03043

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

INSTITUTO DE INVESTIGACIONES EN
MATEMATICAS APLICADAS Y EN SISTEMAS

APLICACION DE METODOS MULTIVARIADOS AL ESTUDIO DEL DESARROLLO SOCIAL EN MEXICO-EL CASO DE CHIAPAS, DISTRITO FEDERAL, ESTADO DE MEXICO Y NUEVO LEON

T E S I N A

QUE PARA OBTENER EL DIPLOMA EN LA ESPECIALIZACION EN

ESTADISTICA APLICADA

PRESENTA:

RAYMUNDO RAMIREZ GOMEZ

DIRECTORA DE LA TESINA:

M. EN C. MARIA ESTHER PEREZ TREJO

MEXICO, D. F. **DICIEMBRE 2004**

A Consuelo, Rodrigo, Alexandra y Diego

Agradecimiento

El presente trabajo fue posible gracias a la orientación y comentarios de mis maestros del I.I.M.A.S: Dr. Ignacio Méndez Ramírez; Dra. Rebeca Aguirre Hernández; M. en C. Patricia Romero Mares; M. en C. Salvador Zamora Muñoz y M. en C. María Esther Pérez Trejo, directora de la tesina. A todos ellos mi admiración y agradecimiento.

RRG

Contenido

Prólogo 4
Objetivo 5
Herramientas Estadísticas 5
Metodología 6
Capítulo 1 7
Introducción 7
Desarrollo frente a crecimiento 7
Antecedentes en el Estudio del Desarrollo Humano 10
Variables Observadas 12
Capítulo 2 16
Resultados del Análisis Estadístico del Desarrollo Social en México 16
2.1 Análisis de Correlación 16
2.2 Análisis de Componentes Principales 17
2.3 Análisis de Factores 21
INTERPRETACIÓN DE LOS FACTORES (Método de Factores Principales) ... 24
CALIFICACIÓN DE LOS FACTORES 26
2.4 Análisis de Conglomerados 27
Conclusiones 37
Capítulo 3 41
Descripción de la Herramientas Estadísticas 41
Empleadas 41
3.1 Análisis de Componentes Principales 41
INTRODUCCIÓN 41
COMPONENTES PRINCIPALES POBLACIONALES 41
Determinación del Número de Componentes Principales 44
Componentes Principales a partir de Variables Estandarizadas 45
Determinación del Número de Componentes para la Matriz de Correlación ρ 47
COMPONENTES PRINCIPALES A PARTIR DE DATOS MUESTRALES 47
Posibles Interpretaciones de las Componentes Principales 48
3.2 Análisis de Factores 49
INTRODUCCIÓN 49
MODELO DE ANÁLISIS DE FACTORES 50
MODELO ORTOGONAL DE FACTORES CON m FACTORES COMUNES .. 51
ESTRUCTURA DE COVARIANZA DEL MODELO ORTOGONAL DE FACTOR 52
Posibles Problemas Numéricos del Modelo de Factores 54
Método de Factores Principales 55
SOLUCIÓN DE FACTORES PRINCIPALES DEL MODELO DE FACTORES 58
Un Enfoque Modificado – la Solución por Factores Principales 59
Método de Máxima Verosimilitud 61
Elementos Adicionales a Considerar en la Elección del Número de Factores 63
ROTACIÓN DE FACTORES 63
CALIFICACIONES DE LOS FACTORES 65
Método de Mínimos Cuadrados Ponderados 66
Método de Regresión 68
PERSPECTIVAS Y ESTRATEGIA DEL ANÁLISIS DE FACTORES 69

3.3 Análisis de Conglomerados 69
INTRODUCCIÓN 69
MEDIDAS DE SIMILITUD (O PROXIMIDAD) 70
MÉTODOS JERÁRQUICOS DE AGRUPAMIENTO 73
Liga Completa 76
Liga Promedio 76
MÉTODOS NO JERÁRQUICOS DE AGRUPACIÓN 77
Método de K-Medias 78
Bibliografía 80
Anexos 82
Anexo 1: Base de datos 82
Anexo 2: Descripción de las variables derivadas 82
Anexo 3: Matrices de residuos 82

Prólogo

"El desarrollo humano entraña mucho más que el simple aumento o disminución del ingreso nacional. Significa crear un entorno en el que las personas puedan hacer plenamente realidad sus posibilidades y vivir en forma productiva y creadora de acuerdo con sus necesidades e intereses"[1]

A pesar de la opulencia alcanzada en muchas regiones del globo durante el último siglo; el avance en cuanto a derechos humanos y libertades políticas, así como el incremento en la longevidad de la población y acceso a servicios de salud y seguridad social, la población mundial aún sufre de un gran número de privaciones. De acuerdo al Programa de las Naciones Unidas para el Desarrollo (PNUD, 2001), de los 4,800 millones de habitantes en países en desarrollo, más de 850 millones son analfabetos, casi 1,000 millones carecen de fuentes de agua potable y 2,400 millones no tienen acceso a servicios sanitarios básicos. Cerca de 325 millones de niños no asisten a la escuela. Además 11 millones de niños menores de 5 años mueren cada año, es decir, más de 30,000 niños cada día, por causas que podrían evitarse. Alrededor de 1,200 millones de personas viven con menos de 1 dólar al día (PPA en dólares de EE.UU. de 1993[2]). Estas privaciones no se limitan a los países en desarrollo. En los miembros de la OCDE más de 130 millones de personas padecen de pobreza de ingreso, 34 millones se encuentran desempleados y la tasa media de analfabetismo funcional de adultos, alcanza el 15 %.

En el caso de México, la desigualdad en el ingreso y desarrollo humano entre la población, sigue siendo abismal. Cifras oficiales revelan que 45.9 % de los hogares están en situación de pobreza, al obtener un ingreso per cápita de alrededor de 2 dólares diarios. En contraste, el 10 % de los hogares más ricos alcanzaron un ingreso per cápita de al menos 26 dólares al día. En el 10 % de los hogares más ricos, la tasa de mortalidad infantil es de 1.4 %, mientras que en el 10 % más pobre es 3.1 %. El 10 % de los hogares más pobres tiene un porcentaje de personas de más de 12 años sin instrucción igual a 19.3 %, contrastando con 1.1 % en el 10 % de la población de mayores ingresos.[3]

Superar estos problemas ha sido un aspecto central en el ejercicio del desarrollo. Amartya Sen[4], ganador del premio Nóbel en Economía en 1998, sostiene que el desarrollo humano consiste en la ampliación de las "*libertades*" de la persona humana; es decir de su capacidad individual y "*poder para*" procurarse una vida plena y saludable, de acuerdo a sus propios valores.

---

[1] Programa de las Naciones Unidas para el Desarrollo (PNUD), Informe Sobre Desarrollo Humano (2001)
[2] PPA, dólares. Paridades del poder adquisitivo. Determinan el número de unidades de la moneda de un país necesarias para adquirir la misma canasta representativa de bienes y servicios que un dólar EEUU adquiriría en los Estados Unidos de Norteamérica. El PPA permite hacer comparación del nivel real de precios entre países de la misma manera que los índices convencionales de precios permiten hacer comparaciones del valor real en el tiempo; de otra manera, el tipo de cambio normal puede sobrevalorar o subvalorar el poder adquisitivo.
[3] Programa de las Naciones Unidas para el Desarrollo, Informe Sobre Desarrollo Humano México (2002)
[4] Amartya Sen. 1999. *Development as Freedom*. Alfred A. Knopf, New York

Por otra parte, esta capacidad individual ineludiblemente se encuentra calificada y constreñida por las oportunidades económicas, políticas y sociales a que tenga acceso la persona. Sen señala la importancia en el reconocimiento tanto de la libertad individual como de la influencia que las fuerzas sociales ejercen, en la extensión y grado de expresión de las libertades humanas. Para mitigar y en última instancia remediar los problemas que aquejan a la humanidad, la expansión de la libertad individual debe constituirse como compromiso social. Bajo el enfoque de Sen, el acrecentar las libertades humanas, representa tanto el fin primordial, como el medio principal para alcanzar el desarrollo. El desarrollo consiste en la eliminación de diferentes tipos de restricciones en las libertades de la persona y que la limitan en su capacidad de elección y en el ejercicio de sus capacidades. La eliminación de estas restricciones, argumenta Sen, es precisamente el elemento constitutivo del desarrollo.

La relación existente entre desarrollo y libertad se entiende de manera más plena mediante la comprensión de las interrelaciones existentes entre diferentes clases de libertades y la capacidad instrumental de algunas de éstas para generar o promover a la vez, otro tipo de libertades. Por ejemplo, existe suficiente evidencia empírica en cuanto a que las libertades políticas y económicas se refuerzan entre sí. Por otra parte, las libertades sociales de educación y cuidado de la salud complementan en la persona las oportunidades de participación económica y política. A fin de guiar el proceso de desarrollo de manera integral y coherente, es necesario encontrar las interrelaciones empíricas existentes entre las distintas clases de libertades y los arreglos o instituciones disponibles en la sociedad para su mayor y más conveniente expansión.

Sen se concentra particularmente en la función e interconexiones existentes entre ciertas libertades instrumentales, incluyendo *oportunidades económicas, libertades políticas, facilidades sociales, garantías de transparencia y protección civil.* En la procuración de estas libertades, se examina una amplia gama de organismos y arreglos institucionales, incluyendo al Estado, los mercados de bienes y servicios, el sistema jurídico, los partidos políticos, los medios, las organizaciones no gubernamentales y foros públicos de discusión, entre otros.

Objetivo

En el presente trabajo, se tiene como objetivo, construir, bajo el enfoque de Sen, mediante el estudio de una muestra de cuatro entidades federales, a nivel municipal, una definición operacional de *Desarrollo Social*, aplicable al caso de México. Lo anterior, en base a las libertades enunciadas en el párrafo anterior, las cuales no son directamente observables, empleando para esto, variables de fácil medición directa e información estadística disponible (variables observables).

Herramientas Estadísticas

Con la finalidad anterior, se utilizará el *análisis de factores* como herramienta estadística, previa identificación, mediante la técnica de *componentes principales* de un número reducido de variables ortogonales que representen de manera satisfactoria la variabilidad de las variables originales.

En base a los resultados anteriores, mediante la técnica de *análisis de conglomerados*, se pretende identificar municipios con niveles similares de desarrollo, y establecer diferencias entre éstos, mediante su clasificación y ordenamiento. Será interesante identificar municipios tanto con carencias en mayor o menor grado de libertades

individuales, como aquellos que aún con capacidades desarrolladas, carezcan de las suficientes oportunidades para su actualización.

Finalmente, como corolario del análisis anterior se pretende identificar mediante el uso del análisis multivariado, un conjunto de condiciones tanto coadyuvantes, como limitantes del desarrollo social, así como las interconexiones empíricas existentes entre las diferentes clases de libertades.

Metodología

El estudio es observacional, ya que se emplean las cifras publicadas por INEGI, principalmente el Censo General de Población del 2000. El estudio es asimismo transversal, tomando como unidades de estudio los municipios y delegaciones comprendidos en los estados de Nuevo León, Chiapas, México y Distrito Federal. Estos estados, fueron seleccionados, a fin de incluir en la muestra poblaciones con niveles avanzados de desarrollo (DF y Nuevo León), con nivel medio de desarrollo (Estado de México) y con atraso (Chiapas). Lo anterior, de acuerdo al orden resultante de estudios previos y cuyas conclusiones se resumen en la Tabla 1.1 del presente trabajo. Como variables explicativas se incluyeron las reportadas en el Censo General de Población del 2000, IFE y CONAPO, y que reflejan alguna de las dimensiones de las libertades instrumentales, concomitantes del desarrollo social propuestas por Sen.

La muestra, originalmente consistente en los 305 municipios y delegaciones que conforman a las cuatro entidades bajo estudio, se redujo a 290 unidades, en vista de que no se dispone de información censal para la totalidad. Los 15 municipios faltantes corresponden al estado de Chiapas, en donde existen regiones incomunicadas o que no reunían las condiciones de seguridad necesarias para el levantamiento del censo. La muestra expresa grados de desarrollo que en conjunto, reflejan la enorme diversidad nacional existente tanto en cuanto a capacidades individuales, como estructuras institucionales. El Distrito Federal y el estado de Nuevo León representan a las entidades federativas con mayor desarrollo. Por otra parte, el estado de Chiapas, se clasifica como el de menor desarrollo. Por su parte el Estado de México caracteriza a una entidad con una muy alta dispersión en cuanto a su condición de desarrollo, ya que ésta cuenta tanto con poblaciones con desarrollo humano equiparable al del estado de Chiapas, como con municipios cuyo desarrollo social es comparable con los municipios y delegaciones del Distrito Federal y el estado de Nuevo León. Con base al interés en los resultados obtenidos y la disponibilidad de recursos, el presente estudio puede ampliarse subsecuentemente a las 32 entidades que conforman la federación.

Los resultados para el caso de México no son en su mayoría comparables con los de otros países, ya que existe un sinnúmero de diferencias en la información censal particular de cada país. Baste señalar las existentes no solamente en la calidad de los sistemas educativos en los distintos países, sino también en sus estructuras políticas, económicas y sociales. Así por ejemplo la educación básica en México no es comparable con la de Finlandia, ni el nivel del debate público en nuestras Cámaras con el del Parlamento Inglés. Sin embargo el enfoque metodológico multivariado que aquí se aplica puede servir de base en la realización de estudios en otras poblaciones.

Capítulo 1

Introducción

Desarrollo frente a crecimiento

Crecimiento y desarrollo son dos términos que se confunden con mucha frecuencia, aún cuando no son conceptos análogos. Crecimiento en sentido estricto es una propiedad que tiene que ver con el incremento en tamaño o en número. Puede haber crecimiento sin desarrollo y desarrollo sin crecimiento.

Ackoff (1974, 1981) y Garajedaghi & Ackoff (1986) definen desarrollo en la persona como el proceso mediante el cual ésta incrementa su habilidad y su deseo para satisfacer sus propias necesidades legítimas y las de los demás. Es un incremento en la capacidad y en el potencial de la persona; no en sus posesiones. Es una característica que depende en mayor medida de la motivación, conocimientos y entendimiento del individuo, que de la magnitud de su fortuna o de su nivel de vida. Una persona que obtiene el premio mayor de la lotería no adquiere por este hecho un mayor desarrollo. De la misma manera para México y sus grandes empresas públicas y privadas, el haber tenido acceso a recursos financieros prácticamente ilimitados con motivo del auge petrolero de finales de los setenta y principios de los ochenta, no produjo un mayor desarrollo en sus instituciones y organizaciones.

Lo anterior no significa que el acceso o la disponibilidad de recursos sea irrelevante en mejorar la calidad de vida de los individuos. La medida en que la gente pueda mejorar su calidad de vida dependerá tanto de los recursos disponibles como de su estado o condición de desarrollo. Obviamente se puede hacer más con recursos que sin ellos. Sin embargo, aquellos con mayor grado de desarrollo mejorarán su calidad de vida en mayor medida con menos recursos que aquellos que, aun contando con recursos ilimitados, carezcan de desarrollo. La escasez de recursos puede limitar la calidad de vida de la persona, pero no su desarrollo. El desarrollo es una cualidad para alcanzar una calidad de vida superior y no la calidad de vida en sí misma.

El problema del desarrollo social debe contemplarse desde el punto de vista de los sistemas sociales, ya que los estados y entidades políticas autónomas y soberanas, al estar constituidas por individuos e instituciones con interacciones entre sí y con propósitos propios, actúan como sistemas sociales. Su objeto principal, es el de estimular y facilitar el desarrollo de sus miembros, así como el de los sistemas sociales superiores e instituciones que las contienen y de los cuales son parte. Debido a que el desarrollo de los individuos y de los estados requiere fundamentalmente del aprendizaje, es decir del desarrollo y uso del conocimiento, éste no puede imponerse a los demás, ni puede ser impuesto por los demás. Por lo tanto el auto-desarrollo, es la única clase de desarrollo posible. Para estimular y facilitar el desarrollo de los demás, se debe inducir el aprendizaje y la motivación adecuada. El desarrollo nacional por lo tanto, no es una cuestión de qué hacen los gobiernos por los ciudadanos, sino de qué manera los gobiernos inducen el desarrollo de los gobernados.

Para estimular y facilitar el desarrollo de la sociedad, sus miembros, deben identificar aquellos *ideales* cuya consecución es necesaria para acrecentar sus propias capacidades

y oportunidades. Así por ejemplo, los antiguos filósofos griegos identificaron las siguientes cuatro: la *verdad*, la *abundancia*, el *bien* y la *belleza*.

El desarrollo es una propiedad *emergente* de los sistemas sociales Garajedaghi (1999). Así como el éxito, la calidad y la vida entre otras características, corresponden a manifestaciones del sujeto en su conjunto, el desarrollo en un sistema social, es el resultado del conjunto de procesos e interacciones que tienen lugar en el sistema de manera continua. Para comprender una propiedad emergente, en este caso *el desarrollo*, es necesario comprender los procesos que lo generaron. Si estos procesos cesan, el fenómeno dejara de existir

En ocasiones las propiedades emergentes, como el desarrollo, no pueden medirse directamente; sin embargo pueden medirse algunas de sus manifestaciones.

Amartya Sen (1999), define desarrollo como "el proceso de expansión de las *libertades* de la persona". El enfocarse en las libertades humanas contrasta con puntos de vista estrechos del desarrollo, tales como el crecimiento del producto interno bruto, del ingreso, la industrialización, el avance tecnológico o la modernización de las estructuras sociales. Los aspectos anteriores, si bien representan medios importantes en la expansión de las libertades de la sociedad, ésta adicionalmente depende de otros factores como los sistemas educativos e instituciones para el cuidado de la salud o las garantías existentes en el ejercicio de los derechos civiles (como la libertad de expresión, el derecho al escrutinio de la función pública y al sufragio, etc.). Estas capacidades se amplían mediante la aplicación de políticas congruentes y al mismo tiempo, la calidad y efectividad de la acción pública se enriquece en la medida en que aumentan las propias capacidades de la sociedad en su conjunto. El éxito de una sociedad debe ser evaluado desde este punto de vista, por las libertades sustantivas de que disfruta la sociedad.

La pobreza desde el punto de vista de Sen no corresponde solamente a la privación de un mínimo ingreso necesario para la satisfacción de las necesidades materiales de la persona. La pobreza en su medida más amplia se entiende como la carencia de libertades que limitan al individuo para alcanzar su potencial. Este enfoque requiere por lo tanto de una base informativa mucho más amplia que la simple medición del ingreso. Así por ejemplo el desempleo o la falta de oportunidades no significa para la persona desempleada una mera deficiencia en su ingreso que pueda compensarse mediante las transferencias del estado o subsidios. Es también una fuente de debilitamiento del individuo que afecta su libertad, iniciativa y conocimientos o habilidades. El desempleo contribuye a la exclusión social, a la pérdida de autoestima, confianza en sí mismo y tiene severas repercusiones adversas en la salud.

De la misma manera, un funcionamiento deficiente de los mercados, no solamente restringe el desarrollo económico de la sociedad, sino que atenta adversamente en contra de la libertad del individuo al limitarlo en su capacidad para realizar transacciones en los mercados de bienes y servicios y concurrir espontáneamente al mercado laboral.

Desde el punto de vista de la evaluación del desarrollo, ya que existe una gran diversidad de libertades, el peso relativo de cada una de ellas deberá hacerse explícito en función de los valores y la cultura de la sociedad. Las libertades individuales son un

producto social, existiendo una relación en dos sentidos en cuanto a: (1) el orden social existente para expandir las libertades individuales y (2) el uso de las libertades individuales no solamente para mejorar las respectivas vidas de los individuos, sino para mejorar y tornar más efectivo el orden social.

La perspectiva del desarrollo como libertad tiene máximas implicaciones en el entendimiento del proceso de desarrollo, así como de los medios para conseguirlo. Desde el punto de vista evaluativo esto implica la necesidad de determinar los requerimientos del desarrollo en términos de la eliminación de aquellas carencias de libertad presentes en la sociedad. El proceso de desarrollo bajo este enfoque si bien se relaciona con el proceso mismo de formación de capital físico y humano, su alcance y cobertura se extiende más allá.

La motivación en la propuesta del desarrollo como libertad no es la de establecer un ordenamiento estricto entre las naciones o entre las regiones, sino la de llamar la atención acerca de los aspectos que merezcan mayor importancia dentro del proceso de desarrollo particular de las comunidades bajo estudio.

La expansión de las libertades, como ya se ha comentado, contiene una doble faceta: (1) como fin primordial y (2) como medio principal para alcanzar el desarrollo. Sen las denomina respectivamente "*función constitutiva*" y "*función instrumental*" de la libertad en el proceso de desarrollo. La primera establece intrínsecamente el objetivo primordial del desarrollo y sitúa a la persona en el plano más elevado como sujeto por excelencia del mismo. El valor instrumental de la libertad concierne a la forma en que los distintos tipos de derechos, oportunidades y beneficios contribuyen a acrecentar la libertad humana en general y por lo tanto en la promoción del desarrollo

En particular, Sen investiga cinco clases libertades bajo la perspectiva instrumental. Estas incluyen *(1) libertades políticas, (2) facilidades económicas, (3) oportunidades sociales, (4) garantías de transparencia y (5) seguridad social.*

Las libertades políticas, en términos generales (incluyendo los llamados derechos civiles) se refieren a las oportunidades de los ciudadanos para determinar quién gobierna y bajo qué principios; a la posibilidad de escrutinio y crítica a las autoridades y a la libertad de expresión política y existencia de una prensa sin censura, etc. Algunas variables asociadas a esta categoría incluyen la participación electoral de los ciudadanos, la presencia de medios de comunicación y la transparencia en cuanto a los procesos de rendición de cuentas de los funcionarios públicos.

Facilidades económicas son las oportunidades que tienen los individuos para utilizar los recursos económicos disponibles en la sociedad, tanto para el consumo, la producción como para el intercambio de bienes y servicios. Los derechos económicos de la persona dependen de los recursos que posea o que se encuentren disponibles para su uso, así como de las condiciones de intercambio, tales como los precios relativos y el funcionamiento de los mercados. Las siguientes variables, entre otras, se relacionan con este conjunto de facilidades: empleo, activos físicos productivos, ahorros (activos financieros), funcionamiento de los mercados (índices de precios locales), precios relativos del trabajo y de los bienes y servicios – condiciones de intercambio, PIB per cápita, crecimiento del PIB, distribución del ingreso.

Oportunidades sociales, corresponden a aquellos arreglos en la sociedad relativos a educación, salud, infraestructura, etc. que influyen en el bienestar de la población al mejorar sus condiciones de vida. Algunas variables asociadas a esta dimensión pudieran ser: la esperanza de vida al nacer, el índice de incidencia de bajo peso al nacer, acceso a los servicios de salud, acceso a servicios públicos (electricidad, agua, drenaje, servicios sanitarios), educación básica, media y superior, igualdad de géneros (en cuanto a oportunidades, educación, empleo e ingreso).

Garantías de transparencia, se refieren a las condiciones de apertura que esperan los individuos en la sociedad. Es decir a la capacidad de interacción ínter subjetiva bajo garantías de acceso a información veraz, lúcida y confiable. Lo anterior a fin de propiciar la confianza en la sociedad, evitar la corrupción, la irresponsabilidad en el manejo de las finanzas públicas y privadas, así como las transacciones subrepticias o el despojo de los derechos de los particulares. Se puede medir mediante: índices de delincuencia, corrupción, indicadores de transparencia y respeto a los derechos de propiedad, etc.

Seguridad social, son mecanismos para proveer una red de protección social para prevenir que la población o ciertos sectores de la población se conviertan en segmentos vulnerables frente a eventos catastróficos o adversos. Puede evaluarse mediante la existencia y cobertura de fondos de desempleo y retiro, así como fondos para la ayuda de los sectores mas necesitados de la población.

Estas libertades instrumentales, directamente acrecientan las capacidades de las personas e indirectamente se suplementan una con otra, reforzándose entre sí. En el presente estudio se identificarán dichas interconexiones.

Antecedentes en el Estudio del Desarrollo Humano

El antecedente más importante en la medición del desarrollo humano, indudablemente corresponde al esfuerzo realizado por el Programa de las Naciones Unidas para el Desarrollo (PNUD). En 1990 PNUD propone el Índice de Desarrollo Humano (IDH), escogiendo tres dimensiones básicas para la medición:

Una vida larga y saludable medida por la esperanza de vida al nacer.
Conocimientos, medidos por la tasa de alfabetización de adultos (con una ponderación de dos tercios) y la combinación de matriculación primaria, secundaria y terciaria (con ponderación de un tercio).
Un nivel de vida decoroso, medido por el PIB per cápita (en dólares corregidos por el poder adquisitivo PPA).

Antes de calcular el IDH, se crea un índice para calcular cada uno de estos tres componentes – esperanza de vida, educación y PIB-. Se seleccionan valores mínimos y máximos respecto de cada uno de los tres indicadores.

El resultado de cada componente se expresa como un valor entero entre 0 y 1 aplicando la siguiente fórmula general:

$$\text{Índice del componente} = \frac{\text{valor efectivo} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}}$$

Los valores máximos y mínimos establecidos por el PNUD son los siguientes:

INDICADOR	VALOR MÁXIMO	VALOR MÍNIMO
Esperanza de vida al nacer (años)	85	25
Tasa de alfabetización de adultos (%)	100	0
Tasa bruta combinada de matriculación (%)	100	0
PIB per cápita (dólares PPA)	40,000	100

Después de obtener el índice de cada dimensión, se calcula el IDH como simple promedio de los índices de los componentes.

Además del IDH, PNUD ha desarrollado tres índices complementarios: Índice de Pobreza Humana (IPH), el cual mide las privaciones de la población en las mismas dimensiones del desarrollo humano básico (IDH). El Índice de Desarrollo Relativo al Género (IDG), que mide el progreso en las mismas dimensiones y utiliza los mismos indicadores que el IDH, pero refleja las desigualdades del progreso entre el hombre y la mujer. Mientras mayor sea la disparidad de género en el desarrollo humano básico, más bajo será el IDG de un país respecto al IDH. Finalmente el Índice de Potenciación de Género (IPG) el cual mide el grado en que la mujer puede participar activamente en la vida económica y política de su país.

PNUD a partir de 1990 publica anualmente el IDH. El Informe Sobre Desarrollo Humano del 2003 muestra a México con un IDH de 0.800, ocupando la clasificación 55 dentro de un conjunto de 175 países. El valor del índice corresponde al del estrato superior de los países con "desarrollo medio". En América Latina, México se encuentra por debajo del valor de los índices alcanzados por Trinidad y Tobago (0.802), Cuba (0,806), Chile (0.831), Costa Rica (0.832), Uruguay (0.834) y Argentina (0.849), quienes ocupan las posiciones: 54, 52, 43, 42, 40 y 34 respectivamente.

A nivel regional, en México, el PNUD publicó por primera vez el Informe sobre Desarrollo Humano México 2002, en donde se realiza una estimación del IDH en cada una de las 32 entidades federativas. Como antecedentes, también se tienen los estudios realizados por De la Torre (1977), quien introduce su propia estimación de la esperanza de vida por entidades. Jarque y Medina (1998) agregan como componente de la dimensión de salud, el porcentaje de viviendas con agua potable. Ramírez (1999), introduce la tasa de mortalidad infantil y el porcentaje de viviendas con servicios públicos en vez de las variables de esperanza de vida e ingresos, mientras que el Consejo Estatal de Población del estado de Guanajuato, Coespo (2000) utiliza la asistencia a la escuela como uno de los sustitutos de la matriculación. Por otra parte, Conapo (2001) y García-Verdú (2002) consideran a la asistencia escolar como único sustituto de la tasa de matriculación en la estimación del logro educativo.

Dentro de un enfoque diferente, Pérez Trejo et al (2003) determinan las componentes principales del factor *Desarrollo* y un ordenamiento jerárquico y clasificación de las entidades federativas en función de su grado de desarrollo social.

La clasificación del desarrollo de las entidades federativas de acuerdo a algunos de los autores mencionados, se muestran en la Tabla 1.1.

Tabla 1.1
Lugar que Ocupan las Entidades en base a su Desarrollo de acuerdo a: PNUD, García-Verdú y CONAPO

	PNUD	García-Verdú	CONAPO
Aguascalientes	7	10	5
Baja California	6	7	4
Baja California Sur	3	8	9
Campeche	10	6	10
Chiapas	32	32	32
Chihuahua	4	4	7
Coahuila	5	5	3
Colima	12	13	11
Distrito Federal	1	1	1
Durango	16	15	16
Estado de México	17	17	15
Guanajuato	24	21	24
Guerrero	30	30	30
Hidalgo	27	24	28
Jalisco	13	14	14
Michoacán	28	26	27
Morelos	15	16	17
Nayarit	22	22	20
Nuevo León	2	2	2
Oaxaca	31	31	31
Puebla	25	25	25
Querétaro	14	11	13
Quintana Roo	9	3	6
San Luis Potosí	21	20	22
Sinaloa	18	18	18
Sonora	8	9	8
Tabasco	20	23	21
Tamaulipas	11	12	12
Tlaxcala	23	27	23
Veracruz	29	28	29
Yucatán	19	19	19
Zacatecas	26	29	26

Variables Observadas

A continuación se describen las variables usadas en este trabajo y que reflejan el fenómeno del *desarrollo* entendido por Sen. Estas variables representan la materia prima para aplicar los métodos multivariados y cumplir con los objetivos del estudio. Su elección se efectuó *a priori*, en base a dos criterios; el primero como medida en la realización de las libertades de la población y en segundo lugar, atendiendo a la disponibilidad de información estadística consistente y confiable a nivel municipal. Algunas de estas variables se miden de manera directa, mientras que otras son

"variables derivadas", cuya obtención se describe en el Anexo 2. Adicionalmente, se incluye un conjunto de variables demográficas, mediante las cuales se espera ampliar la caracterización de los municipios.

Variable	Descripción	Nombre
1. Índice de participación ciudadana	Número de votos emitidos en las elecciones presidenciales del 2000, dividido entre la población mayor de 18 años. Esta variable está asociada con el ejercicio de los derechos políticos de los ciudadanos, mediante el sufragio. (IFE[5])	PARELEC
2. Población económicamente activa ocupada	Proporción de la población económicamente activa ocupada respecto a la población total. (INEGI[6])	PEAOC
3. Índice de Pobreza	$P = H[I + (1-I)G]$[7], en donde H representa a la proporción de la población con ingresos inferiores a la "línea de pobreza" (en este caso definida como 1 salario mínimo), I es la brecha de pobreza, es decir la diferencia en el ingreso total de la población "pobre" respecto al ingreso correspondiente a un salario mínimo y G es el coeficiente Gini de la curva de Lorenz aplicada al sector "pobre". *(Ver Anexo 2 – Variables Derivadas).* Datos tomados de (INEGI)	P
4. y 5. Índice(s) de propiedad de accesorios y equipos duraderos	Se obtiene mediante la reducción de diez variables originales que representaron la tenencia de aparatos electrodomésticos, electrónicos y equipo de transporte (INEGI), a dos nuevas variables "componentes principales" que en conjunto representan el 93.6 % de la varianza total de las diez variables originales. (Ver Anexo 2)	IPAD1 IPAD2
6. PIB per cápita	Ingreso por habitante en $US dólares, ajustado por el poder adquisitivo (CONAPO[8])	PIBCAP
7. Vivienda	Representa la proporción de viviendas existentes, ponderadas por la calidad de dichas viviendas, mediante índice en escala ordinal, del 1 al 6, en base al material y calidad de la construcción, en donde la calificación (1) representa la calidad	VIVIENDA

[5] Instituto Federal Electoral (IFE). Elecciones Presidenciales del 2000
[6] Instituto Nacional de estadística, Geografía e Informática (INEGI). Censo General de Población 2000
[7] Sen, Amartya. 1976. Econometrica, **44**, 219-231
[8] Consejo Nacional de Población (CONAPO), Secretaría de Gobernación

	mínima y (6) la máxima y por la proporción de viviendas propias en el municipio. Esta variable es adimensional. (Anexo 1, Datos de INEGI)	
8. Ocupantes	Número de ocupantes por vivienda. Se obtiene dividiendo el número de viviendas existentes entre la población total, en cada municipio. (Datos de INEGI)	OCUPANTES
9. Servicios básicos	Se refiere al acceso a los servicios de *drenaje*, *servicios sanitarios*, *energía eléctrica* y *agua entubada*. Se calcula multiplicando la proporción de hogares que cuentan con cada uno de dichos servicios, respecto al total de hogares en cada municipio. (Datos de INEGI – Ver anexo 2)	SERBAC
10. Mortalidad infantil	Tasa de mortalidad infantil; número de fallecimientos por cada 1000 nacimientos. Este índice se utiliza como una medida del índice de esperanza de vida, por su elevada correlación con ésta. (Datos de INEGI)	MORIN
11. Fecundidad	Tasa global de fecundidad. Corresponde al número de nacimientos por mujer en edad reproductiva. (Datos de INEGI)	FECUNDI
12. Discapacidad	Proporción de la población con discapacidad, respecto al total de la población municipal. (Datos de INEGI)	PDISC
13. Edad	Edad mediana de la población municipal. (Datos de INEGI)	EDAD
14. Índice de masculinidad	Mide la proporción de habitantes del género masculino, en relación a los habitantes del género femenino. (Datos de INEGI[2])	IMASC
15. Migración	Proporción de la población proveniente de otras entidades. (Datos de INEGI)	MIGRANTES
16. Analfabetismo	Proporción de la población analfabeta respecto al total. (Datos de INEGI)	SINSTRUCCION
17. Educación básica	Proporción de la población con instrucción primaria y secundaria. (Datos de INEGI)	EDUBAS
18. Educación media y superior	Proporción de la población con educación media (preparatoria, normal, técnica) y superior (licenciatura). (Datos de INEGI)	EDUSUP
19. Equidad de género	Se mide como el cociente de la proporción de analfabetismo masculino, respecto al femenino. (Datos de INEGI)	RAGEN
20. Seguridad social	Se mide como la proporción de la población afiliada a alguna de las instituciones de seguridad y procuración de servicios médicos (IMSS, ISSTE, Fuerzas Armadas). (Datos de INEGI)	SSS

21. Delincuencia	Es el número de delitos consignados a nivel municipal, por cada 1000 habitantes. (Datos de INEGI)	DEL

La variable (1), es un indicador de las libertades políticas de la población. Las variables (2) a la (7) se encuentran asociadas con las facilidades económicas disponibles en los municipios. Las variables (9), (10) y (16) a la (19), con las oportunidades sociales. Las variables (20) y (21) se asocian a la seguridad social; finalmente, las variables (11) a la (15) representan condiciones demográficas en los municipios. A nivel municipal, no se encontraron variables indicadoras de las libertades asociadas a garantías de transparencia. Sin embargo, como se verá mas adelante, las veintiuna variables consideradas muestran un alto grado de correlación entre sí.

La inclusión de las 21 variables arriba indicadas, permite resumir cerca de cuarenta variables originales. Dentro de estas se incluyen, por ejemplo las existencias por habitante de los diez bienes duraderos y electrodomésticos más comunes, incluyendo equipo de transporte; la calidad de la vivienda, enunciada en términos de los materiales de construcción empleados, de acuerdo a seis distintas categorías y el nivel de instrucción referido al género. La base de datos, la cual se incluye en el Anexo 2, considera a cuatro entidades, 305 municipios y una población total de cerca de 30 millones de habitantes, de acuerdo al siguiente resumen.

Entidad	Núm. Municipios (Delegaciones)	Población
Estado de México	122	13,096,686
Nuevo León	50	3,834,141
Distrito Federal	16	8,605,239
Chiapas	117	3,920,892
Total	305	29,456,958

La muestra, representa aproximadamente el 30 % de la población total de los Estados Unidos Mexicanos. Su composición, según se comentó con anterioridad, refleja una gran diversidad geográfica y étnica, representando al mismo tiempo poblaciones muy heterogéneas en cuanto a su estado de desarrollo.

Capítulo 2

Resultados del Análisis Estadístico del Desarrollo Social en México[9]

En el presente capítulo se describen los resultados del análisis estadístico del desarrollo social en México, en relación a los estados de Chiapas, Estado de México, Nuevo León y el Distrito Federal. El análisis multivariado, como ya se indicó en el Prólogo, comprende la aplicación de los métodos de *componentes principales*, *análisis de factores* y *análisis de conglomerados*, previo análisis de *correlación* de la matriz de datos.

Los fundamentos estadísticos de los métodos multivariados empleados, se explican en el Capítulo 3.

2.1 Análisis de Correlación

Como fase previa al análisis multivariado, se determinó la correlación existente entre las variables utilizadas en el análisis estadístico. De la Tabla 2.1.1, en donde se muestran los coeficientes de correlación de Pearson para cada par de variables originales, podemos establecer *a priori* la posibilidad de reducir la dimensionalidad del conjunto de datos originales en un nuevo conjunto de variables no correlacionadas. Lo anterior, debido a que un gran número de dichas variables muestran un coeficiente de correlación razonablemente alto ($> |0.5|$), entre sí.

Adicionalmente puede observarse que el sentido de la asociación lineal corresponde al esperado; de esta manera, la variable SINSTRUCCION que mide la proporción analfabeta de la población, muestra una correlación negativa con los índices de educación básica y superior (EDUBAS y EDUSUP), la proporción de la población económicamente activa (PEAOC) y el ingreso (PIBCAP), la disponibilidad de servicios básicos y de seguridad social (SERBAS y SSS), así como el ejercicio de los derechos ciudadanos a través del voto (PARELEC). Por otra parte, la correlación de SINSTRUCCION, es positiva con las variables FECUNDI, OCUPANTES, MORIN, RAGEN y P, entre otras, las cuales miden tasa global de fecundidad, número de ocupantes por vivienda, mortalidad infantil, discriminación por género y pobreza, respectivamente.

La variable EDUSUP, por su parte está asociada positivamente a elevados índices de ocupación, seguridad y salud, y negativamente al índice de pobreza (P), la tasa global de fecundidad y el índice de mortalidad infantil.

A su vez, las variables anteriores, también muestran asociaciones significativas entre sí; la tasa de fecundidad está positivamente correlacionada con el índice de mortalidad infantil y el índice de pobreza, así como el número de ocupantes por hogar.

---

[9] El análisis estadístico se realizó en su totalidad mediante el empleo de la versión 14.0 de Minitab®

Tabla 2.1.1

Matriz de Correlación de Pearson de las Variables Observadas

	PDISC	SINSTRU	EDUBAS	EDUSUP	PEAOC	FECUNDI	OCUP	MIGR	MORIN	EDAD	INMASC	SSS	VIVIENDA	SERBAS	IPAD1	IPAD2	PARELEC	PIBCAP	DEL	RAGEN1	P
PDISC	1.00	-0.30	0.36	0.15	0.19	-0.32	-0.55	0.21	-0.17	0.52	0.16	0.28	-0.06	0.15	-0.38	0.36	0.36	0.23	0.28	-0.33	-0.26
SINSTRUCCION		1.00	-0.36	-0.70	-0.51	0.75	0.57	-0.62	0.66	-0.73	0.16	-0.70	-0.45	-0.74	0.81	-0.14	-0.37	-0.56	-0.44	0.70	0.88
EDUBAS			1.00	-0.14	0.13	-0.24	-0.23	0.09	0.01	0.27	0.35	0.08	0.08	0.01	-0.04	0.74	-0.04	-0.14	0.07	-0.57	-0.24
EDUSUP				1.00	0.74	-0.75	-0.56	0.60	-0.63	0.70	-0.46	0.76	0.34	0.83	-0.85	-0.41	0.27	0.82	0.45	-0.42	-0.72
PEAOC					1.00	-0.68	-0.56	0.51	-0.52	0.65	-0.12	0.71	0.17	0.75	-0.70	-0.22	0.12	0.68	0.26	-0.41	-0.56
FECUNDI						1.00	0.76	-0.57	0.71	-0.84	0.17	-0.67	-0.31	-0.76	0.81	-0.10	-0.23	-0.60	-0.38	0.61	0.74
OCUPANTES							1.00	-0.53	0.61	-0.87	-0.02	-0.64	-0.17	-0.59	0.81	-0.32	-0.36	-0.57	-0.45	0.49	0.63
MIGRANTES								1.00	-0.54	0.55	-0.06	0.62	0.18	0.64	-0.66	-0.03	0.20	0.51	0.35	-0.38	-0.66
MORIN									1.00	-0.69	-0.06	-0.64	-0.21	-0.67	0.73	-0.05	-0.24	-0.58	-0.32	0.47	0.60
EDAD										1.00	-0.05	0.69	0.31	0.70	-0.91	0.21	0.43	0.68	0.53	-0.55	-0.72
INMASC											1.00	-0.08	-0.25	-0.22	0.21	0.56	-0.12	-0.29	-0.19	-0.03	0.24
SSS												1.00	0.34	0.77	-0.81	0.00	0.24	0.66	0.37	-0.44	-0.78
VIVIENDA													1.00	0.41	-0.36	0.02	0.26	0.18	0.22	-0.19	-0.52
SERBAS														1.00	-0.80	-0.12	0.23	0.64	0.35	-0.47	-0.77
IPAD1															1.00	0.00	-0.46	-0.80	-0.55	0.62	0.83
IPAD2																1.00	0.05	-0.36	-0.03	-0.30	-0.15
PARELEC																	1.00	0.28	0.34	-0.21	-0.36
PIBCAP																		1.00	0.47	-0.32	-0.57
DEL																			1.00	-0.27	-0.45
RAGEN1																				1.00	0.58
P																					1.00

El ingreso per cápita (PIBCAP), no muestra una correlación elevada con el resto de las variables, lo cual apunta hacia la conjetura de que el crecimiento económico no es condición necesaria y suficiente para promover el desarrollo en términos de las variables observadas, descritas con anterioridad.

Los resultados empíricos anteriores; es decir, el grado y la dirección en la asociación entre variables, concuerdan con la idea central de la teoría del *"desarrollo como libertades"*. Sen [5] plantea una doble faceta en la interpretación de éstas. Por una parte, la expansión de las capacidades de la persona establece en sí misma la medida y el objetivo del desarrollo. En este sentido la medición de dichas capacidades constituye la forma de *evaluación* del éxito o fracaso de una sociedad. Esta posición valorativa difiere de los enfoques normativos tradicionales, centrados en otras variables como utilidad o ingreso real.

Por otra parte, la libertad no solamente representa la base para evaluar a la sociedad, sino que también se convierte en el principal determinante para promover la iniciativa individual y la *efectividad* de la sociedad en su conjunto. Una persona que disfruta de libertades individuales (capacidades y oportunidades), indudablemente se convierte en agente[10] de cambio en la sociedad. El desarrollo en este sentido es por lo tanto fin y medio, de donde deriva su doble función *constitutiva* e *instrumental*. De esta manera, la satisfacción de las libertades más básicas, expresadas en términos de indicadores de salud, vivienda y educación, conduce hacia nuevas libertades, como las expresadas a través de oportunidades de empleo, ingresos superiores y una mayor participación en la vida pública; todas estas variables asociadas, a la vez con una mayor disponibilidad de servicios tanto de infraestructura básica como de seguridad y protección.

Finalmente, en vista de que existen correlaciones altas entre las variables, podemos esperar que la estructura de covarianza (correlación) de datos pueda ser representada mediante un número reducido de nuevas variables no correlacionadas o *componentes principales*.

2.2 Análisis de Componentes Principales

A continuación se describen los resultados del análisis de componentes principales. Este análisis se realiza con fines exploratorios, con el propósito de determinar el número aproximado de componentes principales; es decir la dimensionalidad de los datos y subsecuentemente el número de *factores* a emplear en la interpretación de la covarianza de los datos originales.

Puesto que las variables originales no se encuentran medidas en unidades homogéneas, se procedió a transformarlas, mediante estandarización, a fin de evitar que una variable expresada en un orden de magnitud superior al resto (i.e. PIBCAP), pudiese tener una influencia desproporcionada en los análisis subsecuentes.

