

ALFREDO SANCHEZ MUÑOZ
M. Phil. en Geografía, M. Sc. en Planificación Urbano-Regional
Profesor de Geografía Regional en la Universidad de Concepción, Chile

**CLASIFICACION NUMERICA EN GEOGRAFIA:
UN EJEMPLO APLICADO
A LAS REGIONES DE CHILE**

I INTRODUCCION

Clasificación es parte de un proceso de reducción de datos que nos permite más bien hablar o pensar en términos de individuos. Así, a manera de ejemplo, el lenguaje es un invento clasificatorio importante creado por la mente humana que, junto con comunicarnos, nos posibilita distinguir entre un lenguaje y otro. En el campo científico la clasificación nos ayuda también a formular hipótesis en nuestro trabajo creativo; algunos científicos van aún más lejos y argumentan que el nivel o grado de desarrollo de una disciplina se mide por su capacidad de establecer relaciones que conlleven a obtener una clasificación. En esta materia, indudablemente las ciencias exactas, como la química y física, han logrado importantes avances, mientras en otros campos del quehacer científico, como en el caso de las ciencias sociales y otras disciplinas preocupadas de los estudios conductuales de la mente humana, deben dedicar esfuerzos mayores para lograr éxitos en materias de clasificación.

El propósito más importante de un sistema de clasificación es llegar a una generalización inductiva acerca de los objetos que nosotros

queremos estudiar y comprender. Así, cada esfuerzo por lograr una clasificación lleva implícito una conceptualización teórica, por ejemplo, si queremos conocer más de cerca los centros urbanos y su crecimiento, debemos primero decidir qué entendemos por centro urbano y, una vez definida nuestra materia objeto de estudio, podemos entrar a la primera etapa de definir un método de agrupamiento apropiado.

II COMO CLASIFICAR

La clasificación es un sistema de agrupamiento de objetos y sucesos (eventos) en clases sobre la base de las propiedades o relaciones que ellos tienen en común. Desde ese punto de vista, un método de clasificación puede adoptar dos direcciones:

- a) una subdivisión lógica de una población: usando criterios previamente definidos por el investigador;
- b) un nivel de aglomeración individual: que define clases de acuerdo a un criterio de agrupamiento determinado.

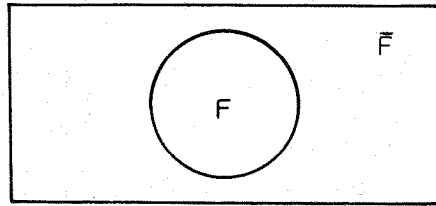


Fig. 1: Diagrama de Venn

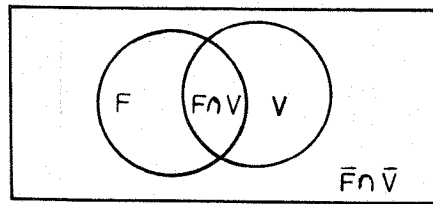


Fig: 2

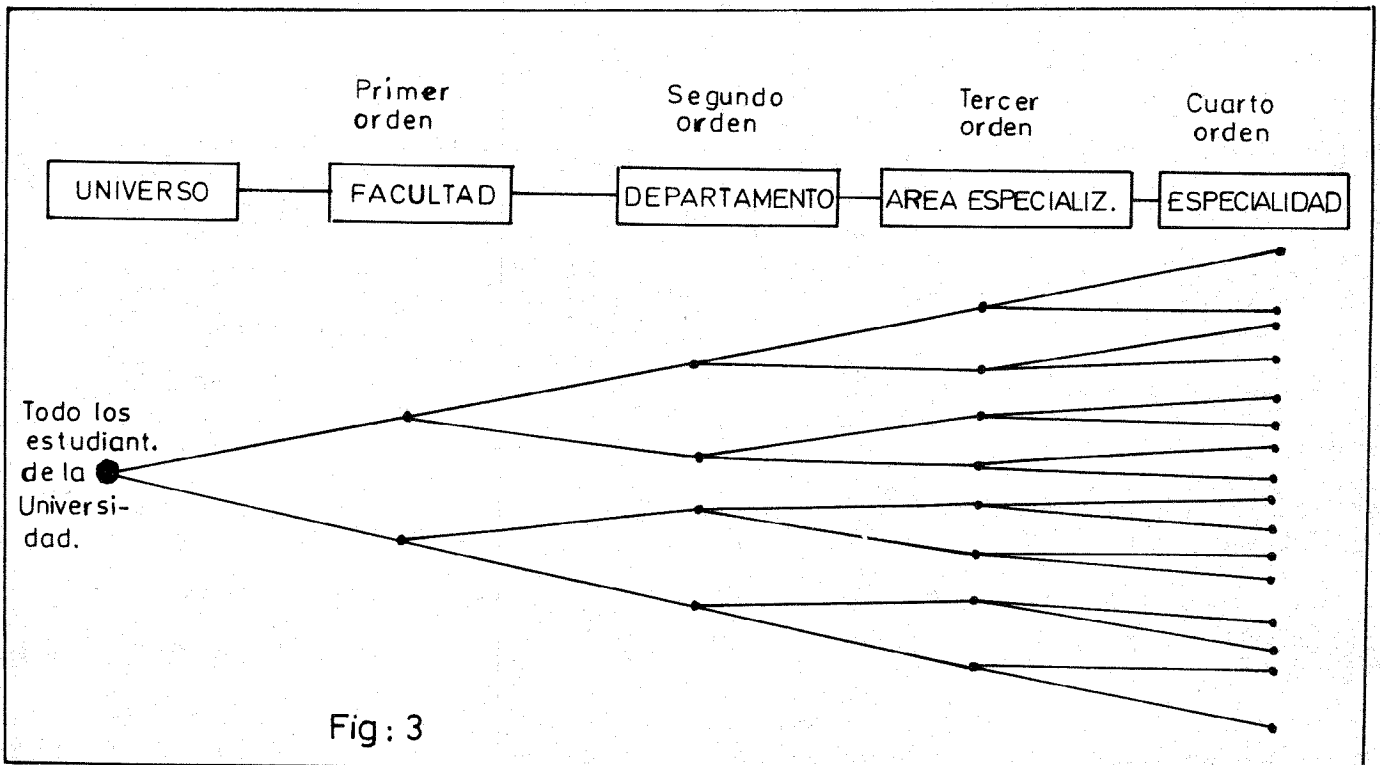


Fig: 3

2.1. Métodos de clasificación siguiendo un criterio lógico de subdivisión

Para ello adoptamos el diagrama de Venn, donde la población estará representada a manera de ejemplo por los estudiantes de un colegio. En este caso el rectángulo (fig. 1) muestra la población estudiantil y el círculo del centro representa a los estudiantes que les agrada fumar (F). Los puntos fuera del círculo comprenden la población que no le agrada fumar (F). En consecuencia, a cada estudiante del ejemplo considerado se le asigna un set F y su complemento será \bar{F} .

Desde un punto de vista estrictamente clasificatorio los estudiantes fueron divididos según un solo criterio, su agrado por fumar. Los atributos implícitos en este criterio son:

- i) cada individuo se ajusta a una u otra clase;
- ii) estas clases son mutuamente excluyentes, es decir, a los miembros de la población les agrada o no fumar.

Este sencillo criterio de clasificación puede también ser redefinido agregándole un segundo criterio adicional, como ejemplo el agrado de beber vino. De esta manera, al introducir un segundo criterio clasificatorio, el diagrama de Venn ofrece ahora cuatro posibilidades de decisión (fig. 2). Los estudiantes que les agrada fumar y no les agrada beber vino (F), los que beben vino (V), los que beben vino y fuman ($F \cap V$) y aquellos estudiantes que no les agrada fumar ni beber ($\bar{F} \cap \bar{V}$).

En nuestro segundo ejemplo las clases fueron también definidas *a priori*, a través de un criterio lógico de selección, donde cada estudiante pudo decidir una de las cuatro posibilidades. Cuando tenemos dos atributos (agrado por fumar y beber), hay cuatro soluciones posibles, si incorporamos un tercer atributo a nuestro diagrama, como por ejemplo el agrado por comer empanadas, estaríamos definiendo ocho clases de casos (2^3), si los atributos fueran cuatro tendríamos dieciséis clases (2^4), y así sucesivamente.

