Tecnociencia 2005, Vol. 7, N° 1.

APLICACIÓN DEL ANÁLISIS DISCRIMINANTE A UN CONJUNTO DE DATOS VINÍCOLAS MEDIANTE EL PAQUETE ESTADÍSTICO SPSS v10

[1]Maylí Z. Pozo Díaz, [2]Gonzalo I. Carrasco O.
[1]Universidad de Panamá, Facultad de Ciencias Naturales Exactas y Tecnología, Escuela de Estadística. E-mail:maylipozo@hotmail.com
[2]Universidad de Panamá, Facultad de Ciencias Naturales Exactas y Tecnología, Departamento de Estadística. Centro de Investigación y Consultoría Estadística.
E-mail: gcarrasco27@hotmail.com

RESUMEN
En este trabajo se muestran y explican los principales elementos que se relacionan con el procedimiento para llevar a cabo el análisis discriminante y la aplicación del mismo utilizando el paquete estadístico SPSS versión 10 (SPSS v10). Aplicando este análisis a los datos originales se muestra, a través de la matriz de clasificación, que el 100% de los vinos fueron clasificados correctamente por medio de las funciones discriminantes canónicas, por lo que se comprueba que estas poseen un alto poder discriminante y pueden ser utilizadas para futuras pruebas de clasificación.

PALABRAS CLAVES
Análisis discriminante, puntuaciones discriminante, lambda de wilks, correlación canónica, análisis multivariado, M de Box, valores propios.

ABSTRACT
This work shown and explain the main elements related with the procedure to carry out the discriminant analysis and the application of the same one using the statistical software SPSS version 10 (SPSSv10). Applying this analysis to the original data, the classification matrix shown that the 100% of wines was classified correctly using the canonical discriminant functions, this verify those functions have a high discriminant power and it can be used for future tests of classifications.

KEYWORDS
Discriminant analysis, discriminant scores, lambda wilks, canonical correlation, multivaried analysis, M's Box, eigenvalues.

INTRODUCCIÓN

El análisis discriminante es un método estadístico por el cual se busca conocer cuales variables, medidas en objetos o individuos, contribuyen en mayor medida a la diferencia de los grupos a los cuales pertenecen dichos objetos o individuos. Con estos atributos medidos se forma una combinación lineal creando una nueva variable que permitirá la máxima separación entre los grupos y la mínima separación dentro de cada grupo.

En este trabajo desarrollamos el modelo estadístico, las condiciones para la aplicación del análisis, la estimación e interpretación de las funciones discriminantes, los métodos de clasificación y la validación de los resultados.

MATERIALES Y MÉTODOS

La base de datos "Datos de reconocimiento de vinos" son el resultado de un análisis químico de vinos producidos en una región de Italia, pero de diferentes variedades cultivadas o grupos. Esta base de datos es la utilizada para la aplicación del método a través del paquete estadístico SPSS v10 permitiendo así explicar cada uno de los elementos que son expuestos en la teoría.

DEFINICIÓN Y MODELO ESTADÍSTICO

Consideremos g grupos o poblaciones $\mathbf{G_1, G_2,...G_g}$ cada una de ellas distribuida $\mathbf{N_p\ (\mu^i,\Sigma^i)}$ para $\mathbf{i=1,\ 2,...,\ g}$. Con matriz de covarianza $\mathbf{\Sigma^1=\Sigma^2=...=\Sigma^g=\Sigma}$ y de rango completo.

Cada uno de estos grupos posee características específicas y únicas, es decir son mutuamente excluyente. La pertenencia de un individuo a un grupo depende de este conjunto de características o variables de predicción. Entonces, se puede definir al análisis discriminante como un procedimiento multivariable utilizado para discriminar y clasificar. Sea $\mathbf{X^i n^i xp}$ una matriz de datos multivariados de orden $\mathbf{n^i xp}$ que representa las observaciones de la i-ésima muestra, $\mathbf{i=1,\ 2,...,\ g}$, donde para cada $\mathbf{1 \leq i \leq g}$, $\mathbf{1 \leq j \leq n^i}$ cada fila (individuo) $\mathbf{X^i_j}$ de $\mathbf{X^i n^i xp}$, corresponde al j-ésimo vector aleatorio de orden $\mathbf{1xp}$. Entonces $\mathbf{X_{nxp}}$, es la matriz que representa el conjunto de **g** matrices $\mathbf{X^i}$ de datos multivariados (Rodríguez 1999).