Los resultados del ajuste inicial del modelo de componentes principales, sin restricción en cuanto al número de componentes, se muestran a continuación:

[10] Sen [5] entiende el concepto de agencia en su sentido más amplio; es decir, como el individuo que forma parte del público y que participa en la actividad económica, política y social de su comunidad.

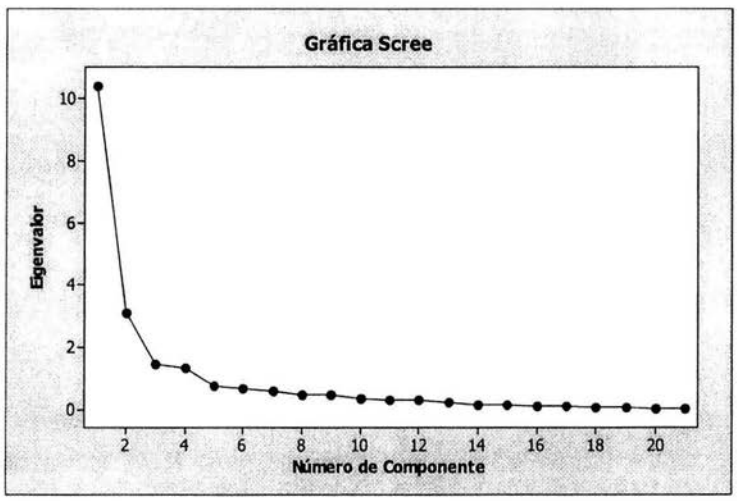
Tabla 2.2.1
Análisis de Componentes Principales

	1	2	3	4	5	6	7	8	9
Eigenvalor	10.375	3.089	1.456	1.338	0.746	0.654	0.590	0.484	0.446
Proporción	0.494	0.147	0.069	0.064	0.036	0.031	0.028	0.023	0.021
Acumulada	0.494	0.641	0.710	0.774	0.810	0.841	0.869	0.892	0.913
	10	11	12	13	14	15	16	17	18
Eigenvalor	0.346	0.319	0.285	0.228	0.158	0.124	0.112	0.089	0.062
Proporción	0.016	0.015	0.014	0.011	0.008	0.006	0.005	0.004	0.003
Acumulada	0.930	0.945	0.958	0.969	0.977	0.983	0.988	0.992	0.995
	19	20	21						
Eigenvalor	0.042	0.031	0.025						
Proporción	0.002	0.001	0.001						
Acumulada	0.997	0.999	1.000						

De la Tabla 2.2.1 se desprende que la primer componente representa el 49.4 % de la varianza total de los datos; ocho componentes resumen prácticamente la variabilidad de los mismos, al reproducir aproximadamente el 90 % de la varianza total del conjunto de 21 variables originales. Las cuatro primeras componentes principales tienen eigenvalores superiores a la unidad y en conjunto representan el 77.4 % de la varianza total de las variables originales. El resultado anterior, también se puede observar en la Gráfica 2.2.1, en donde puede advertirse que la contribución en la explicación de la variabilidad de los datos decrece, hasta volverse poco significativa, posiblemente a partir del 5° o 6° eigenvalor.

Como corolario de lo anterior, se corrobora la factibilidad de expresar los datos, mediante un nuevo conjunto reducido de variables componentes principales.

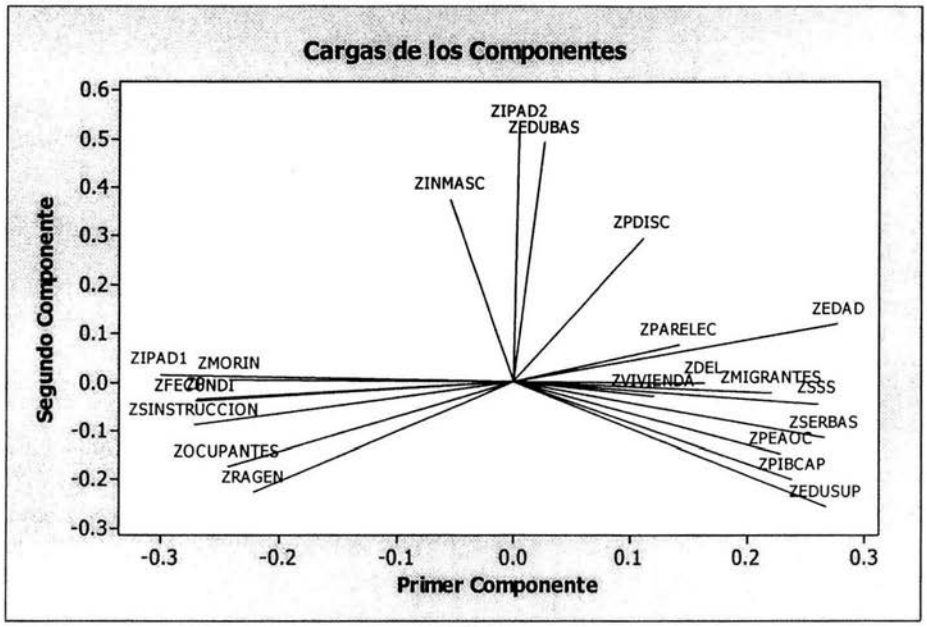
Gráfica 2.2.1
Determinación Gráfica del Número de Componentes Principales



La Gráfica 2.2.2, representa las coordenadas de las cargas de las dos primeras componentes principales, las cuales representan el 64 % de la variabilidad total de los datos y por lo tanto, para este caso particular, la consideramos lo suficientemente adecuada para representar los datos. En ésta se muestra la contribución de la i-ésima

variable en la j-ésima componente. A partir de las coordenadas de las variables en el plano formado por las dos primeras componentes, se pueden identificar tres grandes grupos de variables. Las variables que aparecen en el cuadrante inferior izquierdo se puede decir que se encuentran asociadas a niveles inferiores de "desarrollo", mientras que las que se ubican en el cuadrante inferior derecho, con estados de desarrollo superior. Ambos conjuntos de variables muestran una asociación importante con la primera componente. Por otra parte, al centro de la gráfica, en la mitad superior, se ubican cuatro variables (INMASC, IPAD2, EDUBAS y PDISC), todas ellas fuertemente vinculadas a la segunda componente.

Gráfica 2.2.2



En la Tabla 2.3.3 se reproducen tanto las cargas de las componentes de los primeros cuatro eigenvectores correspondientes a los eigenvalores obtenidos como resultado del ajuste del modelo de componentes principales, como los coeficientes de correlación[11] entre las primeras cuatro componentes Y_j $(j=1,\cdots,4)$ y las 21 variables X_i $(i=1,\cdots,21)$, debido a que dichos coeficientes de correlación contribuyen en la interpretación de las componentes; en este caso principalmente, la primera.

Puesto que la primera componente muestra coeficientes de correlación muy elevados (>|0.7|), con signo positivo con aquellas variables asociadas a un alto grado de desarrollo y con el signo contrario a las relacionadas con condiciones generales de atraso, el análisis de la correlación entre las variables y la primera componente,

[11] $\rho_{Y_j,X_i}=\dfrac{e_{ji}\sqrt{\lambda_j}}{\sqrt{\sigma_{ii}}}$

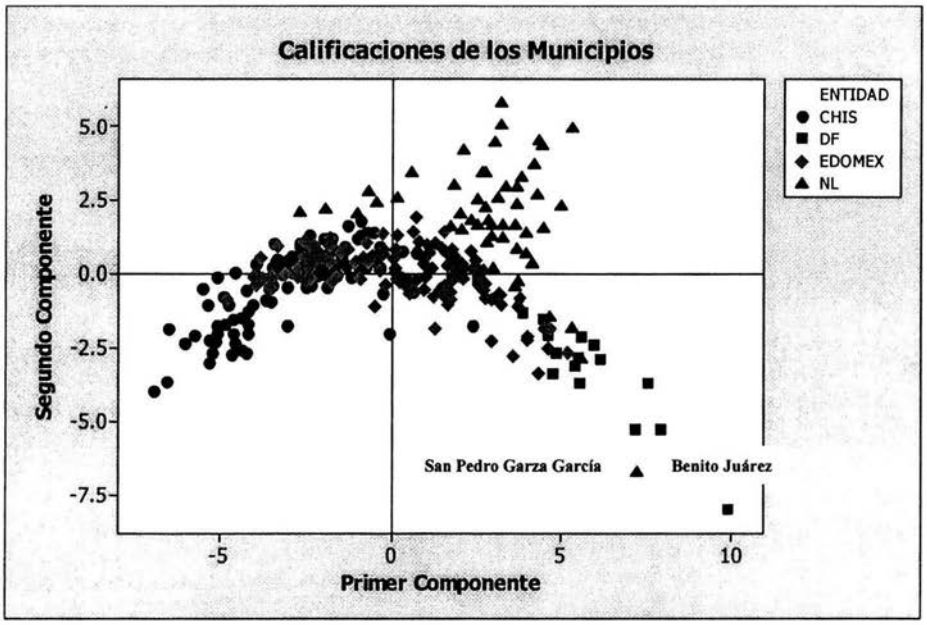
concuerda con el análisis de las cargas. En vista de lo anterior, de manera preliminar, interpretaremos a la primera componente como la "nueva variable" o componente capaz de medir el *desarrollo*. La confirmación de esta interpretación, así como la correspondiente a las tres componentes restantes, se difiere hasta la siguiente sección.

Tabla 2. 2.2
Cargas de las Componentes Principales y sus Coeficientes de Correlación con las Variables Originales

Variable	PC_1	PC_2	PC_3	PC_4	Rho_1	Rho_2	Rho_3	Rho_4
ZPDISC	0.110	0.295	0.197	0.447	0.356	**0.518**	0.238	**0.517**
ZSINSTRUCCION	-0.272	-0.088	0.256	0.111	**-0.875**	-0.155	0.309	0.128
ZEDUBAS	0.025	0.492	-0.167	-0.124	0.081	**0.865**	-0.202	-0.143
ZEDUSUP	0.267	-0.256	0.007	-0.001	**0.861**	-0.450	0.008	-0.001
ZPEAOC	0.230	-0.148	0.265	-0.167	**0.740**	-0.260	0.320	-0.193
ZFECUNDI	-0.271	-0.039	-0.065	0.119	**-0.872**	-0.068	-0.078	0.138
ZOCUPANTES	-0.244	-0.174	-0.237	-0.167	**-0.786**	-0.306	-0.286	-0.193
ZMIGRANTES	0.221	-0.023	0.077	-0.140	**0.711**	-0.040	0.093	-0.162
ZMORIN	-0.242	0.005	-0.200	0.179	**-0.779**	0.009	-0.241	0.207
ZEDAD	0.277	0.119	0.099	0.153	**0.893**	0.209	0.119	0.177
ZINMASC	-0.055	0.375	0.355	-0.212	-0.177	**0.660**	0.428	-0.245
ZSSS	0.260	-0.044	0.089	-0.150	**0.838**	-0.077	0.107	-0.174
ZVIVIENDA	0.121	-0.028	-0.589	-0.128	0.390	-0.050	**-0.711**	-0.148
ZSERBAS	0.265	-0.115	-0.009	-0.213	**0.855**	-0.202	-0.011	-0.246
ZIPAD1	-0.301	0.016	-0.034	-0.085	**-0.970**	0.027	-0.041	-0.098
ZIPAD2	0.003	0.523	-0.020	-0.113	0.010	**0.920**	-0.024	-0.131
ZPARELEC	0.142	0.076	-0.246	0.490	0.458	0.134	-0.297	**0.567**
ZPIBCAP	0.239	-0.199	0.172	0.160	**0.769**	-0.350	0.208	0.185
ZDEL	0.163	-0.003	-0.105	0.429	**0.526**	-0.005	-0.127	**0.496**
ZRAGEN	-0.222	-0.227	0.170	0.124	**-0.714**	-0.398	0.205	0.143
ZP	-0.270	-0.036	0.263	0.112	**-0.869**	-0.063	0.317	0.130

Resulta interesante resaltar la posición de los municipios respecto a los dos primeros componentes principales, la cual se muestra en la Gráfica 2.2.3. Al ubicar a los individuos en el espacio conformado por las dos primeras componentes principales, se observa la formación de conglomerados de unidades bien identificadas. Las unidades (delegaciones en el caso del DF y municipios en las otras tres entidades), tienen una calificación alta en el primer componente para el DF y un gran número de municipios del estado de NL; y baja en el caso de Chiapas, así como aproximadamente la mitad de los municipios del Estado de México. NL y el DF contrastan visiblemente en relación a la calificación correspondiente al segundo componente, mientras que la calificación en el segundo componente correspondiente a los municipios del Estado de México y una parte importante de los del estado de Chiapas es muy cercana a cero. En la misma Gráfica 2.2.3 se aprecian individuos con una calificación atípicamente elevada en ambas componentes; en el caso de la primera componente, con signo positivo y con signo negativo en la segunda. Estos municipios corresponden a la delegación Benito Juárez en el DF y el municipio de San Nicolás Garza García en el estado de NL.

Gráfica 2.2.3



En la siguiente sección, mediante el ajuste del *modelo de factores*, trataremos de generar un conjunto de variables subyacentes o *factores*, que expliquen la estructura de correlación de las variables observadas, asimismo, procuraremos desarrollar una interpretación de los mismos.

2.3 Análisis de Factores

El análisis de componentes principales conduce a la elección de entre cuatro y ocho componentes para explicar la variabilidad de la muestra de datos de las 21 variables originales. Para la determinación de factores, se seleccionan los primeros cuatro, apoyados por una parte en la práctica comúnmente aceptada, de utilizar un número de factores cuyos eigenvalores excedan la unidad y por otra parte, por la dificultad de dar una interpretación razonable a un número mayor de factores.

El procedimiento estadístico consistió en la evaluación de los factores empleando los dos métodos más comúnmente utilizados; es decir el método de las *componentes principales* y el de *máxima verosimilitud.* En ambos casos, el cálculo de las cargas de los factores se realizó sin rotación de los ejes y con rotación *varimax*.

El desarrollo del análisis, así como los resultados más sobresalientes se describe a continuación.

De la Tabla 2.3.1, resulta evidente[12] que las variables SINSTRUCCIÓN, EDUSUP, PEAOC, FECUNDI, OCUPANTES, MIGRANTES, MORIN, EDAD, SSS, SERBAS, IPAD1, PICAP, RAGEN y P, definen al Factor1. Sin embargo, en la interpretación de la

[12] Para fines de interpretación, se considera que una variable es influyente, si |carga | > 0.5

variable DEL (con carga sin rotación de 0.526 en el Factor1), es preciso referirse a las entradas de la misma tabla con rotación de los ejes, ya que en éstas se muestra con carga elevada (0.651) en el Factor3 y moderada o baja en el resto de los factores, permitiendo una mejor interpretación del efecto de esta variable. El Factor2 está definido por las variables EDUBAS, INMASC e IPAD2. La Variable PDISC, al rotar los ejes, aparece claramente, como definitoria del Factor3.

Tabla 2.3.1
Cargas de los Factores – Método de Factores Principales

	Cargas de los Factores sin Rotación				Cargas con Rotación Varimax				
Variable	*Factor1*	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*	*Comunalidad*
ZPDISC	0.356	**0.518**	0.238	**0.518**	0.169	0.383	**0.697**	0.240	0.719
ZSINSTRUCCION	**-0.875**	-0.155	0.308	0.129	**-0.701**	-0.186	-0.276	**0.546**	0.901
ZEDUBAS	0.081	**0.865**	-0.202	-0.143	-0.041	**0.867**	0.119	-0.220	0.815
ZEDUSUP	**0.861**	-0.450	0.009	-0.001	**0.816**	-0.420	0.220	-0.232	0.945
ZPEAOC	**0.740**	-0.260	0.320	-0.193	**0.847**	-0.180	0.021	0.061	0.755
ZFECUNDI	**-0.872**	-0.068	-0.078	0.137	**-0.834**	-0.117	-0.216	0.183	0.790
ZOCUPANTES	**-0.786**	-0.306	-0.286	-0.193	**-0.693**	-0.270	**-0.517**	-0.104	0.830
ZMIGRANTES	**0.711**	-0.040	0.092	-0.162	**0.715**	0.017	0.107	-0.134	0.542
ZMORIN	**-0.779**	0.009	-0.242	0.207	**-0.834**	-0.065	-0.088	0.020	0.709
ZEDAD	**0.893**	0.209	0.120	0.177	**0.751**	0.176	**0.534**	-0.086	0.888
ZINMASC	-0.177	**0.660**	0.428	-0.246	-0.015	**0.711**	-0.168	0.419	0.710
ZSSS	**0.838**	-0.077	0.107	-0.174	**0.841**	-0.013	0.136	-0.157	0.750
ZVIVIENDA	0.390	-0.050	**-0.711**	-0.148	0.172	-0.028	0.060	**-0.804**	0.681
ZSERBAS	**0.855**	-0.202	-0.011	-0.247	**0.855**	-0.118	0.057	-0.291	0.832
ZIPAD1	**-0.970**	0.027	-0.041	-0.099	**-0.844**	0.034	-0.445	0.201	0.953
ZIPAD2	0.010	**0.920**	-0.025	-0.131	-0.058	**0.922**	0.101	-0.030	0.864
ZPARELEC	0.458	0.134	-0.297	**0.567**	0.116	-0.016	**0.734**	-0.291	0.637
ZPIBCAP	**0.769**	-0.350	0.208	0.185	**0.727**	-0.365	0.358	0.019	0.791
ZDEL	**0.526**	-0.005	-0.126	0.496	0.269	-0.126	**0.651**	-0.164	0.539
ZRAGEN	**-0.714**	-0.398	0.205	0.143	**-0.569**	-0.426	-0.250	0.403	0.730
ZP	**-0.704**	-0.098	-0.253	**0.555**	**-0.704**	-0.098	-0.253	**0.555**	0.877
Varianza	10.375	3.089	1.456	1.338	8.47	2.9767	2.7236	2.0874	16.2577
% Var. acumulada	49.4	64.1	71.0	77.4	40.3	54.5	67.5	77.4	

Al Factor3, lo determinan, además de PDISC, las siguientes variables: OCUPANTES, EDAD, PARELEC y como ya se indicó, DEL. Finalmente, el Factor4 está delimitado por las variables SINSTRUCCION, VIVIENDA y P. Del resultado del análisis de factores se concluye que no se encontraron factores triviales, bajo ambos métodos.

En el análisis anterior, la rotación de los ejes, ha resultado útil en la determinación de las variables influyentes en cada uno de los factores.

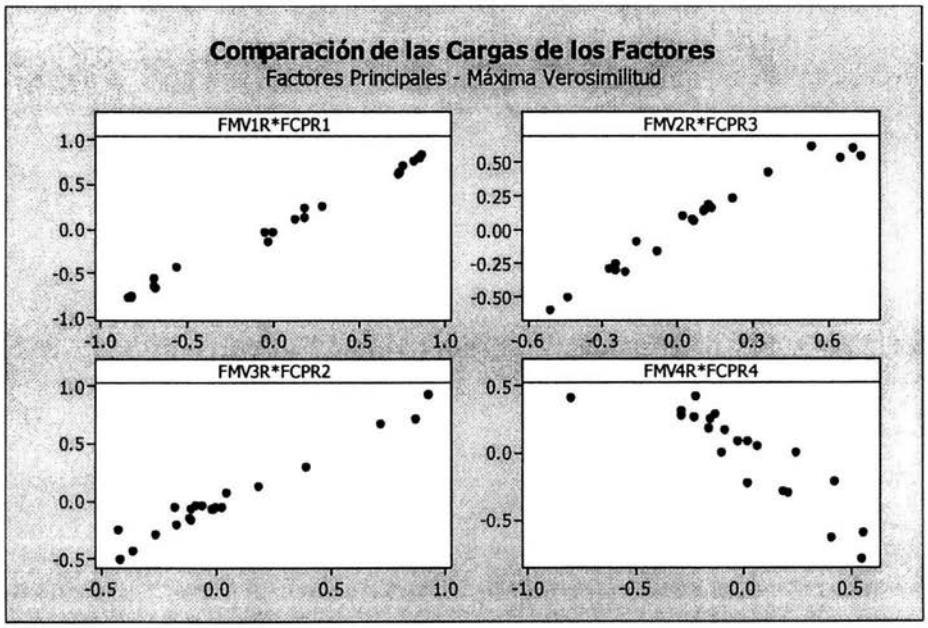
En general, la rotación de los ejes nos permite, sin alterar los resultados del análisis de factores, identificar un patrón de cargas más claro, de manera que cada variable tenga dentro de lo posible, una carga elevada en un factor y baja, moderada o de signo contrario en los demás factores.

Por otra parte, los cuatro factores, representan una comunalidad o varianza explicada por el conjunto de variables del 77.4 %. Dentro de las 21 variables seleccionadas, existen solamente dos (MIGRANTES y DEL) con varianzas específicas notoriamente superiores a las del resto de las variables. Lo anterior significa que estas dos variables son pobremente explicadas por los cuatro factores elegidos.

Tabla 2.3.2
Cargas de los Factores – Método de Máxima Verosimilitud

	Cargas de los Factores sin Rotación				Cargas con Rotación Varimax				
Variable	*Factor1*	*Factor2*	*Factor3*	*Factor4*	*Factor1*	*Factor2*	*Factor3*	*Factor4*	*Comunalidad*
ZPDISC	0.268	-0.009	0.552	0.318	0.128	**0.604**	0.311	0.011	0.478
ZSINSTRUCCION	**-1.000**	0.000	0.000	0.000	-0.567	-0.290	-0.038	**-0.770**	1.000
ZEDUBAS	0.332	**0.703**	0.408	0.052	-0.151	0.189	**0.725**	0.436	0.774
ZEDUSUP	**0.698**	**-0.670**	-0.156	-0.036	**0.770**	0.241	-0.486	0.273	0.962
ZPEAOC	**0.521**	**-0.573**	0.078	-0.270	**0.792**	0.100	-0.192	0.066	0.679
ZFECUNDI	**-0.742**	0.355	-0.269	0.146	**-0.772**	-0.309	-0.052	-0.277	0.771
ZOCUPANTES	**-0.554**	0.364	**-0.666**	-0.039	**-0.672**	-0.600	-0.272	0.014	0.885
ZMIGRANTES	**0.626**	-0.271	0.080	-0.155	**0.614**	0.157	-0.036	0.303	0.495
ZMORIN	**-0.644**	0.359	-0.207	0.256	**-0.760**	-0.164	-0.033	-0.213	0.652
ZEDAD	**0.719**	-0.378	**0.502**	0.098	**0.703**	**0.616**	0.132	0.178	0.922
ZINMASC	-0.172	0.364	**0.522**	-0.273	-0.040	-0.086	0.679	-0.195	0.509
ZSSS	**0.702**	-0.399	0.138	-0.252	**0.798**	0.158	-0.041	0.266	0.734
ZVIVIENDA	0.463	0.019	-0.121	-0.011	0.227	0.062	-0.055	0.413	0.229
ZSERBAS	**0.734**	-0.443	0.005	-0.297	**0.833**	0.076	-0.150	0.319	0.823
ZIPAD1	**-0.809**	0.487	-0.265	-0.048	**-0.785**	**-0.508**	0.089	-0.285	0.964
ZIPAD2	0.130	**0.617**	**0.703**	-0.141	-0.042	0.141	**0.938**	0.100	0.911
ZPARELEC	0.445	-0.078	0.147	0.416	0.117	**0.548**	-0.050	0.287	0.399
ZPIBCAP	**0.548**	**-0.670**	0.027	0.139	**0.639**	0.426	-0.413	0.101	0.770
ZDEL	0.433	-0.253	0.147	0.351	0.247	0.530	-0.138	0.187	0.396
ZRAGEN	**-0.819**	-0.120	-0.183	0.003	-0.441	-0.301	-0.239	**-0.615**	0.719
ZP	**-0.886**	0.140	-0.074	0.091	**-0.650**	-0.252	-0.029	**-0.575**	0.818
Varianza	8.2114	3.5347	2.2721	0.8772	7.1783	2.6756	2.6192	2.4174	14.8905
% Var. acumulada	39.1	55.9	66.7	70.9	34.2	46.9	59.4	70.9	

Gráfica 2.3.1



Nota: FMV_iR y FCP_iR se refieren a los cuatro primeros factores que se obtienen mediante el ajuste del modelo de factores por los métodos de máxima verosimilitud y componentes principales, respectivamente.

La comparación de los resultados del análisis de factores mediante el método de factores principales y máxima verosimilitud, muestra resultados análogos. Al graficar las cargas de los factores equivalentes obtenidas mediante ambos métodos, (Gráfica 2.3.2), se observa que los puntos que definen la posición de la variables, se distribuyen a lo largo de líneas rectas con pendiente aproximada de 45°, con la excepción del Factor4, en donde los signos de las cargas son contrarios. Los coeficientes de correlación entre los factores contrastantes son: 0.997, 0.997, 0.998 y -0.879, respectivamente. El resultado anterior, nos permite suponer que el ajuste del modelo de factores es bastante estable con respecto al método de cálculo de las cargas.

INTERPRETACIÓN DE LOS FACTORES (Método de Factores Principales)

El Factor1 es un factor *bipolar*, debido a que la mitad de las catorce cargas relevantes son positivas y la otra mitad negativas. Las cargas con signo positivo y de magnitud apreciable (> |0.5|), corresponden a las variables observadas que son características de un mayor desarrollo de la sociedad (disponibilidad de servicios de infraestructura básica; SERBAS y de seguridad social SSS, elevados niveles de educación EDUSUP, empleo PEAOC e ingreso PIBCAP, con cargas 0.855, 0.841, 0.816, 0.847 y 0.727, respectivamente). Adicionalmente, la variable MIGRANTES muestra una carga positiva de importancia (0.715), indicando flujo migratorio positivo. Las cargas con valor alto y signo negativo, son aquellas inherentes a la ausencia de desarrollo, fundamentalmente analfabetismo (-0.701), altos índices de pobreza P (-0.704), mortalidad infantil (-0.834) y discriminación por género, RAGEN (-0.569), asociados a tasas de fecundidad elevada (-0.834) y consiguientemente un mayor número de ocupantes por vivienda (-0.693). La variable que describe la participación ciudadana en el gobierno de la comunidad (PARELEC), aunque con signo positivo, tiene relativamente bajo peso en este factor (0.116). El Factor1, debido a que refleja condiciones favorables en cuanto a avance material, educación y bienestar, resulta un factor que pudiera fácilmente identificarse como el factor de que mide el *desarrollo*.

El Factor2 nos describe a una población medianamente educada (EDUBAS) con peso sobresaliente (0.867), un elevado índice de masculinidad (0.711) y una notable propensión al ahorro (o acumulación de riqueza) mediante la adquisición de artefactos duraderos no tradicionales; IPAD2 con carga 0.922. El analfabetismo muestra una carga moderadamente baja y con signo negativo (-0.186), pero también, la educación superior (-0.420). Las variables relativas a los servicios básicos y de salud no muestran pesos altos. Puesto que el ingreso, PIBCAP tiene una carga moderadamente elevada y con signo negativo (-0.365), el Factor2 parece describir un estado caracterizado por pocas oportunidades de empleo y por otra parte, acceso limitado a los servicios públicos. El Factor2 describe un estado de *estancamiento*, que sin representar condiciones de pobreza, no parece describir a una población que ofrezca las oportunidades para propiciar un desarrollo superior.

El Factor3 está definido por cinco variables. La variable dominante es PARELEC (0.734), en donde aparece con la mayor influencia, dentro de los cuatro factores considerados. Las otras cuatro variables que definen el factor son PDISC (0.697), DEL (0.651), EDAD (0.534) y OCUPANTES (-0.517). Por otra parte el ingreso, PIBCAP presenta carga positiva y con magnitud media (0.358). Asimismo, las variables que miden la educación, los servicios y la ocupación no parecen ser influyentes en la definición del factor. El conjunto de las características anteriores, sugiere una población

con edad avanzada y sin acceso a servicios básicos de infraestructura y de salud, lo cual posiblemente explique la influencia de la variable PDISC en la definición del factor. Por otra parte, la relevancia de la variable DEL pudiera corresponder a una población en donde las oportunidades de empleo, medida por PEAOC, no sean relevantes.

El Factor3, por las características anteriores, parece representar un estado de *progreso incipiente,* en donde los partidos políticos en el poder, a pesar de su relativamente elevada capacidad de convocatoria, no han respondido plenamente a la sociedad en la satisfacción de sus necesidades básicas; es decir, salud, educación seguridad y servicios básicos.

El Factor4 está formado por un conjunto ponderado de cargas positivas y negativas; sin embargo en este caso los signos positivos, corresponden, en contraste con el Factor1, a variables asociadas con condiciones de subdesarrollo social. Las variables que definen el factor son SINSTRUCCON, VIVIENDA y el índice de pobreza P, con valores: 0.546, -0.804 y 0.555, respectivamente. Este factor parece describir estados inferiores de desarrollo. La falta de participación electoral PARELEC resulta evidente, debido a su carga, moderadamente elevada y con signo negativo (-0.291). Se podría establecer a priori la existencia de un círculo vicioso en donde el analfabetismo y la falta de instrucción, limitan el ejercicio de los derechos ciudadanos (*libertades políticas*) de la población y consiguientemente, su acceso a servicios de infraestructura y seguridad social y por tanto, falta de interés en la contienda electoral. A este factor puede identificársele como factor de *atraso*.

Los factores (*variables subyacentes no observables*) descritos con anterioridad, pudieran interpretarse bajo el enfoque de las libertades instrumentales propuestas por Sen de la siguiente manera:

El Factor1 expresa el desarrollo en cuanto a *facilidades económicas*, debido al ingreso e índices de pobreza que lo caracterizan. Las condiciones de salud también son favorables, expresadas éstas, en términos de la más baja tasa de mortalidad infantil, índice de fecundidad e igualdad de géneros. Asimismo, el factor describe una gama amplia de *oportunidades sociales* e instrumentos de seguridad social, asociadas a altos niveles de educación superior, servicios básicos y acceso a las instituciones de protección civil. Las *libertades políticas* parecen no estar plenamente desarrolladas, debido al bajo peso que representa la variable que mide la participación electoral, en la descripción del factor.

Por otra parte, el Factor4 describe a una sociedad carente de *oportunidades sociales, facilidades económicas*, acceso limitado a servicios públicos, ignorancia y relativamente poca participación ciudadana en la elección de sus gobernantes. En resumen, una población que, parafraseando a Sen, carece de la...*capacidad de la persona de acrecentar su potencialidad para desempeñar una vida significativa que como sujeto, legítimamente valora.*

Los otros dos factores son de más difícil interpretación. El Factor2 plantea a una comunidad que aunque carece de *facilidades económicas* suficientes, así como de instrumentos adecuados de *seguridad social*, cuenta con una población medianamente educada; tal vez capaz de asumir una mayor participación en la actividad económica, de procurarse las oportunidades necesarias.

Finalmente, el Factor3 sugiere la necesidad de una renovación de la sociedad. La elevada carga en la variable descriptiva de la delincuencia, parece reflejar carencias en cuanto a *seguridad social*.

La Tabla 2.3.3 presenta una la interpretación de los factores bajo la perspectiva de la *falta de libertades* y las condiciones a procurar, a fin de ensancharlas. En la tercera columna se hace énfasis (con base en el análisis de cargas), en aquellas variables que *a priori*, parecen limitar el desarrollo en relación a los factores resultantes del ajuste del modelo y que fueron descritos con anterioridad.

Tabla 2.3.3

FACTOR	Nombre Descriptivo	Oportunidades para el Desarrollo
Factor1	*Desarrollo*	Fomento de una mayor participación ciudadana en el gobierno y desarrollo de la comunidad
Factor2	*Estancamiento*	Desarrollo de infraestructura de servicios básicos y de salud, estímulo a la educación superior y a la creación de oportunidades de empleo
Factor3	*Progreso incipiente*	Desarrollo infraestructura de servicios básicos y de salud y control de la delincuencia. Fortalecimiento de los mecanismos de transparencia y rendición de cuentas de los órganos de gobierno hacia los electores.
Factor4	*Pobreza*	Desarrollo de infraestructura de servicios básicos y de salud, estímulo a la educación básica y al empleo acorde con las capacidades y características actuales de la población

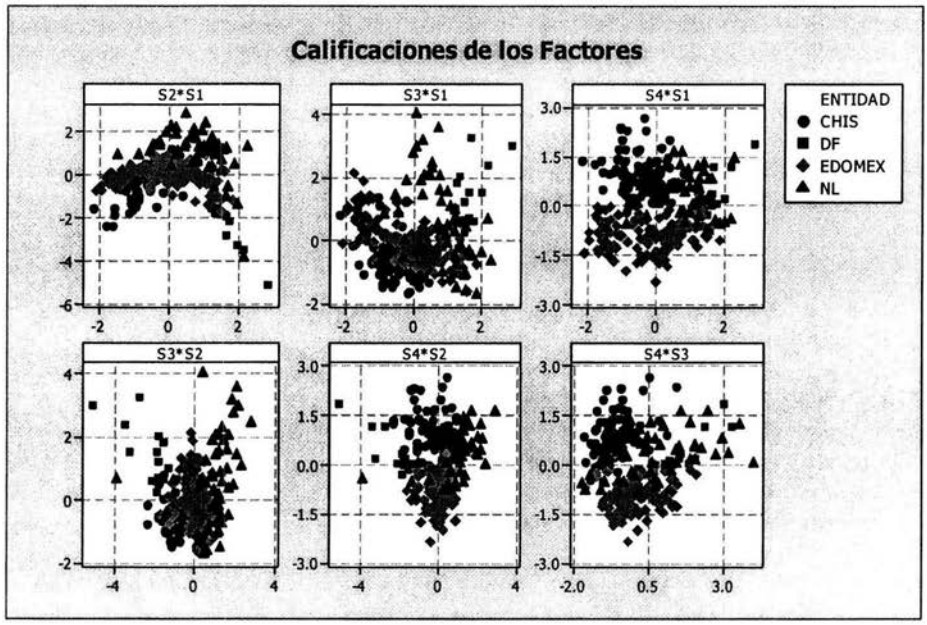
CALIFICACIÓN DE LOS FACTORES

El análisis de factores, como ya se ha comentado, tiene como propósito primordial evaluar si las interrelaciones entre un conjunto de variables observadas se explican en términos de un número pequeño de variables subyacentes no observables o factores; sin embargo, también resulta de utilidad estimar los valores denominados calificaciones de los factores, los cuales se utilizarán como variables en la determinación de grupos de individuos con características semejantes, mediante el ajuste de un modelo de análisis de conglomerados.

En la Gráfica 2.3.3, se representa la posición de cada uno de las unidades (municipios y delegaciones) con respecto a cada pareja de factores. Visualmente se pueden distinguir comunidades con grados característicos de desarrollo. Por ejemplo el espacio S1*S3 muestra una polarización muy evidente entre el estado de Chiapas y el DF. Asimismo, indica un rango de desarrollo medio-bajo para el Estado de México, contrastando con un rango medio-alto en el estado de Nuevo León. En el Análisis de Conglomerados, en la siguiente sección se aborda el problema de clasificación bajo un enfoque multivariado, utilizando las calificaciones de los cuatro factores S1, S2, S3 y S4. En esta sección trataremos de identificar formalmente la formación de conjuntos de unidades o conglomerados.

La posición de los individuos en el plano definido por cada pareja de calificación de factores, es consistente con la estructura resultante en componentes principales.

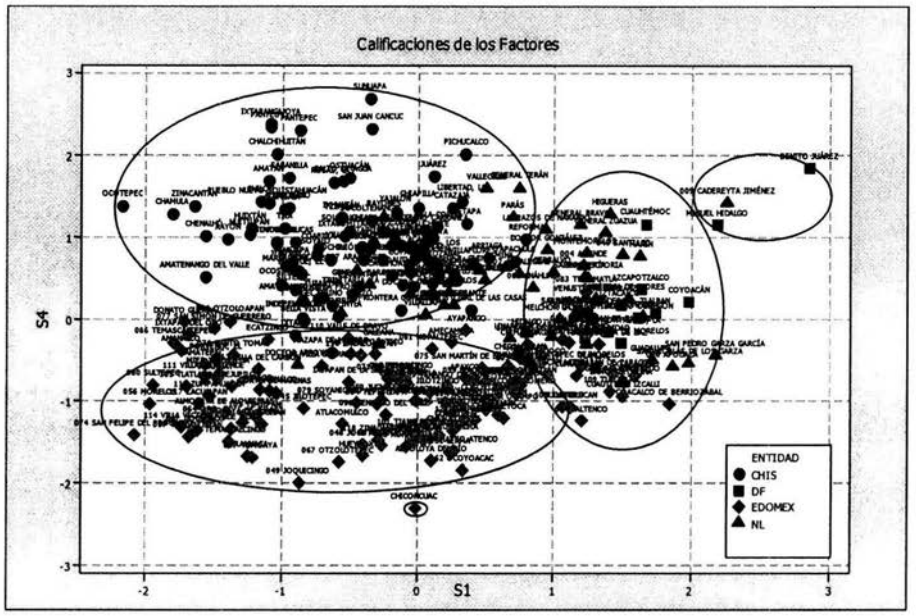
Gráfica 2.3.3



2.4 Análisis de Conglomerados

La inspección visual de la Gráfica 2.4.1 (plano S1-S4), indica la existencia de grupos afines (posiblemente 3 conglomerados) formados por: (1) los municipios del estado de Chiapas y algunos municipios del estado de Nuevo León (2) la mayoría de los municipios del Estado de México y (3) el conjunto de delegaciones del DF, así como posiblemente los municipios con mayor desarrollo (calificación más alta en S1) en los estados de Nuevo León y Estado de México. Por otra parte se observan algunos conjuntos de observaciones atípicas como son el caso de las delegaciones Miguel Hidalgo, Benito Juárez en el DF y el municipio de Cadereyta en el estado de NL (sección superior-derecha) y el municipio de Chiconcuac en el Estado de México (sección inferior-centro). En el análisis de conglomerados subsecuentes, se utilizarán las calificaciones S1, S2, S3 y S4, correspondientes a cada uno de los cuatro factores definidos en la sección anterior.

Gráfica 2.4.1



A continuación trataremos de descubrir, mediante el uso del análisis de conglomerados, la existencia y configuración de la estructura de los datos graficados en el plano S1-S4. En primer lugar se recurre al método de agrupación jerárquica, a fin de proponer un número razonable de grupos en la clasificación, para posteriormente determinar la formación de los grupos, mediante el método de K-medias, el cual minimiza la suma del cuadrado de las distancias en cada grupo, respecto a su correspondiente centroide (media), según se explicará en el capítulo 3.3. El número de grupos será confirmado a través de los métodos multivariados, mediante la inspección de los dendrogramas y las pruebas de hipótesis correspondientes.

La Tabla 2.4.1 reproduce la salida de computadora con la descripción de los diez últimos pasos en la fusión de conglomerados, empleando distancias euclidianas y liga promedio. En la formación de conglomerados, se utilizó el método de agrupamiento por liga promedio, debido a que ofrece un resultado intermedio entre liga simple y liga completa y una solución relativamente robusta a pequeñas alteraciones entre los individuos. Por otra parte, con un número de individuos por clasificar, cercano a las 300 observaciones, resultaría complicado identificar clasificaciones alternas resultantes, en su caso, mediante el empleo de diferentes métodos de clasificación.

Mientras que en cada paso previo al 287 y que conduce a la formación de tres conglomerados, el *nivel de similitud*[13] decrece en aproximadamente un punto porcentual, en este paso se observa un decremento de 66.58 % a 56.85 % (9.73 puntos

---

[13] El nivel de similitud $s(ij)$ entre dos conglomerados *i* y *j*, se define como $s(ij) = 100.\frac{(d_{\max} - d_{ij})}{d_{\max}}$

en donde d_{max} es la máxima distancia entre observaciones y d_{ij} es la distancia entre conglomerados.

porcentuales). Este salto es indicativo de que al nivel de **tres grupos** se tenga posiblemente el "*mejor corte*" en la segmentación de los individuos.

Tabla 2.4.1
Distancia Euclidiana, Liga Promedio
Etapas de Fusión de Grupos

Paso	Número de Conglom.	Nivel Similitud	Nivel Distancia	Conglom. Fusionado		Nuevo Conglom.	Número de obs. en nuevo Conglom.
280	10	73.4963	2.24082	169	179	169	4
281	9	72.1963	2.35073	2	13	2	105
282	8	70.1422	2.52441	2	123	2	124
283	7	70.1011	2.52788	145	229	145	99
284	6	69.7945	2.55380	124	173	124	16
285	5	68.4242	2.66965	2	145	2	223
286	4	67.2616	2.76795	1	2	1	269
287	3	66.5854	2.82512	169	176	169	5
288	2	56.8499	3.64824	1	124	1	285
289	1	36.1720	5.39650	1	169	1	290

Un procedimiento más formal para la determinación del número de grupos a considerar, se obtiene mediante el estadístico tipo F de Beale[14], definido como sigue:

$$F^{*} = \frac{(W_2 - W_1)}{W_1} \cdot \frac{(n-c_1)k_1}{(n-c_2)k_2 - (n-c_1)k_1}$$

$$k_1 = c_1^{-2/p} \quad k_2 = c_2^{-2/p}$$

$$W_1 = \sum_{r=1}^{c1} \sum_{q=1}^{n_r} (\mathbf{x}_{rq} - \bar{\mathbf{x}}_r)(\mathbf{x}_{rq} - \bar{\mathbf{x}}_r)$$

en donde W_1, W_2 representan la dispersión en cada agrupación, $\mathbf{x}_{rq}$ es el vector p-dimensional de observaciones del q-ésimo individuo en el grupo r y $\bar{\mathbf{x}}_r$ es el vector p-dimensional de medias para cada grupo r.

En este caso, c_1 y c_2 corresponden a dos distintas agrupaciones de datos o de observaciones, en donde la primera consta de c_1 grupos y la segunda de c_2, siendo $c_2 < c_1$. Si F^{*} es mayor que un punto crítico F, con $k_2(n-c_2) - k_1(n-c_1)$ grados de libertad para el numerador y $k_1(n-c_1)$ grados de libertad para el denominador, entonces se elegirá la agrupación con mayor número de grupos sobre aquella con menos agrupamientos.

A continuación se compara una estructura de agrupamiento con 4 grupos, respecto a una con 3 grupos. Para este caso:

[14] Everitt B.S. Landau S., Leese M. Cluster Analysis, Arnold, (2001)

$c_1 = 4, \quad c_2 = 3$

$W_1 = 884.373$

$W_2 = 889.743$

$p = 4$

$k_1 = 0.5$

$k_2 = 0.57735$

$$F^* = \frac{(889.743 - 884.373)}{889.743} \cdot \frac{(290-4)0.5}{(290-3)0.57735 - (290-4)0.5} = 0.0380$$

Con $k_2(n-c_2) - k_1(n-c_1) = 22.7$ grados de libertad en el numerador y $k_1(n-c_1) = 143$ grados de libertad en el denominador y con $\alpha = 5\%$, se obtiene un valor crítico[15] de 1.6172>0.0380 y por lo tanto podemos considerar que la estructura con 3 grupos representa adecuadamente los datos.