El diagrama de Venn se usa frecuentemente para representar división de poblaciones; existe también otro procedimiento usado a través de un criterio lógico y consiste en la desagregación en el ordenamiento de clases. En este caso asumimos también, a manera de

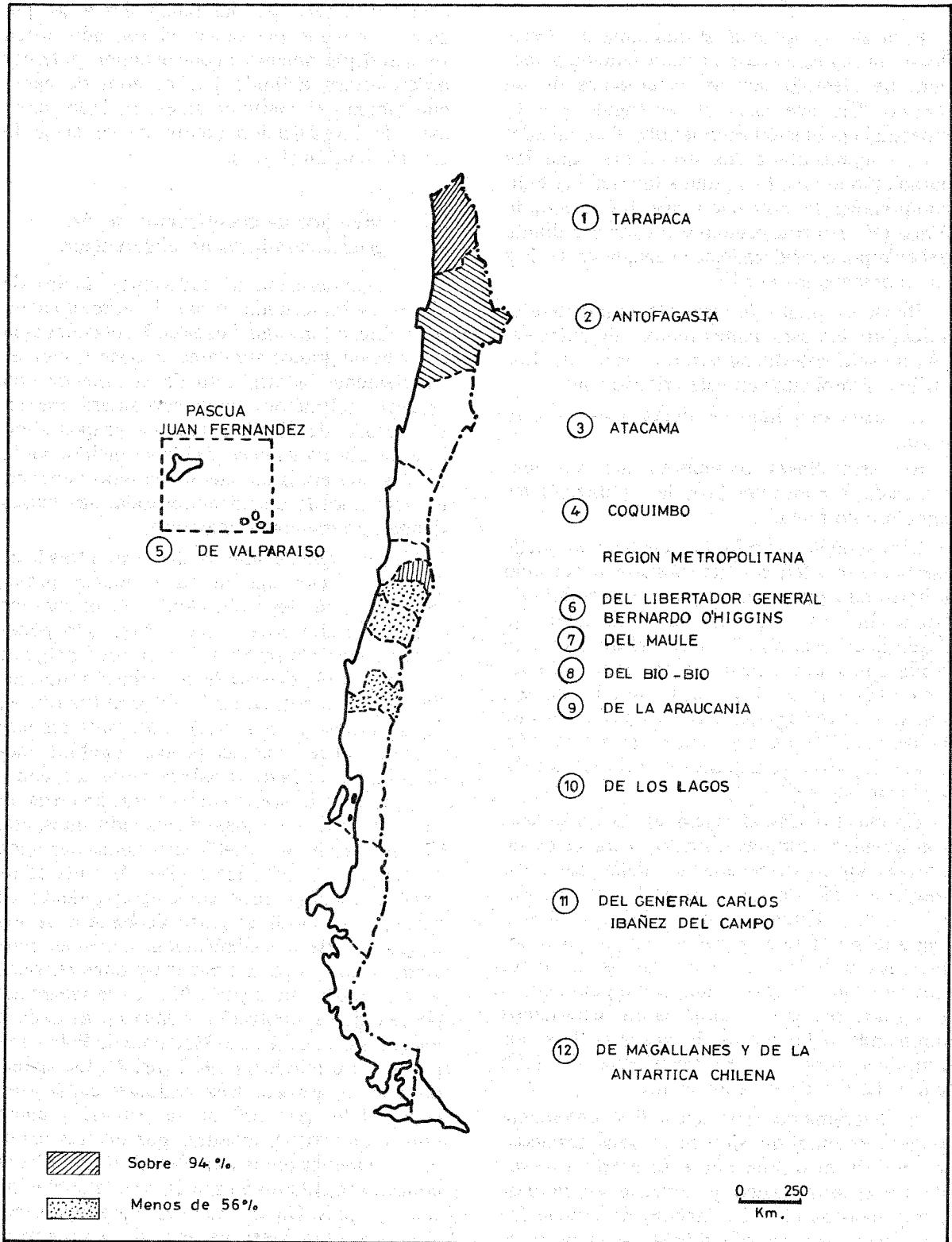
ejemplo, que los estudiantes de una universidad forman la población, el primer orden de clases lo constituyen las facultades a las que cada estudiante pertenece, el segundo orden sería el departamento en que estudian, el tercer orden estará definido por el área de especialización y el cuarto orden estará dado por su área de investigación dentro de su línea de especialización (fig. 3).

2.2. Métodos de clasificación según grupos similares de observación

La aglomeración de individuos dentro de clases es la segunda forma de aproximación para obtener una clasificación. La clasificación a través de grupos similares de observación es simplemente la otra cara de la moneda con respecto del método analizado anteriormente. El método de clasificación por grupos similares de observación es rápido y sencillo, por lo cual se diferencia del método anterior; además en este tipo de clasificación todas las clases tienen a lo menos un miembro.

Una vez que decidimos clasificar a través de la aproximación aglomerativa, nuestro próximo paso será elegir el método por el cual los individuos son comparables, para ello podemos recurrir a los atributos discretos como, por ejemplo, según el sexo de la población (masculino o femenino), según la religión (católicos, protestantes, judíos, etc.). Alternativamente nuestras observaciones pueden también medirse en términos de variables continuas, como es el caso de la preferencia por los ingresos, la salud, educación, trabajo de los individuos, etc. Cuando elegimos variables continuas debemos definir previamente los límites de cada clase antes de iniciar el proceso de agrupamiento. El número de clases elegidas dependerá de los propósitos de la clasificación. Consideremos algunas formas para agrupar las doce regiones y área metropolitana de Chile, en términos del porcentaje de población urbana de acuerdo a los datos del XV Censo Nacional de Población (recuento preliminar) 1982 (fig. 4). De acuerdo con el gráfico, tres regiones concentran sobre el 94 por 100 de su población como población urbana, mientras que las tres regiones con menor porcentaje de población urbana concentran sólo un 56 por 100 de la población total del país. En este caso el mapa representa sólo dos situaciones de acuerdo a un patrón

FIG. 4 : REGIONES CON MAYOR Y MENOR PORCENTAJE DE POBLACION URBANA.



clasificadorio, si introducimos una tercera clasificación (fig. 5), podemos agrupar las regiones de Chile según el mayor, medio y menor porcentaje de población urbana. Con ello mejoramos las posibilidades de análisis en cuanto a la distribución geográfica de las provincias urbanas de Chile.

Es también importante tener en cuenta que todo esfuerzo por clasificar lleva implícito la idea de clarificar. Si con nuestro esfuerzo somos capaces de crear una clasificación donde los mapas representan realmente las relaciones espaciales de los fenómenos en estudio, entonces habremos realizado una buena elección y la clasificación alcanzada servirá a los propósitos de la investigación.

Debemos pensar también que el establecimiento de clasificaciones basadas sobre numerosos atributos o características son también asuntos muy complicados. La mente humana debe dedicar mucho tiempo para tabular y procesar los ítems en términos de numerosos atributos. En consecuencia, cada investigador en sus intentos de clasificación favorece más un aspecto que otro. En este campo, la geografía no es una excepción y en el desarrollo del pensamiento geográfico tenemos el caso que la sola idea de región natural ha sufrido diversas modificaciones en cuanto al concepto de clasificación areal. Sin embargo, con el uso de los computadores se ha abierto todo un campo de posibilidades en esta materia.

III CLASIFICACION NUMERICA

La clasificación numérica es también parte importante de un proceso de reducción de datos siguiendo a través de un método numérico de agrupamiento un criterio de los objetos o individuos. Existen métodos de clasificación numéricos simples hasta los más sofisticados que nos permiten obtener resultados a través de representaciones numéricas ordinales (una variable) o binarias (dos variables). El propósito del presente trabajo se refiere particularmente a la segunda situación, es decir, cuando el investigador dispone de información para dos o más variables. En este caso, el concepto de clasificación engloba un criterio de similitud, por ejemplo los miembros de la clase A son más parecidos entre sí que los miembros de la clase B. Un perro pekinés es más parecido a un maltés que a un

conejo o un gato. Los seres humanos (como cualquier otro animal) están divididos según sus atributos físicos en machos y hembras, puede en ciertas ocasiones darse el caso de un hermafroditismo, es decir, cuando un individuo no puede ser clasificado. Lo mismo ocurre en algunas observaciones que resultan inclasificables.

Una clasificación numérica obtiene por la contigüidad o la unión de los puntos que representan en un diagrama disperso las unidades de observación. Los ejes del diagrama son las características o variables que servirán para representar las ciudades a clasificar, por ejemplo el diagrama de la figura 6 muestra un grupo de unidades (individuos) representados por puntos en un diagrama disperso. El eje vertical corresponde al peso y el eje horizontal al tamaño de los zapatos del grupo de individuos. El grupo A son los niños y el grupo B son soldados. La distancia o contigüidad de los pares de puntos es una forma de medir su similitud en términos de las características consideradas (peso y tamaño de los pies). Así el punto N_1 y N_2 representan un grupo similar (en términos de peso y tamaño de los zapatos), mientras que los puntos N_1 y N_2 son disímil.

En la figura 7 la distancia AB entre los puntos A y B representan dos unidades de observación expresadas por la ecuación:

$$AB^2 = AC^2 + BC^2 = \\ = (X_A - X_B)^2 + (Y_A - Y_B)^2$$

(Diferencia del eje de X al cuadrado + la diferencia del eje de Y al cuadrado y que corresponde al teorema de Pitágoras.)

