] | -0.20528 | -0.05587 | -0.12113 | 0.21142 | -0.24105 | 0.17520 |
| [C12,] | 0.02231 | 0.10764 | 0.08299 | -0.03661 | -0.31953 | -0.29017 |
| [C13,] | 0.00299 | -0.00098 | -0.00348 | 0.00084 | -0.00274 | 0.00181 |

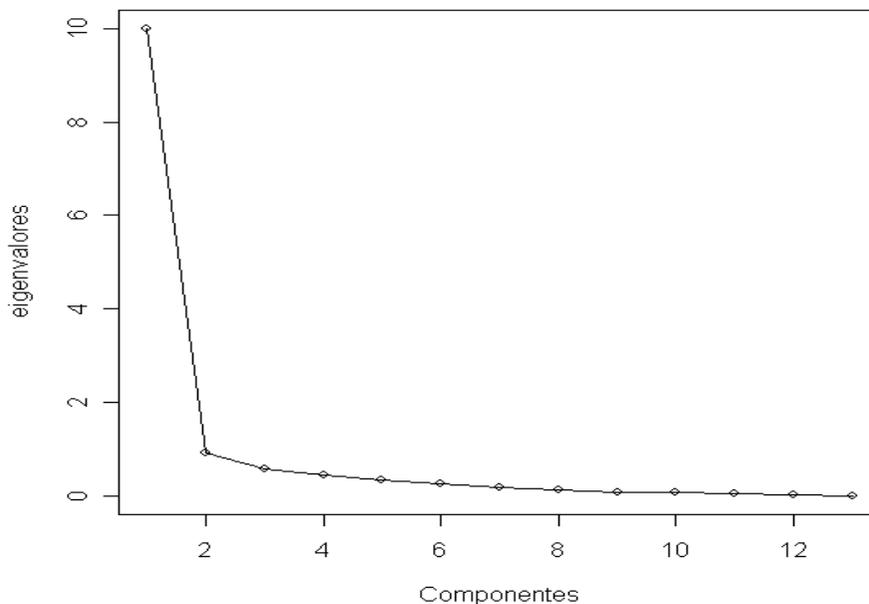
Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Gráfica de rodilla o codo.

En la gráfica se observó que el cambio de la tendencia (caída) es fuerte de la primera a la segunda componente, dado que la primera explicó más del 91% de la variabilidad, posteriormente al quiebre se observa una tendencia casi uniforme (Gráfica 5.22).

Tendencia de componentes

Gráfica 5.22



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Puntuaciones de las componentes principales según entidades federativas

En la componente 1, el Distrito Federal y el Estado de México registraron las cifras negativas más importantes con poco más de 150 millones en su gasto y en la parte positiva destacaron Baja California Sur y Aguascalientes con poco más de 37 millones. En la componente 2, nuevamente la capital del país destacó con poco más de 23 millones seguida de Jalisco con 15 millones. En la parte positiva, sobresalieron Oaxaca, Michoacán y Estado de México con poco más de 14 millones (Cuadro 5.28).

Puntuaciones de los componentes 1 a 7

Cuadro 5. 28

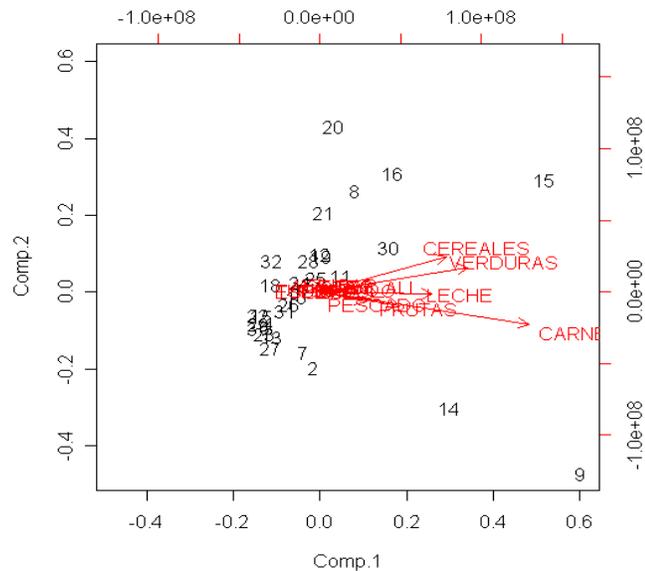
| | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] |
|----|--------------|-------------|--------------|--------------|------------|------------|-------------|
| 1 | 37189760.7 | -2998887.0 | -2097834.19 | 858917.21 | -266623.8 | -937018.5 | 71994.38 |
| 2 | 3544489.6 | -9845653.2 | -2439683.51 | 296588.74 | 9423036.8 | 1684481.1 | 4351799.00 |
| 3 | 38304396.6 | -4700157.4 | -3288020.59 | 3160599.00 | -312922.9 | 652390.5 | -781505.66 |
| 4 | 29314915.5 | -4024724.2 | -120766.41 | 3397918.66 | -2299879.6 | -105799.0 | -1242927.60 |
| 5 | 29931391.0 | -3102045.2 | -2963816.61 | -41901.59 | -2590421.1 | -1478555.3 | 497312.15 |
| 6 | 31310680.3 | -4619925.5 | 515019.05 | 1483125.15 | -584477.1 | -2042430.6 | -101242.82 |
| 7 | 9590010.3 | -7809232.1 | 4180153.82 | 8938493.96 | -4371236.4 | 4995178.7 | -610456.70 |
| 8 | -20611577.3 | 13233629.9 | -4039079.18 | -18279047.04 | 517684.6 | 8261033.1 | 800487.06 |
| 9 | -150643202.3 | -23616052.1 | -10563420.94 | -6678810.33 | -7199731.8 | -2774605.4 | -2251921.46 |
| 10 | 13979617.1 | -560310.1 | -1959615.68 | -3744726.66 | 3575782.0 | 1460770.7 | 2066140.79 |
| 11 | -11899526.1 | 2104080.2 | -69237.39 | 1016837.18 | -2286795.0 | -288037.8 | 2463016.57 |
| 12 | 463989.6 | 4931974.2 | -1710949.77 | 2450030.54 | 7247098.9 | 1924873.0 | -1281943.79 |
| 13 | 27931072.5 | -5892146.1 | -3495375.71 | -685595.48 | 792854.8 | -285255.8 | 1109331.61 |
| 14 | -74275146.6 | -15058261.4 | 11947143.62 | 4085373.07 | 2112419.9 | 5652470.7 | 4756393.10 |
| 15 | -129407838.3 | 14535977.0 | -4720900.71 | 9502628.73 | 9578163.8 | -4154791.8 | -2111143.85 |
| 16 | -41693254.1 | 15503183.1 | 385119.97 | 592806.31 | -6443391.8 | 4210653.7 | 2410142.21 |
| 17 | 10314584.3 | 1002858.0 | -1858451.26 | -253428.34 | -1749149.7 | 1578150.1 | 1818392.17 |
| 18 | 28714867.7 | 979789.1 | -815207.65 | -2620019.32 | -689338.4 | -1189540.8 | -1883124.57 |
| 19 | -817108.7 | 4634337.1 | 10542293.28 | -6863572.58 | 2244539.3 | -7806236.0 | 2140798.83 |
| 20 | -7400100.1 | 21542697.6 | -11364311.28 | 8750745.82 | -6014426.6 | -3364543.4 | 3513450.21 |
| 21 | -2043827.2 | 10400891.1 | 6007941.21 | 652770.69 | -3410718.8 | 5290412.1 | -3434214.28 |
| 22 | 35565824.4 | -2949032.5 | -3842221.71 | 445708.27 | 1104700.4 | 1546987.0 | 987688.47 |
| 23 | 31884170.6 | -5365084.6 | -2422283.45 | 794479.93 | 1497395.3 | 107033.2 | -3324613.13 |
| 24 | 11593604.0 | 1203393.4 | -2519631.70 | -3143241.23 | 622793.0 | -3168745.2 | -789819.41 |
| 25 | 2366828.4 | 1859136.3 | -259648.45 | -2069774.24 | 5979365.0 | 3688106.4 | -4874088.72 |
| 26 | 17878118.9 | -1572269.6 | 387771.26 | -26544.62 | 173029.7 | 887588.8 | 87569.81 |
| 27 | 28372637.2 | -7329190.7 | 352056.10 | 4011367.18 | -457000.8 | 1544862.9 | -584825.98 |
| 28 | 6505085.8 | 4097229.1 | 5122925.19 | -8025206.82 | 248109.3 | -5847609.8 | 258713.64 |
| 29 | 35620388.1 | -4127561.8 | -3217288.41 | 2875627.95 | -709938.2 | -1137291.6 | -378853.83 |
| 30 | -40006763.4 | 5793219.1 | 19979970.70 | 2139665.40 | -3112617.0 | -331447.8 | -2211787.15 |
| 31 | 20281463.0 | -2384413.3 | 5594626.20 | -2346759.95 | -926615.7 | -9463829.6 | 301331.42 |
| 32 | 28140448.9 | 4132551.7 | -1247275.81 | -675055.58 | -1691688.1 | 890746.5 | -1772092.46 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

En la gráfica se observó que destacaron las componentes de carnes, verduras y cereales al sobresalir del resto de los tipos de alimentos.