$$
X_{nxp} = \begin{bmatrix} X^1 \\ X^2 \\ : \\ X^g \end{bmatrix} = \left[\begin{array}{ccc}
x^1_{11} & x^1_{12} \; \ldots & x^1_{1p} \\
x^1_{21} & x^1_{22} \; \ldots & x^1_{2p} \\
: & : & : \\
x^1_{n^1 1} & x^1_{n^1 2} \; \ldots & x^1_{n^1 p} \\
\hline
x^2_{11} & x^2_{12} \; \ldots & x^2_{1p} \\
x^2_{21} & x^2_{22} \; \ldots & x^2_{2p} \\
: & : & : \\
x^2_{n^2 1} & x^2_{n^2 2} \; \ldots & x^2_{n^2 p} \\
\hline
: & : & : \\
\hline
x^g_{11} & x^g_{12} \; \ldots & x^g_{1p} \\
x^g_{21} & x^g_{22} \; \ldots & x^g_{2p} \\
: & : & : \\
x^g_{n^g 1} & x^g_{n^g 2} \; \ldots & x^g_{n^g p}
\end{array}\right]
$$

donde $n = \sum_{i=1}^{g} n^i$

SUPUESTOS O CONDICIONES PARA LA APLICACIÓN DEL ANÁLISIS

- Existencia de dos o más grupos mutuamente excluyentes.
- Mínimo de dos casos por grupo.
- Las variables independientes deben ser cuantitativas y las dependientes cualitativas.
- Las variables seleccionadas en el modelo no deben ser combinaciones lineales de otras.
- La matriz de varianzas y covarianzas iguales en cada grupo.
- Cada muestra proviene de una población normal multivariada.

OBJETIVOS DEL ANÁLISIS

- Seleccionar las variables de predicción que contribuyen en mayor medida a las diferencias entre los grupos.
- Predecir la probabilidad de pertenencia de un individuo a un grupo, basándose en las variables de predicción, por medio de las reglas de clasificación.
- Evaluar la exactitud de la clasificación en una tabla cruzada en donde se compara la pertenencia real de los individuos a un grupo

con la pertenencia, a un grupo, predicha por medio de las reglas de clasificación.

FUNCIÓN DISCRIMINANTE CANÓNICA

Para el aspecto descriptivo del análisis se utilizan las funciones discriminantes canónicas. Estas funciones son combinaciones lineales de las variables clasificatorias originales cuyo propósito es expresar estas variables en una nueva variable. Esta nueva variable está representada por un eje sobre el cual se proyectan los individuos y será la base para establecer qué variables originales tienen mayor poder discriminatorio entre los grupos. Este eje es llamado eje de la función discriminante.

Procedimiento Discriminante de Fisher

Sea

$$\mathbf{E}_{pxp} = \sum_{i=1}^{g} n^i (\mu^i - \mu.)(\mu^i - \mu.)' \quad \text{y} \quad \mathbf{D}_{pxp} = \sum_{i=1}^{g} \sum_{J=1}^{n^i} (x^i_j - \mu^i)(x^i_j - \mu^i)'$$

$\mathbf{E}_{pxp}$ se llama matriz de suma de cuadrados y productos cruzados de las medias entre los grupos y que $\mathbf{D}_{pxp}$ es una matriz no singular que se le llamará matriz de suma de cuadrados y productos cruzados dentro de los grupos.

Sea,

$$\max_{b \neq 0} \lambda = \frac{b'\mathbf{E}\, b}{b'(\mathbf{D}^{-1}\mathbf{E})\, b}$$

es el valor propio más grande de $\mathbf{D}^{-1}\mathbf{E}$. Este valor propio más grande se denota por $\boldsymbol{\lambda_1}$ y este máximo se alcanza cuando $\mathbf{b=b_1}$, en donde $\mathbf{b_1}$ es un vector propio de $\mathbf{D}^{-1}\mathbf{E}$, asociado al valor propio $\boldsymbol{\lambda_1}$, y que contiene los coeficientes discriminantes relacionados a la primera función discriminante. La primera función posee el mayor poder discriminante (Mardia et al., 1979).

