

4.	ALGUNOS MÉTODOS DE CLASIFICACIÓN .....	73
4.1.	Dominio de aplicación.....	73
4.2.	Fundamentos de los métodos.....	73
4.2.1.	Clasificación alrededor de centros móviles.....	74
4.2.2.	Clasificación jerárquica ascendente con el método de Ward.....	76
4.2.3.	Una estrategia de clasificación.....	79
4.2.4.	Caracterización de las clases .....	79
4.3.	Un ejemplo de aplicación: Departamentos – Icfes .....	83
4.3.1.	Introducción .....	83
4.3.2.	Escogencia de una partición.....	83
4.3.3.	Consolidación de la Partición mediante el uso del método de centros móviles.....	85
4.3.4.	Caracterización de las Clases .....	85
4.4.	Ejercicio: Test de personalidad con variables discretizadas .....	87
4.4.1.	Presentación .....	87
4.4.2.	Tablas y gráficos .....	88
4.4.3.	Guía para el análisis .....	92
	BIBLIOGRAFÍA.....	93

### TABLAS Y GRAFICOS

Tabla 4-1:	coordenadas de las muestras de café y partición inicial .....	75
Tabla 4-2:	Iteraciones del proceso de centros móviles para el ejemplo .....	75
Gráfico 4-1:	Movimiento de los centros y clasificación final .....	76
Gráfico 4-2:	árbol de clasificación del ejemplo café.....	78
Tabla 4-3:	Indices de nivel del árbol del ejemplo .....	78
Gráfico 4-3:	Arbol de clasificación para el ejemplo café del capítulo 1 .....	80
Tabla 4-4:	Optimización de la partición del ejemplo del café.....	81
Tabla 4-5:	Caracterización de las clases según las variables continuas .....	82
Tabla 4-6:	Caracterización de las clases según el tipo de agregado.....	82
Gráfico 4-4:	Primer plano factorial mostrando las clases .....	83
Tabla 4-7:	Histograma de índice de nivel. Departamentos - Icfes .....	84
Gráfico 4-5:	Árbol de clasificación jerárquica. Departamentos - Icfes.....	84
Tabla 4-8:	Consolidación de la partición. Departamentos - Icfes .....	85
Gráfico 4-6:	Proyección de los Departamentos sobre el primer plano factorial e identificación de las clases .....	86
Tabla 4-9:	Descripción de la partición – Departamentos - Icfes.....	87

---

Tabla 4-10: Clases vs niveles de Icfes.....	87
Tabla 4-11: Histograma de índices de nivel .....	88
Gráfico 4-7: Arbol de Clasificación del ejercicio.....	89
Tabla 4-12: Consolidación de la partición del ejercicio .....	90
Tabla 4-13: Caracterización de las Clases por las variables activas e ilustrativas del ejercicio .....	91

## 4. ALGUNOS MÉTODOS DE CLASIFICACIÓN

Una forma de sintetizar la información contenida en una tabla multidimensional es mediante la conformación y caracterización de grupos. Los grupos o clases se conforman de manera que los elementos dentro de cada grupo sean lo más homogéneos posibles y que, en cambio, los elementos de diferentes grupos sean lo más diferentes posibles.

En este texto se incluyen los fundamentos necesarios para comprender la utilización de los métodos de clasificación como complemento a los métodos factoriales. Este hecho hace reducir el espectro de los métodos de clasificación. Como criterio de homogeneidad entre grupos se utiliza la inercia, como distancia entre individuos las mismas utilizadas en los métodos factoriales. Se seleccionan y combinan los métodos de clasificación que cumplen con las restricciones puestas en este contexto.

### 4.1. Dominio de aplicación

En el análisis de grandes tablas de datos los métodos de clasificación son útiles reduciendo el tamaño de las tablas agrupando elementos por su parecido. Estos métodos son algorítmicos y por lo tanto se pueden aplicar a cualquier tipo de tabla. Aquí nos restringiremos a tablas de individuos o modalidades (en fila) por variables cuantitativas. La aplicación a las otras tablas: tablas de contingencia, tablas de individuos por variables cualitativas, se hace realizando un análisis factorial previo a la clasificación.

Las tablas de los ejemplos utilizados en los métodos factoriales de los capítulos anteriores son también susceptibles de analizar mediante métodos de clasificación. Para ilustrar los métodos se utilizará el ejemplo de las muestras de café del capítulo 1.

### 4.2. Fundamentos de los métodos

Los métodos de clasificación se pueden dividir en jerárquicos y no jerárquicos. En los no jerárquicos el número de clases se establece previamente y el algoritmo de clasificación asigna los individuos a las clases, partiendo de algunos valores iniciales y buscando optimizar algún criterio establecido de antemano.

En la clasificación jerárquica se construye un "árbol" o "dendrograma", (del griego dendron = árbol), cuyas ramas terminales representan a cada uno de los individuos y el tronco es la clase conformada por todos los individuos. Un dendrograma representa una serie de particiones embebidas, en donde el número de clases decrece a medida que se aumenta la altura del árbol. Para obtener alguna clasificación particular se hace "un corte" en el árbol.

Un árbol se puede construir partiendo del tronco (el conjunto de todos los individuos) y haciendo divisiones sucesivas hasta llegar a cada uno de los individuos o partiendo de las ramas terminales (cada uno de los individuos) y haciendo uniones sucesivas hasta llegar a un grupo con todos los individuos. El primer método se denomina "clasificación jerárquica divisiva" y el segundo "clasificación jerárquica aglomerativa" y es el que más se usa en la práctica.

Los métodos de clasificación requieren de una definición de la distancia o un índice de disimilaridad entre los elementos que se van a clasificar. Si las variables son de tipo continuo la distancia más utilizada es la euclidiana canónica:

$$d(i, l) = \sqrt{\sum_j (x_{ij} - x_{lj})^2}$$

la sumatoria se hace sobre todas las variables (subíndice j), el valor entre paréntesis es la diferencia entre los valores de la variable j que asumen los individuos i y l. Dos individuos se parecen si asumen valores parecidos para todas las variables. Las coordenadas de un análisis factorial son de tipo continuo y por lo tanto es válido utilizar la anterior distancia.

Los dos métodos que se combinan para lograr grupos homogéneos desde el punto de vista estadístico son: el método de centros móviles usando la distancia euclidiana y el método de Ward de clasificación jerárquica aglomerativa.

#### 4.2.1. Clasificación alrededor de centros móviles

Para este procedimiento se necesita conocer el número de clases,  $k$  y dar los centros iniciales de cada clase. Se construye una primera partición (clasificación en  $k$  clases disjuntas), asignando cada elemento a clasificar al centro más cercano. Luego se calcula el centro de gravedad de cada clase y se repite el proceso con los centros de gravedad como nuevos puntos, de ahí el nombre de “centros móviles”. El proceso continua hasta que no haya más cambios o hasta que la inercia intraclases cambie en un valor por debajo de un umbral predeterminado.

De forma análoga al análisis de varianza, la inercia total de la nube de puntos (elementos a clasificar) se divide en inercia intraclases e inercia Interclases. Para un partición en  $k$  clases la expresión es:

$$I = \sum_q p_q d^2(\mathbf{g}_q - \mathbf{g}) + \sum_q \sum_{i \in q} p_i d^2(\mathbf{x}_i - \mathbf{g}_q)$$

donde  $\mathbf{g}$  y  $\mathbf{g}_q$  son los centros de gravedad general y de la clase  $q$  respectivamente. Los pesos (fracción de individuos) se indican con  $p$ .

Se cumple el criterio de homogeneidad si la inercia intraclases es mínima. Esto se logra con el procedimiento descrito pero desafortunadamente no se obtiene un óptimo absoluto sino un óptimo que depende de los puntos iniciales (óptimo local), constituyéndose en uno de los defectos de este método.

Para ilustrar el procedimiento retomemos el ejemplo de las muestras de café utilizando las coordenadas sobre los dos primeros factores. Se construirá una partición en dos clases tomando como puntos iniciales (1,0) y (2,0).

La Tabla 4-1 es la que ingresa al proceso de clasificación, cada individuo esta caracterizado por dos variables cuantitativas (las coordenadas factoriales F1, F2). La columna D1 tiene al distancias al cuadrado al primer centro (que aparece más abajo en la misma tabla) y la D2 al segundo centro. Las columnas C1 y C2 son indicadoras de la pertenencia a la clase. El individuo se asigna a la clase 1 si D1 es menor que D2, de lo contrario se asigna a la clase 2. Se incluye la varianza para cada uno de los factores, cuya suma es la inercia total. Abajo aparecen los centros iniciales (entran) y los centros de gravedad de la partición obtenida (salen), se incluyen la inercias Inter-clases, intra-clases y el % de inercia Inter-clases con respecto a la total. Esta cantidad debe aumentar en cada iteración. Con puntos iniciales dados (1,0) y (2,0) se obtiene una clase con 8 muestras y otra con 2 (Claro40%maíz y Oscuro40%maíz). Los nuevos centros son (-0.78, -0.14) y (3.12, 0.58) y el % de inercia Inter. es: 49.4.

La Tabla 4-2 muestra las tres iteraciones de centros móviles para lograr el optimo con los puntos iniciales dados. En la primera las clases son de tamaño 7 y 3, los centros de gravedad son (-1.09, -0.08) y (2.54, 0.20) y el % inercia Inter. es: 54.8; en la segunda los tamaños son 6 y 4, los centros (-1.43, -0.14) y (2.15, 0.20) y 61.2% de inercia Inter. En la tercera iteración no hay cambios y el proceso termina.