El Distrito Federal es la entidad más alejada y está influenciada fuertemente por las carnes, Jalisco está también próxima. Por su parte el Estado de México está más influenciada por verduras y cereales, Veracruz y Michoacán están próximas a dichas componentes.

Concentración de las entidades según componentes 1 y 2 **Gráfica 5.23**



Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

He ahí los resultados obtenidos entre los dos modelos utilizados, por un lado cuando los hogares contienen a al menos una persona de 65 y más años de edad y por otra parte a los hogares que tienen exclusivamente residentes en dicho rango de edad, y cuyas diferencias son explicadas en el siguiente capítulo de conclusiones, ahora bien surge la pregunta : ¿qué pasaría cuando se tiene el gasto per cápita de los hogares, tanto de toda la población como del exclusivo a este grupo de personas en edad posterior a los 65 años de edad?, las respuestas permitirían identificar otra asignación de gasto en los alimentos o bien, si sigue existiendo semejanza por la concentración de población por entidad y en consecuencia el gasto de los hogares en los mismos productos alimenticios y que podría ser analizado en otro momento.

Capítulo 6. CONCLUSIONES Y CONSIDERACIONES FINALES

6. Conclusiones y consideraciones finales

La principal contribución de este trabajo es la evaluación y comparación del gasto hecho por los hogares del país con al menos un adulto de 65 y más años de edad y de aquellos que solamente son integrados por este segmento de población, utilizando el método de componentes principales, con la intención de resaltar algunos resultados, cómo son: el mínimo número de componentes que explican la mayoría de la variabilidad del modelo, cómo es la matriz de varianzas-covarianzas y la de correlaciones, el interés de determinar el grado de dependencia, cuántas componentes vistas a partir de los valores propios son significativas (superiores a la unidad); y por último, en un razonamiento gráfico de las dos primeras componentes identificar cómo son distintivas las variables en cuanto a la asignación del gasto y cuales entidades federativas están asociadas. Por lo cual resumiremos los resultados obtenidos en las dos siguientes secciones, el de comparativo y la conclusión final.

Comparativo

En el caso de los gastos de los hogares con al menos un residente de 65 y más años de edad, la aplicación del Método de Componentes Principales presentó los siguientes resultados, con un componente se garantizó el 86.38 por ciento de la variabilidad, y con dos componentes se explicó cerca del 90 por ciento.

La matriz vectores propios indicó que la componente 1 tiene factores de afectación negativa casi en las mismas proporciones con excepción de pescado que registra una cifra menor dada por su poca correlación. En cambio la componente 2 tuvo influencia contrastante al registrar factores en forma positiva como negativa, aún cuando los valores de los parámetros son casi similares. La

mayoría de las variables se encontraron con alta correlación, lo que señala que hay dependencia entre variables.

En el caso de la correlación de las componentes principales y las variables originales, se distinguieron dos tipos de asociación sobre todo con la Componente 1, la positiva con Tubérculos, Huevo, Café y Azúcar, y la negativa con Aceites, pescado, Carnes y Especias.

El valor propio de la primera componente fue superior a la unidad, el resto fueron cada vez menos significativos. El vector propio correspondiente a la primera componente tuvo cifras similares y con el mismo signo. La componente uno registró (aún cuando no es un hallazgo de acuerdo a la teoría) la desviación estándar más alta, le siguió en importancia la segunda componente. Por último, sobresalieron del resto de los componentes el tercero. Las entidades que registraron mayores gastos son las que mostraron las cifras de componentes más altas, por señalar Estado de México y Distrito Federal en la mayoría de las variables, lo cual es resultado por la alta concentración de la población (sobre todo con edades inferiores a los 30 años) esto porque el estado de México y el Distrito Federal son las entidades más pobladas en el país (casi una cuarta parte en conjunto).

Para el caso de los hogares donde exclusivamente habitan personas de 65 y más años de edad, la situación es distinta con la aplicación de la técnica:

Una sola componente explicó el 91% de la variabilidad respecto a las variables de insumo, con dos componentes se incrementó a 95 por ciento.

La matriz de varianzas y covarianzas fue positiva, las carnes presentaron las cifras más altas (en covarianzas), le siguieron las verduras, los cereales y las frutas. En la matriz de correlaciones se obtuvo una alta dependencia entre carnes, verduras, leche y cereales.

Para el caso de la correlación de las componentes principales y las variables originales, en esta situación se puede identificar que las variables originales juegan un papel importante de relación con la componente 1, en forma positiva las Carnes, Pescado y el Huevo; y en forma negativa los Cereales y los Tubérculos).

El valor propio correspondiente a la primera componente fue alto, es cercano a las 10 unidades. En la primera componente todos tuvieron el mismo signo y casi todos presentaron la misma representatividad al oscilar entre 0.24 (especies) y 0.30 (verduras y leche), por lo que el gasto aplicado en estos hogares mostró casi la misma proporción. La componente 1 explicó la mayor cantidad de la desviación estándar con poco más de 44 millones de unidades. Al relacionar las componentes 1 y 2 por medio de una gráfica se observó que las carnes, verduras, cereales y frutas sobresalieron del resto de los tipos de alimentos. El Distrito Federal fue la entidad más alejada e influenciada fuertemente por las carnes, Jalisco se mostró próxima. Por su parte el estado de México fue la más influenciada por verduras y cereales.

Conclusión final

Se presentaron distintos comportamientos en cada uno de los modelos: el que considera el gasto de los hogares con al menos un adulto mayor es explicado por un componente en 86% de la variabilidad total y con dos componentes en 90% de la variabilidad. Los cereales y las carnes sobresalen del resto de las 13 variables, así como el pescado. Asimismo, en la relación de la Componente 1 con las variables originales, se distinguieron en forma positiva los productos Tubérculos,

Huevo, Café y Azúcar, las cuales se distinguen por ser productos de necesidades sobre todo de la población joven y productiva. Las entidades que se distinguieron en cuanto a la asignación de gasto son: Distrito Federal y Veracruz. Las correlaciones fueron tan altamente significativas pero sí importantes. Por su parte en el modelo donde sólo se consideró el gasto de los hogares con exclusivamente población de 65 y más años de edad, un solo componente explicó el 91% de la variabilidad y con dos componentes el 95% de la variabilidad, por otra parte, las carnes, verduras, cereales y leche sobresalieron del grueso de las variables.

Aquí la correlación positiva entre la componente 1 y las variables originales la distinción positiva fue con Carnes, Pescado y el Huevo, que están considerados con alto contenido vitamínico y necesario para el adulto mayor. El Distrito Federal estuvo fuertemente influenciado por las carnes, y el estado de México por las verduras. De ahí que comparar un modelo con mayor gasto y que considera más población diferente de 65 años de edad tienen a gastar más en otros tipos de alimentos que aquellos hogares que cuentan exclusivamente con población de 65 y más años de edad. Y es posible señalar que en tiempos no muy lejanos, la posibilidad de gastar en estos tipos de alimentos por parte de la población de 65 y más años de edad se puede modificar por cuestiones asociadas a su salud, ya sea por enfermedades o por que la capacidad de ingerir ciertos alimentos se vuelve más difícil, caso de las carnes.

Asimismo, la aplicación del análisis exploratorio es de gran utilidad para seleccionar el modelo de información óptimo dado, que entre más volumen de información mayor variabilidad se tendrá. Y por otra parte, que la normalidad de los datos se distingue por los valores transformados, y conforme sean más parecidos los datos, ya que entre menos se parezcan la normalidad no se hace presente en forma multivariada.

Por último, señalar que el uso del método de componentes principales aplicado al conjunto de variables de alimentos que se obtienen de la ENIGH permitió determinar aquellas componentes que mayor importancia en el gasto propician en los hogares que tienen población adulta mayor (en hogares con residentes de distintas edades, o en forma exclusiva), por lo que en el futuro, podría dar pauta para que se ponga atención en la adquisición suficiente (producción y costo) de dichos elementos alimenticios (más importantes como son carnes, cereales, verduras y frutas) por la importancia de las necesidades vitamínicas conforme se incrementa la edad de dicha población y por su ritmo de crecimiento observado en el tiempo.

