



UNIVERSIDAD DE PANAMA

VICERECTORIA DE INVESTIGACION Y POSTGRADO

PROGRAMA CENTROAMERICANO DE MAESTRIA EN MATEMATICA

EL ANALISIS DISCRIMINANTE Y SU APLICACIÓN

CARMEN CECILIA RODRIGUEZ MARTINEZ

**TESIS PRESENTADA COMO UNO DE LOS REQUISITOS PARA OPTAR AL
GRADO DE MAESTRO EN CIENCIAS CON ESPECIALIZACION EN
ESTADISTICA MATEMATICA**

PANAMA, REPUBLICA DE PANAMA

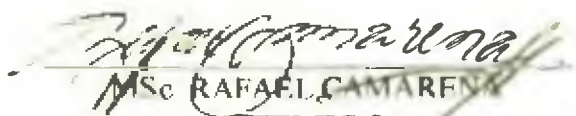
1999

**DIGITALIZADO
DEPTO. DE COMPUTO
SIBIUP**

APROBADO POR:



MSc. AURORA MEJIA
PRESIDENTA



MSc. RAFAEL CAMARENA
MIEMBRO



MSc. PLUTARCO RAMOS
MIEMBRO


REPRESENTANTE DE LA VICERRECTORIA
DE INVESTIGACION Y POSTGRADO

Fecha: 29 de diciembre 1997

322/55

Obsequio

del Autor

T.M.

DEDICATORIA

A *Dios*, por estar a mi lado en cada paso que doy.

A mis padres *Carlos y Zeneida*, quienes ha dedicado su vida a sus hijos y son y serán siempre mi gran inspiración.

A mis hermanos *Vivian, Nellyn, Carlitos, Raúl, y Lucy*.

A mis sobrinos *Arnold, Chuck, Grisnelly y Jacob*.

A mi gran amiga *Lamed G. Mendoza Lámbiz*,
quien siempre me estimula y apoya.

A *J.A.*, recuerda que nunca es tarde para empezar.

Con todo mi amor
Cecy.

AGRADECIMIENTO

Quiero expresar mi agradecimiento, a todas las personas que siempre me han brindado su apoyo para la culminación de este trabajo de investigación.

A la profesora **Aurora Mejía C.**, mi profesora asesora, por su gran ayuda y apoyo.

A la Dra. **Jazmín Arosemena**, por su colaboración en la obtención de los datos para la aplicación del método investigado.

Al profesor **Manuel A. Tejada S.**, por su apoyo y pequeños pero importantes consejos.

A **Nadir** mi gran amigo.

A mis compañeras y compañeros de trabajo **Daniel, Elena, Mitzi, Elisa, y Gonzalo.**

A todos **GRACIAS.**

INDICE GENERAL

INDICE

	Página
RESUMEN.....	1
INTRODUCCIÓN.....	2
 CAPITULO I: ANALISIS DISCRIMINANTE	
Principios de Discriminación.....	6
El Modelo Estadístico.....	7
Matriz Centroide.....	8
Función Discriminante Lineal de Fisher.....	13
La Distancia de Mahalanobis.....	19
La Distancia de Mahalanobis muestral.....	19
 CAPITULO II. CLASIFICACION GENERAL	
Clasificación en una de dos poblaciones Normales Multivariadas con parámetros conocidos.....	24
Función discriminante para dos poblaciones.....	25
Distribución de la Función discriminante.....	25
Procedimiento Intuitivo.....	26
Discriminación cuando las poblaciones son conocidas..	28
Regla discriminante de máxima verosimilitud.....	28
Regla Bayesiana discriminante.....	30
Probabilidad de Asignación.....	31
Regla discriminante aleatoria.....	31
Mejor regla discriminante.....	32
 CAPITULO III: INTERPRETACION DE LA FUNCION DISCRIMINANTE	
Interpretación de las funciones discriminantes.....	34
Cálculo de los puntajes discriminantes.....	34
Diagrama de dos funciones.....	36
Diagrama de una función.....	38

Coeficientes estandarizados.....	39
Estructura de los coeficientes.....	41

CAPITULO IV: ANALISIS E INTERPRETACION DE LOS RESULTADOS

Niveles de Nutrición.....	46
Indicadores de Desnutrición.....	47
Descripción de las variables.....	48
Medias y desviaciones estándar por estado nutricional conocido.....	49
Matriz de correlación total.....	50
Matriz de correlación dentro de grupos.....	52
Funciones de clasificación: Función discriminante lineal de Fisher.....	52
Matriz de clasificación.....	53
Coeficientes discriminantes crudos.....	54
Coeficientes discriminantes estandarizados.....	55
Matriz de estructura total.....	56
Grado de significancia.....	57

CONCLUSIONES Y RECOMENDACIONES.....	68
BIBLIOGRAFIA.....	72

ÍNDICE DE TABLAS Y FIGURAS

INDICE DE TABLAS

		Página
Tabla I.	Niveles de Nutrición.....	46
Tabla II.	Descripción de las Variables.....	48
Tabla III.	Medias y desviaciones estándar por estado nutricional conocido.....	49
Tabla IV.	Matriz de correlación total.....	50
Tabla IVa.	Matriz de correlación total.....	51
Tabla V.	Matriz de correlación dentro de grupos.....	52
Tabla VI.	Funciones de clasificación: Función discriminante lineal de Fisher.....	52
Tabla VII.	Matriz de clasificación.....	53
Tabla VIII.	Coefficientes discriminantes crudos.....	54
Tabla IX.	Coefficientes discriminantes estandarizados.	55
Tabla X.	Matriz de estructura total.....	56
Tabla XI.	Prueba de significancia para las funciones sucesivas.....	57

ÍNDICE DE FIGURAS

	Página
Figura 1.	Distribución de y cuando x pertenece a la población 1 y cuando x pertenece a la población 2..... 27
Figura 2.	Diagrama de dos funciones..... 36
Figura 3.	Histograma de grupo..... 38
Figura 4.	Gráfico de probabilidad de la variable edad, para cada uno de los estados nutricionales..... 59
Figura 5.	Gráfico de probabilidad de la variable peso inicial, para cada uno de los estados nutricionales..... 60
Figura 6.	Gráfico de probabilidad de la variable talla inicial, para cada uno de los estados nutricionales 61
Figura 7.	Gráfico de probabilidad de la variable Hemoglobina(Hb), para cada uno de los estados nutricionales..... 62
Figura 8.	Gráfico de probabilidad de la variable Heritrocito(Hto), para cada uno de los estados nutricionales 63
Figura 9.	Mapa Territorial..... 64

Figura 10.	Funciones discriminantes.....	65
Figura 11.	Funciones discriminantes y centroides de grupo.....	66

RESUMEN

Este trabajo está enfocado a desarrollar la teoría del **Análisis Discriminante** en una forma matricial, tal que el estudio de la misma, pueda ser llevado de una manera menos complicada. Se iniciará con la presentación de los conceptos generales sobre el **Análisis Discriminante**, se demostraran los teoremas que conllevan esta teoría y a todo esto se le dará un enfoque matricial. Por último se presentará el estudio en el que se hará uso del **Análisis Discriminante**. Se realizó un análisis discriminante en el estudio de la desnutrición infantil en estudiantes de escuelas públicas del Distrito de La Chorrera. Datos suministrados por el Centro de Salud Magally.

SUMMARY

This work is focused to develop the theory of the Discriminant Analysis in a matricial form, of such a form that the study of the same, could be taken less complicated. Begin with the presentation of the general concepts on the Discriminant Analysis, the theorems are demonstrated that they bear this theory and to all this will give to you a focus matricial. Lastly will come the study in which uses of the Discriminant Analysis. We carried out an discriminant analysis in the study of the infantile malnutrition in students of public schools of the District of La Chorrera. Data given by the Center of Health Magally Ruiz.

INTRODUCCION

Un estudio de la literatura en el análisis multivariado indica que el desarrollo de la metodología del análisis Discriminante resulta de los esfuerzos de muchos estadísticos diferentes incluyendo a **R.A Pearson, P.C. Mahalanobis, M.S. Bartlett, y C.R. Rao**. Se piensa que la idea de una función Discriminante empezó con **Karl Pearson** alrededor de 1920 cuando propuso el “coeficiente de semejanza racial” que determinaba la distancia estadística entre dos muestras. Descontento con el coeficiente de Pearson, **P.C. Mahalanobis** propone el estadístico D^2 como alternativa. El estadístico D^2 , ahora normalmente llamado distancia de Mahalanobis, se aplicó primeramente a un estudio de mezclas raciales en la India en 1925, pero más recientemente se ha vuelto una parte fundamental de varios procedimientos del método multivariado. El Análisis Discriminante fue introducido por Fisher

en 1936, es una técnica joven, en sus principios se relacionó con investigaciones biológicas y antropométricas.

En sus inicios el Análisis Discriminante fue un método de análisis que se aplicó en las ciencias médicas y biológicas(ecología), despertó mucho interés en las metodologías estadísticas de áreas tales como Negocios, Educación, Ingeniería y Psicología.

Este tipo de análisis es de utilidad cuando la variable dependiente es categórica y las variables independientes o de predicción se presentan en escala de intervalo. Los supuestos en el análisis discriminante son de que cada una de las poblaciones tienen la misma matriz de varianza-covarianza.

CAPÍTULO I

ANÁLISIS DISCRIMINANTE

PRINCIPIOS DE DISCRIMINACIÓN

La familia de técnicas desarrolladas para describir y clasificar individuos, caracterizadas por un número importante de variables, se conoce como **Análisis Discriminante**. El origen de este método, se remonta a los trabajos de Mahalanobis y Fisher (1936).

El problema básico en el **Análisis Discriminante** es el de asignar un individuo desconocido a uno de dos o más grupos o poblaciones¹, en base a observaciones multivariadas. El procedimiento comprende la determinación de varias funciones discriminantes, que son simplemente combinaciones lineales de las variables independientes, que separan o discriminan entre los grupos o poblaciones tanto como sea posible. El número de funciones discriminantes necesarias está dado por un estadístico

de prueba multivariado llamado **Lambda de Wilk's**. Los coeficientes de las funciones discriminantes pueden estandarizarse y luego interpretarse de la misma manera que en la regresión múltiple para extraer conclusiones

¹ Para el caso de más de dos grupos se puede tratar el análisis discriminante como un caso especial de correlación canónica.

acerca de que variables son importantes en la discriminación entre los grupos o poblaciones.

Aunque el **Análisis Discriminante** se emplea con mayor frecuencia para explicar o describir factores que discriminen entre los grupos de interés (Análisis Discriminante Descriptivo. ADD), también puede usarse para clasificar individuos a futuro (Análisis Discriminante Predictivo. ADP). La clasificación comprende el determinar una ecuación de predicción separada, para cada grupo, que da la probabilidad de pertenecer al grupo. Para clasificar un individuo a futuro, se calcula una predicción para cada grupo y el individuo se clasifica como perteneciente al grupo que tiene las características más parecidas o semejantes a las suyas, esta asignación se realiza con un mínimo de error.

EL MODELO ESTADÍSTICO

Consideremos g poblaciones $\Pi^1, \Pi^2, \dots, \Pi^g$ cada una de ellas distribuida $N_p(\mu^i, \Sigma^i)$ para $i=1,2,\dots,g$. Con matriz de covarianza $\Sigma^1 = \Sigma^2 = \dots = \Sigma^g = \Sigma$ y de rango completo.

Sean x_{jk}^i , la j -ésima observación de la población i sobre la variable k con $j=1,2,\dots,n^i$; $i=1,2,\dots,g$; $k=1,2,\dots,p$.

La matriz que representa el conjunto de datos X esta formada por la yuxtaposición de estas g matrices, cada una de ellas con $n = \sum_{i=1}^g n^i$ filas ,

$$X = \left[\begin{array}{c|c} X^1 & n^1 \uparrow \\ X^2 & n^2 \uparrow \\ \vdots & \vdots \uparrow \\ \cdot & \cdot \downarrow \\ \cdot & \cdot \downarrow \\ X^g & n^g \downarrow \end{array} \right]$$

$\leftarrow \quad p \quad \rightarrow$

donde X^i es la matriz de datos de $n^i \times p$ de la i -ésima población con $X^i = (x_{jk}^i)$, $j=1,2,\dots,n^i$; $k=1,2,\dots,p$

Veamos algunas definiciones que nos permitirán ver más claramente como es el modelo estadístico.

DEFINICIÓN I.1. Matriz Centroide

Llamaremos matriz centroide a:

$$C_n = I_n - n^{-1} J_n$$

Donde I_n es la matriz identidad y $J_n = \mathbf{1}\mathbf{1}'$ tiene unos como elementos, ambas de orden $n \times n$.

Propiedades de la Matriz Centroide:

PROPOSICIÓN I.1.

