

## REFERENCIAS

1. VIRIUKOB, V. I. y otros: **Determinación de la red óptima para la exploración preliminar de los yacimientos estratigráficos.** Ed. Nedra, Moscú, 1972.
2. BOGATSKY, V. V. : **Análisis matemático de las redes de exploración.** Gosgeonzdat, Moscú, 1973.
3. VENTSED, E. C. : **Teoría de las probabilidades.** Ed. Nauka, Moscú, 1969.
4. GUDKOV, V. M. : "Determinación y utilización de las características de variabilidad de los indicadores en los yacimientos minerales útiles" en **Métodos matemáticos en geología.** Ed. Nedra, Moscú, 1978.
5. DAVID, D. : **Estadística y análisis de los datos geológicos.** Ed. MIR, Moscú, 1977.
6. KAZDAN, A. B. : **Fundamentos metodológicos para la búsqueda de minerales útiles.** Ed. Nedra, Moscú, 1974.
7. KAZAKOBSKY, D. A. : **Valoración de la exactitud de los resultados en relación con la geometrización y el cálculo de reservas de yacimientos.** Ugletejizdat, Moscú, 1948.
8. MATERON, J. : **Fundamentos de geoestadística aplicada.** Ed. Nedra, Moscú, 1968.
9. OCETSKY, A. I. : **Indicadores de la densidad de las excavaciones de exploración de acuerdo con el carácter del cuerpo investigado.** Ugletejizdat, Moscú, 1961.

CDU : 519.242 : 311.124 : 550 (729.1)

© REVISTA MINERIA Y GEOLOGIA, 1-85

## ANÁLISIS ESTADÍSTICO MULTIVARIADO APLICADO A LA INGENIERÍA GEOLOGICA

Ingeniero Willy R. Rodríguez Miranda, Instructor del Departamento de Geofísica del ISMMMoa

Ingeniera Rosa María Valcarce Ortega, Instructora del Departamento de Geofísica del ISMMMoa

### RESUMEN

El presente trabajo creemos que constituye el primer intento en nuestro país de aplicar las técnicas de clasificación estadística multivariada para resolver tareas concretas en el campo de la ingeniería geológica.

Se demuestra que la utilización del Análisis Euclidiano de Agrupación, resuelve satisfactoriamente tareas como: descripción de calas y construcción de cortes ingeniero-geológicos, las que al ser resueltas por métodos convencionales no están exentas de los criterios subjetivos del interpretador.

Inicialmente, se muestran los aspectos teóricos del Método Euclidiano de Agrupación y posteriormente se discuten los resultados de haber aplicado dicho método de clasificación sobre 31 perfiles litológicos con profundidad promedio de 20 metros, pertenecientes a importantes obras construidas en Ciudad de La Habana.

En nuestro país, que marcha con pasos firmes y seguros por el camino del socialismo, son múltiples y de gran importancia las obras tanto sociales como industriales que se construyen, las cuales exigen cada día más un estudio ingeniero-geológico cuidadoso y profundo.

El presente trabajo muestra cómo interpretar datos ingeniero-geológicos mediante el uso de técnicas de clasificación estadística multivariada, específicamente mediante el Análisis Euclidiano de Agrupación que consiste en clasificar cada muestra de suelo o roca según el grado de similitud que caracteriza a los atributos bajo estudio; ello permite realizar la interpretación de los datos ingeniero-geológicos con la menor influencia posible de los criterios subjetivos del interpretador.

### ABSTRACT

This article represents the attempt made in Cuba to apply the multivariate classification techniques in the solution of special problems in geological engineering.

It is demonstrated that the application of the Euclidian grouping analysis successfully meets tasks such as: description of borings and the construction of engineering geological profiles: which, when designed through conventional methods are not free from subjective interpretations.

Firstly, the theoretical aspects of the Euclidian method are discussed and then are shown the results of the application of such classification methods to 31 lithological profiles, with an average depth of 20 m. All of them are concerned with important facilities constructed in Havana City.