Asimismo, permite identificar aquellas variables que aportan una mínima parte por lo que son susceptibles de ser sustituida (con otras de mayor aportación), por ejemplo en lugar de especias o café agregar pan y tortillas (estas forman parte de los cereales). Y mencionar, que aún cuando se identifican datos extraños o aberrantes se deben de tratar por medio de alguna técnica o bien excluirlos. El uso del método garantiza el mejor tratamiento del conjunto de datos ya que permite encontrar que tan relacionadas están las observaciones en cada componente y el grado de aportación, que no es fácil identificar con otro método.

ANEXOS

Anexos

1. Estructuras de Bases de datos.

En la ENIGH fueron consideradas para la generación de la base P65mas.dbf las siguientes estructuras (tablas):

Tabla: Concentrado

Parte 1

| Nombre decampo | Tipo | Long. | Dec. | Descripción |
|----------------|---------|-------|------|--|
| FOLIO | Carácte | 11 | 0 | Identificador del hogar |
| HOG | Numeri | 10 | 0 | Factor de expansión |
| ESTRATO | C | 1 | 0 | Estrato (tamaño de localidad) |
| CONAPO | C | 1 | 0 | Estrato conapo |
| UBICA_GEO | C | 5 | 0 | Ubicación geográfica |
| TENENCIA | C | 1 | 0 | Tenencia |
| CLASE_HOG | C | 1 | 0 | Clase de hogar |
| EDAD | N | 2 | 0 | Edad del jefe del hogar |
| ED_FORMAL | C | 2 | 0 | Nivel de instrucción del jefe del hogar |
| TAM_HOG | N | 2 | 0 | Población total |
| HOMBRES | N | 2 | 0 | Población total hombres |
| MUJERES | N | 2 | 0 | Población total mujeres |
| TOT_RESI | N | 2 | 0 | Miembros del hogar |
| TOT_HOM | N | 2 | 0 | Miembros del hogar hombres |
| TOT_MUJ | N | 2 | 0 | Miembros del hogar mujeres |
| MAYORES | N | 2 | 0 | Miembros del hogar mayores |
| MENORES | N | 2 | 0 | Miembros del hogar menores |
| P0A11 | N | 2 | 0 | Miembros del hogar de 0 a 11 años |
| P12_64 | N | 2 | 0 | Miembros del hogar de 12 a 64 años |
| P65MAS | N | 2 | 0 | Miembros del hogar de 65 años y más |
| N_OCUP | N | 2 | 0 | Número de ocupados |
| PERING | N | 2 | 0 | Perceptores de ingreso corriente |
| PEROCU | N | 2 | 0 | Perceptores de ingreso corriente monetario ocupados |
| INGTOT | N | 12 | 2 | Ingreso total |
| INGCOR | N | 12 | 2 | Ingreso corriente total |
| INGMON | N | 12 | 2 | Ingreso corriente monetario |
| TRABAJO | N | 12 | 2 | Remuneraciones al trabajo asalariado |
| SUELDOS | N | 12 | 2 | Sueldos y horas extras |
| OTRA_REM | N | 12 | 2 | Otras remuneraciones |
| NEGOCIO | N | 12 | 2 | Renta empresarial |
| NO_AGROP | N | 12 | 2 | Negocios no agropecuarios |
| AGROPE | N | 12 | 2 | Negocios agropecuarios |
| COOPERA | N | 12 | 2 | Cooperativa de producción |
| SOCIE | N | 12 | 2 | Ingresos de sociedades |
| RENTAS | N | 12 | 2 | Renta de la propiedad |
| TRANSFER | N | 12 | 2 | Transferencias |
| JUBILA | N | 12 | 2 | Jubilaciones y pensiones |
| BECA_DON | N | 12 | 2 | Becas y donativos |
| DONATIVO | N | 12 | 2 | Regalos o donativos en dinero provenientes de otros hogares |
| OTROS | N | 12 | 2 | Otros ingresos corrientes |
| GASNOM | N | 12 | 2 | Ingreso corriente no monetario |
| AUTO | N | 12 | 2 | Autoconsumo |
| PAGO | N | 12 | 2 | Pago en especie |
| REGA | N | 12 | 2 | Regalos |
| ESTI | N | 12 | 2 | Estimación del alquiler de la vivienda |
| PERTOT | N | 12 | 2 | Percepciones financieras y de capital totales |
| PERMON | N | 12 | 2 | Percepciones financieras y de capital monetarias |
| RETIRO | N | 12 | 2 | Retiro de inversiones, tandas, cajas de ahorro, etcétera. |
| PRESTAMO | N | 12 | 2 | Préstamos recibidos de personas no miembros del hogar o instituciones; excluye préstamos hipotecarios. |
| OTRAS_PR | N | 12 | 2 | Otras percepciones |
| ERONOM | N | 12 | 2 | Percepciones financieras y de capital no monetarias. |
| GASTOT | N | 12 | 2 | Gasto total |
| GASCOR | N | 12 | 2 | Gasto corriente total |
| GASMON | N | 12 | 2 | Gasto corriente monetario |
| ALIMENTOS | N | 12 | 2 | Alimentos y bebidas consumidas dentro y fuera del |

Tabla: Concentrado

| Nombre del campo | Tipo | Long. | Dec. | Descripción |
|------------------|------|-------|------|---|
| ALI_DENT | N | 12 | 2 | Alimentos y bebidas consumidas dentro del hogar |
| CEREALES | N | 12 | 2 | Cereales |
| CARNES | N | 12 | 2 | Carnes |
| PESCADO | N | 12 | 2 | Pescados y mariscos |
| LECHE | N | 12 | 2 | Leche y sus derivados |
| HUEVO | N | 12 | 2 | Huevo |
| ACEITES | N | 12 | 2 | Aceites |
| TUBÉRCULO | N | 12 | 2 | Tubérculo |
| VERDURAS | N | 12 | 2 | Verduras |
| FRUTAS | N | 12 | 2 | Frutas |
| AZUCAR | N | 12 | 2 | Azúcar |
| CAFÉ | N | 12 | 2 | Café |
| ESPECIAS | N | 12 | 2 | Especias |
| OTRO_ALI | N | 12 | 2 | Otros alimentos |
| BEBIDAS | N | 12 | 2 | Bebidas |
| FUERA_HOG | N | 12 | 2 | Alimentos y bebidas consumidas fuera del hogar |
| TABACO | N | 12 | 2 | Tabaco |
| VESTIDO | N | 12 | 2 | Vestido y calzado |
| VESTIDO_C | N | 12 | 2 | Vestido |
| VES_3YMAS | N | 12 | 2 | Vestido para personas de 4 años y más |
| VES_3MEN | N | 12 | 2 | Vestido para personas de 0 a 3 años |
| CALZADO | N | 12 | 2 | Calzado y su reparación |
| VIVIENDA | N | 12 | 2 | Vivienda, servicios de conservación, energía eléctrica y combustibles |
| ALQUILER | N | 12 | 2 | Alquileres brutos, impuestos, predial y cuotas por servicio de conservación |
| AGUA | N | 12 | 2 | Agua |
| ENERGIA | N | 12 | 2 | Energía eléctrica y combustibles |
| LIMPIEZA | N | 12 | 2 | Artículos y servicios para la limpieza y cuidados de la casa, enseres domésticos, muebles, cristalería, utensilios, domésticos y blancos |
| CUIDADOS | N | 12 | 2 | Artículos y servicios para la limpieza y cuidados de la casa |
| ENSERES | N | 12 | 2 | Enseres domésticos y muebles, cristalería, utensilios domésticos y blancos |
| SALUD | N | 12 | 2 | Cuidados médicos y conservación de la salud |
| ATEN_PRI | N | 12 | 2 | Atención primaria o ambulatoria |
| HOSPITAL | N | 12 | 2 | Atención hospitalaria, servicios médicos y medicamentos durante el embarazo y parto, aparatos ortopédicos, terapéuticos y seguros médicos |
| MEDICA | N | 12 | 2 | Medicamentos sin receta |
| TRANSPORTE | N | 12 | 2 | Transporte, adquisición, mantenimiento y accesorios para vehículos y comunicaciones |
| TRANS_AD | N | 12 | 2 | Transporte, adquisición, mantenimiento y accesorios para vehículos |
| PUBLICO | N | 12 | 2 | Transporte público |
| FORANEO | N | 12 | 2 | Transporte foráneo terrestre, aéreo y otros servicios especiales de transporte |
| VEHICULO | N | 12 | 2 | Adquisición de vehículos de uso particular, mantenimiento y accesorios para vehículos |
| COMUNICA | N | 12 | 2 | Comunicaciones |
| EDUCACION | N | 12 | 2 | Servicios y artículos de educación y esparcimiento; paquetes turísticos y para fiestas, hospedaje y alojamiento |
| EDUCA | N | 12 | 2 | Artículos y servicios de educación |
| ESPARCI | N | 12 | 2 | Artículos y servicios de esparcimiento, paquetes para: fiestas, turísticos, hospedaje y alojamiento |
| PERSONAL | N | 12 | 2 | Artículos y servicios para el cuidado personal, accesorios y efectos personales, otros gastos diversos y transferencias |
| CUIDADO | N | 12 | 2 | Artículos y servicios para el cuidado personal |
| ACCESORIO | N | 12 | 2 | Accesorios y efectos personales |
| TRANSFE | N | 12 | 2 | Otros gastos diversos y transferencias |
| EROTOT | N | 12 | 2 | Erogaciones financieras y de capital totales |
| EROMON | N | 12 | 2 | Erogaciones financieras y de capital monetarias |
| CUOTA | N | 12 | 2 | Cuota pagada por la vivienda propia |

