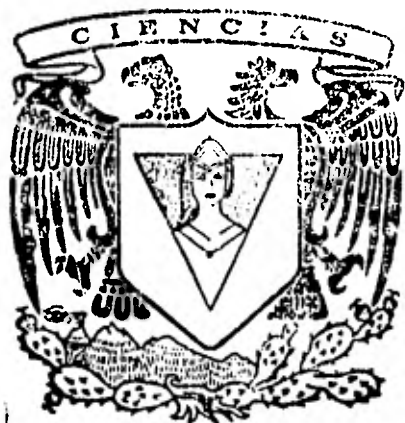


UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS



**EVALUACION DE INFORMACION GEOLOGICA
A TRAVES DE MODELOS ESTADISTICOS.**

T E S I S
QUE PARA OBTENER EL TITULO DE:
A C T U A R I O
P R E S E N T A:
ROCIO HAYDEE CAMERO MARTINEZ

MEXICO, D. F.

1982



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

I N D I C E

| | | |
|--|----|----|
| INTRODUCCION | 1 | |
| CAPITULO 1. CONCEPTOS GEOLOGICOS | | |
| 1.1 GEOLOGIA | 4 | |
| 1.2 ORIGEN Y CARACTERISTICAS DEL PETROLEO. | 6 | |
| 1.3 METODOS EXPLORATORIOS | 12 | |
| 1.3.1 La Exploración Petrolera | 12 | |
| 1.3.2 Métodos Geológicos | 13 | |
| 1.4 GEOQUIMICA ORGANICA | 16 | |
| 1.4.1 Sobre la Materia Orgánica. | 16 | |
| 1.4.2 Estudio Geoquímico - Orgánico. | 19 | |
| 1.4.3 Acerca de la Información | 24 | |
| CAPITULO 2. ANALISIS MULTIVARIADO | | |
| 2.1 CONCEPTOS ESTADISTICOS | 31 | |
| 2.2 ANALISIS DISCRIMINANTE | 36 | |
| CAPITULO 3. APLICACION DEL METODO Y RESULTADOS. . . | | 54 |
| CONCLUSIONES | | 68 |
| SALIDA DEL SUBPROGRAMA DISCRIMINANT ANALYSIS | | 70 |
| BIBLIOGRAFIA | | 99 |

INTRODUCCION

Es de todos conocido que el desarrollo económico de un país, está sujeto en gran parte a sus industrias básicas, mismas que se encargan de satisfacer las necesidades fundamentales de la sociedad; por lo que es muy importante el estudio de las mismas. Dentro de estas industrias se encuentra la petrolera, puesto que ésta proporciona gran cantidad de productos necesarios para el crecimiento de una nación.

México es un país que inició su actividad petrolera a principio de siglo en algunas áreas costeras del Golfo de México. A la fecha nuestro país puede ser considerado como privilegiado en lo relativo a reservas totales de hidrocarburos. Las reservas mencionadas se definen mediante trabajos exploratorios apropiados, por lo que el incremento de las mismas se puede encauzar a través de una política razonable que permita un mayor éxito en la búsqueda de hidrocarburos.

Las áreas que han contribuído con mayores volúmenes de hidrocarburos a la industria petrolera son: la Sonda de Campeche, el Area Mesozoica de Chiapas-Tabasco, la Cuenca de Sabinas en Coahuila, la Plataforma de Tuxpam, la Cuenca de Tampico-Misantla y el Paleocanal de Chicontepec en Veracruz.

De acuerdo con estudios geológicos tanto de superficie como de subsuelo, durante los períodos Jurásico Superior, Cretácico Inferior y Cretácico Medio, se depositaron grandes volúmenes de sedimentos acompañados por abundante materia orgánica, misma que posteriormente dio origen a los hidrocarburos que actualmente se explotan en las áreas mencio

nadas anteriormente.

Favorece a la Exploración el desarrollo de nuevas técnicas, así como la aplicación de nuevos métodos; pues bien, - este trabajo tiene por objeto mostrar la aplicación de un método Estadístico (comprendido dentro del Análisis Multivariado) a una información geológica obtenida mediante análisis geoquímicos, con el fin de verificar los criterios geoquímicos existentes en la exploración petrolera.

El método llamado Función Lineal Discriminante requiere de la existencia previa de la información en grupos, de tal manera que los individuos de un grupo muestren homogeneidad y los individuos de grupos diferentes sean heterogéneos entre sí.

El Análisis Estadístico arrojará una función lineal en base a la cual podrán ser asignadas nuevas observaciones -- (muestras de roca en este caso) a alguno de los grupos originales.

Se trabajarán tres grupos llamados: Pozos Productores, - Areas con Buenas Posibilidades y Pozos Secos.

Así, el objetivo es verificar, a través del Análisis Estadístico, el potencial generador de las rocas estudiadas - definido por los Análisis Geoquímicos en las mismas.

Lo interesante en este estudio no se basa en el hecho de descubrir o evaluar métodos estadísticos para la estimación del grado de incertidumbre en la exploración de los hidrocarburos, ésto es, no se pretende seguir el lineamiento que hasta ahora se ha enfocado, el cual consiste en ver que la probabilidad de las características geológicas percibidas de los datos corresponda a la realidad geológica; sino

más bien en la aplicación de métodos multivariados a información geológica ya obtenida, perteneciente a estudios geoquímicos recientes, para llegar a confirmar que en realidad se trata de una zona; productora, seca o área con buenas posibilidades, y conocer posteriormente la asignación de nueva información de carácter geoquímico a una de tales zonas.

Lo que es necesario justificar ahora es el hecho de haber escogido precisamente este método.

Se conoce que los problemas de Clasificación en Antropología y Botánica llevaron al concepto de Función Discriminante, por otro lado se le considera como uno de los procedimientos más usados en las Ciencias de la Tierra. Finalmente, la importancia del petróleo en esta época y el uso de datos reales y cuantitativos, proporcionan el interés para realizar este trabajo.

La presente Introducción constituye la iniciación de esta Tesis. Enseguida se considera pertinente familiarizarse tanto con los términos y conceptos Geológicos como con las referencias acerca del petróleo (origen y características), etc., Exploración y Geoquímica (de donde se deriva el análisis de la información). La descripción de los términos anteriores constituye el Capítulo 1.

La explicación de los conceptos estadísticos se encuentra dentro del Capítulo 1.

Se pretende detallar en el capítulo 3 las características del subprograma DISCRIMINANT ANALYSIS (perteneciente al SPSS) conforme se presentan los resultados.

Como etapa final, se desarrollan las Conclusiones.

CAPITULO 1

CONCEPTOS GEOLOGICOS

1.1 GEOLOGIA

La Geología es la ciencia que estudia lo referente al -- origen, la estructura y la historia de la Tierra y de sus -- habitantes según su registro en las rocas.

Las rocas indican que los procesos que actualmente se de -- sarrollan en la superficie de la Tierra son los mismos que -- han actuado en el pasado a través de largas etapas. Estos -- procesos han estado regidos por dos tipos de fuerzas: Las -- Fuerzas Externas y las Fuerzas Internas (las primeras tien -- den a destruir las rocas expuestas al agua, aire, energía -- solar, etc., mientras que las segundas actúan desde el inte -- rior de la tierra ocasionando la elevación y depresión de -- grandes áreas superficiales). Entre los fenómenos pertene -- cientes a las fuerzas externas, se encuentran la erosión -- (conjunto de procesos que ocasionan la disolución y trans -- portación de las rocas de un lugar a otro), y la agradación -- (acumulación de las rocas formando estratos rocosos). Den -- tro de las fuerzas internas que afectan a las rocas se en -- cuentran los movimientos tectónicos (del griego Tectonikos -- que significa "de la construcción"). De los procesos que -- integran la tectónica de una área se encuentran;

El Fallamiento. Fracturas y fallas son el resultado de -- los movimientos efectuados cerca de la superficie; esto es, -- toda deformación discontinua que se manifiesta esencialmen -- te por planos de fractura.

El Plegamiento. Se dice que una roca se pliega, cuando -- una superficie de referencia definida antes del plegamiento -- como plana, se transforma en una superficie curva. El anti

clinal es un tipo de plegamiento consistente en la formación de pliegues ascendentes en forma de arco (ver Fig. 1,1-A).

La roca sólida no siempre se encuentra a la misma profundidad, esto es, en diversos lugares se aprecia en la superficie material suelto; suelo en los campos de cultivo, arena y grava en los bancos de las corrientes fluviales, arcilla en los cortes de los caminos; pero a través de excavaciones se puede llegar a la roca firme. En ocasiones la profundidad para llegar a esta roca es muy pequeña y a veces ni siquiera existe, quedando expuestas grandes masas de roca descubierta. Frecuentemente, la regolita (capa constituida por diversos materiales sueltos) es la capa que cubre a la roca que se encuentra bajo la superficie. Las masas de roca que no están cubiertas constituyen los afloramientos.

Las rocas de la litósfera (porción sólida de la Tierra) se pueden clasificar en tres grupos según su origen: Igneas, Sedimentarias y Metamórficas. Se hará mención sólo de las rocas sedimentarias dada su relación en la exploración de los hidrocarburos.

Las rocas sedimentarias se forman usualmente a partir de otras rocas preexistentes, cuando sobre éstas actúan los diferentes agentes geológicos como la erosión, la degradación, la transportación, etc.; cuando los sedimentos son abandonados por el agente cargador sufren el proceso de litificación (consolidación). Los sedimentos a su vez al descansar en capas, forman los estratos. De ahí que la principal característica de las rocas sedimentarias como grupo es que están estratificadas.

La arenisca, la lutita, la caliza, el carbón, la coquina, son ejemplos de rocas sedimentarias.

1.2 ORIGEN Y CARACTERISTICAS DEL PETROLEO

Se le encuentra en el subsuelo impregnando principalmente a formaciones de tipo arenoso o calcáreo. Su estado es líquido, sólido o gaseoso según la temperatura y presión a que se encuentre sometido. Su color varía entre el ámbar y el negro, y su densidad es menor que la del agua.

La composición química del petróleo (de 95 a 99% de carbono e hidrógeno) suministra información como para que la mayoría de los geólogos la consideren de origen orgánico; hasta ahora esta teoría es la más aceptada, tomando como base lo siguiente:

- a) Los estudios de laboratorio de muestras de petróleo han revelado ciertas propiedades ópticas que sólo se encuentran en las sustancias orgánicas.
- b) El contenido de nitrógeno y otras sustancias en el petróleo sólo pueden proceder de materiales orgánicos.
- c) Se han localizado la mayor parte de los yacimientos en lugares donde hace millones de años existieron lagos o mares.

La intervención de esta última fundamentación se justifica al complementar la idea del origen del petróleo: proviene de la descomposición lenta de los restos o residuos de animales o vegetales, una vez que fueron sepultados en el lodo y en la arena de los mares prehistóricos someros principalmente.

Actualmente no se han logrado conocer con precisión los procesos de formación del petróleo. No obstante, se asegura la participación de incrementos en la temperatura y en

la compresión de los sedimentos a través de largos períodos de tiempo.

Detallando aún más, llegamos a la teoría siguiente, que es la más difundida en nuestros días:

- 1) La abundancia de plantas en la superficie y cerca de ésta, así como de organismos marinos simples, constituyen la materia orgánica original. Por mediciones y observaciones se tiene conocimiento de que se produce una cifra aproximada de 400 Kgs. de materia proteica por hectárea -- por año, siendo de 2.5 toneladas en las aguas -- más productivas.
- 2) La materia orgánica llega a acumularse en el -- fondo de cuencas donde el agua se estanca y es -- pobre en oxígeno y en consecuencia ni es des- -- truída por oxidación ni es devorada por los ani- -- males necrófagos, pero sí sufre el ataque y la -- descomposición de materias que separan y libe- -- ran el oxígeno, nitrógeno y otros elementos de- -- jando principalmente al carbono e hidrógeno co- -- mo residuales. Los sedimentos ricos en materia orgánica que actualmente se encuentran en proce- -- so de acumulación contienen gran cantidad de -- bacterias.
- 3) La materia orgánica original queda sepultada ba- -- jo sedimentos más finos que se van depositando- -- posteriormente, obstruyéndose de esta manera, -- las bacterias existentes. A causa de la pre -- sión, el calor y el tiempo se efectúan los cam- -- bios posteriores que convierten la sustancia or- -- gánica en gotitas de petróleo líquido y en mi- -- núsculas burbujas de gas.

- 4) Después de su formación, el petróleo se desplaza desde los lodos y las arcillas, en los cuales se había generado, hacia rocas más porosas, debido a que el peso cada vez mayor originado por los procesos de la compactación gradual de los sedimentos que lo contienen reducen el espacio entre las partículas de roca, expulsándolo hacia las capas cercanas de arena o arenisca donde los poros son mayores.

Los sedimentos en donde el aceite y el gas se han formado, se llaman rocas fuente o rocas generadoras, pudiendo pertenecer a diversas edades geológicas; aquellos porosos y permeables, en los cuales quedan atrapados los hidrocarburos, se conocen como rocas almacenadoras. Por ejemplo las arenas, las areniscas, algunas calizas y dolomitas son tipos de rocas almacenadoras.

- 5) Debido a su menor peso específico que les permite flotar, y quizás por la circulación de las aguas subterráneas, el aceite y el gas se trasladan hacia arriba a través de las rocas porosas y permeables o bien a través de fallas o fracturas hasta alcanzar la superficie, disipándose o deteniéndose a causa de una trampa formando un yacimiento (ver Fig. 1.1).

Así, las estructuras (trampas estructurales) son deformaciones de las rocas yacimiento, mismas que detendrán la migración y ocasionarán su acumulación,

Para que se efectúe el atrapamiento, es necesaria igualmente una roca impermeable para que cubra a la receptora, para evitar el escape del petróleo a la superficie (roca sello).

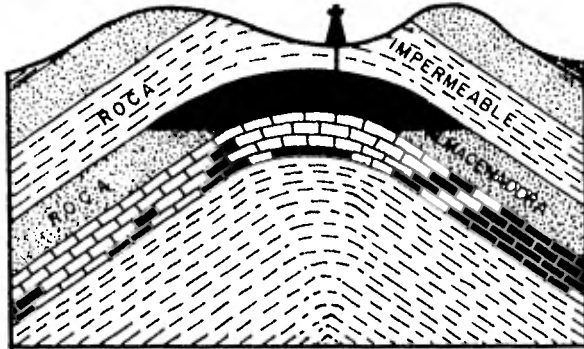
Los anticlinales, los cierres contra falla, los domos salinos y varios tipos de trampas estratigráficas resultan ser buenos receptáculos.

En las diversas partes del mundo donde se encuentra el petróleo, se le refiere siempre a niveles estratigráficos, es decir, a rocas que varían en edad desde el Cámbrico al Terciario Tardío. En algunas áreas se presenta a pocos metros bajo la superficie (Campos de la Faja de Oro-Alazán, Potrero del Llano y Ebano-Pánuco), en otras debe ser extraído desde profundidades de miles de metros (Campos de Chiapas-Tabasco y Golfo de Campeche). A modo de ubicación en el tiempo geológico de que se hace mención, se incluye un cuadro de estos tiempos donde se representan los períodos que intervienen en esta investigación (fig. 1.2).

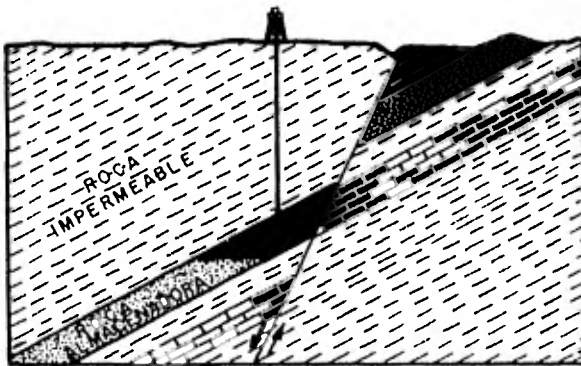
Siendo el aceite y el gas las dos clases principales de petróleo, puede definirse éste como una sustancia gaseosa, líquida o sólida, que se presenta en la naturaleza y que está formado principalmente de compuestos químicos constituidos por carbono e hidrógeno.

Luego, un yacimiento es el resultado de la acumulación subterránea de petróleo o gas en un receptáculo limitado por barreras geológicas.

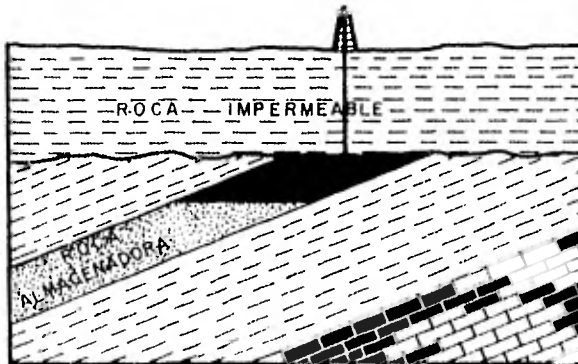
TRES DE LOS TRES TIPOS PRINCIPALES DE TRAMPAS PETROLIFERAS, EL PETROLEO (EN NEGRO) FLOTA SOBRE AGUA, SATURA LA ROCA RECEPTORA FLOTANDO SOBRE EL AGUA SUBTERRANEA Y QUEDA RETENIDO POR UNA CUBIERTA DE LUTITA. EL PETROLEO LLENA SOLO LOS ESPACIOS POROSOS DE LA ROCA.



TRAMPA ESTRUCTURAL A



TRAMPA COMBINADA B



TRAMPA ESTRATIGRAFICA C

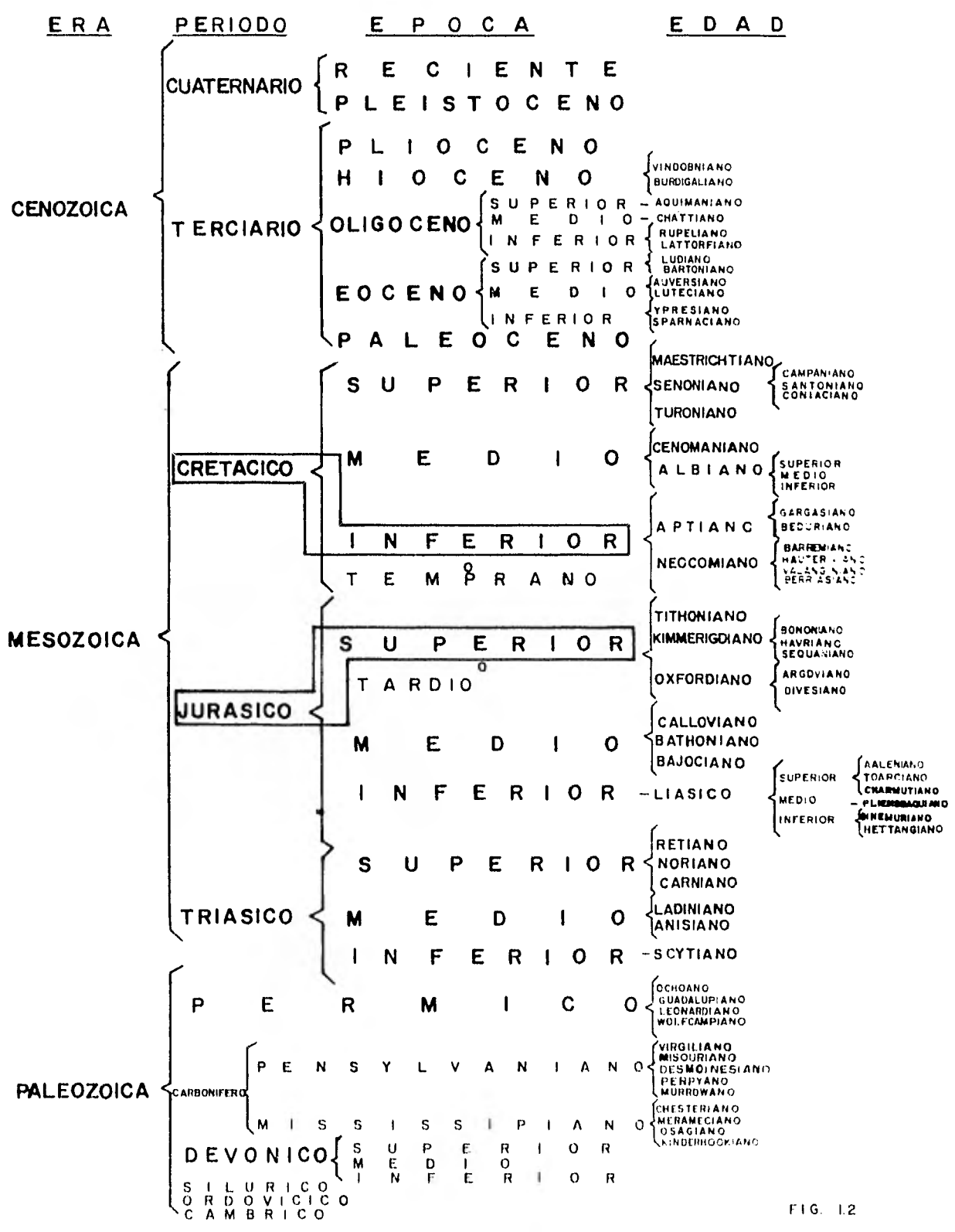


FIG. 1.2

1.3 METODOS EXPLORATORIOS

1.3.1 La Exploración Petrolera

Uno de los fines más importantes que persigue la Geología, es el de la Exploración, así como el desarrollo y la conservación de los recursos naturales. La información geológica es básica para explotar materias primas indispensables para el desarrollo industrial.