Sea C_n una matriz centroide de orden $n \times n$, entonces:

(a) $C' = C$

(b) $CC = C$

(c) $C\mathbf{1} = 0$

d) $CJ = JC = 0$

e) $Cx = x - x\mathbf{1}$

f) $x'Cx = \sum_{i=1}^n (x^i - \bar{x})^2$

Demostración :

Sea $C = I - n^{-1}J$ una matriz centroide de orden $n \times n$. Para demostrar la propiedad a , consideremos la transpuesta de C , esto es :

$$\begin{aligned} C' &= (I - n^{-1}J)' \\ &= (I' - (n^{-1}J)') \\ &= I' - n^{-1}J' \quad \text{por ser } n^{-1} \text{ un escalar} \end{aligned}$$

$= I - n^{-1}J$ la transpuesta de la identidad es igual a la
 I ,
 la transpuesta de J es ella misma, dado que
 J tiene unos como elementos.

Lo que nos indica que C es una matriz simétrica. ✱

Para la propiedad **b**, calculemos el producto de CC es decir

$$\begin{aligned}
 CC &= (I - n^{-1}J)(I - n^{-1}J) \\
 &= II - n^{-1}JI - n^{-1}IJ + (n^{-1})^2 JJ \\
 &= I - n^{-1}J - n^{-1}J + n^{-2}nJ \\
 &= I - 2n^{-1}J + n^{-1}J \\
 &= I - n^{-1}J
 \end{aligned}$$

lo que demuestra que C es idempotente. ✱

Ahora veamos que $C1=0$.

$$\begin{aligned}
 C1 &= (I - n^{-1}J)1 \\
 &= I1 - n^{-1}J1 \\
 &= 1 - n^{-1}n1 \\
 &= 1 - 1 \\
 &= 0
 \end{aligned}$$

es decir C se anula al multiplicarla por el vector 1 a la derecha ✱

En el caso de la parte **d**, sólo demostraremos que $\mathbf{JC} = \mathbf{0}$, ya que el resultado $\mathbf{CJ} = \mathbf{0}$ es inmediato del resultado anterior.

$$\begin{aligned}
 \mathbf{JC} &= \mathbf{J}(\mathbf{I} - \mathbf{n}^{-1}\mathbf{J}) \\
 &= \mathbf{J}\mathbf{I} - \mathbf{J}\mathbf{n}^{-1}\mathbf{J} \\
 &= \mathbf{J} - \mathbf{m}\mathbf{n}^{-1}\mathbf{J} \\
 &= \mathbf{J} - \mathbf{J} \\
 &= \mathbf{0}
 \end{aligned}$$



Ahora consideremos el producto \mathbf{Cx} , para demostrar la propiedad **e**

$$\begin{aligned}
 \mathbf{Cx} &= (\mathbf{I} - \mathbf{n}^{-1}\mathbf{J})\mathbf{x} \\
 &= \mathbf{Ix} - \mathbf{n}^{-1}\mathbf{Jx} \\
 &= \mathbf{x} - \bar{\mathbf{x}}\mathbf{1}
 \end{aligned}$$



Consideremos la forma cuadrática $\mathbf{x}'\mathbf{Cx}$, para la demostración de la propiedad **f**.

$$\begin{aligned}
 \mathbf{x}'\mathbf{Cx} &= \mathbf{x}'(\mathbf{I} - \mathbf{n}^{-1}\mathbf{J})\mathbf{x} \\
 &= \mathbf{x}'[\mathbf{x} - \mathbf{x}\mathbf{1}] \text{ por la propiedad e.} \\
 &= \mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{x}\mathbf{1}
 \end{aligned}$$

$$= \left[\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i \right]$$

$$= \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]$$

A continuación veamos un resultado particular de la propiedad e.

PROPOSICION I.2.

Sea C una matriz centroide, y X de $n \times p$, entonces

$$CX = (x - \bar{x})_{ij}$$

Demostración:

Sea $C = I - n^{-1}J$ y X de $n \times p$.

$$\begin{aligned} \text{Entonces } CX &= (I - n^{-1}J)X \\ &= IX - n^{-1}JX \\ &= X - \bar{X} \end{aligned}$$

donde la j-ésima columna de $X_{n \times p}$ es x_j . X tiene como elementos $x_{ij} - \bar{x}_j$;

donde \bar{x}_j es la media de la columna j de X. *

**DEFINICIÓN I.2 La Función Discriminante Lineal de
Fisher**

Sea

$$y_{n \times 1} = X_{n \times p} a_{p \times 1} = \begin{bmatrix} X^1_a \\ X^2_a \\ \cdot \\ \cdot \\ \cdot \\ X^g_a \end{bmatrix} = \begin{bmatrix} y^1 \\ y^2 \\ \cdot \\ \cdot \\ \cdot \\ y^g \end{bmatrix}$$

donde $y^i = X^i a$, con $i=1,2,\dots,g$, es una combinación lineal de las columnas de X . La media de las n^i observaciones en la variable k en X^i es:

$$\frac{\sum_{j=1}^{n^i} X^i_{jk}}{n^i} = \frac{1' X^i_k}{n^i}$$

con X^i_k como el vector de observaciones de la variable k en la i -ésima población, el vector fila de todas las k medias observadas esta dado por:

$$\bar{X}^i = [\bar{X}^i_1, \bar{X}^i_2, \dots, \bar{X}^i_p] = \frac{1}{n^i} 1' X^i$$

luego entonces,

$$\bar{X}' = (\bar{X}^1, \bar{X}^2, \dots, \bar{X}^p)$$

es el vector de medias de la matriz X .

Entonces consideremos la forma cuadrática:

$$y'Cy = a'X'CXa = a'T'a \quad \text{con } T' = X'CX$$

que puede ser particionada como la suma de cuadrados dentro de grupos

$$\sum_{i=1}^g y^{i'} C^i y^{i'} = \sum_{i=1}^g a' X^{i'} C^i X^i a = a' D a$$

más la suma de cuadrados entre los grupos

$$\sum_{i=1}^g n^i (\bar{y}^i - \bar{y})^2 = \sum_{i=1}^g n^i \left\{ a' (\bar{x}^i - \bar{x}') \right\}^2 = a' E a$$

con

$$D = \sum_{i=1}^g X^{i'} C^i X^i \quad \text{y} \quad E = \sum_{i=1}^g n^i (\bar{x}^i - \bar{x})(\bar{x}^i - \bar{x})'$$

y además, donde \bar{y}^i es la media del i -ésimo sub-vector y^i de y , es decir

$$\bar{y}^i = \frac{1}{n^i} \sum_{j=1}^{n^i} y_j^i, \quad \text{con } C^i \text{ una matriz centroide de } n^i \times n^i.$$

Así, la razón dada por:

$$\frac{a' D a}{a' E a}$$

puede ser maximizada con un vector \mathbf{a} , entonces la función lineal $\mathbf{X}\mathbf{a}$, se llamara **Función Lineal Discriminante de Fisher** o primera variable canónica.

PROPOSICIÓN I.3

Sea $\mathbf{y} = \mathbf{X}\mathbf{a}$, entonces la forma cuadrática $\mathbf{y}'\mathbf{C}\mathbf{y} = \mathbf{a}'\mathbf{T}\mathbf{a}$, es la suma total de cuadrados, que puede ser particionada como la suma de cuadrados dentro de grupos $\mathbf{a}'\mathbf{D}\mathbf{a}$ y la suma de cuadrados entre grupos $\mathbf{a}'\mathbf{E}\mathbf{a}$.

Demostración:

Consideremos la forma cuadrática $\mathbf{y}'\mathbf{C}\mathbf{y}$ con \mathbf{y} de $n \times 1$, y \mathbf{C} de $n \times n$. Sea $\mathbf{1}^i$ un vector de $n \times 1$ con 1 en las correspondientes posiciones de la i -ésima

población y 0 en otro lado, y sea además $\mathbf{\Gamma}^i = \text{diag}(\mathbf{1}^i)$, con $\sum_{i=1}^g \mathbf{\Gamma}^i = \mathbf{I}_{n \times n}$

y también $\sum_{i=1}^g \mathbf{1}^i = \mathbf{1}_{n \times 1}$, sea \mathbf{C}^i una matriz centroide de $n \times n$, luego

$$\mathbf{C}^i = \mathbf{\Gamma}^i - 1/n^i \mathbf{1}^i \mathbf{1}^{i'}$$

Entonces podemos escribir la matriz \mathbf{C} de la siguiente forma,

$$\begin{aligned}
C &= I - n^{-1} 11' \\
&= \sum_{i=1}^g I^i - \sum_{i=1}^g (n^{i-1} 1^i 1^{i'}) + \sum_{i=1}^g n^{i-1} 1^i 1^{i'} - n^{-1} 11' \\
&= \sum_{i=1}^g [I^i - n^{i-1} 1^i 1^{i'}] + \sum_{i=1}^g n^{i-1} 1^i 1^{i'} - n^{-1} 11'
\end{aligned}$$

luego tenemos que

$$\begin{aligned}
y'Cy &= y'I - n^{-1} 11'y \\
&= y' \left[\sum_{i=1}^g (I^i - n^{i-1} 1^i 1^{i'}) + \sum_{i=1}^g n^{i-1} 1^i 1^{i'} - n^{-1} 11' \right] y \\
&= y' \left[\sum_{i=1}^g (I^i - n^{i-1} 1^i 1^{i'}) \right] y + y' \left[\sum_{i=1}^g (n^{i-1} 1^i 1^{i'}) - n^{-1} 11' \right] y \\
&= y' \sum_{i=1}^g C^i y + y' \left[\sum_{i=1}^g \bar{y}^i 1^i - \bar{y} 1 \right] \\
&= \sum_{i=1}^g (y^i{}' C^i y^i) + \sum_{i=1}^g n^i \bar{y}^i{}^2 - n \bar{y}^2 \\
&= \sum_{i=1}^g (a' X^i{}' C^i X^i a) + \sum_{i=1}^g n^i (\bar{y}^i - \bar{y})^2 \\
&= \sum_{i=1}^g (a' X^i{}' C^i X^i a) + \sum_{i=1}^g n^i \{a' \bar{x}^i - a' \bar{x}\}^2 \\
&= \sum_{i=1}^g (a' X^i{}' C^i X^i a) + a' \left\{ \sum_{i=1}^g n^i (\bar{x}^i - \bar{x})(\bar{x}^i - \bar{x})' \right\} a \\
&= a' D a + a' E a
\end{aligned}$$

TEOREMA I.1 Mayores Valores Propios

Sea $y=Xa$ la función discriminante lineal de Fisher, entonces el vector a en la función discriminante lineal de Fischer es el vector propio de $E^{-1}D$ correspondiente a los mayores valores propios.

Demostración:

Sean E y D dos matrices simétricas. Supongase además que E de orden $n \times n$ en donde para $\forall i, j$ $e_{ij} > 0$.

Consideremos la matriz $E^{1/2}$, que denota la raíz cuadrada de la matriz simétrica E , y sea $y=E^{1/2} a$.²

Entonces el máximo de $a'Ea$ sujeta a que $a'Da = 1$ puede escribirse como:

$$\max y'E^{-1/2} D E^{-1/2} y \quad \text{sujeta a } y'y = 1$$

Por otro lado sea

$E^{-1/2}DE^{-1/2} = \Gamma\Lambda\Gamma'$ (teorema de Jordan³) una descomposición espectral de la matriz simétrica

$$E^{-1/2} D E^{-1/2}$$

² Mejía.(1988).

³ Teorema de Jordan: Cualquier matriz simétrica $A_{p \times p}$ puede ser escrita como $A=\Gamma\Lambda\Gamma' = i \sum_{i=1}^p \lambda_i \gamma(i) \gamma'(i)$ donde Λ es la matriz diagonal de valores propios de A y Γ es una matriz ortogonal cuyas columnas son vectores propios estandarizados.

Ahora consideremos la transformación $z = \Gamma'y$. De donde $z'z = y'\Gamma\Gamma'y$ luego podemos escribir

$$\max z'\Lambda z = \max \sum \lambda^i z_i^2 \quad \text{sujeta a } z'z=1$$

Si los valores propios son escritos de forma descendente entonces la igualdad anterior satisface

$$\max \sum \lambda^i z_i^2 \leq \lambda^1 \max \sum z_i^2 = \lambda^1$$

Este límite atiende al hecho de que $z=(1,0,\dots,0)$, esto se da para $y = \gamma^{(1)}$ y para $a = E^{-1/2} \gamma^{(1)}$.

Sabemos que $E^{-1}D$ y $E^{-1/2}DE^{-1/2}$ tienen los mismos valores propios ⁴ y $a = E^{-1/2} \gamma^{(1)}$ es un vector propio de $E^{-1/2}D$ correspondiente a λ^1

Una vez que la función discriminante lineal ha sido calculada, una observación x puede ser asignada a una de las g poblaciones en base a sus “puntajes discriminantes” x_a . Las medias muestrales \bar{x}^i tienen puntajes $\bar{x}^i_a = \bar{y}^i$. Entonces x es asignada a la población cuyo puntaje de medias sean más cercanos a x_a , es decir, se asigna x a Π^j si la distancia de x_a a x_j_a es menor que la distancia de x_a a x_i_a , para todo $i \neq j$

DEFINICIÓN I.3 La Distancia de Mahalanobis

La distancia de Mahalanobis entre 2 poblaciones con medias μ^1 y μ^2 y matriz de covarianza común Σ se define como

$$\Delta^2 = (\mu^1 - \mu^2)' \Sigma^{-1} (\mu^1 - \mu^2)$$

Dado que es muy raro conocer los parámetros de la población, usualmente estos se estiman a través de valores muestrales.