La función discriminante es una función lineal de las **p** variables independientes, es decir

$$A_m = b_0 + b_{1m}X_1 + b_{2m}X_2 + ... + b_{pm}X_p$$

m=1,2,...,k donde **k=min(g-1,p)**, en la cual el **min(g-1,p)** representa el rango de la matriz $\mathbf{D^{-1}E}$. Los **k** ejes discriminantes vienen definidos respectivamente por los vectores $\mathbf{b_1, b_2, ..., b_k}$. Los vectores propios $(\mathbf{b_2, b_3, ..., b_k})$ serán sucesivamente los vectores normalizados de los precedentes valores propios $(\boldsymbol{\lambda_2}, \boldsymbol{\lambda_3}, \ldots, \boldsymbol{\lambda_k})$.

LAMBDA DE WILKS

Mide las desviaciones dentro de cada grupo respecto a las desviaciones totales sin distinguir grupos y toma valores entre **0** y **1** de forma que, cuanto más cerca de **0** esté, mayor es el poder discriminante de la variables consideradas y cuanto más cerca de **1**, menor es dicho poder.

Prueba de igualdad de medias de los grupos.

En esta prueba se mide el poder discriminante de cada una de las variables que se consideran para formar parte del análisis, y se verifica la hipótesis nula: $H_0 : \mu^1_l = \ldots = \mu^g_l$

El Λ_l para la l-ésima variable es

$$\Lambda_l = \frac{\mathbf{D}_{ll}}{\mathbf{T}_{ll}} = \frac{\text{Varianza de la variable l en la matriz de covarianza dentro de los grupos}}{\text{Varianza de la variable l en la matriz de covarianza total}}$$

El estadístico de contraste viene dado por

$$F_l = \frac{(\mathbf{T}_{ll} - \mathbf{D}_{ll})(n-g)}{\mathbf{D}_{ll}(g-1)} \sim F_{g-1,\, n-g}$$

Se rechaza H_0 si $P [F \geq F_{g-1,n-g}]$

Significancia de las Funciones Discriminantes

Una vez calculada las funciones discriminantes, se determina si estas son estadísticamente significativas. Es decir, si el conjunto de las funciones discriminantes separan bien los grupos. En este caso el lambda de wilks viene dado por:

$$\Lambda = \frac{|\mathbf{D}|}{|\mathbf{T}|} = \frac{1}{\prod_{i=1}^{k} (1+\lambda_i)}$$

El número de funciones discriminantes significativas se determina mediante un contraste de hipótesis secuencial. Se tiene **r**, el número de funciones discriminantes significativas. El proceso comienza con **r=0**. En el (r+1)-ésimo paso del algoritmo la hipótesis nula a contrastar es $H_0 = \lambda_{r+1} = \ldots = \lambda_k = 0$ donde **k= min(g-1,p)**.

$$T = \left[n-1-\frac{p+g}{2} \right] \sum_{j=r+1}^{k} \log\,(1+\lambda_i)$$

Es el estadístico de contraste, el cual se distribuye como una $\chi^2_{(p-r)(g-r-1)}$ si H_0 es verdad, con un nivel de significancia α. Se rechaza H_0 si P [$T \geq \chi^2_{(p-r)(g-r-1)}$].

CORRELACIÓN CANÓNICA

$$r^* = \sqrt{\frac{\lambda_i}{1+\lambda_i}}$$

El coeficiente de correlación canónica mide, en términos relativos, el poder discriminante de la m-ésima función discriminante ya que es el porcentaje de la variación total en dicha función que es explicada por las diferencias entre los grupos. Toma valores entre **cero (0)** y **uno (1)**. Valores cercanos a **uno (1)** indican mayor potencia discriminante de la m-ésima función (Salvador 2000).