La industria petrolera es uno de los campos donde los geólogos ejercen su especialidad; siendo precisamente esta industria la involucrada en el presente estudio, el término exploración implica de por sí la actividad petrolera.

La Exploración Petrolera tiene por objeto precisamente determinar cuándo y dónde se reunieron las condiciones necesarias para la formación de un yacimiento.

En un principio el objetivo de los primeros exploradores fué el de encontrar manifestaciones superficiales de petróleo (llamadas chapopoterías) para perforar pozos próximos a estas manifestaciones. Este tipo de búsqueda se considera pobre en bases científicas; los pozos eran localizados utilizando como guía dichas manifestaciones.

Con la introducción de los Métodos Geológicos y Geofísicos aumentaron las posibilidades de éxito en las perforaciones exploratorias, ya que se tuvo mayor conocimiento de las rocas y de las estructuras existentes tanto en la superficie como en el subsuelo. Los trabajos de Geología Superficial y Geoquímica, pertenecen a los Métodos Geológicos y la Gravimetría, Magnetometría, Técnicas Eléctricas y

Sismología constituyen algunos de los Métodos Geofísicos.

La información derivada de todos éstos métodos se analiza para obtener una interpretación constituida en parte -- por un plano integrado donde se involucran todos los datos -- y se decide la localización de la perforación de los Pozos Exploratorios en el lugar más apropiado según la información existente.

1.3.2 Métodos Geológicos

La Geoquímica es una de las ramas de los Métodos Geológicos.

Los estudios Geoquímicos se realizan en diferentes tipos de muestras:

1. Muestras de Afloramientos
2. Muestras de Suelos
3. Muestras del Subsuelo
4. Muestras de Manifestaciones

1. Muestras de Afloramientos: Estas muestras se colectan en los sedimentos que constituyen a las rocas expuestas en la superficie y usualmente deben corresponder a calizas arcillosas o arcillas.

El estudio Geoquímico de este tipo de muestras permite evaluar el potencial generador de una cuenca sedimentaria.

2. Muestras de Suelos: Las muestras de suelos son las -- obtenidas de los sedimentos que son el resultado de la acción de los agentes meteóricos sobre las rocas aflorantes.

El estudio de estas muestras detecta el escape hacia la superficie de los hidrocarburos provenientes de un yacimiento situado en el subsuelo. Se colectan a una profundidad -- de dos metros con un espaciamento reticulado de 800 metros.

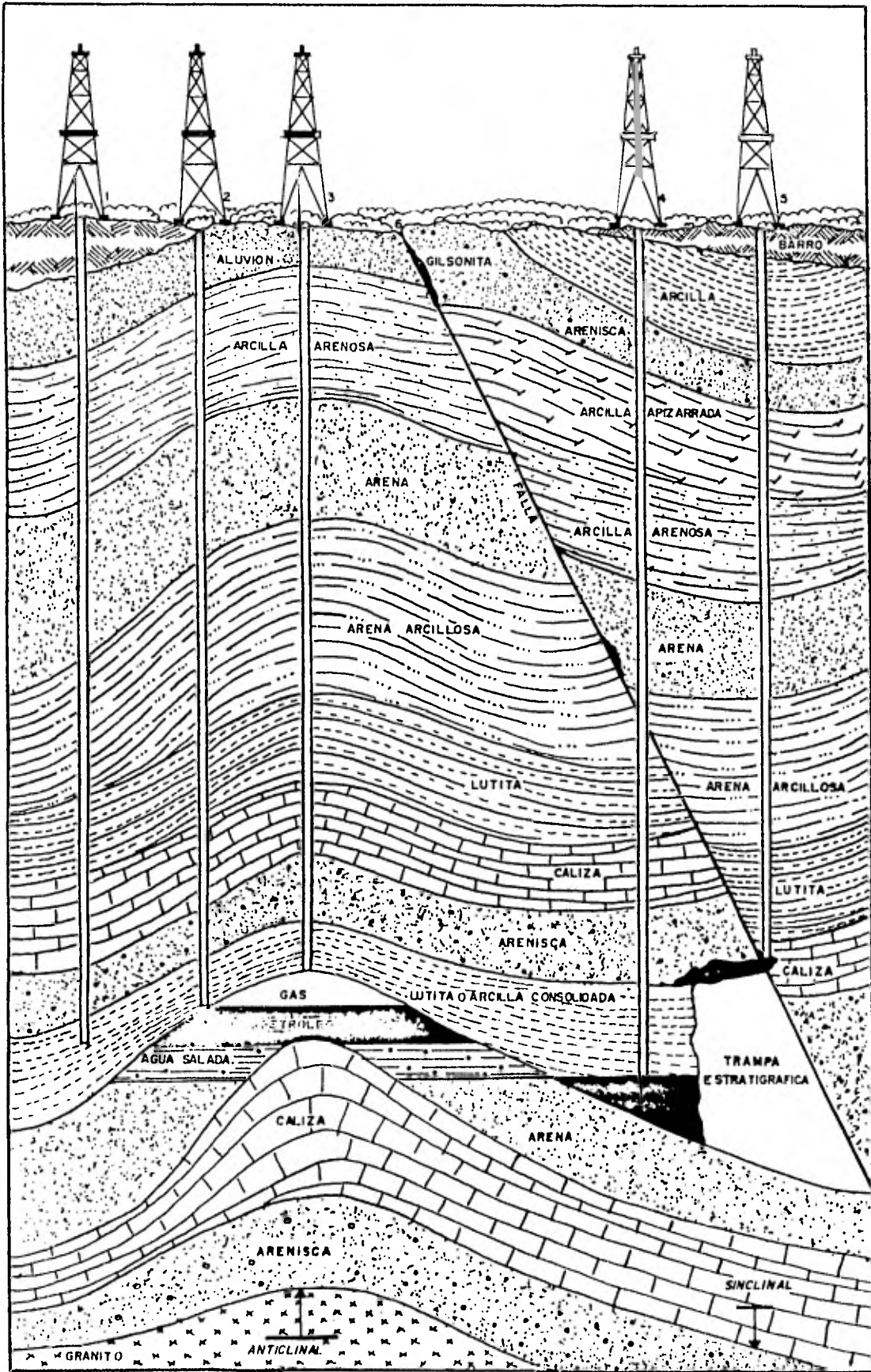
3. Muestras del Subsuelo: Son las muestras que se colectan a medida que avanza la perforación de los pozos y pueden ser de canal y/o de núcleo; el estudio de éstas permite evaluar el potencial generador de las rocas penetradas por la barrena y detecta y define los hidrocarburos que se encuentran entrampados en los yacimientos. Mediante el -- estudio de estas rocas se puede correlacionar entre rocas generadoras e hidrocarburos generados.

Las muestras de canal se toman cada 30 metros y las de núcleo cada vez que se presente una manifestación o bien -- cuando se juzgue conveniente.

4. Muestras de Manifestaciones: Son colectadas en la superficie o en el subsuelo y usualmente el análisis de éstas permite hacer correlaciones entre éllas y las diferentes -- rocas generadoras existentes en la área de estudio. Se colectan todas las manifestaciones que se encuentran en la -- superficie, igualmente se colectan todas aquellas que se -- presentan a medida que se perforan los Pozos Explorato- -- rios (ver Fig. 1.3).

Los datos utilizados en esta investigación proceden del estudio analítico de muestras de roca aflorante, así como de muestras del subsuelo (pozos), pudiéndose decir que corresponden al estudio Geoquímico Orgánico que forma parte de la Exploración del Petróleo.

Considerando que los estudios geoquímicos en nuestro -- país dieron inicio en fechas muy recientes, la información existente es relativamente escasa, por lo que los resultados de este estudio posiblemente puedan ser mejorados al -- incrementarse el volúmen de la información.



CORTE GEOLOGICO DE UN YACIMIENTO PETROLIFERO

POZO 1: SECO O IMPRODUCTIVO A MAYOR PROFUNDIDAD, AGUA SALADA. POZO 2: GRAN PRODUCTOR INICIAL DE PETRO-
LEO. POZO 3: GRAN PRODUCTOR INICIAL DE GAS Y POSTERIORMENTE DE PETROLEO. POZO 4: PRODUCTOR DE PE-
TROLEO POR POCO TIEMPO. POZO 5: PRODUCTOR DE PETROLEO POR UN PERIODO MUY REDUCIDO. 6: AFLORACION
DE GILSONITA (chaporera).

1.4 GEOQUIMICA ORGANICA

1.4.1 Sobre la Materia Orgánica

El fenómeno de la fotosíntesis clorofiliana en los vegetales es la base para la producción de la materia orgánica que los constituye, parte de la cual es preservada e introducida en los sedimentos.

La materia orgánica se divide en cuatro contribuyentes principales: phytoplakton, zooplakton, plantas superiores y bacterias.

Aunque los organismos se distinguen por su estructura química detallada, todos ellos están compuestos básicamente de los mismos constituyentes químicos: lípidos, proteínas, carbohidratos y lignina.

Ciertas épocas geológicas fueron más favorables que otras para la producción de la materia orgánica; el caso por ejemplo del Cretácico Medio (hace aproximadamente 100 millones de años) con anchos mares que invadieron la plataforma continental, propició fuerte producción de materia orgánica en muchas cuencas de sedimentación.

Existen tres etapas en la evolución de la materia orgánica en sedimentos: la Diagénesis, la Catagénesis y la Metagénesis. En la primera el principal agente transformador son las bacterias; en depósitos de sedimentos recientes, los rearrreglos geoquímicos se efectúan a profundidades someras; en esta fase los biopolímeros llegan a constituir los geopolímeros, conocidos también bajo el nombre de kerógeno (materia orgánica insoluble). El kerógeno es la forma más abundante de materia orgánica fósil que se encuentra en forma

dispersa en las rocas sedimentarias, es sólido e insoluble en todos los solventes orgánicos comunes. Durante la Catagénesis se lleva a cabo su degradación termal, la cual es responsable de la generación del aceite y gas. Esta etapa es el resultado del aumento en la temperatura de los estratos durante el sepultamiento en las cuencas sedimentarias.

Si el Kerógeno se observa al microscopio se aprecia que está constituido por partículas, algunas de las cuales conservan formas características del material de origen (al- gas, esporas, polen, tejidos vegetales) pero en general -- presenta aspecto amorfo (fig. 1.4).

EL KEROGENO:

MOLECULAS GIGANTES DE TRES DIMENSIONES QUE SE TRANSFORMAN LENTAMENTE EN PETROLEO Y GAS

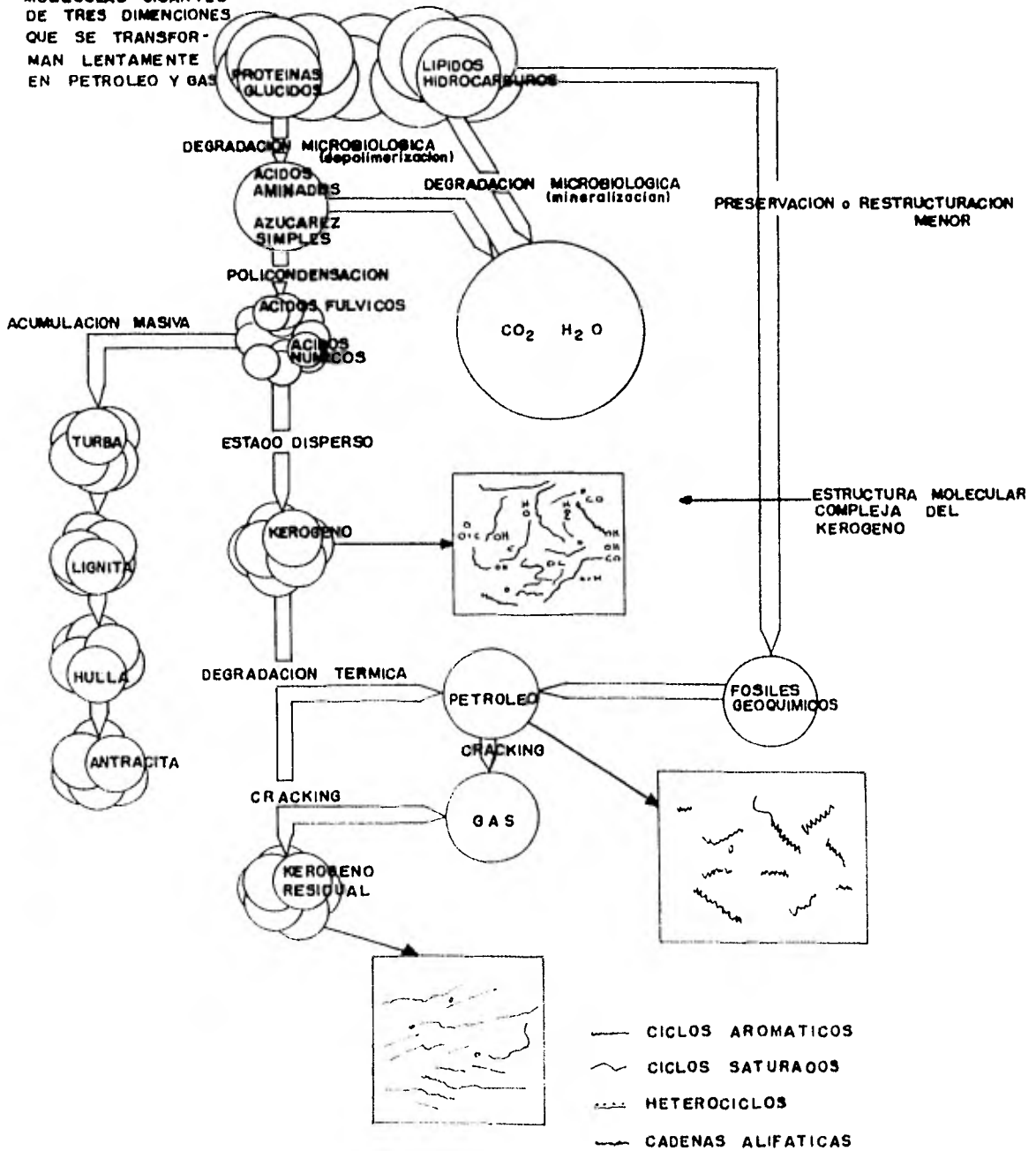


FIG. 1.4

1.4.2 Estudio Geoquímico-Orgánico

Actualmente no existe aún un método directo capaz de detectar yacimientos petroleros en el subsuelo. Primeramente fueron las manifestaciones superficiales las que revelaron la existencia de yacimientos sepultados, después a través de métodos indirectos se determinaban las estructuras subterráneas probables acumuladoras de hidrocarburos, algunas veces con éxito y muchas otras sin él puesto que dichos métodos sólo localizaban estructuras almacenadoras y no su contenido.

En la exploración petrolera son cada vez más numerosos los especialistas encargados de localizar tanto las regiones donde se encuentra la roca generadora bajo condiciones térmicas apropiadas para producir hidrocarburos, así como aquellas donde se localizan las rocas porosas y permeables bajo condiciones estructurales adecuadas para almacenar los hidrocarburos generales.

Los Geoquímicos son los especialistas que analizan y estudian la materia orgánica soluble e insoluble de las rocas sedimentarias pretendiendo conocer el potencial generador y productor de una cuenca. Para tal fin, se determina: la cantidad de materia orgánica en las rocas, si el tipo de la misma es adecuado para generar aceite, gas o ambos, si la historia térmica ha sido favorable para la generación de preservación de los hidrocarburos y si éstos ya han migrado.

De los análisis geoquímicos que se efectúan de manera rutinaria en el estudio de las rocas generadoras, se tienen entre otros: el Contenido de Carbón Orgánico, el Contenido de Carbonatos, el Índice de Madurez Termal, el Tipo y el co

lor de la materia orgánica; estas características fueron utilizadas para desarrollar el presente trabajo. El Contenido de Carbón Orgánico nos indica la cantidad de materia orgánica presente en la roca estudiada, la que en condiciones apropiadas puede generar hidrocarburos. El porcentaje mínimo para considerar a una roca como generadora es de 0.5 para las arcillosas y 0.2 para las calcáreas o calizas; por esta razón es importante definir el contenido de carbonatos en las rocas analizadas.

No obstante la importancia del volumen de la materia orgánica presente en una roca, es inadecuado el considerar solamente dicha riqueza para definir completamente una roca generadora; deben ser consideradas además la madurez termal y el tipo de materia orgánica.

La materia orgánica depositada junto con los sedimentos, al estar sujeta a procesos termales, se transforma, pudiendo de esta manera generar hidrocarburos. El grado de transformación depende de dos factores principales: la temperatura y el tiempo. La Diagénesis cambia la materia orgánica en Kerógeno; éste recién formado es casi incoloro o de color amarillo claro; el incremento en las temperaturas que actúan sobre las rocas modifican el color del Kerógeno en forma gradual a: amarillo, amarillo naranja, café claro, café oscuro y finalmente negro; estos colores indican el grado de transformación que ha sufrido la materia orgánica, utilizándose para fines prácticos valores numéricos para cada color siendo éstos del 1 al 5 (fig. 1.9), correspondiendo éstos valores el calificativo de Índice de Madurez Termal.

La calidad de la materia orgánica contenida en una unidad de roca determina el tipo y la composición química de los hidrocarburos que se generan, durante la Catagénesis.

La materia orgánica que da origen al Kerógeno puede ser de cuatro tipos diferentes, dependiendo de sus constituyentes: algáceo o amorfo, herbáceo, maderáceo y carbónáceo. Cuando predomina el primer y segundo tipo, se considera la roca como generadora de hidrocarburos líquidos, en cambio si se encuentra kerógeno tipo maderáceo-carbónáceo producirá principalmente gas. Cuando existe materia mixta la roca generará petróleo y gas.

En sus primeras etapas de generación las rocas inmaduras producen sólo gas metano por la acción de las bacterias sobre la materia orgánica contenida. La generación significativa de aceite a partir del kerógeno amorfo se inicia con suficiente tiempo y temperatura.

Factor profundidad

Se mencionaba en la Introducción que el Jurásico Superior y Cretácico Inferior son los períodos en los cuales se generaron grandes volúmenes de hidrocarburos; por esa razón fueron seleccionados para efectuar los estudios analíticos de algunas rocas que los integran (se incluyen igualmente formaciones del Cretácico Superior, aún cuando éstas no son tan buenas generadoras como las anteriores). Dichas rocas se encuentran a profundidades muy variables, pudiéndose observar aflorando o bien pueden ser detectadas a más de 4,000 metros de profundidad, dependiendo de diversos factores geológicos que hayan actuado sobre ellas, entre los que se encuentran como más importantes los movimientos tectónicos. En la tabla 1,5 se presentan las columnas formacionales penetradas por los pozos estudiados. Tomando como referencia La Virgen, se confirma lo mencionado anteriormente.

La profundidad de los pozos secos y productores varía -

entre los 50 y los 5,000 metros, en tanto que las de los afloramientos se les da un valor constante de 10 para efecto de cálculos, puesto que en éstos no existe el factor -- profundidad.

P
R
O
F
U
N
D
I
D
A
D

| POZOS PRODUCTORES | | | POZOS SECOS | |
|---|-----------------------------------|---------------------|--------------------------------|--------------------------|
| POZO 3 FORMACION | POZO 1 FORMACION | POZO 2 FORMACION | POZO 2 FORMACION | POZO 1 FORMACION |
| 50 100 Cretacica Superior Austin | | | | |
| 450 Eagle Ford | Del Rio Georgetown Klamichi | | | |
| 720 Buda | | | Austin | |
| 810 Del Rio Georgetown | Tamaulipas Superior | | | |
| 930 Tamaulipas Superior | | | | |
| 1000 1100 La Peña | La Peña | | | |
| 1350 1409 La Peña Cupida | | | Washita Indiferenciado | |
| 1530 1560 1615 La Virgen | | La Virgen | | La Virgen |
| 1723 1785 | La Virgen | Taraises | | La Mula Padilla |
| 1910 | | La Casita | | |
| 2000 2099 2145 2299 | La Mula | La Gloria | | Barril Viejo Menchaca |
| 2470 2500 2611 La Mula | Barril Viejo | | | La Casita |
| 2885 2920 3000 | Menchaca La Casita | | La Peña Cupido La Virgen | |
| 3215 | | | | |
| 4000 | | | Padilla | |
| 4295 | | | | |
| 4805 | | | Taraises | |
| 5000 | | | | |

Fig 15

1.4.3 Acerca de la Información

Se dijo en la Introducción que se requería de la construcción de grupos para la aplicación del método a usar.

Los datos tomados de una región del Noreste de México han sido organizados de tal manera que se ha logrado la configuración de tales grupos a través de la consulta de: Informes de pozos productores, planos de áreas atractivas - según el estudio geoquímico de las posibles rocas generadas de los períodos citados, así como de los informes de los pozos que han resultado secos.

