UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

"ANÁLISIS ESTADÍSTICO DE CONGLOMERADOS. ALGUNAS APLICACIONES."

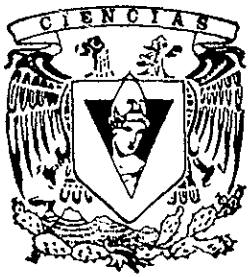
T E S I S

QUE PARA OBTENER EL TÍTULO DE

A C T U A R I O

P R E S E N T A :

ALVARO NOSEDAL SÁNCHEZ



DIRECTOR DE TESIS: ACT. FRANCISCO SÁNCHEZ VILLARREAL

VNIVERSIDAD NACIONAL
AVENOMA DE
MEXICO

MAT. MARGARITA ELVIRA CHÁVEZ CANO
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"Análisis Estadístico de Conglomerados. Algunas Aplicaciones."

realizado por Alvaro Nosedal Sánchez

con número de cuenta 9231316-9 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Act. Francisco Sánchez Villarreal
Propietario	Mat. Julio César Cedillo Sánchez
Propietario	Act. Jaime Vázquez Alamilla
Suplente	Act. Ma. Susana Barrera Ocampo
Suplente	Act. Victor Manuel Solís Najera

Consejo Departamental de Matemáticas

M. en C. José Antonio Flores Díaz

AGRADECIMIENTOS.

Antes que a nadie, quiero agradecerles a mis padres, Ma. Dolores Sánchez Góngora y Víctor Fernando Nosedal Blancas, el amor y el apoyo incondicional que siempre me han brindado. Este pequeño logro, como todos los que he conseguido, simplemente refleja lo grandes que ellos son. ¡No tengo palabras para agradecerles todo lo que han hecho por mí!

A mis hermanos, Benjamin y Jenaro, mis primeros compañeros de juegos y mis fieles amigos en la vida, gracias por su amor, por sus consejos, por sus vivencias, ... ¡En fin!, ¡Gracias por existir!

A mi novia, Erika, por haberme acompañado a lo largo de la realización de este proyecto que felizmente llega a su término. ¡Gracias por tu amor, por tu paciencia y tu comprensión!
Espero que me sigas amando y teniendo paciencia para alcanzar, juntos, al menos otro logro similar a este. ¡Te amo!

A todos mis profesores, por sus conocimientos y por haberme enseñado a amar a la UNAM y , en particular, a la Facultad de Ciencias. Quiero mencionar a algunos de ellos, aquellos que dejaron una huella en mí, no sólo por su calidad académica sino humana.

Al Fís. Mario Cruz Terán, por haber sido el mejor profesor que tuve durante el bachillerato y, de manera muy especial, por la ayuda que me brindó cuando recién había ingresado a la Facultad de Ciencias. ¡Muchas Gracias!

Al Act. Javier Fernández García, quien durante mi primer día en la Facultad dijo: "Su estancia en esta Facultad cambiará sus vidas", ¡Muchas gracias por haber sido parte de ese cambio!. Y gracias por esa pasión que tienes por la enseñanza de las Matemáticas.

Al Mat. Julio César Cedillo Sánchez, un hombre muy brillante, muy sencillo y a quien tengo el privilegio de llamar amigo. Gracias por haber sido mi tutor durante una buena parte de mi estancia en la Facultad, por haber compartido tus conocimientos conmigo, por tus observaciones a este trabajo y , sobre todo, ¡Gracias por tu amistad!.

Al Dr. Miguel Angel García Alvarez, no sólo por ser un magnífico profesor sino, por ser un hombre que a pesar de su enorme valía no ha perdido su humildad. ¡Gracias por su ejemplo!

Al Act. Francisco Sánchez Villlarreal, porque sus cátedras me hicieron darme cuenta de lo interesante que es la Estadística y sus Aplicaciones, por compartir con sus alumnos sus experiencias profesionales y, por supuesto, por haberme regalado mucho de su tiempo y atención en la elaboración de este trabajo de tesis. ¡Muchas Gracias Profesor!

A la Act. Yazmín Bárcenas Orozco, por haber impartido un estupendo curso de Análisis Multivariado pero sobre todo por inculcar en sus alumnos el deseo de superarse y la conciencia de lo mucho que le debemos a nuestra Universidad. ¡Gracias!

Al Act. Jaime Vázquez Alamilla, gracias por las clases tan interesantes que impartiste durante el curso de Análisis Multivariado y por haberte dado tiempo de revisar el presente trabajo. ¡Muchas Gracias!

A la Act. Susana Barrera Ocampo, por haber revisado esta tesis y, en especial, por haberme proporcionado y enseñado el manejo del software que hizo posible la inclusión de las gráficas presentes en este trabajo. ¡Gracias!

Al Act. Víctor Manuel Solís Najera, por haber dedicado parte de su tiempo a revisar este escrito. ¡Gracias!.

Ahora, quiero mencionar a aquellos que compartieron conmigo su existencia durante su estancia en la Facultad de Ciencias y le dieron un sabor característico y especial a esa época de mi vida.
En primer lugar, a uno de mis mejores amigos, Pablito. Gracias por haber compartido conmigo tu forma de entender las Matemáticas, por tu apoyo en los momentos difíciles, por tu ayuda,... ¡Muchas gracias por tu amistad!

A mis amigos : Beto "El Master", Juan "El Amiguito", Erik, Mando, Isa, Edgar, Gerardo, Malú, Bety, Jana, Hugo, Maritza, Yazmín, Jacke, Isa "Tonchez", Mónica, Diana, Aline, Julián, Charlie, César "Ubaldo", Jéssica y Eliel.

A toda mi familia y, en particular a mi tío, el Sr. Rubén Sánchez Góngora, quien,de manera muy minuciosa, revisó el presente escrito durante todas las etapas de su desarrollo. ¡Gracias!

A Dios, por haberme dado la dicha de conocer a todas estas personas. ¡Muchas Gracias!

Finalmente, quiero dedicar de manera muy especial este trabajo a mis abuelos:
A mi abuelita, Doña Sabina Blancas Espejel (q.e.p.d.), a mi abuelita, Doña Consuelo Góngora Hernández (q.e.p.d.), a mi abuelito, Don Efrén Nosedal Montes (q.e.p.d.) y a mi abuelito, Don Delfino Sánchez Zarco, por haberme inculcado siempre el amor al estudio.

Alvaro Nosedal Sánchez.
Agosto del 2000.

ÍNDICE.

CAPÍTULO 1. FUNDAMENTOS MATEMÁTICOS.

CAPÍTULO 2. EL ANÁLISIS DE CONGLOMERADOS Y SU RELACIÓN CON LAS FUNCIONES DISTANCIA.

CAPÍTULO 3. TÉCNICAS JERÁRQUICAS DE CONGLOMERADO.

INTRODUCCIÓN.

Comenzaremos con una pregunta: ¿Qué es el análisis multivariado?. De manera muy general, podríamos decir que es un conjunto de técnicas que nos sirven para estudiar una población o una muestra de individuos con características que se reflejan por medio de más de una variable. Dependiendo del objetivo que se persiga, es decir, de la información que queremos obtener de estos datos, será la técnica que habremos de aplicar.
A mi parecer; el Análisis Estadístico Multivariado no ha recibido la difusión que se merece. Este trabajo intenta ser un vehículo de divulgación de una de las técnicas que lo componen.

Dentro de las técnicas que constituyen el Análisis Estadístico Multivariado cabe mencionar las siguientes: Análisis de Conglomerados, Análisis Multidimensional de Escalas, Análisis de Regresión, Análisis de Discriminante, Análisis de Factores, etc.

La técnica de la que nos habremos de ocupar en el presente trabajo es el ANÁLISIS ESTADÍSTICO DE CONGLOMERADOS. Por ser una técnica de carácter exploratorio tiene bases matemáticas muy sólidas pero propiedades estocásticas inexistentes. El Análisis de Conglomerados es utilizado principalmente para generar hipótesis, no para probarlas.

Ahora bien, ¿Qué tipo de problemas podemos resolver empleando esta técnica? Bueno, imaginemos que tenemos un conjunto de p individuos, de ellos registramos m características (de donde, resulta una matriz de pxm). La pregunta que nos hacemos respecto a estos datos es la siguiente. ¿Existirán grupos bien definidos dentro de este conjunto de datos?. Este es el tipo de problema que se resuelve utilizando Análisis de Conglomerados.

Ahora podrían surgir otras preguntas ¿Para qué sirve encontrar esos grupos?, ¿Qué utilidad tiene encontrar dichos grupos?. Trataremos de dar una respuesta satisfactoria a

estas interrogantes a lo largo del presente trabajo. Sin embargo, como un adelanto a esto, podemos decir que una de las habilidades que necesitó el hombre desde su aparición en la Tierra fue: clasificar. Clasificó las plantas en comestibles, medicinales, venenosas, etc. Es decir, organizó una población en grupos; grupos que ,por cierto, eran desconocidos en un principio. A lo largo del presente haremos mención de las aplicaciones que se le han dado a estas técnicas en muy distintas áreas del conocimiento.

En la mayoría de las ramas de la ciencia, la clasificación es un tarea fundamental. En Biología, por ejemplo, la clasificación de los organismos ha sido un problema de interés desde que se iniciaron las primeras investigaciones.
En un sentido más amplio, un esquema de clasificación puede representar un método conveniente de organización para un conjunto grande de datos. También, dicha clasificación agiliza y facilita la búsqueda de información.

Existen otras aplicaciones donde dicha clasificación pueda servir para propósitos mucho más interesantes. La Medicina nos da un buen ejemplo. Para entender y tratar una enfermedad tiene que ser clasificada y la clasificación tendrá dos propósitos. El primero será: la predicción, separar las enfermedades que requieren diferentes tratamientos, el segundo será dar una base para la investigación, las causas de diferentes tipos de la enfermedad.
Los dos propósitos pueden llevar a clasificaciones distintas y una variedad de clasificaciones alternativas para el mismo conjunto de individuos u objetos siempre habrá de existir.
Los seres humanos, por ejemplo, pueden ser clasificados con respecto al estrato económico al que pertenecen en grupos como: clase baja, clase media y clase alta, o pueden ser clasificados por su consumo anual de alcohol en. bajo, medio y alto. Obviamente clasificaciones diferentes no pueden asignar individuos particulares en grupos iguales. Lo que podemos decir es que, algunas clasificaciones pueden ser más útiles que otras.

Un ejemplo muy claro sería el siguiente. Tenemos un conjunto de libros, primero los clasificamos de acuerdo a su contenido como: diccionarios, novelas, biografías, etc. Ahora procedemos a clasificarlos de acuerdo al color de la pasta del libro. Es claro que la primera clasificación que hicimos es más útil que la primera.
El punto importante sugerido por estos dos ejemplos es que cualquier clasificación es una división de los objetos o individuos en estudio en grupos, dicha clasificación basada en un conjunto particular de reglas (o variables) y no es ni falsa ni verdadera y debe ser juzgada principalmente por la utlilidad de sus resultados.

El contenido de los 5 capítulos que conforman el presente es, a grandes rasgos, el siguiente:

En el primer capítulo hacemos una exposición de los conceptos de la Matemática que nos permitirán desarrollar el Análisis Estadístico de Conglomerados.

En el capítulo 2 del presente, determinamos el tipo de problemas que se pueden abordar con esta Técnica del Análisis Estadístico Multivariado, damos a conocer algunas de las aplicaciones que se han hecho en distintos campos del conocimiento y comenzamos a establecer las relaciones existentes entre los conceptos desarrollados en el capítulo anterior y los del actual.

Una vez expuestos los fundamentos matemáticos necesarios y teniendo claro el tipo de problema que podemos resolver, nos damos a la tarea de dar a conocer una clase particular de procedimientos que en conjunto forman el cuerpo del Análisis Estadístico de Conglomerados, las técnicas o procedimientos jerárquicos. Ellos constituirán nuestro objeto de estudio en el capítulo 3.

En el cuarto capítulo, nos ocuparemos de aplicar algunas de las técnicas expuestas en el capítulo anterior a un conjunto de datos pequeño. De esta manera nos será más transparente la forma en que actúan las técnicas.

En la parte final de trabajo, aplicaremos un Análisis más sofisticado a un conjunto de datos formado por 100 ciudades. Dicho Análisis tendrá como objetivo obtener una estructura de grupos y establecer un pérfil para cada uno de ellos. Este análisis abarcará todo el capítulo 5.

CAPÍTULO 1.

FUNDAMENTOS MATEMÁTICOS.

Introducción.

En este capítulo nos ocuparemos de reconstruir los fundamentos matemáticos necesarios para el Desarrollo del Análisis de Conglomerados.

A primera vista, el capítulo podría parecernos extenso, no lo es tanto. Cada uno de los temas aquí expuestos es prácticamente indispensable para la comprensión de los subsiguientes. Vamos a revisar, de manera muy sucinta, la razón de ser, la finalidad o la necesidad de cada uno de dichos temas.

Estableceremos de manera formal y cuidadosa, la definición y propiedades que debe satisfacer una función para ser llamada distancia. Esto tiene como objeto el poder utilizar más funciones de distancias y no sólo remitirnos a la euclideana. El concepto de distancia es de vital importancia en el Desarrollo del Análisis de Conglomerados ya que nos sirve de medida de la "similitud" o "disimilitud" entre dos elementos en nuestro conjunto de estudio.

Introduciremos los conceptos de valores y vectores propios de una matriz para obtener las condiciones bajo las cuales una matriz es similar a otra y también cuando es diagonalizable. Algunos de estos resultados nos serán útiles para demostrar que una función en particular es distancia (Mahalanobis).

Finalmente, revisamos las formas cuadráticas y llegamos a la extraña coincidencia que la matriz de covarianza puede verse como una forma cuadrática.

Existen diversas formas de introducir en un espacio vectorial el concepto de distancia. Una de ellas, consiste en definirla a partir de una operación conocida como producto interno.

1.1 Producto interno.

Puede decirse que el concepto de producto interno es la generalización, a un espacio vectorial cualquiera, del producto punto en R^n. Cuatro de sus propiedades fundamentales se emplean como los axiomas que debe satisfacer una función dada para ser considerada producto interno, como se enuncia a continuación.

Definición 1.1.
Sea V un espacio vectorial sobre R. Un producto interno en V es una función de $V \times V$ en R que asigna a cada pareja ordenada $(\bar{u}, \bar{v})$ de vectores de V un escalar $\bar{u} \bullet \bar{v}$ perteneciente a R, llamado el producto de $\bar{u}$ y $\bar{v}$, que satisface las siguientes propiedades:

1) $\bar{u} \bullet \bar{v} = \bar{v} \bullet \bar{u}$
2) $\bar{u} \bullet (\bar{v} + \bar{w}) = \bar{u} \bullet \bar{v} + \bar{u} \bullet \bar{w}$
3) $(\alpha \bar{u}) \bullet \bar{v} = \alpha(\bar{u} \bullet \bar{v})$; $\alpha \in R$
4) $\bar{u} \bullet \bar{u} > 0$ si $\bar{u} \neq \bar{0}$

Cuando la función $\bullet$ de la definición es de V x V en R se dice que es un producto interno real.
En este caso también, el campo sobre el que se define el espacio V es el de los números reales.
El ejemplo más conocido de producto interno es, seguramente, el producto escalar en R^n, el cual está definido por

$$\bar{x} \bullet \bar{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

donde $\bar{x} = (x_1, x_2, \ldots, x_n)$, $\bar{y} = (y_1, y_2, \ldots, y_n) \in R^n$.
Otros ejemplos de productos internos son los siguientes:
1) En el espacio V de las funciones reales de variable real, continuas en el intervalo (a, b), la función $\bullet$ definida por

$$f \bullet g = \int_a^b f(x)g(x)dx$$

es un producto interno real.
2) En el espacio V de las matrices de mxn con elementos en R, el siguiente es un producto interno real

$A \bullet B = tr(A^t B)$

Como consecuencia de la definición cualquier producto interno tiene las propiedades que se enuncian a continuación .

Teorema 1.1.

Sea V un espacio vectorial sobre R y sea $\bullet$ un producto interno en V ; entonces, $\forall \bar{u}, \bar{v} \in V$ y $\alpha \in R$:

1) $\bar{u} \bullet (\alpha \bar{v}) = \alpha(\bar{u} \bullet \bar{v})$

2) $\bar{u} \bullet \bar{u} \in R$

3) $\bar{0} \bullet \bar{u} = 0 = \bar{u} \bullet \bar{0}$

4) $\bar{u} \bullet \bar{u} = 0 \Leftrightarrow \bar{u} = \bar{0}$

Demostración.

1) $\bar{u} \bullet (\alpha \bar{v}) = (\alpha \bar{v}) \bullet \bar{u}$ por 1) de la definición de producto interno.

$= \alpha(\bar{v} \bullet \bar{u})$ por 3) de la definición de producto interno.

$= \alpha(\bar{u} \bullet \bar{v})$ por 1) de la definición de producto interno.

2) $\bar{u} \bullet \bar{u} \in R$

Esta propiedad nos indica que $\bar{u} \bullet \bar{u}$ es siempre un número real, es decir, esta operación no es "cerrada". El resultado de multiplicar dos vectores no es otro vector, sino un número real (escalar o elemento del campo). Como nosotros vamos a utilizar sólo el campo de los reales, esta propiedad es, de hecho, consecuencia de la definición de producto interno.

3) Para demostrar esta propiedad vamos a echar mano de las siguientes propiedades de los espacios vectoriales:

a) $\bar{u} + \bar{0} = \bar{0} + \bar{u} = \bar{u}$

b) $\bar{u} + (-\bar{u}) = \bar{0}$, es decir, $\bar{u} - \bar{u} = \bar{0}$

Bien, teniendo estas propiedades en mente podemos expresar al vector cero como sigue:

$\bar{0} = \bar{0} + \bar{0}$ por la propiedad del vector cero

$\bar{0} \bullet \bar{u} = (\bar{0} + \bar{0}) \bullet \bar{u}$ por 2) de la definición de producto interno

$\bar{0} \bullet \bar{u} = \bar{0} \bullet \bar{u} + \bar{0} \bullet \bar{u}$

Como $\bar{0} \bullet \bar{u}$ es un número tenemos lo siguiente

$$\bar{0} \bullet \bar{u} - \bar{0} \bullet \bar{u} = \bar{0} \bullet \bar{u}$$

$$0 = \bar{0} \bullet \bar{u}$$

4) $\Leftarrow$)Supongamos que $\bar{u} = \bar{0}$ y utilizando el resultado anterior tenemos que

$$\bar{u} \bullet \bar{u} = 0$$

y en consecuencia

$$\bar{u} \bullet \bar{u} = 0 \Leftrightarrow \bar{u} = \bar{0}$$ q.e.d.

1.1.1 Norma y distancia.

La idea de magnitud (o tamaño) de un vector se introduce formalmente en un espacio vectorial con el concepto de norma; ésta se puede definir a partir del producto interno como sigue :

Definición 1.2.

Sea V un espacio vectorial sobre R y sea $\bullet$ un producto interno en V. Se llama norma de $\bar{v} \in V$, y se representa con $\| \bar{v} \|$, al número real no negativo definido por

$$\| \bar{v} \| = (\bar{v} \bullet \bar{v})^{1/2}$$

De esta manera, la norma es una función de V en el conjunto de los números reales no negativos.

Empleando el concepto de norma, la desigualdad de Cauchy-Schwarz podrá expresarse como

$$|\bar{u} \bullet \bar{v}|^2 \leq \| \bar{u} \|^2 \| \bar{v} \|^2$$

o bien

$$|\bar{u} \bullet \bar{v}| \leq \| \bar{u} \| \| \bar{v} \|$$

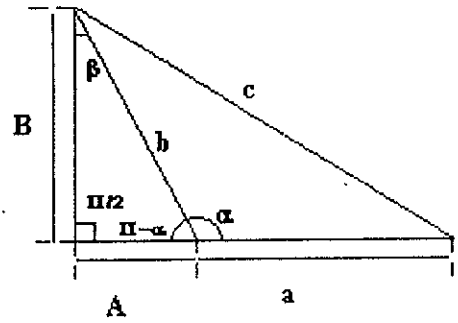
donde se nota que, para cualquier par de vectores, el tamaño del producto es menor o igual que el producto de los tamaños.

Cabe hacer notar que, en la expresión anterior, el producto del lado izquierdo es un producto de vectores; mientras que el del lado derecho es un producto de números reales.

Como se sigue de la definición de norma, la norma de un vector depende del producto interno que se haya elegido. Así, un mismo vector puede tener diferentes normas.

Esta situación, aparentemente contradictoria, es similar al hecho de asignar diferentes números a una misma magnitud, dependiendo de la escala o unidad de medida que se emplee.

Necesitamos demostrar el Teorema de Cauchy-Schwarz, para poder hacerlo, debemos recordar algunos resultados previamente.



Ley de los Cosenos.

Esta ley nos ayuda a calcular la longitud de cualquier lado de un triángulo no rectángulo, siempre y cuando, conozcamos la longitud de 2 de sus lados y del ángulo que forman entre ellos.
Entonces suponemos conocidas las longitudes de los lados a y b, y el ángulo que forman. Queremos conocer la longitud del lado restante, c.
La idea de la deducción de dicha ley, consiste en convertir este triángulo no rectángulo, en un triángulo rectángulo
Primero, sabemos que los ángulos interiores de un triángulo suman π radianes (180 grados). De donde si nos fijamos en la figura:

$$\beta + \frac{\pi}{2} + \pi - \alpha = \pi$$

$$\beta + \frac{\pi}{2} - \alpha = 0$$

$$\beta = \alpha - \frac{\pi}{2}$$

Ahora, volvamos a ver con cuidado nuestra figura. De ahí afirmamos lo siguiente:

$$\operatorname{sen}\beta = \frac{A}{b} \Rightarrow A = b\operatorname{sen}\beta$$

$$\cos\beta = \frac{B}{b} \Rightarrow B = b\cos\beta$$

Ya conocemos los 3 lados del triángulo rectángulo que queríamos, si además utilizamos el Teorema de Pitágoras tendremos:

$$C^2 = (A+a)^2 + B^2$$

$$C^2 = A^2 + 2Aa + a^2 + B^2$$

$$C^2 = b^2\operatorname{sen}^2\beta + 2ab\operatorname{sen}\beta + a^2 + b^2\cos^2\beta$$

$$C^2 = b^2\operatorname{sen}^2\beta + b^2\cos^2\beta + 2ab\operatorname{sen}\beta + a^2$$

$$C^2 = b^2(\operatorname{sen}^2\beta + \cos^2\beta) + 2ab\operatorname{sen}\beta + a^2$$

$$C^2 = b^2 + 2ab\operatorname{sen}\beta + a^2$$

$$C^2 = a^2 + b^2 + 2ab\operatorname{sen}(\alpha - \frac{\pi}{2})$$

Utilizando la fórmula del seno de una suma tenemos que esta última expresión es igual a:

$$\operatorname{sen}(\alpha - \frac{\pi}{2}) = \operatorname{sen}\alpha\cos\frac{\pi}{2} - \operatorname{sen}\frac{\pi}{2}\cos\alpha$$

$$\operatorname{sen}(\alpha - \frac{\pi}{2}) = -\cos\alpha$$

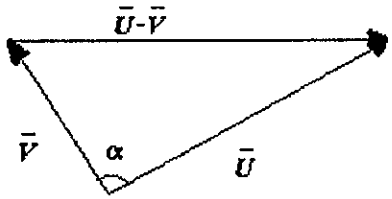
Por tanto, tenemos que la expresión que obtuvimos para C^2 podemos escribirla como:

$$C^2 = a^2 + b^2 + 2ab(-\cos\alpha)$$

$$C^2 = a^2 + b^2 - 2ab\cos\alpha$$

Bien, ahora para que podamos hacer una demostración más corta y elegante de la desigualdad de Cauchy-Schwarz, debemos mostrar una forma alternativa del producto interno

Primero, hagamos un dibujo para tener una idea geométrica del razonamiento que vamos a seguir.



Utilizando nuestra figura y recordando la ley de los cosenos tenemos que:

$$\left\|\bar{u}-\bar{v}\right\|^2 = \left\|\bar{u}\right\|^2 + \left\|\bar{v}\right\|^2 - 2\left\|\bar{u}\right\|\left\|\bar{v}\right\|\cos\alpha \qquad (1$$

Ahora utilicemos la definición de norma de un vector:

$$\left\|\bar{u}-\bar{v}\right\|^2 = (\bar{u}-\bar{v})\bullet(\bar{u}-\bar{v}) = \bar{u}\bullet\bar{u}-\bar{u}\bullet\bar{v}-\bar{v}\bullet\bar{u}+\bar{v}\bullet\bar{v}$$

$$\left\|\bar{u}-\bar{v}\right\|^2 = \left\|\bar{u}\right\|^2 - 2\bar{u}\bullet\bar{v} + \left\|\bar{v}\right\|^2 \qquad (2$$

Igualando 1) y 2)

$$\left\|\bar{u}\right\|^2 + \left\|\bar{v}\right\|^2 - 2\left\|\bar{u}\right\|\left\|\bar{v}\right\|\cos\alpha = \left\|\bar{u}\right\|^2 - 2\bar{u}\bullet\bar{v} + \left\|\bar{v}\right\|^2$$

$$-2\left\|\bar{u}\right\|\left\|\bar{v}\right\|\cos\alpha = -2\bar{u}\bullet\bar{v} \qquad \text{dividiendo entre } -2$$

$$\left\|\bar{u}\right\|\left\|\bar{v}\right\|\cos\alpha = \bar{u}\bullet\bar{v}$$

Finalmente, podemos demostrar la desigualdad de Cauchy-Schwarz

Teorema 1.2. Desigualdad de Cauchy-Schwarz.
Sea V un espacio vectorial sobre R y sea $\bullet$ un producto interno en V; entonces $\forall\bar{u},\bar{v}\in V$

$$\left|\bar{u}\bullet\bar{v}\right|^2 \le \left\|\bar{u}\right\|^2\left\|\bar{v}\right\|^2 \Leftrightarrow \left|\bar{u}\bullet\bar{v}\right| \le \left\|\bar{u}\right\|\left\|\bar{v}\right\|$$

Demostración.
Primero, recordemos la igualdad que acabamos de obtener:

$$\bar{u}\bullet\bar{v} = \left\|\bar{u}\right\|\left\|\bar{v}\right\|\cos\alpha \qquad \text{aplicándole a esta expresión el valor absoluto}$$

$$\left|\bar{u} \bullet \bar{v}\right| = \left\|\bar{u}\right\|\left\|\bar{v}\right\|\left|\cos\alpha\right| \leq \left\|\bar{u}\right\|\left\|\bar{v}\right\| \qquad \text{porque } \left|\cos\alpha\right| \leq 1\text{, entonces}$$

$$\left|\bar{u} \bullet \bar{v}\right| \leq \left\|\bar{u}\right\|\left\|\bar{v}\right\| \qquad \text{q.e.d.}$$

En el siguiente teorema se enuncian las propiedades fundamentales que satisface toda norma, independientemente del producto interno del que provenga.

Teorema 1.3.

Si V es un espacio vectorial con producto interno, entonces $\forall \bar{u}, \bar{v} \in V$ y $\alpha \in R$:

1) $\| \bar{v} \| > 0$

2) $\| \bar{v} \| = 0$ si y sólo si $\bar{v} = \bar{0}$

3) $\| \alpha \bar{v} \| = | \alpha | \, \| \bar{v} \|$

4) $\| \bar{u} + \bar{v} \| \leq \| \bar{u} \| + \| \bar{v} \|$

Demostración.

Sólo vamos a demostrar la propiedad 4) por que nos parece de mayor interés. Por cierto, a esta propiedad se le conoce como la desigualdad del triángulo.

Por la definición de norma de un vector tenemos.

$$\left\|\bar{u}+\bar{v}\right\|^2 = (\bar{u}+\bar{v}) \bullet (\bar{u}+\bar{v}) = \bar{u} \bullet \bar{u} + \bar{u} \bullet \bar{v} + \bar{v} \bullet \bar{u} + \bar{v} \bullet \bar{v}$$

por la definición de norma y por la conmutatividad del producto punto esta expresión queda como

$$\left\|\bar{u}+\bar{v}\right\|^2 = \left\|\bar{u}\right\|^2 + 2\bar{u} \bullet \bar{v} + \left\|\bar{v}\right\|^2 \qquad \text{por la igualdad que obtuvimos del producto punto}$$

$$\left\|\bar{u}+\bar{v}\right\|^2 = \left\|\bar{u}\right\|^2 + 2\left\|\bar{u}\right\|\left\|\bar{v}\right\|\cos\alpha + \left\|\bar{v}\right\|^2 \leq \left\|\bar{u}\right\|^2 + 2\left\|\bar{u}\right\|\left\|\bar{v}\right\| + \left\|\bar{v}\right\|^2$$

porque $\left|\cos\alpha\right| \leq 1$, entonces, si factorizamos

$$\left\|\bar{u}+\bar{v}\right\|^2 \leq \left(\left\|\bar{u}\right\| + \left\|\bar{v}\right\|\right)^2 \qquad \text{si extraemos raíz cuadrada tenemos}$$

$$\left\|\bar{u}+\bar{v}\right\| \leq \left\|\bar{u}\right\| + \left\|\bar{v}\right\| \qquad \text{q.e.d.}$$

Empleando el concepto de norma, podemos introducir en un espacio vectorial el concepto de distancia entre vectores de la siguiente manera.

Definición 1.3.

Sea V un espacio vectorial con producto interno, y sean $\bar{u}, \bar{v} \in V$. Se llama distancia de $\bar{u}$ a $\bar{v}$, y se representa con $d(\bar{u}, \bar{v})$, al número real definido por

$$d(\bar{u}, \bar{v}) = \| \bar{v} - \bar{u} \|$$

De esta manera , la distancia es una función de $V \times V$ en el conjunto de los números reales no negativos y, como puede demostrarse, tiene las propiedades que se enuncian a continuación.

Teorema 1.4.

Si V es un espacio vectorial con producto interno, entonces $\forall \bar{u}, \bar{v}, \bar{w} \in V$ ·

1) $d(\bar{u}, \bar{v}) \geq 0$
2) $d(\bar{u}, \bar{v}) = 0$ si y sólo si $\bar{u} = \bar{v}$
3) $d(\bar{u}, \bar{v}) = d(\bar{v}, \bar{u})$
4) $d(\bar{u}, \bar{w}) \leq d(\bar{u}, \bar{v}) + d(\bar{v}, \bar{w})$

Demostración.

1) Sean $\bar{u}$ y $\bar{v} \in V$

$d(\bar{u}, \bar{v}) = \| \bar{v} - \bar{u} \|$

caso 1) Si $\bar{u} = \bar{v}$ entonces $\bar{v} - \bar{u} = \bar{0}$ de donde $\| \bar{v} - \bar{u} \| = 0$

caso 2) Si $\bar{v} \neq \bar{u}$ entonces $\| \bar{v} - \bar{u} \| > 0$

2) $\Rightarrow$) $d(\bar{u}, \bar{v}) = 0$ entonces $\| \bar{v} - \bar{u} \| = 0$ esto implica que $\bar{v} - \bar{u} = \bar{0}$ por tanto $\bar{u} = \bar{v}$

$\Leftarrow$) $\bar{u} = \bar{v}$ entonces $\bar{v} - \bar{u} = \bar{0}$ obteniendo la norma de esta última expresión

$\| \bar{v} - \bar{u} \| = \| \bar{0} \| = d(\bar{u}, \bar{v}) = 0$

3) $d(\bar{u}, \bar{v}) = \| \bar{v} - \bar{u} \| = \| -1 (\bar{u} - \bar{v}) \| = | 1 | \| \bar{u} - \bar{v} \| = \| \bar{u} - \bar{v} \| = d(\bar{v}, \bar{u})$

4) $d(\bar{u}, \bar{w}) \leq d(\bar{u}, \bar{v}) + d(\bar{v}, \bar{w})$

$d(\bar{u}, \bar{w}) = \| \bar{w} - \bar{u} \| = \| \bar{w} + \bar{v} - \bar{v} - \bar{u} \| = \| (\bar{v} - \bar{u}) + (\bar{w} - \bar{v}) \| \leq \| \bar{v} - \bar{u} \| + \| \bar{w} - \bar{v} \| =$

$d(\bar{u}, \bar{v}) + d(\bar{v}, \bar{w})$

1.2 Valores propios y vectores propios.

A manera de introducción a las eigenestructuras matriciales utilizaremos primero un ejemplo sencillo. Supongan que tenemos una matriz de transformación de 2x2:

$$A = \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix}$$

Ahora, supongamos que queremos encontrar un vector imagen

$$x^* = \begin{bmatrix} x^*_1 \\ x^*_2 \end{bmatrix}$$

que tenga la misma (o, tal vez con sentido opuesto) dirección que el vector preimagen

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

si sólo estamos interesados en mantener la dirección, entonces x^* puede ser representado por

$$x^* = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

donde λ denota a un escalar. Esto significa, que podemos alargar o acortar x, la preimagen, siempre y cuando x^* este en la misma dirección de x.
Si x es transformada bajo A en $x^* = \lambda x$, entonces establecemos lo siguiente:
Los vectores, que bajo una cierta transformación son mapeados en ellos mismos o en múltiplos de ellos mismos, son llamados vectores invariantes bajo esa transformación.
Se sigue, entonces, que tales vectores satisfacen la relación:

$$Ax = \lambda x$$

donde, como habiamos dicho λ es un escalar.

Para ilustrar, suponga que lo intentamos con el vector $x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ para saber si este vector es invariante bajo A:

$$Ax = \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 2 \end{bmatrix} = x^*$$

Y vemos que no es el caso. Porque no cumple lo establecido para vectores invariantes, ya que los componentes de x* no son múltiplos del vector x. Sin embargo, intentemos con el vector $z = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$

$$Az = \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} = 2 \begin{bmatrix} 3 \\ 3 \end{bmatrix} = z^*$$

En el caso de z, si tenemos un vector invariante. Más aún, si probamos con cualquier vector cuyos componentes cumplan la relación de 1:1, que también satisfacen lo establecido para los vectores invariantes.
Donde $\lambda = 2$ es la constante de proporcionalidad.

Ahora bien, ¿Sólo los vectores de la forma $x_i = \begin{bmatrix} k \\ k \end{bmatrix}$ son vectores invariantes?.

Intentemos con otro vector, por ejemplo, $w = \begin{bmatrix} 5 \\ -4 \end{bmatrix}$:

$$A\,w = \begin{bmatrix} -3 & 5 \\ 4 & -2 \end{bmatrix}\begin{bmatrix} 5 \\ -4 \end{bmatrix} = \begin{bmatrix} -35 \\ 28 \end{bmatrix} = -7\begin{bmatrix} 5 \\ -4 \end{bmatrix} = w^*$$

Observamos que también la forma $x_j = \begin{bmatrix} 5k \\ -4k \end{bmatrix}$, funciona también. Pero ¿Habrá otros ? Como veremos, no hay otros que no sean de la forma

$$x_i = \begin{bmatrix} k \\ k \end{bmatrix} \qquad \text{o} \qquad x_j = \begin{bmatrix} 5k \\ -4k \end{bmatrix}$$

Abordemos ahora el problema desde el siguiente punto de vista. Utilicemos la ecuación matricial

$$A\,x = \lambda x$$

que también podemos ver de la siguiente forma (restando λx de ambos lados):

$A\,x - \lambda x = 0$ equivalente a
$A\,x - \lambda I x = 0$

donde I representa a la matriz identidad. Ahora podemos factorizar x para obtener:

$$(A - \lambda I)\,x = 0$$

Una solución trivial es, por supuesto, $x = \bar{0}$. Sin embargo, por lo general, estaremos interesados en soluciones no triviales; esto es, soluciones en las cuales $x \neq \bar{0}$.
Por el momento, permitanme igualar a $(A - \lambda I)$ con B y examinemos que necesitamos de B para el caso en que x no es una solución trivial (i.e. que tengan elementos distintos de cero). La expresión de arriba puede escribirse como

$$Bx = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

lo cual puede ser escrito como el siguiente conjunto de ecuaciones lineales:

$$ax_1 + bx_2 = 0$$
$$cx_1 + dx_2 = 0$$

Después, multiplicamos la primera ecuación por d, la segunda por -b, sumamos ambas ecuaciones y tenemos

$$(ad - bc)\, x_1 = 0$$

Repetimos el procedimiento anterior, multiplicamos la primera ecuación por -c, la segunda por a, y sumando las ecuaciones, y obtenemos

$$(ad - bc)\, x_2 = 0$$

Entonces, si $x_1 \neq 0$ o $x_2 \neq 0$, tenemos la situación en la cual necesariamente $(ad - bc) = 0$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = |B| = |A - \lambda I| = 0$$

Esto último, nos indica que el determinante de $(A - \lambda I)$ debe ser cero si queremos que x $\neq \bar{0}$.

Antes de seguir adelante vamos a definir formalmente los conceptos de vector propio y valor propio

Definición 1.4.

Sea V un espacio vectorial sobre R y sea $A : V \rightarrow V$ un operador lineal. Si existe un vector $\bar{v} \in V ; \bar{v} \neq \bar{0}$ tal que

$$A(\bar{v}) = \lambda \bar{v}$$

para algún escalar $\lambda \in K$, entonces se dice que λ es un valor propio de A y que $\bar{v}$ es un vector propio de **A** correspondiente a λ.

Ahora regresemos a nuestra expresión original de $(A - \lambda I)$, el razonamiento anterior nos dice que necesitamos que el determinante de esta matriz sea cero. Podemos escribir esta matriz de manera explícita como:

$$(A - \lambda I) =$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} = \lambda^2 - \lambda(a_{11} + a_{22}) + a_{11}a_{22} - a_{12}a_{21} = 0$$

Esta última expresión es llamada ecuación característica de la matriz A.

Las raíces de esta ecuación, las cuales se denotan por λ_i, son llamadas valores propios (eigenvalores), y sus vectores asociados x_i se obtienen al sustituir las raíces en

$$(A - \lambda I)\, x_i = 0$$

y resolviendo para x_i. Estos vectores x_i son llamados vectores propios (eigenvectores). Son los vectores invariantes bajo la transformación hecha por la matriz A.

Como indicamos arriba, la ecuación característica de A se define como:

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

Ese mismo determinante se define como

$$|A - \lambda I|$$

y es llamado la función característica de A.

Debe ser claro, entonces, que sólo las matrices cuadradas tienen "eigenestructuras", ya que como sabemos sólo las matrices cuadradas tienen determinantes.

Una de las propiedades más útiles de los vectores propios es la que establece el siguiente teorema:

Teorema 1.5.

Sea $A: V \to V$ un operador lineal y sean $\overline{v_1}$, $\overline{v_2}$, , $\overline{v_k}$ vectores propios de A correspondientes a los valores $\lambda_1, \lambda_2, \ldots, \lambda_k$ respectivamente. Si $\lambda_i \neq \lambda_j \forall i \neq j$, entonces el conjunto ($\overline{v_1}$, $\overline{v_2}$, . . . , $\overline{v_k}$) es linealmente independiente.

Demostración.

Se hará por inducción matemática

1) Si k = 1

considerando de la definición de vector propio , $\overline{v_1} \neq \overline{0}$

$$\alpha \overline{v_1} = \overline{0} \Rightarrow \alpha_1 = 0$$

por lo que $\overline{v_1}$ es linealmente independiente.

2)Suponemos que el conjunto ($\overline{v_1}$, $\overline{v_2}$, . . . , $\overline{v_k}$) es linealmente independiente para k = n

o, equivalentemente:

$$\alpha_1 \overline{v_1} + \alpha_2 \overline{v_2} + \ldots + \alpha_n \overline{v_n} = \overline{0} \Rightarrow \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0 \quad .. \quad (1)$$

Para k = n+1 la ecuación de dependencia es

$$\beta_1 \overline{v_1} + \beta_2 \overline{v_2} + \ldots + \beta_n \overline{v_n} + \beta_{n+1} \overline{v_{n+1}} = \overline{0} \qquad .....(2)$$

aplicando el operador A a ambos miembros tenemos, por la linealidad de A,

$$\beta_1 A(\overline{v_1}) + \beta_2 A(\overline{v_2}) + \ldots + \beta_n A(\overline{v_n}) + \beta_{n+1} A(\overline{v_{n+1}}) = \overline{0}$$

por ser este conjunto de vectores eigenvectores tenemos que

$$\beta_1 \lambda_1(\overline{v_1}) + \beta_2 \lambda_2(\overline{v_2}) + \ldots + \beta_n \lambda_n(\overline{v_n}) + \beta_{n+1} \lambda_{n+1}(\overline{v_{n+1}}) = \overline{0} \qquad ..... \quad .(3)$$

Por otra parte, multiplicando 2 por λ_{n+1} se obtiene

$$\beta_1 \lambda_{n+1}(\overline{v_1}) + \beta_2 \lambda_{n+1}(\overline{v_2}) + \ldots + \beta_n \lambda_{n+1}(\overline{v_n}) + \beta_{n+1} \lambda_{n+1}(\overline{v_{n+1}}) = \overline{0} \qquad .....(4)$$

Restando 4) de 3) y factorizando

$$\beta_1(\lambda_1 - \lambda_{n+1})(\overline{v_1}) + \beta_2(\lambda_2 - \lambda_{n+1})(\overline{v_2}) + \ldots + \beta_n(\lambda_n - \lambda_{n+1})(\overline{v_n}) = \overline{0}$$

$$\beta_1(\lambda_1 - \lambda_{n+1}) = \alpha_1 = 0$$
$$\beta_2(\lambda_2 - \lambda_{n+1}) = \alpha_2 = 0 \quad \text{así hasta}$$
$$\beta_n(\lambda_n - \lambda_{n+1}) = \alpha_n = 0$$

además, $\lambda_1, \lambda_2, \ldots, \lambda_{n+1}$ son diferentes por hipótesis, es decir,

$\lambda_i - \lambda_{n+1} \neq 0, i = 1,2,\ldots,n$

luego $\beta_1 = \beta_2 = \ldots = \beta_n = 0$

de 2, $\beta_{n+1}\bar{v}_{n+1} = 0$ y recordando de la definición de vector propio que $\bar{v}_{n+1} \neq 0$ se tiene que $\beta_{n+1} = 0$ luego, $\forall k \in N$

$\beta_1(\bar{v}_1) + \beta_2(\bar{v}_2) + \ldots + \beta_k(\bar{v}_n) = \bar{0}$ implica que

$$\beta_1 = \beta_2 = \ldots = \beta_k = 0$$

y entonces el conjunto ($\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_k$) es linealmente independiente.

q.e.d.

1.3 Diagonalización y matrices similares.

Debemos saber que la representación matricial de una transformación lineal depende de las bases que se elijan para el dominio y el codominio.

En el caso particular de los operadores suele utilizarse la misma base para el dominio y el codominio. Dicha base ha sido seleccionada, hasta ahora, buscando simplificar al máximo la obtención de la matriz asociada; sin embargo, esto no necesariamente coincide con simplificar al máximo la forma de la matriz que se obtiene (su aspecto).

Que la matriz asociada sea de forma sencilla ofrece ciertas ventajas pues, además de que permite identificar más fácilmente la información contenida en ella, su manejo algebraico se simplifica. Entre los tipos más sencillos de matrices están las diagonales. Por ello, a continuación se tratará el problema de encontrar una representación diagonal para un operador, referida a una misma base del dominio y del codominio .
Con esta última restricción no siempre es posible encontrar una representación diagonal para cualquier operador. Las condiciones bajo las cuales existe tal representación así como las características de ésta, son el objetivo de nuestra búsqueda.

Definición 1.5.

Sea $A = [a_{ij}]$ una matriz de nxn con elementos en R. Se dice que A es una matriz diagonal si $a_{ij} = 0$ para $i \neq j$, y se representa con

$diag(a_{11}, a_{22}, \ldots, a_{nn})$

Definición 1.6.

Sea $B = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\}$ una base de un espacio vectorial V sobre R, y sea $\bar{x} \in V$. Si

$$\bar{x} = \alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2 + \ldots + \alpha_n \bar{v}_n$$

los escalares $\alpha_1, \alpha_2, \ldots, \alpha_n$ se llaman coordenadas de $\bar{x}$ en la base B; y el vector de R^n

$$(\bar{x})_B = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$$

se llama vector de coordenadas de $\bar{x}$ en la base B.

Teorema 1.6.

Sean V y W dos espacios vectoriales con $\dim V = n$ y $\dim W = m$; y sean $A = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\}$ y $B = \{\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_m\}$ bases de V y W, respectivamente.
Si $T: V \to W$ es una transformación lineal, existe una y sólo una matriz $M_B{}^A(T)$,de mxn, tal que

$$M_B{}^A(T)(\bar{v})_A = (T(\bar{v}))_B, \forall \bar{v} \in V$$

Las n columnas de dicha matriz son los vectores
$(T(\bar{v}_1))_B, (T(\bar{v}_2))_B, \ldots, (T(\bar{v}_n))_B$

Demostración.

Sean V y W dos espacios vectoriales con $\dim V = n$ y $\dim W = m$, y sean $A = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\}$ y $B = \{\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_m\}$ dos bases cualesquiera de V y W, respectivamente.
Consideremos una transformación lineal $T: V \to W$.
Si $\bar{v}$ es un vector cualquiera de V, éste puede expresarse unívocamente en términos de la base A como

$$\bar{v} = \alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2 + \ldots + \alpha_n \bar{v}_n \qquad 1)$$

por lo que

$$T(\bar{v}) = T(\alpha_1 \bar{v}_1 + \alpha_2 \bar{v}_2 + \ldots + \alpha_n \bar{v}_n) \quad \text{y como } T \text{ es lineal}$$

$$T(\bar{v}) = \alpha_1 T(\bar{v}_1) + \alpha_2 T(\bar{v}_2) + \ldots + \alpha_n T(\bar{v}_n) \qquad 2)$$

Por otra parte, los vectores $T(\bar{v}_1), T(\bar{v}_2), \ldots, T(\bar{v}_n)$ son elementos de W, por lo que pueden expresarse unívocamente en términos de la base B, es decir

$$T(\bar{v}_1) = a_{11} \bar{w}_1 + a_{21} \bar{w}_2 + \ldots + a_{m1} \bar{w}_m$$

$$T(\bar{v}_2) = a_{12} \bar{w}_1 + a_{22} \bar{w}_2 + \ldots + a_{m2} \bar{w}_m \qquad \text{así hasta}$$

$$T(\bar{v}_n) = a_{1n}\bar{w}_1 + a_{2n}\bar{w}_2 + \ldots + a_{mn}\bar{w}_m \qquad 3)$$

llevando estas expresiones a 2 se tiene que

$$T(\bar{v}) = \alpha_1(a_{11}\bar{w}_1 + a_{21}\bar{w}_2 + \ldots + a_{m1}\bar{w}_m) +$$
$$\alpha_2(a_{12}\bar{w}_1 + a_{22}\bar{w}_2 + \ldots + a_{m2}\bar{w}_m) + \ldots + \alpha_n(a_{1n}\bar{w}_1 + a_{2n}\bar{w}_2 + \ldots + a_{mn}\bar{w}_m)$$

de donde

$$T(\bar{v}) = (\alpha_1 a_{11} + \alpha_2 a_{12} + \ldots + \alpha_n a_{1n})\bar{w}_1 + (\alpha_1 a_{21} + \alpha_2 a_{22} + \ldots + \alpha_n a_{2n})\bar{w}_2 + \ldots +$$
$$(\alpha_1 a_{m1} + \alpha_2 a_{m2} + \ldots + \alpha_n a_{mn})\bar{w}_m$$

Por la definición 6, las sumas encerradas entre paréntesis son las coordenadas de $T(\bar{v})$ en la base B, por lo que podemos escribir

$$\left[T(\bar{v})\right]_B = \begin{bmatrix} \alpha_1 a_{11} + \alpha_2 a_{12} + \ldots + \alpha_n a_{1n} \\ \alpha_1 a_{21} + \alpha_2 a_{22} + \ldots + \alpha_n a_{2n} \\ \ldots \\ \alpha_1 a_{m1} + \alpha_2 a_{m2} + \ldots + \alpha_n a_{mn} \end{bmatrix}$$

o bien como producto

$$\left[T(\bar{v})\right]_B = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & .. \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \ldots \\ \alpha_n \end{bmatrix}.$$

De 1, los escalares $\alpha_1, \alpha_2, \ldots, \alpha_n$ son las coordenadas de $\bar{v}$ en la base A, por lo que

$$\left[T(\bar{v})\right]_B = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} (\bar{v})_A$$

y finalmente

$$\left[T(\bar{v})\right]_B = M_B{}^A(T)(\bar{v})_A$$

donde

$$M_B{}^A(T) = a_{ij}\, i = 1,2,\ldots,m;\, j = 1,2,\ldots,n$$

además, de 3

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{m1} \end{bmatrix} = (T(\overline{v_1}))_B \quad , \begin{bmatrix} a_{12} \\ a_{22} \\ \cdots \\ a_{m2} \end{bmatrix} = (T(\overline{v_2}))_B \ , \ldots , \begin{bmatrix} a_{1n} \\ a_{2n} \\ \cdots \\ a_{mn} \end{bmatrix} = (T(\overline{v_n}))_B$$

por lo que las columnas de $M_B{}^A(T)$ son los vectores de coordenadas, en la base B, de las imágenes de los elementos que integran la base A.

La unicidad no la demostraremos pero, daremos como un hecho que ésta se da. A quien este interesado en conocer la demostración lo remitimos a los textos de álgebra lineal citados en la bibliografía.

Teorema 1.7.
Sean V un espacio vectorial de dimensión n, $A: V \to V$ un operador lineal. Existe una matriz diagonal asociada a A, referida a una base, si y sólo si existe una base de V formada por vectores característicos de A. En tal caso, la matriz asociada a A, referida a esta base, es una matriz diagonal cuyos elementos d_{ii} son los valores característicos correspondientes.