El porcentaje de inercia Inter-clases es una medida de la explicación de la variabilidad lograda por la partición.

En el Gráfico 4-1 se muestra el movimiento de los centros y la partición final obtenida. Los centros están identificados con la letra C y dos números, el primero indica la clase y el segundo la iteración; por ejemplo, C10 es el punto inicial para la clase 1 y C13 el punto final. La pertenencia a la clase está representada por un segmento uniendo al individuo con el centro de la clase (C13 o C23).

El problema del óptimo local se soluciona en parte haciendo varias particiones cambiando los puntos iniciales e identificando las clases que permanecen mas o menos fijas en las distintas corridas. Este procedimiento es denominado formas fuertes y puede consultarse entre otros en Lebart (1995).

**Tabla 4-1: coordenadas de las muestras de café y partición inicial**

MUESTRA	COORDENADAS		ITERACIÓN 0			
	F1	F2	D1	D2	C1	C2
ExcelsoClaro	-2.99	1.02	16.96	25.94	1	0
Claro40Maiz	3.21	0.04	4.89	1.47	0	1
Claro40Cebad	1.39	-0.56	0.47	0.69	1	0
Claro20Maiz	0.98	0.22	0.05	1.09	1	0
Claro20Cebad	-0.35	-0.5	2.07	5.77	1	0
ExcelsoOscur	-3.54	-0.09	20.62	30.70	1	0
Oscuro40Maiz	3.02	1.11	5.31	2.27	0	1
Oscuro40Ceba	0.14	-0.85	1.46	4.18	1	0
Oscuro20Maiz	-0.5	-0.91	3.08	7.08	1	0
Oscuro20Ceba	-1.35	0.52	5.79	11.49	1	0
Varianza	4.6	0.472	Frec.cl		8	2

INERCIA TOTAL 5.072

CENTROS -->	Entran		Salen	
	1	0	-0.78	-0.14
2	0	3.12	0.58	

INERCIA -->	Inter	Intra	%inter
	2.51	2.56	49.43

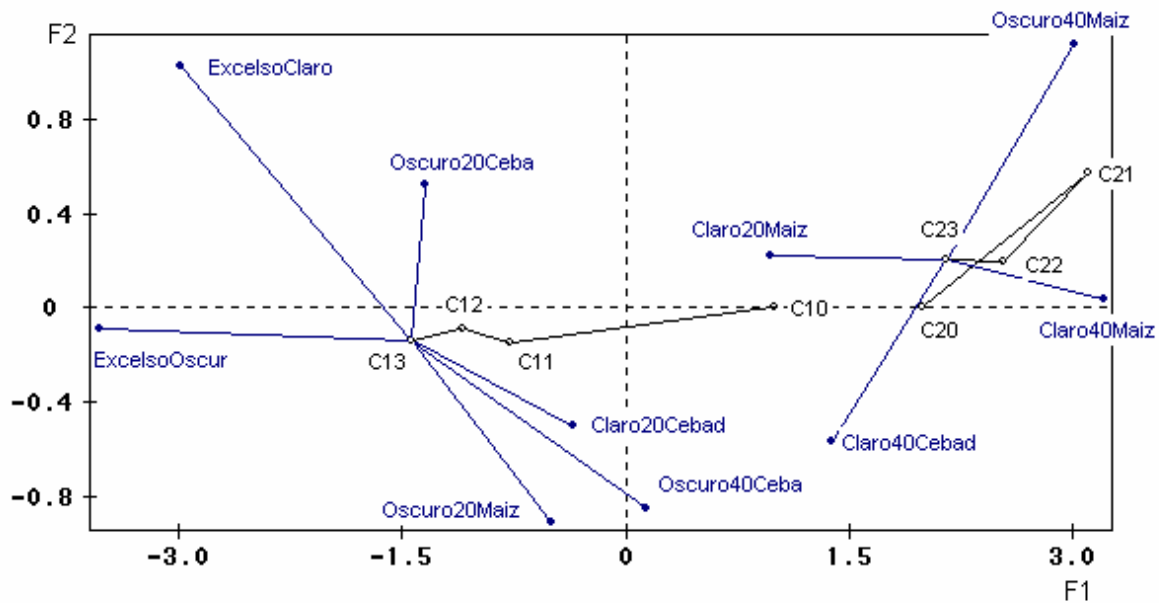
**Tabla 4-2: Iteraciones del proceso de centros móviles para el ejemplo**

MUESTRA	ITERACIÓN 1				ITERACIÓN 2				ITERACIÓN 3			
	D1	D2	C1	C2	D1	D2	C1	C2	D1	D2	C1	C2
ExcelsoClaro	6.25	37.47	1	0	4.84	31.26	1	0	3.76	27.09	1	0
Claro40Maiz	15.93	0.30	0	1	18.48	0.47	0	1	21.58	1.15	0	1
Claro40Cebad	4.87	4.26	0	1	6.36	1.90	0	1	8.14	1.16	0	1
Claro20Maiz	3.22	4.68	1	0	4.37	2.43	0	1	5.94	1.37	0	1
Claro20Cebad	0.31	13.16	1	0	0.72	8.84	1	0	1.30	6.74	1	0
ExcelsoOscur	7.63	44.73	1	0	6.02	37.05	1	0	4.45	32.46	1	0
Oscuro40Maiz	15.99	0.30	0	1	18.29	1.06	0	1	21.37	1.58	0	1
Oscuro40Ceba	1.34	10.88	1	0	2.09	6.86	1	0	2.98	5.15	1	0
Oscuro20Maiz	0.66	15.27	1	0	1.03	10.47	1	0	1.47	8.26	1	0
Oscuro20Ceba	0.77	19.94	1	0	0.43	15.24	1	0	0.44	12.35	1	0
	Frec.cl		7	3	Frec.cl		6	4	Frec.cl		6	4

CENTROS	Entran		Salen		Entran		Salen		Entran		Salen	
		-0.78	-0.14	-1.09	-0.08	-1.09	-0.08	-1.43	-0.14	-1.43	-0.14	-1.43
	3.12	0.58	2.54	0.20	2.54	0.20	2.15	0.20	2.15	0.20	2.15	0.20

INERCIA	Inter	Intra	%inter	Inter	Intra	%inter	Inter	Intra	%inter
		2.78	2.29	54.8	3.11	1.97	61.24	3.11	1.97

Gráfico 4-1: Movimiento de los centros y clasificación final



**4.2.2. Clasificación jerárquica ascendente con el método de Ward**

Los métodos de clasificación jerárquica parten de una matriz de similitudes, disimilitudes o de distancias. Aquí se selecciona la distancia euclidiana canónica. Adicionalmente se requiere la selección de una distancia entre grupos. La distancia entre grupos se denomina criterio de agregación, que además da el nombre al método.

El método de Ward utiliza la distancia entre grupos que cumple con el objetivo de buscar clases que tengan menos inercia intra-clases, como criterio de homogeneidad estadística.

Un procedimiento de clasificación jerárquica ascendente consiste en la construcción sucesiva de particiones del conjunto de elementos que se está clasificando. Se empieza con la partición más fina en donde cada elemento es una clase, es decir que se tienen  $n$  clases, luego se unen los dos elementos más próximos en uno solo. La decisión para saber qué pareja se une primero requiere de la matriz de distancias entre todas las parejas de individuos, de tamaño  $(n,n)$ . Al unir la primera pareja se llega a una partición de  $n-1$  clases una de ellas con dos individuos. Se requiere calcular la distancia entre la nueva clase formada y los individuos restantes, aquí entra a jugar la distancia entre grupos seleccionada. Al unir los dos individuos más próximos se tiene una nueva matriz de distancias de tamaño  $(n-1,n-1)$ , sobre esta matriz se vuelve a seleccionar la pareja más próxima. Se continúa así hasta llegar a una sola clase con todos los individuos. Todo el proceso se puede registrar en un diagrama en forma de árbol denominado dendrograma o árbol de clasificación. Este procedimiento se ilustrará con el método de Ward una vez presentado el criterio de agregación de Ward.

**Distancia de Ward**

La idea del criterio de Ward es unir en cada paso de la clasificación jerárquica los dos grupos que incrementen lo menos posible la inercia intragrupos.