Los parámetros de la partición $c_2 = 3$, se presentan en la Tabla 3.4.2

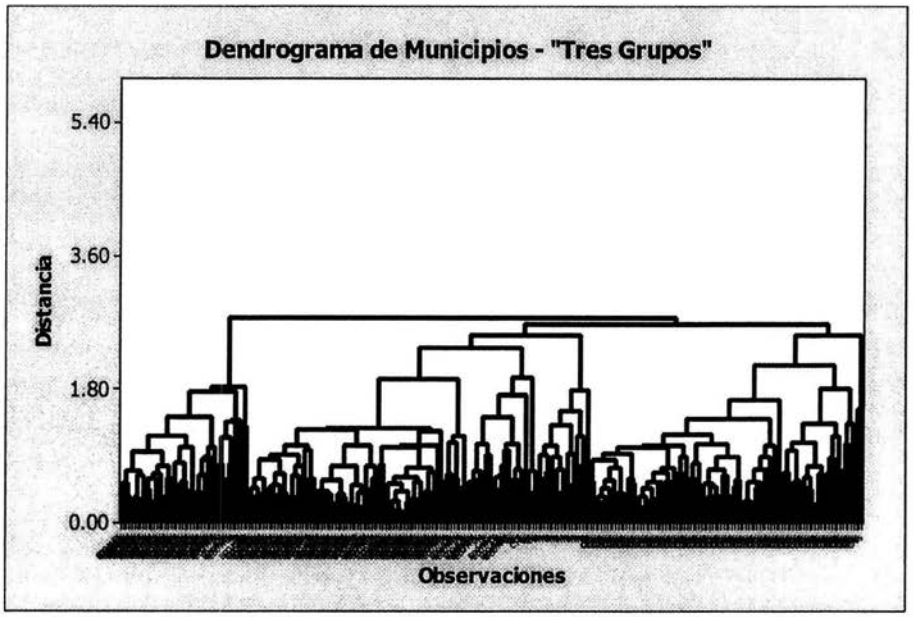
Tabla 2.4.2
Partición Final
Número de Conglomerados: 3

	Número de observaciones	Suma de cuadrados en el conglomerado
Conglomerado1	269	856.217
Conglomerado2	16	21.882
Conglomerado3	5	11.644

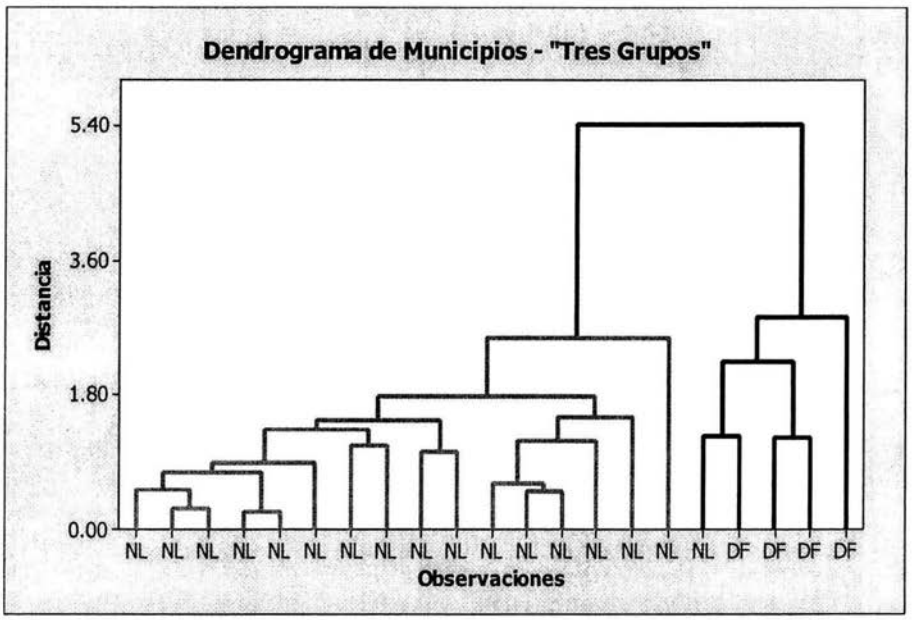
En los dendrogramas que a continuación se muestran, aparece un primer conglomerado con 269 observaciones con municipios de bajo desarrollo, primordialmente de los estados de Chiapas, México y algunas delegaciones del DF; un segundo conglomerado con 16 individuos, pertenecientes todos ellos al estado de Nuevo León y finalmente un tercer conglomerado que contiene a cuatro delegaciones del DF y un municipio de NL.

[15] El valor crítico del estadístico se obtuvo de Excel®

Gráfica 2.4.2



Gráfica 2.4.3



En una siguiente etapa, en la definición de una estructura de agrupación de individuos, se utiliza el método de K-Medias con tres y cuatro agrupaciones. Los resultados de la salida de Minitab (número de observaciones por conglomerado, suma de cuadrados respecto al centroide en los agrupamientos, distancias promedio y máxima, también dentro de los agrupamientos), se muestran a continuación:

Tabla 2.4.3
Análisis de Conglomerados K-medias: S1, S2, S3, S4

Número de Conglomerados: 3

	Número de observaciones	Suma de cuadrados en el conglomerado	Distancia promedio del centroide	Distancia máxima del centroide
Conglomerado1	99	158.569	1.156	2.687
Conglomerado2	102	449.089	1.866	6.386
Conglomerado3	89	131.343	1.157	2.301

Centroides del Conglomerado

Variable	Conglomerado1	Conglomerado2	Conglomerado3	Gran centroide
S1	-0.4704	1.0444	-0.6738	-0.0000
S2	-0.1154	0.0851	0.0309	-0.0000
S3	-0.6269	0.3895	0.2509	0.0000
S4	0.9511	-0.0379	-1.0145	-0.0000

Distancias Entre Centroides de los Conglomerados

	Conglomerado1	Conglomerado2	Conglomerado3
Conglomerado1	0.0000	2.0847	2.1672
Conglomerado2	2.0847	0.0000	1.9820
Conglomerado3	2.1672	1.9820	0.0000

Tabla 2.4.4
Análisis de Conglomerados K-medias: S1, S2, S3, S4

Número de Conglomerados: 4

	Número de observaciones	Suma de cuadrados en el conglomerado	Distancia promedio del centroide	Distancia máxima del centroide
Conglomerado1	42	138.625	1.428	6.296
Conglomerado2	92	163.623	1.208	3.666
Conglomerado3	97	190.993	1.255	3.934
Conglomerado4	59	138.653	1.329	4.050

Centroides del Conglomerado

Variable	Conglomerado1	Conglomerado2	Conglomerado3	Conglomerado4	Gran centroide
S1	-0.7527	0.2367	0.7063	-0.9945	-0.0000
S2	-1.1191	0.7319	-0.3451	0.2227	-0.0000
S3	-0.2157	-0.3866	-0.1403	0.9871	0.0000
S4	1.2926	0.6623	-0.8105	-0.6204	-0.0000

Distancias Entre Centroides de los Conglomerados

	Conglomerado1	Conglomerado2	Conglomerado3	Conglomerado4
Conglomerado1	0.0000	2.1981	2.6751	2.6392
Conglomerado2	2.1981	0.0000	1.9000	2.3038
Conglomerado3	2.6751	1.9000	0.0000	2.1265
Conglomerado4	2.6392	2.3038	2.1265	0.0000

Al igual que en el caso jerárquico, se compara mediante el seudo estadístico de Beale el ajuste para el caso de 3 y 4 agrupamientos. Con $W_1 = 631.894$ y $W_2 = 739.001$, el valor de $F^* = 1.0678 < 1.6172$. El resultado anterior es consistente con la segmentación de los individuos en solamente 3 agrupaciones. Las agrupaciones se muestran en la Tabla 2.4.5.

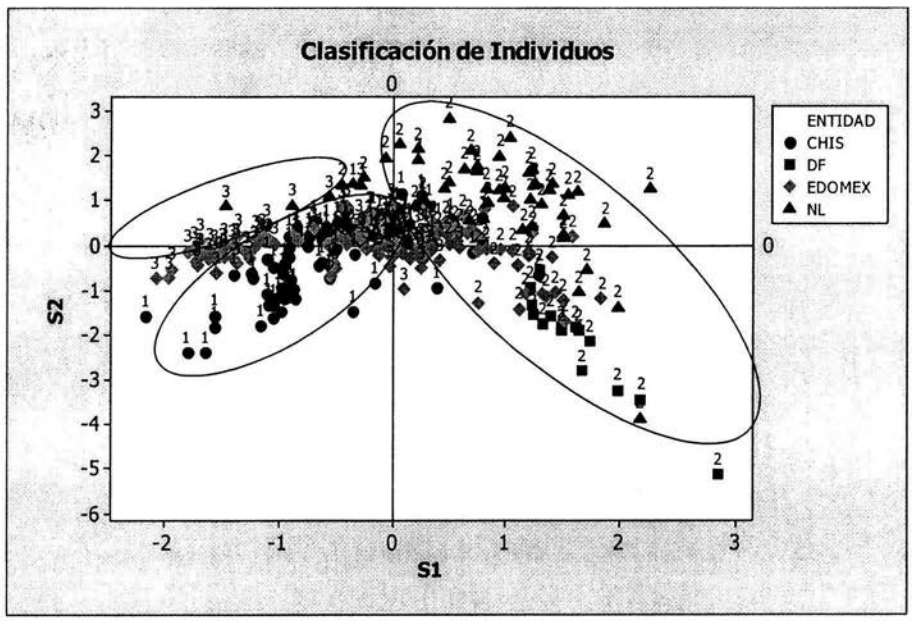
Tabla 2.4.5
Clasificación de Individuos en Tres Grupos

Grupo 1		Grupo 2		Grupo 3	
Municipio	Entidad	Municipio	Entidad	Municipio	Entidad
GENERAL ZARAGOZA	NL	ACOLMAN	EDOMEX	ACAMBAY	EDOMEX
ACACOYAGUA	CHIS	AMECAMECA	EDOMEX	ACULCO	EDOMEX
ACALA	CHIS	APAXCO	EDOMEX	ALMOLOYA DE ALQUISIRAS	EDOMEX
ACAPETAHUA	CHIS	ATIZAPÁN DE ZARAGOZA	EDOMEX	ALMOLOYA DE JUÁREZ	EDOMEX
ALTAMIRANO	CHIS	AYAPANGO	EDOMEX	ALMOLOYA DEL RÍO	EDOMEX
AMATÁN	CHIS	CHALCO	EDOMEX	AMANALCO	EDOMEX
AMATENANGO DE LA FRO.	CHIS	CHICOLOAPAN	EDOMEX	AMATEPEC	EDOMEX
AMATENANGO DEL VA.	CHIS	CHIMALHUACÁN	EDOMEX	ATENCO	EDOMEX
ANGEL ALBINO CORZO	CHIS	COACALCO DE BERRIOZÁ	EDOMEX	ATIZAPÁN	EDOMEX
ARRIAGA	CHIS	COCOTITLÁN	EDOMEX	ATLACOMULCO	EDOMEX
BELLA VISTA	CHIS	CUAUTITLÁN	EDOMEX	ATLAUTLA	EDOMEX
BERRIOZÁBAL	CHIS	CUAUTITLÁN IZCALLI	EDOMEX	AXAPUSCO	EDOMEX
BOCHIL	CHIS	ECATEPEC DE MORELOS	EDOMEX	CALIMAYA	EDOMEX
BOSQUE, EL	CHIS	HUEHUETOCA	EDOMEX	CAPULHUAC	EDOMEX
CACAHOATÁN	CHIS	HUIXQUILUCAN	EDOMEX	CHAPA DE MOTA	EDOMEX
CATAZAJÁ	CHIS	IXTAPALUCA	EDOMEX	CHAPULTEPEC	EDOMEX
CHALCHIHUITÁN	CHIS	JALTENCO	EDOMEX	CHIAUTLA	EDOMEX
CHAMULA	CHIS	MELCHOR OCAMPO	EDOMEX	CHICONCUAC	EDOMEX
CHENALHÓ	CHIS	METEPEC	EDOMEX	COATEPEC HARINAS	EDOMEX
CHIAPA DE CORZO	CHIS	NAUCALPAN DE JUÁREZ	EDOMEX	COYOTEPEC	EDOMEX
CHIAPILLA	CHIS	NEXTLALPAN	EDOMEX	DONATO GUERRA	EDOMEX
CHICOASÉN	CHIS	NEZAHUALCÓYOTL	EDOMEX	ECATZINGO	EDOMEX
CHICOMUSELO	CHIS	NICOLÁS ROMERO	EDOMEX	HUEYPOXTLA	EDOMEX
CHILÓN	CHIS	NOPALTEPEC	EDOMEX	ISIDRO FABELA	EDOMEX
CINTALAPA	CHIS	PAZ, LA	EDOMEX	IXTAPAN DE LA SAL	EDOMEX
COAPILLA	CHIS	SAN MARTÍN DE LAS PIR.	EDOMEX	IXTAPAN DEL ORO	EDOMEX
COMITÁN DE DOMÍNG.	CHIS	TECÁMAC	EDOMEX	IXTLAHUACA	EDOMEX
CONCORDIA, LA	CHIS	TEMAMATLA	EDOMEX	JILOTEPEC	EDOMEX
COPAINALÁ	CHIS	TEOLOYUCÁN	EDOMEX	JILOTZINGO	EDOMEX
ESCUINTLA	CHIS	TEOTIHUACÁN	EDOMEX	JIQUIPILCO	EDOMEX
FRONTERA COMALAPA	CHIS	TEPOTZOTLÁN	EDOMEX	JOCOTITLÁN	EDOMEX
FRONTERA HIDALGO	CHIS	TEXCOCO	EDOMEX	JOQUICINGO	EDOMEX
HUEHUETÁN	CHIS	TEZOYUCA	EDOMEX	JUCHITEPEC	EDOMEX
HUITIUPÁN	CHIS	TLALMANALCO	EDOMEX	LERMA	EDOMEX
HUIXTÁN	CHIS	TLALNEPANTLA DE BAZ	EDOMEX	MALINALCO	EDOMEX
HUIXTLA	CHIS	TOLUCA	EDOMEX	MEXICALTZINGO	EDOMEX
INDEPENDENCIA, LA	CHIS	TULTEPEC	EDOMEX	MORELOS	EDOMEX
IXHUATÁN	CHIS	TULTITLÁN	EDOMEX	OCOYOACAC	EDOMEX
IXTACOMITÁN	CHIS	VALLE DE CHALCO SOL.	EDOMEX	OCUILAN	EDOMEX
IXTAPANGAJOYA	CHIS	ABASOLO	NL	ORO, EL	EDOMEX
JIQUIPILAS	CHIS	AGUALEGUAS	NL	OTUMBA	EDOMEX
JITOTOL	CHIS	ALDAMAS, LOS	NL	OTZOLOAPAN	EDOMEX
JUÁREZ	CHIS	ALLENDE	NL	OTZOLOTEPEC	EDOMEX
LIBERTAD, LA	CHIS	ANÁHUAC	NL	OZUMBA	EDOMEX
MAPASTEPEC	CHIS	APODACA	NL	PAPALOTLA	EDOMEX
MARGARITAS, LAS	CHIS	ARAMBERRI	NL	POLOTITLÁN	EDOMEX
MAZAPA DE MADERO	CHIS	BUSTAMANTE	NL	RAYÓN	EDOMEX
MAZATÁN	CHIS	CADEREYTA JIMÉNEZ	NL	SAN ANTONIO LA ISLA	EDOMEX
METAPA	CHIS	CARMEN	NL	SAN FELIPE DEL PROGRESO	EDOMEX
MOTOZINTLA	CHIS	CERRALVO	NL	SAN MATEO ATENCO	EDOMEX
OCOSINGO	CHIS	CHINA	NL	SAN SIMÓN DE GUERRERO	EDOMEX
OCOTEPEC	CHIS	CIÉNEGA DE FLORES	NL	SANTO TOMÁS	EDOMEX
OCOZOCOAUTLA DE ESPI.	CHIS	DOCTOR COSS	NL	SOYANIQUILPAN DE JUÁREZ	EDOMEX
OSTUACÁN	CHIS	DOCTOR GONZÁLEZ	NL	SULTEPEC	EDOMEX
OSUMACINTA	CHIS	GARCÍA	NL	TEJUPILCO	EDOMEX
OXCHUC	CHIS	GENERAL BRAVO	NL	TEMASCALAPA	EDOMEX
PALENQUE	CHIS	GENERAL ESCOBEDO	NL	TEMASCALCINGO	EDOMEX
PANTELHÓ	CHIS	GENERAL TERÁN	NL	TEMASCALTEPEC	EDOMEX
PANTEPEC	CHIS	GENERAL TREVIÑO	NL	TEMOAYA	EDOMEX
PICHUCALCO	CHIS	GENERAL ZUAZUA	NL	TENANCINGO	EDOMEX
PIJIJIAPAN	CHIS	GUADALUPE	NL	TENANGO DEL AIRE	EDOMEX
PUEBLO NUEVO SOLIST.	CHIS	HERRERAS, LOS	NL	TENANGO DEL VALLE	EDOMEX
RAYÓN	CHIS	HIDALGO	NL	TEPETLAOXTOC	EDOMEX
REFORMA	CHIS	HIGUERAS	NL	TEPETLIXPA	EDOMEX
ROSAS, LAS	CHIS	HUALAHUISES	NL	TEQUIXQUIAC	EDOMEX
SABANILLA	CHIS	JUÁREZ	NL	TEXCALTITLÁN	EDOMEX

SALTO DE AGUA	CHIS	LAMPAZOS DE NARANJO	NL	TEXCALYACAC	EDOMEX
SAN CRISTÓBAL DE	CHIS	LINARES	NL	TIANGUISTENCO	EDOMEX
SAN FERNANDO	CHIS	MARÍN	NL	TIMILPAN	EDOMEX
SAN JUAN CANCUC	CHIS	MELCHOR OCAMPO	NL	TLATLAYA	EDOMEX
SAN LUCAS	CHIS	MINA	NL	TONATICO	EDOMEX
SILTEPEC	CHIS	MONTEMORELOS	NL	VALLE DE BRAVO	EDOMEX
SIMOJOVEL	CHIS	MONTERREY	NL	VILLA DE ALLENDE	EDOMEX
SOCOLTENANGO	CHIS	PARÁS	NL	VILLA DEL CARBÓN	EDOMEX
SOLOSUCHIAPA	CHIS	PESQUERÍA	NL	VILLA GUERRERO	EDOMEX
SOYALÓ	CHIS	RAMONES, LOS	NL	VILLA VICTORIA	EDOMEX
SUCHIAPA	CHIS	RAYONES	NL	XALATLACO	EDOMEX
SUCHIATE	CHIS	SABINAS HIDALGO	NL	XONACATLÁN	EDOMEX
SUNUAPA	CHIS	SALINAS VICTORIA	NL	ZACAZONAPAN	EDOMEX
TAPACHULA	CHIS	SAN NICOLÁS DE LOS GAR	NL	ZACUALPAN	EDOMEX
TAPALAPA	CHIS	SAN PEDRO GARZA GARCÍA	NL	ZINACANTEPEC	EDOMEX
TAPILULA	CHIS	SANTA CATARINA	NL	ZUMPAHUACÁN	EDOMEX
TECPATÁN	CHIS	SANTIAGO	NL	ZUMPANGO	EDOMEX
TENEJAPA	CHIS	VALLECILLO	NL	DOCTOR ARROYO	NL
TEOPISCA	CHIS	VILLALDAMA	NL	GALEANA	NL
TILA	CHIS	ÁLVARO OBREGÓN	DF	ITURBIDE	NL
TONALÁ	CHIS	AZCAPOTZALCO	DF	MIER Y NORIEGA	NL
TRINITARIA, LA	CHIS	BENITO JUÁREZ	DF	IXTAPA	CHIS
TUMBALÁ	CHIS	COYOACÁN	DF	PORVENIR, EL	CHIS
TUXTLA CHICO	CHIS	CUAJIMALPA DE MORELOS	DF		
TUZANTÁN	CHIS	CUAUHTÉMOC	DF		
TZIMOL	CHIS	GUSTAVO A. MADERO	DF		
UNIÓN JUÁREZ	CHIS	IZTACALCO	DF		
VENUSTIANO CARR	CHIS	IZTAPALAPA	DF		
VILLA COMALTITL.	CHIS	MAGDALENA CONTRERAS	DF		
VILLA CORZO	CHIS	MIGUEL HIDALGO	DF		
VILLAFLORES	CHIS	MILPA ALTA	DF		
YAJALÓN	CHIS	TLÁHUAC	DF		
ZINACANTÁN	CHIS	TLALPAN	DF		
		VENUSTIANO CARRANZA	DF		
		XOCHIMILCO	DF		
		TUXTLA GUTIÉRREZ	CHIS		

Las características generales de los tres Grupos que aparecen en la segmentación de la Tabla 2.4.5 se pueden inferir con la asistencia de las Gráficas 2.4.4, 2.4.5 y 2.4.6.

Gráfica 2.4.4

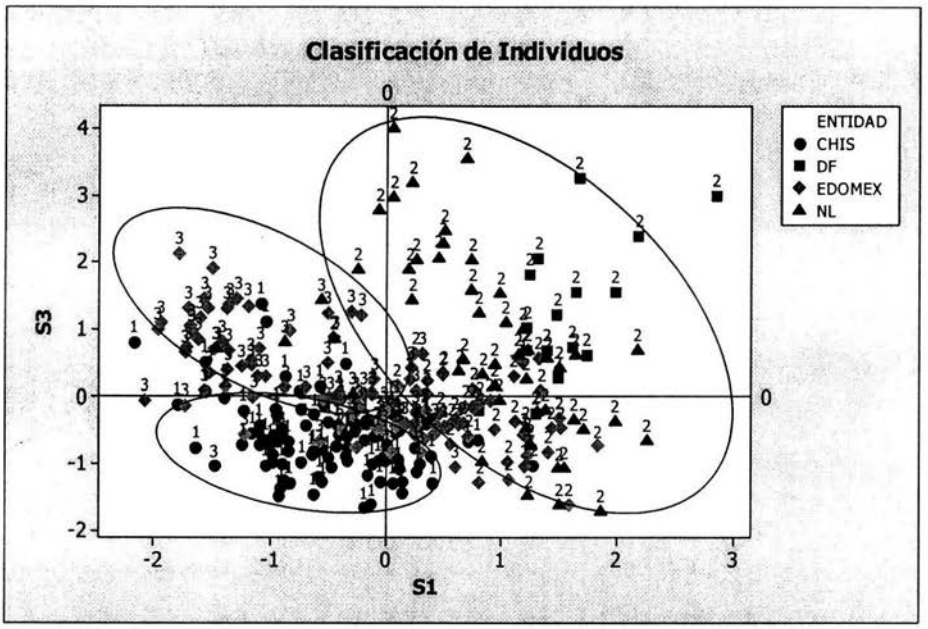


De los 99 individuos que conforman el Grupo1, 98 corresponden al estado de Chiapas y uno, General Zaragoza, al estado de Nuevo León.

Este grupo se caracteriza por un valor de alto a moderadamente alto y con signo negativo en la calificación correspondiente al Factor1 (desarrollo) e igualmente de alta a moderadamente alta, pero con signo positivo en la calificación asociada al Factor4 (pobreza). En el plano [S1, S4], en la Figura 2.4.6, el conglomerado formado por los individuos del Grupo1, se amalgama claramente en el cuadrante superior izquierdo, el cual podríamos definir como el espacio de individuos con máximo atraso.

El Grupo2 comprende a 102 individuos; es decir la totalidad de las 16 delegaciones del Distrito Federal, Tuxtla Gutiérrez, capital del estado de Chiapas, 39 de los 122 municipios del Estado de México (32 %) y 46 de los 50 municipios del estado de Nuevo León (92 %), que figuran en la muestra.

Gráfica 2.4.5



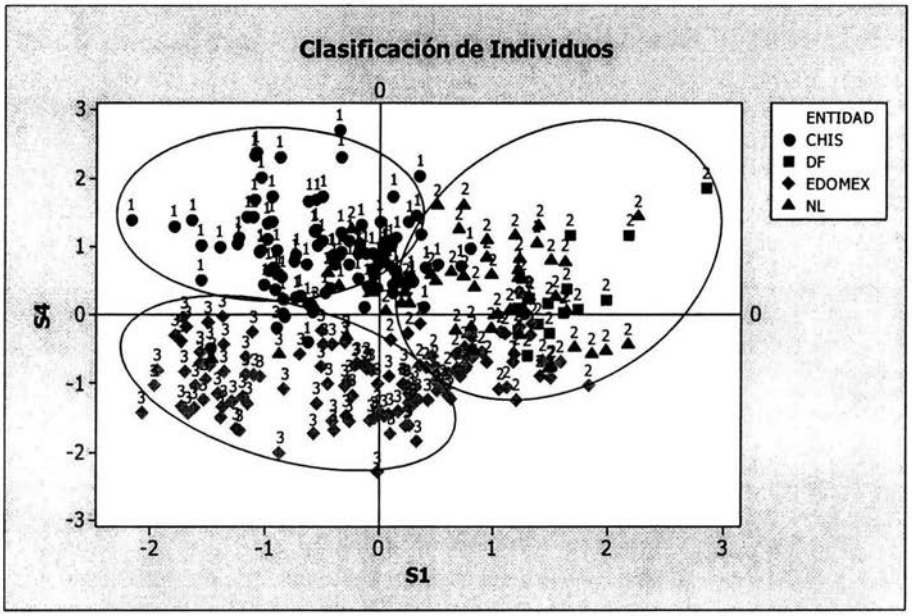
En la misma Figura 2.4.6, el conglomerado formado por este grupo ocupa los cuadrantes inferior y superior derecho; es decir, con una calificación alta a muy alta y con signo positivo, en el Factor1 y de moderadamente baja a moderadamente elevada en el Factor4.

El mismo patrón se observa en la Gráfica 2.4.4 y Gráfica 2.4.5 en donde se señalan los planos [S1, S2] y [S1, S3].

Debido a la posición que ocupa este conglomerado respecto a la calificación relativa al Factor1, se puede concluir que los individuos que lo forman son los que se caracterizan por un mayor desarrollo relativo dentro de la muestra, aunque se tienen presentes elementos moderados de pobreza, progreso incipiente y estancamiento.

El Grupo3 se caracteriza por calificaciones semejantes a las del Grupo1 en cuanto al Factor1 (desarrollo); es decir, de moderada a fuertemente negativa. Sin embargo, es posible atribuir un mayor desarrollo a los individuos de este grupo, respecto a los del Grupo1, debido al signo claramente negativo en la calificación del Factor4 (Gráfica 2.4.6) y positivo en la calificación correspondiente tanto al Factor2, como al Factor3, según puede observarse en las Gráficas 2.4.4 y 2.4.5.

Gráfica 2.4.6



Conclusiones

A partir de las 21 variables seleccionadas para el estudio de la muestra que incluye 290[16] de los 305 municipios o delegaciones comprendidos en los estados de Nuevo León, Estado de México, Chiapas y Distrito Federal, se pueden obtener las siguientes conclusiones.

El desarrollo social puede describirse en términos de cuatro *factores*, los cuales delinean condiciones o estados característicos de desarrollo de la población bajo estudio, de acuerdo al enfoque de "libertades", propuesto por Sen.

Un primer factor, el cual denominamos Factor1, representa un continuo, en donde en un extremo, destacan con alto peso y signo positivo las variables observables asociadas a un nivel elevado de desarrollo, en cuanto a: *facilidades económicas*, *oportunidades sociales* y de *seguridad social* y con el signo opuesto, en el extremo contrario, aquellas variables observables que mejor describen la correspondiente polaridad. En el caso del Factor1, el cual describe el fenómeno del *"Desarrollo"*, se presenta el nivel más alto de desarrollo material y educativo de la población. Destaca sin embargo, mediante este factor, la baja carga asociada a la variable relacionada con la participación ciudadana en la contienda electoral del 2000.

El Factor2, describe un nivel que se ha denominado *"Estancamiento"*, por caracterizarse por grados intermedios de educación, acceso limitado a los servicios de infraestructura básica y seguridad social y oportunidades de empleo e ingreso insuficientes. Este factor, sin representar condiciones de extrema pobreza, tampoco refleja condiciones de bienestar intermedio, equiparable con las que expresa el siguiente factor.

El Factor3 está definido por cinco variables. La variable con mayor peso corresponde a la participación electoral de la población. Las otras cuatro variables que definen el factor son el nivel de discapacidad de la población, la edad y el número de ocupantes por vivienda, con una carga elevada y con signo negativo. Por otra parte el ingreso, aunque con carga positiva, ésta es de magnitud media. Asimismo, las variables que miden la educación, los servicios y la ocupación no parecen ser influyentes en la definición del factor. El conjunto de las características anteriores, parece describir una población con edad avanzada y sin acceso a servicios básicos de infraestructura y de salud, lo cual posiblemente explique la influencia de la variable que mide la discapacidad en la definición del factor. Por otra parte, la relevancia de la variable asociada a la delincuencia, pudiera corresponder a una población en donde las oportunidades de empleo, sean insuficientes. El Factor3, por las características anteriores, así como la carga moderadamente alta en la variable que mide el ingreso, parece representar un estado de *progreso incipiente.*

Finalmente, el Factor4 describe las condiciones de mayor pobreza dentro de la población; es decir ignorancia, carencia de servicios, poca participación ciudadana en la elección de sus gobernantes, elevados índices de pobreza y consiguientemente límites

[16] Los municipios no incluidos, corresponden en su totalidad al estado de Chiapas. Estos municipios fueron excluidos en vista de que no se dispone de información censal de los mismos.

extremos en cuanto a capacidades y oportunidades para el desarrollo de la población. Al Factor4 se le ha denominado *"Pobreza"*.

Las consideraciones anteriores, plantean desde el punto de vista del desarrollo como libertad, la necesidad (y oportunidad) de expandir las capacidades de la población a través de arreglos institucionales que conduzcan a avances en su desarrollo, atendiendo a las características especificas de la población.

De esta manera, las poblaciones que han adquirido un nivel superior en cuanto a su desarrollo material, deberán, a fin de continuar el proceso de expansión de sus libertades, asumir su responsabilidad como agentes de cambio, avanzando en cuanto a su participación en la vida política en sus respectivas comunidades; mediante el ejercicio del voto, y la búsqueda de una mayor transparencia y responsabilidad, así como de rendición de cuentas de los gobernantes y partidos políticos hacia la sociedad, en el ejercicio de la función pública.

En las comunidades con estancamiento, deberá estimularse el avance en la educación superior, el desarrollo de infraestructura en cuanto a servicios básicos, y crear más y mejores oportunidades que favorezcan la actualización de las capacidades la población, principalmente mediante el fomento al empleo.

En las comunidades con progreso incipiente, se requiere fortalecer los servicios públicos y propiciar una respuesta más eficaz de los partidos en el poder, en la satisfacción de las necesidades básicas de la población.

En las comunidades más pobres, se requiere realizar un gran esfuerzo para el combate del analfabetismo y el fomento a la educación básica, la procuración de mejores condiciones de salud y el empleo acorde a las características y potencialidades concretas de la población. El fomento del desarrollo social, abarca por lo tanto a todo el conjunto de actores e instituciones de la sociedad.

Es importante recordar que la expansión de las libertades anteriores, no representa bajo el punto de vista de Sen un mero conjunto instrumental para promover el desarrollo de la población, sino que **representan el desarrollo o libertad de la persona en sí misma**.

El resultado empírico de esta investigación concuerda con la visión de Sen y conduce a una perspectiva amplia y quizás inagotable en la conquista de mayores libertades; es decir de desarrollo. De esta manera, las poblaciones que han alcanzado estados superiores en cuanto a desarrollo educativo y material podrán continuar avanzando en la expansión de sus libertades a través de una mayor participación directa o indirecta en el gobierno de su comunidad. Por otra parte, en aquellas con desarrollo incipiente habrá que atender las libertades más básicas de la población; es decir salud, educación y vivienda, además del empleo.

En otro orden de ideas, al posicionar a los individuos en un contexto multidimensional, por ejemplo con dimensiones contrastantes como (S1, S4), las cuales representan las calificaciones correspondientes a los factores: Factor1 y Factor2, se observan municipios o delegaciones con una calificación notoriamente alta en ambos factores (Miguel Hidalgo y Benito Juárez en el DF y Cadereyta Jiménez en el estado de NL).

Habrá que investigar la naturaleza de este fenómeno aparentemente paradójico, en donde las condiciones de desarrollo se manifiestan como la coexistencia de condiciones en apariencia, mutuamente excluyentes. Al respecto, Garajedaghi (1999)[17] al referirse al principio "*multidimensionalidad*" como una de los principios fundamentales de los sistemas sociales, concluye que esta propiedad produce que las tendencias opuestas no solamente coexistan e interaccionen entre sí, sino que formen relaciones complementarias. De esta manera el *Desarrollo* y *Pobreza* representan dos tendencias complementarias concretas que caracterizan una condición o estado particular con un dominio distinto al de las dos dimensiones unidimensionales consideradas por separado. Ackoff (1978)[18] aborda el mismo problema al referirse a la condición en la que "las partes de un sistema que por separado carecen de factibilidad, pueden dar lugar a conjuntos factibles".

El análisis de conglomerados, mediante el uso de métodos jerárquicos y K-Medias, conduce a la agrupación de los individuos en tres grandes grupos. Al eliminar individuos con observaciones faltantes, se obtiene una muestra de 290 municipios o delegaciones, los cuales se clasifican de la siguiente manera:

	Grupo 1 Desarrollo Bajo	Grupo 2 Desarrollo Alto	Grupo 3 Desarrollo Medio	Total
CHIS	98	1	2	101
EDOMEX	-	39	83	122
NL	1	46	4	51
DF	-	16	-	16
Total	99	102	89	290

En el caso de Chiapas, 98 de los 101 municipios incluidos en la muestra son de bajo desarrollo. Bajo el Grupo 2 de Desarrollo Alto se encuentra Tuxtla Gutiérrez, capital del estado.

El Estado de México muestra un 32 % de sus entidades con desarrollo alto y el 68 % restante con desarrollo medio. Los municipios de mayor desarrollo son los que circundan al norte y al este al DF, así como Toluca, capital del estado.

Los municipios del estado de Nuevo León, el 90 % clasifican como de alto desarrollo, 8 % con desarrollo medio (Doctor Arroyo, Galeana, Iturbide y el municipio de Mier y Noriega) y el 2 % restante (General Zaragoza), con bajo desarrollo. Las entidades de más alto desarrollo corresponden geográficamente a las que se ubican en el noreste del estado.

Finalmente, la totalidad de las delegaciones del DF se pueden clasificar como de alto desarrollo.

Si bien bajo el enfoque del presente trabajo, no se pretende establecer un ordenamiento estricto de los individuos en la muestra en base a su nivel desarrollo, la ubicación de los municipios y delegaciones resultante, en la dimensión de la calificación correspondiente al Factor1 (*desarrollo*), es en general concordante con investigaciones recientemente

[17] Garajedaghi Jamshid. 1999. *Systems Thinking, Managing Chaos and Complexity, A Platform for Designing Business Architecture*. Butterworth Heinemann

[18] Ackoff, Rusell L. 1978. *The Art of Problem Solving*. New York: John Wiley & Sons

realizadas por PNUD, en cuanto al Índice de Desarrollo Humano Municipal en México[19], de reciente publicación. Asimismo, destaca el contraste existente en cuanto al nivel de desarrollo, tanto entre las entidades federales consideradas dentro de la muestra, como entre los municipios y delegaciones que las conforman. Es necesario mencionar nuevamente que los niveles de desarrollo observados, no son comparables con poblaciones ajenas, debido a la existencia de diferencias evidentes, en relación al significado, contenido informático e interpretación de las variables observables.

En cuanto a los métodos multivariados utilizados, el Análisis de Componentes Principales indica que los datos observables pueden reducir su dimensionalidad en un nuevo conjunto de componentes principales, sin gran pérdida de información. El Análisis de Factores muestra una gran utilidad en la identificación de los factores que describen el fenómeno del desarrollo y en la explicación de la estructura de varianza-covarianza de los datos originales. En el Análisis de Factores, no se encontraron factores triviales; el número de factores es concordante con la teoría económica y las entradas de la matriz de residuos para el modelo de cuatro factores, son generalmente bajas (hasta en un orden de magnitud), respecto al modelo de tres factores, lo cual respalda la descripción del fenómeno estudiado en términos de los cuatro factores propuestos.

Por otra parte, el Análisis de Conglomerados resulta una herramienta potente en la clasificación de individuos, debido al nivel de similitud que muestran los individuos incluidos en cada uno de los tres grupos prescritos y posiblemente en una investigación posterior, en el desarrollo de programas comunitarios y diseño de estructuras institucionales específicos que favorezcan el desarrollo de manera particular en los municipios, a través de su transición por *estados*[20] de mayor desarrollo o *"libertad"*.

Finalmente, en general, los resultados empíricos bajo el enfoque multivariado del presente estudio, concuerdan con la perspectiva de Sen. A través de ésta, contemplamos al desarrollo como un proceso de expansión de las *libertades* de la persona y de la sociedad en su conjunto. Este punto de vista contrasta con una visión más estrecha del desarrollo, la cual lo identifica con el crecimiento del producto interno bruto, incremento en el ingreso personal, la industrialización, el avance tecnológico o la modernización de la sociedad, etc. Todos estos factores, aunque indudablemente pueden ser *medios* entre otros elementos para alcanzar el desarrollo, éste dependerá de otros elementos como la salud, la educación y el ejercicio de los derechos ciudadanos, entre otros, los cuales fueron tomados en consideración, mediante la selección del conjunto de variables empleadas en la descripción de los individuos.

---

[19] Programa de las Naciones Unidas para el Desarrollo. 2004. Índice de Desarrollo Humano Municipal en México. Resultados parciales publicados en el periódico *Reforma*. 26 de octubre, 2004

[20] Bunge, Mario.1999. *Sistemas Sociales y Filosofía.* Buenos Aires: Editorial Sudamericana

Capítulo 3

Descripción de la Herramientas Estadísticas Empleadas

3.1 Análisis de Componentes Principales

INTRODUCCIÓN

El análisis de componentes principales tiene como propósito explicar las relaciones entre un conjunto de p variables correlacionadas a través de k combinaciones lineales de las p variables originales $(k \leq p)$. Sus objetivos generales son: (1) reducción de la dimensionalidad de los datos originales y (2) interpretación, en la medida de lo posible, de las nuevas k variables.

Aun cuando se requieran p componentes principales para reproducir la variabilidad total del conjunto de datos, frecuentemente gran parte de esta variabilidad podrá capturarse mediante un número reducido de k componentes principales. En este caso habrá casi siempre tanta información respecto a la variabilidad del conjunto de datos en los k componentes principales, como en la p variables originales. Los k componentes podrán reemplazar a las p variables iniciales, de tal manera que el conjunto de datos originales, consistente en n mediciones sobre p variables se reduce a un conjunto de datos consistente en n mediciones sobre las k componentes principales resultantes.

Si el conjunto de variables originales arriba considerado está mutuamente no correlacionado, entonces las componentes principales son las mismas que las variables originales, pero arregladas en orden decreciente de varianza. Cuando este es el caso, el análisis de componentes principales no es de ayuda, ya que no existe manera de reducir la dimensionalidad del conjunto original de datos.

Un análisis de componentes principales a menudo revela relaciones que previamente no habían sido consideradas y permite realizar interpretaciones que de otra manera no hubieran quedado al descubierto.

El análisis de componentes principales representa un medio más que un fin en sí mismo, ya que frecuentemente se utiliza como un paso preliminar en investigaciones subsecuentes; es decir, se trata de una técnica exploratoria. Por ejemplo, las variables componentes principales pueden servir de entrada en un análisis de regresión múltiple o análisis de conglomerados. Más aún, es un paso previo al análisis de factores considerado más adelante.

COMPONENTES PRINCIPALES POBLACIONALES

Desde un punto de vista algebraico, los componentes principales son combinaciones lineales de p variables aleatorias originales $X_1, X_2, \ldots, X_p$. Geométricamente, estas combinaciones lineales representan la selección de un nuevo sistema de coordenadas, el cual se obtiene mediante la rotación del sistema original con $X_1, X_2, \ldots, X_p$ como los ejes

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

Los componentes principales se calculan a partir de la matriz de covarianza $\boldsymbol{\Sigma}$ (o de la matriz de correlación $\boldsymbol{\rho}$) de $X_1, X_2, \ldots, X_p$. Su desarrollo no requiere del supuesto de normalidad multivariada, excepto en el caso en que se quiera hacer inferencia estadística con dichos componentes.

Supongamos que $\mathbf{X}' = [X_1, X_2, \cdots, X_p]$ es un vector aleatorio p-dimensional con media $\boldsymbol{\mu}$ y con matriz de covarianza $\boldsymbol{\Sigma}$ cuyos eigenvalores son $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.
Cada componente principal Y_j será una combinación lineal de las $X's$, de manera que

$$\begin{aligned} Y_j &= a_{1j}X_1 + a_{2j}X_2 + \cdots + a_{pj}X_p \\ &= \mathbf{a}_j'\mathbf{X} \end{aligned} \qquad (1)$$

donde $\mathbf{a}_j' = [a_{1j}, a_{2j}, \cdots, a_{pj}]$ $j = 1, 2, \cdots, p$ es un vector de constantes

La varianza de Y_j y la covarianza de Y_j, Y_k $(j \neq k)$ son:

$$Var(Y_j) = \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_i \quad j = 1, 2, \cdots, p \qquad (2)$$

$$Cov(Y_jY_k) = \mathbf{a}_i'\boldsymbol{\Sigma}\mathbf{a}_k \quad j, k = 1, 2, \cdots, p \qquad (3)$$

Los componentes principales son aquellas combinaciones lineales no correlacionadas $Y_1, Y_2, \ldots, Y_p$ cuyas varianzas en (2) maximicen la varianza de $\mathbf{a}_j'\mathbf{X}$ $(j = 1, 2, \cdots, p)$.
El primer componente principal es la combinación lineal con máxima varianza. Es decir, los coeficientes $\mathbf{a}_{i1}$ $i = 1, 2, \cdots, p$ serán tales que maximicen $Var(\mathbf{Y}_1) = \mathbf{a}_1'\boldsymbol{\Sigma}\mathbf{a}_1$.
Puesto que ésta puede incrementarse multiplicando cualquier a_1 por una constante, para eliminar esta indeterminación es conveniente restringir el análisis a vectores de norma uno. De acuerdo a lo anterior se define:

Primer componente principal = combinación lineal de $\boldsymbol{a'_1X}$ tal que maximiza
Var $(\boldsymbol{a'_1X})$ sujeto a: $\boldsymbol{a'_1a_1} = 1$
Segundo componente principal = combinación lineal $\boldsymbol{a'_2X}$ que maximiza
Var $(\boldsymbol{a'_2X})$ sujeto a: $\boldsymbol{a'_2a_2} = 1$ y
Cov $(\boldsymbol{a'_1X, a'_2X}) = 0$.

En el i-ésimo paso,
el i-ésimo componente principal = combinación lineal $\boldsymbol{a'_iX}$ que maximiza
Var $(\boldsymbol{a'_iX})$ sujeto a: $\boldsymbol{a'_ia_i} = 1$ y
Cov $(\boldsymbol{a'_iX, a'_kX}) = 0$ para $i > k$

A continuación se enuncian algunos resultados importantes en el análisis de componentes principales. La demostración de estos resultados se puede encontrar en Johnson (2002)[21].

[21] Johnson, Richard A., Wichern, Dean W. 2002. *Applied Multivariate Statistical Analysis, Fifth Edition.* New Jersey: Prentice Hall.

Resultado 1. Sea $\boldsymbol{\Sigma}$ la matriz de covarianza asociada al vector aleatorio $\mathbf{X}' = \left[X_1, X_2, \cdots, X_p\right]$ cuyos eigenvalores son $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, ya que $\boldsymbol{\Sigma}$ es semidefinida positiva y sean $\mathbf{e}_1, \mathbf{e}_2, \cdots \mathbf{e}_p$ sus eigenvectores asociados, $\mathbf{e}_i = \left[e_{1i}, \cdots e_{pi}\right]$ $i = 1, 2, \cdots, p$ y $\mathbf{e}_i'\mathbf{e}_i = 1$. Entonces el i-ésimo componente principal estará dado por

$$Y_i = \mathbf{e}_i'\mathbf{X} = e_{1i}X_1 + e_{2i}X_2 + \cdots + e_{pi}X_p, \quad i = 1, 2, \cdots, p \qquad (4)$$

Con las siguientes características,

$$\begin{aligned} Var(Y_i) &= \mathbf{e}_i'\Sigma\mathbf{e}_i = \lambda_i \quad i = 1, 2, \cdots, p \\ Cov(Y_i, Y_k) &= \mathbf{e}_i'\Sigma\mathbf{e}_k = 0 \quad i \neq k \end{aligned} \qquad (5)$$

Es decir, los vectores $\mathbf{e}_i$ que determinan los coeficientes de las combinaciones lineales para formar las variables componentes principales son los eigenvectores unitarios de la matriz de covarianzas $\boldsymbol{\Sigma}$ de $\mathbf{X}$. El Resultado 1 nos indica que los componentes principales están no correlacionados y tienen varianzas iguales a los eigenvalores de $\boldsymbol{\Sigma}$.