Tabla: Concentrado**Parte 3**

| Nombre de campo | Tipo | Long. | Dec. | Descripción |
|-----------------|------|-------|------|---|
| MATERIAL | N | 12 | 2 | Materiales para reparación, mantenimiento y/o ampliación de la vivienda |
| SERVICIO | N | 12 | 2 | Servicios de reparación, mantenimiento y/o ampliación de la vivienda |
| DEPOSITO | N | 12 | 2 | Depósito de cuenta de ahorros, tandas, cajas de ahorro, etc. |
| TERCEROS | N | 12 | 2 | Préstamos a terceros |
| PAGO_TAR | N | 12 | 2 | Pago por tarjeta de crédito al banco o casa comercial |
| DEUDAS | N | 12 | 2 | |
| MONEDAS | N | 12 | 2 | Compra de monedas, metales preciosos, joyas y obras de arte |
| CASAS | N | 12 | 2 | Compra de casas, condominios, locales y terrenos |
| BALANCE | N | 12 | 2 | Balance negativo en negocios propiedad del hogar |
| OTRA_ERO | N | 12 | 2 | Otras erogaciones |
| SMG | N | 12 | 2 | Salarios mínimos generales |

Tabla Población

| Nombre de campo | Tipo | Long. | Dec. | Descripción |
|-----------------|------|-------|------|---|
| FOLIO | C | 11 | 0 | Identificador del hogar |
| NUM_REN | C | 2 | 0 | Identificador de la persona Apartado 1.3 preg 01 |
| PARENTESCO | C | 3 | 0 | Identificador de la persona 100 Jefe o Jefa 200 Espos(a) 300 Hijo(a), 400 Trabajador(es) doméstico(s), 500 No tiene Parentesco, 601 al 624 Otro parentesco, 700 Huésped |
| SEXO | C | 1 | 0 | Identificador de la persona 1 Hombre, 2 Mujer |
| EDAD | N | 2 | 0 | Identificador de la persona 00 AL 97 |
| .. | | | | |
| .. | | | | |
| .. | | | | |

Tabla: Hogares

| Nombre de campo | Tipo | Long. | Dec. | Descripción |
|-----------------|------|-------|------|---|
| FOLIO | C | 11 | 0 | Identificador del hogar 1,2 Entidad, 3 decena, 4,5,6 consecutivo 001-899, 7 tipo de hogar |
| ESTRATO | C | 1 | 0 | Estrato 1 Localidades de 100 000 habitantes y más 2 Localidades de 15 000 a 99 999 habitantes 3 Localidades de 2 500 a 14 999 habitantes 4 Localidades menores de 2 500 habitantes |
| CONAPO | C | 1 | 0 | Estrato del conapo |
| UBICA_GEO | C | 20 | 0 | Ubicación geográfica |
| MUROS01 | C | | 2 | 0 Material de muros Apartado 1.1 pregunta 01 |
| .. | | | | |
| .. | | | | |

2. Programas utilizados

a) En foxpro

```
1. selec 1***** abrir bases de datos
2. use concen
3. index on folio1 to concen
4. selec 2
5. use pobla
6. index on folio to pobla ***se crean índices de bases de datos
7. close data
8. selec 1
9. use concen index concen*** se abre la base indexada
10. selec 2
11. use pobla index pobla
12. selec 1
13. var1=space(11)
14. var2=space(10)
15. do while not eof() ###procedimiento de búsqueda en la base
16.   var1=folio1
17.   var2=hog
18.   selec 2
19.   seek var1
20.   if found()
21.     do while folio=var1 ###si los folios son iguales
22.       replace fact_aux with var2###se reemplaza el factor
23.       skip
24.       if folio<>var1
25.         Exit****salida
26.       endif
27.     enddo
28.   endif
29.   selec 1
30.   Skip****otro registro
31. enddo
32. close data
33. clear
34. return
35.
```

En el programa gratuito R de comandos estadísticos.

Este programa genera una sección de las estadísticas y gráficas presentadas en el documento.

Com_Princ.R

```
library(foreign)
datos<-read.dbf("Gasto65.dbf")### Gasto65e.dbf son los hogares con exclusivamente
```

```

nr <- nrow(datos)
nc <- ncol(datos)
#ldatos<-log(datos[3:15])### transformación a logaritmos
ldatos<-datos[3:15]
summary(ldatos)### cuadro de estadísticos
(S<-var(ldatos))### matriz de varianzas y covarianzas
(Ro <- cor(ldatos))### Matriz de correlaciones
#####
(ee <- eigen(Ro))### eigenvectores y eigenvalores
(eva <- ee$val)### eigenvalores
(eve <- ee$vec)### eigenvectores
#####
pcS <- princomp(ldatos)## valore de los componentes principales
summary(pcS)#####aquí veo que 2 componentes satisfacen más del 90%
### gráfica de rodilla o codo
plot(eva,type = "o",ylab="eigenvalores",xlab="Componentes",main="Grafica de componentes" )
### matriz de escores o puntuaciones
sco1 <- pcS$score
##sco2 <- as.matrix(scale(datos[3:15],T,F)) %*% eigen(S)$vec
sco2 <- as.matrix(scale(ldatos,T,F)) %*% eigen(S)$vec
###grafica de puntuaciones
plot(sco1[,1],sco1[,2],xlab="Score 1",ylab="Score 2",col = "blue", cex=.6 )
##### gráfica de variables latentes
biplot(pcS)
pcS$scores

x<-as.matrix(ldatos)
y<-as.matrix(eve[,1])
x%*%y
comp1<-as.matrix(x%*%y)
(final<-cbind(x,comp1*-1))
exp(final)##### matriz de datos con eigenvector 1 (para modelo lineal)
###Graficas de boxplot (caja y bigotes)
boxplot(ldatos[3],notch=TRUE, COL="blue")
boxplot(ldatos[1:4],COL="blue", main="Concentr. de Car., Cer., Ver., Lech.")
boxplot(ldatos[5:8],COL="red", main="Concentr. de Otros, Frutas, Pesca, Huevo")
boxplot(ldatos[9:13],COL="orange", main="Concentr. de Azúcar, Aceite, Tuber., Café, Espec.")
boxplot(datos[8:10],notch=TRUE, COL="blue")
boxplot(ldatos,notch=TRUE, ps=0.5, COL="blue", adjust=0.2, main="Variables de gasto alimentario", font=5)
PPP<-eve*sqrt(eva)*1/sqrt(diag(S))### para obtener la matriz del teorema 3, las correlaciones entre las variables originales
y los componentes principales

```

Procedimientos para la generación de gráficas múltiples

```

chisplot <- function(x) {
  if (!is.matrix(x)) stop("x is not a matrix")
  ### determine dimensions
  n <- nrow(x)
  p <- ncol(x)
  #
  xbar <- apply(x, 2, mean)
  S <- var(x)
  S <- solve(S)
  index <- (1:n-0.5)/n
  #
  xcent <- t(t(x) - xbar)
  di <- apply(xcent, 1, function(x,S) x %*% S %*% x,S)
  #
  quant <- qchisq(index,p)
  plot(quant, sort(di), ylab = "Ordered distances",
       xlab = "Chi-square quantile", lwd=2,pch=1)
}

bivden<-function(x, y, ngridx = 30, ngridy = 30, constant.x = 1, constant.y = 1) {
  #x and y are vectors containing the bivariate data
  #ngridx and ngridy are the number of points in the grid
  #
  mx <- mean(x)
  sdX <- sqrt(var(x))