Sean A y B dos grupos o clases no vacías y disjuntas y sean  $p_A, p_B$  y  $\mathbf{g}_A, \mathbf{g}_B$ , los pesos y centros de gravedad de las partes A y B respectivamente. La distancia de Ward entre los dos grupos viene dada por: (la deducción se puede ver en Pardo (1992)).

$$W(A, B) = \frac{p_A p_B}{p_A + p_B} d^2(\mathbf{g}_A - \mathbf{g}_B)$$

Este valor es el incremento de la inercia intra-grupos al unir los grupos A y B en uno solo. En particular para dos individuos  $i$  y  $l$  la distancia de Ward es:

$$W(i, l) = \frac{p_i p_l}{p_i + p_l} d^2(i, l)$$

si los pesos son iguales a  $1/n$  para los dos individuos, la anterior expresión se reduce a:

$$W(i, l) = \frac{1}{2n} d^2(i, l)$$

Es posible calcular la distancia de Ward en un paso, de la construcción del árbol en función de las distancias del paso precedente. Sean A, B y C tres grupos presentes en el mismo paso de construcción del árbol. Si se unen A y B para formar el grupo AB, es necesario calcular la distancia de Ward entre los grupos AB y C. Se conocen las distancias  $W(A,B)$ ,  $W(A,C)$  y  $W(B,C)$ . La distancia  $W(AB,C)$  en función de las anteriores es:

$$W(A \cup B, C) = \frac{(P_A + P_C)W(A, C) + (P_B + P_C)W(B, C) - P_C W(A, B)}{P_A + P_B + P_C}$$

### ***Procedimiento para construir el árbol con el método de Ward***

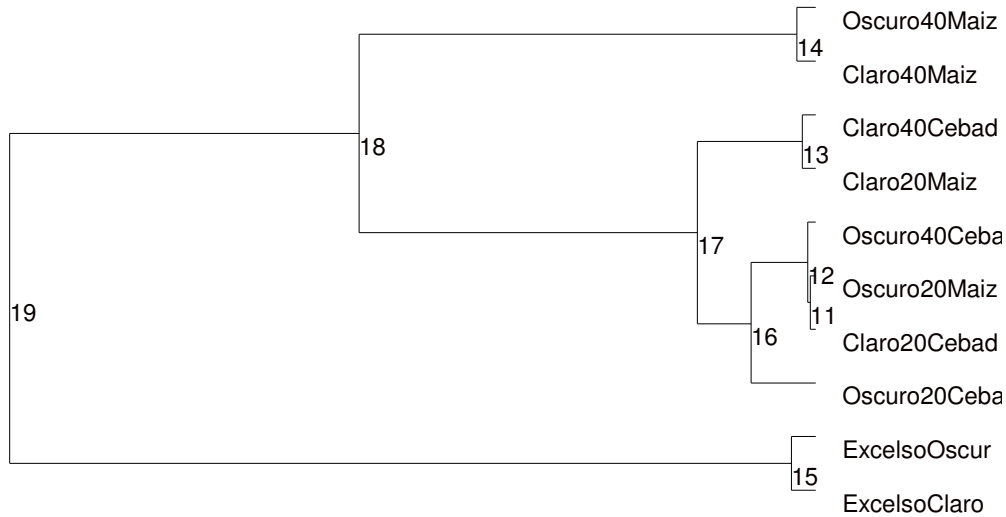
El procedimiento para obtener el árbol de clasificación, utilizando el método de Ward es el siguiente:

1. Calcular las distancias de Ward entre parejas de individuos:  $\frac{1}{2n} d^2(i, l)$ .
2. Seleccionar la pareja de grupos (individuos en el primer paso) que presente la menor distancia de Ward para conformar el nuevo grupo.
3. Calcular las distancias entre todos los grupos y el grupo recién conformado utilizando la fórmula de distancia de Ward o la fórmula de recurrencia.
4. Eliminar las filas y columnas correspondientes a los individuos o grupos unidos y adicionar una fila y una columna para registrar las distancias entre el nuevo grupo y los demás.
5. Repetir el proceso hasta llegar a una sola clase.

El método de Ward se ilustra recurriendo de nuevo al ejemplo de las muestras de café, utilizando como variables las coordenadas sobre los dos primeros ejes factoriales (ver Tabla 4-1) y utilizando la salida del programa SPAD. El Gráfico 4-2 es el árbol de clasificación que resume todo el proceso. Cada unión se denomina nodo, los elementos a clasificar se llaman los nodos terminales. Los nodos aparecen numerados, del 1 al 10 corresponde a los elementos y de 11 en adelante son las uniones- La primera unión es de *Oscuro20Maiz* con *Claro20Cebada* (nodo 11), con un índice de nivel de 0.00953 (Tabla 4-3). Se denomina índice de nivel a la distancia a la que se unen los dos grupos. Esta primera unión corresponde a la distancia de Ward entre los dos individuos, es decir el aumento de la inercia intra-clases al unirlos en una clase. La segunda unión es la del café *Oscuro40Cebada* al grupo anterior a un índice de 0.02268. La tercera unión es la de los cafés *Claro40Cebad* con *Claro20Maiz*, a un índice de 0.03883. La última unión

(nodo 19) corresponde al nodo 18 (8 cafés) con el nodo 15 (2 muestras), con un incremento de la inercia intra de 2.72075.

**Gráfico 4-2: árbol de clasificación del ejemplo café**



**Tabla 4-3: Índices de nivel del árbol del ejemplo**

NUM.	PRIM	BENJ	FRE.	INDICE	HISTOGRAMA DE INDICES DE NIVEL
11	9	5	2	0.00953	*
12	8	11	3	0.02268	*
13	3	4	2	0.03883	**
14	7	2	2	0.05905	**
15	6	1	2	0.07673	***
16	12	10	4	0.21457	*****
17	13	16	6	0.39470	*****
18	14	17	8	1.53502	*****
19	18	15	10	2.72075	*****
SUMA DE LOS INDICES DE NIVEL				5.07185	

Antes de empezar las uniones toda la inercia corresponde a inercia entre-clases (cada individuo es una clase) y a medida que llevan a cabo las uniones va pasando a inercia intra-clases, de modo que al terminar toda la inercia es intra-clases (todos los elementos conforman una clase). Por esta razón en el método de Ward la suma de los índices de nivel es igual a la inercia total.

Los algoritmos de clasificación jerárquica son robustos, es decir que un método para los mismos datos produce los mismos resultados y no requieren de un número de clases preestablecido. Precisamente la mayor utilidad del árbol de clasificación es mostrar la estructura de clases que hay en los datos. En el ejemplo se observa una buena partición en tres clases. La comparación de las alturas de los índices de nivel (histograma de la Tabla 4-3), se constituyen en un buen criterio para decidir cuántas clases tomar para la partición.

Los métodos de clasificación jerárquica tienen sin embargo dos problemas, el primero, debido a la necesidad de la matriz de distancias entre individuos, lo hace exigente en memoria de computador. El

segundo debido al proceso de construcción en forma de particiones anidadas, anidamiento que no permite en general que la inercia intra-clases sea la mínima cuando se obtienen las clases cortando el árbol a alguna altura.

En el ejemplo del café si se corta el árbol para dos clases, se obtiene una con 8 elementos y la otra con 2, los centros de gravedad de las clases son (-3.27, 0.46) y (0.82, -0.12). Al realizar un algoritmo de centros móviles para dos clases con estos puntos iniciales se obtiene una partición con 7 y 3 elementos respectivamente, con lo que consigue un incremento del porcentaje de inercia Inter-clases de 53.6 a 60.3. De esa manera se quita la restricción de anidamiento.

#### **4.2.3. Una estrategia de clasificación**

Los problemas de cada uno de los métodos de clasificación (centros móviles y método de Ward) se solucionan combinando los dos métodos. Utilizándolos con lo métodos factoriales, el procedimiento de clasificación que logra los propósitos buscados desde el punto de vista de la descripción estadística se presenta a continuación.

Realizar un análisis factorial: las coordenadas sobre los ejes factoriales obtenidos son valores de nuevas variables, las cuales son siempre continuas, y son estas las que se utilizan para realizar la clasificación (se debe decidir cuantos ejes se utilizan para la clasificación).

Hacer un reagrupamiento previo cuando hay muchos elementos a clasificar por el método de nubes dinámicas (opcional).

Obtener un árbol de clasificación jerárquica por el método de Ward.

Cortar el árbol para obtener un número de clases adecuado.

Optimizar las clases obtenidas utilizando el método de centros móviles.

Caracterizar las clases mediante las variables activas e ilustrativas.

Utilizar los mapas factoriales como ayuda para caracterizar las clases: proyectar los centros de gravedad de las clases en los planos factoriales u observar los planos factoriales con los elementos identificados según la clase a la que pertenecen.

#### **4.2.4. Caracterización de las clases**

Las clases obtenidas se caracterizan mediante la comparación de las estadísticas descriptivas al interior de las clases con las estadísticas de la población clasificada. Para las variables continuas se compara la media al interior de la clase con la media general. Para las variables nominales se compara el porcentaje de cada modalidad dentro de la clase con el porcentaje general.

Un método de ordenamiento de las variables continuas y modalidades que más caracterizan a una clase es el de realizar pruebas de hipótesis para comparar la estadística de la clase con la estadística en la población.

Para las variables continuas se hace la hipótesis de que la media al interior de la clase es igual a la media general. El valor obtenido para la clase se constituye en la valor muestral con el que se calcula la estadística de prueba bajo el supuesto de que los individuos de la clase son una muestra aleatoria de tamaño  $n_q$  (número de individuos de la clase), tomada sin reposición de la población de tamaño  $n$ . Cuando la estadística de prueba se aproxima a una distribución normal estándar se le da el nombre de valor test. Valores superiores a 2 significan que la variable caracteriza a la clase pues su media dentro de la clase es superior a la media global. Valores inferiores a -2 también caracterizan a la clase pero en el sentido de tener un promedio inferior al promedio general. La comparación de las clases se puede hacer de manera gráfica recurriendo por ejemplo a diagramas box-plot.

Para las variables nominales la caracterización se hace a través de sus modalidades. Ahora el supuesto de selección aleatoria de una muestra lleva al esquema de muestreo de elementos de un lote de tamaño  $n$  con  $n_j$  elementos con la característica de interés. La estadística de prueba en este caso sigue una distribución hipergeométrica., a partir de la cual se obtiene el valor p asociado a la prueba. El valor p se puede recodificar a un cuantil de la distribución normal estándar para obtener el valor test asociado. De manera gráfica la caracterización conjunta de las clases según una variable nominal se puede hacer construyendo los perfiles asociados a la tabla de contingencia que cruza a las clases con las modalidades de la variable nominal.