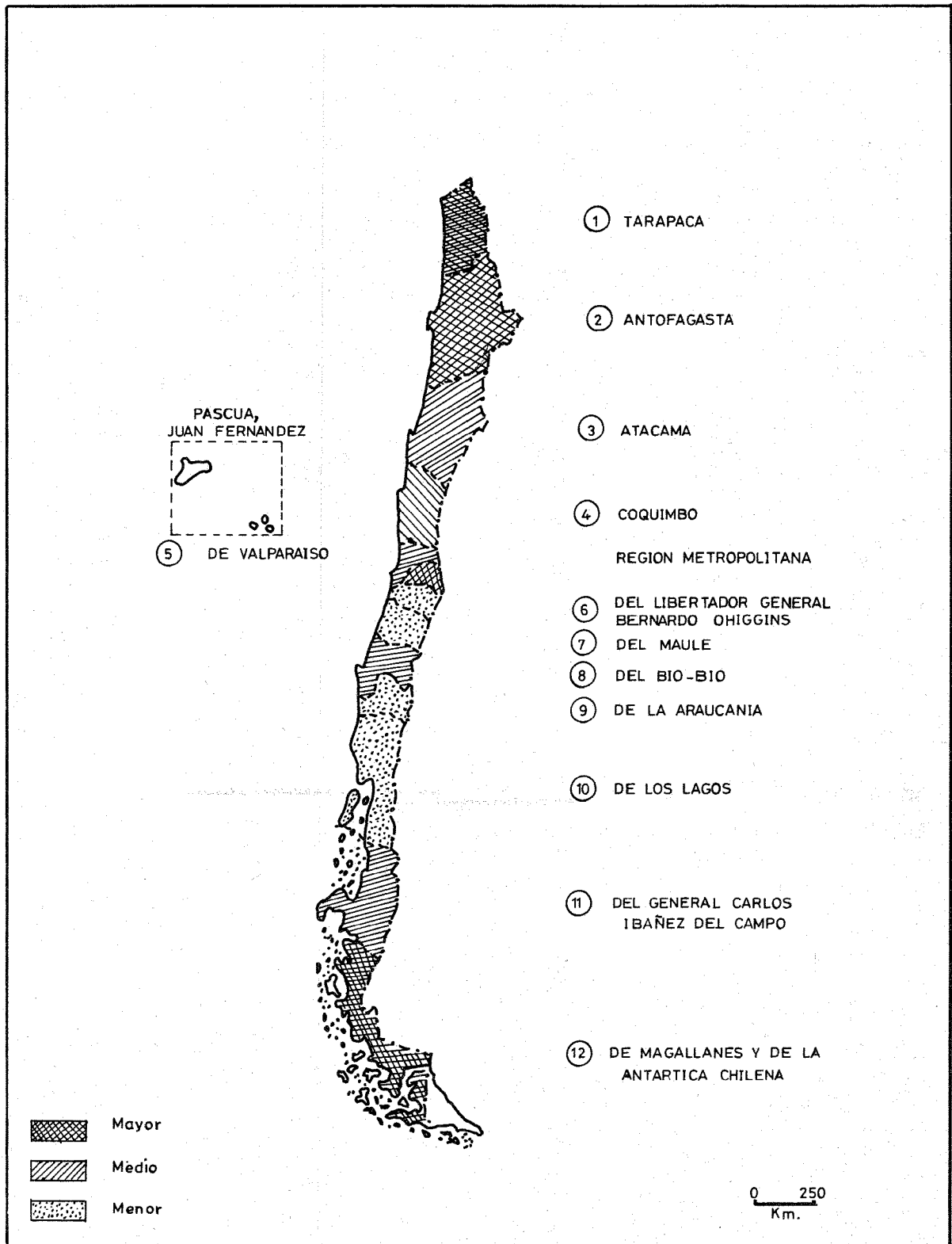
Si trasladamos esta situación al Análisis de Componentes Principales definiremos el ángulo formado por cualquiera de los dos ejes usados para representar las unidades de observación como la correlación entre los pares de variables.

En ciertos casos el coseno $\Theta_{xy} = r_{xy}$, donde Θ_{xy} es el ángulo que separa los dos ejes y r_{xy} es la correlación entre las dos variables o características representadas por los ejes. Si volvemos al ejemplo anterior, tendríamos que x =peso e y =tamaño de los zapatos. En este caso no podríamos sorprendernos de obtener un $r=0,83$, donde $\cos^{-1} 0,83 = 33,9^\circ$. La figura 8 representa un ángulo de $33,9^\circ$.

De acuerdo con el teorema de Pitágoras las variables o caracteres sobre las cuales se basa

FIG. 5 :

REGIONES DE CHILE CON MAYOR, MEDIO Y MENOR PORCENTAJE DE POBLACION URBANA.



la clasificación nos entregan la respuesta correcta, se debe a que el ángulo formado por el triángulo AOB no forman un ángulo recto.

El propósito fundamental del Análisis de Componentes Principales es entregar una posición ortogonal o de sus ejes principales o componentes. Así, los «scores» de las unidades de observación en k componentes principales se usarán normalmente en casos de una clasificación numérica que reemplace las observaciones originales en p variables correlacionadas. Esto nos permite usar el teorema de Pitágoras para el cálculo de distancia en un espacio abstracto k -dimensional, definido por los ejes principales. Estas distancias son una medida de la disimilaridad de cada par de puntos que representan las unidades de observación.

Las clasificaciones pueden ser *jerárquicas* o *nuclear*. En un caso formal, los datos se ordenan de tal forma que cada clase forma parte de una clase mayor y así sucesivamente. Por ejemplo, el perro de Pedro puede ser un poodle, el de Tita un afgano y el perro de Antonio un labrador. Cada perro es miembro de una clase determinada de caninos, pero todos los tipos de caninos considerados en nuestro ejemplo son parte de una clase mayor, los cuadrúpedos, y, a su vez, los cuadrúpedos son animales. Esta misma estructura está representada en forma de un dendrograma (linkage tree), ver figura 9.

3.1. Clasificación jerárquica

El tipo de clasificación jerárquica es la forma más corriente usada para mostrar las relaciones entre organismos. En clasificaciones numéricas los algoritmos matemáticos producen también un dendrograma o diagrama de tipo «ramas de árbol». Dado un dendrograma de la forma que muestra la figura 9, el investigador corta verticalmente el dendrograma de acuerdo a los propósitos de su trabajo, lo que supone determinar los niveles en el ordenamiento jerárquico.

Existen varias formas de producir la estructura de un dendrograma, cada método posee sus características propias. Las cuales serán analizadas en un ejemplo aplicado a las regiones de Chile en el punto 4 del trabajo. A continuación nos preocuparemos de especificar las diferencias que existen entre cada método. El ejemplo de la figura 10 muestra el

método más ampliamente conocido como «grupo promedio». Los siete puntos graficados en un plano bi-dimensional representan espacialmente la posición de siete países. Las coordenadas son ortogonales (x, y) , dichas coordenadas escritas como $x_2, y_1, -x_2y_2, -x_1, y_1$ son los puntajes de cada país en los dos primeros componentes.

La distancia entre los puntos es inversamente proporcional a las similitudes entre los países que dichos puntos representan. Los pares de puntos más cercanos entre sí son los puntos 6 y 5. Estos puntos representan el primer grupo que tiene como centro en punto 5'. La distancia desde 6 a 5 tiene d_1 unidades, esta relación puede representarse como:

$$5 \quad 6 \quad d_1$$

lo que significa que los puntos 5 y 6 se unen a distancia d_1 . A partir de este momento los puntos 5 y 6 desaparecen del diagrama y se reemplazan por el punto 5'. Las distancias desde 5' a los puntos 1, 2, 3, 4 y 7 se calcula y se agrega a la lista de distancias, mientras las distancias desde los puntos 5 y 6 a los puntos 1, 2, 3, 4 y 7 se suprimen. El segundo par de puntos de la nueva lista en unirse es 5' y 7, el centro del nuevo grupo formado por los puntos 5, 6 y 7 es ahora 5'' (ver fig. 10). La distancia entre 5' y 7 es $2/3$ de la distancia de 7 a 5, por cuanto el punto 5 representa dos países, es decir, tiene un peso de 2, mientras el punto 7 tiene sólo un peso de 1 (fig. 11).

La distancia (5', 7) se divide en la proporción 1 : 2.

Este simple pero tedioso procedimiento se repite hasta que todos los puntos se unan en un solo grupo.