DEFINICIÓN I.4 La Distancia de Mahalanobis Muestral

Sean dos muestras de tamaño n^1 y n^2 , donde $n_1 + n_2 = n$.

Entonces la distancia de Mahalanobis muestral es D , definida de la siguiente forma:

$$D^2 = (\bar{X}^1 - \bar{X}^2)' S_u^{-1} (\bar{X}^1 - \bar{X}^2)$$

donde $S_u = (n^1 S^1 + n^2 S^2) / (n-2)$ este es un estimador insesgado de Σ .

La distribución de D^2 sobre un conjunto particular de supuestos esta dado por el teorema I.2.

⁴ Mardia. Teorema A.6.2. (1979).

DEFINICION I.5

Si α puede escribirse como $md'M^{-1}d$ cuando d y M son independientes y distribuidas como $N_p(O,I)$ y $W_p(I,m)$ respectivamente, entonces diremos que α tiene una distribución **T² Hotelling** con parámetros p y m .

$$\alpha \sim T^2_{(p, m)}$$

PROPOSICIÓN I.6

Si M^1 y M^2 son dos matrices aleatorias, tales que $M^1 \sim W(\Sigma, n^1-1)$ y $M^2 \sim W(\Sigma, n^2-1)$ son independientes, entonces:

$$cM = c(M^1 + M^2) \sim W(c\Sigma, n - 2)$$

con $c = n / (n^1 + n^2)$.

Demostración:

Consideremos n^1-1 vectores $c^{1/2}Z_\alpha \sim N_p(0, c\Sigma)$ mutuamente independientes tales que

$$cM^1 = \sum_{\alpha=1}^{n^1-1} c^{1/2}Z_\alpha c^{1/2}Z_\alpha'$$

de igual forma consideremos n^2-1 vectores $c^{1/2}Z_\alpha \sim N_p(0, c\Sigma)$ mutuamente independientes tales que

$$cM^2 = \sum_{\alpha=n^1}^{n^1-1+n^2-1} c^{1/2}Z_\alpha c^{1/2}Z_\alpha'$$

Además como M^1 y M^2 son independientes entonces $\forall 1 \leq \alpha \leq n^1 + n^2 - 2$ los $c^{1/2}Z_\alpha$ son mutuamente independientes.

Consideremos la suma $cM^1 + cM^2$

$$\begin{aligned} cM^1 + cM^2 &= \sum_{\alpha=1}^{n^1-1} c^{1/2}Z_\alpha c^{1/2}Z_\alpha' + \sum_{\alpha=n^1}^{n^1-1+n^2-1} c^{1/2}Z_\alpha c^{1/2}Z_\alpha' \\ &= \sum_{\alpha=1}^{n^1+n^2-2} c^{1/2}Z_\alpha c^{1/2}Z_\alpha' \end{aligned}$$

con $M = M^1 + M^2$ y como $\forall 1 \leq \alpha \leq n^1 + n^2 - 2$ los $c^{1/2}Z_\alpha \sim N_p(0, c\Sigma)$ son mutuamente independientes entonces $cM \sim W(c\Sigma, n-2)$

TEOREMA 1.2

Si X^1 y X^2 son matrices de datos independientes entre si, y si las n^i filas de X^i son independientes y distribuidas $N_p(\mu^i, \Sigma^i)$, $i=1,2$, entonces cuando $\mu^1 = \mu^2$ y $\Sigma^1 = \Sigma^2$,

$$(n^1 n^2 / n) D^2 \sim T^2_{(p, n-2)}$$

Demostración:

Dado que X^i se distribuye $N_p(\mu^i, \Sigma^i)$, para $i=1,2$, la distribución general de $d = \bar{x}^1 - \bar{x}^2$ es normal con media $\mu^1 - \mu^2$ y matriz de covarianza $(1/n^1) \Sigma^1 + (1/n^2) \Sigma^2$. Cuando $\mu^1 = \mu^2$ y $\Sigma^1 = \Sigma^2 = \Sigma$, entonces d se distribuye $N_p(\mathbf{0}, c\Sigma)$, donde $c = n/n^1n^2$.

Si $M^i = n^i S^i$ entonces M^i se distribuye $W_p(\Sigma^i, n^i - 1)$. Luego entonces cuando $\Sigma^1 = \Sigma^2 = \Sigma$,

$$M = (n - 2) S_u = M^1 + M^2 \text{ que se distribuye } W_p(\Sigma, n-2).^5$$

Lo cual nos lleva a concluir que la distribución de cM es $W_p(c\Sigma, n-2)$.

Además, M es independiente de d , y las dos muestras son independientes una de la otra. Por consiguiente

$$(n-2)d'(cM)^{-1}d \text{ se distribuye } T^2(p, n-2).$$

Como $d = \bar{x}^1 - \bar{x}^2$ y $M = (n-2)S_u$ tenemos que

$$(n^1 n^2 / n) (\bar{x}^1 - \bar{x}^2)' S_u^{-1} (\bar{x}^1 - \bar{x}^2) = (n^1 n^2 / n) D^2$$

luego $(n^1 n^2 / n) D^2 \sim T^2(p, n-2)$.

⁵ Mejía. (1988).

CAPÍTULO II
CLASIFICACIÓN GENERAL

CLASIFICACIÓN GENERAL

Clasificación en una de dos poblaciones normales multivariadas con parámetros conocidos.

En el caso de dos poblaciones los supuestos deben ser los siguientes:

- a) Los datos deben provenir de una de dos poblaciones normales multivariadas.
- b) Las matrices de covarianza deben ser iguales.

Sea el vector de observaciones $\mathbf{X} = (X_1, X_2, \dots, X_p)'$ con $\mathbf{X} \sim N(\boldsymbol{\mu}^1, \Sigma)$ o $N(\boldsymbol{\mu}^2, \Sigma)$, se debe considerar la existencia de las dos poblaciones:

$$\Pi^1 \sim N(\boldsymbol{\mu}^1, \Sigma^1)$$

y

$$\Pi^2 \sim N(\boldsymbol{\mu}^2, \Sigma^2)$$

donde

$$\boldsymbol{\mu}^i = ((\mu^i_1, \mu^i_2, \dots, \mu^i_p))' \quad i=1,2.$$

y

$$\Sigma^1 = \Sigma^2 = \Sigma$$

Se debe considerar que $\boldsymbol{\mu}^1$, $\boldsymbol{\mu}^2$, y Σ son conocidas.

DEFINICIÓN II.1. Función Discriminante

La función discriminante esta dada por la siguiente combinación lineal:

$$y' = \mathbf{a}'\mathbf{X}$$

donde \mathbf{a} es un vector de constantes desconocidas.

Para la clasificación se procede de la siguiente forma:

Se clasifica \mathbf{x} en Π^1 si $y \geq h$

o

Se clasifica \mathbf{x} en Π^2 si $y < h$

Donde h es una constante. Luego, debemos determinar los valores de $\mathbf{a}_1, \dots, \mathbf{a}_p$ y h , de forma que las probabilidades de clasificación incorrecta sean las mínimas posibles.

DEFINICIÓN II.2. Distribución de la función discriminante

Si \mathbf{x} pertenece a Π^1 , entonces $y \sim N(\zeta_1, \sigma_y^2)$ donde,

$$\zeta_1 = \sum_{j=1}^p a_j \mu_j^1$$

y

$$\sigma_y^2 = \sum_{i=1}^p \sum_{j=1}^p a_i \sigma_{ij} a_j$$

de la misma forma,

Si \mathbf{x} pertenece a Π^2 , entonces $y \sim N(\zeta_2, \sigma_y^2)$ donde,

$$\zeta_2 = \sum_{j=1}^p a_j \mu_j^2$$

y con la misma varianza σ_y^2 .

DEFINICIÓN II.3. Procedimiento Intuitivo

El procedimiento intuitivo para la determinación de $\mathbf{a}_1, \dots, \mathbf{a}_p$, indica que ζ_1 debe estar suficientemente separado de ζ_2 , tomando en cuenta la varianza de y . Para solucionar esto se define la distancia de Mahalanobis en forma matricial como

$$\Delta^2 = (\boldsymbol{\mu}^1 - \boldsymbol{\mu}^2) \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}^1 - \boldsymbol{\mu}^2)$$

Este estadígrafo fue propuesto por Mahalanobis (1936) como medida de la distancia entre 2 poblaciones. Así, el problema se reduce a encontrar los coeficientes $\mathbf{a}_1, \dots, \mathbf{a}_p$ que maximicen Δ^2 .

Para encontrarlo Fischer (1936) sugirió resolver el siguiente sistema de ecuaciones

$$a = \Sigma^{-1} (\mu^1 - \mu^2)$$

Para determinar la constante h , observemos (Fig.1) que representa la distribución de y cuando pertenece a Π^1 y cuando pertenece a Π^2 .

Se denota la probabilidad de cometer el error de clasificar a x en Π^1 cuando en realidad es de Π^2 como $p(\Pi^1/\Pi^2)$. De manera

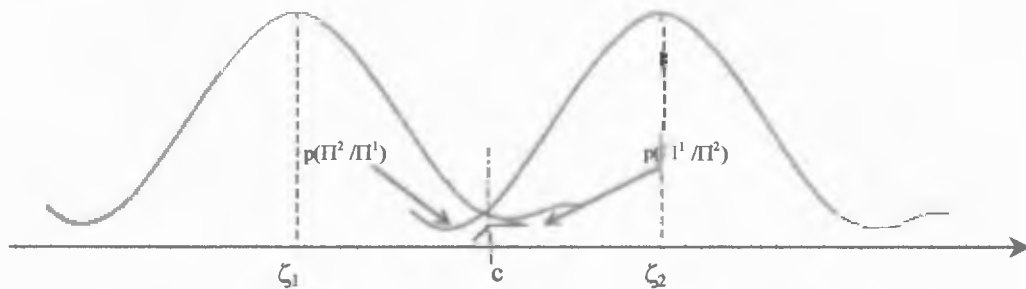


Fig. 1. Distribución de y cuando x pertenece a la población 1 y cuando x pertenece a la población 2.

similar $p(\Pi^2/\Pi^1)$ es el error de clasificar un individuo en Π^2 cuando en realidad pertenece a Π^1 . Intuitivamente se puede encontrar la constante h de tal manera que la suma de estas probabilidades sea mínima. Así se elige h como el promedio de las medias, esto es

$$h = \frac{\zeta^1 + \zeta^2}{2}$$

DISCRIMINACIÓN CUANDO LAS POBLACIONES SON CONOCIDAS

DEFINICIÓN II.4. Regla discriminante de máxima Verosimilitud.

La regla discriminante de máxima verosimilitud para la asignación de una observación \mathbf{x} a una de las poblaciones Π^1, \dots, Π^g poblaciones dice: asigñese \mathbf{x} a la población que observa la mayor verosimilitud para \mathbf{x} , es decir, la regla asigna la observación \mathbf{x} a la población Π^i

$$L^i(\mathbf{x}) = \max_i L^i(\mathbf{x}) \quad 1 \leq i \leq g$$

Si varias funciones de verosimilitud tienen el mismo valor máximo, entonces cualquiera de estas puede ser seleccionada.

TEOREMA II.1

Si la población Π^i distribuida $N_p(\mu^i, \Sigma)$, para $i=1,2,\dots,g$ y Σ definida positiva, entonces la regla discriminante de máxima verosimilitud

asigna \mathbf{x} a Π^j donde $j \in \{1,2,\dots,g\}$ es el valor de i que minimiza el cuadrado de la distancia de Mahalanobis

$$(\mathbf{x} - \mu^i)' \Sigma^{-1} (\mathbf{x} - \mu^i)$$

Demostración:

La distribución normal multivariable para p variables esta dada por:

$$f(\mathbf{x}) = |2 \Pi \Sigma|^{-1/2} \exp \{ -1/2 (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \}$$

donde $\Sigma > \mathbf{0}$.

Y su i -ésima función de verosimilitud esta dada por ,

$$L^i(\mathbf{x}) = |2 \Pi \Sigma|^{-1/2} \exp \{ -1/2 (\mathbf{x} - \mu^i)' \Sigma^{-1} (\mathbf{x} - \mu^i) \},$$

Esta función se maximiza cuando el exponente es minimizado.

PROPOSICION II.1

Cuando $g=2$, la regla discriminante de máxima verosimilitud asigna \mathbf{x} a Π^1 si

$$\alpha'(\mathbf{x} - \mu) > 0$$

donde $\alpha = \Sigma^{-1}(\mu^1 - \mu^2)$ y $\mu = \frac{1}{2}(\mu^1 + \mu^2)$ y a Π^2 en otro caso.