Resultado 2. Sea $\mathbf{X}' = \left[X_1, X_2, \cdots, X_p\right]$ el vector aleatorio con matriz de covarianza $\boldsymbol{\Sigma}$, cuyos eigenvalores son $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ y sean $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p$ sus eigenvectores asociados, $\mathbf{e}_i = \left[e_{1i}, \cdots e_{pi}\right]$ $i = 1, 2, \cdots, p$ y $\mathbf{e}_i'\mathbf{e}_i = 1$.

Sean $\mathbf{Y}_1 = \mathbf{e}_1'\mathbf{X}$, $\mathbf{Y}_2 = \mathbf{e}_2'\mathbf{X}, \cdots, \mathbf{Y}_p = \mathbf{e}_p'\mathbf{X}$, las variables componentes principales. La matriz de covarianza de $\mathbf{Y}$ se denota por $\boldsymbol{\Lambda}$ y está dada por

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_p \end{pmatrix}$$

La matriz es diagonal, debido a que los componentes se han elegido de manera que no estén correlacionados. Entonces los eigenvalores se interpretan como las respectivas varianzas de las componentes y se obtienen las siguientes igualdades:

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} Var(Y_i)$$

en donde σ_{ii} es la varianza de las variables originales. Por lo tanto

$$\sum_{i=1}^{p} Var(X_i) = tr(\boldsymbol{\Sigma}) = tr(\boldsymbol{\Lambda}) = \sum_{i=1}^{p} Var(Y_i)$$

El Resultado 2 es muy importante, ya que indica que la variabilidad total de las variables originales, $\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}$, es decir, la varianza total de la población está dada por

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \lambda_1 + \lambda_2 + \cdots + \lambda_p \qquad (6)$$

y consecuentemente la proporción de la varianza total debida (o explicada por) la k-ésima componente principal es

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \quad k = 1,2,\cdots,p \qquad (7)$$

De manera similar, las primeras m componentes principales explican una proporción

$$\frac{\lambda_1 + \cdots + \lambda_m}{\lambda_1 + \cdots + \lambda_p} \quad m \leq p$$

de la variabilidad total.

Resultado 3 Si $\mathbf{Y_1} = \mathbf{e_1'X}$, $\mathbf{Y_2} = \mathbf{e_2'X}, \cdots, \mathbf{Y_p} = \mathbf{e_p'X}$ son las componentes principales que se obtienen de la matriz $\boldsymbol{\Sigma}$, éstas podrán emplearse en análisis estadísticos subsecuentes. Para lo anterior, será necesario calcular las calificaciones de tales componentes para cada individuo en el conjunto de datos. Estas calificaciones proporcionan las ubicaciones de los individuos en un conjunto de datos con respecto a sus ejes componentes principales. Sea $\mathbf{X_r}$ el vector de variables medidas para la r-ésima unidad experimental. Entonces el valor (calificación) de la j-ésima componente principal, para la r-ésima unidad bajo estudio estará dado, en general, por

$$Y_{rj} = \mathbf{e_j'X_r} \quad j = 1,\cdots,p \text{ componentes}$$
$$r = 1,\cdots,n \quad \text{individuos}$$

Adicionalmente,

$$\rho_{Y_i X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i,k = 1,2,\cdots,p \qquad (8)$$

denota el coeficiente de correlación entre la componente Y_i y la variable X_k.

Aunque las correlaciones de las variables con sus componentes principales en ocasiones son auxiliares en su interpretación, éstas miden únicamente la contribución univariada de una variable individual a una componente.

Determinación del Número de Componentes Principales

Uno de los propósitos fundamentales del análisis de componentes principales, consiste en reducir la dimensionalidad del espacio que contiene los datos. Lo anterior es equivalente a determinar el número de componentes principales que explican una proporción considerable de la variabilidad en el conjunto de datos.

De entre los métodos para elegir el número de componentes principales destacan dos de ellos, ambos basados en los eigenvalores de $\boldsymbol{\Sigma}$. Se define d como la dimensionalidad real de los datos.

Método 1

Supongamos que se desea tomar en cuenta el $\gamma 100\,\%$ de la variabilidad total en las variables originales. En este método para estimar d se considera el cociente

$V = \sum_{j=1}^{k} \lambda_j / \sum_{j=1}^{p} \lambda_j$ para valores sucesivos de $k = 1, \ldots, p$. En este caso d se estima por el menor de los valores de k en el que, por primera vez, V excede $\gamma 100\,\%$.

Método 2

Se utiliza una gráfica *scree* de los eigenvalores. Ésta se construye graficando las parejas $(1,\lambda_1),(2,\lambda_2),\cdots,(p,\lambda_p)$. Cuando los puntos de la gráfica tienden a nivelarse, estos eigenvalores suelen estar suficientemente cercanos a cero, por lo que pueden ignorarse. Por lo tanto en este método se supone que la dimensionalidad del espacio de los datos originales es la que corresponde al orden del eigenvector grande más pequeño.

Componentes Principales a partir de Variables Estandarizadas

Las componentes principales también se pueden obtener a partir de las variables estandarizadas

$$\begin{aligned} Z_i &= \frac{\left(X_i - \mu_i\right)}{\sqrt{\sigma ii}} \quad i = 1, \cdots, p \\ \mu_i &= \text{media de } X_i \\ \sigma_{ii} &= \text{varianza de } X_i \end{aligned} \qquad (9)$$

Utilizando notación matricial,

$$\mathbf{Z} = \left(\mathbf{V}^{1/2}\right)^{-1}\left(\mathbf{X} - \boldsymbol{\mu}\right) \qquad (10)$$

En donde la matriz diagonal $V^{1/2}$ se define

$$\mathbf{V}^{1/2} = \begin{pmatrix} \sqrt{\sigma_{11}} & \cdots & 0 \\ & \ddots & \\ 0 & & \sqrt{\sigma_{pp}} \end{pmatrix}$$

E(**Z**) = 0 y

$Cov(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1}\Sigma(\mathbf{V}^{1/2})^{-1} = \boldsymbol{\rho}$

Los componentes principales de **Z** se obtienen a partir de los eigenvectores de la matriz de correlación $\boldsymbol{\rho}$ de ***X***. Todos los resultados previos aplican con algunas simplificaciones, ya que la varianza de cada Z_i es la unidad. Se continuará utilizando la notación Y_i para referirse al i-ésimo componente principal y $\lambda_i, \mathbf{e}_i$ para los eigenvalores y eigenvectores calculados ya sea a partir de $\boldsymbol{\rho}$ *o* $\boldsymbol{\Sigma}$. *Sin embargo, los valores y vectores propios que se obtienen a partir de* $\boldsymbol{\Sigma}$, *en general, no son iguales a los que se*

obtienen a partir de $\boldsymbol{\rho}$ *y no existe una relación simple para pasar de un resultado a otro.*

Resultado 4. La i-ésima componente principal de las variables estandarizadas $\mathbf{Z}' = [Z_1, Z_2, \cdots, Z_p]$ con $Cov(\mathbf{Z}) = \boldsymbol{\rho}$ está dado por

$$Y_i = \mathbf{e}_i' Z = \mathbf{e}_i' (\mathbf{V}^{1/2})(\mathbf{X} - \boldsymbol{\mu}), \quad i = 1, 2, \cdots, p$$

Adicionalmente

$$\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} Var(Z_i) = p \qquad (11)$$

Y la correlación entre la componente principal Y_i y la variable estandarizada z_k está dada por:

$$\rho_{Y_i Z_k} = e_{ik}\sqrt{\lambda_i} \qquad \text{i, } k = 1, 2, \ldots, p$$

En este caso $\lambda_1, \cdots, \lambda_p$ y $\mathbf{e}_1, \cdots \mathbf{e}_p$ son los eigenvalores y eigenvectores de $\boldsymbol{\rho}$, respectivamente con $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

De la ecuación (11), la varianza total de las variables estandarizadas originales es simplemente *p*, es decir el número total de variables. Utilizando esta ecuación con las observaciones estandarizadas **Z**'s en lugar de las variables originales **X's**, encontramos que la proporción de la varianza total explicada por la k-ésima componente principal sobre $\boldsymbol{\rho}$ está dada por:

$$\frac{\lambda_k}{p}, \quad k = 1, 2, \cdots, p \qquad (12)$$

en donde λ_k son los eigenvalores de $\boldsymbol{\rho}$. De manera similar, la varianza explicada por las *m* primeras componentes principales es:

$$\frac{\lambda_1 + \cdots \lambda_m}{p} \quad m \leq p$$

La estandarización de las variables no es inconsecuente y es conveniente realizar dicha estandarización en los siguientes casos:

1) Cuando las variables originales están medidas en unidades diferentes
2) Cuando las varianzas de las variables originales tengan tamaños muy diferentes entre sí.

Por ejemplo, si X_1 representa el ingreso anual con un rango entre \$10,000 y \$350,000 y X_2 es la razón de (ingreso anual)/(total de activos) con rango entre 0.01 y 0.6, entonces la varianza de los datos resultará casi totalmente atribuible al ingreso. En este caso esperaríamos un solo componente principal (importante) con un peso muy relevante en la variable X_1. Alternativamente si ambas variables se estandarizan, sus magnitudes

subsecuentes serán del mismo orden y X_2 o (Z_2) resultará más relevante en la construcción de las componentes principales.

Determinación del Número de Componentes para la Matriz de Correlación ρ

En general los mismos criterios enunciados para el caso de la matriz de covarianza Σ, son aplicables al caso de la matriz de correlación ρ. Sin embargo una regla empírica adicional, comúnmente utilizado en la práctica del análisis de componentes principales, consiste en seleccionar únicamente los eigenvalores mayores que "1" y se estima que la dimensionalidad del espacio muestral es la del número de eigenvalores mayores que 1. La razón para comparar los eigenvalores con 1 es que, cuando se está realizando el análisis sobre datos estandarizados, la varianza de cada variable estandarizada es igual a 1. Se considera que si una componente principal no puede explicar más variación que una variable por sí misma, entonces es probable que no sea importante, por lo que frecuentemente se ignoran componentes cuyos eigenvalores son menores que 1. En todo caso la decisión por lo que toca a cuántas componentes principales se deben considerar es subjetiva[22]

COMPONENTES PRINCIPALES A PARTIR DE DATOS MUESTRALES

Supongamos que los datos $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ representan una muestra aleatoria de n individuos a los cuales se les miden p-variables (es decir una muestra aleatoria de n vectores p dimensionales) con vector de medias μ y matriz de covarianza Σ. Estos datos producen un vector de medias muestrales $\bar{\mathbf{x}}$, la matriz de covarianza muestral **S** y la matriz de correlación muestral **R**.

De nueva cuenta, el objetivo será el de construir combinaciones lineales no correlacionadas de las variables medidas en la muestra y que representen la mayor proporción posible de la variación de datos muestrales provenientes de la población. Las combinaciones lineales no correlacionadas, se denominan *componentes principales muestrales*.

Si $\mathbf{S} = \{s_{ik}\}$ es la matriz de covarianza muestral de dimensión $p \times p$ con eigenvalores $\hat{\lambda}_1, \cdots, \hat{\lambda}_p$ y sus respectivos eigenvectores, $\hat{\mathbf{e}}_1, \cdots \hat{\mathbf{e}}_p$, la i-ésima componente principal muestral está dada por

$$\hat{y}_i = \hat{\mathbf{e}}_i' \mathbf{x} = \hat{e}_{i1} x_1 + \hat{e}_{i2} x_2 + \cdots + \hat{e}_{ip} x_p, \quad i = 1, 2, \cdots, p$$

En donde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \hat{\lambda}_p \geq 0$ y $\mathbf{x}_i$ es el vector observado en las variables $X_1, X_2, \ldots, X_p$.

Los resultados enunciados para el caso de componentes principales poblacionales, serán aplicables al caso de componentes principales muestrales.

Vectores de Carga de Componentes/Correlaciones de Componentes

[22] Chatfield C., Collins A.J. 1980. Introduction to Multivariate Analysis. Chapman & Hall

Como se mencionó, los eigenvectores de $\boldsymbol{\Sigma}$ que se están utilizando para definir las componentes principales se normalizan para tener longitud de 1; esto es $\mathbf{e}_j'\mathbf{e}_j = 1$ para $j = 1, \cdots, p$. Este hecho puede crear confusión en la interpretación de los componentes principales mediante el examen de los elementos de los eigenvectores que definen dichas componentes. Es decir, los elementos dentro de un eigenvector son comparables entre sí; sin embargo los elementos en eigenvectores diferentes, no lo son. Esto es en virtud de que los eigenvectores se normalizan para tener longitud 1, lo cual requiere que la suma de los cuadrados de los elementos en cada vector debe ser igual a 1. Por consiguiente entre más elementos haya en un solo eigenvector que sean en realidad diferentes de cero, mas pequeño debe ser cada elemento.

A fin de hacer posible las comparaciones entre eigenvectores se cambia su escala multiplicando los elementos en cada eigenvector por la raíz cuadrada del eigenvalor correspondiente. Sea $\mathbf{e}_j^* = \lambda_j^{1/2}\mathbf{e}_j$ para $j = 1, \cdots, p$. Estos vectores son llamados vectores de carga de componentes y son también eigenvectores de $\boldsymbol{\Sigma}$, pero tienen longitudes iguales a λ_j en lugar de tener longitud 1.

Defínase $\mathbf{C} = \left[\mathbf{e}_1^*, \mathbf{e}_2^*, \cdots, \mathbf{e}_p^*\right]$, es decir $\mathbf{C} = \mathbf{E}\boldsymbol{\Lambda}^{1/2}$, en donde $\mathbf{E}$ es la matriz formada por los vectores columna $\mathbf{e}_i$ correspondientes a los eigenvectores λ_i. Los elementos de $\mathbf{C}$ son tales que los coeficientes de los componentes más importantes están escalados para, en general, ser más grandes que aquellos correspondientes a los componentes menos importantes.

Todos los elementos en todas las $\mathbf{e}_j^*$ son comparables entre sí. El i-ésimo elemento de $\mathbf{e}_j^*$ permite calcular la covarianza entre la i-ésima variable original y la j-ésima componente principal, según se indica en (8).

Los vectores de cargas de las componentes principales son particularmente útiles en el caso del análisis multivariado, cuando éstas representan una proporción elevada de la varianza total de las p variables originales, ya que se puede graficar la posición de los individuos respecto a las dos primeras componentes. La idea detrás de esta construcción gráfica consiste en agregar información acerca de las variables a las componentes, determinando su contribución relativa y al mismo tiempo ubicar la posición de los individuos respecto a las primeras dos componentes. Este tipo de representación gráfica, se denomina *bi-plot*.

Posibles Interpretaciones de las Componentes Principales

Una situación común que puede presentarse al realizar un análisis de componentes principales surge cuando todas las variables están correlacionadas de manera positiva. La primera componente principal es entonces una especie de promedio ponderado de las variables y puede considerarse como una medida de tamaño. En particular si se analiza la matriz de correlación, las variables (estandarizadas) tendrán pesos casi iguales.

Para matrices de correlación que contengan tanto elementos positivos como negativos, la posición es menos clara. Al tratar de dar alguna interpretación significativa a un componente en particular, el procedimiento usual parece consistir en examinar el eigenvector correspondiente y tomar las variables para las cuales los coeficientes son

relativamente grandes, ya sean positivos o negativos. Una vez que se ha establecido el subconjunto de variables que son importantes para un componente en particular, el investigador tratará de determinar qué tienen en común dichas variables. Sin embargo, si las componentes son utilizadas para agrupar las variables, con frecuencia es también el caso que estos grupos puedan encontrarse mediante inspección visual directa de la matriz de correlaciones. De cualquier manera, se deberá actuar con cautela al asignar significado a las componentes principales.

3.2 Análisis de Factores

INTRODUCCIÓN

El análisis de factores tiene su origen a principios del siglo XX en los estudios de Karl Pearson, Charles Spearman y otros autores para definir y medir la inteligencia. Debido a esta asociación inicial con conceptos como la inteligencia, el análisis de factores se nutrió y desarrolló originalmente por científicos interesados en la psicometría. Las controversias respecto a la interpretación psicológica de los primeros estudios y la falta de recursos de cómputo con la potencia necesaria, impidieron su desarrollo como método estadístico. El advenimiento de las computadoras de alta velocidad ha generado un renovado interés en los aspectos teóricos y computacionales del análisis de factores. En este renovado ímpetu, muchas de las técnicas y controversias originales han sido abandonadas y resueltas. No obstante, cada aplicación de la técnica deberá ser examinada en base a sus propios méritos, a fin de determinar sus posibilidades de éxito.

El propósito esencial del análisis de factores consiste en describir, en lo posible las relaciones de covarianza entre muchas variables correlacionadas, mediante un número reducido y no observable de cantidades aleatorias, denominadas *factores*. El *modelo de factores* está motivado básicamente por el siguiente argumento: Supongamos que ciertas variables pueden agruparse de acuerdo a sus correlaciones. Es decir, supongamos que todas las variables dentro de un grupo particular están correlacionadas entre sí y tienen una muy baja correlación con variables de grupos distintos. En este caso es concebible que cada grupo de variables represente una construcción subyacente única o *factor*, el cual es responsable de las correlaciones observadas. Por ejemplo, las correlaciones entre un conjunto de calificaciones recolectadas por Spearman en textos clásicos, francés, inglés, matemáticas y música sugerían un *factor* de "inteligencia" subyacente. Un segundo grupo de variables, representando calificaciones sobre condicionamiento físico; por ejemplo resistencia, velocidad, fuerza física y elasticidad, en caso de estar disponible, a través de resultados en pruebas de pista y campo, correspondería a un nuevo *factor*; que en este caso pudiera denominarse "aptitud física".

La estructura anterior, es precisamente la que el análisis de factores pretende confirmar.

El análisis de factores se puede considerar como una extensión del análisis de componentes principales. Ambos representan intentos para lograr una descripción de la estructura de la matriz de covarianza Σ. Sin embargo, la diferencia entre ambos métodos radica en lo siguiente:

El análisis de componentes principales produce una transformación ortogonal de las variables y no depende de un modelo estadístico subyacente. Su interés se centra en la determinación de la verdadera dimensionalidad del conjunto de variables originales

cuando éstas se encuentran correlacionadas entre sí, mediante un nuevo conjunto de variables no correlacionadas. Por su parte el análisis de factores, sí depende de un modelo estadístico subyacente y su interés radica en la explicación de la estructura de covarianza y/o correlación entre las variables bajo estudio. La cuestión primordial en el análisis de factores, es si los datos son consistentes con la estructura teórica prescrita por el conocimiento en la materia bajo estudio.

MODELO DE ANÁLISIS DE FACTORES

El vector aleatorio observado **X**, con p variables tiene media $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$. El modelo de análisis de factores postula que **X** es linealmente dependiente de un nuevo conjunto de variables aleatorias no observables F_1, F_2,..., F_m, denominados *factores* comunes, además de p fuentes adicionales de variación $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p$, llamadas *errores* o *factores específicos*. El modelo de análisis de factores se define de la siguiente manera:

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \varepsilon_p \end{aligned} \qquad (1)$$

o en notación matricial,

$$\underset{(p\,x\,1)}{\mathbf{X} - \boldsymbol{\mu}} = \underset{(p\,x\,m)}{\mathbf{L}}\ \underset{(m\,x\,1)}{\mathbf{F}} + \underset{(p\,x\,1)}{\boldsymbol{\varepsilon}} \qquad (2)$$

El coeficiente l_{ij} es conocido como *carga* de la i-ésima variable en el j-ésimo factor, de tal manera que la matriz **L** es la matriz de cargas de los factores. Hay que notar que el i-*ésimo* factor específico ε_i se encuentra asociado únicamente con la i-*ésima* variable respuesta X_i. Las p desviaciones $X_1 - \mu_1, X_2 - \mu_2, \cdots, X_p - \mu_p$, se expresan en términos de $p + m$ variables aleatorias $F_1, F_2, \cdots F_m, \varepsilon_1, \varepsilon_2, \cdots \varepsilon_p$, las cuales no son observables. Esta característica distingue el modelo de factores (2) del modelo de regresión lineal multivariada en el cual las variables independientes [cuya posición ocupan las **F** en el modelo (2)], pueden ser observadas.

Ya que se tiene un gran número de cantidades no observables, la verificación directa del modelo de factores mediante la observación de $X_1, X_2, \cdots X_p$ resulta imposible. Sin embargo, mediante un conjunto de supuestos adicionales en cuanto a los vectores aleatorios **F** y $\boldsymbol{\varepsilon}$, el modelo en (2) trae consigo un conjunto de relaciones de covarianza que pueden ser verificados Johnson (2002) y cuyos principales resultados se describen en las siguientes dos secciones.

Suponemos que

$$\underset{(m\,x\,1)}{E(\mathbf{F}) = \mathbf{0}}, \quad Cov(\mathbf{F}) = E[\mathbf{F}\mathbf{F}'] = \underset{(m\,x\,m)}{\mathbf{I}}$$

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad Cov(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Psi}_{p\times p} = \begin{pmatrix} \Psi_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Psi_p \end{pmatrix} \tag{3}$$

y que $\boldsymbol{F}$ y $\boldsymbol{\varepsilon}$ son independientes, de tal manera que

$$Cov(\boldsymbol{\varepsilon}, \mathbf{F}) = E(\boldsymbol{\varepsilon}\mathbf{F}') = \underset{(p\,x\,m)}{\mathbf{0}}$$

Estos supuestos y la relación en (2), constituyen el modelo ortogonal de factores.

MODELO ORTOGONAL DE FACTORES CON *m* FACTORES COMUNES

$$\underset{(p\,x\,1)}{\mathbf{X}} = \underset{(p\,x\,1)}{\boldsymbol{\mu}} + \underset{(p\,x\,m)(m\,x\,1)}{\mathbf{L}\ \mathbf{F}} + \underset{(p\,x\,1)}{\boldsymbol{\varepsilon}}$$

μ_i = *media* de la variable *i*
ε_i = i-ésimo *factor específico*
F_j = j-ésimo *factor común* (4)
l_{ij} = carga de la i-ésima variable en el j-ésimo factor

Esta última mide la contribución del q-ésimo factor común a la j-ésima variable respuesta.

Los vectores aleatorios no observables **F** y $\boldsymbol{\varepsilon}$ satisfacen las condiciones indicadas en (3).

El modelo de factores ortogonal implica una estructura de covarianza para **X**. A partir del modelo en (4),

$$(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})' = (\mathbf{LF}+\boldsymbol{\varepsilon})(\mathbf{LF}+\boldsymbol{\varepsilon})'$$
$$= (\mathbf{LF}+\boldsymbol{\varepsilon})\left((\mathbf{LF})' + \boldsymbol{\varepsilon}'\right)$$
$$= \mathbf{LF}(\mathbf{LF})' + \boldsymbol{\varepsilon}(\mathbf{LF})' + \mathbf{LF}\boldsymbol{\varepsilon}' + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$$

de tal manera que

$$\boldsymbol{\Sigma} = Cov(\mathbf{X}) = E(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'$$
$$= \mathbf{L}E(\mathbf{F}\mathbf{F}')\mathbf{L}' + E(\boldsymbol{\varepsilon}\mathbf{F}')\mathbf{L}' + \mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}') + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$$
$$= \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$$

de acuerdo a (3).

Asimismo, por independencia entre $\boldsymbol{\varepsilon}$ y $\mathbf{F}$, $Cov(\boldsymbol{\varepsilon},\mathbf{F}) = E(\boldsymbol{\varepsilon},\mathbf{F}') = \mathbf{0}$. Y por el modelo en (4), $(\mathbf{X}-\boldsymbol{\mu})\mathbf{F}' = (\mathbf{LF}+\boldsymbol{\varepsilon})\mathbf{F}' = \mathbf{LFF}' + \boldsymbol{\varepsilon}\mathbf{F}'$, de tal manera que $Cov(\mathbf{X},\mathbf{F}) = E(\mathbf{X}-\boldsymbol{\mu})\mathbf{F}' = \mathbf{L}E(\mathbf{FF}') + E(\boldsymbol{\varepsilon}\mathbf{F}') = \mathbf{L}$.

El resultado anterior nos indica que la covarianza entre el vector observable $\mathbf{X}$ con p componentes y las $F_1, F_2, \ldots, F_m$ variables no observables, está dada por la matriz de cargas $\boldsymbol{L}$.

ESTRUCTURA DE COVARIANZA DEL MODELO ORTOGONAL DE FACTOR

1. $Cov(\mathbf{X}) = \mathbf{LL}' + \boldsymbol{\Psi}$

   o

   $Var(X_i) = l_{i1}^2 + \cdots + l_{im}^2 + \Psi_i$

   $Cov(X_i, X_k) = l_{i1}l_{k1} + \cdots + l_{im}l_{km}$ (5)

2. $Cov(\mathbf{X},\mathbf{F}) = \mathbf{L}$

   o

   $Cov(X,F) = l_{ij}$

La covarianza del vector $\mathbf{X}$ puede descomponerse en dos componentes; el primero, atribuible a la varianza común a todos los factores y el segundo a la variabilidad específica X_i.

El modelo $\mathbf{X}-\boldsymbol{\mu} = \mathbf{LF}$ es lineal en los factores comunes. Si las p variables respuesta X se encuentran realmente relacionadas con un conjunto de factores subyacentes, pero la relación es no lineal como en $X_1 - \mu_1 = l_{11}F_1F_3 + \varepsilon_1$, $X_2 - \mu_2 = l_{21}F_2F_3 + \varepsilon_2$, etc., entonces la estructura de covarianza $\mathbf{LL}' + \boldsymbol{\Psi}$ dada en (5), es inadecuada. El importante supuesto de linealidad es inherente al planteamiento tradicional del modelo de factor.

La porción de la varianza de la i-ésima variable compartida con las demás variables a través de los m factores comunes, se denomina la i-ésima *comunalidad*. La porción de $Var(X_i) = \sigma_{ii}$ debida específicamente a la variable por sí misma, se llama la *varianza específica*. Denotando la i-ésima comunalidad por h_i^2, de la ecuación (5) se tiene que

$$\begin{aligned} \sigma_{ii} &= l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \Psi_i \\ Var(X_i) &= \text{comunalidad} + \text{varianza especìfica} \end{aligned}$$

o bien

$$\begin{aligned} & h_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 \\ & \text{y} \\ & \sigma_{ii} = h_i^2 + \psi_i, \qquad i = 1,2,\cdots,p \end{aligned} \tag{6}$$

La i-ésima comunalidad es la suma de cuadrados de las cargas de la i-ésima variable sobre los *m* factores comunes.

El modelo de factor supone que las $p + p(p-1)/2 = p(p+1)/2$ varianzas y covarianzas de **X** se pueden reproducir a partir de *pm* cargas de factores l_{ij} y de *p* varianzas específicas Ψ_i. Cuando *m* = *p*, cualquier matriz de covarianza se reproduce exactamente como $\mathbf{LL}'$, de tal manera que $\boldsymbol{\Psi}$ es la matriz **0**. Sin embargo, es precisamente cuando *m* < *p* (es decir, que el número de factores es menor al número de variables), que el análisis de factores resulta de mayor utilidad. En este caso, el modelo de factores proporciona una explicación más "simple" de la covarianza en **X** con un número menor de parámetros que los $p(p+1)/2$ parámetros en $\boldsymbol{\Sigma}$. La solución del modelo de factor, para ser consistente desde el punto de vista de su interpretación estadística, además de representar los datos de manera sucinta (*m* < *p*), deberá generar valores tales que $Var(\varepsilon_i) \geq 0$ y $-1 \leq l_{ij} \leq 1$.

Puesto que las cargas de los factores l_{ij} son las correlaciones entre las variables X_i y los factores F_j, éstos deben tomar valores entre -1 y 1.

No Unicidad de los Factores

Cuando m > 1, existirá ambigüedad inherente asociada al modelo de factor. Supongamos que **T** es cualquier matriz ortogonal de *m x m*, de tal manera que $\mathbf{TT}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$. La expresión (2) se puede escribir

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{LF} + \boldsymbol{\varepsilon} = \mathbf{LTT}'\mathbf{F} + \boldsymbol{\varepsilon} = \mathbf{L}^*\mathbf{F}^* + \boldsymbol{\varepsilon} \qquad (7)$$

En donde

$$\mathbf{L}^* = \mathbf{LT} \quad y \quad \mathbf{F}^* = \mathbf{T}'\mathbf{F}$$

ya que

$$E(\mathbf{F}^*) = \mathbf{T}'E(\mathbf{F}) = \mathbf{0}$$

y

$$Cov(\mathbf{F}^*) = \mathbf{T}'Cov(\mathbf{F})\mathbf{T} = \mathbf{T}'\mathbf{T} = \mathbf{I}_{(m\,x\,n)}$$

Resulta imposible con base en las observaciones en **X**, distinguir entre las cargas en **L** y las cargas en $\mathbf{L}^*$. Es decir los factores **F** y $\mathbf{F}^* = \mathbf{T}'\mathbf{F}$ tienen las mismas propiedades estadísticas y aunque las cargas en $\mathbf{L}^*$ son, en general, diferentes de las cargas en **L**, ambas generan la matriz de covarianza $\boldsymbol{\Sigma}$. Es decir,

$$\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi} = \mathbf{LTT}' + \boldsymbol{\Psi} = (\mathbf{L}^*)(\mathbf{L}^*)' + \boldsymbol{\Psi} \qquad (8)$$

Esta ambigüedad proporciona el razonamiento de la "rotación de factores", ya que las matrices ortogonales corresponden a rotaciones (y reflexiones) del sistema de coordenadas de **X**.

El análisis del modelo de factores procede imponiendo condiciones que permiten estimaciones únicas de $\mathbf{L}$ y $\mathbf{\Psi}$, las cuales se explicarán más adelante. La matriz de cargas se rota subsecuentemente (multiplicándola por una matriz ortogonal), en donde la rotación se determina mediante algún criterio que "facilitará" la interpretación. Una vez que se han obtenido las cargas y varianzas específicas, se identifican los factores y los valores estimados para los factores (denominados calificaciones de los factores).

Dadas las observaciones $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ en p variables correlacionadas, el análisis de factores trata de evaluar si las interrelaciones entre dicho conjunto de variables se explican en términos de un número pequeño de variables subyacentes no observadas o *factores*. Adicionalmente, se desea interpretar estas nuevas variables. En esencia, el problema de construcción del modelo estadístico, consiste en verificar la estructura de covarianza en (5).

La matriz de covarianza muestral $\mathbf{S}$ es un estimador de la covarianza poblacional $\mathbf{\Sigma}$, desconocida. Si los elementos fuera de la diagonal principal de $\mathbf{S}$ son pequeños o si los de la matriz de correlación muestral $\mathbf{R}$ son esencialmente cero, las variables observadas no se encuentran asociadas y el análisis de factores no tiene sentido. En estas circunstancias, los factores *específicos* juegan un papel dominante. Sin embargo, el objetivo del análisis de factores es el de encontrar un número reducido de factores comunes.

Si $\mathbf{\Sigma}$ muestra desviaciones importantes fuera de la diagonal principal, entonces se podrá considerar el modelo de factor como alternativa de análisis de los datos y el problema inicial en este caso consistirá en estimar las cargas de los factores l_{ij} y las varianzas específicas en $\mathbf{\Psi}$. Se consideran a continuación dos de los métodos más empleados en la estimación de los parámetros; el de *componentes principales* (y su método relacionado de *factores principales*) y el método de *máxima verosimilitud*. La solución derivada de ambos métodos puede ser rotada subsecuentemente, con el propósito de simplificar la interpretación de los factores, según se describió anteriormente.

Posibles Problemas Numéricos del Modelo de Factores[23]

Para determinar si existe un conjunto de m factores subyacentes, se determina si existen $\mathbf{L}$ y $\mathbf{\Psi}$ tales que la matriz de correlación $(\mathbf{P})$

$$\mathbf{P} = \mathbf{LL}' + \mathbf{\Psi}$$

El número de cantidades desconocidas en $\mathbf{L}$ y $\mathbf{\Psi}$ es $pm + p = p(m+1)$. El número de cantidades desconocidas en $\mathbf{P}$ es $p(p+1)/2$ (debido a que $\mathbf{P}$ es simétrica). Por lo tanto las ecuaciones del modelo de análisis de factores $\mathbf{\Sigma} = \mathbf{LL}' + \mathbf{\Psi}$, dan como resultado un sistema de $p(p+1)/2$ ecuaciones en $p(m+1)$ incógnitas. Se tienen tres posibles escenarios:

1. Si $p(m+1) > p(p+1)/2$, o de manera equivalente, si $m > p(p+1)/2$, entonces existen más incógnitas que ecuaciones y no existe una solución única

[23] Johnson, Dallas. 2000. *Métodos Multivariados Aplicados al Análisis de Datos*. Thomson Editores

al sistema. En este caso, el análisis de factores aporta una solución más complicada que $\boldsymbol{\Sigma}$ o $\mathbf{P}$.

2. Si $m=(p-1)/2$, entonces el modelo contiene tantos parámetros como elementos de $\mathbf{P}$ y no hay simplificación de las relaciones entre las variables observadas. Puede encontrarse una relación única, pero no necesariamente una con todas las varianzas específicas mayores que cero.
3. Si $m<(p-1)/2$, ésta es la única situación de verdadero interés. En este caso hay menos parámetros en el modelo de factores que elementos de $\mathbf{P}$. Por consiguiente, el modelo puede proporcionar una explicación más simple de las relaciones entre las variables observadas que aquella proporcionada por los elementos de $\mathbf{P}$.

Los dos métodos que se van a describir requieren estimadores preeliminares de las varianzas específicas (o de manera equivalente, de las comunalidades).

Dos estimadores de las comunalidades son:

(1) R^2. Es el cuadrado del coeficiente de correlación múltiple de la i-ésima variable con todas las otras variables (es decir, el porcentaje de la variabilidad en la i-ésima variable que se explica por las otras variables).

(2) El mayor de los valores absolutos de los coeficientes de correlación entre la i-ésima variable y una de las variables restantes.
Cada uno de estos estimadores dará valores mayores de comunalidades cuando x_i esté altamente correlacionado con las otras variables.

Método de Factores Principales

El algoritmo general del método de solución de análisis de factores a través de factores principales, se describe como sigue.

En primer lugar deben estimarse las comunalidades o, equivalentemente, las varianzas específicas. Este proceso es esencialmente equivalente a llevar a cabo un PCA de la matriz de correlaciones (covarianzas) reducida **S*** (**P***), obtenida reemplazando los elementos observados en la diagonal de **S** (**P**) por las comunalidades estimadas. Las primeras k componentes principales se utilizan para obtener los estimadores de las cargas de factores. A partir de (12) se obtienen estimaciones de las varianzas específicas y se considera que el método es adecuado si todas estas estimaciones son no negativas.

La descomposición espectral[24] de $\boldsymbol{\Sigma}$, nos permite su expresión mediante la siguiente ecuación. Sean λ_i los eigenvalores y $\mathbf{e}_i$ los correspondientes eigenvectores de $\boldsymbol{\Sigma}$, con $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_p \geq 0$. Entonces

[24] La descomposición espectral (de A) corresponde a una representación $\mathbf{A} = \lambda_1 \mathbf{u}_1 \mathbf{u}'_1 + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}'_n$ en donde $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ es una base ortonormal de eigenvectores de $\mathbf{A}$ y $\lambda_1, \cdots, \lambda_n$ son los eigenvalores de $\mathbf{A}$

$$\Sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p'$$

$$= \left[\sqrt{\lambda_1}\mathbf{e}_1, \sqrt{\lambda_2}\mathbf{e}_2, \cdots, \sqrt{\lambda_p}\mathbf{e}_p \right] \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1' \\ \sqrt{\lambda_2}\mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_p}\mathbf{e}_p' \end{bmatrix} \qquad (9)$$

Este modelo se ajusta a la estructura de covarianza prescrita para el modelo de análisis de factores, conteniendo tantos factores como variables ($m = p$) y con varianza específica $\Psi = 0$ para toda i. La j-ésima columna de la matriz de cargas está dada por $\sqrt{\lambda_j}\mathbf{e}_j$. Es decir, la matriz de covarianza $\mathbf{\Sigma}$, se puede escribir como sigue:

$$\underset{(pxp)}{\mathbf{\Sigma}} = \underset{(pxp)}{\mathbf{L}} \underset{(pxp)}{\mathbf{L}'} + \underset{(pxp)}{\mathbf{0}} = \mathbf{LL}' \qquad (10)$$

Aunque la representación de $\mathbf{\Sigma}$ mediante el análisis de factores en (10) es exacta, no es de utilidad, ya que emplea tantos factores comunes como variables y no permite ninguna variación en los factores específicos $\boldsymbol{\varepsilon}$ definidos en (4). Se prefieren los modelos que explican la estructura de covarianza en términos de solamente un número reducido de factores. Una alternativa consiste en que cuando los últimos $p - m$ eigenvalores son pequeños[25], ignorar la contribución de $\lambda_{m+1}\mathbf{e}_{m+1}\mathbf{e}_{m+1}' + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p'$ para $\mathbf{\Sigma}$ en (9). Ignorando esta contribución se obtiene la aproximación

$$\mathbf{\Sigma} \doteq \left[\sqrt{\lambda_1}\mathbf{e}_1, \sqrt{\lambda_2}\mathbf{e}_2, \cdots, \sqrt{\lambda_m}\mathbf{e}_m \right] \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1' \\ \sqrt{\lambda_2}\mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_p}\mathbf{e}_m' \end{bmatrix} = \mathbf{LL}' \qquad (11)$$

$(p \times m)(m \times p)$

La aproximación representada en (11) supone que los factores específicos $\boldsymbol{\varepsilon}$ definidos en (4) son de menor importancia respecto a los factores comunes y también pueden ser ignorados en la factorización de $\mathbf{\Sigma}$. Si los factores específicos se incluyen en el modelo, sus varianzas deberán tomarse de los elementos de la diagonal de $\mathbf{\Sigma} - \mathbf{LL}'$ en donde $\mathbf{LL}'$ se define en (11).

Al considerar los factores específicos, la aproximación se convierte en:

[25] En la práctica se consideran pequeños a aquellos eigenvalores inferiores a 1.

$$\Sigma \doteq \mathbf{L}\mathbf{L}' + \Psi$$

$$= \left[\sqrt{\lambda_1}\mathbf{e}_1, \sqrt{\lambda_2}\mathbf{e}_2, \cdots, \sqrt{\lambda_m}\mathbf{e}_m\right] \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1' \\ \sqrt{\lambda_2}\mathbf{e}_2' \\ \vdots \\ \sqrt{\lambda_m}\mathbf{e}_m' \end{bmatrix} + \begin{pmatrix} \Psi_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Psi_p \end{pmatrix} \qquad (12)$$

en donde $\Psi_i = \sigma_{ii} - \sum_{j=1}^{m} l_{ij}^2 \quad i = 1, 2, \cdots, p$.

Para aplicar este enfoque al conjunto de datos $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$, se acostumbra primeramente centrar las observaciones restando la media muestral $\bar{\mathbf{x}}$. Las observaciones centradas

$$\mathbf{x}_j - \bar{\mathbf{x}} = \begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ x_{jp} \end{bmatrix} - \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} = \begin{bmatrix} x_{j1} - \bar{x}_1 \\ x_{j2} - \bar{x}_2 \\ \vdots \\ x_{jp} - \bar{x}_p \end{bmatrix} \quad j = 1, 2, \cdots, n \qquad (13)$$

tienen matriz de covarianza muestral **S**, al igual que las observaciones originales.

En casos en que las unidades de las variables no sean conmensurables, se utilizan las variables estandarizadas

$$\mathbf{z}_j = \begin{bmatrix} \dfrac{(x_{j1} - \bar{x}_1)}{\sqrt{s_{11}}} \\ \dfrac{(x_{j2} - \bar{x}_2)}{\sqrt{s_{22}}} \\ \\ \dfrac{(x_{jp} - \bar{x}_p)}{\sqrt{s_{pp}}} \end{bmatrix} \quad j = 1, 2, \cdots, n$$

cuya matriz de covarianza muestral es la matriz de correlación muestral **R** de las observaciones $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$. La estandarización elimina el problema de que una variable con varianza excesivamente grande tenga una influencia excesiva en la determinación de las cargas de los factores.

La representación en (12) cuando se aplica a la matriz de covarianza **S** o a la matriz de correlación **R**, se conoce como la solución de las *componentes principales*. El nombre proviene del hecho de que las cargas de los factores son coeficientes escalados de las primeras componentes principales muestrales.

SOLUCIÓN DE FACTORES PRINCIPALES DEL MODELO DE FACTORES

El análisis de factores mediante componentes principales de la matriz de covarianza muestral **S** se especifica en términos de los eigenvalores $\hat{\lambda}_1, \hat{\lambda}_2, \cdots \hat{\lambda}_p$ en donde $\hat{\lambda}_1 \geq \hat{\lambda}_2 \cdots \geq \hat{\lambda}_p$ y sus correspondientes eigenvectores $\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \cdots, \hat{\mathbf{e}}_p$. Sea $m < p$ el número de factores comunes. Entonces la matriz de las cargas de los factores estimados $\{\tilde{l}_{ij}\}$ estará dada por

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1, \sqrt{\hat{\lambda}_2}\hat{\mathbf{e}}_2 \cdots, \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_m \right] \qquad (14)$$

Las varianzas específicas estimadas se obtienen a partir de los elementos en la diagonal de la matriz $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$, de tal manera que

$$\tilde{\mathbf{\Psi}} = \begin{pmatrix} \tilde{\Psi}_1 & & 0 \\ & \ddots & \\ 0 & & \tilde{\Psi}_p \end{pmatrix} con\ \Psi_i = s_{ii} - \sum_{j=1}^{m} \tilde{l}_{ij}^2 \qquad (15)$$

Las comunalidades se estiman mediante

$$\tilde{h}_i^2 = \tilde{l}_{i1}^2 + \tilde{l}_{i2}^2 + \cdots + \tilde{l}_{im}^2 \qquad (16)$$

El análisis de factores mediante componentes principales de la matriz de correlación muestral se obtiene a partir de **R**, en lugar de **S**.

En la solución por componentes principales, las cargas estimadas para un factor dado no cambian conforme aumenta el número de factores. Por ejemplo si $m = 1$, $\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1 \right]$ y si $m = 2$, $\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\mathbf{e}_1, \sqrt{\hat{\lambda}_2}\mathbf{e}_2 \right]$ en donde $\tilde{\lambda}_1$ y $\tilde{\lambda}_2$ son los primeros dos eigenvalores más grandes de **S** (o de **R**), y $\hat{\mathbf{e}}_1$ y $\hat{\mathbf{e}}_1$ sus correspondientes eigenvectores.

Por definición de $\tilde{\Psi}_i$, los elementos de la diagonal de **S** son iguales a los elementos de la diagonal de $\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}}$. Sin embargo, los elementos fuera de la diagonal de **S** no se reproducen por $\tilde{\mathbf{L}}\tilde{\mathbf{L}}' + \tilde{\mathbf{\Psi}}$; es decir no son iguales. La cuestión por lo tanto es cómo seleccionar el número de factores *m,* a fin de minimizar los valores fuera de la diagonal en la matriz residual, la cual se definirá más adelante.