Demostración.
Supongamos primero que existe una base de V formada por vectores característicos de A. Esto es,
sea
$B = (\overline{v_1}, \overline{v_2}, \ldots, \overline{v_n})$
una base de V tal que
$A(\overline{v_i}) = \lambda_i \overline{v_i}$, para $i = 1,2,\ldots,n$
entonces, por la definición 6
$(A(\overline{v_1}))_B = (\lambda_1, 0, \ldots, 0)^T$
$(A(\overline{v_2}))_B = (0, \lambda_2, \ldots, 0)^T$ así hasta
$(A(\overline{v_n}))_B = (0, 0, \ldots, \lambda_n)^T$
por lo que, del teorema 6

$$M_B{}^B(A) = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix}$$

y A tiene una representación matricial diagonal, formada con los valores propios de A correspondientes a los vectores propios que constituyen la base de B.

Recíprocamente, supongamos que A tiene una representación diagonal
$D = \text{diag}\ (d_{11}, d_{22}, \ldots, d_{nn})$
referida a la base
$B = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\}$
Entonces, por teorema 6 sabemos que

$(A(\bar{v}_1))_B = (d_{11}, 0, \ldots, 0)^T$
$(A(\bar{v}_2))_B = (0, d_{22}, \ldots, 0)^T$ así hasta
$(A(\bar{v}_n))_B = (0, 0, \ldots, d_{nn})^T$
y en consecuencia, por definición 6
$A(\bar{v}_i) = d_{ii}\bar{v}_i$, para $i = 1,2,\ldots,n$
por lo que los vectores $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n$ son valores característicos de A correspondientes a los valores $d_{11}, d_{22}, \ldots, d_{nn}$.

q.e.d.

Teorema 1.8.
Si dos matrices M y N representan al mismo operador lineal, entonces existe una matriz no singular P tal que $N = P^{-1}MP$.

Demostración.
Sea V un espacio vectorial de dimensión n y sea $T: V \to V$ un operador lineal.
Si M y N son matrices asociadas al operador T referidas a las bases A y B, respectivamente, del espacio V; entonces, para cualquier vector $\bar{v} \in V$:

$$(T(\bar{v}))_A = M(\bar{v})_A \quad \ldots\ldots\ldots 1)$$

$$(T(\bar{v}))_B = N(\bar{v})_B \quad \ldots\ldots\ldots\ldots 2)$$

Por otra parte, si P es la matriz de transición de la base B a la base A, se tiene que

$$P(\bar{v})_B = (\bar{v})_A$$

sustituyendo esta expresión en 1 se llega a

$$(T(\bar{v}))_A = MP(\bar{v})_B$$

Como la matriz de transición es no singular $\exists P^{-1}$ por lo que, de la expresión anterior

$$P^{-1}(T(\bar{v}))_A = P^{-1}MP(\bar{v})_B \quad \ldots\ldots\ldots\ldots 3)$$

Además, como P^{-1} es la matriz de transición de la base A a la base B, se tiene que

$$P^{-1}(T(\bar{v}))_A = (T(\bar{v}))_B$$

por lo que, de 2

$$P^{-1}(T(\bar{v}))_A = N(\bar{v})_B \text{.....4)}$$

entonces de 3 y 4

$$N(\bar{v})_B = P^{-1}MP(\bar{v})_B; \forall \bar{v} \in V$$

y en consecuencia

$$N = P^{-1}MP$$

como se quería.

q.e.d.

Definición 1.7.
Dos matrices A y B de nxn, son similares si existe una matriz no singular C tal que $B = C^{-1}AC$
Se tiene entonces el siguiente teorema

Teorema 1.9.
Dos matrices representan al mismo operador lineal si y sólo si son similares.

Teorema 1.10.
Una matriz A de nxn es similar a una matriz diagonal D si y sólo si existe un conjunto linealmente independiente formado por n vectores propios de A. En tal caso, existe una matriz no singular P tal que $D = P^{-1}AP$ donde D es una matriz diagonal cuyos elementos d_{ii} son los valores propios de A, y P tiene como columnas a n vectores propios de A correspondientes a dichos valores.

Demostración.
Tomando en cuenta que cualquier matriz A de nxn puede considerarse como la representación matricial de un operador lineal T en cierta base, el último teorema es una consecuencia inmediata de los teoremas 6 y 9, y de la definición 5, salvo la afirmación de que P tiene como columnas a n vectores propios de A.
Esto último se demuestra a continuación.
Sea $S = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\}$ un conjunto linealmente independiente formado por n vectores propios de A, correspondientes a los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$, y sea P una matriz cuyas columnas son dichos vectores; esto es

$$P = \left[\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\right]$$

entonces

$$AP = A\left[\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\right]$$

$$AP = \left[A\bar{v}_1, A\bar{v}_2, \ldots, A\bar{v}_n\right]$$

$$AP = \left[\lambda_1\bar{v}_1, \lambda_2\bar{v}_2, \ldots, \lambda_n\bar{v}_n\right]$$

$$= \left[\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\right] \text{ diag } (\lambda_1, \lambda_2, \ldots, \lambda_n)$$

$$AP = P \text{ diag } (\lambda_1, \lambda_2, \ldots, \lambda_n)$$

Como el conjunto S es linealmente independiente, P es no singular y existe P^{-1}. En consecuencia

$$P^{-1}AP = P^{-1}Pdiag(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

$$P^{-1}AP = diag(\lambda_1, \lambda_2, \ldots, \lambda_n) \qquad \text{q.e.d.}$$

1.4 Algunas matrices importantes en el análisis estadístico multivariado.

Antes de abordar el tema de las funciones cuadráticas presentaremos algunas matrices y la forma en que éstas se calculan. Esto obedece a la utilidad que tienen para el desarrollo de las técnicas presentadas y también por la forma en que se relacionan algunas de ellas con el siguiente tema.
En primer lugar, tenemos la siguiente matriz:

1) Suma de cuadrados y productos cruzados.

$B = A'A$, la cual es llamada también el menor producto momento (de A), que da como resultado una matriz simétrica B. Las entradas de la diagonal de la matriz B denotan las sumas de cuadrados por renglones de cada variable, y los elementos que están fuera de la diagonal denotan las sumas de productos cruzados por renglón.

Sea $A = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ . & .. & .. & .. \\ x_{m1} & x_{m2} & .. & x_{mn} \end{bmatrix}$ y $A' = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{m1} \\ x_{12} & x_{22} & \ldots & x_{m2} \\ .. & .. & .. & .. \\ x_{1n} & x_{2n} & .. & x_{mn} \end{bmatrix}$ de donde

$$B = A'A = \begin{bmatrix} \sum_{i=1}^{m} x_{i1}^2 & \sum_{i=1}^{m} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{m} x_{i1}x_{in} \\ \sum_{i=1}^{m} x_{i2}x_{i1} & \sum_{i=1}^{m} x_{i2}^2 & \cdots & \sum_{i=1}^{m} x_{i2}x_{in} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{m} x_{in}x_{i1} & \sum_{i=1}^{m} x_{in}x_{i2} & \cdots & \sum_{i=1}^{m} x_{in}^2 \end{bmatrix}$$

2) Matriz corregida por media (SSCP).

Podemos también expresar las sumas de cuadrados y de productos cruzados como desviaciones respecto a las medias. La matriz de sumas de cuadrados y productos cruzados de corrección de medias es frecuentemente llamada la matriz SSCP (sums of squares and cross products) y es expresada en notación como:

$$S = A'A - \frac{1}{m}(A'I)(I'A)$$

donde I denota un vector de "unos" y "m" denota el número de observaciones. El último término en el lado derecho de la ecuación representa el término de corrección y es una generalización de la forma escalar usual para calcular sumas de cuadrados respecto a la media:

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{m}$$

donde $x = X - \overline{X}$, es decir, x denota la forma desviación de la media.

Vamos a obtener S de manera explícita:

$$(A'I) = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} x_{i1} \\ \sum_{i=1}^{m} x_{i2} \\ \cdots \\ \sum_{i=1}^{m} x_{in} \end{bmatrix}$$

$$(I'A) = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ .. & .. & .. & .. \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_{i=1}^{m} x_{i1} & \sum_{i=1}^{m} x_{i2} & \dots & \sum_{i=1}^{m} x_{in} \end{bmatrix}$$

$$(A'I)(I'A) = \begin{bmatrix} \left(\sum_{i=1}^{m} x_{i1}\right)^2 & \sum_{i=1}^{m} x_{i1} \sum_{i=1}^{m} x_{i2} & \dots & \sum_{i=1}^{m} x_{i1} \sum_{i=1}^{m} x_{in} \\ \sum_{i=1}^{m} x_{i2} \sum_{i=1}^{m} x_{i1} & \left(\sum_{i=1}^{m} x_{i2}\right)^2 & \dots & \sum_{i=1}^{m} x_{i2} \sum_{i=1}^{m} x_{in} \\ .. & .. & .. & .. \\ \sum_{i=1}^{m} x_{in} \sum_{i=1}^{m} x_{i1} & \sum_{i=1}^{m} x_{in} \sum_{i=1}^{m} x_{i2} & \dots & \left(\sum_{i=1}^{m} x_{in}\right)^2 \end{bmatrix}$$

$$-\frac{1}{m}(A'I)(I'A) = \begin{bmatrix} -\frac{1}{m}\left(\sum_{i=1}^{m} x_{i1}\right)^2 & -\frac{1}{m}\sum_{i=1}^{m} x_{i1} \sum_{i=1}^{m} x_{i2} & \dots & -\frac{1}{m}\sum_{i=1}^{m} x_{i1} \sum_{i=1}^{m} x_{in} \\ -\frac{1}{m}\sum_{i=1}^{m} x_{i2} \sum_{i=1}^{m} x_{i1} & -\frac{1}{m}\left(\sum_{i=1}^{m} x_{i2}\right)^2 & \dots & -\frac{1}{m}\sum_{i=1}^{m} x_{i2} \sum_{i=1}^{m} x_{in} \\ .. & .. & .. & .. \\ -\frac{1}{m}\sum_{i=1}^{m} x_{in} \sum_{i=1}^{m} x_{i1} & -\frac{1}{m}\sum_{i=1}^{m} x_{in} \sum_{i=1}^{m} x_{i2} & \dots & -\frac{1}{m}\left(\sum_{i=1}^{m} x_{in}\right)^2 \end{bmatrix}$$

$$S = A'A - \frac{1}{m}(A'I)(I'A) =$$

$$= \begin{bmatrix} \sum_{i=1}^{m} x_{i1}^2 - \frac{1}{m}\left(\sum_{i=1}^{m} x_{i1}\right)^2 & \sum_{i=1}^{m} x_{i1}x_{i2} - \frac{1}{m}\sum_{i=1}^{m} x_{i1} \sum_{i=1}^{m} x_{i2} & \dots & \sum_{i=1}^{m} x_{i1}x_{in} - \frac{1}{m}\sum_{i=1}^{m} x_{i1} \sum_{i=1}^{m} x_{in} \\ \sum_{i=1}^{m} x_{i2}x_{i1} - \frac{1}{m}\sum_{i=1}^{m} x_{i2} \sum_{i=1}^{m} x_{i1} & \sum_{i=1}^{m} x_{i2}^2 - \frac{1}{m}\left(\sum_{i=1}^{m} x_{i2}\right)^2 & \dots & \sum_{i=1}^{m} x_{i2}x_{in} - \frac{1}{m}\sum_{i=1}^{m} x_{i2} \sum_{i=1}^{m} x_{in} \\ .. & .. & .. & .. \\ \sum_{i=1}^{m} x_{in}x_{i1} - \frac{1}{m}\sum_{i=1}^{m} x_{in} \sum_{i=1}^{m} x_{i1} & \sum_{i=1}^{m} x_{in}x_{i2} - \frac{1}{m}\sum_{i=1}^{m} x_{in} \sum_{i=1}^{m} x_{i2} & \dots & \sum_{i=1}^{m} x_{in}^2 - \frac{1}{m}\left(\sum_{i=1}^{m} x_{in}\right)^2 \end{bmatrix}$$

La matriz S, como podemos observar, es una matriz simétrica.

De manera alternativa, si las medias por columnas se sustraen de A desde el principio, resulta la matriz corregida por media A_d, entonces

$$S = A_d' A_d$$

la matriz A_d puede obtenerse de A de la siguiente manera:

$$A_d = A - I\bar{a}'$$

donde I es un vector de unos y $\bar{a}'$ es un vector de medias. El vector de medias se obtiene, a su vez de la siguiente manera:

$$\bar{a}' = I' A / m$$

donde en este caso I' es un vector renglón .

3) Matriz de covarianza.

La matriz de covarianza se obtiene de la matriz SSCP (corregida por media) simplemente dividiendo cada entrada de S por el escalar m, el tamaño de la muestra. Esto es :

$$C = \frac{1}{m} S$$

$$= \begin{bmatrix} \frac{1}{m}\sum_{i=1}^{m} x_{i1}^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m} x_{i1}\right)^2 & \frac{1}{m}\sum_{i=1}^{m} x_{i1}x_{i2} - \frac{1}{m^2}\sum_{i=1}^{m} x_{i1}\sum_{i=1}^{m} x_{i2} & \cdots & \frac{1}{m}\sum_{i=1}^{m} x_{i1}x_{in} - \frac{1}{m^2}\sum_{i=1}^{m} x_{i1}\sum_{i=1}^{m} x_{in} \\ \frac{1}{m}\sum_{i=1}^{m} x_{i2}x_{i1} - \frac{1}{m^2}\sum_{i=1}^{m} x_{i2}\sum_{i=1}^{m} x_{i1} & \frac{1}{m}\sum_{i=1}^{m} x_{i2}^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m} x_{i2}\right)^2 & \cdots & \frac{1}{m}\sum_{i=1}^{m} x_{i2}x_{in} - \frac{1}{m^2}\sum_{i=1}^{m} x_{i2}\sum_{i=1}^{m} x_{in} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{m}\sum_{i=1}^{m} x_{in}x_{i1} - \frac{1}{m^2}\sum_{i=1}^{m} x_{in}\sum_{i=1}^{m} x_{i1} & \frac{1}{m}\sum_{i=1}^{m} x_{in}x_{i2} - \frac{1}{m^2}\sum_{i=1}^{m} x_{in}\sum_{i=1}^{m} x_{i2} & \cdots & \frac{1}{m}\sum_{i=1}^{m} x_{in}^2 - \frac{1}{m^2}\left(\sum_{i=1}^{m} x_{in}\right)^2 \end{bmatrix}$$

Notemos que la covarianza, es meramente un producto cruzado corregido por medias ponderado. Los elementos en la diagonal de C son, por supuesto, varianzas. Y además que la matriz de covarianza es simétrica. (En algunas aplicaciones desearemos obtener un estimador insesgado de la matriz de covarianza poblacional; de ser así, usamos el divisor m-1 en lugar de m).

Una forma alternativa de obtener la matriz de covarianza sería:

$$C = 1/m A_d' A_d$$

4) Matriz de correlación.

La correlación entre dos variables, y y x, se obtiene como:

$$r_{xy} = \frac{\sum yx}{\sqrt{\sum y^2}\sqrt{\sum x^2}}$$

donde y y x ya incluyen la desviación con respecto a la media (como más arriba).

No es de sorprender que, R la matriz de correlación esté relacionada con S, la matriz SSCP, y C, la matriz de covarianza. Ahora bien, regresemos a S. Las entradas de la diagonal principal de S representan las sumas de cuadrados corregidas por medias de las variables. Si tomamos la raíz cuadrada de estas entradas y ponemos los recíprocos de estas raíces cuadradas en una matriz diagonal tenemos:

$$D = \begin{bmatrix} 1/\sqrt{\sum x_1^2} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\sum x_2^2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/\sqrt{\sum x_n^2} \end{bmatrix}$$

entonces, al premultiplicar y postmultiplicar S por D podemos obtener la matriz de correlación R.

$R = DSD$

la matriz de arriba es la que se deriva de las correlaciones entre cada par de variables. Otra forma de obtener a R es, por lo tanto:

$R = 1/mA'_S A_S$

donde $A_S = A_d D$

1.5 Formas cuadráticas.

En análisis multivariado uno se encuentra con frecuencia que el mapeo de algún vector involucra una función cuadrática más que una lineal *. A primera vista puede parecer sorprendente que el álgebra lineal sea importante para este tipo de situación. Ha llegado el momento de considerar las funciones cuadráticas y, en particular, las formas cuadráticas.

1.5.1 Formas lineales.

Si tenemos un conjunto de variables x_i y un conjunto de coeficientes a_i, una forma lineal puede ser escrita en notación escalar como

$$g(\bar{x}) = a_1x_1 + a_2x_2 + \dots + a_nx_n = \sum_{i=1}^{n} a_i x_i$$

en la cual todas las x_i, como vemos, son de primer grado. En notación vectorial tenemos

$$g(x) = a^T x = (a_1, a_2, \ldots, a_n)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

el cual, por supuesto, da como resultado un escalar, una vez que le asignamos un valor numérico a a y a x.

*Claramente, la idea de varianza de una variable involucra una función cuadrática, y la varianza representa un concepto central en el análisis estadístico.

Ahora, suponga que consideramos un conjunto de muchas formas lineales, con la matriz de coeficientes dada por A y el vector de constantes dado por c. Entonces tenemos

$$Ax = c = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \ldots \\ c_n \end{bmatrix}$$

Este, por supuesto, representa un conjunto de ecuaciones lineales simultáneas. Entonces, una forma lineal es simplemente una función lineal en un conjunto de variables x_i.

1.5.2 Formas bilineales.

Las formas bilineales involucran solamente una pequeña extensión de las formas lineales. Aquí tenemos dos conjuntos de variables x_i y y_j, cada una de primer grado, como se ilustra en :

$$f(x,y) = x_1y_1 + 6x_2y_1 - 4x_3y_1 + 2x_1y_2 + 3x_2y_2 + 2x_3y_2$$

en la cual exactamente una x_i y una y_j (cada una de primer grado) aparece en cada término. De manera, más general, las expresiones de este tipo pueden escribirse en notación escalar como:

$$f(x,y) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_iy_j$$

y son llamadas formas bilineales en x_i y en y_j. Si escribimos los vectores $x' = (x_1, x_2, \ldots, x_m)$ y $y' = (y_1, y_2, \ldots, y_n)$, una forma bilineal involucra términos

en los cuales cualquier combinación posible de los vectores componentes se forma. En notación matricial podemos escribir una forma bilineal como

$$f(x,y) = x' Ay$$

En el ejemplo numérico de arriba, tenemos

$a_{11} = 1; a_{12} = 2; a_{21} = 6; a_{22} = 3; a_{31} = -4; a_{32} = 2$

y la función puede expresarse como

$$f(x,y) = (x_1, x_2, x_3)\begin{bmatrix} 1 & 2 \\ 6 & 3 \\ -4 & 2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= (x_1 + 6x_2 - 4x_3, 2x_1 + 3x_2 + 2x_3)\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= x_1y_1 + 6x_2y_1 - 4x_3y_1 + 2x_1y_2 + 3x_2y_2 + 2x_3y_2$$

La matriz A es llamada la matriz de la forma bilineal, y la determina la forma bilineal completamente. Note que, en general, A no necesariamente es cuadrada.

Al asignar diferentes valores a x y y uno obtiene diferentes valores de la forma bilineal, cada uno de los cuales es un escalar. El conjunto de todos esos escalares, para un dominio dado de x y y, es el rango de la forma bilineal.

1.5.3 Formas cuadráticas.

Ahora, vamos a particularizar la forma bilineal al caso en que $x = y$ En este caso asumimos que y puede ser reemplazada por x y, dada su misma dimensionalidad, la matriz de coeficientes A será cuadrada en lugar de rectangular. Por ejemplo

$f(x_1, x_2) = 2x_1^2 + 5x_1x_2 + 3x_1x_2 + 6x_2^2$

también puede ser escrita en notación matricial como:

$$f(\bar{x}) = (x_1, x_2)\begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (2x_1 + 5x_2, 3x_1 + 6x_2)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= 2x_1^2 + 5x_1x_2 + 3x_1x_2 + 6x_2^2$$

y el resultado es, otra vez, un escalar, toda vez que se le asignen valores numéricos a x_1 y x_2.

A modo de definición formal, una forma cuadrática es una función polinomial de $x_1, x_2, ..., x_n$ que es homogénea y de segundo grado Por ejemplo, en el caso de dos variables, tenemos

$f(x_1, x_2) = x_1^2 + 6x_1x_2 + 9x_2^2$

en la cual el vector $x' = (x_1, x_2)$ es mapeado de un espacio bidimensional a un espacio unidimensional.

En general, una forma cuadrática en n dimensiones puede ser escrita en notación escalar como

$$q(\bar{u}) = \sum_{i,j}^{n} a_{ij} u_i u_j$$

donde $\bar{u}' = (u_1, u_2, \ldots, u_n)$, los a_{ij} son coeficientes reales y los $u_i u_j$ son las preimágenes del mapeo. Si $i = j$, obtenemos los términos cuadráticos $a_i {u_i}^2$, y si $i \neq j$, obtenemos los productos cruzados $a_{ij} u_i u_j$.

Por homogénea nos referimos a que los términos son de la forma descrita arriba y, en particular, no hay términos lineales en los $u_i's$ ni tampoco constantes. Mientras que la función

$v = {x_1}^2 + 2{x_2}^2 + x_1 x_2 + x_1 + 3x_2$

es un polinomio de segundo grado, no es una forma cuadrática ya que los últimos dos términos no son de la forma general $a_{ij} u_i u_j$.

Las formas cuadráticas son de particular interés en el análisis multivariado en la medida que nos ocupemos de lo que le pasa a las varianzas y covarianzas bajo distintas funciones lineales de un conjunto de datos multivariados.

En efecto, todas las matrices de productos cruzados empleadas en el análisis multivariado, tales como las matrices de covarianza y correlación, son ilustraciones de formas cuadráticas. En este caso las entradas de la diagonal son una medida de la dispersión de una sola variable, y las entradas fuera de la diagonal son una medida de la forma en que covarian entre sí un par de variables.

1.5.4 Tipos de formas cuadráticas.

Las formas cuadráticas pueden ser clasificadas de acuerdo a la naturaleza de los valores propios de la matriz asociada a la forma cuadrática:

1. Si todos los λ_i son positivos, la forma cuadrática es definida positiva.
2. Si todos los λ_i son negativos, la forma cuadrática es definida negativa.
3. Si todos los λ_i son no negativos (positivos o cero), la forma cuadrática es semidefinida positiva.
4. Si todos los λ_i son no positivos (cero o negativos), la forma cuadrática es semidefinida negativa.
5. Si los λ_i son una mezcla de valores positivos, cero, y negativos, la forma cuadrática es indefinida.

En el análisis multivariado estamos interesados por lo general en formas cuadráticas que sean positivas definidas o semidefinidas positivas. Por ejemplo, si una matriz simétrica es de la forma producto-momento (o $A'A$ o AA'), entonces es positiva

definida o semidefinida positiva. Ya que muchas matrices son de esta forma, los casos de las formas positivas definidas o semipositivas definidas son los de mayor interés para nosotros en el análisis multivariado.

1.5.5 Relación entre formas cuadráticas y matrices de transformación.

Suponga que tenemos la siguiente transformación

$$u = Xv$$

donde X, cuyos renglones son cosenos directores, mapea v, considerado como vector columna, en el vector u en algún espacio de interés.
Para ilustrar, sea

$$X = \begin{bmatrix} 0.8 & 0.6 \\ 0.71 & 0.71 \end{bmatrix}$$

Entonces, si $v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, tenemos

$$u = Xv = \begin{bmatrix} 0.8 & 0.6 \\ 0.71 & 0.71 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.71 \end{bmatrix}$$

Ahora suponga que queremos encontrar la norma al cuadrado de u
La norma al cuadrado de u se definió como $u'u$. Dado v y la transformación lineal X, obtenemos la siguiente expresión:

$$u'u = (Xv)'(Xv) = v'X'Xv$$

Pero ahora observamos que $X'X$ es justamente el menor producto-momento de X que hemos discutido previamente. Podemos denotar esto como A. Entonces, tenemos

$$u'u = v'Av = (1,0)\begin{bmatrix} 1.14 & 0.98 \\ 0.98 & 0.86 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$u'u = 1.14$$

Esto es, $A = A_d' A_d = S(X)$ es la matriz de la forma cuadrática que obtiene la norma al cuadrado de v bajo la transformación lineal A_d.

CAPÍTULO 2.

EL ANÁLISIS DE CONGLOMERADOS Y SU RELACIÓN CON LAS FUNCIONES DISTANCIA.

Introducción.

En este capítulo vamos a explicar la naturaleza de los problemas que intentamos resolver con el análisis de conglomerados. Luego, hablaremos un poco acerca de las áreas en las que se han aplicado este tipo de técnicas.

Después, ahondaremos en el estudio de las Funciones Distancia. Propondremos algunas funciones como "candidatos" a Funciones Distancia y, más adelante en el capítulo, demostraremos que cumplen con los requisitos para ser Funciones Distancia. Dentro de esta misma parte del capítulo, será necesaria la inclusión de algunos conceptos del álgebra lineal. Dichos conceptos nos serán útiles para demostrar que la función propuesta por Mahalanobis, es en efecto una función distancia. Finalmente, introduciremos otra técnica del análisis multivariado: el análisis de componentes principales.

2.1 Análisis de conglomerados .

Académicos e investigadores de mercado se encuentran frecuentemente con situaciones que pueden resolverse mejor definiendo grupos homogéneos de objetos, éstos pueden ser individuos, compañías, productos o incluso comportamientos. Las opciones de estrategia basadas en la identificación de grupos en la población tales como una segmentación y objetivos de mercado no serían posibles sin una metodología. Esta misma necesidad es encontrada en otras áreas, que van desde las

ciencias físicas (clasificación de varios grupos animales, insectos, mamíferos, etc.) hasta las ciencias sociales (análisis de varios pérfiles psiquiátricos). En todos los casos, el analista busca una estructura natural entre las observaciones basado en un perfil multivariado.

La técnica más usada para este propósito es el análisis de conglomerados. El análisis de conglomerados es una técnica para agrupar individuos u objetos dentro de conglomerados de tal forma que esos objetos en el mismo conglomerado son más parecidos uno de otro que lo que se parecerían dos objetos tomados de dos distintos conglomerados .

Hay muchas razones por las cuales el análisis de conglomerados puede ser útil. En primer lugar, puede ser una interrogante el encontrar los "verdaderos" grupos.
Por ejemplo, en psiquiatría siempre ha existido un gran desacuerdo acerca de la clasificación de pacientes depresivos, y el análisis de conglomerados ha sido usado para definir los grupos "objetivo". En segundo lugar, el análisis de conglomerados puede ser útil para reducir los datos. Por ejemplo, un gran número de ciudades puede ser potencialmente usado como mercado de prueba para un nuevo producto pero sólo es posible hacer el estudio en algunas. Si las ciudades pueden ser agrupadas en un número pequeño de grupos entonces uno de los miembros de cada grupo puede ser usado para la prueba de mercado. Por otro lado, si el análisis de conglomerados genera agrupamientos inesperados entonces esto puede por si sólo sugerir relaciones dignas de ser investigadas.

2.1.1 Usos de los conglomerados.

Biología. Las clasificaciones formales de animales y plantas datan desde los tiempos de Aristóteles, pero el sistema moderno de clasificación se debe esencialmente a Lineo (1753). Cada especie pertenece a una serie de conglomerados de tamaño cada vez mayor con un número decreciente de características comunes. Por ejemplo, el hombre pertenece a los primates, mamíferos, cordados, vértebrados y animales.

Este árbol, el cual fue originalmente desarrollado para nombrar ciertos objetos de manera consistente, tuvo una significación física en las teorías evolutivas de Darwin, las cuales establecían que el hombre, por ejemplo, tiene ancestros en varios niveles del árbol. El hombre tiene un ancestro en común con el mono, el conejo, la rana, el pez y el mosquito.

El árbol se usa para almacenar y diseminar el conocimiento. Por ejemplo, los vertebrados tienen columna vertebral, simetría bilateral, cuatro extremidades, cabeza, dos ojos y boca, corazón con circulación sanguinea, hígado y otras propiedades en común. Una vez que has visto a un vertebrado, ya has visto a todos. No es necesario registrar estas propiedades de manera separada para cada especie.

Las técnicas usadas en la taxonomía merecen ser imitadas en otras áreas. La estructura de árbol (arborescente) es ahora utilizada como una estructura estandard de conglomerado. Las funciones de nombrar objetos y de almacenar información de manera barata se han generalizado también a otras áreas. De hecho, es en la construcción de los árboles que se vuelve difícil generalizar los métodos de la taxonomía animal y vegetal.

Medicina. El problema de clasificación principal en medicina es como clasificar las enfermedades. Las técnicas numéricas de clasificación sólo han tenido un pequeño impacto en la taxonomía de enfermedades, quizás porque los datos médicos y especialmente las historias clínicas no son fácilmente asimiladas dentro del estandard de estructuras de datos que fue desarrollado en un principio para la taxonomía biológica.

Un tipo particular de clasificación dentro de una enfermedad es la identificación de las etapas de severidad, por ejemplo, de una enfermedad renal.

Hay otros usos de la clasificación en medicina además de la clasificación directa de las enfermedades. Los grupos sanguíneos son una clasificación de la sangre.

Finalmente, en epidemiología, las enfermedades pueden ser conglomeradas de acuerdo a su patrón de distribución en el espacio y tiempo.

Psiquiatría. Las enfermedades de la mente son más elusivas que las enfermedades del cuerpo, y la clasificación de los desordenes mentales se encuentra en un estado incierto. Hay un acuerdo en cuanto a la existencia de la paranoia, esquizofrenia y la depresión pero un criterio para el diagnóstico no está disponible.

Una dificultad característica de la clasificación de enfermedades mentales es la subjetividad y el carácter variable de los síntomas.

Las técnicas numéricas han ganado más aceptación en esta área que en el diagnóstico médico. Una de las primeras contribuciones de los conglomerados, hecha por Zubin, es un método para descubrir subgrupos de pacientes esquizofrénicos.

Arqueología y antropología. El investigador de campo encuentra un gran número de objetos tales como herramientas de piedra, objetos funerarios, estatuas ceremoniales, o cráneos que le gustaría dividir en grupos de objetos similares, cada grupo producido por la misma civilización.

Algunos estudios, relativamente recientes se han llevado a cabo para clasificar cráneos y objetos tales como broches, herramientas de piedra y de cobre.

Fitosociología. Esta disciplina se interesa en la distribución que tienen en el espacio las especies animales y vegetales. Mantiene la misma relación con la taxonomía que la existente entre la epidemiología y la clasificación de enfermedades.

Los datos típicos consisten en el número de especies en distintos cuadrantes. Al conglomerar se detectan los cuadrantes que son similares y se identifican como un mismo tipo de habitat.

Ligüistica. En esta rama del conocimiento, se usa la proporción de palabras coincidentes dentro de una lista de 196 significados como una medida de distancia entre dos lenguas, ésto con el objeto de reconstruir un árbol de evolución de las lenguas.

2.2 Funciones de distancia.

Un gran número de problemas multivariados pueden ser vistos como problemas en términos de "distancia" entre las observaciones, o entre muestras de observaciones o entre poblaciones de observaciones. En nuestro caso, para el desarrollo de las técnicas del análisis de conglomerados, es de vital importancia éste concepto. Porque las funciones de distancia nos proporcionarán un criterio para medir las similitud o disimilitud entre los objetos que estemos estudiando. Por ejemplo, consideremos datos donde lo que se mide es la mandíbula en perros, lobos, chacales y dingos, ¿Qué tan lejos está uno de estos grupos de los restantes? La idea es la siguiente, que si dos animales tienen una media similar en las medidas de sus mandíbulas entonces están "cerca", mientras que si tienen medidas muy diferentes en la media de las medidas de las mandíbulas entonces están alejados entre sí. Este es el tipo de concepto de distancia que será usado a lo largo de este capítulo.

Un gran número de distancias han sido propuestas y usadas en el análisis multivariado. Aquí sólo mencionaremos algunas de las más usadas. Es pertinente decir que medir distancias es un tema donde, en cierta medida, la arbitrariedad no puede evitarse por completo.

Medidas de similitud. El concepto de similitud es fundamental para el análisis de conglomerados. La similitud interobjeto es una medida de la correspondencia o de la semejanza entre los objetos que serán conglomerados. Aquí, las características que definen la similitud son primero especificadas. Entonces, las características son combinadas dentro de una medida de similitud calculada para todo par de objetos. En este sentido, cualquier objeto puede ser comparado con otro objeto mediante la medida de similitud. El procedimiento de análisis de conglomerados entonces procede a agrupar los objetos similares dentro de conglomerados.

La similitud interobjeto puede ser medida de distintas formas, pero hay tres métodos que dominan las aplicaciones dentro del análisis de conglomerados: medidas de correlación, de distancia, y de asociación. Cada uno de los métodos representa una perspectiva particular sobre la similitud, dependiendo de sus objetivos como del tipo de datos. Las medidas de correlación al igual que las de distancia requieren de una métrica en los datos, mientras que las medidas de asociación son para datos no-métricos.

Medidas de correlación. La medida de similitud interobjeto que probablemente nos viene primero a la mente es el coeficiente de correlación entre un par de objetos medidos sobre distintas variables. En efecto, en lugar de correlacionar dos conjuntos de variables, invertimos los objetos de la matriz X de variables entonces las columnas representan los objetos y los renglones representan las variables. Entonces, el coeficiente de correlación entre las dos columnas de números es la correlación (o similitud) entre los perfiles de los dos objetos. Una alta correlación indica similitud y una baja correlación indica ausencia de la misma.

Las medidas de correlación representan la similitud por la correspondencia de patrones en las características (variables X). Una medida de correlación de similitud no ve la magnitud de los valores, en su lugar ve los patrones de los valores. Pero las medidas de correlación son raramente usadas porque en la mayoria de las aplicaciones de análisis de conglomerados se hace énfasis en las magnitudes de los objetos, no en el patrón de los valores.

Medidas de distancia. Mientras que las medidas de correlación tienen un razón intuitiva de ser y son usadas en muchas otras técnicas multivariadas, no son las más usadas para medir la similitud en análisis de conglomerados. Las medidas de distancia de similitud, las cuales representan la similitud como la proximidad de las observaciones hacia otra observación sobre todas las variables en la variable de conglomerado, son las medidas de similitud más frecuentemente usadas.

Existen muchas medidas de distancia. La que se usa más comúnmente es la distancia euclideana. La distancia euclideana entre dos puntos es la longitud de la hipotenusa de un triángulo rectángulo. Este concepto es fácilmente generalizable a un espacio de dimensión n. La distancia euclideana es usada para calcular muchas medidas específicas, una es la distancia euclideana simple y otra es el cuadrado de ésta, o el valor absoluto de la misma, otra es la distancia tomada como la suma de los cuadrados de las diferencias sin aplicarle la raíz cuadrada. La distancia cuadrada euclideana tiene la ventaja de no tomar la raíz cuadrada, lo que hace más rápidos los cálculos, y es recomendado como medida de distancia para los métodos de conglomerado del centroide y Ward (en el siguiente capítulo hablaremos con mayor detalle acerca de los distintos métodos disponibles para llevar a cabo este tipo de análisis).

Tenemos muchas opciones que no se basan en la distancia euclideana. Una de las medidas más ampliamente usadas implican reemplazar el cuadrado de las diferencias

por la suma de los valores absolutos de las diferencias. Este procedimiento es llamado función de distancia absoluta, o distancia city-block. El usar la distancia city-block puede ser apropiado bajo ciertas circunstancias, pero causa muchos problemas. Uno de ellos es el supuesto de que las variables no están correlacionadas, pues si lo están, los conglomerados no son válidos.

Un problema que enfrentan todas las medidas de distancia que usan datos no estandarizados son las inconsistencias entre los conglomerados solución cuando la escala es cambiada.

Medidas de asociación. Las medidas de asociación de similitud son usadas para comparar objetos cuyas características son medidas sólo en términos no-métricos (medidas ordinales o nominales). Como ejemplo, los encuestados pueden responder si o no a un cierto número de preguntas. Una medida de asociación puede reflejar el grado de acuerdo o correspondencia entre cada par de encuestados. La forma más simple de medida de asociación puede ser el porcentaje de veces que hubo correspondencia (ambos encuestados respondieron que si o no a una misma pregunta) sobre todas las preguntas. Extensiones de este simple coeficiente de correspondencia han sido desarrollados para variables nominales multicategóricas y hasta para medidas ordinales.

2.2.1 Distancias entre observaciones individuales.

Para comenzar, acerca de la medida de distancias, consideremos el caso más simple donde hay n individuos, cada uno de ellos tiene valores para p variables $X_1, X_2, \ldots, X_p$. Los valores para el individuo i pueden ser denotados por $x_{i1}, x_{i2}, \ldots, x_{ip}$ y los del individuo j por $x_{j1}, x_{j2}, \ldots, x_{jp}$.

El problema es medir la distancia entre el individuo i y el individuo j.

Los casos para dos y tres variables sugieren que una distancia euclideana generalizada sería:

$$d(\bar{x}, \bar{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} = \|\bar{x} - \bar{y}\|$$

la cual debe servir, de manera satisfactoria, para muchos propósitos.

De esa ecuación de distancia, se nota claramente que si una de las variables medidas varía mucho más que las otras entonces ésta habrá de dominar el cálculo de las distancias. Por ejemplo, para tomar un caso extremo, suponga que se comparan a n hombres y que X_1 denota su estatura; y que las otras variables son dimensiones de dientes, con todas las medidas hechas en milímetros. Las diferencias de estatura serán del orden de 20 a 30 milímetros mientras las diferencias en las dimensiones de los dientes serán del orden de uno o dos milímetros. Un cálculo simple de d_{ij} proveerá

distancias entre individuos que esencialmente serán diferencias de estatura solamente, con las diferencias en dientes siendo prácticamente ignoradas. Claramente, se tiene un problema de escala.

En la práctica es usualmente deseable, para todas las variables, que tengan aproximadamente la misma influencia en el cálculo de la distancia. Esto se logra mediante una estandarización en las variables. Esto puede hacerse, por ejemplo, dividiendo cada variable por su desviación estandard para los n individuos que serán sometidos a comparación.

2.2.1.1 Demostraciones de las funciones distancia.

Sean $\bar{x} = (x_1, x_2, \ldots, x_n)$, $\bar{y} = (y_1, y_2, \ldots, y_n)$ y $\bar{z} = (z_1, z_2, \ldots, z_n)$; $\bar{x}, \bar{y}$ y $\bar{z} \in R^n$.

Queremos demostrar que la función definida como:

$\bar{x} \bullet \bar{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$

es un producto interno.

Tenemos que demostrar lo siguiente:

1) $\bar{x} \bullet \bar{x} \geq 0$ y $\bar{x} \bullet \bar{x} = 0 \Leftrightarrow \bar{x} = 0$

2) $\bar{x} \bullet \bar{y} = \bar{y} \bullet \bar{x}$

3) $\bar{z} \bullet (\bar{x} + \bar{y}) = \bar{z} \bullet \bar{x} + \bar{z} \bullet \bar{y}$

4) $(\alpha \bar{x} \bullet \bar{y}) = \alpha(\bar{x} \bullet \bar{y})$

Demostración.

1) Sea $\bar{x} \in R^n$ por definición

$\bar{x} \bullet \bar{x} = x_1 x_1 + x_2 x_2 + \ldots + x_n x_n = x_1^2 + x_2^2 + \ldots + x_n^2$ claramente $x_i^2 \geq 0$

$i = 1,2,\ldots,n$ de donde si al menos una $x_i > 0 \Rightarrow \bar{x} \bullet \bar{x} > 0$

ahora bien, si $\bar{x} \bullet \bar{x} = 0 \Rightarrow x_1^2 + x_2^2 + \ldots + x_n^2 = 0$, es decir, tenemos una suma de términos no negativos igualada con cero entonces cada término debe ser igual a cero, esto es, $x_i = 0$, $i = 1,2,\ldots,n$

$\therefore \bar{x} = 0$

si $\bar{x} = 0$ resulta inmediato que $\bar{x} \bullet \bar{x} = 0$

2) Por definición

$\bar{x} \bullet \bar{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$ por la conmutatividad de la multiplicación entre reales tenemos que

$x_i y_i = y_i x_i \ , \ i = 1,2,\ldots,n$ de donde

$$\bar{x} \bullet \bar{y} = \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} y_i x_i = \bar{y} \bullet \bar{x}$$

3) Por definición

$$\bar{z} \bullet (\bar{x} + \bar{y}) = \sum_{i=1}^{n} z_i (x_i + y_i) = \sum_{i=1}^{n} (z_i x_i + z_i y_i)$$

por la asociatividad de la multiplicación de los números reales

$$\sum_{i=1}^{n} (z_i x_i + z_i y_i) = \sum_{i=1}^{n} z_i x_i + \sum_{i=1}^{n} z_i y_i = \bar{z} \bullet \bar{x} + \bar{z} \bullet \bar{y}$$

por la linealidad de la suma y por la definición.

$\therefore \bar{z} \bullet (\bar{x} + \bar{y}) = \bar{z} \bullet \bar{x} + \bar{z} \bullet \bar{y}$

4) Sean $\bar{x}$ y $\bar{y} \in R^n$ y $\alpha \in R$

$$(\alpha \bar{x} \bullet \bar{y}) = \sum_{i=1}^{n} \alpha x_i y_i \qquad \text{por definición}$$

$$= \alpha \sum_{i=1}^{n} x_i y_i \qquad \text{por la linealidad de la suma}$$

$$= \alpha (\bar{x} \bullet \bar{y}) \qquad \text{por la definición de producto punto}$$

$\therefore (\alpha \bar{x} \bullet \bar{y}) = \alpha (\bar{x} \bullet \bar{y})$ q.e.d.

Ahora bien, consideremos el vector $(\bar{x} - \bar{y})$ y apliquémosle el producto punto anteriormente visto:

$$(\bar{x} - \bar{y}) \bullet (\bar{x} - \bar{y}) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2$$

esta última expresión nos recuerda algo, a saber :

$$d(\bar{x}, \bar{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} = \|\bar{x} - \bar{y}\|$$

es decir, hemos demostrado, formalmente, que la distancia euclideana es, en efecto, una función de distancia .

Otra función de distancia que habremos de utilizar, es la siguiente:

$$D(\bar{x}, \bar{y}) = |x_1 - y_1| + |x_2 - y_2| + \ldots + |x_n - y_n| = \sum_{i=1}^{n} |x_i - y_i|$$

Esta distancia se conoce como Manhattan o "city block". Ahora bien, debemos demostrar formalmente que estamos tratando con una función distancia, para lo cual debemos probar, como ya hemos hecho antes, lo siguiente:

1) $D(\bar{x},\bar{y}) > 0 ; \bar{x} \neq \bar{y}$

2) $D(\bar{x},\bar{y}) = 0 \Leftrightarrow \bar{x} = \bar{y}$

3) $D(\bar{x},\bar{y}) = D(\bar{y},\bar{x})$

4) $D(\bar{x},\bar{y}) \leq D(\bar{x},\bar{z}) + D(\bar{z},\bar{y})$

Demostración.

1) Sean $\bar{x}$ y $\bar{y} \in R^n$ $\bar{x} \neq \bar{y}$, por lo que $x_i - y_i \neq 0$ para alguna $i = 1.2\ldots,n$

Recordemos que una de las propiedades del valor absoluto, es.
Si $a \neq 0$ y $a \in R \Rightarrow |a| > 0$, de donde:
$||x_i - y_i| > 0$ para alguna $i = 1.2\ldots,n$
Como estamos sumando términos no negativos, es claro que:
$\sum_{i=1}^{n} |x_i - y_i| > 0$ y por la definición de la distancia de Manhattan tenemos que
$D(\bar{x},\bar{y}) > 0$

2) $\Rightarrow) D(\bar{x},\bar{y}) = 0$
Por definición
$D(\bar{x},\bar{y}) = \sum_{i=1}^{n} |x_i - y_i| = 0$, en palabras tenemos una suma de términos no negativos igual a cero, de donde
$|x_i - y_i| = 0 \quad \forall i = 1,2,\ldots,n$; recordemos la siguiente propiedad del valor absoluto :
Si $|a| = 0 \Leftrightarrow a = 0$; de donde :
$x_i - y_i = 0 \quad \forall i = 1,2,\ldots,n$; entonces
$x_i = y_i \quad \forall i = 1,2,\ldots,n$
$\therefore \bar{x} = \bar{y}$

$\Leftarrow)$ Si $\bar{x} = \bar{y} \Rightarrow x_i = y_i \quad \forall i = 1,2,\ldots,n$
$\Rightarrow x_i - y_i = 0 \Rightarrow |x_i - y_i| = 0$; de donde
$\sum_{i=1}^{n} |x_i - y_i| = 0$; y por la definición de distancia de Manhattan tenemos
$D(\bar{x},\bar{y}) = 0$

3) Sean $\bar{x} = (x_1, x_2, \ldots, x_n)$ y $\bar{y} = (y_1, y_2, \ldots, y_n) \in R^n$

$|x_i - y_i| = |-(y_i - x_i)| = |-1||y_i - x_i| = |y_i - x_i|$

$\therefore |x_i - y_i| = |y_i - x_i| \quad \forall i = 1,2,\ldots,n$, de donde

$\sum_{i=1}^{n} |x_i - y_i| = \sum_{i=1}^{n} |y_i - x_i|$ y por la definición de distancia de Manhattan

$D(\bar{x}, \bar{y}) = D(\bar{y}, \bar{x})$

4) Sean $\bar{x} = (x_1, x_2, \ldots, x_n)$, $\bar{y} = (y_1, y_2, \ldots, y_n)$ y $\bar{z} = (z_1, z_2, \ldots, z_n) \in R^n$

$|x_i - y_i| = |x_i + 0 - y_i| = |x_i + z_i - z_i - y_i| = |(x_i - z_i) + (z_i - y_i)|$

y por la desigualdad del triángulo en los reales

$|(x_i - z_i) + (z_i - y_i)| \leq |x_i - z_i| + |z_i - y_i|$

$\therefore |x_i - y_i| \leq |x_i - z_i| + |z_i - y_i| \quad \forall i = 1,2,\ldots,n$, de donde

$\sum_{i=1}^{n} |x_i - y_i| \leq \sum_{i=1}^{n} |x_i - z_i| + |z_i - y_i|$ por la linealidad de la suma

$\sum_{i=1}^{n} |x_i - y_i| \leq \sum_{i=1}^{n} |x_i - z_i| + \sum_{i=1}^{n} |z_i - y_i|$

utilizando la definición de distancia de Manhattan

$D(\bar{x}, \bar{y}) \leq D(\bar{x}, \bar{z}) + D(\bar{z}, \bar{y})$ q.e.d.

2.2.2 Distancias entre poblaciones y muestras.

Muchas medidas han sido propuestas para la distancia entre dos poblaciones multivariadas cuando la información de las medias, varianzas y covarianzas de las poblaciones está disponible. Aquí, sólo nos ocuparemos de la distancia de Mahalanobis.

Suponga que tenemos g poblaciones y las distribuciones multivariadas en estas poblaciones son conocidas para p variables $X_1, X_2, \ldots, X_P$ .Una medida que si toma en cuenta las correlaciones entre las variables es la distancia de Mahalanobis,

$$D_{ij}^{\ 2} = \sum\sum (\mu_{ri} - \mu_{rj}) v^{rs} (\mu_{si} - \mu_{sj}),$$

donde la media de la variable X_K en al i-ésima población es μ_{ki} y, v^{rs} es el elemento en el r-ésimo renglón y la s-ésima columna de la inversa de la matriz de covarianza para las p variables. Esto es una forma cuadrática y puede escribirse de la siguiente forma:

$$D_{ij}^{\ 2} = (\mu_i - \mu_j)' V^{-1} (\mu_i - \mu_j), \qquad 1)$$

donde

$$\mu_i = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \cdots \\ \mu_{pi} \end{bmatrix}$$

es el vector de medias para la i-ésima población y (análogamente para j) V es la matriz de covarianza. Esta medida puede ser sólo calculada si la matriz de covarianza poblacional es la misma para todas las poblaciones.
La distancia de Mahalanobis se usa frecuentemente para medir la distancia de una única observación multivariada al centro de la población de la cual proviene. Si $x_1, x_2, \ldots, x_p$ son los valores de $X_1, X_2, \ldots, X_P$ para el individuo, con los correspondientes valores de la media $\mu_1, \mu_2, \ldots, \mu_p$, entonces

$$D^2 = (\bar{x} - \bar{\mu})' V^{-1} (\bar{x} - \bar{\mu}), \qquad 2)$$

donde $\bar{x}' = (x_1, x_2, \ldots, x_p)$ y $\bar{\mu}' = (\mu_1, \mu_2, \ldots, \mu_p)$. Igual que antes V denota la matriz de covarianza poblacional y V^{-1} la inversa de dicha matriz.

El valor de D^2 puede ser pensado como un residual multivariado para la observación $\bar{x}$. Un residual bajo este contexto significa una medida de que tan lejos esta la observación $\bar{x}$ del centro de las distribuciones de todos los valores, tomando en cuenta todas las variables en consideración.

Las ecuaciones 1) y 2) pueden obviamente ser usadas con una muestra de datos si se estima la medias, varianzas y covarianzas poblacionales en lugar de los verdaderos parámetros.

Vamos a recordar algunos conceptos básicos del álgebra lineal. Todo esto encaminado a demostrar que la distancia de Mahalanobis es, en efecto, una función distancia.

2.3 Diagonalización ortogonal; matrices simétricas.

Antes de abordar este problema, la diagonalización de una matriz, debemos dar algunas definiciones importantes.

Definición 2.1.

Sea V un espacio con producto interno y sea $S = \{\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n\}$ un conjunto de vectores de V. Se dice que S es un conjunto ortogonal cuando

$\bar{v}_i \bullet \bar{v}_j = 0, \ \forall i \neq j$

si además $\|\bar{v}_i\| = 1, \forall i$, el conjunto S es ortonormal.

Definición 2.2.

Una matriz cuadrada A con la propiedad de que $A^{-1} = A^t$ se le denomina matriz ortogonal.

Definición 2.3.