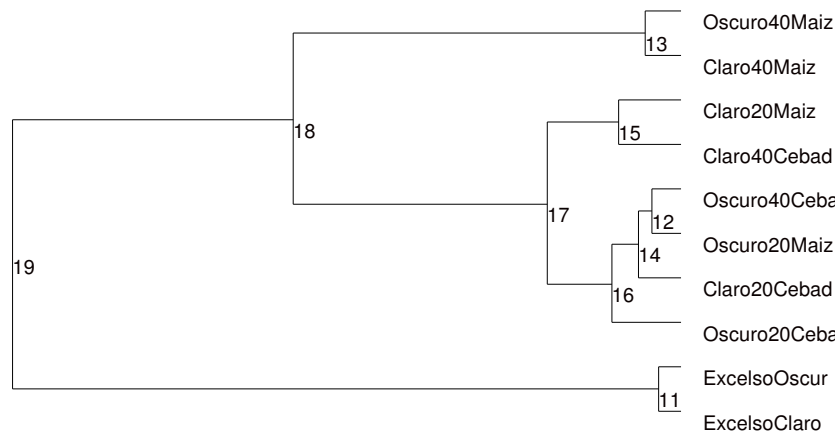
Para ilustrar todo el proceso retomemos el ejemplo completo de las muestras de café del capítulo 1.

Realizar un análisis factorial: el ACP normado del capítulo 1 con notas de degustación como variables continuas activas, la apreciación global como variable continua ilustrativa y la variable nominal tipo de agregado como ilustrativa. Para la clasificación utilizaremos todos los ejes factoriales.

Hacer un agrupamiento previo: obviamente no es necesario un agrupamiento previo pues no se tienen sino 10 individuos a clasificar. En general se utiliza el agrupamiento previo cuando se tiene un número muy grande de individuos y se sobrepasa la capacidad de cálculo al hacer una clasificación directa.

Obtener un árbol de clasificación jerárquica por el método de Ward: Gráfico 4-3.

**Gráfico 4-3: Arbol de clasificación para el ejemplo café del capítulo 1**



Cortar el árbol para obtener un número de clases adecuado: el Gráfico 4-3 muestra que se obtiene una buena partición en tres clases cortando el árbol entre los nodos 17 y 18.

Optimizar las clases obtenidas utilizando el método de centros móviles: en la Tabla 4-4 se muestra que la partición obtenida al cortar el árbol permanece estable con un porcentaje de inercia Inter.-clases de 73.1. Nótese que la inercia total es 6, igual al número de variables activas en el análisis.

Tabla 4-4: Optimización de la partición del ejemplo del café

INERTIES	INERTIES		EFFECTIFS		POIDS		DISTANCES	
	AVANT	APRES	AVANT	APRES	AVANT	APRES	AVANT	APRES
INTER-CLASSES	4.3875	4.3875						
INTRA-CLASSE								
CLASSE 1 / 3	0.0868	0.0868	2	2	2.00	2.00	11.1255	11.1255
CLASSE 2 / 3	1.3803	1.3803	6	6	6.00	6.00	0.1989	0.1989
CLASSE 3 / 3	0.1454	0.1454	2	2	2.00	2.00	10.2156	10.2156
TOTALE	6.0000	6.0000						
QUOTIENT (INERTIE INTER / INERTIE TOTALE) : AVANT ... 0.7313 APRES ... 0.7313								

La partición en tres clases coincide entonces con la que se puede ver en el árbol (Gráfico 4-3) es decir:

{*ExcelsoOscur, ExcelsoClaro*}

{*Claro20Cebad, Oscuro40Ceba, Oscuro20Maiz, Claro40Cebad, Claro20Maiz, Oscuro20Ceba*}

{*Claro40Maiz, Oscuro40Maiz*}

1. Caracterizar las clases mediante las variables activas e ilustrativas: en la Tabla 4-5 se comparan las medias de las clases con respecto a la media general para todas las variables continuas y en la Tabla 4-6 la frecuencia del tipo de agregado al interior de la clase con respecto a la frecuencia global. El interés aquí es mostrar los elementos para la lectura pues el ejemplo es de tamaño muy pequeño.

La clase 1 es la de mejor apreciación, la clase 2 tiene valores parecidos a los promedio y la clase 3 tiene valores inferiores al promedio

En la Tabla 4-6 se puede leer: la clase 1 tiene las dos muestras con agregado de Nada. La clase dos contiene todas las muestras con agregado de cebada (4), que son el 67% de la clase. La clase 3 es de dos muestras con agregado de maíz. En relación a contaminación del café con cereales se puede concluir que el agregado de maíz afecta más la calidad que el agregado de cebada. Los cafés sin agregado son los de mejor calidad.

2. Utilizar los mapas factoriales como ayuda para caracterizar las clases: en el Gráfico 4-1 aparecen los centros de gravedad de las clases y el punto de cada individuo se ha reemplazado por el número de la clase a la que pertenece. Se aprecia de manera global la caracterización de las 10 muestras de café. Hacia el centro aparece la clase más numerosa (6), al lado izquierdo la clase de mejor calidad (sin agregado) y en el extremo derecho la clase de menos calidad construida por las dos muestras con 50% de maíz.

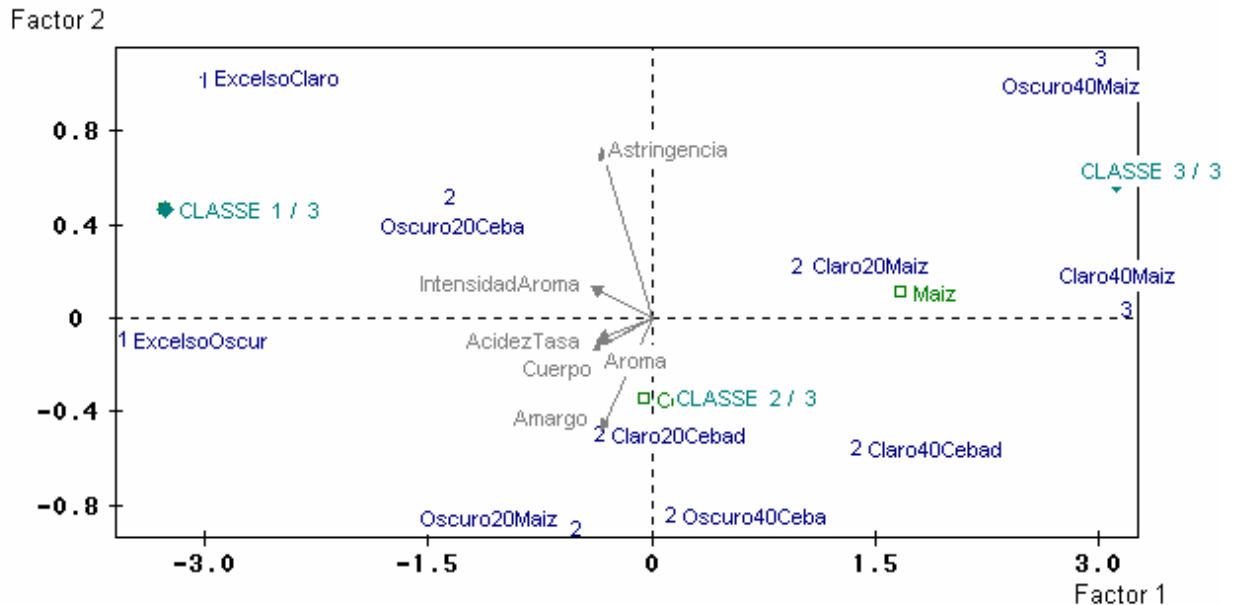
Tabla 4-5: Caracterización de las clases según las variables continuas

V. TEST	PROBA	MOYENNES CLASSE GENERALE		ECARTS TYPES CLASSE GENERAL		VARIABLES CARACTERISTIQUES NUM. LIBELLE	IDEN
CLASSE 1 / 3 ( POIDS = 2.00 EFFECTIF = 2 )							
2.43	0.007	5.29	4.97	0.07	0.19	8.Astringencia	Astr
2.39	0.008	7.69	6.85	0.03	0.53	3.IntensidadAroma	Inte
2.23	0.013	5.07	4.60	0.05	0.32	6.AcidezTasa	Acid
2.10	0.018	7.43	6.64	0.03	0.57	9.Impresion	Impr
1.92	0.027	7.21	6.46	0.21	0.59	4.Aroma	Arom
1.71	0.043	5.11	4.83	0.07	0.25	7.Amargo	Amar
1.34	0.090	6.91	6.67	0.07	0.27	5.Cuerpo	Cuer
CLASSE 2 / 3 ( POIDS = 6.00 EFFECTIF = 6 )							
0.89	0.187	6.74	6.67	0.18	0.27	5.Cuerpo	Cuer
0.19	0.425	6.49	6.46	0.35	0.59	4.Aroma	Arom
0.06	0.476	4.83	4.83	0.21	0.25	7.Amargo	Amar
-0.15	0.438	4.59	4.60	0.14	0.32	6.AcidezTasa	Acid
-0.23	0.410	6.81	6.85	0.20	0.53	3.IntensidadAroma	Inte
-0.49	0.314	6.56	6.64	0.45	0.57	9.Impresion	Impr
-1.33	0.091	4.90	4.97	0.13	0.19	8.Astringencia	Astr
CLASSE 3 / 3 ( POIDS = 2.00 EFFECTIF = 2 )							
-0.80	0.212	4.87	4.97	0.09	0.19	8.Astringencia	Astr
-1.51	0.066	6.07	6.64	0.17	0.57	9.Impresion	Impr
-1.78	0.037	4.53	4.83	0.07	0.25	7.Amargo	Amar
-2.04	0.021	4.17	4.60	0.17	0.32	6.AcidezTasa	Acid
-2.11	0.017	6.10	6.85	0.08	0.53	3.IntensidadAroma	Inte
-2.15	0.016	5.62	6.46	0.20	0.59	4.Aroma	Arom
-2.43	0.008	6.24	6.67	0.02	0.27	5.Cuerpo	Cuer