La secuencia para cada paso es:

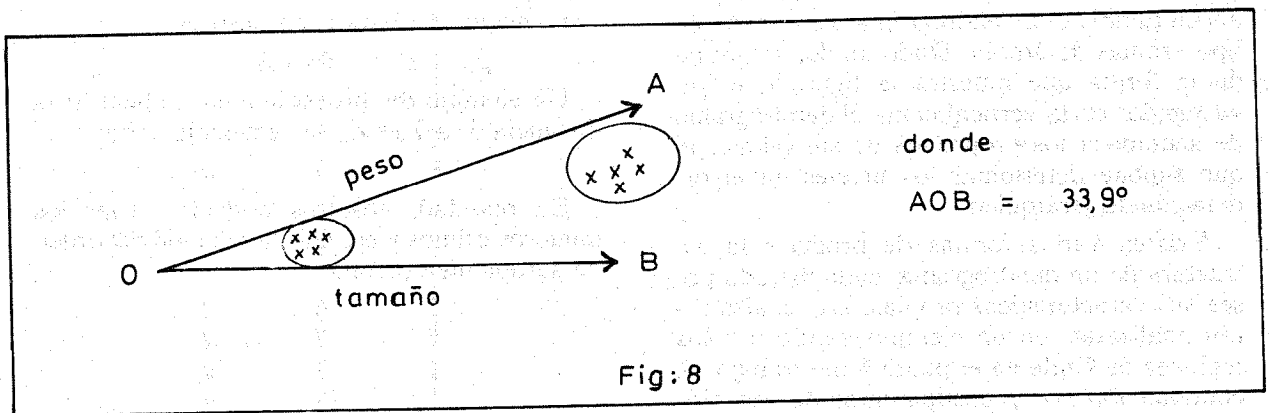
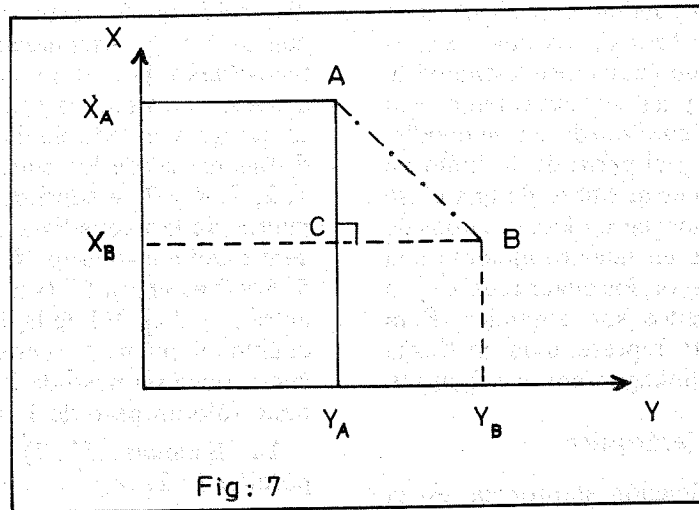
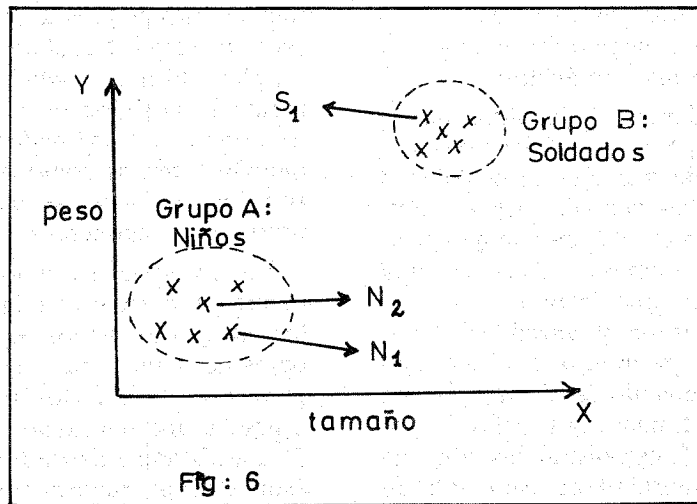
$$p_1 \quad p_1 \quad dp \quad 1p_2$$

Un ejemplo del procedimiento indica: Si la distancia 5' a 7 es d_2 la secuencia sería:

$$5' \quad 7 \quad d_2$$

En realidad, nosotros podemos dejar los números primos y escribir la relación del orden de agrupamiento como:

$$\begin{array}{ccc} 5 & 6 & d_1 \\ 5 & 7 & d_2 \\ 1 & 2 & d_3 \\ 1 & 3 & d_4 \\ 1 & 4 & d_5 \\ 1 & 5 & d_6 \end{array}$$



hasta completar los 6 pasos para $n = 7$ unidades de observación. La figura 12 muestra el mismo procedimiento representado gráficamente.

Es posible cortar el dendrograma con una línea vertical, de esta forma reconocemos que los países de Ecuador, Bolivia y Paraguay forman un solo grupo y los otros países: Brasil, Venezuela, Chile y Argentina corresponden a un segundo grupo de países.

3.2. Clasificación nuclear

Con respecto a la clasificación nuclear, dicho método no compromete cualquier idea de jerarquización o relaciones entre grupos. Sino que cada grupo es una entidad separada y distinta de la misma, como lo son las categorías de uso de suelo en un mapa. Más aún, la clasificación de tipo nuclear depende de la escala de estudio como también del nivel de relaciones, es decir, si éste es jerárquico o no. De tal forma que si consideramos a manera de ejemplo una clasificación de uso de la tierra podemos considerar la siguiente representación (ver fig. 13). Si queremos subdividir un área en regiones con características similares, ejemplo: bosque, suelos cultivables, urbanos, etcétera, sin comprometer la idea de jerarquización, nosotros usaremos el método de clasificación nuclear.

La clasificación nuclear es mucho más simple en algoritmos matemáticos que el método jerárquico descrito anteriormente. El concepto de distancia se usa como una medida de disimilitud, pero en lugar de calcular la distancia entre todos los n puntos, las distancias entre los n puntos y m grupos seleccionados se obtienen reduciendo el número de distancias desde:

$$\frac{n(n-1)}{2} \text{ a } nm.$$

Cada punto está asignado al grupo cuyo centro está más unido a él, tal como lo muestra el diagrama A de la figura 14.

La segunda representación es para encontrar el centro medio de cada uno de los grupos. Estos centros medios aparecen indicados por estrellas (★) en el diagrama A de la figura 14. Puede ocurrir que algunos puntos estén más

cercanos al centro en un grupo que en otro; centro al cual pertenecen dichos puntos. Así estos puntos se mueven en el grupo cuyo centro está más cercano (ver diagrama B de la figura 14).

La forma de calcular la posición de los grupos centros y el movimiento de los puntos para agrupar aquellos centros más cercanos continúa hasta que el número de movimientos es menor o igual a un valor especificado. Es posible realizar este proceso inicial con diferentes números de centros. Por ejemplo, sería muy interesante probar el procedimiento con $m = 10, 9, 8 \dots, 4$, cuando no conocemos cuántos grupos hay.

El éxito de cualquier grupo está en la comparación entre el número de grupos y la densidad de los grupos. De tal forma que sería más apropiado seleccionar el menor número posible de grupos compactos. En el diagrama C de la figura 15 tenemos tres tipos de grupos, de los cuales dos de ellos muestran un agrupamiento disperso, más de tres grupos el agrupamiento es espúreo.

La composición de un grupo está medida por la distancia promedio desde los puntos de cada grupo y su centro. Para eliminar los efectos positivos y negativos de las distancias se usa la raíz media al cuadrado, que se calcula como:

$$\left(\frac{\xi d^2}{n_p} \right)^{1/2}$$

donde n_p es el número de puntos en cada grupo. La distancia promedio de la raíz media al cuadrado para todos los m grupos es un indicador del grado de consistencia de la solución. De tal forma que, si m varía, el diagrama puede representarse como la relación que existe entre el promedio de la raíz media al cuadrado y m .

Esta situación está graficada en la letra c de la figura 15. Hay un considerable aumento en la composición cuando m va de 1 a 2 y 2 a 3, aunque el mayor beneficio se establece cuando m es mayor que 3.

Cuando nosotros sabemos cuantos grupos existen, debemos escoger el grupo centro inicial usando números aleatorios que son dados por medio de un programa de computación. Los ejemplos en que se usa el método nuclear para llegar a una clasificación incluyen nu-

