



Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS

ALGUNAS APLICACIONES DEL ANALISIS DISCRIMINANTE
A LA MEDICINA Y A LA BIOLOGIA

TESIS

Que para obtener el título de

ACTUARIO

presenta

ALMA ROSA VERA SOLIS

1 9 8 2



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
PARTE I.	
CAPITULO 1. EL ANALISIS DISCRIMINANTE.	
1.1 Introducci3n al An3lisis Discriminante.	4
1.2 Principales definiciones y notaciones.	5
CAPITULO 2. SOLUCIONES POR MEDIO DE LA TEORIA DE DECISIONES.	
2.1 Regla no aleatorizada.	7
2.2 Procedimiento propuesto por Anderson.	10
2.3 Regla Aleatorizada.	12
2.4 Soluci3n de un problema de clasificaci3n cuando las probabilidades a priori son desconocidas.	16
2.5 Desventajas de las soluciones por medio de la teoria de decisiones.	19
CAPITULO 3. SOLUCIONES CLASICAS AL PROBLEMA DE LA DISCRIMINACION PARA DOS POBLACIONES.	
3.1 Discriminaci3n en dos poblaciones con H ₀ mucedasticidad.	20
3.1.1 La funci3n lineal de Fisher.	20
3.1.2 Otros procedimientos para la obtenci3n de funciones lineales.	23
3.1.3 Procedimientos empleados en dos poblaciones Normales con par3metros desconocidos.	26

3.1.4 Solución al problema de clasificación por medio del cociente de verosimilitudes.	28
3.1.5 Aplicación del Análisis de regresión a la discriminación.	31
3.2 Discriminación en poblaciones con Heterocedasticidad.	35
CAPITULO 4. DISCRIMINACION EN K POBLACIONES.	
4.1 Discriminación en k poblaciones normales con homocedasticidad.	37
4.1.1 Aplicación de la teoría de decisiones al problema de Discriminación de k poblaciones Normales con Homocedasticidad.	37
4.2 Solución al problema de Discriminación para k poblaciones mediante la derivación de un conjunto de funciones lineales.	41
CAPITULO 5. SELECCION DE VARIABLES.	
5.1 Métodos de Selección de Variables.	47
5.2 Criterios de Selección de Variables.	49
PARTE II.	
CAPITULO 1. APLICACION DEL ANALISIS DISCRIMINANTE A LA BIOLOGIA.	
1.1 La potencialidad de los mutágenos químicos en la inducción de mutaciones útiles en la agricultura.	53
1.2 Experimento realizado para evaluar los efectos del MMS sobre el Triticale.	55

1.3 Efectos del MMS en la generación MM1.	56
1.3.1 Datos obtenidos para la generación MM1.	56
1.3.2 Aplicación del Análisis Discriminante al experimento MM1.	62
1.3.3 Principales resultados obtenidos para la generación MM1.	62
1.3.4 Análisis de los resultados obtenidos para la generación MM1.	81
1.3.5 Caracterización de los grupos en base al mapa territorial.	84
1.4 Efectos del MMS en la generación MM2.	86
1.4.1 Datos obtenidos para la generación MM2.	86
1.4.2 Aplicación del Análisis Discriminante a la generación MM2.	111
1.4.3 Resultados obtenidos para la generación MM2.	113
1.4.4 Análisis de los resultados obtenidos para la generación MM2.	118
1.4.5 Caracterización de los grupos en base al mapa territorial.	121
1.5 Conclusiones.	123

CAPITULO 2. PRIMER EJEMPLO DE LA APLICACION DEL ANALISIS DISCRIMINANTE A LA MEDICINA.

2.1 Estudio sobre las cifras de amilasa y lipasa sérica en niños lactantes y escolares mexicanos.	124
2.2 Investigación realizada para la determinación de las cifras de amilasa y lipasa séricas en niños mexicanos.	125
2.3 Datos recopilados durante la investigación.	127

2.4 Aplicación del Análisis Discriminante a la investigación.	138
2.5 Principales resultados obtenidos.	140
2.5.1 Resultados obtenidos para el grupo de lactantes y escolares.	140
2.5.2 Resultados obtenidos para los lactantes subdivididos en los grupos de hombres y mujeres.	142
2.5.3 Resultados obtenidos para los escolares subdivididos en grupos de hombres y mujeres.	144
2.6 Análisis de los resultados obtenidos.	146
2.6.1 Análisis de los resultados obtenidos para los grupos de lactantes y escolares.	146
2.6.2 Análisis de los resultados obtenidos para los lactantes subdivididos en los grupos de hombres y mujeres.	147
2.6.3 Análisis de los resultados obtenidos para los escolares subdivididos en los grupos de hombres y mujeres.	148
2.7 Caracterización de los grupos en base al mapa territorial.	150
2.7.1 Caracterización de los lactantes y escolares.	150
2.7.2 Caracterización de los lactantes.	151
2.7.3 Caracterización de los escolares.	152
2.8 Conclusiones.	153

CAPITULO 3. SEGUNDO EJEMPLO DE LA APLICACION DEL ANALISIS DISCRIMINANTE A LA MEDICINA.

3.1 La asociación entre la enfermedad articular degenerativa y la Diabetes Mellitus.	154
--	-----

3.2 Estudio realizado para encontrar la existencia de una posible relación entre la Diabetes Mellitus y la Enfermedad articular degenerativa.	156
3.3 Datos recopilados durante el estudio realizado.	159
3.4 Aplicación del Análisis Discriminante al estudio realizado.	166
3.5 Principales resultados obtenidos.	167
3.6 Análisis de los resultados obtenidos.	171
3.7 Caracterización de los grupos en base al mapa territorial.	174
3.8 Conclusiones.	176
BIBLIOGRAFIA.	177

I N T R O D U C C I O N

El uso de herramientas matemáticas no es exclusivo de investigaciones de este tipo, por lo contrario pueden ser aplicadas a diversos campos de la ciencia. Existen métodos muy útiles cuyas bases matemáticas son procesos sofisticados y muy elaborados, esto restringe a los investigadores para hacer uso de ellos, sin embargo su aplicación por medio de paquetes de computadora es fácil y nos puede proporcionar mejores resultados que los métodos convencionales. Un ejemplo de ello es el Análisis Discriminante, que es una aplicación del Análisis -- Multivariado.

Cuando una población se encuentra dividida en varios grupos, el Análisis Discriminante nos permite, a partir de variables que posee un determinado individuo clasificarlo o caracterizar a los grupos por medio de estas variables con un mínimo de error posible. También puede contribuir a la obtención de conclusiones sobre diferencias en los grupos que forman una población. En el Análisis Discriminante las variables que ayudan a la discriminación se estudian simultáneamente porque de esta manera se tiene una mayor información que la que podrían proporcionar individualmente.

Existen varios enfoques para obtener métodos de discriminación, pero nosotros solo mencionaremos algunos de ellos, - ya que el propósito del presente trabajo no es desarrollar un tratado matemático del Análisis Discriminante, sino proveer de ejemplos prácticos que den una idea del tipo de problemas en - el que puede ser utilizado, dando una sencilla explicación de sus fundamentos para hacerlo accesible a aquellas personas que no tengan conocimientos profundos del Análisis Multivariado.

La aplicación del Análisis Discriminante requiere un - considerable número de cálculos, por ello es indispensable el uso de paquetes de computación que los efectúen con rapidez y precisión. Tres ejemplos prácticos, dos pertenecientes al campo de la Medicina y uno al de la Biología serán analizados por medio del paquete SPSS (Statistical Package for Social Sciences). Debe hacerse notar que al utilizarse los paquetes debe investigarse antes los procedimientos para los cuales fueron - hechos y los supuestos que se deben cumplir, ya que de esta manera podrán aplicarse con eficiencia a la práctica.

Esperamos que el trabajo realizado en páginas subsentes dé un panorama del Análisis Discriminante y motive a usarlo.

P A R T E I

C A P I T U L O 1

EL ANALISIS DISCRIMINANTE

1.1 INTRODUCCION AL ANALISIS DISCRIMINANTE.

El problema de clasificación de un individuo dentro de algún conjunto de poblaciones, o de la caracterización de valores que tomarían los elementos pertenecientes a k grupos puede resolverse por medio del Análisis Discriminante. La situación anterior puede presentarse dentro de muy diversas ramas, por ello, el Análisis Discriminante puede ser aplicado a muy variadas investigaciones aunque no sean de tipo estadístico, como es el caso de investigaciones antropológicas, étnicas, riesgos, selección de personal, Medicina, Biología, etc.

En una investigación donde se encuentran las observaciones divididas en dos o más grupos, se determinan las variables en las cuales se espera que estos grupos difieren y se les llaman "variables discriminantes", de manera que si tenemos una nueva observación, esta puede clasificarse dentro de algún grupo en base a los valores que tomen las variables discriminantes.

La elaboración de un modelo de clasificación debe minimizar los efectos de mala clasificación. Existen varios procedimientos que llevan a cabo la discriminación, estos pueden dividirse en:

- a) Soluciones por medio de la teoría de decisiones.
- b) Soluciones clásicas.
- c) Soluciones Bayesianas.
- d) Soluciones no paramétricas o de distribución libre.

1.2. PRINCIPALES DEFINICIONES Y NOTACIONES.

En el presente trabajo serán utilizadas las siguientes definiciones y notaciones:

$$X' = (X_1, X_2, \dots, X_p)$$

Es el vector que contiene los valores de las p variables discriminantes que caracterizan a un individuo X .

p

Es el número de variables discriminantes.

k

Es el número de grupos en que se encuentra dividido el espacio muestral.

G_i

Es la población i -ésima del espacio muestral.

q_i

Es la probabilidad a priori de que un individuo provenga de la población G_i ($\sum q_i = 1$).

R_i

Es la región de clasificación para el grupo G_i .

$f_i(x)$

Es la función de densidad de G_i .

$C(j/i)$, $j \neq i$

Es el costo de mala clasificación de un individuo que pertenece al grupo i como proveniente del grupo j . $C(j/i) \geq 0$.

$C(i/i) = 0$

Se define para $i = 1, \dots, k$.

C A P I T U L O 2.

SOLUCIONES POR MEDIO DE LA TEORIA DE DECISIONES.

2.1 REGLA NO ALEATORIZADA.

Por medio de este procedimiento se obtienen k regiones, tal que la intersección de cualquier par de ellas es vacía, y minimizan los costos asociados de mala clasificación de un individuo. Para una observación dada, el conjunto de los valores que toman las variables discriminantes forma un vector el cual cae dentro de algunas de las regiones definidas, entonces la clasificación se lleva a cabo asignando la observación como proveniente de la población que tiene el mismo índice de la región a la cual pertenece el vector de las variables discriminantes.

OBTENCION DE LA REGLA NO ALEATORIZADA.

Se supondrá durante el desarrollo de este método que las probabilidades a priori de pertenecer a una población (q_i , $i = 1, 2, \dots, k$), y que los costos de mala clasificación de un individuo ($C(j/i)$, $i = 1, 2, \dots, k$, $j \neq i$) son conocidos.

Sean G_1, G_2, \dots, G_k , k poblaciones con funciones de densidad $f_1(x), f_2(x), \dots, f_k(x)$ respectivamente, se determinarán R_1, R_2, \dots, R_k , k regiones tales que:

$$\bigcup_{i=1}^k R_k = Y \quad \text{y} \quad R_i \cap R_j = \emptyset \quad \forall i \neq j$$

Ω es el espacio de observaciones), y que minimicen la esperanza de la pérdida generada por la mala clasificación de individuos.

Si R_i es la región de clasificación de G_i , entonces - la probabilidad de clasificar en forma acertada una observación dado que proviene de la población G_i es:

$$P(i/i) = \int_{R_i} f_i(x) dx$$

donde $dx = dx_1, dx_2, \dots, dx_p$.

La probabilidad de mala clasificación de una observación que proviene de G_i como de G_j es:

$$P(j/i) = \int_{R_j} f_i(x) dx$$

La probabilidad de extraer una observación de G_i y se clasifique en forma correcta es:

$$q_i P(i/i) = q_i \int_{R_i} f_i(x) dx$$

La probabilidad de extraer una observación de G_i y se clasifique incorrectamente dentro de G_j es:

$$q_i P(j/i) = q_i \int_{R_j} f_i(x)$$

entonces la esperanza de pérdida producida por los costos de mala clasificación de un individuo proveniente de la población G_i es:

$$C_i = C(1/i)P(1/i) + C(2/i) + \dots + C(k/i)P(k/i) =$$

$$= \sum_{j=1}^k C(j/i)P(j/i)$$

y la pérdida total esperada es:

$$C = q_1 C_1 + q_2 C_2 + \dots + q_k C_k = \sum_{i=1}^K q_i C_i$$

$$C = \sum_{i=1}^K q_i C_i = \sum_{i=1}^K q_i \left\{ \sum_{j=1}^K \int_{R_j} C(j/i) f_i(x) dx \right\} =$$

$$= \sum_{j=1}^K \int_{R_j} \sum_{i=1}^K q_i C(j/i) f_i(x) dx = \sum_{j=1}^K \int_{R_j} S_j(x) dx$$

$$\text{donde: } S_j(x) = \sum_{i=1}^K q_i C(j/i) f_i(x)$$

Las k regiones R_1, R_2, \dots, R_k que minimizan C se construyen de tal forma que:

$$X \in R_i \quad \Rightarrow \quad S_i(X) \leq S_j(X) \quad \forall \quad j=1, 2, \dots, k$$

un individuo cuyo vector discriminante está dentro de la región R_i se clasifica como proveniente del grupo G_i .

La partición R_1, R_2, \dots, R_k se obtiene mediante la aplicación del siguiente lema donde $f_i(x) = -S_i(x)$ y $-C$ se maximiza.

Lema. Sea R_1, R_2, \dots, R_k una división del espacio G en k regiones mutuamente exclusivas. Sea $R_1^*, R_2^*, \dots, R_k^*$ una división en k regiones mutuamente exclusivas tales que:

$$x \in R_i^* \quad \Rightarrow \quad f_i(x) \geq f_j(x) \quad j=1, 2, \dots, k$$

con $f_1(x), f_2(x), \dots, f_k(x)$ integrables en R con respecto a \underline{u} na medida ν , entonces:

$$\int_{R_1^*} f_1 \, d\nu + \int_{R_2^*} f_2 \, d\nu + \dots + \int_{R_k^*} f_k \, d\nu \geq \int_{R_1} f_1 \, d\nu + \int_{R_2} f_2 \, d\nu + \dots + \int_{R_k} f_k \, d\nu$$

Si el mínimo de S_1, S_2, \dots, S_k se alcanza en dos o más índices, puede entonces escogerse cualquiera de estos y asignarse la observación a la región correspondiente.

2.2 PROCEDIMIENTO PROPUESTO POR ANDERSON.

Se obtiene una partición R_1, R_2, \dots, R_k la cual minimice la esperanza de pérdida total al clasificar una observación dentro de la población i -ésima, es decir minimizar:

$$\sum_{i=1}^k q_i \left\{ \sum_{j=1}^k \int_{R_j} C(j/i) f_i(x) dx \right\}$$

Como en el caso de la regla no aleatorizada, supondremos que

Las probabilidades apriori son también conocidas.

La probabilidad condicional de que una observación provenga de la población i -ésima dado el vector discriminante x es:

$$P(G_i/x) = \frac{q_i f_i(x)}{\sum_{r=1}^k q_r f_r(x)}$$

La esperanza de pérdida debida a la clasificación de la observación dentro de alguna población j es:

$$\sum_{\substack{i=1 \\ i \neq j}}^k \frac{q_i f_i(x)}{\sum_{r=1}^k q_r f_r(x)} C(j/i) = \sum_{\substack{i=1 \\ i \neq j}}^k P(G_i/x) C(j/i) =$$

$$= \left\{ \sum_{\substack{i=1 \\ i \neq j}}^k q_i f_i(x) C(j/i) \right\} \frac{1}{M} = L_j$$

donde $M = \sum_{r=1}^k q_r f_r$ es una constante.

Se escoge el índice del mfn $\{L_1, L_2, \dots, L_k\}$ y el vector x se asigna a R_j , y de esto como proveniente de G_j , en caso de que varios índices tengan el mínimo es indiferente utilizar cualquiera de ellos. El procedimiento se efectúa para cada una de las observaciones y así se obtienen R_1, R_2, \dots, R_k .

Utilizar la regla no aleatorizada o el procedimiento sugerido por Anderson es indiferente porque las particiones generadas por los dos métodos son iguales, ya que:

$\text{Min } \{L_j, j=1,2,\dots, k\}$ es equivalente a:

$$\text{Min}_j \sum_{\substack{i=1 \\ i \neq j}}^k q_i f_i(x) C(j/i)$$

$$\text{pero: } \text{Min}_j \sum_{\substack{i=1 \\ i \neq j}}^k q_i f_i(x) C(j/i) = \text{Min}_j \sum_{i=1}^k q_i f_i(x) C(j/i) =$$

$$= \text{Min}_j S_j(x)$$

que es el paso en el que asignabamos en la regla no aleatorizada la observación x a R_j .

El procedimiento sugerido por Anderson nos conduce a la minimización de la pérdida total esperada:

$$C = \sum_{i=1}^k q_i \left\{ \sum_{j=1}^k \int_{R_j} C(j/i) f_i(x) dx \right\}$$

puesto que, como se acaba de mencionar, la partición que se obtiene es la misma que la de la regla no aleatorizada.

2.3 REGLA ALEATORIZADA.

Se desea encontrar la mejor función vectorial que indique a cual población debe ser asignado el vector x , después de luego, minimizando los costos debidos a la mala clasificación.

OBTENCION DE LA REGLA ALEATORIZADA.

Sea q_i la probabilidad a priori de extraer una observación de la población G_i con función de densidad $f_i(x)$ para $i = 1, 2, \dots, k$ y $C(j/i)$ los costos de mala clasificación de un individuo como perteneciente a G_j dado que proviene de G_i , se determinará la función vectorial $H : \mathbb{R}^p \rightarrow \mathbb{R}^k$ definida como:

$$H(x) = (\lambda_1(x), \lambda_2(x), \dots, \lambda_k(x))$$

donde la probabilidad de que x se asigne a G_i dados los valores de x es:

$$P(i/x) = \lambda_i(x) \quad i = 1, 2, \dots, k$$

por consiguiente:

$$\lambda_i(x) \geq 0 \quad i = 1, 2, \dots, k \quad \text{y} \quad \sum_{i=1}^k \lambda_i(x) = 1$$

$H(x)$ debe minimizar la esperanza de pérdida total generada por los costos de mala clasificación.

Una vez determinada $H(x)$ se encuentra el índice r tal que:

$$\text{Max} \{ \lambda_1(x), \lambda_2(x), \dots, \lambda_k(x) \} = \lambda_r(x)$$

y se asigna el vector x como proveniente de la población G_r .

Se considerará durante esta sección que las probabilidades a priori (q_i), y los costos de mala clasificación ($C(j/i)$) son conocidos.

La esperanza de la pérdida asociada a la mala clasificación de una observación que proviene de la población G_i es:

$$C_i = \int_{\Omega} C(1/i) \lambda_1(x) f_i(x) + C(2/i) \lambda_2(x) f_i(x) + \dots + C(k/i) \lambda_k(x) f_i(x) dx$$

donde $dx = dx_1 dx_2 \dots dx_p$

y $\Omega =$ Espacio de observaciones.

$$C_i = \int_{\Omega} \sum_{j=1}^k C(j/i) \lambda_j(x) f_i(x) dx = \sum_{j=1}^k \int_{\Omega} C(j/i) \lambda_j(x) f_i(x) dx$$

La pérdida total esperada de mala clasificación es:

$$C = q_1 C_1 + q_2 C_2 + \dots + q_k C_k$$

$$\sum_{i=1}^k q_i C_i = \sum_{i=1}^k q_i \left\{ \sum_{j=1}^k \int_{\Omega} C(j/i) \lambda_j f_i(x) dx \right\} = q_1 \sum_{j=1}^k \int_{\Omega} C(j/1) \lambda_j f_1(x) dx +$$

$$+ q_2 \sum_{j=1}^k \int_{\Omega} C(j/2) \lambda_j f_2(x) dx + \dots + q_k \sum_{j=1}^k \int_{\Omega} C(j/k) \lambda_j f_k(x) dx =$$

$$= \sum_{i=1}^k \int_{\Omega} q_i C(1/i) \lambda_1 f_i(x) dx + \dots + \sum_{i=1}^k \int_{\Omega} q_i C(k/i) \lambda_k(x) f_i(x) dx =$$

$$= \int_{\Omega} S_1(x) \lambda_1(x) dx + \int_{\Omega} S_2(x) \lambda_2(x) dx + \dots + \int_{\Omega} S_k(x) \lambda_k(x) dx =$$

$$= \sum_{i=1}^k \int_{\Omega} S_i(x) \lambda_i(x) dx$$

Lema.- Sean f_1, f_2, \dots, f_k funciones integrables sobre S con respecto a una medida ν y $\phi_i(x) \geq 0$ para $i = 1, 2, \dots, k$ tal que

$\sum_{i=1}^k \theta_i(x) = 1$. Escógase las $\theta_{i_{r+1}}^* = \dots = \theta_{i_k}^* = 0$ y $\theta_{i_1}^* > \dots$,

$\theta_{i_r}^* > 0$ tal que $\theta_1^* + \dots + \theta_k^* = 1$ cuando $f_{i_1} = \dots = f_{i_r} \geq f_{i_{r+1}} \geq \dots \geq f_{i_k}$

entonces

$$\sum_{i=1}^k \int_{\Omega} f_i \theta_i^* dv \geq \sum_{i=1}^k \int_{\Omega} f_i \theta_i dv$$

Utilizaremos el lema anterior para encontrar $H(x) = (\lambda_1(x), \dots, \lambda_k(x))$. Ordenemos las $S_j(x)$, $j=1, 2, \dots, k$ en tal forma:

$$S_{i_{r-1}}(x) = S_{i_r} \leq S_{i_{r+1}}(x) \leq \dots \leq S_{i_k}(x) \Rightarrow -S_{i_1} = \dots =$$

$$= -S_{i_r}(x) \geq -S_{i_{r+1}}(x) \geq \dots \geq -S_{i_k}(x) \text{ y escogemos } \lambda_{i_1}(x) + \dots +$$

$$\lambda_{i_r}(x) = 1 \text{ con } \lambda_{i_1}, \dots, \lambda_{i_r} \geq 0 \text{ y } \lambda_{i_{r+1}}(x) = \dots = \lambda_{i_k}(x) = 0.$$

Aplicaremos el lema con $f_i(x) = -S_i(x)$ para $i=1, 2, \dots, k$ y $\lambda_i(x) = \theta_i^*(x)$ para $i=1, 2, \dots, k$, de esta manera vamos a maximizar $-C$ que equivale a minimizar C .

En el caso que $S_i(x) \leq S_j(x)$, $j=1, 2, \dots, k$, nos percatamos rapidamente que debemos escoger $\lambda_i(x) = 1$ y $\lambda_j(x) = 0$ $j=1, 2, \dots, k, i \neq j$. Como $\lambda_i(x) = P(i/x)$ entonces el vector se asigna a la población G_i .

El problema más general que puede presentarse es aquel donde $\min\{S_j(x), j=1, 2, \dots, k\} = S_{i_1} = \dots = S_{i_r}$ (es decir hay varias $S_j(x)$ que son mínimos de $\{S_1(x), S_2(x), \dots, S_k(x)\}$). La pregunta que haríamos ahora sería: ¿A qué población debemos asignar

el vector discriminante x ?, la respuesta que damos es la siguiente: tomaremos el subconjunto de $\{\lambda_j(x), j=1, \dots, k\}$ tal que

$$\lambda_{i_1}(x), \dots, \lambda_{i_r}(x) > 0 \text{ y } \lambda_{i_1} + \dots + \lambda_{i_r} = 1.$$

Podemos dentro de este subconjunto dar el valor mayor a la $\lambda_j(x)$ que queramos y forzar a las restantes con cantidades pequeñas para que la suma sea 1, supongamos que hemos elegido aquella cuyo índice es r . -- Considerando que $\lambda_r(x)$ es la mayor probabilidad de clasificar a un individuo dentro de la población r dado su vector discriminante, asignaremos a la observación x como proveniente de la población G_r . En resumen diremos que un individuo con un vector discriminante x puede clasificarse dentro de cualquier población G_j donde j es cualquier índice de $S_{i_1} = \dots = S_{i_r} = \text{Min}\{S_j(x), i=1, \dots, k\}$.

2.4 Solución de un problema de clasificación cuando las probabilidades a priori son desconocidas.

Hasta ahora nos hemos concretado a tratar el caso en que las probabilidades a priori ($q_i, i=1, \dots, k$) son conocidas, este tipo de situaciones podría decirse que son muy ideales y que no siempre se presentan, por ello debemos considerar el problema en el cual $q_i, i=1, \dots, k$ son poco relevantes o desconocidas.

Durante esta sección desarrollaremos un procedimiento para resolver el problema.

Definición. $P(j/i, R)$ es la probabilidad de mala clasificación de un individuo dentro de la población G_j dado que proviene de G_i obtenida por medio de un procedimiento o regla de decisión R .

Sea G_1, G_2, \dots, G_k, k poblaciones que se distribuyen con funciones de densidad $f_1(x), f_2(x), \dots, f_k(x)$ respectivamente entonces la pérdida esperada asociada a la aplicación de un procedimiento R de la clasificación de un individuo que proviene de la población G_i es:

$$C(i,R) = \sum_{\substack{j=1 \\ j \neq i}}^k C(j/i)P(j/i,R) = \sum_{j=1}^k C(j/i)P(j/i,R) \quad i=1, \dots, k$$

La igualdad de las dos sumatorias se sigue del hecho ya antes mencionado de que $C(i/i) = 0 \quad \forall i=1, 2, \dots, k$. Tendremos k valores -- $C(1,R), C(2,R), \dots, C(k,R)$ para cada procedimiento R , y en base a ellos podemos comparar dos procedimientos o reglas de decisión R y R^* , para ello utilizaremos las siguientes definiciones.

DEFINICION. Un procedimiento R es al menos tan bueno como R^* si $C(i,R) \leq C(i,R^*) \quad i=1, 2, \dots, k$.

DEFINICION. Un procedimiento R es mejor que R^* si $C(i,R) < C(i,R^*) \quad i=1, 2, \dots, k$, donde al menos una de las desigualdades es estricta.

DEFINICION. Dos procedimientos R y R^* son equivalentes si $C(i,R) = C(i,R^*) \quad i=1, 2, \dots, k$.

DEFINICION. Un procedimiento R es admisible si no existe un procedimiento R^* mejor que R .

DEFINICION. Dos procedimientos R y R^* no son comparables si $C(i,R) \leq C(i,R^*)$ para r índices y para los restantes se tienen que $C(i,R) \geq C(i,R^*)$.

DEFINICION. Una clase de procedimientos es completa si para cada procedimiento R fuera de la clase hay un procedimiento R^* dentro de la clase que es mejor.

La solución del problema con probabilidad a priori desconocidas se encuentra en la búsqueda de reglas de decisión admisibles, es decir un conjunto de procedimientos que sean los mejores en el sentido de su definición. Si la clase de reglas admisibles solo -

tienen un elemento éste es el mejor y es el que debe elegirse, pero si es vasto debemos encontrar una clase de reglas de decisión completa. Una clase completa que no contiene una subclase que sea también completa se le llama completa mínima⁽¹⁾ sobre la cual de acuerdo con algún criterio se elige determinado procedimiento.

Un principio usado como criterio de elección entre los procedimientos de una clase mínima completa es el minimax que generalmente conduce a un procedimiento único.

DEFINICION. Un procedimiento minimax es una regla de decisión R tal que $\max_{i=1,2,\dots,k} \{C(i,R)\} = \min_{d \in \Delta} \max_{i=1,2,\dots,k} C(i,R)$

donde Δ = espacio de reglas admisibles.

DEFINICION. R es un procedimiento de Bayes con respecto a las probabilidades a priori q_1, q_2, \dots, q_k si:

$$\sum_{i=1}^k q_i C(i,R) \leq \sum_{i=1}^k q_i C(i,R^*)$$

donde R^* es cualquier otro procedimiento. Los procedimientos de Bayes juegan un papel importante en las soluciones por medio de la teoría de decisiones, ya que puede ser demostrado⁽²⁾ que:

- a) Los procedimientos de Bayes son admisibles con respecto a q_1, q_2, \dots, q_k , con $q_i > 0$.

(1) Para una información más amplia acerca de las clases mínimas completas y sus características consultar Introducción al Análisis Multivariado (Anderson).

(2) Consultar Introducción al Análisis Multivariado (Anderson) para una información más amplia acerca de los procedimientos de Bayes y demostraciones concernientes a las proposiciones mencionadas.

- b) Cada procedimiento admisible es un procedimiento de Bayes con respecto a q_1, q_2, \dots, q_k .
- c) La clase de procedimientos de Bayes forman una clase completa.

2.5 DESVENTAJAS DE LAS SOLUCIONES POR MEDIO DE LA TEORIA DE DECISIONES.

Entre los principales tropiezos con que se encuentran este tipo de procedimientos podemos mencionar:

- a) La dificultad para establecer los costos de mala clasificación.
- b) Es difícil el conocimiento profundo de las poblaciones y por ello sus funciones de densidad no están completamente determinadas.

Las desventajas anteriores han motivado la búsqueda de diferentes soluciones al problema.

C A P I T U L O 3

SOLUCIONES CLASICAS AL PROBLEMA DE LA DISCRIMINACION
PARA DOS POBLACIONES.

Por medio de este tipo de soluciones se determina a cual población de k distintas posibles $G_i(x)$ con función de densidad $f_i(x)$ para $i=1,2,\dots,k$, proviene un individuo cuyo vector discriminante es $x=(x_1, x_2, \dots, x_p)'$. El procedimiento es obtener una partición del espacio de observaciones Ω en k regiones ajenas R_1, R_2, \dots, R_k . Si $x \in R_j$, entonces x se clasifica dentro la población G_j , y se asume que $x \sim f_j(x)$. La mayor parte de las veces se asume que $f_i(x)$ es la densidad Normal Multivariada ⁽¹⁾, aunque puede usarse cualquier otra.

Haciendo uso de la distribución Normal Multivariada se considerarán dos casos:

- a) Caso de poblaciones con Homocedasticidad.
- b) Caso de poblaciones con Heterocedasticidad.

3.1 DISCRIMINACION EN DOS POBLACIONES CON HOMOCEDASTICIDAD.

3.1.1 LA FUNCION LINEAL DE FISHER.

En esta sección derivaremos la función lineal de Fisher, y más adelante veremos como podemos obtener funciones lineales proporcionales a ella por medio de otros procedimientos.

(1) Esto es debido al importante papel que juega esta distribución dentro del Análisis Multivariado. Para efectos prácticos es la más usada. Si se desea información referente a otras densidades consultar Análisis Discriminante (Aranda & Ignacio Méndez).

Sean G_1 y G_2 , dos poblaciones cuyas funciones de densidad son $N(\mu(1), \Sigma)$ y $N(\mu(2), \Sigma)$, es decir su vector de medias es distinto, pero su matriz de covarianzas Σ es igual en ambas poblaciones. Se busca una función lineal $\lambda'x$ y en base a ella se clasifica el vector discriminante $x=(x_1, x_2, \dots, x_p)'$ dentro de alguna de las dos poblaciones.

$$\text{Sea: } \lambda' = (\lambda_1, \lambda_2, \dots, \lambda_p)$$

$$d = \lambda'(x(1) - x(2))$$

$$\text{donde } x(1) \sim N(\mu(1), \Sigma) \quad \text{y} \quad x(2) \sim N(\mu(2), \Sigma)$$

d es considerada por Fisher como un indicador de la distancia existente entre las dos poblaciones G_1 y G_2 , de esta manera definida la distancia esperada es:

$$E(d) = E(\lambda'(x(1) - x(2))) = \lambda'(\mu(1) - \mu(2))$$

se desea encontrar el vector óptimo λ' en el sentido que maximice $E(d)$.