CÁLCULO DE LA PUNTUACIONES DISCRIMINANTES

Para cada individuo se calcula una puntuación discriminante correspondiente a cada una de las funciones discriminantes. Cada individuo tendrá tantas puntuaciones según el número de funciones discriminantes. Esta puntuación se obtiene reemplazando en la ecuación $A_{jm} = b_0 + b_1X_{1j} + b_2X_{2j} + ... + b_pX_{pj}$ donde, $\mathbf{A_{jm}}$ es la puntuación discriminante del individuo **j** en la función m-ésima, $\mathbf{1 \leq m \leq k}$ (Gondar 2001).

Con estas puntuaciones se permite establecer, para cada individuo, un punto en el plano gráfico formado por las funciones discriminantes. Cada función discriminante actúa como eje de coordenadas. El centroide de grupo esta constituido por las medias de estas puntuaciones, en cada una de las funciones, para un grupo en particular. Habrán tantos centroides como grupos.

MATRIZ DE ESTRUCTURA

Es una matriz que contiene, por filas, los coeficientes de correlación de las funciones discriminantes, con las variables originales. De esta forma es posible detectar el peso o importancia de cada una de las variables originales con respecto a las funciones discriminantes.

REGLA DE CLASIFICACIÓN DE BAYES (PROBABILIDAD A POSTERIORI)

Esta regla permite calcular la probabilidad de pertenencia de cada individuo, con una puntuación discriminante obtenida, a cada grupo y asignar al individuo al grupo que presente mayor probabilidad. La probabilidad estimada de que un individuo **j**, con puntuaciones discriminantes $\mathbf{A=(a_{j1}, a_{j2},..., a_{jk})}$, pertenezca al grupo i es :

$$P(G_i/A) = \frac{P(G_i)\,\left|M^i\right|^{-1/2} e^{-(\chi^2)^i/2}}{\sum_{i=1}^{g} P(G_i)\,\left|M^i\right|^{-1/2} e^{-(\chi^2)^i/2}}$$

El individuo será clasificado en el grupo para el que la probabilidad *a aposteriori* sea la mayor, será clasificado en $\mathbf{G_i}$ si

$$\mathbf{P(G_i/A) = máx\ \{ P(G_1/A), P(G_2/A), ..., P(G_g/A)\}}\ \text{(Ferrán 2001)}$$

RESULTADOS Y DISCUSIÓN

Tenemos 178 observaciones de vinos relativas a 13 variables cuantitativas clasificadas en 3 variedades cultivadas. Las 13 variables que miden las características químicas de los vinos son: $\mathbf{X_1}$: Alcohol. $\mathbf{X_2}$: Ácido Málico. $\mathbf{X_3}$: Ceniza. $\mathbf{X_4}$: Alcalinidad de la ceniza.

X_5: Magnesio. X_6: Total de Fenoles.X_7: Flavenoides X_8: Fenoles no flavenoides. X_9: Proantocianinas. X_{10}: Intensidad del Color. X_{11}: Tonalidad. X_{12}: OD280/OD315 de los vinos diluidos. X_{13}: Prolina.

De las 13 variables, 11 se distribuyen normalmente en cada uno de los grupos. La variable ácido málico no se distribuye normalmente en la Variedad 1 y tampoco en la Variedad 2; y la variable magnesio no tiene una distribución normal en la Variedad 2. Sin embargo, si determinamos la prueba de normalidad a ambas variables sin separar los grupos observamos que la variable ácido málico no se distribuye normalmente, por lo que se toma la decisión de no incluir esta variable como variable del análisis.

Para generar los autovalores y las funciones discriminantes canónicas se seleccionó aleatoriamente el 92.1% de la muestra de vinos (muestra de análisis) que corresponde a 164 observaciones. La clasificación de estos casos seleccionados posteriormente se comparo, con la clasificación de los vinos que no fueron seleccionados (muestra de validación).

La prueba de igualdad de medias nos demuestra que considerando las variables de forma individual, sus valores de significancia son menores a 0.05, por lo tanto, las medias de las 12 variables son diferentes en los tres grupos. (Cuadro 1).

Cuadro 1. Prueba de igualdad de las medias de los grupos.