Tabla 4-6: Caracterización de las clases según el tipo de agregado

V. TEST	PROBA	POURCENTAGES			MODALITES		IDEN	POIDS
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES	DES VARIABLES		
CLASSE 1 / 3								
2.01	0.022	100.00	100.00	20.00	Nada	Agredado	Nada	2
-0.43	0.333	0.00	0.00	40.00	Cebada	Agredado	Ceba	4
-0.43	0.333	0.00	0.00	40.00	Maiz	Agredado	Maiz	4
CLASSE 2 / 3								
1.47	0.071	100.00	66.67	40.00	Cebada	Agredado	Ceba	4
-1.11	0.133	0.00	0.00	20.00	Nada	Agredado	Nada	2
CLASSE 3 / 3								
1.11	0.133	50.00	100.00	40.00	Maiz	Agredado	Maiz	4
-0.43	0.333	0.00	0.00	40.00	Cebada	Agredado	Ceba	4

Gráfico 4-4: Primer plano factorial mostrando las clases



### 4.3. Un ejemplo de aplicación: Departamentos – Icfes

#### 4.3.1. Introducción

Al hacer el ACS se logra visualizar la formación de grupos de departamentos que se conforman alrededor de centros de gravedad constituidos por las categorías de calidad. Sin embargo, algunos departamentos no se sabe si están mas cerca de un centro de gravedad que de otro. No se puede olvidar que lo que podemos observar en un plano factorial es una proyección plana de la nube de puntos y en consecuencia algunos aspectos de la nube de puntos pueden no ser visibles. Los métodos de clasificación permiten, partiendo de las coordenadas factoriales ya calculadas, precisar cuantos grupos de individuos se pueden formar que tengan la doble condición: muy homogéneos los integrantes de cada grupo, y cada grupo muy diferenciado de los demás.

Del ACS se sabe, entre otras cosas, que el perfil Bog+Cund es bastante atípico por su estructura muy diferenciada de los demás perfiles y por su peso relativo, el solo constituye el 22.2% del total. Se puede comprobar, al hacer una primera clasificación, que este perfil se constituye en una Clase aparte y dificulta ver con mayor claridad como se relacionan los demás perfiles. Por esa razón, antes de hacer la clasificación se corrió de nuevo el ACS, pero ahora Bog+Cund se dejó como fila ilustrativa.

#### 4.3.2. Escogencia de una partición

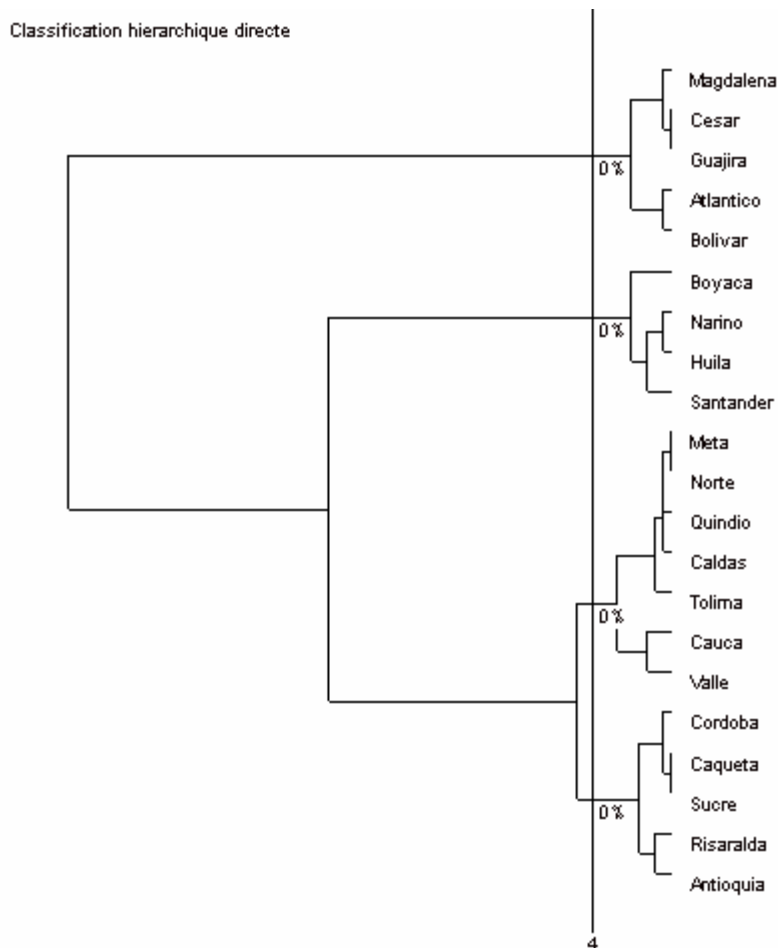
El histograma de índice de nivel (Tabla 4-7) muestra que hay saltos importantes en los cuatro primeros índices. Esto significa que particiones en 2, 3, 4 y hasta 5 clases pueden mostrar la doble condición señalada anteriormente. Puesto que es preciso escoger una partición particular nos inclinamos por la partición en cuatro clases ya que el último salto más importante se da entre el tercero y el cuarto nivel.

**Tabla 4-7: Histograma de índice de nivel. Departamentos - Icfes**

NUM.	AINE	BENJ	EFF.	POIDS	INDICE	HISTOGRAMME DES INDICES DE NIVEAU
22	8	11	2	247.00	0.00015	*
23	13	15	2	352.00	0.00039	*
24	6	19	2	162.44	0.00039	*
25	23	16	3	430.00	0.00061	*
26	9	24	3	308.44	0.00078	*
27	12	22	3	428.00	0.00086	*
28	14	10	2	378.00	0.00092	*
29	2	3	2	679.00	0.00126	**
30	25	5	4	631.00	0.00145	**
31	17	1	2	748.00	0.00183	**
32	30	20	5	897.00	0.00222	***
33	7	21	2	761.00	0.00276	***
34	28	18	3	680.00	0.00297	****
35	26	31	5	1056.44	0.00396	*****
36	4	34	4	919.00	0.00460	*****
37	27	29	5	1107.00	0.00481	*****
38	32	33	7	1658.00	0.00693	*****
39	38	35	12	2714.44	0.01207	*****
40	36	39	16	3633.44	0.04450	*****
41	37	40	21	4740.44	0.07768	*****
SOMME DES INDICES DE NIVEAU =				0.17114		

Al árbol de clasificación jerárquica o dendrograma (Gráfico 4-5), se le ha adicionado una recta vertical que indica el corte para la formación de cuatro clases. En el grafico se puede observar cuales son los departamentos que estarían originalmente en cada clase, es decir antes de hacer la partición y la consolidación de las clases, o sea antes de ejecutar el procedimiento de centros móviles.

**Gráfico 4-5: Árbol de clasificación jerárquica. Departamentos - Icfes**



### 4.3.3. Consolidación de la partición mediante el uso del método de centros móviles.

La información de la Tabla 4-8 muestra en que consiste el proceso de consolidación. La primera parte muestra las iteraciones. Como puede observarse no hay cambios en las proporciones de inercia, lo cual indica que las clases originales constituyen clases fuertes. Esto se puede comprobar en la segunda parte. En ella se ve que las clases antes y después de la consolidación tienen el mismo número de individuos. La última parte sirve para ubicar los centros de gravedad de las clases y caracterizarla.

Tabla 4-8: Consolidación de la partición. Departamentos - Icfes

AUTOUR DES 4 CENTRES DE CLASSES, REALISEE PAR 10 ITERATIONS A CENTRES MOBILES												
PROGRESSION DE L'INERTIE INTER-CLASSES												
ITERATION	I. TOTALE	I. INTER	QUOTIENT									
0	0.17114	0.13425	0.78444									
1	0.17114	0.13425	0.78444									
2	0.17114	0.13425	0.78444									

INERTIES	INERTIES		EFFECTIFS		POIDS		DISTANCES	
	AVANT	APRES	AVANT	APRES	AVANT	APRES	AVANT	APRES
INTER-CLASSES	0.1343	0.1343						
INTRA-CLASSE								
CLASSE 1 / 4	0.0070	0.0070	5	5	1056.44	1056.44	0.0585	0.0585
CLASSE 2 / 4	0.0144	0.0144	7	7	1658.00	1658.00	0.0168	0.0168
CLASSE 3 / 4	0.0085	0.0085	4	4	919.00	919.00	0.2878	0.2878
CLASSE 4 / 4	0.0071	0.0071	5	5	1107.00	1107.00	0.2550	0.2550
TOTALE	0.1711	0.1711						

COORDENADAS Y VALORES-TEST DESPUÉS DE LA CONSOLIDACION													
CLASSES				VALEURS-TEST					COORDONNEES				
IDEN - LIBELLE	EFF.	P. ABS		1	2	3	4	0	1	2	3	4	0
COUPURE 'a' DE L'ARBRE EN 4 CLASSES													
aa1a - CLASSE 1 / 4	5	1056.44		0.9	-3.1	0.4	-0.8	0.0	0.12	-0.21	0.02	-0.02	0.00
aa2a - CLASSE 2 / 4	7	1658.00		-1.1	-0.8	0.6	0.9	0.0	-0.12	-0.04	0.02	0.01	0.00
aa3a - CLASSE 3 / 4	4	919.00		-3.1	2.0	-0.6	0.0	0.0	-0.51	0.15	-0.03	0.00	0.00
aa4a - CLASSE 4 / 4	5	1107.00		3.4	1.9	-0.6	-0.4	0.0	0.49	0.13	-0.02	-0.01	0.00