Nótese que si λ'_0 es una solución entonces $c\lambda'_0$ también lo será, es decir $E(d)$ no es acotada, para que lo sea se introduce una restricción en la varianza de d :

$$\text{Var}(d) = 2 \lambda' \Sigma \lambda$$

y la restricción a la que se sujeta es:

$$\lambda' \Sigma \lambda = 1$$

Resolveremos el problema utilizando multiplicadores de Lagrange, pero por conveniencia la función a maximizar no será $E(d)$ sino $(E(d))^2$ entonces ⁽¹⁾ :

$$L = \lambda' \delta \delta \lambda - \eta (\lambda' \Sigma \lambda - 1) \quad \text{donde } \delta = (\mu(1) - \mu(2))$$

$$\frac{\partial L}{\partial \lambda} = 2 \delta \delta' \lambda - 2\eta \Sigma \lambda$$

(1) Esto es a causa de simplificar las cosas no teniendo que tomar en cuenta el signo de $E(d)$.

$$\frac{\partial L}{\partial \eta} = (\lambda' \Sigma \lambda - 1)$$

igualando ambas derivadas a cero formamos el sistema de ecuaciones

$$(1) \delta \delta' \lambda - \eta \Sigma \lambda = 0$$

$$(2) \lambda' \Sigma \lambda - 1 = 0$$

resolviendo tenemos:

$$\begin{aligned} \text{de (1)} \quad \lambda' (\delta \delta' \lambda - \eta \Sigma \lambda) &= 0 \\ \lambda' \delta \delta' \lambda &= \eta \end{aligned}$$

pero también de (1)

$$\delta \delta' \lambda - \eta \Sigma \lambda = (\delta \delta' - \eta \Sigma) \lambda = 0$$

este es un sistema de ecuaciones lineales que tiene solución distinta de la trivial $\Leftrightarrow |\delta \delta' - \eta \Sigma| = 0$, de donde puede verse que η es una raíz característica de $\delta \delta'$ en la métrica de Σ y λ es el vector característico asociado. Se escoge la mayor raíz característica que es la que maximiza L y se obtiene su λ correspondiente.

$$\text{Sea} \quad \tau = \delta' \lambda$$

$$\text{de (1)} \quad \delta \delta' \lambda = \eta \Sigma \lambda \quad \delta \tau = \eta \Sigma \lambda$$

$$\delta = \frac{\eta}{\tau} \Sigma \lambda$$

$$\frac{\tau}{\eta} \delta \Sigma^{-1} \delta = \lambda \quad \Rightarrow \quad \lambda' = \frac{\tau}{\eta} \delta' \Sigma^{-1}$$

Multiplicando λ' por el vector discriminante x' obtenemos la función lineal de Fisher

$$\lambda' x = \left(\frac{\tau}{\eta} \delta' \Sigma^{-1} \right) x = \frac{\tau}{\eta} (\mu_{(1)} - \mu_{(2)})' \Sigma^{-1} x$$

Esta fue la forma en que Fisher introdujo el Análisis Discriminante.

3.1.2 OTROS PROCEDIMIENTOS PARA LA OBTENCION DE FUNCIONES LINEALES.

Existen otros caminos de encontrar funciones lineales del vector discriminante x , tales como la encontrada por Anderson mediante la regla no aleatorizada o por la aplicación del procedimiento de Wald⁽¹⁾. Cualquiera de ellos conduce a la misma función lineal, nosotros la derivaremos por la regla no aleatorizada.

Sean G_1, G_2 dos poblaciones normales multivariadas con funciones de densidad $N(\mu_{(1)}, \Sigma)$ y $N(\mu_{(2)}, \Sigma)$, y $x' = (x_1, x_2, \dots, x_p)$ un vector discriminante, entonces la regla -- de decisión mediante la regla no aleatorizada es:

$$x \text{ se clasifica en } R_1 \text{ si } S_1(x) \leq S_2(x)$$

$$x \text{ se clasifica en } R_2 \text{ si } S_1(x) > S_2(x)$$

donde $S_j(x) = \sum_{i=1}^2 q_i C(j/i) f_i(x)$ y $C(j/j) = 0$ para $j=1, 2$ --

así pues R_1 y R_2 se escogen de acuerdo a:

$$R_1: q_2 C(1/2) f_2(x) \leq q_1 C(2/1) f_1(x)$$

$$R_2: q_2 C(1/2) f_2(x) > q_1 C(2/1) f_1(x)$$

como $C(1/2)q_2$ y $q_1 C(2/1)$ son mayores que cero tenemos:

$$R_1: \frac{f_1(x)}{f_2(x)} \geq \frac{q_2 C(1/2)}{q_1 C(2/1)}$$

$$R_2: \frac{f_1(x)}{f_2(x)} < \frac{q_2 C(1/2)}{q_1 C(2/1)}$$

(1) Wald considera dos poblaciones normales multivariadas G_1 y G_2 y una observación $x' = (x_1, \dots, x_p)'$. La función lineal se obtiene a través de la aplicación del lema de Neyman-Pearson en la prueba de la hipótesis; $H_0: x$ proviene de G_1 v.s $H_a: x$ proviene de G_2 si se desea seguir paso por paso el procedimiento descrito anteriormente véase Análisis Discriminante (Aranda e Ignacio Méndez, IMAS. Comunicaciones Técnicas, Serie Naranja)

$$f_i(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu_{(i)})' \Sigma^{-1} (x - \mu_{(i)})\right\}$$

$$\begin{aligned} \frac{f_1(x)}{f_2(x)} &= \frac{\exp\left\{-\frac{1}{2}(x - \mu_{(1)})' \Sigma^{-1} (x - \mu_{(1)})\right\}}{\exp\left\{-\frac{1}{2}(x - \mu_{(2)})' \Sigma^{-1} (x - \mu_{(2)})\right\}} = \\ &= \exp\left\{x' \Sigma^{-1} (\mu_{(1)} - \mu_{(2)}) - \frac{1}{2} (\mu_{(1)} + \mu_{(2)})' \Sigma^{-1} (\mu_{(1)} - \mu_{(2)})\right\} \\ &= \dots\dots\dots(1) \end{aligned}$$

tomando logaritmos y efectuando algunas operaciones algebraicas en la expresión se tiene que las mejores regiones de clasificación son:

$$R_1 : x' \Sigma^{-1} (\mu_{(1)} - \mu_{(2)}) \geq \ln k + \frac{1}{2} (\mu_{(1)} + \mu_{(2)})' \Sigma^{-1} (\mu_{(1)} - \mu_{(2)})$$

$$R_2 : x' \Sigma^{-1} (\mu_{(1)} - \mu_{(2)}) < \ln k + \frac{1}{2} (\mu_{(1)} + \mu_{(2)})' \Sigma^{-1} (\mu_{(1)} - \mu_{(2)})$$

donde :

$$k = \frac{q_2 C(1/2)}{q_1 C(2/1)}$$

de esta manera hemos encontrado la función lineal

$$x' \Sigma^{-1} (\mu_{(1)} - \mu_{(2)})$$

la cual solo difiere de la obtenida por Fisher por una constante de proporcionalidad τ/η .

En caso de que las probabilidades a priori sean desconocidas se busca el punto c , en que la esperanza de pérdida producida por la mala clasificación de una observación sea la misma, para ello se utiliza la prueba de la hipótesis:

$$H_0 : x \in G_1 \quad \text{v.s} \quad H_a : x \in G_2$$

y a partir de ella encontramos:

$$P\left\{\frac{f_1(x)}{f_2(x)} < k' / H_0\right\} = P(2/1)$$

$$P \left\{ \frac{f_1(x)}{f_2(x)} > k' / H_a \right\} = P(1/2)$$

El cociente $f_1(x)/f_2(x)$ ya ha sido obtenido en (1) de donde:

$$P \left\{ x' \Sigma^{-1} (\mu(1) - \mu(2)) - \frac{1}{2} (\mu(1) + \mu(2))' \Sigma^{-1} (\mu(1) - \mu(2)) < \ln k' / H_0 \right\} = P(2/1)$$

$$P \left\{ x' \Sigma^{-1} (\mu(1) - \mu(2)) - \frac{1}{2} (\mu(1) + \mu(2))' \Sigma^{-1} (\mu(1) - \mu(2)) > \ln k' / H_a \right\} = P(1/2)$$

sea $c = \ln k'$

$$U = x' \Sigma^{-1} (\mu(1) - \mu(2)) - \frac{1}{2} (\mu(1) + \mu(2))' \Sigma^{-1} (\mu(1) - \mu(2))$$

buscaremos la distribución de U bajo ambas hipótesis. Si $x \sim N(\mu(1), \Sigma)$ entonces:

$$\begin{aligned} E(U) &= \mu(1)' \Sigma^{-1} (\mu(1) - \mu(2)) - \frac{1}{2} (\mu(1) + \mu(2))' \Sigma^{-1} (\mu(1) - \mu(2)) = \\ &= \frac{1}{2} (\mu(1) - \mu(2))' \Sigma^{-1} (\mu(1) - \mu(2)) \end{aligned}$$

$$\begin{aligned} \text{Var}(U) &= E \left\{ (U - E(U))(U - E(U))' \right\} = E \left\{ (\mu(1) - \mu(2))' \Sigma^{-1} (x - \mu(1))(x - \mu(1))' \Sigma^{-1} (\mu(1) - \mu(2)) \right\} = (\mu(1) - \mu(2))' \Sigma^{-1} (\mu(1) - \mu(2)) \end{aligned}$$

definamos: $\alpha = (\mu(1) - \mu(2))' \Sigma^{-1} (\mu(1) - \mu(2))$

por consiguiente U se distribuye como $N(\frac{1}{2} \alpha, \alpha)$.

Si $x \sim N(\mu(2), \Sigma)$ entonces:

$$E(U) = \frac{1}{2} (\mu(2) - \mu(1))' \Sigma^{-1} (\mu(1) - \mu(2)) = -\frac{1}{2} \alpha$$

$$\text{Var}(U) = (\mu(1) - \mu(2))' \Sigma^{-1} (\mu(1) - \mu(2)) = \alpha$$

de donde la distribución de U es $N(-\frac{1}{2} \alpha, \alpha)$

Las probabilidades de mala clasificación son:

$$P(2/1) = \int_{-\infty}^c \frac{1}{(2\pi\alpha)^{1/2}} \exp\left\{-\frac{1}{2}\left(u - \frac{1}{2}\alpha\right)^2\right\} du =$$

$$= \int_{-\infty}^{(c - \frac{1}{2}\alpha)/\alpha^{1/2}} \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}y^2} dy$$

$$P(1/2) = \int_c^{\infty} \frac{1}{(2\pi\alpha)^{1/2}} \exp\left\{-\frac{1}{2}\left(u + \frac{1}{2}\alpha\right)^2\right\} du =$$

$$= \int_{\frac{(c + \frac{1}{2}\alpha)}{\alpha^{1/2}}}^{\infty} \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}y^2} dy$$

y el punto c se elige en tal forma que $C(1/2)P(1/2) = C(2/1)P(2/1)$. Una vez determinado el punto c la regla de decisión es:

Clasificar a la observación como proveniente de G_1 si $U \geq c$ ó clasificar a la observación como proveniente de G_2 si $U < c$.

3.1.3 PROCEDIMIENTOS EMPLEADOS EN DOS POBLACIONES NORMALES CON PARAMETROS DESCONOCIDOS.

Los métodos que hemos utilizado hasta ahora en las poblaciones con homocedasticidad, suponen que sus distribuciones son conocidas. Si los parámetros que determinan la distribución no se encuentran completamente establecidos, se usan sus estimadores (generalmente aquellos que son insesgados y consistentes).

Supóngase que $x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}$ pertenecen a la población G_1 con función de densidad $N(\mu_{(1)}, \Sigma)$ y $x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)}$ provienen de G_2 cuya función de densidad es $N(\mu_{(2)}, \Sigma)$, donde los parámetros $\mu_{(1)}, \mu_{(2)}, \Sigma$ no son conocidos. Deseamos clasificar una observación x dentro de alguna de las dos poblaciones. En este caso los estimadores de los parámetros que pueden ser empleados son:

$$\hat{\mu}_{(1)} = \frac{\sum_{i=1}^{N_1} x_i^{(1)}}{N_1} \quad \hat{\mu}_{(2)} = \frac{\sum_{i=1}^{N_2} x_i^{(2)}}{N_2}$$

$$S = \sum_{i=1}^{N_1} (x_i^{(1)} - \hat{\mu}_{(1)})(x_i^{(1)} - \hat{\mu}_{(1)})' + \sum_{i=1}^{N_2} (x_i^{(2)} - \hat{\mu}_{(2)})(x_i^{(2)} - \hat{\mu}_{(2)})'$$

$$\hat{\Sigma} = \frac{1}{N_1 + N_2 - 2} S$$

Sustituyendo el valor de Σ en U definida en la sección 3.1.2 :

$$U = x' \hat{\Sigma}^{-1} (\hat{\mu}_{(1)} - \hat{\mu}_{(2)}) - \frac{1}{2} (\hat{\mu}_{(1)} + \hat{\mu}_{(2)})' \hat{\Sigma}^{-1} (\hat{\mu}_{(1)} - \hat{\mu}_{(2)})$$

U debe usarse para llevar a cabo la clasificación de igual manera que se hacía en los casos en que las probabilidades a priori son conocidas y cuando no lo son. Aunque el uso de U en la misma forma cuando los parámetros son conocidos no se justifica, se espera que se obtengan buenos resultados. Al valor:

$$(\hat{\mu}_{(1)} - \hat{\mu}_{(2)})' \hat{\Sigma}^{-1} (\hat{\mu}_{(1)} - \hat{\mu}_{(2)})$$

se le llama distancia generalizada de Mahalanobis.

3.1.4 SOLUCION DEL PROBLEMA DE CLASIFICACION POR MEDIO DEL COCIENTE DE VEROSIMILITUDES.

Sean $x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}$, N_1 observaciones que provienen de la población G_1 , y $x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)}$, N_2 observaciones de la población G_2 . Las funciones de densidad de las poblaciones G_1 y G_2 son $N(\mu_{(1)}, \Sigma)$ y $N(\mu_{(2)}, \Sigma)$ respectivamente, los parámetros $\mu_{(1)}, \mu_{(2)}, \Sigma$ no son conocidos. Considérese una observación x la cual se desea clasificar dentro de alguna de las poblaciones G_1 o G_2 , la solución se encuentra planteando el problema por medio de la prueba de la hipótesis siguiente:

$$H_0 : \begin{array}{l} x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}, x \\ x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)} \end{array} \begin{array}{l} \text{han sido extraídas de } G_1 \text{ y} \\ \text{pertenecen a } G_2. \end{array}$$

contra la hipótesis alternativa:

$$H_a : \begin{array}{l} x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)} \\ x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)}, x \end{array} \begin{array}{l} \text{han sido extraídas de } G_1 \text{ y} \\ \text{han sido extraídas de } G_2. \end{array}$$

Tomando la hipótesis H_0 tenemos que los estimadores de máxima verosimilitud de $\mu_{(1)}, \mu_{(2)}, \Sigma$ son:

$$\hat{\mu}_{(1)} = \left\{ x + \sum_{i=1}^{N_1} x_i^{(1)} \right\} \frac{1}{N_1 + 1} = \frac{N_1 \bar{x}^{(1)} + x}{N_1 + 1}$$

$$\hat{\mu}_{(2)} = \frac{\sum_{i=1}^{N_2} x_i^{(2)}}{N_2} = \bar{x}^{(2)}$$

$$\hat{\Sigma}_{H_0} = \left\{ \sum_{i=1}^{N_1} (x_i^{(1)} - \hat{\mu}_{(1)})(x_i^{(1)} - \hat{\mu}_{(1)})' + (x - \hat{\mu}_{(1)})(x - \hat{\mu}_{(1)})' + \right.$$

$$+ \sum_{i=1}^{N_2} (x_i^{(2)} - \hat{\mu}(2))(x_i^{(2)} - \hat{\mu}(2))' \left\{ \frac{1}{N_1 + N_2 + 1} \right.$$

vamos a escribir $\hat{\Sigma}_{H_0}$ en otra forma, pero para ello utilizaremos los siguientes resultados obtenidos a través de simples cálculos algebraicos.

$$\sum_{i=1}^{N_1} (x_i^{(1)} - \hat{\mu}(1))(x_i^{(1)} - \hat{\mu}(1))' = \sum_{i=1}^{N_1} (x_i^{(1)} - \bar{x}^{(1)})(x_i^{(1)} - \bar{x}^{(1)})' + N_1(\bar{x}^{(1)} - \hat{\mu}(1))(\bar{x}^{(1)} - \hat{\mu}(1))' \dots\dots\dots(1)$$

$$N_1(\bar{x}^{(1)} - \hat{\mu}(1))(\bar{x}^{(1)} - \hat{\mu}(1))' + (x - \hat{\mu}(1))(x - \hat{\mu}(1))' = \frac{N_1}{N_1 + 1} (x - \bar{x}^{(1)})(x - \bar{x}^{(1)})' \dots\dots\dots(2)$$

$$\sum_{i=1}^{N_2} (x_i^{(2)} - \hat{\mu}(2))(x_i^{(2)} - \hat{\mu}(2))' = \sum_{i=1}^{N_2} (x_i^{(2)} - \bar{x}^{(2)})(x_i^{(2)} - \bar{x}^{(2)})' \dots\dots(3)$$

de (1), (2) y (3) resulta :

$$\hat{\Sigma}_{H_0} = \left\{ \frac{1}{N_1 + N_2 + 1} \right\} \left\{ \sum_{j=1}^2 \sum_{i=1}^{N_j} (x_i^{(j)} - \bar{x}^{(j)})(x_i^{(j)} - \bar{x}^{(j)})' + \frac{1}{N_1 + 1} (x - \bar{x}^{(1)})(x - \bar{x}^{(1)})' \right\} = \left\{ C + \frac{N_1}{N_1 + 1} (x - \bar{x}^{(1)})(x - \bar{x}^{(1)})' \right\} \left\{ \frac{1}{N_1 + N_2 + 1} \right.$$

donde :

$$C = \sum_{j=1}^2 \sum_{i=1}^{N_j} (x_i^{(j)} - \bar{x}^{(j)})(x_i^{(j)} - \bar{x}^{(j)})'$$

tomando la hipótesis alternativa, encontramos en forma similar:

$$\hat{\mu}_{(1)} = \bar{x}^{(1)} \quad \hat{\mu}_{(2)} = \frac{1}{N_1 + 1} (N_1 \bar{x}^{(1)} + x)$$

$$\hat{\Sigma}_{H_a} = \frac{1}{N_1 + N_2 + 1} \left\{ C + \frac{N_2}{N_2 + 1} (x - \bar{x}^{(2)})(x - \bar{x}^{(2)})' \right\}$$

entonces:

$$\begin{aligned} \max_{H_0} L(x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}, x, x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)}, \mu_{(1)}, \mu_{(2)}, \Sigma) = \\ = (2\pi)^{-\frac{1}{2}p(N_1+N_2+1)} |\hat{\Sigma}_{H_0}|^{-\frac{1}{2}(N_1+N_2+1)} e^{-\frac{1}{2}p(N_1+N_2+1)} \end{aligned}$$

$$\begin{aligned} \max_{H_a} L(x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}, x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)}, x, \mu_{(1)}, \mu_{(2)}, \Sigma) = \\ = (2\pi)^{-\frac{1}{2}p(N_1+N_2+1)} |\hat{\Sigma}_{H_a}|^{-\frac{1}{2}(N_1+N_2+1)} e^{-\frac{1}{2}p(N_1+N_2+1)} \end{aligned}$$

$$\lambda = \frac{\max_{H_0} L(x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}, x, x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)}, \mu_{(1)}, \mu_{(2)}, \Sigma)}{\max_{H_a} L(x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}, x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)}, x, \mu_{(1)}, \mu_{(2)}, \Sigma)}$$

$$= \frac{|\hat{\Sigma}_{H_0}|^{-\frac{1}{2}(N_1+N_2+1)}}{|\hat{\Sigma}_{H_a}|^{-\frac{1}{2}(N_1+N_2+1)}} = \frac{\left| C + \frac{N_2}{N_2+1} (x - \bar{x}^{(2)})(x - \bar{x}^{(2)})' \right|^t}{\left| C + \frac{N_1}{N_2+1} (x - \bar{x}^{(1)})(x - \bar{x}^{(1)})' \right|^t}$$

$$= \frac{\left| 1 + \frac{N_2}{N_2+1} (x - \bar{x}^{(2)})' C^{-1} (x - \bar{x}^{(2)}) \right|^t}{\left| 1 + \frac{N_1}{N_2+1} (x - \bar{x}^{(1)})' C^{-1} (x - \bar{x}^{(1)}) \right|^t} \quad \text{donde: } t = \frac{1}{2}(N_1+N_2+1)$$

las mejores regiones de clasificación son:

$$R_1 : \lambda \geq k$$

$$R_2 : \lambda < k$$

la desventaja de este procedimiento es no conocer la distribución de λ .

3.1.5 APLICACION DEL ANALISIS DE REGRESION A LA DISCRIMINACION.

Sea G_1 y G_2 dos poblaciones con funciones de densidad $N(\mu(1), \Sigma)$ y $N(\mu(2), \Sigma)$. Supongamos que $x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}$, son N_1 observaciones extraídas de G_1 y $x_{N_1+1}^{(2)}, \dots, x_{N_1+N_2}^{(2)}$ son N_2 observaciones de G_2 . Se desea clasificar en G_1 o en G_2 a un individuo cuyo vector discriminante es $x = (x_1, x_2, \dots, x_p)'$.

Por medio de este procedimiento, obtendremos una función lineal utilizando la analogía existente entre la Discriminación y el Análisis de Regresión.

Sea:

$$\bar{x}^{(1)} = \frac{\sum_{i=1}^{N_1} x_i^{(1)}}{N_1} \quad \bar{x}^{(2)} = \frac{\sum_{i=1}^{N_2} x_{N_1+i}^{(2)}}{N_2}$$

$$\bar{x} = \frac{N_1 \bar{x}^{(1)} + N_2 \bar{x}^{(2)}}{N_1 + N_2}$$

y definamos la variable artificial Y , tal que:

$$Y = \begin{cases} \lambda_1 & \text{si } x \in G_1 \\ \lambda_2 & \text{si } x \in G_2 \end{cases}$$

donde λ_1 y λ_2 son constantes.

Si consideramos que el vector discriminante x cumple con la suposición:

$$E(x) = \alpha + \beta Y \quad \dots \dots \dots (1)$$

donde :

$$\alpha = \frac{1}{\lambda_1 - \lambda_2} (\lambda_1 \mu_2 - \lambda_2 \mu_1)$$

$$\beta = \frac{1}{\lambda_1 - \lambda_2} (\mu_1 - \mu_2)$$

entonces si el individuo x pertenece a la población G_1 , $E(x)=\mu_1$. Si x proviene de G_2 , $E(x)=\mu_2$. La ecuación (1) es por lo tanto la regresión de x en Y , sin embargo nuestro problema, es completamente al contrario, tenemos un vector discriminante y no conocemos si pertenece a la población 1 ó a la 2, por ello deseamos encontrar el valor de Y por medio de su regresión inversa en base a las mediciones $x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}, x_{N_1+1}^{(2)}, \dots, x_{N_1+N_2}^{(2)}$, aunque no nos estamos basando estrictamente en las hipótesis necesarias sobre las cuales descansa la regresión. Vamos a usar el modelo lineal general de regresión $Y=x*\beta*+u$. Supongamos que $E(Y)=x*\beta*$

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{N_1} \\ Y_{N_1+1} \\ \vdots \\ Y_{N_1+N_2} \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_1 \\ \vdots \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_2 \end{pmatrix} \quad x^* = \begin{pmatrix} 1 & x_1^{(1)} \\ 1 & x_2^{(1)} \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & x_{N_1}^{(1)} \\ \vdots & x_{N_1+1}^{(2)} \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & x_{N_1+N_2}^{(2)} \\ 1 & \dots \dots \dots x_{N_1+N_2}^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & x_{11}^{(1)} & x_{12}^{(1)} & \dots & x_{1p}^{(1)} \\ 1 & x_{21}^{(1)} & x_{22}^{(1)} & \dots & x_{2p}^{(1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \dots & \dots & \dots & x_{N_1+N_2}^{(2)} \end{pmatrix}$$

$$e = (1, 1, \dots, 1)^t \quad x^* = (e, x^t) \quad x = (x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}, x_{N_1+1}^{(2)}, \dots, x_{N_1+N_2}^{(2)})$$

$$\beta^* = (\beta_0, \beta_1, \dots, \beta_p)^t = \begin{pmatrix} \beta_0 \\ \beta \end{pmatrix}$$

β^* es un vector de constantes desconocidas, el cual se estima por medio de mínimos cuadrados y la solución correspondiente es:

$$\beta^* = (x^{*t} x^*)^{-1} x^{*t} Y$$

Tos valores de λ_1 y λ_2 es de acuerdo al siguiente producto:

$$x^{*t} Y = \begin{pmatrix} N_1 \lambda_1 + N_2 \lambda_2 \\ \lambda_1 \bar{x}^{(1)} + \lambda_2 \bar{x}^{(2)} \end{pmatrix}$$

con el fin de simplificar los cálculos se eligen λ_1 y λ_2 de tal forma que $N_1\lambda_1 + N_2\lambda_2 = 0$. Se ha observado que los mejores resultados se obtienen cuando los valores de λ_1 y λ_2 son:

$$\lambda_1 = \frac{N_2}{N_1 + N_2} \quad \lambda_2 = -\frac{N_1}{N_1 + N_2}$$

las ecuaciones normales de regresión son:

$$(x^{*'}x^*)\beta^* = x^{*'}Y$$

entonces:

$$(x^{*'}x^*)\beta^* = \begin{pmatrix} N_1\lambda_1 + N_2\lambda_2 \\ \frac{N_1N_2(\bar{x}^{(1)} - \bar{x}^{(2)})}{N_1 + N_2} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{N_1N_2(\bar{x}^{(1)} - \bar{x}^{(2)})}{N_1 + N_2} \end{pmatrix} \dots\dots\dots(1)$$

$$(x^{*'}x^*)\beta^* = \begin{pmatrix} e'e\beta_0 + e'x'\beta \\ xe\beta_0 + xx'\beta \end{pmatrix} \dots\dots\dots(2)$$

de (1) y de (2) se tiene

$$e'e\beta_0 + e'x'\beta_0 = 0 \quad \text{entonces} \quad \beta_0 = -\frac{e'x'\beta}{e'e}$$

$$xe\beta_0 + xx'\beta = \frac{N_1N_2}{N_1 + N_2} (\bar{x}^{(1)} - \bar{x}^{(2)})$$

$$\frac{xe(-e'x'\beta)}{e'e} + xx'\beta = \frac{N_1N_2(\bar{x}^{(1)} - \bar{x}^{(2)})}{N_1 + N_2}$$

$$x(I - \frac{1}{e'e} ee')x'\beta = \frac{N_1N_2}{N_1 + N_2} (\bar{x}^{(1)} - \bar{x}^{(2)})$$

Sea:

$e_{N_1} = (1, 1, \dots, 1)'$ un vector de N_1 componentes.

$e_{N_2} = (1, 1, \dots, 1)'$ un vector de N_2 componentes.

$$x_1 = (x_1^{(1)}, x_2^{(1)}, \dots, x_{N_1}^{(1)}) \quad x_2 = (x_1^{(2)}, x_2^{(2)}, \dots, x_{N_2}^{(2)})$$

$$x = (x_1 \mid x_2)$$

$$x(I - \frac{1}{e'e} ee')x' = (x_1 \mid x_2)(I - \frac{ee'}{N_1+N_2}) \begin{pmatrix} x_1' \\ \vdots \\ x_2' \end{pmatrix} =$$

$$= (x_1 \mid x_2) \begin{pmatrix} x_1' \\ \vdots \\ x_2' \end{pmatrix} - \frac{1}{N_1 + N_2} (x_1 \mid x_2) \begin{pmatrix} e_{N_1} \\ \vdots \\ e_{N_2} \end{pmatrix} (e_{N_1}' \mid e_{N_2}') \begin{pmatrix} x_1' \\ \vdots \\ x_2' \end{pmatrix} =$$

$$= x_1(I_{N_1} - \frac{1}{N_1} e_{N_1} e_{N_1}')x_1' + x_2(I_{N_2} - \frac{1}{N_2} e_{N_2} e_{N_2}')x_2' + \frac{N_1 N_2}{N_1+N_2}(\bar{x}^1 - \bar{x}^2)(\bar{x}^1 - \bar{x}^2)'$$

de donde:

$$x(I - \frac{1}{e'e} ee')x'\beta = \{ x_1(I_{N_1} - \frac{1}{N_1} e_{N_1} e_{N_1}')x_1' + x_2(I_{N_2} - \frac{1}{N_2} e_{N_2} e_{N_2}')x_2' + \frac{N_1 N_2}{N_1+N_2}(\bar{x}^{(1)} - \bar{x}^{(2)})(\bar{x}^{(1)} - \bar{x}^{(2)})' \} \beta = \frac{N_1 N_2}{N_1+N_2}(\bar{x}^1 - \bar{x}^2) \dots \dots \dots (3)$$

si se define:

$$S_0 = x_1(I_{N_1} - \frac{1}{N_1} e_{N_1} e_{N_1}')x_1' + x_2(I_{N_2} - \frac{1}{N_2} e_{N_2} e_{N_2}')x_2'$$

$$c^2 = \frac{N_1 N_2}{N_1+N_2} \quad d = (\bar{x}^{(1)} - \bar{x}^{(2)})$$

la ecuación (3) puede entonces escribirse en la forma:

$$(S_0 + c^2 d d') = \frac{N_1 N_2}{N_1+N_2} d$$

la solución del vector β puede escribirse también en la forma (1)

$$\beta = \frac{c^2}{1 + c^2 d' S_0^{-1} d} S_0^{-1} d \dots \dots \dots (4)$$

el fin de encontrar una expresión de este estilo es mostrar que

(1) Esta expresión puede ser encontrada mediante el resultado de la pag. 23 de Análisis multivariado de Press.

la función discriminante de Fisher es proporcional a la función lineal encontrada por medio de regresión. Puede demostrarse algebraicamente que:

$$S_0 = (N_1 + N_2 - 2)S = \sum_{i=1}^2 \sum_{j=1}^{N_i} (x_j^{(i)} - \bar{x}^{(i)})(x_j^{(i)} - \bar{x}^{(i)})'$$

si la matriz de covarianzas Σ es desconocida entonces su estimador es S , es decir $\hat{\Sigma} = S$, y la función lineal de Fisher está dada por:

$$\alpha'x = \frac{1}{\lambda} (\mu(1) - \mu(2))' \Sigma^{-1} x = \frac{1}{\lambda} (\bar{x}^{(1)} - \bar{x}^{(2)})' S^{-1} x = \frac{1}{\lambda} d' S^{-1} x$$

multiplicando (4) por x'

$$x' \beta = \beta' x = \frac{c^2}{1 + c^2 d' S_0^{-1} d} d' \{(N_1 + N_2 - 2)S\}^{-1} x$$

de donde se ve que ambas funciones son proporcionales, por consiguiente ambos enfoques gufan al mismo procedimiento de discriminación.