Si el número de factores comunes no está determinado por consideraciones a priori, tales como la teoría referente al fenómeno bajo estudio o antecedentes de otros investigadores, la selección de *m* se basa en los eigenvalores estimados de la misma manera que en el caso del análisis de componentes principales. El procedimiento de selección está dado como sigue:

Considerar la matriz residual

$$\mathbf{S}-\left(\tilde{\mathbf{L}}\tilde{\mathbf{L}}'+\tilde{\mathbf{\Psi}}\right) \qquad (17)$$

que resulta por la aproximación de **S** por la solución de componentes principales. Los elementos de la diagonal son cero y si el resto de los elementos es también pequeño, podemos considerar de manera subjetiva que el modelo de m factores es apropiado. Analíticamente se tiene que:

$$\text{Suma de las entradas de}\left(\mathbf{S}-\left(\tilde{\mathbf{L}}\tilde{\mathbf{L}}'+\tilde{\mathbf{\Psi}}\right)\right)\text{elevadas al cuadrado} \le \hat{\lambda}_{m+1}^2+\cdots+\hat{\lambda}_p^2 \qquad (18)$$

Consecuentemente un valor pequeño para la suma de cuadrados de los eigenvalores despreciados implica un valor pequeño de la suma de los errores al cuadrado en la aproximación.

Idealmente, la contribución de los primeros factores a las varianzas muestrales de las variables deberá ser grande. La contribución a la varianza muestral s_{ii} del primer factor común es $\tilde{l}_{i1}^2$. La contribución a la varianza muestral total es $s_{11}+s_{22}+\cdots+s_{pp}=tr(\mathbf{S})$, del primer factor común es

$$\tilde{l}_{11}^2+\tilde{l}_{21}^2+\cdots+\tilde{l}_{p1}^2=\left(\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right)'\left(\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1\right)=\hat{\lambda}_1$$

ya que el eigenvector $\hat{\mathbf{e}}_1$ tiene longitud unitaria. En general,

$$\begin{array}{lll} \text{(Proporción de la} & & \\ \text{Varianza muestral} & = \dfrac{\hat{\lambda}_j}{s_{11}+s_{22}+\cdots+s_{pp}} & \text{para análisis de factores sobre } \mathbf{S} \\ \text{total debida al j-ésimo} & & \\ \text{factor)} & = \dfrac{\hat{\lambda}_j}{p} & \text{para análisis de factores sobre } \mathbf{R} \end{array} \qquad (19)$$

El criterio (19) frecuentemente se utiliza como elemento heurístico para determinar el número apropiado de factores comunes. El número de factores comunes retenidos en el modelo se incrementa hasta que la "proporción adecuada" del total de la varianza muestral haya sido explicada.

Otra convención frecuentemente encontrada en los paquetes de computadora es fijar m igual al número de eigenvalores de **R** mayores que uno si se utiliza la matriz de correlación muestral o igual a un número positivo de eigenvalores de **S** si se emplea la matriz de covarianza muestral. Estos resultados no deben de aplicarse de manera indiscriminada. Por ejemplo, $m = p$ si se sigue la regla para **S**, ya que todos los eigenvalores se espera sean positivos en el caso de muestras grandes. El mejor enfoque consiste en retener un número reducido de factores, suponiendo que éstos proporcionen una interpretación satisfactoria de los datos y permitan un ajuste adecuado a **S** o **R**.

Un Enfoque Modificado – la Solución por Factores Principales

En ocasiones se utiliza una modificación al enfoque de componentes principales. Se describe el razonamiento en términos del análisis de factores de **R**, aunque el procedimiento es también aplicable al caso de **S**. Si en el modelo de factores $\mathbf{P}=\mathbf{L}\mathbf{L}'+\mathbf{\Psi}$ los parámetros se encuentra debidamente especificados, m factores

comunes deberán tomar en cuenta tanto los elementos fuera de la diagonal de $\mathbf{P}$, como las porciones de comunalidad incluidas en los elementos diagonales

$$\rho_{ii} = 1 = h_i^2 + \Psi_i$$

Si la contribución de los factores específicos Ψ_i se elimina de la diagonal o equivalentemente, el 1 se reemplaza por h_i^2, la matriz resultante es $\mathbf{P} - \mathbf{\Psi} = \mathbf{L}\mathbf{L}'$.

Supongamos ahora que se dispone de los estimados iniciales Ψ_i^* de las varianzas específicas. Entonces reemplazando el i-ésimo elemento de la diagonal de $\mathbf{R}$ por $h_i^{*2} = 1 - \Psi_i^*$, obtenemos una matriz muestral de correlación reducida $\mathbf{R}_r$

$$\mathbf{R}_r = \begin{bmatrix} h_1^{*2} & r_{12} & \cdots & r_{1p} \\ r_{12} & h_2^{*2} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & h_p^{*2} \end{bmatrix}$$

Ahora, aparte de la variación muestral, todos los elementos de la matriz de correlación muestral $\mathbf{R}$, deberán estar representados por los m factores comunes. En particular, $\mathbf{R}_r$ se descompone como

$$\mathbf{R}_r \doteq \mathbf{L}_r^* \mathbf{L}_r^{*'} \qquad (20)$$

en donde $\mathbf{L}_r^* = \{l_{ij}^*\}$ son las cargas estimadas.

El método de los factores principales emplea los estimados

$$\begin{aligned} L_r^* &= \left[\sqrt{\hat{\lambda}_1^*}\hat{\mathbf{e}}_1^*, \sqrt{\hat{\lambda}_2^*}\hat{\mathbf{e}}_2^*, \cdots, \sqrt{\hat{\lambda}_m^*}\hat{\mathbf{e}}_m^* \right] \\ \Psi_1^* &= 1 - \sum_{j=1}^{m} l_{ij}^{*2} \end{aligned} \qquad (21)$$

en donde $\hat{\lambda}_i^* \quad i = 1, 2, \cdots, m$ son los m eigenvalores más grandes calculados de la matriz $\mathbf{R}_r$ y $\hat{\mathbf{e}}_i^*$ los correspondientes eigenvectores. A la vez, las comunalidades se estimarían mediante

$$\tilde{h}_i^{*2} = \sum_{j=1}^{m} l_{ij}^{*2} \qquad (22)$$

La solución de factores principales se obtiene iterativamente, empleando las comunalidades estimadas en (22) como los estimados iniciales para la siguiente iteración.

En el proceso de solución por factores principales, la consideración de los eigenvalores estimados $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \cdots, \hat{\lambda}_p^*$, resulta auxiliar en el número de factores comunes a retener, en base a la proporción de la varianza explicada.

Una complicación adicional es que ahora algunos de los eigenvalores podrán ser negativos, debido al uso de los estimados iniciales de comunalidad. Idealmente se deberá considerar el número de factores comunes iguales al rango de la matriz poblacional reducida. Desafortunadamente, este rango no siempre se puede determinar

correctamente de $\mathbf{R_r}$, por lo que se requieren consideraciones adicionales, basadas en el criterio del investigador.

Aunque existen muchas opciones para los estimados iniciales de las varianzas específicas, la elección mas común, cuando se trabaja con la matriz de correlación es $\Psi_i^* = 1/r^{ii}$, en donde r^{ii} es el i-ésimo elemento en la diagonal de $\mathbf{R}^{-1}$. Los estimados iniciales de la comunalidad resultan

$$h_i^{*2} = 1 - \Psi_i^* = 1 - \frac{1}{r^{ii}} \qquad (23)$$

Que es igual al cuadrado del coeficiente de correlación múltiple entre Xi y las otras $p - 1$ variables. La relación con el coeficiente de correlación múltiple, significa que h_i^{*2} se puede calcular cuando $\mathbf{R}$ no sea de rango completo. Para factorizar $\mathbf{S}$, las estimaciones iniciales de la varianza utilizan s^{ii}, es decir los elementos de la diagonal de la matriz $\mathbf{S}^{-1}$. Aunque el método de los componentes principales para $\mathbf{R}$ puede considerarse como el método de los factores principales con estimados iniciales de comunalidad igual a la unidad o varianzas específicas iguales a cero, ambos métodos son filosóficamente y geométricamente distintos. En la práctica, sin embargo ambos producen con frecuencia factores de carga comparables, si se tiene un número grande de variables y un número pequeño de factores comunes.

Método de Máxima Verosimilitud

Si se supone que los factores comunes $\mathbf{F}$ y los factores específicos $\boldsymbol{\varepsilon}$ se encuentran normalmente distribuidos, entonces se podrán obtener los estimadores de máxima verosimilitud de las cargas de los factores y las varianzas específicas. Cuando la distribución conjunta de $\mathbf{F_j}$ y $\boldsymbol{\varepsilon_j}$ es normal, las observaciones $\mathbf{X_j - \mu = LF_j + \varepsilon_j}$ se distribuyen normalmente y la función de verosimilitud a maximizar es:

$$\begin{aligned} L(\boldsymbol{\mu},\boldsymbol{\Sigma}) &= (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} e^{-(1/2)tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x_j}-\bar{\mathbf{x}})(\mathbf{x_j}-\bar{\mathbf{x}})' + n(\bar{\mathbf{x}}-\boldsymbol{\mu})(\bar{\mathbf{x}}-\boldsymbol{\mu})'\right)\right]} \\ &= (2\pi)^{-\frac{(n-1)p}{2}} |\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}} e^{-(1/2)tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\mathbf{x_j}-\bar{\mathbf{x}})(\mathbf{x_j}-\bar{\mathbf{x}})'\right)\right]} \\ &\times (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-(n/2)\left((\bar{\mathbf{x}}-\boldsymbol{\mu})'\Sigma^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})\right)} \end{aligned} \qquad (24)$$

la cual depende de $\mathbf{L}$ y $\boldsymbol{\Psi}$ a través de $\boldsymbol{\Sigma} = \mathbf{LL}' + \boldsymbol{\Psi}$. Este modelo aún no está bien definido, debido a la multiplicidad de valores de L que posibilitan las transformaciones ortogonales. Se desea definir la unicidad de $\mathbf{L}$ mediante la imposición de la restricción

$$\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta} \quad \text{matriz diagonal} \qquad (25)$$

Los estimadores de máxima verosimilitud de $\hat{\mathbf{L}}$ y $\hat{\boldsymbol{\Psi}}$ se obtienen por maximización de (24) mediante métodos numéricos. Se cuenta con paquetes que realizan dichos cálculos.

Un resultado fundamental del procedimiento de máxima verosimilitud se resume a continuación

Resultado 1. Sea $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ una muestra aleatoria de una población $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ en donde $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ es la matriz de covarianza para m factores comunes en el modelo (4). Los estimadores de máxima verosimilitud $\hat{\mathbf{L}}, \hat{\boldsymbol{\Psi}}$, y $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ maximizan (24) sujeto a $\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}$ matriz diagonal.

Es posible demostrar[26] que los estimadores de máxima verosimilitud de las comunalidades están dados por:

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2 \quad i = 1, 2, \cdots, p \qquad (26)$$

de tal manera que

$$\begin{pmatrix}\text{(Proporción de la varianza}\\ \text{Muestral total debida al}\\ \text{j-ésimo factor)}\end{pmatrix} = \frac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \cdots + \hat{l}_{pj}^2}{s_{11} + s_{22} + \cdots s_{pp}} \qquad (27)$$

Cuando el análisis de máxima verosimilitud se realiza para variables estandarizadas a partir de la matriz de correlación,

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2 \quad i = 1, 2, \cdots, p \qquad (28)$$

se denominan los estimadores de máxima verosimilitud de las comunalidades y se evalúa la importancia de cada factor

$$\begin{pmatrix}\text{(Proporción de la varianza}\\ \text{- estandarizada muestral total}\\ \text{debida al j-ésimo factor)}\end{pmatrix} = \frac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \cdots + \hat{l}_{pj}^2}{p} \qquad (29)$$

En donde los $\hat{l}_{ij}$ denotan los elementos de $\hat{\mathbf{L}}_z$.

Determinación del Número de Factores

Existen técnicas para evaluar que tan bien ajusta a los datos el modelo con un número particular de factores comunes. Muchas técnicas son procedimientos informales que se basan más en la experiencia y la intuición que en algún modelo muestran formal. Algunas de estas técnicas son:

(1) Uno de los criterios más populares para indicar el número de factores es quedarse únicamente con aquellos factores asociados a los eigenvalores mayores a 1 (cuando se analiza P).
(2) Otro método informal es la llamada 'prueba scree'. Se realiza una gráfica SCREE de los eigenvalores y se elige el número de factores correspondiente al punto donde los eigenvalores empiezan a decrecer para formar una línea casi horizontal.

[26] Johnson Richard A., Wichern, Dean W. 2002. *Applied Multivariate Statistical Analysis*. Prentice Hall

En nuestro caso empleamos ambos métodos informales

Elementos Adicionales a Considerar en la Elección del Número de Factores[27]

1. No se incluyen factores triviales. Los factores triviales son aquellos que tienen una y sola una de las variables originales cargando sobre el factor. En general, las variables que cargan sólo sobre un factor, no están correlacionadas con las demás variables del conjunto de datos y esas variables, por sí mismas, son características subyacentes. En estos casos, probablemente lo mejor es eliminar esas variables y volver a iniciar el análisis de factores. Esto no significa que estas variables no sean importantes, sólo significa que son características de la población independientes de las que se están midiendo por las otras variables. No tiene sentido crear factores para esas variables, cuando se pueden emplear ellas mismas.

2. Algunos paquetes producirán matrices de diferencias entre las correlaciones observadas entre las variables y aquellas que se producen por la solución del modelo de análisis de factores. Si estas diferencia son pequeñas, podrá existir la posibilidad de reducir la cantidad de factores; por otra parte, si algunas diferencias son grandes (quizás mucho mayor que 0.25), entonce podría ser necesario incrementar el número de factores.

ROTACIÓN DE FACTORES

Si $\hat{\mathbf{L}}$ es la matriz de dimensión $p \times m$ de las cargas de los factores estimadas mediante cualquier método (componentes principales, máxima verosimilitud, etc.) entonces

$$\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}, \quad \text{en donde } \mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I} \qquad (30)$$

es una matriz $p \times m$ de cargas "rotada". Mas aún, la matriz de covarianza (o de correlación) estimada permanece invariante, ya que

$$\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} = \hat{\mathbf{L}}\mathbf{T}\mathbf{T}'\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}} = \hat{\mathbf{L}}^*\hat{\mathbf{L}}^{*\prime} + \hat{\boldsymbol{\Psi}} \qquad (31)$$

La ecuación (31) indica que la matriz residual, $\mathbf{S}_n - \hat{\mathbf{L}}\hat{\mathbf{L}}' - \hat{\boldsymbol{\Psi}} = \mathbf{S}_n - \hat{\mathbf{L}}^*\hat{\mathbf{L}}^{*\prime} - \hat{\boldsymbol{\Psi}}$, permanece sin cambios. Mas aún, las varianzas específicas $\hat{\Psi}_i$ y consecuentemente las comunalidades $\hat{h}_i^2$, no se alteran. Por lo tanto desde un punto de vista matemático, es indistinto si se determina $\hat{\mathbf{L}}$ o $\hat{\mathbf{L}}^*$.

Puesto que las cargas de los factores originales pueden no ser directamente interpretables, se acostumbra rotarlas, para obtener una estructura "adecuada" a la interpretación. Idealmente, se debe buscar un patrón de cargas tal que cada variable tenga un peso alto en un factor y un peso bajo o moderado en los factores restantes. Esta

[27] Chatfield C., Collins A.J. 1980. Introduction to Multivariate Analysis. Chapman & Hall

estructura, sin embargo no siempre es posible de obtener. A continuación se describen métodos gráficos y analíticos para la rotación de factores. Cuando m = 2 o cuando los factores comunes se consideran de dos en dos, la transformación a una estructura más simple puede determinarse gráficamente. Los factores comunes no correlacionados se consideran como vectores unitarios en ejes perpendiculares de coordenadas. Una gráfica de cada par de cargas de factores (l_{i1}, l_{i2}), genera p puntos, cada uno de los cuales corresponde a una variable. La rotación de los ejes coordenados un ángulo ϕ genera nuevas cargas l_{ij}^{*}, las cuales se obtienen mediante la relación

$$\hat{\mathbf{L}}^{*} = \hat{\mathbf{L}}\,\mathbf{T} \qquad (32)$$

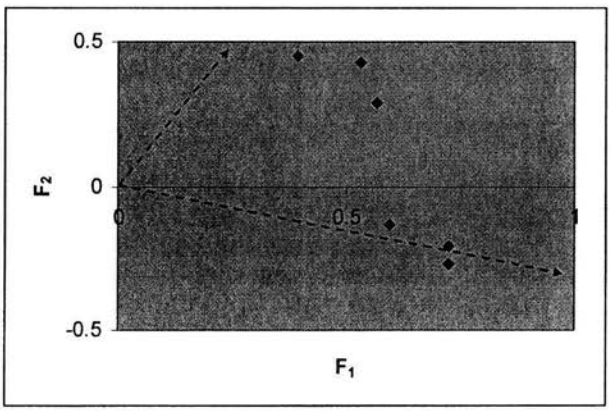
en donde

$$T = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$$ en sentido de las manecillas del reloj

$$\mathrm{T} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$ en sentido contrario a las manecillas del reloj

La relación en (32) raramente se utiliza en un análisis gráfico de dos dimensiones. En este caso los conglomerados de variables son observables visualmente y la identificación de factores comunes puede efectuarse sin necesidad de inspeccionar las magnitudes de los factores rotados, sino rotando los ejes originales un ángulo ϕ tal que los nuevos ejes coordenados pasen a través de los conglomerados de las variables desplegadas en el plano de dos dimensiones F_1, F_2 (*ver Gráfica 3.2.1*).

Gráfica 3.2.1
Rotación de Factores



Por otra parte cuando m > 2 la orientación de los ejes no es fácilmente identificable por lo que se requiere determinar la magnitud de los factores rotados, a fin de lograr una adecuada interpretación de las variables originales. A continuación se describe un procedimiento para la elección de una matriz ortogonal **T** que satisface el objetivo de obtención de una estructura conducente a la interpretación de factores.

Kaiser [22] sugirió una medida analítica de la estructura simple conocida como el *criterio de rotación varimax*. Se define $\tilde{l}_{ij}^{*} = \tilde{l}_{ij} / \hat{h}_i$ como los coeficientes rotados escalados por la raíz cuadrada de las comunalidades. El procedimiento varimax selecciona la transformación ortogonal **T** que maximiza

$$V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p} l_{ij}^{*4} - \left(\sum_{i=1}^{p} l_{ij}^{*2}\right)^2 / p\right] \qquad (33)$$

El escalamiento de los coeficientes rotados $\hat{l}_{ij}^{*}$ tiene el efecto de proporcionar a las variables con comunalidad más pequeña, un peso relativamente mayor en la determinación de la estructura simple. Una vez determinada la matriz de transformación **T**, las cargas $\hat{l}_{ij}^{*}$ se multiplican por $\hat{h}_i$ de tal manera que las comunalidades originales sean preservadas.

La ecuación (33) puede interpretarse de la siguiente manera

$$V \propto \sum_{j=1}^{m}\left(\begin{array}{c}\text{varianza de los cuadrados de las}\\ \text{cargas escaladas del j-ésimo factor}\end{array}\right) \qquad (35)$$

Efectivamente la maximización de V equivale a distribuir los cuadrados de las cargas en cada factor lo máximo posible, por lo que se esperaría encontrar grupos tanto de coeficientes grandes, como de coeficientes despreciables en cada columna de la matriz rotada $\hat{\mathbf{L}}^{*}$.

Existen algoritmos en los paquetes estadísticos (SAS, SPSS, MINITAB, S-Plus, etc.) para determinar la rotación varimax. Las rotaciones de las cargas de los factores que se obtienen a partir de procedimientos distintos (factores principales, máxima verosimilitud, etc.) en general no van a coincidir. De la misma manera, el patrón de cargas rotadas puede cambiar considerablemente si se incluyen factores comunes adicionales en la rotación. En caso de existir un solo factor dominante, generalmente éste podrá quedar enmascarado por la rotación ortogonal. Por otra parte, es posible mantener fijo el factor dominante, rotando solamente el resto de los factores.

CALIFICACIONES DE LOS FACTORES

En análisis de factores, generalmente el interés se centra en los parámetros del modelo de factores. Sin embargo, los valores estimados de los factores comunes para cada individuo, denominados *calificaciones de los factores*, también podrán requerirse. Estas cantidades se utilizan tanto para fines de clasificación, como variables en análisis subsecuentes.

Las calificaciones de los factores no son estimaciones de parámetros desconocidos en el sentido usual. Éstas más bien son estimaciones del valor que toma cada factor $\mathbf{F}_j$, $j = 1, 2, \cdots, m$, para cada individuo. Es decir las calificaciones de los factores $\hat{\mathbf{f}}_j$ = estimado de los valores $\mathbf{f}_j$ en $\mathbf{F}_j$ (para el j-ésimo caso).

La estimación se complica por el hecho de que ε no se conoce y L se estima. Para superar esta dificultad se han desarrollado varios procedimientos heurísticos. A continuación se describen dos de ellos, ambos con los siguientes dos elementos en común:

1. Tratan las cargas de los factores estimadas $\hat{l}_{ij}$ y las varianzas específicas $\hat{\Psi}_i$ como si fuesen los valores verdaderos.
2. Implican transformaciones lineales de las variables originales, generalmente centradas o estandarizadas. Típicamente las cargas rotadas estimadas más que las cargas originales estimadas, son las que se utilizan para calcular las calificaciones de los factores. Las fórmulas de cálculo dadas a continuación no cambian cuando se sustituyen en el cálculo las cargas rotadas por las cargas sin rotar.

Método de Mínimos Cuadrados Ponderados

Supongamos que el vector de medias $\boldsymbol{\mu}$, las cargas de los factores $\mathbf{L}$ y las varianzas específicas son conocidas para el modelo de factor

$$\underset{(px1)}{\mathbf{X}} - \underset{(px1)}{\boldsymbol{\mu}} = \underset{(pxm)}{\mathbf{L}} \underset{(mx1)}{\mathbf{F}} + \underset{(px1)}{\boldsymbol{\varepsilon}}$$

Adicionalmente, se considera a los factores $\boldsymbol{\varepsilon}' = \left[\varepsilon_1, \varepsilon_2, \cdots \varepsilon_p\right]$ como errores. Puesto que las $Var(\varepsilon_i) = \Psi_i, \quad i = 1, \cdots, p$ no son iguales, Bartlett sugirió el uso de la técnica de mínimos cuadrados ponderados para estimar los valores de los factores comunes.
La suma del cuadrado de los errores ponderada por el recíproco de sus varianzas, es

$$\sum_{i=1}^{p} \frac{\varepsilon_i^2}{\Psi_i} = \boldsymbol{\varepsilon}' \boldsymbol{\Psi}^{-1} \boldsymbol{\varepsilon} = (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f})' \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu} - \mathbf{L}\mathbf{f}) \qquad (36)$$

Bartlett propuso elegir los estimadores $\hat{\mathbf{f}}$ de $\mathbf{f}$ para minimizar (36). La solución está dada por:

$$\hat{\mathbf{f}} = \left(\mathbf{L}' \boldsymbol{\Psi}^{-1} \mathbf{L}\right)^{-1} \mathbf{L}' \boldsymbol{\Psi}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \qquad (37)$$

En base a (37) se consideran los estimados $\hat{\mathbf{L}}$, $\hat{\boldsymbol{\Psi}}$, y $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$, como los valores verdaderos y se obtienen las calificaciones de los factores para el j-ésimo individuo como

$$\hat{\mathbf{f}}_j = \left(\hat{\mathbf{L}}' \hat{\boldsymbol{\Psi}}^{-1} \hat{\mathbf{L}}\right)^{-1} \hat{\mathbf{L}}' \hat{\boldsymbol{\Psi}}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \quad j = 1, 2, \cdots, n \qquad (38)$$

Cuando $\hat{\mathbf{L}}$ y $\hat{\boldsymbol{\Psi}}$ se determinan por el método de máxima verosimilitud, estos estimados deberán satisfacer la condición de unicidad $\hat{\mathbf{L}}' \hat{\boldsymbol{\Psi}}^{-1} \mathbf{L} = \boldsymbol{\Delta}$, matriz diagonal. Entonces se tendrá el siguiente estimador para las calificaciones de los factores:

$$\begin{aligned} \hat{\mathbf{f}}_j &= \left(\hat{\mathbf{L}}' \hat{\boldsymbol{\Psi}}^{-1} \hat{\mathbf{L}}\right)^{-1} \hat{\mathbf{L}}' \hat{\boldsymbol{\Psi}}^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) \\ &= \hat{\boldsymbol{\Delta}}^{-1} \hat{\mathbf{L}}' \hat{\boldsymbol{\Psi}}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), \quad j = 1, 2, \cdots, n \end{aligned} \qquad (39)$$

o bien si se parte de la matriz de correlación

$$\hat{\mathbf{f}}_j = \left(\hat{\mathbf{L}}_z'\hat{\boldsymbol{\Psi}}_z^{-1}\hat{\mathbf{L}}_z\right)^{-1}\hat{\mathbf{L}}_z'\hat{\boldsymbol{\Psi}}_z^{-1}\mathbf{z}_j$$
$$= \hat{\boldsymbol{\Delta}}_z^{-1}\hat{\mathbf{L}}_z'\hat{\boldsymbol{\Psi}}_z^{-1}\mathbf{z}_j, \quad j = 1,2,\cdots,n$$

en donde $\mathbf{z}_j = \mathbf{D}^{-1/2}\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)$ y $\hat{\boldsymbol{\rho}} = \hat{\mathbf{L}}_z\hat{\mathbf{L}}_z' + \hat{\boldsymbol{\Psi}}_z$

Las calificaciones de los factores generadas por (39) tienen un vector de medias muestral igual a **0** y covarianzas muestrales iguales a cero[28].

Si se utilizan las cargas rotadas $\hat{\mathbf{L}}^* = \hat{\mathbf{L}}\mathbf{T}$ en lugar de las cargas originales en (39), las calificaciones subsecuentes de los factores $\hat{f}_j^*$, se relacionan con $\hat{\mathbf{f}}_j$ mediante la expresión $\hat{\mathbf{f}}_j^* = \mathbf{T}'\hat{\mathbf{f}}_j$, $j = 1,\cdots,n$.

Cuando las cargas de los factores se estiman por el método de componentes principales, se acostumbra generar las calificaciones de los factores mediante el método ordinario de mínimos cuadrados. Implícitamente esto equivale a suponer que todas las Ψ_i son iguales o casi iguales. Las calificaciones de los factores se determinan mediante

$$\hat{\mathbf{f}}_j = \left(\hat{\mathbf{L}}'\hat{\mathbf{L}}\right)^{-1}\hat{\mathbf{L}}'\left(\mathbf{x}_j - \bar{\mathbf{x}}\right)$$

o

$\hat{\mathbf{f}}_j = \left(\hat{\mathbf{L}}_z'\hat{\mathbf{L}}_z\right)^{-1}\hat{\mathbf{L}}_z'\mathbf{z}_j$ para los datos estandarizados. Puesto que

$\mathbf{L} = \left[\sqrt{\hat{\lambda}_1}\mathbf{e}_1, \sqrt{\hat{\lambda}_2}\mathbf{e}_2, \cdots, \sqrt{\hat{\lambda}_m}\mathbf{e}_m\right]$ se tiene

$$\hat{f}_j' = \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}}\hat{\mathbf{e}}_1'\left(\mathbf{x}_j - \bar{\mathbf{x}}\right) \\ \frac{1}{\sqrt{\hat{\lambda}_2}}\hat{\mathbf{e}}_2'\left(\mathbf{x}_j - \bar{\mathbf{x}}\right) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}}\hat{\mathbf{e}}_m'\left(\mathbf{x}_j - \bar{\mathbf{x}}\right) \end{bmatrix} \qquad (40)$$

Para estas calificaciones de los factores,

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{f}_j = \mathbf{0} \quad \text{(media muestral)}$$

y

$$\frac{1}{\text{n-1}}\sum_{j=1}^{n}\hat{\mathbf{f}}_j\hat{\mathbf{f}}_j' = \mathbf{I} \quad \text{(covarianza muestral)}$$

[28] Johnson, Richard A., Wichern, Dean W. 2002. *Applied Multivariate Statistical Analysis*. Prentice Hall

De la ecuación (40), se observa que $\hat{\mathbf{f}}_j$ corresponde al vector de los primeros m componentes principales (escalados), evaluados en $\mathbf{x}_j$.

Método de Regresión

Partiendo nuevamente del modelo original de factores $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\,\mathbf{F} + \boldsymbol{\varepsilon}$, se tratan inicialmente a la matriz de cargas $\mathbf{L}$ y a la matriz de varianzas específicas $\boldsymbol{\Psi}$ como conocidas. Cuando los factores comunes F y los factores específicos (o errores) se encuentran normalmente distribuidos de manera conjunta con medias y covarianzas dadas por (3), la combinación lineal $\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\,\mathbf{F} + \boldsymbol{\varepsilon}$ se distribuye $N_p\left(\mathbf{0}, \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}\right)$. Mas aún, la distribución de $\left(\mathbf{X} - \boldsymbol{\mu}\right)$ y $\mathbf{F}$ es $N_{m+p}\left(\mathbf{0}, \boldsymbol{\Sigma}^*\right)$, en donde

$$\Sigma^* = \underset{(m+p)(m+p)}{\begin{bmatrix} \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi} & \mathbf{L} \\ \mathbf{L}' & \mathbf{I} \end{bmatrix}} \qquad (41)$$

y $\mathbf{0}$ es un vector de (m + p) x 1. La distribución condicional de F dado que X es normal multivariada con media

$$E\left(\mathbf{F}/\mathbf{x}\right) = \mathbf{L}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}\right) = \mathbf{L}'\left(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}\right)^{-1}\left(\mathbf{x} - \boldsymbol{\mu}\right) \quad (42)$$

y

covarianza

$$Cov\left(\mathbf{F}/\mathbf{x}\right) = \mathbf{I} - \mathbf{L}'\boldsymbol{\Sigma}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}'\left(\mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}\right)^{-1}\mathbf{L} \qquad (43)$$

Las cantidades $L'\left(LL' + \Psi\right)^{-1}$ en (42) son los coeficientes en una regresión multivariada de los factores en las variables. La estimación de estos coeficientes produce calificaciones de los factores que son análogas a las estimaciones de los valores de la media condicional en un análisis de regresión multivariada. Consecuentemente, dado un vector de observaciones $\mathbf{x}_j$ y tomando los estimadores de máxima verosimilitud de $\hat{\mathbf{L}}$ y $\hat{\boldsymbol{\Psi}}$ como los valores verdaderos, el vector correspondiente a la calificación del j-ésimo factor está dado por

$$\hat{\mathbf{f}}_j = \hat{\mathbf{L}}'\hat{\boldsymbol{\Sigma}}^{-1}\left(\mathbf{x}_j - \bar{\mathbf{x}}\right) = \hat{\mathbf{L}}'\left(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}\right)^{-1}\left(\mathbf{x}_j - \bar{\mathbf{x}}\right), \quad j = 1, 2, \cdots, n \qquad (44)$$

El cálculo de $\hat{\mathbf{f}}_j$ en (44) se simplifica usando la matriz identidad como sigue:

$$\underset{(mxp)\ (pxp)}{\hat{\mathbf{L}}'\left(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\Psi}}\right)^{-1}} = \underset{(mxm)\quad (mxp)(pxp)}{\left(\mathbf{I} + \hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}\hat{\mathbf{L}}\right)^{-1}\hat{\mathbf{L}}'\hat{\boldsymbol{\Psi}}^{-1}} \qquad (45)$$

PERSPECTIVAS Y ESTRATEGIA DEL ANÁLISIS DE FACTORES

En el ajuste del modelo de factores, deberán tomarse una serie de decisiones, dentro de las cuales probablemente la más importante corresponda a la elección *m*, el número de factores comunes. Aunque para muestras grandes, se cuenta con pruebas estadísticas de adecuación del modelo, estos métodos son adecuados solamente cuando los datos se encuentran normalmente distribuidos. Más aún, la prueba de hipótesis respecto a la suficiencia en el número de factores *m*, seguramente rechazará la hipótesis de nulidad cuando *m* es pequeña si el número de variables y de observaciones es grande. Sin embargo, es en este caso cuando el modelo de factor muestra su mayor utilidad. Generalmente, la elección de *m* se basa en una combinación de los siguientes tres aspectos (1) la proporción de la varianza muestral explicada, (2) el conocimiento de la materia bajo estudio y (3) la coherencia de los resultados.

La elección del método de solución y del tipo de rotación son decisiones menos cruciales.

Actualmente, el análisis de factores continúa teniendo aspectos muy subjetivos y ninguna estrategia en particular resultará incontrovertible.

3.3 Análisis de Conglomerados

INTRODUCCIÓN

Una técnica exploratoria importante en el análisis de datos multivariados es aquella que permite la identificación de agrupamientos "naturales" entre individuos con características similares. El estudio de dichos agrupamientos proporciona una manera informal de identificar observaciones atípicas, así como sugerir hipótesis en cuanto a relaciones existentes entre los individuos bajo estudio.

La formación de grupos o conglomerados es una técnica distinta a otros métodos multivariados de clasificación como puede ser el caso del análisis discriminante. Estos últimos se refieren a un número conocido de grupos o categorías definidas a priori, de tal manera que el objetivo operacional consiste en la asignación de nuevas observaciones a dichos grupos. En el análisis de conglomerados no existe una hipótesis en cuanto al número y estructura de los grupos a considerar. El agrupamiento se realiza en base a las similitudes[29] o distancias (diferencias) entre individuos. La entrada requerida para el análisis de conglomerados, corresponde a las medidas de similitud entre individuos o bien los datos a partir de los cuales éstas puedan ser determinadas.
En síntesis, el objetivo básico del análisis de conglomerados consiste en el descubrimiento de agrupamientos naturales entre objetos (o entre variables). Con este

---

[29] El nivel de similitud $s(ij)$ entre dos conglomerados *i* y *j* se define como

$s(ij) = 100.\dfrac{(d_{\max} - d_{ij})}{d_{\max}}$, en donde $d_{\max}$ es la máxima distancia entre observaciones y d_{ij} es la distancia entre conglomerados

fin deberá definirse en primer lugar una escala cuantitativa que permita medir la asociación (o similitud) entre los objetos bajo estudio. En la siguiente sección se discuten algunas medidas de similitud. Posteriormente se discutirán algunos de los algoritmos comúnmente empleados en la clasificación de individuos en conglomerados. Estos procedimientos son complementarios a los métodos gráficos de clasificación, como las caras de Chernoff, diagramas de estrellas y curvas de Andrews empleadas con frecuencia en las exploraciones preliminares de datos.

MEDIDAS DE SIMILITUD (O PROXIMIDAD)

Para obtener una estructura o agrupamiento de individuos a partir de un conjunto complejo de datos, se requiere generalmente de una medida de "proximidad" o de "similitud". Las cuestiones más importantes a considerar son: la naturaleza de las variables (discretas, continuas, binarias), la escala de medición (nominal, ordinal, intervalo, razón) y el conocimiento acerca de la materia en estudio.

Cuando se desea agrupar a los individuos o casos, la medida de proximidad generalmente se encuentra definida por algún tipo de distancia. Por otra parte, las variables generalmente se agrupan sobre la base de sus coeficientes de correlación o medidas análogas de asociación.

Distancias y Coeficientes de Similitud para Pares de Individuos

La distancia euclidiana entre dos observaciones (individuos) p-dimensionales $\mathbf{x}' = [x_1, x_2, \cdots, x_p]$ y $\mathbf{y}' = [y_1, y_2, \cdots, y_p]$, está dada por

$$\begin{aligned} d(\mathbf{x},\mathbf{y}) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_p - y_p)^2} \\ &= \sqrt{(\mathbf{x}-\mathbf{y})'(\mathbf{x}-\mathbf{y})} \end{aligned} \qquad (1)$$

Otra medida de distancia, es la métrica Minkowski

$$d(\mathbf{x},\mathbf{y}) = \left[\sum_{i=1}^{p} |x_i - y_i|^m\right]^{1/m} \qquad (2)$$

Para $m = 1$, $d(\mathbf{x},\mathbf{y})$ mide la distancia *"en manzana urbana"* entre dos puntos en p-dimensiones. Esta distancia también se conoce como *Manhattan*. Para $m = 2$, $d(\mathbf{x},\mathbf{y})$ se convierte en la distancia euclidiana. En general, variando m se altera el peso dado a las diferencias máxima y mínima. La distancia Manhattan describe distancias en una configuración rectilínea y es más robusta contra observaciones atípicas que la distancia euclidiana.

Otras medidas de "distancia" o disimilaridad se encuentran dadas por la métrica de Canberra y el coeficiente de Czekanowski. Ambas medidas se encuentran definidas únicamente para el caso de variables no negativas. La distancia de Canberra puede utilizarse para variables discretas o continuas, o para una mezcla de ambas.

Métrica de Canberra:

$$d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{p} \frac{|x_i - y_i|}{(x_i + y_i)} \qquad (3)$$

Coeficiente de Czekanowski:

$$d(\mathbf{x},\mathbf{y}) = 1 - \frac{2\sum_{i=1}^{p} \min(x_i, y_i)}{\sum_{i=1}^{p} (x_i + y_i)} \qquad (4)$$

Para el agrupamiento de objetos, siempre que sea posible deberán emplearse las distancias *"verdaderas"* o que satisfacen las siguientes propiedades:

$$d(P,Q) = d(Q,P)$$
$$d(P,Q) > 0 \text{ si } P \neq Q$$
$$d(P,Q) = 0 \text{ si } P = Q$$
$$d(P,Q) \leq d(P,R) + d(R,Q) \quad \text{(desigualdad triangular)}$$

en donde P y Q son dos puntos y R representa a otro punto intermedio

Por otra parte, la mayoría de los algoritmos para la formación de conglomerados aceptarán distancias asignadas de manera subjetiva, aún cuando no satisfagan, por ejemplo la desigualdad triangular.

Cuando los objetos no pueden representarse mediante un conjunto de mediciones *p*-dimensionales, cada par de objetos frecuentemente se compara en base a la presencia o ausencia de cierta característica. Los objetos similares contienen un mayor número de características en común que los individuos no similares. La presencia o ausencia de alguna característica puede describirse matemáticamente introduciendo una variable binaria, la cual toma el valor de 1 si la característica está presente y de 0 cuando no lo está. Sea x_{ij} la calificación (1 ó 0) de la j-ésima variable binaria en el i-ésimo individuo y x_{kj} la calificación (nuevamente 1 ó 0) de la j-ésima variable en el k-ésimo individuo, $j = 1, 2, \cdots, p$. Consecuentemente

$$(x_{ij} - x_{kj})^2 = \begin{cases} 0 & \text{si } x_{ij} = x_{kj} = 1 \text{ ó } x_{ij} = x_{kj} = 0 \\ 1 & \text{si } x_{ij} \neq x_{kj} \end{cases} \qquad (5)$$

y la distancia euclidiana $\sum_{j=1}^{p} (x_{ij} - x_{kj})^2$, proporciona la frecuencia del número de calificaciones distintas. Una distancia grande corresponde a un número elevado de calificaciones diferentes; es decir, individuos no similares.

Aunque las distancias basadas en (5) se pueden utilizar para medir la similitud, esta medición adolece del defecto de que las calificaciones 1, 1 y 0, 0, tienen el mismo peso. Por ejemplo, puede resultar más relevante desde el punto de vista de la agrupación, cuando existan dos individuos con la misma característica (1, 1) a cuando existan dos individuos con ausencia de dicha característica (0, 0). En este último caso, la similitud (ausencia de la característica) puede ser por completo irrelevante. Un caso típico se

tiene en la medicina, cuando resulta irrelevante la comparación entre dos individuos con *ausencia de enfermedad.* Con el propósito de tomar en consideración diferencias en el tratamiento entre ambas clases de concordancia entre individuos, se han sugerido diversos esquemas en la definición de coeficientes de similitud. Para introducir dichos esquemas, las frecuencias se pueden agrupar en una tabla de contingencia como la que se muestra a continuación:

Tabla 3.3.1

		Individuo k		
		1	0	Totales
Individuo i	1	a	b	$a+b$
	0	c	d	$c+d$
Totales		$a+c$	$b+d$	$p=a+b+c+d$

En la Tabla 3.3.1, a representa la frecuencia de las calificaciones 1 – 1, b la correspondiente a 0 – 1 y así sucesivamente.

En la Tabla 3.3.2 se define una lista de coeficientes de similitud comúnmente empleados, cuando se utilizan frecuencias.

Los coeficientes 1, 2 y 3 en la Tabla 3.3.2 se encuentran monotónicamente relacionados. Supongamos que se determina el coeficiente 1 para dos tablas de contingencia, Tabla I y Tabla II. En este caso si tenemos que $(a_I + d_I)/p \geq (a_{II} + d_{II})/p$, también tendremos que $2(a_I + d_I)/[2(a_I + d_I) + b_I + c_I] \geq 2(a_{II} + d_{II})/[2(a_{II} + d_{II}) + b_{II} + c_{II}]$ y el coeficiente 3 será también al menos tan grande para la Tabla I que para la Tabla II. Los coeficientes 5, 6 y 7 también retienen su orden relativo.

Tabla 3.3.2

Medida	Característica
1. $\frac{a+d}{p}$	Mismo peso para calificaciones 1 – 1 y 0 - 0
2. $\frac{2(a+d)}{2(a+d)+b+c}$	Doble peso para calificaciones 1 – 1 y 0 - 0
3. $\frac{a+d}{a+d+2(b+c)}$	Doble peso para pares con calificación distinta
4. $\frac{a}{p}$	No se incluyen calificaciones 0 – 0 en el numerador
5. $\frac{a}{a+b+c}$	No se incluyen calificaciones 0 – 0 en el numerador o en el denominador (Las calificaciones 0 – 0, se tratan como irrelevantes)

6. $\frac{2a}{2a+b+c}$	No se incluyen calificaciones 0 – 0 en el numerador o denominador. Doble peso para las calificaciones 1 – 1.
7. $\frac{a}{a+2(b+c)}$	No se incluyen calificaciones 0 – 0 en el numerador o denominador. Doble peso para calificaciones diferentes
8. $\frac{a}{b+c}$	Razón de calificaciones iguales a calificaciones diferentes, con exclusión de calificaciones 0 – 0.

La monotonicidad es importante, debido a que ciertos procedimientos de agrupamiento no se afectan si se cambia la definición de similitud de tal manera que el orden relativo de similitud se mantenga. Los procedimientos jerárquicos de liga simple y liga completa que se discuten más adelante no se ven afectados. Para estos métodos, cualquier elección de los coeficientes 1, 2 y 3 producirá los mismos agrupamientos. De la misma manera, la elección de los coeficientes 5, 6 y 7 proporcionará los mismos conglomerados.

Hasta ahora se ha descrito la construcción de distancias y similitudes. Siempre será posible construir similitudes a partir de distancias. Por ejemplo, podrá definirse

$$\tilde{s}_{ik} = \frac{1}{1+d_k} \qquad (6)$$

En donde $0 < s_{ik} \leq 1$ corresponde a la similitud entre los individuos i y k y d_k es la distancia correspondiente.

Sin embargo, no siempre podrán construirse distancias a partir de similitudes que satisfagan las condiciones arriba enunciadas. Según fue demostrado por Gower [23, 24], lo anterior puede llevarse a cabo únicamente cuando la matriz de similitudes es *definida no negativa*. Con la condición de *definida no negativa* y estableciendo la escala de máxima similitud de tal manera que $\tilde{s}_{ii} = 1$

$$d_{ik} = \sqrt{2(1-\tilde{s}_{ik})} \qquad (7)$$

tiene las propiedades de una distancia.

Para resumir esta sección se hace notar que existen varias maneras de medir la similitud entre pares de objetos. El método más comúnmente empleado en la práctica corresponde al uso de distancias [fórmulas (1) a (5)] o los coeficientes de la Tabla 2.3.2 para agrupar *individuos*. En ocasiones, también se podrán emplear frecuencias simples. Los métodos anteriormente descritos contienen un alto grado de subjetividad. A continuación se discuten algunos métodos para la formación de conglomerados.

MÉTODOS JERÁRQUICOS DE AGRUPAMIENTO

Ante la imposibilidad de examinar todas las alternativas posibles de agrupamiento en un conjunto de individuos, se han desarrollado una variedad de algoritmos, cuyo propósito consiste en encontrar un arreglo "razonable", sin necesidad de evaluar la totalidad de opciones posibles.