Se dice que una matriz cuadrada A es ortogonalmente diagonalizable si existe una matriz ortogonal P tal que $P^{-1}AP(= P^tAP)$ sea diagonal; se dice que la matriz P diagonaliza ortogonalmente a A.

Teorema 2.1.

Si A es una matriz de nxn, entonces las proposiciones que siguen son equivalentes:

a) A es ortogonalmente diagonalizable.

b) A tiene un conjunto ortonormal de n eigenvectores.

Demostración.

$\Rightarrow$) Supuesto que A es ortogonalmente diagonalizable, existe una matriz ortogonal P tal que $P^{-1}AP(= P^tAP)$ es diagonal. Como se indica en la demostración del teorema 10, los n vectores columna de P son eigenvectores de A. Como P es ortogonal, estos vectores columna son ortonormales de modo que A tiene n eigenvectores ortonormales.

$\Leftarrow$) Supóngase que A tiene un conjunto ortonormal de n eigenvectores $\{p_1, p_2, \ldots, p_n\}$ como se indica en la demostración del teorema 10, la matriz P tiene a estos eigenvectores como columnas, diagonaliza a A. Debido a que estos eigenvectores son ortonormales, P es ortogonal y, por lo tanto, diagonaliza ortogonalmente a A.

En la demostración del teorema anterior se indica que una matriz A de nxn, ortogonalmente diagonalizable, es ortogonalmente diagonalizada por cualquier matriz P de nxn cuyas columnas formen un conjunto ortonormal de eignevectores de A. Sea D la matriz diagonal

$$D = P^{-1}AP$$

Por tanto, $A = PDP^{-1}$

o, como P es ortogonal, $A = PDP^t$

Por tanto, $A^t = (PDP^t)^t = (DP^t)^t P^t = (P^t)^t D^t P^t = PD^t P^t$

pero como D es diagonal $D = D^t$

finalmente resulta $A^t = PDP^t = A$

Así entonces, se ha demostrado que una matriz ortogonalmente diagonalizable es simétrica. El recíproco también es verdadero pero omitiremos su demostración dentro de éste trabajo.
Enunciemos formalmente este resultado.

Teorema 2.2.
Si A es una matriz de nxn, entonces las proposiciones siguientes son equivalentes:
a) A es ortogonalmente diagonalizable.
b) A es simétrica.

De donde, tenemos los siguientes resultados en el caso de que A sea una matriz simétrica.
1) Si una matriz simétrica A puede ser escrita como el producto

$$A = TDT'$$

donde D es diagonal con todas sus entradas no-negativas y T es una matriz ortogonal de eigenvectores, entonces

$$A^{1/2} = TD^{1/2}T'$$

y en este caso $A^{1/2}A^{1/2} = A$

2) Si una matriz simétrica A^{-1} puede ser escrita como el producto

$$A^{-1} = TD^{-1}T'$$

donde D^{-1} es diagonal con todas sus entradas no-negativas y T es una matriz ortogonal de eigenvectores, entonces

$$A^{-1/2} = TD^{-1/2}T'$$

y en este caso $A^{-1/2}A^{-1/2} = A^{-1}$.

Ahora, definiremos la raíz cuadrada de una matriz diagonal. Las entradas en la diagonal deberán ser no negativas.

Parece natural que definamos dicha matriz (la raíz cuadrada) como la raíz cuadrada de los elementos existentes en la diagonal; ejemplo:

$$D=\begin{bmatrix}1 & 0 & 0\\ 0 & 4 & 0\\ 0 & 0 & 9\end{bmatrix}$$

buscamos $D^{1/2}$, de modo que; $D^{1/2}D^{1/2}=D$. Entonces :

$$D^{1/2}=\begin{bmatrix}1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix}$$

comprobémoslo

$$D^{1/2}D^{1/2}=\begin{bmatrix}1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 0 & 4 & 0\\ 0 & 0 & 9\end{bmatrix}=D$$

Definición 2.4.
La raíz cuadrada de una matriz diagonal se define como $D^{1/2}D^{1/2}=D$ donde los elementos de $D^{1/2}$ son $\sqrt{d_{ii}}$ y se tiene el supuesto de que todas las d_{ii} en D son no-negativas.

Ahora, vamos a recordar el concepto de inversa de una matriz.

Definición 2.5.
Sea A una matriz de *nxn* con elementos en R. Una matriz X se dice que es inversa de A si y sólo si

$$XA=I_n=AX$$

y se representa con A^{-1}.

Cabe hacer notar que la igualdad $XA=AX$ sólo es posible cuando A y X son matrices cuadradas del mismo orden; en consecuencia, para que una matriz A tenga inversa es condición necesaria que sea cuadrada. Además, la inversa deberá ser también cuadrada y del mismo orden que A.
En lo que se refiere a unicidad, se puede demostrar que la inversa de una matriz cuadrada (si existe) es única, como lo establece el siguiente teorema, en el que se enuncian además otras propiedades importantes de la inversa.

Teorema 2.3.
Si A y B son dos matrices no singulares del mismo orden y $\lambda \in R$, entonces :

1) A^{-1} es única.

2) $(A^{-1})^{-1} = A$

3) $(AB)^{-1} = B^{-1}A^{-1}$

4) $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$, si $\lambda \neq 0$

Demostración.

Se demuestran a continuación 1) y 3) por ser de mayor interés las demostraciones. Omitimos las demostraciones de los incisos 2) y 4).

1) Sea A una matriz de *nxn* no singular, y sean X, Y dos inversas de A; entonces por la definición de inversa tenemos :

$XA = I_n = AX$ y $YA = I_n = AY$

Por otra parte

$X = XI_n$

$X = X(AY)$ por hipótesis

$X = (XA)Y$ por la distributividad del producto de matrices

$X = I_nY$ por hipótesis

$\therefore X = Y$

y en consecuencia la inversa es única.

3) Sean A y B dos matrices de *nxn* no singulares. Por tanto, existen A^{-1} y B^{-1} y puede formarse el producto $B^{-1}A^{-1}$

para el cual se tiene que

$(B^{-1}A^{-1})(AB) = (B^{-1}A^{-1})[(A)(B)]$

$= [(B^{-1}A^{-1}A)]B$ por la distributividad del producto de matrices

$= [B^{-1}(A^{-1}A)]B$ por la distributividad del producto de matrices

$= (B^{-1}I_n)B$ por la definición de matriz inversa

$= B^{-1}B$ por definición de matriz identidad

$\therefore (B^{-1}A^{-1})(AB) = I_n$

en forma análoga se puede demostrar que:

$(AB)(B^{-1}A^{-1}) = I_n$

y en consecuencia, de la definición de matriz inversa se tiene que $B^{-1}A^{-1}$ es la inversa de AB; esto es

$B^{-1}A^{-1} = (AB)^{-1}$ q.e.d.

Bien, por fin, demostraremos que la función propuesta por Mahalanobis es una función distancia.
Recordemos que dicha función está definida de la siguiente forma: $D = (\overline{y_1} - \overline{y_2})' S^{-1} (\overline{y_1} - \overline{y_2})$ donde S es la matriz de covarianza muestral y $\overline{y_1}$ y $\overline{y_2}$ son los vectores de medias muestrales. Esta función resulta ser una forma cuadrática. Ahora sí, pasemos a la demostración.

2.3.1 Distancia de Mahalanobis.

Sabemos por el teorema 10 que

$$S = PDP^{-1}$$

ahora bien, D es una matriz diagonal y si $\lambda_i > 0 \; \forall i = 1,2,\ldots,n$ (i.e. S es positiva definida) entonces podemos expresar a D como:

$$D = D^{1/2} D^{1/2}$$

de donde S puede expresarse de la siguiente forma:

$$S = PD^{1/2} D^{1/2} P^{-1}$$

Pero si S es simétrica y, de hecho lo es (como lo mostramos en el capítulo anterior) podemos afirmar algo más acerca de la matriz P. Lo que podemos afirmar acerca de P es que, no es sólo singular sino que también es ortogonal. Teniendo esto en mente S quedaría expresada de la siguiente manera:

$$S = TD^{1/2} D^{1/2} T'$$

pues bien, sigamos adelante. Recordemos que la distancia de Mahalanobis está expresada en términos de la matriz S^{-1} de donde utilizando la última descomposición:

$$S^{-1} = (TD^{1/2} D^{1/2} T')^{-1}$$

$$S^{-1} = TD^{-1/2} D^{-1/2} T' \qquad (D^{-1/2})' = D^{-1/2} \text{ porque es una matriz diagonal.}$$

ahora vamos a sustituir esta expresión en la Función de Mahalanobis:

$$(\overline{y_1} - \overline{y_2}) S^{-1} (\overline{y_1} - \overline{y_2})' =$$

$$(\overline{y_1} - \overline{y_2}) TD^{-1/2} D^{-1/2} T' (\overline{y_1} - \overline{y_2})' =$$

$$\overline{u} = (\overline{y_1} - \overline{y_2}) T \Rightarrow \overline{u}' = T' (\overline{y_1} - \overline{y_2})'$$

$$\overline{u} D^{-1/2} D^{-1/2} \overline{u}' = \overline{u} D^{-1/2} (D^{-1/2})' \overline{u}' =$$

$$\overline{v} = \overline{u} D^{-1/2} \Rightarrow \overline{v}' = (D^{-1/2})' \overline{u}'$$

$$= \bar{v}\,\bar{v}' = ||\bar{v}||^2$$

Y como ya habíamos demostrado en el Capítulo 1 que la norma al cuadrado de un vector es una función distancia, esto termina la demostración.

q.e.d.

2.4 La elección de las variables.

La elección inicial del conjunto particular de medidas a usar para describir a cada individuo al cual se le someterá a alguna técnica de conglomerado constituye el marco de referencia dentro del cual se establecerán los conglomerados; esta elección se supone que refleja el juicio del investigador, la relevancia que tiene una variable para la clasificación. Consecuentemente la primera pregunta acerca de la elección de las variables es si son o no son relevantes para el tipo de clasificación que se este haciendo. Por ejemplo, si la clasificación de los enfermos mentales es la meta, útil para estudiar los efectos de los distintos tratamientos, probablemente nos es importante incluir variables tales como: altura, peso y otras estadísticas vitales, ya que esto podría traer consigo el resultado de tener simplemente conglomerados formados por hombres y mujeres.

Es importante tener en mente que, la elección inicial de variables es por sí misma una categorización de datos para la cual, hay, solamente, guías tanto estadísticas como matemáticas, muy limitadas.

La siguiente pregunta que puede ser considerada es· ¿Cuántas variables deben medirse en cada individuo?

En muchos de los casos, probablemente, en teoría no hay un número límite de variables que puedan utilizarse para producir una clasificación. En la práctica, por supuesto, muchas pueden considerarse irrelevantes para el propósito que se tiene, y otra restricción importante que surgirá serán consideraciones de carácter económico como de tiempo. En lo que concierne a cuales variables a medir, no hay, por lo general, una base teórica para determinar el número de variables a usar y el problema debe de ser resuelto empíricamente.
Para el desarrollo de nuestros análisis, nos será de mucha utilidad emplear el Análisis de Componentes Principales. Esta utilidad la descubriremos mediante la exposición del objetivo que persigue.

2.5 Análisis de Componentes Principales.

El análisis de componentes principales es uno de los métodos multivariados más útiles. El objeto del análisis es tomar p variables $X_1, X_2, .., X_P$ y encontrar combinaciones de éstas para producir índices $Z_1, Z_2, \ldots, Z_P$ que no estén correlacionados.

La ausencia de correlación es una propiedad útil porque significa que los índices están midiendo "diferentes" dimensiones en los datos. Sin embargo, los índices también están ordenados de manera que Z_1 nos da la mayor proporción de varianza, Z_2 nos da la segunda proporción de varianza más grande y así sucesivamente. Esto es, $\text{var}(Z_1) \geq \text{var}(Z_2) \geq \ldots \geq \text{var}(Z_P)$, donde la $\text{var}(Z_t)$ denota la varianza de Z_t en el conjunto de datos considerado. Las $Z_{t's}$ reciben el nombre de componentes principales. Cuando se hace un análisis de componentes principales existe siempre la esperanza que las varianzas de la mayoría de los índices sean tan bajas como para ser despreciada. En ese caso la variación en el conjunto de datos puede ser adecuadamente descrita por unas cuantas variables Z con varianzas que sean más significativas.

Logramos entonces economizar, ya que la variación en las p $X's$ originales se tiene con un número más pequeño de variables Z.

Debe hacerse hincapié que del análisis de componentes principales no siempre resulta que, un número grande de variables originales se reduce a un número menor de variables transformadas. En efecto, si el número original de variables no están correlacionadas entonces el análisis no hace absolutamente nada. Los mejores resultados se obtienen cuando el número original de variables están altamente correlacionadas, positiva o negativamente. Si es este el caso es muy sensato pensar que 20 ó 30 variables originales pueden ser representadas adecuadamente mediante 2 ó 3 componentes principales. Si este deseable estado de las cosas se da entonces los componentes principales importantes serán de interés como una forma de subrayar las dimensiones en los datos. Sin embargo, también será de valor saber que hay redundancia en las variables originales, con la mayoria de ellas midiendo las mismas cosas.

2.5.1 Procedimiento para el Análisis de Componentes Principales.

Un análisis de componentes principales comienza con datos en p variables para n individuos. El primer componente principal es la combinación lineal de las variables $X_1, X_2, \ldots, X_P$

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$$

que varía tanto como sea posible para los individuos, sujeto a la restricción

$$a_{11}^{\ 2} + a_{12}^{\ 2} + \ldots + a_{1p}^{\ 2} = 1$$

entonces la varianza de Z_1, $\text{var}(Z_1)$, es tan grande como sea posible dada esta restricción para las constantes a_{ij}. La restricción es introducida porque de otra manera la $\text{var}(Z_1)$ se incrementaría simplemente aumentando cualquiera de los valores a_{ij}.

el segundo componente principal

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_P$$

es tal que la $\text{var}(Z_2)$ es tan grande como es posible sujeta a la restricción

$$a_{21}^{\ 2} + a_{22}^{\ 2} + \ldots + a_{2p}^{\ 2} = 1$$

y también la condición que Z_1 y Z_2 no estén correlacionadas.

El tercer componente principal

$$Z_3 = a_{31}X_1 + a_{32}X_2 + \ldots + a_{3p}X_P$$

es tal que la $\text{var}(Z_3)$ es tan grande como es posible sujeta a la restricción

$$a_{31}^{\ 2} + a_{32}^{\ 2} + \ldots + a_{3p}^{\ 2} = 1$$

y también Z_3 no está correlacionada con Z_2 y Z_1. Los demás componentes principales se definen continuando de misma forma. Si hay p variables entonces puede haber hasta p componentes principales.

Para usar los resultados del análisis de componentes principales no es necesario conocer como se derivan las ecuaciones para los componentes principales. Sin embargo, es útil el entender la naturaleza de las ecuaciones por sí mismas. De hecho un análisis de componentes principales sólo implica encontrar los eigenvalores de la matriz de covarianza muestral.

La forma de obtener la matriz de covarianza fue definida en el primer capítulo. Como ya mostramos, la matriz es simétrica y tiene la forma

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix}$$

donde el elemento c_{ii} es la varianza de X_i y c_{ij} es la covarianza de las variables X_i y X_j.

Las varianzas de los componentes principales son los eigenvalores de la matriz C. Hay p de estos valores, algunos de ellos pueden ser cero. Los eigenvalores negativos no son posibles para una matriz de covarianza (C es una matriz simétrica).

Suponiendo que los eigenvalores están ordenados como $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$, entonces λ_i corresponde al i-ésimo componente principal

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \ldots + a_{ip}X_P$$

en particular, $\text{var}(Z_i) = \lambda_i$ y las constantes $a_{i1}, a_{i2}, \ldots, a_{ip}$ son los elementos del eigenvector correspondiente.

Una propiedad importante de los eigenvalores es que suman lo mismo que los elementos de la diagonal de la matriz C (la traza de C). Esto es:

$$\lambda_1 + \lambda_2 + \ldots + \lambda_p = c_{11} + c_{22} + \ldots + c_{pp}$$

Ya que c_{ii} es la varianza de X_i y λ_i es la varianza de Z_i, esto significa que la suma de las varianzas de los componentes principales es igual a la suma de las variables originales. Por tanto, de alguna manera, los componentes principales contienen toda la variación existente en los datos originales.

Para poder evitar que una variable tenga una influencia mucho mayor que las demás en los componentes principales es usual estandarizar las variables $X_1, X_2, \ldots, X_P$ para que tengan medias de cero y varianzas de uno desde el principio del análisis.
Entonces la matriz C toma la forma

$$C = \begin{bmatrix} 1 & c_{12} & \ldots & c_{1p} \\ c_{21} & 1 & \ldots & c_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ c_{p1} & c_{p2} & \ldots & 1 \end{bmatrix}$$

donde $c_{ij} = c_{ji}$ es la correlación entre X_i y X_j. En otras palabras, el análisis de componentes principales se lleva a cabo sobre la matriz de correlación. En ese caso, la suma de los elementos de la diagonal, y por tanto la suma de los eigenvalores, es igual a p, el número de variables.

Los pasos necesarios en un análisis de componentes principales pueden enunciarse de la siguiente forma.

1. Se comienza por estandarizar las variables $X_1, X_2, \ldots, X_P$ para que tengan media cero y varianza unitaria. Esto es usual, pero se omite en algunos casos

2. Se calcula la matriz de covarianza C. Esta es una matriz de correlación si el paso número uno ha sido efectuado.

3. Se encuentran los eigenvalores $\lambda_1, \lambda_2, \ldots, \lambda_p$ y los correspondientes eigenvectores $a_1, a_2, \ldots, a_p$. Los coeficientes de i-ésimo componente principal están dados por a_i mientras que la varianza está dada por λ_i.

4. Desprecie cualquier componente que sólo contribuya con una pequeña proporción de la variación de los datos. Por ejemplo, si se empieza con 20 variables puede llegarse a la conclusión que los primeros componentes contribuyen con el 90% de la varianza total. Tomando esto en consideración los otros 17 componentes principales pueden ser entonces ignorados.

Ejemplo. Medidas corporales de gorriones hembra.

Ya hemos mencionado que es lo que se pretende al llevar a cabo un análisis de componentes principales. Ahora, hagamos un ejemplo para una mayor comprensión.
Los datos que habremos de utilizar fueron reunidos después de una severa tormenta el 1o. de febrero de 1898, un cierto número de gorriones moribundos fueron llevados al laboratorio de biología de la Universidad de Brown, en Rhode Island. Subsecuentemente, cerca de la mitad de las aves murió y Hermon Bumpus vio esto como una oportunidad para estudiar el efecto de la selección natural sobre las aves. El tomó ocho medidas morfológicas de cada una de las aves y las pesó. Los resultados para 5 de las variables se muestran en la tabla 1 del apéndice.

Por supuesto, el desarrollo de los métodos del análisis multivariado apenas habían comenzado en 1898 cuando Bumpus estaba haciendo su estudio. Sin embargo, sus métodos de análisis fueron bastantes finos. Muchos autores han vuelto a analizar sus datos y, en general, han confirmado sus conclusiones.

Es apropiado comenzar con el paso 1) de los cuatro que consta el análisis de conglomerados. La estandarización de las medidas asegura que todas tendrán la misma importancia en el análisis. Omitir la estandarización significaría que las variables X_1 y X_2, las cuales varían más dentro de las 49 aves, dominarían los componentes principales.

La matriz de covarianza para las variables estandarizadas es la matriz de correlación, la cual mostramos a continuación:

matriz de correlación	X1	X2	X3	X4	X5
X1	1	0.73496422	0.66181194	0.64528412	0.60512465
X2	0.73496422	1	0.67374109	0.76850868	0.52901381
X3	0.66181194	0.67374109	1	0.7631899	0.52627007
X4	0.64528412	0.76850868	0.7631899	1	0.60664925
X5	0.60512465	0.52901381	0.52627007	0.60664925	1

Los eigenvalores de esta matriz son:

	Eigenvalor	% total de la Varianza	Eigenvalor Acumulado	% Acumulado
1	3.61597834	72.31956689	3.615978345	72.3195669
2	0.53150408	10.63008153	4.147482421	82.9496484
3	0.38642455	7.728490981	4.53390697	90.6781394
4	0.30156552	6.031310351	4.835472488	96.7094498
5	0.16452751	3.290550245	5	100

Observese que la suma de los eigenvalores es cinco, la suma de los términos de la diagonal en la matriz de correlación (la traza de la matriz de correlación). Los eigenvectores correspondientes se muestran en la siguiente tabla:

	factor 1	factor 2	factor 3	factor 4	factor 5
X1	0.45179893	0.05072137	0.69047023	-0.42041399	-0.3739091
X2	0.46168085	-0.29956355	0.34054844	0.54786307	0.53008046
X3	0.45054161	-0.32457242	-0.45449265	-0.60629605	0.34279226
X4	0.47073887	-0.18468403	-0.410935	0.38827811	-0.65166652
X5	0.39767537	0.87648935	-0.1784558	0.06887199	0.19243414

Los eigenvectores mostrados están estandarizados, de donde, la suma del cuadrado de sus componentes es igual a uno para cada uno de ellos. Estos eigenvectores nos proporcionan los coeficientes de los componentes principales.

El eigenvalor para un componente principal indica la varianza con la que contribuye del total de 5.000. Entonces el primer componente principal contribuye con (3.61597834/5.000)100%=72.3195669%,el segundo con el 10.63008153%, el tercero con el 7.728490981%, el cuarto con el 6.031310351% y el quinto con el 3.290550245%. Claramente, el primer componente es mucho más importante que cualquiera de los otros.

Otra forma de entender la importancia relativa de cada componente principal es en términos de su varianza en comparación con la varianza de las variables originales. Después de la estandarización las variables originales tienen, todas varianza 1.0. Entonces, el primer componente principal tiene una varianza 3.61597834 veces la

varianza de cualquiera de las originales. Sin embargo, el segundo componente principal tiene una varianza de sólo 0.53150408 veces la de alguna de las variables originales. Los otros componentes principales contribuyen aún con menos varianza

El primer componente principal es :

$$Z_1 = 0.452X_1 + 0.462X_2 + 0.451X_3 + 0.471X_4 + 0.398X_5$$

donde de X_1 a X_5 tenemos variables estandarizadas;

el siguiente componente principal es:

$$Z_2 = -0.051X_1 + 0.300X_2 + 0.325X_3 + 0.185X_4 - 0.877X_5$$

Observemos que una ventaja adicional de poder reducir nuestros datos a sólo dos componentes principales, es la de poder representarlos gráficamente de manera más clara para nosotros.

CAPÍTULO 3.

TÉCNICAS JERÁRQUICAS DE CONGLOMERADO.

Introducción.

En este capítulo, nos daremos a la tarea de presentar las distintas técnicas existentes para llevar a cabo un análisis de conglomerados. Cabe mencionar que sólo nos ocuparemos de las técnicas jerárquicas de conglomerado y, dentro de estas estudiaremos sólo los métodos aglomerativos. Se ilustrará la manera en que funcionan cada una de las distintas técnicas, utilizando datos sencillos para facilitar su comprensión.

Finalmente, mostraremos las ventanas que despliega el paquete STATISTICA, al llevar un análisis de esta naturaleza. Esto último para que el lector se vaya familiarizando con dichas ventanas pues, a partir de este capítulo, las veremos con frecuencia.

3.1 Técnicas jerárquicas.

En una clasificación jerárquica los datos no son particionados en un número en especial de clases o conglomerados en un sólo paso. En lugar de eso la clasificación consiste en una serie de particiones que pueden ir desde un solo conglomerado que contenga a todos los individuos, hasta n conglomerados, cada uno de ellos, conteniendo a un solo individuo. Las técnicas jerárquicas de conglomerado se pueden subdividir en métodos aglomerativos, los cuales, proceden a hacer una serie sucesiva de fusiones de n individuos en grupos; y métodos divisivos, los cuales separan los n individuos sucesivamente en agrupamientos más "finos". Ambas formas de agrupamientos jerárquicos pueden ser vistos como un intento de encontrar el paso más

eficiente, bajo un cierto criterio definido, en cada etapa en la subdivisión progresiva o en la síntesis de datos. Con tales métodos, las divisiones o fusiones una vez hechas son irrevocables, de manera que cuando un algoritmo aglomerativo ha unido a dos individuos éstos no pueden ser separados de manera subsecuente, y cuando un algoritmo divisivo ha hecho una separación, los individuos separados no pueden reunirse en un paso sucesivo.

Ahora bien, ya que todos los algoritmos aglomerativos jerárquicos reducen los datos a un solo conglomerado que contiene a todos los individuos, y los algoritmos divisivos finalmente separarán al conjunto entero de datos en n grupos cada uno de ellos conteniendo a un solo individuo, el investigador que desee una solución con un número "óptimo" de conglomerados, necesitará detenerse en un paso en particular. El problema de decidir cuál es el número correcto de conglomerados se discutirá más adelante.

Las clasificaciones jerárquicas pueden ser representadas por un diagrama bi-dimensional conocido como dendograma el cual ilustra las fusiones o divisiones hechas en cada etapa sucesiva del análisis. La estructura de dichos diagramas se asemeja a un árbol de evolución y es en las aplicaciones biológicas que las clasificaciones jerárquicas son probablemente más importantes (sin embargo, como veremos después, los métodos jerárquicos de conglomerado pueden ser aplicados en muchas otras áreas).

3.1.1 Métodos aglomerativos.

Un procedimiento aglomerativo jerárquico de conglomerado produce una serie de particiones de datos $P_n, P_{n-1}, \ldots, P_1$. El primero, P_n, consiste de n conglomerados con un único miembro, el último P_1, consiste de un solo grupo que contiene a todos los n individuos. La forma de operar de todos los métodos de este estilo es similar.

A cada etapa particular los métodos unen a los individuos o grupos de individuos que se encuentran más cercanos (o más similares). Las diferencias de los métodos surgen gracias a las distintas formas de definir la distancia (o similitud) entre un individuo y un grupo que contiene varios individuos, o entre dos grupos de individuos. Algunas técnicas jerárquicas aglomerativas se describirán en detalle y por conveniencia dicha descripción será en términos de medidas de distancia.

3.2 Objetivos del análisis de conglomerados.

El objetivo primario del análisis de conglomerados es particionar un conjunto de objetos en dos o más grupos basados en la similitud de los objetos para un conjunto especificado de características (variable de conglomerado). El uso más tradicional del análisis de conglomerados ha sido para propósitos exploratorios. En un modo

exploratorio, el análisis de conglomerados es más frecuentemente usado para desarrollar una clasificación objetiva de los objetos en estudio. Pero no sólo eso. La capacidad que tiene el análisis de conglomerados de particionar puede también generar o confirmar hipótesis que estén relacionadas con la estructura de los objetos. Si se define una estructura para un conjunto de objetos, se puede aplicar el análisis de conglomerados y se compara el resultado derivado de éste con la estructura propuesta anteriormente.

En cualquier aplicación, los objetivos del análisis de conglomerados no pueden ser separados de la selección de las variables usadas para caracterizar los objetos sometidos al conglomerado. Sin importar si el objetivo es la exploración o confirmación. Los conglomerados obtenidos pueden solamente reflejar la estructura inherente de los datos dada la definición de las variables.

¿Qué son las variables? Las variables son las cosas que medimos, controlamos o manipulamos en nuestra investigación. Estas difieren en muchos aspectos, de manera más notoria en el rol que juegan en nuestra investigación y en el tipo de medidas que se les puede aplicar

Investigación experimental vs. investigación correlacional. La mayoría de las investigaciones empíricas pertenecen de manera clara a una de esas dos categorías: En la investigación correlacional nosotros no influenciamos (o al menos, tratamos de no hacerlo) a ninguna de las variables sino que solamente las medimos y buscamos relaciones (correlaciones) entre algún conjunto de variables, como la presión arterial y el nivel de colesterol. En la investigación experimental, manipulamos algunas variables y entonces medimos los efectos de esta manipulación sobre las otras variables; por ejemplo, un investigador puede aumentar de manera artificial la presión sanguínea y entonces medir el nivel de colesterol. El análisis de datos en la investigación experimental también deriva en el cálculo de "correlaciones" entre variables, especificamente, entre esas que fueron manipuladas y esas afectadas por la manipulación. Sin embargo, los datos experimentales pueden, potencialmente, proveer de mejor información cualitativa: sólo los datos experimentales pueden demostrar de manera concluyente las relaciones causales entre variables.
Por ejemplo, si encontramos que siempre que cambiamos una variables A entonces la variable B cambia, entonces podemos concluir que "A influencia a B". Los datos obtenidos de una investigación correlacional pueden ser sólo "interpretados" en términos causales basándose en algunas teorías que tenemos, pero los datos correlacionales no pueden ayudarnos a probar de manera concluyente la causalidad.

Escalas de medición. Las variables difieren entre sí en que tan bien pueden ser medidas, es decir, en que tan medible es la información que la escala de medida nos da. Existe obviamente algún error involucrado en cualquier manera de medir, el cual determina "la cantidad de información" que podemos obtener. Otro factor que determina la cantidad de información que puede darnos una variable es su "tipo de

escala de medición" De manera específica son clasificados como a) nominales, b) ordinales, c) intervalares o d) ratio (relación o razón).

a) Las variables nominales permiten sólo una clasificación cualitativa Esto es, no pueden ser medidas sólo en términos de si un individuo pertenece a alguna de las categorías distintivas, pero no podemos cuantificar o ni siquiera dar un orden jerárquico a esas categorías. Por ejemplo, todo lo que podemos decir es si 2 son diferentes en términos de la variable *A* (en otras palabras, que son de diferente raza) , pero no podemos decir cual de los dos tiene "más " de la cualidad representada por la variable. Ejemplos típicos de variables nominales son: género, raza, color, ciudad, etc.
b) Las variables ordinales nos permiten establecer un orden entre los individuos que medimos en términos de cual tiene menos y cual más de la cualidad representada por la variable, pero de igual manera que la anterior, no nos permite decir "que tanto más". Un ejemplo típico de una variable ordinal es el estrato socioeconómico de las familias. Por ejemplo, sabemos que la clase media-alta es más alta que la media pero nosotros no podemos decir, 18% más alta. De la misma manera existe esta distinción entre escalas nominales, ordinales e intervalares representa un buen ejemplo de una variable ordinal. Por ejemplo, podemos decir que la medida nominal nos da menos información que la medida ordinal, pero no podemos decir "que tanto menos" o como comparamos esta diferencia con la diferencia entre escalas ordinales e intervalares.

c) Las variables intervalares nos permiten no sólo establecer un orden a los objetos que medimos, sino que también nos permiten cuantificar y comparar las medidas de las diferencias entre ellos. Por ejemplo, la temperatura, como se mide en grados Celsius o Fahrenheit, constituyen una escala intervalar. Podemos decir que una temperatura de 40 grados es más alta que una temperatura de 30 grados, y que un incremento de 20 a 40 grados es el doble de un incremento que va de 30 a 40 grados.

d) Las variables de razón son muy similares a las variables intervalares; y además de todas las propiedades de éstas últimas, tienen una característica adicional un punto de cero absoluto, entonces este tipo de variables nos permiten hacer afirmaciones tales como que *X* es dos veces mayor que *Y*. Ejemplos típicos de escalas de razón son medidas de tiempo y espacio. Por ejemplo, la temperatura medida en grados Kelvin es una escala de razón, no sólo podemos decir que una temperatura de 200 grados es más alta que una de 100 grados, podemos correctamente afirmar que es dos veces más alta. Las escalas intervalares no tienen la propiedad de razón.

3.3 Diseño de la investigación en el análisis de conglomerados.

Una vez definido el objetivo y seleccionadas las variables, el investigador debe hacerse tres preguntas antes de empezar el proceso de partición: 1) ¿Se pueden detectar los outliers?, si la respuesta es sí, entonces, ¿Deben eliminarse?, 2) ¿Cómo debe medirse la similitud entre los objetos? y 3) ¿Deben estandarizarse los datos?. Para estas preguntas existen distintas soluciones. Sin embargo, ninguna de ellas ha

sido evaluada de manera suficiente para dar una respuesta definitiva a las mismas, y desafortunadamente, a distintas soluciones corresponden distintos resultados (para el mismo conjunto de datos).

3.3.1 Detección de outliers.

En esta técnica, nos apoyaremos principalmente en el uso de los dendogramas para que nos sugieran que puntos pueden ser outliers. Los identificaremos por puntos que se unirán a los conglomerados solución en las etapas finales del análisis. Una vez identificados estos candidatos, se pueden analizar sus vectores de características para decidir si en efecto son outliers.

3.3.2 Estandarización de datos.

Una vez elegida la medida de similitud, el investigador debe hacerse otra pregunta: ¿Deben estandarizarse los datos antes de calcular las similitudes?. Para contestar esta pregunta, el investigador debe tomar en cuenta varios aspectos. Primero, la mayoria de las distancias son muy sensibles a una diferencia de escalas o de magnitud entre las variables.

La forma más común de estandarizar es la conversión de cada variable a calificaciones estandard (también conocidos como calificaciones *Z*) al restar la media y dividir a cada variable por su desviación estandard. Esta es la forma general de una función de distancia normalizada, que utiliza una medida dada en términos de una distancia euclideana para una transformación normal de los datos por renglón. Este proceso convierte cada renglón de datos en un valor estandard con media cero y desviación estandard unitaria. Esta transformación, elimina el sesgo introducido por las diferencias en escalas de muchos atributos o variables usados en el análisis.

Hasta ahora hemos estandarizado sólo variables. ¿Qué respecto a estandarizar casos?, ¿Porqué tendríamos que hacer esto?. Tomemos un ejemplo simple. Supongamos que hemos recolectado un número de ratings sobre una escala de diez puntos acerca de la importancia de varios atributos en su decisión de compra para un producto. Podemos aplicar análisis de conglomerados y obtener conglomerados, pero una posibilidad muy distinta es que obtuvieramos conglomerados de gente que dijo que todo es importante, algunos diciendo que todo tenía poca importancia, y posiblemente algunos conglomerados intermedios. ¿Lo que estamos viendo son efectos del estilo de respuesta en los conglomerados?. Los efectos del estilo de respuesta son patrones sistemáticos de respuesta a un conjunto de preguntas, tales como los que dicen a todo que si (responden favorablemente a todas las preguntas) o los que dicen que no a todo (responden desfavorablemente a todas las preguntas).

Si queremos identificar grupos de acuerdo a su estilo de respuesta, entonces no es apropiado estandarizar.
Pero en la mayoria de los casos lo que se desea es medir la importancia relativa de una variable con respecto a otra. En otras palabras, es el atributo uno más o menos importante que los otros atributos, y ¿Pueden encontrarse conglomerados de encuestados con patrones similares de importancia?.

3.4 Algoritmos para conglomerar.

Otra pregunta que debe contestarse es: ¿Qué procedimiento debe usarse para poner en grupos a los objetos similares?. Eso es, ¿Qué algoritmo de conglomerado o qué conjunto de reglas es el más apropiado?. Esta no es una pregunta sencilla. El criterio esencial de todos los algoritmos, sin embargo, es que intentan maximizar las diferencias entre conglomerados relativo a la variación interior de los conglomerados.

Los algoritmos más comúnmente usados para el análisis de conglomerados pueden ser clasificados en dos categorías generales: jerárquicos y no-jerárquicos. En el presente trabajo abordaremos solamente los procedimientos jerárquicos de conglomerado.

3.4.1 Algunos procedimientos jerárquicos de conglomerado.

Como ya lo habíamos mencionado anteriormente; los procedimientos jerárquicos de conglomerado involucran la construcción de una jerarquía de estructura similar a las ramas de un árbol. Hay dos tipos de procedimientos jerárquicos de conglomerado, básicamente, aglomerativos y divisivos. En los métodos aglomerativos, cada objeto u observación empieza con su propio conglomerado. En pasos subsecuentes, los dos conglomerados más cercanos (o individuos) son combinados en un nuevo conglomerado agregado, entonces se reduce en uno el número de conglomerados en cada paso. En algunos casos, un tercer individuo une los primeros dos en un conglomerado. En otros, dos grupos de individuos formados en una etapa anterior puede unirse en un nuevo conglomerado. Eventualmente, todos los individuos son agrupados en un sólo conglomerado grande; por esta razón, los procedimientos aglomerativos son a veces llamados métodos "buildup" (de remanentes o excesos). Una característica importante de los procedimientos jerárquicos es que los resultados de una etapa anterior son siempre heredados dentro de los resultados de una etapa posterior, lo cual provoca su semejanza con un árbol.

Cuando el proceso de conglomerado procede en la dirección opuesta a los métodos aglomerativos, se le llama método divisivo. En los métodos divisivos, comenzamos con un conglomerado grande que contiene todas las observaciones (objetos). En pasos sucesivos, las observaciones que son más disimilares se separan y se ponen en conglomerados más pequeños. Este proceso continua hasta que cada observación es un

conglomerado en sí mismo. Nosotros sólo hablaremos de los métodos aglomerativos más usados.

Cinco procedimientos aglomerativos populares para obtener conglomerados son: 1) single linkage, 2) complete linkage, 3) average linkage, 4) Ward's method y 5) centroid. Estos métodos difieren en como se miden las distancias entre conglomerados.

3.4.1.1 Single linkage.

El procedimiento de single linkage está basado en distancias mínimas. Encuentra los dos individuos (objetos) separados por la distancia más corta y los ponen en el primer conglomerado. Entonces la siguiente distancia más corta es encontrada, y un tercer individuo se une a los primeros dos para formar un conglomerado o un nuevo conglomerado de dos individuos es formado. El proceso continua hasta que todos los individuos están en un solo conglomerado. Este procedimiento también es conocido como la aproximación del "vecino más cercano".

La distancia entre cualesquiera dos conglomerados es la más corta distancia de cualquier punto en un conglomerado a cualquier punto en otro conglomerado. Dos conglomerados son fusionados en cualquier etapa por la distancia más corta entre dos elementos o por el lazo más fuerte entre ellos. Los problemas surgen, sin embargo, cuando los conglomerados están pobremente delineados. En tales casos, los procedimientos de single linkage forman largas, cadenas en forma de serpiente, y eventualmente todos los individuos son puestos en una cadena. Individuos en lados opuestos de una cadena pueden ser muy disimilares.

Si C_1 y C_2 son dos conglomerados, entonces la distancia entre ellos se define como la más pequeña disimilitud entre un miembro de C_1 y un miembro de C_2, es decir:

$$d_{C_1C_2} = min\{d_{rs}: r \in C_1, s \in C_2\}$$

donde r denota al objeto r. Mostramos el proceso de fusión con el siguiente ejemplo. Sea

	C_1	C_2	C_3	C_4
C_1	0	7	1	9
C_2	7	0	6	3
C_3	1	6	0	8
C_4	9	3	8	0

La mínima d_{rs} es $a_1 = d_{13} = 1$, entonces los objetos 1 y 3 son juntados y nuestros conglomerados ahora son: (1,3), (2) y (4). Ahora

$d_{(2)(1,3)} = min\{d_{21}, d_{23}\} = d_{23} = 6$

$d_{(4)(1,3)} = min\{d_{41}, d_{43}\} = d_{43} = 8$

y la matriz distancia para los conglomerados es:

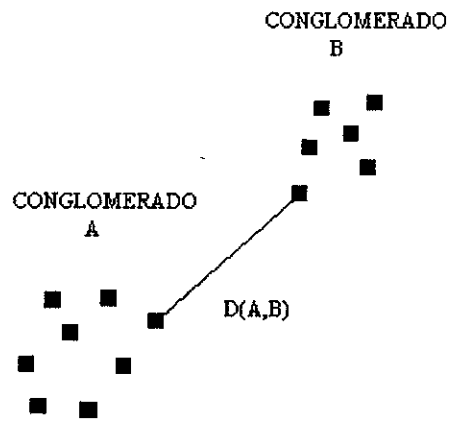
	C_(1,3)	C_2	C_4
C_(1,3)	0	6	8
C_2	6	0	3
C_4	8	3	0

La entrada más pequeña es $a_2 = d_{24} = 3$, entonces los objetos 2 y 4 son juntados y nuestros conglomerados se convierten en : (1,3) y (2,4).

Finalmente, estos dos conglomerados son juntados para dar como resultado un solo conglomerado, a saber: (1,2,3,4). Notemos que

$d_{(1,3)(2,4)} = min\{d_{(1,3)(2)}, d_{(1,3)(4)}\} = d_{(1,3)(2)} = 6$

La característica que define este método es que la distancia entre grupos se define como aquella que sea más pequeña entre un par de individuos, donde sólo se consideran pares de individuos cada uno de ellos proveniente de un grupo distinto. Esta medida de la distancia entre dos grupos se ilustra en la siguiente figura.

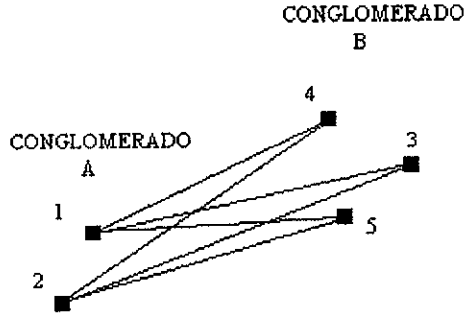


3.4.1.2 Average linkage.

El método de average linkage, también llamado de promedios, comienza igual que el single linkage (y que el complete linkage, como veremos más adelante) pero el criterio de conglomerado es la distancia promedio de todos los individuos en un conglomerado

a todos los individuos en otro conglomerado. Tales técnicas no dependen de los valores extremos, como el single y el complete linkage, y la partición está basada en todos los miembros de los conglomerados más que en un único par de miembros extremos.

Pues bien, como ya habíamos dicho arriba, aquí la distancia entre dos conglomerados se define como el promedio de las distancias entre todos los pares de individuos que se forman de un individuo a los individuos que pertenecen a otro conglomerado. Dicha medida se ilustra con la siguiente figura.



$$d_{AB} = \frac{1}{6}(d_{14} + d_{13} + d_{15} + d_{24} + d_{23} + d_{25})$$

Para ejemplificar la forma en que opera este método habremos de utilizar la siguiente matriz distancia:

	C_1	C_2	C_3	C_4	C_5
C_1	0	2	6	10	9
C_2	2	0	5	9	8
C_3	6	5	0	4	5
C_4	10	9	4	0	3
C_5	9	8	5	3	0

Al aplicar el método de average linkage a la matriz D_1, del ejemplo de single linkage, en la primera etapa, como con el single linkage, se forma un conglomerado que contiene a los individuos 1 y 2. (Espero nos sea claro que si la distancia entre los objetos 1 y 2 es la mínima si sacamos su promedio, es decir, si la dividimos entre 1, que es el número de elementos en este caso, sigue siendo mínima).

Ahora, calculamos un nuevo conjunto de distancias, a saber:

$$d_{(1,2)3} = \frac{1}{2}(d_{13} + d_{23}) = \frac{1}{2}(6+5) = 5.5$$

$$d_{(1,2)4} = \frac{1}{2}(d_{14} + d_{24}) = \frac{1}{2}(10+9) = 9.5$$

$$d_{(1,2)5} = \frac{1}{2}(d_{15} + d_{25}) = \frac{1}{2}(9+8) = 8.5$$

Si acomodamos estos resultados en una matriz, llamémosle matriz D_2, tenemos:

	C_(1,2)	C_3	C_4	C_5
C_(1,2)	0	5.5	9.5	8.5
C_3	5.5	0	4	5
C_4	9.5	4	0	3
C_5	8.5	5	3	0

La entrada más pequeña, o en otras palabras la distancia más pequeña entre dos conglomerados (objetos en este caso) es la formada entre los individuos 4 y 5. La distancia promedio entre los dos, los dos conglomerados formados por dos miembros está dada por:

$$d_{(1,2)(4,5)} = \frac{1}{4}(d_{14} + d_{15} + d_{24} + d_{25}) = \frac{1}{4}(10+9+9+8) = 9$$

$$d_{3(4,5)} = \frac{1}{2}(d_{34} + d_{35}) = \frac{1}{2}(10+9) = 9.5$$

Haciendo uso de estos datos podemos formar una tercera matriz distancia, que llamaremos, necesariamente, D_3, y quedaría formada así:

	C_(1,2)	C_3	C_(4,5)
C_(1,2)	0	5.5	9
C_3	5.5	0	9.5
C_(4,5)	9	9.5	0

Observemos que en D_3 la entrada más pequeña es $d_{(1,2,3)}$ y entonces formamos un nuevo conglomerado con estos elementos, es decir, con los individuos (objetos) 1, 2 y 3. También formamos un nuevo grupo de distancias :

$$d_{(1,2,3)(4,5)} = \frac{1}{6}(d_{14} + d_{15} + d_{24} + d_{25} + d_{34} + d_{35}) = \frac{1}{6}(10+9+9+8+4+5) = 7.5$$

Con esta información formemos, nuevamente, una matriz a la que llamaremos D_4 y es como sigue:

	C_(1,2,3)	C_(4,5)
C_(1,2,3)	0	7.5
C_(4,5)	7.5	0

En el último paso, como ya nos suponemos, estos dos últimos conglomerados se unen para formar uno solo.

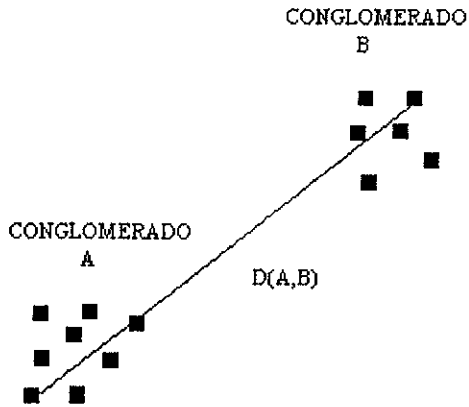
3.4.1.3 Complete linkage.

El procedimiento de complete linkage es similar al de single linkage exceptuando por el criterio de conglomerado que, en este caso se basa en una distancia máxima. Por esta razón, se refiere uno a este método como la aproximación del vecino más lejano o como un método de diámetro. La distancia máxima entre individuos en cada conglomerado representa la más pequeña esfera que puede contener a todos los objetos en conglomerados.

Este método es llamado complete linkage porque todos los objetos en un conglomerado están unidos uno a otro por alguna distancia máxima o por una mínima similitud. Podemos decir que la similitud entre grupos es igual al diámetro del grupo. Esta técnica elimina el problema del serpenteo detectado en el single linkage.

El complete linkage o método del vecino más lejano es lo opuesto al método de single linkage (también conocido como el vecino más cercano) en el sentido de que la distancia entre grupos, ahora, se define como aquella que exista entre el par más distante de individuos, cada uno perteneciente a un grupo distinto.

La forma de medir se ejemplifica en la siguiente figura.



Como ya habiamos mencionado, este método es el opuesto al del vecino más cercano en este nuevo método la distancia entre dos conglomerados es definida en términos de la mayor disimilitud entre un miembro de C_1 y C_2, es decir:

$$d_{C_1C_2} = max\{d_{rs}: r \in C_1, s \in C_2\}$$

donde r denota al objeto r. Mostramos el proceso de fusión con el siguiente ejemplo. Sea la siguiente matriz distancia:

	C_1	C_2	C_3	C_4
C_1	0	7	1	9
C_2	7	0	6	3
C_3	1	6	0	8
C_4	9	3	8	0

Este método comienza como el single linkage; la distancia mínima es $d_{13} = 1$, entonces juntamos los objetos 1 y 3, de esta manera nuestros conglomerados resultan: (1,3), (2) y (4). Ahora

$$d_{(2)(1,3)} = max\{d_{21}, d_{23}\} = d_{21} = 7$$

$$d_{(4)(1,3)} = max\{d_{41}, d_{43}\} = d_{41} = 9$$

y la matriz distancia para los conglomerados es

	C_(1,3)	C_2	C_4
C_(1,3)	0	7	9
C_2	7	0	3
C_4	9	3	0

La entrada más pequeña es $d_{(4)(2)} = 3$, entonces los objetos 4 y 2 son juntados y nuestros conglomerados quedan como: (1,3) y (4,2).

Finalmente, estos dos últimos conglomerados son juntados para dar como resultados un solo conglomerado (1,2,3,4). Notemos que :
$d_{(1,3)(4,2)} = max\{d_{(1,3)(4)}, d_{(1,3)(2)}\} = d_{(1,3)(4)} = 9$

Los tres métodos descritos arriba operan de manera directa sobre la matriz distancia y no necesitan tener acceso a la información que nos proporcionan los valores originales de las variables de los individuos. Como veremos más adelante existen métodos que si hacen uso de los datos originales obtenidos de la población en estudio.

3.4.1.4 Ward's Method.

En el método de Ward la distancia entre dos conglomerados es la suma de cuadrados entre los dos conglomerados sumada sobre todas las variables. En cada etapa en el procedimiento de conglomeración, la suma de cuadrados dentro del conglomerado es minimizada sobre todas las particiones (conjunto completo de conglomerados separados o disjuntos) que se obtiene al combinar dos conglomerados de una etapa anterior. Este procedimiento tiende a combinar conglomerados con un número pequeño de observaciones. También tiende a producir conglomerados con aproximadamente el mismo número de observaciones.

Este método de conglomerado propuesto por Ward busca formar las particiones $P_n, P_{n-1}, \ldots, P_1$ de manera que se minimice la pérdida asociada con cada agrupamiento (o unión) y cuantificar esa pérdida en una forma que sea fácilmente interpretable. A cada paso del análisis, la unión de todo posible par de conglomerados es considerada y los dos conglomerados cuya fusión da como resultado un mínimo incremento en "pérdida de información" son combinados. La pérdida de información es definida por Ward en términos un criterio de error de las sumas de cuadrados, (por sus siglas en inglés ESS).

El razonamiento que está detrás de la propuesta de Ward puede ser ilustrado de manera más simple si consideramos datos en una variable.