	Lambda de Wilks	F	gl1	gl2	Sig.
Alcohol	.40	117.08	2	161	**.000**
Ceniza	.90	8.93	2	161	**.000**
Alcalinidad de la Ceniza	.69	35.86	2	161	**.000**
Magnesio	.89	9.86	2	161	**.000**
Total de Fenoles	.48	85.11	2	161	**.000**
Flavenoides	.27	208.77	2	161	**.000**
Fenoles No	.74	28.32	2	161	**.000**
Proantocianin	.73	28.78	2	161	**.000**
Intensidad del	.41	111.99	2	161	**.000**
Tonalida	.44	100.18	2	161	**.000**
OD280/OD315 de Vinos	.30	181.44	2	161	**.000**
Prolina	.30	180.38	2	161	**.000**

En el cuadro 2 se contrasta la igualdad de matrices de varianzas-covarianzas por medio de la M de Box, en donde se indica que el valor p=0.000 ($\leq$ 0.05) y por lo tanto se rechaza la hipótesis nula de igualdad de varianzas-covarianzas. "La razón por la cual se rechaza la igualdad de matrices es porque para las muestras grandes la prueba de las matrices de varianzas-covarianzas es tan poderosa que casi siempre rechaza la igualdad de esas matrices. Sin embargo, la diferencias entre las matrices de covarianzas pueden no ser suficientemente grandes como para tener alguna importancia práctica real."(Johnson 2000).

Cuadro 2. Resultados de la prueba M de Box.

M de Box		702.24
F	Aprox.	4.014
	gl1	156
	gl2	58933.47
	Sig.	**.000**

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

Los autovalores o valores propios asociados con la primera y segunda funciones discriminantes canónicas, se muestran en el cuadro 3. Son dos funciones discriminantes debido a que el min(3-1,12)=2. El valor propio asociado a la primera función es 9.195 y explica el 71.3% de la varianza en los datos. Este primer valor propio es muy alto, por lo tanto esta función aporta mucha información y discriminará muy bien. Con respecto a las correlaciones canónicas ambas con valores grandes, cercanos a uno, esto quiere decir que la segunda función discrimina bien a pesar que solo explique el 28.7% de la varianza.

Cuadro 3. Autovalores.

Funcion	Autovalor	% de Varianza	% Acumulado	Correlación Canónica
1	9.195[a]	71.3	71.3	.950
2	3.700[a]	28.7	100.0	.887

a. Se han empleado las 2 primeras funciones discriminantes canónicas en el análisis (Manzano 2001).

El cuadro 4 presenta los coeficientes de las funciones discriminantes (coeficientes no estandarizados), con los que se construyen las funciones discriminantes canónicas. Con estas funciones se obtendrán las puntuaciones discriminantes para cada individuo.

Cuadro 4. Coeficientes de las funciones canónicas discriminantes.

	Función	
	1	2
Alcohol	.421	.950
Ceniza	.526	2.414
Alcalinidad de la Ceniza	-.173	-.140
Magnesio	.00033	-.0032
Total de Fenoles	-.731	-.107
Flavenoides	1.629	-.557
Fenoles No Flavenoides	.916	-1.335
Proantocianinas	-.262	-.295
Intensidad del Color	-.303	.251
Tonalidad	1.123	-2.109
OD280/OD315 de los Vinos Diluidos	1.241	.099
Prolina	.003	.003
(Constante)	-9.896	-14.215

coeficientes no estandarizados

Ahora se determina si las funciones generadas son estadísticamente significativas, es decir si el conjunto de las funciones permite que las medias de los grupos estén separadas. En el cuadro Lambda de Wilks (Cuadro 5) se observa que el valor de la lambda de wilks para el conjunto formado por las dos funciones es 0.021; el segundo valor obtenido cuando se elimina la primera función es 0.213, y corresponde a la segunda función. Ambos valores son cercanos a cero (0) por lo que se concluye que ambas funciones tienen un alto poder discriminatorio. Esto se corrobora al ver que los p-valores asociados al estadístico chi-cuadrado son menores que 0.05 (columna "sig") por lo que se rechaza las hipótesis nulas de $\lambda_1=\lambda_2=0$ y $\lambda_2=0$, y concluimos que ambas funciones son significativas.

Cuadro 5. Lambda de Wilks.