Se cuenta entonces con datos analíticos de pozos productores, áreas con buenas posibilidades - o prospectos - y pozos secos. De acuerdo a ésto se formaron los siguientes grupos:

| | | |
|--------------------------------|---------|--|
| | | Pozo 1 : 39 muestras de roca |
| GRUPO 1 | | Pozo 2 : 29 muestras |
| POZOS PRODUCTORES | | Pozo 3 : 80 muestras |
| | | <hr/> |
| | | 148 muestras de roca |
| | GRUPO 2 | Prospecto 1 : 1a. Sección: 35 muestras |
| AREAS CON BUENAS POSIBILIDADES | | 2a. Sección: 24 muestras |
| | | 3a. Sección: 24 muestras |
| | | <hr/> |
| | | 83 Muestras de roca |
| | | Prospecto 2; 1a. Sección: 37 muestras |
| | | 2a. Sección: 48 muestras |
| | | 3a. Sección: 12 muestras |
| | | <hr/> |
| | | 97 muestras de roca |
| | GRUPO 3 | Pozo 1: 43 muestras de roca |
| POZOS SECOS | | Pozo 2: 110 muestras de roca |
| | | 153 muestras de roca |
| | | <hr/> |
| | | 148+180+153 = 481 muestras de roca |

Los pozos productores son aquellos de los cuales se ha obtenido producción comercial de hidrocarburos; en los pozos secos no se ha obtenido dicha productividad. Las muestras del grupo dos corresponden a rocas aflorantes colectadas en áreas que de antemano han sido consideradas con buenas posibilidades de producir hidrocarburos.

Cabe hacer notar que las características o condiciones necesarias para la evaluación potencial generadora de la roca no son utilizadas para la separación de los datos en grupos, sino más bien para realizar una discriminación, -- con una categorización posterior de nuevas muestras de roca a los grupos ya formados.

Esto es, la Profundidad, la Madurez Termal, el Contenido de Materia Orgánica (M.O.), el Contenido de Carbonatos, el Tipo y el Color^{1/} de la M.O. serán las variables que intervendrán como mediadoras para llegar a formar una combinación lineal (ver 3.1) en la cual se podrán ubicar los valores de las muestras de roca estudiadas posteriormente a los grupos referidos.

A las seis variables usadas en este trabajo le fueron asignados valores numéricos con el fin de aplicar el método estadístico seleccionado.

La información sobre el Índice de Madurez, el Tipo y el Color de la M.O. se presenta en un formato como el de la fig. 1.6.

^{1/}El color de la materia orgánica es una característica que está implícita en el Índice de Madurez Termal, no obstante, para el efecto de los cálculos estadísticos se consideró cada una como variable.

En la tabla 1.7 se muestra un ejemplo de cómo se indican los valores obtenidos en el estudio de laboratorio.

Como puede apreciarse, en las tres características de interés (fig. 1.6) se especifican las divisiones necesarias para la asignación numérica correspondiente.

FIG. 1.7 20 PRIMERAS OBSERVACIONES PERTENECIENTES AL POZO I DEL GRUPO I (POZOS PRODUCTORES) EN LAS VARIABLES PROFUNDIDAD, INDICE DE MADUREZ Y TIPO Y COLOR DE LA MATERIA ORGANICA.

HOJA DE TRABAJO PARA EL EXAMEN VISUAL DEL KEROGENO

| No MUESTRA | PROFUNDIDAD | POBLACION INDIGENA (INTERPRETADA) | | | | CARACTERISTICAS GENERALES | | | POBLACION (ES) RE-CIRCULADA Y/O RETRAJADA | | RESUMEN DEL TIPO DE MATERIA ORGANICA | |
|------------|-------------|-----------------------------------|-------------------|-------------------|-------------------|---------------------------|----------------------|---|---|-------------------|--------------------------------------|------|
| | | TIPO DE MAT ORGANICA | INDICE-DE-MADUREZ | INDICE DE MADUREZ | INDICE DE MADUREZ | COLOR DE LA MAT ORGANICA | EDO. DE LA MAT. ORG. | % | TIPO DE MAT. ORGANICA | INDICE DE MADUREZ | TIPO | M.O. |
| 1 | 505 m | | | 2.78 | | | | | | 4 | 2 | |
| 2 | 550 | | | 3.10 | | | | | | 4 | 2 | |
| 3 | 620 | | | 3.17 | | | | | | 4 | 2 | |
| 4 | 870 | | | 3.17 | | | | | | 4 | 2 | |
| 5 | 1110 | | | 3.55 | | | | | | 4 | 2 | |
| 6 | 1615 | | | 3.90 | | | | | | 5 | 2 | |
| 7n | 1356 | | | 3.50 | | | | | | 4 | 2 | |
| 8n | 1981 | | | 3.90 | | | | | | 5 | 2 | |
| 9n | 1990 | | | 3.90 | | | | | | 5 | 2 | |
| 10n | 1999 | | | 3.90 | | | | | | 5 | 2 | |
| 11n | 2008 | | | 3.90 | | | | | | 5 | 3 | |
| 12n | 2015 | | | 3.98 | | | | | | 5 | 3 | |
| 13n | 2024 | | | 3.45 | | | | | | 4 | 2 | |
| 14n | 2033 | | | 3.98 | | | | | | 5 | 2 | |
| 15n | 2044 | | | 4.25 | | | | | | 5 | 3 | |
| 16n | 2053 | | | 3.98 | | | | | | 5 | 2 | |
| 17n | 2118 | | | 3.98 | | | | | | 5 | 2 | |
| 18n | 2127 | | | 3.85 | | | | | | 5 | 2 | |
| 19n | 2136 | | | 3.98 | | | | | | 5 | 3 | |
| 20n | 2145 | | | 4.40 | | | | | | 5 | 3 | |

no núcleo

FIG. 1.7

Tanto en el Índice de Madurez como en el Tipo y Color de la M.O. (fig. 1.6) se utiliza una escala de 1 a 5.

Fig. 1.8

| ALGACEO - AMORFO | | | | HERBACEO | | MADEROSO | CARBONACEO | POBLACION NO DEFINIDA | |
|----------------------|----------------|-------------------|-----------------|--------------------|--------------------------------|--------------------------------|-------------------------------|-----------------------|--------------------------|
| ALGAS IDENTIFICABLES | AMORFO-SIROPES | RELIQUIAS AMORFAS | ESBORAS - POLEN | HERBACEO DEGRADADO | RESTOS CUTICULARES MEMBRANOSOS | RESTOS MADEROSOS ESTRUCTURADOS | RESTOS CARBONACEOS INERTINITA | PROBITUMEN | RESTOS NO IDENTIFICABLES |
| 1 | | | | 2 | | 3 | 4 | 5 | |

El color de la M.O. lo caracterizan de amarillo claro - verdoso a negro en ocho divisiones, las cuales reduciendo a cinco se transforma en el dibujo siguiente (encontrándose la etapa madura en los números 3 y 4):

Fig. 1.9

| AMARILLO | | AMARILLO NARANJA | | CAFE CLARO | | CAFE OSCURO | | NEGRO | |
|------------------------|----------|------------------|--------------|------------|------|-------------|-------------|-------|-------|
| AMARILLO CLARO VERDOSO | AMARILLO | AMARILLO NARANJA | CAFE NARANJA | CAFE CLARO | CAFE | CAFE OSCURO | CAFE OSCURO | NEGRO | NEGRO |
| 1 | | 2 | | 3 | | 4 | | 5 | |

Se presentan las primeras 10 observaciones utilizadas en el análisis:

| Profundidad | Carbón Orgánico | Contenido de Carbonatos | Grupo | Índice de Madurez | tipo M.O. | Color M.O. |
|-------------|-----------------|-------------------------|-------|-------------------|-----------|------------|
| 0505 | 016 | 867 | 1 | 278 | 2 | 4 |
| 0550 | 011 | 933 | 1 | 310 | 2 | 4 |
| 0620 | 020 | 904 | 1 | 317 | 2 | 4 |
| 0870 | 007 | 992 | 1 | 317 | 2 | 4 |
| 1110 | 182 | 838 | 1 | 355 | 2 | 4 |
| 1615 | 028 | 915 | 1 | 390 | 2 | 5 |
| 1356 | 032 | 754 | 1 | 350 | 2 | 4 |
| 1981 | 009 | 956 | 1 | 390 | 3 | 5 |
| 1990 | 024 | 844 | 1 | 390 | 2 | 5 |
| 2008 | 033 | 796 | 1 | 390 | 3 | 5 |

Nota: El carbón orgánico, el contenido de carbonatos e índice de madurez son cantidades de dos decimales, un decimal y dos decimales respectivamente. Las cifras anteriores no lo especifican por entrar de esa manera a la computadora (ver pág. 1 de la salida del subprograma DISCRIMINANT ANALYSIS),

CAPITULO 2

ANALISIS MULTIVARIADO

2.1 CONCEPTOS ESTADISTICOS

El Análisis de Datos consistente de sólo una variable individual medida en cada unidad observacional da lugar a un análisis de datos univariados. Al tratar con un conjunto de medidas en cada unidad observacional, esto es, con un conjunto de varias y distintas variables en un número de individuos u objetos; se estará dando lugar al Análisis Multivariado de Datos. Así, el Análisis Multivariado se dedica al estudio simultáneo de variables aleatorias que están correlacionadas entre sí. Estas variables en conjunto aportan generalmente más al análisis que el estudio individual de la variable.

El campo donde el Análisis Multivariado puede aplicarse, es muy extenso: por ejemplo, en la investigación educacional psicológica el análisis multivariado trata con un grupo (o grupos diversos) de individuos donde cada uno de los cuales poseen valores en dos o más variables tal como pruebas mentales y otras medidas. La colección de medidas (ancho y largo) de pétalos de plantas de dos especies ejemplifica la aplicación en la Botánica.

Un contexto amplio para las medidas múltiples se encuentra también en la Geología. Datos apropiados pueden estar constituidos por análisis químicos donde las variables podrían ser composiciones porcentuales o partes por millón de elementos rastreados tales como medidas en ríos, descarga o carga de sedimentos, profundidad, sólidos disolventes, pH y oxígeno contenido y variables paleontológicas como --

puede ser el número de medidas hechas en especímenes de un organismo.

En el caso de éste trabajo las variables están constituidas por la Profundidad, el Contenido de Carbón Orgánico, el Contenido de Carbonatos, el Índice de Madurez Termal y el Tipo y Color de la Materia Orgánica; en cada muestra de roca.

La construcción de diversos métodos multivariados es una solución a los diversos problemas prácticos reales que van presentándose en las distintas áreas.

Dichos métodos han sido desarrollados y estudiados de una forma organizada y sistemática. En la mayoría de las situaciones reales es común la presencia de un modelo multivariado normal (aunque muchas veces funciona una metodología sin necesidad de apoyarse rigurosamente en una teoría acerca de la Distribución Normal).

Serán expresados algunos conceptos de la Distribución Normal definidos en un espacio multidimensional.

La variable aleatoria (v.a.) definida en varias dimensiones se encuentra dentro del campo multivariado. Esto mismo ocurre para las funciones de densidad, funciones de distribución, valores esperados, etc.; cuando son definidos en el espacio p-dimensional.

La extensión de la varianza de un vector aleatorio X de p-componentes es la matriz de varianzas y covarianzas:

$$E\left\{[X-E(X)] [X-E(X)]'\right\} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdot & \cdot & \cdot & \cdot & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdot & \cdot & \cdot & \cdot & \sigma_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{p1} & \sigma_{p2} & \cdot & \cdot & \cdot & \cdot & \sigma_{pp} \end{bmatrix} = \Sigma.$$

La función de densidad Normal Univariada está definida - por:

$$ke^{-\frac{1}{2}\xi(x-\beta)^2} = ke^{-\frac{1}{2}(x-\beta)\xi(x-\beta)} \dots (2.1)$$

donde ξ es positiva y k es tal que la integral de (2.1) sobre el eje de las x 's, es la unidad.

Se tiene el análogo para la función de densidad Normal - Multivariada al expresar:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_p \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_p \end{bmatrix} \quad \text{y} \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdot & \cdot & a_{1p} \\ a_{21} & a_{22} & \cdot & \cdot & a_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{p1} & a_{p2} & \cdot & \cdot & a_{pp} \end{bmatrix}$$

- X es un vector de las variables escalares x 's
- b es un vector de las constantes b 's
- A es una matriz simétrica definida positiva

El término $\xi(x-\beta)^2 = (x-\beta)\xi(x-\beta)$ es reemplazado por la forma cuadrática:

$$(X-b)' A(X-b) = \sum_{i,j=1}^p a_{ij} (x_i - b_i)(x_j - b_j)$$

Por lo tanto, la función de densidad de una distribución normal p -variada es;

$$f(x_1, \dots, x_p) = Ke^{-\frac{1}{2}(X-b)'A(X-b)} \dots (2.2)$$

donde

$$K = \sqrt{|A|} (2\pi)^{-\frac{1}{2}p}$$

De la búsqueda del significado de b y A (a través de encontrar los dos primeros valores esperados o momentos del vector aleatorio $X = (X_1, \dots, X_p)$ se tiene que:

$$\mu = E(X) = b$$

$$\Sigma = E(X-\mu)(X-\mu)' = A^{-1}$$

"Si la densidad de un vector aleatorio p-dimensional X es (2.2), el valor esperado de X es b y la matriz de varianzas y covarianzas es A^{-1} o lo que es lo mismo, dado un vector μ y una matriz definida positiva Σ ; se tiene la función de densidad Normal Multivariada

$$(2\pi)^{-\frac{1}{2}p} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)}$$

con μ esperanza de X y Σ matriz de varianzas y covarianzas"^{2/}

El i-ésimo elemento de μ es aún la media de X_i , el elemento i-ésimo diagonal de la matriz general Σ es todavía -

^{2/} ANDERSON, T.W. p.17 [1]

la i -ésima varianza; y ahora el ij -ésimo elemento σ_{ij} de Σ se muestra como la covarianza del i -ésimo y j -ésimo componente de X . Si todas las $1/2p(p-1)$ covarianzas son cero, -- las p componentes de X se distribuyen independientemente.

El exponente

$$(X-\mu)' \Sigma^{-1} (X-\mu)$$

de la densidad Normal Multivariada especifica la ecuación de un elipsoide en el espacio p -dimensional cuando se iguala a alguna constante positiva d .

La familia de elipsoides generados al variar d tiene como centro común μ . La forma y orientación de los elipsoides están determinados por Σ .

2.2 ANALISIS DISCRIMINANTE

El problema de la Discriminación tiene cabida en situaciones donde a un individuo se le miden diversas características para ser identificado como elemento perteneciente a alguna de las posibles clases o poblaciones previamente determinadas.

Situaciones como las anteriores son:

- 1) Determinación de sexo de huesos humanos fósiles basándose en mediciones antropométricas en cráneos.
- 2) Identificación de un individuo como perteneciente a uno de los grupos étnicos en base a mediciones de estatura, altura del individuo sentado, longitud de la cabeza, largo de la nariz, perímetro torácico, etc.
- 3) Diagnósticos de personas que padecen o no una enfermedad particular (Diabéticos y no Diabéticos) basados en diferentes pruebas clínicas para determinar la cantidad de colesterol en la sangre, de urea, presión sanguínea, etc.
- 4) Pertenencia a una de ciertas clases de neurosis determinada para un individuo en base a medidas como estado de ansiedad, sentimiento de culpa, etc.
- 5) Detección de reflujo gastroesofágico de importancia en niños que posiblemente requieran una operación para corregir el problema en base a: la frecuencia de reflujo por doce horas, duración del reflujo con pH menor que 4, % de tiempo con pH menor que 4, y todas estas medidas cuando el individuo está despierto, dormido, sentado y acostado boca arriba.
- 6) Diferenciar la edad entre pacientes de tetralogía de Fallot en base a mediciones de ciertos aspectos de la curva electrocardiográfica.

- 7) Discriminar entre tres clases de áreas exploratorias petroleras: pozos productores, áreas con buenas posibilidades y pozos secos, según mediciones geoquímicas realizadas en muestras de roca (siendo precisamente el problema a desarrollar en este estudio).

En cada situación anterior se pueden encontrar características afines:

- a) Se desea determinar la pertenencia de un individuo en una de las poblaciones definidas a priori, esto es, no se desea identificar a las poblaciones sino discriminar a los individuos.
- b) Las variables o número de mediciones requeridas para caracterizar a un individuo sirven como indicadoras, es decir, se trata de mediciones que se espera varíen entre las poblaciones y permitan identificar la población de la que proviene el individuo.
- c) Por las características mismas del problema existe la posibilidad de cometer el error de no asignar correctamente al individuo, lo que acarreará un costo (o pérdida). Reglas de clasificación óptimas han sido estudiadas para resolver este problema.^{3/}

Se decía anteriormente que en diferentes situaciones se plantea la posición de un elemento a una determinada población y que para esto es necesario contar con individuos pertenecientes a sus respectivas poblaciones, y a los que se les tendrán que medir p variables (en escala ordinal, de intervalo o de relación, de preferencia cuantitativa).

^{3/} ANDERSON, T.W. p. 133. [1].

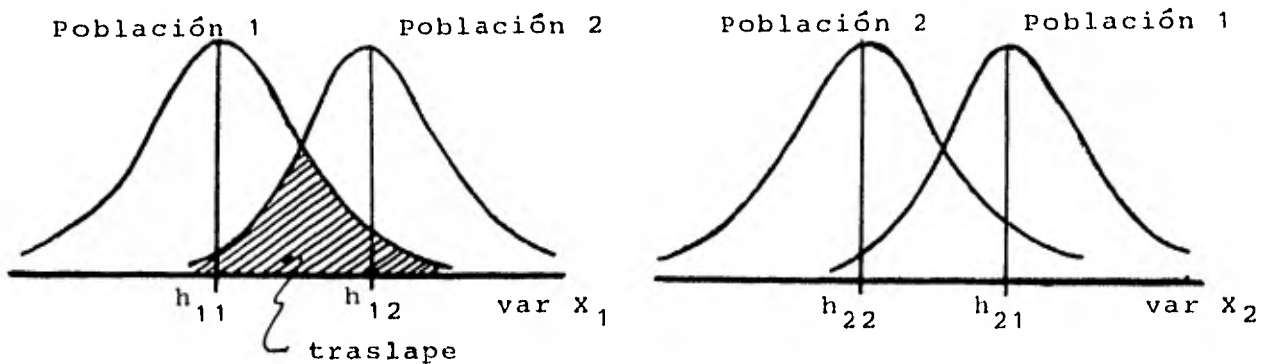
Siendo así el objetivo es conocer si los valores para -- un individuo están más cercanos en conjunto a los de una u otra población.

El Análisis Discriminante Lineal es un procedimiento que en base a combinaciones lineales de un conjunto de variables, se marcan diferencias entre diversos grupos.

En otras palabras, el problema de estudiar la dirección de diferencias de grupos es, equivalentemente, un problema de encontrar largas diferencias en medias de grupos.

Una manera de introducir intuitivamente la metodología y el concepto de Discriminación, es la siguiente:

Se suponen dos poblaciones y dos variables X_1 , X_2 cuyos valores se presentan con traslape; al considerar las variables en forma individual se tendrá para la variable X_1 y la variable X_2 respectivamente:



Al querer efectuar una discriminación o predicción de un elemento con valores X_1 y X_2 a una de las poblaciones por separado se tendrá un gran porcentaje de errores en la discriminación. En cambio si el estudio de las dos variables se hace simultáneamente se contará con un número menor de errores, puesto que al pasar de una a dos dimensiones los --

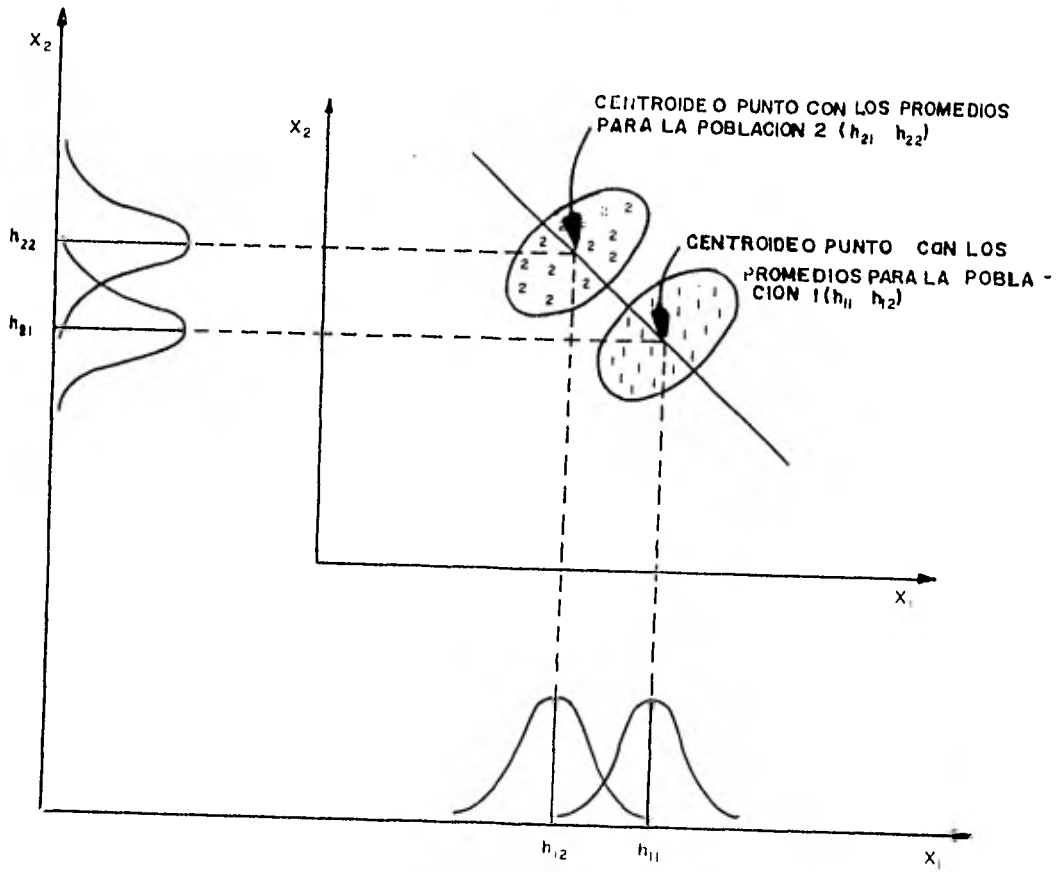


FIG. 2.1 Una separación adecuada entre las poblaciones 1 y 2 no puede hacerse usando una u otra variable X_1, X_2 . Sin embargo es posible encontrar una orientación a lo largo de la cual los dos grupos están separados a la más considerando las variables X_1 y X_2 conjuntamente.