La estadística aplicada ha pasado a desempeñar un papel de primera importancia en las investigaciones actuales, y la ingeniería geológica también encuentra una poderosa herramienta de trabajo en esta ciencia.

### ASPECTOS TEORICOS DEL METODO EUCLIDIANO DE AGRUPACION

Antes de referirnos propiamente a los aspectos teóricos del Método Euclidiano de Agrupación, debemos definir los términos coeficiente de distancia euclidiano y centroide [1, 4].

Coeficiente de distancia euclidiano ( $d_{ij}$ )

Es una medida de similitud definida como:

$$d_{ij} = \frac{\sum_{k=1}^p (X_{ik} - X_{jk})^2}{p}^{1/2}$$

donde:

- p: número de atributos
- $X_{ik}$ : atributo k de la muestra elemental i
- $X_{jk}$ : atributo k de la muestra elemental j

Este coeficiente  $d_{ij}$  se interpreta de la siguiente manera:

$$d_{ij} = \begin{cases} 0: & \text{similitud perfecta entre las muestras analizadas} \\ \infty: & \text{disimilitud perfecta entre las muestras analizadas} \end{cases}$$

### Centroide

Punto que tiene como coordenadas el valor medio de cada uno de los atributos que caracterizan a la población muestral multivariada. Se define como:

$$Y_{kj} = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ij}$$

donde:

- $n_k$ : número de muestras elementales de la población muestral k
- $X_{ij}$ : muestras elementales de la población muestral k
- j: atributo analizado

Si la población muestral está formada por q atributos entonces el centroide tendrá q que serán:

$$Y_{k1}, Y_{k2}, \dots, Y_{kq}$$

y nos dará los valores más probables de los atributos estudiados de esa población muestral multivariada.

El método Euclidiano de Agrupación parte de suponer que las n muestras elementales pertenecientes al espacio euclidiano p dimensional ( $E^p$ ) pueden ser divididas en subpoblaciones y que puede existir un número máximo ( $C_{max}$ ) y un número mínimo ( $C_{min}$ ) de grupos de poblaciones estadísticas. Para definir la distribución de las muestras elementales por grupos se realizan los siguientes pasos [1, 4]:

1. Se calculan los centroides de cada grupo aleatoriamente a partir de los atributos.

(Si  $C_{max} = k$ , se generan k centroides aleatoriamente).

2. Cada una de las muestras elementales se ubican en el grupo para el cual el coeficiente de distancia euclidiano muestra-centroide, sea menor.
3. Cuando los n individuos (muestras elementales), han sido asignados a cada uno de los  $C_{max}$  grupos se reevalúan las coordenadas de los centroides.
4. Se chequean las distancias de cada individuo a cada uno de los nuevos centroides, asignándose nuevamente cada muestra elemental a aquel grupo para el cual el coeficiente de distancia euclidiano muestra-centroide sea menor, modificándose inmediatamente las coordenadas de los centroides de los dos grupos afectados.
5. El número de Grupos (C) se reduce en uno (a menos que  $C = C_{min}$ ). Esto se realiza fundiendo aquel par de grupos para los cuales el coeficiente de distancia euclidiano entre sus centroides sea menor. Posteriormente se calculan las coordenadas del centroide del nuevo grupo formado.
6. Una vez realizadas las agrupaciones para  $C_{min} \leq C \leq C_{max}$  disponemos para cada grupo de la desviación medio cuadrática puntos-centroides ( $S_k$ ) definida como:

$$S_k = \left[ \frac{1}{n-C} \sum_{i=1}^n d_{ik}^2 \right]^{1/2}$$

donde:

- C:  $C_{max}$
- n: número de elementos del grupo k
- $d_{ik}$ : coeficiente de distancia euclidiano entre la muestra elemental i y el grupo k

A partir de esta magnitud es posible estimar la suma de los cuadrados de los residuos definida como:

$$SSR_k = (n - C) S_k^2$$

donde:

- $SSR_k$ : Suma de los cuadrados de los residuos. Esta magnitud expresa cuantitativamente la concentración de los puntos alrededor del centroide. A mayor  $SSR_k$ , menos concentrados están los puntos alrededor del centroide

Es necesario escoger el número de poblaciones estadísticas óptimas para realizar la clasificación. Para resolver este problema nos planteamos la siguiente hipótesis básica:

- $H_0$ : La agrupación con N2 grupos y con N1 grupos ( $N2 > N1$ ) produce ajustes estadísticos equivalentes; o sea el modelo estadístico con N1 grupos es tan bueno como con N2 grupos.