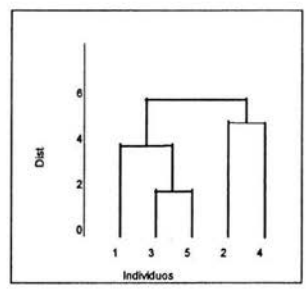
Las técnicas jerárquicas de agrupamiento proceden ya sea mediante una serie de uniones sucesivas o bien mediante una serie de divisiones sucesivas. Los *métodos jerárquicos por aglomeración*, inician considerando a cada individuo como un grupo. Por lo tanto al comienzo existirán tantos conglomerados como individuos. Los objetos más parecidos entre sí serán los primeros en agruparse y estos grupos iniciales se fusionarán con otros individuos o grupos de acuerdo a su similitud. Eventualmente, conforme decrece la similitud entre individuos, todos los subgrupos se unen en un conglomerado único.

Los *métodos jerárquicos por división* operan en la dirección opuesta. Se tiene un grupo inicial formado por todos los individuos, el cual se divide en dos subgrupos de tal manera que los objetos en un subgrupo se encuentren "alejados" de los objetos en el otro grupo considerado. Estos subgrupos se dividen subsiguientemente en no similares; el proceso continúa hasta que existan tantos subgrupos como objetos – es decir hasta que cada objeto represente en sí mismo un grupo.

Los resultados de ambos métodos, por aglomeración o por división se despliegan mediante un diagrama bi-dimensional denominado dendrograma o árbol jerárquico. Según se explica más adelante, el dendrograma ilustra las uniones o divisiones que se han llevado a cabo en los niveles sucesivos. En el dendrograma, se grafican distancias o similitudes en el eje vertical e individuos en el horizontal. El dendrograma muestra de esta manera las distancias o nivel de similitud al que se realiza la fusión, Figura 2.3.1.

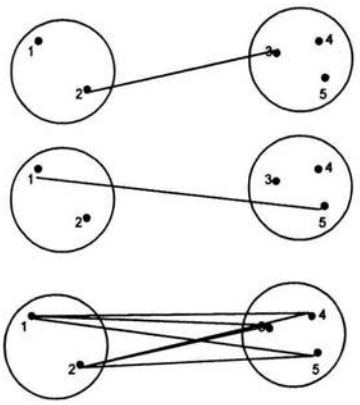
En la Figura2.3.1, los individuos 3 y 5 se fusionan a la distancia 2; a su vez éstos se unen con el individuo 1 en la distancia 3. Por otra parte, los individuos 2 y 4 se unen entre sí a la distancia 5, para amalgamarse con el conglomerado formado por los individuos 1, 3 y 5 a la distancia 6.

Figura 3.3.1
Dendrograma para Distancias entre Cinco Individuos



A continuación se explican los métodos jerárquicos por aglomeración; particularmente los *métodos de liga.*

Los tres tipos de liga más utilizados son: la liga simple (mínima distancia o vecino más cercano), liga completa (máxima distancia o vecinos distantes) y liga media (distancia promedio). La integración de conglomerados mediante estos tres criterios se ilustra esquemáticamente en la Figura 2.3.2.



d_{23} liga simple

d_{15} liga completa

$$\frac{d_{13}+d_{14}+d_{15}+d_{23}+d_{24}+d_{25}}{6}$$

liga promedio

Figura 3.3.2
Distancias entre Conglomerados, Según Tipo de liga

De la figura anterior se observa que la liga simple resulta cuando los grupos se fusionan de acuerdo a la distancia entre los miembros más cercanos. La liga completa ocurre cuando los grupos se fusionan de acuerdo a la distancia entre sus miembros más lejanos. En el caso de la liga promedio, la fusión ocurre de acuerdo a la distancia promedio entre cada par de miembros en sus respectivos conjuntos.

A continuación se mencionan los pasos comprendidos en el algoritmo jerárquico de agrupamiento de N objetos (individuos o variables).

1. Iniciar con N conglomerados, cada uno conteniendo una sola entidad y una matriz simétrica $N\,x\,N$ de distancias (o similitudes) $\mathbf{D} = \{d_{ik}\}$
2. Identificar en la matriz el par más cercano (o más parecido) de agrupamientos. Definir la distancia entre los grupos "más próximos" U y V como la distancia d_{UV}
3. Unir los conglomerados U y V. Asignar al nuevo conglomerado el nombre (UV). Actualizar las entradas en la matriz de distancias mediante (a) eliminando las filas y columnas correspondientes a los conglomerados U y V y (b) agregando una nueva columna que proporcione las distancias entre el conglomerado (UV) y el resto de los conglomerados.
4. Repetir los pasos 2 y 3 un total de $N - 1$ veces. (Todos los objetos deberán formar un solo conglomerado al concluir la ejecución del algoritmo). Registrar la identidad de los conglomerados que se unen y los niveles (distancias o similitudes) en las cuales la fusión se lleva a cabo.

En el paso 3 del algoritmo, puede usarse el método de la liga simple para calcular la distancia entre (UV) y cualquier otro conglomerado W, se calcula mediante

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \qquad (8)$$

En este caso las cantidades d_{UW} y d_{VW}, son las distancias entre los vecinos más cercanos en los conglomerados U y W y los conglomerados V y W respectivamente.

Los resultados del agrupamiento mediante la liga simple, se pueden mostrar gráficamente mediante el dendrograma o diagrama de árbol. Las ramas en el árbol representan los conglomerados. Las ramas se unen (o fusionan) en los nodos cuyas posiciones en términos de un eje de distancias (o similitud) indican el nivel en el que ocurrió la fusión.

Puesto que la liga simple une conglomerados a través de la distancia más corta entre ellos, esta técnica no permite discernir entre conglomerados pobremente separados. La tendencia de la liga simple de unir conglomerados con estructura de cadena (encadenamiento), puede producir agrupamientos equivocados, cuando los extremos opuestos en el encadenamiento, son de hecho objetos o individuos no similares.

La formación de conglomerados mediante el método de liga simple resulta invariante a la asignación de distancias (similitud) que proporciona el mismo ordenamiento que las distancias originales. En particular cualquier conjunto de similitudes seleccionado de la Tabla 2 conducirá al mismo agrupamiento.

Liga Completa

La formación de conglomerados mediante el método de liga completa procede de manera similar a la de la liga simple, con una importante excepción: en cada etapa las distancias (similitudes) entre los dos elementos, uno de cada conglomerado, corresponden a la distancia mayor. Por consiguiente, la liga completa garantiza que todos los elementos dentro de un conglomerado se encuentren dentro de una distancia máxima (o mínima similitud) con respecto a los otros elementos en el conglomerado.

El algoritmo general de aglomeración inicia nuevamente encontrando la mínima distancia en $\mathbf{D} = \{d_{ik}\}$ y fusionando los objetos correspondientes, tales como U y V, para obtener el agrupamiento (UV). En el paso 3 del algoritmo general, la distancia entre (UV) y cualquier otro conglomerado W se calcula mediante

$$d_{(UV)W} = \max\{d_{UV}, d_{VW}\} \qquad (9)$$

En este caso d_{UW} y d_{VW} representan las distancias entre los miembros más distantes en los conglomerados U y W y los conglomerados V y W, respectivamente

Liga Promedio

La liga promedio trata la distancia entre dos conglomerados como las distancias promedio entre todos los pares de elementos, en donde un elemento del par pertenece a un conglomerado y el otro elemento al otro conglomerado.

Nuevamente la entrada en la ejecución del algoritmo pueden ser las distancias o similitudes. El método de liga promedio procede de acuerdo al algoritmo general. Se inicia mediante la identificación de la distancia en la matriz $\mathbf{D} = \{d_{ik}\}$ para encontrar a los objetos más cercanos (similares) – por ejemplo U y V, estos elementos se fusionan

para formar el conglomerado (UV). En el paso 3 del algoritmo general de aglomeración, la distancia entre (UV) y otro conglomerado W se determina mediante

$$d_{(UV)W} = \frac{\sum_i \sum_k d_k}{N_{(UV)} N_W} \qquad (10)$$

En donde d_{ik} es la distancia entre el objeto i del conglomerado (UV) y el objeto k en el conglomerado W y $N_{(UV)}$ y N_W son el número de elementos en los conglomerados (UV) y W respectivamente. Ward [25] consideró procedimientos jerárquicos de agrupación basados en minimizar "la pérdida de información" de los grupos fusionados.

La mayoría de los métodos de agrupamiento no considera el tratamiento formal de fuentes de error y variación en los procedimientos jerárquicos. Esto significa que los métodos de formación de conglomerados son sensibles a las observaciones atípicas.
En los métodos jerárquicos no se prevé la reasignación de objetos que hayan sido agrupados incorrectamente en las etapas previas. Consecuentemente la configuración final de conglomerados debe examinarse para verificar su sensibilidad.

Para cualquier problema en particular, es conveniente intentar distintos métodos de agrupación y dentro de cada método, distintas maneras de asignar las distancias (similitudes). Si el resultado en los diferentes métodos empleados es aproximadamente consistente, entonces podrá concluirse la probable presencia de agrupamientos "naturales". De lo contrario, estaremos hablando de una jerarquización de individuos.

La estabilidad de las soluciones jerárquicas en ocasiones puede verificarse aplicando el algoritmo a los datos antes y después de introducir pequeñas perturbaciones en la matriz de datos. Si los conglomerados son claramente identificables, en general los agrupamientos obtenidos antes y después de la perturbación, deberán coincidir.

Los empates en la matriz de distancias podrán producir soluciones múltiples al problema de agrupamiento; es decir los dendrogramas correspondientes a diferentes tratamientos de las similitudes (o distancias) podrán ser diferentes, particularmente en los niveles inferiores. Esta situación no es inherente a algún método en particular, más bien, las soluciones múltiples ocurren en ciertos conjuntos de datos. En esta circunstancia, se deberán comparar e interpretar las distintas soluciones obtenidas.

Algunos conjuntos de datos y métodos jerárquicos producen inversiones. Una inversión ocurre cuando un objeto se une a un conglomerado existente a una distancia inferior (mayor similitud) que a la que ocurrió la consolidación previa.
Las inversiones pueden ocurrir cuando no existe una estructura clara de agrupamiento y generalmente se encuentran asociadas con dos algoritmos de asociación jerárquica conocidos como el método del centroide y el método de la mediana. Los métodos jerárquicos anteriormente descritos, en general no tienden a producir inversiones.

MÉTODOS NO JERÁRQUICOS DE AGRUPACIÓN

Las técnicas no jerárquicas de agrupamiento están diseñadas para agrupar *individuos* en un conjunto de K conglomerados. El número de K conglomerados puede especificarse anticipadamente o bien como parte del procedimiento de formación de conglomerados.

Debido a que no se requiere determinar la matriz de distancias (similitudes) y los datos básicos no requieren ser almacenados durante el proceso de cálculo, los métodos no jerárquicos pueden aplicarse al caso de bases de datos mucho mayores que los procedimientos jerárquicos.

Los métodos no jerárquicos pueden iniciar (1) mediante una partición inicial de los individuos en grupos o (2) un conjunto inicial de puntos que representarán los núcleos de los conglomerados resultantes. Una buena elección del punto de inicio deberá estar libre de sesgos evidentes. Una forma de iniciar el proceso, es seleccionando aleatoriamente los individuos iniciales dentro del conjunto de objetos o bien partir aleatoriamente los individuos en grupos iniciales[30].

A continuación se discute el método no jerárquico mas comúnmente empleado, conocido como el procedimiento de K-medias.

Método de K-Medias

MacQueen [26] sugirió el término de K-medias par describir un algoritmo desarrollado por él mismo que asigna cada individuo al conglomerado con el centroide (media) más cercano. De acuerdo a esta versión simple, el proceso se compone de tres pasos:

1. Partir los individuos en K conglomerados iniciales
2. Proceder a través de la lista de individuos asignando el individuo al conglomerado cuyo centroide (media) sea mas cercana. (La distancia generalmente se calcula usando distancia euclidianas con observaciones ya sea estandarizadas o no estandarizadas). Recalcular el centroide tanto para el conglomerado receptor del nuevo individuo, como para el conglomerado que cede dicho individuo.
3. Repetir el paso 2, hasta que no haya más asignaciones

En lugar de iniciar con una partición de todos los individuos en K grupos preliminares en el paso 1, se pueden especificar K centroides iniciales y de allí proceder al paso 2.
La asignación final de individuos a los conglomerados dependerá hasta cierto punto de la partición inicial o de la selección inicial de puntos. La experiencia sugiere que los mayores cambios en la asignación ocurren en los primeros pasos.

Para verificar la estabilidad del agrupamiento, es conveniente correr nuevamente el algoritmo utilizando una partición alternativa de individuos o bien otros centroides iniciales. Una vez que se han determinado los conglomerados, su interpretación se facilita ordenando la lista de individuos, de tal manera que aquellos que aparecen en el primer conglomerado se muestran primero, los correspondientes al segundo conglomerado, en seguida y así sucesivamente. Una tabla especificando los centroides (medias) y las varianzas en centroides también puede ser de utilidad para delinear las diferencias entre los conglomerados.

Es necesario señalar que la importancia de cada una de las variables en la formación de conglomerados debe juzgarse bajo la perspectiva multivariada. La totalidad de las

[30] Johnson, Richard A., Wichern, Dean W. 2002. Applied Multivariate Statistical Analysis. New Jersey: Prentice Hall

variables (observaciones multivariadas) determina las medias de los conglomerados y la reasignación de individuos. Adicionalmente los valores de los estadísticos descriptivos que miden la importancia de las variables individuales, son funciones del número de conglomerados y de la configuración final de dichos conglomerados.

Finalmente, es conveniente señalar que existen argumentos importantes en contra de la fijación del número de conglomerados K de manera anticipada, incluyendo los siguientes:

1. Si dos o más puntos iniciales inadvertidamente corresponden al mismo conglomerado, sus conglomerados resultantes no estarán claramente diferenciados.
2. La existencia de observaciones atípicas podrá producir al menos un grupo con individuos muy dispersos entre sí.
3. Aún cuando se sabe que la población consiste de K grupos, el método de muestreo empleado puede ser de tal manera que los datos de los individuos menos comunes no aparezcan en la muestra. En este caso, el forzar los datos en K grupos conduciría a la existencia de conglomerados con poco sentido.

En caso de que el algoritmo de cálculo requiera al usuario la especificación de K, es conveniente repetir el cálculo, empleando varias opciones.

Bibliografía

1. Richard A. Johnson, Dean W. Wichern 2002. *Applied Multivariate Statistical Analysis, 5th Edition.* Prentice Hall.
2. Dallas E. Johnson 2000. *Métodos Multivariados Aplicados al Análisis de Datos.* Internacional Thomson Editores.
3. David C. Lay 2001. *Algebra Lineal y sus Aplicaciones, 2ª Edición.* Prentice Hall,
4. Programa de las Naciones Unidas para el Desarrollo (PNUD) 2002 Informe Sobre Desarrollo Humano México.
5. Sen Amartya 1999. *Development as Freedom*, New York: Alfred A. Knopf,.
6. Sen Amartya 1982. *Choice, Welfare and Measurement.* Harvard University Press.
7. Méndez Ramírez Ignacio, Namihira Guerrero Delia, Moreno Altamirano Laura, Sosa de Martínez Cristina 1990. *El Protocolo de la Investigación, Lineamientos para su Elaboración y Análisis, 2ª Edición.* Trillas.
8. Ackoff, Russell L. Gupta, Shiv K. J., Minas Sayer 1962. *Scientific Method.* John Wiley & Sons Inc.
9. Ackoff Russell L. 1974. *Redesigning the Future.* New York: John Wiley & Sons.
10. Ackoff, Rusell L. 1978. *The Art of Problem Solving.* New York: John Wiley & Sons
11. Ackoff Russell L. 1981. *Creating the Corporate Future.* New York: John Wiley & Sons.
12. Gharajedaghi Jamshid, Ackoff, Russell L. (1986). *A Prologue to National Development Planning.* Greenwood Press.
13. Garajedaghi Jamshid. 1999. *Systems Thinking, Managing Chaos and Complexity, A Platform for Designing Business Architecture.* Butterworth Heinemann
14. INEGI (Instituto Nacional de Estadística, Geografía e Informática) 2000. Tabulados de la Muestra Censal del XII Censo General de Población.
15. Chatfield C., Collins A.J. 1980. *Introduction to Multivariate Analysis.* Chapman & Hall.
16. Everitt B.S., Landau S. Leese M. 2001. *Cluster Analysis.* Arnold
17. De la Torre, Rodolfo. 1997. "Indicadores de desarrollo regional con información limitada". En Gabriel Martínez, ed., *Pobreza y política social en México.* Lecturas del Trimestre Económico, 85. México: Fondo de Cultura Económica.
18. Jarque, Carlos M., y Medina Fernando. 1998. *Índices de desarrollo humano en México 1960-1990.* Santiago de Chile: CEPAL (Comisión Económica para América Latina).
19. García-Verdú, Rodrigo. 2002. "The Human Development Index and its Application to Status in México". Dirección de Estudios Económicos. México: Banco de México.
20. Bunge, Mario.1999. *Sistemas Sociales y Filosofía.* Buenos Aires: Editorial Sudamericana

21. Rencher, A.C. Interpretation of Canonical Discriminant Functions, Canonical Variates and Principal Components. *The American Statistician*, **46**, (1992). 217-225
22. Kaiser, H.F. The Varimax Criterion for Analytic Rotation in Factor Analysis. *Psychometrica*, **23**, (1958), 187-200
23. Gower, J.C. "Some Distance Properties of Latent Roots and Vector Methods Used in Multivarite Analysis." *Biometrika* **53** (1966), 325-338
24. Gower, J.C., and D.J. Hand. Biplots. London: Chapman and Hall, 1996
25. Ward, Jr., J. H. "Hierarchical Grouping to Optimize an Objective Function". *Journal of the American Statistical Association*, **58** (1963), 236-244
26. MacQueen, J.B. "Some Methods for Classification and Analysis of Multivariate Observations". *Proceedings of 5th Berkely Simposium on Mathematical Statistics and Probability*, **1**, Berkely, CA: University of California Press (1967), 281-297
27. Pérez Trejo, et. al. (2004) "Análisis Multivariado del Desarrollo Integral en México". Ponencia Presentada en el XIV Encuentro de Estadísticas Cuba-México. Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (I.I.M.A.S.) e Instituto de Cibernética, Matemática y Física de la Habana (ICIMAF). La Habana, Cuba (23-27 Feb.)

Anexos

Anexo 1: Base de datos

Anexo 2: Descripción de las variables derivadas

Anexo 3: Matrices de residuos

Anexo 1
Base de Datos

	NUM	POBA	POBT	PDISC	SINSTRUCCION	EDUBAS	EDUSUP	PEAOC	FECUNDI	OCUPANTES	MIGRANTES	MORIN	EDAD	INMASC	SSS	VIVIENDA	SERBAS	IPAD1	IPAD2	PARELEC	PIBCAP	DEL	RAGEN1	P (para q grande)
001 ACAMBAY	1	26642	58389	0.019901009	0.103786672	0.561218723	0.05889808	0.221514326	4.2394	4.734655755	0.013718337	0.15	18	93.32	0.077428968	3.52	0.07	1.85	0.07	0.65	1,478	0.63	1.20	0.5424
002 ACOLMAN	2	35728	61250	0.024995918	0.028653061	0.581044898	0.184163265	0.336244898	2.6717	4.381510318	0.035673469	0.10	23	95.95	0.392457143	3.71	0.73	-1.67	0.42	0.70	4,546	1.04	1.05	0.0999
003 ACULCO	3	18804	38827	0.015813738	0.091611507	0.605892807	0.046359492	0.283024699	4.2387	4.726598099	0.018028691	0.15	18	96.20	0.162361243	3.49	0.08	1.46	0.24	0.73	2,059	0.98	1.16	0.3865
004 ALMOLOYA DE ALQUISIR/	4	7080	15584	0.019314682	0.079183778	0.58226386	0.052040554	0.191029261	4.1583	4.673666667	0.013988706	0.13	18	91.59	0.067825975	3.48	0.14	0.83	0.32	0.67	2,059	1.09	1.05	0.4604
005 ALMOLOYA DE JUÁREZ	5	56996	110591	0.013952311	0.088886076	0.607454494	0.065828142	0.261296127	3.7163	4.865936945	0.012369904	0.13	20	98.53	0.162834227	3.83	0.10	1.39	0.17	0.71	4,168	8.93	1.15	0.2995
006 ALMOLOYA DEL RÍO	6	4821	8873	0.008114505	0.042150344	0.484728953	0.159472557	0.332018483	3.2533	4.718573046	0.01780683	0.09	22	90.53	0.212104136	3.81	0.83	-0.83	-0.11	0.74	3,919	1.35	1.07	0.1659
007 AMANALCO	7	9985	21095	0.014079166	0.123441574	0.611614127	0.032377341	0.217207869	4.3705	4.947965475	0.006304812	0.19	17	98.88	0.043564826	3.03	0.07	2.44	-0.19	0.71	1,208	1.47	1.24	0.5419
008 AMATEPEC	8	14185	30141	0.02564613	0.132775953	0.548289705	0.053747387	0.200026542	3.8245	4.619437939	0.016688232	0.14	18	92.74	0.090109817	3.51	0.07	0.93	0.32	0.74	1,444	0.96	1.09	0.5585
009 AMECAMECA	9	26362	45255	0.016926306	0.034051486	0.589305049	0.176864435	0.325599381	2.8529	4.524533806	0.016837918	0.11	23	93.74	0.328383604	2.86	0.68	-0.88	0.07	0.76	8,274	2.28	1.04	0.1670
010 APAXCO	10	13502	23734	0.022499368	0.049296368	0.606345327	0.140642117	0.339260133	2.6592	4.429122469	0.022583635	0.09	22	97.40	0.383753265	3.72	0.62	-1.03	0.29	0.72	3,762	0.17	1.07	0.1305
011 ATENCO	11	19873	34435	0.016901408	0.028227095	0.595063162	0.161608828	0.334862785	2.7269	4.544049248	0.043008567	0.10	23	98.47	0.34737912	3.86	0.61	-0.83	0.28	0.66	3,528	1.36	1.04	0.1208
012 ATIZAPÁN	12	4412	8172	0.009667156	0.047479197	0.6139256	0.116005874	0.348507097	3.3779	4.94026975	0.012236907	0.12	21	96.30	0.156754772	3.65	0.79	-0.30	0.09	0.63	3,959	1.47	1.07	0.1371
013 ATIZAPÁN DE ZARAGOZA	13	282372	467886	0.013270327	0.025540409	0.505945893	0.276469054	0.372208187	2.3713	4.123051742	0.057757659	0.08	24	95.54	0.503261478	3.69	0.88	-4.43	-1.51	0.69	7,619	1.17	1.04	0.0689
014 ATLACOMULCO	14	38161	76750	0.012742671	0.073980456	0.547921824	0.1432443	0.296534202	3.5309	4.772058334	0.023661238	0.14	19	91.88	0.260078176	3.64	0.32	0.12	-0.19	0.68	3,975	0.90	1.17	0.2380
015 ATLAUTLA	15	13480	25950	0.017109827	0.067013487	0.596184971	0.090134875	0.288208092	3.1349	4.954657873	0.016377649	0.14	21	94.98	0.14894027	2.98	0.39	1.12	0.28	0.64	2,146	0.46	1.07	0.3642
016 AXAPUSCO	16	11074	20516	0.015256385	0.060050692	0.616738156	0.094024176	0.29625658	3.0623	4.394128525	0.028758043	0.13	22	99.30	0.225190096	3.41	0.42	-0.16	0.80	0.73	2,707	0.29	1.06	0.2417
017 AYAPANGO	17	3248	5947	0.016815201	0.033125946	0.616445267	0.123759879	0.330418699	2.7941	4.33515625	0.038674962	0.10	22	98.90	0.249537582	2.86	0.68	-0.62	0.31	0.77	3,038	2.02	1.02	0.1712
018 CALIMAYA	18	19870	35196	0.013325378	0.058245255	0.55315945	0.115211956	0.311200136	2.9408	4.71509774	0.009773838	0.12	22	96.53	0.350778498	3.92	0.69	-0.35	0.17	0.68	3,198	1.19	1.04	0.1780
019 CAPULHUAC	19	16281	28808	0.0107609	0.034053041	0.573590669	0.18918356	0.350701194	2.8731	4.698499318	0.011420439	0.09	22	96.74	0.201645376	3.78	0.79	-1.09	0.18	0.59	6,433	0.80	1.05	0.1667
025 CHALCO	20	113187	217972	0.013423742	0.039679408	0.599925679	0.120923788	0.312462151	2.8419	4.479098526	0.095971042	0.09	21	97.31	0.321422018	3.21	0.66	-0.61	0.25	0.60	3,612	5.02	1.05	0.1050
026 CHAPA DE MOTA	21	11337	22828	0.009681093	0.104170317	0.610697389	0.041571754	0.243604346	4.0700	4.833482342	0.006921325	0.15	19	97.76	0.10920799	3.74	0.08	1.65	0.20	0.77	3,425	1.49	1.23	0.3542
027 CHAPULTEPEC	22	3235	5735	0.014646905	0.030340017	0.522057541	0.161987794	0.34646905	3.4004	4.663511188	0.024934612	0.11	22	91.68	0.46190061	3.69	0.73	-0.56	0.10	0.73	3,809	0.87	1.04	0.1469
028 CHIAUTLA	23	11370	19620	0.016666667	0.027420999	0.593017329	0.188124363	0.355708461	2.8898	4.654063344	0.031141692	0.09	23	94.76	0.252395515	3.78	0.61	-1.16	0.27	0.77	4,171	1.07	1.03	0.1731
029 CHICOLOAPAN	24	44213	77579	0.015738795	0.036478944	0.596965674	0.158883203	0.353446165	2.7328	4.353797579	0.06433442	0.09	22	96.51	0.351486871	3.14	0.78	-1.80	0.34	0.77	4,381	1.04	1.05	0.0905
030 CHICONCUAC	25	10370	17972	0.009681727	0.024927665	0.572835522	0.190629869	0.348931671	2.6610	5.034472599	0.013910527	0.09	23	94.14	0.143612286	4.25	0.76	-0.70	-0.16	0.81	3,789	1.89	1.04	0.1983
031 CHIMALHUACÁN	26	259633	490772	0.013841458	0.038736929	0.631117912	0.106301908	0.332103298	3.0274	4.543889231	0.066862005	0.10	20	98.40	0.297612741	3.09	0.61	-0.54	0.81	0.51	4,462	1.24	1.06	0.0868
020 COACALCO DE BERRIOZ/	27	148360	252555	0.013177328	0.0095108	0.451858011	0.327025796	0.359070302	2.0380	4.068992751	0.09753915	0.06	25	94.79	0.581924729	4.04	0.95	-5.06	-1.38	0.71	6,238	0.81	1.03	0.0602
021 COATEPEC HARINAS	28	16053	35068	0.019961218	0.095642751	0.59478727	0.040806433	0.275407779	4.1024	4.791562177	0.017565872	0.14	17	91.82	0.059798107	3.35	0.20	0.75	0.30	0.61	2,248	0.88	1.04	0.2871
022 COCOTITLÁN	29	6133	10205	0.016364527	0.028417442	0.579715826	0.216266536	0.351494366	2.3961	4.244740232	0.01910828	0.11	24	97.50	0.524154826	3.66	0.71	-1.05	0.22	0.83	4,380	1.67	1.05	0.1104
023 COYOTEPEC	30	19544	35358	0.009700775	0.048956389	0.598902653	0.140477403	0.315798405	3.0279	4.775900277	0.030375021	0.12	21	98.60	0.453362747	3.62	0.65	-0.26	0.14	0.68	3,629	1.22	1.07	0.0937
024 CUAUTITLÁN	31	43162	75836	0.011419379	0.019528984	0.487156496	0.258373332	0.351772245	2.3875	4.126159	0.079012606	0.08	24	96.58	0.538846986	3.60	0.87	-3.62	-0.73	0.63	6,063	10.39	1.04	0.0774
121 CUAUTITLÁN IZCALLI	32	273711	453298	0.013337804	0.016702037	0.487908616	0.313237649	0.365920432	2.1013	4.154472803	0.059942907	0.07	24	95.73	0.592958716	3.89	0.89	-4.94	-1.64	0.72	6,816	2.29	1.03	0.0649
032 DONATO GUERRA	33	12718	28006	0.013997001	0.142576591	0.564986074	0.028636721	0.222702278	5.0976	5.319157393	0.00710562	0.16	17	98.88	0.045026066	2.82	0.03	2.84	-0.29	0.73	1,772	2.36	1.31	0.4195
033 ECATEPEC DE MORELOS	34	957036	1622697	0.014422902	0.025069375	0.559689825	0.214131166	0.357921411	2.4920	4.275037567	0.070385291	0.08	23	96.74	0.451754086	3.45	0.80	-3.04	-0.29	0.67	4,621	1.88	1.04	0.0911
034 ECATZINGO	35	3908	7916	0.017306721	0.063668519	0.643127842	0.057099545	0.293329965	4.7478	5.149558124	0.01061142	0.17	18	100.35	0.133779687	2.88	0.17	1.82	0.15	0.70	1,810	1.14	1.09	0.4924
035 HUEHUETOCA	36	21354	38458	0.013313225	0.034167143	0.619741016	0.141401009	0.335404857	2.9191	4.525173502	0.077097093	0.10	21	103.92	0.471397369	3.84	0.72	-1.30	0.26	0.85	4,315	1.09	1.05	0.0664
036 HUEYPOXTLA	37	18473	33343	0.012716312	0.077857421	0.667336472	0.062261944	0.311399694	2.9344	4.855183846	0.016375251	0.13	21	100.54	0.224214978	4.12	0.32	0.49	0.65	0.70	3,715	0.15	1.06	0.1728
037 HUIXQUILUCAN	38	114756	193468	0.009924122	0.029519094	0.524712097	0.236829863	0.375860607	2.3702	4.390731789	0.076126284	0.08	23	90.60	0.340376703	3.63	0.73	-3.42	-1.64	0.70	10,066	1.94	1.03	0.0500
038 ISIDRO FABELA	39	4407	8168	0.010039177	0.059255632	0.612512243	0.092923604	0.317091087	3.0396	4.531100478	0.018364349	0.11	22	99.51	0.243143976	3.41	0.26	0.19	0.51	0.71	3,131	0.98	1.02	0.2055
039 IXTAPALUCA	40	153721	297570	0.011681285	0.02730786	0.542796653	0.167526968	0.323207985	2.7857	4.25862834	0.160516853	0.09	22	97.26	0.435786538	3.58	0.74	-2.07	0.12	0.54	4,598	1.07	1.04	0.0752
040 IXTAPAN DE LA SAL	41	15046	30529	0.017262275	0.068918078	0.567853516	0.094139998	0.302990599	3.6071	4.463650794	0.019784467	0.13	19	90.20	0.200497887	3.17	0.35	-0.54	-0.09	0.69	3,346	1.38	1.05	0.1979
041 IXTAPAN DEL ORO	42	2953	6425	0.013229572	0.107859922	0.598599222	0.033929961	0.229105058	5.0483	4.707046714	0.013540856	0.17	17	101.41	0.062879377	2.95	0.10	2.21	-0.01	0.87	1,108	1.40	1.12	0.7052
042 IXTLAHUACA	43	58981	115165	0.01242565	0.099092606	0.607250467	0.074953328	0.260999436	3.5891	5.049498206	0.010272218	0.14	19	92.46	0.201354578	3.81	0.09	1.30	0.24	0.70	2,464	2.74	1.21	0.2904
044 JALTENCO	44	18172	31629	0.013848051	0.018400835	0.54377312	0.22454077	0.355370072	2.7520	4.409887173	0.084953682	0.09	23	97.05	0.452274811	4.03	0.76	-2.85	-0.45	0.65	5,041	0.22	1.03	0.0649
045 JILOTEPEC	45	35878	68336	0.01571646	0.066436432	0.621414774	0.087977055	0.290871576	3.4031	4.613243147	0.019799227	0.14	20	98.05	0.247629361	3.67	0.16	0.21	0.03	0.75	3,732	4.61	1.10	0.2755
046 JILOTZINGO	46	8169	15086	0.013853904	0.059459101	0.613151266	0.100225375	0.340779531	2.8989	4.539215686	0.022338592	0.11	22	98.14	0.328185072	3.53	0.48	-0.27	0.18	0.78	4,384	0.53	1.07	0.1081
047 JIQUIPILCO	47	27309	56614	0.016268061	0.104673756	0.582806373	0.058254142	0.208711626	4.1703	4.742740068	0.014766665	0.15	18	92.59	0.100770128	3.72	0.11	1.84	0.13	0.73	2,112	1.45	1.19	0.4512
048 JOCOTITLÁN	48	27951	51979	0.012928298	0.062294388	0.600473268	0.121222032	0.312472345	3.0245	4.661786929	0.013851748	0.13	21	95.29	0.319917659	3.92	0.29	0.23	0.13	0.72	2,999	0.94	1.09	0.2276
049 JOQUICINGO	49	5668	10720	0.014458955	0.068190299	0.644123134	0.077145522	0.268190299	3.2096	4.770174611	0.019869403	0.16	21	93.22	0.157835821	3.92	0.53	0.86	0.35	0.72	2,165	0.47	1.07	0.2690
050 JUCHITEPEC	50	10597	18968	0.013443695	0.064529734	0.619042598	0.113876002	0.316216786	3.1018	4.537317134	0.010385913	0.13	21	96.97	0.15489245	3.41	0.50	-0.17	0.17	0.72	2,813	0.74	1.04	0.2084
051 LERMA	51	55889	99870	0.010683889	0.048993692	0.576078903	0.160278362	0.319665565	2.9589	4.683921413	0.031641133	0.11	22	97.92	0.396225093	4.04	0.67	-0.68	0.07	0.69	4,434	2.84	1.08	0.1029
052 MALINALCO	52	10743	21712	0.015475313	0.069443626	0.605287399	0.053058217	0.258290346	4.1316	4.787386527	0.013817244	0.13	20	95.90	0.094417833	3.33	0.25	1.13	0.27	0.72	1,864	2.76	1.05	0.3715
053 MELCHOR OCAMPO	53	21708	37716	0.014131933	0.029987273	0.599082617	0.170537703	0.345794888	2.7039	4.614862316	0.027150281	0.09	22	96.42	0.448085693	3.57	0.70	-1.45	0.10	0.69	4,341	1.99	1.04	0.0763
054 METEPEC	54	117541	194463	0.011683456	0.018656505	0.439091241	0.345088783	0.373685483	2.1467	4.254357872	0.046646406	0.08	24	93.59	0.561572124	3.72	0.86	-4.47	-1.69	0.71	9,816	0.87	1.03	0.0649
055 MEXICALTZINGO	55	5302	9225	0.013333333	0.038590786	0.555447154	0.15403794	0.362384824	2.9416	4.783378016	0.011490515	0.11	22	95.36	0.226558266	3.92	0.70	-0.56	-0.00	0.67	4,419	1.95	1.08	0.1615
056 MORELOS	56	12480	26971	0.012124133	0.134440696	0.54121093	0.038115012	0.195098439	4.7309	4.596949476	0.009195061	0.18	19	91.66	0.06955619	3.53	0.05	2.11	-0.06	0.74	1,252	0.93	1.29	0.5983
057 NAUCALPAN DE JUÁREZ	57	541016	858711	0.014999226	0.028724449	0.52439645	0.262018304	0.386694709	2.4663	4.058938579	0.059038489	0.08	24	93.98	0.510455788	3.21	0.79	-4.36	-1.56	0.72	8,211	2.28	1.04	0.0620
059 NEXTLALPAN	58	10287	19532	0.015871391	0.028926889	0.609359001	0.125332787	0.337087856	3.1817	4.574252533	0.055805857	0.10	20	97.05	0.217233258	3.44	0.57	-0.51	0.19	0.64	6,471	0.72	1.06	0.1097
058 NEZAHUALCÓYOTL	59	770392	1225972	0.016923714	0.026401908	0.554439253	0.23856907	0.383848897	2.2888	4.211326229	0.052689621	0.09	25	94.48	0.431165638	3.19	0.83	-3.54	-0.19	0.75	6,151	2.78	1.04	0.0732
060 NICOLÁS ROMERO	60	154238	269546	0.01343741	0.038542586	0.606178537	0.147781826	0.340220222	2.6156	4.324299425	0.032807016	0.09	22	97.86	0.475521803	3.57	0.69	-1.85	0.17	0.68	4,389	1.10	1.05	0.0691
061 NOPALTEPEC	61	4242	7512	0.017305644	0.057374867	0.58972311	0.123801917	0.327742279	3.0869	4.172678888	0.034078807	0.13	23	101.94	0.226837061	3.48	0.50	-1.02	0.46	0.76	3,166	0.00	1.04	0.2516
062 OCOYOACAC	62	29170	49643	0.011119392	0.04052938	0.576395464	0.176198054	0.344540016	2.7854	4.744593921	0.019197067	0.11	23	96.35	0.372439216	4.18	0.73	-0.78	0.08	0.70	4,522	1.17	1.07	0.0935
063 OCUILAN	63	12894	25969	0.012736158	0.07333872	0.638154604	0.051714187	0.271422525	4.1333	5.012279984	0.011735734	0.16	19	96.74	0.100696448	3.64	0.19	1.48	0.20	0.60	1,811	2.65	1.09	0.5022
064 ORO, EL	64	13945	30411	0.01483016	0.084541778	0.566998783	0.078557101	0.225378975	4.1980	4.684351018	0.021538259	0.16	17	92.16	0.161783565	3.36	0.15	0.63	-0.08	0.81	2,119	3.32	1.20	0.3277
065 OTUMBA	65	15990	29097	0.014743788	0.053785614	0.621266797	0.106643297	0.315874489	2.9051	4.555407875	0.031893322	0.11	22	98.37	0.193318899	3.50	0.44	-0.45	0.23	0.68	2,936	3.85	1.05	0.2703
066 OTZOLOAPAN	66	2316	5196	0.017898383	0.123941493	0.586797537	0.031562741	0.254811393	4.2090	5.034126163	0.015973826	0.12	16	100.85	0.065627406	2.95	0.16	2.05	0.17	0.94	4,309	2.12	1.14	0.5061
067 OTZOLOTEPEC	67	30305	57583	0.009256204	0.08278485	0.605942726	0.092371012	0.292395325	3.7064	5.136145468	0.011010194	0.13	20	95.86	0.238872584	3.85	0.37	0.92	0.24	0.65	3,052	0.71	1.13	0.1564
068 OZUMBA	68	12813	23592	0.015852831	0.03683452	0.597151577	0.14242116	0.327229569	2.9617	4.816609735	0.016276704	0.11	21	93.16	0.192692438	3.15	0.46	-0.00	0.07	0.70	3,098	1.02	1.07	0.2533
069 PAPALOTLA	69	2019	3469	0.019313923	0.027385414	0.541654656	0.228596137	0.350821562	2.7038	4.536986301	0.021043528	0.06	23	96.21	0.279042952	3.64	0.77	-1.77	0.42	0.94	4,739	0.00	1.03	0.0996
070 PAZ, LA	70	120468	212694	0.015289571	0.030541529	0.581619604	0.168556706	0.353550171	2.7686	4.322317205	0.099184744	0.09	22	97.03	0.384510141	3.18	0.66	-2.00	0.31	0.57	6,031	1.57	1.05	0.0676
071 POLOTITLÁN	71	6131	11065	0.021690014	0.053953909	0.637596023	0.10456394	0.341256213	3.3542	4.335378323	0.028468143	0.10	22	97.38	0.344871216	3.51	0.41	-1.11	0.61	0.90	3,610	2.44	1.03	0.1651