### 4.3.4. Caracterización de las Clases

La última parte del análisis de clasificación está constituido por la caracterización de las clases, en este caso por las categorías de calidad de la educación. En la Tabla 4-9 se puede ver por ejemplo que la Clase 1, que está ubicada en la parte inferior del plano factorial (ver Gráfico 4-6) está asociada positivamente a la categoría *Bajo* y negativamente a la categoría *Medio*. Esto puede explicarse observando los siguientes porcentajes:

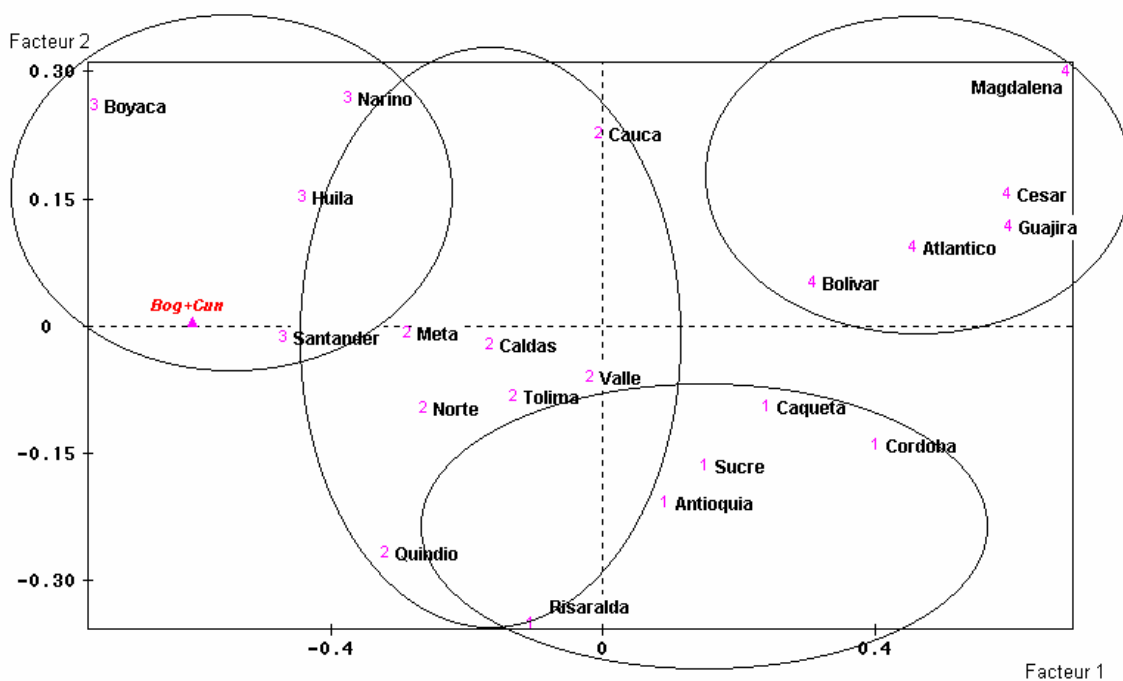
- La Clase 1 tiene el 22.29% del total de planteles, pero de los planteles clasificados como *Bajo* a esta clase pertenece el 27.62 %, es decir, esta por encima del perfil promedio. Por el contrario, de los clasificados como *Medio* solo el 14.43 % pertenecen a esta clase, o sea, por debajo del perfil promedio.

Por otra parte, del total de planteles, el 45.9 % son clasificados como *Bajo*, mientras en la Clase 1 el 57 % están clasificados como *Bajo*. Además, del total de planteles, el 24.4 % son clasificados como *Medio*, pero en esta clase solo el 15.8 %.

Lo anterior nos permite afirmar que la Clase 1 se caracteriza porque en ella es dominante la categoría *Bajo* y por tanto atraída por ella, y al mismo tiempo es deficitaria en la categoría *Medio*, y por tanto rechazada ella. La Clase 2 es atraída por *Medio* y rechazada por *Infer*. La Clase 3 es atraída por *Medio* y *Alto* y rechazada por *Bajo* e *Infer*, la Clase 4 es atraída por *Infer* y rechazada por *Alto* y *Medio*.

Otra forma de interpretar la tabla de descripción de la partición (Tabla 4-10) es observar que el es esencialmente un resumen de la tabla de contingencia que cruza Clases con Categorías (Tabla 4-10).

**Gráfico 4-6: Proyección de los Departamentos sobre el primer plano factorial e identificación de las clases**



**Tabla 4-9: Descripción de la partición – Departamentos - Icfes**

CARACTERIZACION DE LAS CLASES POR LAS CATEGORÍAS (FRECUENCIAS)										
CLASE 1 / 4										
V. TEST	PROBA	POURCENTAGES			NUM .	LIBELLE	FREQUENCES	CARACTERISTIQUES	IDEN	POIDS
		CLA/FRE	FRE/CLA	GLOBAL						
		22.29					CLASE 1 / 4		aa1a	1056
8.10	0.0000	27.62	56.89	45.90	4	Bajo		Bajo	2176	
-7.59	0.0000	14.43	15.81	24.41	3	Medio		Medi	1157	
CLASE 2 / 4										
		34.98					CLASE 2 / 4		aa2a	1658
3.86	0.0001	39.76	27.74	24.41	3	Medio		Medi	1157	
-6.13	0.0000	25.71	12.48	16.98	5	Infer		Infe	805	
CLASE 3 / 4										
		19.39					CLASE 3 / 4		aa3a	919
14.71	0.0000	35.00	44.07	24.41	3	Medio		Medi	1157	
4.53	0.0000	28.43	12.40	8.46	2	Alto		Alto	401	
-9.35	0.0000	13.60	32.21	45.90	4	Bajo		Bajo	2176	
-10.57	0.0000	7.08	6.20	16.98	5	Infer		Infe	805	
CLASE 4 / 4										
		23.35					CLASE 4 / 4		aa4a	1107
16.24	0.0000	46.83	34.06	16.98	5	Infer		Infe	805	
-4.40	0.0000	14.71	5.33	8.46	2	Alto		Alto	401	
-12.28	0.0000	10.80	11.29	24.41	3	Medio		Medi	1157	

**Tabla 4-10: Clases vs niveles de Icfes**

Frecuencia  Porcentaje  Perfil fil  Perfil col	SUPER	ALTO	MEDIO	BAJO	INFER	Total
Clase1	39	83	167	601	164	1054
	0.82	1.75	3.52	12.68	3.46	22.25
	3.70	7.87	15.84	57.02	15.56	
	19.60	20.70	14.43	27.62	20.37	
Clase2	74	145	460	772	207	1658
	1.56	3.06	9.71	16.29	4.37	34.99
	4.46	8.75	27.74	46.56	12.48	
	37.19	36.16	39.76	35.48	25.71	
Clase3	47	114	405	296	57	919
	0.99	2.41	8.55	6.25	1.20	19.40
	5.11	12.40	44.07	32.21	6.20	
	23.62	28.43	35.00	13.60	7.08	
Clase4	39	59	125	507	377	1107
	0.82	1.25	2.64	10.70	7.96	23.36
	3.52	5.33	11.29	45.80	34.06	
	19.60	14.71	10.80	23.30	46.83	
Total	199	401	1157	2176	805	4738
	4.20	8.46	24.42	45.93	16.99	100.00

**4.4. Ejercicio: Test de personalidad con variables discretizadas**

**4.4.1. Presentación**

En el ejercicio realizado anteriormente (sección 3.4), sobre los datos del test de personalidad con variables discretizadas usando AFCM se discutió acerca de la posibilidad de construir tipologías de individuos. Los métodos de clasificación pueden ahora ser usados para precisar esta idea. A continuación se presentan las

salidas básicas que pueden ayudar a tal examen. Al final, después de las tablas y gráficos, encontrará una guía para realizar el análisis.