```

```

my <- mean(y)
sdy <- sqrt(var(y))
#scale x and y before estimation
x <- scale(x)
y <- scale(y)
#
den <- matrix(0, ngridx, ngridy)
#
#find possible value for bandwidth
#
n <- length(x)
#
hx <- constant.x * n^(-0.2)
hy <- constant.y * n^(-0.2)
h <- hx * hy
hsqrt <- sqrt(h)
#
seqx <- seq(range(x)[1], range(x)[2], length = ngridx)
seqy <- seq(range(y)[1], range(y)[2], length = ngridy)
#
for(i in 1:n) {
  X <- x[i]
  Y <- y[i]
  xx <- (seqx - X)/hsqrt
  yy <- (seqy - Y)/hsqrt
  den <- den + outer(xx, yy, function(x, y)
    exp(-0.5 * (x^2 + y^2)))
  }
  den <- den/(n * 2 * pi * h)
  seqx <- sdx * seqx + mx
  seqy <- sdy * seqy + my
  result <- list(seqx = seqx, seqy = seqy, den = den)
  result
}
chiplot<-function(x,y,vlabs=c("X","Y"),matrix="NO") {
n<-length(x)
ind<-numeric(length=n)
for(i in 1:n) {
  for(j in (1:n)[-i]) if(x[i]>x[j]&y[i]>y[j]) ind[i]<-ind[i]+1
}
ind<-ind/(n-1)
#
ind1<-numeric(length=n)
for(i in 1:n) {
  for(j in (1:n)[-i]) if(x[i]>x[j]) ind1[i]<-ind1[i]+1
}
ind1<-ind1/(n-1)
#
ind2<-numeric(length=n)
for(i in 1:n) {
  for(j in (1:n)[-i]) if(y[i]>y[j]) ind2[i]<-ind2[i]+1
}
ind2<-ind2/(n-1)
s<-sign((ind1-0.5)*(ind2-0.5))
chi<-(ind-ind1*ind2)/sqrt(ind1*(1-ind1)*ind2*(1-ind2))
lambda<-4*s*pmax((ind1-0.5)^2,(ind2-0.5)^2)
thresh<-4*(1/(n-1)-0.5)^2
if(matrix=="NO") {
  par(mfrow=c(1,2))
  plot(x,y,xlab=vlabs[1],ylab=vlabs[2])
  plot(lambda[abs(lambda)<thresh],chi[abs(lambda)<thresh],ylim=c(-1,1),xlab="lambda",
  ylab="Chi")
  abline(h=1.78/sqrt(n))
  abline(h=-1.78/sqrt(n))
}
if(matrix=="YES") {
  plot(lambda[abs(lambda)<thresh],chi[abs(lambda)<thresh],ylim=c(-1,1))
  abline(h=1.78/sqrt(n))
  abline(h=-1.78/sqrt(n))
}
#

```

```

bvbox<-function(a, d = 7, mtitle = "Bivariate Boxplot",
method = "robust",xlab="X",ylab="Y")
{
  p <- length(a[, 1])
  if(method == "robust") {
    param <- biweight(a[, 1:2])
    m1 <- param[1]
    m2 <- param[2]
    s1 <- param[3]
    s2 <- param[4]
    r <- param[5]
  }
  else {
    m1 <- mean(a[, 1])
    m2 <- mean(a[, 2])
    s1 <- sqrt(var(a[, 1]))
    s2 <- sqrt(var(a[, 2]))
    r <- cor(a[, 1:2])[1, 2]
  }
  x <- (a[, 1] - m1)/s1
  y <- (a[, 2] - m2)/s2
  e <- sqrt((x * x + y * y - 2 * r * x * y)/(1 - r * r))
  e2 <- e * e
  em <- median(e)
  emax <- max(e[e2 < d * em * em])
  r1 <- em * sqrt((1 + r)/2)
  r2 <- em * sqrt((1 - r)/2)
  theta <- ((2 * pi)/360) * seq(0, 360, 3)
  xp <- m1 + (r1 * cos(theta) + r2 * sin(theta)) * s1
  yp <- m2 + (r1 * cos(theta) - r2 * sin(theta)) * s2
  r1 <- emax * sqrt((1 + r)/2)
  r2 <- emax * sqrt((1 - r)/2)
  theta <- ((2 * pi)/360) * seq(0, 360, 3)
  xpp <- m1 + (r1 * cos(theta) + r2 * sin(theta)) * s1
  ypp <- m2 + (r1 * cos(theta) - r2 * sin(theta)) * s2
  maxx1 <- max(xpp)
  minx1 <- min(xpp)
  maxy1 <- max(ypp)
  miny1 <- min(ypp)
  b1 <- (r * s2)/s1
  a1 <- m2 - b1 * m1
  y1 <- a1 + b1 * minx1
  y2 <- a1 + b1 * maxx1
  b2 <- (r * s1)/s2
  a2 <- m1 - b2 * m2
  x1 <- a2 + b2 * miny1
  x2 <- a2 + b2 * maxy1
  maxx <- max(c(a[, 1], xp, xpp, x1, x2))
  minx <- min(c(a[, 1], xp, xpp, x1, x2))
  maxy <- max(c(a[, 2], yp, ypp, y1, y2))
  miny <- min(c(a[, 2], yp, ypp, y1, y2))
  plot(a[, 1], a[, 2], xlim = c(minx, maxx), ylim = c(miny, maxy), xlab =xlab, ylab =ylab,
  lwd = 2, pch = 1)
  text(a[, 1]+ 0.02*(max(a[,1])- min(a[,1])), a[, 2],as.character(1:length(a[,1])),cex=0.75)
  lines(xp, yp, lwd = 2)
  lines(xpp, ypp, lty = 2, lwd = 2)
  segments(minx1, y1, maxx1, y2, lty = 3, lwd = 2)
  segments(x1, miny1, x2, maxy1, lty = 4, lwd = 2)
}
biweight<-function(a,const1=9,const2=36,err=0.0001) {
#a is data matrix with two cols.
#const1=common tuning constant
#const2=bivariate tuning constant
#err=convergence criterion.
  x<-a[,1]
  y<-a[,2]
  n<-length(x)
  mx<-median(x)
  my<-median(y)
  madx<-median(abs(x-mx))

```

```

mady<-median(abs(y-my))
if(madx != 0) { ux<-(x-mx)/(const1*madx)
  ux1<-ux[abs(ux)<1]
  tx<-mx+(sum((x[abs(ux)<1]-mx)*(1-ux1*ux1)^2)/
    sum((1-ux1^2)^2))
  sx<- sqrt(n)*sqrt(sum((x[abs(ux)<1]-mx)^2*
    (1-ux1*ux1)^4))/abs(sum((1-ux1*ux1)*
    (1-5*ux1*ux1)))
  }
else { tx<-mx
  sx<-sum(abs(x-mx))/n
  }
if(mady != 0) { uy<-(y-my)/(const1*mady)
  uy1<-uy[abs(uy)<1]
  ty<-my+(sum((y[abs(uy)<1]-my)*(1-uy1*uy1)^2)/
    sum((1-uy1^2)^2))
  sy<- sqrt(n)*sqrt(sum((y[abs(uy)<1]-my)^2*
    (1-uy1*uy1)^4))/abs(sum((1-uy1*uy1)*
    (1-5*uy1*uy1)))
  }
else { ty<-my
  sy<-sum(abs(y-my))/n
  }
z1<-(y-ty)/sy+(x-tx)/sx
z2<-(y-ty)/sy-(x-tx)/sx
mz1<-median(z1)
mz2<-median(z2)
madz1<-median(abs(z1-mz1))
madz2<-median(abs(z2-mz2))
if(madz1 != 0) { uz1<-(z1-mz1)/(const1*madz1)
  uz11<-uz1[abs(uz1)<1]
  tz1<-mz1+(sum((z1[abs(uz1)<1]-mz1)*(1-uz11*uz11)^2)/
    sum((1-uz11^2)^2))
  sz1<- sqrt(n)*sqrt(sum((z1[abs(uz1)<1]-mz1)^2*
    (1-uz11*uz11)^4))/abs(sum((1-uz11*uz11)*
    (1-5*uz11*uz11)))
  }
else { tz1<-mz1
  sz1<-sum(abs(z1-mz1))/n
  }
if(madz2 != 0) { uz2<-(z2-mz2)/(const1*madz2)
  uz21<-uz2[abs(uz2)<1]
  tz2<-mz2+(sum((z2[abs(uz2)<1]-mz2)*(1-uz21*uz21)^2)/
    sum((1-uz21^2)^2))
  sz2<- sqrt(n)*sqrt(sum((z2[abs(uz2)<1]-mz2)^2*
    (1-uz21*uz21)^4))/abs(sum((1-uz21*uz21)*
    (1-5*uz21*uz21)))
  }
else { tz2<-mz2
  sz2<-sum(abs(z2-mz2))/n
  }
esq<-((z1-tz1)/sz1)^2+((z2-tz2)/sz2)^2
w<-numeric(length=n)
c2<-const2
for(i in 1:10) {
w[esq<const2]<-(1-esq[esq<const2]/const2)^2
w[esq>=const2]<-0
l<-length(w[w==0])
if(l<0.5*n) break
else const2<-2*const2
}
tx<-sum(w*x)/sum(w)
sx<-sqrt(sum(w*(x-tx)^2)/sum(w))
ty<-sum(w*y)/sum(w)
sy<-sqrt(sum(w*(y-ty)^2)/sum(w))
r<-sum(w*(x-tx)*(y-ty))/(sx*sy*sum(w))
const2<-c2
wold<-w
for(i in 1:100) {
  z1<-(y-ty)/sy+(x-tx)/sx/sqrt(2*(1+r))