FIG. 2.1

puntos en el plano (que son las combinaciones de los valores de X_1 y X_2) se agrupan más entre sí, dependiendo de la población a que pertenezcan.

La gráfica de la fig. 2.1 muestra el área común de las dos poblaciones a lo largo de las variables X_1 y X_2 .

El objetivo de una función lineal discriminante es transformar un conjunto de medidas de una muestra en un "score" discriminante. Dicho "score" o variable transformada representa la posición de la muestra a lo largo de la línea definida por la función lineal discriminante. Al proyectar todas las observaciones de las poblaciones 1 y 2 sobre la línea D se trata de minimizar la varianza dentro de las poblaciones y maximizar la distancia entre los centros de los grupos.

D es precisamente la nueva variable producida por el análisis discriminante (fig. 2.2), a_1 y a_2 son tal que para D , existe el mínimo traslape entre las dos poblaciones y la distancia máxima entre los promedios de D (δ en este caso).

Este proceso es precisamente lo que se muestra en la gráfica de la fig. 2.2, donde los puntos de cada población están proyectados en una línea D' a través de los centroides o en una línea paralela a ésta. La ecuación de esa línea es la función lineal discriminante ($D = a_1 X_1 + a_2 X_2$).

Una vez calculada D , se podría calcular para un elemento de población desconocida, su respectivo valor D ocasionando ahora una D^* en base a sus valores X_1^* y X_2^* .

Así,

$$D^* = a_1 X_1^* + a_2 X_2^*$$

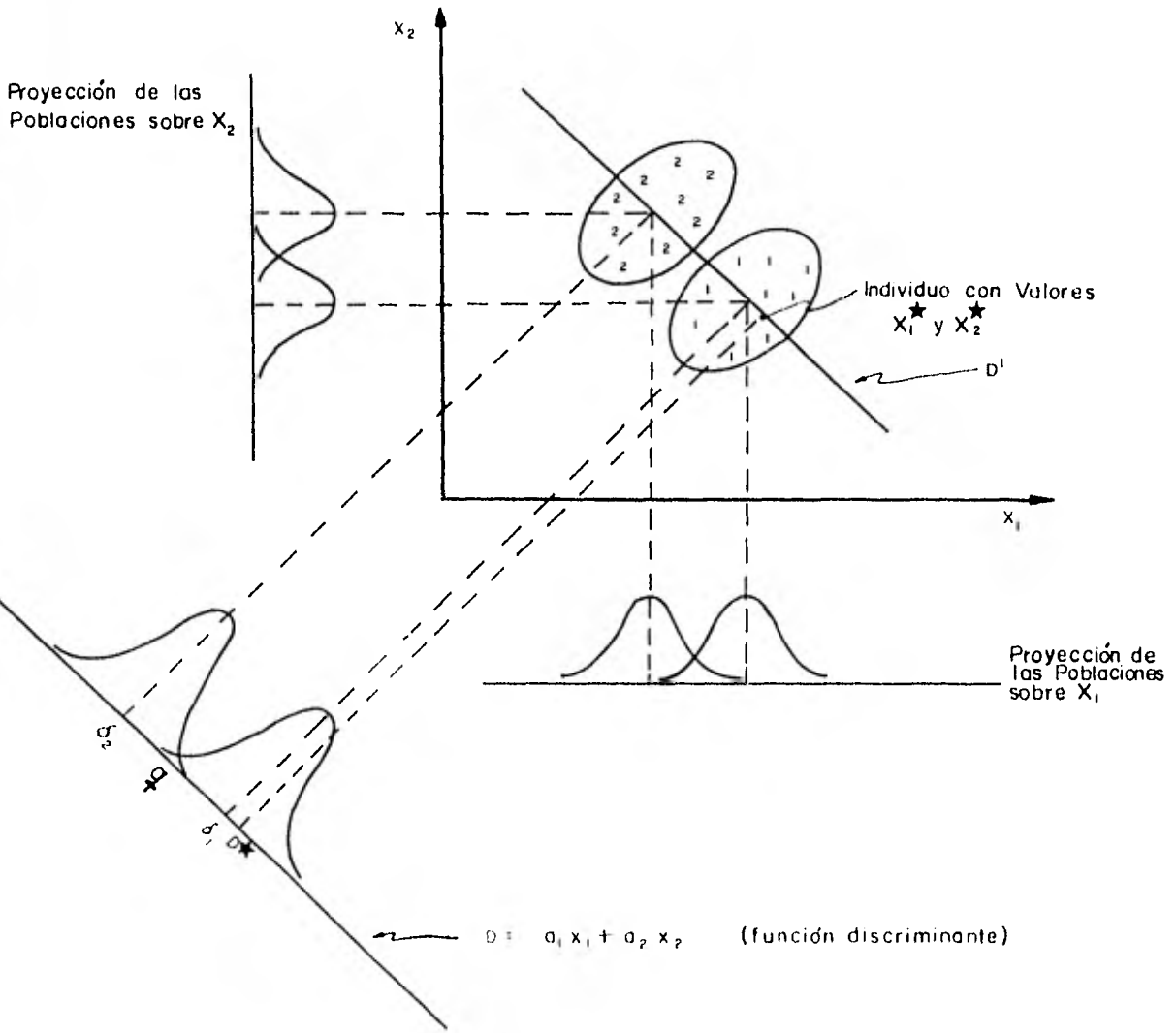


FIG. 2.2

será el valor que determine a qué población pertenece ese nuevo elemento.

El criterio para lo anterior será:

Si $D^* < q$ el elemento pertenece a la población 2
Si $D^* > q$ el elemento pertenece a la población 1

donde q es el punto en el cual la probabilidad de que un individuo pertenezca a cualquiera de las dos poblaciones es la misma.

Para el caso de tres poblaciones, el razonamiento es semejante.

Si se tienen tres o más poblaciones y tres o más variables se pueden construir dos o más funciones discriminantes.

Por ejemplo, para un caso de cuatro poblaciones y tres variables, se tendría el Plano Discriminante (D_1, D_2) de la fig. 2.3.

La primer función discriminante:

$$D_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3$$

es la línea donde se proyectan los puntos produciendo una distancia máxima entre los centroides y un mínimo traslape entre las cuatro poblaciones proyectadas sobre esa línea D_1 .

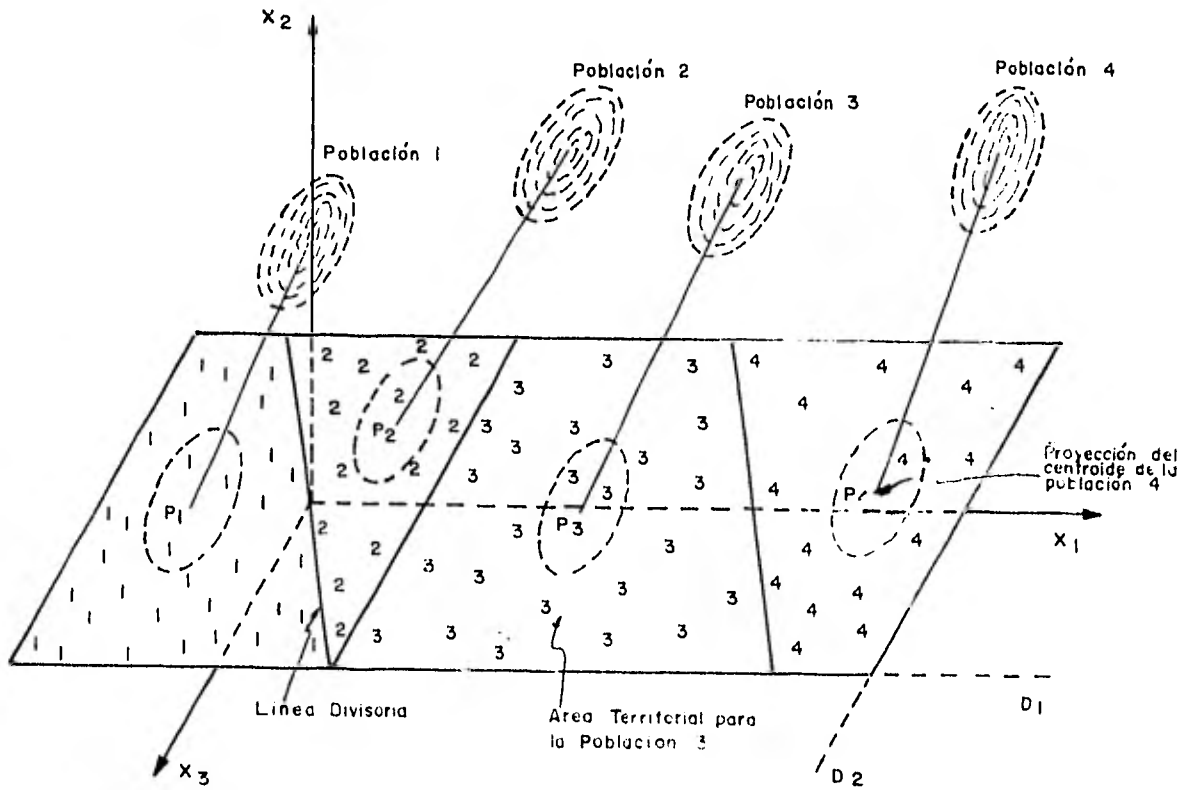
Un nuevo elemento con valores X_1^* , X_2^* y X_3^* tendrá que sustituirse en D_1 y D_2 para ver en qué área territorial (entendiéndose por área territorial a las regiones donde son proyectados los puntos P_1 , P_2 , P_3 y P_4 en este caso, y que se encuentran separadas por las líneas divisorias las cuales constituyen a su vez el equivalente de q en la fig. 2.2) --

del plano discriminante cae el punto formado con las coor
de
nadas:

$$D_1^* = a_{11}X_1^* + a_{12}X_2^* + a_{13}X_3^*$$

$$D_2^* = a_{21}X_1^* + a_{22}X_2^* + a_{23}X_3^*$$

donde a_{11} , a_{12} , a_{13} , a_{21} , a_{22} y a_{23} se les llama coeficien-
te no estandarizados en las funciones discriminantes.



Plano discriminante para 4 poblaciones y 3 variables

Fig. 2.3

De entre las diversas metodologías existentes para calcular la función lineal discriminante de ciertos grupos o poblaciones, se detallará en esta tesis la que se considera de mayor accesibilidad para continuar el entendimiento de este método multivariado.

Dicha técnica^{4/} funciona sólo para dos grupos por lo -- que no es la que se aplica a la información existente en -- este estudio, ya que esta información está constituida por tres grupos. No obstante, en un principio, se trató de -- usar un programa basado en esta técnica a través de un análisis de grupos de dos a dos. Esto se deshechó debido a -- que el programa no es tan general de tal manera que pudiese aceptar casos tan especiales como el de un determinante cercano a cero (-.5336E-24).

Se tratará entonces de complementar la idea de Análisis Discriminante a través del método siguiente.

De acuerdo a párrafos anteriores, se desea construir un componente lineal o índice para sumarizar las observaciones de los grupos en una escala unidimensional que discrimine entre las poblaciones por alguna medida de separación máxima.

• Por lo tanto, el primer paso para determinar una combinación lineal de un conjunto de variables, tal que las medias de los grupos en dicha combinación lineal difieran entre sí; es decidir un criterio para medir tales diferencias de medias de grupos.

^{4/} La notación usada en esta metodología es independiente a la utilizada anteriormente.

En la notación matricial, se debe resolver una ecuación de la forma:

$$[S_p^2]_{m \times m} \cdot [\lambda]_{m \times 1} = [D]_{m \times 1} \quad \dots \quad (2.4)$$

donde

- $[S_p^2]$: es una matriz conjunta de (mxm) de varianzas y covarianzas de m variables.
- $[\lambda]$: es un vector columna (mx1) de λ 's desconocidas que representan los coeficientes de la ecuación discriminante (esto es, las lambdas son el equivalente a las betas : β 's en las ecuaciones de regresión).
- $[D]$: es el vector columna de las m diferencias entre las medias de los dos grupos.

Resolviendo (2.4) por inversión, se tiene:

$$[\lambda] = [S_p^2]^{-1} \cdot [D]$$

las entradas de estas matrices se determinan de la siguiente forma:

$$D_j = \bar{A}_j - \bar{B}_j = \frac{\sum_{i=1}^{n_a} A_{ij}}{n_a} - \frac{\sum_{i=1}^{n_b} B_{ij}}{n_b}$$

estas medias multivariadas de los grupos A y B pueden formar dos vectores:

$$\begin{bmatrix} D_1 \\ D_2 \\ \vdots \\ D_m \end{bmatrix} = \begin{bmatrix} \bar{A}_1 \\ \bar{A}_2 \\ \vdots \\ \bar{A}_m \end{bmatrix} - \begin{bmatrix} \bar{B}_1 \\ \bar{B}_2 \\ \vdots \\ \bar{B}_m \end{bmatrix}$$

A_{ij} es la i -ésima observación de la variable j en el grupo A

\bar{A}_j es la media o promedio de las n_a observaciones del grupo A.

Análogamente para B.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_a 1} & a_{n_a 2} & \dots & a_{n_a m} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n_b 1} & b_{n_b 2} & \dots & b_{n_b m} \end{bmatrix}$$

$$[S_p^2]_{m \times m} = \frac{[SPA] + [SPB]}{n_a + n_b - 2}$$

Matriz conjunta de varianzas y covarianzas

donde

$[SPA]$ y $[SPB]$ son las matrices de suma de cuadrados y-

productos cruzados para el grupo A y B respectivamente definidas por:

$$SPA_{jk} = \sum_{i=1}^{n_a} (A_{ij}A_{ik}) - \frac{\sum_{i=1}^{n_a} A_{ij} \sum_{i=1}^{n_a} A_{ik}}{n_a} \quad \begin{matrix} j=1, k=1, \dots, m \\ j=2, k=1, \dots, m \\ \dots \\ j=m, k=1, \dots, m \end{matrix}$$

A_{ij} denota la i -ésima observación de la variable j en el grupo A

A_{ik} denota la i -ésima observación de la variable k en el grupo A

A_{kk} es la suma de cuadrados de la variable k cuando $j=k$

Análogamente para el grupo B.

Finalmente encontrada $[\lambda]$, cuyos elementos constituyen las entradas en la ecuación (2.5), se podrá conocer más acerca de la ecuación de la función discriminante:

$$R = \lambda_1 \psi_1 + \lambda_2 \psi_2 + \dots + \lambda_m \psi_m \quad (2.5)$$

Los "scores discriminantes" pueden calcularse para las observaciones individuales protegiendo de esa manera sus coordenadas multivariadas sobre la función discriminante:

Por ejemplo, para el caso de dos variables:

Sea R_{iA} el "score discriminante" de la observación i para el grupo A,

Así

$$R_{iA} = \sum_{j=1}^2 \lambda_j a_{ij} \quad i=1, \dots, n_a$$

$$R_{1A} = \lambda_1 a_{11} + \lambda_2 a_{12}$$

$$R_{2A} = \lambda_1 a_{21} + \lambda_2 a_{22}$$

.
.
.

$$R_{n_a A} = \lambda_1 a_{n_a 1} + \lambda_2 a_{n_a 2}$$

son los "scores discriminantes" para el grupo A.

El análogo para el grupo B.

A R_0 se le conoce como el Índice Discriminante debido a que es el punto sobre la Línea Discriminante que se encuentra exactamente a la mitad entre el centro del grupo A y el centro del grupo B; y su valor es conocido cuando se lleva a cabo la sustitución de las medias de los grupos, es decir:

$$R_0 = \lambda_1 \left(\frac{\bar{A}_1 + \bar{B}_1}{2} \right) + \lambda_2 \left(\frac{\bar{A}_2 + \bar{B}_2}{2} \right)$$
$$R_0 = \lambda_1 \psi_1 + \lambda_2 \psi_2 \quad , \quad \psi_j = \frac{\bar{A}_j + \bar{B}_j}{2}$$

Los centros de los dos grupos originales a lo largo de la función discriminante están dados por:

$$R_A = \lambda_1 \bar{A}_1 + \lambda_2 \bar{A}_2 \quad \text{para el grupo A}$$

$$R_B = \lambda_1 \bar{B}_1 + \lambda_2 \bar{B}_2 \quad \text{para el grupo B}$$

Tanto los puntos anteriores (R_0, R_A, R_B) como cada "score discriminante" de cada grupo pueden localizarse sobre la función discriminante. Por ejemplo, la figura 2.4 muestra este hecho y además puede apreciarse que algunos puntos pertenecientes al grupo A se encuentran en el grupo B a un lado de R_0 ; y algunos puntos de B se localizan en el lado del grupo A. A este tipo de puntos se les considera mal asignados por la Función Discriminante.

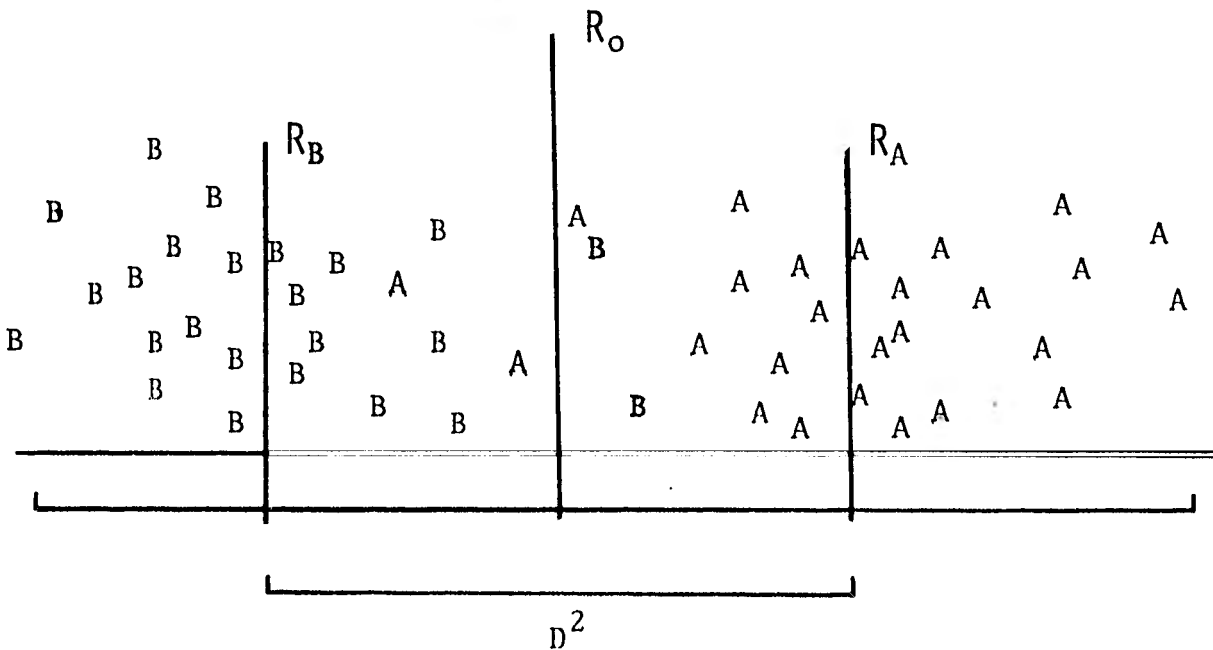


Fig. 2.4

Las entradas o elementos de $[\lambda]$ son usados también para asignar nuevos individuos en uno y otro grupo, dependiendo del lado del Índice Discriminante en el cual caigan.

Supongamos entonces un tercer grupo de observaciones (C) que se desea sean asignadas a una de las poblaciones originales.

Para determinar a qué grupo (A ó B) pertenecen, se obtienen los "scores discriminantes" para tales observaciones, es decir:

$$R_{1C} = \lambda_1 c_{11} + \lambda_2 c_{12} = \theta_1$$

$$R_{2C} = \lambda_1 c_{21} + \lambda_2 c_{22} = \theta_2$$

.
.
.

$$R_{n_c C} = \lambda_1 c_{n_c 1} + \lambda_2 c_{n_c 2} = \theta_{n_c}$$

Así, si

- $\theta_i < R_0$ la observación pertenece al grupo B, $i=1, \dots, n_c$
- $\theta_i > R_0$ la observación pertenece al grupo A,
- $\theta_i = R_0$ se considera que puede pertenecer tanto al grupo A como al B.

Para probar la significancia de la separación entre los dos grupos (A y B), puede desarrollarse una prueba a partir de la estadística T^2 de Hotelling.