**4.4.2. Tablas y gráficos**

**Tabla 4-11: Histograma de índices de nivel**

CLASSIFICATION HIERARCHIQUE (VOISINS RECIPROQUES)						
SUR LES 5 PREMIERS AXES FACTORIELS						
DESCRIPTION DES 50 NOEUDS D'INDICES LES PLUS ELEVES						
NUM.	AINE	BENJ	EFF.	POIDS	INDICE	HISTOGRAMME DES INDICES DE NIVEAU
2168	2090	2161	53	53.00	0.00240	**
2169	2095	2074	19	19.00	0.00258	**
2170	2077	2034	28	28.00	0.00264	**
2171	2128	2137	31	31.00	0.00269	**
2172	2099	2097	32	32.00	0.00271	**
2173	2130	2156	43	43.00	0.00277	**
2174	2075	2102	19	19.00	0.00282	**
2175	2145	2105	44	44.00	0.00287	**
2176	2153	2150	50	50.00	0.00288	**
2177	2121	2144	38	38.00	0.00291	**
2178	2116	2119	37	37.00	0.00306	**
2179	2141	2152	51	51.00	0.00314	**
2180	2169	2136	48	48.00	0.00336	**
2181	2149	2154	43	43.00	0.00349	**
2182	2171	2081	48	48.00	0.00358	**
2183	2129	2118	35	35.00	0.00359	**
2184	2165	2147	44	44.00	0.00369	**
2185	2108	2159	42	42.00	0.00376	**
2186	2172	2138	51	51.00	0.00389	**
2187	2173	2122	57	57.00	0.00411	**
2188	2158	2162	61	61.00	0.00412	**
2189	2174	2184	63	63.00	0.00459	***
2190	2168	2103	71	71.00	0.00461	***
2191	2148	2166	64	64.00	0.00498	***
2192	2182	2142	64	64.00	0.00502	***
2193	2177	2157	67	67.00	0.00529	***
2194	2178	2124	58	58.00	0.00545	***
2195	2160	2183	66	66.00	0.00568	***
2196	2180	2140	69	69.00	0.00620	***
2197	2155	2194	91	91.00	0.00640	****
2198	2193	2146	90	90.00	0.00666	****
2199	2186	2181	94	94.00	0.00734	****
2200	2179	2187	108	108.00	0.00811	****
2201	2185	2163	75	75.00	0.00827	****
2202	2170	2195	94	94.00	0.01043	*****
2203	2191	2188	125	125.00	0.01088	*****
2204	2164	2167	71	71.00	0.01161	*****
2205	2175	2190	115	115.00	0.01234	*****
2206	2204	2176	121	121.00	0.01362	*****
2207	2200	2203	233	233.00	0.01741	*****
2208	2192	2189	127	127.00	0.01957	*****
2209	2206	2201	196	196.00	0.02031	*****
2210	2205	2196	184	184.00	0.02379	*****
2211	2197	2207	324	324.00	0.03078	*****
2212	2202	2208	221	221.00	0.03879	*****
2213	2210	2198	274	274.00	0.04257	*****
2214	2213	2199	368	368.00	0.04814	*****
2215	2212	2209	417	417.00	0.08299	*****
2216	2214	2211	692	692.00	0.10756	*****
2217	2215	2216	1109	1109.00	0.16915	*****
SOMME DES INDICES DE NIVEAU =					0.99845	

**Gráfico 4-7: Arbol de Clasificación del ejercicio**

Classification hierarchique directe

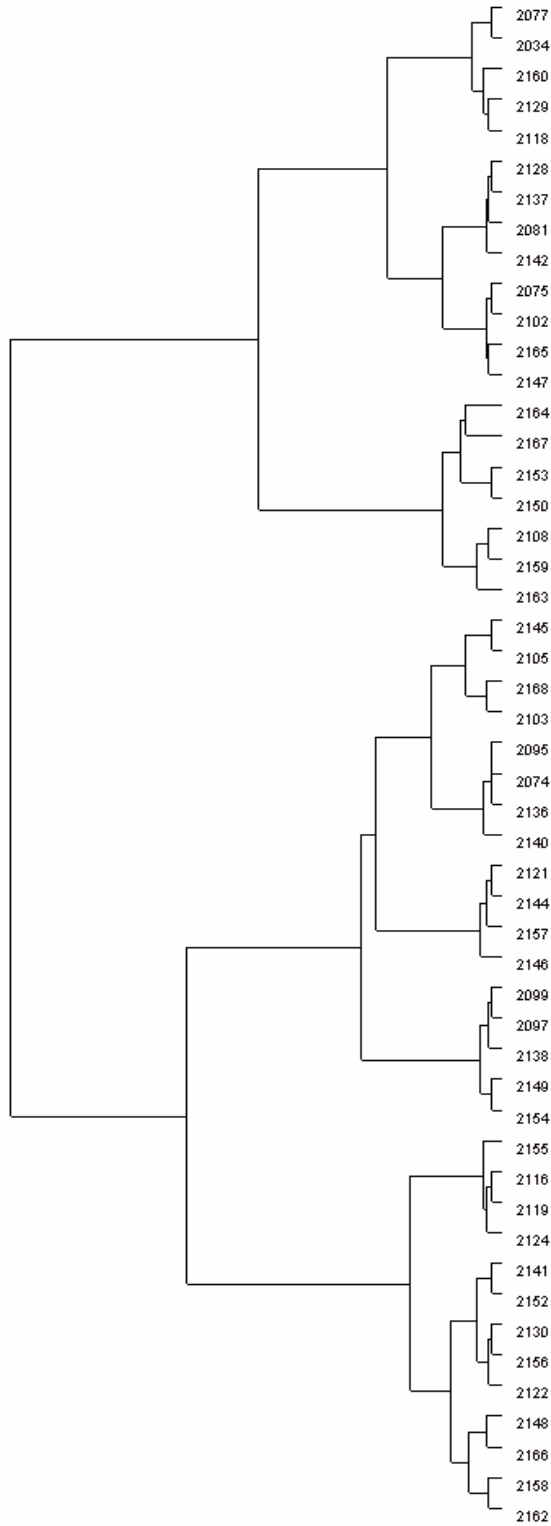


Tabla 4-12: Consolidación de la partición del ejercicio

PARTITION PAR COUPURE D'UN ARBRE HIERARCHIQUE  
 COUPURE 'a' DE L'ARBRE EN 4 CLASSES  
 FORMATION DES CLASSES (INDIVIDUS ACTIFS)  
 DESCRIPTION SOMMAIRE

CLASSE	EFFECTIF	POIDS	CONTENU
aa1a	324	324.00	1 A 13
aa2a	368	368.00	14 A 30
aa3a	196	196.00	31 A 37
aa4a	221	221.00	38 A 50

COORDONNEES ET VALEURS-TEST AVANT CONSOLIDATION  
 AXES 1 A 5

CLASSES			VALEURS-TEST					COORDONNEES					DISTO.
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	DISTO.
COUPURE 'a' DE L'ARBRE EN 4 CLASSES													
aa1a - CLASSE 1 / 4	324	324.00	-10.9	-18.9	10.8	1.4	3.3	-0.28	-0.43	0.20	0.03	0.06	0.30
aa2a - CLASSE 2 / 4	368	368.00	-15.2	16.2	-6.5	-1.7	-1.9	-0.35	0.33	-0.11	-0.03	-0.03	0.25
aa3a - CLASSE 3 / 4	196	196.00	11.4	10.8	2.1	4.8	12.6	0.40	0.34	0.05	0.12	0.32	0.40
aa4a - CLASSE 4 / 4	221	221.00	19.4	-7.9	-6.5	-4.3	-13.6	0.64	-0.23	-0.16	-0.10	-0.32	0.60

CONSOLIDATION DE LA PARTITION  
 AUTOUR DES 4 CENTRES DE CLASSES, REALISEE PAR 10 ITERATIONS A CENTRES MOBILES  
 PROGRESSION DE L'INERTIE INTER-CLASSES

ITERATION	I. TOTALE	I. INTER	QUOTIENT
0	0.99845	0.35971	0.36026
1	0.99845	0.40619	0.40682
2	0.99845	0.40880	0.40943
3	0.99845	0.41016	0.41079
4	0.99845	0.41087	0.41151
5	0.99845	0.41158	0.41222
6	0.99845	0.41261	0.41325
7	0.99845	0.41287	0.41351
8	0.99845	0.41300	0.41364

ARRET APRES L'ITERATION 8 L'ACCROISSEMENT DE L'INERTIE INTER-CLASSES  
 PAR RAPPORT A L'ITERATION PRECEDENTE N'EST QUE DE 0.030 %.  
 DECOMPOSITION DE L'INERTIE  
 CALCULEE SUR 5 AXES.

INERTIES	INERTIES		EFFECTIFS		POIDS		DISTANCES	
	AVANT	APRES	AVANT	APRES	AVANT	APRES	AVANT	APRES
INTER-CLASSES	0.3597	0.4130						
INTRA-CLASSE								
CLASSE 1 / 4	0.1550	0.2017	324	396	324.00	396.00	0.3029	0.2798
CLASSE 2 / 4	0.2420	0.1955	368	347	368.00	347.00	0.2491	0.3214
CLASSE 3 / 4	0.0995	0.0954	196	187	196.00	187.00	0.3952	0.5695
CLASSE 4 / 4	0.1422	0.0928	221	179	221.00	179.00	0.5956	0.7217
TOTALE	0.9985	0.9985						

QUOTIENT (INERTIE INTER / INERTIE TOTALE) : AVANT ... 0.3603  
 APRES ... 0.4136

COORDONNEES ET VALEURS-TEST APRES CONSOLIDATION  
 AXES 1 A 5

CLASSES			VALEURS-TEST					COORDONNEES					DISTO.
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	DISTO.
COUPURE 'a' DE L'ARBRE EN 4 CLASSES													
aa1a - CLASSE 1 / 4	396	396.00	-14.6	-21.0	5.2	1.9	3.7	-0.32	-0.41	0.08	0.03	0.06	0.28
aa2a - CLASSE 2 / 4	347	347.00	-12.1	22.0	-4.5	-0.8	-4.5	-0.29	0.47	-0.08	-0.01	-0.08	0.32
aa3a - CLASSE 3 / 4	187	187.00	15.9	6.5	3.5	3.8	16.1	0.57	0.21	0.10	0.10	0.42	0.57
aa4a - CLASSE 4 / 4	179	179.00	18.2	-7.0	-4.6	-5.4	-15.6	0.68	-0.23	-0.13	-0.14	-0.42	0.72

Tabla 4-13: Caracterización de las Clases por las variables activas e ilustrativas del ejercicio