```

```

z2<-(y-ty)/sy-(x-tx)/sx/sqrt(2*(1+r))
esq<-z1*z1+z2*z2
for(j in 1:10) {
  w[esq<const2]<-(1-esq[esq<const2]/const2)^2
  w[esq>=const2]<-0
  l<-length(w[w==0])
  if(l<0.5*n) break
  else const2<-2*const2
}
tx<-sum(w*x)/sum(w)
sx<-sqrt(sum(w*(x-tx)^2)/sum(w))
ty<-sum(w*y)/sum(w)
sy<-sqrt(sum(w*(y-ty)^2)/sum(w))
r<-sum(w*(x-tx)*(y-ty))/(sx*sy*sum(w))
term<-sum((w-wold)^2)/(sum(w)/n)^2
if(term<-err) break
else {wold<-w
const2<-c2
}
}
param<-c(tx,ty,sx,sy,r)
param
}
panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col="red", ...)
}

panel.reg <- function(x,y)
{
  abline(lsf(x,y)$coef,lwd=2)
  lines(lowess(x,y),lty=2,lwd=2)
  points(x,y)
}
### Aquí termina procedimiento multivariado
#####
library(xlsReadWrite)
datos<-read.xls(paste("hogp65.xls"),sheet=1)
nr <- nrow(datos)
nc <- ncol(datos)
datos<- as.matrix(datos[,3:15])
#datos<-log(datos)### Caso transformados a logaritmos
summary(datos)

pairs(datos,main="Gráfica multivariada",diag.panel=panel.hist,panel=panel.reg)
cor(datos)
round(cor(datos),2)
###
library(TeachingDemos)
faces(datos, main="Hogares exclusiv. pob 65 y más años") ### grafica de caras o imagenes
# Prueba de normalidad univariada
for(i in 1:nc)
{
  print(c(i,dimnames(datos)[[2]][i]))
  print(shapiro.test(datos[,i]))
}
par(mfrow = c(3, 5))
nomc <-
c("Carnes","Cereales","Verduras","Leche","Otros_Ali","Frutas","Pescado","Huevo","Azúcar","Aceites","Tuberculo","Café","Especias")
for(i in 1:nc)
{
  qqnorm(datos[,i],main=nomc[i])
  qqline(datos[,i],col=2,lwd=3,lty=3)
}

```

```

# Prueba de normalidad multivariada
###datos<-as.numeric(datos)
nr <- nrow(datos)
nc <- ncol(datos)
dd <- matrix(0,nr,nr)
mm <- colMeans(datos)
sii <- solve(var(datos))
#####
for(i in 1:nr)
  for(j in 1:nr)
    dd[i,j] <- (datos[i,]-mm)%*%sii%*(datos[j,]-mm)

Ap <- sum(dd^3)/(nr*nr)
Kp <- sum(diag(dd)^2)/nr

1 - pchisq(nr*Ap/6, nc*(nc+1)*(nc+2)/6)
1 - pnorm(Kp,nc*(nc+2), 8*nc*(nc+2)/nr)
cor(datos)
#####
# Detectando outliers
si <- solve(var(datos))
mm <- colMeans(datos)
dd <- rep(0,nr)
for(i in 1:nr)
  dd[i] <- t(datos[i,]-mm)%*%si%*(datos[i,]-mm)
which(dd > 9)
dd <- sort(dd)
ii <- ((1:nr)-0.5)/nr
qq <- qchisq(ii,nc)
plot(qq,dd,xlab="Qc,4 (j-0.5)/nr",ylab="d^2(j)",main="Grafica de probabilidad ji-cuadrada")
abline(coef=c(0,1))
boxplot(datos,notch=TRUE, ps=0.5, COL="blue", adjust=0.2, main="Variables de gasto alimentario", font=5)
##### para obtener la matriz del teorema 3
(PPP1<-eveS%*%sqrt(evaS)%*%1/sqrt(diag(S)))###vector de coeficientes de correlación
(PPP1A<-eveS*sqrt(evaS)*1/sqrt(diag(S)))### matriz de coeficientes de correlación entre
#####variables originales y componentes principales
###
#####F I N

```

3. Cuadros estadísticos.

Puntuaciones de los componentes en los hogares con al menos un adulto mayor (en términos reales)

Parte1

| | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 |
|----|------------|--------|--------|--------|--------|--------|--------|
| 1 | 19.5982 | 0.3119 | 1.1918 | 0.7857 | 0.8596 | 0.6050 | 0.9540 |
| 2 | 2.1142 | 2.0422 | 2.0872 | 0.7314 | 0.7611 | 1.6429 | 0.9493 |
| 3 | 15631.1235 | 5.2968 | 1.9080 | 0.5228 | 1.1600 | 1.3833 | 0.8457 |
| 4 | 44.9446 | 0.6715 | 0.4084 | 1.5935 | 0.9224 | 2.4497 | 0.8340 |
| 5 | 0.8921 | 0.5657 | 0.8186 | 1.6885 | 1.2883 | 1.4806 | 1.7417 |
| 6 | 62.6553 | 1.4010 | 1.0544 | 1.6400 | 0.8753 | 0.4214 | 0.7919 |
| 7 | 0.0883 | 1.9475 | 0.2609 | 0.8312 | 0.9930 | 0.7860 | 0.9179 |
| 8 | 0.2023 | 0.4230 | 1.7682 | 1.2499 | 1.6006 | 0.9577 | 1.2518 |
| 9 | 0.0012 | 1.5713 | 1.6245 | 0.9996 | 0.4824 | 2.0979 | 0.6902 |
| 10 | 3.8044 | 0.6897 | 1.3314 | 1.4834 | 1.3972 | 1.1260 | 1.4842 |
| 11 | 0.1295 | 0.6779 | 1.4697 | 0.5921 | 0.6368 | 0.6313 | 1.0226 |
| 12 | 0.1981 | 4.7278 | 1.0217 | 0.7876 | 1.2744 | 0.3990 | 1.0753 |
| 13 | 1.3127 | 1.1559 | 1.0121 | 0.6257 | 1.6228 | 1.2012 | 0.8491 |
| 14 | 0.0405 | 1.2747 | 0.8806 | 1.0155 | 1.2968 | 1.4263 | 1.1689 |
| 15 | 0.0008 | 1.5351 | 1.2996 | 0.8536 | 0.7509 | 1.0560 | 0.9766 |
| 16 | 0.0646 | 1.1374 | 2.6259 | 0.7785 | 1.4292 | 1.0542 | 0.9722 |
| 17 | 2.7111 | 0.8447 | 1.3751 | 1.0863 | 1.7777 | 1.2267 | 0.4507 |
| 18 | 2.0678 | 1.7457 | 0.7998 | 2.5831 | 0.4063 | 0.9774 | 1.1525 |
| 19 | 0.1033 | 0.7054 | 1.5216 | 1.7584 | 0.7264 | 0.7179 | 0.9055 |
| 20 | 0.0424 | 0.9864 | 0.6298 | 0.5430 | 0.9736 | 1.1047 | 0.9590 |
| 21 | 0.0604 | 0.7731 | 1.1056 | 0.7296 | 1.6760 | 0.8411 | 0.7861 |
| 22 | 16.0037 | 0.8375 | 2.9380 | 1.1595 | 0.4155 | 0.8970 | 0.9776 |
| 23 | 63.2743 | 1.5773 | 0.6084 | 0.8620 | 0.7526 | 0.7906 | 0.9324 |
| 24 | 0.4849 | 0.4244 | 1.2150 | 1.2276 | 1.0467 | 1.1230 | 1.1446 |
| 25 | 1.4142 | 1.7730 | 1.2201 | 1.6384 | 0.9720 | 0.9533 | 1.8852 |
| 26 | 0.6397 | 1.9506 | 1.0986 | 1.6252 | 1.8732 | 1.2382 | 1.2898 |
| 27 | 7.5998 | 1.8799 | 0.4091 | 1.4668 | 1.7347 | 1.1893 | 0.7198 |
| 28 | 0.3252 | 0.7854 | 0.8384 | 2.2950 | 1.1748 | 0.8681 | 1.2090 |
| 29 | 4.7504 | 0.8255 | 0.4152 | 0.1865 | 0.7034 | 1.1718 | 1.8830 |
| 30 | 0.0089 | 1.6426 | 0.5807 | 1.0907 | 1.3816 | 0.7283 | 0.9066 |
| 31 | 2.3740 | 0.1362 | 0.3720 | 0.9469 | 0.6415 | 0.8789 | 0.5904 |
| 32 | 33.6690 | 0.1804 | 1.8714 | 0.6003 | 1.6743 | 0.9564 | 1.1796 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