Supongamos, por ejemplo, que 10 individuos tienen los siguientes datos (2, 6, 5, 6, 2, 2, 2, 0, 0, 0) en una variable en particular. La pérdida de información que resultaría de tratar los diez datos como un grupo con una media de 2.5 está representada por la ESS, es decir:

$$ESS = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Para este ejemplo,

ESS (de un grupo)= $(2-2.5)^2 + (6-2.5)^2 + \ldots + (0-2.5)^2 = 50.5$

De manera similar sucede si los 10 individuos son clasificados de acuerdo a sus datos dentro de 4 grupos:
{0,0,0},{2,2,2,2},{5},{6,6}, la ESS puede ser evaluada como la suma de cuatro sumas de cuadrados de errores separadas ESS (cuatro grupos) = ESS (grupo 1)+ESS (grupo 2)+ESS (grupo 3)+ESS (grupo 4)= 0.0

3.4.1.5 Centroid.

En el método del centroide la distancia entre dos conglomerados es la distancia (usualmente el cuadrado de la euclideana o la euclideana simplemente) entre sus centroides. Los centroides de los conglomerados son los valores medios de las observaciones sobre las variables en la variable de conglomerado. En este método, cada vez que los individuos son agrupados, un nuevo centroide es calculado. Los centroides de los conglomerados migran a medida que las fusiones van teniendo lugar. En otras palabras, hay un cambio en un centroide de un conglomerado cada vez que un nuevo individuo o grupo de individuos se adiciona al conglomerado existente. Estos métodos son los más usuales entre los biólogos pero pueden producir resultados confusos y desordenados con frecuencia. La confusión surge debido a los reversals, esto es, casos en los cuales la distancia entre los centroides de un par puede ser menor que la distancia entre los centroides de otro par fusionado en una combinación anterior. La ventaja de este método es que se ve menos afectado por los outliers que los otros métodos jerárquicos.

Con este método, los grupos una vez formados son representados por sus valores medios para cada variable, esto es, su vector de medias, y la distancia inter-conglomerados es ahora definida en términos de distancia entre tales vectores

El uso de la media implica estrictamente, que las variables están en una escala intervalar; el método es sin embargo, usado con frecuencia para otro tipo de variables. Para ilustrar como opera el conglomerado por centroides, lo aplicaremos al siguiente grupo de datos bivariados:

INDIVIDUO	VAR 1	VAR 2
1	1	1
2	1	2
3	6	3
4	8	2
5	8	0

Suponga que la distancia euclideana es elegida como la distancia inter-individuo dada la siguiente matriz distancia ·

	C_1	C_2	C_3	C_4	C_5
C_1	0	1	5.38516474	7.07106781	7.07106781
C_2	1	0	5.09901953	7	7.28010988
C_3	5.38516474	5.09901953	0	2 23606801	3.60555124
C_4	7.07106781	7	2.23606801	0	2
C_5	7.07106781	7.28010988	3.60555124	2	0

Llamemos a esta matriz distancia D_1, pues bien D_1 muestra que d_{12} es la entrada más pequeña y los individuos 1 y 2 se fusionan para formar un grupo. El vector de medias del grupo es calculado, (1.0,1.5), y una nueva matriz distancia es calculada. (Es fácil, veamos que es lo que hacemos: primero sumamos las x's, es decir , 1+1=2; este resultado lo dividimos entre el número de elementos sumados, o sea , 2 ; de donde resulta la primera entrada del vector que es igual a 1. Hacemos lo mismo con las y's . Sumamos 1+2=3, y este resultado lo dividimos entre el número de elementos sumados: 2. Resulta de esta operación la segunda entrada del vector de medias que es igual a 1.5).

Ahora calcularemos una nueva matriz distancia utilizando la distancia euclideana, pero con los siguientes datos:

INDIVIDUO	VAR1	VAR2
(1,2)	1	1.5
3	6	3
4	8	2
5	8	0

	C_(1,2)	C_3	C_4	C_5
C_(1,2)	0	5.22015333	7.01783419	7.15891075
C_3	5.22015333	0	2.23606801	3.60555124
C_4	7.01783419	2.23606801	0	2
C_5	7.15891075	3.60555124	2	0

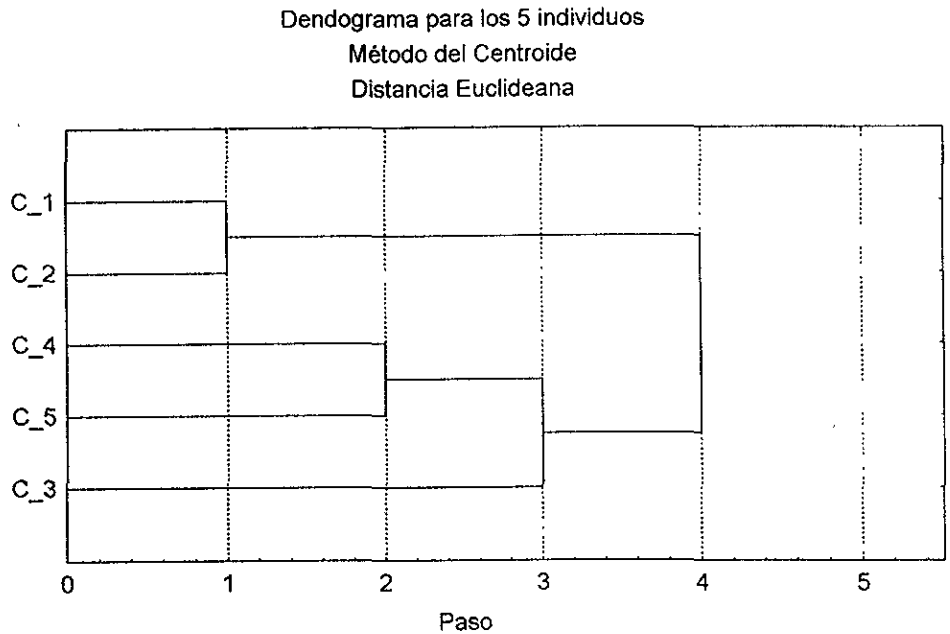
Llamémosle a esta última matriz de distancia D_2, observemos que en D_2 la entrada más pequeña es d_{45} y los individuos 4 y 5, por tanto, se unen para formar un segundo grupo, a esta última matriz le calculamos el vector de medias, que resulta ser (8 0,1.0) (ver explicación en el parráfo anterior).

Calculamos, otra vez, una nueva matriz distancia, pero ahora con los siguientes elementos.

INDIVIDUO	VAR1	VAR2
(1,2)	1	1.5
3	6	3
(4,5)	8	1

	C_(1,2)	C_3	C_(4,5)
C_(1,2)	0	5.22015333	7.01783419
C_3	5.22015333	0	2.82842708
C_(4,5)	7.01783419	2.82842708	0

En D_3 la distancia más pequeña es $d_{(4,5)3}$ y entonces los *individuos 3,4 y 5 son* fusionados en un conglomerado de tres miembros. La etapa final consiste en la fusión de los dos grupos, o conglomerados, restantes en uno solo. El dendograma resultante aparece, también a continuación.



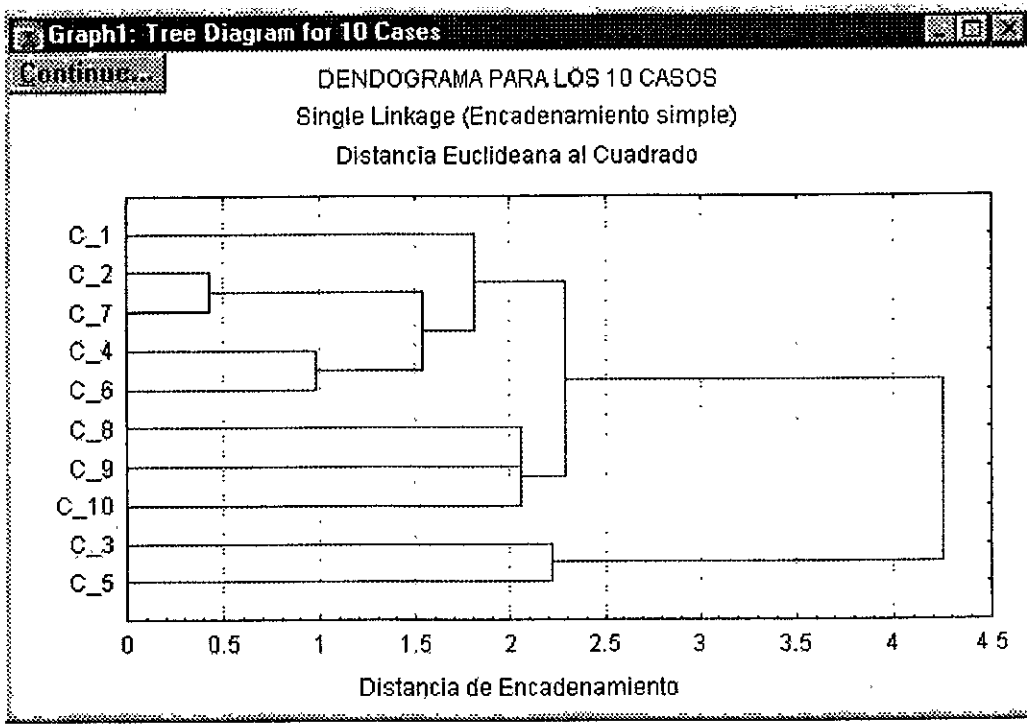
3.5 ¿Cuántos conglomerados deben formarse?

Un tema de mucho interés respecto a las técnicas de conglomerado es el cómo elegir el número de conglomerados con los que se quedará el investigador al final del proceso. Hay muchos criterios y líneas a seguir propuestas para la solución del problema. Desafortunadamente, ningún procedimiento típico, objetivo de selección existe. Las distancias entre los conglomerados en los pasos sucesivos puede servir como una guía útil, y el analista decide parar cuando esta distancia sobrepasa un valor especificado o cuando las distancias sucesivas entre pasos hacen un salto repentino. Estas distancias son llamadas algunas veces medidas de variabilidad del error. Además, alguna conceptualización intuitiva de la relación teórica puede sugerir un número natural de

conglomerados. En el análisis final, sin embargo, probablemente es mejor calcular un número distinto de conglomerados de solución (dos, tres, cuatro, etc.) entonces decidir entre las soluciones alternativas mediante el uso de un criterio a priori, juicio práctico, sentido común o fundamentos teóricos. Por otro lado, uno puede comenzar este proceso especificando algún criterio basado en consideraciones prácticas tales como:"Mis hallazgos serán más manejables y fáciles de comunicar si tengo 3 ó 6 conglomerados", y entonces resolver para este número de conglomerados y seleccionar la mejor alternativa después de evaluar esas soluciones predefinidas.
Los conglomerados solución serán mejorados mediante la restricción de la solución de acuerdo a aspectos conceptuales del problema.

3.5.1 Las particiones de una jerarquia, el problema del número de grupos.

Sucede con frecuencia, que cuando se usan las técnicas de conglomerado en la práctica, que el investigador no está interesado en la jerarquía completa sino que sólo en una o dos particiones obtenidas de esta. En el conglomerado jerárquico, las particiones se obtienen cortando un dendograma o seleccionando una de las soluciones en la secuencia anidada de conglomerados que forma la jerarquía. En aplicaciones particulares será interesante tratar de determinar cual de todas las posibles particiones produce el mejor ajuste a los datos; esencialmente esto significa decidir cual es el número apropiado de conglomerados para los datos. Un método informal, pero que se usa con frecuencia, para este propósito es examinar las diferencias en el nivel de fusión en el dendograma. Cambios grandes indican un particular número de conglomerados. Consideremos por ejemplo, el dendograma mostrado a continuación.



Este muestra una diferencia grande en el nivel entre dos grupos, en la etapa final en la cual todos los individuos están en un solo conglomerado. Esto sería tomado como evidencia para considerar la solución de dos grupos como muy relevante. Aunque este procedimiento es comúnmente usado y puede ser útil, también trae consigo la posibilidad de influencia de expectativas hechas a priori.

Soluciones más formales al problema del número de conglomerados en el contexto de los métodos de conglomerados jerárquicos han sido sugeridos por muchos autores. Duda y Hart (1973), por ejemplo, proponen un criterio de razón, E(2)/E(1), donde E(2) es la suma de los errores al cuadrado intraconglomerados cuando los datos están particionados en dos conglomerados, y E(1) da los errores al cuadrado cuando sólo un conglomerado está presente. La hipótesis de un conglomerado es rechazada si la razón es más pequeña que un valor crítico especificado.

Calinski y Harabasz (1974) también sugieren un índice para un número de conglomerados basado en la suma de términos al cuadrado, a saber

$$\frac{traza(B)/(m-1)}{traza(W)/(n-m)}$$

donde B y W son la suma de cuadrados entre e intra conglomerados y de productos cruzados, y m es el número de grupos. El máximo valor del índice en la jerarquía es tomado para indicar el número correcto de grupos.

$W = \frac{1}{n-m}\sum_{i=1}^{m}(n_i - 1)C_i$ donde C_i es la matriz de covarianza muestral del i-ésimo grupo

$B = \frac{1}{m-1}\sum_{i=1}^{m} n_i(\bar{y}_i - \bar{y})(\bar{y}_i - \bar{y})^t$ donde $\bar{y} = \sum_{i=1}^{m} n_i \bar{y}_i / n$; $\bar{y}_i$ es la media para la i-ésimo grupo y n_i el número de individuos en el i-ésimo grupo.

Mojena (1977) sugiere un procedimiento basado en las medidas relativas de los diferentes niveles de fusión en el dendograma. En detalle, la propuesta es seleccionar el número de grupos correspondientes a la primera etapa en el dendograma que satisfaga

$$\alpha_{j+1} > \bar{\alpha} + ks_\alpha$$

donde $\alpha_0, \alpha_1, .., \alpha_{n-1}$ son los niveles de fusión correspondientes a las etapas con n, $n-1$, ..., 1 conglomerados. Los términos $\bar{\alpha}$ y s_α son, respectivamente, la media y la desviación estandard insesgada de los valores α y k es una constante. Mojena sugiere que los valores de k se encuentren en un rango de 2.75 a 3 5 para mejores resultados.

Milligan y Cooper (1985) reportan una investigación de índices para el número de grupos y encontraron que los tres antes mencionados están entre los más satisfactorios, aunque ellos sugieren que el valor de k para la regla de Mojena debería ser 1.25.

3.6 Interpretación de los conglomerados.

La etapa de interpretación involucra examinar la variable de conglomerado para nombrar o asignar una etiqueta que describa exactamente la naturaleza de los conglomerados. Para aclarar este proceso, tendremos que esperar un poco. Abordaremos esta etapa del análisis con mayor detalle en el capítulo 5.

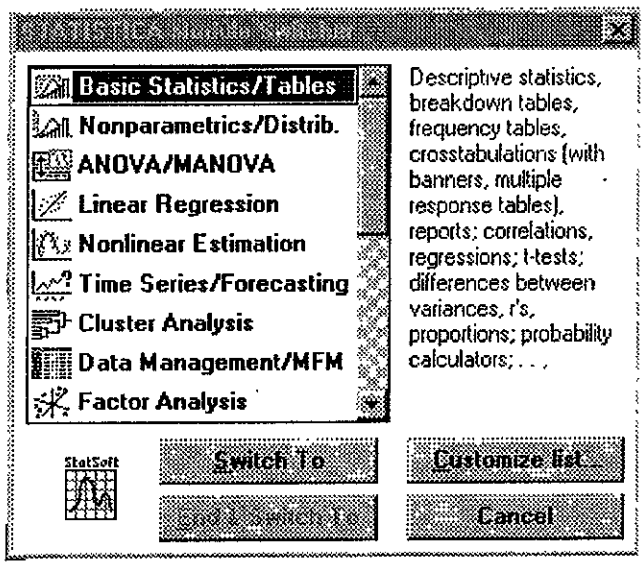
3.7 Los elementos básicos del módulo Cluster Analysis (Análisis de Conglomerados) de STATISTICA.

Ya hablamos un poco acerca de la teoría que envuelve al análisis de conglomerados pero, existe también un poco de teoría desde un punto de vista más operativo. Nos

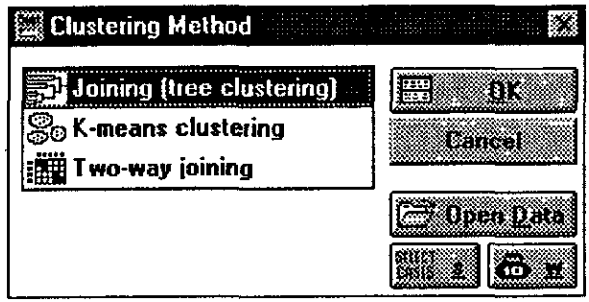
referimos a ciertas características del módulo Cluster Analysis de STATISTICA. Vamos a revisar algunas de ellas.

En esta sección se describen cada una de las ventanas que STATISTICA despliega a lo largo del análisis en el orden en que estas aparecen. En cada una de ellas se tiene el botón OK que permite pasar a la siguiente ventana y el botón CANCEL que regresa a la anterior.

Panel de inicio.
Al seleccionar un módulo, una breve descripción de los procedimientos que realiza aparece junto al menú.En nuestro caso, el módulo de interés es el de Cluster Analysis. En esta caja de diálogo hacemos click sobre la opción Cluster Analysis, y después sobre el botón Switch To (Cambiar a)



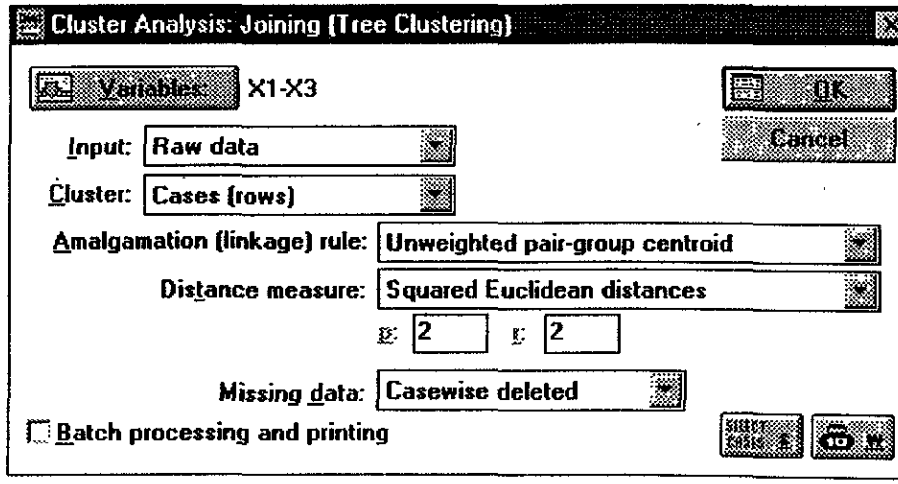
Al activar por primera vez el módulo Cluster Analysis se despliega el correspondiente panel de inicio que permite definir las condiciones iniciales para el análisis:



La opción que habremos de utilizar a lo largo de los siguientes capítulos es *Joining* (tree clustering) pues, como ya habiamos mencionado anteriormente utilizaremos solamente técnicas jerárquicas.

Joining (tree clustering). El propósito de este algoritmo es unir objetos en conglomerados que van creciendo de manera sucesiva, usando alguna medida de similitud o distancia. Un resultado típico de este tipo de conglomeración es un árbol jerárquico.

Como la opción Joining está seleccionada, hacemos click en el botón OK. Aparecerá la siguiente caja de diálogo.



Variables. Al escoger esta opción aparecerá una ventana para seleccionar variables. Note que las variables que se seleccionen aquí serán interpretadas por el programa como dimensiones si en la caja de diálogo Cluster Analysis la opción Cluster indica

"cases"; y serán interpretadas como objetos si en la caja de diálogo Cluster Analysis la opción Cluster indica "variables".

Input Data. Este botón nos permite seleccionar ya sea Raw Data (datos por renglones) o una matriz (seleccionando Distance Matrix) como información de entrada para el Análisis de Conglomerados. La matriz accesada puede ser una matriz de correlación, o una matriz distancia (disimilitud) con números que indican las distancias o disimilitudes entre los objetos. Si la matriz accesada es una matriz de correlación (la cual indica la similitud y cercanía entre objetos), ésta es convertida a distancias antes de que comience el análisis.

Cluster Cases or Variables. Este botón permite al usuario seleccionar entre conglomerar Variables (columnas) o conglomerar Cases (renglones).

Amalgamation (linkage) rule. Uno de los principales parámetros que guían el proceso de unión (tree-clustering) es la regla de encadenamiento, esto es, la regla que determina cuando dos conglomerados serán unidos (encadenados o amalgamados). Hay siete diferentes reglas de unión disponibles: Single Linkage, Complete Linkage, Unweighted Pair group average, weighted pair-group average, unweighted pair group centroid, weighted group centroid (median) y el Método de Ward.. La opción por default es el single linkage (también llamado método del vecino más cercano).

a) single linkage. En este método la distancia entre dos conglomerados se determina por la distancia de los dos objetos que se encuentren más cercanos (vecinos más cercanos) en los diferentes conglomerados. Esta regla, de alguna manera, encadena objetos para formar conglomerados, y los conglomerados resultantes tienden a representar largas cadenas.

b) complete linkage. En este método, las distancias entre los conglomerados se determinan mediante la distancia más grande entre cualesquiera dos objetos en los diferentes conglomerados (es decir, por los vecinos más lejanos). Este método funciona bastante bien en situaciones en que los objetos forman, de hecho, de manera natural distintas "nubes". Si los conglomerados tienden a formas elongadas o de una cadena, entonces el método es inapropiado.

c) unweigthed pair-group average (promedios no ponderados). En este método, la distancia entre dos conglomerados se calcula como la distancia promedio entre todos los objetos en dos diferentes conglomerados. Este método es además muy eficiente cuando los objetos forman "nubes" distintas de manera natural, sin embargo, funciona igual de bien con conglomerados de tipo cadena o elongados.

d) weighted pair-group average (promedios ponderados). Este método es idéntico al anterior, excepto porque en los cálculos, la medida de los respectivos conglomerados (es decir, el número de objetos contenidos en ellos) es usado como ponderador.

Entonces, este método (más que el anterior) debe ser usado cuando la medida de los conglomerados se sospecha que es en extremo desigual.

e) unweighted pair-group centroid (centroide no ponderado). El centroide de un conglomerado es el punto promedio en el espacio definido por sus dimensiones. De alguna forma, es el centro de gravedad de su respectivo conglomerado. En este método la distancia entre dos conglomerados se define como la diferencia entre sus centroides.

f) weighted pair-group centroid (centroide ponderado o median). Este método es idéntico al anterior, excepto que se introduce una ponderación dentro de los cálculos para tomar en consideración las diferencias existentes entre las medidas de los conglomerados (es decir, el número de elementos contenidos en ellos). Entonces, cuando existen (o cuando sospechamos que las hay) diferencias considerables en las medidas de los conglomerados, este método es preferible al expuesto anteriormente.

g) Ward's method. Este método es distinto a todos los métodos porque utiliza un análisis de la varianza encaminado a evaluar las distancias entre los conglomerados. En pocas palabras, este método busca minimizar la suma de cuadrados de cualesquiera dos conglomerados (hipotéticos) que pueden ser formados a cada paso. En general, se considera este método como muy eficiente, sin embargo, tiene tendencia a crear conglomerados pequeños.

Distance measure. El algoritmo de unión empieza por el cálculo de una matriz distancia entre los objetos que habrán de ser conglomerados. Hay seis diferentes medidas de distancia que pueden ser calculadas en la opción Raw Data (ver arriba): cuadrado de la distancia euclideana, distancia euclideana, city block (Manhattan), distancia de Chebychev, Potencia de distancia y Porcentaje de Desacuerdo. No abordaremos la distancia 1-Pearson r.

Si la opción de Potencias de distancias (Power Distances) se selecciona, entonces podemos especificar los dos parámetros p y r para la potencia de la distancia en las cajas de edición debajo de esta botón.

a) Euclidean distances (distancia euclideana). Esta es, probablemente, el tipo más común de distancia elegida. Simplemente es la distancia geométrica en el espacio multidimensional. Se calcula como sigue:

$$D(x,y) = \left[\sum_{i=1}^{n} (x_i - y_i)^2 \right]^{1/2}$$

b) Squared euclidean distances (cuadrado de la distancia euclideana). Podemos necesitar elevar al cuadrado la distancia euclideana estandard para darle un mayor peso a los objetos que se encuentran más alejados. Esta distancia se calcula:

$$D(x,y) = \left[\sum_{i=1}^{n} (x_i - y_i)^2 \right]$$

c) City-block (Manhattan) distances. Esta distancia es simplemente la diferencia promedio entre las dimensiones. En muchos casos, esta medida de la distancia nos da resultados similares a la distancia Euclideana Simple. Sin embargo, notemos que en esta medida,el efecto de diferencias únicas muy grandes es disminuido (ya que no están al cuadrado). La distancia de Manhattan se calcula como:

$$D(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$

d) Chebychev distance metric. Esta medida de la distancia puede ser apropiada en los casos en que uno quiere definir dos objetos como diferentes en cualquiera de sus dimensiones (o variables). La distancia de Chebychev se calcula como:

$$D(x,y) = max|x_i - y_i|$$

e)Power distance. Esta distancia se calcula como sigue:

$$D(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^{p} \right)^{\frac{1}{r}}$$

donde r y p son parámetros definidos por el usuario. El parámetro p controla el peso que se le da a las mayores diferencias entre las dimensiones de los individuos, el parámetro r controla el peso a las mayores diferencias entre los individuos. Si r y p son iguales a 2, entonces esta distancia se convierte en la distancia euclideana.

f) Percent disagreement. Esta medida es particularmente útil si los datos para las dimensiones (variables) incluyen también algunos de naturaleza categórica. Esta distancia se calcula como:

$$D(x,y) = (\#de\_x_i \neq y_i) / i$$

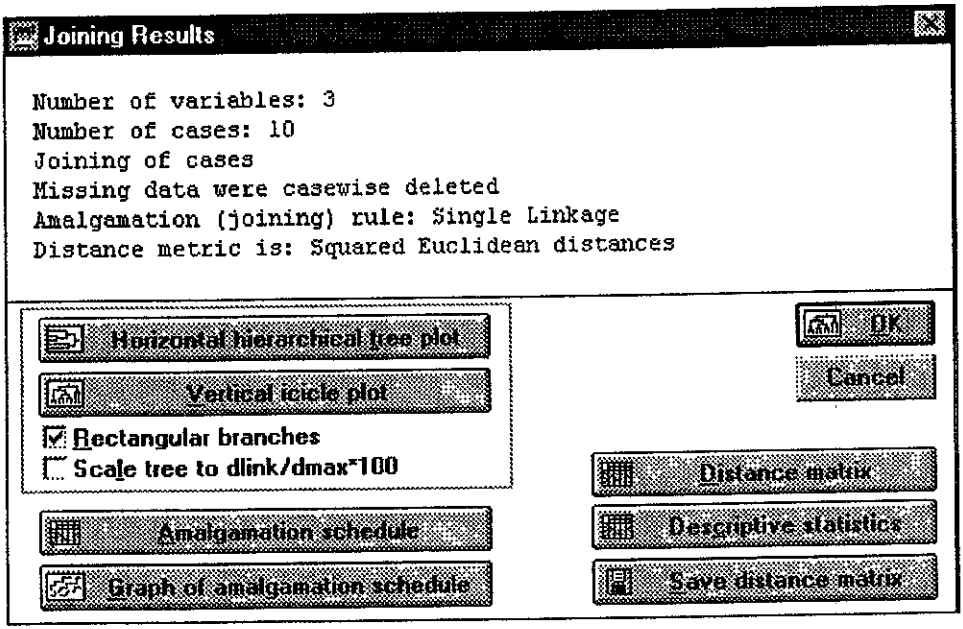
Missing data: Esta opción nos permite seleccionar uno de dos posibles tratamientos a los datos faltantes o perdidos.

a) casewise deletion of missing data: Cuando se selecciona esta opción , un caso será eliminado del análisis se tiene datos faltantes en cualquiera de las variables seleccionadas para el análisis.

b) substition by means. Cuando esta opción es seleccionada, los datos faltantes son sustituidos por las medias correspondientes a sus dimensiones. Notemos que la sustitución por las medias es efectuada después de que el programa ha determinado cuales son las variables (lógicas, es decir dimensiones) y cuales son los casos (lógicos, es decir objetos) para el análisis respectivo. Por ejemplo, cuando se unen casos (tree clustering), lógicamente, las variables se convierten en casos y los casos en variables. En ese caso, una matriz distancia será calculada para los casos, tratando a las variables como dimensiones. Cuando el programa calcula las medias para sustituir los datos faltantes, se calcularán las medias para cada caso, para todas las variables.

Batch processing and printing. Esta opción está sólo disponible si en la ventana de salida fue seleccionado(a): printer, disk file, y o text/output en la caja de diálogo Page/Output Setup.

Ahora, si hacemos click en OK una vez hecha nuestra elección aparece la pantalla de Joining Results



Horizontal hierarchical tree plot. Este botón da como resultado un diagrama horizontal de árbol, el cual resume como se van formando los conglomerados de manera sucesiva.

Vertical icicle plot. Al hacer click en este botón obtenemos como resultado un diagrama de árbol, pero a diferencia del botón anterior, éste se presenta de manera vertical (en el eje vertical).

Rectangular branches. Podemos elegir el desplegado de ambos tipos de diagrama de árbol (ver arriba) con ramas rectangulares (si se selecciona la opción) o diagonales (si no se selecciona la opción).

Scale tree to dlink/ dmax*100. Cuando seleccionamos esta opción, el diagrama de árbol será escalado en una escala estandarizada. De otra manera, si no se selecciona, la escala estará basada en la distancia de encadenamiento del Startup Panel.

Amalgamation Schedule. Al hacer click en este botón aparecerá una hoja cuadriculada con los datos de la amalgamation schedule (ruta de conglomeración). La primera columna de la hoja contendrá las distancias de encadenamiento a la cual los respectivos conglomerados son formados (como se indican en los renglones respectivos), y cada fila contiene los nombres de los objetos (casos o variables) que comprenden el conglomerado respectivo.

Graph of amalgamation schedule. Esta opción hará aparecer una gráfica de línea (similar a una escalera) donde aparecen las distancias entre dos pasos consecutivos del proceso de encadenamiento. Esta gráfica es útil para identificar etapas (pasos) en donde muchos conglomerados son formados a una distancia muy parecida. Esto puede indicar una discontinuidad natural en términos de las distancias entre los objetos observados.

Distance matrix. Esta opción hace que aparezca en pantalla una hoja estandard con la matriz distancia. Esta matriz puede ser guardada usando la opción Save Distance Matrix (ver abajo).

Descriptive statistics. Esta opción desplegará una pequeña hoja estandard (parecida a las de Excel) con medias y desviaciones estandard para cada objeto incluido en el análisis de conglomerados (es decir, para cada caso o variable, dependiendo de la elección hecha en la caja de dialogo Cluster en el Panel de inicio (startup panel).

Save distance matrix. Este botón nos permite guardar la matriz de distancia actual (ver arriba) como un "matrix file" (archivo matriz) standard de STATISTICA. Este archivo puede ser accesado después con los módulos Cluster Analysis o Multidimensional Scaling.

Estas últimas opciones de las cuales hablamos, nos quedarán más claras haciendo uso de un ejemplo. Dicho ejemplo y la utilización detallada de las opciones son el objeto de atención de nuestro siguiente capítulo.

CAPÍTULO 4.

LAS OPCIONES DE CLUSTER ANALYSIS (ANÁLISIS DE CONGLOMERADOS) DE STATISTICA.

Introducción.

Ahora bien, en este capítulo intentaremos ilustrar con un conjunto de datos el uso de las opciones de STATISTICA para la realización de un Análisis de Conglomerados. A lo largo del capítulo, analizaremos nuestro conjunto de datos utilizando 3 distintas técnicas de las expuestas en el capítulo anterior y haremos una comparación entre los distintos resultados que obtenemos en este caso particular. El objetivo que perseguimos, por ahora, es familiarizarnos con el uso de las distintas opciones que tenemos para realizar un análisis de este tipo. Por ahora, no pretendemos llegar a ninguna conclusión. En el capítulo final, haremos más hincapié en las consideraciones pertinentes y en la interpretación de los resultados.

4.1 Clasificación de comidas.

Una empresa de Investigación de Mercados realizó un proyecto relativo a la percepción de diferentes tipos de comida:
1. Japonesa , 2. Cantonesa (China), 3. Szechuan (China), 4. Francesa, 5. Mexicana, 6. Mandarina (China), 7. Americana (E.U.A.), 8. Española, 9. Italiana y 10. Griega.
La opinión de 50 entrevistados se basó en su respuesta a 3 preguntas incorporadas en una escala bipolar de 7 puntos, cuyos extremos son:
X1=Sencilla (1) o Condimentada (7), X2=Ligera o Pesada (7), X3= Baja en Calorías o Alta en Calorias (7),
La tabla de promedios para las 3 variables es la siguiente:

	X1	X2	X3
JAPONESA	2.8	3.2	3.4
CANTONESA	2.6	5.3	5.4
SZECHUAN	6.6	3.6	3.0
FRANCESA	3.5	4.5	5.1
MEXICANA	6.4	4.3	4.3
MANDARINA	3.4	4.1	4.2
AMERICANA	2.3	5.8	5.7
ESPAÑOLA	4.7	5.4	4.9
ITALIANA	4.6	6.0	6.2
GRIEGA	5.3	4.7	6 0

Bueno, y ahora con estos datos vamos a realizar un análisis de conglomerados utilizando el cuadrado de la distancia euclideana y los siguientes métodos de conglomeración:

a) Encadenamiento simple.
b) Promedios (entre conglomerados).
c) Centroide.

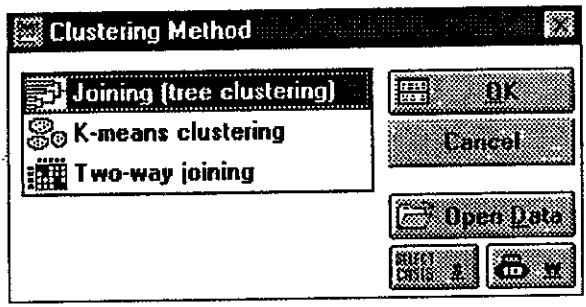
4.1.1 Encadenamiento simple.

Pues bien, vamos a llevar a cabo el desarrollo de este análisis paso por paso. Primero, veamos de que manera aparecen ante nosotros los datos de las diez distintas comidas en STATISTICA:

Data: COMIDA.STA 12v * 10c

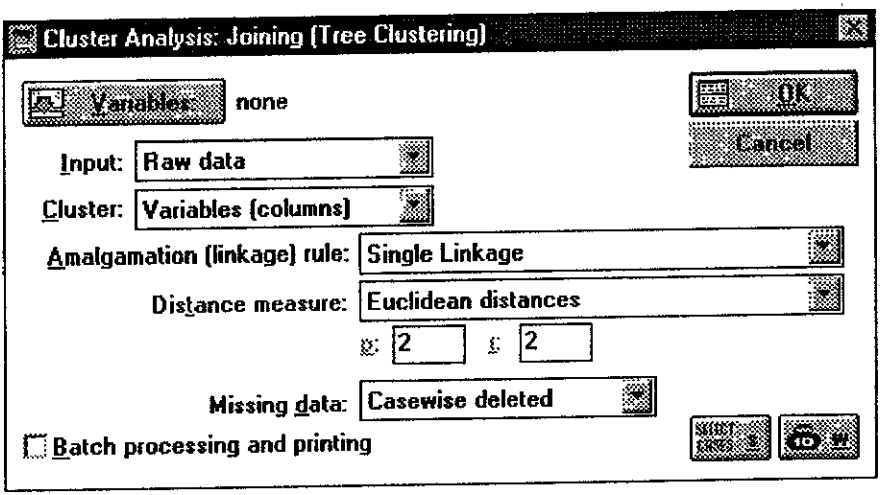
#	1 X1	2 X2	3 X3	4 VAR4	5 VAR5	6 VAR6	7 VAR7	8 VAR8	9 VAR9
1	2.800	3.200	3.400						
2	2.600	5.300	5.400						
3	6.600	3.600	3.000						
4	3.500	4.500	5.100						
5	6.400	4.300	4.300						
6	3.400	4.100	4.200						
7	2.300	5.800	5.700						
8	4.700	5.400	4.900						
9	4.600	6.000	6.200						
10	5.300	4.700	6.000						

Ahora bien, al seleccionar la opción Analysis de la barra de menú principal aparece en la pantalla la siguiente caja de diálogo:



En este caso, vamos a elegir la primera opción de las tres posibles: Joining (tree clustering).

Para hacerlo sólo debemos hacer click en dicha opción y luego en el botón de OK de la caja de diálogo. Al hacer esto aparecerá en nuestra pantalla otra caja de diálogo, que presentamos a continuación:



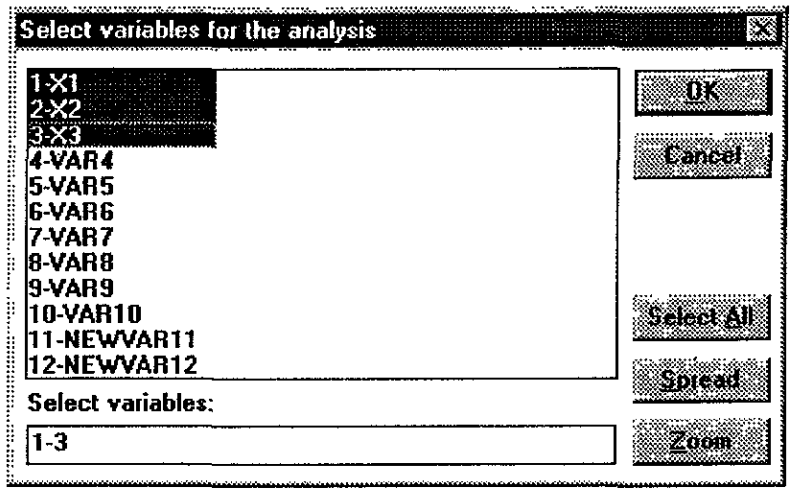
Esta es la pantalla que aparece por default en esta opción pero nosotros en Cluster elegiremos la opción Cases (rows) en vez de Variables (columns). Para lograr que se desplieguen las distintas opciones disponibles hacemos click en la flecha que se encuentra inscrita en los pequeños rectángulos de la derecha.

En el caso de la elección de la Amalgamation (linkage) rule que viene a ser para nosotros el método de conglomeración que habremos de utilizar, no lo cambiamos lo dejamos en Single Linkage (vecino más cercano).

Pasando ahora a la cuestión de que distancia utilizar, nosotros en esta ocasión queremos utilizar el cuadrado de la euclideana (opción que habremos de buscar haciendo, como ya dijimos anteriormente, click en el rectángulo de la derecha que tiene una flecha inscrita).

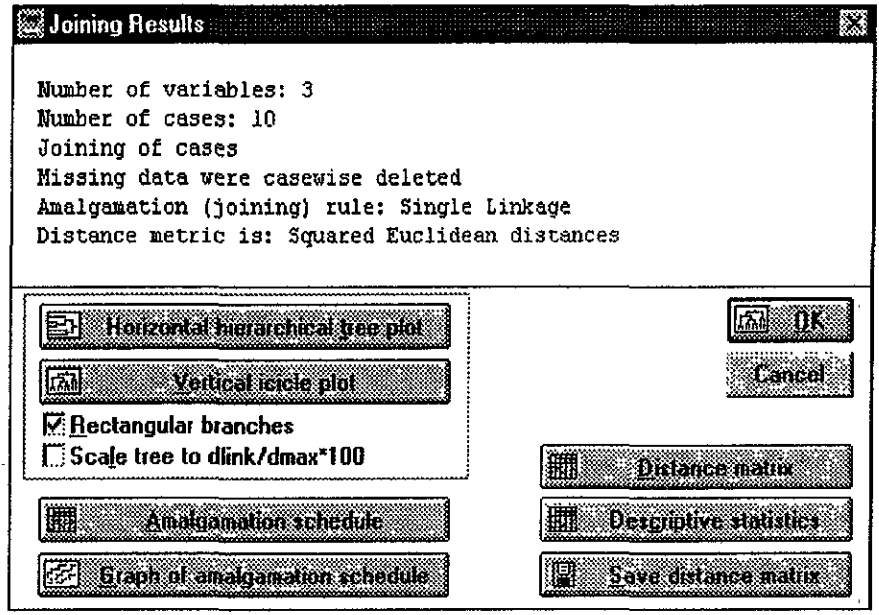
En este caso, no tenemos datos faltantes, de donde, la regla que impongamos para los datos faltantes es intrascendente para este análisis.

Dejamos la opción de Variables hasta el final porque al hacer click en este botón aparece otra caja de diálogo, que mostramos a continuación:

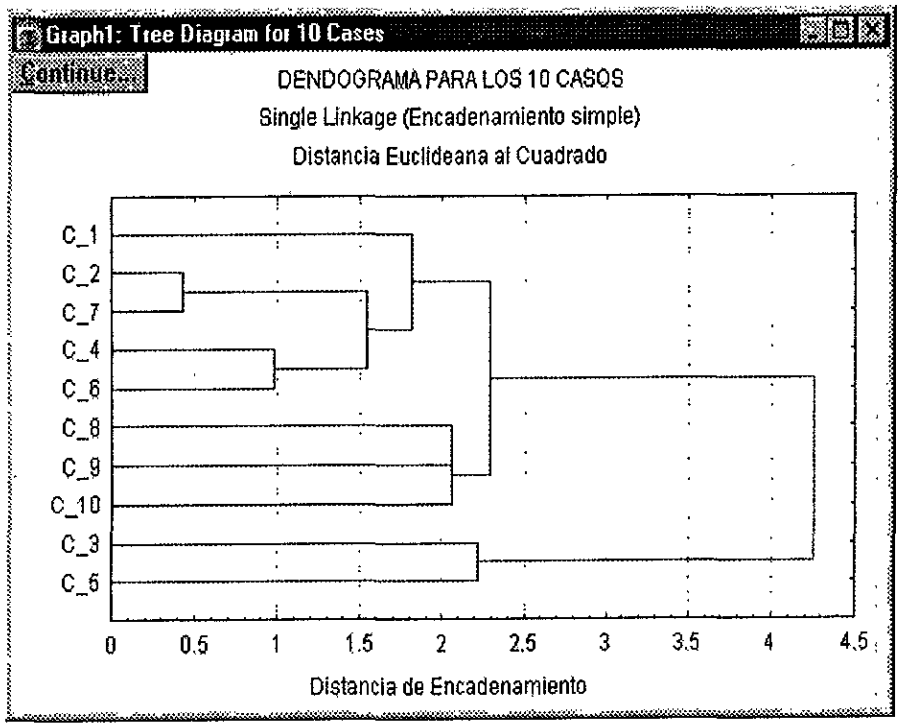


Pues bien, una vez que hemos definido las condiciones bajo las cuales se llevará a cabo el análisis, hacemos click en el botón de OK de la caja de diálogo. Al hacer esto aparecerán los resultados del análisis de conglomerados bajo las condiciones que definimos.

En la parte superior de la ventana Joining Results, podemos observar una especie de resumen de las condiciones bajo las que se efectuó el análisis.



Veamos que sucede al hacer click en el botón de Horizontal Hierarchical Tree. Lo que vemos aquí es el dendograma correspondiente al análisis hecho con los datos de las 10 distintas comidas.



La forma de interpretarlo es la siguiente. Por una parte vemos que en el eje vertical están representados los conglomerados (que consisten en un sólo objeto al principio) del 1 al 10. En el eje horizontal, distancias entre conglomerados.

Veamos con cuidado el dendograma. Los primeros conglomerados en ser unidos fueron el C2 y el C7, es decir, la comida cantonesa y la americana, había una distancia entre ellos menor a 0.5 unidades.

La explicación podría ser la siguiente. Se condimentan de manera similar. La semejanza que existe entre los platillos pertenecientes a estas dos cocinas es el sabor dulce que tienen la mayoría de ellos. Ambas cocinas hacen uso de distintos tipos de carne así que la cantidad de calorías que aportan los platillos de la cocina cantonesa y americana deben ser parecidos al igual que la pesadez o ligereza de los mismos.

Por otra parte, si sólo nos fijamos en su vector de características nos daríamos cuenta de que las calificaciones no difieren mucho entre sí. La calificación que más varía es X2. Dicha variación es de 0.5 unidades.

Después, en el "segundo paso" los conglomerados que se unieron fueron C4 y C6, es decir, la francesa y la mandarina, había una distancia entre ellos menor a 1 unidad.

Veamos que tienen en común estas dos cocinas. Por una parte en la comida francesa se hace uso de ajo y cebolla, entre otras cosas, para dar sabor a los platillos. La comida mandarina también utiliza esta última y en lugar de ajo, utiliza otro tipo de especias. Por tanto, los condimentos son similares. El número de calorías que aportan no debe ser muy distinto.

En este caso, los vectores de características vuelven a ser parecidos entre sí pero, en la tercera calificación, X3, se puede observar una diferencia de 0.9 unidades.

En el "tercer paso" los conglomerados que se fusionaron fueron aquellos que se formaron en los dos pasos anteriores, ambos formados por dos objetos, a saber, C2 y C7, uno y, C4 y C6, el otro. De esta unión nace un conglomerado formado por cuatro elementos : cantonesa, americana, francesa y mandarina, había entre ellos una distancia un poco mayor a 1.5 unidades.

El comentario que podemos hacer en esta etapa del análisis es, que parece natural el hecho de que eventualmente la comida cantonesa y la comida mandarina estén en un mismo conglomerado, ya que ambas son comidas chinas y deben tener muchas semejanzas entre sí.

En el "cuarto paso", los conglomerados que se unen, difieren bastante en cuanto a "tamaño" se refiere, se fusionan el conglomerado C1 (con un solo elemento) y el conglomerado formado en el paso anterior (C2, C4, C6 y C7) resultando un conglomerado de 5 elementos, que son: japonesa, cantonesa, americana, francesa y mandarina. Entre ellos la distancia existente era de 1.8 unidades.

Veamos que semejanzas pueden existir entre estas distintas comidas. La comida japonesa utiliza vino para ciertos platillos, por ejemplo, el mirín que es un vino dulce de sabor similar al jerez, con la diferencia de que esta hecho a base de arroz. Por otra parte, la comida japonesa hace uso de una pasta de rábano picante llamada wasabe, la cual, puede compararse con el sabor que tienen la pasta de chile y algunas salsas utilizadas en la elaboración de comida china. Aquí encontramos la semejanza que existe entre el sabor de la comida japonesa y el de las comidas cantonesa y mandarina. La similitud en cuanto a la aportación de calorías empieza a perderse un poco, ya que la comida japonesa utiliza sobre todo pescado.

En el "quinto paso", los conglomerados que se unieron fueron C8, C9 y C10, esto es, española, italiana y griega. La distancia que existía entre ellos era de 2.1 unidades.
A estas alturas del análisis, a diferencia de lo que sucedía en los primeros pasos del mismo, los vectores de características ya no son tan similares entre sí. Veamos, en la primera calificación, X1, tenemos una mínima de 4.6 unidades y una máxima de 5.3 unidades; en cuanto a la segunda calificación, X2, tenemos una mínima de 4.7 unidades y una máxima de 6 unidades y, finalmente, en la tercera calificación, X3, tenemos una mínima 4.9 unidades y una máxima de 6.2 unidades. ¿Cómo es posible que estos tres elementos formen un sólo grupo?. En primer lugar, el criterio de unión,

a estas alturas, ya no es tan estricto y, por otra parte, como todas las calificaciones tienen la misma importancia las diferencias en una calificación se pueden compensar con las semejanzas de las otras.

En el "sexto paso", se unieron los conglomerados C3 y C5, es decir, szechuan y mexicana, y la distancia que había entre ellos era de aproximadamente 2.2 unidades. Aquí las semejanzas que podemos apuntar son: el uso de distintos tipos de carne (cerdo, pollo, res, etc.) en la elaboración de los platillos y, la utilización del picante en ambas cocinas.

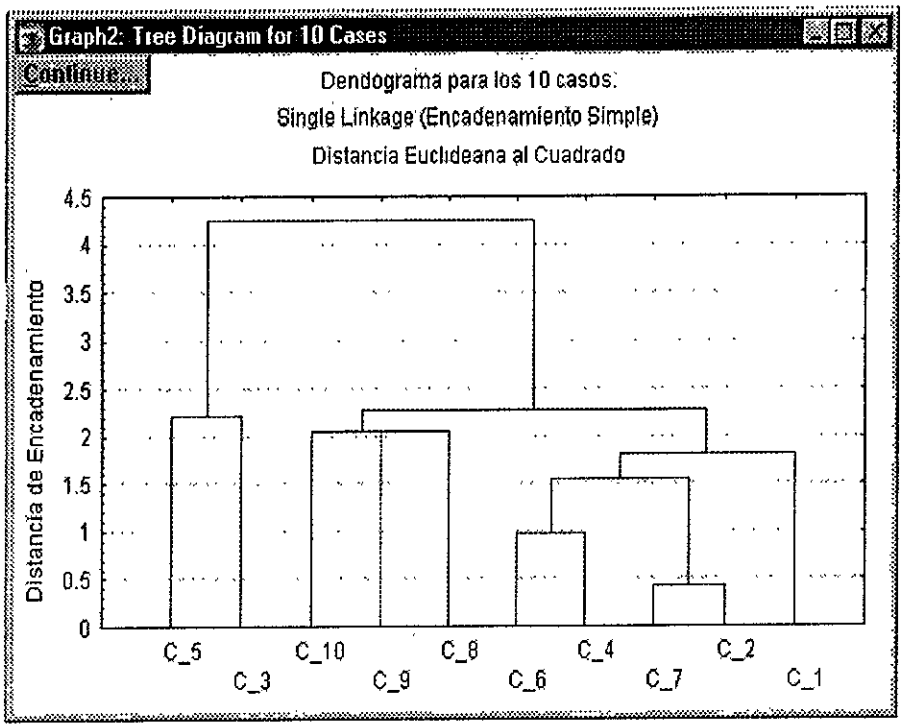
Si observamos el vector de características, podemos afirmar que la gente entrevistada concibe a ambas comidas como muy condimentadas y bastante similares en cuanto a la facilidad con que se digieren (pesada o ligera). En este renglón particular, se les considera en un término medio a ambas.