Demostración:

Se puede observar que $L_1 > L_2$ si y solo si,

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

$$\cancel{(x' \Sigma^{-1} x)} - (x' \Sigma^{-1} \mu^1) - (\mu^1' \Sigma^{-1} x) + (\mu^1' \Sigma^{-1} \mu^1) < \cancel{(x' \Sigma^{-1} x)} - (x' \Sigma^{-1} \mu^2) - (\mu^2' \Sigma^{-1} x) + (\mu^2' \Sigma^{-1} \mu^2).$$

$$0 < (x' \Sigma^{-1} \mu^1) + (\mu^1' \Sigma^{-1} x) - (\mu^1' \Sigma^{-1} \mu^1) - (x' \Sigma^{-1} \mu^2) - (\mu^2' \Sigma^{-1} x) + (\mu^2' \Sigma^{-1} \mu^2).$$

$$0 < (x' \Sigma^{-1} (\mu^1 - \mu^2) + (\mu^1' - \mu^2') \Sigma^{-1} x - (\mu^1' \Sigma^{-1} \mu^1) + (\mu^2' \Sigma^{-1} \mu^2).$$

$$0 < 2 (\mu^1 - \mu^2)' \Sigma^{-1} x - (\mu^1' \Sigma^{-1} \mu^1) + (\mu^2' \Sigma^{-1} \mu^2).$$

$$0 < 2 (\mu^1 - \mu^2)' \Sigma^{-1} x - (\mu^1 - \mu^2)' \Sigma^{-1} (\mu^1 + \mu^2).$$

$$0 < (\mu^1 - \mu^2)' \Sigma^{-1} \{2x - 2/2 (\mu^1 + \mu^2)\}.$$

$$0 < \{ \Sigma^{-1} (\mu^1 - \mu^2) \}' \{x - 1/2 (\mu^1 + \mu^2)\}.$$

Luego $\alpha'(x - \mu) > 0$.

DEFINICIÓN II.5. Regla Bayesiana Discriminante

Si las poblaciones Π^1, \dots, Π^g tienen probabilidad a priori π^i para $i = 1, \dots, g$, entonces la regla bayesiana discriminante (con respecto a π^j) asigna una observación x a la población para la cual la función

$$\pi^j L^j(x)$$

es máxima.

La Regla Bayesiana con respecto a probabilidades a priori π^1, \dots, π^g es definida por:

$$\Phi^j(x) = \begin{cases} 1 & \text{si } \pi^j L^j(x) = \max_i \pi^i L^i(x) \\ 0 & \text{e. o. c} \end{cases}$$

excepto para los casos donde x alcanza el máximo para más de una población.

DEFINICIÓN II.6. Probabilidad de asignación.

La probabilidad de asignar un individuo a la población Π^i , cuando en realidad proviene de Π^j , esta dada por,

$$\Phi^{ij} = \int_{\Pi^j} p^i L^i(x) dx$$

Si un individuo es en realidad de Π^i , la probabilidad de asignación correcta es Φ^{ii} y la probabilidad de mala asignación es $1 - \Phi^{ii}$.

DEFINICIÓN II.6. Regla discriminante aleatoria d.

Una regla discriminante aleatoria d envuelve la asignación de una observación x a la población j con probabilidad $\Phi^j(x)$, donde p^1, \dots, p^g son

funciones no negativas cada una definida en \mathbf{R}^p , que satisfacen $\sum_{j=1}^g \Phi^j(x)=1$

$\forall x$.

DEFINICIÓN II.7. Mejor regla discriminante

Una regla discriminante d con probabilidad de asignación correcta $\{\Phi^{ii}\}$ es mejor que otra regla d' con probabilidad $\{\Phi^{ii*}\}$ si

$$\Phi^{ii} \geq \Phi^{ii*} \quad \forall i=1,\dots,g$$

d es mejor que d' si al menos una de las desigualdades es estricta. Si d es una regla para la cual no hay otra mejor, se dice que d es admisible.

CAPÍTULO III

INTERPRETACION DE LAS FUNCIONES

DISCRIMINANTES

INTERPRETACION DE LAS FUNCIONES DISCRIMINANTES

Una vez derivadas las funciones discriminantes, podemos realizar una interpretación de su significado. Este significado se encuentra:

(1) Examinando las posiciones relativas de los casos en la base de datos y los grupos centroides y

(2) Estudiando la relación entre las variables individuales y las funciones.

Cuando existe más de una función, la pregunta es si se requieren todas ellas.

3.1 Cálculo de los Puntajes Discriminantes

Las funciones discriminantes definen un espacio p -dimensional en el que se pueden localizar casos. Por ejemplo: una primera función discriminante define uno de los ejes. Si pensamos en un espacio típico tridimensional, la primera función ocupará lógicamente el eje horizontal. Las reglas para la derivación de la segunda función discriminante requieren que la misma sea perpendicular a la primera función, esta estaría

en el eje vertical, de manera que representen informaciones completamente diferentes (las dos funciones no están correlacionadas),. La tercera función debe ser perpendicular a ambas, se debe notar que la dirección de una función es arbitraria. Para cambiar los signos de todos los coeficientes en una función dada, simplemente invierta la dirección del los ejes. Esta nueva dirección es tan "buena" como la anterior. La única razón para preferir una dirección a la otra, sería en la situación en que ciertos casos pertenecen "intuitivamente" a un extremo particular.

Los coeficientes son usados para calcular la posición de un caso particular en el espacio discriminante. La función es la siguiente:

$$f_{ij} = aX^i_j \quad (1)$$

donde f_{ij} denota los puntajes discriminantes para el caso j en la primera función, y X^i_j de $p \times 1$, representa los valores originales del caso j en la i -ésima población, de cada una de las variables dicriminatorias. Las ecuaciones para las otras funciones son similares, sustituyendo únicamente los valores correspondientes de i y j .

3.2 Diagrama de Dos Funciones.

Cuando hay dos funciones discriminantes, podemos trazar fácilmente la posición del centroide y de los casos. Por ejemplo si tenemos tres funciones, un diagrama de dos funciones es de

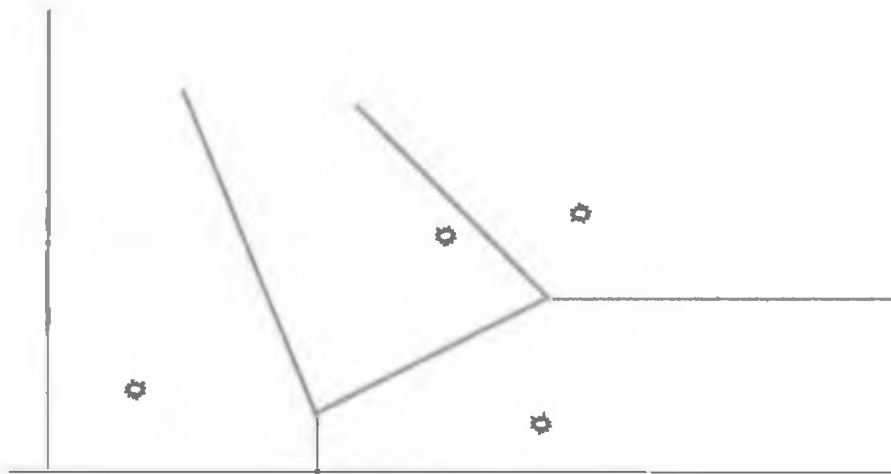


Figura 2. Diagrama de dos funciones discriminantes.

mucha utilidad, sobre todo cuando consideramos que las primeras dos funciones son más importantes. (Fig.2) representa un diagrama de dos funciones. Los asteriscos denotan cuatro grupos centroides, mientras los números simbolizan casos de cada uno de los cuatro grupos.

Un examen de la (Fig.2) nos dice que los grupos son bastante distintos debido a que los centroides están bien separados y no hay ningún

solapamiento de los casos individuales, aunque hay algunos casos en donde se presenta un acercamiento de grupos (el grupo A se acerca al grupo D). Las líneas dibujadas en este diagrama representan líneas de clasificación, llamemoslés "territorios".

El diagrama puede ser muy útil aunque exista un pequeño solapamiento entre los grupos. Cuando los grupos se vuelven menos distinguibles, sobre todo si el número de casos es grande, diagramar todos los casos produce una confusión. En esa situación sería más factible examinar un diagrama con sólo los centroides o examinar diagramas separados para cada grupo.

Cuando el número de funciones discriminantes aumenta, tendremos progresivamente más dificultad en visualizar las ubicaciones de los centroides. No obstante, se podría elaborar un modelo por computadora para tres funciones, pero en el caso de cuatro o más se dificulta la representación gráfica. Dado que generalmente las primeras dos funciones tienen más poder de discriminación, una diagrama basado solamente en estas dos funciones puede ser muy informativo, si las funciones no son suficientemente claras.

3.3 Diagrama de Una Función

Cuando hay únicamente una función discriminante, se pueden representar los casos a lo largo de una línea recta. De esa forma podríamos observar que partes de la función están "ocupadas", pero no nos daría una buena representación de la

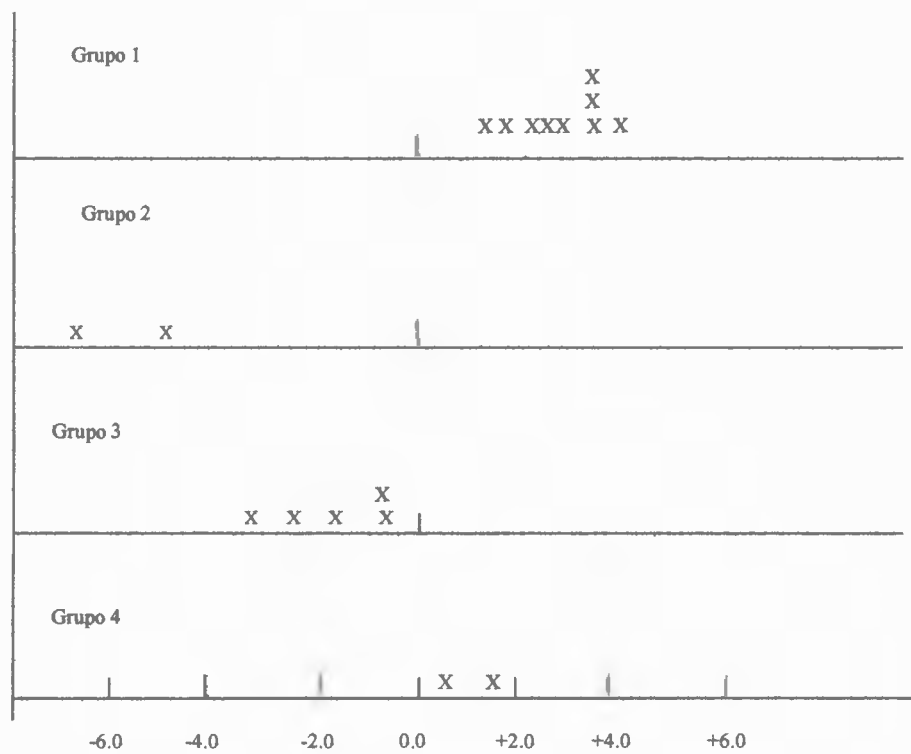


Fig. 3. Histograma de grupo. Cada x denota un individuo.

densidad de puntos, sobre todo cuando hay un gran número de casos. Una estrategia alternativa es preparar un histograma para cada grupo.

Primeramente, dividimos la línea continua, en pequeños intervalos con una desviación estandar de 0.1, por ejemplo, o cualquiera que parezca apropiado. Entonces una "X," u otro símbolo representará un caso dado, se coloca en cada intervalo que contenga un caso. Para el segundo y casos subsecuentes se colocará dentro de un intervalo, las "X" se apilan una sobre la otro, de manera que la altura de la pila denota el número de casos en ese intervalo. Esta es una forma rápida para observar la densidad y distribución de los grupos. Colocando los histogramas de cada grupo en una misma figura, podemos comparar las posiciones relativas de los grupos.

3.4 Coeficientes Estandarizados

Si enfocamos nuestra atención a los casos individuales y al centroide del grupo para descubrir la contribución de las variables individuales, debemos ver los coeficientes estandarizados; ya que los coeficientes no estandarizados nos indican la contribución **absoluta** de

una variable en la determinación del puntaje discriminante, esta información puede ser engañosa cuando la unidad de medida cambia de una variable a otra (cuando las desviaciones estándares no están representadas en las mismas unidades de medida). Si queremos saber la importancia relativa de la variable, necesitamos ver los coeficientes estandarizados.