Para verificar esta hipótesis realizamos una prueba Fisher y calculamos el estadístico F como:

$$F = \left[ \frac{SSR(N1) \cdot SSR(N2)}{SSR(N2)} \right] / \left[ \frac{(n - N1) \left( \frac{N2^2}{N1} \right) - 1}{n - N2} \right]$$

donde:

- n: número total de puntos, número total de muestras elementales bajo estudio
- p: número de atributos
- N2 y N1: número de grupos que estamos analizando
- $SSR(N1)$  y  $SSR(N2)$ : suma de los cuadrados de los residuos con N1 grupos y N2 grupos respectivamente

$$F > F_{1-\alpha}(v_1, v_2)$$

donde:

$v_1$ : número de grados de libertad del numerador

$$v_1 = p(N2 - N1)$$

$v_2$ : número de grados de libertad del denominador

$$v_2 = p(n - N2)$$

$\alpha$ : nivel de significación

Si F cae en RC se rechaza la hipótesis  $H_0$  y se acepta entonces que trabajar con N2 grupos mejora la clasificación estadística.

Para seleccionar cuál es el número adecuado de grupos para clasificar las muestras se realiza la prueba de hipótesis anterior, haciendo:

$$N1 = l \quad l = C_{min}, \dots, C_{max}$$

$$N2 = l + 1$$

y se valora de forma sucesiva si el uso del grupo es tan eficiente como el uso de (l + 1) grupos, teniendo entonces que:

$$\hat{F} = \left[ \frac{SSR(l) - SSR(l+1)}{SSR(l+1)} \right] / \left[ \frac{n-1 \left( \frac{l+2^2/p}{l} \right) - 1}{n-l-1} \right]$$

siendo la RC:

$$\hat{F} > F_{1-\alpha}(v_1, v_2)$$

$v_1 = p$

$$v_2 = p(n - l - 1)$$

Aquel caso en que el incremento del número de grupos hace que F caiga en la RC se tomará como el número de grupos óptimos para clasificar las muestras elementales bajo estudio.

En el presente trabajo la aplicación de este método se realizó mediante algoritmo GRUPOS, que es uno de los algoritmos de agrupación euclidiana más eficaz [3].

Debe destacarse que este método de clasificación estadística multivariada encuentra aplicación en:

1. Mapeo o construcción de mapas pronósticos.
2. Verificación de la existencia de diferencias significativas entre dos o más tipos de poblaciones geológicas. En este trabajo el método es utilizado para determinar la existencia de diferencias significativas entre las rocas desde el punto de vista ingeniero-geológico.

### APLICACIONES

El método de clasificación estadística fue aplicado sobre 31 perfiles litológicos con profundidad promedio de 20 metros, pertenecientes a las siguientes obras construidas en Ciudad de La Habana [2].

- Edificios de moldes deslizantes de 20 plantas (San Rafael y Espada)