En el "séptimo paso", los conglomerados que se unieron fueron aquellos formados por los conglomerados C1, C2, C7, C4 y C6 ; y por otra parte , C8, C9 y C10, es decir, resulta un conglomerado formado por 8 elementos, a saber: japonesa, cantonesa, americana, francesa, mandarina, española, italiana y griega. Existía entre ellos una distancia, aproximadamente, de 2.3 unidades. Las similitudes que existen ya son menos aún.

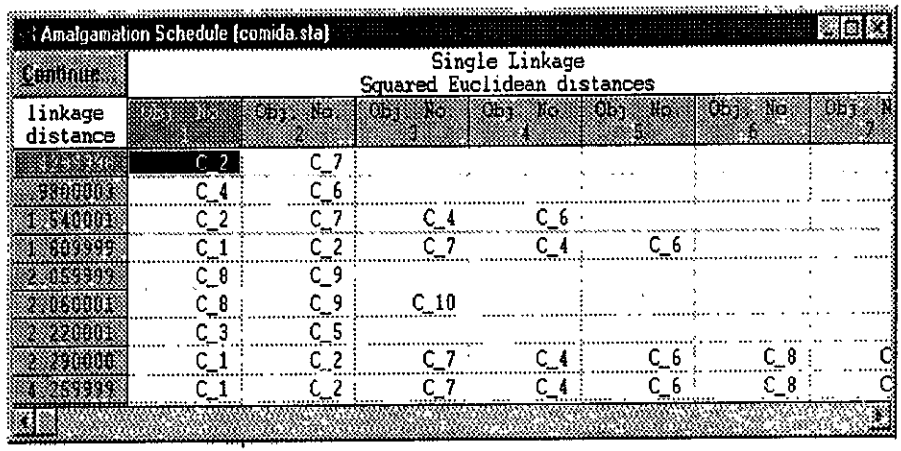
Finalmente, y como es de esperarse en un método aglomerativo, todos los conglomerados son unidos en uno, es decir, tenemos un conglomerado formado por los 10 elementos.

Al hacer click en el botón Continue ... regresamos a la caja de diálogo de Joining Results. Veamos ahora que pasa cuando hacemos click en el botón de Vertical icicle Plot.

Si observamos de manera cuidadosa esta gráfica descubriremos que, los cubos de hielo son una versión "vertical" de los dendogramas. Si lo analizamos, nos daremos cuenta de que esta gráfica nos proporciona la misma información que el dendograma y el único cambio estriba en la presentación de la misma. Así que podemos utilizar de manera indistinta cualquiera de ellos (cuestión de gustos).



Nuevamente, hacemos click en el botón Continue... para regresar a la caja de diálogo de Joining Results. Ahora veamos que resulta al hacer click sobre el botón Amalgamation Schedule (Ruta de Conglomeración).



Amalgamation Schedule (comida.sta)

Continue...

Single Linkage
Squared Euclidean distances

linkage distance	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6	Obj. No. 7
[illegible]	C_2	C_7					
.9800003	C_4	C_6					
1.540001	C_2	C_7	C_4	C_6			
1.669999	C_1	C_2	C_7	C_4	C_6		
2.059999	C_8	C_9					
2.060001	C_8	C_9	C_10				
2.220001	C_3	C_5					
2.290000	C_1	C_2	C_7	C_4	C_6	C_8	C
4.259999	C_1	C_2	C_7	C_4	C_6	C_8	C

Si analizamos la información que nos brinda este arreglo matricial nos daremos cuenta de que es, en esencia, la misma que nos dan las gráficas de los dendogramas y los cubos de hielo, con la diferencia de que en esta "matriz" conocemos de manera mucho más precisa las distancias a las cuales se formaron sucesivamente los conglomerados.

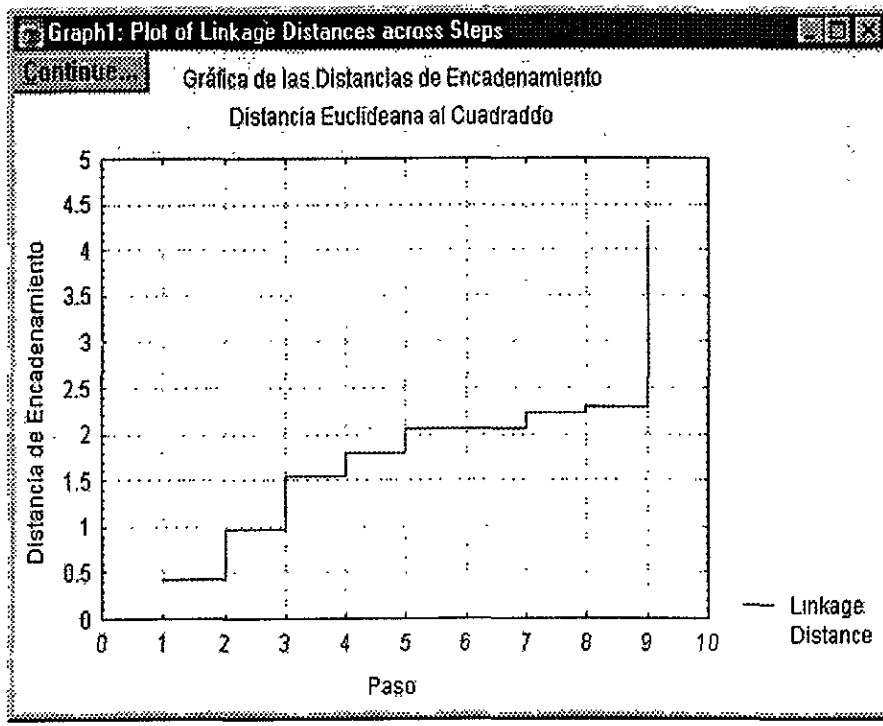
Por ejemplo, a una distancia de 0.4299998 se formó el primer conglomerado de dos elementos, se fusionaron C2 y C7 (justamente igual que en las gráficas de dendogramas y cubos de hielo).

Luego a una distancia de 0.9800003 se formó el segundo conglomerado de dos elementos, se fusionaron C4 y C6 (de la misma manera que en las gráficas). Y así podríamos seguir hasta llegar a tener todos los elementos incluídos en un solo conglomerado y, todas las etapas precedentes coincidirían.

Además, esta información puede resultar muy útil en el caso de que optemos por el uso de algún criterio que indique el número "natural" de conglomerados que deben formarse.

Hagamos click en Continue... y al aparecer la caja de Joining results, veamos que sucede al seleccionar la opción Graph of Amalgamation Schedule (Gráfica de la Ruta de Conglomeración).

Esta gráfica nos sirve para visualizar las diferencias existentes entre las distancias que habían de una "fusión" a otra, es decir, de la unión de conglomerados a la siguiente. Esta gráfica nos puede dar una idea de que "tan natural" resulta fusionar un cierto conglomerado con otro. En nuestro caso la última fusión, ya no resulta muy natural ¿No?. Esto nos podría sugerir la existencia de dos grupos o conglomerados "naturales".



Hagamos click en Continue... y veamos que se despliega en pantalla al seleccionar Distance Matrix.

En la matriz distancia se muestran las distancias, dos a dos, que existen entre los diez objetos (conglomerados) en el primer paso de nuestro análisis.
Observemos que, en efecto, la entrada más pequeña es la (7,2) es decir $d_{C2C7} = 0.4$, que fueron los primeros conglomerados en fusionarse.

Squared Euclidean distances (comida.sta)

Continue...	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10
C_1	0	8.5	14.8	5.1	15.0	1.8	12.3	10.7	18.9	15.3
C_2	8.5	.0	24.7	1.5	16.7	3.5	.4	4.7	5.1	8.0
C_3	14.8	24.7	.0	14.8	2.2	11.9	30.6	10.5	20.0	11.9
C_4	5.1	1.5	14.8	.0	9.1	1.0	3.5	2.3	4.7	4.1
C_5	15.0	16.7	2.2	9.1	.0	9.1	21.0	4.5	9.7	4.3
C_6	1.8	3.5	11.9	1.0	9.1	.0	6.4	3.9	9.0	7.2
C_7	12.3	.4	30.6	3.5	21.0	6.4	.0	6.6	5.6	10.3
C_8	10.7	4.7	10.5	2.3	4.5	3.9	6.6	.0	2.1	2.1
C_9	18.9	5.1	20.0	4.7	9.7	9.0	5.6	2.1	.0	2.2
C_10	15.3	8.0	11.9	4.1	4.3	7.2	10.3	2.1	2.2	.0

Nuevamente, hacemos click en Continue... y veamos que sucede al seleccionar la opción Descriptive statistics.

Esta opción, a mi juicio, no tiene mucho sentido en nuestro caso particular, ya que no nos proporciona ninguna información. Por ejemplo, 3.13 significa que se tiene "más o menos" sencillez, ligereza y bajas calorías.

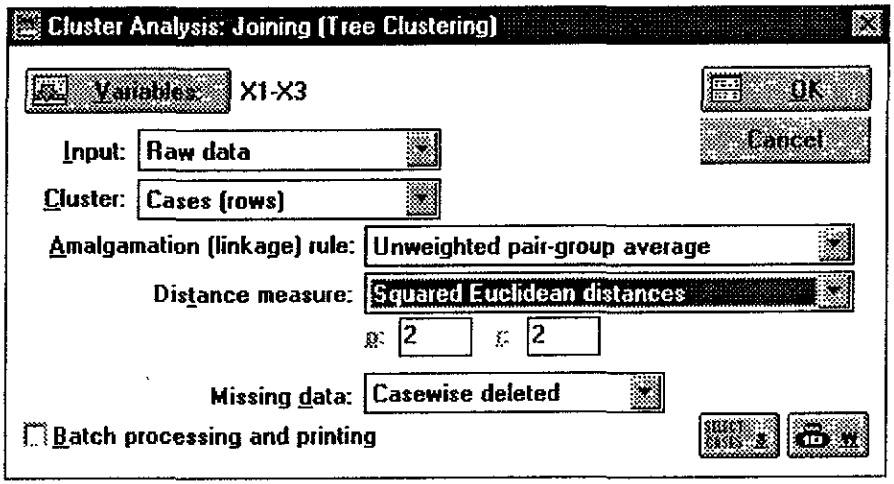
Means and Standard Deviations (comida.sta)

Continue...	Mean	St. dev.
C_1	3.133333	.305505
C_2	4.433333	1.588500
C_3	4.400000	1.928730
C_4	4.366667	.808290
C_5	5.000000	1.212435
C_6	3.900000	.435890
C_7	4.600000	1.992486
C_8	5.000000	.360555
C_9	5.600000	.871780
C_10	5.333333	.650641

Ahora veamos los resultados que obtenemos llevando a cabo un análisis de manera similar pero, usando un método de conglomerado distinto:

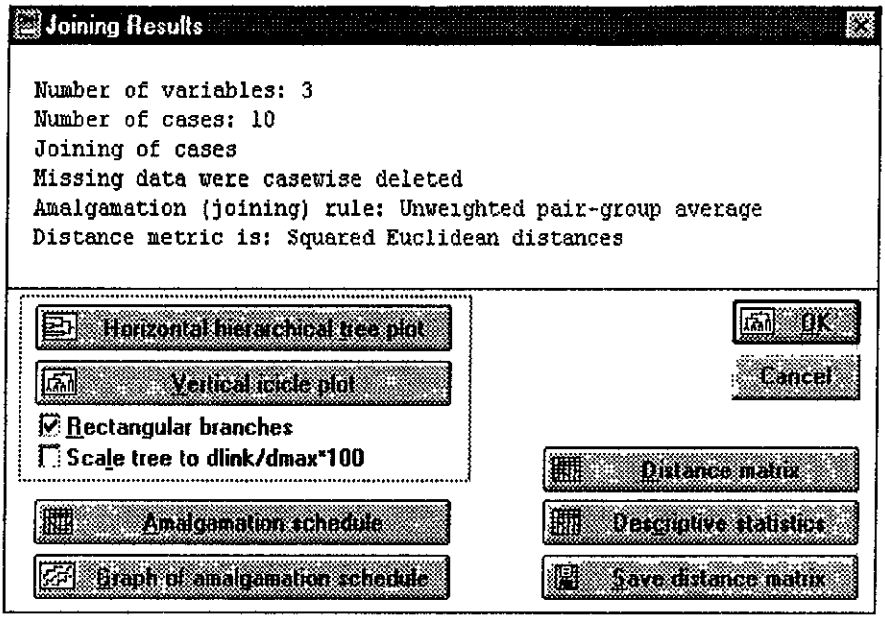
4.1.2 Promedios entre conglomerados.

Para hacer este análisis es necesario cambiar la opción de single linkage (vecino más cercano) por unweighted pair group average, como se muestra a continuación.

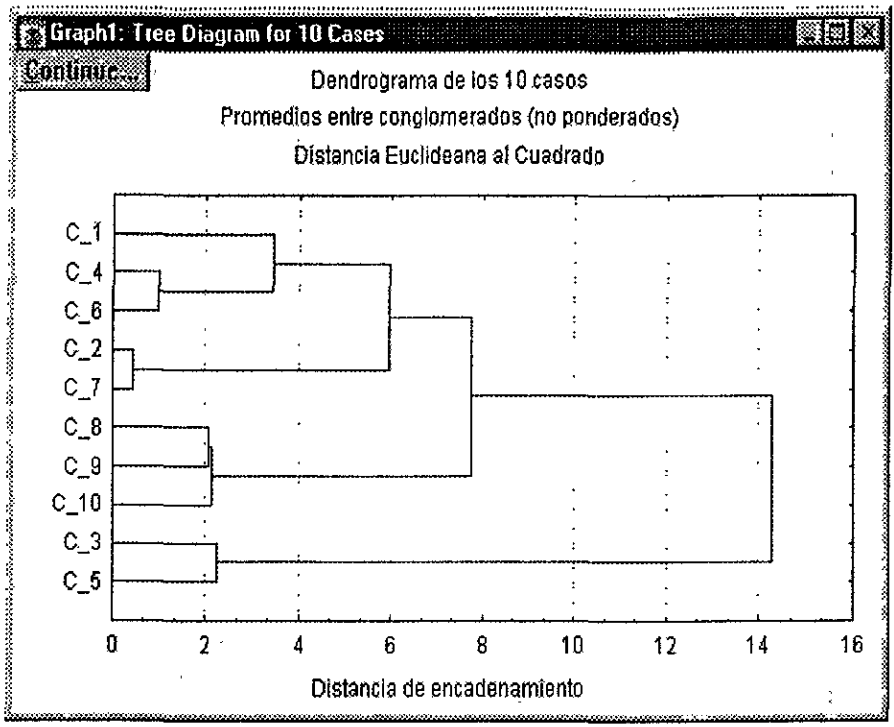


Hacemos click en el botón OK y aparecerá la ventana de Joining results correspondiente a un análisis hecho con estas características.

Analizaremos cada una de las opciones de esta ventana.



Empezaremos por ver que resultados obtuvimos en el dendograma (Horizontal hierarchical tree plot).



En primer lugar, observemos que las distancias de encadenamiento han aumentado en relación a las que obtuvimos utilizando el método del vecino más cercano.
Ahora bien, compararemos los resultados obtenidos con este método y con el anterior.

En el primer paso, utilizando este método, se fusionan C2 y C7. Lo mismo sucedió con el método del vecino más cercano.

En el segundo paso, utilizando este método, se fusionan C4 y C6. Lo mismo sucedió con el método del vecino más cercano. Como veremos a continuación, las diferencias se darán a partir del tercer paso.

En el tercer paso, usando el método de promedios, se fusionan C8 y C9. Por otra parte, usando el vecino más cercano, se fusionaban C2, C7, C4 y C6.

En el cuarto paso, usando promedios, se fusionan C8, C9 y C10. Utilizando el vecino más cercano se fusionan C1, C2, C7, C4 y C6.

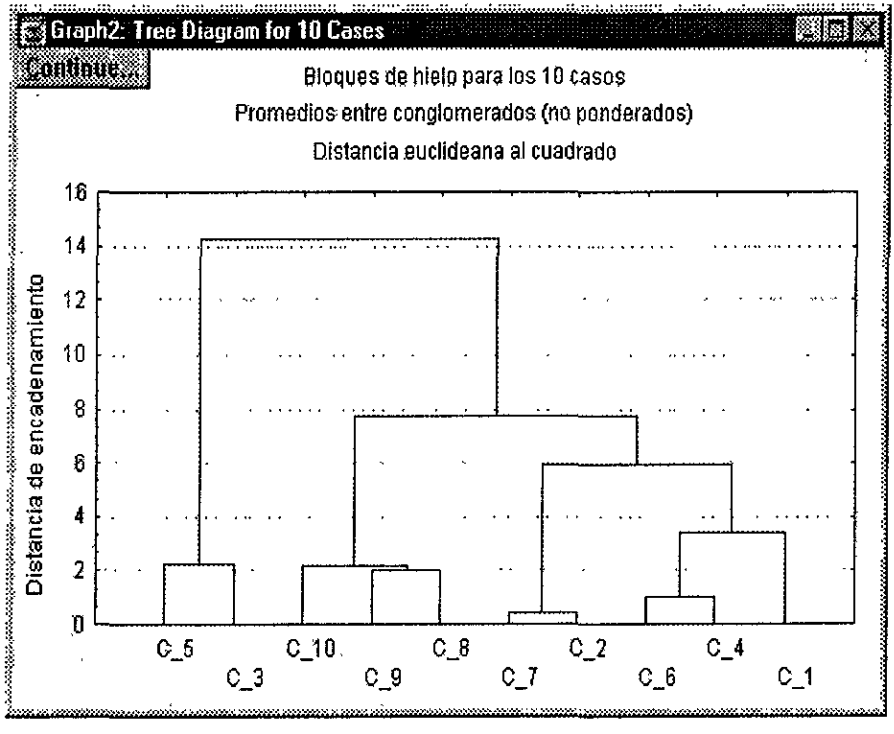
En el quinto paso, usando promedios, se fusionan C3 y C5. Utilizando el vecino más cercano C8, C9 y C10.

En el sexto paso, usando promedios, se fusionan C1, C4,C6, C2 y C7. Utilizando el vecino más cercano C3 y C5.

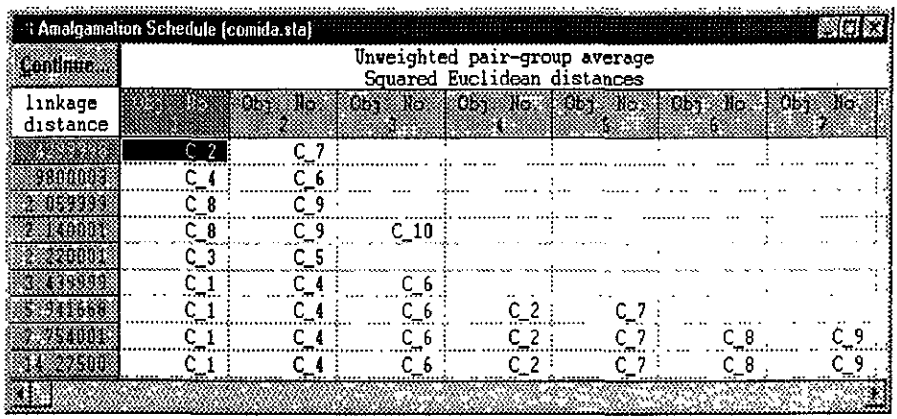
En el séptimo paso, usando promedios, se fusionan C1, C4, C6, C2, C7, C8, C9 y C10. Utilizando el vecino más cercano C1, C2, C7, C4, C6, C8, C9 y C10. En este paso se vuelven a encontrar ambos métodos.

Finalmente, en ambos casos, todos los elementos forman un solo conglomerado.

Ahora veamos que tenemos en *los bloques de hielo (Vertical icicle plot)*. Nuevamente, hacemos la misma observación respecto a esta opción que la hecha en el método anterior: la información que los dendogramas y los cubos de hielo nos dan es la misma, su única diferencia estriba en el formato que tienen (horizontal y vertical respectivamente).

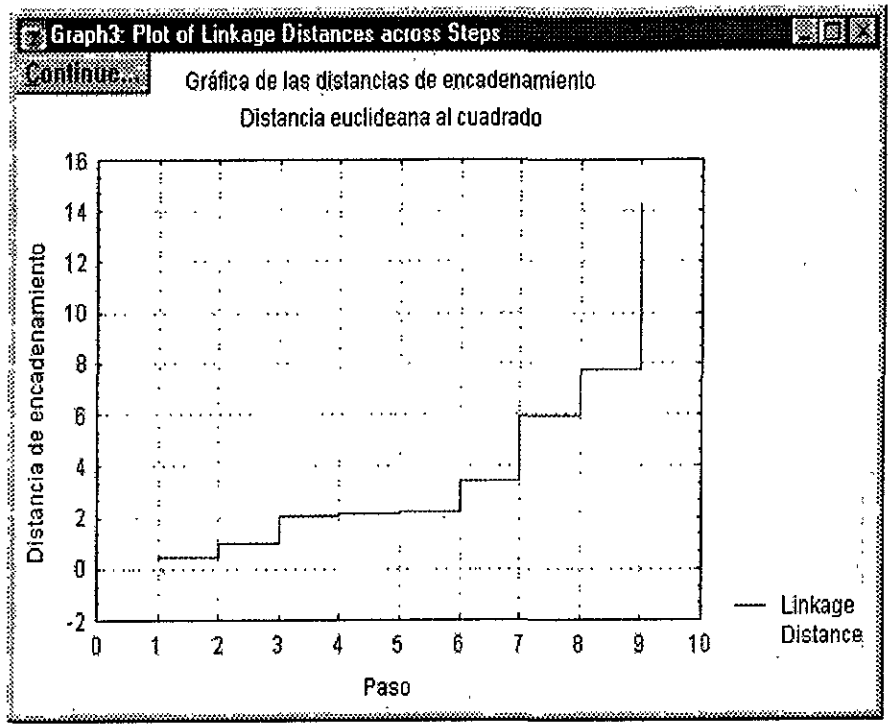


Aquí, en la opción de amalgamation schedule (ruta de conglomeración), tenemos a un mayor nivel de precisión las distancias de encadenamiento que se encuentran tanto en el dendograma como en la gráfica de cubos de hielo.
También podemos constatar que las diferencias se dan a partir del tercer paso. Cabe señalar la enorme diferencia entre las distancias finales 14.27500 utilizando este método y, 4.259999 utilizando el anterior.



Hacemos click en el botón de Continue...y luego en el botón de graph amalgamation schedule.

Como ya habíamos mencionado antes, esta gráfica puede sugerirnos separaciones naturales. En nuestro caso particular, en el noveno paso podemos observar una "separación natural" o dicho en otros términos observamos una discontinuidad muy grande en la altura de los escalones (¿no es cierto?).



Ahora, volvemos a hacer click en Continue... y veamos que sucede con la opción Distance Matrix.

Squared Euclidean distances (comida.sta)

Continue...

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10
C_1	0	8.5	14.8	5.1	15.0	1.8	12.3	10.7	18.9	15.3
C_2	8.5	.0	24.7	1.5	16.7	3.5	.4	4.7	5.1	8.0
C_3	14.8	24.7	.0	14.8	2.2	11.9	30.6	10.5	20.0	11.9
C_4	5.1	1.5	14.8	.0	9.1	1.0	3.5	2.3	4.7	4.1
C_5	15.0	16.7	2.2	9.1	.0	9.1	21.0	4.5	9.7	4.3
C_6	1.8	3.5	11.9	1.0	9.1	0	6.4	3.9	9.0	7.2
C_7	12.3	.4	30.6	3.5	21.0	6.4	.0	6.6	5.6	10.3
C_8	10.7	4.7	10.5	2.3	4.5	3.9	6.6	.0	2.1	2.1
C_9	18.9	5.1	20.0	4.7	9.7	9.0	5.6	2.1	.0	2.2
C_10	15.3	8.0	11.9	4.1	4.3	7.2	10.3	2.1	2.2	.0

Si observamos cuidadosamente, nos daremos cuenta de que esta matriz distancia es idéntica a la anterior. ¿Por qué?. Pues bien, la matriz distancia en el primer paso NO se ve afectada por el método de conglomerado que hayamos elegido. Recordemos como funciona el proceso. Primero, elegimos la función distancia (en nuestro caso fue la distancia euclideana al cuadrado). Una vez hecha esta elección calculamos la distancia que existe entre cada par de elementos involucrados en el análisis.

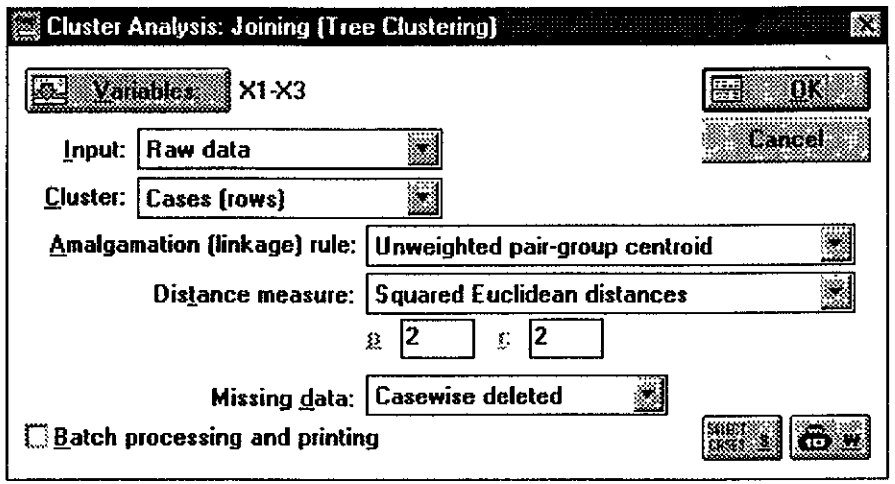
Bien, hasta este momento, no hemos involucrado para nada al método de conglomerado. Entonces, ¿Cuándo se ve involucrado el método de conglomerado?, Después de que hemos calculado la matriz distancia, ya que una vez que hemos calculado ésta, hacemos uso de la información que tiene para que, en base al método de conglomerado vayamos uniendo a los elementos.

En pocas palabras, el método de conglomerado es un criterio que utiliza la matriz distancia para establecer la similitud (parecido) o disimilitud (diferencia) que existe entre dos elementos y concluir si pertenecen a un mismo grupo o no.

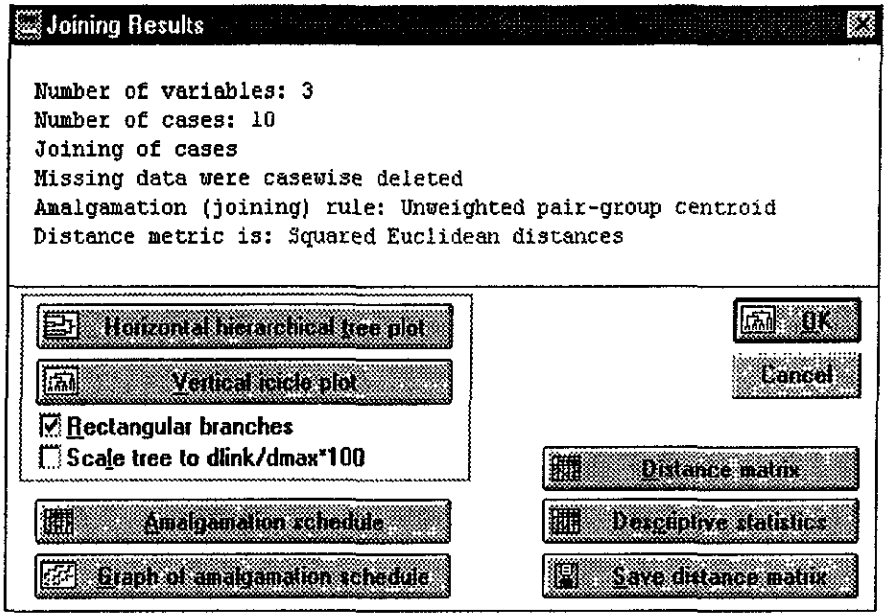
Estadísticas descriptivas. Ya no las presentamos por las razones anteriormente expuestas. Además de que son las mismas que las presentadas con el método anterior.

4.1.3 Método del centroide.

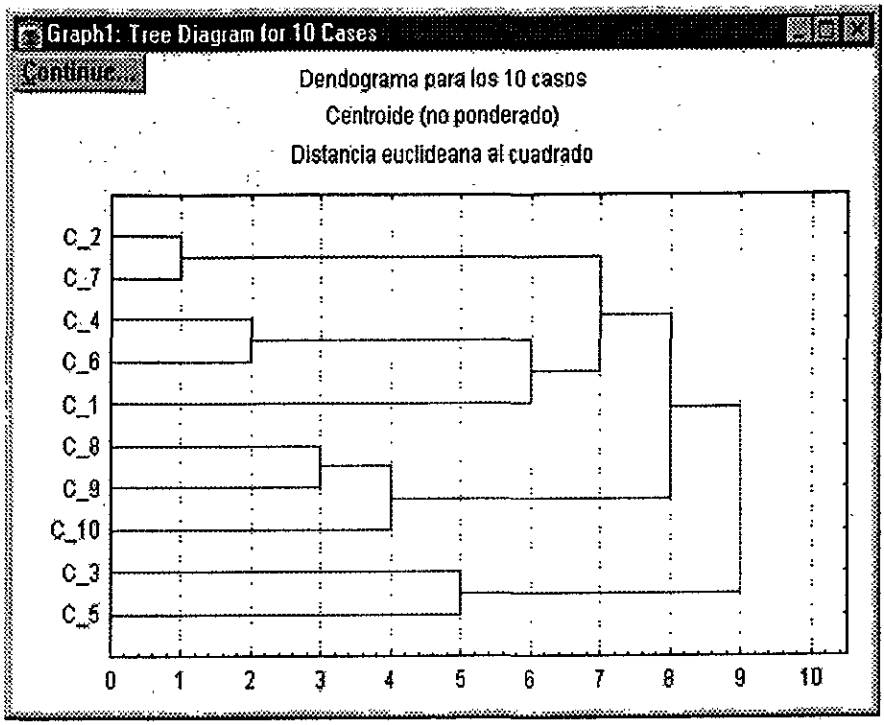
Ahora veamos los resultados que obtenemos haciendo el mismo análisis pero utilizando el método del centroide (unweighted pair-group centroid).



Como ya nos podemos imaginar, hacemos click en el botón OK y aparece la pantalla de Joining Results.



Aquí la distancia de encadenamiento también es distinta a las dos anteriores.



Otra vez, en el primer paso, como en los dos métodos anteriores, los conglomerados que se fusionan son C2 y C7.

En el segundo paso, coinciden también los tres métodos, los conglomerados que se fusionan son C4 y C6.

En el tercer paso, coinciden este método, centroide, y el método de promedios, en ambos se fusionan los conglomerados C8 y C9.

En el cuarto paso, coinciden este método y el de promedios, en ambos se fusionan los conglomerados C8, C9 y C10.

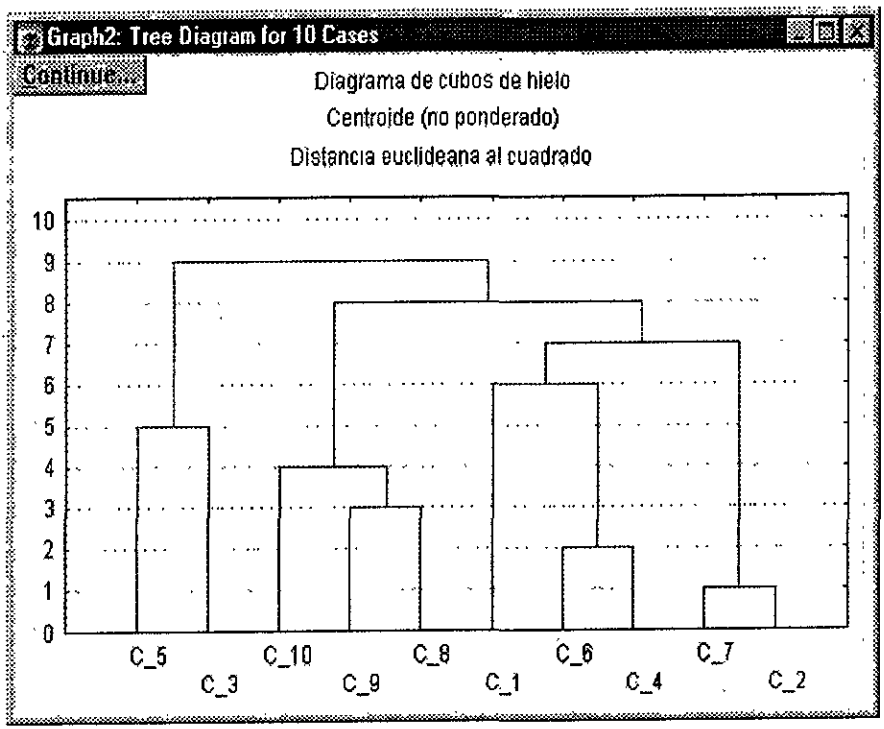
En el quinto paso, coinciden, nuevamente, este método y el de promedios, en los dos métodos se unen los conglomerados C3 y C5.

En el sexto paso, coinciden, una vez más, estos dos métodos (centroide y promedios) formando un conglomerado con C1,C4 y C6.

En el séptimo paso, coinciden los tres métodos, en los tres se forma un conglomerado con los siguientes elementos:C1, C2, C4, C6 y C7.
En el octavo paso, coinciden también los tres métodos, se forma un conglomerado con los siguientes elementos: C1, C2, C4, C6, C7, C8, C9 y C10.

Y finalmente, también coinciden, al formar un solo conglomerado con todos los elementos incluídos en él.

Ya no queda mucho que decir respecto a esta opción excepto que, en lo personal, me parece más clara la información presentada en este formato que en el dendograma convencional.



Nuevamente, hacemos click en Continue... para seguir explorando las opciones restantes. Veamos que podemos observar en la opción Amalgamation Schedule (ruta de conglomeración).

Amalgamation Schedule (comida.sta)

Continue...

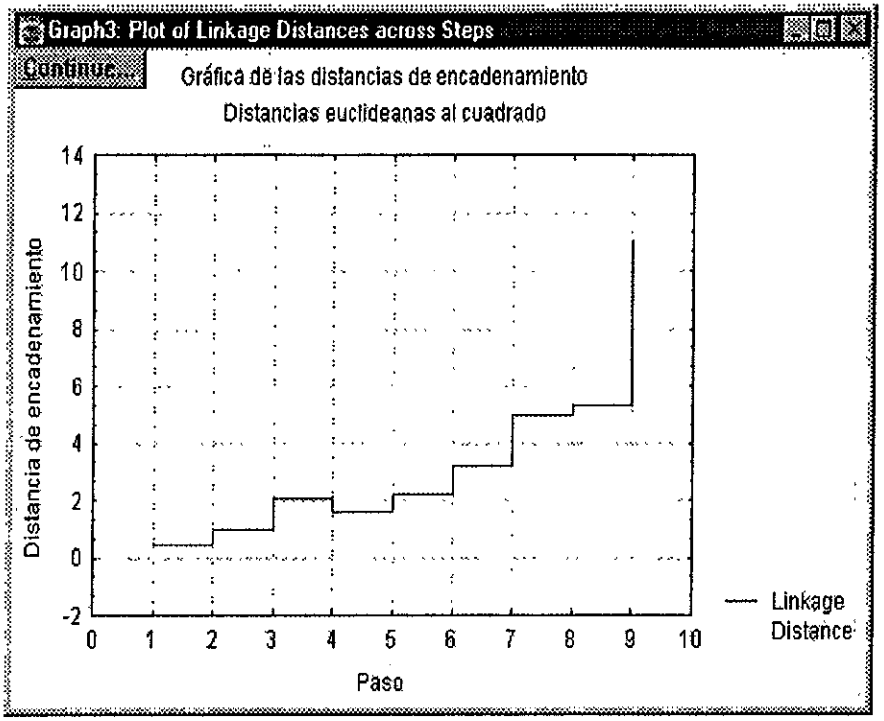
Unweighted pair-group centroid
Squared Euclidean distances

linkage distance	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6	Obj. No. 7
1.625001	C_2	C_7					
1.900003	C_4	C_6					
2.059999	C_8	C_9					
1.625001	C_8	C_9	C_10				
2.200001	C_3	C_5					
3.194999	C_1	C_4	C_6				
4.960835	C_1	C_4	C_6	C_2	C_7		
5.241916	C_1	C_4	C_6	C_2	C_7	C_8	C_9
11.11783	C_1	C_4	C_6	C_2	C_7	C_8	C_9

Si observamos detenidamente la columna de linkage distance de la opción Amalgamation Schedule (ruta de conglomeración) encontraremos algo nuevo. En los métodos anteriores los valores de la columna linkage distance iban siempre aumentando, y en este caso observamos un comportamiento distinto. Revisemos el valor 3 de la tabla, es igual a 2.059999 y el valor 4 es igual a 1.625001, ¿Qué fue lo que sucedió?. La explicación la daremos en el siguiente párrafo donde veremos como se traduce este comportamiento en la Gráfica de distancias de encadenamiento (Graph of Linkage Distance).

Hagamos click en Continue... y después seleccionemos la opción Graph of amalgamation schedule.

Observemos la gráfica.



Algo interesante que destacar en esta gráfica es lo que se observa entre el paso 4 y 5. La distancia de encadenamiento del paso 3 al 4 es mayor que la distancia existente del paso 4 al 5. ¿Algo está mal? La respuesta es que no. A este tipo de fenómeno se le llama reversal. Un reversal nos da la siguiente información: al unirse nuevos elementos a los conglomerados los centroides se ven afectados de tal manera que se encuentran más cerca entre sí que en el paso anterior.

Hagamos click en Continue... y luego seleccionemos la opción Distance Matriz. Esto nos servirá para cerciorarnos de que, en efecto, la matriz distancia es la misma sin importar el método que utilicemos.

Squared Euclidean distances (comida.sta)

Continue...	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10
C_1	.0	8.5	14.8	5.1	15.0	1.8	12.3	10.7	18.9	15.3
C_2	8.5	.0	24.7	1.5	16.7	3.5	.4	4.7	5.1	8.0
C_3	14.8	24.7	.0	14.8	2.2	11.9	30.6	10.5	20.0	11.9
C_4	5.1	1.5	14.8	.0	9.1	1.0	3.5	2.3	4.7	4.1
C_5	15.0	16.7	2.2	9.1	.0	9.1	21.0	4.5	9.7	4.3
C_6	1.8	3.5	11.9	1.0	9.1	.0	6.4	3.9	9.0	7.2
C_7	12.3	.4	30.6	3.5	21.0	6.4	0	6.6	5.6	10.3
C_8	10.7	4.7	10.5	2.3	4.5	3.9	6.6	.0	2.1	2.1
C_9	18.9	5.1	20.0	4.7	9.7	9.0	5.6	2.1	0	2.2
C_10	15.3	8.0	11.9	4.1	4.3	7.2	10.3	2.1	2.2	.0

CAPÍTULO 5.

UTILIZACIÓN DEL ANÁLISIS DE CONGLOMERADOS EN LA DEMOGRAFÍA.

Introducción.

Recordemos que en el capítulo 4 estudiamos la forma de como funciona el Análisis de Conglomerados. Esos ejemplos nos permitieron ver la manera en que se van formando los grupos y la evolución de los algoritmos. Teniendo estas ideas en mente, vamos a abordar un problema de mayor interés.

En este capítulo seremos más minuciosos, y analizaremos con mayor detalle la forma en que llevaremos a cabo nuestro análisis. De hecho, el propósito principal de este capítulo, será el establecimiento de una serie de pasos que debemos seguir para llevar a cabo un buen análisis de conglomerados.

En este capítulo intentaremos hacer uso de la técnica del análisis de conglomerados con un conjunto de datos más complejo que el que utilizamos en el capítulo pasado, tanto en magnitud como en interpretación.

5.1 Clasificación de 100 Ciudades de acuerdo a ciertas características demográficas.

Este conjunto de datos corresponde a un folleto publicado por el INEGI titulado: "Programa de 100 Ciudades Indicadores Sociodemográficos y Económicos" De el tomamos algunos, no todos los datos que se dan a conocer ahí.

Imaginemos que tenemos la siguiente situación: Un demógrafo nos da el conjunto de datos que aparece a continuación :

Variables

ABREVIATURA	EQUIVALENCIA
POBTOTAL (V1)	POBLACION TOTAL
%_15ALFA (V2)	%DE LA POBLACION DE 15 AÑOS Y MAS ALFABETA
%POBSSEC (V3)	% DE LA POBLACION EN EL SECTOR SECUNDARIO
%POBSTER (V4)	% DE LA POBLACION EN EL SECTOR TERCIARIO
TOTVPHAB (V5)	TOTAL DE VIVIENDAS PARTICULARES HABITADAS
PROMOCVP (V6)	PROMEDIO DE OCUPANTES EN VIVIENDAS PARTICULARES
%MANUFAC (V7)	% DEL PERSONAL OCUPADO PROMEDIO EN EL SECTOR MANUFACTURERO
%EXTRACT (V8)	% DEL PERSONAL OCUPADO PROMEDIO EN LA INDUSTRIA EXTRACTIVA
%CONSTRU (V9)	% DEL PERSONAL OCUPADO PROMEDIO EN LA INDUSTRIA DE LA CONSTRUCCION
%ELECTRI (V10)	% DEL PERSONAL OCUPADO PROMEDIO EN EL SECTOR ELECTRICO
%PESCA (V11)	% DEL PERSONAL OCUPADO PROMEDIO EN EL SECTOR PESQUERO
%COMERCIO (V12)	% DEL PERSONAL OCUPADO PROMEDIO EN EL SECTOR COMERCIO
%SERVICIO (V13)	% DEL PERSONAL OCUPADO PROMEDIO EN EL SECTOR SERVICIOS
%TRANCOM (V14)	% DEL PERSONAL OCUPADO PROMEDIO EN EL SECTOR TRANSPORTE Y COMUNICACIONES

De la siguiente lista de ciudades:

CIUDAD	NOMBRE
1	AGUASCALIENTES (AGS)
2	ENSENADA (ENSE)
3	MEXICALI (MXL)
4	TECATE (TECA)
5	TIJUANA (TIJU)
6	CD. CONSTITUCION (CONS)
7	LA PAZ (PAZ)
8	SAN JOSE DEL CABO (CABO)
9	CAMPECHE (CAMP)
10	CD. DEL CARMEN (CARM)
11	CD. ACUÑA (ACUÑ)

12 MONCLOVA (MOCL)
13 PIEDRAS NEGRAS (PNEG)
14 SALTILLO (SALT)
15 TORREON (TORR)
16 COLIMA (COL)
17 MANZANILLO (MANZ)
18 SAN CRISTOBAL DE LAS CASAS (SCRIS)
19 TAPACHULA (TAPA)
20 TUXTLA GUTIERREZ (TUXT)
21 CD. JUAREZ (JUAR)
22 CUAUHTEMOC (CUAUH)
23 CHIHUAHUA (CHIH)
24 DELICIAS (DELI)
25 HIDALGO DEL PARRAL (HGOP)
26 DURANGO (DGO)
27 CELAYA (CELA)
28 GUANAJUATO (GTO)
29 IRAPUATO (IRAP)
30 LEON (LEON)
31 MOROLEON (MLEO)
32 SALAMANCA (SALM)
33 SAN MIGUEL DE ALLENDE (SMIG)
34 ACAPULCO (ACAP)
35 CHILPANCINGO (CHILP)
36 IGUALA (IGUA)
37 IZTAPA ZIHUATANEJO (ZIHUA)
38 PACHUCA (PACH)
39 TULA (TULA)
40 TULANCINGO (TLAC)
41 CD. GUZMAN (GUZM)
42 LAGOS DE MORENO (LAGO)
43 PUERTO VALLARTA (PVALL)
44 TOLUCA (TOLU)
45 VALLE DE BRAVO (VBRA)
46 APATZINGAN (APAT)
47 LAZARO CARDENAS (LCAR)
48 MORELIA (MORE)
49 PATZCUARO (PATZ)
50 URUAPAN (URUA)
51 ZAMORA (ZAMO)
52 CUAUTLA (CTLA)
53 CUERNAVACA (CVCA)
54 TEPIC (TEPI)

55 LINARES (LINA)
56 BAHIAS DE HUATULCO (BHUA)
57 OAXACA (OAXA)
58 SALINA CRUZ (CRUZ)
59 TUXTEPEC (TUXT)
60 TEHUACAN (THUA)
61 QUERETARO (QRO)
62 SAN JUAN DEL RIO (SJR)
63 CANCUN (CANC)
64 COZUMEL (COZU)
65 CHETUMAL (CHET)
66 CD. VALLES (CVALL)
67 SAN LUIS POTOSI (SLP)
68 CULIACAN (CULC)
69 GUASAVE (GUAS)
70 LOS MOCHIS (MOCH)
71 MAZATLAN (MAZA)
72 AGUA PRIETA (APRT)
73 CD. OBREGON (OBRE)
74 GUAYMAS (GUAY)
75 HERMOSILLO (HERM)
76 NAVOJOA (NAVO)
77 NOGALES (NOGA)
78 SAN LUIS RIO COLORADO (SLRC)
79 CARDENAS (CARD)
80 FRONTERA (FRON)
81 VILLA HERMOSA (VHER)
82 CD. MANTE (MANT)
83 CD. VICTORIA (VICT)
84 MATAMOROS (MATA)
85 NUEVO LAREDO (NLAR)
86 REYNOSA (REYN)
87 TAMPICO (TAMP)
88 APIZACO (APIZ)
89 TLAXCALA (TLAX)
90 COATZACOALCOS (COAT)
91 CORDOBA (CORD)
92 MARTINEZ DE LA TORRE (MTOR)
93 POZA RICA (POZA)
94 TUXPAN (TUXP)
95 VERACRUZ (VERA)
96 JALAPA (JALA)
97 MERIDA (MERI)

98 VALLADOLID (VALL)
99 FRESNILLO (FRES)
100 ZACATECAS (ZACT)

El listado de estos datos lo podemos consultar en el apéndice. Ver Tabla 2.
Lo que le gustaría que le dijeramos es cuáles de las ciudades se parecen entre sí dadas las características observadas en el folleto.

Quizás podríamos preguntarnos de que manera le sería útil esta información. Pues bien, la respuesta sería más o menos la siguiente: Es importante tanto para la política poblacional como para la política de desarrollo que se tenga en el país.

Por ejemplo, existen algunos estudios que asocian niveles de ingreso y de educación y grado de industralización con los niveles de fecundidad. Existen estudios que nos indican desde hace algún tiempo acerca de que mayores niveles de analfabetismo están íntimamente relacionados con mayores niveles de fecundidad.

Los mayores niveles de fecundidad corresponden a las mujeres que tienen menor nivel educativo. Si se piensa que los matrimonios tienden a producirse, en la generalidad de los casos, entre personas de una misma clase social, cabría esperar también una asociación entre la educación del marido y la fecundidad de las mujeres. En otras palabras, a menor nivel educativo del marido mayor es la fecundidad de las mujeres.

Ahora consideremos otro aspecto de las poblaciones: la salud. La distribución geográfica de la población implica la existencia de distintos grupos de población expuestos a diversos riesgos ambientales y, por lo tanto, la presencia de diferenciales en niveles y estructura de salud.

Como en el caso del tamaño de la población y sobre todo en sus aspectos dinámicos (migraciones) que implican la modificación de la densidad, la concentración, el hacinamiento, etc., la distribución geográfica actúa indirectamente a través de la modificación del ambiente físico, biológico y social, aumentando, disminuyendo o generando nuevos riesgos sin constituir una causa directa de los diversos estados de salud.

Además las migraciones significan el movimiento de gente que tiene ciertas características de salud ya adquiridas y que al trasladarse a otras áreas pueden introducir en ellas cambios en la estructura de la morbilidad.

En cuanto a la composición de la población, puede decirse que, en general, la pertenencia de un individuo a un grupo dado de población le determina un cierto grado de susceptibilidad y de exposición a riesgos para la salud, que son inherentes a dicho grupo.

La composición económica, laboral, educacional, social y cultural de la población cuya información se obtiene en los censos, por lo cual son designadas como características demográficas, se presentan en grupos de población asociados a distintas formas de vida y de comportamiento social. Por lo tanto, pueden ser considerados como grupos que tienen distintos grados de exposición a los diversos riesgos de enfermar y de morir. Así, puede concebirse la existencia de diferencias importantes en el nivel y estructura de la salud entre una población cuya actividad más importante es la minería, o bien entre profesionales y obreros, etc.

Finalmente, consideremos otro factor de vital importancia para la dinámica poblacional: la educación. Un aumento en el nivel de instrucción puede traducirse en una disminución del nivel de mortalidad, pues en general las personas más instruidas tienen una mejor comprensión de las normas higiénicas y de la utlilidad de su aplicación en la prevención de las enfermedades, principalmente las de la primera infancia.

La natalidad tal vez tienda a disminuir pues con la instrucción pueden transmitirse pautas culturales que estén en conflicto con el ideal de un número elevado de hijos.
La intensidad de las corrientes migratorias también podría sufrir la influencia de la expansión de la educación. Se ha observado que muchas veces los migrantes de las zonas rurales a las urbanas tienen un nivel de instrucción más alto que los no-migrantes rurales, lo que indicaría una mayor propensión a migrar entre los más instruidos.

A su vez, los cambios en la mortalidad, la natalidad y la migración alteran el ritmo de crecimiento de la población, su distribución territorial y la composición de la misma. Estos cambios pueden afectar el proceso de planificación de la educación.
Un cambio en el ritmo de crecimiento de la población puede significar una variación en la proporción de recursos asignados al sector educacional dentro de un plan general de desarrollo. Los cambios de la distribución territorial de la población afectan la localización de las escuelas y la distribución del personal docente.