3.2 DISCRIMINACION EN POBLACIONES CON HETEROCEDASTICIDAD.

No profundizaremos en el tratamiento de este caso, sólo mencionaremos los lineamientos generales que se proponen para su solución.

Considérese dos poblaciones Normales Multivariadas G_1 y G_2 cuyas funciones de densidad son $N(\mu(1), \Sigma(1))$ y $N(\mu(2), \Sigma(2))$ donde $\mu(1) \neq \mu(2)$ $\Sigma(1) \neq \Sigma(2)$ ($\Sigma(1)$ y $\Sigma(2)$ son las matrices de covarianzas no singulares). El procedimiento consiste en buscar el procedimiento lineal óptimo $\beta'x$ donde $\beta = (\beta_1, \dots, \beta_p)'$ $\neq (0, 0, \dots, 0)'$ y $x = (x_1, x_2, \dots, x_p)'$ es el vector discriminante del individuo que se desea clasificar dentro de alguna

de las poblaciones G_1 ó G_2 . La regla de decisión es⁽¹⁾:

clasificar a x dentro de G_1 si $\beta'x \leq c$

clasificar a x dentro de G_2 si $\beta'x > c$

de acuerdo con lo expuesto anteriormente las probabilidades de mala clasificación son:

$$P(1/2) = P(\beta'x \leq c / x \in G_2)$$

$$P(2/1) = P(\beta'x > c / x \in G_1)$$

entonces los problemas que pueden presentarse son:

- a) Minimizar la probabilidad de mala clasificación cuando la otra se especifica.
- b) Minimizar la máxima probabilidad de mala clasificación.
- c) Minimizar las probabilidades de mala clasificación cuando las probabilidades a priori de pertenecer a G_1 ó G_2 son conocidas.

Para cada uno de los casos debe encontrarse un procedimiento que lo solucione de acuerdo con el planteamiento que se especifique. Las soluciones de dichos problemas se encuentran dentro del conjunto de procedimientos admisibles.

(1) Enfoque propuesto por Anderson y Bahadur.

CAPITULO 4 .

DISCRIMINACION EN K POBLACIONES.

4.1 DISCRIMINACION EN K POBLACIONES NORMALES CON HOMOCE-
DASTICIDAD.

El tratamiento de un problema de discriminación con k poblaciones, es simplemente una generalización de los procedimientos aplicados en capítulos anteriores a los casos en que sólo se consideraban dos poblaciones Normales con Homocedasticidad, por ello se aplicará la generalización de esos métodos y esto resultará relativamente fácil.

4.1.1 APLICACION DE LA TEORIA DE DECISIONES AL PROBLEMA DE
DISCRIMINACION DE K POBLACIONES NORMALES CON HOMOCE-
DASTICIDAD.

Sea G_1, G_2, \dots, G_k , k poblaciones Normales Multivariadas, entonces la función de densidad de cada una de ellas está dada por:

$$f_i(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_{(i)})' \Sigma^{-1} (x-\mu_{(i)})}$$

es decir que sus vectores de medias son distintos, pero sus matrices de covarianzas son iguales, Supóngase que q_i es la probabilidad a priori de extraer un individuo de la población G_i para $i=1,2,\dots,k$ y $C(j/i)$ el costo de mala clasificación de un individuo dentro de la población G_j dado que pertenece a G_i . Aplicando cualquiera de los procedimientos desarrollados en el capítulo I (Regla Aleatorizada, Regla no aleatorizada o la solución propuesta por Anderson), no importa cual de ellos se e-

liza, ya que todos ellos conducen a la misma solución, se encuentran las regiones de decisión:

$$R_i : x \in R_i \quad \text{si } S_i(x) \leq S_j(x), \quad j=1,2,\dots,k, \quad i \neq j$$

donde

$$S_j(x) = \sum_{r=1}^k q_r C(j/r) f_r(x)$$

en otras palabras el individuo que tiene el vector discriminante x se asigna a la región R que tiene el mismo índice en el cual la función $S_i(x)$ es mínima. Con el objeto de simplificar los cálculos también supondremos que:

$$C(j/i) = c \quad i, j = 1, 2, \dots, k, \quad i \neq j$$

y desde luego que:

$$C(i/i) = 0 \quad i = 1, 2, \dots, k$$

tomense los índices tales que:

$$\sum_{\substack{r=1 \\ r \neq i}}^k q_r c f_r(x) < \sum_{\substack{r=1 \\ r \neq j}}^k q_r c f_r(x) \quad j=1, \dots, k, \\ i \neq j$$

desde que el lado izquierdo de la desigualdad es un mínimo

$$\sum_{\substack{r=1 \\ r \neq i}}^k q_r f_r(x) c - \sum_{\substack{r=1 \\ r \neq i, j}}^k q_r f_r(x) c <$$

$$< \sum_{\substack{r=1 \\ r \neq j}}^k q_r f_r(x) c - \sum_{\substack{r=1 \\ r \neq i, j}}^k q_r f_r(x) c$$

$$q_j f_j(x) < q_i f_i(x)$$

en la misma forma repitiendo el procedimiento para cada $S_j(x)$

mayor que $S_j(x)$, concluimos que x se asigna a R_i si i es el índice para el cual $q_i f_i(x)$ es un máximo.

$$\Rightarrow R_i : q_i f_i(x) > q_j f_j(x) \quad j=1,2,\dots,k, \quad i \neq j$$

$$R_i : \frac{f_i(x)}{f_j(x)} > \frac{q_j}{q_i} \quad j=1,2,\dots,k, \quad i \neq j$$

$$R_i : \ln \frac{f_i(x)}{f_j(x)} > \ln \frac{q_j}{q_i} \quad j=1,2,\dots,k, \quad i \neq j$$

defínase la función:

$$U_{ij} = \ln \frac{f_i(x)}{f_j(x)}$$

en nuestro caso $f_i(x)$ es una Normal Multivariada y por ello el valor de $U_{ij}(x)$ es:

$$\begin{aligned} U_{ij}(x) &= -\frac{1}{2} (x-\mu_{(i)})' \Sigma^{-1} (x-\mu_{(i)}) + \\ &+ \frac{1}{2} (x-\mu_{(j)})' \Sigma^{-1} (x-\mu_{(j)}) = \\ &= (x - \frac{1}{2}(\mu_{(i)} + \mu_{(j)}))' \Sigma^{-1} (\mu_{(i)} - \mu_{(j)}) \end{aligned}$$

resumiendo lo expuesto anteriormente, tenemos que las regiones de clasificación R_i están definidas por aquellas x que satisfacen :

$$R_i : U_{ij} > \ln \frac{q_j}{q_i} \quad j=1,2,\dots,k, \quad i \neq j$$

puede demostrarse fácilmente que $U_{ij} = -U_{ji}$

Nuevamente volvemos a encontrarnos con el caso en que las probabilidades a priori no son conocidas, entonces las regiones R_i están definidas en la forma:

$$R_i : U_{ij} > d_i - d_j \quad j=1,2,\dots,k, \quad i \neq j$$

Las constantes d_i , $i = 1,2,\dots,k$ se determinan por medio del procedimiento minimax de tal manera que todas las probabilidades de clasificación correcta $P(i/i)$ sean iguales.

Para evaluar la probabilidad de clasificación correcta necesitamos conocer la distribución de la variable aleatoria:

$$U_{ij} = (x - \frac{1}{2}(\mu(i) + \mu(j)))' \Sigma^{-1} (\mu(i) - \mu(j))$$

si la observación x se extrae de la población G_i entonces la distribución de U_{ij} es Normal Multivariada con los parámetros que estimaremos a continuación:

$$E(U_{ij}) = \frac{1}{2}(\mu(i) - \mu(j))' \Sigma^{-1} (\mu(i) - \mu(j))$$

$$\text{Var}(U_{ij}) = (\mu(i) - \mu(j))' \Sigma^{-1} (\mu(i) - \mu(j))$$

si definimos: $\alpha_{ij} = (\mu(i) - \mu(j))' \Sigma^{-1} (\mu(i) - \mu(j))$, la distribución de U_{ij} es:

$$N(\frac{1}{2}\alpha_{ij}, \alpha_{ij})$$

y la probabilidad de clasificar correctamente a un individuo es

$$P(i/i) = \int_{d_i - d_k}^{\infty} \dots \int_{d_i - d_{i+1}}^{\infty} \int_{d_i - d_{i-1}}^{\infty} \dots \int_{d_i - d_1}^{\infty} f_i du_{i1} \dots du_{ik}$$

donde $f_i(x)$ es la densidad de U_{ij} , $j=1,2,\dots,k$, $j \neq i$. Desde luego también puede presentarse la situación en la cual los parámetros son desconocidos, el problema se resuelve si se dispone de una muestra de cada población, de esta manera es posible estimar los parámetros y sustituirlos en U_{ij} . Si $x_1^{(i)}, x_2^{(i)}, \dots, x_{N_i}^{(i)}$ son observaciones extraídas de la población G_i , los estimadores

muestrales son:

$$\hat{\mu}(i) = \left(\sum_{\alpha=1}^{N_i} x_{\alpha}^{(i)} \right) \frac{1}{N_i} \quad \text{y} \quad S = \hat{\Sigma}$$

donde S está definida por:

$$\left(\sum_{i=1}^k N_i - k \right) S = \sum_{i=1}^k \sum_{\alpha=1}^k (x_{\alpha}^{(i)} - \bar{x}^{(i)})(x_{\alpha}^{(i)} - \bar{x}^{(i)})'$$

la distribución de U_{ij} son distintos para los casos en los cuales los parámetros son conocidos y cuando no lo son, pero si $N_i \rightarrow \infty$ las distribuciones se aproximan, por ello sólo puede usarse la sustitución de los parámetros estimados cuando el tamaño de la muestra sea suficientemente grande.

4.2 SOLUCION AL PROBLEMA DE DISCRIMINACION PARA K POBLACIONES MEDIANTE LA DERIVACION DE UN CONJUNTO DE FUNCIONES LINEALES.

Por medio de este procedimiento el problema de discriminación para k poblaciones se resuelve encontrando la óptima combinación lineal del vector discriminante x y en base a ella se clasifica dentro de alguna de las k poblaciones (procedimiento sugerido por Fisher).

Sean G_1, G_2, \dots, G_k , k poblaciones Normales Multivariadas cuyas funciones de densidad son $N(\mu_{(i)}, \Sigma)$ respectivamente, los parámetros $\mu_{(i)}, \Sigma$ no son conocidos, y $\lambda'x$ una función lineal del vector discriminante $x = (x_1, x_2, \dots, x_p)'$. Para determinar el vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)'$ se considera el análisis de varianza de la función lineal $\lambda'x$ (referente a la hipótesis de la igualdad de medias poblacionales), por esta razón, λ se escoge de tal forma que maximice la estadística F calculada. si es máxima se rechaza con mayor probabilidad la igualdad de medias y por

lo tanto la igualdad de grupos. La estadística F calculada está dada por:

$$F = \frac{\frac{\lambda' A \lambda}{k-1}}{\frac{\lambda' W \lambda}{\left(\sum_{i=1}^k N_i \right) - k}} = \frac{(\lambda' A \lambda) \left\{ \left(\sum_{i=1}^k N_i \right) - k \right\}}{\lambda' W \lambda (k-1)}$$

donde:

N_i es el número de individuos que pertenecen a la población G_i .

A es la matriz de suma de cuadrados y productos entre poblaciones.

W es la matriz de sumas de cuadrados y productos dentro de poblaciones.

consideremos las siguientes definiciones:

x_{hij} es la variable j -ésima del vector discriminante x del individuo i -ésimo, de la población h -ésima.

$$\bar{x}_{.j.} = \frac{\sum_{h=1}^k \sum_{i=1}^{N_h} x_{hij}}{\sum_{h=1}^k N_h}$$

es la media de la variable j para todas las observaciones.

$$\bar{x}_{hj.} = \frac{\sum_{i=1}^{N_h} x_{hji}}{N_h}$$

es la media de la variable j del vector x para la población h .

de acuerdo con las definiciones dadas anteriormente podemos escribir el elemento s, t de la matriz $A_{p \times p}$ y $W_{p \times p}$ de la siguiente

te manera:

$$W_{p \times p} = \sum_{h=1}^k \sum_{i=1}^{N_h} (x_{hsi} - \bar{x}_{hj.})(x_{hti} - \bar{x}_{ht.}) \quad s, t=1, 2, \dots, p$$

$$A_{p \times p} = \sum_{h=1}^k (\bar{x}_{hs.} - \bar{x}_{.s.})(\bar{x}_{ht.} - \bar{x}_{.t.}) \quad s, t=1, 2, \dots, p$$

maximizar la estadística F es equivalente a maximizar el cociente

$$\frac{\lambda' A \lambda}{\lambda' W \lambda}$$

y a su vez es equivalente a:

$$\max_{\lambda} \frac{\lambda' A \lambda}{\lambda' (W+A) \lambda}$$

esta última expresión es la que se utiliza solamente con el fin de simplificar los cálculos. Analizando este cociente nos damos cuenta de que si λ es un vector solución, también lo será el vector $c\lambda$, donde c es una constante arbitraria, en este sentido el problema no está completamente determinado, por ello se sujeta a la restricción:

$$\lambda' (W+A) \lambda = 1$$

usando multiplicadores de Lagrange tenemos:

$$L = \lambda' A \lambda - \alpha \{ \lambda' (W+A) - 1 \}$$

$$\frac{\partial L}{\partial \lambda} = 2A\lambda - 2\alpha(W+A)\lambda \quad \dots\dots\dots(1)$$

$$\frac{\partial L}{\partial \alpha} = - \{ \lambda' (W+A) \lambda - 1 \}$$

igualando estas derivadas a cero y de (1):

$$\{A - \alpha(W+A)\}\lambda = 0$$

$$(W+A)^{-1}\{A - \alpha(W+A)\}\lambda = 0$$

$$\{(W+A)^{-1}A - \alpha I\}\lambda = 0$$

este sistema de ecuaciones tiene soluciones diferentes de la trivial si:

$$|(W+A)^{-1}A - \alpha I| = 0$$

de donde α es una raíz característica de $(W+A)^{-1}A$.

No lo demostraremos, pero en general el número de raíces características positivas distintas que se obtienen son:

$$r = \min \{p, k-1\} = \text{rango}(A)$$

para cada raíz característica α_j se encuentra asociado un vector característico λ_j , por lo que habrá r vectores característicos ortogonales y en consecuencia el número máximo de funciones lineales discriminante será r . Las raíces características se ordenan en forma decreciente $\alpha_1, \alpha_2, \dots, \alpha_r$ y conforme a cada una de ellas se van obteniendo los vectores característicos y las funciones lineales correspondientes (a estas funciones se les llaman discriminantes), debemos aclarar a este respecto que no siempre es necesario derivar todas ellas, ya que un número menor de ellas pueden describir los grupos adecuadamente. Para determinar la importancia de las funciones discriminantes podemos basarnos en:

$$\rho_z = \frac{\alpha_z}{\sum_{i=1}^r \alpha_i}$$

es el porcentaje de la raíz característica z de la suma total de las raíces características. Aquellas funciones discriminantes cuyo ρ_z es pequeño se deshechan, en una investigación particular

debe establecerse a partir de cual valor deben removerse las funciones discriminantes, ya que no hay una regla fija para decidir cuando ρ_2 es suficientemente pequeña⁽¹⁾.

A las variables $v_1 = \lambda_1'x$, $v_2 = \lambda_2'x, \dots, v_r = \lambda_r'x$ se les llaman variables canónicas. En base a las funciones discriminantes $\lambda_1'x, \lambda_2'x, \dots, \lambda_r'x$ se pueden definir nuevos ejes en un espacio geométrico donde cada caso se considera como un punto. Las nuevas coordenadas de un individuo con vector discriminante x está dada por $(\lambda_1'x, \lambda_2'x, \dots, \lambda_r'x)$, es decir cada función discriminante determina una coordenada. El vector $\bar{x}(h) = (\bar{x}_{h1}, \bar{x}_{h2}, \dots, \bar{x}_{hr})$ contiene las medias de las variables discriminantes para el grupo h y sus coordenadas con respecto a los nuevos ejes son $(\lambda_1'\bar{x}(h), \dots, \lambda_r'\bar{x}(h))$, esta es la localización más típica de un caso en el grupo h , y se le denomina el centroide de la población h . El cuadrado de la distancia entre el vector $\bar{x}(h)$ y un vector discriminante cualquiera x en base a los nuevos ejes es el cuadrado de la distancia entre el vector $(\lambda_1'x, \lambda_2'x, \dots, \lambda_r'x)$ y el centroide del grupo h , expresando lo anterior en forma matemática tenemos:

$$\sum_{i=1}^r (\lambda_i'x - \lambda_i'\bar{x}(h))^2 = \sum_{i=1}^r |\lambda_i'(x - \bar{x}(h))|^2$$

si x es una observación que se desea clasificar dentro de alguno de los grupos la regla de decisión es:

$$x \in G_i \quad \sum_{j=1}^r (\lambda_j'(x - \bar{x}(i)))^2 = \text{Min}_h \sum_{j=1}^r (\lambda_j'(x - \bar{x}(h)))^2$$

$\bar{x}(h)$ es una estimación de $\mu(h)$.

Los valores de las funciones discriminantes que se obtie

(1) Existen otros criterios para juzgar la importancia de las funciones discriminantes. Si se desea una información acerca de ellos consultar Análisis Discriminante de Aranda y Méndez (Comunicaciones técnicas. IIMAS).

nen para el conjunto de casos y los coeficientes de las funciones discriminantes se utilizan para caracterizar diferencias entre los grupos.

CAPITULO 5 .

SELECCION DE VARIABLES.

Un investigador inicialmente basándose en individuos de quienes se conoce el grupo al cual pertenecen propone un conjunto de variables x_1, x_2, \dots, x_t para llevar a cabo la discriminación, eligiéndolas de tal manera que sea en estas variables donde difieran los grupos. Existen dos formas en que pueden usarse x_1, x_2, \dots, x_t :

- 1) El método directo: se incluye todo el conjunto de variables propuestas.
- 2) El método escalonado (stepwise): usa solamente un subconjunto de las variables x_1, x_2, \dots, x_t tomando en cuenta su poder discriminante.

Cualquiera de los métodos que se desee utilizar dará como resultado la selección de x_1, x_2, \dots, x_p , para $p < t$ variables que formarán el vector discriminante de quien dependerá la clasificación de un individuo. En la siguiente sección presentaremos una descripción detallada de cada uno de los métodos de selección de variables.

5.1 METODOS DE SELECCION DE VARIABLES.

El método directo. Se introducen todas las variables propuestas por el investigador dentro del Análisis Discriminante.

El método escalonado (stepwise). Para llevar a cabo la selección de un subconjunto de variables de x_1, x_2, \dots, x_t se siguen los siguientes pasos:

- 1) Se escoge un criterio de selección.
- 2) Se busca la variable que tenga el más alto valor en el criterio de selección. Supongamos que es x_j .
- 3) x_j es apareada con cada una de las variables $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_t$.
- 4) Se calcula para cada subconjunto de variables el criterio de selección y se escoge aquel que tiene el mayor.
- 5) En el subconjunto elegido se busca si existe alguna ó algunas variables redundantes y se elimina la peor⁽¹⁾.
- 6) El procedimiento es terminado cuando se han seleccionado todas las variables del conjunto o ninguna variable proporciona un mejor valor al comparar la puntuación del criterio obtenido en el paso 4 durante la iteración anterior y la actual, en caso contrario se continua con el paso siguiente.
- 7) El grupo de variables seleccionadas finalmente es apareado con las restantes sin tomar en cuenta la que acaba de eliminarse, aunque en iteraciones pos-

(1) La redundancia de una variable se presenta como una consecuencia de la inclusión, ya que la última variable que se introduce junto con otras ya incluidas puede contener la información de algunas variables ya seleccionadas, tornándose éstas últimas innecesarias, es decir redundantes, por ello cada una de las variables del grupo elegido, se prueba para determinar si su contribución es redundante, de donde se desprendería que no fuera lo suficientemente buena para la discriminación.

teriores si se considera y puede ser introducida si satisface el criterio de selección de variables y es la menos redundante dentro del subconjunto que se está juzgando en esta etapa.

8) Se regresa al paso 4.

Una vez que el procedimiento ha concluido hemos encontrado el conjunto de variables que se utilizarán para llevar a cabo la discriminación.

Hay razones especiales para aplicar alguno de los métodos descritos de selección de variables en un determinado problema. El método directo es adecuado, cuando por razones de tipo teórico se considera que todas las variables son importantes y se desea incluirlas en la discriminación, por el contrario, el método escalonado es recomendable cuando el conjunto x_1, x_2, \dots, x_t tiene información superflua o inútil de la diferencia de grupos y entonces un subconjunto de las variables propuestas es tan bueno como éste.

5.2 CRITERIOS DE SELECCION DE VARIABLES.

Entre los criterios de selección de variables se encuentran:

a) EL CRITERIO DE LA MINIMA LAMBDA DE WILKS.

Se consideran los cocientes multivariados F sobre la prueba de la diferencia de los centroides.

CRITERIO DE SELECCION. Se escoge aquella variable que al ser agregada a las anteriores maximiza el cociente F . Una variable que maximiza a F también minimiza la λ de Wilks.

b) EL CRITERIO DE MAHALANOBIS.

Está basado en la distancia generalizada de Mahalano-

bis.

CRITERIO DE SELECCION. Se elige la variable que produce la máxima distancia de Mahalanobis entre dos grupos más cercanos.

c) CRITERIO MINIMAX.

Se toman en cuenta todos los posibles cocientes F - multivariados para la prueba de la distancia de Mahalanobis entre pares de grupos.

CRITERIO DE SELECCION. Se elige la variable que maximiza el más pequeño cociente F.

d) CRITERIO MINDRESID.

Se define el valor:

$$R = \sum_{\substack{i=1 \\ i \neq j}}^k \frac{1}{1 + \frac{D_{ij}}{4}}$$

D_{ij} : es la distancia de Mahalanobis entre el grupo i y el grupo j.

CRITERIO DE SELECCION. Se elige la variable que minimiza el valor R.

P A R T E I I

Los ejemplos prácticos que se muestran en este trabajo son analizados a través del paquete de computación SPSS (Statistical Package for Social Sciences). El método implementado en nuestra rutina es el expuesto en la sección 4.2. Las suposiciones bajo las cuales se establece el modelo es que las variables discriminantes tienen una distribución Normal Multivariada con parámetros desconocidos.

C A P I T U L O 1

APLICACION DEL ANALISIS DISCRIMINANTE A LA BIOLOGIA.

1.1 LA POTENCIALIDAD DE LOS MUTAGENOS QUIMICOS EN LA INDUCCION DE MUTACIONES UTILES EN LA AGRICULTURA.

La evolución se basa en tres factores:

- a) La mutación que es la causa esencial de toda variación en los seres vivos.
- b) La hibridación que produce una serie de patrones genéticos.
- c) La selección natural que permite la reproducción de individuos portadores de patrones genéticos ventajosos para la especie en términos de supervivencia.

El hombre ha podido manipular estos factores con el fin de obtener mejor cantidad y calidad de satisfactores de sus necesidades. Por medio de la mutagénesis experimental pueden obtenerse mutaciones útiles con bastante frecuencia para lograr el mejoramiento de las plantas, desgraciadamente ha sido escasamente aplicada y ha sido reelegada a una alternativa poco útil. La mutagénesis experimental es importante como una solución a los problemas del deterioro de recursos genéticos, ya que el uso de esta metodología implica la posibilidad de alterar la base hereditaria y obtener genotipos nuevos no existentes en la naturaleza.

La inducción de mutaciones puede ser producida por -

medio de agentes físicos o químicos. Los procedimientos para explorar la potencialidad de los agentes químicos para la obtención de mutaciones en plantas superiores es relativamente sencilla y no precisa de instalaciones costosas ni sofisticadas y con un número reducido de equipo de laboratorio es posible realizar experimentos que eleven la frecuencia de mutaciones en la población tratada y eventualmente aislar mutantes - que resuelvan el problema específico. El metil metano Sulfo-nato (MMS) es un mutágeno químico susceptible de ser utilizado en diversas condiciones experimentales de manera que pueda - elevarse su efectividad y eficiencia mutagénica.

El triticale que es un híbrido del trigo y centeno - es un cultivo con grandes potencialidades agronómicas que no ha podido adaptarse a una amplia gama de ambientes, ni ha podido por diversas causas competir con cereales tradicionales, así pues es una planta que puede ser mejorada la ampliación - de su base genética y esto puede ser logrado mediante la inducción artificial de mutaciones.

OBJETIVO. El propósito de nuestro estudio es evaluar los efectos del MMS sobre el triticale.

Entre las posibles alteraciones que puede sufrir el triticale proveniente de una semilla tratada con el mutágeno se encuentran:

- a) Altura de la planta.
- b) Número de hijuelos (es el número de ramitas que salen de la semilla).
- c) El peso del grano de la planta (el peso de todas las semillas de la planta).
- d) Modificaciones en generaciones posteriores.

1.2. EXPERIMENTO REALIZADO PARA EVALUAR LOS EFECTOS DEL MMS SOBRE EL TRITICALE.

Para efectuar el análisis concienzudo de las mutaciones inducidas por el MMS en este cereal se siguió el siguiente programa: Se seleccionaron visualmente las semillas de triticale con buen llenado de grano, poco arrugamiento y embrión bien definido, más tarde fueron sometidas a condiciones de laboratorio de 20°C de temperatura y se les aplicó el mutágeno a las concentraciones de 0.0, 0.1, 0.20, 0.30, 0.40, 0.50 p/vol/hora. Una vez que hubo pasado un tiempo prudente desde la germinación de la planta se efectuaron las mediciones correspondientes a las variables mencionadas en los incisos a, b, c. A esta generación de plantas de Triticale se le denominó "Generación MM1". Para medir las modificaciones que sufren generaciones posteriores que provienen de las semillas tratadas con el MMS, se efectuó la germinación de las semillas obtenidas en la generación MM1, y se midieron las variables ya mencionadas, agregando una más, llamada de maduración, la cual se refiere al tiempo en el cual aparecen las semillas en la planta, es decir pueden hacer su aparición en forma muy precoz, normal, tardío o muy tardío. A estas nuevas plantas se les llamó "Generación MM2".

La investigación antes descrita fue realizada por el Dr. Abraham Rubluo Islas en el laboratorio de genética de la UNAM, quien proporcionó los datos resumidos en las siguientes páginas.

1.3 EFECTOS DEL MMS EN LA GENERACION MM1.

1.3.1 DATOS OBTENIDOS PARA LA GENERACION MM1.

EXPLRIMENTO N°1

CONCENTRACION	ALTURA	N. DE HIJUELOS	PESO
.50	75	15	9.0
.60	60	10	8.8
.60	65	16	9.3
.60	60	16	0.1
.60	65	17	15.0
.60	60	15	4.9
.60	75	17	10.9
.60	70	22	0.5
.60	75	20	9.8
.60	60	15	5.3
.20	70	17	8.8
.20	65	9	6.3
.20	60	18	8.8
.20	75	13	7.2
.20	70	14	11.0
.20	72	23	11.0
.20	72	24	9.9
.20	70	12	9.2
.20	60	19	7.2
.20	65	13	3.4
.30	72	15	9.5
.30	69	14	11.4
.30	65	11	7.1
.30	72	8	5.4
.30	85	12	9.3
.30	69	18	8.0
.30	75	22	5.2
.30	73	19	7.3
.30	70	20	10.3
.30	54	6	3.8
.40	77	15	9.4
.40	69	10	5.9
.40	75	14	9.1
.40	58	17	4.3
.40	70	22	13.3
.40	58	15	10.0
.40	60	22	0.0
.40	60	9	0.4
.40	70	22	9.5
.40	65	18	4.0

EXPERIMENTO MM1

CONCENTRACION

ALTURA

N DE HIJUELOS

PESO

.50	70	22	4.6
.50	76	14	4.9
.40	75	19	6.7
.50	69	17	8.7
.50	60	20	2.5
.50	54	11	6.6
.50	60	14	6.4
.10	67	12	2.8
.10	66	15	11.0
.10	60	11	7.1
.10	71	9	8.7
.10	60	29	7.3
.10	68	19	2.7
.10	53	21	9.1
.10	60	22	5.0
.10	65	19	8.2
.10	62	6	8.5
.50	75	23	9.5
.30	60	19	8.3
.30	65	12	5.9
.30	68	11	8.0
.30	64	13	11.3
.30	68	14	6.1
.30	68	17	10.1
.30	70	12	7.5
.20	70	18	10.9
.20	48	8	3.6
.20	72	14	10.0
.20	70	14	10.0
.20	68	12	16.5
.20	72	15	13.2
.20	70	20	12.5
.20	69	14	9.9
.20	76	8	10.7
.20	71	10	11.2
.50	70	7	8.9
.50	71	10	8.3
.50	79	12	12.5
.50	80	20	4.0
.50	85	9	11.6

EXPERIMENTO MM1

CONCENTRACION	ALTURA	N DE RIJUELOS	PLSO
.50	85	13	11.7
.50	85	13	14.2
.50	81	7	11.1
.40	75	12	11.5
.40	70	12	9.2
.40	84	11	15.5
.40	72	4	5.5
.40	82	11	14.9
.40	84	12	16.2
.40	87	7	13.7
.40	80	19	13.8
.10	75	12	13.3
.10	73	11	9.4
.10	85	9	10.1
.10	76	16	12.1
.10	82	9	18.5
.10	90	7	12.3
.10	85	14	18.4
.10	81	13	6.3
.10	80	8	14.2
.10	93	8	7.5
.00	65	12	9.3
.00	64	12	10.7
.00	80	6	11.2
.00	64	13	10.9
.00	83	6	17.0
.00	77	11	19.7
.00	86	10	11.5
.00	92	11	19.9
.00	85	12	11.4
.00	86	10	17.0
.40	65	0	4.0
.40	79	13	8.2
.40	69	10	5.5
.40	70	10	10.7
.40	70	9	6.5
.40	80	13	11.9
.40	70	9	9.1
.40	66	17	4.7
.40	72	15	8.0

EXPERIMENTO MM1

CONCENTRACION	ALTURA	N DE HIJULLOS	PESO
.40	79	20	11.8
.50	86	5	6.1
.50	92	9	13.9
.50	88	5	9.3
.50	87	7	11.7
.50	90	8	12.3
.50	89	9	15.0
.50	84	13	13.6
.20	80	18	5.5
.20	75	10	11.6
.20	95	9	14.8
.20	78	13	14.9
.20	82	10	16.4
.20	82	13	12.7
.20	92	10	17.9
.20	100	12	13.7
.20	85	8	15.1
.20	90	10	13.1
.30	76	9	13.3
.30	90	7	12.1
.30	95	12	19.0
.30	100	8	15.9
.30	92	8	17.1
.30	110	7	14.3
.30	80	10	16.8
.30	85	14	19.2
.00	81	9	17.1
.00	82	11	13.3
.00	89	10	12.8
.00	87	7	15.5
.00	85	19	17.9
.00	86	20	19.9
.00	90	6	18.1
.00	97	14	13.3
.00	97	10	11.9
.00	76	12	21.8
.10	86	6	8.3
.10	72	25	20.4
.10	76	10	17.9
.10	77	11	13.9

EXPERIMENTO MM1

CONCENTRACION	ALTURA	N. DE HIJUELOS	PESO
.10	82	18	13.4
.10	93	11	17.7
.10	88	11	15.6
.10	87	11	18.1
.10	77	14	12.0
.10	90	11	19.9

1.3.2 APLICACION DEL ANALISIS DISCRIMINANTE AL EXPERIMENTO MM1

En base a nuestro objetivo hemos subdividido en grupos los distintos tratamientos aplicados, en la siguiente forma:

GRUPO	TRATAMIENTO CORRESPONDIENTE
1	0.00
2	0.10
3	0.20
4	0.30
5	0.40
6	0.50

El grupo testigo, es aquel que no se le aplica ningún tipo de tratamiento, pero sirve para hacer inferencias de lo que sucederfa en condiciones completamente normales.