Similaridad decreciente

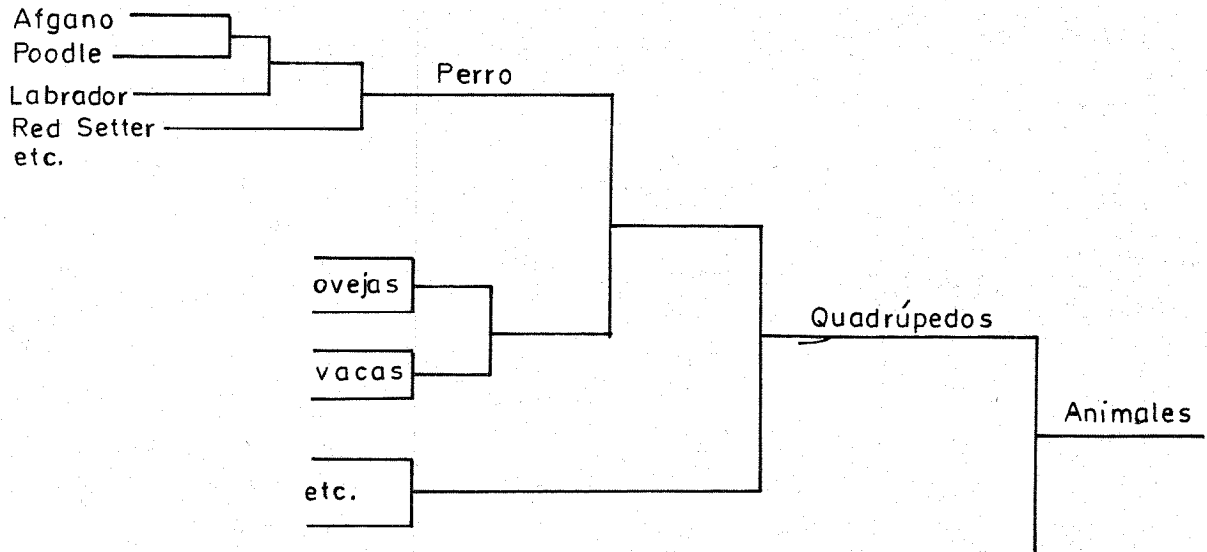


Fig: 9

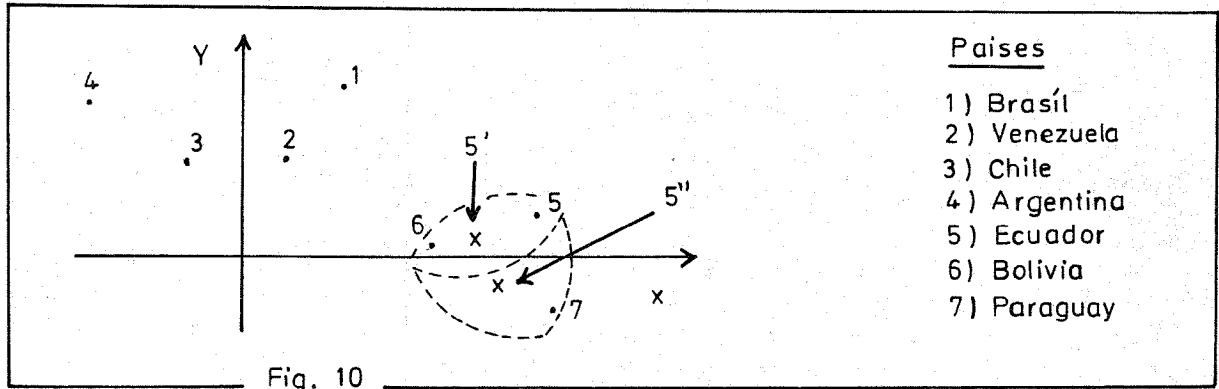
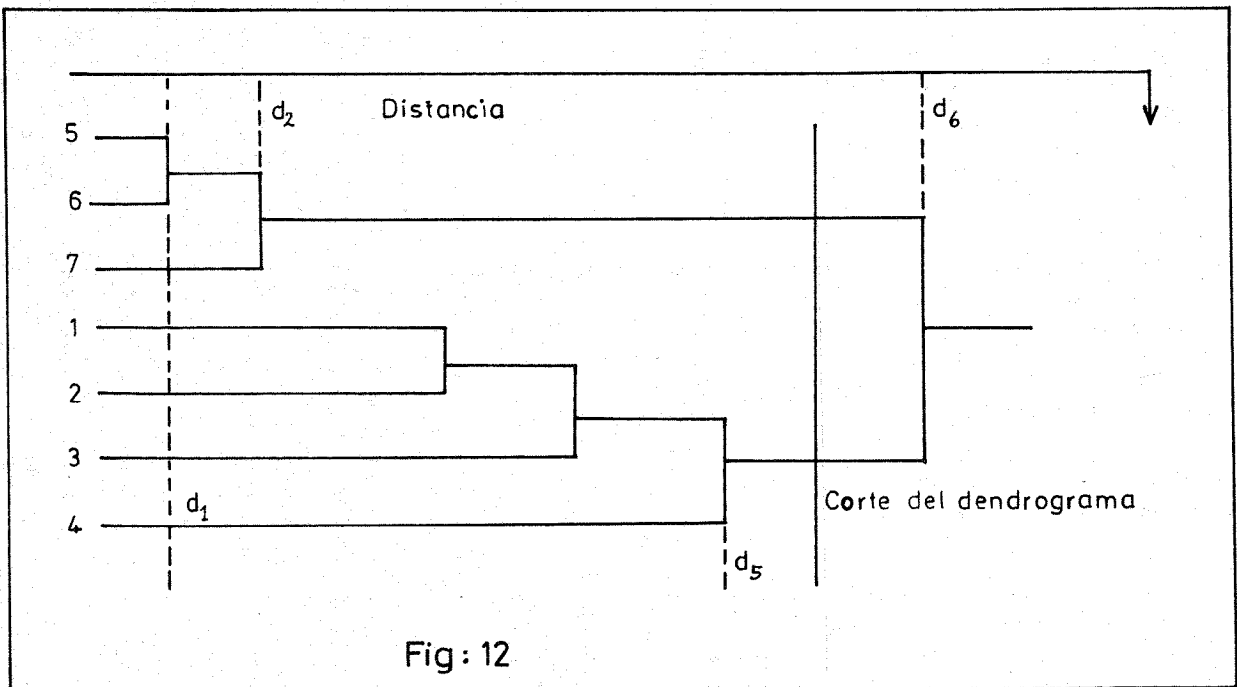
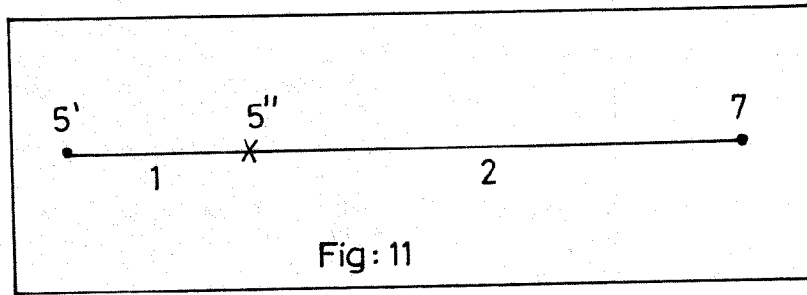


Fig. 10



merosos tópicos geográficos tanto físicos como humanos. Todo depende finalmente del punto de vista del autor y de las limitaciones de tiempo para usar el computador.

IV EJEMPLOS DE CLASIFICACION JERARQUICA APLICADA A LAS REGIONES DE CHILE

Tal como lo vimos en el punto tres del trabajo, el método de agrupamiento permite combinar tipos de similaridad de acuerdo a las distintas propiedades originales del espacio. El conocimiento de las características de cada método permitirá a los lectores decidir qué tipo de algoritmo es el más apropiado a sus intereses de clasificación.

La estrategia general que sirve de base a cada uno de los métodos de agrupamiento es similar y puede representarse de la siguiente forma: los pares de puntos más próximos i y j (los cuales representan simples objetos o grupos de objetos) en un espacio Euclidean p dimensional, están combinados dentro de un grupo simple. La distancia de todos los otros puntos de este grupo reemplazan sus distancias en i y j . El proceso se repite hasta que todos los puntos hayan sido incorporados dentro de un simple grupo. Los algoritmos usados dependen de los diferentes nombres dados por los autores para alcanzar sus propósitos en un proceso clasificatorio. La relación usada por el programa de computación es de la forma:

$$d_{km} = \alpha_i d_{im} + \alpha_j d_{jm} + \beta d_{ij} + \gamma |d_{im} - d_{jm}|$$

donde m es uno de los puntos restantes en el Espacio Euclidiano, los valores de d_{im} , d_{jm} son valores conocidos. Las diferentes combinaciones de agrupamiento están determinadas por los algoritmos que asignan distintos valores α_i , α_j , β y γ . Cuando $\gamma = 0$, entonces la fila de los valores mínimos de d_{ij} está determinada para cada uno de los $(n-1)$ el proceso de agrupamiento es repetitivo, en consecuencia no es posible realizar el agrupamiento en el dendrograma. Los valores de: $\alpha_i + \alpha_j + \beta \geq 1$. Y corresponden a los llamados: Vecino más cercano, Vecino más lejano, Centroide, Mediana, Simple promedio, Grupo promedio, Error Ward's (método Ward).

La tabla 1 muestra los puntajes (score) usando Análisis de Componentes Principales

en un estudio realizado con ocho variables para las doce regiones de Chile más su área metropolitana. A partir de la tabla 1 se usó el Hierarchical Cluster Analysis del doctor P. Mather, que proporciona la matriz de distancia, la consecuencia de pares para cada región y el dendrograma para cada uno de los algoritmos más usados en la clasificación jerárquica.

TABLA 1

Puntaje (score) de los componentes principales para las regiones de Chile usando 8 variables

Región	1	2	3
1	-0,5782	1,0137	1,1587
2	-0,7423	0,3171	0,6024
3	-0,3597	0,2555	1,3099
4	0,8058	-0,4851	0,7947
5	-0,5400	-0,4758	0,9078
6	0,8197	0,1944	0,8204
7	0,9401	-0,4781	-0,7232
8	-0,1422	-0,6833	0,4240
9	1,2847	-1,1659	-1,2117
10	0,9299	-0,5592	-0,6636
11	0,5971	1,2147	-0,4217
12	-0,4947	2,3102	-1,7889
13	-2,5203	-1,4584	-1,2087

4.1. Método del Vecino más cercano

Es una de las técnicas más antiguas de clasificación y se le conoce también como método de agrupamiento simple. La distancia entre dos grupos (i) y (j) está definida como la distancia entre sus dos miembros más cercanos. El agrupamiento entre objetos y grupos o entre grupos está establecido por la simple unión entre pares de puntos. La figura 16 muestra la matriz de distancia o de similaridad y la tabla 2 contiene la secuencia de pares obtenidos a través del programa de computación.