- Edificio Cayo Hueso

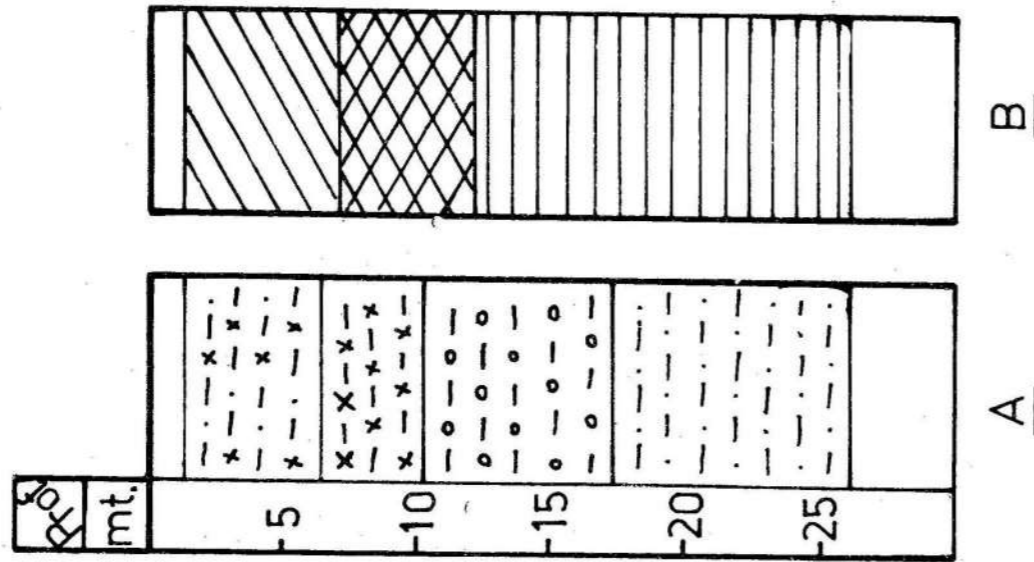
- Edificio no. 2 de 20 plantas, Cayo Hueso

- Edificio de 20 plantas, Tulipán y Loma

Los atributos estudiados fueron: límite líquido (LL), límite plástico (LP), número de golpes (NG) y humedad natural (W).

Nuestro objetivo era describir nuevamente las calas, separando los estratos que desde el punto de vista ingeniero-geológico presentaran propiedades diferentes y comparar esta clasificación con la descripción litológica tacto-visual que ya existía. Podemos decir que los mejores modelos de clasificación se obtuvieron con 3 grupos estadísticos para una fiabilidad del 90%, reflejándose, de manera general, correspondencia entre la descripción de las calas a partir del método euclidiano de agrupación y las descripciones ya existentes de las mismas, pero en la descripción propuesta por nosotros numerosos estratos, que habían sido considerados diferentes al ser descritos por métodos tacto-visuales, pertenecen a una misma población estadística, o sea, tienen propiedades ingeniero-geológicas similares. Además, mediante la aplicación de esta técnica de clasificación estadística podemos caracterizar los valores medios de los atributos bajo estudio en cada estrato ingeniero-geológico [4].

En las Figuras 1 y 2 pueden verse de forma gráfica los aspectos anteriores.



Grupo	Simb.	$\bar{N}G$	$\bar{W}$	$\bar{L}L$	$\bar{L}P$	distancia media puntos centroides
1		10	0,4	0,88	0,31	2,9
2		34	0,36	0,87	0,34	4,8
3		21	0,39	0,92	0,34	1,7

A - Descripción tacto-visual  
 B - Descripción por el Método Euclidiano de Agrupación

LEYENDA PARA LA DESCRIPCIÓN TACTO VISUAL:

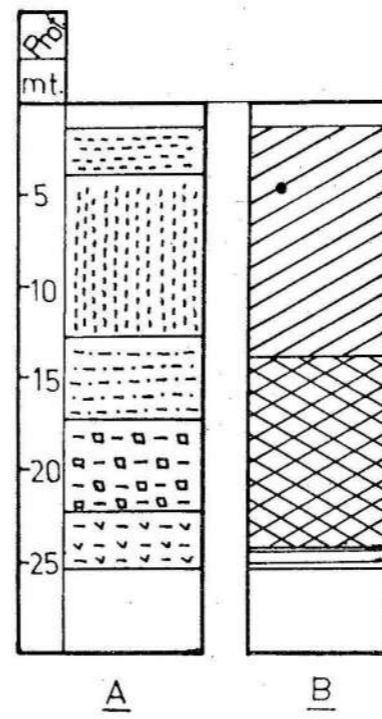
Arcilla calcárea veteada en gris y carmelita, fisurada y compacta, manchas amarillas y de oxidación, bolsones arenosos y algunos de carbonato.