Las variables en las cuales se espera que existan diferencias entre los grupos para la generación MM1 son:

L1	altura de la planta en cm
L2	número de hijuelos.
L3	peso del grano en gramos.

1.3.3 PRINCIPALES RESULTADOS OBTENIDOS PARA LA GENERACION MM1.

Los principales resultados se reportan en las siguientes páginas.

GROUP COUNTS

	GROUP 1	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6	TOTAL
COUNT	30,000	30,000	30,000	26,000	29,000	21,000	166,000

MEANS

	GROUP 1	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6	TOTAL
L1	77.8333	76.6000	74.1333	75.6510	72.2759	78.0000	73.5783
L3	12.7933	11.8800	10.8067	10.5038	9.2862	9.1933	10.8349

STANDARD DEVIATIONS

	GROUP 1	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6	TOTAL
L1	11.8807	10.9972	10.8524	11.0994	7.7731	10.9955	11.0241
L3	4.3023	5.2190	4.9310	4.5071	4.8212	4.7948	4.7927
L3	4.8845	4.7599	3.6240	4.3685	3.7875	4.0012	4.3688

MILK'S LAMBDA (U-STATISTIC) AND UNIVARIATE F-RATIO WITH 9 AND 160 DEGREES OF FREEDOM

VARIABLE MILK'S LAMBDA F

L1	0.9682	1.0495
L2	0.5087	0.4310
L3	0.0106	3.1468

CUADRO 1

MUTAGENOS

WITHIN GROUPS COVARIANCE MATRIX

	L1	L2	L3
L1	121.3482		
L2	20.3482	23.3726	
L3	29.8367	-4.5524	17.8974

WITHIN GROUPS CORRELATION MATRIX

	L1	L2	L3
L1	1.0000		
L2	-0.3820	1.0000	
L3	0.6402	-0.2226	1.0000

TOTAL COVARIANCE MATRIX

	L1	L2	L3
L1	121.5302		
L2	20.4395	22.9696	
L3	30.1851	-4.2943	19.0584

CUADRO 2

ANALYSIS NUMBER 1
F FOR INCLUSION 1.00000
F FOR DELETION 1.00000
TOLERANCE LEVEL 0.00100
MAXIMUM STEPS 6

SOLUTION METHOD = STEPWISE. SELECT VARIABLE WHICH WILL
MINIMIZE MILKS' LAMBDA.

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

GROUP	1	GROUP	2	GROUP	3	GROUP	4	GROUP	5	GROUP	6
	0.16667		0.16667		0.16667		0.16667		0.16667		0.16667

CUADRO 3

Se ha considerado en el problema planteado que las probabilidades a priori de pertenecer a cualquier grupo son iguales.

La selección de una variable además de cumplir con el criterio de selección depende de:

- a) Tolerancia
- b) F de entrada
- c) F para remover

La F de entrada es la prueba de la significancia estadística del incremento de la separación del centroide agregado por una variable y la separación producida por las variables introducidas previamente. La variable en cuestión es seleccionada si su cociente F es suficientemente grande. Se considera suficientemente grande si es mayor que la F para inclusión.

La pérdida del poder discriminante de una variable es una consecuencia de la inclusión de variables. La contribución de una variable a la discriminación se mide a partir de la F para remover, la cual es una prueba parcial multivariada del poder discriminante agregado para una variable determinada unida con un grupo de variables. Si la F para remover es menor que un valor especificado llamado F de rechazo (F for deletion), la variable se excluirá del análisis.

Una variable que toma valores muy pequeños podría acarrear problemas en los cálculos que se toman en cuenta, esto -- conduciría a estimaciones imperfectas, por ello si sus valores son menores que cierta Tolerancia esta variable no será escogida para ser introducida dentro del análisis.

La lambda de Wilks nos proporciona información acerca del poder discriminante de una variable en forma inversa, es decir mientras mayor sea este valor menor es su poder discriminatorio. En la tabla sumaria, en la parte donde se especifica las variables que se incluyen en cada paso, aparece una columna correspondiente a la lambda de Wilks, este valor pertenece al conjunto que contiene a las variables incluidas en pasos anteriores y la que se anexa en el n-ésimo paso. Así por ejemplo tenemos para la variable L1 que es la primera en ser incluida una lambda de 0.91062, al anexar la variable L3 la lambda es de --- 0.8569.

Al terminar la lista de variables seleccionadas para llevar a cabo la discriminación la lambda debe ser pequeña, para considerar que el conjunto elgido posee un buen poder discriminatorio (en este caso no sucede así). La lambda de Wilks puede ser transformada en una χ^2 para efectuar una prueba de significancia estadística. El nivel de significancia al que se lleva a cabo esta prueba aparece en la columna siguiente, lo cual indica que para la variable L1 existe una probabilidad de 0.01 de que haya una lambda menor que la que se tiene y para la variable L1 y L3 es de 0.006.

Al derivar las funciones discriminantes, también se utiliza la lambda de Wilks como un indicador de la información que contienen. Cuando ninguna función ha sido derivada, solamente se tiene la lambda correspondiente a las variables seleccionadas para la discriminación, como una parte de este poder se pierde al hacer uso de la primera función, la lambda de Wilks empieza a crecer, esto sucede de la misma manera de función en función, hasta que las últimas presentan un valor muy elevado, a partir de este hecho podemos decidir hasta que función nos es útil considerar. En nuestro caso, cuando hay una sola función la lambda se ha elevado a 0.9689 el cual es un valor muy grande lo que se traduce en un bajo poder discriminatorio en la función que ha sido derivada, y su nivel de significancia es de 0.278.

La V de Rao es una medida de la distancia generalizada entre todos los grupos, mientras mayor es su valor hay una separación más grande entre todos los grupos. El valor de la V de Rao al igual que el de la lambda de Wilks corresponde a las variables incluidas en pasos anteriores y la que se agrega en el n-ésimo. En nuestro ejemplo el valor para L1 es de 15.70407 y para L1 y L3 de 26.03524. El incremento de la V de Rao debido a la inclusión de variables se distribuye como una χ^2 con un grado de libertad cuando hay un gran número de casos.

VARIABLE ENTERED ON STEP NUMBER 2.. L1

WILKS' LAMBDA 0.85690 APPROXIMATE F 2.55157 DEGREES OF FREEDOM 318.00 SIGNIFICANCE 0.008
 RAD'S V 28.03524 CHANGE IN V 10.33117

F MATRIX - DEGREES OF FREEDOM: 2, 159

GROUP	1	2	3	4	5
GROUP 1	0.30623				
GROUP 2	1.24363	0.52750			
GROUP 3	2.02316	0.20795	0.43191		
GROUP 4	5.08315	2.75421	0.89334	0.73716	
GROUP 5	8.40538	5.65142	4.75798	2.33449	3.02919

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

VARIABLE	ENTER	REMOVE
L1	3.92126	4.99126
L3	3.14081	4.12951

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

VARIABLE	TOLERANCE	ENTER	CRITERION
L2	0.85322	0.23176	0.67963

F LEVEL INSUFFICIENT FOR FURTHER COMPUTATION

SUMMARY TABLE

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F TO ENTER OR REMOVE	NUMBER INCLUDED	WILKS' LAMBDA	SIG.	RAD'S V	CHANGE IN RAD'S V	SIG. OF CHANGE
1	L1		3.14081	1	0.91062	0.010	15.70407	15.70407	0.008
2		L3	3.04128	2	0.85888	0.008	28.03524	10.33117	0.008

DISCRIMINANT	EIGENVALUE	PERCENTAGE	CORRELATION	FUNCTIONS	LAMBDA	CHI-SQUARE	DF	SIGNIFICANCE
1	0.13061	80.26	0.340	1	3.1570	24.052	10	0.008
	0.03211	19.74	0.178	1		5.089	4	0.277

CUADRO 4

En este cuadro se presenta el último paso de selección de variables, la matriz de pruebas de significancia F para la distancia de Mahalonobis entre grupos, una tabla resumen de la forma como fueron introducidas las variables, e información sobre las funciones discriminantes derivadas.

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

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
L1	-0.71422	1.06628
L3	1.27850	-0.11181

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
L1	-0.06479	0.09672
L3	0.29236	-0.02561
CONSTANT	1.72344	-7.03265

CUADRO 5

Solamente dos funciones discriminantes fueron derivadas. Los coeficientes estandarizados se derivan de tal manera que los valores discriminantes producidos están en forma estándar, es decir que sobre todos los casos en el análisis, el valor de la función tendrá media cero y desviación estándar de 1. Cuando se ignora su signo representan la contribución relativa a la variable asociada en esa función, el signo denota si es positiva o negativa.

Los coeficientes no estandarizados forman la función discriminante a partir de la cual se calculan los valores discriminantes.

MUTAGENS

CENTROIDS OF GROUPS IN REDUCED SPACE

		FUNC 1	FUNC 2
GROUP	1	0.45335	0.12926
GROUP	2	0.23986	0.07205
GROUP	3	0.08534	0.13904
GROUP	4	0.10136	0.01579
GROUP	5	0.23960	0.27976
GROUP	6	0.65523	0.27781

CUADRO 6

Este cuadro presenta la localización más típica de un individuo perteneciente al grupo i , para $i=1, 2, \dots, 6$.

CUADRO 7

SUBJECT	SEQNUM	MISSING VALUES	ACTUAL GROUP	HIGHEST PROBABILITY GROUP	HIGHEST PROBABILITY	GROUP	HIGHEST	DISCRIMINANT FUNCTIONS
A	1	0	*	6	0.947	1	0.187	0.500
A	2	0	*	6	0.937	1	0.221	0.413
A	3	0	*	6	0.639	1	0.221	0.377
A	4	0	*	6	0.534	1	0.206	0.233
A	5	0	*	6	0.319	1	0.247	0.319
A	6	0	*	6	0.492	1	0.187	0.729
A	7	0	*	6	0.798	1	0.181	0.955
A	8	0	*	6	0.778	1	0.224	0.929
A	9	0	*	6	0.584	1	0.180	0.865
A	10	0	*	6	0.519	1	0.192	0.613
A	11	0	*	6	0.559	1	0.209	0.363
A	12	0	*	6	0.559	1	0.209	0.363
A	13	0	*	6	0.500	1	0.192	0.711
A	14	0	*	6	0.476	1	0.183	0.553
A	15	0	*	6	0.476	1	0.183	0.553
A	16	0	*	6	0.600	1	0.235	0.323
A	17	0	*	6	0.600	1	0.235	0.323
A	18	0	*	6	0.722	1	0.244	0.524
A	19	0	*	6	0.689	1	0.188	0.524
A	20	0	*	6	0.232	1	0.248	0.391
A	21	0	*	6	0.332	1	0.248	0.591
A	22	0	*	6	0.680	1	0.184	0.443
A	23	0	*	6	0.334	1	0.237	0.434
A	24	0	*	6	0.500	1	0.209	0.363
A	25	0	*	6	0.488	1	0.209	0.363
A	26	0	*	6	0.488	1	0.209	0.363
A	27	0	*	6	0.488	1	0.209	0.363
A	28	0	*	6	0.488	1	0.209	0.363
A	29	0	*	6	0.213	1	0.250	0.189
A	30	0	*	6	0.333	1	0.250	0.189
A	31	0	*	6	0.333	1	0.250	0.189
A	32	0	*	6	0.333	1	0.250	0.189
A	33	0	*	6	0.222	1	0.174	0.666
A	34	0	*	6	0.222	1	0.174	0.666
A	35	0	*	6	0.222	1	0.174	0.666
A	36	0	*	6	0.222	1	0.174	0.666
END OF DATA				1	0.000	2	0.167	0.833

END OF DATA INPUT READ COUNT = 166

BLK # TEA # SEQU NUM

MISSING

AGRUOP

HIGHEST PROBABILITY

GROUP HIGHEST

DISCRIMINANT SCORES

BLK #	TEA #	SEQU NUM	MISSING	AGRUOP	HIGHEST PROBABILITY	GROUP HIGHEST	DISCRIMINANT SCORES
0000	A	11	0	3	0.979	0.203	0.233
0000	A	12	0	3	0.357	0.221	0.442
0000	A	13	0	3	0.991	0.264	0.435
0000	A	14	0	3	0.890	0.194	0.337
0000	A	15	0	3	0.704	0.195	0.450
0000	A	16	0	3	0.788	0.189	0.377
0000	A	17	0	3	0.516	0.247	0.515
0000	A	18	0	3	0.307	0.266	0.555
0000	A	19	0	3	0.688	0.331	0.480
0000	A	20	0	3	0.930	0.186	0.333
0000	A	21	0	3	0.788	0.331	0.480
0000	A	22	0	3	0.930	0.186	0.333
0000	A	23	0	3	0.204	0.194	0.527
0000	A	24	0	3	0.722	0.220	0.490
0000	A	25	0	3	0.999	0.181	0.437
0000	A	26	0	3	0.907	0.197	0.424
0000	A	27	0	3	0.440	0.439	0.390
0000	A	28	0	3	0.969	0.202	0.450
0000	A	29	0	3	0.440	0.439	0.390
0000	A	30	0	3	0.824	0.226	0.427
0000	A	31	0	3	0.685	0.260	0.334
0000	A	32	0	3	0.876	0.186	0.447
0000	A	33	0	3	0.772	0.186	0.457
0000	A	34	0	3	0.332	0.293	0.333
0000	A	35	0	3	0.711	0.371	0.333
0000	A	36	0	3	0.498	0.170	0.433
0000	A	37	0	3	0.527	0.195	0.333
0000	A	38	0	3	0.527	0.195	0.333
0000	A	39	0	3	0.527	0.195	0.333
0000	A	40	0	3	0.527	0.195	0.333
0000	A	41	0	3	0.527	0.195	0.333
0000	A	42	0	3	0.527	0.195	0.333
0000	A	43	0	3	0.527	0.195	0.333
0000	A	44	0	3	0.527	0.195	0.333
0000	A	45	0	3	0.527	0.195	0.333
0000	A	46	0	3	0.527	0.195	0.333
0000	A	47	0	3	0.527	0.195	0.333
0000	A	48	0	3	0.527	0.195	0.333
0000	A	49	0	3	0.527	0.195	0.333
0000	A	50	0	3	0.527	0.195	0.333

END OF DATA INPUT READ COUNT = 166

SUBJECT NUMBER	MESSAGES	AGROUP	HIGHEST PROBABILITY	GROUP HIGHEST	DISCRIMINANT SCORES
NONA	21	4	0.926	3	1.000
NONA	22	4	0.779	3	1.000
NONA	23	4	0.676	3	1.000
NONA	24	4	0.677	3	1.000
NONA	25	4	0.529	3	1.000
NONA	26	4	0.528	3	1.000
NONA	27	4	0.422	3	1.000
NONA	28	4	0.421	3	1.000
NONA	29	4	0.424	3	1.000
NONA	30	4	0.423	3	1.000
NONA	31	4	0.427	3	1.000
NONA	32	4	0.429	3	1.000
NONA	33	4	0.427	3	1.000
NONA	34	4	0.429	3	1.000
NONA	35	4	0.427	3	1.000
NONA	36	4	0.429	3	1.000
NONA	37	4	0.427	3	1.000
NONA	38	4	0.429	3	1.000
NONA	39	4	0.427	3	1.000
NONA	40	4	0.429	3	1.000
NONA	41	4	0.427	3	1.000
NONA	42	4	0.429	3	1.000
NONA	43	4	0.427	3	1.000
NONA	44	4	0.429	3	1.000
NONA	45	4	0.427	3	1.000
NONA	46	4	0.429	3	1.000
NONA	47	4	0.427	3	1.000
NONA	48	4	0.429	3	1.000
NONA	49	4	0.427	3	1.000
NONA	50	4	0.429	3	1.000
NONA	51	4	0.427	3	1.000
NONA	52	4	0.429	3	1.000
NONA	53	4	0.427	3	1.000
NONA	54	4	0.429	3	1.000
NONA	55	4	0.427	3	1.000
NONA	56	4	0.429	3	1.000
NONA	57	4	0.427	3	1.000
NONA	58	4	0.429	3	1.000
NONA	59	4	0.427	3	1.000
NONA	60	4	0.429	3	1.000
NONA	61	4	0.427	3	1.000
NONA	62	4	0.429	3	1.000
NONA	63	4	0.427	3	1.000
NONA	64	4	0.429	3	1.000
NONA	65	4	0.427	3	1.000
NONA	66	4	0.429	3	1.000
NONA	67	4	0.427	3	1.000
NONA	68	4	0.429	3	1.000
NONA	69	4	0.427	3	1.000
NONA	70	4	0.429	3	1.000
NONA	71	4	0.427	3	1.000
NONA	72	4	0.429	3	1.000
NONA	73	4	0.427	3	1.000
NONA	74	4	0.429	3	1.000
NONA	75	4	0.427	3	1.000
NONA	76	4	0.429	3	1.000
NONA	77	4	0.427	3	1.000
NONA	78	4	0.429	3	1.000
NONA	79	4	0.427	3	1.000
NONA	80	4	0.429	3	1.000
NONA	81	4	0.427	3	1.000
NONA	82	4	0.429	3	1.000
NONA	83	4	0.427	3	1.000
NONA	84	4	0.429	3	1.000
NONA	85	4	0.427	3	1.000
NONA	86	4	0.429	3	1.000
NONA	87	4	0.427	3	1.000
NONA	88	4	0.429	3	1.000
NONA	89	4	0.427	3	1.000
NONA	90	4	0.429	3	1.000
NONA	91	4	0.427	3	1.000
NONA	92	4	0.429	3	1.000
NONA	93	4	0.427	3	1.000
NONA	94	4	0.429	3	1.000
NONA	95	4	0.427	3	1.000
NONA	96	4	0.429	3	1.000
NONA	97	4	0.427	3	1.000
NONA	98	4	0.429	3	1.000
NONA	99	4	0.427	3	1.000
NONA	100	4	0.429	3	1.000

END OF DATA INPUT READ COUNT = 166

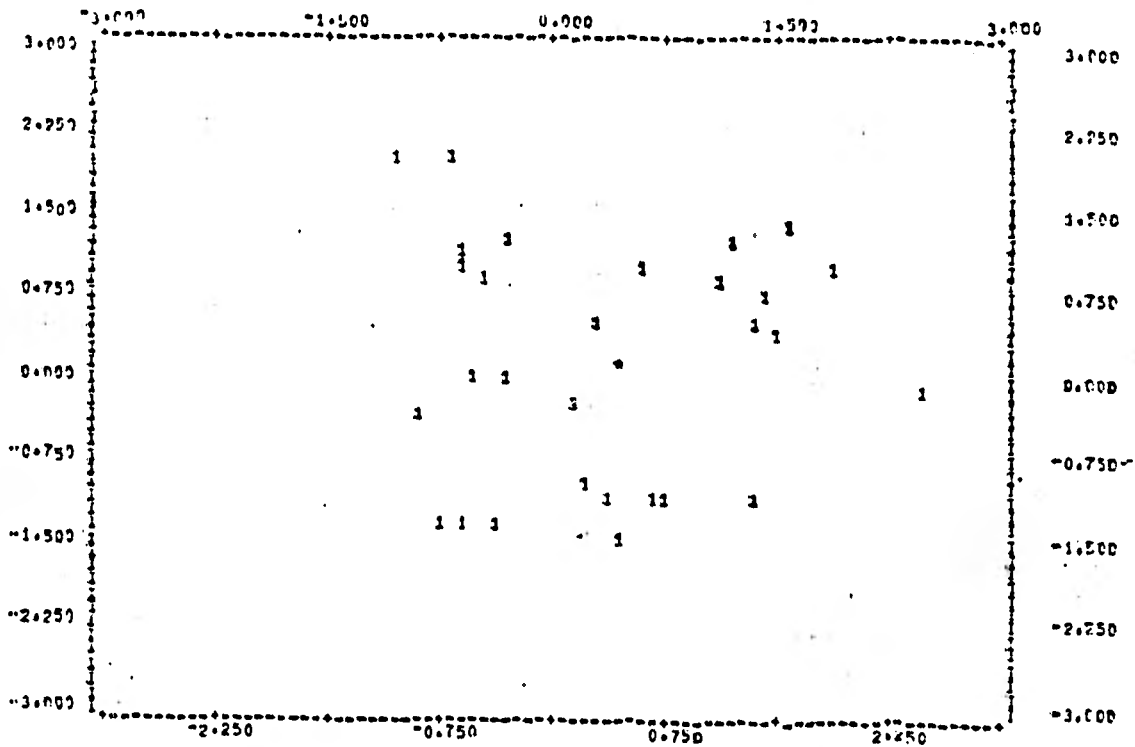
SLIP ID	SEQ NUM	VARIABLES	AGENTS	HIGHEST PROBABILITY	GROUP HIGHEST	DISCRIMINANT FUNCTIONS
000001	000001	000001	000001	000001	000001	000001
000002	000002	000002	000002	000002	000002	000002
000003	000003	000003	000003	000003	000003	000003
000004	000004	000004	000004	000004	000004	000004
000005	000005	000005	000005	000005	000005	000005
000006	000006	000006	000006	000006	000006	000006
000007	000007	000007	000007	000007	000007	000007
000008	000008	000008	000008	000008	000008	000008
000009	000009	000009	000009	000009	000009	000009
000010	000010	000010	000010	000010	000010	000010
000011	000011	000011	000011	000011	000011	000011
000012	000012	000012	000012	000012	000012	000012
000013	000013	000013	000013	000013	000013	000013
000014	000014	000014	000014	000014	000014	000014
000015	000015	000015	000015	000015	000015	000015
000016	000016	000016	000016	000016	000016	000016
000017	000017	000017	000017	000017	000017	000017
000018	000018	000018	000018	000018	000018	000018
000019	000019	000019	000019	000019	000019	000019
000020	000020	000020	000020	000020	000020	000020

Con el objeto de observar que tan buena es la discriminación se presentan los resultados referentes a la clasificación de los casos cuyos grupos son conocidos.

$P(G/x)$ es la probabilidad condicional de que un individuo se clasifique en el grupo G dado su vector discriminante x . Se presentan las dos probabilidades condicionales más altas.

$P(x/G)$ es la probabilidad de que un miembro del grupo predicho esté tan lejos del centroide como el caso que se considera. Si este valor es pequeño indica la posibilidad de que este caso no pertenezca a la población de la cual provienen los grupos especificados.

DISCRIMINANT FUNCTION VS. DISCRIMINANT 2 (VERTICAL). * INDICATES / GROUP IDENTIFIED.

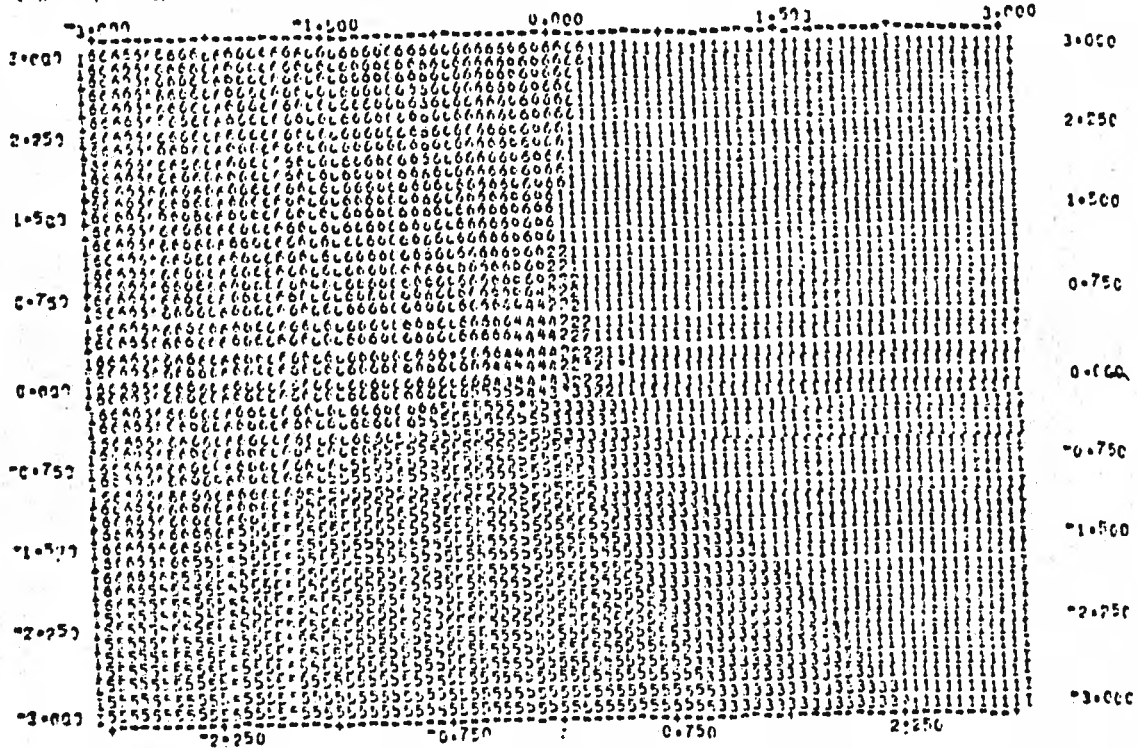


CUADRO 8

La gráfica anterior muestra donde quedarían situados los elementos cuyo grupo conocido es el 1 utilizando las funciones discriminantes. De la misma manera se efectúa la localización de los miembros pertenecientes a los otros grupos.

MAPA TERRITORIAL

MAPA TERRITORIAL DE LAS FUNCIONES DISCRIMINANTES 1 (HORIZONTAL) VS. DISCRIMINANTE 2 (VERTICAL). * INDICATES A GROUP



CUADRO 9

En base a las dos primeras funciones discriminantes, el mapa territorial muestra las regiones de clasificación de cada uno de los grupos en los cuales se encuentra dividido el espacio muestral. Las regiones se encuentran denotadas -- por su número respectivo y los asteriscos representan la localización de los centroides correspondientes.

PREDICTION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP					
		GP. 1	GP. 2	GP. 3	GP. 4	GP. 5	GP. 6
GROUP 1	30.	40.12% 40.0%	0.0%	16.5% 16.7%	3.1% 3.3%	16.5% 16.7%	23.7% 23.3%
GROUP 2	30.	40.12% 40.0%	6.2% 6.7%	6.2% 6.7%	3.1% 3.3%	23.7% 23.3%	20.6% 20.0%
GROUP 3	30.	23.7% 23.3%	6.2% 6.7%	20.6% 20.0%	3.1% 3.3%	26.8% 26.7%	20.6% 20.0%
GROUP 4	26.	19.5% 19.2%	0.0%	19.5% 19.2%	0.0%	34.8% 34.6%	26.7% 26.3%
GROUP 5	29.	17.5% 17.2%	4.0% 13.8%	2.0% 6.9%	0.0%	10.0% 34.5%	8.0% 27.6%
GROUP 6	21.	14.3%	4.8%	0.0%	0.0%	23.8%	57.1%

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 25.30%

CUADRO 10

1.3.4 ANALISIS DE LOS RESULTADOS OBTENIDOS PARA LA GENERACION MMI.

Las variables que tienen poder discriminatorio son:

- L1 altura de la planta
- L3 peso del grano en gramos.

L2 la variable correspondiente al número de hijuelos no tiene poder discriminatorio, es decir, la aplicación del mutágeno no afectó en forma notable esta variable para lograr diferencias entre los grupos (véase el cuadro #4 donde se presenta el último paso de selección de variables).

Entre las características deseables a través de la aplicación del mutágeno, es la obtención de alturas mínimas. El grupo que tiene la altura promedio más baja es el grupo 5 correspondiente a la concentración de 0.4 p/vol/hora con 72.28 cm, mientras que la altura máxima se presentó en la concentración de 0.5 p/vol/hora con una media de 78 cm (cuadro #1).

El peso del grano alcanzó un promedio mínimo de 9.133 grs a la concentración de .5 p/vol/hora/ y la media máxima se obtuvo en el grupo testigo con 12.79 grs (cuadro #1).

Utilizando las variables discriminantes L1 y L3 los grupos que presentan una mayor diferenciación son:

El grupo testigo y la aplicación del mutágeno a las concentraciones de 0.4 y 0.5.

La aplicación del mutágeno a la concentración de 0.1 y la concentración de 0.5.

La aplicación de la concentración de 0.2 y la de 0.5.

Dos funciones discriminantes fueron derivadas. Las funciones discriminantes estandarizadas son:

$$D1 = -0.71422L1 + 1.27850L3$$

$$D2 = 1.06628L1 - 0.11181L3$$

en base a ellas podemos observar que la variable L3 es la que contribuye mayormente a la discriminación en la función D1, mientras que en D2 es L1 la que proporciona la mayor contribución.

$$D1 = -0.06479L1 + 0.29286L3 + 1.72344$$

$$D2 = 0.09672L1 - 0.02561L3 - 7.03265$$

son las funciones discriminantes no estandarizadas utilizadas para llevar a cabo la clasificación de un nuevo caso. D1 tiene una importancia relativa de 80.26% y D2 de 19.74%. La correlación canónica de estas funciones mediante las cuales se mide su habilidad de discriminación es muy baja⁽¹⁾, ya que para D1 tiene un valor de 0.340 y para D2 es de 0.176, por lo que se espera que la clasificación de un caso no sea lo suficientemente buena.

La clasificación de un individuo x se lleva a cabo sustituyendo los valores de las variables discriminantes L1 y L3 en las funciones D1 y D2 con coeficientes no estandarizados, estos valores se unen para formar el punto (D1,D2) para determinar la región a la cual pertenece y en base a ella asignar al individuo x a algún grupo. Los individuos cuyo gru

(1) Esta correlación canónica es obtenida para m-1 variables ficticias indicadoras de pertenencia a los m grupos y las funciones discriminantes derivadas.

po es conocido se clasificaron utilizando las funciones discriminantes D1 y D2 no estandarizadas para observar que tan buenos son los resultados obtenidos (véase el cuadro #7), el resumen de estas clasificaciones se presenta en el cuadro #10 donde se reportan los principales resultados siguientes:

- 1) El grupo mejor clasificado es el correspondiente a la concentración de 0.5 p/vol/hora, con 57.1% de casos - clasificados correctamente.
- 2) El grupo 4 es el peor clasificado, ya que 0.0% de los casos se clasificaron correctamente.
- 3) Las fuentes mayores de error se encuentran en:
 - a) La clasificación de individuos del grupo 2 como pertenecientes al grupo 1, cuyo porcentaje es del 40%.
 - b) La clasificación de individuos del grupo 4 como pertenecientes al grupo 5, con un porcentaje de 34.6%

Estos errores son debidos al parecido que existe entre estos grupos (cuadro# 4).

- 4) Tomando en cuenta el total de casos, solamente se clasificó correctamente el 25.30% de ellos, el cual es un porcentaje muy bajo.