Finalmente la Figura 17 representa el dendrograma para cada una de las 12 regiones de Chile y su Area Metropolitana. El dendrograma puede cortarse según los propósitos del autor o de acuerdo al criterio de similaridad que tenga cada grupo que se desee formar.

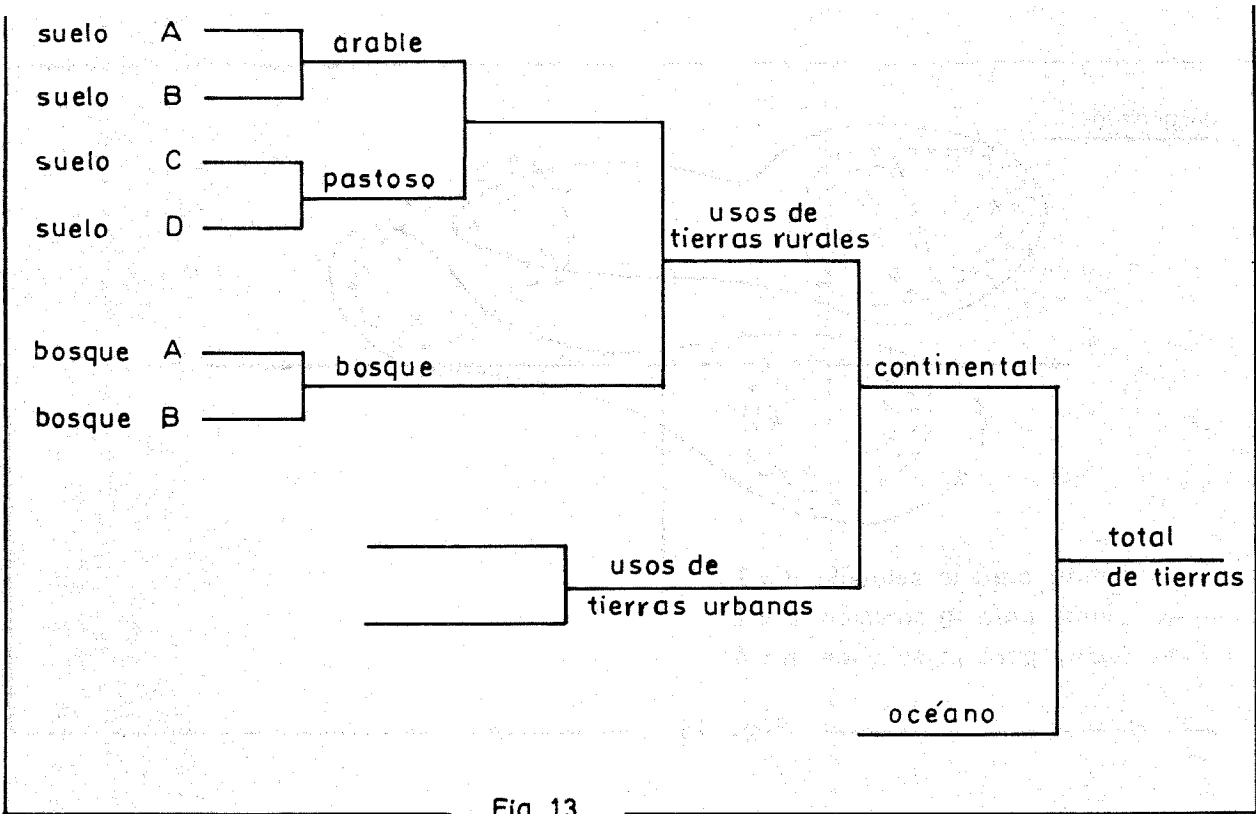


Fig. 13

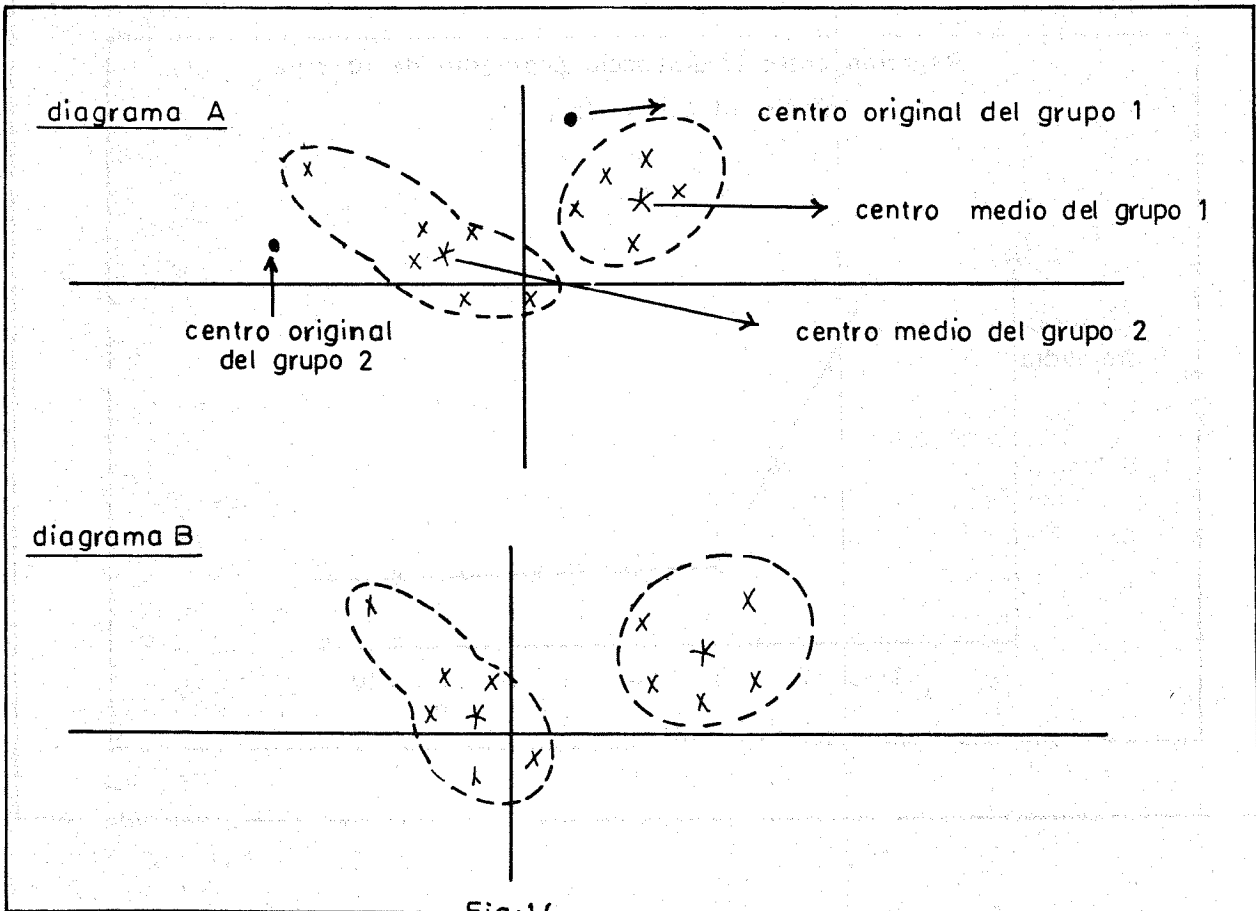
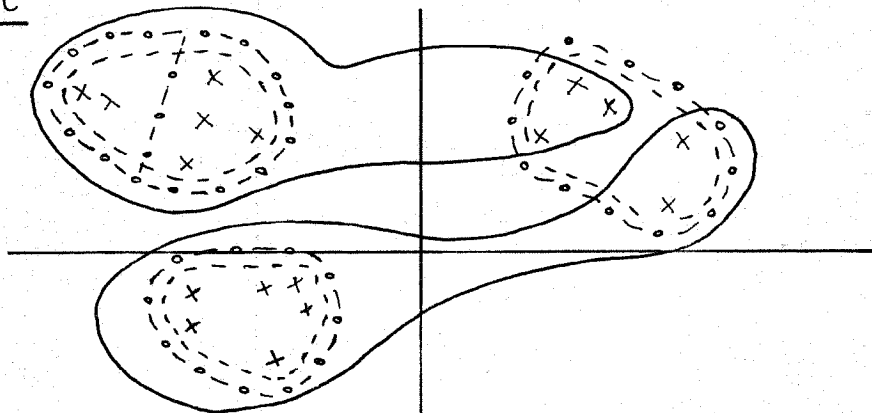


Fig:14

diagrama C



- Límite para la solución $n = 3$
- Límite para la solución $n = 2$
- o-o Límite para la solución $n = 4$

Fig: 15

Relación entre la distancia promedio de la raíz media al cuadrado y m .