Los coeficientes estandarizados que se obtienen por medio de la siguiente ecuación

$$u_k = v_k \sqrt{n-g} \quad \text{y} \quad u_0 = - \sum_{k=1}^p u_k X_k \quad (2)$$

si los datos originales tienen todos desviaciones estándar de 1.0, lo que se puede lograr estandarizando los datos reales. En lugar de convertir los datos reales y recalcular los coeficientes, podemos calcular los coeficientes estandarizados (w^k), de los coeficientes no estandarizados (u^k) utilizando la siguiente transformación:

$$w_k = u_k \sqrt{\frac{q_{kk}}{n-g}} \quad (3)$$

donde

q_{kk} = la suma de cuadrados para la variable k ,

n = el número del total de casos,

g = el número de grupos.

Los coeficientes estandarizados son útiles, porque pueden usarse para determinar que variables contribuyen más en la función discriminante. Esto se hace para examinar la magnitud de los coeficientes estandarizados (ignorando el signo). La variable cuyo coeficiente es más grande, representa la variable de mayor contribución.

Los puntajes discriminantes se pueden calcular por medio de los coeficientes estandarizados; sin embargo, tendríamos que multiplicar por los datos reales estandarizados. Serían puntajes z .

3.5 Estructura de los Coeficientes

Para determinar la relación entre una variable simple y una función discriminante, podemos observar la correlación del producto-momento entre las dos. Estas correlaciones se llaman "estructura total de los coeficientes."⁴ Las correlaciones, se pueden considerar como los cosenos

⁴ Klecka (1980).

que forman los ángulos entre las variables y la función. Así, para conocer estos coeficientes, debemos saber la estructura geométrica en el espacio de los datos. Los coeficientes de la estructura total se pueden conseguir de dos maneras. Una es a través de un programa de computadora que calcule el puntaje discriminante para cada función usando la siguiente ecuación

$$f_{ij} = \mathbf{u}'\mathbf{X}_j^i \quad (4)$$

donde f_{ij} denota los valores o puntajes de la función discriminante canónica para el caso j en el grupo i y \mathbf{X}_j^i representan los valores originales en la variable discriminatoria X_i para el caso j del grupo i , y \mathbf{u}^k son los coeficientes que producen la función. Con la ecuación anterior y el empleo de la correlación de Pearson se calculan las correlaciones entre las funciones y las variables. Alternativamente, podemos calcular los coeficientes estandarizados de la función discriminante canónicas en base a la matriz T , dichos coeficientes son de la siguiente forma:

$$w_{ik}^* = \frac{v_{ik}^i \sqrt{t_{kk}}}{\sqrt{\sum_{k=1}^p \sum_{m=1}^p v_{ik}^i v_{im}^i t_{km}}} \quad (5)$$

donde w_{ik}^* es el coeficiente para la función i y la variable k . (El denominador de esta ecuación es una constante y necesita ser evaluada

sólo una vez). Los coeficientes de la estructura del total se obtienen de la siguiente ecuación:

$$s^i_m = \sum_{k=1}^p r_{mk} w^i_k = \sum_{k=1}^p \frac{t_{mk} w^i_k}{\sqrt{t_{mm} t_{kk}}} \quad (6)$$

donde s^i_m es el coeficiente de la estructura total (correlación) para la variable m y la función i y r_{mk} es la correlación total entre las variables m y k .

Un coeficiente de la estructura nos indica cuan estrechamente una variable y una función están relacionadas. Cuando la magnitud absoluta del coeficiente es muy grande (cerca de + 1.0 o - 1.0), la función lleva casi la misma información como la variable. Cuando el coeficiente es casi cero o cercano, tienen muy poco en común. Podemos nombrar una función en base a los coeficientes de la estructura para destacar aquellas variables que tiene los coeficientes más altos.

CAPÍTULO IV

ANALISIS E INTERPRETACION DE LOS RESULTADOS

ANALISIS E INTERPRETACION DE LOS RESULTADOS

Los investigadores han utilizado el análisis discriminante en una gran variedad de áreas. En su primera exploración Fisher(1936), buscó resolver problemas en el área de la antropología física y la biología. En las ciencias sociales, algunas de las primeras aplicaciones trataron con pruebas psicológicas y educativas (Tatsuoka y Tiedeman, 1954). Los psicólogos han hecho un uso extenso del análisis discriminante, especialmente en áreas en las cuales se realizan pruebas de personal y educación.

En la presente investigación utilizaremos datos concernientes a mediciones sobre el estado nutricional de 187 estudiantes varones de una escuela pública en La Chorrera. En la misma, se pretende identificar claramente los grupos o estados nutricionales. Los mismos, pertenecen a uno de los 3 grupos o estados nutricionales, de acuerdo a la siguiente tabla.

Tabla I. Niveles de Nutrición.

Grupo	Número de Casos	Estado Nutricional
C	106	Normal
D	64	Desnutrición aguda leve
E	17	Desnutrición aguda moderada
Total	187	Cantidad total de casos

Fuente: Centro de Salud Magally Ruiz.

Las categorías de estado nutricional, se definen de acuerdo a la *clasificación de la relación peso/talla de niños* y la misma se obtiene interpolando las curvas de peso/talla elaboradas por el Centro Nacional de Estadísticas de Salud de los Estados Unidos (NCHS) y utilizadas internacionalmente por recomendación de la OMS, y en particular por el Programa de Salud Escolar del Ministerio de Salud, las mismas están separadas por sexo y clasificadas de la siguiente manera:

Sobrepeso.....	> + 2 DS
Riesgo al sobrepeso.....	> + 1 DS
Normal.....	± 1 DS
Desnutrición aguda leve.....	< - 1 DS
Desnutrición aguda moderada.....	< - 2 DS
Desnutrición grave.....	< - 3 DS

4.1. Indicadores de Desnutrición

(a) Antropométricos

La antropometría es la técnica más usada para la estimación cuantitativa del estado nutricional. Las medidas antropométricas requieren un equipo relativamente simple, este examen puede ser completado rápidamente, y llevado a cabo por personal no técnico, después de un período corto de entrenamiento, aunque, la antropometría por si sola, no puede definir el estado nutricional. Las medidas del cuerpo pueden, en algunas ocasiones, detectar la desnutrición del niño, y permite distinguir entre la desnutrición aguda y una crónica. Para la evaluación final del tipo y severidad de la desnutrición, se requieren evaluaciones clínicas.

(b) Examen Clínico

La anemia nutricional se debe a una reducción de los glóbulos rojos circulantes o a la disminución del contenido de hemoglobina. Estas anemias se deben a una deficiencia de hierro y folatos. Por ello, las mediciones de las variables hemoglobina y heritrocitos a través de exámenes de sangre contribuyen a establecer el estado nutricional del niño.

Para el estudio se utilizaron 5 variables, la cuales se presentan en la tabla II. Un estudiante puede asignarse a uno de los 3 estados nutricionales de acuerdo a las mediciones de las variables que se presentan en dicha tabla.

Tabla II. Descripción de las Variables

Variable	Descripción
X_1	Edad: medida en años y meses
X_2	Peso Inicial: en libras
X_3	Talla Inicial: en centímetros
X_4	Hb: Nivel de Hemoglobina
X_5	Hto: Nivel de Heritrocitos

Primeramente se estudiará si las variables tienen una distribución multinormal en cada estado nutricional.

Como se observa las figura (Fig. 4,5,6,7, y 8) de las 5 variables estudiadas, 4 de ellas tienen una distribución normal; sin embargo, la gráfica de la variable edad nos indica que esta no tiene una distribución normal, esto se debe a que los grupos de edades son homogéneos.

La tabla III, muestra las medias y desviaciones estándares de los 3 estados nutricionales. En la misma podemos observar que,

Tabla III. Medias y desviaciones estándar por Estado Nutricional Conocido

Estado Nutricional	Variable									
	Edad (en años y meses)		Peso Inicial (libras)		Talla Inicial (cms)		Hb (mg/100ml)		Hto (%)	
	Media	Desv	Media	Desv.	Media	Desv.	Media	Desv.	Media	Desv.
C	6.78	0.99	50.62	8.40	1.20	0.06	11.32	1.17	33.73	2.67
D	6.80	0.93	45.34	7.08	1.20	0.07	11.19	1.08	33.65	3.25
E	6.53	0.79	42.12	6.60	1.23	0.08	11.19	1.76	33.70	5.29
General	6.76	0.95	47.94	8.36	1.20	0.06	11.26	1.20	33.70	3.18

aparentemente los grupos están más separados en términos del peso inicial que de las restantes 4 variables. Las diferencias entre las medias de edades en los tres grupos nutricionales son pequeñas, al igual que las desviaciones estándares de los mismos. De igual manera, en el caso de la variable peso se observa una diferencia marcada entre los grupos, más sin embargo para esta última, la desviación estándar es relativamente alta.

La tabla IV muestra la matriz de correlación total, y en ella las asociaciones entre las variables existentes.

Tabla IV. Matriz de Correlación Total

Variable	Edad	Peso Inicial	Talla Inicial	Hb	Hto
Edad	1.000				
Peso Inicial	0.526	1.000			
Talla Inicial	0.607	0.723	1.000		
Hb	0.230	0.180	0.210	1.000	
Hto	0.190	0.150	0.170	0.900	1.000

Así podemos observar que las variables antropométricas muestran una mayor relación entre ellas, al igual que las variables clínicas, más sin embargo, se observa poca relación entre estas últimas con las primeras. Esto nos indica que se deben utilizar las variables antropométricas separadas de las variables clínicas, es decir elaborar una función discriminante para las variables antropométricas y otra para las variables clínicas, en el caso de estas últimas hemos decidido no utilizarlas ya que solamente contamos con dos, y a nuestro criterio estas dos variables no

son suficientes para discriminar. En la tabla IVa, se presenta la matriz de correlación total con las variables que utilizaremos en nuestro estudio.

Tabla IVa. Matriz de Correlación Total

Variable	Edad	Peso Inicial	Talla Inicial
Edad	1.000		
Peso Inicial	0.536	1.000	
Talla Inicial	0.607	0.723	1.000

Notece que la matriz de correlación dentro de grupos enfatiza con más claridad las asociaciones existentes, que en el caso de la matriz de correlación total. La tabla V (Matriz de correlación dentro de grupos).

La matriz de correlación dentro de los grupos (tabla V) muestra correlaciones relativamente bajas entre las variables. Es decir que hay muy poca probabilidad de encontrar problemas de multicolinealidad¹.

¹ Klein dice: “La intercorrelación de variables no es necesariamente un problema, a menos que sea muy elevada en relación con el grado global de correlación múltiple.”

Aunque se observa que entre las variables Talla inicial y Peso inicial la correlación es un poco alta .

Tabla V. Matriz de Correlación Dentro de Grupos

Variable	Edad	Peso Inicial	Talla Inicial
Edad	1.000		
Peso Inicial	0.554	1.000	
Talla Inicial	0.623	0.823	1.000

Tabla VI. Funciones de Clasificación: Función Discriminante Lineal de Fisher.

Variable	C p=0.5668	D p=0.3422	E P=0.0909
Edad	-5.763	-5.596	-6.363
Peso Inicial	-3.417	-3.708	-3.996
Talla Inicial	674.379	702.307	743.807
Constante	-299.251	-320.528	-354.586

$$y^1 = -299.251 - 5.763X_1 - 3.417X_2 + 674.379X_3$$

$$y^2 = -320.528 - 5.596X_1 - 3.708X_2 + 702.307X_3$$

$$y^3 = -354.586 - 6.363X_1 - 3.996X_2 + 743.807X_3$$

Si deseamos evaluar a un nuevo estudiante, para saber en que estado nutricional estaría clasificado, debemos obtener las mediciones de las 3 variables discriminatorias y evaluarlas en las tres funciones.

La tabla VII muestra los resultados de la clasificación de los casos de acuerdo a la función discriminante. El porcentaje de casos que se clasificaron correctamente (razón de aciertos), es la razón del

Tabla VII. Matriz de Clasificación

Grupo Predicción Observado	Porcentaje Correcto	C P=0.5668	D p=0.3422	E p=0.0909
C	92.45	98	8	0
D	75.00	15	48	1
E	41.18	0	10	7
Total	81.82	113	66	8

total de los valores en la diagonal principal al número de casos de la matriz de clasificación $(98+48+7)/187=0.8182$, o sea 81.82%, es decir que el 81.82% de los casos fueron clasificados correctamente.

La tabla VIII muestra las coeficientes crudos o coeficientes no estandarizados de las funciones discriminantes.

Tabla VIII. Coeficientes Discriminantes Crudos

Variable	Función 1 y^1	Función 2 y^2
Edad	0.1037	-1.2191
Peso Inicial	0.2243	0.0480
Talla Inicial	-25.9933	12.9005
Constante	18.6222	-9.5964
Valor propio	0.7312	0.0235
Proporción acumulada	0.9688	1.0000
Correlación Canónica	0.6651	0.1566

Los valores propios asociados con estas funciones son 0.7312 que representa el 96.88% de la varianza explicada y 0.0235, que representa solamente el 3.12% de la varianza. El primer valor propio es bastante alto, por lo tanto la primera función discriminante aporta más información y produce una mejor clasificación.