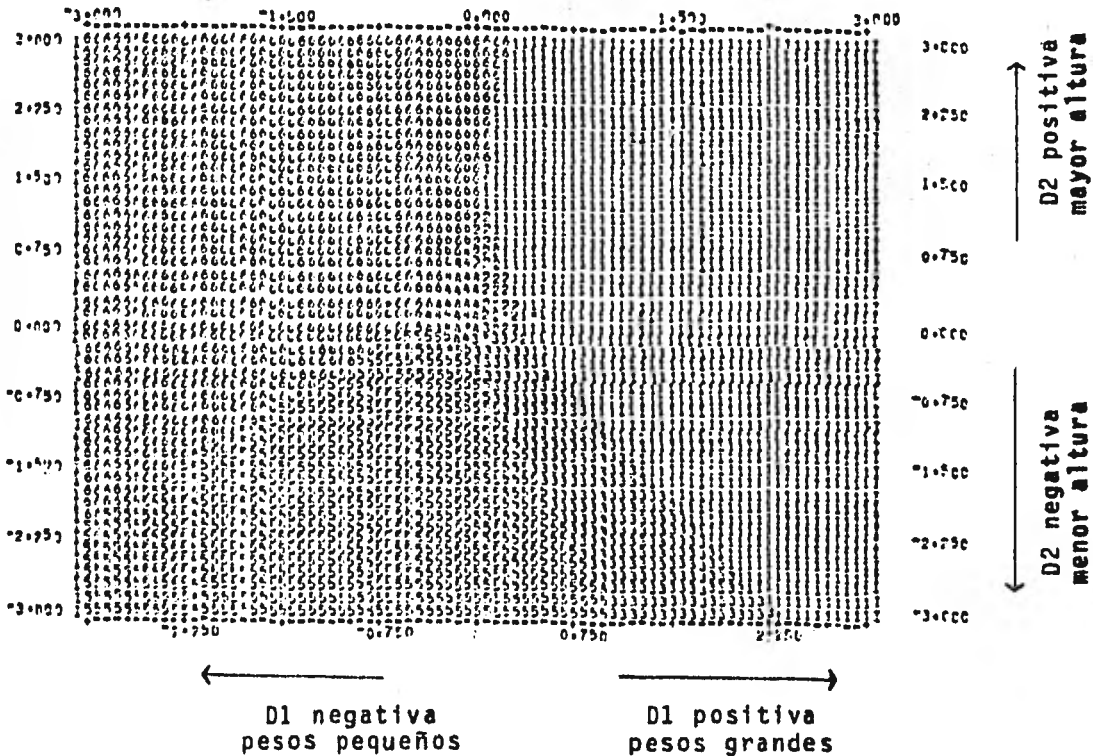
1.3.5 CRACTERIZACION DE LOS GRUPOS EN BASE AL MAPA TERRITORIAL

El mapa territorial contiene las regiones de clasificación correspondientes a cada grupo. Cada una de estas regiones está marcada con el número del grupo que están definiendo. El mapa territorial se interpreta a partir de las funciones -- discriminantes estandarizadas:

$$D1 = -0.71422 L1 + 1.2785 L3$$

$$D2 = 1.06628 L1 - 0.011181 L3$$

MAPA DE DISCRIMINANTE SCORE 1 (HORIZONTAL) vs. DISCRIMINANTE SCORE 2 (VERTICAL). * INDICATES A CRUI



La caracterización de los grupos se hace en base al mapa territorial, aunque debe aclararse que esta es la interpretación que se daría si las funciones discriminantes tuvieran una buena habilidad para la discriminación, la detallamos con el fin de que se observe la manera como se ha analizado el mapa territorial, no tomando en cuenta que sea completamente válida por la razón ya expuesta.

De esta manera presentaremos las características -- principales que distinguen a cada uno de los grupos en la siguiente tabla resumen:

GRUPO	CONCENTRACION CORRESPONDIENTE	PESO *	ALTURA *
1	0	Grande	Grande
2	.1	Grande	Media
3	.2	Grande	Baja
4	.3	Pequeño	Media
5	.4	Pequeño	Baja
6	.5	Pequeño	Grande

* Estos valores son los estandarizados.

1.4 EFECTOS DEL MMS EN LA GENERACION MM2

1.4.1 DATOS OBTENIDOS PARA LA GENERACION MM2

CLAVES USADAS PARA LA MADUREZ DE LA PLANTA A LA
APARICION DE LA SEMILLA

1 MUY PRECOZ

2 PRECOZ

3 NORMAL

4 TARDIO

5 MUY TARDIO

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.00	2	100	18	16.0
.00	4	104	15	5.0
.00	4	105	14	5.0
.00	4	102	30	43.5
.00	4	105	16	30.0
.00	2	107	16	12.0
.00	2	108	27	51.0
.00	3	107	29	35.5
.00	3	104	17	19.0
.00	3	105	34	55.5
.00	4	105	22	19.0
.00	3	108	25	38.0
.00	4	106	22	39.5
.00	2	102	19	22.5
.00	3	102	18	11.0
.00	4	105	26	33.0
.00	4	100	18	22.0
.00	3	92	17	22.0
.00	3	105	18	22.5
.00	3	103	22	35.6
.00	4	92	13	6.5
.00	4	100	12	18.5
.00	4	98	10	9.0
.00	3	103	22	35.5
.00	3	106	16	28.5
.00	3	100	13	28.0
.00	4	106	19	5.0
.00	4	105	26	35.0
.00	4	110	18	18.0
.00	4	104	23	33.0
.00	4	97	19	14.0
.00	2	100	14	13.0
.00	3	104	19	26.5
.00	3	107	22	37.5
.00	4	103	20	21.0
.00	2	110	33	38.5
.00	3	108	22	27.8
.00	3	112	22	28.5
.00	4	110	28	.5
.00	4	107	16	13.5

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.00	3	112	22	30.0
.00	3	112	29	37.5
.00	2	100	12	18.0
.00	4	102	9	2.5
.00	4	100	32	25.5
.00	2	94	13	14.5
.00	3	104	26	31.5
.00	3	106	23	37.5
.00	4	104	16	15.0
.00	3	107	25	42.0
.00	2	105	13	25.5
.00	3	110	17	34.0
.00	2	117	22	28.5
.00	4	104	26	33.3
.00	4	110	27	39.5
.00	4	110	19	32.5
.00	3	105	23	35.0
.00	4	110	22	34.0
.00	4	104	17	15.5
.10	3	110	17	37.5
.10	3	100	10	10.0
.10	3	100	18	27.5
.10	3	92	9	14.5
.10	4	100	16	16.0
.10	4	100	28	42.5
.10	4	70	4	2.0
.10	2	85	6	4.5
.10	3	103	21	34.5
.10	4	104	12	19.5
.10	3	100	7	15.0
.10	4	110	30	53.5
.10	3	103	28	30.5
.10	4	108	22	10.5
.10	4	109	22	20.5
.10	4	107	19	24.5
.10	3	104	6	15.0
.10	3	105	16	12.0
.10	4	95	20	27.5
.10	4	100	23	37.5
.10	4	103	25	43.0

EXPERIMENTO MM2

CONCENTRACION	MADUPLZ	ALTURA	N DE MIJUELOS	PESO
.10	4	98	22	28.0
.10	3	100	35	36.5
.10	3	102	24	26.5
.10	4	95	19	23.0
.10	3	100	25	33.5
.10	3	100	27	15.0
.10	3	95	17	36.0
.10	3	102	28	55.5
.10	4	105	24	14.5
.10	3	91	13	13.0
.10	4	100	27	33.0
.10	3	96	32	36.0
.10	3	90	8	37.0
.20	3	95	16	25.0
.20	3	105	24	27.0
.20	4	105	16	28.0
.20	3	100	23	24.0
.20	4	103	20	23.5
.20	3	90	16	9.0
.20	3	103	23	28.5
.20	4	105	16	10.5
.20	3	104	24	37.0
.20	4	96	10	18.0
.20	4	105	28	30.5
.20	4	102	32	26.5
.20	4	100	17	15.0
.20	4	94	21	22.0
.20	4	100	16	14.0
.20	4	102	19	33.0
.20	4	110	28	26.0
.20	4	100	21	27.0
.20	3	106	24	31.5
.20	3	98	21	25.0
.20	4	90	20	15.0
.20	4	98	20	30.0
.20	4	99	23	26.5
.20	3	99	22	29.5
.20	4	101	17	23.0
.20	4	103	23	20.0
.20	4	111	36	56.5

EXPERIMENTO MMZ

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.20	4	106	28	28.0
.20	4	104	16	14.0
.20	3	112	30	45.5
.20	4	106	18	12.5
.20	3	106	28	43.0
.20	3	110	28	39.0
.20	3	104	22	28.0
.20	3	104	25	47.0
.20	3	115	24	45.0
.20	4	112	8	15.5
.20	4	107	30	49.0
.20	4	110	26	38.5
.20	3	108	27	37.0
.20	3	106	20	33.0
.20	3	65	12	2.5
.20	3	104	22	22.5
.20	3	115	30	37.0
.20	4	110	18	17.5
.20	3	110	19	34.5
.20	3	107	26	51.0
.20	4	110	27	52.0
.20	4	103	5	55.0
.20	3	115	13	8.5
.20	3	100	24	41.0
.20	4	110	14	22.0
.20	3	110	25	33.0
.20	4	110	12	11.0
.20	5	99	18	12.0
.20	4	110	30	43.0
.30	4	116	19	18.5
.30	4	106	27	12.0
.30	4	100	20	2.5
.30	4	110	14	3.5
.30	4	119	23	13.0
.30	4	110	15	12.0
.30	4	108	26	23.5
.30	3	120	24	24.5
.30	4	125	25	19.0
.30	4	117	38	22.5
.30	4	100	16	10.5

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.30	4	124	33	27.5
.30	4	105	29	12.5
.30	3	100	12	10.6
.30	4	118	26	27.5
.30	3	108	11	15.5
.30	3	110	26	24.0
.30	3	100	17	4.0
.30	4	110	24	2.5
.30	5	115	28	17.0
.30	4	109	22	10.0
.30	4	104	22	12.0
.30	4	95	16	10.5
.30	5	125	32	12.0
.30	4	126	9	.5
.30	4	102	16	25.0
.30	4	102	16	18.0
.30	4	104	12	10.5
.30	4	103	19	22.0
.30	4	104	23	32.5
.30	3	92	8	11.5
.30	2	104	17	34.5
.30	3	99	13	19.5
.30	3	104	16	21.0
.30	3	98	12	37.0
.30	4	103	18	33.0
.30	3	104	22	22.5
.30	3	106	26	33.5
.30	4	104	20	25.5
.30	3	100	28	41.0
.30	2	60	16	8.0
.30	4	98	14	13.5
.30	3	107	21	25.5
.30	3	100	25	25.5
.30	3	100	26	12.5
.30	4	107	24	17.5
.30	3	90	5	1.5
.30	3	104	16	29.0
.30	2	100	20	27.0
.30	3	97	20	32.0
.40	3	95	23	31.5

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.40	4	77	6	.5
.40	2	90	12	13.5
.40	4	95	29	31.5
.40	3	60	8	10.5
.40	3	98	18	18.0
.40	3	70	1	24.5
.40	2	95	17	20.5
.40	3	94	26	38.5
.40	3	87	7	14.0
.40	3	95	33	40.0
.40	3	55	12	1.5
.40	4	104	16	11.0
.40	3	102	93	38.5
.40	2	85	8	5.5
.40	3	105	28	21.0
.40	4	104	12	16.0
.40	4	97	11	13.5
.40	3	102	27	17.0
.40	4	109	27	45.0
.40	4	105	19	42.5
.40	4	104	16	.5
.40	4	102	25	13.0
.40	3	112	20	20.0
.40	3	104	16	9.5
.40	3	95	19	15.0
.40	3	108	22	21.0
.40	3	102	19	26.5
.40	3	109	28	33.0
.40	3	104	22	20.0
.40	3	110	33	42.0
.40	4	106	23	42.0
.40	3	104	30	40.0
.40	3	109	28	31.5
.40	3	106	27	47.0
.40	3	105	14	14.0
.40	3	105	16	21.0
.50	3	85	10	9.0
.50	3	96	17	12.0
.50	3	92	14	21.5
.50	4	79	12	6.0

EXPERIMENTO MM2

CONCLNTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.50	3	97	21	34.0
.50	4	100	26	37.0
.50	4	103	22	29.0
.50	3	100	22	29.0
.50	3	100	19	22.0
.50	4	96	19	20.0
.50	4	90	23	31.5
.50	3	90	14	23.5
.50	3	103	28	41.0
.50	3	93	21	22.5
.50	3	75	5	4.0
.50	3	104	24	11.5
.50	3	104	35	32.5
.50	3	75	3	1.0
.50	4	90	8	2.5
.50	4	100	34	20.0
.50	3	100	26	31.0
.50	4	102	26	29.0
.50	4	105	22	17.0
.50	4	103	19	18.0
.50	3	100	22	24.5
.50	4	100	14	9.0
.50	3	103	24	11.0
.50	3	100	9	10.5
.50	3	107	19	20.0
.60	3	110	31	51.5
.60	3	105	7	13.5
.60	3	106	20	23.0
.60	4	105	16	18.0
.60	3	107	22	37.0
.60	3	105	20	38.0
.60	3	115	19	41.0
.60	3	104	20	27.0
.60	4	108	24	25.0
.60	3	100	9	13.0
.60	3	104	17	28.5
.60	3	102	22	25.5
.60	3	100	18	22.0
.60	3	103	18	18.0
.60	3	102	16	26.0

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.00	3	97	22	22.0
.00	4	110	26	39.0
.00	3	95	11	7.0
.00	4	115	18	22.0
.00	4	102	23	11.0
.00	4	100	10	15.5
.00	4	112	23	36.5
.00	4	107	28	24.5
.00	3	102	24	32.5
.00	4	104	42	35.0
.00	3	102	17	.5
.00	4	100	8	20.5
.00	4	107	33	25.5
.00	4	105	26	29.5
.00	4	103	37	30.5
.00	4	109	30	18.0
.00	3	98	21	20.0
.00	4	100	28	16.0
.00	3	103	27	25.0
.00	4	104	26	29.0
.00	4	100	27	25.5
.00	2	100	22	20.0
.00	4	100	30	21.0
.00	3	109	31	21.5
.00	3	106	23	12.0
.00	3	106	32	23.0
.00	3	104	37	21.0
.00	3	100	23	20.5
.10	4	105	29	33.0
.10	3	101	12	23.0
.10	3	105	23	36.5
.10	3	95	8	10.5
.10	3	95	16	20.5
.10	3	94	12	19.0
.10	3	108	22	24.5
.10	3	100	15	10.5
.10	4	104	24	40.5
.10	4	102	27	32.0
.10	2	82	26	10.0
.10	4	100	20	31.0

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N° DE HIJUELOS	PESO
.10	4	104	26	23.5
.10	3	105	21	26.0
.10	3	100	22	22.0
.10	3	102	18	7.5
.10	4	103	21	17.0
.10	4	105	16	18.5
.10	2	100	22	31.0
.10	4	106	24	33.0
.10	3	110	22	26.5
.10	3	105	26	24.0
.10	3	106	27	33.0
.10	3	102	28	27.0
.10	3	100	22	27.0
.10	3	100	12	30.5
.10	3	104	18	39.5
.10	1	91	10	45.0
.10	3	103	12	11.0
.10	3	100	12	12.5
.10	3	106	20	29.5
.10	3	103	25	30.5
.10	3	100	22	34.5
.10	3	100	19	25.0
.10	3	107	24	35.0
.10	2	97	13	33.0
.10	3	104	19	32.5
.10	3	107	29	36.5
.10	3	103	32	35.0
.10	2	105	24	49.0
.10	2	98	11	9.0
.10	2	90	16	23.0
.10	3	102	29	39.0
.10	3	105	32	41.0
.20	4	100	22	23.0
.20	3	110	24	30.0
.20	4	109	30	28.0
.20	2	93	17	6.0
.20	3	107	34	17.5
.20	1	75	13	10.0
.20	3	90	26	10.5
.20	4	104	22	12.0

EXPERIMENTO MM2

CONCENTRACION	MADURLZ	ALTURA	N DE HIJUELOS	PEISO
.20	4	110	38	26.0
.20	4	102	16	31.5
.20	4	105	37	26.0
.20	2	75	28	15.0
.20	4	60	3	.5
.20	1	78	9	7.5
.20	3	100	29	37.0
.20	4	107	22	18.0
.20	4	110	23	25.5
.20	3	110	34	25.0
.20	2	80	13	.5
.20	4	111	23	9.5
.20	4	112	25	10.0
.20	3	109	20	26.5
.20	4	100	27	16.0
.20	3	106	33	28.0
.20	3	105	23	18.5
.20	4	93	24	18.5
.20	3	108	25	52.0
.20	3	104	23	45.0
.20	4	110	28	47.0
.20	3	108	30	47.0
.20	3	108	26	56.5
.20	4	109	27	44.0
.20	2	113	29	58.5
.20	2	115	17	35.5
.20	3	111	27	29.0
.20	3	111	30	44.5
.20	4	110	17	57.0
.20	4	110	23	23.0
.20	4	108	25	45.0
.20	4	110	26	43.0
.20	3	102	20	42.0
.20	3	105	28	37.5
.20	4	105	18	21.5
.20	3	105	25	43.0
.20	4	103	26	35.5
.20	3	109	26	53.0
.20	4	107	18	8.0
.20	3	107	17	30.0

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.20	2	104	19	15.0
.20	3	103	19	27.5
.20	3	108	12	15.0
.20	2	105	21	32.0
.20	3	106	10	27.0
.20	2	103	14	32.0
.20	3	107	19	33.0
.30	3	100	25	43.0
.30	2	99	13	22.0
.30	1	80	8	9.0
.30	3	95	11	20.0
.30	3	105	19	35.5
.30	3	107	13	24.0
.30	2	95	10	26.0
.30	3	96	7	13.0
.30	3	101	23	38.0
.30	2	100	22	29.5
.30	3	101	24	44.0
.30	3	97	22	30.5
.30	2	95	16	25.0
.30	2	105	21	32.0
.30	2	102	19	35.5
.30	4	95	12	21.5
.30	3	93	9	19.0
.30	3	100	18	26.5
.30	3	94	18	31.0
.30	3	90	18	24.0
.30	3	101	20	28.0
.30	3	95	16	25.0
.30	3	95	8	12.0
.30	3	95	15	17.5
.30	3	90	12	14.5
.30	3	80	10	8.0
.30	3	120	13	17.5
.30	3	100	11	19.5
.30	2	87	12	13.5
.30	3	99	11	9.0
.30	2	95	9	15.5
.30	1	95	15	15.0
.30	2	104	17	30.0

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.30	3	106	13	23.0
.30	3	104	23	27.0
.30	3	108	18	35.0
.30	4	115	19	31.0
.30	3	109	22	45.5
.30	3	109	15	26.0
.30	4	105	18	27.5
.30	4	106	19	29.5
.30	4	107	24	30.0
.30	3	112	22	37.0
.30	3	110	21	32.0
.30	4	105	20	27.0
.30	4	105	23	32.5
.30	3	96	5	2.0
.30	3	104	17	34.5
.30	3	110	15	34.0
.30	3	110	24	40.0
.30	3	103	11	16.0
.30	3	110	28	19.5
.30	4	109	19	29.5
.30	2	108	17	31.0
.30	3	96	10	19.0
.30	3	108	17	33.0
.30	3	109	24	38.0
.30	3	110	24	30.0
.30	3	95	11	14.0
.30	4	100	24	27.0
.30	3	109	22	28.0
.30	2	105	18	25.0
.30	4	101	16	28.5
.40	3	100	18	18.0
.40	3	95	9	8.5
.40	3	102	23	25.0
.40	4	95	15	13.0
.40	3	100	18	30.5
.40	3	104	26	34.5
.40	3	103	19	28.5
.40	3	105	16	20.5
.40	3	102	24	28.5
.40	4	103	19	19.5

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.40	4	101	29	25.5
.40	3	100	19	16.5
.40	4	95	24	16.0
.40	3	103	27	29.0
.40	3	106	17	30.5
.40	3	103	16	21.5
.40	3	105	18	23.0
.40	4	105	35	36.0
.40	4	107	24	47.5
.40	2	105	18	17.5
.40	3	106	26	35.0
.40	3	108	36	39.0
.40	3	102	34	34.5
.40	4	108	24	26.5
.40	3	103	16	25.0
.40	3	108	25	19.5
.40	3	105	11	13.0
.40	3	107	19	31.0
.40	4	105	9	11.0
.40	3	103	21	22.0
.40	3	100	6	6.5
.40	3	98	18	17.0
.40	3	81	14	12.0
.40	2	82	8	3.0
.40	3	97	10	12.0
.40	3	100	23	22.0
.40	3	98	9	5.5
.40	2	85	4	1.5
.40	4	100	11	6.0
.40	4	100	18	16.0
.40	4	103	17	12.0
.40	4	97	8	2.0
.40	4	104	33	22.5
.40	3	109	27	22.0
.40	3	106	16	17.0
.40	3	102	19	24.0
.40	3	107	19	19.5
.40	3	108	30	38.0
.40	3	70	5	2.0
.40	3	86	6	1.5

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJULLOS	PESO
.40	4	82	17	23.0
.40	3	95	16	14.0
.40	2	65	11	6.0
.40	3	96	23	15.0
.50	3	102	27	30.0
.50	3	104	20	23.5
.50	3	108	23	33.0
.50	3	105	12	15.5
.50	3	104	22	27.5
.50	2	100	22	11.0
.50	3	100	21	36.5
.50	3	104	19	32.5
.50	4	90	13	2.0
.50	3	96	18	15.0
.50	3	103	13	31.0
.50	2	103	17	22.5
.50	3	105	15	18.0
.50	3	106	24	22.0
.50	3	108	25	32.0
.50	3	96	18	9.0
.50	2	97	13	12.0
.50	3	110	23	30.0
.50	3	105	12	11.0
.50	3	92	6	12.5
.50	3	105	24	23.0
.50	4	103	17	7.0
.50	3	98	17	31.0
.50	3	98	22	35.0
.50	3	108	25	28.0
.50	3	97	23	27.5
.50	3	106	19	25.5
.50	2	95	18	30.5
.50	3	100	17	19.5
.50	3	109	20	28.5
.50	4	107	26	40.0
.50	4	106	21	40.0
.50	4	106	16	21.5
.50	4	100	28	30.0
.50	3	100	20	23.0
.50	3	97	19	20.5

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PF50
.50	3	102	22	27.5
.50	4	103	28	28.5
.50	3	104	23	27.5
.50	3	100	19	26.5
.50	3	99	17	22.0
.50	3	100	18	27.0
.50	3	103	16	27.0
.50	4	109	19	34.5
.50	4	94	4	6.0
.50	3	90	7	7.0
.50	4	91	16	11.0
.50	3	102	24	32.0
.50	3	92	11	12.0
.60	3	105	9	16.0
.60	3	107	22	33.0
.60	2	110	23	38.5
.60	2	109	12	23.5
.60	3	110	10	49.0
.60	3	110	24	42.0
.60	3	100	12	21.5
.60	3	106	25	50.0
.60	4	110	23	36.5
.60	4	108	20	25.5
.60	3	104	26	23.0
.60	3	110	25	38.5
.60	3	107	24	37.0
.60	4	113	18	26.5
.60	3	106	28	44.0
.60	3	108	18	13.0
.60	3	102	16	12.0
.60	3	97	26	26.0
.60	3	104	22	36.0
.60	3	106	26	40.5
.60	3	110	26	42.5
.60	3	106	25	33.0
.60	3	107	22	42.0
.60	3	100	17	26.0
.60	4	100	17	15.5
.60	4	107	17	23.5
.60	3	105	10	18.0

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PFSO
.00	4	107	17	28.0
.00	4	109	16	18.0
.00	4	103	10	9.0
.00	3	110	17	20.5
.00	4	106	16	31.5
.00	4	110	19	26.5
.00	4	103	18	26.5
.00	4	108	13	18.0
.00	4	107	16	31.5
.00	4	106	16	23.5
.00	4	106	16	25.0
.00	4	105	16	20.0
.00	3	107	17	21.5
.00	2	108	18	32.5
.00	3	111	19	22.5
.00	4	103	16	23.0
.00	4	100	17	18.0
.00	4	115	19	35.5
.00	3	103	19	41.0
.00	4	106	16	34.0
.00	3	107	5	20.5
.00	3	105	16	26.0
.00	3	105	19	26.5
.00	3	108	18	27.0
.00	4	93	6	17.5
.00	4	109	12	22.5
.00	4	103	20	40.0
.00	4	110	16	23.0
.10	3	90	19	8.0
.10	4	97	16	9.5
.10	4	100	17	13.5
.10	4	94	16	14.6
.10	4	94	13	15.5
.10	3	102	11	14.0
.10	4	102	18	13.5
.10	4	100	20	15.5
.10	4	100	17	15.0
.10	4	95	16	11.0
.10	4	103	19	8.0
.10	4	90	14	5.0

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.10	4	98	17	14.5
.10	4	85	10	4.5
.10	4	103	22	12.0
.10	4	100	20	18.5
.10	4	103	23	20.0
.10	4	107	24	16.0
.10	4	100	17	16.0
.10	3	104	18	24.5
.10	4	100	19	6.5
.10	3	90	7	6.5
.10	4	110	34	17.0
.10	4	95	23	17.0
.10	4	100	28	30.5
.10	4	106	16	17.5
.10	3	105	19	17.0
.10	3	100	12	18.0
.10	3	102	20	31.0
.10	3	110	23	25.0
.10	4	100	14	12.5
.10	3	105	19	35.0
.10	4	104	19	20.5
.10	3	103	19	27.0
.10	2	105	16	21.5
.10	2	95	17	17.5
.10	4	100	16	18.0
.10	3	95	11	8.0
.10	3	100	25	19.5
.10	1	67	6	8.0
.10	3	60	10	12.0
.10	2	90	12	9.0
.10	3	95	7	8.5
.10	3	106	23	41.5
.10	4	109	29	38.0
.10	4	104	19	20.5
.10	3	100	19	27.5
.10	3	104	21	38.0
.10	3	100	14	12.5
.10	3	105	18	30.0
.10	3	103	28	26.5
.10	3	103	16	15.5

EXPERIMENTO MMZ

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.20	3	95	15	13.0
.20	3	90	20	25.5
.20	4	100	28	29.0
.20	3	77	16	9.0
.20	3	98	23	15.5
.20	3	104	30	33.5
.20	4	97	18	16.0
.20	3	100	28	43.0
.20	3	97	12	14.0
.20	3	100	26	30.0
.20	3	91	25	23.0
.20	4	94	25	19.5
.20	3	100	13	15.5
.20	3	96	22	13.0
.20	3	95	23	25.0
.20	4	94	27	17.0
.20	3	85	13	7.5
.20	1	85	14	10.0
.20	2	65	5	2.0
.20	3	96	14	13.0
.20	3	102	16	25.0
.20	3	100	27	36.5
.20	3	95	19	21.0
.20	3	102	22	51.0
.20	2	94	18	15.5
.20	3	94	10	8.5
.20	4	100	20	27.5
.20	3	98	13	9.0
.20	3	100	7	8.5
.20	3	99	14	21.5
.20	3	106	32	52.5
.20	4	92	17	19.5
.20	4	106	24	37.0
.20	3	100	33	42.0
.20	4	106	26	35.0
.20	3	103	26	38.5
.20	3	108	13	17.0
.20	3	100	26	40.0
.20	3	108	27	33.5
.20	4	110	29	24.0

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.20	4	103	33	31.0
.20	4	102	20	15.0
.20	4	110	39	41.0
.20	4	102	34	28.0
.20	4	102	26	31.0
.20	3	97	21	31.0
.20	3	100	8	16.5
.20	4	103	21	33.0
.20	3	104	21	36.0
.20	4	92	17	20.5
.20	4	109	23	32.0
.20	4	102	18	20.5
.20	4	100	20	40.0
.20	4	103	17	32.0
.20	4	104	20	44.5
.20	4	103	26	45.0
.20	3	97	26	36.0
.20	4	103	21	26.0
.20	3	98	16	17.5
.30	3	100	16	16.0
.30	3	100	19	30.0
.30	3	100	25	41.0
.30	3	103	23	41.5
.30	3	100	19	28.5
.30	3	102	21	34.5
.30	2	100	10	7.5
.30	2	60	18	11.0
.30	4	103	22	44.5
.30	3	108	17	39.5
.30	3	102	26	47.0
.30	3	102	22	40.0
.30	3	102	19	30.5
.30	3	103	20	23.5
.30	3	97	19	26.5
.30	3	35	14	21.0
.30	4	90	17	15.5
.30	4	100	12	6.5
.30	3	102	23	21.0
.30	3	102	19	31.0
.30	3	95	19	25.5

EXPERIMENTO MMZ

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.30	4	90	11	16.0
.30	3	93	18	28.0
.30	3	93	9	15.0
.30	4	102	22	23.0
.30	3	92	18	24.0
.30	4	100	14	27.5
.30	3	100	22	36.0
.30	4	103	24	50.5
.30	4	97	16	18.5
.30	3	95	22	24.0
.30	3	78	16	19.0
.30	4	100	26	41.0
.30	3	99	15	18.0
.30	3	96	20	38.0
.30	3	102	20	41.0
.30	3	100	22	32.5
.30	4	102	27	43.0
.30	4	104	36	36.5
.30	4	102	18	22.5
.30	3	107	20	35.5
.30	3	100	26	47.5
.30	4	98	25	27.5
.30	4	101	22	37.5
.30	4	107	22	40.5
.30	3	100	20	37.0
.30	3	100	20	38.0
.30	3	100	14	26.0
.30	3	78	12	15.0
.30	2	100	19	29.0
.30	3	93	16	23.0
.30	3	104	26	36.0
.40	4	101	12	20.0
.40	4	102	15	18.5
.40	3	75	10	14.0
.40	4	103	16	31.0
.40	3	107	9	16.0
.40	3	107	14	22.5
.40	3	110	16	32.0
.40	4	100	13	28.5
.40	3	106	20	43.5

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.40	3	105	16	26.0
.40	3	100	12	30.5
.40	3	101	11	17.0
.40	3	100	18	30.5
.40	4	100	12	26.0
.40	3	99	12	22.5
.40	3	90	18	17.0
.40	3	96	10	17.0
.40	3	96	11	6.5
.40	4	103	17	16.5
.40	2	60	9	5.5
.40	3	105	17	34.0
.40	3	105	17	22.5
.40	2	105	10	13.0
.40	3	103	14	24.5
.40	3	97	10	18.0
.40	3	100	16	21.0
.40	4	100	17	35.0
.40	4	106	14	31.0
.40	3	108	16	32.5
.40	3	108	21	38.0
.40	4	109	8	18.5
.40	2	100	12	27.0
.40	3	103	14	26.5
.40	4	100	11	20.5
.40	4	117	25	31.0
.40	3	104	24	23.0
.40	2	60	6	34.5
.40	2	97	14	26.0
.40	2	90	12	14.0
.40	3	100	13	36.0
.40	2	106	17	26.5
.40	3	100	14	14.0
.40	4	104	17	32.0
.40	4	100	14	35.5
.40	3	90	10	16.0
.40	2	95	14	22.0
.40	3	100	27	34.0
.40	4	102	15	24.5
.40	4	100	8	1.5

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.40	4	102	16	23.5
.40	4	106	25	30.0
.40	3	95	16	19.0
.40	3	100	15	32.0
.40	3	93	14	15.0
.40	3	75	21	4.0
.40	1	75	3	3.0
.40	2	102	16	30.5
.40	3	100	18	30.0
.40	4	102	17	36.5
.40	3	104	22	43.5
.40	3	100	10	27.0
.40	3	100	18	31.5
.40	3	99	24	22.5
.40	2	90	15	22.0
.40	3	90	17	30.0
.40	3	97	7	13.0
.40	4	100	18	30.0
.40	3	91	12	15.0
.50	4	88	11	14.5
.50	3	92	13	19.5
.50	3	100	26	31.0
.50	2	98	14	13.0
.50	4	95	16	25.0
.50	4	97	24	28.5
.50	3	104	27	24.5
.50	4	97	16	4.5
.50	3	95	12	22.5
.50	4	98	12	9.0
.50	5	102	16	3.0
.50	5	100	33	14.0
.50	4	90	3	4.0
.50	3	99	18	17.0
.50	2	100	8	9.5
.50	3	75	2	2.0
.50	3	60	17	7.0
.50	3	105	17	22.5
.50	4	100	20	22.0
.50	4	105	16	15.0
.50	4	65	3	22.0

EXPERIMENTO MM2

CONCENTRACION	MADUREZ	ALTURA	N DE HIJUELOS	PESO
.50	5	100	19	8.5
.50	4	103	17	29.5
.50	4	103	9	13.0
.50	4	90	15	16.0
.50	4	115	11	15.0
.50	3	100	12	30.0
.50	4	85	15	22.5
.50	3	97	18	36.5
.50	2	91	16	26.0
.50	2	92	23	25.0
.50	3	97	15	22.0
.50	3	100	28	44.0
.50	3	96	6	5.0
.50	3	100	24	30.0
.50	3	91	18	12.5
.50	3	96	11	13.0
.50	2	99	14	11.0
.50	2	104	18	27.0
.50	3	96	16	18.5
.50	3	88	18	18.0
.50	3	100	18	25.5
.50	3	92	12	24.5
.50	2	102	22	31.5
.50	4	97	19	20.5
.50	2	99	7	13.5
.50	3	100	19	24.0
.50	3	97	12	21.0
.50	3	100	16	24.0
.50	4	96	13	17.0
.50	2	90	11	12.0
.50	2	100	19	25.0
.50	2	85	11	12.0
.50	2	85	12	55.0
.50	3	75	13	9.5
.50	3	99	23	26.0
.50	3	85	12	9.0
.50	3	98	29	26.0

1.4.2 APLICACION DEL ANALISIS DISCRIMINANTE A LA GENERACION MM2

Recuerde que la generación MM2 son las plantas que provienen de las semillas que se originaron de las plantas cuyas - semillas fueron tratadas a distintas concentraciones de mutágeno⁽¹⁾. Como hemos mencionado anteriormente, el estudio de esta segunda generación es con el objeto de analizar las modificaciones que sufren las generaciones posteriores de las semillas tratadas con MMS. La aplicación del Análisis Discriminante dentro de esta fase de la investigación se llevó a cabo tomando en --- cuenta seis grupos definidos de la siguiente forma:

GRUPO	TRATAMIENTO AL CUAL FUERON SOMETIDAS LAS SEMILLAS QUE DIERON ORIGEN A LAS PLANTAS DE QUIENES PROVIENEN LAS ACTUALMENTE ESTUDIADAS.
1	0
2	.1
3	.2
4	.3
5	.4
6	.5

(1) Estas semillas ya no son sometidas a ninguna concentración MMS.