Todo esto que hemos dicho sólo pretende esbozar la importancia que puede tener el encontrar ciertos grupos de ciudades dadas características demográficas y, no sólo tomando en cuenta su proximidad geográfica.

¿No sería mejor tener un plan a nivel nacional y olvidarnos de hacer estos grupos? La respuesta es: NO. De qué serviría dedicarle a cierta región una inversión dirigida a planificación familiar si la fecundidad de dicha zona es ya de por si baja. Un plan regional nos ayudaría a hacer un mejor uso de los recursos que se tengan disponibles para cada región y haciendo uso de ellos en los puntos más relevantes en el desarrollo de dicha región. En una palabra optimizar el uso de estos recursos.

5.2 Etapas del Análisis.

Pues bien, comencemos por establecer una serie de pasos para llevar a cabo el análisis.
ETAPA 1. OBJETIVOS DEL ANÁLISIS.
ETAPA 2. DISEÑO DEL ANÁLISIS.
ETAPA 3. SUPUESTOS DEL ANÁLISIS.
ETAPA 4. ELECCIÓN DEL ALGORITMO.
ETAPA 5. INTERPRETACIÓN DE LOS CONGLOMERADOS.
ETAPA 6. VALIDACIÓN DE LOS CONGLOMERADOS.

Sobre la marcha nos iremos dando cuenta en que consiste cada una de las etapas. Comencemos.

5.2.1 Objetivos del análisis.

Necesitamos formar un número manejable de grupos de ciudades similares entre sí. Este número nos será sugerido al aplicar un criterio específico (a este número habremos de llamarle el "número natural de grupos" existente en el conjunto de datos) previamente expuesto en el capítulo 3. Nos referimos al criterio de Mojena, si el número obtenido no nos es de utilidad propondremos un número en base a otro tipo de consideraciones, por ejemplo, algún número en particular que le fuera útil al demógrafo, algún estudio similar hecho previamente, etc.

Una vez obtenidos los grupos obtendremos un representante de cada uno de ellos para etiquetarlos y compararlos entre sí.

5.2.2 Diseño del análisis de conglomerados.

Dentro de esta etapa existen 3 pasos que son los siguientes:
a) Detección de outliers. Podemos llevar a cabo el análisis completo y dejar que este nos sugiera la existencia de outliers. Esta es la forma en que proponemos la detección de dichos puntos.
Otra opción es llevar a cabo primero un análisis que arroje información respecto a la existencia de los outliers. Una vez detectados éstos debemos decidir si se eliminan o no estos elementos.
b) Elección de la medida de similitud. Nosotros hemos elegido la distancia euclideana al cuadrado debido a que nuestras variables son métricas y a que su cálculo es más sencillo.
c) Estandarización. Debido a que nuestras variables miden características tan distintas en magnitud como: población y porcentaje de la población dedicado a la industria (una en miles y la otra menor a la unidad, respectivamente) hemos decidido estandarizar.

De lo contrario, como ya habíamos mencionado en el capítulo 3, dominaría la influencia de las variables de mayor magnitud sobre todas las demás. Los datos estandarizados aparecen en la tabla 3 del apéndice.

Como un primer acercamiento, vamos a utilizar las 14 variables que tenemos.

5.2.3 Supuestos del análisis.

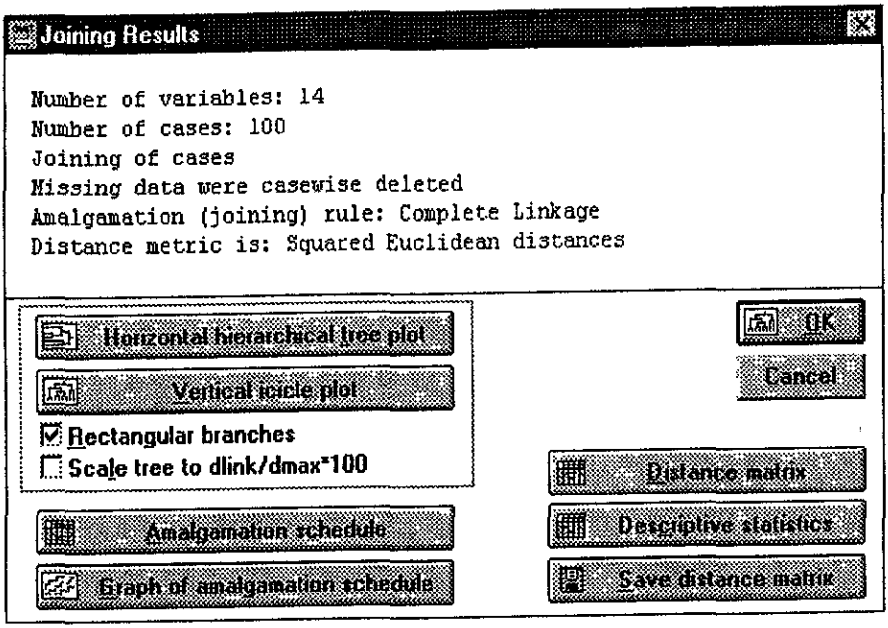
Representatividad de la muestra. En nuestro caso no aplica pues no utilizamos una muestra de la población que nos interesa analizar sino que todos los elementos de interés están en la muestra.

5.2.4 Elección del algoritmo.

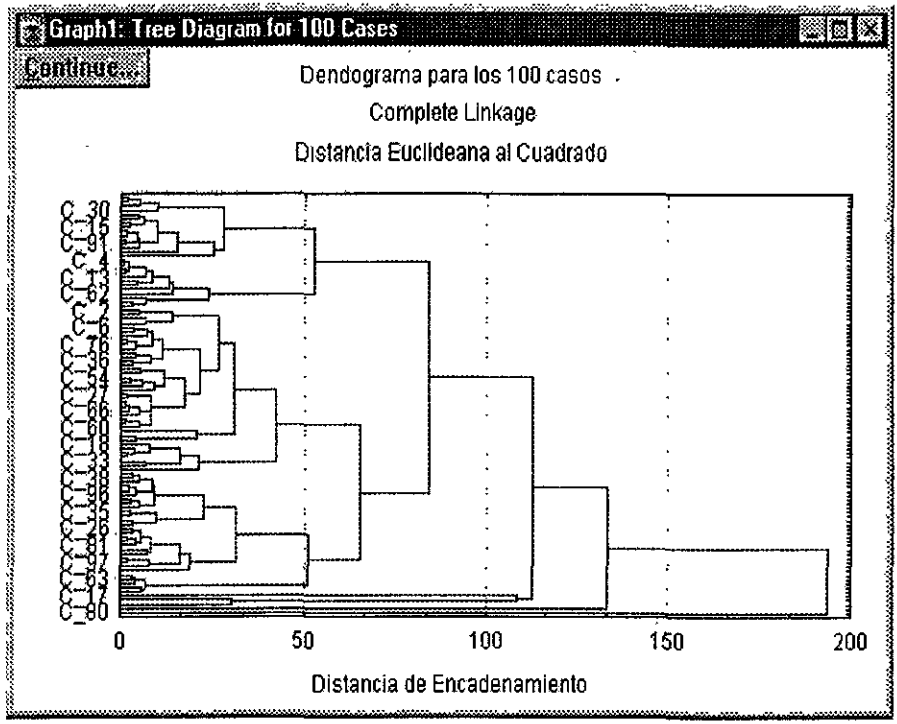
En este paso definimos que algoritmo habremos de usar y justificamos su uso. Nosotros utilizaremos complete linkage (vecino más lejano).
Las razones de dicha elección son las siguientes: Utilizando la lógica y viendo rápidamente los datos que tenemos del Programa de 100 Ciudades, podemos darnos cuenta de que los elementos que conforman este grupo son parecidos entre sí. Si utilizaramos el método de Single Linkage (vecino más cercano) podría (de hecho sucede) dar como resultado una clasificación poco útil para nosotros. ¿A qué nos referimos con poco útil? A que el método, por su misma naturaleza, nos diera como resultado de su aplicación un solo grupo (recordemos el problema de serpenteo que se presenta utilizando este método).

¿Qué pasaría si utilizaramos el método de Ward? Probablemente, como nos dice la teoría, obtendríamos como resultado de su aplicación a este caso, muchos grupos pequeños, es decir, al contrario de lo que pasaría con el método de Single Linkage (vecino más cercano), tendríamos demasiados grupos. Como consecuencia nuestros esfuerzos habrán sido prácticamente inútiles pues, nosotros buscamos hacer la información más manejable.

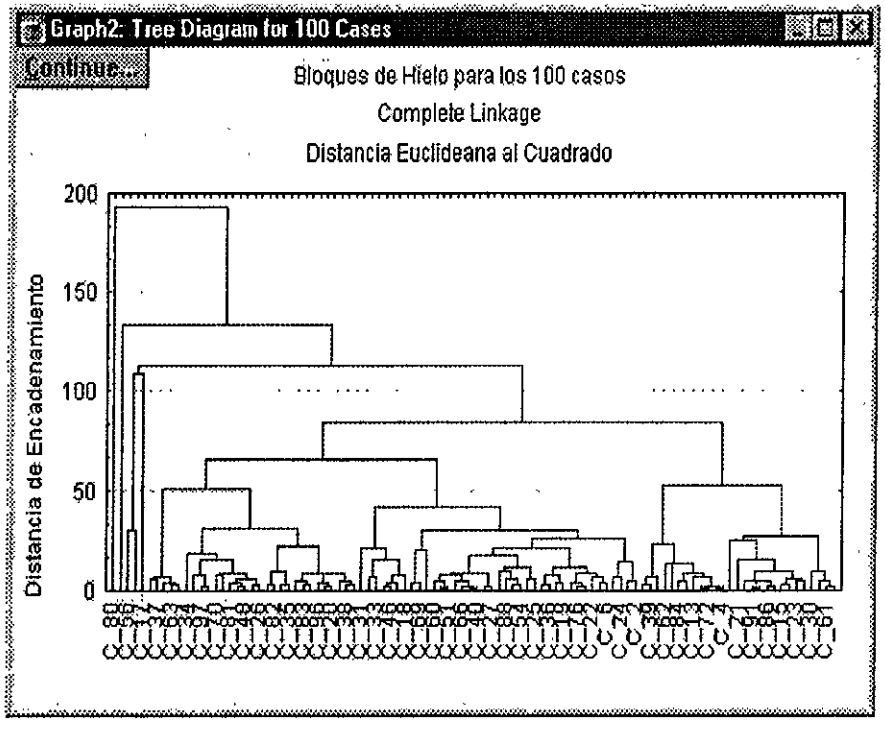
Por su parte, el método de Complete Linkage (vecino más lejano), de acuerdo con la teoría, elimina el problema del serpenteo que existe utilizando Single Linkage (vecino más cercano); y tenemos la esperanza de que nos de como resultado, a diferencia de lo que podría suceder con el Método de Ward, un número manejable de conglomerados como solución.
Una vez establecidas las condiciones del análisis veamos que resultados obtenemos de su aplicación.



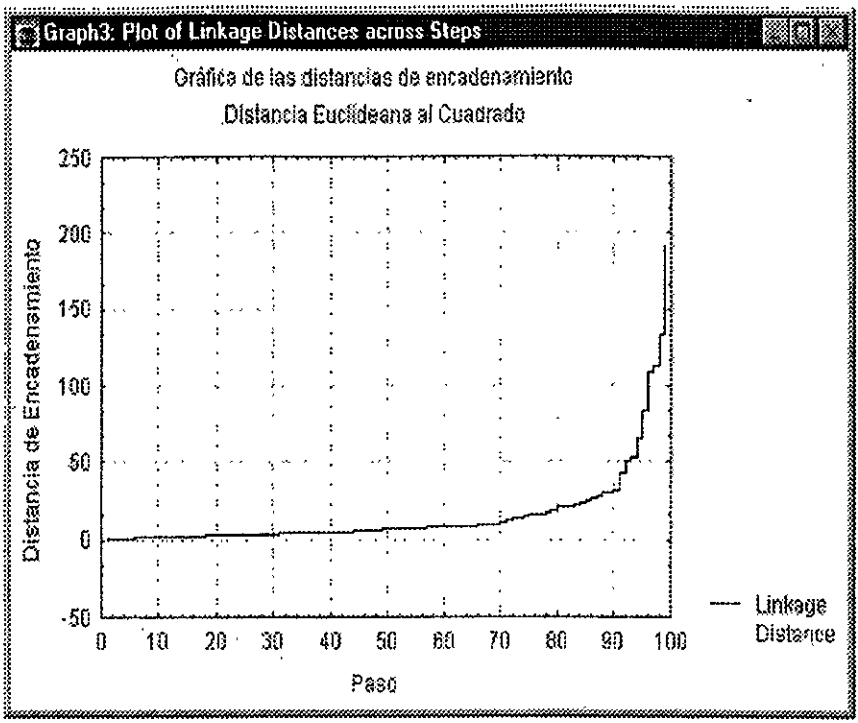
En el siguiente dendograma podemos observar cuatro agrupaciones principales. Una de ellas (ver la parte inferior de la gráfica) nos llama más la atención debido al número tan reducido que la integran. Por una parte debemos averiguar que los hace tan distintos al resto del conjunto de datos y por otra, que semejanzas existen entre ellos.



En la siguiente gráfica observamos lo mismo que en el dendograma. Aunque, a mi parecer, la información en este formato se puede "leer" de manera mucho más clara y sencilla.



Lo que podemos comentar acerca de la gráfica que aparece a continuación, nos referimos a la gráfica de distancias de encadenamiento, es lo siguiente. Los saltos que observamos en la gráfica son pequeños y más o menos regulares hasta que la distancia de 50. Ahora, regresemos a la gráfica anterior o al dendograma. ¿Qué sucede a la distancia de 50? La respuesta es que a esta distancia se empieza a observar en el dendograma (o en los cubos de hielo) una diferenciación más clara entre los conglomerados que se han formado.



Vamos a utilizar el criterio de Mojena (ver capítulo 3) para obtener el número natural de grupos que existen en nuestro conjunto de datos.

Primero hacemos uso de la información de la amalgamation schedule (Ruta de conglomeración), para ser precisos necesitamos las distancias de encadenamiento (linkage distances) a lo largo de todo el proceso. Esta información la encontraremos en la tabla 4 del apéndice. Una vez que las tenemos debemos calcular la media ($\overline{\alpha}$) y la desviación standard (S_α) del conjunto de distancias de encadenamiento.

Como resultado de nuestros cálculos obtenemos los siguientes valores:

$\overline{\alpha} = 21.0291019$

$S_\alpha = 32.5012783$

Ahora bien, recordemos la forma matemática del criterio de Mojena: $\alpha_{j+1} > \overline{\alpha} + kS_\alpha$ donde $k=1.25$.

Una vez hecho esto calculemos el umbral establecido por el criterio de Mojena

$\overline{\alpha} + kS_\alpha = 61.6556998$

Bueno, ahora debemos regresar a la amalgamation schedule (ruta de conglomeración, tabla 4) para encontrar el α_j ó distancia de encadenamiento que cumpla con el criterio de Mojena.

Buscamos en toda la tabla y encontramos la siguiente distancia de encadenamiento $\alpha_j = 65.34326$ es la primera distancia de encadenamiento mayor que el umbral, pero el criterio establece que debe ser la siguiente distancia de encadenamiento, por tanto $\alpha_{j+1} = 83.88204$ es la distancia que buscamos.

Para saber que elementos formarán los conglomerados solución, es necesario apoyarnos en la información que nos brindan las gráficas de los dendogramas y los cubos de hielo, además de aquella contenida en la amalgamation schedule (Ruta de Conglomeración).

Pues bien, una vez que sabemos que $\alpha_{j+1} = 83.88204$, la localizamos aproximadamente en el eje vertical de la gráfica de Bloques de Hielo (o en el dendograma) y nos fijamos en las agrupaciones que se hallan formado antes de esta distancia (umbral). Si nos fijamos bien en la gráfica, la primera agrupación se produce a una distancia aproximada de 64 unidades, buscamos la distancia exacta en la Ruta de Conglomeración y, la distancia exacta es 65.34326. Si recorremos de izquierda a derecha el renglón correspondiente a dicha distancia, tendremos todos los elementos pertenecientes al grupo.

La siguiente agrupación se forma a una distancia aproximada de 51 unidades, buscamos en la Ruta de Conglomeración la distancia exacta, resulta ser 52.88701. Nuevamente, al recorrer el renglón de izquierda a derecha encontramos a los elementos agrupados en esta distancia.

Por último, a una distancia aproximada de 30 unidades, se observa una pequeña agrupación formada, según la gráfica, por 2 elementos. Buscamos en la Ruta de Conglomeración, la distancia exacta es de 30.16358, y los elementos son la ciudad 28 y 56.

Siguiendo este procedimiento llegamos a los siguientes resultados:
Grupo 1. (individuos)
1,61,44,30,3,23,5,15,14,86,53,91,90,21,67,4,11,72.77,13,12,84,58,62,32,39,47.
Grupo 2. (individuos)
2,71,24,74,6,78,22,55,76,99,19,92,36,41,25,93,54,85,88,89,27,29,40,50,66,49,51,59,
60,10,69,79,18,45,46,98,33,42,31,7,38,9,20,57,96,16,83,100,35,65,82,94,26,73,48,75,
81,68,70,87,97,95,34,8,63,43,37,64.
Grupo 3 (individuos) 28,56.

5.2.5 Interpretación de los conglomerados.

Calculamos las medias de cada uno de los conglomerados que obtuvimos. Le damos interpretación a cada uno de los valores de los individuos representantes de su grupo (principales características y diferencias).

VECTORES DE MEDIAS UTILIZANDO LOS VALORES ORIGINALES Y LAS 14 VARIABLES

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
M1	402164 52	92 71	41.36	47.49	81433.63	4 82	51 57	1.32	3 33	2.28	0 58	21 34	16.51	3.06
M2	209866.46	89.79	24 07	54.74	42752.24	4.89	20.11	2.05	3.80	0.83	2 26	36 06	29 51	5 38
M3	65907.50	81.80	22 90	49 70	11338 50	5 25	7 55	10 23	4 70	0.00	7.47	18.07	22 70	29 30

Podemos dividir a los grupos de la siguiente manera:
G1. Poblacion total alta.
G2. Poblacion total media.
G3. Poblacion total baja.

En cuanto al total de viviendas particulares habitadas (totvphab) en:
G1. Total alto de vph
G2. Total medio de vph
G3. Total bajo de vph

Ahora vamos a dividir a los grupos de acuerdo a su vocación en:
G1. Principalmente ocupado en el sector manufacturero.
G2. Principalmente dedicado al sector comercio y servicios.
G3. Principalmente ocupado en el sector comunicaciones y transportes y destacado en ocupación en el sector servicios.

5.2.6 Validación de los conglomerados.

Para validar nuestros resultados vamos a aplicarle a los datos un análisis de componentes principales, elegiremos un número determinado de componentes (aquellos que reunan más del 90 % de la varianza total) y estos factores se suponen no están correlacionados entre sí. A estos datos habremos de practicarles el mismo tipo de análisis que previamente aplicamos a los datos originales. Una vez hecho esto compararemos los resultados.

Como una forma adicional de comparación vamos a incluir, en este caso particular, dos mapas de la República Mexicana. Cada uno de ellos reflejando los resultados del análisis correspondiente.
Además, en esta etapa haremos uso de la distancia de Mahalanobis para calcular las distancias entre los elementos representantes de cada grupo solución. La distancia existente entre ellos será un parámetro que nos indique que tan distintos son entre sí.

5.2.6.1 Utilización de Componentes principales para la validación.

Cabe la posibilidad de que dentro de las 14 variables que consideramos en el análisis anterior existen algunas que nos dan información redundante o, en otras palabras, están correlacionadas.
Para resolver este tipo de problemas podemos hacer uso de la técnica que presentamos en el capítulo 2: Análisis de Componentes Principales.

Recordemos los pasos a seguir para llevar a cabo dicho análisis.

Paso 1. Vamos a estandarizar los datos. Lo cual es muy conveniente en nuestro caso debido a la disparidad que existe entre las medidas de las variables involucradas en el estudio. Veamos un ejemplo: Población Total vs. Porcentaje de alfabetización. De no llevar a cabo la estandarización de los datos el efecto (el peso) de la variable población total anularía el efecto del porcentaje de alfabetización. Este paso ya fue efectuado previamente.

Paso 2. Obtenemos la matriz de covarianza. Dado que se llevó a cabo primero la estandarización de los datos esta matriz es una matriz de correlación.

Covariances (ciliestd.sta)

Confirm...

Casewise deletion of MD
N=100

Variable	%POBTOT	%ALFA	%POBSEC	%POBTER	%POBTPAB	%PROMOVP	%MANUFAC	%AGRICT	%CONSTRU	%ELECTRI	%PESCA	%COMERCI	%SERVICI	%TRANSP
%POBTOT	1.00	.32	.24	.20	.99	-.01	.22	-.06	.37	-.06	-.15	-.15	-.19	.02
%ALFA	.32	1.00	.26	.31	.35	-.56	.24	-.02	.30	.11	-.16	-.28	-.12	-.11
%POBSEC	.24	.26	1.00	-.38	.22	.03	.08	-.06	-.13	.09	-.33	-.49	-.59	-.34
%POBTER	.20	.31	-.38	1.00	.23	-.42	-.47	.04	.47	-.05	-.27	.18	.51	.35
%POBTPAB	.99	.35	.22	.23	1.00	-.10	.21	-.05	.38	-.05	-.15	-.16	-.17	.01
%PROMOVP	-.01	-.56	.03	-.42	-.10	1.00	.01	-.08	-.16	-.10	.13	.22	-.20	.00
%MANUFAC	.22	.24	.08	-.47	.21	.01	1.00	-.17	-.19	.02	-.24	-.52	-.66	-.38
%AGRICT	-.06	-.02	-.06	.04	-.05	-.08	-.17	1.00	-.04	.00	-.01	-.01	-.01	.18
%CONSTRU	.37	.30	-.13	.47	.38	-.16	-.19	-.04	1.00	-.08	-.12	.00	-.09	.33
%ELECTRI	-.06	.11	.09	-.05	-.05	-.10	.02	.00	-.08	1.00	.02	-.34	-.12	-.11
%PESCA	-.15	-.16	-.33	.27	-.15	.13	-.24	-.01	-.12	.02	1.00	-.21	-.11	-.01
%COMERCI	-.15	-.28	-.49	.18	-.16	.22	-.52	-.01	.00	-.34	-.21	1.00	.30	.00
%SERVICI	-.19	-.12	-.59	.51	-.17	-.20	-.66	-.01	-.09	-.12	-.11	.30	1.00	.07
%TRANSP	.02	-.11	-.34	.35	.01	.00	-.38	.18	.33	-.11	-.01	.00	.07	1.00

Paso 3. Encontramos los eigenvalores y eigenvectores correspondientes a esta matriz de correlación. Recordemos que la suma de los eigenvalores dará como resultado el número de variables involucradas en el estudio. En nuestro caso 14.

Eigenvalues (citiestd.sta)

Extraction: Principal components

Value	Eigenval	% Total variance	Cum. Eigenval	Cum. %
1	3.47141	24.79580	3.47141	24.7958
2	2.874829	20.53449	6.34624	45.3303
3	1.622882	11.59201	7.96912	56.9223
4	1.349172	9.63694	9.31830	66.5593
5	1.105036	7.89312	10.42333	74.4524
6	.903541	6.45386	11.32687	80.9062
7	.873091	6.23637	12.19996	87.1426
8	.678118	4.84370	12.87808	91.9863
9	.409831	2.92737	13.28791	94.9137
10	.337425	2.41018	13.62534	97.3238
11	.229838	1.64170	13.85518	98.9655
12	.111656	.79754	13.96683	99.7631
13	.030282	.21630	13.99711	99.9794
14	.002887	.02062	14.00000	100.0000

Paso 4. En este último paso decidiremos cuales de los componentes principales son de utilidad. En otras palabras si podemos reducir el número original de variables perdiendo un mínimo de infomación. Esta reducción de variables nos podría ayudar a que la información sea más manejable.
La elección del número de componentes principales es un tanto arbitraria y depende exclusivamente del investigador. En nuestro caso, vamos a conformarnos con los 8 primeros componentes principales. Con éstos tenemos un poco más del 91% de la varianza.

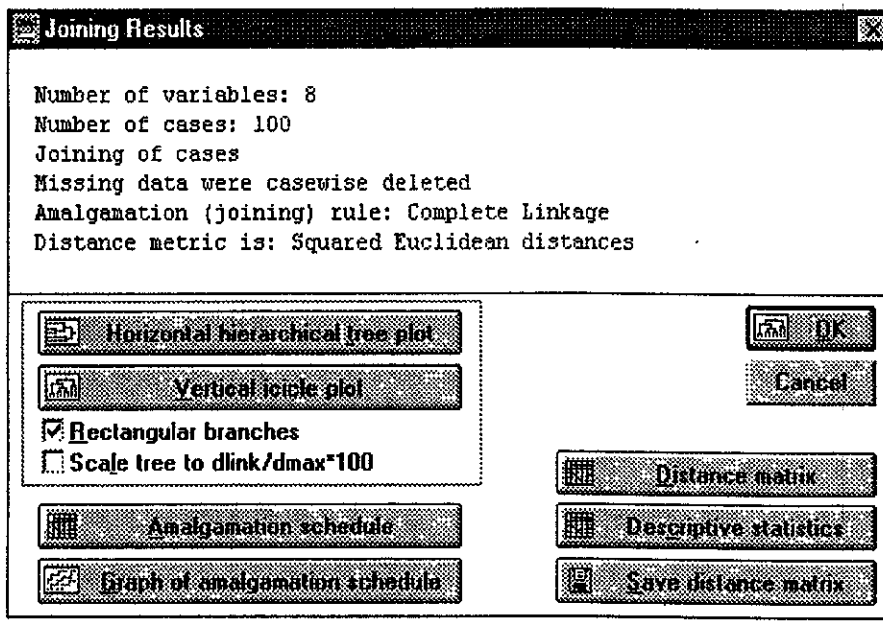
Esto significa que tendremos ahora una matriz de datos de 100x8 en vez de tener una matriz de datos de 100x14 como teníamos antes.

Recordemos que esta matriz de 100x8 la obtendremos multiplicando la matriz original de datos por la matriz de factores (formada por los eigenvectores de la matriz de covarianza). El producto de esta multiplicación será una matriz de 100x14, de la cual sólo habremos de tomar las primeras 8 columnas. A estas nuevas variables las llamaremos Z's. Estos datos están disponibles en la tabla 5 del apéndice.

A continuación presentamos la matriz de eigenvectores estandarizados.

EIGENVECTORES ESTANDARIZADOS

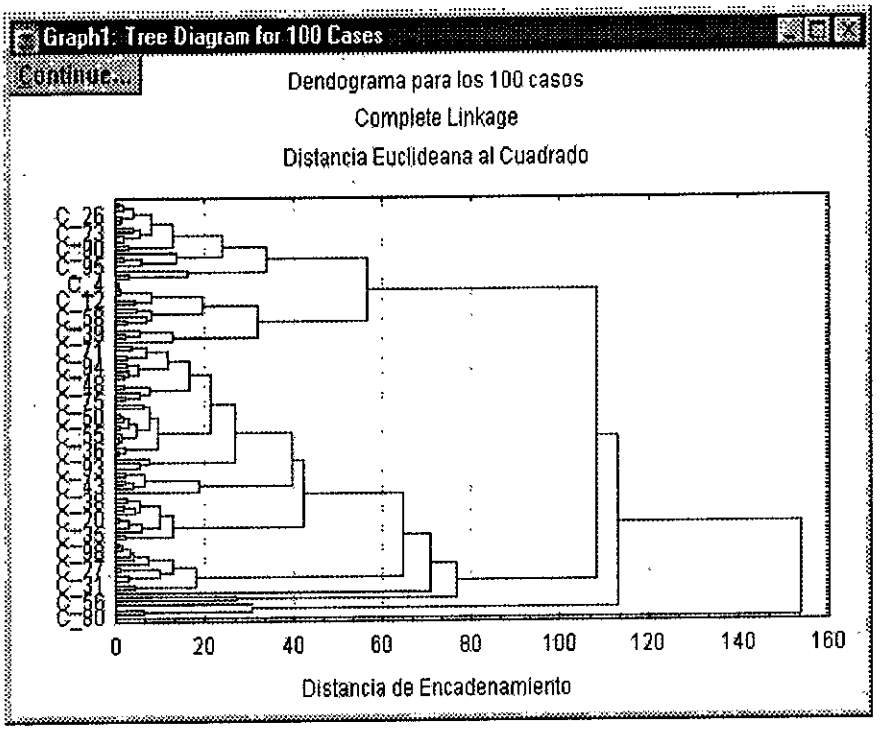
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Factor 11	Factor 12	Factor 13	Factor 14
V1	0.248	-0.387	-0.361	-0 063	0.175	-0.306	-0.089	-0.108	0 093	-0 032	0 042	0 006	-0.007	-0 706
V2	0 213	-0.340	0 353	0.057	0.070	0.041	0 251	0 285	0.272	0.646	0.255	0.060	-0 006	0.022
V3	0.478	0 076	0 031	0 166	-0.216	0.058	-0.038	-0 095	-0.157	0.132	-0.166	-0 779	0 019	-0 007
V4	-0.221	-0 456	0 142	0 102	-0 030	0.067	-0.101	-0 098	-0 155	0 192	-0.783	0 107	-0 034	0.008
V5	0 245	-0.408	-0.314	-0.061	0 182	-0 311	-0 057	-0 099	0 119	-0 137	0 035	-0 035	-0 010	0 702
V6	-0.018	0 289	-0 556	-0 073	-0.011	0 025	-0.308	0 042	-0 225	0 651	-0 011	0.152	-0 010	0.087
V7	0 493	0.117	0 011	0.149	-0 110	0 117	0 091	-0.134	-0 008	-0 065	-0 150	0 436	0.672	0 011
V8	-0.072	-0 017	0 094	-0 237	-0.656	-0 637	0 110	0.189	-0 177	0 040	-0.001	0 059	0.096	0 002
V9	-0 015	-0 412	-0 142	-0 197	-0.070	0.422	0 000	0 411	-0 544	-0 172	0 229	-0 060	0 192	-0 005
V10	0 097	0.045	0 385	-0 266	0 159	-0 099	-0 789	0 234	0.123	-0 044	-0.009	-0 044	0 195	-0 003
V11	-0 085	0 184	-0 007	-0 599	0 437	-0 101	0 403	0 060	-0 013	0 093	-0.267	-0 235	0 312	-0 009
V12	-0 325	0.006	-0 308	0.409	-0 035	-0.058	-0.005	0.499	0 401	-0 063	-0.123	-0 235	0 380	-0.013
V13	-0 381	-0.121	0 170	0 277	0.197	-0.229	-0.069	-0.431	-0 310	0 169	0.342	-0 167	0.429	0 010
V14	-0.201	-0.191	-0 121	-0.395	-0.424	0 355	-0.091	-0 402	0 452	0.084	0 108	-0 129	0.192	0 004

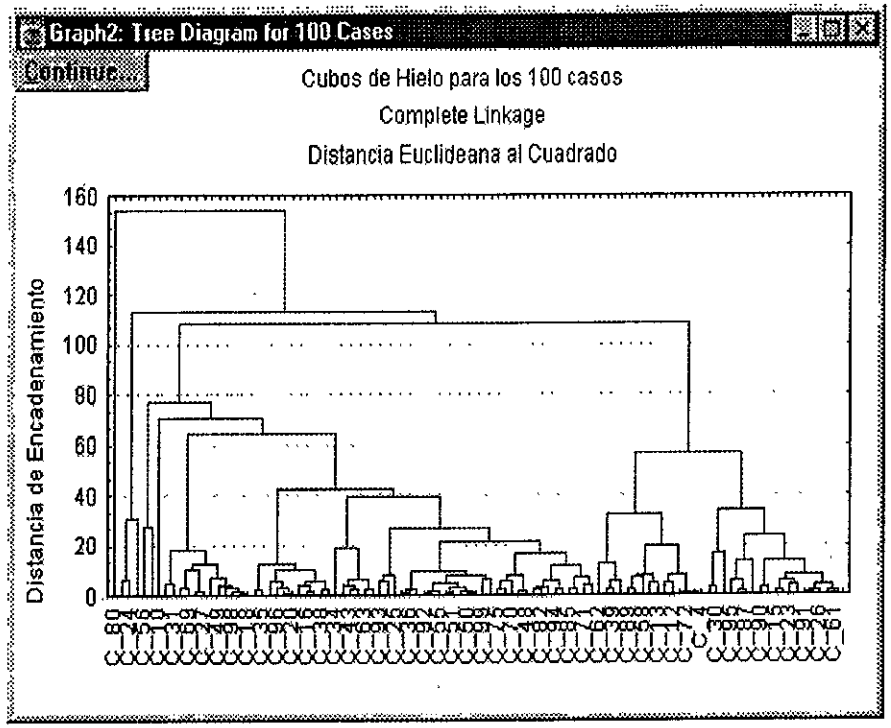


Hagamos una comparación de manera "visual" entre esta gráfica y la que obtuvimos anteriormente haciendo uso de todas las variables. Existen similitudes entre ellas. Para empezar coincide el número de conglomerados solución, es decir, 3 grupos. Y por otro

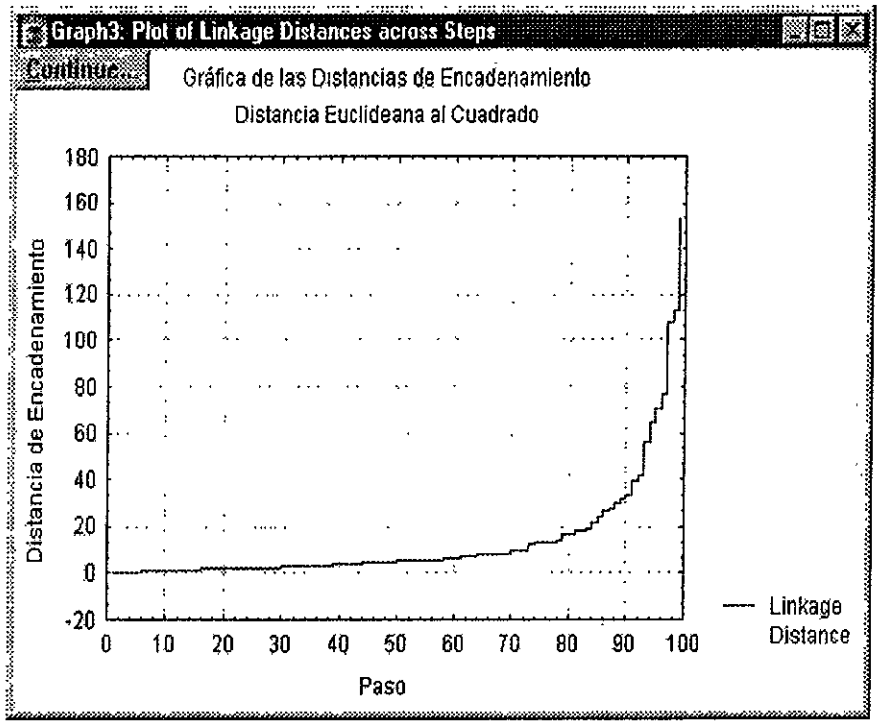
lado, sigue existiendo un grupo, (ver la parte inferior de la gráfica) que es muy reducido y parece estar muy distante del resto.

Finalmente, en esta gráfica parece que hay una más clara diferenciación entre los conglomerados.





En la siguiente gráfica, correspondiente a las distancias de encadenamiento, los saltos que se observan en ella son pequeños. Esto hasta antes de llegar a la distancia de 40. A partir de esa distancia comienzan a ser más pronunciados. Si consultamos el dendograma correspondiente (o los cubos de hielo) nos daremos cuenta que a partir de esa distancia se empieza a observar una mayor diferenciación entre los conglomerados que se han formado.



Vamos a utilizar el criterio de Mojena (ver capítulo 3) para obtener el número natural de grupos que existen en nuestro conjunto de datos.

Primero hacemos uso de la información de la amalgamation schedule (ruta de conglomeración), para ser precisos necesitamos las distancias de encadenamiento (linkage distances) a lo largo de todo el proceso. Dicha información se encuentra en la tabla 6 del apéndice. Una vez que las tenemos debemos calcular la media ($\overline{\alpha}$) y la desviación standard (S_α) del conjunto de distancias de encadenamiento.

Como resultado de nuestros cálculos obtenemos los siguientes valores:

$\overline{\alpha} = 25.3405761$

$S_\alpha = 31.4661159$

Ahora bien, recordemos la forma matemática del criterio de Mojena:

$\alpha_{j+1} > \overline{\alpha} + kS_\alpha$ donde $k=1.25$.

Una vez hecho esto calculemos el umbral establecido por el criterio de Mojena

$\overline{\alpha} + kS_\alpha = 64.673221$

Bueno, ahora debemos regresar a la amalgamation schedule (ruta de conglomeración) para encontrar el α_j ó distancia de encadenamiento que cumpla con el criterio de Mojena.
Buscamos en toda la tabla y encontramos la siguiente distancia de encadenamiento $\alpha_j = 70.87754$ es la primera distancia de encadenamiento mayor que el umbral, pero el criterio establece que debe ser la siguiente distancia de encadenamiento, por tanto $\alpha_{j+1} = 76.81706$ es la distancia que buscamos.

Para saber qué elementos formarán los conglomerados solución, es necesario apoyarnos en la información que nos brindan las gráficas de los dendogramas y los cubos de hielo, además de aquella contenida en la amalgamation schedule (Ruta de conglomeración).

Una vez que tenemos que $\alpha_{j+1} = 76.81706$, observamos la gráfica de bloques de hielo (o el dendograma) y calculamos donde quedaría ubicada aproximadamente esta distancia de encadenamiento. Una vez ubicado este punto fijamos en los grupos que se formaron antes de esa distancia. La primera agrupación se forma a una distancia aproximada de 70, buscamos en la amalgamation schedule y, encontramos que la distancia exacta es de 70.87754, de esta agrupación quedarían excluídos dos elementos que, a una distancia de 76.81706 se habrían unido al grupo. Debido a la cercanía de estos dos elementos con el grupo formado a la distancia de 70.87754 y a su lejanía con las otras agrupaciones formadas, consideramos prudente incluir estos dos elementos, a saber, 28 y 56.

El siguiente, es el que se forma a una distancia próxima a 60 unidades revisamos la amalgamation schedule para precisar que distancia exacta es, en este caso resulta 56.51686 unidades. Recordemos, que si recorremos de izquierda a derecha el renglón correspondiente a esta distancia, podremos identificar a todos y cada uno de los elementos conglomerados hasta esa distancia, éstos elementos conformarán el G1.

El último en formarse, es el que se observa a la izquierda de la gráfica de cubos de hielo, se forma a una distancia aproximada de 30 unidades. Buscamos esta distancia en la amalgamation schedule, y la distancia exacta es 30.28623 unidades, y el grupo esta formado por los elementos 17, 24 y 74.

Utilizando este criterio llegamos a los siguientes resultados:

Grupo 1. (individuos) 1, 61, 14, 26, 53, 91, 3, 23, 5, 15, 86, 90, 67, 87, 97, 95, 21, 30, 44, 4, 11, 72, 77, 12, 84, 13, 78, 58, 60, 89, 32, 39, 47 y 62.

Grupo 2. (individuos) 2, 71, 54, 85, 6, 94, 52, 82, 88, 48, 68, 70, 73, 75, 19, 99, 40, 50, 66, 51, 59, 55, 76, 92, 22, 36, 41, 25, 81, 93, 8, 63, 37, 43, 64, 34, 7, 38, 9, 16, 100, 20, 57, 96, 83, 35, 65, 18, 46, 98, 45, 49, 33, 27, 29, 69, 79, 31, 42, 10, 28 y 56.

Grupo 3. (individuos) 17, 24 y 74.

VECTORES DE MEDIAS UTILIZANDO LOS VALORES PRODUCTO DEL ANÁLISIS DE COMPONENTES PRINCIPALES

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
M1	397875 74	92 76	38.99	49.19	81571.35	4 81	47 69	1.19	3 66	2 16	0 77	23 06	18 01	3 45
M2	187658.89	89 01	23 30	54.64	37646 02	4 92	18.14	2.50	3 68	0.27	2 30	36 61	30.09	6,42
M3	123995.33	93.43	24 87	53 93	26281.67	4.67	14 98	3 34	1 60	21.09	6.46	22.10	27 35	3.07

Podemos dividir a los grupos de la siguiente manera:

G1. Poblacion total alta.
G2. Poblacion total media.
G3. Poblacion total baja.

En cuanto al total de viviendas particulares habitadas (totvphab) en:

G1. Total alto de vph
G2. Total medio de vph
G3. Total bajo de vph

Ahora vamos a dividir a los grupos de acuerdo a su vocación en:

G1. Principalmente ocupado en el sector manufacturero.
G2. Principalmente dedicado al sector comercio y servicios.
G3. Principalmente ocupado en el sector servicios y el más destacado en ocupación en el sector eléctrico.

5.2.6.2 Utilización de la Distancia de Mahalanobis para la validación.

A continuación vamos a hacer un análisis utilizando la distancia de Mahalanobis. Dicho análisis consistirá en tomar el vector de medias de cada uno de los conglomerados que obtuvimos a partir de los dos análisis que realizamos previamente.

Haremos lo siguiente, una vez obtenido el vector de medias de cada uno de los conglomerados que obtuvimos calcularemos la distancia de Mahalanobis que existe entre cada uno de ellos.

Podemos preguntarnos: ¿ Porqué utilizar la distancia de Mahalanobis y no otra? La respuesta es sencilla, esta función de distancia nos ayuda a conocer cuán lejos o cerca se encuentra una población de otra.

Esto podría proveernos de cierta evidencia de que tan válida es la separación que hicimos al formar estos conglomerados.

Si quisieramos hacer un estudio más minucioso, que el presentado en este trabajo sólo habremos de proponer y no a desarrollar, sería lo siguiente. Una vez obtenidos los vectores de medias podemos hacer una prueba de diferencia de medias multivariada. Esto con la intención de ver si en realidad los individuos pertenecen a poblaciones distintas.

5.2.6.2.1 Cálculo de la Distancia de Mahalanobis para el Análisis Original.

Pues bien, calculemos la distancia de Mahalanobis. Basándonos en los resultados que obtuvimos en el análisis de conglomerados con el que comenzamos el capítulo tenemos 3 grupos distintos. Dichos grupos están formados como sigue:

Grupo 1. (individuos) 1, 61, 44, 30, 3, 23, 5, 15, 14, 86, 53, 91, 90, 21, 67, 4, 11, 72, 77, 13, 12, 84, 58, 62, 32, 39 y 47.

Grupo 2. (individuos)2, 71, 24, 74, 6, 78, 22, 55, 76, 99, 19, 92, 36, 41, 25, 93, 54, 85, 88, 89, 27, 29, 40, 50, 66, 49, 51, 59, 60, 10, 69, 79, 18, 45, 46,98, 33, 42, 31, 7, 38, 9, 20, 57, 96, 16, 83, 100, 35, 65, 82, 94, 26, 73, 48, 75, 81, 68, 70, 87. 97, 95, 34, 8, 63, 43, 37 y 64.

Grupo 3 (individuos) 28 y 56.

Ahora, regresando a los 3 grupos que estamos estudiando. Para cada uno de estos grupos debemos calcular la matriz de covarianza. Una vez realizado este proceso tendremos 3 matrices de covarianza. Nosotros solamente necesitamos una, ¿Qué vamos a hacer? Pues vamos a utilizar las tres para formar una matriz de covarianza combinada (pooled). Dicha matriz tiene la siguiente forma:

$$C = \frac{\sum_{i=1}^{m} (n_i - 1) C_i}{\sum_{i=1}^{m} (n_i - 1)}$$

Ya casi terminamos de describir el proceso a seguir, ¿Qué nos falta?. Recordemos que la distancia de Mahalanobis hace uso de la matriz inversa de la covarianza, entonces habremos de calcular la inversa.
Finalmente, multiplicamos esta matriz por los vectores diferencia entre todas las medias.