072 RAYÓN	72	5348	9024	0.010970745	0.049534574	0.524379433	0.14616578	0.336546986	2.9385	4.615062762	0.012965426	0.11	23	97.98	0.496675532	3.94	0.78	-0.31	0.03	0.73	7,835	1.44	1.07	0.1450
073 SAN ANTONIO LA ISLA	73	5916	10321	0.01182056	0.048541808	0.547136905	0.139612034	0.323999612	3.0339	4.825728155	0.009785873	0.12	22	97.27	0.387462455	3.72	0.79	0.22	-0.10	0.77	3,575	0.29	1.03	0.1192
074 SAN FELIPE DEL PROGRE	74	78531	177287	0.010908865	0.137201261	0.577464789	0.029330972	0.207144348	5.6967	5.554563107	0.004709877	0.19	16	94.37	0.033465511	3.67	0.02	2.74	-0.31	0.60	2,703	0.47	1.38	0.4150
075 SAN MARTÍN DE LAS PIR/	75	11135	19694	0.016248604	0.03539149	0.6069869	0.133746319	0.342439322	2.8585	4.339306493	0.037778003	0.10	23	99.37	0.230780949	3.52	0.65	-1.17	0.34	0.71	4,316	0.46	1.03	0.2224
076 SAN MATEO ATENCO	76	33170	59647	0.009120325	0.036431002	0.56829346	0.159806864	0.344241957	3.0315	4.716267547	0.027746576	0.09	21	96.71	0.349606854	3.96	0.49	-1.02	0.13	0.72	4,353	1.53	1.05	0.1002
077 SAN SIMÓN DE GUERREF	77	2511	5436	0.033296542	0.101545254	0.549300957	0.068432671	0.215599706	4.2598	4.708373436	0.01986755	0.15	18	93.80	0.15397351	3.24	0.16	1.61	0.21	0.71	1,742	2.39	1.11	0.4477
078 SANTO TOMÁS	78	4120	8592	0.020833333	0.099045624	0.600442272	0.059590317	0.217760708	4.0015	4.840649429	0.009660149	0.15	18	97.65	0.161060074	2.90	0.25	0.97	0.35	0.85	1,901	1.51	1.10	0.3069
079 SOYANIQUILPAN DE JUÁI	79	5458	10007	0.01948636	0.050264815	0.656840212	0.076946138	0.299190567	2.7703	4.231117142	0.025082442	0.12	22	99.50	0.298191266	3.81	0.18	-0.42	0.66	0.78	2,740	2.50	1.05	0.2591
080 SULTEPEC	80	12888	27592	0.016019136	0.14609307	0.562155697	0.038851841	0.154066396	4.5479	5.08329988	0.006632357	0.15	18	94.47	0.080603073	3.40	0.02	2.48	-0.32	0.63	1,215	4.57	1.20	0.5430
081 TECÁMAC	81	98554	172813	0.014998872	0.028464294	0.562017904	0.194233073	0.340992865	2.6061	4.310655867	0.062153889	0.09	23	96.39	0.388118949	3.43	0.81	-2.33	-0.18	0.68	4,915	0.87	1.04	0.0747
082 TEJUPILCO	82	43660	95032	0.017415186	0.126220641	0.516184022	0.066914303	0.185158683	4.4379	4.957462351	0.017583551	0.13	17	94.79	0.116129304	3.39	0.18	1.00	0.14	0.64	2,326	2.94	1.12	0.3122
083 TEMAMATLA	83	4749	8840	0.013687783	0.027375566	0.578280543	0.155316742	0.331561086	3.0109	4.191251272	0.111425339	0.09	22	98.30	0.434841629	2.59	0.69	-1.38	0.44	0.72	4,871	0.68	1.05	0.0982
084 TEMASCALAPA	84	15875	29307	0.016787798	0.047463063	0.635923158	0.093424779	0.314975944	3.1335	4.42923175	0.038557341	0.12	21	99.67	0.281912171	3.70	0.45	-0.20	0.89	0.70	3,091	0.55	1.06	0.1669
085 TEMASCALCINGO	85	29208	61974	0.013263627	0.106076742	0.56978733	0.066043825	0.216703779	4.0643	5.041339797	0.012634331	0.14	18	93.94	0.095540065	3.80	0.13	1.54	0.11	0.67	1,925	0.74	1.22	0.3772
086 TEMASCALTEPEC	86	14764	31192	0.019203642	0.089766607	0.619133111	0.039914081	0.222140292	4.5790	4.848479814	0.007694281	0.15	18	98.32	0.102911003	3.09	0.07	1.92	0.01	0.64	2,002	10.87	1.12	0.6250
087 TEMOAYA	87	34037	69306	0.00992699	0.096470724	0.63625083	0.045739186	0.262747814	4.0091	5.501352792	0.007820391	0.15	18	94.79	0.079993074	3.72	0.17	2.11	-0.06	0.62	2,125	0.48	1.20	0.3351
088 TENANCINGO	88	41411	77531	0.013968606	0.061755943	0.595232875	0.127716655	0.324051025	3.2330	4.749596018	0.021320504	0.12	20	94.95	0.242922186	3.22	0.45	-0.13	-0.13	0.59	3,647	4.79	1.05	0.1770
089 TENANGO DEL AIRE	89	4887	8486	0.018501061	0.03582371	0.611124205	0.145651662	0.347277869	2.5439	4.396504642	0.032052793	0.12	23	96.48	0.283407966	3.40	0.74	-0.77	0.35	0.79	3,378	1.77	1.02	0.1724
090 TENANGO DEL VALLE	90	34959	65119	0.014634746	0.073388719	0.601836638	0.106266988	0.304795835	3.1634	4.717341482	0.011947358	0.13	21	94.77	0.247838572	3.67	0.58	0.05	-0.05	0.55	3,084	3.87	1.06	0.2302
091 TEOLOYUCÁN	91	37057	66556	0.01398822	0.036014785	0.591366669	0.156079091	0.334605445	3.0184	4.506436554	0.038719274	0.11	22	99.22	0.412359517	3.70	0.66	-1.22	0.38	0.66	4,234	0.75	1.05	0.0765
092 TEOTIHUACÁN	92	25435	44653	0.014265559	0.03309968	0.59088975	0.168029024	0.342059884	2.7726	4.4216205	0.046939735	0.09	22	98.65	0.352652677	3.53	0.66	-1.53	0.43	0.75	4,321	0.94	1.04	0.1127
093 TEPETLAOXTOC	93	12657	22729	0.013770953	0.042456773	0.594614809	0.141757226	0.328347046	2.8591	4.564118896	0.02657398	0.11	22	98.44	0.260064235	3.69	0.54	-0.75	0.28	0.65	3,419	0.79	1.04	0.1784
094 TEPETLIXPA	94	9261	16863	0.012393999	0.044594675	0.632271838	0.1037182	0.276285358	2.6832	4.805471125	0.024135682	0.13	22	97.71	0.162130107	3.43	0.40	0.58	0.48	0.66	2,828	1.30	1.05	0.3759
095 TEPOTZOTLÁN	95	35271	62280	0.016457932	0.035019268	0.566490045	0.187973667	0.356374438	2.5875	4.32269374	0.043208092	0.10	22	96.66	0.505266536	3.59	0.66	-2.54	-0.20	0.73	5,345	1.11	1.04	0.0872
096 TEQUIXQUIAC	96	15890	28067	0.019702854	0.035949692	0.651761856	0.121744397	0.340827306	2.7423	4.485102881	0.020557951	0.10	22	98.69	0.287169986	3.96	0.56	-0.63	0.29	0.63	3,767	0.29	1.04	0.1464
097 TEXCALTITLÁN	97	7489	16370	0.014722053	0.089920586	0.596334759	0.060720831	0.183445327	4.3772	4.925625203	0.013194869	0.14	17	93.57	0.088271228	3.47	0.12	1.61	0.10	0.76	2,733	2.50	1.08	0.4291
098 TEXCALYACAC	98	2181	3997	0.015511634	0.024518389	0.586690018	0.158118589	0.327745809	2.9826	4.455864571	0.01701276	0.11	23	96.80	0.276207155	3.64	0.81	-0.74	0.08	0.81	8,739	0.50	1.06	0.2295
099 TEXCOCO	99	120427	204102	0.013978305	0.029024703	0.525712634	0.247807469	0.341309737	2.4510	4.343951678	0.05741247	0.10	23	99.19	0.404155765	3.54	0.69	-2.32	-0.31	0.63	5,706	5.16	1.04	0.1154
100 TEZOYUCA	100	10892	18852	0.018300446	0.026787609	0.580468916	0.180352217	0.335349035	2.5317	4.433333333	0.051612561	0.09	23	98.90	0.363091449	3.65	0.69	-1.52	0.47	0.71	4,125	1.64	1.04	0.1029
101 TIANGUISTENCO	101	31671	58381	0.011630496	0.043387403	0.59070588	0.141518645	0.325790925	3.1845	4.925549009	0.019338483	0.13	21	95.46	0.246689648	3.70	0.62	-0.09	-0.16	0.65	6,923	1.22	1.06	0.1881
102 TIMILPAN	102	7371	14512	0.021775083	0.08303473	0.589649945	0.074007718	0.274255788	3.6093	4.307289964	0.018122933	0.15	21	95.55	0.160281147	4.00	0.18	0.40	0.54	0.81	2,285	1.24	1.13	0.4038
103 TLALMANALCO	103	25126	42507	0.016256146	0.022819771	0.559625473	0.216576093	0.335333004	2.4715	4.329293146	0.02583104	0.09	24	95.42	0.543157598	3.29	0.78	-1.90	0.14	0.73	4,572	1.11	1.03	0.0682
104 TLALNEPANTLA DE BAZ	104	466334	721415	0.017346465	0.025664839	0.508278869	0.296327357	0.385174969	2.1862	4.05555362	0.060116576	0.08	26	94.32	0.553519126	3.41	0.84	-4.59	-1.11	0.76	6,812	3.91	1.04	0.0726
105 TLATLAYA	105	16579	36100	0.020747922	0.141108033	0.546038781	0.05767313	0.163711911	4.0390	4.751858205	0.020277008	0.13	18	93.88	0.087534626	3.54	0.04	1.38	0.29	0.77	1,613	1.08	1.11	0.5512
106 TOLUCA	106	386246	666596	0.012614837	0.039890728	0.510537117	0.251659176	0.349178513	2.5651	4.401353885	0.028314001	0.11	23	93.16	0.453081027	3.65	0.66	-2.42	-0.92	0.70	14,046	2.13	1.07	0.0769
107 TONATICO	107	6248	11502	0.025039124	0.060337333	0.629455747	0.083985394	0.292731699	2.9925	4.077992278	0.034341854	0.12	23	91.67	0.149626152	3.19	0.53	-1.86	0.25	0.75	4,255	1.04	1.05	0.2222
108 TULTEPEC	108	51324	93277	0.012586168	0.026823333	0.55711483	0.19444236	0.331410744	2.7031	4.36618391	0.065771841	0.09	22	96.94	0.473546533	3.76	0.81	-2.62	-0.26	0.63	4,838	1.09	1.04	0.0847
109 TULTITLÁN	109	242297	432141	0.01408568	0.020449344	0.529477647	0.227819161	0.349038855	2.4700	4.212190256	0.086869795	0.08	23	96.67	0.54187175	3.78	0.87	-3.27	-0.37	0.66	4,855	1.37	1.04	0.0780
110 VALLE DE BRAVO	110	27996	57375	0.013525054	0.065446623	0.541681917	0.11135512	0.294065359	3.2815	4.460431019	0.014187364	0.13	20	96.74	0.198448802	2.76	0.47	-0.07	-0.24	0.77	3,245	6.69	1.06	0.1684
122 VALLE DE CHALCO SOLID	111	174303	323461	0.014604543	0.041430033	0.640735668	0.101693867	0.345645998	2.8587	4.481367173	0.094601822	0.09	20	99.02	0.313707062	3.07	0.80	-0.73	0.84	0.58	4,890	1.19	1.06	0.0856
111 VILLA DE ALLENDE	112	18885	40164	0.012573449	0.120182253	0.605567175	0.02380241	0.269171397	4.8974	5.342637151	0.005875909	0.17	17	97.61	0.062618265	3.38	0.03	2.93	-0.34	0.64	1,849	1.97	1.30	0.4839
112 VILLA DEL CARBÓN	113	18384	37993	0.016292475	0.102676809	0.583054773	0.05485221	0.268286263	4.3311	4.933444956	0.013081357	0.14	19	100.28	0.11715316	3.48	0.12	1.59	0.17	0.74	2,113	1.74	1.15	0.4066
113 VILLA GUERRERO	114	24774	50829	0.017037518	0.069684629	0.641130064	0.047984418	0.316276142	4.2921	4.652721677	0.00924669	0.13	18	95.11	0.074091562	3.04	0.26	0.66	0.35	0.63	3,383	1.26	1.02	0.1984
114 VILLA VICTORIA	115	34019	74043	0.010412868	0.137730778	0.570425293	0.024728874	0.241319233	5.4321	5.526863354	0.005753414	0.16	17	98.78	0.071526005	3.81	0.01	2.53	-0.24	0.67	1,909	1.06	1.28	0.3587
043 XALATLACO	116	10091	19182	0.011938275	0.046971119	0.613387551	0.098894797	0.308257742	3.1156	4.960551724	0.014753415	0.14	20	95.65	0.166093212	3.72	0.56	0.67	0.41	0.56	3,635	1.46	1.07	0.2028
115 XONACATLÁN	117	23543	41402	0.010313511	0.060576784	0.597314139	0.137819429	0.324501232	2.8169	4.806300349	0.013646684	0.12	22	97.00	0.220448288	3.91	0.51	0.16	-0.04	0.68	4,263	1.33	1.08	0.1359
116 ZACAZONAPAN	118	1833	3797	0.012378193	0.112983935	0.570713721	0.076112721	0.265209376	3.7234	4.920943135	0.055570187	0.13	19	102.08	0.201474849	3.12	0.44	0.42	0.05	0.81	3,017	1.58	1.07	0.2096
117 ZACUALPAN	119	7598	16101	0.023725234	0.104279237	0.58139246	0.049934787	0.150860195	4.1026	4.750325945	0.015961742	0.13	19	90.95	0.146574747	3.48	0.05	1.48	0.10	0.64	1,342	1.43	1.13	0.3727
118 ZINACANTEPEC	120	66232	121850	0.017513336	0.064095199	0.60679524	0.116290521	0.290406237	3.1997	4.941574936	0.008370948	0.12	21	97.39	0.360467788	4.00	0.44	0.20	-0.07	0.63	3,567	0.91	1.11	0.1160
119 ZUMPAHUACÁN	121	7227	15372	0.015482696	0.118787406	0.615014312	0.030184751	0.241347905	4.3161	5.206214689	0.006830601	0.16	17	90.84	0.06167057	3.11	0.11	2.09	-0.07	0.65	1,423	1.11	1.11	0.4369
120 ZUMPANGO	122	56469	99774	0.013149718	0.044500571	0.588530078	0.161705454	0.332822178	2.9277	4.693470286	0.035881091	0.10	22	97.13	0.332581634	3.71	0.74	-0.92	0.17	0.69	4,167	3.58	1.06	0.1038
001 ABASOLO	123	1481	2514	0.019093079	0.031026253	0.604614161	0.134447096	0.34089101	3.2632	3.953074434	0.033412888	0.08	23	105.06	0.653540175	3.69	0.64	-2.39	1.14	0.72	9,764	1.59	0.98	0.0786
002 AGUALEGUAS	124	2921	4390	0.033712984	0.043507973	0.664464692	0.103189066	0.327790433	2.4536	3.213953488	0.023917995	0.09	30	103.05	0.388610478	3.78	0.60	-4.94	2.29	0.79	15,823	1.37	0.99	0.1898
003 ALDAMAS, LOS	125	1637	2464	0.034902597	0.036931818	0.711850649	0.053977273	0.236607143	2.0603	3.167832168	0.050324675	0.08	31	103.97	0.117694805	3.16	0.55	-5.44	2.11	0.81	3,762	5.28	0.99	0.3828
004 ALLENDE	126	17239	27773	0.019155295	0.027832787	0.640154107	0.150469881	0.348143881	2.9219	3.687185759	0.021063623	0.08	25	102.84	0.618946459	2.91	0.63	-4.89	0.82	0.70	8,988	1.51	1.00	0.0694
005 ANÁHUAC	127	10857	18524	0.017112935	0.038112719	0.622381775	0.140736342	0.342582596	2.8111	3.815911976	0.036007342	0.08	23	101.44	0.406715612	2.93	0.49	-3.23	1.18	0.72	7,770	2.32	0.99	0.1031
006 APODACA	128	162568	283497	0.012056565	0.014582165	0.559109973	0.210033969	0.377446675	3.0011	4.180832559	0.040720008	0.06	22	102.28	0.753242539	3.89	0.94	-3.64	0.68	0.53	9,461	1.13	1.01	0.0485
007 ARAMBERRI	129	8769	14840	0.025808625	0.09245283	0.66819407	0.071226415	0.321293801	3.5634	4.144533486	0.005525606	0.12	24	106.14	0.160107817	3.52	0.10	1.06	0.59	0.67	2,504	1.68	1.04	0.6472
008 BUSTAMANTE	130	2124	3499	0.030008574	0.043726779	0.614747071	0.143755359	0.354672764	2.6960	3.626623377	0.018005144	0.11	26	99.94	0.444412689	3.37	0.64	-3.17	1.71	0.89	5,521	2.86	1.01	0.1880
009 CADEREYTA JIMÉNEZ	131	47837	75059	0.017546197	0.027378462	0.60980029	0.188158649	0.38870755	2.7386	3.801627133	0.119426052	0.08	25	113.22	0.663744521	3.03	0.62	-3.87	0.60	0.57	12,729	4.22	1.00	0.0581
010 CARMEN	132	3964	6644	0.018061409	0.032962071	0.600993377	0.147802529	0.367399157	2.5141	3.968292683	0.043949428	0.10	23	101.95	0.731035521	3.12	0.72	-3.11	1.16	0.71	7,979	1.35	1.01	0.0761
011 CERRALVO	133	5948	9343	0.023654073	0.036819009	0.662956224	0.123300867	0.33725784	2.4821	3.410887397	0.063684042	0.08	27	102.58	0.368511185	3.23	0.67	-4.91	1.62	0.59	8,957	9.63	1.00	0.0901
013 CHINA	134	7067	11540	0.026516464	0.042807626	0.658665511	0.11507799	0.343327556	2.7185	3.605602061	0.032842288	0.07	26	103.78	0.39965338	3.26	0.37	-3.66	1.37	0.79	7,882	6.93	0.99	0.1862
012 CIÉNEGA DE FLORES	135	6411	11204	0.017583006	0.031149589	0.627543734	0.122009996	0.369689397	3.3498	4.033666297	0.056229918	0.09	22	104.86	0.702963227	3.28	0.85	-2.83	0.85	0.59	8,676	0.71	1.01	0.0540
014 DOCTOR ARROYO	136	18731	33721	0.019609614	0.103021856	0.646689007	0.073692951	0.271492542	3.6226	4.362748978	0.012722043	0.12	22	100.21	0.153939682	3.93	0.04	0.43	0.90	0.65	2,264	3.17	1.02	0.6785
015 DOCTOR COSS	137	1471	2246	0.021371327	0.049421193	0.694568121	0.068121104	0.29430098	2.8057	3.28100471	0.029385574	0.09	32	104.74	0.28539626	3.13	0.36	-3.64	2.47	1.04	6,206	4.01	0.99	0.2545
016 DOCTOR GONZÁLEZ	138	1924	3185	0.02166405	0.055259027	0.665620094	0.068759812	0.318367347	2.9864	3.483981693	0.049293564	0.08	26	104.95	0.529984301	2.89	0.47	-3.11	2.02	0.66	5,772	2.51	0.95	0.0899
017 GALEANA	139	23362	39519	0.015460918	0.072522078	0.666767884	0.085857436	0.240896784	3.2664	4.232682448	0.011437536	0.12	24	103.44	0.249626762	3.48	0.17	-0.14	1.02	0.58	4,027	1.62	1.01	0.4097
018 GARCÍA	140	16300	28974	0.014737351	0.043590806	0.636743287	0.086077173	0.331711189	3.3773	4.108411215	0.043970456	0.10	21	102.98	0.657555049	3.68	0.50	-1.07	1.22	0.56	6,289	1.62	1.03	0.0584
020 GENERAL BRAVO	141	3655	5799	0.020865666	0.046559752	0.670460424	0.107259872	0.327987584	2.6889	3.454488778	0.054492154	0.07	27	111.57	0.285566477	3.34	0.45	-4.09	1.39	0.76	20,646	2.76	1.00	0.1407
021 GENERAL ESCOBEDO	142	133384	233457	0.01186514	0.020624783	0.586133635	0.17318821	0.360186244	2.9845	4.217982504	0.037338782	0.07	22	101.70	0.690144224	3.98	0.68	-2.80	0.68	0.49	8,515	1.52	1.02	0.0440
022 GENERAL TERÁN	143	10143	15475	0.034184168	0.046785137	0.684394184	0.116768982	0.334668821	2.5041	3.523532153	0.015896607	0.09	28	99.81	0.517673667	2.50	0.37	-3.78	1.64	0.66	6,930	1.36	1.00	0.1947
023 GENERAL TREVIÑO	144	1109	1699	0.021777516	0.037080636	0.625662154	0.150676869	0.280753384	2.1351	3.16502947	0.044732195	0.09	30	104.70	0.2866392	3.20	0.60	-5.65	1.89	0.80	4,877	7.65	0.99	0.2782
024 GENERAL ZARAGOZA	145	2983	5576	0.013809182	0.072991392	0.68687231	0.051470588	0.298063128	4.2830	4.39452496	0.008456241	0.11	20	108.92	0.099892396	3.23	0.13	1.90	0.17	0.62	2,191	0.36	1.06	0.6701
025 GENERAL ZUAZUA	146	3698	6033	0.019890602	0.048897729	0.657384386	0.096966683	0.402618929	3.4251	3.984574111	0.058345765	0.09	23	103.54	0.683739433	2.63	0.61	-2.85	1.02	0.72	8,394	0.83	1.02	0.0512

026 GUADALUPE	147	425563	670162	0.017439067	0.018573121	0.529492869	0.280017369	0.39661455	2.4658	4.248057859	0.02655925	0.07	24	98.75	0.700775335	3.98	0.92 -	4.43 -	0.48	0.62	14,241	2.06	1.01	0.0416
027 HERRERAS, LOS	148	1908	2795	0.032200358	0.039713775	0.716636852	0.086940966	0.320572451	2.2048	3.204628502	0.024329159	0.07	33	103.27	0.191413238	3.50	0.64 -	5.29	1.97	0.93	6,817	5.72	0.98	0.2154
047 HIDALGO	149	8814	14275	0.03474606	0.033765324	0.572049037	0.194535902	0.354255692	2.6398	3.968160045	0.014220665	0.09	24	99.09	0.768966725	3.89	0.76 -	3.59	1.03	0.77	8,555	1.61	1.02	0.0572
028 HIGUERAS	150	866	1371	0.025528811	0.051057622	0.623632385	0.13785558	0.41210795	2.7788	3.552845528	0.061998541	0.08	27	105.24	0.517140773	2.95	0.50 -	3.49	1.17	0.80	8,191	0.00	0.97	0.1155
029 HUALAHUISES	151	4045	6413	0.031654452	0.036488383	0.647902698	0.162014658	0.312958054	2.4635	3.655794991	0.017464525	0.11	27	93.98	0.464213317	3.13	0.42 -	3.26	1.17	0.78	5,325	1.09	1.00	0.2013
030 ITURBIDE	152	2036	3484	0.022675086	0.064867968	0.693168772	0.055683123	0.312858783	3.8864	3.869714286	0.008036739	0.11	24	100.00	0.242250287	3.20	0.21	0.61	0.53	0.86	3,408	0.86	1.01	0.4580
031 JUÁREZ	153	37227	66497	0.014647277	0.024482307	0.626404199	0.112125359	0.343699716	3.1878	4.115455294	0.032706242	0.07	22	102.57	0.607110095	3.68	0.46 -	1.83	1.46	0.42	7,667	1.50	1.01	0.0544
032 LAMPAZOS DE NARANJO	154	3100	5305	0.021112158	0.039396795	0.593213949	0.115928369	0.324222432	2.6305	3.759911894	0.02262017	0.09	24	102.56	0.432233742	2.58	0.47 -	3.59	1.33	0.66	6,902	0.00	0.97	0.1513
033 LINARES	155	42091	69205	0.020157503	0.029665487	0.635691063	0.170117766	0.346607904	2.5585	4.094006347	0.020143053	0.09	24	97.47	0.546607904	3.38	0.50 -	2.56	0.85	0.65	8,802	3.70	1.01	0.1518
034 MARÍN	156	2822	4719	0.018436109	0.036024581	0.63212545	0.128417038	0.37995338	3.0503	3.888235294	0.077134986	0.07	23	106.16	0.616656071	2.96	0.78 -	3.41	0.95	0.69	8,993	2.54	1.00	0.0503
035 MELCHOR OCAMPO	157	801	1215	0.027983539	0.019753086	0.691358025	0.110288066	0.319341564	2.3641	3.232686981	0.04691358	0.07	31	108.05	0.288888889	3.97	0.81 -	5.99	2.37	0.83	8,028	0.00	0.99	0.1347
036 MIER Y NORIEGA	158	3746	7078	0.017942922	0.121079401	0.672506358	0.023594236	0.233257982	3.9796	4.652925532	0.008053122	0.14	19	100.97	0.033625318	3.72	0.01	1.29	0.49	0.71	1,011	1.70	1.03	0.8175
037 MINA	159	2959	5049	0.029906912	0.059021588	0.652208358	0.088730442	0.347593583	3.0876	3.95280716	0.039611804	0.11	23	106.17	0.601901367	3.37	0.33 -	0.69	1.22	0.67	5,925	0.59	0.98	0.1447
038 MONTEMORELOS	160	33570	52741	0.021122087	0.030488614	0.594717582	0.204717393	0.357274227	2.4331	3.75219106	0.038148689	0.08	25	99.36	0.586014675	2.91	0.48 -	3.74	0.77	0.66	12,766	3.32	1.01	0.1074
039 MONTERREY	161	747572	1110997	0.022367297	0.022984761	0.496780819	0.320482414	0.402776965	2.3463	4.097734007	0.039706678	0.08	26	97.20	0.654935162	3.70	0.90 -	4.86 -	0.90	0.66	14,769	4.53	1.01	0.0532
040 PARÁS	162	814	1226	0.041598695	0.048939641	0.634584013	0.123980424	0.345840131	2.2080	3.078167116	0.07177814	0.09	34	111.38	0.360522023	3.71	0.60 -	5.18	2.23	0.95	7,699	6.53	1.00	0.1971
041 PESQUERÍA	163	6813	11321	0.017489621	0.040809116	0.654270824	0.090981362	0.370461973	2.9210	3.937319885	0.041250773	0.09	24	106.93	0.653917498	3.14	0.51 -	2.30	1.61	0.67	7,503	0.97	1.01	0.0773
042 RAMONES, LOS	164	4081	6237	0.025974026	0.043931377	0.726150393	0.059804393	0.27994228	2.9833	3.461088356	0.028058361	0.10	30	101.00	0.208593875	3.07	0.42 -	3.54	2.27	0.62	5,056	3.37	0.99	0.2077
043 RAYONES	165	1559	2613	0.023344814	0.094527363	0.670876387	0.040566399	0.325679296	3.0335	3.63832853	0.009950249	0.10	26	108.37	0.118637581	2.98	0.18 -	0.27	1.06	0.99	7,235	0.77	1.07	0.5104
044 SABINAS HIDALGO	166	19951	32329	0.021343067	0.02811717	0.588697454	0.211605679	0.353707198	2.8194	3.720629699	0.025487952	0.07	25	98.40	0.535958427	3.37	0.75 -	5.51	0.62	0.72	9,500	2.10	1.01	0.0711
045 SALINAS VICTORIA	167	11374	19024	0.015506728	0.030067283	0.630519344	0.111227923	0.354814971	3.1802	3.984575835	0.051040791	0.08	23	107.35	0.640191337	3.16	0.36 -	1.97	0.96	0.64	8,772	0.74	1.01	0.0513
046 SAN NICOLÁS DE LOS GA	168	324274	496878	0.016494995	0.011562194	0.460237724	0.375369809	0.39917042	2.2161	4.205631668	0.030681576	0.06	25	98.44	0.748147433	4.12	0.96 -	5.82 -	1.53	0.67	14,478	2.10	1.01	0.0356
019 SAN PEDRO GARZA GAR	169	83447	125978	0.013542047	0.0140977	0.390480678	0.417239518	0.414477131	1.8220	4.296537258	0.065816571	0.06	25	86.36	0.514272333	3.95	0.93 -	6.61 -	3.88	0.71	32,877	4.24	1.00	0.0376
048 SANTA CATARINA	170	138290	227026	0.014857329	0.023402606	0.588016351	0.199333997	0.393981306	2.9005	4.351975778	0.026697383	0.07	23	99.94	0.730356875	4.00	0.83 -	3.40	0.08	0.56	9,745	1.92	1.02	0.0415
049 SANTIAGO	171	23734	36812	0.019504509	0.02724655	0.570194502	0.22748017	0.374524612	2.4467	3.637033248	0.020455286	0.08	26	101.29	0.647044442	2.98	0.73 -	5.26	0.21	0.72	11,739	1.58	1.01	0.0614
050 VALLECILLO	172	1348	2169	0.04472107	0.040571692	0.73121254	0.045643154	0.334716459	2.4154	3.293075684	0.046565237	0.08	29	107.96	0.381742739	3.09	0.36 -	3.24	2.37	0.85	5,115	0.92	1.01	0.2292
051 VILLALDAMA	173	2703	4247	0.025665175	0.032493525	0.62326348	0.16647045	0.332234518	2.6519	3.357142857	0.017894965	0.07	28	99.11	0.477748999	3.41	0.54 -	4.45	1.57	0.84	9,370	16.48	0.99	0.1142
010 ALVARO OBREGÓN	174	458562	687020	0.016428925	0.024463626	0.511449448	0.310636657	0.421839248	2.0335	3.969101962	0.039083287	0.08	26	91.06	0.500177579	3.45	0.87 -	5.34 -	1.63	0.75	19.27	2.55	1.03	0.0994
002 AZCAPOTZALCO	175	304901	441008	0.020389653	0.01740558	0.480129612	0.367367485	0.415699942	1.8545	3.842097349	0.047359685	0.08	28	90.99	0.631330951	3.14	0.88 -	6.02 -	1.57	0.85	19.70	4.06	1.03	0.0994
014 BENITO JUÁREZ	176	272165	360478	0.022442424	0.008524792	0.30117788	0.575874256	0.484048957	1.3550	3.023208244	0.071025139	0.06	33	80.18	0.581242129	2.76	0.95 -	11.45 -	4.84	0.83	17.65	5.31	1.02	0.0712
003 COYOACÁN	177	453228	640423	0.020583271	0.016484417	0.406892007	0.445266644	0.441150615	1.6887	3.743634486	0.045832208	0.07	28	88.36	0.556643031	3.70	0.90 -	7.52 -	3.30	0.81	18.82	3.28	1.02	0.0913
004 CUAJIMALPA DE MORELC	178	93945	151222	0.01251802	0.024454114	0.515705387	0.273624208	0.402666279	2.2888	4.273798494	0.062609938	0.09	24	90.57	0.425672191	3.42	0.83 -	4.28 -	1.69	0.70	21.76	3.06	1.03	0.0905
015 CUAUHTÉMOC	179	361970	516255	0.022700022	0.015658928	0.45081791	0.396552092	0.45210797	1.7921	3.322936254	0.06170594	0.08	29	88.07	0.53300965	2.75	0.90 -	7.42 -	2.12	0.80	18.09	12.39	1.03	0.1096
005 GUSTAVO A. MADERO	180	834142	1235542	0.020074591	0.021920744	0.514297369	0.322004432	0.402443624	2.0187	3.97374458	0.044579626	0.08	27	92.93	0.533994797	3.29	0.87 -	4.95 -	1.15	0.80	20.36	7.79	1.03	0.1156
006 IZTACALCO	181	280103	411321	0.019009484	0.019099438	0.494757136	0.340179568	0.419545805	1.9587	3.93416377	0.036621033	0.08	27	91.03	0.542872841	3.21	0.87 -	5.38 -	1.67	0.81	20.07	3.64	1.03	0.1114
007 IZTAPALAPA	182	1126899	1773343	0.016904231	0.025003623	0.54704589	0.261850076	0.397972079	2.3269	4.142950804	0.043558409	0.08	25	95.06	0.46732358	3.48	0.85 -	3.98 -	0.71	0.70	20.39	5.81	1.04	0.1330
008 MAGDALENA CONTRERA	183	144358	222050	0.017829318	0.025016888	0.523350597	0.283661338	0.413861743	2.1235	4.02977194	0.034721909	0.08	25	92.12	0.498473317	3.54	0.81 -	4.65 -	1.65	0.78	20.28	1.86	1.03	0.1131
016 MIGUEL HIDALGO	184	252239	352640	0.019943852	0.014493534	0.423755104	0.426687273	0.449529265	1.5910	3.528343144	0.082406987	0.07	30	83.18	0.539431148	2.69	0.90 -	7.74 -	2.88	0.81	19.07	6.48	1.02	0.0826
009 MILPA ALTA	185	57996	96773	0.012121149	0.036807787	0.58839759	0.199766464	0.367902204	2.5294	4.344626115	0.024376634	0.08	23	98.00	0.339071849	3.54	0.67 -	1.73	0.41	0.64	22.19	4.11	1.04	0.2369
011 TLAHUAC	186	185291	302790	0.014359787	0.023197596	0.542970376	0.254162951	0.37383335	2.4190	4.153873944	0.041883814	0.08	24	94.94	0.48250933	3.69	0.85 -	3.48 -	0.12	0.65	21.11	2.28	1.04	0.1322
012 TLALPAN	187	383749	581781	0.015851325	0.022181886	0.472837717	0.352123909	0.420276702	2.0358	3.927545353	0.048855153	0.07	26	92.84	0.486863614	3.57	0.78 -	5.74 -	2.37	0.75	19.14	2.44	1.03	0.1104
017 VENUSTIANO CARRANZA	188	316626	462806	0.02188606	0.017113002	0.500138287	0.343779035	0.416651902	1.9179	3.799861893	0.037674533	0.08	28	89.98	0.522478101	3.17	0.90 -	5.70 -	1.15	0.83	20.04	6.49	1.03	0.1149
013 XOCHIMILCO	189	234150	369787	0.016571702	0.0260934	0.513268449	0.302530916	0.395460089	2.2558	4.195786951	0.035082358	0.07	25	96.78	0.458902017	3.42	0.74 -	3.75 -	1.16	0.88	19.43	7.97	1.03	0.1487
001 ACACOYAGUA	190	6994	14189	0.024103179	0.103319473	0.628585524	0.049968285	0.299809712	3.7727	5.122826087	0.015998309	0.12	17	105.25	0.060837268	2.92	0.26	2.60 -	0.16	0.58	2,146	0.00	1.15	0.5913
002 ACALA	191	13900	24754	0.013654359	0.172860952	0.564636018	0.071988366	0.28165145	3.0419	4.206139599	0.005817242	0.10	21	99.74	0.146400582	3.16	0.51	1.91	0.25	0.51	2,273	0.08	1.12	0.7331
003 ACAPETAHUA	192	13254	25154	0.018525881	0.125149082	0.606821977	0.064005725	0.299236702	3.2784	4.420688432	0.016299594	0.12	19	102.32	0.130317246	2.74	0.19	1.62	0.33	0.57	2,332	0.08	1.12	0.5122
113 ALDAMA	193	1396	3635	0.004126547	0.255570839	0.291059147	0.001375516	0.240165062	5.3317	4.786392405	0	0.13	16	94.91	0.016781293	2.91	0.01	*	*	NA	752	0.00	1.91	0.9230
004 ALTAMIRANO	194	7531	21948	0.017814835	0.150492072	0.375523966	0.019637325	0.216466193	4.6805	5.189012426	0.005148533	0.09	16	99.87	0.128576636	2.87	0.13	3.36 -	0.65	0.60	2,158	0.00	1.45	0.8140
005 AMATÁN	195	8582	18778	0.016135904	0.180796677	0.545532006	0.01384599	0.270848866	5.1630	5.252951096	0.008041325	0.16	16	104.75	0.011769091	2.64	0.08	3.75 -	0.74	0.48	862	0.00	1.49	0.9027
006 AMATENANGO DE LA FR(	196	12938	26094	0.012301679	0.13949567	0.624588028	0.027324289	0.281673948	4.2261	5.31710311	0.006055032	0.16	17	101.06	0.285736185	2.98	0.30	2.81 -	0.19	0.56	1,601	0.00	1.17	0.8143
007 AMATENANGO DEL VALLI	197	2696	6559	0.008690349	0.220003049	0.426284495	0.007013264	0.292422625	4.0868	4.682426405	0.000152462	0.13	18	94.69	0.050312548	3.66	0.10	3.77 -	0.75	0.85	636	0.00	1.82	0.9618
008 ANGEL ALBINO CORZO	198	10999	21848	0.0111223	0.154155987	0.581838155	0.05222446	0.321768583	3.9270	5.371598219	0.004622849	0.14	18	103.96	0.040598682	3.18	0.41	2.44 -	0.03	0.67	2,699	0.00	1.18	0.6290
009 ARRIAGA	199	22278	37989	0.018584327	0.095922504	0.585827476	0.142830819	0.336465819	2.6931	3.987633443	0.023269894	0.10	23	97.56	0.25246782	3.07	0.31 -	0.11	0.81	0.52	3,505	0.03	1.09	0.4442
010 BEJUCAL DE OCAMPO	200	3141	6673	0.012438184	0.103851341	0.672111494	0.013187472	0.259553424	5.5413	6.103669725	0.002997153	0.16	16	102.46	0.039112843	2.94	0.04	*	*	0.36	563	0.00	1.26	0.9037
011 BELLA VISTA	201	8947	18205	0.010876133	0.074814611	0.704476792	0.025542433	0.275693491	4.3725	5.7718014	0.002966218	0.16	17	99.99	0.114144466	2.83	0.16	3.16 -	0.50	0.46	4,845	0.00	1.13	0.9110
114 BENEMÉRITO DE LAS AM	202	6332	14436	0.011706844	0.134663342	0.561651427	0.01426981	0.278054863	5.0526	5.393233083	0.057148795	0.11	15	105.03	0.019950125	2.85	0.01	3.38 -	0.50	NA	2,789	0.00	1.22	0.6974
012 BERRIOZÁBAL	203	14996	28719	0.008879139	0.128068526	0.556809081	0.080956858	0.324523834	3.7355	4.670336894	0.020578711	0.12	18	101.88	0.168355444	2.93	0.35	1.26 -	0.14	0.51	3,052	0.00	1.20	0.4690
013 BOCHIL	204	10555	22722	0.011266614	0.161605492	0.517075962	0.069448112	0.286902561	4.7443	4.911680283	0.002728633	0.12	16	100.14	0.090220931	2.95	0.32	2.61 -	0.32	0.57	1,865	0.00	1.29	0.7007
014 BOSQUE, EL	205	5529	14993	0.006002801	0.183619022	0.402321083	0.015073701	0.194490762	4.3963	5.12135506	6.66978E-05	0.11	16	99.99	0.047155339	3.25	0.56	3.31 -	0.52	0.53	1,330	0.00	1.63	0.7474
015 CACAHOATÁN	206	21101	39033	0.014500551	0.118438245	0.587041734	0.098019624	0.283759895	3.3228	4.889224192	0.012912151	0.13	20	98.97	0.1877642	2.71	0.54	1.65	0.32	0.59	2,276	0.00	1.14	0.6236
016 CATAZAJÁ	207	9327	15709	0.015277866	0.129989178	0.637468967	0.087911388	0.321026163	2.5090	4.399661781	0.023553377	0.11	22	106.32	0.091539682	2.82	0.29	1.61	0.45	0.57	1,929	0.00	1.10	0.7642
022 CHALCHIHUITÁN	208	5237	12256	0.004650783	0.251305483	0.458061358	0.003426893	0.32131201	5.7440	4.83009901	0	0.14	14	101.84	0.081347911	2.76	0.00	3.77 -	0.77	0.57	960	0.00	1.78	0.9460
023 CHAMULA	209	27560	59005	0.008643335	0.313295483	0.413897127	0.003287857	0.315227523	4.8863	4.781940128	0.000389797	0.21	16	90.78	0.005897805	3.30	0.01	3.64 -	0.66	0.55	957	0.00	2.08	0.8778
024 CHANAL	210	3048	7568	0.011760042	0.214587738	0.432610994	0.015988372	0.277748414	5.0829	5.739761092	0.000132135	0.19	15	98.53	0.051664905	2.93	0.00	*	*	0.68	224	0.00	1.57	0.9765
025 CHAPULTENANGO	211	3243	6965	0.014788227	0.15850682	0.564967696	0.036180905	0.256568557	5.9112	5.363987635	0.00373295	0.15	16	98.38	0.021392678	2.75	0.24	*	*	0.64	1,057	0.00	1.38	0.8405
026 CHENALHÓ	212	10358	27331	0.007281109	0.187516007	0.422962936	0.007134755	0.254692474	5.3586	5.381554738	0.000182942	0.13	15	100.52	0.091471223	2.89	0.03	3.86 -	0.84	0.81	2,700	0.29	1.65	0.9122
027 CHIAPA DE CORZO	213	34776	60620	0.013229957	0.127977565	0.565176509	0.116281755	0.317485978	2.8897	4.366429402	0.004965358	0.09	21	99.25	0.237743319	3.23	0.36	1.16	0.49	0.52	2,865	0.03	1.12	0.5218
028 CHIAPILLA	214	2960	5242	0.02079359	0.230827928	0.560091568	0.024990462	0.333651278	2.7878	4.18174475	0.004196871	0.13	21	107.36	0.045211751	3.55	0.62	2.24	0.23	0.52	1,112	0.00	1.07	0.8550
029 CHICOASÉN	215	2408	4345	0.020023015	0.109781358	0.639585731	0.043498274	0.28009206	2.8839	4.455216017	0.044188723	0.12	20	106.51	0.231070196	3.09	0.49	1.72	0.40	0.63	2,416	0.00	1.19	0.4763
030 CHICOMUSELO	216	12581	24994	0.00904217	0.141994079	0.618948548	0.032527807	0.266663999	3.4414	5.220470786	0.003160759	0.13	18	101.39	0.040089622	2.91	0.18	2.53 -	0.09	0.66	1,312	0.04	1.19	0.7896
031 CHILÓN	217	28671	77686	0.004672657	0.197500193	0.381136884	0.020621476	0.236336019	4.7981	5.681220742	0.000630744	0.07	16	99.55	0.026838813	2.76	0.03	3.73 -	0.72	0.75	1,538	0.01	1.88	0.8742
017 CINTALAPA	218	35016	64013	0.012747411	0.111977255	0.602846297	0.095371253	0.302094887	3.1976	4.566235632	0.007748426	0.10	20	99.46	0.145361099	3.08	0.28	1.18	0.47	0.53	2,421	0.00	1.10	0.6336
018 COAPILLA	219	3701	7217	0.012470556	0.138146044	0.604267701	0.032284883	0.308022724	4.2741	4.339975845	0.001801302	0.13	18	102.61	0.04004434	2.91	0.24	3.23 -	0.39	0.70	1,508	0.00	1.20	0.8268
019 COMITÁN DE DOMÍNGUE	220	59952	105210	0.016538352	0.112251687	0.571552134	0.125529693	0.366391027	2.9204	4.322241609	0.019532364	0.13	21	92.63	0.232373349	3.03	0.34 -	0.16 -	0.02	0.51	5,411	0.10	1.13	0.4504
020 CONCORDIA, LA	221	20680	39770	0.011566507	0.166582851	0.571259744	0.037943173	0.295876289	3.7431	5.000252845	0.002589892	0.12	18	105.65	0.039376414	3.14	0.24	2.35	0.16	0.41	1,668	0.00	1.13	0.7456
021 COPAINALÁ	222	10357	19298	0.015079283	0.105917712	0.611617784	0.083272878	0.319618613	3.5444	4.38993135	0.005130065	0.11	19	100.64	0.089387501	3.06	0.35	2.30 -	0.06	0.49	2,160	0.00	1.11	0.7019
032 ESCUINTLA	223	14349	28064	0.017281927	0.111245724	0.597491448	0.08245439	0.282354618	3.8771	4.97099128	0.011616306	0.14	18	98.54	0.18097919	2.88	0.33	2.28 -	0.03	0.50	2,214	0.00	1.14	0.5961
033 FRANCISCO LEÓN	224	2351	5236	0.008212376	0.18487395	0.535905271	0.010122231	0.258594347	6.8338	5.694656489	0.000572956	0.12	15	102.71	0.011650115	2.50	0.05	*	*	0.66	585	0.00	1.42	0.9079
034 FRONTERA COMALAPA	225	26698	52168	0.009507744	0.096131728	0.642616163	0.063353013	0.292995706	3.3846	4.841903873	0.01056203	0.13	18	95.65	0.103358381	2.84	0.35	1.82	0.30	0.37	2,633	0.04	1.12	0.6408
035 FRONTERA HIDALGO	226	5867	10917	0.015388843	0.139873592	0.564715581	0.095722268	0.24649629	3.0585	4.674398625	0.009526427	0.12	20	99.18	0.173582486	2.45	0.17	1.78	0.32	0.55	2,390	0.00	1.11	0.5624