Se espera que estos grupos difieran en las variables:

- | | |
|----|---|
| L1 | La madurez de la planta a la apa
rición de la semilla. |
| L2 | Altura de la planta (cm). |
| L3 | Número de hijuelos |
| L4 | Peso del grano en gramos. |

1.4.3 RESULTADOS OBTENIDOS PARA LA GENERACION MM2

FILE NCNAME (CREATION DATE = 07/30/80)

GROUP COUNTS

	GROUP 1	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6	TOTAL
COUNT	157.0000	130.0000	169.0000	165.0000	160.0000	136.0000	517.0000

MEANS

	GROUP 1	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6	TOTAL
L1	3.3503	3.2769	3.3609	3.2242	3.1563	3.2132	3.2650
L2	105.0382	100.0308	101.7811	101.2727	98.6256	97.2147	100.2157
L3	20.2102	19.1846	21.2533	18.6848	17.6825	17.7228	19.2228
L4	26.0363	23.1200	27.3580	24.7339	22.4594	20.9853	24.2580

113

STANDARD DEVIATIONS

	GROUP 1	GROUP 2	GROUP 3	GROUP 4	GROUP 5	GROUP 6	TOTAL
L1	0.6392	0.6707	0.6857	0.6658	0.6199	0.6601	0.6497
L2	4.4832	6.9315	9.1062	10.5451	10.2471	8.3958	8.5763
L3	6.4926	6.6330	6.7268	6.6693	5.2806	6.5983	6.5823
L4	10.7797	11.4320	13.4214	10.2769	11.0815	10.1171	11.2549

WILKS' LAMBDA (U-STATISTIC) AND UNIVARIATE F-RATIO WITH 3 AND 911 DEGREES OF FREEDOM

VARIABLE WILKS' LAMBDA F

L1	0.9868	2.4308
L2	0.9281	14.1227
L3	0.9527	9.0516
L4	0.9648	6.6408

WITHIN GROUPS COVARIANCE MATRIX

	L1	L2	L3	L4
L1	0.4318			
L2	1.3957	75.3554		
L3	0.7180	24.8100	49.4106	
L4	-0.2003	38.6798	41.7473	129.5221

WITHIN GROUPS CORRELATION MATRIX

	L1	L2	L3	L4
L1	1.0000			
L2	0.2447	1.0000		
L3	0.1555	0.4066	1.0000	
L4	-0.0268	0.3915	0.5219	1.0000

TOTAL COVARIANCE MATRIX

	L1	L2	L3	L4
L1	0.4352			
L2	1.5292	80.7531		
L3	0.6232	27.3040	51.5822	
L4	-0.0644	42.7860	44.6727	133.5100

CUADRO 12

VARIABLE ENTERED ON STEP NUMBER 3.4 L4

MILK² LAMBDA 0.69345 APPROXIMATE F 6.96674 DEGREES OF FREEDOM 15 SIGNIFICANCE 0.000
 RACES V 105.59466 CHANGE IN V 5.62967 2509.75 0.344

F MATRIX - DEGREES OF FREEDOM: 3, 909

GROUP	2	GROUP 1	GROUP 2	GROUP 3	GROUP 4	GROUP 5
GROUP	2	8.25618				
GROUP	3	5.13310	4.61491			
GROUP	4	5.29489	1.68646	6.33817		
GROUP	5	14.50718	1.44832	10.69312	2.72672	
GROUP	6	18.51888	2.33264	12.51924	5.54981	0.67533

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

VARIABLE	ENTRY CRITERION	F TO REMOVE
L2	14.12271	5.82669
L3	5.13548	1.98567
L4	1.09826	1.09826

F LEVEL INSUFFICIENT FOR FURTHER COMPLETION

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

VARIABLE	TOLERANCE	F TO ENTER	ENTRY CRITERION
L1	0.60508	0.91417	0.04930

CUADRO 13

115

SUMMARY TABLE

STEP NUMBER	VARIABLE ENTERED	REMOVED	F TO ENTER OR REMOVE	NUMBER INCLUDED	MILK ² LAMBDA	SIG.	RACES V	CHANGE IN RACES V	SIG. OF CHANGE
1	L2		14.12271	1	0.52806	0.000	70.61353	70.61353	0.000
2	L3		5.13548	2	0.55289	0.000	100.28470	29.67117	0.000
3	L4		1.09826	3	0.69345	0.000	105.69466	5.62997	0.344

DISCRIMINANT FUNCTION	EIGENVALUE	RELATIVE PERCENTAGE	CANONICAL CORRELATION	FUNCTIONS DERIVED	MILK ² LAMBDA	CHI-SQUARE	DF	SIGNIFICANCE
1	0.68249	71.07	0.274	1	0.8973	102.692	134	0.000
2	0.02960	25.469	0.170	2	0.9673	30.274	3	0.000
3	0.00378	3.23	0.061	3	0.9863	3.424	3	0.321

REMAINING COMPUTATIONS WILL BE BASED ON 3 DISCRIMINANT FUNCTIONS

CUADRO 14

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2	FUNC 3
L2	0.79291	0.76846	-0.27154
L3	0.21135	-0.95764	-0.74176
L4	0.17507	-0.19427	1.19438

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2	FUNC 3
L2	0.08824	0.08551	-0.03026
L3	0.02943	-0.13334	-0.10328
L4	0.01515	-0.01681	0.10337
CONSTANT	-9.82936	-5.64758	2.53067

CUADRO 15

CENTROIDS OF GROUPS IN REDUCED SPACE

		FUNC 1	FUNC 2	FUNC 3
GROUP	1	0.42798	0.20219	-0.04392
GROUP	2	-0.08822	-0.04024	-0.08752
GROUP	3	0.21014	-0.32301	0.01726
GROUP	4	0.03111	0.10547	0.09222
GROUP	5	-0.26706	0.05361	0.04342
GROUP	6	-0.39442	0.01541	-0.05054

CUADRO 16

PREDICTION RESULTS -

ACTUAL GROUP		NO. OF CASES	PREDICTED GROUP MEMBERSHIP					
-----			GP. 1	GP. 2	GP. 3	GP. 4	GP. 5	GP. 6
GROUP 1	157.	70. 44.6%	13. 8.3%	46. 29.3%	12. 7.6%	6. 3.8%	10. 6.4%	
GROUP 2	130.	21. 16.2%	11. 8.5%	41. 31.5%	13. 10.0%	6. 4.6%	38. 29.2%	
GROUP 3	169.	39. 23.1%	6. 3.6%	78. 46.2%	8. 4.7%	5. 3.0%	33. 19.5%	
GROUP 4	165.	52. 31.5%	3. 1.8%	39. 23.6%	22. 13.3%	12. 7.3%	37. 22.4%	
GROUP 5	160.	34. 21.3%	9. 5.6%	33. 20.6%	32. 20.0%	9. 5.6%	43. 26.9%	
GROUP 6	138.	19. 14.0%	10. 7.2%	33. 24.0%	15. 11.0%	10. 7.2%	42. 30.0%	

117

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 26.6%

CUADRO 17

1.4.4 ANALISIS DE LOS RESULTADOS OBTENIDOS PARA LA GENERACION MM2

Las variables introducidas dentro del análisis son:

- L2 altura de la planta
- L3 número de hijuelos
- L4 peso del grano de la planta

L1 la variable correspondiente a la madurez de la planta a la aparición de la semilla no fue incluida.

En forma general las variables altura, número de hijuelos y peso incrementaron el promedio obtenido en la generación MM2 con respecto a la generación MM1 (cuadro #11).

La altura máxima se obtuvo en el grupo testigo con un promedio de 105.04 cm y la mínima a la concentración de 0.5 --- p/vol/hora con un media de 97.51 cm (cuadro #11).

La variable correspondiente al número de hijuelos presentó la media mínima en el grupo 6 con 17 hijuelos (cuadro 11).

La variable peso alcanzó su valor mínimo de 20.99 y -- máximo de 27.35 grs en los grupos 6 y 3 respectivamente.

Utilizando las variables L2, L3 y L4 dentro del análisis, los grupos que presentan una mayor diferenciación son:
 El grupo testigo y los grupos 5 y 6.
 El grupo 3 y los grupos 5 y 6 (cuadro 13).

Se derivaron solamente tres funciones discriminantes. Las funciones discriminantes estandarizadas son:

$$DI = 0.79291 L2 + 0.21135 L3 + 0.17507 L4$$

$$D2 = 0.7684 L2 - 0.95764L3 - 0.19427 L4$$

$$D3 = -0.27194L2 - 0.74176L3 + 1.19438 L4$$

Las variables que proporcionan una mayor contribución a la discriminación se presentan a continuación:

FUNCIÓN	VARIABLE
D1	L2
D2	L3
D3	L4

La clasificación de un nuevo caso se efectúa a partir de las funciones discriminantes no estandarizadas:

$$D1 = 0.08824 L2 + 0.02943 L3 + 0.01515 L4 - 9.82936$$

$$D2 = 0.08551 L2 - 0.13334 L3 - 0.01681 L4 - 5.64758$$

$$D3 = -0.03026 L2 - 0.10328 L3 + 0.10337 L4 + 2.53067$$

La importancia relativa de estas funciones son para D1 de 71.07%, para D2 de 25.69% y de 3.23% para D3. Las correlaciones canónicas asociadas a cada una de las funciones es -- respectivamente de 0.276, 0.170 y 0.061 que son demasiado bajas, por lo que nuevamente se espera que la clasificación de un nuevo caso dentro de la generación MM2 no sea lo suficientemente buena (cuadro 14).

Los resultados relativos de la clasificación de nuevos casos se encuentran en la tabla resumen del cuadro 17 donde se reporta:

1) Los grupos mejor clasificados son el grupo testigo -- con 44.6% y el grupo 3 con 46% de casos clasificados correctamente, puede observarse que estos eran los grupos que presentaban las mayores diferencias con respecto de los otros (cuadro

13).

2) Los grupos peor clasificados fueron el grupo 5 y el grupo 2 con 5.6% y 8.5% de casos clasificados correctamente.

3) Las mayores fuentes de error se encuentran en:

a) La clasificación de individuos del grupo 4 como perteneciente al grupo 1 con 31.5% de los casos.

b) La clasificación de individuos del grupo 1 como pertenecientes al grupo 3 con 29.3% de los casos

c) La clasificación de individuos del grupo 2 como pertenecientes al grupo 3 con 31.5% y el grupo 6 con 29.2% de los casos.

4) Tomando en cuenta el número total de casos solamente se clasificó correctamente el 26.06% de ellos, el cual es un porcentaje muy bajo de buena clasificación.

Con respecto a esta tabla resumen de características debe hacerse la misma aclaración que en la sección 1.3.5

TABLA RESUMEN DE CARACTERISTICAS

GRUPO	CONCENTRACION CORRESPONDIENTE *	ALTURA *	# DE HIJUELOS *	PESO *
1	0	Baja	Pequeño	Pequeño
2	.1	Baja	Grande	Grande
3	.2	Grande	Grande	Grande
4	.3	Pequeña	Pequeño	Pequeño
5	.4	Pequeña	Pequeño	Pequeño
6	.5	Pequeña	---	---

* Estos valores son los estandarizados.

1.5 CONCLUSIONES

Durante el estudio realizado se ha destacado primordialmente el hecho de que existe un gran parecido entre los grupos, es decir que tienen un gran traslape. Una consecuencia del traslape de grupos es la baja habilidad de las funciones derivadas para la discriminación y de ello se desprende directamente el gran porcentaje de error en la clasificación de un nuevo caso.

El objetivo de nuestra investigación es evaluar los efectos del metil-metano-sulfonato en las plantas de Triticale. El traslape que se observa en los grupos nos permite concluir que las alteraciones que sufren las plantas que provienen de las semillas tratadas a las diferentes concentraciones de mutágeno son casi nulas y esto prevalece al menos en una generación posterior (generación MM2).

Las mayores modificaciones que pueden obtenerse respecto al grupo testigo se logran aplicando la concentración de mutágeno correspondiente al grupo que presenta la mayor separación posible.

A partir de los resultados arrojados durante el análisis el investigador puede tomar alguna determinación, ya sea utilizar solamente aquellas concentraciones que producen las mayores alteraciones y de alguna de ellas, escoger aquella que produzca las características deseadas, o bien no utilizar este mutágeno y sustituirlo por otro que mejore los resultados.

Al tener 6 grupos se aumenta la similitud de grupos vecinos, quizá se colapsan algunos grupos, teniendo: testigo, dosis bajas, dosis altas, se tenga una mayor habilidad discriminadora con correlaciones canónicas mayores.

C A P I T U L O 2

PRIMER EJEMPLO DE LA APLICACION
DEL ANALISIS DISCRIMINANTE A LA MEDICINA2.1 ESTUDIO SOBRE LAS CIFRAS DE AMILASA Y LIPASA SERICA EN NI--
ÑOS LACTANTES Y ESCOLARES MEXICANOS.

La lipasa en presencia de sales biliares, desdobra las grasa en ácidos grasos y gliserol. La amilasa actúa sobre el almidón hidrolizándolo y convirtiéndolo a la forma de maltosa.

La amilasa se encuentra presente en el páncreas, glándulas salivales, hígado, músculos, tejido adiposo, saliva, sangre, orina, heces, leche y semen. Se detecta por primera vez en el suero de los niños entre las edades de 1 y 2 meses, y al año de edad alcanzan el valor inferior de las cifras normales del adulto. En la pancreatitis aguda, la amilasa en el suero está casi siempre elevada dentro de las 2 a 12 horas del comienzo de la enfermedad y alcanza su máximo de uno a dos días y puede regresar a las cifras normales casi siempre en dos o tres días siguientes y en forma menos frecuente de cuatro a seis días, aunque en casos raros puede persistir el aumento después de 165 --días. Las elevaciones de la amilasa en el suero pueden producirse en varias enfermedades caracterizadas por edema e inflama-ción del páncreas, aumento de la presión intraductal o pérdida de enzimas a través de la pared intestinal en el líquido peritoneal.

La lipasa sérica sufre incrementos en sus niveles si los pacientes padecen enfermedades de tipo pancreático, ictericia obstructiva por cálculos, tumores y cirrosis del hígado, hepatitis por virus, obstrucción intestinal y peritonitis.

Se han efectuado estudios en el extranjero relativos a la actividad amilasa y lipasa séricas en niños, y son estos los valores citados en la literatura médica, a partir de los cuales se realizan comparaciones de patrones de normalidad. El Dr. Miguel Piñero León especialista en pediatría llevó a cabo una investigación para determinar las cifras de amilasa y lipasa en el suero de niños mexicanos lactantes y escolares con el fin de dar un panorama más apegado a la realidad infantil de nuestro país.

La determinación de las enzimas se efectuó en niños asistentes a la consulta externa de pediatría DIF, a causa de -- que no se disponía de niños completamente sanos para tal investigación.

2.2 INVESTIGACION REALIZADA PARA LA DETERMINACION DE LAS CIFRAS DE AMILASA Y LIPASA SERICAS EN NIÑOS MEXICANOS.

Se tomaron 400 niños de la consulta externa de pediatría del Hospital del Niño DIF (diversas especialidades). Se hizo la toma de sangre para biometría u otro estudio con plasma o suero, evitando que estos niños fueran pacientes con patología renal, pancreática o tuvieran alguna enfermedad que pudiera alterar sus niveles de amilasa y lipasa en su suero.

El grupo de 400 niños estaba integrado de la siguiente manera:

200 escolares de 7 a 12 años cumplidos de edad (100 -- hombres y 100 mujeres)

200 lactantes de 0 a 23 meses de edad (100 hombres y -

100 mujeres).

Se resumió a cada paciente su edad, sexo, talla, actividad amilasa, actividad lipasa y su diagnóstico clínico.

OBJETIVO:

Analizar las cifras de la actividad amilasa y lipasa, y de esta manera obtener marcos de referencia para -- comparar las determinadas en un paciente dado, pudiendo situarlo dentro de la anormalidad o anormalidad, - especialmente en situaciones de difícil diagnóstico.

2.3 DATOS RECOPIRADOS DURANTE LA INVESTIGACION.

LACTANTES

PACIENTE	EDAD	SEXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
1	.50	M	9.00	74.00	103.70	.000
2	1.83	M	12.00	85.00	74.07	.000
3	.83	M	9.00	73.00	44.44	.000
4	.16	F	5.40	54.00	35.55	.390
5	.41	F	5.50	60.00	29.02	.000
6	1.91	F	11.50	80.00	103.70	.050
7	1.16	M	8.30	76.50	118.51	.409
8	1.75	M	11.00	81.00	124.44	.530
9	.07	M	9.00	76.50	47.40	.520
10	1.35	M	9.00	75.00	62.90	1.041
11	1.83	F	9.50	77.00	100.00	.000
12	1.33	M	9.50	78.00	94.81	.000
13	1.17	M	10.00	75.00	65.18	.000
14	1.50	M	9.50	75.00	82.90	.520
15	.75	F	6.80	70.50	14.61	.000
16	1.67	F	11.00	85.00	53.33	.000
17	1.25	M	9.50	71.00	59.25	.000
18	1.50	M	10.00	79.00	29.02	.000
19	1.25	M	9.50	75.00	77.41	.434
20	1.92	F	9.00	74.00	75.00	.000
21	.92	M	9.50	70.00	60.00	.000
22	.92	M	7.50	65.00	105.00	.000
23	1.25	F	9.80	76.00	62.50	.000
24	.17	M	5.50	60.00	12.50	.652
25	1.00	F	8.00	69.00	50.00	.000
26	1.42	F	10.00	81.00	67.50	.084
27	1.50	F	.00	.00	175.00	.416
28	1.83	F	9.00	82.00	109.09	.206
29	1.75	M	10.70	82.00	157.57	.009
30	.75	M	9.00	72.00	121.21	.203
31	1.17	F	4.00	53.00	77.50	.409
32	1.42	F	6.50	61.00	48.48	.000
33	1.58	M	9.50	76.00	218.18	1.000
34	1.50	M	9.20	77.00	169.69	.198
35	.25	M	4.50	56.00	96.96	.000
36	.83	M	9.00	75.00	157.57	.384
37	1.67	M	.00	.00	101.61	.061
38	1.92	F	12.00	86.00	150.30	.000
39	1.42	F	8.00	75.00	157.57	.192
40	.92	M	8.70	76.00	96.96	.378

LACTANTES

PACIENTE	EDAD	SEXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
41	.92	M	9.30	72.00	133.33	.185
42	1.50	M	12.00	83.00	189.09	.576
43	1.25	M	10.50	80.00	157.57	.192
44	.17	F	5.50	57.00	109.09	.378
45	1.00	M	9.50	78.00	133.33	.000
46	.75	M	7.00	66.00	116.36	.456
47	.56	M	8.00	71.00	133.33	.000
48	1.58	M	9.50	76.00	206.06	.000
49	.42	F	5.90	.00	121.21	.000
50	.92	M	9.50	77.00	138.18	.225
51	.33	F	6.00	61.00	13.33	.000
52	.33	F	5.00	59.00	16.00	.229
53	.83	F	6.50	67.00	21.30	.652
54	1.08	M	.00	.00	53.53	.000
55	.50	F	8.00	66.00	26.66	.000
56	1.08	F	8.00	71.00	106.66	.903
57	1.50	F	.83	82.00	112.00	.520
58	1.17	M	9.90	80.00	66.66	1.442
59	1.42	F	10.30	81.00	106.66	.000
60	1.25	M	10.00	81.00	120.00	.200
61	1.33	F	8.90	75.00	93.33	2.083
62	1.92	M	11.00	88.00	120.00	1.730
63	.83	M	8.50	70.00	82.66	.943
64	.67	M	6.90	63.00	58.66	.000
65	.50	M	8.00	68.00	106.66	.000
66	.42	M	7.00	69.00	133.00	1.851
67	.33	M	6.15	58.00	38.09	1.008
68	1.75	F	10.00	77.00	63.49	.409
69	.42	F	7.00	65.50	38.09	.211
70	1.25	M	9.70	77.00	76.19	.390
71	.33	M	7.00	63.00	43.17	.467
72	.50	F	6.00	65.00	43.17	.000
73	.33	F	6.00	59.00	38.09	.000
74	.58	F	6.00	69.00	68.57	.000
75	.75	F	7.00	66.00	83.60	.231
76	1.92	M	10.00	80.00	88.88	.450
77	1.25	F	9.50	71.00	95.00	.000
78	1.92	F	.00	.00	87.50	.000
79	.83	M	10.30	80.00	180.00	.000
80	.50	F	7.00	64.00	50.00	.000

LACTANTES

PACIENTE	EDAD	SLXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
81	1.17	M	10.10	76.50	82.50	.000
82	1.00	M	11.50	76.50	100.00	.520
83	.50	M	6.50	67.00	50.00	.255
84	.17	F	4.00	51.00	43.50	.000
85	1.07	F	11.00	80.00	133.30	.000
86	1.25	F	8.55	70.00	107.09	.125
87	.33	M	6.40	63.00	35.89	.000
88	.07	F	6.50	67.00	107.09	.000
89	1.25	M	9.60	78.00	151.20	.000
90	1.33	F	8.50	75.00	112.62	.000
91	1.83	M	10.00	81.00	71.79	.000
92	.42	F	6.20	64.00	66.00	.000
93	1.00	F	10.00	76.00	107.09	.247
94	.25	M	4.50	56.00	46.15	.384
95	.83	M	7.90	72.50	82.00	.000
96	.33	M	5.00	60.00	30.70	.000
97	1.42	M	9.50	77.00	107.09	.255
98	1.92	F	11.00	86.00	133.30	.132
99	.92	F	8.50	71.00	58.97	.000
100	1.33	M	10.00	79.00	82.00	.260
101	.58	F	6.00	60.00	30.70	.125
102	.83	F	4.00	55.00	94.87	1.562
103	1.08	F	7.90	76.00	35.09	.000
104	1.92	F	10.70	81.00	13.30	.270
105	.83	M	8.50	72.00	65.30	.000
106	1.92	F	11.70	84.00	26.00	.000
107	1.25	M	9.20	74.00	53.30	.135
108	1.17	F	7.30	70.00	50.00	.135
109	1.92	F	10.10	80.00	56.00	.264
110	.83	F	6.00	64.00	26.00	.135
111	.83	M	7.20	69.00	53.30	.265
112	1.92	F	10.50	78.00	48.00	.259
113	.83	F	6.40	67.00	13.30	.000
114	.83	F	9.40	72.00	53.30	.255
115	1.92	F	13.80	89.00	106.00	.130
116	1.00	F	7.00	72.00	13.30	.515
117	1.92	F	13.50	88.00	101.30	.000
118	1.42	M	10.00	80.00	32.00	.471
119	.67	F	6.00	66.00	48.00	.000
120	1.92	F	13.00	87.00	40.00	.231

LACTANTES

PACIENTE	EDAD	SEXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
121	1.58	M	10.50	81.00	154.00	.427
122	.63	M	8.00	63.00	8.00	.423
123	1.58	M	.00	.00	120.00	.000
124	.75	M	.00	.00	80.00	.000
125	.67	F	8.60	.00	58.00	.431
126	.75	M	7.80	73.00	48.00	.000
127	1.67	F	.00	.00	40.00	.427
128	.75	F	9.00	75.00	8.00	.213
129	.67	F	.00	.00	80.00	.000
130	1.17	F	12.50	73.00	150.30	.000
131	1.00	F	10.70	74.00	72.72	.000
132	.42	M	.00	.00	133.33	.000
133	1.92	F	10.50	77.00	235.15	.240
134	1.33	M	7.10	70.00	84.84	.126
135	.33	F	5.50	57.50	84.84	.000
136	.42	M	5.00	60.00	67.87	.127
137	.75	M	7.00	70.00	145.45	.000
138	.08	M	.00	.00	150.30	.000
139	.08	M	.00	.00	109.09	.000
140	1.25	M	9.90	75.00	140.00	.000
141	1.33	F	10.00	75.00	138.16	.000
142	.67	M	6.00	68.00	116.36	.250
143	1.50	F	13.50	90.00	109.09	.000
144	.33	M	6.00	59.00	109.09	.252
145	1.08	M	6.70	70.00	72.72	1.243
146	.67	F	7.18	72.00	96.96	.000
147	1.83	M	13.00	89.00	150.30	.123
148	.75	F	8.00	71.00	72.72	.000
149	1.42	F	8.70	79.00	111.51	.000
150	.83	F	10.00	68.00	116.36	.000
151	.58	F	.00	.00	77.41	.618
152	1.92	F	8.50	77.00	123.87	.000
153	1.58	M	11.50	77.00	69.67	.000
154	1.42	F	9.40	77.00	82.58	.000
155	.25	F	5.40	59.00	25.00	.000
156	1.92	F	11.70	80.00	77.41	.000
157	1.58	M	12.00	82.00	56.77	.250
158	1.58	M	7.00	73.00	51.61	.119
159	1.33	M	10.00	78.00	64.51	.480
160	1.17	F	9.50	75.00	72.25	.235

LACTANTES

PACIENTE	EDAD	SLXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
161	1.92	F	15.70	95.00	87.74	.000
162	.25	F	4.90	57.00	64.51	.000
163	1.92	F	7.00	69.00	72.25	.000
164	.42	M	8.00	64.00	46.45	.000
165	.50	F	6.50	62.50	12.90	.000
166	.75	M	7.70	78.00	20.64	.000
167	1.92	F	.00	.00	201.29	.000
168	1.92	F	12.40	88.00	51.01	.245
169	.33	M	6.00	60.00	69.67	.875
170	.83	M	9.00	71.00	12.90	.000
171	1.67	M	11.50	81.00	90.32	.000
172	.25	F	3.90	50.00	98.00	.000
173	.75	M	7.10	72.00	77.41	1.884
174	1.92	F	8.50	75.00	20.64	.000
175	.58	M	7.00	66.00	77.41	.000
176	1.92	M	15.00	88.00	33.54	.245
177	1.58	F	10.00	77.00	98.00	.000
178	1.58	F	7.00	74.00	108.30	.367
179	1.50	M	12.70	85.00	103.22	.000
180	.56	F	7.54	69.00	51.01	.236
181	.25	M	5.20	56.00	25.00	.000
182	1.25	F	11.10	83.00	46.45	.000
183	1.25	F	.00	.00	77.41	.000
184	1.00	F	9.00	72.00	62.56	.467
185	.67	F	6.70	64.00	123.67	.242
186	1.63	F	10.40	85.00	103.22	.238
187	.58	M	8.20	70.00	67.69	.213
188	.17	M	5.00	58.00	25.00	.000
189	.63	M	9.50	70.00	59.35	.000
190	1.92	F	11.50	88.00	141.93	.000
191	1.92	F	10.50	82.00	56.77	.000
192	.50	F	4.00	53.00	56.77	.000
193	.17	M	5.00	55.50	77.41	.000
194	.67	F	3.80	56.00	38.70	.000
195	1.58	M	9.90	79.00	92.90	.000
196	.92	F	6.00	68.00	64.51	1.196
197	1.92	F	13.50	89.00	72.25	.124
198	1.92	M	.00	.00	116.12	.000
199	1.92	M	11.50	90.00	64.51	.000
200	.33	M	5.20	54.00	123.67	.000

ESCOLARES

PACIENTE	EDAD	SEXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
201	11.00	M	33.00	141.00	88.88	.434
202	10.00	M	23.00	121.00	53.33	1.041
203	10.00	M	38.00	140.00	44.44	.223
204	9.00	F	24.90	120.00	112.50	.000
205	8.00	M	32.00	132.00	125.00	.840
206	7.00	F	18.50	117.00	100.00	.635
207	9.00	F	24.00	121.00	75.00	.961
208	12.00	M	39.00	142.00	212.50	.000
209	10.00	M	29.50	140.00	150.00	.000
210	10.00	M	31.00	136.00	130.00	.403
211	7.00	F	21.50	120.00	150.00	.434
212	11.00	F	31.30	144.00	150.00	.217
213	10.00	F	27.70	139.00	150.00	.806
214	9.00	M	.00	.00	218.18	.203
215	7.00	M	15.80	108.00	109.69	1.171
216	9.00	F	25.00	125.00	193.93	.576
217	7.00	F	15.00	140.00	181.81	.195
218	7.00	M	21.00	118.00	198.78	.669
219	9.00	F	27.50	120.00	218.18	.000
220	7.00	M	.00	.00	198.78	.000
221	10.00	F	21.70	111.00	109.69	1.327
222	9.00	M	38.30	140.00	145.45	.000
223	10.00	F	21.50	120.00	181.81	.000
224	8.00	M	25.00	121.00	133.33	.000
225	9.00	F	41.50	140.00	93.33	.000
226	7.00	M	23.00	123.00	16.00	.227
227	10.00	F	23.50	120.00	122.66	.227
228	12.00	F	42.00	150.00	106.66	.221
229	10.00	M	25.00	124.00	93.33	.223
230	9.00	F	20.00	105.00	48.00	.000
231	12.00	M	35.50	140.00	106.66	.227
232	12.00	M	30.30	136.00	93.33	.793
233	10.00	M	41.50	105.00	80.00	1.923
234	12.00	M	43.50	142.50	181.33	1.020
235	10.00	F	41.00	138.00	80.00	.500
236	7.00	M	22.50	120.00	85.33	.000
237	10.00	F	34.50	147.00	85.33	.740
238	10.00	M	67.60	152.00	93.33	.980
239	10.00	F	27.10	126.00	170.66	1.530
240	11.00	F	29.50	135.00	136.00	1.886