Es posible calcular una distancia (caso; 2 variables) entre las dos medias multivariadas por simple sustracción de R_A y R_B , lo cual es equivalente a sustituir el vector de diferencias entre las medias de los grupos:

$$\begin{aligned} D^2 &= \lambda' (\underline{A} - \underline{B}) = \lambda' [D] \\ &= [\lambda_1, \lambda_2] \begin{bmatrix} \bar{A}_1 - \bar{B}_1 \\ \bar{A}_2 - \bar{B}_2 \end{bmatrix} = \\ &= \lambda_1(\bar{A}_1 - \bar{B}_1) + \lambda_2(\bar{A}_2 - \bar{B}_2) \end{aligned}$$

A esta distancia se le conoce como Distancia de MAHALANO BIS o Distancia Generalizada D^2 , y es una medida de la separación entre las dos medias multivariadas expresadas en unidades de varianza conjunta.

La prueba T^2 en estos términos es:

$$T^2 = \frac{n_a n_b}{n_a + n_b} D^2$$

que transformada en una prueba F tendremos:

$$F = \left(\frac{n_a + n_b - m - 1}{(n_a + n_b - 2) m} \right) \left(\frac{n_a n_b}{n_a + n_b} \right) D^2$$

con m y $(n_a + n_b - m - 1)$ grados de libertad.

nota: ver DAVIS, JOHN. p.451 [5].

La suposición inicial o hipótesis de que las dos mues--

tras realmente pertenecen a poblaciones distintas puede ser juzgada por una prueba de Hipótesis Estadística utilizando precisamente la distancia de Mahalanobis.

Se quiere conocer si existe distancia entre las medias - de grupos, por lo tanto se formula la hipótesis en términos contrarios, es decir, se propone:

$$H_0: [D_j] = 0$$

Así podremos rechazar la hipótesis bajo la existencia de la distancia entre las medias de los dos grupos (error tipo I) a un nivel de significancia $\alpha = 0.05$, esto es, tener una confianza del 95% de tomar la decisión acertada; entonces

$$H_0: [D_j] = 0 \quad \text{vs.} \quad H_1: [D_j] > 0$$

Existe además un procedimiento muy simple para medir la contribución directa o importancia de las variables (más no, la interacción entre ellas).

La contribución relativa de la variable j para la distancia entre las dos medias de grupo está dada por:

$$E_j = \frac{\lambda_j D_j}{D^2}$$

Se vuelve a hacer hincapié en que todas estas medidas obtenidas por el método explicado anteriormente no son las utilizadas en la información existente. Algunas cifras coinciden con las producidas por el paquete SPSS (Statistical - - Package for the Social Sciences) que fue el medio para obtener los resultados finales.

CAPITULO 3

APLICACION DEL METODO Y RESULTADOS

La aplicación del Análisis Multivariado a datos reales - requiere gran número de cálculos, por lo que gracias a paquetes de programas de biblioteca para computadoras se facilita el uso de técnicas multivariadas. Es precisamente por uno de estos paquetes de biblioteca: SPSS (Statistical Package for the Social Sciences) que se obtuvieron los resultados finales.

Entre los varios subprogramas del SPSS se encuentra el de DISCRIMINANT ANALYSIS cuyo contexto incluye diversos métodos de Selección de Variables para llegar al mejor conjunto de variables discriminantes. Entre los criterios disponibles se encuentra la Lambda mínima de Wilks, la Distancia de Mahalanobis entre los grupos, la mayor V de Rao, - - etc., fundamentados cada uno en su respectivo método.

Se escogió el de la Distancia Mínima de Mahalanobis entre los grupos, debido a que refiere algo a la explicación que se hizo del método.

Para un mejor entendimiento de la salida del subprograma de Discriminante, se irá haciendo noción de las características de cada concepto conforme se van describiendo los resultados.

Primeramente se obtienen los casos de observaciones pertenecientes a cada grupo (148 para el grupo 1: pozos productores, 180 para el grupo 2: áreas con buenas posibilidades y pozos secos: grupo 3 con 153 observaciones). Ver página-3 del subprograma DISCRIMINANT ANALYSIS (D.A.)

A ésto le siguen algunas estadísticas: medias, desviaciones estandar, pruebas de hipótesis con F y matrices de correlaciones, varianzas y covarianzas. La llamada "WITHIN GROUPS COVARIANCE MATRIX" es precisamente la matriz obtenida por el método referido en el capítulo 2 si se trataran sólo dos grupos ($[S_p^2]$ = matriz conjunta de varianzas y covarianzas). Ver p. 4 del subprograma D.A.

Enseguida se lleva a cabo un método de Selección de variables. Cuando existen muchas variables es posible que no todas aporten una discriminación satisfactoria, por lo que los cálculos no son ejecutados, muchas veces, sobre todas las variables simultáneamente sino sobre variables individuales tomadas consecutivamente en orden decreciente de poder de discriminación. El Método de "Stepwise" selecciona primeramente la mejor variable discriminante según un criterio. Una segunda entra en combinación con la primera; y la tercera y subsecuentes son seleccionadas simi- larmente de acuerdo a su aptitud para una discriminación más fuerte.

En adelante los párrafos en letras mayúsculas corresponden a los resultados de esta investigación.

EN ESTE ESTUDIO TODAS LAS VARIABLES RESULTAN ELEGIBLES- PARA EL ANALISIS. Ver P. 5-10 del subprograma D.A.

Entre las diversas maneras de encontrar funciones de clasificación (llamadas así por el subprograma D.A.) ^{5/}, se encuentra aquella donde son obtenidas de la matriz conjunta de varianzas y covarianzas de un grupo (se deriva una función para cada grupo).

^{5/} Ya que en esta investigación se les llamaría funciones de asignación.

$$C_1 = -40.26332 + 0.00071(V1) + 3.63424(V2) + 0.26800(V3) + \\ + 0.13793(V4) + 5.65884(V5) + 9.97424(V6)$$

$$C_2 = -34.01229 - 0.00281(V1) + 3.23731(V2) + 0.25532(V3) + \\ + 1.14108(V4) + 5.20298(V5) + 8.92407(V6)$$

$$C_3 = -48.00053 + 0.00306(V1) + 4.60952(V2) + 0.30136(V3) + \\ + 0.94749(V4) + 5.49742(V5) + 9.26531(V6)$$

$$C_i = c_{i1}V_1 + c_{i2}V_2 + \dots + c_{ip}V_p + c_{i0}$$

donde:

C_i es el "score de clasificación" para el grupo i -- ($i=1,2,3$ en este caso), las c_{ij} ($j=1,\dots,6$) son los coeficientes de clasificación con c_{i0} constante y

- V1 : PROFUNDIDAD EN METROS
- V2 : CARBON ORGANICO
- V3 : CONTENIDO DE CARBONATOS
- V4 : INDICE DE MADUREZ
- V5 : TIPO DE MATERIA ORGANICA
- V6 : COLOR DE LA MATERIA ORGANICA

Los coeficientes fueron los estimados por el subprograma (ver P.11 del subprograma D.A.)

A través de la asignación (realizada por las funciones de clasificación) puede ser medido el grado de una buena discriminación al observar la proporción de las asignaciones correctas (p.13-24 subp. D.A.) Por lo tanto el propósito de asignar todos los casos es para ver cómo están discriminando las variables. Si se encuentra una gran proporción de observaciones mal asignadas, entonces puede ser -- que las variables seleccionadas sean pobres en el sentido discriminante. Si sólo existe un pequeño porcentaje de observaciones mal asignadas puede ser que los grupos sean similares en algunas de sus características.

DE LAS MUESTRAS DE ROCA PERTENECIENTES A LOS POZOS PRODUCTORES SOLO HUBO 40 OBSERVACIONES MAL ASIGNADAS, MIENTRAS QUE EN EL GRUPO 2 HUBO CERO Y EN EL GRUPO 3, 48; LO CUAL SE CONSIDERA UNA BUENA DISCRIMINACION DEBIDO A LA BAJA PROPORCION DE OBSERVACIONES MAL ASIGNADAS. (Ver p. 28 del subprograma D.A.).

El número máximo de funciones discriminantes es menor que el número de grupos o igual al número de variables discriminantes, lo que sea más pequeño.

El subprograma del SPSS provee información necesaria para juzgar la importancia de las funciones discriminantes.

Los eigenvalores y sus correlaciones canónicas denotan el porcentaje relativo de cada función para separar los grupos (el eigenvalor por tratarse de una medida de importancia relativa de la función, puesto que la suma de los eigenvalores es una medida del total de la varianza existente en las variables discriminantes y un eigenvalor individual revelaría un porcentaje de éste último; y la correlación canónica por tratarse de la medida que nos indica qué tan relacionadas están la función y la "variable grupo" o variables indicadoras, lo cual es justamente otra medida de la función para discriminar entre grupos).

La Lambda de Wilks es una medida inversa del poder discriminante de las variables originales. La lambda mayor es la de menor información. La lambda puede transformarse en una estadística X^2 para una prueba fácil de significancia estadística.

POR TRATARSE DE TRES GRUPOS Y SEIS VARIABLES, EL ANALISIS ESTA BASADO EN DOS FUNCIONES DISCRIMINANTES. EL PORCENTAJE RELATIVO EN LA PRIMERA FUNCION DISCRIMINANTE INDICA ENTONCES QUE SE HA EXPLICADO UN 96.93% DE LA DISCRIMINACION, EL CUAL ES MUY BUEN PORCENTAJE. LA CORRE-

LACION ENTRE LAS VARIABLES: V1, V2, V3, V4, V5 Y V6 Y LA PRIMER FUNCION DISCRIMINANTE ES ALTA: 0.867. ESTOS VALORES PARA LA SEGUNDA FUNCION SON BAJOS PUES SOLO SE EXPLICA UN 3.07% DE LA DISCRIMINACION Y SU CORRELACION CANONICA ES DE 0.296, EN LAS FUNCIONES DERIVADAS LA 1a. LAMBDA DE WILKS ES DE 0.2267 LO CUAL SIGNIFICA QUE EXISTE UN GRAN PODER DISCRIMINANTE Y UN ALTO GRADO DE SEPARACION DE CENTROIDES. LA LAMBDA DE 0.9125 SE DESCARTA COMPLETAMENTE DADO SU ALTO VALOR (Ver p.11 del subprograma (D.A.).

Los coeficientes estandarizados de la función discriminante son los de mayor importancia en el análisis. Cada coeficiente representa la contribución relativa, ya sea negativa o positiva según el signo de su variable asociada. Aquellas variables de mayor coeficiente son precisamente las que aportan mayor poder discriminante (esto es, las que difieren más de un grupo a otro),

Esta relación entre las variables y los coeficientes puede ser analizada en el mapa territorial (fig. 3.1).

Los coeficientes no estandarizados no se interpretan, sin embargo son utilizados para el cálculo de un nuevo elemento, esto es, cuando se trata de casos que requieran categorizarse a una de las poblaciones originales.

Así, para los coeficientes estandarizados, sus funciones son:

$$D_1 = -0.97168(V1) - 0.11384(V2) - 0.13811(V3) + 0.02986(V4) - 0.03218(V5) - 0.04602(V6) \dots (3.1)$$

$$D_2 = 0.18989(V1) - 0.41179(V2) - 0.52409(V3) - 0.95278(V4) + 0.31851(V5) + 0.98013(V6)$$

EN D_1 LOS COEFICIENTES DE MAYOR APORTACION SON LOS ASOCIADOS CON LAS VARIABLES V1, V2 y V3, MIENTRAS QUE EN D_2 SON LOS ASOCIADOS CON V6, V4 y V3. (p.12-subp.D.A.)

Las funciones discriminantes no estandarizadas se expresan enseguida:

$$D_1 = 1.68143 - 0.00071(V1) - 0.15817(V2) - 0.00530(V3) + 0.03954(V4) \\ - 0.04068(V5) - 0.05663(V6)$$

$$D_2 = -0.13225 + 0.00014(V1) - 0.57217(V2) - 0.02010(V3) - 1.26175(V4) + \\ + 0.40264(V5) + 1.20628(V6)$$

Para la asignación de un nuevo elemento, basta sustituir en D_1 y D_2 anteriores sus valores respectivos de V1 a V6 -- produciendo un punto: (D_1^*, D_2^*) , el cual al localizar su posición en el mapa territorial (ver fig. 3.1 y pág. 27 del subprograma DA.) se conocerá el grupo al que pertenece.

La examinación de las gráficas permite estudiar la separación de los grupos y sus localizaciones relativas.

DE ESTA MANERA, SE OBSERVA QUE LOS TRES GRUPOS FORMARON-UNA Y, PERMANECIENDO EL GRUPO 2 DEL LADO DERECHO, EL 3 - EN EL LADO IZQUIERDO, MIENTRAS QUE EL GRUPO 1 OCUPÓ EL - CENTRO. (p.27-subp.D.A.).

La última tabla en la salida del subprograma de DISCRIMINANT ANALYSIS es un sumario del total de casos que cayeron en los diferentes grupos. Por ejemplo, de las 40 observaciones mal asignadas del grupo 1, 23 cayeron en el grupo 2- y 17 en el tercero. El 100% de asignación correcta lo constituyó el grupo 2 (puesto que no tiene observaciones en los otros grupos); y el tercer grupo cuenta con 47 observaciones

mal asignadas en el grupo 1 y con una en el grupo 2. De és to puede concluirse que el grupo de mejores asignaciones - fue el 2 siguiéndole el grupo 1 y finalmente el grupo 3; - ocasionando un muy buen grado de correcta asignación: - - 81.70%. (p. 28-subp. D.A).

En el mapa territorial puede hacerse una interpretación de las funciones discriminantes, que vendrá a complementar las anteriores.

Esta interpretación se hace en base a las variables cuyos coeficientes sean los de mayor valor.

Nota: Para una mejor comprensión de la fig. 3.1, ver MEN-
DEZ & RODRIGUEZ [12].

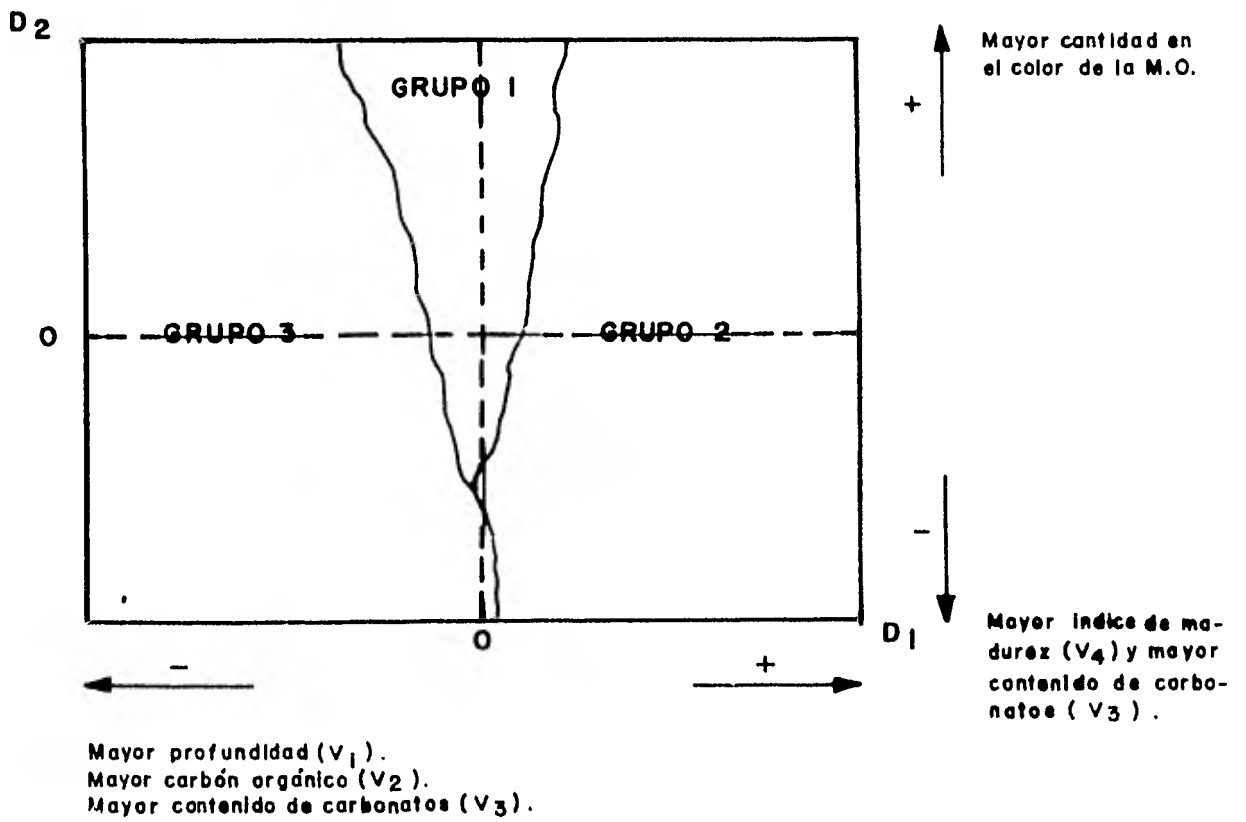


Fig. 3.1

De lo anterior se entiende lo siguiente:

EN EL GRUPO 1 (POZOS PRODUCTORES) TANTO LA PROFUNDIDAD (V1), EL CARBON ORGANICO (V2), EL CONTENIDO DE CARBONATOS (V3), COMO EL INDICE DE MADUREZ TERMAL Y COLOR DE LA MATERIA ORGANICA (M.O.) (V4, V6 RESPECTIV.), ES MENOR.

Las medias de las variables V1, V2 y V3 del grupo 1, -- coinciden, aunque no completamente, con el adjetivo "menor" respecto a las medias de los grupos 2 y 3 en estas mismas variables (ver P. 3 del superprograma D.A). Ahora bien, - el grado de "menor" en la profundidad (V1) indica que intervinieron (en promedio) pozos de menor profundidad que los que constituyen los pozos secos.

Las calizas y arcillas son las rocas con los porcentajes más altos de materia orgánica. Sus límites mínimos para considerarlas generadoras de hidrocarburos son 0.2 y -- 0.5% respectivamente; así el grado de menor en V2 (Contenido de Carbón Orgánico) y V3 (Contenido de Carbonatos) se explica en estos porcentajes pequeños, esto es, se indica que las rocas en estudio alcanzan los límites mínimos para generar hidrocarburos.

La aportación pequeña de la variable V6 y V4 (Color de M.O. e Índice de Madurez Termal respectivamente) corresponde : en V6 al intervalo (2,3 - fig. 1.9), en donde aproximadamente se encuentra la etapa madura de la roca (color amarillo /naranja y café claro), y al intervalo II, III, - IV (moderadamente inmaduro, moderadamente maduro, maduro) - en el Índice de Madurez Termal (fig. 1.6).

Tales cantidades pequeñas de V6 y V4 concuerdan con la presencia de los pozos productores de gas tratados en este estudio, puesto que en las primeras etapas de generación de la materia orgánica, las rocas inmaduras producen sólo gas metano, como se mencionó anteriormente.

EN EL GRUPO 2 (AREAS CON BUENAS POSIBILIDADES) RESULTAN MAYORES: EL CONTENIDO DE CARBONATOS (V3), EL INDICE DE MADUREZ TERMAL (V4) Y EL COLOR DE LA M.O. (V6).

En las áreas con buenas posibilidades se encuentran características propias del grado de las rocas metamorfizadas (ya que V6 y V4 mayor implica un color cercano al negro y una temperatura alta). Lo que significa posible generación de gases ligeros. Debido a que el contenido de carbonatos (V3), es alto, una área con buenas posibilidades se aseguraría como tal; puesto que se situaría en los mayores porcentajes de los valores mínimos (0.2, 0.5) del carbón orgánico.

LAS VARIABLES V1, V2, V3, V4 y V6 REPRESENTARON VALORES MAYORES EN EL GRUPO 3, QUE CORRESPONDE A POZOS SECOS.

De los pozos secos que constituyen la información para el tercer grupo (pozos secos) se encuentra el más profundo (4985 metros) del número total de pozos considerado en el estudio; por este motivo el análisis de V1 en el mapa territorial arroja un valor grande para esta variable. Las variables restantes (V2, V3, V4 y V6) indican también una aportación mayor. Para expresar una justificación a esto último es necesario una intervención simultánea de las variables; esto es, el hecho de que el grupo 3 revele valores mayores para las variables V2 a V6 indica que habiendo gran cantidad de Carbón Orgánico y Carbonatos, la temperatura fue demasiado alta, de tal manera que ocasionó el acercamiento al color negro lo que indica la "posibilidad" (ya que son pozos secos) de producir gas metano (gas ligero). El grado alto del Índice de Madurez Termal implica un estado de alteración severa o metamorfosis en la roca,-

produciendo generalmente gases ligeros.

Por lo tanto, se encontraron las características de las rocas metamorfizadas en el grupo de los pozos secos.

En términos de las variables de mayor distinción o poder discriminante entre los grupos se tiene que:

LOS POZOS PRODUCTORES (GRUPO 1) SE DISTINGUEN DE LAS --
AREAS CON BUENAS POSIBILIDADES Y POZOS SECOS (GRUPO 2 Y --
GRUPO 3 JUNTOS) EN BASE AL CONTENIDO DE CARBONATOS, INDI --
CE DE MADUREZ TERMAL Y COLOR DE LA MATERIA ORGÁNICA; --
MIENTRAS QUE SOLO LA PROFUNDIDAD Y EL CARBÓN ORGÁNICO --
CONSTITUYEN LAS VARIABLES DE MAYOR PODER DISCRIMINANTE --
ENTRE LOS POZOS PRODUCTORES Y LOS POZOS SECOS.