La correlación canónica asociada con esta función es 0.6651. El cuadrado de esta correlación es 0.4424, lo que indica que el 44.24% de la varianza en la población C (estado nutricional normal), es explicada por medio de esta función; sin embargo en la función 2, el coeficiente de correlación canónico respectivo es 0.1566, cuyo cuadrado es 0.0245, es decir, la función dos sólo explica o representa el 2.45% de la varianza en la población D (desnutrición aguda leve).

La tabla IX presenta los coeficientes estandarizados. La interpretación de los resultados se basa en un estudio detallado de los coeficientes estandarizados de la función discriminante.

Tabla IX. Coeficientes Discriminantes Estandarizados

Variable	Función 1	Función 2
Edad	0.0983	-1.1550
Peso Inicial	1.7274	0.3699
Talla Inicial	-1.6163	0.8343
Valor propio	0.7312	0.0235
Proporción acumulada	0.9688	1.0000

Estos coeficientes estandarizados muestran un nivel relativamente alto para la variable Peso inicial en la función 1; mientras que para la función 2, la variable edad muestra un coeficiente relativamente más alto.

Tabla X. Matriz de Estructura Total

Variable	Función 1	Función 2
Edad	0.0484●	-0.4299
Peso Inicial	0.4515●	0.4163
Talla Inicial	-0.1332	0.4188●

Examinando la matriz de estructura total, se llega a la misma conclusión para ambas funciones. Para la interpretación de las funciones se coloca un símbolo (●) a las variables cuyos coeficientes resultaron más altos en la respectiva función, estas variables se agrupan dentro de la función. Así tenemos que las variables edad y peso, tiene un símbolo (●) en la función 1, porque éstas variables poseen un coeficiente más alto dentro de la función 1 que para la 2. Es decir, que las variables se relacionan, sobre

todo, con la función 1. Por otro lado, las variables talla inicial se relacionan principalmente con la función 2, como se indican con el símbolo.

Grado de Significancia

Para probar la hipótesis nula de igual centroide para todos los grupos, se toman en cuenta ambas funciones en forma simultánea. Luego se procede a excluir una función a la vez y las medias de las funciones que restan se prueban paso a paso. La tabla XI muestra las pruebas de significancia de las funciones discriminantes.

Tabla XI. Prueba de significancia para las funciones sucesivas.

# de Funciones removidas	Lambda de Wilks	Chi-cuadrado	gl	Probabilidad asociada
0	0.5643	104.6918	6	< 0.000001
1	0.9770	4.2589	2	0.118920

La columna “# de funciones removidas” indica la cantidad de funciones que se han excluido para el análisis. El 0 indica que no se ha eliminado ninguna función. El valor de λ de Wilks es 0.5643, con un valor para la chi-cuadrada de 104.6918 y 6 grados de libertad, y con una probabilidad menor que 0.000001, que es significativa. De esta forma, podemos decir que las dos funciones en conjunto discriminan significativamente entre los tres grupos. El número 1 indica que se a excluido una función discriminante, en este caso se ha eliminado la primera función discriminante. Al observar los resultados de eliminar la primera función, vemos que el λ asociado con la segunda función es de 0.9770, con una probabilidad de 0.1189, lo cual no es significativo al nivel de 0.01 ni de 0.05. Luego podemos decir que la segunda función no aporta información importante en cuanto a diferencias entre los grupos.

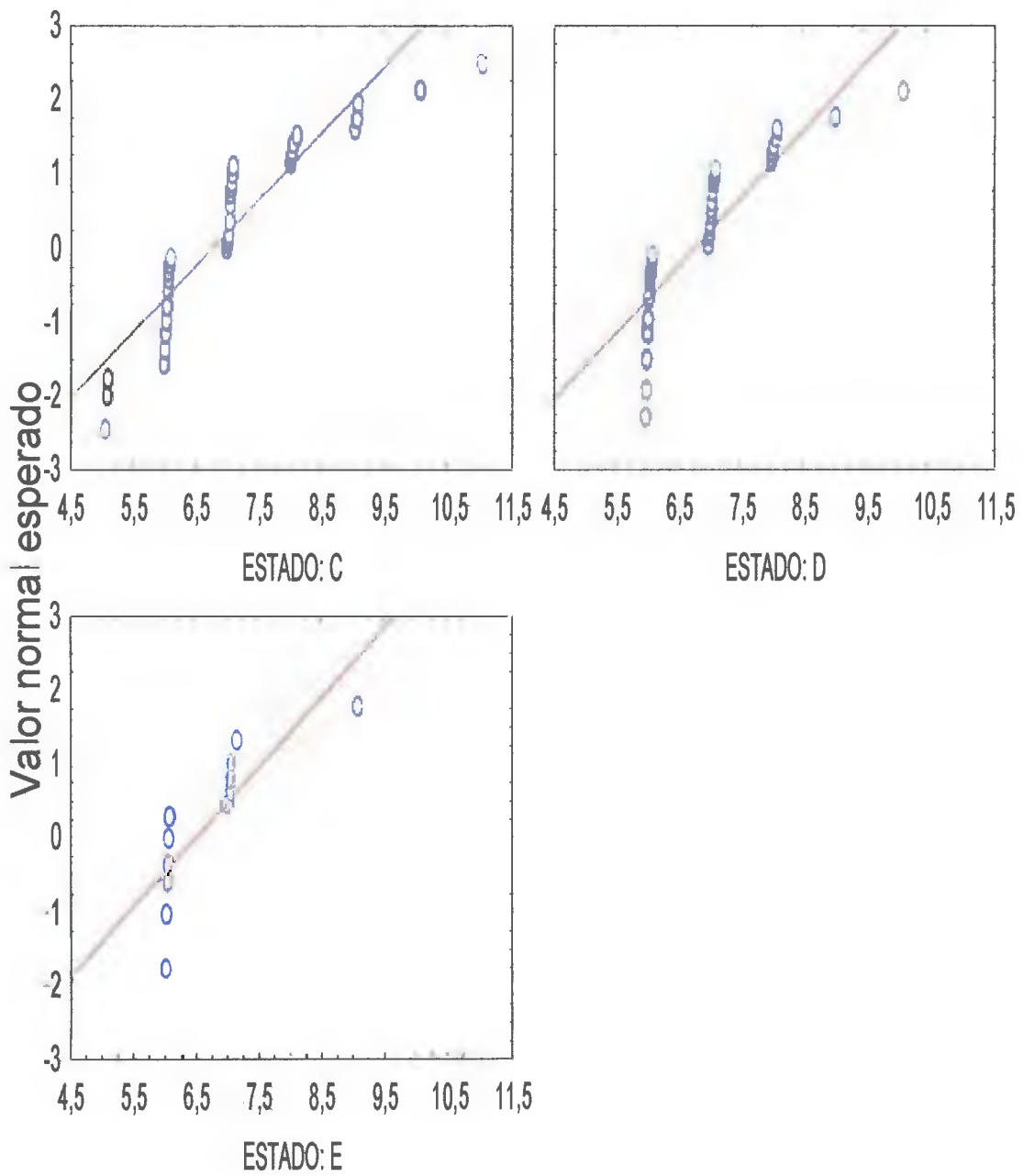


Fig. 4. Gráfico de probabilidad de la variable edad para cado uno de los estados nutricionales.

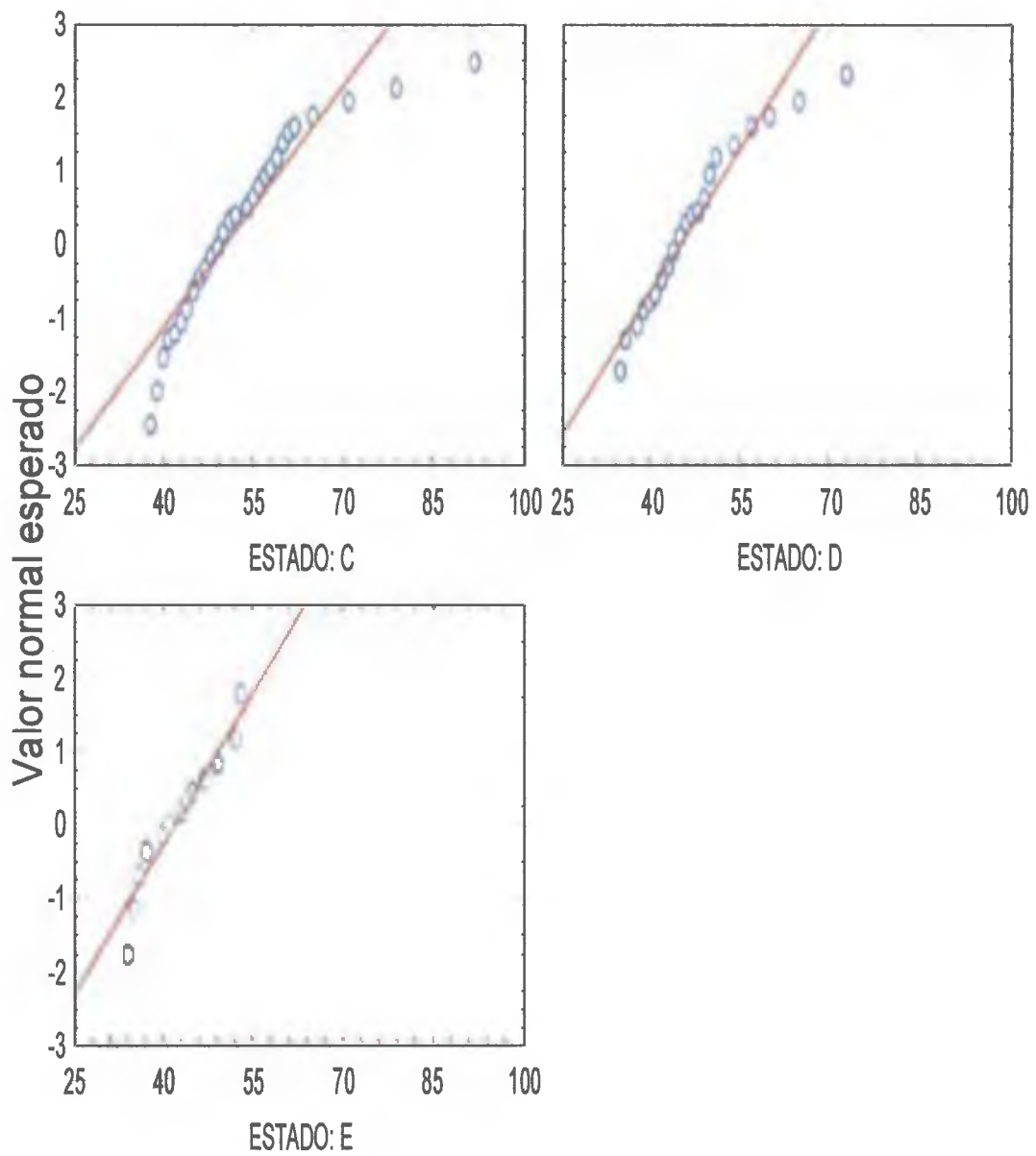


Fig. 5. Gráfico de probabilidad de la variable peso para cado uno de los estados nutricionales.

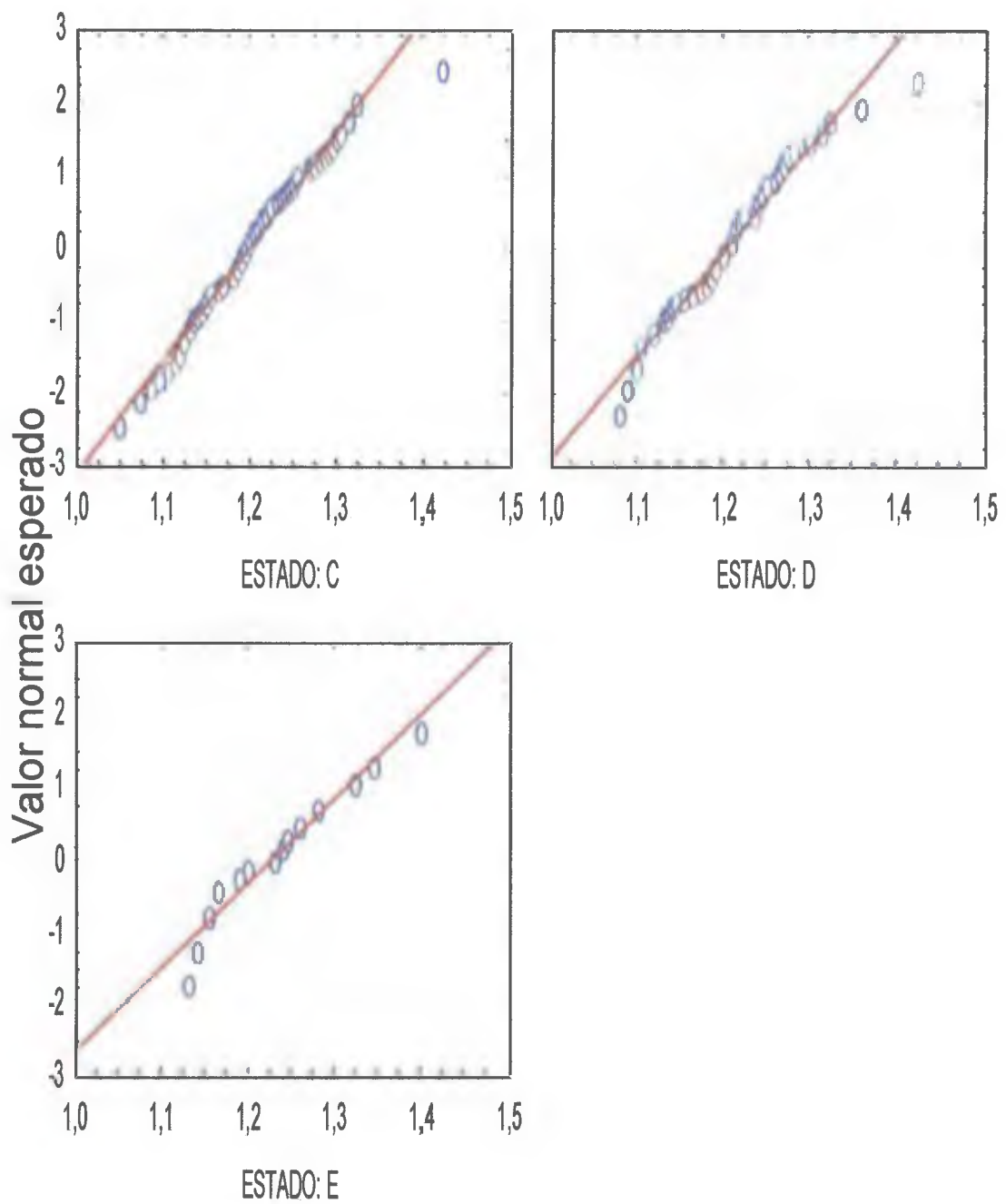


Fig. 6. Gráfico de probabilidad de la variable talla para cado uno de los estados nutricionales.