LSCOLARES

PACIENTE	EDAD	SEXO	PESO	ESTATURA	AC. AMILASA	AC. IPASA
241	12.00	M	38.00	146.00	112.00	.396
242	9.00	M	23.00	128.00	100.00	.000
243	11.00	F	27.10	135.00	96.00	3.682
244	11.00	M	54.00	148.00	153.33	.200
245	8.00	M	18.50	122.00	80.00	.000
246	11.00	M	26.00	132.00	100.00	.000
247	7.00	M	17.00	118.00	190.47	.400
248	10.00	F	27.00	140.00	106.60	1.190
249	9.00	M	25.00	135.00	159.68	.423
250	12.00	M	38.00	142.50	165.67	.000
251	9.00	M	.00	.00	152.00	1.612
252	10.00	F	36.00	134.00	114.28	.000
253	8.00	M	18.00	121.50	157.46	1.209
254	12.00	F	46.30	144.00	101.58	.000
255	8.00	M	19.50	123.00	119.36	.423
256	10.00	M	29.50	134.00	114.28	.000
257	11.00	F	23.50	123.00	147.30	.000
258	8.00	F	24.00	120.00	91.42	.423
259	8.00	F	33.00	133.00	119.36	.393
260	8.00	F	20.00	120.50	126.98	.201
261	7.00	F	17.00	122.00	152.00	.390
262	7.00	M	18.10	111.00	159.68	.403
263	7.00	M	19.00	123.00	165.07	.609
264	11.00	F	25.50	125.00	165.07	3.500
265	9.00	M	.00	.00	68.57	.000
266	12.00	M	38.50	147.00	190.47	.833
267	9.00	F	26.50	126.00	165.07	.694
268	10.00	F	42.20	146.50	190.47	.231
269	10.00	M	29.60	131.00	88.88	.000
270	9.00	F	25.00	126.50	88.88	.000
271	8.00	M	17.70	.00	63.49	.688
272	9.00	F	21.10	104.00	114.28	.000
273	10.00	M	33.00	135.00	152.38	.000
274	9.00	M	30.00	138.00	111.74	.450
275	9.00	F	32.90	132.00	152.38	.694
276	11.00	F	30.50	146.00	88.88	.231
277	10.00	M	24.80	136.00	101.58	.675
278	10.00	M	29.50	135.50	58.41	.462
279	9.00	M	23.50	123.50	124.44	.235
280	10.00	F	27.50	132.50	68.57	.000

ESCOLARES

PACIENTE	EDAD	SEXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
261	11.00	F	29.20	136.00	93.90	.229
262	9.00	F	30.00	141.50	76.19	.467
263	9.00	F	27.50	127.50	139.68	.000
264	10.00	M	30.00	130.50	144.76	.000
265	7.00	M	21.50	118.00	170.15	.000
266	9.00	F	19.50	118.00	144.76	.000
267	12.00	F	41.00	146.00	134.60	.490
268	9.00	M	25.00	130.50	139.68	.450
269	10.00	F	20.00	115.00	167.61	.000
290	7.00	F	23.50	125.00	126.98	.000
291	11.00	F	30.50	136.00	117.50	.000
292	11.00	F	26.00	138.50	150.00	.231
293	7.00	M	20.00	114.50	125.00	.231
294	8.00	M	28.00	136.50	117.50	.000
295	7.00	F	23.70	121.50	100.00	.231
296	12.00	F	39.50	150.00	100.00	.467
297	9.00	F	28.00	131.00	132.50	.231
298	9.00	M	18.00	115.00	120.00	.229
299	7.00	F	23.20	120.00	125.00	.000
300	8.00	F	21.50	122.00	100.00	.000
301	8.00	F	22.50	127.50	60.00	.000
302	8.00	M	24.00	130.00	95.00	.000
303	9.00	M	25.00	126.00	112.50	.000
304	10.00	F	33.40	148.00	125.00	.000
305	12.00	F	36.00	145.50	100.00	.000
306	9.00	M	29.50	130.50	112.50	.000
307	9.00	F	24.50	126.00	105.00	.000
308	9.00	F	29.20	129.00	75.00	.510
309	12.00	M	42.50	142.00	100.00	.505
310	9.00	M	40.40	130.00	100.00	.000
311	12.00	M	35.50	147.00	150.00	.252
312	10.00	M	23.50	127.00	105.00	.250
313	10.00	F	23.00	133.50	112.50	.757
314	10.00	F	22.00	125.50	100.00	.382
315	12.00	M	33.50	150.00	150.00	.510
316	9.00	M	18.30	118.00	145.00	.500
317	12.00	M	30.00	138.00	120.00	.765
318	9.00	F	32.30	125.00	130.00	.255
319	7.00	F	24.40	119.00	100.00	.382
320	7.00	F	23.50	127.00	92.50	.505

ESCOLARES

PACIENTE	EDAD	SEXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
321	9.00	F	32.50	128.50	100.00	.257
322	8.00	F	22.50	123.50	225.00	1.530
323	7.00	M	27.00	126.00	151.20	.515
324	10.00	F	32.00	143.00	82.05	.000
325	8.00	M	25.00	125.00	87.94	.384
326	12.00	F	31.50	136.50	107.69	.000
327	10.00	F	25.50	130.00	112.82	.000
328	9.00	M	.00	.00	189.74	.255
329	12.00	F	31.50	139.00	82.05	.257
330	11.00	F	37.00	150.00	61.55	.245
331	8.00	F	23.50	125.00	69.23	.250
332	7.00	M	24.50	121.00	107.69	.000
333	10.00	F	29.50	137.00	82.05	.380
334	7.00	M	19.50	112.00	82.05	.000
335	12.00	M	30.50	136.00	69.23	.000
336	10.00	F	36.50	142.00	82.05	.398
337	8.00	F	26.00	123.50	158.46	.403
338	11.00	M	33.00	140.00	94.87	.398
339	10.00	M	26.10	131.00	45.33	.407
340	11.00	M	25.00	123.50	48.00	.135
341	11.00	M	33.20	138.00	74.66	.133
342	9.00	F	26.00	125.00	120.00	.661
343	7.00	F	22.50	125.50	98.66	.270
344	9.00	F	22.50	134.00	72.00	.263
345	9.00	M	25.00	126.50	58.66	.000
346	9.00	M	29.20	135.00	80.00	.263
347	9.00	M	36.50	145.50	98.66	.132
348	12.00	M	31.10	140.50	85.33	.427
349	10.00	M	39.00	140.00	106.66	.000
350	7.00	M	21.00	118.50	146.66	.434
351	8.00	M	.00	.00	58.66	.000
352	11.00	F	44.50	.00	85.33	.427
353	8.00	M	27.60	130.00	98.66	.208
354	10.00	M	21.50	121.50	53.33	.210
355	9.00	F	23.80	123.00	98.66	.210
356	10.00	F	24.60	125.00	80.00	.211
357	12.00	M	37.00	155.50	53.33	.210
358	11.00	F	35.00	147.00	192.00	.000
359	11.00	M	32.50	137.50	34.66	.000
360	8.00	F	18.80	116.00	74.66	.213

ESCOLARES

PACIENTE	EDAD	SEXO	PESO	ESTATURA	AC. AMILASA	AC. LIPASA
301	7.00	F	21.80	121.00	100.00	.201
302	9.00	M	22.10	125.00	58.60	.211
303	8.00	F	.00	.00	133.33	.000
304	8.00	M	21.10	123.00	169.69	.000
305	11.00	M	33.70	135.50	116.36	.367
306	9.00	F	31.00	133.00	172.12	.000
307	12.00	F	32.00	140.00	30.96	.367
308	10.00	M	27.00	136.00	98.00	.240
309	12.00	M	69.00	161.00	98.00	.000
310	12.00	M	27.20	134.00	118.70	.238
311	12.00	M	23.30	110.50	123.87	.000
312	7.00	F	27.00	130.00	51.61	.480
313	9.00	M	26.40	127.50	129.03	.000
314	11.00	F	31.00	144.00	77.41	.000
315	9.00	M	21.40	121.00	51.61	.000
316	10.00	M	21.00	123.50	33.54	.000
317	10.00	F	40.70	138.50	77.41	.480
318	9.00	F	20.90	109.00	69.67	.000
319	12.00	M	47.00	155.50	51.61	.000
320	10.00	M	10.50	115.00	129.03	.000
321	11.00	F	38.00	138.00	64.51	.000
322	11.00	M	29.70	136.00	64.51	.000
323	9.00	M	26.00	130.00	36.12	.000
324	11.00	F	39.30	153.00	116.12	.118
325	10.00	M	33.80	134.00	90.32	.000
326	9.00	F	25.60	131.00	51.61	.000
327	7.00	M	15.00	110.00	108.38	.000
328	8.00	F	22.50	117.00	77.41	.000
329	7.00	M	23.00	128.00	20.64	.000
330	8.00	F	29.60	130.00	90.32	.480
331	11.00	M	40.50	142.00	123.87	.358
332	9.00	M	27.00	130.00	126.45	.000
333	9.00	F	26.00	128.50	90.32	.721
334	12.00	F	48.00	145.00	123.87	.242
335	12.00	F	27.50	136.00	149.67	.233
336	12.00	F	38.00	142.00	116.12	.000
337	9.00	F	35.30	136.00	103.22	.358
338	9.00	F	30.00	128.00	64.51	.000
339	7.00	F	24.30	123.50	180.64	.480
460	9.00	F	24.50	130.50	141.93	.000

2.4 APLICACION DEL ANALISIS DISCRIMINANTE A LA INVESTIGACION

Para cumplir con nuestro objetivo se han analizado los siguientes puntos:

- a) Las diferencias existentes entre los lactantes y escolares en la actividad de la amilasa y la actividad de la lipasa
- b) Considerando a los lactantes, encontrar las diferencias que se presentan en hombres y mujeres en la actividad de la amilasa y la lipasa.
- c) Considerando solamente a los escolares, encontrar las diferencias que se presentan en hombres y mujeres en la actividad de la amilasa y la lipasa.

Para tratar el punto (a) hemos subdividido el número total de casos de la forma siguiente:

GRUPO	CASOS QUE LO INTEGRAN
1	Lactantes
2	Escolares

Las variables que se desean analizar son V6 y V7 correspondientes a la actividad de la amilasa y a la actividad de la lipasa respectivamente.

En el punto (b) los casos que forman parte de los lactantes se agrupan de la manera siguiente:

GRUPO	CASOS QUE LO INTEGRAN
1	Hombres
2	Mujeres

Se desea incluir dentro del análisis las variables: peso (V4), estatura (V5), actividad de la amilasa (V6), actividad de la lipasa (V7).

En el punto (c) los individuos que integran a los escolares se les agrupa de igual forma, y también se quieren incluir las mismas variables que en el punto b.

2.5 PRINCIPALES RESULTADOS OBTENIDOS.

2.5.1 Resultados obtenidos para los grupos de lactantes y escolares.

MEANS

	GROUP 1	GROUP 2	TOTAL
V2	1.0893	0.5230	0.7960
V3	1.4973	1.5131	1.5053
V4	18.5601	22.7694	18.8600
V5	72.0262	130.0566	102.0660
V6	81.5099	111.0871	97.0740
V7	0.2562	0.3277	0.2829

STANDARD DEVIATIONS

	GROUP 1	GROUP 2	TOTAL
V2	0.5447	1.5680	4.3881
V3	0.5010	0.5511	0.5006
V4	7.4351	8.3220	11.8197
V5	23.2971	13.9801	39.0532
V6	43.7728	40.6656	44.0333
V7	0.7667	0.4911	0.4448

CUADRO 1

F MATRIX - DEGREES OF FREEDOM: 2, 371

GROUP 2 GROUP 1
24.86476

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

VARIABLE	UNIT OF MEASUREMENT	F TO ENTER
V6	48.71042	45.26710
V7	3.09776	1.07230

ALL FLIGHT VARIABLES INCLUDED

CUADRO 2

SUMMARY TABLE

Step Number	Variable entered	Variable removed	F to enter or removed	number included	Wilk's lambda	sig	Rao's V	Change in Rao's V	Sig. of Change
1	V6		48.71542	1	0.88421	0.000	48.71542	48.71542	0.000
2	V7		1.01236	2	0.88180	0.000	49.86344	1.14802	0.284

Discriminant Function	eigenvalue	Relative percentage	Canonical correlation	Functions derived	Wilk's lambda	Chi-square	DF	Significance.
1	0.13404	100.00	0.344	0	0.8818	46.667	2	0.000

CUADRO 3

REMAINING COMPUTATIONS WILL BE BASED ON 1 DISCR.

PREDICTION RESULTS -

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
V6	0.96859
V7	0.14455

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP	
		GP. 1	GP. 2
GROUP 1	200.	127.03.5%	73.03.5%
GROUP 2	200.	73.36.5%	127.63.5%

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
V6	0.02156
V7	0.32495
CONSTANT	-2.14400

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 63.50%

CENTROIDS OF GROUPS IN REDUCED SPACE

	FUNC 1
GROUP 1	-0.35076
GROUP 2	0.33607

CUADRO 5

CUADRO 4

2.5.2 RESULTADOS OBTENIDOS PARA LOS LACTANTES SUBDIVIDIDOS EN LOS GRUPOS DE HOMBRES Y MUJERES.

MEANS

	GROUP 1	GROUP 2	TOTAL
V2	1.8171	1.1665	1.0893
V4	8.7223	8.4363	8.5801
V5	70.6030	77.1808	72.4262
V6	68.4552	74.4864	61.5090
V7	0.3896	0.1710	0.2362

STANDARD DEVIATIONS

	GROUP 1	GROUP 2	TOTAL
V2	3.5050	3.5783	3.7467
V4	2.1951	2.7385	2.4381
V5	5.5065	10.0476	6.2971
V6	4.1728	11.3793	3.7728
V7	0.4227	0.3364	0.3867

CUADRO 6

F MATRIX - DEGREES OF FREEDOM: 2, 180

GROUP	GROUP 1
2	4.03131

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

VARIABLE	F INY. CRITERION	F IN DECVT
V6	3.92523	3.92523
V7	0.25261	4.11775

F LEVEL INSUFFICIENT FOR FURTHER CORRELATION

CUADRO 7

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	V7		5.25251	1	0.97180	0.022	5.25251	5.25251	0.022
2	V6		3.92523	2	0.95106	0.011	9.31409	4.06158	0.044

Discriminant function	eigenvalue	Relative percentage	canonical correlation	functions derived	Wilks' lambda	Chi-square	DF	Significance
1	0.05146	100.00	0.221	0	0.9511	9.032	2	0.011

CUADRO 8

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
V6	0.65422
V7	0.69405

PREDICTION RESULTS -

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1
V6	0.01475
V7	1.70500
CONSTANT	-1.64211

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP	
		GP. 1	GP. 2
GROUP 1	100.	50.0%	50.0%
GROUP 2	100.	35.0%	65.0%

CENTROIDS OF GROUPS IN REDUCED SPACE

	FUNC 1
GROUP 1	0.21942
GROUP 2	-0.22103

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 57.50%

CUADRO 9

CUADRO 10

2.5.3 RESULTADOS OBTENIDOS PARA LOS ESCOLARES SUBDIVIDIDOS EN LOS GRUPOS DE HOMBRES Y MUJERES.

MEANS

	GROUP 1	GROUP 2	TOTAL
V2	9.5914	9.4592	9.5236
V4	29.1032	28.7357	28.7094
V5	131.0054	130.7143	130.8560
V6	109.6517	114.2033	111.9871
V7	0.2780	0.3748	0.3277

STANDARD DEVIATIONS

	GROUP 1	GROUP 2	TOTAL
V2	1.6826	1.5003	1.5888
V4	9.5861	6.9447	8.3226
V5	11.2693	10.6541	10.9001
V6	42.2971	39.1433	40.6658
V7	0.3489	0.5935	0.4911

CUADRO 11

F MATRIX - DEGREES OF FREEDOM: 1, 189
 GROUP 2 GROUP 1
 1.86324

ESCOLARES, GRUPOS: HOMBRES Y MUJERES

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

VARIABLE	ENTRY CRITERION	F TO REMOVE
V7	1.86324	1.86324

F LEVEL INSUFFICIENT FOR FURTHER COMPUTATION

CUADRO 12

SUMMARY TABLE

step number	variable entered removed	F to enter or remove	number included	Wilks' lambda	Sig. Rao's V	Change in Rao's V	Sig. of change	
1	V7	1.86324	1	0.99024	0.170	1.86324	0.172	
Discriminant function	eigenvalue	Relative percentage	Canonical correlation	Functions derived	Wilks' lambda	Chi-square	DF	significance
1	0.00093	100.00	0.031	0	0.9991	0.176	1	0.675

CUADRO 13

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

V7	1.00000
----	---------

UNSTANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS

V7	2.03689
CONSTANT	-0.06728

CENTROIDS OF GROUPS IN REDUCED SPACE

GROUP 1	-0.10116
GROUP 2	0.07600

CUADRO 14

ESCOLARES, GRUPOS INGRESOS Y PROGRES

PREDICTION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GP. 1	GROUP MEMBERSHIP GP. 2
GROUP 1	100.	64.0%	36.0%
GROUP 2	100.	61.0%	39.0%

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 51.50%

CUADRO 15

2.6 ANALISIS DE LOS RESULTADOS OBTENIDOS.

2.6.1 ANALISIS DE LOS RESULTADOS OBTENIDOS PARA LOS GRUPOS DE LACTANTES Y ESCOLARES.

Ambas variables V6 y V7 fueron incluidas dentro del análisis.

Tomando en consideración los 400 casos la media obtenida para la actividad de la amilasa es de 97.074 U y para la actividad de la lipasa es de 0.2829 U⁽¹⁾.

El promedio de la actividad de la amilasa en los lactantes es de 81.509 que es menor a la que presentan los escolares con 111.9871. La actividad de la lipasa tiene una media para los lactantes de 0.7362, la cual es también menor que la obtenida para el grupo de escolares con 0.3227.

La diferencia existente entre los grupos de lactantes y escolares es notable haciendo uso de las variables discriminantes V6 y V7.

Se derivó una sola función discriminante. La estandarizada es:

$$DI = 0.96659 V6 + 0.14455 V7$$

de donde observamos que la actividad de la amilasa proporciona la mayor contribución a la discriminación.

La función discriminante no estandarizada es:

$$DI = 0.02159 V6 + 0.32495 V7 - 2.1848$$

(1) Los valores promedios citados en medicina son para los adultos en la actividad de la amilasa de 110 U y para la lipasa es de 0.525.

Es por demás decir que D1 tiene una importancia relativa del 100%. La relación canónica de esta función es de 0.344 la cual es bastante baja.

Los resultados relativos a la clasificación de nuevos casos se encuentra en la tabla resumen del cuadro #5 donde se observa:

- 1) Ambos grupos tienen un porcentaje de clasificación correcta de 63.5%.
- 2) Ambos grupos tienen el porcentaje de clasificación incorrecta de 36.5%.
- 3) Del grupo total de casos un 63.5% de ellos fue clasificado correctamente.

2.6.2 ANALISIS DE LOS RESULTADOS OBTENIDOS PARA LOS LACTANTES SUBDIVIDIDOS EN LOS GRUPOS DE HOMBRES Y MUJERES.

Las variables incluidas dentro del análisis son V6 y V7.

Para los hombres, el promedio para la actividad de la amilasa es de 88.455 U y de la actividad lipasa es de 0.30 U, estas medias son mayores que las obtenidas para las mujeres cuyos valores son de 74.4864 y .1710 respectivamente.

En los lactantes los grupos de hombres y mujeres son poco diferentes entre sí (cuadro 7).

La función discriminante estandarizada que se derivó es:

$$D1 = 0.65422 V6 + 0.69405 V7$$

La contribución a la discriminación de las dos variables es casi la misma.

La función discriminante no estandarizada está dada por:

$$D1 = 0.01495 V6 + 1.795 V7 - 1.64211$$

Es obvio que su importancia relativa es de 100%. La correlación canónica asociada a esta función es de 0.221.

En los resultados relativos a la clasificación de nuevos casos (cuadro 10) se reporta lo siguiente:

- 1) El grupo mejor clasificado es el de las niñas con un 65% de casos clasificados correctamente.
- 2) El grupo peor clasificado es el correspondiente al de los niños con el 50% de casos clasificados en forma correcta.
- 3) Solamente un 57.5% de los casos totales fueron clasificados correctamente.

2.6.3 ANALISIS DE LOS RESULTADOS OBTENIDOS PARA LOS ESCOLARES SUBDIVIDIDOS EN LOS GRUPOS DE HOMBRES Y MUJERES.

La única variable seleccionada por el método stepwise es V7 correspondiente a la actividad de la lipasa. Su media para los hombres es de 0.3489 U que es menor que la de las mujeres con 0.5935.

La diferencia existente entre estos grupos es casi nu
la.

La función discriminante estandarizada derivada es:

$$D1 = V7$$

y la no estandarizada es:

$$D1 = 2.03640 V7 - 0.66728$$

La correlación canónica asociada a esta función es de 0.031 el cual es un valor pésimo.

Los resultados correspondientes a la clasificación de nuevos individuos se encuentra en el cuadro 15 donde se resume lo siguiente:

- 1) El grupo peor clasificado es el 2 con 39% de casos -- clasificados correctamente
- 2) El grupo mejor clasificado es el 1 con 64% de casos - clasificados correctamente.
- 3) Solamente se clasificaron en forma correcta el 51.5% del total de los casos.

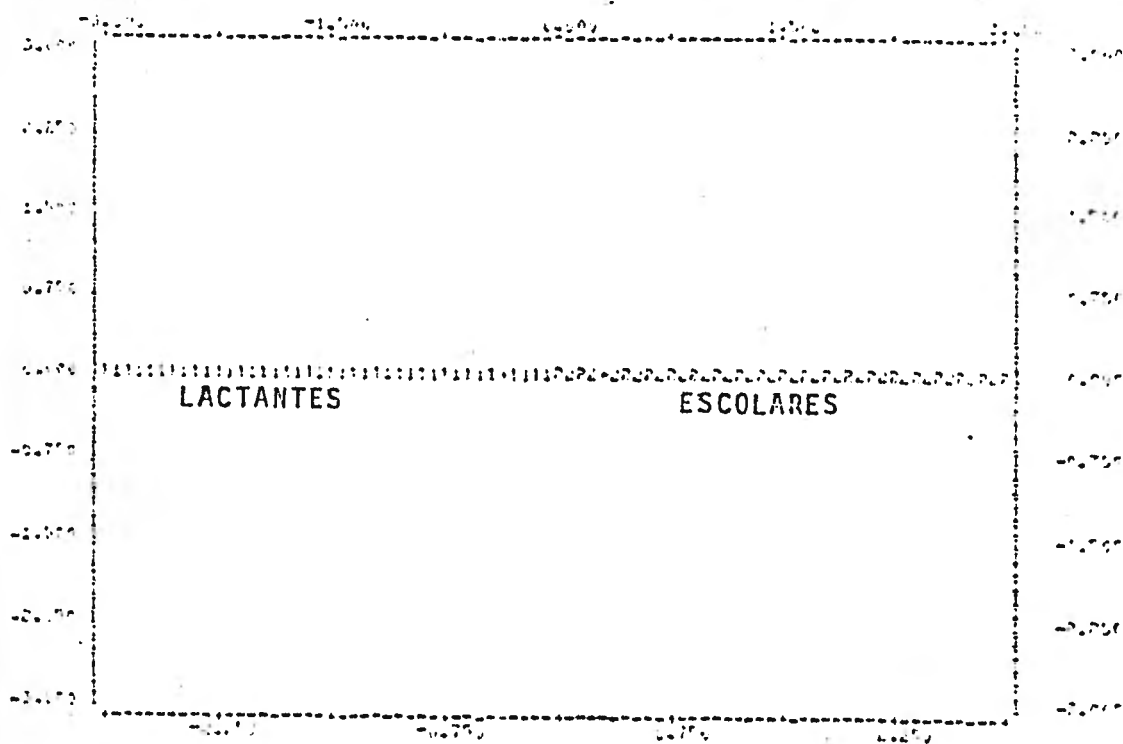
2.7 CARACTERIZACION DE LOS GRUPOS EN BASE AL MAPA TERRITORIAL.

2.7.1 CARACTERIZACION DE LOS LACTANTES Y ESCOLARES.

GRUPOS: Lactantes y escolares.

Función discriminante estandarizada en base a la cual se efectúa la caracterización:

$$D1 = 0.96659 V6 + 0.14455 V7$$



←
D1 negativa
Valores pequeños
de V6 y V7

→
D1 positiva
Valores grandes
de V6 y V7

DESCRIPCION DE LOS GRUPOS. El grupo de los lactantes presenta valores pequeños en las enzimas de amilasa y lipasa séricas. El grupo de los escolares se caracteriza por valores grandes en la actividad de la amilasa y lipasa séricas.

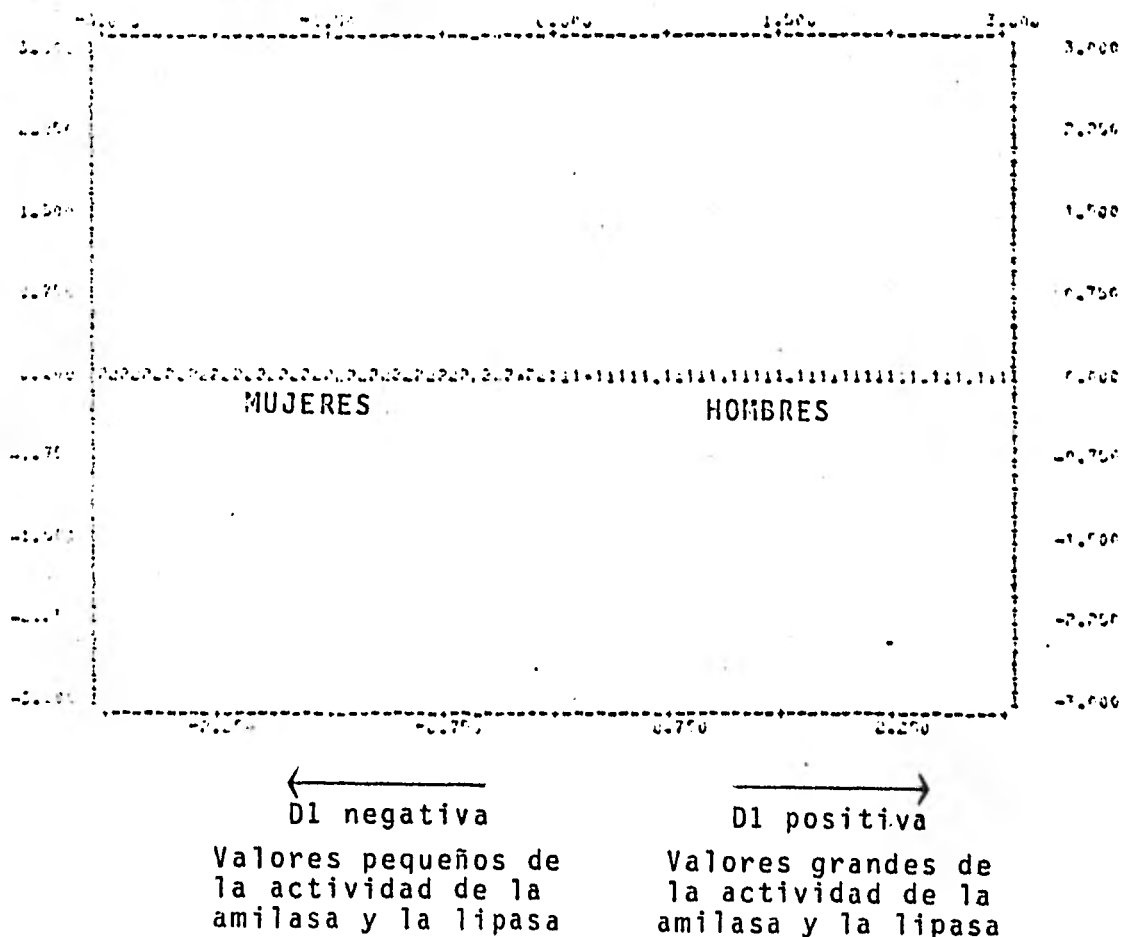
2.7.2 CARACTERIZACION DE LOS LACTANTES.

GRUPOS: Hombres y Mujeres.

Función discriminante estandarizada a partir de la cual se efectúa la caracterización:

$$D1 = 0.65422 V6 + 0.69405 V7$$

DESCRIPCION DE LOS GRUPOS. El grupo de hombres se caracteriza por valores grandes en la actividad de la amilasa y lipasa séricas. El grupo de mujeres presenta los valores más pequeños en estas variables.



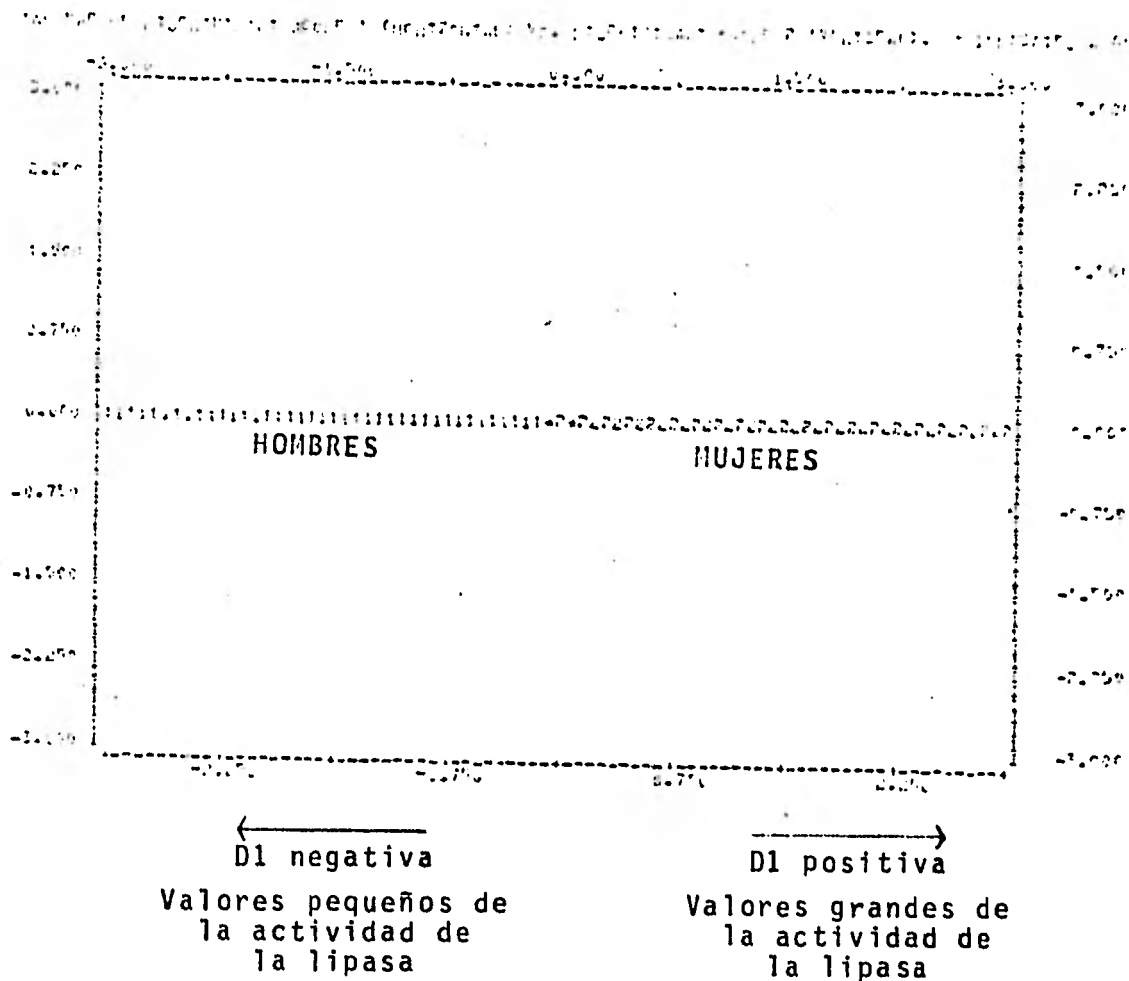
DESCRIPCION DE LOS GRUPOS. El grupo de hombres se caracteriza por valores grandes en la actividad de la amilasa y lipasa séricas. El grupo de mujeres presenta los valores más pequeños en estas variables.