036 GRANDEZA, LA	227	2746	5969	0.013905177	0.078405093	0.660077065	0.039202546	0.289160663	5.9935	6.174093264	0.003685709	0.14	16	101.38	0.059641481	2.84	0.08	*	*	0.53	681	0.00	1.21	0.8776
037 HUEHUETÁN	228	16684	31464	0.012140859	0.126017035	0.55682685	0.111333588	0.27294686	2.7280	4.742930591	0.007818459	0.13	19	98.30	0.189581744	2.53	0.12	1.72	0.26	0.59	2,912	0.10	1.15	0.4861
039 HUITIUPÁN	229	8846	20041	0.008332918	0.193802704	0.542138616	0.016316551	0.265056634	4.6639	5.359093392	0.001397136	0.15	16	102.03	0.011526371	2.80	0.04	3.70	- 0.70	0.60	876	0.10	1.50	0.9005
038 HUIXTÁN	230	8507	18630	0.020665593	0.190982287	0.544766506	0.012936125	0.306387547	4.7540	5.699967459	0.000429415	0.15	17	97.58	0.143370907	2.93	0.05	3.82	- 0.79	0.56	1,193	0.00	1.45	0.8998
040 HUIXTLA	231	26720	48476	0.016729928	0.091798003	0.599801964	0.125340375	0.314279231	2.7413	4.408948339	0.017390049	0.12	20	97.22	0.253197459	2.71	0.38	0.90	0.40	0.59	3,890	0.04	1.10	0.3777
041 INDEPENDENCIA, LA	232	16743	32245	0.011009459	0.129880602	0.662211196	0.009799969	0.323305939	3.1931	4.856448428	0.001643666	0.16	18	94.26	0.032997364	3.15	0.09	2.73	- 0.11	0.50	1,122	0.00	1.22	0.9054
042 IXHUATÁN	233	4238	8877	0.013292779	0.166272389	0.543764785	0.03694942	0.304832714	3.8696	4.931476323	0.004843979	0.15	17	100.07	0.149149467	2.73	0.39	3.15	- 0.43	0.55	2,198	0.00	1.44	0.7947
043 IXTACOMITÁN	234	4677	9143	0.021765285	0.117685661	0.573444165	0.08542054	0.283604944	3.9724	4.898332437	0.018046593	0.14	18	97.47	0.091654818	2.48	0.33	2.53	- 0.09	0.55	2,264	0.00	1.16	0.6077
044 IXTAPA	235	9266	18533	0.009874278	0.148383964	0.582636378	0.047482868	0.281767658	4.7916	4.932817915	0.000269789	0.11	18	98.87	0.109480386	3.50	0.34	3.12	- 0.39	0.69	1,692	0.05	1.24	0.8191
045 IXTAPANGAJOYA	236	2307	4707	0.045039303	0.136180157	0.559804546	0.022094752	0.280220948	4.5596	5.248044693	0.018908009	0.13	17	99.87	0.019545358	2.35	0.17	3.66	- 0.69	0.59	1,300	0.00	1.26	0.8155
046 JIQUIPILAS	237	20185	34937	0.012508229	0.109082062	0.633740733	0.083206915	0.318115465	2.8427	4.222559495	0.008787246	0.09	22	102.02	0.142656782	3.29	0.36	1.34	0.64	0.50	2,261	0.00	1.08	0.7557
047 JITOTOL	238	5903	13076	0.009483022	0.180559804	0.526537167	0.0426736	0.235851942	5.8731	5.091334895	0.001835424	0.09	15	102.60	0.261012542	3.10	0.34	3.23	- 0.66	0.74	1,350	0.00	1.46	0.8074
048 JUÁREZ	239	11013	19956	0.016987372	0.118761275	0.594257366	0.083684105	0.31208659	2.7743	4.551937629	0.024503909	0.12	20	101.17	0.087091602	2.11	0.19	1.60	0.38	0.54	3,359	0.00	1.12	0.5642
049 LARRAINZAR	240	5827	16538	0.004837344	0.168218648	0.410992865	0.011125892	0.226145846	5.2774	5.202469136	6.04668E-05	0.15	16	99.59	0.001632604	3.16	0.04	*	*	0.78	481	0.00	1.50	0.9358
050 LIBERTAD, LA	241	3179	5288	0.010022693	0.108736762	0.649205749	0.057299546	0.340393343	2.2575	3.957767722	0.023827534	0.08	24	100.00	0.056354009	2.42	0.24	1.11	0.81	0.70	3,607	0.00	1.14	0.7650
051 MAPASTEPEC	242	20055	39055	0.021559339	0.121700166	0.604813724	0.058225579	0.277429266	3.5798	4.520670781	0.010958904	0.12	18	101.94	0.227755729	2.80	0.25	1.40	0.36	0.52	3,489	0.03	1.14	0.4176
115 MARAVILLA TENEJAPA	243	4721	11147	0.005831165	0.154032475	0.565443617	0.005203194	0.30420741	5.5964	5.613265306	0.010137257	0.10	15	105.97	0.017762627	2.83	0.01	*	*	NA	563	0.00	1.33	0.9323
052 MARGARITAS, LAS	244	37651	86413	0.0093273	0.176512793	0.505201764	0.028155486	0.302963674	4.3667	5.575248933	0.002916228	0.15	17	98.78	0.154860958	3.02	0.08	3.09	- 0.57	0.55	1,411	0.01	1.34	0.8342
116 MARQUÉS DE COMILLAS	245	3432	8580	0.017365967	0.140909091	0.507342657	0.018997669	0.251282051	5.0397	5.502094972	0.034149184	0.11	15	107.80	0.025058275	2.94	0.01	3.59	- 0.65	NA	1,201	0.00	1.28	0.7774
053 MAZAPA DE MADERO	246	3620	7180	0.006128134	0.070612813	0.654874652	0.029247911	0.258077994	3.9802	5.762096774	0.001392758	0.11	18	99.28	0.197771588	2.95	0.27	2.92	- 0.37	0.57	1,325	0.00	1.17	0.8244
054 MAZATÁN	247	13647	24079	0.017650235	0.119980066	0.595747332	0.086589975	0.304165455	2.7545	4.212979247	0.015449147	0.09	22	102.01	0.183313261	2.97	0.07	1.53	0.52	0.59	2,507	0.08	1.10	0.4832
055 METAPA	248	2710	4794	0.014601585	0.111597831	0.571547768	0.12745098	0.282644973	2.2902	4.462184874	0.018773467	0.11	22	98.02	0.274926992	2.37	0.39	1.04	0.53	0.61	7,815	0.00	1.16	0.4089
056 MITONTIC	249	3243	7602	0.009602736	0.305840568	0.395422257	0.008550381	0.191133912	6.0594	4.954966887	0	0.24	15	94.33	0.064456722	2.97	0.11	*	*	0.50	913	0.00	1.76	0.8162
117 MONTECRISTO DE GUER	250	2347	5086	0.014353126	0.150412898	0.574714904	0.013173417	0.282736925	4.9285	5.769495413	0.000786473	0.14	16	107.25	0.095163193	2.84	0.29	3.27	- 0.40	NA	3,448	0.00	1.32	0.6335
057 MOTOZINTLA	251	29796	59675	0.012876827	0.086580376	0.643490605	0.060542797	0.272016701	4.5936	5.588638498	0.006680585	0.14	18	101.23	0.113903967	2.94	0.35	2.49	- 0.26	0.48	4,645	0.00	1.15	0.6648
058 NICOLÁS RUIZ	252		3135	0	0	0	0	0			0	-	N.A.	99.94	0	#¡DIV/0!	#¡DIV/0!	*	*	NA	n.d.	0.00	#¡DIV/0!	#¡DIV/0!
059 OCOSINGO	253	50011	146696	0.004867208	0.132259912	0.390249223	0.034227245	0.20779026	4.7445	5.394636972	0.004942193	0.09	16	101.00	0.082422152	2.96	0.15	3.42	- 0.63	0.70	2,298	0.00	1.39	0.7319
060 OCOTEPEC	254	4229	9271	0.015532305	0.260489699	0.429403516	0.018336749	0.266098587	6.7696	5.010298103	0.005177435	0.24	16	99.59	0.021141193	2.73	0.11	3.98	- 0.90	0.71	1,402	0.00	1.52	0.8536
061 OCOZOCOAUTLA DE ESF	255	33818	65673	0.011648623	0.120902045	0.580025277	0.071338297	0.326252798	4.0310	4.696247209	0.006029875	0.10	18	99.46	0.099919297	3.01	0.40	1.74	0.29	0.50	2,341	0.00	1.16	0.6629
062 OSTUACÁN	256	8196	17026	0.020674263	0.142076824	0.556149419	0.034652884	0.275578527	3.9445	5.029037037	0.027311171	0.13	17	103.81	0.054152473	2.34	0.15	3.01	- 0.33	0.55	1,648	0.12	1.18	0.7251
063 OSUMACINTA	257	1650	3132	0.021711367	0.079501916	0.644316731	0.043103448	0.311302682	3.3335	4.269863014	0.05715198	0.12	19	101.16	0.190613027	2.74	0.59	1.94	0.31	0.69	2,506	0.00	1.16	0.6640
064 OXCHUC	258	17120	37887	0.01918864	0.161295431	0.580014253	0.024757833	0.330007654	4.7333	5.574762837	2.63943E-05	0.11	16	100.56	0.122706997	2.75	0.01	3.93	- 0.88	0.65	1,141	0.00	1.38	0.9458
065 PALENQUE	259	42117	85464	0.010647758	0.13328419	0.54199429	0.081940934	0.272910231	3.5314	4.73846771	0.021400624	0.10	18	99.48	0.140012169	2.64	0.24	2.22	- 0.05	0.55	2,098	0.00	1.19	0.6124
066 PANTELHÓ	260	6868	16262	0.019739269	0.252060017	0.378489731	0.012975034	0.262882794	5.0367	5.28530772	0.00079941	0.15	16	100.72	0.10982659	2.69	0.09	3.68	- 0.70	0.45	1,194	0.00	1.63	0.8804
067 PANTEPEC	261	4150	8566	0.021130049	0.249824889	0.479220173	0.014125613	0.3565258	5.2612	4.669764899	0.004669624	0.18	17	100.51	0.444314733	2.66	0.19	3.58	- 0.70	0.67	1,708	0.00	1.55	0.8435
068 PICHUCALCO	262	16049	29357	0.015873557	0.131212317	0.531832272	0.113397145	0.333412815	2.8340	4.597427346	0.026365092	0.11	20	100.77	0.131893586	2.06	0.25	1.76	0.16	0.59	6,966	0.34	1.11	0.5622
069 PIJIJIAPAN	263	25562	46949	0.013440116	0.124539394	0.600673071	0.075592664	0.298302413	3.0054	4.291076357	0.01243903	0.11	20	100.46	0.096934972	3.07	0.20	1.48	0.52	0.58	2,637	0.09	1.08	0.4603
070 PORVENIR, EL	264	5407	11641	0.012284168	0.080749077	0.676316468	0.029464823	0.27901383	5.4983	6.816206694	0.002319388	0.15	16	101.16	0.025770982	2.99	0.07	3.64	- 0.67	0.59	2,111	0.00	1.23	0.8999
072 PUEBLO NUEVO SOLISTA	265	11120	24405	0.015734481	0.21958615	0.474984634	0.036795739	0.268674452	5.6965	5.201297297	0.014300348	0.15	16	97.53	0.044785905	2.75	0.27	3.23	- 0.66	0.51	1,137	0.00	1.50	0.8384
073 RAYÓN	266	3244	6870	0.014556041	0.186462882	0.515720524	0.048908297	0.26783115	6.1789	5.074759437	0.004075691	0.19	16	101.11	0.083842795	2.63	0.27	3.12	- 0.40	0.64	1,632	0.00	1.34	0.6841
074 REFORMA	267	19027	34809	0.015866696	0.075497716	0.602315493	0.129104542	0.301875952	2.7874	4.362158074	0.078715275	0.10	20	100.95	0.322531529	2.42	0.54	0.26	0.44	0.65	5,128	0.23	1.09	0.3310
075 ROSAS, LAS	268	11016	21100	0.018293839	0.227819905	0.480900474	0.040473934	0.322701422	3.9046	4.554712893	0.003554502	0.19	19	98.65	0.113886256	2.67	0.28	2.02	0.03	0.48	1,732	0.00	1.17	0.7621
076 SABANILLA	269	9086	21156	0.016118359	0.180468898	0.375543581	0.012478729	0.265031197	4.6718	5.213813179	0.001796181	0.16	16	101.49	0.175978446	2.79	0.16	3.80	- 0.77	0.49	802	0.00	1.52	0.8986
077 SALTO DE AGUA	270	23161	49300	0.010973631	0.215436105	0.513955375	0.033691684	0.279959432	3.7810	5.363350903	0.006085193	0.10	17	100.38	0.134077079	2.80	0.03	3.59	- 0.64	0.47	1,157	0.00	1.65	0.8655
118 SAN ANDRÉS DURAZNAL	271	1362	3423	0.004966404	0.21121823	0.404615834	0.003797838	0.257084429	4.6840	5.055932203	0	0.12	15	96.27	0.001168566	3.75	0.22	3.84	- 0.81	NA	1,357	0.00	2.40	0.8335
078 SAN CRISTÓBAL DE LAS	272	72143	132421	0.011070752	0.109438835	0.489574916	0.186986958	0.360426216	3.1254	4.55976758	0.016606127	0.12	20	92.76	0.251463137	3.13	0.54	- 0.06	- 0.29	0.56	5,073	0.07	1.17	0.3805
079 SAN FERNANDO	273	14658	26436	0.015736117	0.14090634	0.599334241	0.050953246	0.334279013	3.4688	4.424206816	0.010251173	0.13	20	103.87	0.114805568	3.32	0.31	1.84	0.23	0.48	4,625	0.00	1.15	0.5715
112 SAN JUAN CANCUC	274	9158	20688	0.006622196	0.264742846	0.419518561	0.009715777	0.369102862	4.4336	4.853857946	0.000193349	0.09	17	99.36	0.381718871	2.79	0.00	3.61	- 0.65	0.65	1,584	0.00	1.73	0.9241
110 SAN LUCAS	275	2800	5673	0.015688348	0.20500617	0.535871673	0.00987132	0.309688948	4.3667	4.645507007	0.002291556	0.16	18	103.48	0.01674599	3.38	0.25	3.04	- 0.35	0.56	666	0.00	1.24	0.9120
119 SANTIAGO EL PINAR	276	697	2174	0.010579577	0.254829807	0.263109476	0.00275989	0.155473781	6.2419	5.936241611	0	0.22	13	96.56	0	3.35	0.01	*	*	NA	683	0.00	4.34	0.8538
080 SILTEPEC	277	15350	32457	0.009150568	0.136549897	0.629386573	0.022429676	0.264719475	4.6051	6.103018868	0.00073944	0.12	16	105.28	0.130079798	2.84	0.14	3.47	- 0.66	0.59	762	0.00	1.23	0.8780
081 SIMOJOVEL	278	11604	31615	0.006199589	0.188201803	0.359291476	0.032990669	0.252886288	5.1524	5.286582809	0.001549697	0.12	16	99.12	0.09286731	2.90	0.22	3.29	- 0.71	0.68	1,626	0.00	1.51	0.7534
082 SITALÁ	279	2166	7987	0.022661825	0.187554777	0.183548266	0.004757731	0.213096281	4.8905	5.141899441	0	0.10	16	99.48	0.034305747	2.85	0.02	*	*	1.36	1,029	0.00	1.73	0.8414
083 SOCOLTENANGO	280	8127	15171	0.01423769	0.193658955	0.56153187	0.031573397	0.30821963	3.2280	4.527103923	0.001120559	0.16	19	101.72	0.226089249	2.86	0.21	1.65	0.42	0.49	1,921	0.00	1.16	0.7170
084 SOLOSUCHIAPA	281	3808	7784	0.01053443	0.145298047	0.564491264	0.051772867	0.291495375	4.0604	5.061929596	0.009249743	0.12	17	101.24	0.037384378	2.50	0.28	3.06	- 0.36	0.59	1,620	0.00	1.23	0.7588
085 SOYALÓ	282	3998	7767	0.015063731	0.178061027	0.587099266	0.039011201	0.28839964	3.8064	4.480324074	0.001802498	0.15	18	103.32	0.06793614	3.12	0.66	2.90	- 0.25	0.76	3,798	0.00	1.30	0.8476
086 SUCHIAPA	283	8638	15890	0.014537445	0.149276274	0.55059786	0.119509125	0.291567023	3.4913	4.687592538	0.00188796	0.08	20	102.52	0.158086847	3.40	0.46	1.53	0.40	0.53	2,715	0.00	1.09	0.5849
087 SUCHIATE	284	16076	30251	0.012495455	0.133053453	0.554163499	0.077121417	0.29899838	3.4136	4.395912409	0.034544313	0.13	19	97.96	0.141020132	2.46	0.28	1.34	0.55	0.49	2,824	0.07	1.12	0.3907
088 SUNUAPA	285	980	1936	0.030475207	0.146694215	0.55268595	0.049070248	0.325413223	3.8696	4.989664083	0.007231405	0.09	18	105.30	0.133780992	2.08	0.07	3.29	- 0.42	0.67	1,767	0.00	1.15	0.7758
089 TAPACHULA	286	157178	271674	0.014458505	0.079915634	0.545609812	0.179119091	0.339417832	2.7666	4.237345972	0.031361117	0.11	22	94.02	0.356677488	2.64	0.40	- 0.22	0.25	0.53	5,464	0.11	1.09	0.3161
090 TAPALAPA	287	1787	3639	0.025831272	0.128057159	0.611156911	0.032426491	0.350645782	4.6381	4.414841849	0.009343226	0.15	17	93.67	0.324814509	2.87	0.26	3.37	- 0.65	0.81	1,075	0.00	1.35	0.8695
091 TAPILULA	288	5210	10349	0.017392985	0.148130254	0.536573582	0.092472703	0.312687216	3.9164	4.730151446	0.009372866	0.15	18	98.22	0.094308629	2.71	0.51	2.32	- 0.05	0.50	2,603	0.39	1.16	0.5917
092 TECPATÁN	289	18957	38383	0.013235026	0.124846937	0.582679832	0.048016049	0.276737097	3.9213	4.706775268	0.016700102	0.12	17	99.73	0.11656202	2.62	0.33	2.60	- 0.18	0.52	1,987	0.05	1.16	0.6991
093 TENEJAPA	290	14127	33161	0.00566931	0.170863364	0.525255571	0.010464099	0.40704442	5.4620	5.824531516	0.000482494	0.15	16	96.51	0.006242273	3.01	0.05	3.75	- 0.75	0.53	1,107	0.06	1.46	0.9601
094 TEOPISCA	291	12868	26996	0.007075122	0.224403615	0.47358868	0.025596385	0.262557416	4.4801	4.94	0.001407616	0.15	17	96.05	0.04871092	3.21	0.32	2.74	- 0.38	0.44	1,335	0.00	1.47	0.7708
096 TILA	292	26990	58153	0.011366568	0.204735783	0.51627603	0.037968806	0.286124534	4.8034	5.457020197	0.001564838	0.15	17	99.92	0.107698857	2.99	0.17	3.68	- 0.70	0.52	3,488	0.09	1.56	0.8526
097 TONALÁ	293	44970	78438	0.016420612	0.101838395	0.597363523	0.124531477	0.309773324	3.0687	4.185613601	0.018116219	0.11	22	99.03	0.18414544	3.16	0.33	0.60	0.69	0.59	3,371	0.03	1.06	0.4241
098 TOTOLAPA	294	2779	5513	0.010883367	0.257935788	0.493923454	0.011427535	0.280972247	3.8511	5.007292616	0	0.17	18	101.13	0.011790314	3.50	0.27	*	*	0.71	573	0.00	1.07	0.9343
099 TRINITARIA, LA	295	32477	59686	0.010421204	0.131873471	0.668263914	0.018027678	0.306018162	2.9162	4.639563012	0.003819991	0.14	20	95.10	0.042589552	2.99	0.09	2.00	0.23	0.42	1,270	0.00	1.15	0.8346
100 TUMBALÁ	296	12018	26866	0.006476588	0.226271123	0.448485074	0.029628527	0.260180153	4.1093	5.504477612	0.000967766	0.11	18	97.38	0.029963523	2.90	0.07	3.77	- 0.76	0.55	913	0.07	1.81	0.8679
102 TUXTLA CHICO	297	18659	33467	0.012848478	0.136193863	0.564854932	0.118952998	0.251113037	2.8001	4.841097063	0.005557714	0.10	20	98.72	0.194669376	2.65	0.16	1.80	0.33	0.55	8,116	0.03	1.16	0.5056
101 TUXTLA GUTIÉRREZ	298	266598	434143	0.013942411	0.051939108	0.494033072	0.279534623	0.383477334	2.2670	4.048713733	0.025652838	0.08	23	92.54	0.416623094	3.22	0.69	- 2.13	- 0.33	0.53	2,441	0.22	1.06	0.2149
103 TUZANTÁN	299	12028	23180	0.014667817	0.103451251	0.646807593	0.058455565	0.275496117	3.3221	4.936938863	0.009275237	0.16	18	100.81	0.090250216	2.75	0.24	2.10	0.11	0.60	1,892	0.00	1.15	0.5992
104 TZIMOL	300	6646	11925	0.020880503	0.18591195	0.599496855	0.021048218	0.332830189	2.8319	4.218194741	0.00360587	0.18	20	99.18	0.153962264	2.91	0.24	1.99	0.33	0.58	1,594	0.00	1.17	0.7519
105 UNIÓN JUÁREZ	301	7743	13934	0.02246304	0.116334147	0.613032869	0.097674752	0.29783264	3.1405	4.888380282	0.026840821	0.17	20	99.89	0.196435482	2.79	0.48	1.84	0.23	0.61	2,232	0.00	1.15	0.7111
106 VENUSTIANO CARRANZA	302	28620	52833	0.012113641	0.196108493	0.519088449	0.059508262	0.301553953	3.1857	4.386685791	0.00232809	0.13	20	102.38	0.149565612	3.12	0.39	1.55	0.41	0.42	2,084	0.00	1.13	0.6553
071 VILLA COMALTITLÁN	303	13358	26706	0.016887591	0.12075938	0.597318955	0.062962101	0.280948102	3.5251	4.842412097	0.012806111	0.12	18	102.43	0.073204523	2.74	0.15	2.12	0.09	0.55	2,186	0.07	1.14	0.6056
107 VILLA CORZO	304	37083	68685	0.012651962	0.153570649	0.553803596	0.072635947	0.315134309	3.2908	4.479236371	0.003523331	0.09	20	101.84	0.083074907	3.02	0.29	1.63	0.44	0.44	2,141	0.03	1.12	0.7257
108 VILLAFLORES	305	47962	85957	0.012448085	0.121316472	0.567458148	0.112300336	0.332515095	2.9932	4.37145635	0.005177007	0.09	20	99.74	0.142350245	3.08	0.50	0.88	0.67	0.48	2,729	0.00	1.09	0.6195
109 YAJALÓN	306	13246	26044	0.007948088	0.200545231	0.453079404	0.089655967	0.299915528	4.0326	4.916057673	0.004953156	0.11	19	98.02	0.08961757	2.64	0.31	2.33	- 0.14	0.50	3,214	0.08	1.32	0.5657
111 ZINACANTÁN	307	13851	29754	0.008603885	0.289574511	0.418666398	0.00436916	0.288532634	5.2953	5.033682373	6.72179E-05	0.17	16	92.30	0.007461182	3.55	0.04	3.49	- 0.55	0.42	1,195	0.00	2.44	0.8381

Anexo 2
DESCRIPCIÓN DE VARIABLES DERIVADAS

1. Índice de Pobreza (P)

La medida más comúnmente empleada en la medición de la pobreza consiste simplemente en contar el número de *"pobres"* entre el total de los individuos [individuos con un ingreso por debajo de algún criterio definido como *la línea de pobreza* (z)], estimando la proporción de dichos individuos (*H*), como porcentaje de la población total.

Esta métrica, resulta incompleta, en virtud de por lo menos las dos siguientes deficiencias: (1) la proporción de individuos pobres (*H*), puede permanecer sin cambios, aún cuando la brecha entre el ingreso medio de dicha población y la línea de pobreza sufra un aumento considerable y (2) el índice *H* también resulta insensible a la distribución del ingreso dentro del sector pobre de la población. Una transferencia de recursos de los más pobres entre los pobres, hacia los menos pobres, mantendría a *H* invariante o incluso pudiera indicar un efecto favorable; sin duda, este un resultado perverso. Sen[31], propone un índice (*P*), que satisface, entre otras, los siguientes dos axiomas:

AXIOMA DE MONOTONICIDAD. Dado todo lo demás constante, una disminución en el ingreso de cualquier persona cuyo ingreso se encuentre por debajo de la línea de pobreza, deberá incrementar el valor de *P*.

AXIOMA DE TRANSFERENCIA. Dado todo lo demás constante, la transferencia de ingreso de cualquier persona cuyo ingreso se encuentre por debajo de la línea de pobreza, hacia cualquiera otra persona con un ingreso mayor, deberá incrementar el valor de *P*.

Este índice para poblaciones grandes, se define de la siguiente manera

$$P = H\left[I + (1 - I)G\right]$$

en donde

[31] Econometrica, **44** (March 1976), 219-31, publicado en Amartya Sen. 1999. *Choice, Welfare and Measurement*. Harvard University Press, Third Printing

P = índice de pobreza

$H = \frac{q}{n}$ razón del número de pobres

q = número de personas con ingreso inferior al de la línea de pobreza (z)

n = población total

$I = \sum_{i \in S(z)} g_i / qz$ razón de brecha en el ingreso

$g_i = z - y_i$ diferencia entre ingreso de la persona i (y_i) y la línea de pobreza (z)

$G = 1 + \frac{1}{q} - \frac{2}{q^2 m} \sum_{i=1}^{q} y_i (q + 1 - i)$ coeficiente de Gini

m= ingreso medio de la población pobre

Así como la razón H indica la proporción de la población con ingresos inferiores al de la línea de pobreza, la razón de brecha en el ingreso mide la proporción en que su ingreso se aparta del correspondiente a la línea de pobreza. Mientras que la razón H es insensible al grado de pobreza de la persona, la brecha de ingreso I lo es al número de personas involucradas.
Por otra parte, el coeficiente de Gini es una medida de la desigualdad en la distribución del ingreso (en este caso referida al sector pobre de la población) y que puede tomar valores entre [0,1], en donde el primero representa perfecta igualdad en el ingreso y 1 total desigualdad.

Al igual que el coeficiente de Gini, el índice de pobreza se encuentra dentro del intervalo [0,1], con P = 0 cuando la totalidad de la población tiene un ingreso superior a z y P = 1, si todos tienen un ingreso nulo. En la práctica, P nunca será igual a 1, debido a que existen necesidades de subsistencia (de tal manera que para cada i: $y_i > 0$), así como porque inclusive en las economías muy pobres, para algunos i: $y_i > z$.
Cuando todos los individuos de la población pobre tienen el mismo ingreso, i.e.: G = 0, entre mas bajo sea el ingreso de este sector, P se aproximará al valor de H y entre mayor sea la proporción de los pobres, P se aproximará al valor de I.

En nuestro caso (n = 306 observaciones), el valor mínimo (0.0356, mínima pobreza) corresponde al municipio de San Nicolás de los Garza en el estado de Nuevo León y el máximo (0.9765 máxima pobreza) al municipio de Chanal en el estado de Chiapas.

2. Índice de Propiedad de Accesorios Electrodomésticos y Equipos Duraderos (IPAD1, IPAD2)

El Censo General de Población 2000, INEGI reporta la tenencia de bienes duraderos, en términos del número de artefactos disponibles en cada municipio en cuanto a los siguientes objetos: radio, televisión, video-casetera, licuadora, refrigerador, lavadora, teléfono, calefacción, automóvil y computadora. Como medida de la riqueza material de la población y como expresión de las "libertades económicas" descritas con anterioridad, se pretende resumir dicha información mediante un número reducido de nuevas variables componentes principales a incluir en los análisis multivariados subsecuentes. Para lo anterior, se obtuvo el cociente de las existencias de cada uno de

los bienes mencionados, entre la población total de cada municipio, utilizando los datos estandarizados como entrada para el ajuste del modelo.

El modelo de componentes principales, permitió explicar 93.6 % de la varianza total de estas diez variables mediante solamente dos nuevas variables no correlacionadas. Los resultados se muestran en la Tabla A2.1.

Análisis de Componentes Principales: ZRADIO, ZTV, ZVC, ZLIC, ZREFRI, ZLAV, ZTEL, ZCAL, ZAUTO y ZCOMP

Tabla A2.1

Eigenvalores de la Matriz de Covarianza
294 casos utilizados, 13 casos contienen datos faltantes

Eigenvalor	8.2935	0.8000
Proporción	0.854	0.082
Acumulada	0.854	0.936

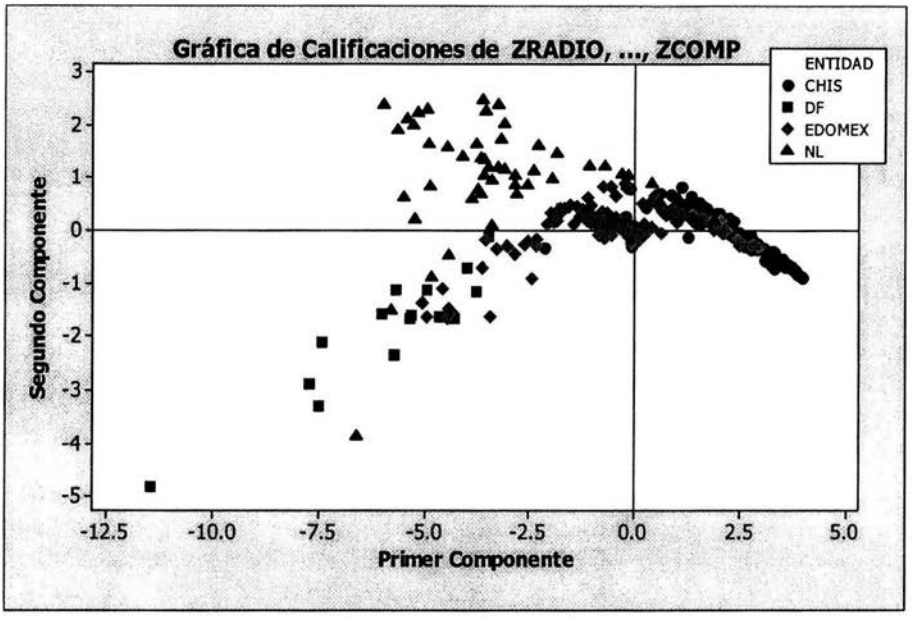
En la Tabla A2.2, se muestran los eigenvectores correspondientes a los dos primeros eigenvalores de la matriz de covarianza. La primera variable componente representa un promedio ponderado en la tenencia de los bienes considerados. Los coeficientes tienen el mismo signo y su magnitud se encuentra dentro de un intervalo relativamente estrecho [0.257, 0.334]. La segunda variable componente principal puede interpretarse como un contraste entre la propiedad de los artículos con mayor tradición en el mercado (i.e.: radio, televisión, licuadora, refrigerador, lavadora y automóvil – signo "+") y con signo contrario, aquellos artículos de más reciente introducción, en donde destacan las computadoras (-0.730), así como accesorios que en algunas poblaciones pueden ser innecesarios (calefacción) o bien (el teléfono), en donde aún existen regiones en la república en las cuales no se dispone de acceso a este servicio.

Tabla A2.2

Variable	PC1	PC2
ZRADIO	-0.318	0.218
ZTV	-0.310	0.274
ZVC	-0.334	-0.229
ZLIC	-0.309	0.168
ZREFRI	-0.326	0.240
ZLAV	-0.331	0.192
ZTEL	-0.329	-0.303
ZCAL	-0.328	-0.166
ZAUTO	-0.313	0.223
ZCOMP	-0.257	-0.730

En los análisis subsecuentes, se utilizan las calificaciones que se obtienen de los dos primeros componentes. La posición de los municipios respecto estas dos componentes se muestra en la grafica A2.1.

Gráfica A2.1



3. Vivienda

Este indicador se determina como el producto de tres variables observables: (1) proporción de viviendas dentro del total de viviendas existentes en los municipios correspondientes a cada una de (2) seis categorías que más adelante se enuncian, a cada una de las cuales se le asigna un valor en el rango [1,6] para designar su calidad (1 mínima calidad, 6 máxima calidad) y (3) proporción de viviendas propias. El indicador VIVIENDA resultante es adimensional y puede tomar cualquier valor dentro del rango [0,6], en donde "0" representa un mínimo de bienestar (viviendas no propias o de mínima calidad) y "6" el máximo correspondiente (viviendas propias de la máxima calidad en cuanto a materiales de construcción). La clasificación de viviendas en cuanto a sus materiales de construcción, se indica a continuación:

Material de Construcción	**Calificación**
Material de desecho	1
Lámina de cartón	2
Lámina de asbesto y metálica	3
Palma, tejamanil y madera	4
Teja	5
Losa de concreto, tabique y ladrillo	6

El índice de vivienda se determina de la siguiente manera:

Sean,

$$\mathbf{P} = \begin{pmatrix} p_{11} & \cdots & p_{16} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{n6} \end{pmatrix}_{n\times 6}$$

P = proporción de viviendas con la característica *i*, $i = 1,6$

$$\mathbf{C} = \begin{pmatrix} c_{11} \\ \vdots \\ c_{61} \end{pmatrix}_{6\times 1}$$

C = Índice de calidad de vivienda

$$\mathbf{V} = \begin{pmatrix} v_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & v_{n6} \end{pmatrix}_{n\times n}$$

V = Proporción de viviendas propias (matriz diagonal)

$$\textbf{VIVIENDA} = \mathbf{V}\times\mathbf{P}\times\mathbf{C}$$

En el caso de los municipios seleccionados (n = 306), los valores mínimo, máximo y medio son 2.0638, 4.2483 y 3.2464 respectivamente. El primero corresponde al municipio de Pichucalco en el estado de Chiapas y el segundo al municipio de Chiconcuac en el Estado de México.

4. Servicios Básicos

Este índice resulta de multiplicar la frecuencia relativa de viviendas con los siguientes servicios: drenaje, servicios sanitarios, energía eléctrica y agua entubada. El valor del índice se encuentra dentro del intervalo [0,1]. En el caso de la muestra de datos seleccionada (n = 306), se tiene un mínimo de 0.0005 en el municipio de Chalchihuitán en el estado de Chiapas y un máximo de 0.9621 en el municipio de San Nicolás de los Garza en el estado de Nuevo León . El valor medio es de 0.41761.

A fin de obtener una mayor sensibilidad en la medición, se exploró, como posible variable alternativa el producto de la sumatoria de la unidad a cada una de las proporciones resultantes. Esta última variable tiene una correlación de Pearson de 0.984 con la primera, por lo que se considera que son equivalentes.

Anexo 3

Matrices de Residuos con Tres y Cuatro Factores

Matriz MRES3

0.548618	0.049425	-0.092850	0.0233189	-0.084920	0.085018	-0.046572
0.049425	0.115713	-0.065259	-0.0173131	-0.013070	-0.006694	-0.092276
-0.092850	-0.065259	0.205069	0.0113058	0.106846	-0.037564	0.072393
0.023319	-0.017313	0.011306	0.0554434	-0.013402	-0.026309	0.013610
-0.084920	-0.013070	0.106846	-0.0134021	0.282607	-0.011771	0.051569
0.085018	-0.006694	-0.037564	-0.0263090	-0.011771	0.229007	0.002933
-0.046572	-0.092276	0.072393	0.0136100	0.051569	0.002933	0.206710
-0.075834	-0.038797	0.005572	-0.0445885	-0.065239	0.048088	0.047819
0.158742	0.038461	0.024277	0.0132082	0.042104	0.028999	-0.071499
0.044615	0.057750	-0.029536	0.0030187	-0.022369	-0.024592	-0.066205
-0.198117	-0.012370	-0.030062	-0.0111200	0.044347	0.064504	0.099867
-0.047423	-0.013552	-0.015338	-0.0197592	0.040102	0.082312	0.049035
-0.023137	0.089652	-0.061693	-0.0284397	0.073374	0.007318	-0.041216
-0.090957	-0.013850	0.003226	-0.0079814	0.056185	-0.017905	0.054461
-0.037150	-0.022711	0.021122	0.0037710	0.043100	-0.038223	0.039123
-0.114849	0.028982	-0.062042	-0.0240063	0.010302	-0.017588	-0.029612
0.207770	0.068111	-0.120321	-0.0244131	-0.056334	0.118950	-0.080669
0.068701	0.006124	0.050448	-0.0076473	-0.048202	0.058720	-0.007111
0.109292	0.065814	-0.047379	-0.0205068	-0.114912	0.091922	-0.055610
0.040092	0.070372	-0.068702	-0.0576664	-0.005227	0.058205	-0.065487
0.053461	0.017765	0.025150	0.0138467	-0.000979	-0.025558	-0.027365

-0.075834	0.158742	0.044615	-0.198117	-0.047423	-0.023137	-0.090957
-0.038797	0.038461	0.057750	-0.012370	-0.013552	0.089652	-0.013850
0.005572	0.024277	-0.029536	-0.030062	-0.015338	-0.061693	0.003226
-0.044588	0.013208	0.003019	-0.011120	-0.019759	-0.028440	-0.007981
-0.065239	0.042104	-0.022369	0.044347	0.040102	0.073374	0.056185
0.048088	0.028999	-0.024592	0.064504	0.082312	0.007318	-0.017905
0.047819	-0.071499	-0.066205	0.099867	0.049035	-0.041216	0.054461
0.484720	0.015021	-0.099092	0.054194	0.003407	-0.042714	0.027890
0.015021	0.334215	0.033770	-0.093949	-0.005281	-0.069873	-0.018098
-0.099092	0.033770	0.143701	-0.049594	-0.084690	0.037131	-0.043185
0.054194	-0.093949	-0.049594	0.350244	0.095903	0.170917	0.085116
0.003407	-0.005281	-0.084690	0.095903	0.280057	0.067447	0.024140
-0.042714	-0.069873	0.037131	0.170917	0.067447	0.340729	0.040498
0.027890	-0.018098	-0.043185	0.085116	0.024140	0.040498	0.228849
0.020711	-0.031440	-0.037850	0.021019	0.011676	-0.007029	0.034845
-0.001417	-0.050197	0.002676	-0.008875	0.053228	0.047209	0.045813
-0.062012	0.043812	0.088715	-0.011860	-0.063689	-0.097901	-0.097610
-0.084264	0.066249	0.032405	-0.007625	-0.044709	-0.008703	-0.101052
-0.032347	0.059153	0.060078	-0.032597	-0.085907	-0.093993	-0.132605
0.019037	-0.022463	0.070007	0.028591	0.047855	0.123418	-0.014334
-0.066587	-0.004381	0.061409	-0.020218	-0.071313	0.038438	-0.018182

-0.0371505	-0.114849	0.207770	0.068701	0.109292	0.040092	0.053461
-0.0227105	0.028982	0.068111	0.006124	0.065814	0.070372	0.017765
0.0211216	-0.062042	-0.120321	0.050448	-0.047379	-0.068702	0.025150
0.0037710	-0.024006	-0.024413	-0.007647	-0.020507	-0.057666	0.013847
0.0430995	0.010302	-0.056334	-0.048202	-0.114912	-0.005227	-0.000979
-0.0382227	-0.017588	0.118950	0.058720	0.091922	0.058205	-0.025558
0.0391228	-0.029612	-0.080669	-0.007111	-0.055610	-0.065487	-0.027365
0.0207112	-0.001417	-0.062012	-0.084264	-0.032347	0.019037	-0.066587
-0.0314405	-0.050197	0.043812	0.066249	0.059153	-0.022463	-0.004381
-0.0378502	0.002676	0.088715	0.032405	0.060078	0.070007	0.061409
0.0210190	-0.008875	-0.011860	-0.007625	-0.032597	0.028591	-0.020218
0.0116765	0.053228	-0.063689	-0.044709	-0.085907	0.047855	-0.071313
-0.0070286	0.047209	-0.097901	-0.008703	-0.093993	0.123418	0.038438
0.0348452	0.045813	-0.097610	-0.101052	-0.132605	-0.014334	-0.018182
0.0567860	-0.003313	-0.028291	-0.034572	-0.041394	-0.038173	-0.005264
-0.0033128	0.152942	-0.087386	-0.053135	-0.037648	0.077532	-0.060584

-0.0282911	-0.087386	0.684211	0.067554	0.098374	0.151404	0.075002
-0.0345720	-0.053135	0.067554	0.243438	0.070258	-0.014262	0.028837
-0.0413936	-0.037648	0.098374	0.070258	0.707073	0.078793	0.070633
-0.0381733	0.077532	0.151404	-0.014262	0.078793	0.290289	-0.012302
-0.0052642	-0.060584	0.075002	0.028837	0.070633	-0.012302	0.139952

Matriz MRES4

0.280767	-0.017275	-0.018882	0.0240137	0.015147	0.013920	0.053199
-0.017275	0.099104	-0.046840	-0.0171401	0.011848	-0.024399	-0.067431
-0.018882	-0.046840	0.184642	0.0111140	0.079212	-0.017930	0.044840
0.024014	-0.017140	0.011114	0.0554416	-0.013662	-0.026125	0.013351
0.015147	0.011848	0.079212	-0.0136617	0.245223	0.014790	0.014295
0.013920	-0.024399	-0.017930	-0.0261246	0.014790	0.210135	0.029416
0.053199	-0.067431	0.044840	0.0133512	0.014295	0.029416	0.169546
0.008054	-0.017908	-0.017594	-0.0448061	-0.096579	0.070355	0.016572
0.051545	0.011767	0.053881	0.0134862	0.082152	0.000545	-0.031569
-0.046811	0.034983	-0.004288	0.0032558	0.011787	-0.048860	-0.032150
-0.071037	0.019275	-0.065156	-0.0114496	-0.003129	0.098236	0.052531
0.042374	0.008809	-0.040136	-0.0199921	0.006554	0.106147	0.015586
0.053692	0.108784	-0.082910	-0.0286390	0.044672	0.027711	-0.069834
0.036790	0.017961	-0.032052	-0.0083127	0.008460	0.016004	0.006876
0.013865	-0.010007	0.007033	0.0036387	0.024041	-0.024681	0.020120
-0.047163	0.045837	-0.080734	-0.0241819	-0.014985	0.000378	-0.054824
-0.085608	-0.004946	-0.039303	-0.0236521	0.053269	0.041077	0.028611
-0.026848	-0.017669	0.076835	-0.0073995	-0.012506	0.033358	0.028480
-0.147666	0.001827	0.023581	-0.0198403	-0.018915	0.023716	0.040104
-0.034108	0.051895	-0.048212	-0.0574740	0.022493	0.038509	-0.037849
-0.013383	0.001120	0.043609	0.0140201	0.023993	-0.043300	-0.002466

0.008054	0.051545	-0.046811	-0.071037	0.042374	0.053692	0.036790
-0.017908	0.011767	0.034983	0.019275	0.008809	0.108784	0.017961
-0.017594	0.053881	-0.004288	-0.065156	-0.040136	-0.082910	-0.032052
-0.044806	0.013486	0.003256	-0.011450	-0.019992	-0.028639	-0.008313
-0.096579	0.082152	0.011787	-0.003129	0.006554	0.044672	0.008460
0.070355	0.000545	-0.048860	0.098236	0.106147	0.027711	0.016004
0.016572	-0.031569	-0.032150	0.052531	0.015586	-0.069834	0.006876
0.458447	0.048594	-0.070458	0.014394	-0.024717	-0.066776	-0.012119
0.048594	0.291313	-0.002820	-0.043089	0.030658	-0.039125	0.033029
-0.070458	-0.002820	0.112494	-0.006217	-0.054039	0.063356	0.000419
0.014394	-0.043089	-0.006217	0.289952	0.053299	0.134466	0.024508
-0.024717	0.030658	-0.054039	0.053299	0.249952	0.041689	-0.018688
-0.066776	-0.039125	0.063356	0.134466	0.041689	0.318691	0.003855
-0.012119	0.033029	0.000419	0.024508	-0.018688	0.003855	0.167922
0.004734	-0.011024	-0.020437	-0.003185	-0.005426	-0.021662	0.010514
-0.022616	-0.023108	0.025780	-0.040988	0.030536	0.027794	0.013531
0.029871	-0.073603	-0.011424	0.127331	0.034667	-0.013750	0.042312
-0.054339	0.028009	-0.000209	0.037708	-0.012676	0.018703	-0.055481
0.048129	-0.043685	-0.027631	0.089315	0.000239	-0.020288	-0.010053
0.042275	-0.052159	0.044680	0.063794	0.072731	0.144702	0.021054
-0.045652	-0.031132	0.038593	0.011495	-0.048904	0.057611	0.013698

0.0138646	-0.047163	-0.085608	-0.026848	-0.147666	-0.034108	-0.013383
-0.0100069	0.045837	-0.004946	-0.017669	0.001827	0.051895	0.001120
0.0070335	-0.080734	-0.039303	0.076835	0.023581	-0.048212	0.043609
0.0036387	-0.024182	-0.023652	-0.007399	-0.019840	-0.057474	0.014020
0.0240408	-0.014985	0.053269	-0.012506	-0.018915	0.022493	0.023993
-0.0246814	0.000378	0.041077	0.033358	0.023716	0.038509	-0.043300
0.0201203	-0.054824	0.028611	0.028480	0.040104	-0.037849	-0.002466
0.0047338	-0.022616	0.029871	-0.054339	0.048129	0.042275	-0.045652
-0.0110235	-0.023108	-0.073603	0.028009	-0.043685	-0.052159	-0.031132
-0.0204370	0.025780	-0.011424	-0.000209	-0.027631	0.044680	0.038593
-0.0031847	-0.040988	0.127331	0.037708	0.089315	0.063794	0.011495
-0.0054265	0.030536	0.034667	-0.012676	0.000239	0.072731	-0.048904
-0.0216616	0.027794	-0.013750	0.018703	-0.020288	0.144702	0.057611
0.0105144	0.013531	0.042312	-0.055481	-0.010053	0.021054	0.013698

0.0470696	-0.016204	0.027586	-0.016374	0.007547	-0.024041	0.007467
-0.0162044	0.135838	-0.013249	-0.028989	0.027286	0.096282	-0.043692
0.0275860	-0.013249	0.362872	-0.037101	-0.183074	0.070133	0.001788
-0.0163737	-0.028989	-0.037101	0.209354	-0.021405	-0.040731	0.004992
0.0075469	0.027286	-0.183074	-0.021405	0.460565	0.007611	0.006508
-0.0240411	0.096282	0.070133	-0.040731	0.007611	0.269735	-0.030819
0.0074669	-0.043692	0.001788	0.004992	0.006508	-0.030819	0.123271