Puntuaciones de los componentes en los hogares con al menos un adulto mayor (en términos reales)
Parte 2

| | Comp.8 | Comp.9 | Comp.10 | Comp.11 | Comp.12 | Comp.13 |
|----|--------|--------|---------|---------|---------|---------|
| 1 | 1.4944 | 1.3902 | 0.8883 | 0.8513 | 1.0486 | 1.1167 |
| 2 | 0.6753 | 0.8274 | 1.1828 | 1.0834 | 0.9570 | 1.1205 |
| 3 | 0.9404 | 1.5880 | 1.0186 | 0.8466 | 0.9004 | 0.9464 |
| 4 | 0.9288 | 0.8384 | 0.7993 | 1.0340 | 1.0760 | 0.8306 |
| 5 | 1.0616 | 1.0623 | 0.8494 | 0.8339 | 0.9849 | 1.0802 |
| 6 | 0.8276 | 0.8874 | 1.0644 | 0.9136 | 0.9650 | 0.9593 |
| 7 | 0.8834 | 0.8479 | 1.1051 | 0.7863 | 1.0426 | 0.9585 |
| 8 | 0.9324 | 1.3058 | 1.0187 | 1.2174 | 0.9857 | 1.0292 |
| 9 | 1.0508 | 1.0566 | 0.9389 | 1.1164 | 0.8978 | 1.0260 |
| 10 | 1.2598 | 0.9382 | 1.3208 | 0.9476 | 1.0340 | 1.0537 |
| 11 | 0.9055 | 1.0087 | 0.6918 | 0.9343 | 0.9972 | 0.9295 |
| 12 | 0.9043 | 1.0323 | 1.0014 | 1.2238 | 1.0386 | 0.9675 |
| 13 | 0.8337 | 1.0085 | 1.0791 | 0.9724 | 1.0334 | 1.1057 |
| 14 | 0.9250 | 1.5825 | 1.1291 | 1.2905 | 1.2504 | 1.1148 |
| 15 | 1.1854 | 1.0257 | 1.1206 | 1.1033 | 0.8613 | 0.9186 |
| 16 | 1.3201 | 1.1020 | 0.6425 | 0.8316 | 1.0891 | 0.9230 |
| 17 | 0.8666 | 0.9246 | 1.3621 | 0.8614 | 1.2011 | 1.0455 |
| 18 | 1.2426 | 1.1628 | 1.2254 | 0.8445 | 1.0878 | 0.9930 |
| 19 | 0.5202 | 0.7906 | 1.0305 | 0.9404 | 0.8420 | 1.0962 |
| 20 | 1.4112 | 0.7650 | 1.1338 | 0.6452 | 0.9407 | 1.0820 |
| 21 | 1.3838 | 0.7529 | 1.1691 | 1.2814 | 1.2985 | 0.9146 |
| 22 | 1.2968 | 0.7445 | 0.9667 | 1.0762 | 1.1138 | 0.8523 |
| 23 | 1.5128 | 0.6980 | 1.2369 | 1.4051 | 0.9669 | 1.3414 |
| 24 | 0.6972 | 0.8925 | 0.7517 | 0.7597 | 1.1272 | 1.1282 |
| 25 | 0.8601 | 1.2321 | 1.0267 | 1.1858 | 1.0907 | 1.0097 |
| 26 | 1.8093 | 0.8474 | 1.1318 | 0.9875 | 0.7703 | 0.9602 |
| 27 | 1.1430 | 0.9596 | 0.6481 | 1.3007 | 0.9340 | 0.9823 |
| 28 | 0.7965 | 0.9339 | 1.1269 | 0.9886 | 0.9936 | 0.8616 |
| 29 | 0.7892 | 0.8910 | 1.1037 | 1.0215 | 1.0210 | 0.9144 |
| 30 | 0.9669 | 1.5952 | 0.9340 | 0.8630 | 0.9492 | 0.9839 |
| 31 | 1.1626 | 1.4554 | 1.1658 | 1.1673 | 0.8649 | 0.9736 |
| 32 | 0.8325 | 0.9192 | 1.1157 | 1.2111 | 0.8475 | 0.9377 |

Fuente: INEGI. Encuesta Nacional de Ingreso y Gasto de los Hogares 2006. Bases de datos.

4. Notas.

A. Componentes principales, parte II. (Un resumen)

Definición:

Una matriz simétrica A tiene un valor característico o eigenvalor λ y un vector característico o eigenvector asociado \tilde{x} si

$$A\tilde{x} = \lambda\tilde{x}, \text{ para } \tilde{x} \neq \tilde{0}$$

De esto se tiene que

$$A\tilde{x} - \lambda\tilde{x} = \tilde{0} \Leftrightarrow (A - \lambda I)\tilde{x} = \tilde{0}$$

Los eigenvalores de A se obtienen encontrando las raíces del polinomio que resulta de

$$|A - \lambda I| = 0$$

Los eigenvectores de A se obtienen resolviendo la ecuación

$$(A - \lambda_i I)\tilde{x} = \tilde{0} \quad (i = 1, \dots, p)$$

donde p es el número de valores característicos.

El análisis de componentes principales (ACP) es una técnica utilizada para explicar la estructura de las varianzas con pocas combinaciones lineales de las variables de insumo. Se pretende que la varianza del sistema original de variables sea explicada, casi toda, por un pequeño número de componentes.

Las componentes principales son variables incorrelacionadas que tienen sentido cuando las variables de insumo sí están correlacionadas.

La primera componente contribuye con la máxima cantidad para explicar la suma de las varianzas de las p variables. La segunda componente hace lo mismo con la varianza residual, después de que el efecto de la primera es removido de los datos; y así

sucesivamente hasta que se obtiene toda la varianza de los datos originales, es decir, se tendrán tantas componentes como variables observadas se hayan considerado. No obstante, en la práctica solamente se retienen unas cuantas componentes principales, aquellas que expliquen el mayor porcentaje de la varianza total (cuando juntas acumulen alrededor del 70%).

Dado que la primera componente principal explica la mayor proporción de varianza, suele utilizarse como un índice cuyo nombre es acorde a la naturaleza del problema, y puede servir como variable de estratificación. La gráfica de esta componente contra la segunda permite visualizar las clasificaciones de individuos u objetos, donde los puntos cercanos indican características similares.

Resultado 1:

Si \tilde{X} es un vector aleatorio cuya distribución tiene un vector de medias poblacionales $\tilde{\mu}$ y matriz de varianzas y covarianzas $\tilde{\Sigma}$. Adicionalmente, si de $\sum \tilde{e}_i = \lambda_i \tilde{e}_i$ se obtiene $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, entonces la i-ésima componente principal es

$$Y_i = \tilde{e}_i^t \tilde{X} = \sum_{j=1}^p e_{ij} X_j$$

donde \tilde{e}_i son vectores ortonormales.

Además,

$$\text{Var}(Y_i) = \lambda_i, \text{Cov}(Y_i, Y_j) = 0 \quad (i \neq j; i, j = 1, \dots, p)$$

Resultado 2:

Bajo las mismas condiciones anteriores,

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

La proporción de la variabilidad total debido a la i-ésima componente principal es:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad (i = 1, \dots, p)$$

También es posible realizar un ACP con base en la matriz de correlaciones, en tal situación la proporción de la varianza explicada por Y_i es:

$$\frac{\lambda_i}{p} \quad (i = 1, \dots, p)$$

Por lo anterior, el índice con base en la primera componente principal viene dado como:

$$Y_1 = e_{11} x_1 + e_{21} x_2 + \dots + e_{p1} x_p$$

Para el caso de utilizar la matriz de varianzas y covarianzas los valores de X se dan en las unidades originales, mientras que si se utiliza la matriz de correlaciones se toman los valores estandarizados.

B. APLICACIONES DEL ANÁLISIS DE COMPONENTES PRINCIPALES

En la actualidad la técnica de análisis de componentes principales tiene aplicación prácticamente en cualquier ciencia experimental. Algunas aplicaciones son las siguientes.

Exploración de datos

Reducir la dimensionalidad y facilitar el análisis de un conjunto de datos multidimensionales (describir un conjunto de datos).

Análisis de regresión

Cuando las variables predictoras son correlacionadas se puede presentar el problema de la multicolinealidad (algunas variables son aproximadamente combinación lineal de otras) y, por tanto, la estimación de los coeficientes puede alejarse de la realidad. Una solución razonable es utilizar CP en sustitución de todas o algunas de las variables independientes.

Construcción de un indicador

La primera CP puede utilizarse como un índice para resumir las características generales del fenómeno estudiado.

Estudios morfológicos

Para determinar el tamaño y forma de individuos u objetos.

Selección de variables

Para seleccionar un subconjunto de variables aproximadamente independientes de un conjunto mayor de variables correlacionadas.