Arcilla verdosa fisurada, dura, manchas amarillas, bolsones arenosos, cristales de yeso, algunas intercalaciones de fragmentos de areniscas.

Arcilla carmelita muy fisurada y muy compacta, vetas grises y bolsones de carbonato.

Arcilla gris oscuro a verdoso, muy dura y fisurada con intercalaciones arenosas.

Fig. 1. Descripción de calas. Cala no. 3.



Grupo	Simb.	$\bar{N}G$	$\bar{W}$	$\bar{L}L$	$\bar{L}P$	distancia media puntos centroides
1		15	0,38	0,84	0,33	3,8
2		29	0,35	0,83	0,36	4,1
3		45	0,39	0,79	0,37	5,3

A. Descripción tacto-visual  
 B. Descripción por el Método Euclidiano de Agrupación

FIG. II. CALA NUMERO 16

Arcilla verde grisáceo con zonas amarillentas, con bolsones blancos de carbonato y bolsones arenosos, compacta.

Arcilla verde grisáceo con zonas amarillentas, con bolsones blancos de carbonato, con bolsones arenosos finos, fisurada muy compacta.

Arcilla verdoso carmelitosa, muy fisurada con manchas de óxido, con bolsones de marga, color crema, dura.

Arcilla verdosa oscura, carmelitosa, fisurada con manchas de óxido, con bolsones de marga color crema, de muy compacta a dura, al final con zonas menos finas, limosa, densa.

Arcilla gris oscuro, con cristales de pirita, dura, fisurada, con bolsones arenosos finos, limosos gris oscuro.

Fig. 2. Descripción de calas. Cala no. 16.

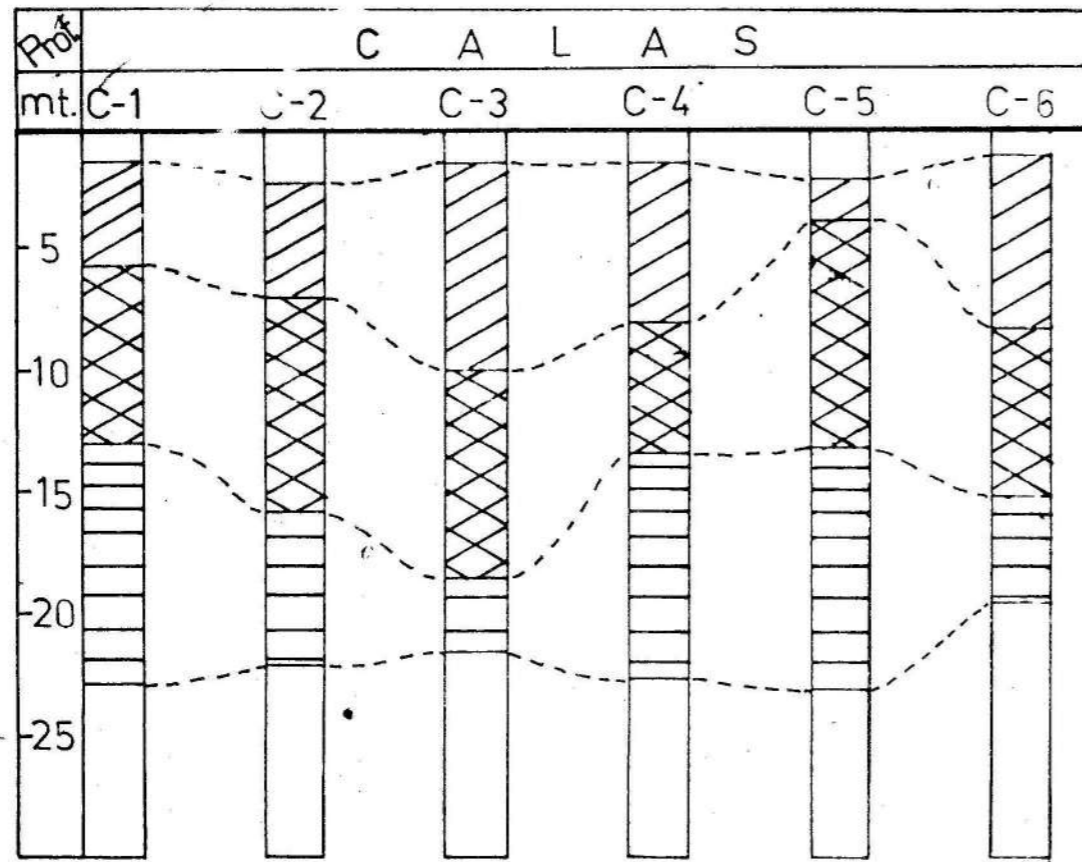
Si conocemos la posición de las calas bajo estudio es posible construir el corte ingeniero-geológico correlacionando los horizontes que presentan el mismo comportamiento, o sea, que pertenecen al mismo grupo estadístico. Sabemos que esta tarea por métodos convencionales está impregnada por los criterios subjetivos del interpretador y que el análisis estadístico de los atributos se realiza de forma univariada. Mediante el Análisis Euclidiano de Agrupación realizamos el estudio multivariado de los atributos y redujimos el carácter subjetivo de la interpretación, lo cual en zonas de gran complejidad ingeniero-geológica resulta de mucha importancia. En la Figura 3 puede verse un ejemplo de construcción de corte ingeniero-geológico mediante el Método Euclidiano de Agrupación [ 4 ].