Respecto a la asignación de nuevas observaciones en base a las funciones discriminantes encontradas, se tiene lo siguiente:

Para los Pozos Secos y Areas con Buenas Posibilidades -- existe gran número de información analizada geoquímicamente, suceso que no ocurre con los Pozos Productores (la única información de este último grupo pertenece a los pozos productores presentados en este estudio; y no de hidrocarburos -- líquidos, sino de gas). La causa de ésto se debe al gran -- interés del explorador petrolero sobre las áreas improductivas como medida complementaria de la Exploración, orientán-dole hacia lugares probables de hidrocarburos.

Se obtuvieron las observaciones o muestras de roca si -- guientes para ser categorizadas:

Pozos Secos

| V1 | V2 | V3 | | V4 | V5 | V6 |
|-------------|-----------------|-------------------------|-------|--------------------|--------------|---------------|
| Profundidad | Carbón Orgánico | Contenido de Carbonatos | Grupo | Indice de Madurez. | Tipo de M.O. | Color de M.O. |
| 1) 130 | 0.23 | 91.0 | 3 | 2.35 | 1 | 3 |
| 2) 250 | 0.22 | 74.8 | 3 | 2.35 | 3 | 3 |
| 3) 430 | 0.20 | 71.4 | 3 | 3.13 | 1 | 4 |
| 4) 1080 | 0.31 | 88.9 | 3 | 3.19 | 1 | 4 |
| 5) 1400 | 3.91 | 47.3 | 3 | 3.50 | 1 | 5 |
| 6) 1990 | 2.08 | 52.1 | 3 | 3.90 | 3 | 5 |

Areas con Buenas Posibilidades

| V1 | V2 | V3 | | V4 | V5 | V6 |
|--------|------|------|---|------|----|----|
| 7) 10 | 0.47 | 17.7 | 2 | 4.30 | 3 | 5 |
| 8) 10 | 0.82 | 96.6 | 2 | 3.60 | 2 | 4 |
| 9) 10 | 1.18 | 82.1 | 2 | 3.85 | 2 | 5 |
| 10) 10 | 0.04 | 88.8 | 2 | 4.25 | 1 | 5 |
| 11) 10 | 0.07 | 97.6 | 2 | 3.99 | 1 | 5 |

Puesto que las Funciones Discriminantes no estandarizadas son:

$$D_1 = 1.68143 - 0.00071(V1) - 0.15817(V2) - 0.00530(V3) + 0.03954(V4) - 0.04068(V5) - 0.05663(V6)$$

$$D_2 = -0.13225 + 0.00014(V1) - 0.57217(V2) - 0.02010(V3) - 1.26175(V4) + 0.40264(V5) + 1.20628(V6)$$

Para la nueva observación 1) tenemos:

$$D_1^* = 1.68143 - 0.00071(130) - 0.15817(0.23) - 0.00530(91) + 0.03954(2.35) - 0.04068(1) - 0.05663(3) = 0.95280$$

$$D_2^* = -0.13225 + 0.00014(130) - 0.57217(0.23) - 0.02010(91) - 1.26175(2.35) + 0.40264(1) + 1.20628(3) = -1.01838$$

de donde

$$(D_1^*, D_2^*)_1 = (0.95280, -1.01838)$$

es el punto 1) a ser categorizado en el mapa territorial (p.27 del subprograma DISCRIMINANT ANALYSIS).

Realizando los cálculos para los puntos restantes, tendremos:

| Grupo Original | | Se categoriza en el grupo: |
|----------------|---|----------------------------|
| 3 | $(D_1^*, D_2^*)_1 = (0.95280, -1.01838)$ | 2 |
| 3 | $(D_1^*, D_2^*)_2 = (0.873682, 0.13504)$ | 2 |
| 3 | $(D_1^*, D_2^*)_3 = (0.822636, -0.34314)$ | 2 |
| 3 | $(D_1^*, D_2^*)_4 = (0.253360, -0.74254)$ | 1 |
| 3 | $(D_1^*, D_2^*)_5 = (-0.367145, -1.10625)$ | 3 ó 1 |
| 3 | $(D_1^*, D_2^*)_6 = (-0.587578, 0.22752)$ | 1 |
| 2 | $(D_1^*, D_2^*)_7 = (1.271012, 1.058255)$ | 2 |
| 2 | $(D_1^*, D_2^*)_8 = (0.867115, -1.453589)$ | 2 |
| 2 | $(D_1^*, D_2^*)_9 = (0.840278, -0.477278)$ | 2 |
| 2 | $(D_1^*, D_2^*)_{10} = (1.041578, -0.867014)$ | 2 |
| 2 | $(D_1^*, D_2^*)_{11} = (0.979913, -0.733004)$ | 2 |

De esta manera las once muestras de roca pertenecientes a Pozos Secos y Areas con Buenas Posibilidades fueron categorizadas en un mayor porcentaje a éste último grupo. De las seis observaciones pertenecientes al grupo 3, un 50% cayó en el grupo 2, un 33.3% en el grupo 1, esto es, 2 observaciones y una observación se puede asignar tanto al grupo 3 como al grupo 1; mientras que las cinco muestras de roca pertenecientes al grupo 2 se asignaron en un 100% al mismo grupo.

CONCLUSIONES

Se ha detallado la interpretación de los resultados en base al mapa territorial y en relación a las variables; encontrando como variables de mayor poder discriminante a V3, V4 y V6 entre los grupos 1 y 2 y 3 juntos; y a V1 y V2 para distinguir al grupo 1 del grupo 3.

Los Pozos Productores, fue el grupo que conservó características propias. En los prospectos y secos se presentaron características de las rocas metamorfizadas y en consecuencia con probabilidad de producir gases ligeros, lo que confirma aún más el calificativo de áreas con buenas posibilidades para el grupo 2; más no así para los pozos secos.

En resumen, el porcentaje relativo es de 96.93%, la correlación canónica de 86.7%; porcentajes ocasionados por un muy buen grado de asignación (81.70%), lo que a su vez confirma la intervención de las variables geoquímicas usadas, como principales para determinar la potencialidad generadora de hidrocarburos de una muestra de roca.

En la categorización de nuevas observaciones o muestras de roca, el grupo 2 conservó el 100% de asignación correcta, mientras que el grupo 3 resultó con otros porcentajes de asignación.

Esta y conclusiones anteriores pueden irse depurando a través de investigaciones posteriores en los mismos campos (Geología y Estadística).

Por otra parte, pueden conjuntarse como posibles fuentes de error las razones siguientes: el trabajar con una información pionera en el campo de la Geoquímica Orgánica dentro de la Exploración Petrolera (lo que la hacer ser no signifi

ficativa en un porcentaje alto); y el crear los valores numéricos para las variables Índice de Madurez Termal y Tipo-y Color de M.O., en base a áreas sombreadas (fig. 1.7), lo-cual puede ocasionar errores de exactitud al hacer dicha --transformación.

El objetivo de este trabajo se ha confirmado, esto es, -la discriminación obtenida da lugar a reafirmar los aspec--tos geoquímicos, al ser utilizados los elementos estadísti-cos.



DISTRIBUTED FOR THE BURROUGHS B6700 BY THE
 SOCIAL SCIENCE DATA SERVICE
 UNIVERSITY OF CALIFORNIA, DAVIS
 DEFAULT WORKSPACE FOR THIS RUN.. 20000 WORDS
 RUN NAME ANALYSIS DISCRIMINANT
 FILE NAME ROCA
 VARIABLE LIST V1,V2,V3,GRUPO,V4,V5,V6
 INPUT FORMAT FIXED(F4.0,F3.2,F3.1,F1.0,F3.2,2F1.0)

ACCORDING TO YOUR INPUT FORMAT, VARIABLES ARE TO BE READ AS FOLLOWS:

| VARIABLE | FORMAT | RECORD | COLUMNS |
|----------|--------|--------|---------|
| V1 | F 4. 0 | 1 | 1-4 |
| V2 | F 3. 2 | 1 | 5-7 |
| V3 | F 3. 1 | 1 | 8-10 |
| GRUPO | F 1. 0 | 1 | 11-11 |
| V4 | F 3. 2 | 1 | 12-14 |
| V5 | F 1. 0 | 1 | 15-15 |
| V6 | F 1. 0 | 1 | 16-16 |

THE INPUT FORMAT PROVIDES FOR 7 VARIABLES TO BE READ FROM 1 RECORDS ('CARDS') PER CASE.
 A MAXIMUM OF 16 'COLUMNS' ARE USED ON A RECORD

N OF CASES 431
 INPUT MEDIUM CARD
 READ INPUT DATA
 END OF DATA INPUT, READ COUNT = 431

ANALISIS DISCRIMINANTE

12/02/81

PAGE 2

DISCRIMINANT GRUPOS=GRUPO(1,3)/VARIABLES=V1 TO V3,V4 TO V6/
ANALISIS=V1 TO V3,V4 TO V5/
METHOD=MAHAL/
OPTIONS 5,6,7,8,9,10,11,12,15,15,17,18,19
STATISTICS ALL



***** WARNING ***** OPTIONS 15 THRU 19 AND STATISTICS 7, 8 ARE NOT YET IMPLEMENTED AND WILL BE IGNORED.

***** THIS DISCRIMINANT ANALYSIS REQUIRES 1278 WORDS OF WORKSPACE *****

END OF DATA INPUT, READ COUNT = 461

ANALISIS DISCRIMINANTE

12/02/81

PAGE 3



FILE ROCA (CREATION DATE = 12/02/81)

GROUP COUNTS

| | GROUP 1 | GROUP 2 | GROUP 3 | TOTAL |
|-------|----------|----------|----------|----------|
| COUNT | 148.0000 | 180.0000 | 155.0000 | 483.0000 |

MEANS

| | GROUP 1 | GROUP 2 | GROUP 3 | TOTAL |
|----|-----------|---------|-----------|-----------|
| V1 | 1712.6687 | 10.0000 | 2305.8366 | 1423.2133 |
| V2 | 0.3751 | 0.4114 | 0.5550 | 0.4459 |
| V3 | 69.8746 | 74.8723 | 85.7317 | 76.7248 |
| V4 | 3.5473 | 2.8845 | 4.0234 | 3.5524 |
| V5 | 2.3851 | 2.0944 | 2.1830 | 2.2121 |
| V6 | 4.5333 | 3.7444 | 4.7255 | 4.2974 |

STANDARD DEVIATIONS

| | GROUP 1 | GROUP 2 | GROUP 3 | TOTAL |
|----|----------|---------|----------|-----------|
| V1 | 761.5337 | 0.0000 | 991.4396 | 1368.5813 |
| V2 | 0.5529 | 0.6333 | 0.9201 | 0.7197 |
| V3 | 26.4777 | 23.7147 | 17.9915 | 26.0771 |
| V4 | 0.4640 | 0.7330 | 0.4813 | 0.7551 |
| V5 | 0.6652 | 1.0374 | 0.4790 | 0.7911 |
| V6 | 0.3005 | 0.9402 | 0.4477 | 0.8125 |



WILKS' LAMBDA (U-STATISTIC) AND UNIVARIATE F-RATIO WITH 2 AND 478 DEGREES OF FREEDOM

| VARIABLE | WILKS' LAMBDA | F |
|----------|---------------|----------|
| V1 | 0.2610 | 675.6200 |
| V2 | 0.9338 | 2.5990 |
| V3 | 0.9541 | 11.5088 |
| V4 | 0.5976 | 150.7644 |
| V5 | 0.9763 | 3.7463 |
| V6 | 0.7117 | 35.7924 |

WITHIN GROUPS COVARIANCE MATRIX

| | V1 | V2 | V3 | V4 | V5 | V6 |
|----|-------------|---------|----------|--------|--------|--------|
| V1 | 490942.7403 | | | | | |
| V2 | -40.2533 | 3.5143 | | | | |
| V3 | -1381.3323 | -7.2077 | 551.4935 | | | |
| V4 | 182.7033 | 0.0542 | -1.5799 | 0.3422 | | |
| V5 | 59.3301 | 0.0035 | -7.4263 | J.1254 | 0.6135 | |
| V6 | 163.1512 | 0.0412 | -3.5366 | 0.2894 | 0.0329 | 0.4714 |

WITHIN GROUPS CORRELATION MATRIX

| | V1 | V2 | V3 | V4 | V5 | V6 |
|----|---------|---------|---------|--------|--------|--------|
| V1 | 1.0000 | | | | | |
| V2 | -0.0801 | 1.0000 | | | | |
| V3 | -0.0772 | -0.3333 | 1.0000 | | | |
| V4 | 0.4458 | 0.1272 | -0.1058 | 1.0000 | | |
| V5 | 0.0727 | J.0151 | -0.3714 | 0.2736 | 1.0000 | |
| V6 | 0.3370 | 0.0337 | -0.2017 | J.7203 | J.1540 | 1.0000 |

TOTAL COVARIANCE MATRIX

| | V1 | V2 | V3 | V4 | V5 | V6 |
|----|--------------|---------|----------|--------|--------|--------|
| V1 | 1873014.3663 | | | | | |
| V2 | 19.7193 | 3.5180 | | | | |
| V3 | 2106.3307 | -6.7605 | 680.0169 | | | |
| V4 | 746.2427 | 0.0793 | -0.0837 | 0.5702 | | |
| V5 | 104.6774 | J.0052 | -7.7440 | 0.1476 | 0.6258 | |
| V6 | 663.2741 | 0.0577 | -2.6716 | 0.4913 | 0.1155 | 0.5607 |

ANALYSIS DISCRIMINANTE

12/02/81

PAGE 5

FILE RJCA (CREATION DATE = 12/02/81)



----- DISCRIMINANT ANALYSIS -----

ANALYSIS NUMBER 1
F FOR INCLUSION 1.00000
F FOR DELETION 1.00000
TOLERANCE LEVEL 0.00100
MAXIMUM STEPS 12

SOLUTION METHOD - STEPWISE. SELECT VARIABLE WHICH WILL
MAXIMIZE MINIMUM MAHALANOBIS DISTANCE BETWEEN GROUP PAIRS.

NOTE: THE ENTRY CRITERIA CURRENTLY BEING PRINTED ARE (USABLE) INTERMEDIATE RESULTS.
PRIOR PROBABILITIES -

GROUP 1 GROUP 2 GROUP 3
0.33333 0.33333 0.33333

VARIABLE ENTERED ON STEP NUMBER 1.. VI

WILKS' LAMBDA 0.26103 APPROXIMATE F 676.61997 DEGREES OF FREEDOM 2 SIGNIFICANCE 0.000
RAO'S J 1353.23993 CHANGE IN V 1353.23993 474.00 0.000

F MATRIX - DEGREES OF FREEDOM: 1, 473

GROUP 2 GROUP 1 GROUP 2
GROUP 3 479.63437 1316.73327
GROUP 3 183.11451

ANALYSIS DISCRIMINANTE

12/02/71

PAGE 6



----- VARIABLES IN THE ANALYSIS -----

| VARIABLE | ENTRY CRITERION | F TO REMOVE |
|----------|-----------------|-------------|
| V1 | 2.43407 | 676.61977 |

----- VARIABLES NOT IN THE ANALYSIS -----

| VARIABLE | TOLERANCE | F TO ENTER | ENTRY CRITERION |
|----------|-----------|------------|-----------------|
| V2 | 0.79354 | 4.15180 | 0.00257 |
| V3 | 0.99434 | 12.10393 | 0.03304 |
| V4 | 0.80177 | 0.58606 | 0.67543 |
| V5 | 0.99477 | 4.73506 | 0.01278 |
| V6 | 0.84510 | 4.58396 | 0.07739 |

 VARIABLE ENTERED ON STEP NUMBER 2.. V3

| WILKS' LAMBDA | 0.24341 | APPROXIMATE F | 240.02115 | DEGREES OF FREEDOM | SIGNIFICANCE |
|---------------|------------|---------------|-----------|--------------------|----------------|
| RAO'S V | 1599.03962 | CHANGE IN V | 43.79969 | 4 2 | 0.000 0.000 |

F MATRIX - DEGREES OF FREEDOM: 2, 477

| GROUP | 2 | GROUP 1 | 239.33296 | GROUP 2 | 574.93297 |
|-------|---|-----------|-----------|---------|-----------|
| GROUP | 3 | 108.04340 | | | |

----- VARIABLES IN THE ANALYSIS -----

| VARIABLE | ENTRY CRITERION | F TO REMOVE |
|----------|-----------------|-------------|
| V1 | 2.43409 | 677.43574 |
| V3 | 2.87440 | 12.11893 |

----- VARIABLES NOT IN THE ANALYSIS -----

| VARIABLE | TOLERANCE | F TO ENTER | ENTRY CRITERION |
|----------|-----------|------------|-----------------|
| V2 | 0.83261 | 13.06536 | 2.57625 |
| V4 | 0.79615 | 1.02058 | 2.45401 |
| V5 | 0.86002 | 1.41824 | 2.37277 |
| V6 | 0.85410 | 3.14253 | 2.50457 |

ANALYSIS DISCRIMINANTE

12/02/81

PAGE 7



VARIABLE ENTERED ON STEP NUMBER 3.. V2

| | | | | | |
|---------------|------------|---------------|-----------|--------------------|--------------|
| WILKS' LAMBDA | U.23549 | APPROXIMATE F | 153.29792 | DEGREES OF FREEDOM | SIGNIFICANCE |
| RAO'S V | 1455.03126 | CHANGE IN V | 67.01164 | 6 | 0.000 |
| | | | | 2 | -0.000 |

F MATRIX - DEGREES OF FREEDOM: 3, 476

| | | | | |
|-------|-------|-----------|-----------|---|
| | GROUP | 1 | GROUP | 2 |
| GROUP | 2 | 159.84443 | | |
| GROUP | 3 | 84.23301 | 467.27045 | |

----- VARIABLES IN THE ANALYSIS -----

| VARIABLE | ENTRY CRITERION | F TO REMOVE |
|----------|-----------------|-------------|
| V1 | 2.43407 | 637.71517 |
| V2 | 3.37537 | 13.05536 |
| V3 | 2.87340 | 21.27337 |

----- VARIABLES NOT IN THE ANALYSIS -----

| VARIABLE | TOLERANCE | F TO ENTER | ENTRY CRITERION |
|----------|-----------|------------|-----------------|
| V4 | 0.77326 | 0.16216 | 2.91645 |
| V5 | 0.84073 | 1.13536 | 5.39711 |
| V6 | 0.65216 | 3.32053 | 2.34993 |

ANALYSIS DISCRIMINANTE

12/01/51

PAGE 3



VARIABLE ENTERED ON STEP NUMBER 4... V6

| | | | | | |
|---------------|------------|---------------|-----------|--------------------|--------------|
| WILKS' LAMBDA | 0.23224 | APPROXIMATE F | 127.66442 | DEGREES OF FREEDOM | SIGNIFICANCE |
| RAO'S V | 1475.55976 | CHANGE IN V | 9.60353 | 8 | 0.003 |
| | | | | 2 | 0.003 |

F MATRIX - DEGREES OF FREEDOM: 4, 475

| | | | |
|---------|-----------|-----------|---------|
| | | GROUP 1 | GROUP 2 |
| GROUP 2 | 122.31507 | | |
| GROUP 3 | 63.66931 | 351.72377 | |

----- VARIABLES IN THE ANALYSIS -----

| VARIABLE | ENTRY CRITERION | F TO REMOVE |
|----------|-----------------|-------------|
| V1 | 2.43409 | 414.56776 |
| V2 | 3.37537 | 13.22976 |
| V3 | 2.37340 | 19.35291 |
| V6 | 3.40573 | 3.32053 |

----- VARIABLE NOT IN THE ANALYSIS -----

| VARIABLE | TOLERANCE | F TO ENTER | ENTRY CRITERION |
|----------|-----------|------------|-----------------|
| V4 | 0.41885 | 3.45717 | 3.38149 |
| V5 | 0.33504 | 0.84329 | 3.37639 |



VARIABLE ENTERED ON STEP NUMBER 3... V4

| | | | | | |
|---------------|------------|---------------|-----------|--------------------|--------------|
| WILKS' LAMBDA | 0.22820 | APPROXIMATE F | 103.34616 | DEGREES OF FREEDOM | SIGNIFICANCE |
| RAO'S V | 1483.26620 | CHANGE IN V | 7.50645 | 10 | 0.000 |
| | | | | 2 | 0.022 |

F MATRIX - DEGREES OF FREEDOM: 5, 474

| | | |
|---------|----------|-----------|
| | GROUP 1 | GROUP 2 |
| GROUP 2 | 48.62474 | |
| GROUP 3 | 51.91333 | 250.84703 |

----- VARIABLES IN THE ANALYSIS -----

| VARIABLE | ENTRY CRITERION | F TO REMOVE |
|----------|-----------------|-------------|
| V1 | 2.43409 | 514.35346 |
| V2 | 3.37537 | 11.17125 |
| V3 | 2.87340 | 17.30453 |
| V4 | 3.47733 | 3.47717 |
| V6 | 3.40673 | 6.65269 |

----- VARIABLE 11 IN THE ANALYSIS -----

| VARIABLE | TOLERANCE | F TO ENTER | ENTRY CRITERION |
|----------|-----------|------------|-----------------|
| V5 | 0.74695 | 2.24635 | 3.40703 |