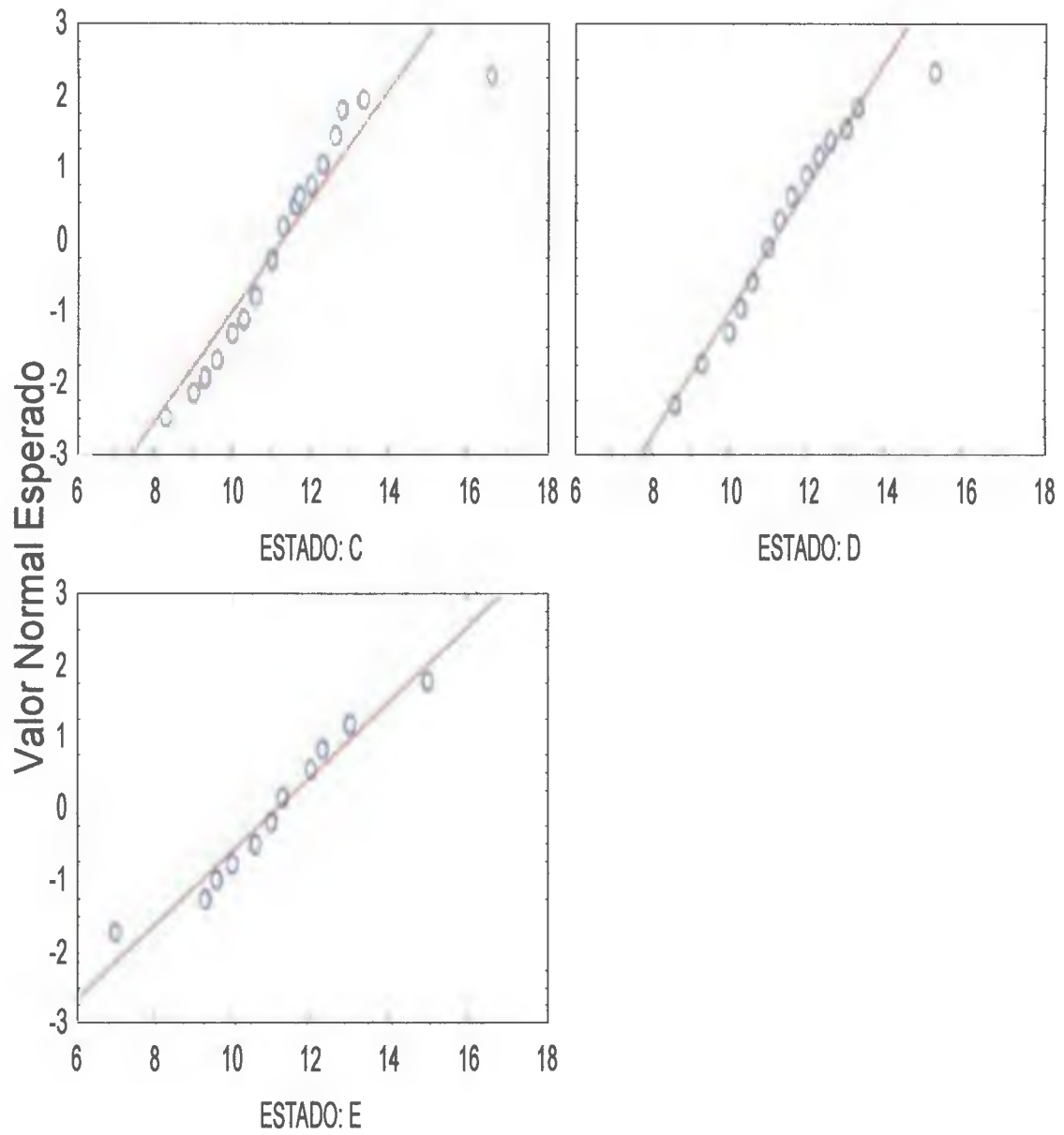


Fig. 7. Gráfica de probabilidad de la variable hemoglobina(Hb) para cada uno de los estados nutricionales.

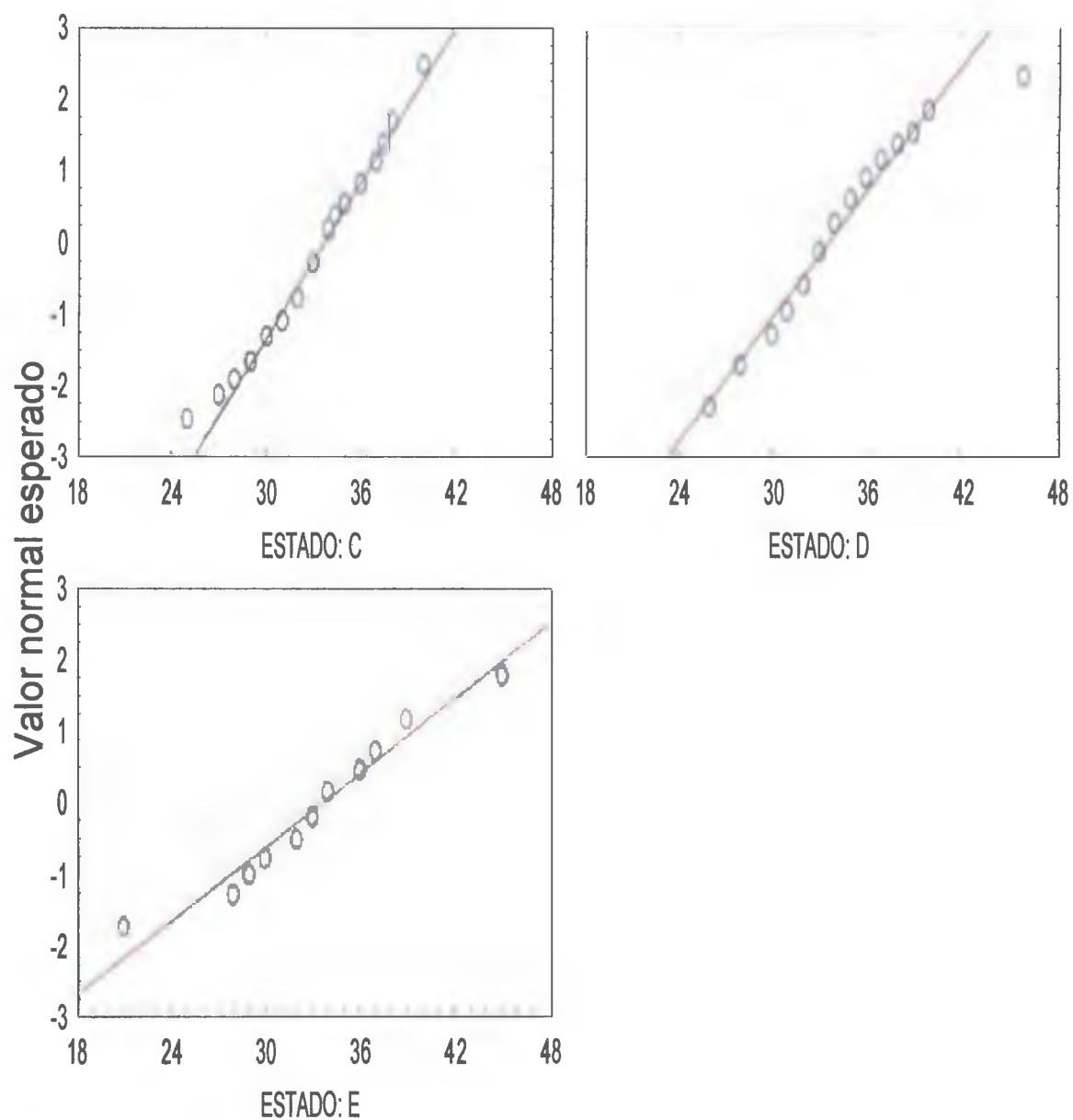


Fig. 8. Gráfica de probabilidad de la variable heritrocitos (Hto) para cado uno de los estados nutricionales.