Debe destacarse que este método de clasificación estadística multivariada ha sido aplicado en la clasificación de suelos y en la clasificación de datos geofísicos y geomorfológicos con resultados positivos.

### CONCLUSIONES

1. Con la aplicación del Análisis Euclidiano de Agrupación es posible la construcción de mapas pronósticos y la verificación de la existencia de diferencias significativas entre dos o más tipos de poblaciones geológicas de forma rápida y eficiente.
2. Los atributos deben ser seleccionados correctamente, de forma tal que reflejen las características más importantes que permitan la diferenciación de las poblaciones muestrales.
3. El presente trabajo es sólo el inicio de la aplicación de algunas técnicas de clasificación estadística multivariada a la ingeniería geológica, por lo que se hace necesario investigar más en este sentido.





LEYENDA

Grupo	Simb.	$\bar{N}_G$	$\bar{W}$	$\bar{L}_L$	$\bar{L}_P$	distancia media puntos centroides
1		10	0.4	0.88	0.31	2.9
2		34	0.36	0.87	0.34	4.8
3		2.1	0.39	0.92	0.34	1.7

- arcilla fisurada con bolsones de carbonato, consistencia media a firme
- arcilla arenosa con bolsones de carbonato y marga, consistencia muy firme
- arcilla calcárea con bolsones de carbonato, consistencia dura

Fig. 3. Corte ingeniero-geológico propuesto. Obra 20 Plantas San Rafael y Espada.

AGRADECIMIENTOS

Queremos agradecer la colaboración de los ingenieros geofísicos RN Doctor en Ciencias José R. Alfonso Roche, MSc, Profesor Titular del Departamento de Geofísica del ISPJAE y Carlos Sacasas León, Profesor Auxiliar del Departamento de Geofísica del ISPJAE que nos iniciaron en el estudio del análisis estadístico multivariado y sus aplicaciones.

REFERENCIAS

1. ALFONSO ROCHE, J. R.: "Tratamiento numérico de datos geológicos". Curso de Posgrado, ISPJAE, Ciudad de La Habana, 1981.
2. GRINACO, DAP: "Perfiles litológicos de las obras de Ciudad de La Habana, Cayo Hueso, Cayo Hueso 20 plantas no. 2, Moldes Deslizantes 20 plantas San Rafael y Espada, 20 plantas Tulipán y Loma". Informes Técnicos, Ciudad de La Habana, 1972.
3. IBM: System/360 Scientific Subroutine Package, Versión III. IBM Corp., 1970.
4. VALCARCE ORTEGA, R. M.: "Análisis estadístico multivariado aplicado a la ingeniería geológica". Trabajo de Diploma, ISPJAE, Ciudad de La Habana, 1982.