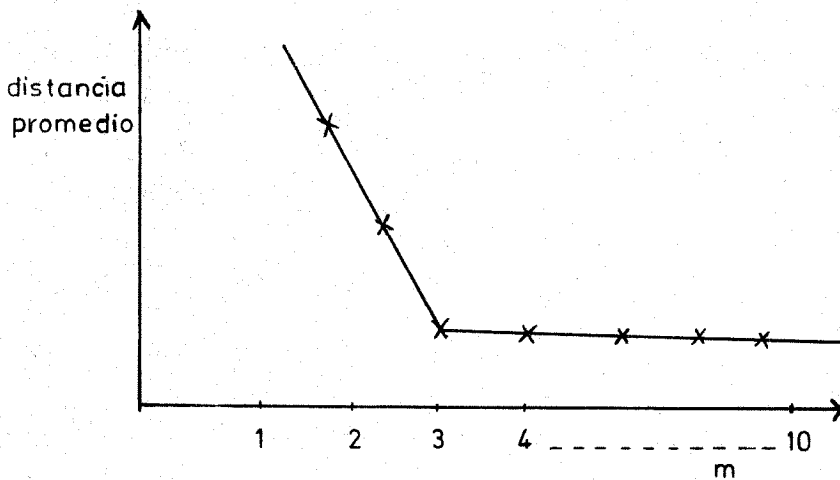


Fig: 16

FIG. 16 MATRIZ DE DISTANCIA

	1	2	3	4	5	6	7	8
1	0.0000							
2	0.7446	0.0000						
3	0.6130	0.5299	0.0000					
4	1.2735	1.1675	0.9527	0.0000				
5	0.8476	0.6368	0.5252	0.9479	0.0000			
6	1.1448	1.4930	1.2061	0.9160	1.2677	0.0000		
7	1.5319	1.5369	1.3841	0.9062	1.2892	0.9198	0.0000	
8	1.0327	0.8239	0.7944	0.8183	0.7109	1.0644	1.1000	0.0000
9	2.0087	1.7383	1.7268	1.0104	1.6213	1.4993	0.8676	1.2700
10	1.5224	1.3985	1.3032	0.7024	1.2065	1.0030	0.3606	0.9843
11	1.1289	1.2194	1.2355	1.0190	1.3652	1.0609	1.1181	1.2290
12	1.4419	1.4935	1.6871	2.0085	1.8308	1.9129	1.8026	1.8315
13	2.0682	1.8024	2.0634	2.4707	1.7314	2.5744	2.4125	1.8358

	9	10	11	12	13
9	0.0000				
10	0.6259	0.0000			
11	1.3973	1.0322	0.0000		
12	2.1550	1.8007	1.1901	0.0000	
13	2.6157	2.3862	2.5315	2.3690	0.0000

Tabla 2: Secuencia de pares obtenidos

REG.	REG.	DISTANCIA
7	10	0.361
3	5	0.525
2	3	0.530
1	2	0.613
7	9	0.626
4	7	0.702
1	8	0.711
1	4	0.818
1	6	0.916
1	11	1.032
1	12	1.190
1	13	1.731

COPHENETIC CORRELATION COEFFICIENT = 0.84609

FIG. 17 DENDOGRAMA DE CHILE

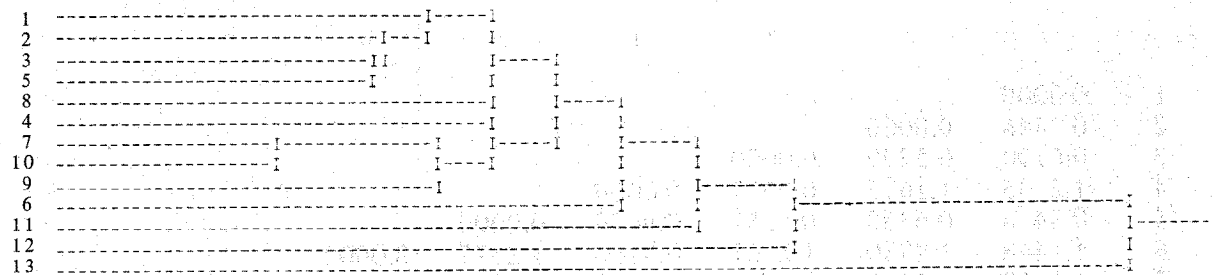


FIG. 18 DENDOGRAMA DE LAS 12 REGIONES DE CHILE

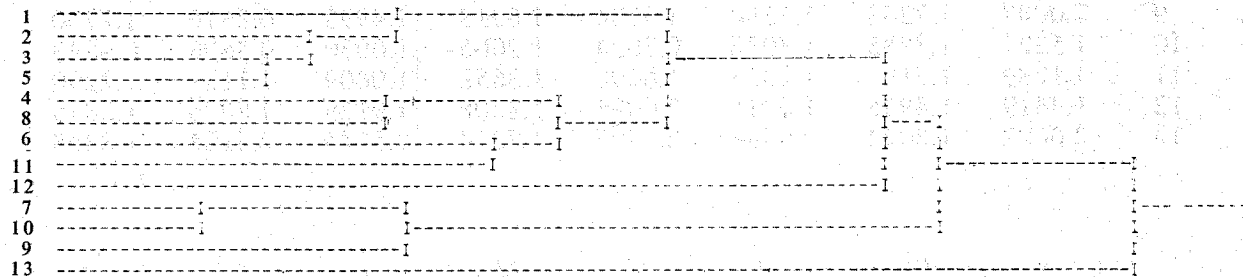


FIG. 19 DENDOGRAMA DE LAS REGIONES DE CHILE

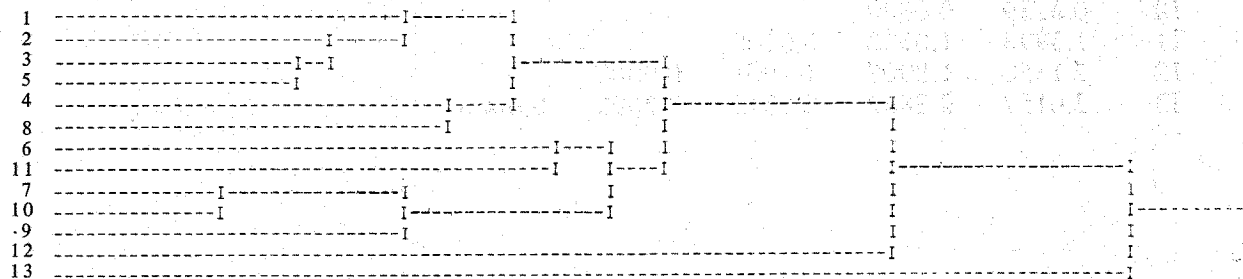


FIG. 20 DENDOGRAMA PARA LAS REGIONES DE CHILE

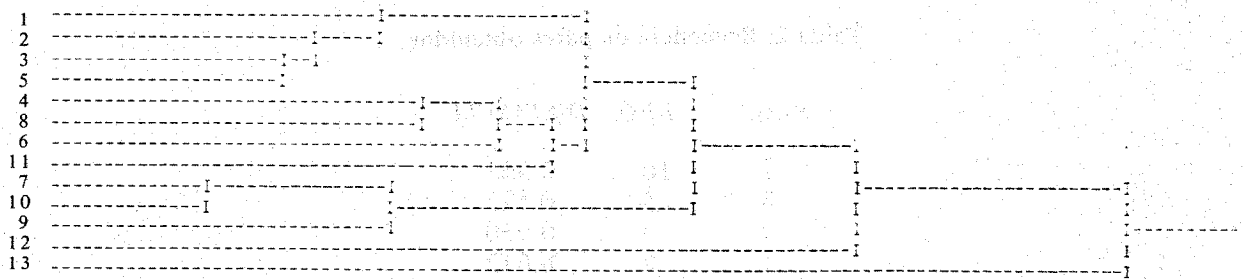


FIG. 21 DENDOGRAMA PARA LAS 12 REGIONES DE CHILE

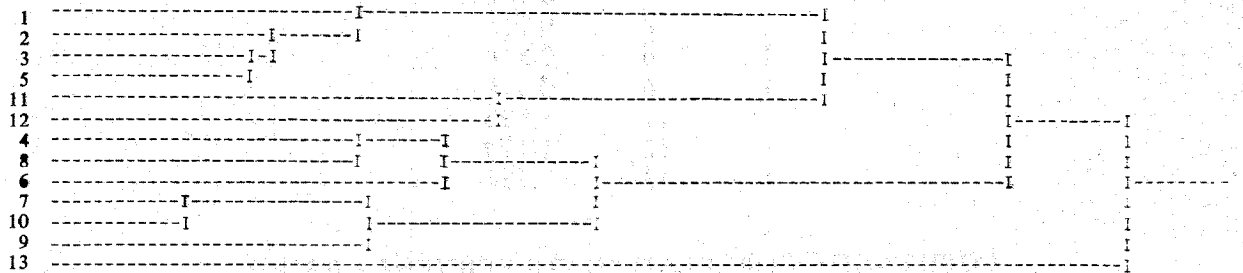


Tabla 3: Secuencia de pares obtenidos

REG.	REG.	DISTANCIA
7	10	0.361
3	5	0.525
2	3	0.637
4	8	0.818
1	2	0.848
7	9	0.868
6	11	1.061
4	6	1.229
1	4	1.493
1	12	2.009
1	7	2.155
1	13	2.616