2.7.3 CARACTERIZACION DE LOS ESCOLARES.

GRUPOS: Hombres y mujeres.

Función discriminante estandarizada a partir de la cual se efectúa la caracterización:

$$D1 = V7$$



DESCRIPCION DE LOS GRUPOS. El grupo de los niños escolares se caracteriza por valores pequeños en la actividad de la lipasa y el de las niñas por valores grandes en esta variable.

2.8 CONCLUSIONES

Considerando los grupos de lactantes y escolares, las diferencias encontradas en la actividad de la amilasa y la lipasa son notorias, siendo los valores de ambas enzimas superiores en los escolares con respecto a los lactantes.

Analizando a los lactantes en base al sexo, se -- observa que estos grupos no presentan grandes diferencias, es decir que están traslapados, pero a pesar de ello en -- las dos actividades enzimáticas los hombres se caracterizan por tener los valores más elevados.

En los escolares, en cuanto al sexo existen diferencias casi nulas en la actividad de la amilasa y con respecto a la actividad de la lipasa estas no son significativas. La correlación canónica de la función discriminante derivada es extremadamente baja, y por esta razón, no podemos concluir que el grupo de niños se caracteriza por presentar los valores más pequeños que los de las niñas en la actividad de la lipasa, sino simplemente que las diferencias presentadas son irrelevantes en esta enzima.

C A P I T U L O 3

SEGUNDO EJEMPLO DE LA APLICACION
DEL ANALISIS DISCRIMINANTE A LA MEDICINA3.1 ASOCIACION ENTRE LA ENFERMEDAD ARTICULAR DEGENERATIVA Y LA
DIABETES MELLITUS.

El páncreas es una glándula situada por abajo del estómag^o y a su derecha, se extiende desde el arco duodenal hacia el fondo del estómago. En el seno del tejido pancreático se encuentran los islotes de Langerhans, cuya función es sintetizar la insulina. La insulina es una sustancia que se vierte en el torrente sanguíneo y es utilizada por los tejidos para el aprovechamiento de la glucosa circulante o glucógeno almacenado.

La diabetes Mellitus es una enfermedad cuya causa se desconoce pero su defecto básico se debe a que el páncreas no produce la necesaria cantidad de insulina, por tal motivo se eleva el azúcar en la sangre, es un trastorno metabólico complejo que suele establecerse de manera genética. Los síntomas de la diabetes Mellitus varían mucho según que la insuficiencia de la insulina sea relativa o absoluta. La diabetes Mellitus puede clasificarse en varias fases clínicas:

- 1) Periodo prediabético. No se identifican anomalías metabólicas pero el sujeto es genéticamente diabético.
- 2) Fase de diabetes química. Fase durante la cual el paciente es asintomático pero se identifican anomalías de la tolerancia de la glucosa en pruebas.
- 3) Periodo de diabetes patente. El síndrome puede ser ---

asintomático o acompañarse de síntomas atribuibles de manera directa a la falta de insulina o de un conjunto de manifestaciones que ocurren aisladamente o en combinación y afectan a los riñones, retina, sistema vascular y neurológico. Es motivo de controversias que estén relacionadas de manera casual con una deficiencia de insulina.

El diagnóstico se hace fundándose en datos de laboratorio.

La enfermedad articular degenerativa (EAD) es una anomalía de las articulaciones diartroideas (aquellas que cuentan con movimiento), caracterizada por erosión y disgregación fibrilar y que no es directamente atribuible a un proceso inflamatorio. Es la enfermedad articular más común, su frecuencia aumenta según la edad. Muchos individuos son asintomáticos o casi asintomáticos. La enfermedad sintomática con poca frecuencia origina incapacidad intensa, y ocurre con mayor facilidad en las articulaciones que soportan peso. Cuando la enfermedad guarda relación con traumatismos, defectos posturales, obesidad o artritis inflamatoria previa, se le llama de tipo secundaria. Si no se identifican factores desencadenantes, excepto la herencia se le llama de tipo primaria.

Se piensa que la glucosa tiene un papel importante en el metabolismo de los glucosa-aminoglicanos, constituyentes fundamentales de la sustancia amorfa del tejido conectivo --- principalmente a nivel de cartílago, de aquí que sea probable que una disminución de su utilización por deficiencia absoluta o relativa de insulina en los pacientes diabéticos, produzca alteraciones degenerativas articulares en etapas más tempranas y de mayor intensidad que los observados en la población general. Existen por lo menos diez padecimientos reumáticos que se asocian con la diabetes Mellitus. La enfermedad articular dege

nerativa es el padecimiento reumatológico más común y aparentemente se presenta con mayor frecuencia y severidad en el diabético. La relación entre ambas enfermedades ha sido poco estudiada y no se han determinado las características de tales padecimientos en el mismo paciente.

OBJETIVO: Establecer la existencia de una posible relación entre la diabetes Mellitus y la Enfermedad Articular degenerativa.

3.2 ESTUDIO REALIZADO PARA ENCONTRAR LA EXISTENCIA DE UNA POSIBLE RELACION ENTRE LA DIABETES MELLITUS Y LA ENFERMEDAD ARTICULAR DEGENERATIVA.

Se tomó una población de 150 pacientes del Hospital General de México el cual estaba integrado de la siguiente manera:

GRUPO 1. 50 pacientes diabéticos.

GRUPO 2. 50 pacientes con Enfermedad Articular Degenerativa.

GRUPO 3. 50 pacientes control.

GRUPO DE DIABETICOS. Está formado por individuos cuyo diagnóstico es el de diabetes Mellitus, de ambos sexos y mayores de 40 años. Los pacientes deberfan estar bajo control médico y se excluyeron aquéllos con otros padecimientos endócrinos.

GRUPO DE PACIENTES CON EDAD. Está formado por sujetos con Enfermedad Articular Degenerativa primaria, de ambos -

sexos, mayores de 40 años. Los pacientes deberían estar bajo tratamiento médico con analgésicos. Se excluyeron los pacientes con enfermedades congénitas o hereditarias, y enfermedad articular degenerativa secundaria.

GRUPO CONTROL. Lo integran pacientes en los mismos promedios de edad y frecuencia por sexo en relación a los grupos anteriores. Se incluyeron sujetos sanos voluntarios o enfermos sin padecimientos óseos, articulares del tejido conectivo, endocrinológicos o neurológicos.

En el estudio de cada uno de los pacientes de los 3 grupos se incluyó: historia clínica y exploración física completa. A cada uno de los 150 pacientes se les valoraron nueve áreas anatómicas correspondientes a las siguientes articulaciones:

- | | |
|--------------------------------------|-------|
| 1) Interfalanges distales de manos | (IFD) |
| 2) Interfalanges proximales de manos | (IFP) |
| 3) Metacarpofalángicas | (MCF) |
| 4) Trapeciometacarpianas | (TMC) |
| 5) Muñecas | (MS) |
| 6) Columna cervical | (CC) |
| 7) Columna Lumbar | (CL) |
| 8) Coxofemorales | (CX) |
| 9) Rodillas | (RX) |

Tres personas examinaron las placas al azar, sin conocer el grupo al que pertenecían con el fin de que no se influenciaran por ello, y, establecieron las siguientes graduaciones de enfermedad articular degenerativa:

- | | |
|---------|---------------------------------------|
| GRADO 1 | Sin evidencia de EAD. |
| GRADO 2 | Evidencia de cambios mínimos de EAD. |
| GRADO 3 | Evidencia de cambios moderados de EAD |

GRADO 4 Evidencia de cambios severos de EAD.

A todos los enfermos se les aplicó un cuestionario para investigar manifestaciones clínicas de enfermedad articular degenerativa tales como historia de artralgias, rigidez matutina, dolor articular al movimiento, crepitación, dolor articular post-reposo, limitación articular al movimiento y Nódulos de Heberden. Hubo sin embargo poca especificidad en las respuestas, por lo que el análisis estadístico se basó sólo en las anomalías radiográficas.

Este estudio fue realizado por el Dr. Gabor Katona del Hospital General de S.S.A. Los datos proporcionados se encuentran detallados en las siguientes páginas.

3.3 DATOS RECOPIADOS DURANTE EL ESTUDIO REALIZADO

Los subgrupos de edad considerados son:

SUBGRUPO	EDAD
1	de 40 a 50 años
2	de 51 a 60 años
3	61 años o más.

GRAFICO DE DIABETICOS

ERADC	JFC	IFF	MCF	TFC	MS	CC	CL	CX	RX	5. DE	EDAD
3	1	1	2	2	3	3	3	2	2		1
3	2	2	2	2	2	3	3	3	2		1
3	3	3	2	2	3	3	2	2	2		1
3	4	3	2	2	3	3	2	3	2		1
3	5	4	2	2	3	2	2	3	2		1
3	6	5	2	2	3	2	2	3	2		1
3	7	5	2	2	3	2	2	3	2		1
3	8	6	2	2	3	2	2	3	2		1
3	9	6	2	2	3	2	2	3	2		1
3	10	7	2	2	3	2	2	3	2		1
3	11	7	2	2	3	2	2	3	2		1
3	12	8	2	2	3	2	2	3	2		1
3	13	8	2	2	3	2	2	3	2		1
3	14	9	2	2	3	2	2	3	2		1
3	15	9	2	2	3	2	2	3	2		1
3	16	10	2	2	3	2	2	3	2		1
3	17	10	2	2	3	2	2	3	2		1
3	18	11	2	2	3	2	2	3	2		1
3	19	11	2	2	3	2	2	3	2		1
3	20	12	2	2	3	2	2	3	2		1
3	21	12	2	2	3	2	2	3	2		1
3	22	13	2	2	3	2	2	3	2		1
3	23	13	2	2	3	2	2	3	2		1
3	24	14	2	2	3	2	2	3	2		1
3	25	14	2	2	3	2	2	3	2		1
3	26	15	2	2	3	2	2	3	2		1
3	27	15	2	2	3	2	2	3	2		1
3	28	16	2	2	3	2	2	3	2		1
3	29	16	2	2	3	2	2	3	2		1
3	30	17	2	2	3	2	2	3	2		1
3	31	17	2	2	3	2	2	3	2		1
3	32	18	2	2	3	2	2	3	2		1
3	33	18	2	2	3	2	2	3	2		1
3	34	19	2	2	3	2	2	3	2		1
3	35	19	2	2	3	2	2	3	2		1
3	36	20	2	2	3	2	2	3	2		1
3	37	20	2	2	3	2	2	3	2		1
3	38	21	2	2	3	2	2	3	2		1
3	39	21	2	2	3	2	2	3	2		1
3	40	22	2	2	3	2	2	3	2		1
3	41	22	2	2	3	2	2	3	2		1
3	42	23	2	2	3	2	2	3	2		1
3	43	23	2	2	3	2	2	3	2		1
3	44	24	2	2	3	2	2	3	2		1
3	45	24	2	2	3	2	2	3	2		1
3	46	25	2	2	3	2	2	3	2		1
3	47	25	2	2	3	2	2	3	2		1
3	48	26	2	2	3	2	2	3	2		1
3	49	26	2	2	3	2	2	3	2		1
3	50	27	2	2	3	2	2	3	2		1
3	51	27	2	2	3	2	2	3	2		1
3	52	28	2	2	3	2	2	3	2		1
3	53	28	2	2	3	2	2	3	2		1
3	54	29	2	2	3	2	2	3	2		1
3	55	29	2	2	3	2	2	3	2		1
3	56	30	2	2	3	2	2	3	2		1
3	57	30	2	2	3	2	2	3	2		1
3	58	31	2	2	3	2	2	3	2		1
3	59	31	2	2	3	2	2	3	2		1
3	60	32	2	2	3	2	2	3	2		1
3	61	32	2	2	3	2	2	3	2		1
3	62	33	2	2	3	2	2	3	2		1
3	63	33	2	2	3	2	2	3	2		1
3	64	34	2	2	3	2	2	3	2		1
3	65	34	2	2	3	2	2	3	2		1
3	66	35	2	2	3	2	2	3	2		1
3	67	35	2	2	3	2	2	3	2		1
3	68	36	2	2	3	2	2	3	2		1
3	69	36	2	2	3	2	2	3	2		1
3	70	37	2	2	3	2	2	3	2		1
3	71	37	2	2	3	2	2	3	2		1
3	72	38	2	2	3	2	2	3	2		1
3	73	38	2	2	3	2	2	3	2		1
3	74	39	2	2	3	2	2	3	2		1
3	75	39	2	2	3	2	2	3	2		1
3	76	40	2	2	3	2	2	3	2		1
3	77	40	2	2	3	2	2	3	2		1
3	78	41	2	2	3	2	2	3	2		1
3	79	41	2	2	3	2	2	3	2		1
3	80	42	2	2	3	2	2	3	2		1
3	81	42	2	2	3	2	2	3	2		1
3	82	43	2	2	3	2	2	3	2		1
3	83	43	2	2	3	2	2	3	2		1
3	84	44	2	2	3	2	2	3	2		1
3	85	44	2	2	3	2	2	3	2		1
3	86	45	2	2	3	2	2	3	2		1
3	87	45	2	2	3	2	2	3	2		1
3	88	46	2	2	3	2	2	3	2		1
3	89	46	2	2	3	2	2	3	2		1
3	90	47	2	2	3	2	2	3	2		1
3	91	47	2	2	3	2	2	3	2		1
3	92	48	2	2	3	2	2	3	2		1
3	93	48	2	2	3	2	2	3	2		1
3	94	49	2	2	3	2	2	3	2		1
3	95	49	2	2	3	2	2	3	2		1
3	96	50	2	2	3	2	2	3	2		1
3	97	50	2	2	3	2	2	3	2		1
3	98	51	2	2	3	2	2	3	2		1
3	99	51	2	2	3	2	2	3	2		1
3	100	52	2	2	3	2	2	3	2		1

PACIENTES CON ENFERMEDAD ARTICULAR DEGENERATIVA

EFACC	IFE	IFF	HCF	TYC	MS	CC	CL	CN	RN	S.DE	EDAD
3	3	3	2	2	0	3	2	2	1		3
4	4	4	2	2	2	3	2	3	3		3
4	3	3	3	2	4	3	3	2	2		3
4	4	3	2	4	3	3	3	2	2		3
4	3	3	3	3	3	4	3	2	3		3
2	2	2	2	2	2	2	2	2	2		3
3	3	2	2	2	2	3	3	3	2		3
4	2	2	2	2	2	3	2	3	2		3
4	2	2	2	4	2	4	3	3	2		3

GELPO CONTROL

EFACC
J
N
N
N
N
N
N
N
N
N
N
N
N
N
N

IFC
C
C
C
C
C
C
C
C
C
C
C
C
C
C

IFF
C
C
C
C
C
C
C
C
C
C
C
C
C
C

MCF
O
O
O
O
O
O
O
O
O
O
O
O
O
O

TFC
C
C
C
C
C
C
C
C
C
C
C
C
C
C

MS
2
0
3
2
0
2
1
2
0
2

CC
2
2
3
2
2
2
3
2
2
3

CL
3
0
0
2
2
2
2
2
2
2

CX
2
0
0
2
1
0
1
1
0
1

RX
2
0
0
2
2
2
2
1
2
2

SIDE EDOA
3
3
3
3
3
3
3
3
3
3

3.4 APLICACION DEL ANALISIS DISCRIMINANTE AL ESTUDIO REALIZADO

El numero total de pacientes se subdividió en la siguiente forma:

GRUPO	CASOS QUE LO INTEGRAN
1	Diabéticos
2	Pacientes con EAD
3	Control

Las variables que se desean incluir dentro del análisis se etiquetaron:

V1	Grado de la enfermedad articular degenerativa
V2	IFD
V3	IFP
V4	MCF
V5	TMC
V6	MS
V7	CC
V8	CL
V9	CX
V10	RX
V11	Subgrupo de edad al -- que pertenecen.

Los principales resultados obtenidos se encuentran en las páginas siguientes.

3.5 PRINCIPALES RESULTADOS OBTENIDOS.

MEANS

	GROUP 1	GROUP 2	GROUP 3	TOTAL
V1	2.0000	2.3200	2.0000	2.7267
V2	2.4000	2.9000	2.8800	2.6867
V3	2.1000	2.3000	2.3400	2.6867
V4	2.3000	2.1000	2.0800	2.1733
V5	2.1000	2.3000	2.0800	2.1733
V6	2.6000	2.2000	2.6000	2.5267
V7	2.4000	2.2000	2.6000	2.5267
V8	2.4000	2.2000	2.7600	2.4000
V9	2.2000	2.2000	2.4000	2.1533
V10	2.2000	2.2000	2.7000	2.1533
V11	2.2000	2.1200	2.7000	2.1546
			1.9600	2.0800

STANDARD DEVIATIONS

	GROUP 1	GROUP 2	GROUP 3	TOTAL
V1	0.6999	0.6227	0.8000	0.8967
V2	1.1000	0.6907	1.1540	1.3134
V3	0.9000	0.6024	1.0600	1.2593
V4	0.9000	0.6024	0.3400	1.1637
V5	1.1000	0.7103	1.0500	1.1627
V6	0.9000	0.8200	0.9781	1.0840
V7	0.9000	0.8224	0.9581	0.9762
V8	0.9000	0.6069	1.0529	1.0345
V9	1.1000	0.6481	0.9381	1.0731
V10	1.1000	0.8100	1.0012	1.2017
V11	0.7585	0.7998	0.8391	0.7987

CUADRO 1

F MATRIX - DEGREES OF FREEDOM: 8, 140

GROUP	GROUP 1	GROUP 2
GROUP	5.69987	3.20014
GROUP	19.6078	

CUADRO 2

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	V3		112.23358	1	0.39573	0.000	224.46717	224.46717	0.000
2	V4		13.39418	2	0.33438	0.000	277.30815	52.84099	0.000
3	V5		6.63871	3	0.30633	0.000	315.36339	38.05524	0.000
4	V8		2.49410	4	0.29607	0.000	330.01423	14.65084	0.001
5	V10		1.83462	5	0.28866	0.000	341.95066	11.93643	0.003
6	V9		1.50694	6	0.28266	0.000	347.34594	5.39528	0.067
7	V1		1.53086	7	0.27666	0.000	350.81140	3.46547	0.177
8	V11		1.35673	8	0.27140	0.000	353.94795	3.13658	0.208

Discriminant function	eigenvalue	relative percentage	canonical correlation	functions derived	Wilks' lambda	chi-square	DF	significance
1	2.28675	94.97	0.834	0	0.2714	187.149	16	0.000
2	0.12106	5.03	0.329	1	0.8920	18.399	7	0.022

CUADRO 4

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

VARIABLE	ENTRY CRITERION	F TO REMOVE
V1	1.53086	1.77988
V3	112.23358	23.48028
V4	13.39418	13.81666
V5	6.63871	6.86427
V8	2.49410	2.66807
V9	1.50694	1.11631
V10	1.83462	1.92086
V11	1.35673	1.35673

F LEVEL INSUFFICIENT FOR FURTHER CORRELATION

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

VARIABLE	TOLERANCE	F TO ENTER	ENTRY CRITERION
V2	0.39573	1.81608	9.49173
V6	0.27703	1.81191	1.92210
V7	0.27435	1.10512	0.29208

CUADRO 3

STAIR WINDING DISCRETE FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
V1	0.01124	-0.000205
V3	-0.00117	-1.000007
V4	-0.00117	-1.000007
V5	0.00117	1.000007
V8	-0.00117	-1.000007
V9	-0.00117	-1.000007
V10	0.00117	1.000007
V11	-0.00117	-1.000007

CUADRO 5

UNSTABILIZED DISCRETE FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
V1	0.001452	-1.00145
V3	-0.001452	-1.00145
V4	-0.001452	-1.00145
V5	0.001452	1.00145
V8	-0.001452	-1.00145
V9	-0.001452	-1.00145
V10	0.001452	1.00145
V11	-0.001452	-1.00145
CONSTANT	1.001452	1.00145

CENTROIDS OF GROUPS IN REDUCED SPACE

	FUNC 1	FUNC 2
GROUP 1	-0.000003	-0.45196
GROUP 2	-0.000003	-0.31265
GROUP 3	0.000003	-0.13833

CUADRO 6

PREDICTION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP		
		GP. 1	GP. 2	GP. 3
GROUP 1	50	27 54.0%	14 28.0%	9 18.0%
GROUP 2	50	11 22.0%	39 78.0%	0 0.0%
GROUP 3	50	3 6.0%	0 0.0%	47 94.0%

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED: 75.33%

CUADRO 7

3.6 ANALISIS DE LOS RESULTADOS OBTENIDOS

Las variables que presentan poder discriminatorio son:

V1	Grado de la enfermedad articular degenerativa
V2	IFP
V4	MCF
V5	TMC
V8	CL
V9	CX
V10	RX
V11	Subgrupo de edad al que pertenecen.

Las variables no introducidas dentro del análisis son:

V2	IFD
V6	MS
V7	CC

Vease el cuadro 3.

Tomando en cuenta las variables que se introdujeron durante el Análisis Discriminante, se produce una ligera separación entre el grupo de diabéticos y el grupo de enfermos con EAD. La separación entre el grupo de diabéticos y el control, es menor que el existente entre el grupo de enfermos con EAD y el control (cuadro 2).

Se derivaron dos funciones discriminantes. Las funciones discriminantes con coeficientes estandarizados son:

$$D1 = 0.0132 V1 - 0.66810 V3 - 0.58909V4 + 0.34244 V5 - 0.19591V8 - \\ -0.08356V9 + 0.1848 V10 - 0.00241V11$$

$$D2 = -0.89793V1 + 1.44487 V3 - 1.37166V4 + 0.02396 V5 - 0.03853V8 + \\ +0.67695V9 - 0.01826V10 + 0.45473V11$$

Las funciones discriminantes con coeficientes no estandarizados son:

$$D1 = 0.01452V1 - 0.053055V3 - 0.53376V4 + 0.29451 V5 - 0.18937V8 - \\ -0.07787V9 + 0.15386V10 - 0.00302V11 + 1.33561$$

$$D2 = -1.00143V1 + 1.1474 V3 - 1.24282V4 + 0.02061 V5 - 0.03724V8 + \\ +0.63085V9 + 0.01519V10 + 0.56937V11 + 0.12384$$

En la función discriminante D1 las variables que proporcionan mayor contribución a la discriminación corresponden a V3, V4 y V5. En la segunda función discriminante la proporcionan V1, V3, V4, V9 y V11.

La importancia relativa de la función D1 es de 94.97 % y de D2 de 5.03. La correlación canónica asociada a la primera función es de 0.834 que es bastante buena y para la segunda de 0.329 (cuadro 4).

Los resultados relativos a la clasificación de nuevos casos se presentan en el cuadro 7 donde se reporta:

- 1) El grupo mejor clasificado es el grupo 3 con el 94% de casos clasificados correctamente.
- 2) El grupo peor clasificado es el grupo 1 con 54.0% de casos clasificados correctamente.
- 3) Las mayores fuentes de error en la clasificación se encuentran en:

a) La clasificación de individuos del grupo 1 como perteneciente al grupo 2 con el 28% de los casos.

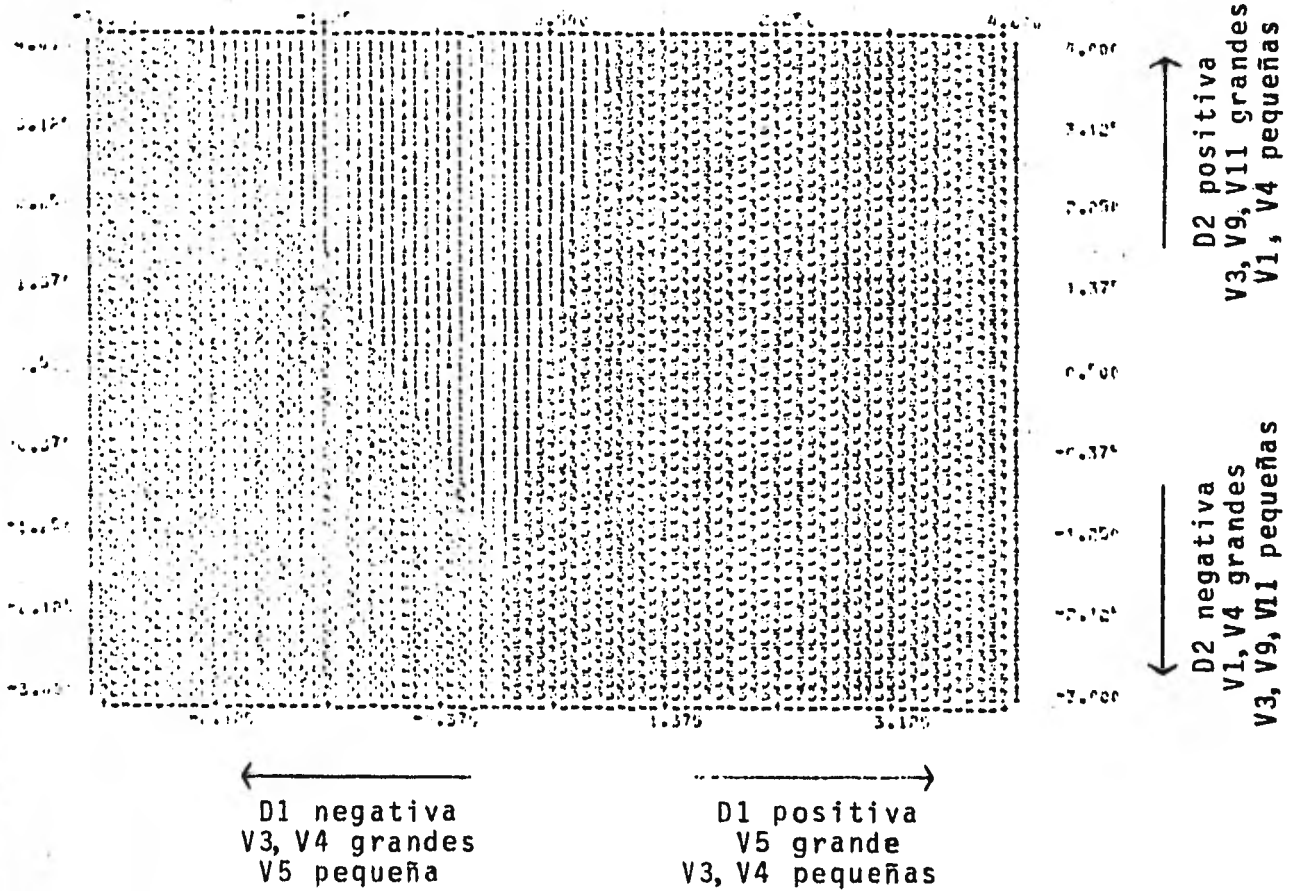
b) La clasificación de individuos del grupo 2 como provenientes del grupo 1 con un 22.0% de los casos.

Estos errores son debidos al parecido existente entre los grupos 1 y 2.

4) Tomando en cuenta el número total de casos, se clasificó correctamente el 75.33%.

3.7 CARACTERIZACION DE LOS GRUPOS EN BASE AL MAPA TERRITORIAL

La interpretación se efectuará a partir de las funciones discriminantes D1 y D2.



DESCRIPCION DE LOS GRUPOS

GRUPO DE DIABETICOS (grupo 1). Se caracteriza por presentar valores grandes en:

V3 Articulaciones metacarpolángicas

V9 Rodillas

y pequeños en:

V5 Muñecas

GRUPO DE ENFERMOS CON EAD (grupo 2). Presenta los valores mayores en:

VI Interfalanges distales de manos

V4 Articulaciones trapeciometacarpianas

y pequeños en:

V5 Muñecas

V9 Rodillas

GRUPO CONTROL (grupo 3). Presenta valores grandes en:

VI Interfalanges distales de manos

y pequeños en:

V3 Metacarpofalángicas

V4 Trapeciometacarpianas

3.8 CONCLUSIONES

La semejanza existente entre el grupo de diabéticos y de enfermos con EAD, y la separación de estos grupos con el control muestran que los diabéticos sufren la enfermedad EAD, en forma más frecuente y severa que las personas que no padecen este tipo de enfermedad.

Los diabéticos presentan mayores afectaciones en las articulaciones metacarpofalángicas y en las rodillas, en cambio los enfermos con EAD las presentan en las interfalanges distales de manos y en las articulaciones trapeciometacarpianas.

B I B L I O G R A F I A

- 1) Afifi, A & Azen S, P. Statistical Analysis, A Computer Oriented Approach. USA, Academic Press, Inc, 1972.
- 2) Anderson, T.W. An Introduction to Multivariate Statistical Analysis. New York, John Wiley & Sons, 1958.
- 3) Aranda Ordaz, Francisco Javier & Méndez, Ignacio. Análisis - Discriminante. México, IIMAS, 1977 (Comunicaciones Técnicas, Serie Naranja, No. 51).
- 4) Bock, D. Multivariate Statistical Methods in the Behavioral Research. USA, McGraw Hill, 1975.
- 5) Bolch, Ben & Cliff, J. Multivariate Statistical Methods for Business and Economics. USA, Prentice-Hall, Inc, 1974.
- 6) Dhrymes, Phoebus J. Econometrics: Statistical Foundations - and Applications. New York, Harper and Row, 1977.
- 7) Marriot, F.H.C. The Interpretation of Multiple Observations. London, Academic Press, 1974.
- 8) Maxwell, A.E. Multivariate Analysis in Behavioral Research. London, Chapman and Hall, 1977.

- 9) Méndez, Ignacio & Rodríguez, Silvia. Dos Ejemplos de Análisis Discriminante en Medicina. México, IIMAS, 1978. (Comunicaciones Técnicas, Serie Naranja, No. 179).
- 10) Morrison, Donald F. Multivariate Statistical Methods. USA, McGraw Hill, 1967.
- 11) Nie, Norman H. Et Al. Spss: Statistical Package for the Social Sciences. 2 ed. New York, McGraw Hill, 1977.
- 12) Overall, John & Klett, James. Applied Multivariate Analysis. USA, McGraw Hill, 1972.
- 13) Srivastava, Khatri. An-Introduction to Multivariate Statistics. New York, Elsevier North Holland, Inc, 1979.
- 14) Steven, Arnold. The Theory of Linear Models and Multivariate Analysis. USA, John Willey & Sons, Inc, 1981.
- 15) Tatsuoka, Maurice M. Analysis Multivariate. New York, McGraw Hill, 1974.