Función discriminante 2

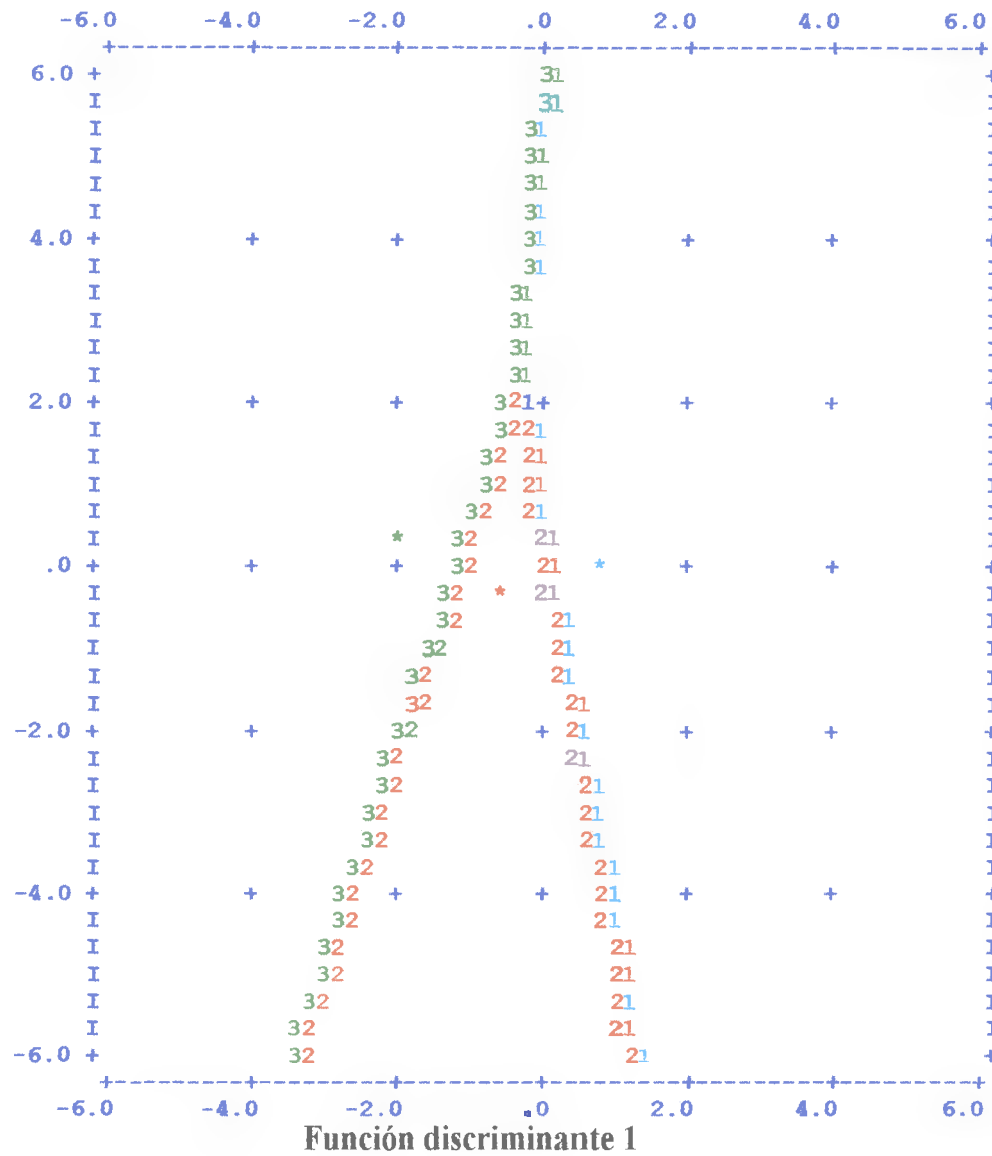
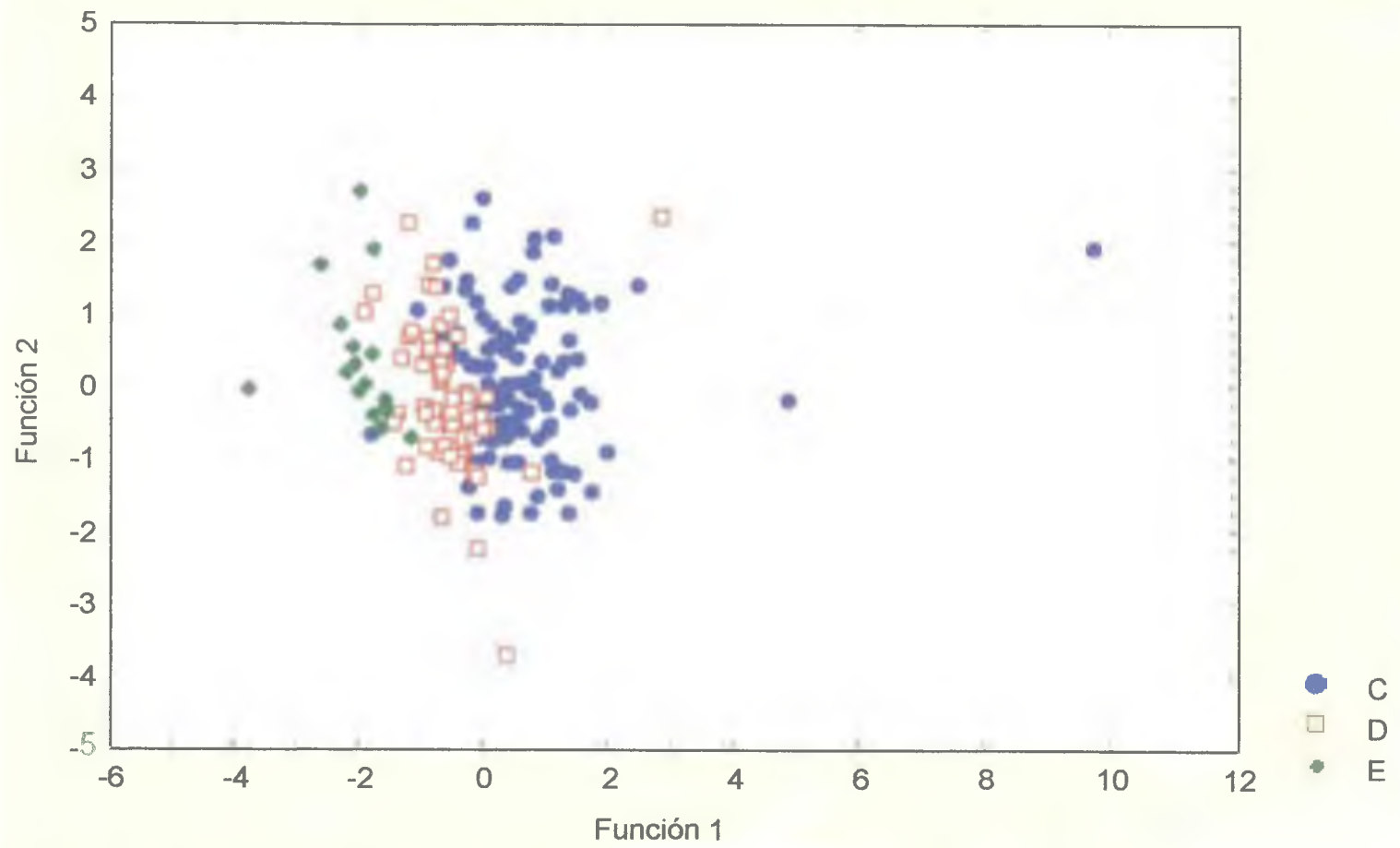


Fig. 9. Mapa Territorial.



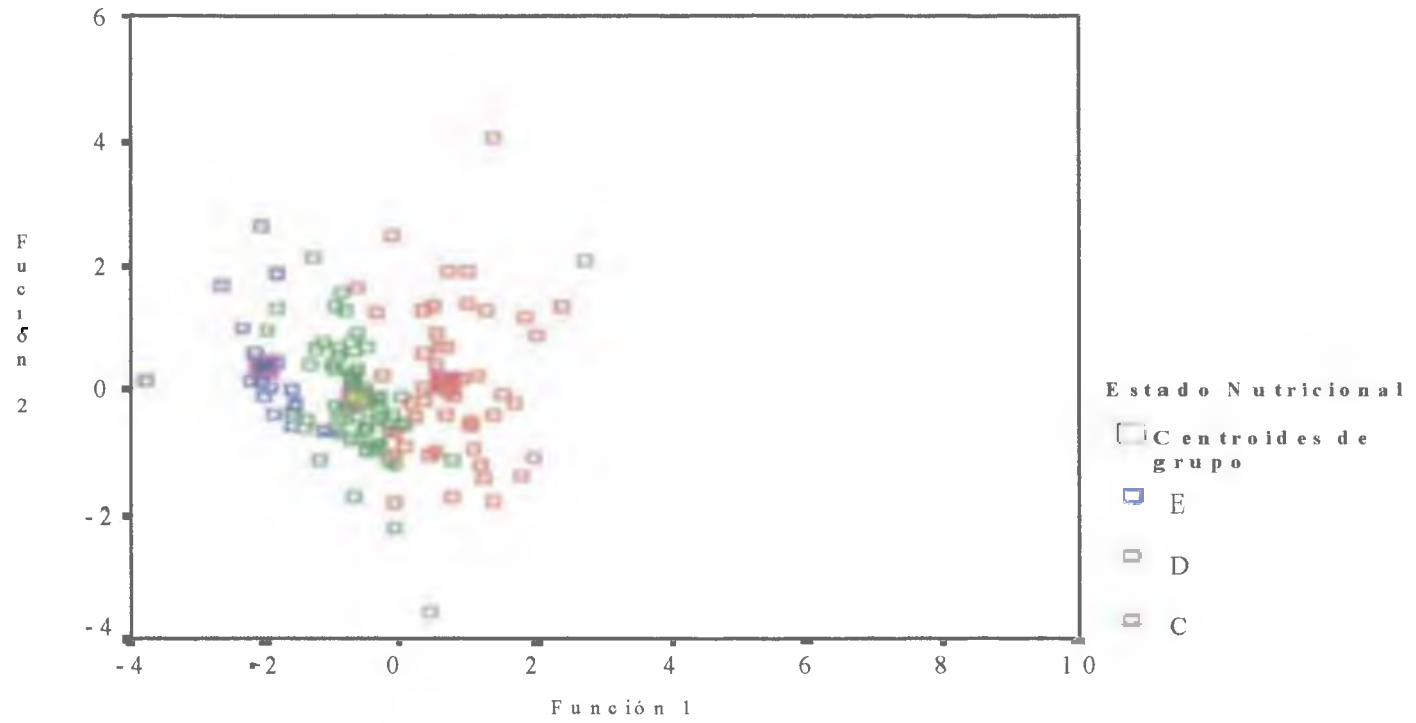


Fig. 11. Funciones discriminantes y centroides de grupo.

CONCLUSIONES Y RECOMENDACIONES

1. El **Análisis Discriminante**, es una técnica estadística multivariada para la clasificación de individuos u objetos dentro de grupos mutuamente exclusivos y completos, sobre la base de un conjunto de variables independientes.
2. El **Análisis Discriminante** envuelve la derivación de combinaciones lineales de las variables independientes, tales que discriminan entre los grupos definidos a priori de tal forma que los errores de clasificación sean mínimos.
3. La función discriminante calculada, es una combinación lineal de las medidas originales, en este caso edad, peso y talla, en la cual, la varianza entre los grupos C, D y E es máxima (con el propósito de resaltar las diferencias entre los grupos), y la varianza dentro de los grupos es mínima (con el fin de que la dispersión dentro de los grupos este bien definida).
4. El porcentaje de casos clasificados correctamente fue de 81.82%, es decir con estos datos y la función discriminante encontrada hay un 18.18% de clasificación errónea.

5. Dado que solamente tenemos tres grupos, debemos utilizar dos funciones discriminantes. Los valores propios asociados con estas funciones son de 0.7312, que representa el 96.88% de la varianza explicada y 0.0235, que representa solamente el 2.45% de la varianza.
6. La utilización de este método de análisis multivariado permite obtener una función que discrimina el estado nutricional de los niños de 6 a 7 años en promedio. Esto significa que solamente con tener las mediciones de las variables que se estudiaron y evaluándolas en las funciones discriminantes podemos determinar en qué estado nutricional se encuentran los niños de estas edades.
7. Recomendamos la utilización de las variables clínicas, Hemoglobina(Hb) y Heritrocitos(Hto), en conjunto con otras como lo son concentración de *vitamina A*, *colesterol* y *albúmina*, y poder definir una función que discrimine los estados nutricionales.
8. Se recomienda profundizar este estudio ampliando el tamaño de la muestra, ya que el presente, sólo hizo uso de 187 datos, debido a que

en la etapa de depuración de los mismos, existían algunos datos no confiables o incompletos.

9. ~~Es necesario ampliar el número de variables antropométricas, entre las variables que recomendamos están: *circunferencia de la cabeza, circunferencia del brazo, y espesor del pliegue cutáneo.*~~
10. ~~Será de gran utilidad realizar este estudio en áreas rurales, en donde se pueda evaluar mucho mejor el fenómeno de la desnutrición.~~

BIBLIOGRAFÍA

ANDERSON, T. W. 1958. An Introduction to multivariate Statistical Analysis. 1st.ed. J. WILLEY and SONS, INC., New York, 374 págs.

BOLCH, B.W. y HUANG, C.J. 1974. Multivariate Statistical Methods for Business and Economics. 1st.ed. PRENTICE-HALL, INC., New Jersey, 323 págs.

COOLEY, W.W. y LOHNES, P.R. 1962. Multivariate Data Analysis. 2da.ed. ROBERT E. KRIEGER PUBLISHING COMPANY, Florida, 361 págs.

CUADRAS, C.M. 1996. Métodos de Análisis Multivariante. 1st.ed. EUB, S.L., Barcelona, 644 págs.

DILLON, W. y GOLDSTEIN, M. 1984. Multivariate Analysis methods and applications. 1st.ed. J. WILLEY and SONS, INC., New York, 520 págs.

FLURY, BERNHARD. 1993. Multivariate Statistics: A practical approach. 2da.ed. CHAPMAN AND HALL, New York, 293 págs.

HARRIS, R.J. 1975. A Primer of Multivariate Statistics. 1st.ed. ACADEMIC PRESS, New York, 323 págs.

HUBERTY, C.J. 1994. Applied Discriminant Analysis. 1st.ed. J. WILLEY and SONS, INC., New York, 466 págs.

JOHNSON, R.A Y WICHERN, D.W. 1982. Applied Multivariate Statistical Analysis. 1st.ed. PRENTICE-HALL, INC., New Jersey, 594 págs.

KENDALL, MAURICE. 1980. *Multivariate Analysis*. 2da.ed. CHARLES GRIFFIN & COMPANY LTD, London, 209 págs.

KLECKA, W.R. 1980. *Discriminant Analysis*. 1st.ed. SAGE PUBLICATIONS, INC., Newbury Park, California, 71 págs.

MANLY, BRYAN F.J. 1986. *Multivariate Statistical Methods: A Primer*. 2da.ed. CHAPMAN AND HALL, New York, 155 págs.

MARDIA, K.V., KENT, J.T. y BIBBY, J.M. 1988. *Multivariate Analysis*. 6ta.ed. ACADEMIC PRESS, London, 521 págs.

MEJÍA, C. AURORA. 1988. , Panamá. Tesis. Universidad de Panamá, Panamá. 90 págs.

MUIRHEAD, R.J. 1982. *Aspects of Multivariate Statistical Theory*. 2da.ed. JOHN WILEY & SONS, INC., New York, 595 págs.

SEARLE, SHAYLE R. 1982. *Matrix Algebra Useful for Statistics*. 1st.ed. JOHN WILEY & SONC INC., New York, 433 págs.

TOMASSONE, R., DANZART, M., DAUDIN, J.J. y MASON, J.P. 1988. *Discrimination et Classement*. 1st.ed. EDITORIAL MASSON, Paris, 172 págs.