Análisis de factores

El ACP es uno de los métodos utilizados para la extracción de factores principales.

Análisis de conglomerados

Al graficar la primera contra la segunda CP se pueden observar grupos de unidades homogéneas. Por otro lado, la primera CP también puede utilizarse como variable de estratificación.

Control de calidad

Para decidir si algún artículo de un lote es defectuoso o no con base en un vector de dimensiones. En caso de que éstas se encuentren correlacionadas, entonces puede suceder que el error tipo I (es decir, la probabilidad de parar un proceso cuando en realidad funciona bien) no sea el real. Una solución de esto es utilizar CP, ya que son independientes.

C. Relación entre Análisis de Componentes principales (ACP), Análisis de Factores (AF) y Análisis de Conglomerados (AC)

ACP vs. AF

El AF se apoya en el ACP para extraer los factores o variables latentes (hay otras técnicas de extracción).

ACP vs. AC

El AC puede ser aplicado a la 1ª CP para conformar grupos de unidades homogéneas. Esto es de mucha aplicación porque, además, el puntaje da idea de las diferencias entre las unidades analizadas (cuando la 1ª CP extrae una “buena” cantidad de información de las variables originales). Cuando se toman dos o más CP también se pueden hacer clasificaciones de unidades, similar al caso de AC, sólo que este último es preferible, sobre todo para grandes conjuntos de datos.

AF vs. AC

Mientras el AC sirve para clasificar individuos u objetos (para el caso del INEGI casi siempre son unidades geográficas), el AF sirve para agrupar variables.

NOTA: Debemos tener cuidado en el uso de estas técnicas, por ejemplo, el AF también ha sido utilizado para clasificar individuos u objetos y para generar “scores”, así lo documentan algunos libros de texto; pero esto era en el pasado, cuando estaban en proceso de desarrollo otras técnicas.

GLOSARIO



Glosario

Cuartiles. Son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales. **Cálculo de los cuartiles.** Se ordenan los datos de menor a mayor y se busca el lugar que ocupa el cuartil mediante la siguiente expresión:

$$\frac{k \cdot N}{4}, k = 1, 2, 3$$

Dato extraño, atípico o aberrante (outlier). Es un registro mayor o menor de lo esperado que se detecta por tener un residuo que es un valor “inusual”, muy grande o muy pequeño en relación con la distribución asociada a los residuos.

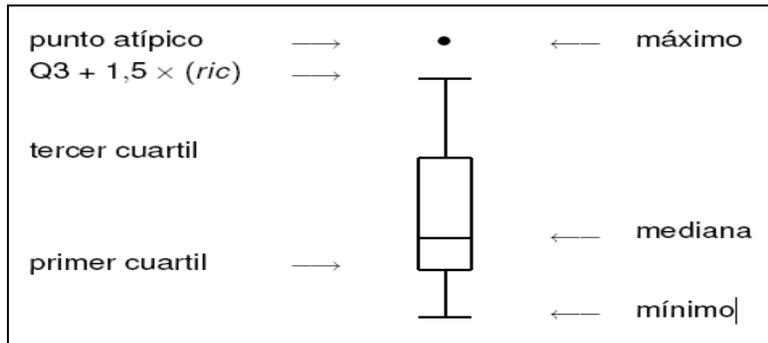
Diagrama de caja y bigotes (Box-and-Whisker). Es un resumen visual de la distribución (comportamiento) de una variable que provee detalles acerca de si uno o ambos extremos de la distribución contienen valores inusualmente grandes o pequeños. El rango intercuartil se usa para construir el diagrama de caja o “boxplot”.

Este tipo de diagramas tiene varias versiones, pero en general se representa el rango intercuartil por una caja o rectángulo, de modo que los extremos del rectángulo están ubicados en el primer y tercer cuartil, tal como se indica en la figura 3.1.

Dentro del rectángulo se indica por una línea o un punto la ubicación de la mediana. Fuera del rectángulo se dibujan dos segmentos, llamados ‘bigotes’ que llegan hasta los datos más lejanos que estén a una distancia menor o igual a 1,5 x

(**ric**) del rectángulo, donde **ric** representa el rango intercuartil. Cualquier punto que no esté incluido en este rango se representa individualmente y se considera un punto atípico (outlier).

Diagrama de caja o boxplot



Prueba de Shapiro-Wilks. Cuando la muestra es como máximo de tamaño 50 se puede contrastar la normalidad con la prueba de Shapiro-Wilks. Para efectuarla se calcula la media y la varianza muestral, S^2 , y se ordenan las observaciones de menor a mayor.

A continuación se calculan las diferencias entre: el primero y el último; el segundo y el penúltimo; el tercero y el antepenúltimo, etc. y se corrigen con unos coeficientes tabulados por Shapiro y Wilk. El estadístico de prueba es:

$$W = \frac{D^2}{nS^2}$$

donde D es la suma de las diferencias corregidas.

BIBLIOGRAFÍA

Bibliografía

1. INEGI (2006). Encuesta Nacional de Ingreso y gasto de los Hogares. Bases de datos.
2. INEGI (2006). Encuesta Nacional de Ocupación y Empleo, IV trimestre de 2006.
3. INEG. Censos de Población y Vivienda, varios años.
4. INEGI (2005). Censo de Población y Vivienda. Estados Unidos Mexicanos. Tabulados Definitivos.
5. HAM CHANDE, ROBERTO (2001). Vejez y dependencia / Paradigmas y nuevos contratos sociales. Demos. Carta demográfica sobre México. DemoS, No. 014.
6. HAM CHANDE, ROBERTO (1993). **Envejecimiento demográfico y seguridad social / la insuficiencia de las pensiones por vejez**, DemoS, No. 006.
7. MONTES DE OCA, VERONICA (1996). **La tercera edad / situaciones sociales de los viejos**, DemoS, No. 009.
8. CAMPOSORTEGA CRUZ, SERGIO (1988). **Dinámica y estructura de la población / El inicio de una nueva era demográfica**, DemoS, No. 001.
9. PARTIDA BUSH, VIRGILIO (1991). **La sobrevivencia de los viejos / vivir más cuesta más**, DemoS, No. 004.
10. CANALES, ALEJANDRO I. (2001). **Hacia el envejecimiento demográfico. De la transición demográfica al envejecimiento de la población**. DemoS, No. 014.
11. GOMES D'ACONCEICAO, MARIA CRISTINA (2001). **Desigualdad social de la vejez / condiciones socioeconómicas de la tercera edad**, DemoS, No. 014.
12. MONTEERRUBIO GOMEZ, MA. ISABEL; LOZANO ASCENCIO, RAFAEL (2001). **Salud y Vejez. Utilización y disponibilidad de los servicios de salud en México**. DemoS, No. 014.
13. ORDORICA, MANUEL (2001). **Supervivencia y muerte en la población mayor / grandes cambios en las causas de muerte de los mayores**, DemoS, No. 014.
14. NEGRETE SALAS, MARIA EUGENIA (2001). **Envejecimiento y migración / distribución geográfica de la población mayor**, DemoS, No. 014.
15. WELTI CHANES, CARLOS (2001). **Economía y envejecimiento / cambios socioeconómicos y sobrevivencia de la población mayor**, DemoS, No. 014.
16. PALMA CABRERA, JOSE LUIS (2002). **El apoyo familiar / transferencia de y para la población mayor dentro y fuera del hogar**, DemoS, No. 015.
17. Green, P. E. (1988). **Analyzing Multivariate data**. The Drydes Press, Illinois.

18. Peña, D. (2002). **Análisis de datos Multivariantes**. Mc Graw Hill. Madrid.
19. Johnson, R.; Wichern, D. (1999). **Applied multivariate statistical analysis**. Prentice Hall.
20. **Economía del Envejecimiento**. www.eumed.net/cursecon/2/vejez.htm
21. **Tercera Edad**. <http://es.wikipedia.org/>
22. http://www.estadisticaparatodos.com/index_archivos/page0004.htm
23. **Tema 1. Introducción al análisis exploratorio de datos**. <http://www.unizar.es/ice/rec-info/spss/Introduccion-AED.pdf>
24. **The R Foundation for Statistical Computing**. <http://www.r-project.org/>
25. Vences Rivera, José. **Componentes principales**. Un resumen. Notas.
26. Maestría en estadística Oficial. CIMAT-INEGI (2008). **Notas, apuntes y procedimiento en R de la clase Modelos Estadísticos III**.
27. Parra Olivares, Javier (2008). Análisis Exploratorio y análisis confirmatorio de datos. http://www2.scielo.org.ve/scielo.php?script=sci_arttext&pid=S1315-00062002000100007&lng=es&nrm=iso
28. Sin referencia de autor. (2009). Homogeneidad de los errores. Datos atípicos. http://dm.udc.es/asignaturas/estadistica2/sec4_6.html