MATRIZ DE COVARIANZA 1

1.718	0.077	-0.425	0 560	1.653	0.457	-0 558	-0.019	0 470	-0 318	-0.077	0 488	0 339	0.374
0.077	0.502	0 037	0 162	0.130	-0 429	0 039	-0 093	0 114	-0 207	-0.037	0 010	0 063	-0.027
-0 425	0.037	0 472	-0 350	-0.450	0 001	0 392	-0.036	-0 270	0.025	-0 016	-0 297	-0 189	-0.163
0.560	0 162	-0.350	0 483	0 570	-0.049	-0 342	-0.028	0 394	-0.289	0 003	0 302	0.220	0.157
1 653	0.130	-0 450	0 570	1 627	0 271	-0 561	0 005	0 465	-0 310	-0.073	0 479	0.352	0.338
0.457	-0 429	0 001	-0.049	0 271	1 136	-0 089	-0 111	0 095	0 142	-0 028	0.081	-0 053	0.224
-0.558	0.039	0.392	-0.342	-0 561	-0 089	0.468	-0 085	-0 353	-0 003	-0.016	-0.328	-0.223	-0.166
-0 019	-0.093	-0 036	-0.028	0 005	-0 111	-0.085	0 333	-0.003	-0 001	0.034	-0 001	-0.017	-0 056
0 470	0.114	-0.270	0 394	0.465	0 095	-0.353	-0 003	0 837	-0 088	0 041	0.175	0.126	0.089
-0.318	-0.207	0.025	-0 289	-0 310	0.142	-0.003	-0 001	-0 088	0 915	0 003	-0.197	-0 163	-0.025
-0.077	-0.037	-0 016	0.003	-0 073	-0 028	-0 016	0 034	0 041	0 003	0 047	-0 020	-0 013	-0.014
0 488	0 010	-0 297	0.302	0 479	0 081	-0 328	-0.001	0 175	-0 197	-0.020	0 332	0.205	0 150
0.339	0 063	-0.189	0 220	0 352	-0 053	-0.223	-0 017	0 126	-0 163	-0 013	0.205	0.192	0 066
0.374	-0 027	-0 163	0.157	0 338	0.224	-0 166	-0.056	0 089	-0 025	-0 014	0 150	0 066	0 185

MATRIZ DE COVARIANZA 2

0.521	0 216	0 005	0 234	0.537	-0 101	0.019	0 002	0.364	0 021	0.009	-0 107	-0 119	0 121
0.216	1.027	-0 005	0.447	0 247	-0.574	-0 030	0 077	0 363	0 149	0.033	-0.259	-0 038	0.191
0.005	-0 005	0.347	-0 134	0.002	0 181	0.211	0 024	-0 095	0.029	-0.108	0.038	-0 240	-0 129
0.234	0 447	-0.134	1.067	0.270	-0 570	-0.240	-0 067	0.494	0.017	-0.147	-0.186	0.366	0 373
0.537	0 247	0.002	0 270	0 561	-0 166	0.009	0 009	0 390	0 029	0 011	-0.124	-0 110	0 132
-0 101	-0.574	0 181	-0.570	-0 166	0.923	0 205	-0 127	-0 312	-0.084	-0.041	0 282	-0 281	-0.201
0.019	-0 030	0.211	-0 240	0.009	0 205	0.313	-0 108	-0 116	0.014	-0.045	-0.022	-0 320	-0 126
0.002	0 077	0 024	-0 067	0 009	-0.127	-0 108	1.126	0 092	-0.028	0 087	-0.196	-0 230	-0 072
0 364	0.363	-0 095	0 494	0.390	-0 312	-0 116	0 092	1.038	0 027	-0.033	-0 142	-0 240	0.430
0.021	0.149	0 029	0 017	0.029	-0.084	0.014	-0 028	0 027	0 310	0.080	-0 151	-0.052	-0 026
0.009	0 033	-0 108	-0 147	0.011	-0.041	-0 045	0.087	-0 033	0 080	0 246	-0 158	-0 007	-0.024
-0.107	-0.259	0.038	-0 186	-0.124	0 282	-0 022	-0 196	-0 142	-0 151	-0.158	0 622	-0.178	-0 105
-0.119	-0 038	-0 240	0.366	-0 110	-0.281	-0 320	-0.230	-0 240	-0.052	-0 007	-0 178	0 945	-0 071
0 121	0.191	-0 129	0.373	0 132	-0.201	-0 126	-0 072	0 430	-0.026	-0 024	-0 105	-0.071	0.633

MATRIZ DE COVARIANZA 3

-0 568	-0.644	-0.368	0.062	-0 583	0 296	-0 831	1 078	-0 042	1.513	0 273	-0.919	-0 030	1.889
-0 644	-0.437	-0.328	0 210	-0 662	0.307	-1 055	1.573	0.030	2 474	0 156	-1 141	0.021	1 755
-0.388	-0 328	-0 163	0 071	-0.383	0 290	-0 587	0 942	0 123	0 987	0 097	-0.640	-0 025	1 260
0 062	0 210	0.071	0.062	0 065	-0.072	0 041	0 016	0.011	0 249	-0.095	0 057	0.035	-0.429
-0 583	-0.662	-0.383	0.065	-0.598	0 293	-0 851	1 090	-0.057	1.574	0.284	-0 941	-0.028	1.925
0 296	0 307	0 290	-0.072	0 293	0 082	0 413	-0.296	0 291	-1 358	-0 201	0.457	-0.040	-0 610
-0.831	-1.055	-0.587	0.041	-0 851	0.413	-1.170	1.402	-0 129	1 994	0 468	-1.303	-0 057	2.868
1.078	1 573	0 942	0 016	1.090	-0 296	1 402	-1 236	0.530	-2.603	-0.795	1 584	0.058	-3.614
-0.042	0.030	0.123	0 011	-0.057	0.291	-0.129	0 530	0.346	-0 274	-0 107	-0 132	-0 047	0 437
1 513	2 474	0.987	0.249	1.574	-1 358	1 994	-2 603	-0 274	-1 035	-0.967	2 264	0.325	-6 750
0 273	0 156	0.097	-0 095	0.284	-0 201	0.468	-0 795	-0 107	-0 967	-0.028	0 503	0.001	-0.812
-0.919	-1 141	-0.640	0.057	-0.941	0.457	-1 303	1.584	-0 132	2 264	0 503	-1 450	-0.059	3 141
-0 030	0 021	-0 025	0.035	-0 028	-0.040	-0 057	0 058	-0 047	0 325	0 001	-0 059	0 019	-0 045
1.889	1 755	1.260	-0.429	1.925	-0 610	2 868	-3.614	0.437	-6 750	-0 812	3 141	-0 045	-5.290

MATRIZ DE COVARIANZA COMBINADA

0.826	0 160	-0.121	0 320	0 819	0 060	-0 157	0 019	0 384	-0 040	-0 009	0.039	0 008	0 228
0 160	0 852	0.000	0.364	0 196	-0 516	-0 033	0 062	0 288	0 100	0.017	-0 204	-0 009	0 164
-0.121	0 000	0.370	-0 189	-0 130	0 134	0 244	0 027	-0 138	0 048	-0 079	-0 068	-0.222	-0 109
0.320	0.364	-0.189	0.886	0.348	-0 417	-0 262	-0 054	0 456	-0 062	-0 105	-0 047	0.319	0 297
0 819	0 196	-0.130	0 348	0 828	-0 036	-0 165	0 031	0 401	-0 031	-0 006	0.024	0 018	0 226
0 060	-0.516	0.134	-0.417	-0.036	0.964	0.129	-0.126	-0.188	-0 049	-0 041	0 231	-0.214	-0.093
-0.157	-0 033	0 244	-0 262	-0 165	0.129	0 324	-0.070	-0.181	0 051	-0 026	-0.133	-0 288	-0 074
0.019	0 062	0 027	-0.054	0 031	-0.126	-0.070	0.860	0.075	-0 075	0 053	-0 105	-0 166	-0 142
0.384	0.288	-0.138	0.456	0.401	-0.188	-0.181	0.075	0.968	-0.011	-0 014	-0 055	-0.136	0 337
-0.040	0 100	0.048	-0.062	-0 031	-0.049	0 051	-0.075	-0.011	0.447	0 037	-0.113	-0 075	-0.167
-0.009	0 017	-0.079	-0 105	-0 006	-0.041	-0 026	0.053	-0 014	0.037	0.186	-0 106	-0 008	-0.038
0 039	-0 204	-0 068	-0 047	0 024	0 231	-0 133	-0.105	-0 055	-0 113	-0 106	0 499	-0 071	0 034
0.008	-0 009	-0.222	0.319	0.018	-0.214	-0.288	-0.166	-0.136	-0.075	-0 008	-0 071	0.720	-0 033
0 228	0 164	-0.109	0.297	0 226	-0.093	-0.074	-0 142	0 337	-0.167	-0.038	0.034	-0 033	0 385

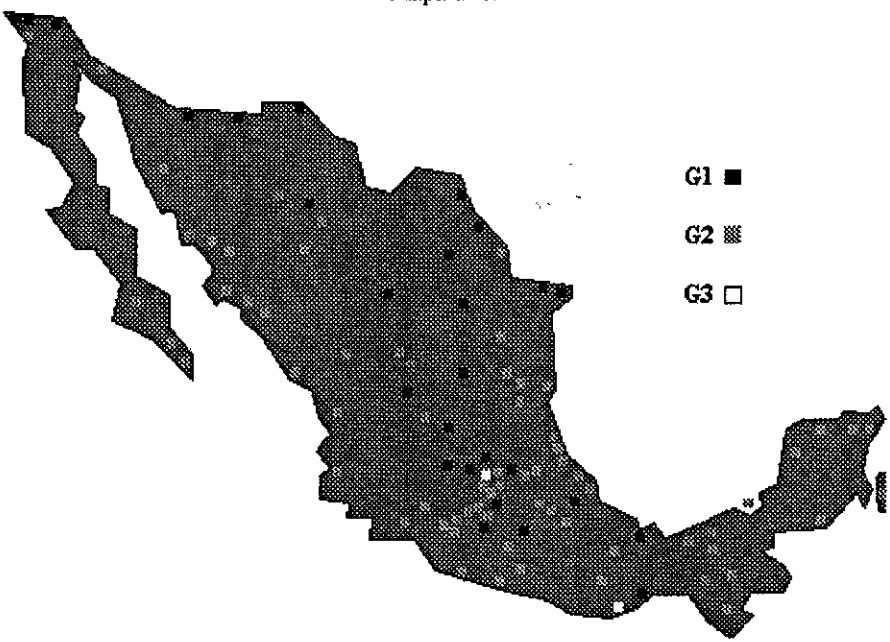
MATRIZ INVERSA DE COVARIANZA COMBINADA

199	-5	-1	-1	-194	-26	3654	1073	692	938	1569	2069	2204	1051
-5	3	-1	0	5	2	1688	498	319	434	726	955	1021	491
-1	-1	7	0	1	-1	-834	-245	-157	-213	-354	-469	-502	-240
-1	0	0	4	0	1	1716	505	323	441	740	971	1036	497
-194	5	1	0	190	25	-3243	-952	-614	-833	-1392	-1837	-1955	-932
-26	2	-1	1	25	6	280	83	53	73	121	158	170	83
3654	1688	-834	1716	-3243	280	8235135	2426693	1554561	2121295	3541545	4659720	4980195	2395660
1073	498	-245	505	-952	83	2426693	715089	458091	625096	1043607	1373106	1467543	705946
692	319	-157	323	-614	53	1554561	458091	293459	400440	668544	879624	940121	452232
938	434	-213	441	-833	73	2121295	625096	400440	546431	912270	1200303	1282855	617105
1569	726	-354	740	-1392	121	3541545	1043607	668544	912270	1523063	2003930	2141750	1030263
2069	955	-469	971	-1837	158	4659720	1373106	879624	1200303	2003930	2636632	2817966	1355547
2204	1021	-502	1036	-1955	170	4980195	1467543	940121	1282855	2141750	2817966	3011776	1448778
1051	491	-240	497	-932	83	2395660	705946	452232	617105	1030263	1355547	1448778	696926

MATRIZ DISTANCIA DE MAHALANOBIS

0	18.8328908	80.3534264
18.8328908	0	32.8589835
80.3534264	32.85898349	0

Mapa uno.



5.2.6.2.2 Cálculo de la Distancia de Mahalanobis para el Análisis con Componentes Principales.

Veamos que resulta de llevar a cabo el proceso anterior aplicado a los datos que obtuvimos producto de la combinación de las técnicas del análisis de conglomerados y de componentes principales.

Nuevamente, al igual que con los datos originales, con estos datos resultado del análisis de componentes principales, obtenemos 3 grupos. Los 3 grupos a los que nos referimos son los siguientes:

Grupo 1 (individuos) 1, 61, 14, 26, 53, 91, 3, 23, 5, 15, 86, 90, 67, 87, 97, 95, 21, 30, 44, 4, 44, 72, 77, 12, 84, 13, 78, 58, 60, 89, 32, 39, 47 y 62.

Grupo 2 (individuos) 2, 71, 54, 85, 6, 94, 52, 82, 88, 48, 68, 70, 73, 75, 19, 99, 40, 50, 66, 51, 59, 55, 76, 92, 22, 36, 41, 25, 81, 93, 8, 63, 37, 43, 64, 34, 7, 38, 9, 16, 100, 20, 57, 96, 83, 35, 65, 18, 46, 98, 45, 49, 33, 27, 29, 69, 79, 31, 42, 10, 28 y 56.

Grupo 3. (individuos) 17, 24 y 74.

Otra vez, obtenemos las matrices covarianza correspondientes a cada grupo. Tenemos 3 matrices covarianza las utilizamos para obtener una sola combinándolas. A esta matriz le calculamos su inversa. Multiplicamos y postmultiplicamos por los vectores diferencia a la matriz inversa.

MATRIZ DE COVARIANZA 1

1.2370	0.7344	0.2209	0.0395	-0.1999	0.0818	-0.0040	-0.1927
0.7344	3.2938	1.3283	0.2472	-0.7219	0.7334	-0.0582	-0.1286
0.2209	1.3283	1.7443	0.1933	-0.3446	0.5110	0.4072	0.1722
0.0395	0.2472	0.1933	0.3347	-0.0332	0.0874	0.3439	-0.1051
-0.1999	-0.7219	-0.3446	-0.0332	0.3331	-0.1544	-0.1275	0.0269
0.0818	0.7334	0.5110	0.0874	-0.1544	0.6058	0.0838	0.0969
-0.0040	-0.0582	0.4072	0.3439	-0.1275	0.0838	0.8016	-0.0680
-0.1927	-0.1286	0.1722	-0.1051	0.0269	0.0969	-0.0680	0.2796

MATRIZ DE COVARIANZA 2

0.8233	0.1380	-0.2972	0.0040	-0.1102	-0.1175	0.0793	0.4580
0.1380	2.4506	-0.7131	0.3961	0.0360	-0.3155	-0.1652	-0.0563
-0.2972	-0.7131	1.1463	0.0569	0.0647	-0.1641	0.3510	-0.1788
0.0040	0.3961	0.0569	1.2558	0.6849	-0.1229	-0.1188	0.1191
-0.1102	0.0360	0.0647	0.6849	1.1027	0.0801	-0.0444	-0.2235
-0.1175	-0.3155	-0.1641	-0.1229	0.0801	0.9752	-0.1550	-0.1258
0.0793	-0.1652	0.3510	-0.1188	-0.0444	-0.1550	0.2680	0.1471
0.4580	-0.0563	-0.1788	0.1191	-0.2235	-0.1258	0.1471	0.9528

MATRIZ DE COVARIANZA 3

0.031	0.009	-0.034	0.044	0.044	0.078	0.109	-0.001
0.009	-0.127	-0.972	0.398	-0.255	0.186	0.859	-0.257
-0.034	-0.972	-5.178	2.613	-2.001	0.960	4.293	-1.526
0.044	0.398	2.613	-1.067	0.857	-0.361	-2.128	0.723
0.044	-0.255	-2.001	0.857	-0.490	0.447	1.834	-0.525
0.078	0.186	0.960	-0.361	0.447	0.003	-0.620	0.289
0.109	0.859	4.293	-2.128	1.834	-0.620	-3.347	1.296
-0.001	-0.257	-1.526	0.723	-0.525	0.289	1.296	-0.434

MATRIZ DE COVARIANZA COMBINADA

0.9692	0.3401	-0.113	0.0161	-0.139	-0.047	0.049	0.2248
0.3408	2.6896	0.0028	0.3376	-0.224	0.0532	-0.123	-0.08
-0.113	0.0009	1.3077	0.1109	-0.083	0.0753	0.3809	-0.06
0.0154	0.3164	-0.005	0.9674	0.3821	-0.028	0.1322	0.0078
-0.138	-0.217	-0.023	0.4016	0.833	-0.01	-0.116	-0.118
-0.046	0.0463	0.0297	-0.03	-0.012	0.8372	-0.031	-0.058
0.0506	-0.121	0.383	0.0352	-0.063	-0.07	0.4329	0.0762
0.2271	-0.062	0.035	-0.005	-0.095	-0.06	0.0004	0.7285

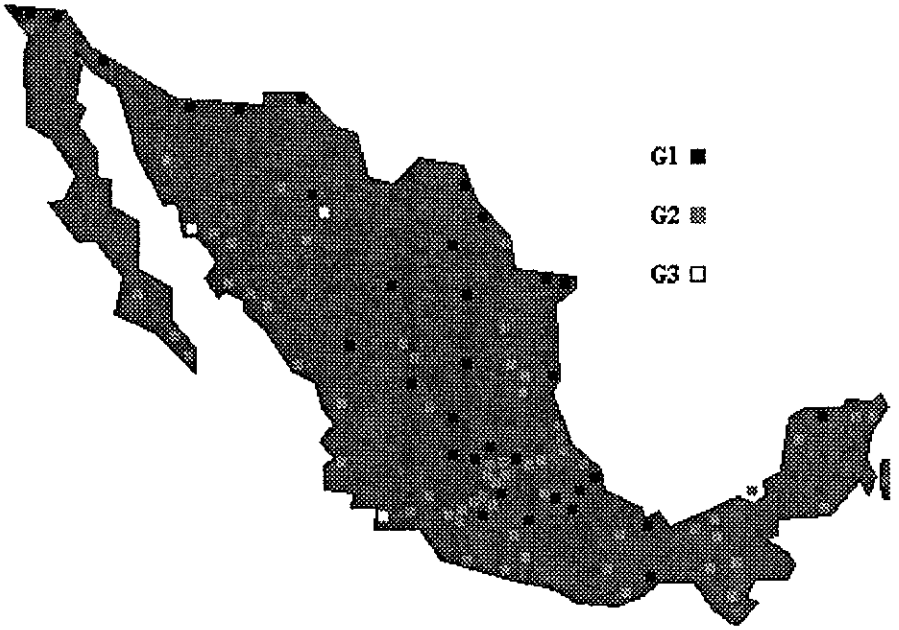
INVERSA DE LA MATRIZ COVARIANZA COMBINADA

1.226	-0.162	0.2113	-0.034	0.1365	0.0103	-0.322	-0.322
-0.186	0.4663	-0.14	-0.259	0.2421	0.0085	0.4208	0.0956
0.1573	-0.026	1.0338	-0.14	0.1337	-0.148	-0.867	0.1355
0.0683	-0.292	0.3001	1.4649	-0.795	-0.054	-1.019	-0.07
0.034	0.2814	-0.285	-0.818	1.7303	0.1026	1.049	0.1837
0.0375	-0.024	-0.061	0.0455	0.0076	1.2196	0.1192	0.0655
-0.259	0.2116	-1.025	-0.158	0.2163	0.327	3.4311	-0.277
-0.397	0.1241	-0.167	-0.097	0.1915	0.117	0.3153	1.5035

MATRIZ DISTANCIA DE MAHALANOBIS

0	15.4798968	54.2992956
15.4798968	0	58.6320418
54.2992956	58.6320418	0

mapa dos.



Utilizando la información que nos brindan los mapas uno y dos podemos decir que ambos análisis coinciden bastante.

Comparando los resultados de los 2 analisis realizados, encontramos que los siguientes elementos no coinciden en su clasificación :78, 89, 60, 26, 87, 97, 95, 52, 28, 56, 24 y 74.

De este tipo de comparación también se desprende que el grupo 3 es totalmente distinto para cada uno de los análisis. Por otra parte, en el primer análisis quedan sin clasificar los siguientes elementos: 17, 52 y 80. En el segundo análisis queda sin clasificar sólo el elemento: 80.

Ahora, sólo resta responder una pregunta: ¿A qué se deben las diferencias? Principalmente, a que los datos a los que se les aplicó el análisis de componentes principales NO recogen el 100% de la información de los datos originales. Recordemos que aportan un poco más del 90% de la varianza. Quizás la información que tenían componentes restantes aportan ciertas evidencias para asignar esos individuos en distintos conglomerados solución.

A los casos 17, 52 y 80 los vamos a catalogar como outliers. Entonces los dejamos sin clasificar.

Finalmente, gracias a esta etapa de validación, a las coincidencias que encontramos en los dos distintos análisis y a la interpretación que pudimos dar de los resultados, podemos decir que los hallazgos que hemos hecho, la estructura que hemos encontrado, pueden ser útiles y confiables.

Grupo 1.(individuos) 1, 61, 44, 30, 3, 23, 5, 15, 14, 86, 53, 91, 90, 21, 67, 4, 11, 72, 77, 13, 12, 84, 58, 62, 32, 39 y 47.

Grupo 2.(individuos) 2, 71, 24, 74, 6, 78, 22, 55, 76, 99, 19, 92, 36, 41, 25, 93, 54, 85, 88, 89, 27, 29, 40, 50, 66, 49, 51, 59, 60, 10, 69, 79, 18, 45, 46, 98, 33, 42, 31, 7, 38, 9, 20, 57, 96, 16, 83, 100, 35, 65, 82, 94, 26, 73, 48, 75, 81, 68, 70, 87, 97, 95, 34, 8, 63, 43, 37 y 64.

Grupo 3 (individuos) 28 y 56.

Outliers (individuos) 17,52 y 80.

CONCLUSIONES.

Primero quisiera hacer enfásis en la sencillez de la idea generadora que motivó el desarrollo de esta técnica. Sin embargo, como suele suceder con problemas que intentan modelarse matemáticamente, al querer generalizar esta idea o crear u algoritmo que permita llegar a la automatización nos encontramos con la necesidad de echar mano de conceptos un poco complicados. En nuestro caso tuvimos que echar mano de los espacios vectoriales y de ciertas propiedades que le son propias.

Si repasamos los primeros 2 capítulos del presente trabajo, aquellos que tratan principalmente de los fundamentos y de su relación con la técnica, nos daremos cuenta que estas herramientas se habían desarrollado hace ya varios años. ¿Porqué la aplicación de estas técnicas es tan reciente? La respuesta, como era de esperarse, es que gracias al desarrollo de los equipos de cómputo las Técnicas del Análisis Estadístico Multivariado han podido ser aplicadas a problemas de la vida real. Sin la ayuda de la computadora, en muchos casos (por no decir en todos), la aplicación de las técnicas sería muy tardada y en otros tantos inaccesible.

Nosotros, como pudieron observar a lo largo del presente, hicimos uso del paquete STATISTICA. Considero que las gráficas y los elementos que nos proporciona son de extrema utilidad para llevar a cabo un análisis de esta naturaleza Sin embargo, como todo, tiene algunas limitaciones, por ejemplo: no incluye la Distancia de Mahalanobis; y otra es el hecho de que si utilizaramos como entrada una matriz distancia el máximo de individuos que podemos manejar sería 56.

Por otra parte, mi experiencia a lo largo del desarrollo del presente me lleva a formular las siguientes sugerencias y también a hacer algunas recomendaciones:

1- Consideré muy importante el análisis de los dendogramas para tener una idea de la utilidad que pueden tener los resultados producto de una combinación de variables, función distancia y método de conglomerado.

2- Me parece conveniente y recomendable aplicar a un mismo conjunto de datos más de una técnica de conglomerado y comparar cuál nos brinda los resultados más útiles.

3- Por último, yo encontré muy útil el uso del criterio de Mojena para encontrar el número de conglomerados solución (naturales) y por supuesto los elementos que los formaban. A este respecto, recomiendo que se utilice la constante $k = 1.25$, ya que su aplicación nos llevó a resultados más útiles.

Finalmente, considero que aún falta mucho que desarrollar en esta técnica y que el presente no es un estudio exhaustivo, sino más que nada introductorio. Algunos temas

que no abordamos y que serían interesantes en futuros trabajos relativos a esta Técnica serían los siguientes:

1. Vectores de características mixtos. Para abordar este problema habría que proponer medidas de similitud adecuadas.

2. Incluir como una técnica previa al Análisis Estadístico de Conglomerados, o como parte del proceso de validación al Análisis Multidimensional de Escalas.

BIBLIOGRAFÍA.

Anton H.(1996), Introducción al Álgebra Lineal. Limusa. México, D.F.

Chatfield C. & Collins A. (1980), Introduction to Multivariate Analysis. Chapman and Hall. London.

Everitt B. S. (1993), Cluster Analysis. Halsted. New York.

Gnanadesikan R. (1977), Methods for Statistical Data Analysis of Multivariate Observations. John Wiley. New York.

Green E. P. (1976), Mathematical Tools for Applied Multivariate Analysis. Academic Press. New York.

Hair J. F., Anderson R. E. & Tatham R. E. (1987), Multivariate Data Analysis. With Readings. MacMillan. New York.

Hartigan J. A. (1975), Clustering Algoriths. John Wiley. New York.

Hasser N. B., LaSalle J. P. & Sullivan J. A., Análisis Matemático II. Trillas. México, D. F.

Manly B. F. (1986), Multivariate Statistical Methods. A primer. Chapman & Hall. London.

Purcell E. J. & Varberg D. (1987), Cálculo con Geometría Analítica. Prentice Hall. México, D. F.

Seber G. A. F. (1984), Multivariate Observations. John Wiley. New York.

Solar E. & Spenziale L.(1991), Apuntes de Álgebra Lineal. Limusa. Limusa. México, D.F.

Zupan J. (1982), Clustering of Large Data Sets. Research Studies. Chichester, England.

TABLA 1

IND.	X1	X2	X3	X4	X5
1	156	245	31.6	18.5	20.5
2	154	240	30.4	17.9	19.6
3	153	240	31	18.4	20.6
4	153	236	30.9	17.7	20.2
5	155	243	31.5	18.6	20.3
6	163	247	32	19	20.9
7	157	238	30.9	18.4	20.2
8	155	239	32.8	18.6	21.2
9	164	248	32.7	19.1	21.1
10	158	238	31	18.8	22
11	158	240	31.3	18.6	22
12	160	244	31.1	18.6	20.5
13	161	246	32.3	19 3	21.8
14	157	245	32	19.1	20
15	157	235	31.5	18.1	19.8
16	156	237	30.9	18	20.3
17	158	244	31.4	18.5	21.6
18	153	238	30.5	18.2	20.9
19	155	236	30.3	18.5	20.1
20	163	246	32.5	18.6	21.9
21	159	236	31.5	18	21.5
22	155	240	31.4	18	20.7
23	156	240	31.5	18.2	20.6
24	160	242	32.6	18.8	21.7
25	152	232	30.3	17.2	19.8
26	160	250	31.7	18.8	22.5
27	155	237	31	18.5	20
28	157	245	32.2	19.5	21.4
29	165	245	33.1	19.8	22.7
30	153	231	30.1	17.3	19.8
31	162	239	30 3	18	23.1
32	162	243	31.6	18.8	21.3
33	159	245	31.8	18.5	21.7
34	159	247	30.9	18.1	19
35	155	243	30.9	18.5	21.3
36	162	252	31.9	19.1	22.2
37	152	230	30.4	17.3	18.6
38	159	242	30.8	18.2	20.5
39	155	238	31.2	17.9	19.3
40	163	249	33.4	19.5	22.8
41	163	242	31	18.1	20.7
42	156	237	31.7	18.2	20.3
43	159	238	31.5	18.4	20.3
44	161	245	32.1	19.1	20.8
45	155	235	30.7	17.7	19.6
46	162	247	31.9	19.1	20.4
47	153	237	30.6	18.6	20.4
48	162	245	32.5	18.5	21.1
49	164	248	32.3	18.8	20.9

TABLA 2

IND	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
1	547366	95	36.6	54 5	100569	5 4	42 79	0 11	4 02	0 33	0 04	26.14	18 65	7.92
2	259979	93.5	22 3	54 1	58460	4.3	26 65	0 5	3 63	0	12 69	26.05	27 02	3 46
3	601938	95 1	26 1	52 1	131515	4 5	32 49	0.42	3 77	6.94	0 96	26 27	21 29	5 88
4	51557	94.1	49.4	38 2	11414	4 4	65 54	0	0 79	0	0	17 98	14 56	1 13
5	747381	95 6	38	56 5	161338	4 5	39 57	0 03	3 63	0	0.02	27 29	27 09	2 35
6	74346	91 4	15 2	47 2	15151	4 8	20 07	0	2 07	0	8 47	32 69	30.78	5 93
7	160970	95 7	19 9	66.3	34198	4 6	12 15	5 42	5.35	2.22	4.06	34.44	22.93	13 43
8	43920	94 2	20 3	62 2	9427	4 4	5 27	0	1.68	0	1.95	21.19	64 35	5 55
9	173645	90 7	21 7	60 2	37405	4 5	17	0 97	9 24	2 88	5.06	34.89	24.13	6 04
10	136034	87 6	21 4	41	27997	4 7	7 34	34.73	4 3	0	13 58	19 18	19 85	1 03
11	56336	95	51.1	39.8	11984	4 6	68 16	0	0.53	0	0.17	12.24	18 9	0
12	240056	96.2	48 7	45 9	51316	4 7	58 5	0.9	4 54	8 33	0	16 1	9 72	1 9
13	98185	95 5	44 7	48.7	21554	4 5	48.86	0 05	1.63	0	0	28.56	20.32	0 56
14	469166	94 7	41.4	50 3	94766	4 9	49 67	0 44	4.08	0	0	23.46	17 74	4 58
15	791891	95.6	33 2	53.7	158231	5	39 81	0.46	4.6	1.77	0 03	27 57	22.68	3 09
16	154347	94 2	21 8	65.9	32987	4 6	11.97	0 65	11 25	0	0.55	34 75	31 92	8 92
17	92863	91 6	22 1	60.2	19835	4.6	2 48	9 65	0 08	35.06	3 55	16 27	29 99	2 91
18	89335	74.8	23.6	59 2	16580	5 3	15 82	0	1 03	0	0	49 45	29.44	4 26
19	222405	83 3	15.5	54 2	44217	4 9	12 21	0 23	3.99	0	3.55	43.09	34.18	2 75
20	295608	89.1	19.1	72 5	61825	4 7	10 75	0.17	11 26	0	0	39 12	31 01	7 69
21	798499	96.3	49.4	44 9	170585	4.5	62 86	0 02	2.95	0	0	15 95	16 28	1.93
22	112589	96 2	25.3	47	24756	4.5	11.16	0	0 18	0	0	46 98	39.01	2 67
23	530783	97 2	36.7	65.1	118870	4.4	43.77	0 39	6 72	0	0	24 25	19.62	5 24
24	104014	94 9	30.7	52 3	22576	4.8	23 79	0 05	3 24	12.98	0 03	26.6	29 56	3 75
25	90647	95 7	32	58 6	18950	4.7	25 6	11 86	5 26	0	0	34 2	21 51	1 58
26	413635	85.8	28 8	60 3	82600	5	34 37	0 01	3 87	0	0.06	33 33	22 28	6 11
27	310569	88	26 3	55 4	56465	5.4	36 7	0	2 54	0	0	31.26	24 4	5 1
28	119170	88.4	31.8	53 4	20390	5.5	4 93	20 45	9 39	0	0.18	17.45	22 77	24 83
29	362915	86 9	31 6	51 3	63851	5.6	35 04	0 41	4 13	0	0	35 46	23 71	1 23
30	867920	88 7	50 3	43 8	145100	5 9	48 29	0 04	3 32	0	0 02	25 39	18 86	6 08
31	94901	87	40 2	36 9	16921	5 3	21 91	0	0	0	0	62 41	15.4	0.29
32	204311	84.9	39 6	38 4	37700	5 4	54 25	0 03	1.96	10.48	0	18 49	13 06	1 73
33	110692	73 9	33 4	36 1	18384	5 8	26.14	0	0 7	0	2 99	32 97	36 13	1.07
34	593212	87 4	18	70 1	122622	4 8	7 52	0 05	2 08	0	1 99	36 66	47 84	3.86
35	165107	82 8	18 8	58.1	31737	5 1	9 74	0 23	9 65	0	0	40 21	26 19	14
36	101067	86 2	23	62.4	20922	4 8	24 6	0 59	0.15	0	0 16	47 3	25 49	1.7
37	63366	84 6	14 5	65 4	13581	4.6	2 01	1 39	0 77	0	1.78	23 56	67.96	2 53
38	201450	94 2	26 5	68 5	42511	4 7	16 08	10 96	7.08	0	0	35.85	22 72	7 32
39	124912	91.2	50 4	30 6	24619	5	59 16	1 27	0 69	15.72	0 08	11 77	8 33	2.98
40	92570	87 8	29 7	57 7	18031	5 1	25 97	1 01	0.17	0	0	42 76	27 39	2 69
41	74068	90 1	28	59.5	14864	5	10 6	0	0	0	0	49 36	38 46	1 58
42	106157	84.1	34 2	33 2	18499	5.7	39.75	0 11	0.82	0	0 25	37 04	21 38	0.65
43	111457	94	15 6	73 9	22614	4.7	2.95	0	0 3	0	0 3	31 25	63.12	2 08
44	819915	90.6	36 9	53 7	154319	5.3	50 57	0 04	1 52	0	0 04	25.2	14 56	8 06
45	36135	83.3	30.8	41.6	6926	5 2	14.73	0	0.44	0	0 22	45 73	38.87	0
46	100926	79.9	17 9	46 7	18927	5.2	16.12	0 05	0	0	0	49 13	31.13	3 58
47	134969	87	41.5	40 3	26989	4.8	53.43	3 57	0 88	12 9	2 37	13 25	11 86	1 73
48	492901	91 6	25 9	63 7	94133	5	20 62	0 03	6.87	0	0 03	35 86	26 9	9 89
49	66736	84.2	26 6	48 4	11960	5 6	24 23	0	0 33	0	4 26	35 4	28 41	7 37
50	217068	88	27 7	55 9	41192	5 2	22.53	0.18	2 15	0	0 02	41.16	29 82	4 15
51	185445	85	22 6	44.7	35025	5 2	34 26	0 05	3.37	0	0	36 56	23 55	2 22

TABLA 2

IND	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
52	180573	88 5	23.7	58 2	37952	4 7	13 53	0 42	0,01	0	0 71	39 25	31 9	14 18
53	511779	91 2	33 7	58.1	109657	4 6	35.73	1.11	3 77	2.4	0.05	28 61	23 48	4 87
54	241463	93 6	25,5	60.5	49730	4 7	22 74	0.44	3 38	1.71	0 11	34 77	29 42	7 43
55	61569	93.9	25.6	44.8	12620	4.8	20 68	0 95	1 69	0	0.67	46 19	29 14	0 67
56	12645	75 2	14	46	2287	5	10 16	0	0	0	14 75	18 69	22 62	33 77
57	322317	91.3	23	70 7	64501	4 9	16 82	0 11	7 7	0	0	41 35	24 54	9 48
58	65707	90 2	38 6	50.9	13752	4 7	46 38	2 59	7 63	0	8.49	18 52	14 5	1 92
59	110136	84 2	24.7	41.4	22197	4 9	40 71	0	2 48	0	0	32 04	22 89	2 08
60	155563	86.5	40.4	44 8	29816	5 2	48 29	0 11	2 58	0	0	27 32	20 15	1 56
61	555491	90 1	37 8	53.1	104760	5 2	46 52	0 24	4,3	0	0,03	24 61	19 57	4 73
62	126555	85 5	47.1	36 9	23504	5 3	70 28	0 06	0.2	0	0 13	16 8	10 97	1 55
63	187431	93 2	17.1	73 3	41213	4 3	5 04	0 48	4 21	1 08	1 79	24 19	58 53	4.68
64	44903	87 7	15 4	69.5	9642	4.6	2 76	0	1.14	0	5 56	31 12	49 91	9 49
65	172563	86 9	15 2	51 6	35587	4 8	16 75	0 61	7 73	0	2 03	39 77	22 39	10 73
66	130939	87 9	24.4	51 8	26827	4 9	28 67	2 29	1 26	0	0 48	35,31	24 79	7 2
67	658712	94 1	35,3	58,3	128750	5 1	36 7	0 2	15 78	2,75	0	23 44	17 95	3 18
68	601123	91.1	19.8	51 8	113544	5 2	14 45	0 07	6 12	0	2 16	39	27 49	10 73
69	258130	88 9	14	30.1	46297	5 4	15 18	0	0.87	0	14 36	32 48	29 85	7.26
70	303558	93 1	18 8	48 6	58729	5 1	20.84	0 02	7 29	0	9 91	35 32	23 76	3 05
71	314345	94 7	21.1	63 2	66772	4 6	14 73	0.53	3 54	4 77	7.75	26 97	39 38	2 32
72	39120	96.3	48 2	39 8	8410	4.5	76 53	0 87	0	0	0	13 86	8 02	0 72
73	311443	94 8	22 5	57.5	64473	4.8	24 9	0 05	5 6	0	0 26	42 59	24 21	2 39
74	175109	93 8	21.8	49.3	38434	4 8	18 68	0 33	1 47	15 24	15 81	23 42	22 49	2 55
75	448966	96.1	26.5	60	94093	4 7	21 28	0 5	6 51	0	0 72	38 64	25 19	7 17
76	122061	92 1	22.3	50	23861	5.1	24 1	0 75	2 12	0	0	49 8	21 35	1 87
77	107938	97 2	46.8	49 3	22498	4.5	73.3	0	1 24	0	0	13 37	10 96	1 13
78	110530	94.5	24 7	45	23031	4 6	32 31	0	1 33	0	2 52	31 99	31 1	0 74
79	172635	87.7	20.1	35 5	32047	5 4	27 91	0	2 14	0	8 59	38 23	21 48	1 65
80	70053	84 6	11.8	30.4	12802	5 5	4 7	0	0	0	70 84	18 12	7 78	0 076
81	386776	92.2	23 6	59	80336	4.8	14 96	17 32	6 35	1 28	0,41	31,46	20 71	7.5
82	116174	89.8	20.1	52 6	25255	4.8	23 14	0 99	3 95	0	0.17	34 52	22 39	14 85
83	207923	95 2	22	69.4	44525	4 8	10 17	0,62	14,81	0	0	33 09	27 14	14 17
84	303293	94 2	46.8	43 2	66902	4 5	64 27	0,12	2 38	0	1 32	17 16	13 56	1 2
85	219468	95.1	33 3	60.3	45241	4.8	36 2	0	1 13	0	0	25 37	30 43	6 88
86	376676	93.4	38.8	48 6	81891	4 6	45 51	7 46	3 25	0	0	24 78	17 85	1 15
87	648598	93 9	29 6	55 9	144954	4 5	23.69	0 8	9 17	3 66	4 27	29 89	26 34	2 18
88	51744	94.6	29	60 9	10648	4 8	21 38	0	0 93	0	0	42 19	24 89	10 61
89	153729	93 5	32.3	53 5	29366	5 2	45 22	0 25	1 47	5 09	0	25 14	16 99	5 84
90	514074	89.4	40 1	45.5	110164	4 6	41 21	14 57	3 14	0	1 75	21,92	15 9	1 5
91	513914	89	29 6	52 1	107551	4 7	38 35	0 65	2 49	0	0 15	31 23	21 58	5 56
92	103089	86 1	17 4	40.9	21684	4 7	21 25	0	3.08	0	1 6	46 79	26 9	0 38
93	263264	88 9	30 6	51 1	54517	4 7	18 8	21,99	2,03	0	0 27	32 3	22 37	2 25
94	143187	88	20 9	43 2	29468	4.8	13 26	0	2 77	0	7 77	32 15	29 76	14 3
95	522198	93.2	25 3	65 9	121345	4 3	21 88	2 88	5 1	0	2 95	28 78	26 37	12 05
96	310564	92.3	21 2	71.3	69896	4.4	11 87	0	7 47	0	0	35 09	31 65	14 13
97	664882	93	28 1	64 8	143606	4 6	23 18	0 91	11 23	3 2	0 73	31 37	23 4	5 99
98	64618	75 7	24 6	39 7	12434	5 2	17 22	1 42	0	0	0	46 54	31 83	2 99
99	160181	91 5	19 1	40 1	29952	5 3	14 39	9 21	0 65	0	0 62	46 31	27 18	1 63
100	191326	94 4	24.1	63 5	38637	5.2	8 68	5 92	11 29	2,53	0 02	35 24	23 74	12 58

TABLA 3

IND	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
1	1.38	0.91	0 80	0.19	1 13	1 43	0 82	-0 36	0 13	-0 26	-0 33	-0 51	-0.64	0.51
2	0 02	0.61	-0 63	0 15	0.15	-1.60	-0.07	-0 29	0 02	-0.33	1 30	-0.52	0.12	-0 34
3	1 64	0.93	-0.25	-0 04	1.85	-1 05	0 25	-0.31	0.06	1 16	-0 21	-0 30	-0 40	0 12
4	-0 97	0 73	2.08	-1 37	-0.94	-1 32	2 08	-0.39	-0 82	-0 33	-0.33	-1.31	-1 02	-0 78
5	2 33	1.03	0.94	0 38	2.54	-1 05	0 64	-0.38	0.02	-0 33	-0.33	-0 40	0 13	-0.55
6	-0 86	0.20	-1 34	-0 51	-0.86	-0 22	-0 44	-0.39	-0 44	-0 33	0.76	0 13	0.47	0.13
7	-0 45	1.05	-0 87	1 32	-0.41	-0 77	-0 88	0.63	0 52	0 15	0.19	0 30	-0 25	1 55
8	-1 00	0.75	-0 83	0 92	-0.99	-1 32	-1 26	-0.39	-0.56	-0 33	-0.08	-1 00	3.54	0.06
9	-0 39	0 06	-0.69	0.73	-0 34	-1 05	-0.61	-0.20	1 66	0.29	0.32	0 33	-0 14	0 15
10	-0 57	-0.54	-0.72	-1.10	-0 56	-0 50	-1.14	6.13	0 21	-0.33	1.41	-1 19	-0.53	-0 80
11	-0 95	0.91	2.25	-1 22	-0 93	-0.77	2.23	-0 39	-0 89	-0.33	-0 31	-1.87	-0.62	-0.99
12	-0 08	1.15	2.01	-0.63	-0 02	-0.50	1.69	-0 22	0 28	1.46	-0.33	-1 49	-1 46	-0.63
13	-0 75	1.01	1.61	-0.37	-0 71	-1.05	1.16	-0.38	-0 57	-0.33	-0.33	-0.27	-0 49	-0 89
14	1.01	0.85	1.28	-0 21	0.99	0.05	1 20	-0 30	0.15	-0.33	-0 33	-0 77	-0.73	-0.13
15	2 54	1.03	0 46	0 11	2.47	0 33	0 66	-0 30	0 30	0 05	-0 33	-0 37	-0 27	-0.41
16	-0 48	0.75	-0.68	1.28	-0 44	-0.77	-0 89	-0 26	2 25	-0.33	-0.26	0 33	0 57	0 70
17	-0 77	0.24	-0.65	0 73	-0 75	-0 77	-1 41	1 43	-1 03	7 19	0 12	-1.48	0.39	-0 44
18	-0.79	-3 06	-0 50	0.64	-0.82	1 15	-0 67	-0 39	-0.75	-0 33	-0 33	1 77	0.34	-0.19
19	-0.16	-1.39	-1 31	0 16	-0.18	0.05	-0 87	-0 34	0.12	-0 33	0 12	1 15	0 78	-0 47
20	0 19	-0.25	-0.95	1 91	0.23	-0.50	-0 95	-0 35	2 25	-0 33	-0 33	0 76	0 49	0 46
21	2 57	1.17	2 08	-0 73	2.75	-1.05	1 93	-0.38	-0.18	-0 33	-0.33	-1 51	-0 86	-0.63
22	-0.68	1.15	-0.33	-0 53	-0.63	-1.05	-0.93	-0.39	-1.00	-0.33	-0 33	1 53	1 22	-0.49
23	1.30	1.34	0 81	0.25	1.55	-1 32	0 88	-0.31	0 92	-0 33	-0.33	-0 70	-0.55	0.00
24	-0.72	0.89	0.21	-0.02	-0 68	-0.77	-0 23	-0.38	-0 10	2 46	-0.33	-0.47	0.36	-0 26
25	-0.78	1.05	0.34	0.58	-0 77	-0.50	-0.13	1.84	0 49	-0 33	-0.33	0.28	-0 38	-0 69
26	0 75	1.07	0.02	0.74	0.71	0.33	0.36	-0 38	0 09	-0.33	-0.33	0 19	-0.31	0.16
27	0.26	-0.47	-0 03	0.27	0.10	1.43	0.48	-0 39	-0 30	-0.33	-0.33	-0 01	-0 12	-0 03
28	-0.65	-0.39	0 32	0.08	-0.73	1.70	-1.27	3.45	1.71	-0.33	-0 31	-1.36	-0 27	3.72
29	0.51	-0 68	0 30	-0.12	0 28	1.98	0.39	-0 31	0.16	-0.33	-0.33	0.40	-0 18	-0 76
30	2 90	-0 33	2 17	-0.83	2 16	2 80	1 13	-0.38	-0.08	-0.33	-0 33	-0.58	-0.81	0 16
31	-0.76	-0 66	1 16	-1.49	-0 81	1 15	-0 33	-0 39	-1.05	-0 33	-0 33	3.04	-0 94	-0 94
32	-0 24	-1 08	1.10	-1.35	-0 33	1 43	1 46	-0.38	-0.47	1 92	-0 33	-1.26	-1 15	-0 67
33	-0 69	-3 24	0 48	-1 57	-0 78	2 53	-0 10	-0.39	-0.84	-0 33	0.05	0 16	0 96	-0 79
34	1 60	-0.58	-1.06	1 68	1 64	-0.22	-1 13	-0.38	-0.44	-0 33	-0.08	0.52	2 03	-0 26
35	-0 43	-1.49	-0.98	0.53	-0 47	0 60	-1 01	-0.34	1 78	-0 33	-0.33	0 87	0 05	1 66
36	-0 73	-0 82	-0 56	0.94	-0 72	-0 22	-0.19	-0.27	-1 01	-0 33	-0.31	1 56	-0.02	-0 67
37	-0.91	-1.13	-1.41	1.23	-0 89	-0.77	-1.44	-0 12	-0 82	-0 33	-0.10	-0.76	3.87	-0 51
38	-0 26	0 75	-0 21	1.53	-0.22	-0 50	-0.66	1.67	1.02	-0.33	-0.33	0 44	-0 27	0.39
39	-0.62	0 16	2 18	-2 09	-0.64	0 33	1 73	-0 15	-0.85	3.04	-0 32	-1 92	-1 59	-0.43
40	-0 77	-0 51	0 11	0.49	-0 79	0 60	-0 11	-0 20	-1.00	-0 33	-0 33	1.12	0.16	-0.48
41	-0 86	-0 05	-0.06	0.67	-0 86	0 33	-0 96	-0 39	-1.05	-0 33	-0.33	1.76	1 17	-0 69
42	-0.71	-1.23	0.56	-1 85	-0 78	2.25	0.65	-0.36	-0.81	-0 33	-0.30	0.56	-0 39	-0 87
43	-0.68	0 71	-1 30	2 04	-0.68	-0 50	-1 38	-0 39	-0.96	-0.33	-0 29	-0 01	3 43	-0 60
44	2.67	0.05	0.83	0.11	2 38	1 15	1.25	-0.38	-0 60	-0 33	-0 33	-0 60	-1 02	0.53
45	-1.04	-1 39	0 20	-1 04	-1.05	0.88	-0 73	-0 39	-0.92	-0 33	-0 31	1 41	1 21	-0.99
46	-0 73	-2.06	-1 07	-0.56	-0 77	0 88	-0.66	-0.38	-1.05	-0 33	-0 33	1.74	0.50	-0.32
47	-0 57	-0 66	1.29	-1.17	-0 58	-0.22	1.41	0.28	-0 79	2 44	-0.03	-1 77	-1 26	-0 67
48	1 12	0.24	-0 27	1.07	0.98	0 33	-0.41	-0.38	0 97	-0.33	-0.33	0.42	0.11	0 88
49	-0 90	-1.21	0 00	-0 39	-0 93	1.98	-0.21	-0 39	-0 95	-0 33	0.21	0.40	0 25	0 40
50	-0 18	-0.47	-0.09	0 32	-0 25	0 88	-0 30	-0 35	-0 42	-0 33	-0 33	0.96	0 38	-0 21
51	-0.33	-1 06	-0.60	-0 75	-0.39	0 88	0 35	-0 38	-0.06	-0 33	-0 33	0 51	-0 19	-0 57
52	-0.36	-0.37	-0.49	0.54	-0.33	-0.50	-0.80	-0 31	-1.05	-0 33	-0 24	0 77	0 57	1.70
53	1.21	0.16	0.51	0 53	1.34	-0.77	0.43	-0.18	0 06	0 19	-0.33	-0 27	-0 20	-0 07
54	-0 07	0.63	-0 31	0.76	-0.05	-0.50	-0.29	-0.30	-0.06	0.04	-0 32	0 33	0 34	0 41
55	-0.92	0 69	-0 30	-0.74	-0 91	-0 22	-0.40	-0.21	-0.55	-0 33	-0 25	1 45	0 32	-0 87
56	-1.15	-2 98	-1.46	-0.62	-1 15	0.33	-0.99	-0 39	-1 05	-0 33	1.57	-1.24	-0 28	5 41
57	0 31	0 18	-0 56	1.74	0 29	0 05	-0.62	-0 36	1 21	-0.33	-0.33	0 98	-0.10	0.80
58	-0 90	-0 03	1.00	-0.16	-0 89	-0 50	1.02	0.10	1 19	-0 33	0.76	-1.26	-1.02	-0 63
59	-0.69	-1.21	-0 39	-1 06	-0.69	0.05	0.71	-0.39	-0 32	-0.33	-0.33	0 07	-0 27	-0 60
60	-0.48	-0 76	1.18	-0.74	-0.51	0 88	1.13	-0 36	-0 29	-0.33	-0 33	-0 40	-0 51	-0 70
61	1.42	-0 05	0 92	0.05	1.23	0 88	1 03	-0 34	0 21	-0.33	-0 33	-0.66	-0 56	-0 10
62	-0.61	-0 96	1 85	-1.49	-0.66	1.15	2.34	-0 37	-0.99	-0.33	-0.32	-1.42	-1.35	-0 70
63	-0 32	0 56	-1.15	1.98	-0.25	-1 60	-1 27	-0.30	0.19	-0 10	-0 10	-0 70	3.01	-0 11
64	-1.00	-0.52	-1 32	1 62	-0.98	-0.77	-1.40	-0.39	-0 72	-0 33	0.38	-0 02	2 22	0 81
65	-0 40	-0.68	-1.34	-0.09	-0 38	-0.22	-0.62	-0.27	1.22	-0.33	-0.07	0 82	-0 30	1 04
66	-0 59	-0.49	-0.42	-0.07	-0.59	0.05	0.04	0 04	-0 68	-0.33	-0 27	0 39	-0 08	0 37
67	1.91	0.73	0.67	0 55	1.78	0.60	0 48	-0.35	3 58	0 26	-0.33	-0 78	-0.71	-0 39
68	1 63	0 14	-0.88	-0 07	1 43	0 88	-0.75	-0.37	0 75	-0 33	-0 06	0 75	0 17	1 04

TABLA 3

IND	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
69	0.01	-0 29	-1 46	-2 14	-0 13	1 43	-0 71	-0 39	-0 79	-0 33	1 52	0 11	0 38	0 39
70	0.22	0.54	-0 98	-0.37	0 16	0 60	-0 41	-0.38	1 09	-0 33	0 94	0 39	-0 18	-0.42
71	0.28	0 85	-0 75	1 02	0 34	-0 77	-0 73	-0 29	-0 01	0 69	0 66	-0 43	1 25	-0 55
72	-1.03	1 17	1 96	-1.22	-1 01	-1 05	2.69	-0.22	-1 05	-0.33	-0 33	-1 71	-1 62	-0.86
73	0 26	0.87	-0 61	0 48	0.29	-0.22	-0 17	-0 38	0.59	-0 33	-0.30	1 10	-0.13	-0 54
74	-0 38	0.67	-0.68	-0.31	-0 36	-0.22	-0 51	-0 32	-0.62	2 94	1.70	-0 78	-0.29	-0 51
75	0.91	1 13	-0.21	0 71	0.98	-0.50	-0 37	-0 29	0.86	-0 33	-0.24	0 71	-0.04	0 37
76	-0.63	0 34	-0.63	-0.24	-0.65	0.60	-0 21	-0 24	-0.43	-0 33	-0 33	1 80	-0.40	-0 64
77	-0.70	1 34	1.82	-0 31	-0.68	-1.05	2 51	-0.39	-0 69	-0.33	-0 33	-1 76	-1 35	-0 78
78	-0.69	0.81	-0.39	-0.72	-0.67	-0 77	0 24	-0.39	-0 66	-0.33	-0 01	0 06	0 50	-0.85
79	-0 39	-0.52	-0 85	-1 63	-0 46	1 43	0.00	-0.39	-0 42	-0 33	0 77	0 67	-0 38	-0 68
80	-0 88	-1.13	-1 68	-2 11	-0 91	1 70	-1.29	-0.39	-1 05	-0 33	8 76	-1 49	-1 64	-0.98
81	0 62	0.36	-0.50	0 62	0 66	-0 22	-0 72	2 87	0 81	-0 05	-0 28	0 01	-0 45	0.43
82	-0.66	-0 11	-0.85	0 01	-0.62	-0 77	-0 27	-0 20	0.11	-0.33	-0 31	0 31	-0 30	1.82
83	-0.23	0 95	-0.66	1.61	-0.17	-0.77	-0 98	-0 27	3.30	-0.33	-0.33	0 17	0 13	1 69
84	0.22	0 75	1.80	-0.89	0.35	-1.05	2 01	-0 36	-0.35	-0.33	-0.16	-1 39	-1 11	-0 77
85	-0 17	0.93	0.47	0 74	-0.16	-0 77	0 46	-0 39	-0.72	-0.33	-0 33	-0 59	0 44	0 31
86	0.57	0.60	1 02	-0.37	0 69	-0.77	0 97	1 01	-0 10	-0.33	-0 33	-0 64	-0 72	-0 78
87	1 86	0.69	0 10	0.32	2 16	-1 05	-0.24	-0 24	1 64	0 46	0 22	-0 14	0 06	-0 58
88	-0 97	0.83	0 04	0 80	-0 96	-0.22	-0 36	-0 39	-0 78	-0 33	-0 33	1 06	-0.07	1 02
89	-0 48	0 61	0 37	0.09	-0 53	0 88	0.96	-0 34	-0 62	0 76	-0 33	-0 61	-0 80	0 11
90	1 22	-0 19	1.15	-0 67	1.35	-0 77	0.73	2 35	-0 13	-0.33	-0 11	-0 92	-0 89	-0 71
91	1.22	-0.27	0.10	-0.04	1 29	-0 50	0 58	-0 26	-0 32	-0 33	-0 31	-0 01	-0 37	0 06
92	-0.72	-0 84	-1.12	-1.11	-0.70	-0.50	-0 37	-0 39	-0 15	-0.33	-0 13	1 51	0 11	-0 92
93	0 03	-0.29	0.20	-0.14	0.06	-0.50	-0 51	3 74	-0.45	-0.33	-0 30	0 09	-0.30	-0 57
94	-0.53	-0 47	-0.77	-0.89	-0.52	-0.22	-0.81	-0 39	-0 24	-0.33	0 67	0 08	0 37	1 72
95	1.26	0 56	-0.33	1.28	1.61	-1.60	-0.34	0 16	0.45	-0.33	0 05	-0 25	0 06	1 29
96	0 26	0 38	-0 74	1.79	0 41	-1.32	-0 90	-0 39	1 14	-0 33	-0 33	0 36	0 55	1 69
97	1.93	0.52	-0 05	1 17	2 13	-0 77	-0 26	-0 21	2.24	0 36	-0 24	0 00	-0 21	0 14
98	-0 91	-2.88	-0 40	-1 23	-0 92	0.88	-0 59	-0 12	-1 05	-0.33	-0 33	1 48	0 56	-0 43
99	-0.45	0.22	-0 95	-1 19	-0 51	1 15	-0 75	1 34	-0 86	-0 33	-0 25	1 46	0 14	-0 69
100	-0 31	0 79	-0.45	1.05	-0.36	0.88	-1.07	0 73	2 26	0.21	-0 33	0 38	-0 18	1 39