COPHENETIC CORRELATION COEFFICIENT = 0.77727

Tabla 4: Secuencia de pares obtenidos

REG.	REG.	DISTANCIA
7	10	0.361
3	5	0.525
2	3	0.524
1	2	0.665
7	9	0.735
1	8	0.747
4	7	0.798
4	6	1.005
4	11	0.895
1	4	1.062
1	12	1.579
1	13	2.087

COPHENETIC CORRELATION COEFFICIENT = 0.87583

*** I = 3

0.5236520171

0.5251600146

LINKAGE TREE CANNOT BE DRAWN - BACKWARD LINKS PRESENT

Tabla 5: Secuencia de pares obtenidos

<i>REG.</i>	<i>REG.</i>	<i>DISTANCIA</i>
7	10	0.361
3	5	0.525
2	3	0.524
1	2	0.669
7	9	0.735
4	8	0.818
4	7	0.881
4	6	0.966
4	11	0.934
1	4	1.009
1	12	1.388
1	13	2.139

COPHENETIC CORRELATION COEFFICIENT = 0.86904

*** I = 3

0.5236520171

0.5251600146

LINKAGE TREE CANNOT BE DRAWN - BACKWARD LINKS PRESENT

Tabla 6: Secuencia de pares obtenidos

<i>REG.</i>	<i>REG.</i>	<i>DISTANCIA</i>
7	10	0.361
3	5	0.525
2	3	0.583
1	2	0.735
7	9	0.747
4	8	0.818
1	4	0.963
6	11	1.061
6	7	1.167
1	6	1.285
1	12	1.741
1	13	2.238

COPHENETIC CORRELATION COEFFICIENT = 0.88789

4.2. Método del Vecino más lejano:

Es el proceso recíproco de (i). La mayor distancia entre cada par de grupos está definida por la mayor distancia entre sus miembros, ver Tabla 3. Se le conoce también con el nombre de Agrupamiento Completo.

En ambos métodos (Vecino más cercano y Vecino más lejano), las propiedades del espacio sufren una deformación opuesta lo que ha motivado a los autores a crear algunos métodos alternativos los cuales se basan en el promedio de similitud o disimilitud entre los grupos medios. La Figura 18, muestra el Dendograma para las Regiones de Chile siguiendo el criterio del Vecino más lejano.

4.3. Método Centroide:

En este método la distancia entre los grupos se define por la distancia entre el centro (de gravedad) de los grupos. En este caso el primer grupo se une a la misma distancia que en el caso del Vecino más cercano, luego la distancia entre el primer y segundo grupo debe ordenarse a partir del centro de gravedad del primer grupo y su distancia con el segundo.

Cuando no se cumple la condición:

$$\alpha_i + \alpha_j + \beta \geq 1$$

los valores mínimos se hacen repetitivos, en consecuencia no es posible que el programa de computación construya el dendograma. Tal como ocurre en nuestro ejemplo de las 12 Regiones de Chile. La Tabla 4, muestra la distancia en que se unen las 12 Regiones de Chile.

4.4. Método de la Mediana:

Tiene mucha relación con el el método Centroide. En este caso, es importante precisar el peso que se le otorga a cada grupo, pues si el peso difiere considerablemente entre dos grupos, el que tenga un mayor peso anulará las características del grupo con menor peso. Para evitar esta situación, Gower (1969) introdujo

una modificación otorgándoles a (i) y (j) igual peso.

La Tabla 5, muestra la forma en que cada par de región se unen. Al igual que en el caso del método centroide no es posible obtener el dendograma.

4.5. Método del Grupo Promedio:

Es tal vez, el método más ampliamente usado por la geografía para clasificar grupos de observación. La similitud o disimilitud entre cada grupo está definida como el promedio aritmético de las similitudes entre cada par de miembros de (i) y de (j). La Tabla 6 muestra la secuencia de pares según el grupo promedio y la Figura 19, representa el dendograma para las regiones de Chile.

4.6. Método del Promedio simple:

Este método tiene tanta relación con el grupo promedio, como el método de la Mediana lo tiene con el centroide. En este caso, el peso de cada par de grupos se obtiene usando los promedios aritméticos. Los grupos que se unen tienen el mismo peso respectivamente, esto significa los mismos valores para n_i y n_j . Tal como se observa en la Tabla 7. El método de promedio simple contribuye a mantener las características de los grupos más pequeños. La Figura 20, representa el dendograma para las regiones de Chile siguiendo el agrupamiento de promedio simple.

4.7. Método Error Ward's:

Este método propuesto por Ward (1963) permite minimizar las combinaciones dentro de cada grupo. Según Ward, dos grupos que se combinen en cualquier nivel, su unión produce a los menos un aumento en la suma de los cuadrados del grupo. Este método se conoce también con el nombre de método Ward. La Tabla 8 nos muestra las secuencias de pares para cada región con respectivas y distancia. Mientras que la Figura 21 representa el dendograma obtenido para las regiones de Chile.

Tabla 7: Secuencia de pares obtenidos

<i>REG.</i>	<i>REG.</i>	<i>DISTANCIA</i>
7	10	0.361
3	5	0.525
2	3	0.583
1	2	0.737
7	9	0.747
4	8	0.818
4	6	0.990
4	11	1.100
1	4	1.171
1	7	1.422
1	12	1.761
1	13	2.362

COPHENETIC CORRELATION COEFFICIENT = 0.87924

Tabla 8: Secuencia de pares obtenidos

<i>REG.</i>	<i>REG.</i>	<i>DISTANCIA</i>
7	10	0.361
3	5	0.525
2	3	0.605
1	2	0.815
4	8	0.818
7	9	0.848
4	6	1.045
11	12	1.190
4	7	1.450
1	11	2.043
1	4	2.519
1	13	2.836

COPHENETIC CORRELATION COEFFICIENT = 0.62230

CONCLUSIONES

Los aspectos espaciales de la geografía han recibido gran impulso en todo lo referente a patrones y relaciones, en particular entre los elementos que forman parte de la superficie de la tierra. Gran parte de las distintas formas geométricas tienen también un contexto especial. Los métodos de agrupamiento matemático y el análisis geográfico del espacio se han estudiado intensamente en los últimos años, y con ello la geografía ha ganado mucho en precisión y profundidad.

El desarrollo del análisis estadístico multivariado con el apoyo de los computadores nos permite simplificar los análisis de las complejas interrelaciones de muchas variables, las cuales dan a cada región de un país un aspecto particular. Sin duda, que la geografía regional ha aumentado en valor y precisión en este campo.

Un gran número de métodos matemáticos

pueden usarse para reducir un alto número de variables hasta transformarlas en una familia de variables, cuyos resultados pueden agruparse en unidades matemáticas homogéneas. El método de Análisis de Componentes Principales y el Análisis Factorial se usan, hoy día, para reducir un alto número de variables en una pequeña familia de variables llamadas componentes o factores, de acuerdo al método que se siga. A partir de los «scores» es posible clasificar en términos de similaridad de acuerdo a los propósitos del investigador.

Para el caso de nuestro estudio aplicado a las 12 regiones de Chile y su área metropolitana, se siguió un criterio de similaridad, que varía de acuerdo al algoritmo matemático del programa que se use. De tal forma que el dendrograma adopta distintas formas sin que varíe la importancia de cada región clasificada. Siguiendo este mismo criterio, se puede intentar regionalizar un área que reúna condiciones de similaridad e información adecuada a estos propósitos.

BIBLIOGRAFIA

1. ABLER, R.; ADAMS, J. and COULD, P. (1977): *Spatial Organization*. Prentice-Hall International Editions, London.
2. BERRY, B. J. (1964): Approaches to Regional Analysis: A synthesis. *Annals of the Association of American Geographers*. N.º 54, United States, pp. 2-11.
3. BRADSHAW, R. y ESTEBANEZ, J. (1978): *Técnicas de Cuantificación en Geografía*. T. Flores Editor, Madrid.
4. COLE, J. P. and KING, C. A. M. (1968): *Quantitative Geography*. John Wiley, London.
5. EBDON, D. (1977): *Statistics in Geography. A Practical Approach*. Oxford.
6. HAMMOND, R. and McCULLAGH, P. (1978): *Quantitative Techniques in Geography. An Introduction*. Second Edit., Clarendon Press, London.
7. JOHNSTON, R. J. (1978): *Multivariate Statistical Analysis in Geography*. Longan Edit., London.
8. MATHER, P. M. and DOORNKAMP, J. C. (1970): A Multivariate Analysis in Geography. Transactions Publications. *The Institute of British Geographers*. N.º 15.
9. MATHER, P. M. (1976): *Computational Methods of Multivariate Analysis in Physical Geography*. J. Wiley, London.
10. RUMMEL, R. J. (1970): *Applied Factor Analysis*. Northwestern University Press, United States.
11. SANCHEZ, A. (1981): *Regional Development in Chile from 1940 to 1970 and Future Prospects*. Thesis Degree, University of Nottingham, England.
12. SANCHEZ, A. (1982): *Las Provincias de Chile: Un estudio multivariado usando Análisis de Componentes Principales*. Documento de Trabajo, N.º 5, Universidad de Chile, Santiago.
13. TAYLOR, P. (1977): *Quantitative Methods in Geography, An Introduction to Spatial Analysis*. United States.