ANALYSIS DISCRIMINANTE

* 2/02/81

PAGE 10



 VARIABLE ENTERED ON STEP NUMBER 6.. V5

| | | | | | |
|---------------|------------|---------------|----------|--------------------|--------------|
| WILKS' LAMBDA | 0.22575 | APPROXIMATE F | 86.72079 | DEGREES OF FREEDOM | SIGNIFICANCE |
| RAO'S V | 1471.40112 | CHANGE IN V | 3.13492 | 12 | 0.000 |
| | | | | 2 | 0.017 |

F MATRIX - DEGREES OF FREEDOM: 0, 473

| | | | | |
|-------|-------|----------|-----------|---|
| | GROUP | 1 | GROUP | 2 |
| GROUP | 2 | 83.29325 | | |
| GROUP | 3 | 43.31633 | 234.03391 | |

----- VARIABLES IN THE ANALYSIS -----

| VARIABLE | ENTRY CRITERION | F TO REMOVE |
|----------|--------------------|----------------|
| V1 | 2.43409 | 513.59118 |
| V2 | 3.37537 | 10.35377 |
| V3 | 2.37840 | 13.03336 |
| V4 | 3.47930 | 4.85537 |
| V5 | 3.40703 | 2.24635 |
| V6 | 3.40673 | 7.69332 |

ALL ELIGIBLE VARIABLES INCLUDED

----- VARIABLES NOT IN THE ANALYSIS -----

| VARIABLE | TOLERANCE | F TO ENTER | ENTRY CRITERION |
|----------|-----------|---------------|--------------------|
|----------|-----------|---------------|--------------------|

ANALISIS DISCRIMINANTE

12/02/81

PAGE 11

FILE ROCA (CREATION DATE = 12/02/81)



----- DISCRIMINANT ANALYSIS -----

SUMMARY TABLE

| STEP NUMBER | VARIABLE ENTERED | VARIABLE REMOVED | F TO ENTER OR REMOVE | NUMBER INCLUDED | WILKS' LAMBDA | SIG. | RAO'S V | CHANGE IN RAO'S V | SIG. OF CHANGE |
|-------------|------------------|------------------|----------------------|-----------------|---------------|-------|------------|-------------------|----------------|
| 1 | V1 | | 576.61997 | 1 | 0.26103 | 0.000 | 1357.23993 | 1357.23993 | 0.000 |
| 2 | V3 | | 12.10893 | 2 | 0.24841 | 0.000 | 1391.83962 | 45.79969 | 0.000 |
| 3 | V2 | | 13.06536 | 3 | 0.23549 | 0.000 | 1471.05126 | 67.01164 | 0.000 |
| 4 | V6 | | 3.32053 | 4 | 0.23224 | 0.000 | 1471.65970 | 9.60850 | 0.008 |
| 5 | V4 | | 3.45717 | 5 | 0.22890 | 0.000 | 1471.25620 | 7.60645 | 0.022 |
| 6 | V5 | | 2.24635 | 6 | 0.22675 | 0.000 | 1471.40112 | 8.13492 | 0.017 |

CLASSIFICATION FUNCTION COEFFICIENTS

| | GROUP 1 | GROUP 2 | GROUP 3 |
|----------|-----------|-----------|-----------|
| V1 | 0.00371 | -0.00261 | 0.00306 |
| V2 | 3.03424 | 3.25731 | 4.40752 |
| V3 | 0.26303 | 0.25532 | 0.30130 |
| V4 | 0.13793 | 1.14104 | 0.74749 |
| V5 | 5.45884 | 5.20293 | 5.47742 |
| V6 | 9.97424 | 3.92407 | 9.26531 |
| CONSTANT | -40.26332 | -34.01229 | -43.00053 |

| DISCRIMINANT FUNCTION | EIGENVALUE | RELATIVE PERCENTAGE | CANONICAL CORRELATION | FUNCTIONS DERIVED | WILKS' LAMBDA | CHI-SQUARE | DF | SIGNIFICANCE |
|-----------------------|------------|---------------------|-----------------------|-------------------|---------------|------------|----|--------------|
| 1 | 3.02410 | 70.93 | 0.857 | 1 | 0.2287 | 705.606 | 12 | 0.000 |
| 2 | 1.09594 | 1.07 | 0.296 | 1 | 0.211 | 43.561 | 5 | 0.000 |

REMAINING COMPUTATIONS WILL BE BASED ON 2 DISCRIMINANT FUNCTION(S)

ANALYSIS DISCRIMINANTE

12/01/81

PAGE 12



STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

| | FUNC 1 | FUNC 2 |
|----|----------|----------|
| V1 | -0.97168 | 0.18789 |
| V2 | -0.11384 | -0.41177 |
| V3 | -0.13811 | -0.52409 |
| V4 | 0.02746 | -0.95273 |
| V5 | -0.03213 | 0.31851 |
| V6 | -0.04632 | 0.93013 |

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

| | FUNC 1 | FUNC 2 |
|----------|----------|----------|
| V1 | -0.00071 | 0.00014 |
| V2 | -0.15317 | -0.57217 |
| V3 | -0.00530 | -0.02010 |
| V4 | 0.03934 | -1.25175 |
| V5 | -0.04068 | 0.40264 |
| V6 | -0.05463 | 1.20623 |
| CONSTANT | 1.63143 | -0.15225 |

CENTROIDS OF GROUPS IN REDUCED SPACE

| | | FUNC 1 | FUNC 2 |
|-------|---|----------|----------|
| GROUP | 1 | -0.17737 | 0.43918 |
| GROUP | 2 | 1.02943 | -0.15042 |
| GROUP | 3 | -1.03957 | -0.24787 |



| SUBFIL | CASE SEQNUM | MISSING VALUES | ACTUAL GROUP | HIGHEST PROBABILITY | | 2ND HIGHEST GROUP | DISCRIMINANT SCORES | | |
|--------|-------------|----------------|--------------|---------------------|---------------|-------------------|---------------------|--------|-------|
| | | | | GROUP | P(X/G) P(G/X) | | FUNC 1 | FUNC 2 | |
| ROCA | 1. | | 1 | **** | 2 | 0.993 | 0.271 | 0.140 | 0.227 |
| ROCA | 2. | | 1 | **** | 2 | 0.993 | 0.252 | 0.149 | 0.275 |
| ROCA | 3. | | 1 | **** | 2 | 0.987 | 0.285 | 0.149 | 0.247 |
| ROCA | 4. | | 1 | | 1 | 0.930 | 0.478 | 0.145 | 0.215 |
| ROCA | 5. | | 1 | | 1 | 0.516 | 0.239 | -0.106 | 0.153 |
| ROCA | 6. | | 1 | | 1 | 1.000 | 0.201 | -0.104 | 0.159 |
| ROCA | 7. | | 1 | | 1 | 0.930 | 0.209 | -0.104 | 0.228 |
| ROCA | 8. | | 1 | | 1 | 0.998 | 0.370 | -0.156 | 0.186 |
| ROCA | 9. | | 1 | | 1 | 1.000 | 0.323 | -0.127 | 0.186 |
| ROCA | 10. | | 1 | | 1 | 1.000 | 0.725 | -0.333 | 0.112 |
| ROCA | 11. | | 1 | | 1 | 0.999 | 0.710 | -0.469 | 0.176 |
| ROCA | 12. | | 1 | | 1 | 0.999 | 0.513 | -0.447 | 0.145 |
| ROCA | 13. | | 1 | | 1 | 1.000 | 0.354 | -0.238 | 0.148 |
| ROCA | 14. | | 1 | | 1 | 0.998 | 0.310 | -0.364 | 0.175 |
| ROCA | 15. | | 1 | | 1 | 0.994 | 0.132 | -0.271 | 0.330 |
| ROCA | 16. | | 1 | | 1 | 1.000 | 0.347 | -0.257 | 0.193 |
| ROCA | 17. | | 1 | | 1 | 1.000 | 0.773 | -0.363 | 0.195 |
| ROCA | 18. | | 1 | | 1 | 0.994 | 0.365 | -0.264 | 0.225 |
| ROCA | 19. | | 1 | | 1 | 0.997 | 0.300 | -0.310 | 0.264 |
| ROCA | 20. | | 1 | | 1 | 0.999 | 0.122 | -0.266 | 0.121 |
| ROCA | 21. | | 1 | | 1 | 0.999 | 0.363 | -0.262 | 0.224 |
| ROCA | 22. | | 1 | **** | 3 | 0.997 | 0.417 | -0.335 | 0.170 |
| ROCA | 23. | | 1 | | 1 | 0.999 | 0.203 | -0.375 | 0.123 |
| ROCA | 24. | | 1 | | 1 | 0.999 | 0.256 | -0.501 | 0.119 |
| ROCA | 25. | | 1 | | 1 | 0.993 | 0.405 | -0.586 | 0.124 |
| ROCA | 26. | | 1 | **** | 3 | 0.994 | 0.268 | -0.460 | 0.171 |
| ROCA | 27. | | 1 | **** | 3 | 0.993 | 0.267 | -0.625 | 0.134 |
| ROCA | 28. | | 1 | **** | 3 | 0.993 | 0.266 | -0.510 | 0.134 |
| ROCA | 29. | | 1 | **** | 3 | 0.993 | 0.374 | -0.384 | 0.134 |
| ROCA | 30. | | 1 | **** | 3 | 1.000 | 0.366 | -0.633 | 0.146 |
| ROCA | 31. | | 1 | **** | 3 | 0.999 | 0.369 | -0.300 | 0.177 |
| ROCA | 32. | | 1 | **** | 3 | 0.999 | 0.351 | -0.319 | 0.193 |
| ROCA | 33. | | 1 | **** | 3 | 0.999 | 0.171 | -0.316 | 0.133 |
| ROCA | 34. | | 1 | **** | 3 | 0.999 | 0.187 | -0.031 | 0.173 |
| ROCA | 35. | | 1 | **** | 3 | 0.999 | 0.141 | -0.199 | 0.199 |
| ROCA | 36. | | 1 | **** | 3 | 0.999 | 0.427 | -0.327 | 0.195 |
| ROCA | 37. | | 1 | **** | 3 | 1.000 | 0.083 | -0.269 | 0.158 |
| ROCA | 38. | | 1 | **** | 3 | 0.999 | 0.044 | -0.363 | 0.174 |
| ROCA | 39. | | 1 | | 1 | 1.000 | 0.106 | -0.075 | 0.178 |
| ROCA | 40. | | 1 | | 1 | 0.997 | 0.261 | -0.170 | 0.301 |
| ROCA | 41. | | 1 | | 1 | 0.997 | 0.274 | -0.195 | 0.114 |
| ROCA | 42. | | 1 | | 1 | 0.997 | 0.240 | -0.191 | 0.177 |
| ROCA | 43. | | 1 | | 1 | 1.000 | 0.083 | -0.102 | 0.194 |
| ROCA | 44. | | 1 | | 1 | 0.999 | 0.200 | -0.277 | 0.163 |
| ROCA | 45. | | 1 | | 1 | 0.993 | 0.302 | -0.241 | 0.145 |
| ROCA | 46. | | 1 | | 1 | 0.993 | 0.253 | -0.277 | 0.140 |
| ROCA | 47. | | 1 | | 1 | 0.993 | 0.253 | -0.277 | 0.140 |

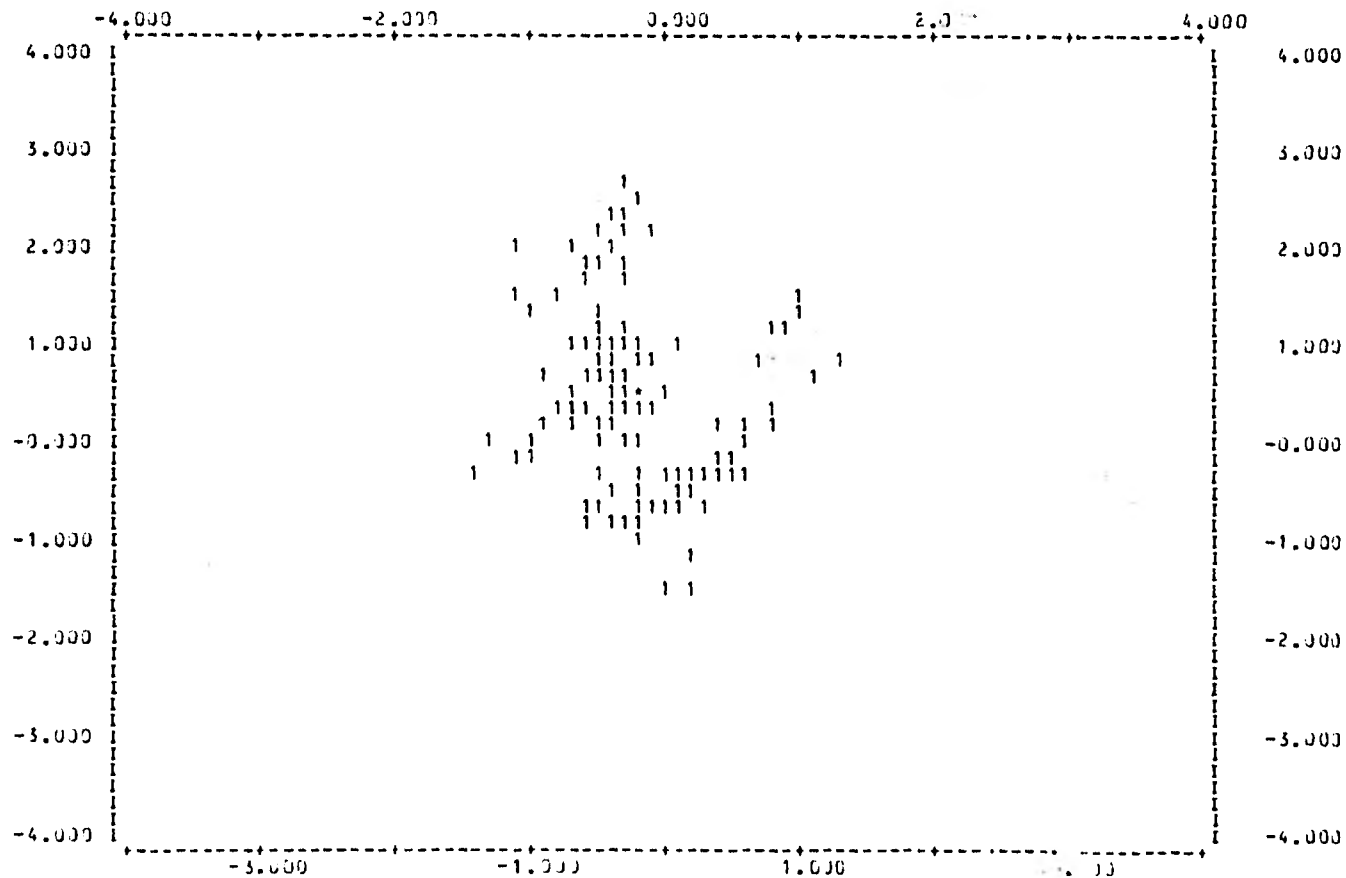
ANALISIS DISCRIMINANTE

| | | | | | | | | | |
|------|------|---|---|-------|-------|---|-------|-------|-------|
| ROCA | 102. | 0 | 1 | 0.797 | J.455 | 2 | 0.122 | 0.122 | 3.222 |
| ROCA | 103. | 0 | 1 | 0.785 | J.452 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 104. | 0 | 1 | 0.750 | J.653 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 105. | 0 | 1 | 0.000 | J.836 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 106. | 0 | 1 | 0.000 | J.441 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 107. | 0 | 1 | 0.000 | J.844 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 108. | 0 | 1 | 0.791 | J.461 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 109. | 0 | 1 | 0.716 | J.627 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 110. | 0 | 1 | 0.770 | J.689 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 111. | J | 1 | 0.663 | J.699 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 112. | 0 | 1 | 0.775 | J.711 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 113. | 0 | 1 | 0.000 | J.811 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 114. | 0 | 1 | 0.000 | J.855 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 115. | 0 | 1 | 0.000 | J.903 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 116. | 0 | 1 | 0.000 | J.630 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 117. | 0 | 1 | 0.000 | J.767 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 118. | 0 | 1 | 0.000 | J.779 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 119. | 0 | 1 | 0.000 | J.755 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 120. | J | 1 | 0.790 | J.629 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 121. | J | 1 | 0.761 | J.604 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 122. | 0 | 1 | 0.748 | J.563 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 123. | 0 | 1 | 0.755 | J.552 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 124. | 0 | 1 | 0.744 | J.556 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 125. | 0 | 1 | 0.742 | J.555 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 126. | 0 | 1 | 0.000 | J.717 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 127. | 0 | 1 | 0.739 | J.747 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 128. | 0 | 1 | 0.000 | J.631 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 129. | 0 | 1 | 0.000 | J.733 | 2 | 0.000 | 0.000 | 3.541 |
| ROCA | 130. | 0 | 1 | 0.733 | J.766 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 131. | 0 | 1 | 0.771 | J.300 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 132. | 0 | 1 | 0.706 | J.766 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 133. | 0 | 1 | 0.732 | J.702 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 134. | 0 | 1 | 0.731 | J.756 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 135. | 0 | 1 | 0.734 | J.644 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 136. | 0 | 1 | 0.733 | J.751 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 137. | 0 | 1 | 0.737 | J.695 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 138. | 0 | 1 | 0.732 | J.537 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 139. | 0 | 1 | 0.788 | J.503 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 140. | 0 | 1 | 0.787 | J.511 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 141. | 0 | 1 | 0.786 | J.510 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 142. | 0 | 1 | 0.776 | J.535 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 143. | 0 | 3 | 0.795 | J.560 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 144. | 0 | 1 | 0.775 | J.622 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 145. | J | 1 | 0.660 | J.733 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 146. | J | 3 | 0.735 | J.605 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 147. | J | 3 | 0.519 | J.600 | 2 | 0.122 | 0.122 | 3.541 |
| ROCA | 148. | 0 | 3 | 0.647 | J.614 | 2 | 0.122 | 0.122 | 3.541 |

END OF DATA INPUT, READ COUNT = 481



PLOT OF DISCRIMINANT SCORE 1 (HORIZONTAL) VS. DISCRIMINANT SCORE 2 (VERTICAL). * INDICATES A GROUP CENTROID.
THE SYMBOL '1' DENOTES A CASE FROM GROUP 1 000000000000000000