TABLA 4

Distancia de union	Obj No 1	Obj No 2	Obj No 3	Obj No 4	Obj No 5	Obj No 6	Obj No 7	Obj No 8	Obj No 9	Obj No 10	Obj No 11	Obj No 12	Obj No 13	Obj No 14	Obj No 15	Obj No 16	Obj No 17	Obj No 18	Obj No 19	Obj No 20	Obj No 21	Obj No 22	Obj No 23	Obj No 24	Obj No 25	Obj No 26	Obj No 27	Obj No 28	Obj No 29	Obj No 30	Obj No 31	Obj No 32	Obj No 33	Obj No 34
3.563198	C_45	C_46	C_88																														
3.750458	C_6	C_78																															
3.947809	C_55	C_76	C_90																														
3.968157	C_8	C_83	C_43																														
4.011094	C_19	C_92																															
4.063237	C_20	C_57	C_96																														
4.166864	C_59	C_79																															
4.229012	C_45	C_75	C_81																														
4.454648	C_14	C_86	C_53	C_91																													
4.672895	C_12	C_94																															
4.833643	C_2	C_71																															
4.872152	C_51	C_59	C_60																														
4.928876	C_1	C_61	C_44																														
4.971014	C_7	C_38	C_9																														
4.978143	C_16	C_69	C_100																														
5.198125	C_14	C_86	C_53	C_91	C_90																												
5.214497	C_3	C_23																															
5.253981	C_40	C_50	C_66	C_49																													
5.691899	C_25	C_93																															
5.813502	C_26	C_73	C_48	C_75	C_81																												
5.936016	C_37	C_94																															
5.978704	C_54	C_65	C_88																														
6.257276	C_3	C_23	C_5	C_15																													
6.718958	C_24	C_74																															
6.856208	C_33	C_42																															
6.963189	C_37	C_76	C_47																														
6.982993	C_8	C_52	C_43	C_37	C_64																												
6.994571	C_4	C_11	C_72	C_77	C_13																												
7.319838	C_6	C_78	C_22																														
7.537370	C_68	C_70																															
7.711878	C_97	C_97	C_95																														
7.890753	C_18	C_45	C_46	C_96																													
8.083372	C_19	C_92	C_36	C_41																													
8.195647	C_40	C_50	C_66	C_49	C_51	C_59	C_60																										
8.378973	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_59	C_60																								
8.472943	C_26	C_73	C_48	C_75	C_81	C_69	C_70																										
8.598165	C_8	C_78	C_22	C_55	C_76	C_99																											
8.898092	C_7	C_38	C_9	C_20	C_57	C_96																											
8.930274	C_4	C_11	C_73	C_77	C_13	C_12	C_84																										
9.094437	C_7	C_38	C_9	C_20	C_57	C_96	C_16	C_83	C_100																								
9.370983	C_54	C_65	C_88	C_89																													
9.553535	C_35	C_65	C_87	C_94																													
10.00985	C_1	C_61	C_44	C_30																													
10.15419	C_3	C_23	C_5	C_15	C_14	C_86	C_53	C_91	C_90																								
11.34719	C_6	C_78	C_22	C_55	C_76	C_99	C_19	C_92	C_36	C_41																							
11.79575	C_25	C_93	C_54	C_65	C_88	C_89																											
13.44560	C_4	C_11	C_72	C_77	C_13	C_12	C_84	C_56																									
14.07536	C_4	C_11	C_72	C_77	C_13	C_12	C_84	C_56	C_82																								
14.25269	C_3	C_23	C_24	C_74																													
15.51207	C_7	C_73	C_5	C_15	C_14	C_86	C_53	C_91	C_90	C_21																							
15.85503	C_26	C_73	C_48	C_75	C_81	C_68	C_70	C_67	C_97	C_95																							
16.09697	C_18	C_45	C_46	C_93	C_33	C_42																											
17.49455	C_29	C_93	C_54	C_65	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_59	C_60																		
19.04309	C_26	C_73	C_48	C_75	C_81	C_68	C_70	C_67	C_97	C_95	C_34																						
20.77995	C_10	C_69	C_79																														
21.15839	C_18	C_45	C_46	C_93	C_33	C_42	C_31																										
21.69370	C_6	C_78	C_22	C_55	C_76	C_99	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_65	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_59	C_60								
22.37155	C_7	C_38	C_9	C_20	C_57	C_96	C_16	C_83	C_100	C_35	C_65	C_92	C_84																				
24.02064	C_4	C_11	C_72	C_77	C_13	C_12	C_84	C_56	C_82	C_32	C_39	C_47																					
25.14945	C_3	C_23	C_5	C_15	C_14	C_86	C_53	C_91	C_90	C_21	C_67																						
26.50627	C_2	C_71	C_24	C_74	C_6	C_78	C_22	C_55	C_76	C_99	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_65	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_59	C_60	C_10	C_69	C_79	
27.84558	C_1	C_61	C_44	C_30	C_3	C_23	C_5	C_15	C_14	C_86	C_53	C_91	C_90	C_21	C_67																		
30.15358	C_28	C_56																															
30.67549	C_2	C_71	C_24	C_74	C_6	C_78	C_22	C_55	C_76	C_99	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_65	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_59	C_60	C_10	C_69	C_79	C_45
31.00993	C_7	C_38	C_9	C_20	C_57	C_96	C_16	C_83	C_100	C_35	C_65	C_92	C_84	C_26	C_73	C_48	C_75	C_81	C_68	C_70	C_97	C_29	C_40	C_50	C_66	C_49	C_51	C_59	C_60	C_10	C_69	C_79	C_16
42.21890	C_2	C_71	C_24	C_74	C_6	C_78	C_22	C_55	C_76	C_99	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_65	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_59	C_60	C_10	C_69	C_79	C_45
50.72868	C_7	C_38	C_9	C_20	C_57	C_96	C_16	C_83	C_100	C_35	C_65	C_92	C_84	C_26	C_73	C_48	C_75	C_81	C_68	C_70	C_97	C_97	C_95	C_34	C_8	C_63	C_43	C_97	C_64				
52.89701	C_1	C_61	C_44	C_30	C_3	C_23	C_5	C_15	C_14	C_86	C_53	C_91	C_90	C_21	C_67	C_4	C_11	C_72	C_77	C_13	C_12	C_64	C_58	C_62	C_32	C_39	C_47						
60.34329	C_2	C_71	C_24	C_74	C_6	C_78	C_22	C_55	C_76	C_99	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_65	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_59	C_60	C_10	C_69	C_79	C_45
83.98204	C_1	C_61	C_44	C_30	C_3	C_23	C_5	C_15	C_14	C_86	C_53	C_91	C_90	C_21	C_67	C_4	C_11	C_72	C_77	C_13	C_12	C_64	C_58	C_62	C_32	C_39	C_47	C_7	C_71	C_24	C_74	C_6	C_78
108.9053	C_17	C_28	C_56																														
117.9238	C_1	C_61	C_44	C_30	C_3	C_23	C_5	C_15	C_14	C_86	C_53	C_91	C_90	C_21	C_67	C_4	C_11	C_72	C_77	C_13	C_12	C_64	C_58	C_62	C_32	C_39	C_47	C_7	C_71	C_24	C_74	C_6	C_78
133.3239	C_1	C_61	C_44	C_30	C_3	C_23	C_5	C_15	C_14	C_86	C_53	C_91	C_90	C_21	C_67	C_4	C_11	C_72	C_77	C_13	C_12	C_64	C_58	C_62	C_32	C_39	C_47	C_7	C_71	C_24	C_74	C_6	C_78
193.4340	C_1	C_61	C_44	C_30	C_3	C_23	C_5	C_15	C_14	C_86	C_53	C_91	C_90	C_21	C_67	C_4	C_11	C_72	C_77	C_13	C_12	C_64	C_58	C_62	C_32	C_39	C_47	C_7	C_71	C_24	C_74	C_6	C_78

TABLA 4

Distancia de unión	Obj. No 35	Obj. No 36	Obj. No 37	Obj. No 38	Obj. No 39	Obj. No 40	Obj. No 41	Obj. No 42	Obj. No 43	Obj. No 44	Obj. No 45	Obj. No 46	Obj. No 47	Obj. No 48	Obj. No 49	Obj. No 50	Obj. No 51	Obj. No 52	Obj. No 53	Obj. No 54	Obj. No 55	Obj. No 56	Obj. No 57	Obj. No 58	Obj. No 59	Obj. No 60	Obj. No 61	Obj. No 62	Obj. No 63	Obj. No 64	Obj. No 65	Obj. No 66	Obj. No 67	Obj. No 68
3.563168																																		
3.750458																																		
3.947809																																		
3.998157																																		
4.011094																																		
4.063237																																		
4.198084																																		
4.226612																																		
4.464646																																		
4.872865																																		
4.893643																																		
4.872129																																		
4.926675																																		
4.971014																																		
4.978143																																		
5.198125																																		
5.214487																																		
5.253881																																		
5.581895																																		
5.813502																																		
5.836816																																		
5.918704																																		
6.257276																																		
6.718959																																		
6.858308																																		
6.982189																																		
6.953983																																		
6.984511																																		
7.318838																																		
7.527378																																		
7.711879																																		
7.950752																																		
8.065072																																		
8.195647																																		
8.373973																																		
8.422847																																		
8.556163																																		
8.909092																																		
8.630274																																		
9.034437																																		
9.372983																																		
9.653525																																		
10.00995																																		
10.15419																																		
11.34710																																		
11.78535																																		
13.44589																																		
14.07539																																		
14.25289																																		
15.51327																																		
15.95023																																		
16.09897																																		
17.49455																																		
19.04309																																		
20.77865																																		
21.15933																																		
21.69370																																		
22.97155																																		
24.07094																																		
25.14645																																		
26.80572																																		
27.84556																																		
29.16058																																		
30.67649																																		
31.90383																																		
42.21980	C_46	C_98	C_33	C_42	C_31																													
50.72868																																		
52.88701																																		
55.34226	C_46	C_98	C_33	C_42	C_31	C_7	C_38	C_9	C_20	C_57	C_86	C_18	C_83	C_100	C_25	C_85	C_82	C_94	C_26	C_73	C_48	C_75	C_81	C_88	C_70	C_87	C_37	C_35	C_34	C_5	C_63	C_43	C_37	C_64
80.89204	C_55	C_76	C_89	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_85	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_58	C_60	C_10	C_69	C_79	C_18	C_45	C_46	C_98	C_33	C_42	C_31	C_7	C_38
108.8063																																		
117.9238	C_55	C_76	C_89	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_85	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_58	C_60	C_10	C_69	C_79	C_18	C_45	C_46	C_98	C_33	C_42	C_31	C_7	C_38
133.3230	C_55	C_76	C_89	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_85	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_58	C_60	C_10	C_69	C_79	C_18	C_45	C_46	C_98	C_33	C_42	C_31	C_7	C_38
193.4340	C_55	C_76	C_89	C_19	C_92	C_36	C_41	C_25	C_93	C_54	C_85	C_88	C_89	C_27	C_29	C_40	C_50	C_66	C_49	C_51	C_58	C_60	C_10	C_69	C_79	C_18	C_45	C_46	C_98	C_33	C_42	C_31	C_7	C_38

TABLA 4

Distancia de unión	Obj. No 69	Obj. No 70	Obj. No. 71	Obj. No. 72	Obj. No 73	Obj. No 74	Obj. No 75	Obj. No 76	Obj. No 77	Obj. No 78	Obj. No 79	Obj. No 80	Obj. No 81	Obj. No 82	Obj. No 83	Obj. No 84	Obj. No. 85	Obj. No 86	Obj. No 87	Obj. No 88	Obj. No 89	Obj. No 90	Obj. No 91	Obj. No 92	Obj. No 93	Obj. No 94	Obj. No 95	Obj. No 96	Obj. No 97	Obj. No 98	Obj. No 99	Obj. No 100
3.583186																																
3.750458																																
3.947809																																
3.999157																																
4.011094																																
4.063227																																
4.190164																																
4.228012																																
4.454646																																
4.572865																																
4.633947																																
4.872129																																
4.929875																																
4.971014																																
4.978143																																
5.159175																																
5.214487																																
5.259881																																
5.591895																																
5.813502																																
5.836815																																
5.918704																																
6.257275																																
6.718968																																
6.855308																																
6.863196																																
6.953983																																
6.964571																																
7.310899																																
7.537770																																
7.711879																																
7.860752																																
8.063072																																
8.199647																																
8.379873																																
8.472847																																
8.556163																																
8.609032																																
8.630274																																
9.034432																																
9.373983																																
9.653525																																
10.00995																																
10.15419																																
11.34710																																
11.79525																																
12.44580																																
14.07539																																
14.25298																																
15.51327																																
15.85323																																
16.09892																																
17.49455																																
19.04309																																
20.77665																																
21.18033																																
21.68370																																
22.37156																																
24.02084																																
25.14645																																
26.80622																																
27.04556																																
30.16358																																
30.67649																																
31.00383																																
42.21990																																
50.72868																																
52.98701																																
65.34326																																
93.99204	C_9	C_20	C_57	C_96	C_16	C_83	C_100	C_35	C_65	C_92	C_94	C_26	C_73	C_48	C_75	C_81	C_68	C_70	C_87	C_97	C_95	C_34	C_8	C_63	C_43	C_37	C_64					
108.8053																																
112.9238	C_9	C_20	C_57	C_96	C_16	C_83	C_100	C_35	C_65	C_92	C_94	C_26	C_73	C_48	C_75	C_81	C_68	C_70	C_87	C_97	C_95	C_34	C_8	C_63	C_43	C_37	C_64	C_17	C_28	C_56		
133.3230	C_9	C_20	C_57	C_96	C_16	C_83	C_100	C_35	C_65	C_92	C_94	C_26	C_73	C_48	C_75	C_81	C_68	C_70	C_87	C_97	C_95	C_34	C_8	C_63	C_43	C_37	C_64	C_17	C_28	C_56	C_52	
193.4340	C_9	C_20	C_57	C_96	C_16	C_83	C_100	C_35	C_65	C_92	C_94	C_26	C_73	C_48	C_75	C_81	C_68	C_70	C_87	C_97	C_95	C_34	C_8	C_63	C_43	C_37	C_64	C_17	C_28	C_56	C_52	C_80

TABLA 5

IND	Z1	Z2	Z3	Z4	Z5	Z6	Z7
1	1.86949919	-0.96130906	-1.42853267	-0.19307691	-0.05366869	0.15778078	-0.3407651
2	-0.09921359	-0.59372633	1.11011546	-0 62849333	1.12220974	-0.12571682	1 41549174
3	1.46751343	-1.95797727	0.14993455	-0.49707255	0.89523597	-0.77254565	-0.67311611
4	3.03744682	1.68588807	1.81280999	0.8051418	-0.62638831	0.85709477	1.21193402
5	2.33293389	-2.60358446	-0.55380584	0.58710433	0.98911365	-1.21726737	0.34276346
6	-1.4238316	0.86158934	0.57472929	-0.26151531	0.65691183	0.31717206	0.78412843
7	-1.48142014	-1.4526261	0.95984764	-0.93569324	-0.89143727	0.60050656	0.28871398
8	-2.54977108	-0.71726872	2.58461774	0.91886693	0.9108471	-0.20016211	0.50394099
9	-1.02817278	-1.11206131	0.64025224	-0.51726015	0.17881497	0.97754456	0.16176482
10	-0.89297328	1.06452684	1.05954209	-2.9027521	-3.11480681	-3.95664568	1.75529034
11	3.26682132	1.75526344	1.86275234	0.80528389	-0.46730851	0.73512927	1.07893567
12	3.51395728	0.32194242	1.4956099	-0.32625234	-0.29197977	0.67722489	-0.55218243
13	1.78211144	0.7502642	1.47042918	1.21999076	-0.26503025	0.54510258	0.96463296
14	2.4943022	-0.75907141	-0.42770123	0 15518447	-0.05995445	0.10671057	0.29174433
15	2.32289614	-2.25072113	-1.40531393	0.04073346	0.95658589	-1.09750782	-0.32487933
16	-1.58506923	-2.03644748	0.59927526	-0.07678849	-0.24384398	1.51425808	0.33072649
17	-0.45913798	0.51841244	4.72343698	-2.52035528	0.67647138	-1.91982107	-5.21972801
18	-2.41150062	1.84997068	-1.62758932	1.14108949	-0.15931018	0.08333101	-1.04690961
19	-2.0631286	0.31554563	-0.7728803	0.54864792	0.74741467	-0.24116095	-0.13164226
20	-1.82312442	-2.42007744	-0.40465184	0.07920842	0.0646801	1.06259124	-0.17630934
21	4.65595898	-1.878484	-0.55712068	0.18162604	0.61634265	-1.03007702	0.61017093
22	-1.38186046	0.23214366	1.11277584	1.70581508	0.56437845	-0.38759385	0.72320108
23	2.21960698	-2.31087256	0.15609296	0.029033	0.19627752	0.12862551	0.69849413
24	0.21757873	0.1348175	2.38442708	-0.1340733	0 50575234	0.2679183	-1.57206018
25	-0.12784001	-0.26863278	1.17484129	0.32356807	-1.50497651	-0.52135486	0.86314902
26	0.64849006	-1.23866115	-0.49369918	0.36304671	0.17741244	0.15945965	0.10883808
27	0.21232786	0.43272066	-1.17153558	0.35570751	0.02516036	0.1790256	-0.47506778
28	-1.4286221	-0.50799712	-0.72969756	-3.1673852	-4.37662226	0.37109987	-0.45906698
29	0.4842734	0.64891227	-1.85106799	0.60849413	0 23112334	-0.08534668	-0.63436757
30	3.3946355	-0.37909436	-3.56536159	-0.22682884	0.12884227	-0.75209016	-1.09342777
31	-0.229488	2.59114164	-1.51639089	1.90485554	-0 2748351	-0.0099495	-0.11317616
32	2.37747083	2.36316412	-0.07680686	-0.56375458	-0.01198736	0.29962577	-1.98695457
33	*-0.81977166*	*3.54131667*	*-2.09663594*	*0.58372899*	*0.23875656*	*-0.29870276*	*-1 11332549*
34	-1.63583963	-2.16579064	-0.86224459	0.79456049	1.52616512	-1.60259285	-0.47763247
35	-2.26925114	-0.51804285	-1.39269293	-0.60341737	-0.78791374	1.65827714	-0.67520723
36	-1.43038903	0.7846992	0.03239495	1.47113195	0.08656367	-0.09221344	0.03624265
37	-3.44748093	-0.02468419	1.69024186	1.0833824	1.08088433	-0.92567521	-0.13026288
38	-0.96723606	-1.56437241	0.56712018	-0.2306919	-1.5572244	-0.21195408	0.46211516
39	3.7347797	2.55278101	1.74133879	-0.99543458	-0.3826688	0.38105851	-2.07314794
40	-0.91143582	1.1565799	-0.1720462	1.29750955	-0.15044152	-0.00103424	-0.13175748
41	-1.92732261	0.74178681	0.19312782	1.8974433	0.37491957	-0.28177123	-0.08526321
42	0.50690303	3.10963437	-1.60116749	0.83735041	-0.12998635	0.15181396	-0.47033507
43	-3.07354451	-1.00239556	1.85999155	1.67207729	1.25121091	-0.75669802	0.06462929
44	2.7312314	-1.3290674	-2.40676828	-0.30313571	0.25675423	-0.81473184	-0.53410355
45	-1.53971743	2.32889981	-0.56075401	1.64840795	0.33093983	-0.32589738	-0.34266838
46	-2.18507156	2.01975861	-1.24953163	1.13650947	0.1913492	-0.23951002	-0.57227642
47	2.56911532	2.1095357	1.65969095	-0.99647125	-0.3109511	-0.11336303	-1.56432285
48	-0.34756849	-2.02156346	-1.18536902	-0.09728831	0.07712127	0.35740568	-0.3355619
49	-1.06846459	2.18479679	-1.12711059	0.13902509	-0.17284082	0.429253	-0.51372476
50	-0.86868334	0.49918124	-0.77069144	0.93996262	0.11441716	0.01532444	-0.33819332
51	-0.31863467	1.33030487	-1.00807664	0 56685538	0.13970571	0.25099778	-0.20638714
52	-1.76535128	-0.13537157	0.08031388	0.22787396	-0.48337643	0.33487977	-0.06746996
53	1.25165024	-1.48318762	-0.16812138	0 15086083	0 29934301	-0.49613583	-0.2145471
54	-0.61172765	-0.86260509	0.53959709	0.39503879	0.05217694	0.29707743	-0.02484131
55	-0.88265402	1.0039087	0.51652628	1.40173235	0.24789961	-0.03356697	0.61292582
56	-2.95199273	1.79094932	-0.94597304	-3 77465782	-1.53049995	2.09024254	-0.27450856
57	-1.20132667	-1.89624002	-0.69339796	0.14960167	-0.18212623	0.89575922	-0.21106371
58	1.3788731	0.71529817	0.88750487	-0.72542628	-0.39959044	1.18196347	1.05762183
59	0.03055838	1.72624098	-0.25446629	0.57633161	-0.01884643	0.38866069	0.1227422

TABLA 5

IND	Z1	Z2	Z3	Z4	Z5	Z6	Z7
60	1.35004493	1.70748126	-0.47564837	0.64915128	-0 3251118	0.51241441	-0.07184188
61	2.02435728	-0.68651094	-1.42279986	0.01926267	0.10980405	-0.11453848	-0.28587112
62	2.98455289	2.89103976	-0.40284477	0.3295486	-0.71572594	0.69017563	0 04577492
63	-2.48406847	-2.00677977	2.16307304	0.71340665	1.06409437	-0.32025214	0.21033915
64	-3.28585555	-0.32473568	1.31218837	0.26200871	0.52213145	0.15715063	0.15510091
65	-1.64359777	-0.33996373	-0.67080266	-0.50118148	-0.34388827	1.16774914	0.07952554
66	-0.73042532	0.81029601	-0.02536078	0.33387526	-0.49731295	0.19407415	0.09946191
67	2.11682134	-3.0473072	-1.49414478	-0.84129399	0.41596406	0.84894566	-0.59978978
68	-0.5284245	-1.67305287	-2.11738422	-0.5072479	0.50156631	-0.18110534	-0.40288317
69	-1.06667952	1.82376671	-1.29069842	-1.25176224	1.22533856	-0.41496092	0.46606219
70	-0.46163753	-0.29000454	-0.78643713	-0.62464469	1.02897931	0 30130778	0.60339941
71	-0.79043906	-1.2531265	1.28704824	-0.18541938	1.47638945	-0.65292928	-0.07979405
72	3.66808937	1.83927	1.96063527	0.61168989	-0.82972809	0.9266807	1.35831221
73	-0.33792375	-1.04444183	-0.24772739	0.81876612	0.48422461	0.16451036	0.3578556
74	0.09443517	0.86432846	1.97401376	-1.91796263	1.78167144	-0.45429521	-1.29864917
75	-0.01541637	-2.13196295	-0.41807107	0.23203154	0.27094065	0.14743857	0 29836606
76	-0.89255308	0.89331117	-0.46851814	1.20460851	0.07637449	0.14905104	0.19703885
77	3.41488549	0 87909746	1.90927995	0.6147305	-0.57456978	0 99056071	1 20903315
78	-0.08644794	0.73391074	1.16482428	0.89952059	0.60104766	-0.02777173	0.90779065
79	-0.39270679	2.09857205	-1.24589378	-0.14988503	0.78463511	-0.05012181	0.29929915
80	-1.1087848	4.69062598	-0.88116199	-6.29907005	4.40196014	-0.77595151	3.43622351
81	-0.44970486	-1.56439685	-0.11069519	-1.11860354	-1.91476795	-1.6495363	0.19401033
82	-1.20838998	-0.19692046	0.27008224	-0.42658546	-0.92453571	1.22757598	0 27642
83	-1.51835771	-3.05449886	0.24595249	-0.81804812	-0.74323809	2.28816677	0.20966081
84	3.40893048	0.37081756	0.8321781	0.36189388	-0.12474043	0.30256921	0.9854682
85	0.41471151	-0.54677632	1.15295537	0.63017887	-0.08156292	0.30423882	0.4721673
86	2.06506883	-0.39065967	0.35992601	0.1991661	-0.68719324	-0.93302206	0.74668324
87	1.17534543	-2.81452562	-0.4318929	-0.4492819	1.24251089	-0.67349186	-0.09797169
88	-1.11464886	0.08674743	0.66695019	0.75996252	-0.65105228	0.93661945	0.29187367
89	1.10924787	0.84793372	0.49519359	-0.06618636	-0.27885986	0.66854541	-0.70582717
90	2.26027204	-0.48710011	-0.22270104	-0.60589544	-1.32978949	-2.21468739	0.7030993
91	1.06528082	-0.86013981	-0.83191676	0.23263765	0.25680606	-0.52263578	0.08168335
92	-1.33128837	1.28773212	-0.20831539	1.01908147	0.52342798	0.00013234	0 40247954
93	-0.2227853	0.13332099	0 40745133	-0.3962507	-2 40967561	-2 7604506	0 65901805
94	-1.49577684	0.60756265	-0.14489279	-1.00104737	-0 08589734	0.73338202	0.3761398
95	-0.0062386	-2.89080765	0.04762406	-0.66469041	-0.05240583	-0.29991909	0.45624808
96	-1.58217532	-2.68365205	0.32222672	-0.20989423	-0.27937296	0.99391592	0.11062017
97	0.79958291	-3.53672354	-0.84636508	-0.55046111	0.60965949	-0.03551295	-0.46649317
98	-1.88095448	2.79924222	-1.37848054	1.05899486	-0.15499518	-0.37632661	-0 67280632
99	-1.25454861	1.41560341	-0.66387466	0.59488436	-0 56399723	-1 40609919	0.20846327
100	-1.35411452	-1.62661437	-0.34711043	-1.02494481	-1.27247128	1.19657527	-0.54965428

TABLA 5

IND	Z8	Z9	Z10	Z11	Z12	Z13	Z14
1	-0.4229629	0.21726375	1.31644249	-0.07131984	0.21407608	-0.01469193	-0.02948968
2	-0.0777421	0.20874474	-0.5664574	-0 09361097	0.09762011	0.05223006	-0.03236386
3	0.05318106	1.13037352	-0.57641251	0.38869863	0.23673959	0.01931331	0.07087146
4	-0.39633596	0.07684494	-0.39688561	0.17581083	-0.29291051	0.06361837	-0.06917732
5	-0.66122634	0.39713078	-0.3192378	0.03913909	-0.37516439	-0.00156774	0.0814158
6	-0.013226	0.44113194	0.11245107	0.53112454	0.51497911	0.03730895	-0.00724361
7	0.43911583	0.82617048	0.45514268	-0.40220304	0.11790377	-0.08577018	-0.01343207
8	-1.90792134	-0.8483988	0.69584158	0.96855861	-0.25180021	0.05268264	-0.03040039
9	1.01756658	-0.41819731	-0.80010436	-0.14882009	-0 06850993	0.11859226	-0.0657021
10	1.47791574	-1.79850387	-0.54208521	0.51905185	0.35211001	-0.55340134	-0.05317149
11	-0.77239909	-0.45056194	0.21478783	0.20546623	-0.15924557	0.06040223	-0.01193383
12	0.49627064	-0.18295503	0.09522878	-0.04151047	-0.24447087	0.09432275	0.02134452
13	0.15297471	0.13200974	0.08375797	-0.2115241	-0.50282934	0.03936132	-0.0383644
14	-0.33821229	0.0496205	0.32639045	0.02644809	-0.08277956	0.01981445	0.01437063
15	-0.19735726	0.29209066	0.34284822	0.25312859	0.20157322	-0.00644166	-0.00330161
16	0.75576631	-0.74743943	-0.00032833	0.22829185	-0.11900289	0.12744401	-0.02018864
17	1.17867638	0.35049704	-0.02622977	-0.31236998	0.01977469	-0.06003991	-0.0383793
18	-0.25116182	-0.32002878	-0.9593265	-1.37114914	-0.15585186	-0.0286818	0.00155375
19	0.20809859	-0.28012543	-0.90695585	-0.07431607	0.20392548	0.03697276	-0.04886534
20	0.63870213	-0.88731714	-0.56340432	-0.50471794	0.04577417	0.09873399	-0.02118783
21	-1.01865823	0.3992615	-0.48390386	0.32230556	-0.3685291	0.01921761	0.06630728
22	0.47968283	1.13778269	0.29101948	0.90016579	-0.64900181	0 01417713	-0.04296322
23	-0.06969544	0.20564862	-0.38901309	0.2067808	-0.11738767	0.05416559	0.09303382
24	0.55018048	0.24029616	0.15525364	0.38852063	-0.2744105	0.10671806	-0.01943775
25	1.40224488	-0.62673979	0.47902728	-0.31649523	0.00745818	-0.17646902	-0.01050961
26	0.07608813	0.48232732	0.80611225	-0.31643889	0.35865331	-0.02320795	0.03005999
27	-0.43882096	-0.23589196	0.62252848	-0.42170182	0.561994	-0.02226236	0.01557315
28	-0.56731602	-1.04860739	1.30305626	0.85349208	-0.46711511	-0.2733527	0.09764779
29	0.23157938	-0.78065252	0.60860711	-0.24543044	0.27411966	0.01237323	-0.01341855
30	-0.98463073	-0.2543485	1.17156811	0.0951879	-0.61729504	0.00193236	-0.29810646
31	1.80118615	1.13102351	0.0670382	-0.16169388	-1.35325449	0.01444223	-0.01477052
32	-0.055683	-0.53508124	-0.14177127	-0.00471743	0.36729858	0.08232369	0.03851716
33	-1.05439999	-1.54155963	-0.31198627	0.27320483	-0.43903992	0.06951749	0.07932671
34	-1.26679183	-0.12153744	-0.18580717	-0.46737781	0.00196528	-0.05158642	0.01662005
35	0.18243144	-0.41790142	-0.69455388	0.06545933	-0.06511839	0.10559451	-0.01653943
36	0.38796162	0.50180962	-0.44224459	-1.29632859	0.24735888	-0.07443683	-0.03085255
37	-2.2517321	-1.58861026	-0.13938363	0.33884992	0.12172904	-0.0001262	-0.01653035
38	1.01727339	-0.38306991	0.34251179	-0.66395041	0.03488606	-0.17672835	0 00236006
39	0.17405637	0.19834947	-0 01052003	0.51999856	-0.33911457	0.10431132	0.01543462
40	0.1373921	0.18970961	0.37901199	-0.85831947	-0.08262043	-0.06041811	0.03009801
41	0.29841082	0.27519193	0.68474009	-0.49226238	-0.61723871	-0.02837814	0.02077797
42	0.23713967	-0.38950297	0.41400463	0.46052458	0.10005319	0.03374016	0.1052861
43	-1.37174796	-0.72218284	1.22953021	-0.07169983	0.23843832	-0.03919489	0.02781293
44	-1.21221759	0.80074073	0.40092936	-0.48217113	0.3671409	-0.0823501	-0.10202522
45	0.07158972	-0.38817769	-0.11379265	0.45517663	-0.79307221	0 04997184	0.027099
46	0.02807302	0.34416084	-0.73625301	-0.106986	0.20071929	-0.0198462	-0.00269614
47	-0.05082663	-0.22704369	-0.74181638	-0.20489386	0.1017643	0.02723413	-0.03478067
48	-0.12482907	0.13099362	0.26417075	-0.1966684	-0.03804234	0.02332505	-0.06835432
49	-0.60980385	-0.09939433	0.83309996	-0.25330796	-0.06711027	-0.00584069	0.12287998
50	0.02838126	0.02224798	0.40867487	-0.3530437	-0.0805706	-0.01215575	0.01623023
51	0.26896518	-0.28716716	-0.42456624	0.21300866	0.67991555	0.03445483	0 0076203
52	-1.0762134	1.42569219	-0.06372011	-0.22780446	-0.37753462	-0.05844375	-0.0184491
53	-0.41287604	0.27403831	-0.55709412	-0.36383175	-0.14203763	-0.00759307	0.029079
54	-0.07336289	0 51349241	0.25790923	-0.14709198	0 0285875	-0.00221807	-0.00944229
55	1.11057283	0.59988427	0.27129128	0.57439633	-0 1769181	0 0140847	-0.01692242
56	-3.4284606	1.8215094	-0.95365512	0.01279579	-0.12895685	-0.00013001	-0.01171875
57	0.49700798	0.08292078	0.15150934	-0.80519897	0.04445278	0.02038525	-0.0122227
58	0.48196214	-1.41598442	-0.42010492	-0.46046303	-0.0870688	0.10379577	-0.03667333
59	-0.06423575	-0.22103869	-1.01007574	0 27967022	0.64758708	0.04598479	-0.01883354

TABLA 5

IND	Z8	Z9	Z10	Z11	Z12	Z13	Z14
60	-0.25276646	-0.72052405	0.00716542	-0.20298308	-0.04977281	0.04664497	0.03382813
61	-0.6064345	-0.3226092	0.26142901	-0.25087463	0.15069626	0.00753877	-0.05660055
62	-0.85837868	-0.63614192	-0.00808896	-0.29462291	0.33268497	0.02225986	0.05818177
63	-1.37648393	-0.98595531	0.17031146	0.13710033	0.0260755	0.03567694	-0.0240572
64	-1.53664518	-0.13144264	0.15543112	-0.45922373	-0.05340031	-0.01090094	-0.0292572
65	0.58410035	0.2505178	-0.90081988	0.39272707	0.36251323	0.0781692	-0.03531799
66	-0.25866135	0.47999957	-0.14960272	-0.16806215	0.32742389	-0.05918328	0.00423882
67	1.19023801	-1.86081241	-0.05244581	0.42579394	0.01127316	0.2076315	-0.03823544
68	-0.01176628	0.6578966	0.26968324	0.72458946	0.07371609	0.02881451	-0.08030482
69	-0.21883601	0.69286699	0.57044888	1.50906029	0.3447864	0.05503222	-8.479E-05
70	1.11280122	-0.29192546	0.34668406	0.51097443	0.36031551	0.1034962	-0.00702473
71	-0.1730779	-0.24108235	0.34929597	-0.08013507	0.05411225	0.03826426	0.01080485
72	-0.30919641	0.19446677	0.01061185	-0.12947544	0.38010716	0.0072949	-0.03423717
73	1.11714809	0.31283065	0.12357113	-0.02084186	0 32068919	0 02670414	0.0061584
74	0.84069358	0.59298042	0.31514015	-0.10596591	0.07128563	0.0932081	-0.00635499
75	0.53875276	0.55509042	0.18448917	0.09712552	-0.16534497	0.02589589	0.01531911
76	1.37637569	0.75349622	0.41952388	-0.07703046	0.32954203	-0.02690944	0.02533865
77	-0.43060342	-0.04168893	0.22285241	-0.53009874	0.43413661	0.01391615	-0.02328955
78	0.15422192	0.29170645	0.05083149	0.67862447	0.3266183	0.03432208	-0.03039478
79	0.73661996	0.15649433	0.22228727	0.64580994	0.51313142	0.04994136	0.0405727
80	0.77161046	-0.32909272	0.71815992	-1.30749125	-0.51817874	0.15019269	0.01798262
81	0.91857239	-0.34338853	-0.01501242	-0.00342927	0.31363486	-0.29918219	0.01520032
82	-0.35349045	1.12274125	-0.52046513	0.27280876	0.25029727	0.00300924	-0.03191361
83	0.86729485	-0.73832065	-0 01142079	0.28476545	-0.20184739	0.16475025	-0.0090316
84	-0.44608128	-0.00245098	-0.43369481	-0.00243609	-0.08438067	0.04862894	0.01927609
85	-0.98081786	0.39239921	0.52754307	-0.31342203	-0.00827933	-0.02335802	-0.00735561
86	0.17209209	-0.19017416	-0.33808331	0.00566626	0.01337874	-0.11131625	0.03451405
87	0.61459152	-0.34692415	-0.82006484	0.39767435	-0.41694155	0.1137634	0.11406401
88	0.05531479	1.3060272	0.84576606	-0.56964349	-0.31290782	-0.04439992	0.00644669
89	-0.02178942	0.32875321	1.00192117	-0.41323601	0.64487612	-0.0145001	0.07456504
90	0.02389866	-0.48626484	-0.90854168	0.01271663	-0.17557273	-0.24305762	0.01610742
91	-0.58518152	0.61393072	-0.77372228	-0.14458079	0.18593544	-0.02782226	0.00085287
92	1.01026071	0.34527077	-1.22344589	0.60332635	0.25944924	0.06251965	-0.06861409
93	0.77886134	-0.54224658	-0.37813438	-0.11889176	-0.05452041	-0.41542839	-0.03329613
94	-0.78463765	0.91723966	-0.27121071	0.8654381	-0.35177949	0.06357056	-0.02425922
95	-0.78654129	0.82783152	-0.63528271	-0.34243986	-0.14583159	-0.05597774	0.1299358
96	-0.43320012	0.4692954	-0.39231693	-0.33260586	-0.21086837	0 03697024	0.01343513
97	0.60442191	-0.43192663	-0.74331726	-0.05574261	-0.19441981	0.09681286	0.0642418
98	-0.2407567	-0.13370891	-1.25935654	0.11401291	-0.33277403	-0.0017201	-0.01053897
99	1.27236895	0.46206203	0.73965114	0.87158502	0.41281269	-0.18512727	0.04628846
100	1.21550701	-0.63015621	1.01543356	0.2275969	-0.07726779	0.01064099	0.0499438

TABLA 6

Distancia de unión	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No.	Obj. No.	Obj. No	Obj. No
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
5.419718	C_7	C_38	C_9	C_16	C_100																
5.438824	C_81	C_93																			
5.482388	C_70	C_73	C_75																		
5.574488	C_32	C_39	C_47																		
5.597477	C_3	C_23	C_5	C_15																	
5.764996	C_87	C_97	C_95																		
5.828564	C_20	C_57	C_96	C_83																	
6.203174	C_19	C_99																			
6.206764	C_24	C_74																			
6.466206	C_8	C_63	C_37	C_43	C_64																
6.934991	C_2	C_71	C_54	C_85																	
7.002265	C_58	C_60	C_89																		
7.351754	C_18	C_46	C_98	C_45	C_49	C_33															
7.637473	C_25	C_81	C_93																		
7.658294	C_48	C_68	C_70	C_73	C_75																
7.765815	C_19	C_99	C_40	C_50	C_66	C_51	C_59	C_55	C_76	C_92											
7.906781	C_1	C_61	C_14	C_26	C_53	C_91	C_3	C_23	C_5	C_15											
8.040079	C_13	C_78	C_58	C_60	C_89																
8.072192	C_4	C_11	C_72	C_77	C_12	C_84															
9.628334	C_19	C_99	C_40	C_50	C_66	C_51	C_59	C_55	C_76	C_92	C_22	C_36	C_41								
9.971753	C_7	C_38	C_9	C_16	C_100	C_20	C_57	C_96	C_83												
10.07124	C_27	C_29	C_69	C_79																	
11.84281	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88												
12.69656	C_18	C_46	C_98	C_45	C_49	C_33	C_27	C_29	C_69	C_79											
12.79260	C_32	C_39	C_47	C_62																	
12.84991	C_7	C_38	C_9	C_16	C_100	C_20	C_57	C_96	C_83	C_35	C_65										
12.98905	C_1	C_61	C_14	C_26	C_53	C_91	C_3	C_23	C_5	C_15	C_86	C_90									
13.50225	C_67	C_87	C_97	C_95																	
16.27379	C_21	C_30	C_44																		
16.42724	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88	C_48	C_68	C_70	C_73	C_75							
18.13549	C_18	C_46	C_98	C_45	C_49	C_33	C_27	C_29	C_69	C_79	C_31	C_42									
18.54976	C_8	C_63	C_37	C_43	C_64	C_34															
19.41036	C_4	C_11	C_72	C_77	C_12	C_84	C_13	C_78	C_58	C_60	C_89										
21.24588	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59
23.73859	C_1	C_61	C_14	C_26	C_53	C_91	C_3	C_23	C_5	C_15	C_86	C_90	C_67	C_87	C_97	C_95					
26.62459	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59
27.29744	C_28	C_56																			
30.28623	C_17	C_24	C_74																		
31.85452	C_4	C_11	C_72	C_77	C_12	C_84	C_13	C_78	C_58	C_60	C_89	C_32	C_39	C_47	C_62						
33.79808	C_1	C_61	C_14	C_26	C_53	C_91	C_3	C_23	C_5	C_15	C_86	C_90	C_67	C_87	C_97	C_95	C_21	C_30	C_44		
39.61636	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59
42.10278	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59
56.51686	C_1	C_61	C_14	C_26	C_53	C_91	C_3	C_23	C_5	C_15	C_86	C_90	C_67	C_87	C_97	C_95	C_21	C_30	C_44	C_4	C_11
64.54491	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59
70.87754	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59
76.81706	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59
108.0842	C_1	C_61	C_14	C_26	C_53	C_91	C_3	C_23	C_5	C_15	C_86	C_90	C_67	C_87	C_97	C_95	C_21	C_30	C_44	C_4	C_11
112.9913	C_1	C_61	C_14	C_26	C_53	C_91	C_3	C_23	C_5	C_15	C_86	C_90	C_67	C_87	C_97	C_95	C_21	C_30	C_44	C_4	C_11
153.7985	C_1	C_61	C_14	C_26	C_53	C_91	C_3	C_23	C_5	C_15	C_86	C_90	C_67	C_87	C_97	C_95	C_21	C_30	C_44	C_4	C_11

TABLA 6

Distancia de unión	Obj_No	Obj_No	Obj_No.	Obj_No.	Obj_No	Obj_No	Obj_No.	Obj_No	Obj_No	Obj_No	Obj_No	Obj_No	Obj_No.	Obj_No.	Obj_No.	Obj_No	Obj_No.	Obj_No.	Obj_No.	Obj_No	Obj_No.
	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
5.419718																					
5.438824																					
5 482388																					
5 574488																					
5.597477																					
5 764996																					
5.828564																					
6.203174																					
6 206764																					
6 466206																					
6 934991																					
7.002265																					
7 351754																					
7.637473																					
7.658294																					
7 765815																					
7 906781																					
8 040079																					
8 072192																					
9.628334																					
9.971753																					
10 07124																					
11.84281																					
12 69656																					
12 79260																					
12.84991																					
12.98905																					
13 50225																					
16 27379																					
16.42724																					
18 13549																					
18 54976																					
19.41036																					
21 24588	C_55	C_76	C_92	C_22	C_36	C_41															
23 73859																					
26 62459	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81	C_93												
27.29744																					
30 28623																					
31.85452																					
33 79808																					
39 61636	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81	C_93	C_8	C_63	C_37	C_43	C_64	C_34						
42 10278	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81	C_93	C_8	C_63	C_37	C_43	C_64	C_34	C_7	C_38	C_9	C_16	C_100	C_20
56 51686	C_72	C_77	C_12	C_84	C_13	C_78	C_58	C_60	C_89	C_32	C_39	C_47	C_62								
64 54491	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81	C_93	C_8	C_63	C_37	C_43	C_64	C_34	C_7	C_38	C_9	C_16	C_100	C_20
70 87754	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81	C_93	C_8	C_63	C_37	C_43	C_64	C_34	C_7	C_38	C_9	C_16	C_100	C_20
76 81706	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81	C_93	C_8	C_63	C_37	C_43	C_64	C_34	C_7	C_38	C_9	C_16	C_100	C_20
108 0842	C_72	C_77	C_12	C_84	C_13	C_78	C_58	C_60	C_89	C_32	C_39	C_47	C_62	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82
112 9913	C_72	C_77	C_12	C_84	C_13	C_78	C_58	C_60	C_89	C_32	C_39	C_47	C_62	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82
153 7985	C_72	C_77	C_12	C_84	C_13	C_78	C_58	C_60	C_89	C_32	C_39	C_47	C_62	C_2	C_71	C_54	C_85	C_6	C_94	C_52	C_82

TABLA 6

Distancia de unión	Obj. No	Obj. No.	Obj. No	Obj. No.	Obj. No	Obj. No.	Obj. No	Obj. No	Obj. No.	Obj. No	Obj. No	Obj. No.	Obj. No.	Obj. No	Obj. No.	Obj. No.	Obj. No.	Obj. No	Obj. No	Obj. No
	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
5 419718																					
5.438824																					
5.482388																					
5.574488																					
5 587477																					
5 764996																					
5 828564																					
6 203174																					
6.206764																					
6.466206																					
6.934991																					
7.002265																					
7.351754																					
7.637473																					
7.658294																					
7.765815																					
7.906781																					
8.040079																					
8.072192																					
9 628334																					
9 971753																					
10 07124																					
11 84281																					
12 69656																					
12 79260																					
12 84991																					
12 98905																					
13.50225																					
16.27379																					
16 42724																					
18 13549																					
18.54976																					
19 41036																					
21.24588																					
23.73859																					
26.62459																					
27 29744																					
30 28623																					
31.85452																					
33 79808																					
39.61636																					
42.10278	C_57	C_96	C_83	C_35	C_65																
56 51686																					
64.54491	C_57	C_96	C_83	C_35	C_65	C_18	C_46	C_98	C_45	C_49	C_33	C_27	C_29	C_69	C_79	C_31	C_42				
70 87754	C_57	C_96	C_83	C_35	C_65	C_18	C_46	C_98	C_45	C_49	C_33	C_27	C_29	C_69	C_79	C_31	C_42	C_10			
76 81706	C_57	C_96	C_83	C_35	C_65	C_18	C_46	C_98	C_45	C_49	C_33	C_27	C_29	C_69	C_79	C_31	C_42	C_10	C_28	C_56	
108.0842	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81
112.9913	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81
153 7985	C_88	C_48	C_68	C_70	C_73	C_75	C_19	C_99	C_40	C_50	C_66	C_51	C_59	C_55	C_76	C_92	C_22	C_36	C_41	C_25	C_81

TABLA 6

Distancia de unión	Obj. No.	Obj. No.	Obj. No	Obj. No.	Obj. No	Obj. No.	Obj. No	Obj. No	Obj. No	Obj. No	Obj. No.	Obj. No	Obj. No	Obj. No.	Obj. No	Obj. No	Obj. No.	Obj. No	Obj. No	Obj. No.	Obj. No
	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
5.419718																					
5.438824																					
5.482388																					
5.574488																					
5.597477																					
5.764996																					
5 828564																					
6.203174																					
6 206764																					
6 466206																					
6 934991																					
7.002265																					
7 351754																					
7 637473																					
7.658294																					
7.765815																					
7 906781																					
8.040079																					
8 072192																					
9 628334																					
9.971753																					
10 07124																					
11 84281																					
12 69656																					
12 79260																					
12.84991																					
12 98905																					
13.50225																					
16 27379																					
16 42724																					
18 13549																					
18 54976																					
19 41036																					
21 24588																					
23 73859																					
26 62459																					
27 29744																					
30 28623																					
31.85452																					
33.79808																					
39.61636																					
42.10278																					
56 51686																					
64 54491																					
70 87754																					
76.81706																					
108.0842	C_93	C_8	C_63	C_37	C_43	C_64	C_34	C_7	C_38	C_9	C_16	C_100	C_20	C_57	C_96	C_83	C_35	C_65	C_18	C_46	C_98
112 9913	C_93	C_8	C_63	C_37	C_43	C_64	C_34	C_7	C_38	C_9	C_16	C_100	C_20	C_57	C_96	C_83	C_35	C_65	C_18	C_46	C_98
153 7985	C_93	C_8	C_63	C_37	C_43	C_64	C_34	C_7	C_38	C_9	C_16	C_100	C_20	C_57	C_96	C_83	C_35	C_65	C_18	C_46	C_98

TABLA 6

Distancia de unión.	Obj. No.	Obj. No.	Obj. No.	Obj. No	Obj. No.	Obj No	Obj No	Obj No	Obj. No.	Obj No	Obj. No.	Obj. No.	Obj No.	Obj. No.	Obj. No.	Obj. No.
	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
5 419718																
5 438824																
5 482388																
5.574488																
5.597477																
5 764996																
5 828564																
6 203174																
6 206764																
6 466206																
6 934991																
7.002265																
7.351754																
7 637473																
7 658294																
7.765815																
7.906781																
8.040079																
8 072192																
9 628334																
9 971753																
10 07124																
11.84281																
12.69656																
12 79260																
12 84991																
12 98905																
13 50225																
16 27379																
16 42724																
18 13549																
18 54976																
19 41036																
21 24588																
23 73859																
26 62459																
27 29744																
30 28623																
31 85452																
33.79808																
39 61636																
42 10278																
56 51686																
64 54491																
70 87754																
76 81706																
108 0842	C_45	C_49	C_33	C_27	C_29	C_69	C_79	C_31	C_42	C_10	C_28	C_56				
112 9913	C_45	C_49	C_33	C_27	C_29	C_69	C_79	C_31	C_42	C_10	C_28	C_56	C_17	C_24	C_74	
153 7985	C_45	C_49	C_33	C_27	C_29	C_69	C_79	C_31	C_42	C_10	C_28	C_56	C_17	C_24	C_74	C_80