DESCRIPTION DE PARTITION(S) DESCRIPTION DE LA COUPURE 'a' DE L'ARBRE EN 4 CLASSES CARACTERISATION DES CLASSES PAR LES MODALITES CARACTERISATION PAR LES MODALITES DES CLASSES OU MODALITES DE COUPURE 'a' DE L'ARBRE EN 4 CLASSES CLASSE 1 / 4									
V.TEST	PROBA	POURCENTAGES			MODALITES		IDEN	POIDS	
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES	DES VARIABLES			
35.71 CLASSE 1 / 4							aa1a	396	
15.82	0.000	81.61	45.96	20.11	EST5	Estabilidad	EST5	223	
14.03	0.000	76.44	43.43	20.29	RES5	Responsabilidad	RES5	225	
11.36	0.000	67.21	41.41	22.00	SOC2	Sociabilidad	SOC2	244	
8.65	0.000	62.87	32.07	18.21	CAU5	Cautela	CAU5	202	
8.25	0.000	60.18	33.59	19.93	ASE2	Asendencia	ASE2	221	
6.89	0.000	56.48	30.81	19.48	COM3	Comprension	COM3	216	
6.81	0.000	54.12	34.85	22.99	ASE3	Asendencia	ASE3	255	
6.12	0.000	54.55	28.79	18.85	EST4	Estabilidad	EST4	209	
5.79	0.000	52.65	30.05	20.38	SOC3	Sociabilidad	SOC3	226	
5.01	0.000	46.37	41.92	32.28	Ten1	TensoImpulsivo-Relajado	Ten1	358	
4.79	0.000	50.00	27.78	19.84	VIT5	Vitalidad	VIT5	220	
2.63	0.004	44.06	22.47	18.21	ORI2	Originalidad	ORI2	202	
2.42	0.008	42.79	24.75	20.65	Dom1	Dominacion-Subordinacion	Dom1	229	
CLASSE 2 / 4									
V.TEST	PROBA	POURCENTAGES			MODALITES		IDEN	POIDS	
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES	DES VARIABLES			
31.29 CLASSE 2 / 4							aa2a	347	
12.54	0.000	69.61	40.92	18.39	ASE5	Asendencia	ASE5	204	
12.10	0.000	67.62	40.92	18.94	SOC5	Sociabilidad	SOC5	210	
11.55	0.000	65.88	40.06	19.03	SOC4	Sociabilidad	SOC4	211	
7.98	0.000	55.24	33.43	18.94	ASE4	Asendencia	ASE4	210	
6.70	0.000	50.44	32.85	20.38	EST2	Estabilidad	EST2	226	
6.60	0.000	50.22	32.56	20.29	COM5	Comprension	COM5	225	
6.25	0.000	47.51	35.73	23.53	RES3	Responsabilidad	RES3	261	
5.61	0.000	47.71	29.97	19.66	ORI5	Originalidad	ORI5	218	
5.20	0.000	46.54	29.11	19.57	COM4	Comprension	COM4	217	
4.31	0.000	43.35	29.11	21.01	CAU4	Cautela	CAU4	233	
4.28	0.000	44.08	26.80	19.03	RES2	Responsabilidad	RES2	211	
4.11	0.000	51.65	13.54	8.21	Tra5	Tranquilo-Reservado	Tra5	91	
3.79	0.000	41.88	28.24	21.10	ORI4	Originalidad	ORI4	234	
3.50	0.000	39.60	34.01	26.87	Suf1	Autosufic-Suficiencia	Suf1	298	
3.15	0.001	40.45	25.65	19.84	ORI3	Originalidad	ORI3	220	
2.83	0.002	39.46	25.36	20.11	VIT3	Vitalidad	VIT3	223	
2.71	0.003	38.82	26.51	21.37	EST3	Estabilidad	EST3	237	
2.67	0.004	39.09	24.78	19.84	VIT4	Vitalidad	VIT4	220	
2.58	0.005	39.13	23.34	18.67	CAU2	Cautela	CAU2	207	
CLASSE 3 / 4									
V.TEST	PROBA	POURCENTAGES			MODALITES		IDEN	POIDS	
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES	DES VARIABLES			
16.86 CLASSE 3 / 4							aa3a	187	
16.23	0.000	56.03	69.52	20.92	CAU1	Cautela	CAU1	232	
14.02	0.000	50.00	63.64	21.46	COM1	Comprension	COM1	238	
10.65	0.000	42.42	52.41	20.83	RES1	Responsabilidad	RES1	231	
7.87	0.000	35.32	44.39	21.19	ORI1	Originalidad	ORI1	235	
7.64	0.000	35.98	41.18	19.30	EST1	Estabilidad	EST1	214	
7.08	0.000	33.77	41.18	20.56	VIT1	Vitalidad	VIT1	228	
5.51	0.000	34.48	26.74	13.07	Ten5	TensoImpulsivo-Relajado	Ten5	145	
3.98	0.000	35.53	14.44	6.85	Dom5	Dominacion-Subordinacion	Dom5	76	
3.90	0.000	26.24	31.02	19.93	ASE2	Asendencia	ASE2	221	
3.21	0.001	23.83	32.62	23.08	Ego1	EgoFuerte-EgoDebil	Ego1	256	
CLASSE 4 / 4									
V.TEST	PROBA	POURCENTAGES			MODALITES		IDEN	POIDS	
		CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES	DES VARIABLES			
16.14 CLASSE 4 / 4							aa4a	179	
21.94	0.000	70.32	86.03	19.75	ASE1	Asendencia	ASE1	219	
17.42	0.000	59.63	72.63	19.66	SOC1	Sociabilidad	SOC1	218	
8.47	0.000	35.74	46.93	21.19	ORI1	Originalidad	ORI1	235	
7.37	0.000	33.33	43.02	20.83	RES1	Responsabilidad	RES1	231	
6.39	0.000	30.67	40.78	21.46	COM1	Comprension	COM1	238	
5.49	0.000	28.95	36.87	20.56	VIT1	Vitalidad	VIT1	228	
5.08	0.000	28.50	34.08	19.30	EST1	Estabilidad	EST1	214	
3.08	0.001	24.32	25.14	16.68	Suf5	Autosufic-Suficiencia	Suf5	185	
2.47	0.007	21.83	27.93	20.65	Dom1	Dominacion-Subordinacion	Dom1	229	

#### 4.4.3. Guía para el análisis

1. Analizando la Tabla 4-11, que se sugiere en cuanto al número de clases que se pueden formar.
2. Sobre el Gráfico 4-7 señale con un línea recta por donde se debe cortar el árbol para formar cuatro clases?. Porqué en este caso no interesa saber quienes son los individuos?
3. Después de decidir la formación de cuatro clases se hace la consolidación de la partición. Examine el proceso de consolidación. Que tan estables resultaron ser las clases que salieron del proceso de clasificación jerárquica? En que datos es preciso fijar la atención para observar el proceso de clasificación?
4. La Tabla 4-13 contiene los elementos para una caracterización de las cuatro clases. Cuales son las modalidades activas que más fuertemente caracterizan la Clase 1. Cual es el porcentaje de individuos en esta clase? En qué porcentaje están representadas las modalidades que la caracterizan? Las modalidades más importantes para esta clase cómo están representadas en el conjunto de la población? Y cómo están representadas en esta clase?
5. Haga un análisis similar para cada una de las clases.
6. Con la información de la Tabla 4-13, y a manera de ejemplo ilustrativo, llene las celdas que pueda de una tabla de contingencia que cruce las clases con las modalidades de la variable *estabilidad*.
7. Haga un resumen del análisis.

## BIBLIOGRAFÍA

- CABARCAS G. (1999). *Uso de técnicas exploratorias para describir los perfiles de calidad regionales de a educación media en Colombia*. Reporte Interno. Universidad Nacional de Colombia Medellín.
- CHARUM J., PARDO C.E., MONTENEGRO A. y OLAYA D.L. (2000). *Recursos humanos para la ciencia y la tecnología. Convocatoria de centros y grupos de investigación-1998*. Informe interno. Observatorio de Ciencia y Tecnología de Colombia. Bogotá.
- CRIVISQUI, E. (1993). *Análisis factorial de correspondencias un instrumento de investigación en ciencias sociales*. Laboratorio de Informática Social Universidad Católica de Asunción, Asunción.
- DUARTE R., SUAREZ M., MORENO E. y ORTIZ P (1996). *Análisis multivariado por componentes principales, de cafés tostados y molidos adulterados con cereales*. Cenicafé 47(2): 65-76.
- ESCOFIER, B. y PAGÉS, J. (1992) *Análisis factoriales simples y múltiples, objetivos, métodos e interpretación*. Servicio Editorial Universidad del País Vasco. Bilbao..
- GREENACRE M. (1984) – *Theory and Applications of Correspondence Analysis*. Academic press, London.
- LEBART L., MORINEAU A. y PIRON M.. (1995) *Statistique exploratoire multidimensionnelle*. Dunod, París.
- LOAIZA G. (2001). *Comparación entre los test de personalidad PPG-IPG y 16PF*. Trabajo final para optar al título de Especialista en Estadística. Universidad Nacional de Colombia. Medellín.
- PADILLA, L.S. (1999). *Análisis multivariado de un grupo de estudiantes del colegio Calasanz Bogotá según el rendimiento académico, las aptitudes mentales y los factores familiares*. Trabajo final para optar al título de Especialista en Estadística. Universidad Nacional de Colombia. Bogotá.
- PARDO, C.E. (1992). *Análisis de la Aplicación del Método de Ward de Clasificación al caso de Variables Cualitativas*. Tesis para optar al título de M.Sc. Estadística. Universidad Nacional de Colombia. Bogotá.