| CASE SUBFIL | CASE SEGNJM | MISSING VALUES | ACTUAL GROUP | HIGHEST PROBABILITY GROUP | HIGHEST P(X/G) | P(G/X) | 2ND HIGHEST GROUP | 2ND HIGHEST P(G/X) | DISCRIMINANT FUNC 1 | DISCRIMINANT FUNC 2 | CORES |
|----------------|----------------|-------------------|-----------------|------------------------------|-------------------|--------|----------------------|-----------------------|------------------------|------------------------|-------|
| ROCA | 149. | U | 2 | 2 | 0.914 | J.931 | 1 | 0.019 | 1.022 | 1.524 | |
| ROCA | 150. | U | 2 | 2 | 0.919 | J.933 | 1 | 0.017 | 1.045 | 1.507 | |
| ROCA | 151. | U | 2 | 2 | 0.936 | J.934 | 1 | 0.016 | 1.033 | 1.596 | |
| ROCA | 152. | U | 2 | 2 | 0.937 | J.932 | 1 | 0.038 | 0.964 | 1.544 | |
| ROCA | 153. | U | 2 | 2 | 1.000 | J.919 | 1 | 0.080 | 0.870 | 1.508 | |
| ROCA | 154. | U | 2 | 2 | 0.915 | J.957 | 1 | 0.031 | 0.921 | 1.506 | |
| ROCA | 155. | U | 2 | 2 | 0.919 | J.953 | 1 | 0.042 | 0.955 | 1.734 | |
| ROCA | 156. | U | 2 | 2 | 1.000 | J.950 | 1 | 0.000 | 0.953 | 1.637 | |
| ROCA | 157. | U | 2 | 2 | 0.973 | J.957 | 1 | 0.033 | 0.977 | 1.535 | |
| ROCA | 158. | U | 2 | 2 | 1.000 | U.954 | 1 | 0.046 | 0.966 | 1.533 | |
| ROCA | 159. | U | 2 | 2 | 1.000 | J.953 | 1 | 0.077 | 0.919 | 1.518 | |
| ROCA | 160. | U | 2 | 2 | 1.000 | J.954 | 1 | 0.045 | 0.963 | 1.533 | |
| ROCA | 161. | U | 2 | 2 | 0.974 | J.974 | 1 | 0.030 | 1.000 | 1.963 | |
| ROCA | 162. | U | 2 | 2 | 0.933 | J.962 | 1 | 0.037 | 0.990 | 1.665 | |
| ROCA | 163. | U | 2 | 2 | 0.932 | J.984 | 1 | 0.016 | 1.044 | 1.527 | |
| ROCA | 164. | U | 2 | 2 | 0.937 | J.957 | 1 | 0.042 | 0.951 | 1.750 | |
| ROCA | 165. | U | 2 | 2 | 1.000 | J.954 | 1 | 0.046 | 0.966 | 1.537 | |
| ROCA | 166. | U | 2 | 2 | 0.934 | J.966 | 1 | 0.034 | 1.026 | 1.533 | |
| ROCA | 167. | U | 2 | 2 | 0.947 | J.973 | 1 | 0.027 | 1.051 | 1.526 | |
| ROCA | 168. | U | 2 | 2 | 0.960 | J.975 | 1 | 0.023 | 1.023 | 1.568 | |
| ROCA | 169. | U | 2 | 2 | 0.939 | J.974 | 1 | 0.026 | 1.024 | 1.570 | |
| ROCA | 170. | U | 2 | 2 | 0.936 | U.974 | 1 | 0.026 | 1.037 | 1.504 | |
| ROCA | 171. | U | 2 | 2 | 0.934 | J.974 | 1 | 0.026 | 1.037 | 1.514 | |
| ROCA | 172. | U | 2 | 2 | 1.000 | J.952 | 1 | 0.047 | 0.957 | 1.524 | |
| ROCA | 173. | U | 2 | 2 | 1.000 | J.953 | 1 | 0.047 | 0.964 | 1.532 | |
| ROCA | 174. | U | 2 | 2 | 1.000 | J.956 | 1 | 0.044 | 0.957 | 1.431 | |
| ROCA | 175. | U | 2 | 2 | 0.939 | J.949 | 1 | 0.051 | 0.913 | 1.710 | |
| ROCA | 176. | U | 2 | 2 | 1.000 | J.930 | 1 | 0.070 | 0.922 | 1.611 | |
| ROCA | 177. | U | 2 | 2 | 1.000 | J.931 | 1 | 0.063 | 0.929 | 1.536 | |
| ROCA | 178. | U | 2 | 2 | 1.000 | J.930 | 1 | 0.070 | 0.919 | 1.733 | |
| ROCA | 179. | U | 2 | 2 | 1.000 | J.927 | 1 | 0.072 | 0.904 | 1.577 | |
| ROCA | 180. | U | 2 | 2 | 0.974 | J.970 | 1 | 0.030 | 0.995 | 1.770 | |
| ROCA | 181. | U | 2 | 2 | 1.000 | J.933 | 1 | 0.046 | 0.992 | 1.575 | |
| ROCA | 182. | U | 2 | 2 | 1.000 | J.934 | 1 | 0.065 | 0.949 | 1.533 | |
| ROCA | 183. | U | 2 | 2 | 0.930 | J.959 | 1 | 0.041 | 1.069 | 1.329 | |
| ROCA | 184. | U | 2 | 2 | 1.000 | J.924 | 1 | 0.076 | 0.943 | 1.435 | |
| ROCA | 185. | U | 2 | 2 | 0.918 | J.977 | 1 | 0.033 | 1.067 | 1.452 | |
| ROCA | 186. | U | 2 | 2 | 1.000 | J.926 | 1 | 0.032 | 1.042 | 1.433 | |
| ROCA | 187. | U | 2 | 2 | 1.000 | J.963 | 1 | 0.032 | 1.027 | 1.612 | |
| ROCA | 188. | U | 2 | 2 | 1.000 | J.967 | 1 | 0.035 | 1.013 | 1.513 | |
| ROCA | 189. | U | 2 | 2 | 1.000 | J.965 | 1 | 0.032 | 1.033 | 1.513 | |
| ROCA | 190. | U | 2 | 2 | 1.000 | J.967 | 1 | 0.035 | 1.021 | 1.513 | |
| ROCA | 191. | U | 2 | 2 | 1.000 | J.963 | 1 | 0.035 | 1.013 | 1.513 | |
| ROCA | 192. | U | 2 | 2 | 1.000 | J.963 | 1 | 0.035 | 1.004 | 1.513 | |
| ROCA | 193. | U | 2 | 2 | 1.000 | J.967 | 1 | 0.035 | 1.004 | 1.513 | |
| ROCA | 194. | U | 2 | 2 | 1.000 | J.974 | 1 | 0.026 | 1.429 | 1.513 | |
| ROCA | 195. | U | 2 | 2 | 1.000 | J.967 | 1 | 0.032 | 1.329 | 1.597 | |

ANALISIS DISCRIMINANTE

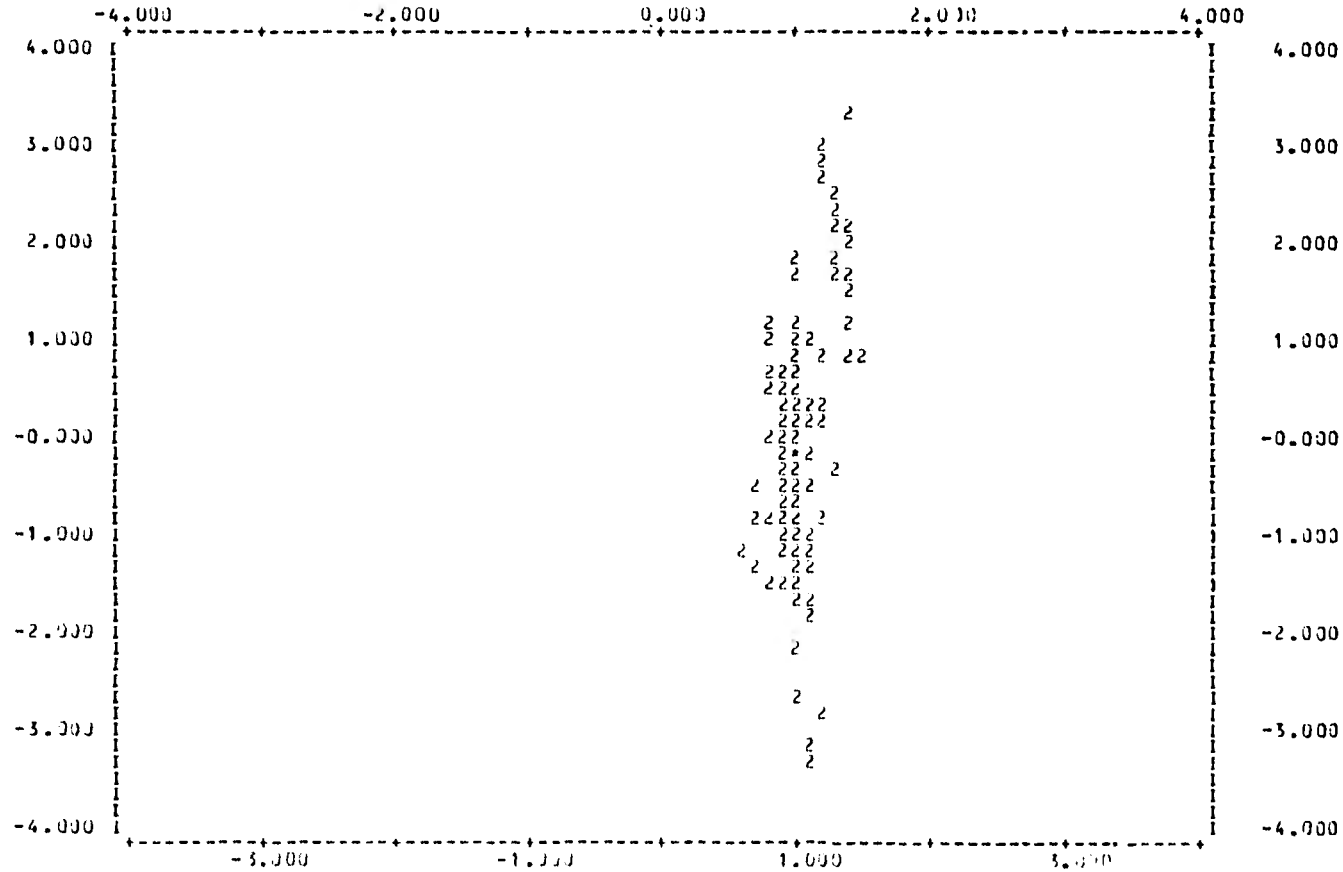
| | | | | | | | | | | |
|------|------|---|---|-------|----|-------|---|-------|-------|------|
| ROCA | 250. | J | N | 0.000 | J. | 0.950 | 1 | 0.069 | 0.946 | 0.15 |
| ROCA | 251. | J | N | 0.000 | J. | 0.963 | 1 | 0.132 | 0.868 | 0.18 |
| ROCA | 252. | J | N | 0.000 | J. | 0.868 | 1 | 0.102 | 0.898 | 0.10 |
| ROCA | 253. | J | N | 0.000 | J. | 0.659 | 1 | 0.341 | 0.659 | 0.34 |
| ROCA | 254. | J | N | 0.000 | J. | 0.743 | 1 | 0.257 | 0.743 | 0.26 |
| ROCA | 255. | J | N | 0.000 | J. | 0.743 | 1 | 0.257 | 0.743 | 0.26 |
| ROCA | 256. | J | N | 0.000 | J. | 0.202 | 1 | 0.798 | 0.202 | 0.80 |
| ROCA | 257. | J | N | 0.000 | J. | 0.704 | 1 | 0.296 | 0.704 | 0.30 |
| ROCA | 258. | J | N | 0.000 | J. | 0.729 | 1 | 0.271 | 0.729 | 0.27 |
| ROCA | 259. | J | N | 0.000 | J. | 0.859 | 1 | 0.141 | 0.859 | 0.14 |
| ROCA | 260. | J | N | 0.000 | J. | 0.732 | 1 | 0.268 | 0.732 | 0.27 |
| ROCA | 261. | J | N | 0.000 | J. | 0.300 | 1 | 0.700 | 0.300 | 0.70 |
| ROCA | 262. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 263. | J | N | 0.000 | J. | 0.427 | 1 | 0.573 | 0.427 | 0.57 |
| ROCA | 264. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 265. | J | N | 0.000 | J. | 0.427 | 1 | 0.573 | 0.427 | 0.57 |
| ROCA | 266. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 267. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 268. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 269. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 270. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 271. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 272. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 273. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 274. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 275. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 276. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 277. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 278. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 279. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 280. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 281. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 282. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 283. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 284. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 285. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 286. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 287. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 288. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 289. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 290. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 291. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 292. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 293. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 294. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 295. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 296. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 297. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 298. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 299. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 300. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 301. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 302. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |
| ROCA | 303. | J | N | 0.000 | J. | 0.313 | 1 | 0.687 | 0.313 | 0.69 |

1.71 2/31

PAGE 19



PLOT OF DISCRIMINANT SCORE 1 (HORIZONTAL) VS. DISCRIMINANT SCORE 2 (VERTICAL). * INDICATES A GROUP CENTROID.
THE SYMBOL 2 DENOTES A CASE FROM GROUP 2 0100010001000100



ANALISIS DISCRIMINANTE

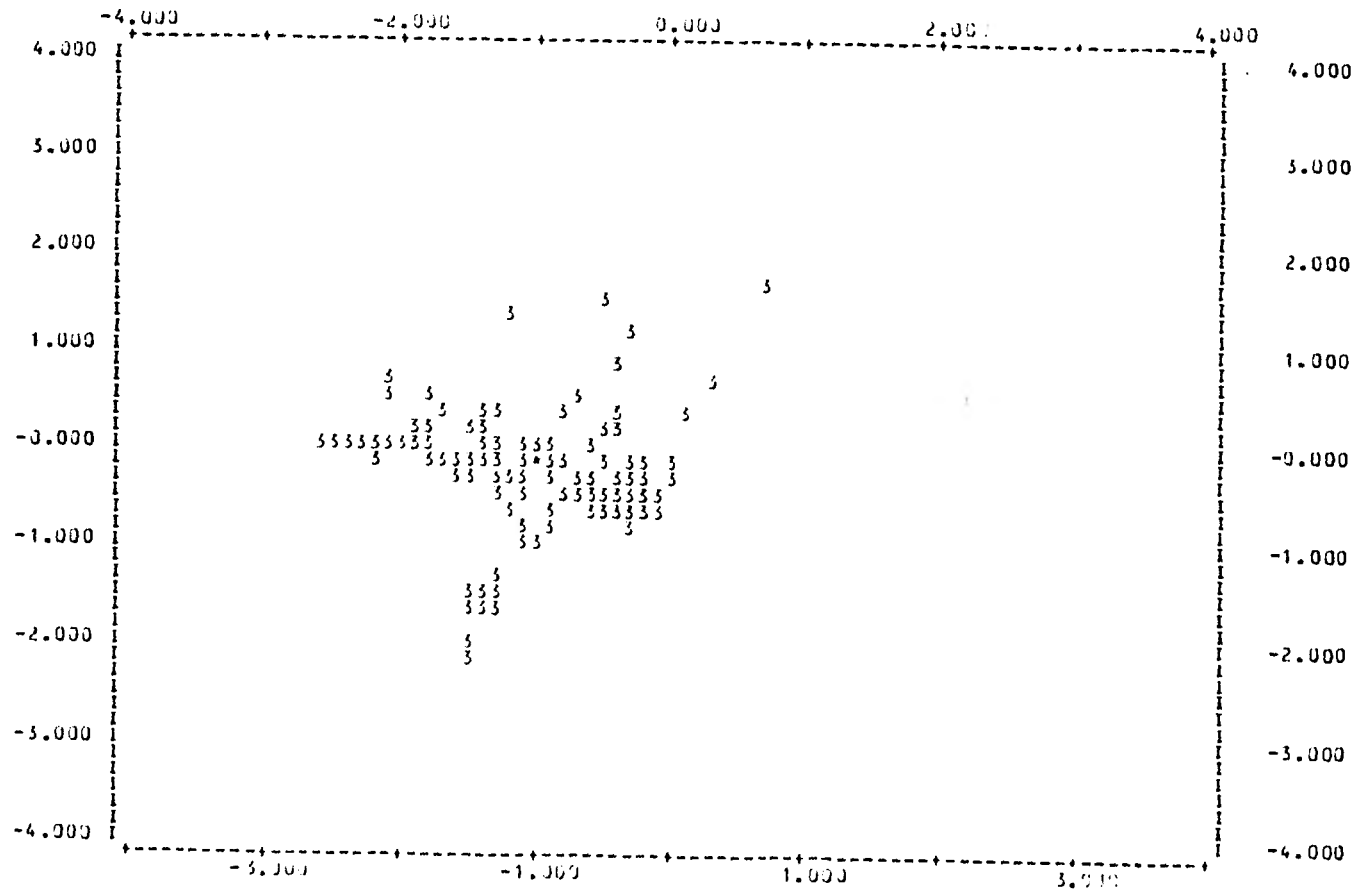
12/02/11

PAGE 22



| CASE SUBFIL | SEQNUM | MISSING VALUES | ACTUAL GROUP | HIGHEST PROBABILITY | | 2ND HIGHEST GROUP | DISCRIMINANT SCORES | | |
|----------------|--------|-------------------|-----------------|---------------------|---------------|----------------------|---------------------|----------|-----|
| | | | | GROUP | P(X/G) P(G/X) | | FUNC 1 | FUNC 2 | |
| ROCA | 329. | 0 | 3 | 0 | .3771 | 0 | 0 | -10.1133 | 111 |
| ROCA | 330. | 0 | 3 | 0 | .3776 | 0 | 0 | -10.1148 | 111 |
| ROCA | 331. | 0 | 3 | 0 | .3776 | 0 | 0 | -10.1141 | 111 |
| ROCA | 332. | 0 | 3 | 0 | .3699 | 0 | 0 | -10.2219 | 111 |
| ROCA | 333. | 0 | 3 | 0 | .355 | 0 | 0 | -10.117 | 111 |
| ROCA | 334. | 0 | 3 | 0 | .368 | 0 | 0 | -10.2260 | 111 |
| ROCA | 335. | 0 | 3 | 0 | .368 | 0 | 0 | -10.2267 | 111 |
| ROCA | 336. | 0 | 3 | 0 | .395 | 0 | 0 | -10.2255 | 111 |
| ROCA | 337. | 0 | 3 | 0 | .397 | 0 | 0 | -10.2240 | 111 |
| ROCA | 338. | 0 | 3 | 0 | .399 | 0 | 0 | -10.2254 | 111 |
| ROCA | 339. | 0 | 3 | 0 | .392 | 0 | 0 | -10.3377 | 111 |
| ROCA | 340. | 0 | 3 | 0 | .392 | 0 | 0 | -10.3367 | 111 |
| ROCA | 341. | 0 | 3 | 0 | .377 | 0 | 0 | -10.4422 | 111 |
| ROCA | 342. | 0 | 3 | 0 | .369 | 0 | 0 | -10.3777 | 111 |
| ROCA | 343. | 0 | 3 | 0 | .369 | 0 | 0 | -10.3353 | 111 |
| ROCA | 344. | 0 | 3 | 0 | .369 | 0 | 0 | -10.3353 | 111 |
| ROCA | 345. | 0 | 3 | 0 | .369 | 0 | 0 | -10.3353 | 111 |
| ROCA | 346. | 0 | 3 | 0 | .369 | 0 | 0 | -10.4455 | 111 |
| ROCA | 347. | 0 | 3 | 0 | .369 | 0 | 0 | -10.4455 | 111 |
| ROCA | 348. | 0 | 3 | 0 | .369 | 0 | 0 | -10.4455 | 111 |
| ROCA | 349. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 350. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 351. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 352. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 353. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 354. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 355. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 356. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 357. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 358. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 359. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 360. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 361. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 362. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 363. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 364. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 365. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 366. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 367. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 368. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 369. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 370. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 371. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 372. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 373. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 374. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |
| ROCA | 375. | 0 | 3 | 0 | .366 | 0 | 0 | -10.4455 | 111 |

PLOT OF DISCRIMINANT SCORE 1 (HORIZONTAL) VS. DISCRIMINANT SCORE 2 (VERTICAL). * INDICATES A GROUP CENTROID.
THE SYMBOL 'S' DENOTES A CASE FROM GROUP 3 0000000000000000



ANALISIS DISCRIMINANTE
END OF DATA INPUT, READ COUNT = 481

12/02/81

PAGE 26



TERRITORIAL MAP OF DISCRIMINANT SCORE 1 (HORIZONTAL) VS. DISCRIMINANT SCORE 2 (VERTICAL). * INDICATES A GROUP CE



ANALYSIS DISCRIMINANTE

1 / 27 / 81

PAGE 23



PREDICTION RESULTS -

| ACTUAL GROUP | NO. OF CASES | PREDICTED GROUP MEMBERSHIP | | |
|--------------|--------------|----------------------------|----------------|---------------|
| | | GP. 1 | GP. 2 | GP. 3 |
| GROUP 1 | 142. | 103. 73.0% | 23. 15.5% | 17. 11.5% |
| GROUP 2 | 130. | 0. 0.0% | 130. 100.0% | 0. 0.0% |
| GROUP 3 | 155. | 47. 30.7% | 1. 0.7% | 105. 68.6% |

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 81.73%

BIBLIOGRAFIA

1. ANDERSON, T.W. An Introduction to Multivariate Statical Analysis. New York, John Wiley & Sons, - Inc. 1958.
2. ARANDA ORDAZ, FRANCISCO JAVIER & MENDEZ, IGNACIO. Análisis Discriminante. México. IIMAS, 1977 (Comunicaciones Técnicas, serie naranja- 151).
3. COOLEY, WILLIAM W & LOHNES PAUL R. Multivariate Data Ana^lysis. New York, John Wiley & Sons, Inc. 1966.
4. CRAIG, ALLEN T. & HOGG, ROBERT V. Introduction to Mathe^matical Statistics. 3a. ed. New York, -- Collier MacMillan, 1970.
5. DAVIS, JOHN C. Statistics and Data Analysis in Geology.- New York, John Wiley & Sons, Inc. 1973.
6. DOBRIN, MILTON B. Introducción a la Prospección Geofísi- ca. 2a. ed. Ediciones Omega, S.A. Barce- lona 1961.
7. DRAPER, N.R. & SMITH, H. Applied Regression Analysis. -- New York, John Wiley & Sons, Inc. 1966.
8. GEOCHEM LABORATORIES, Inc. Source Rock Evaluation Refe- rence Manual.
9. HARBAUGH, JOHN W. & DOVETON JOHN. H. & DAVIS, JOHN C. --

Probability Methods in Oil Exploration.-
Wiley-Interscience.

10. LONGWELL, CHESTER R. & FLINT, RICHARD F. Geología Física. 1a. ed. Limusa - Wiley, S.A., México, 1965.
11. MATTAUER, MAURICE. Las Deformaciones de los Materiales de la Corteza Terrestre. Ed. Omega, S.A.- Barcelona, 1976.
12. MENDEZ, IGNACIO & RODRIGUEZ SILVIA. Dos ejemplos de aplicación de Análisis Discriminante en Medicina. México, IIMAS, 1978. (Comunicaciones Técnicas, serie naranja, 179).
13. MORRISON, DONALD F. Multivariate Statistical Methods. - 2a. ed. New York, McGraw-Hill. 1967.
14. NIE, NORMAN H. & BENT, O.H. & HULL, C.H. SPSS: Statistical Package for the Social Sciences. 2a. ed. New York, McGraw-Hill. 1977.
15. RUEDA GAXIOLA, JAIME. Los Métodos Palinológico y Geoquímico en la Exploración Petrolera. Subdirección de Tecnología de Exploración.- IMP.
16. TATSUOKA, MAURICE T. Multivariate Analysis. New York, John Wiley & Sons, Inc. 1971.
17. TISSOT, B.P. & WELTE, DH. Petroleum Formation and Occurrence. Springer-Verlag, Berlín Heidelberg-New York, 1978.

18. ZARAZUA MUCIÑO, HUMBERTO. El Petróleo. PEMEX. Dirección General. Biblioteca Central. México, 1969.