Contraste de las funciones	Lambda de Wilks'	Chi-cuadrado	gl	Sig.
1 through 2	.021	601.71	24	.000
2	.213	240.65	11	.000

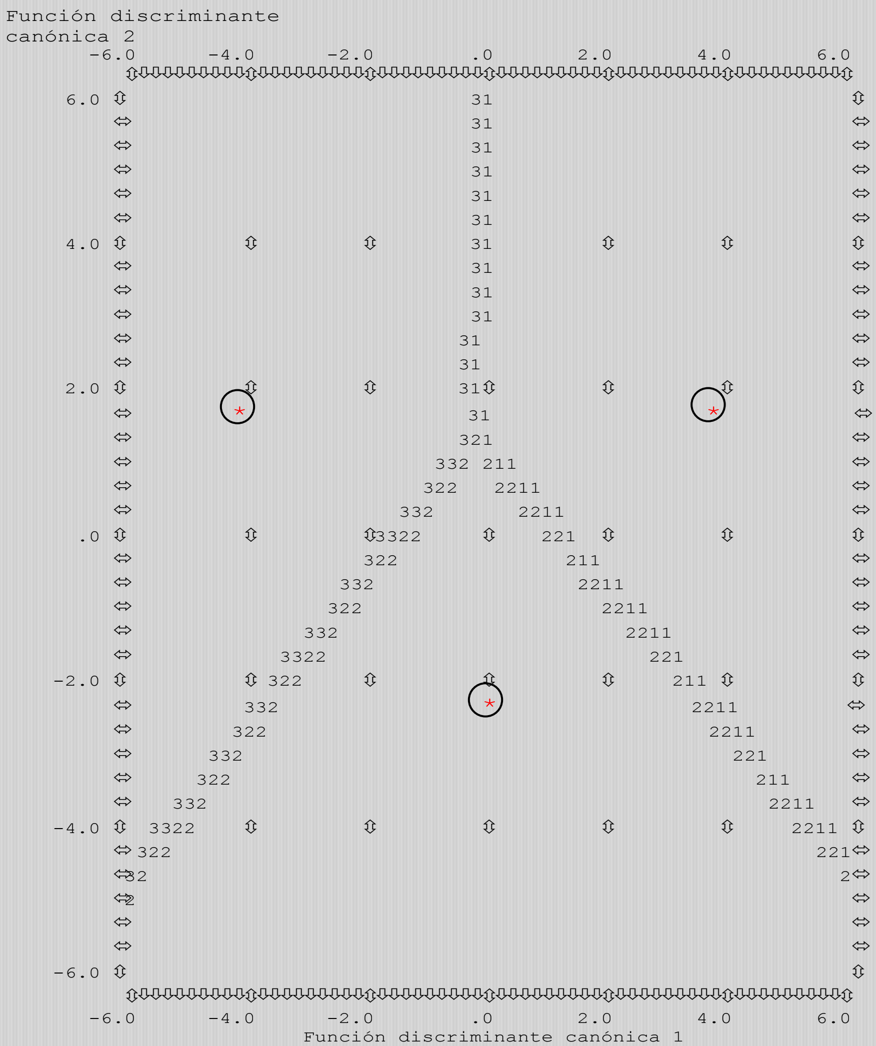

Fig. 1. Mapa Territorial.

Los centroides están representados gráficamente en el mapa territorial (Fig. 1). Como se puede observar, cada eje es representado por las funciones discriminantes (con coeficientes no estandarizados), en donde los números 1, 2 y 3 delimitan el espacio de cada grupo y los asteriscos (*) representan los centroides de cada grupo. Se puede observa cómo discriminan perfectamente, en conjunto ambas funciones, entre los grupos manteniendo a los centroides alejados unos de otros. Sin embargo, se puede ver que los centroides del grupo 1 y 3 están casi alineados, por lo que se puede decir que la función 2 no hace una buena discriminación entre estos dos grupos, a pesar de ser significativa.

En la matriz de estructura se observan las correlaciones que hay entre las variables y cada una de las 2 funciones. Las variables están agrupadas dentro de cada función según el tamaño de la correlación y están marcadas con un asterisco (*).

La variable flavenoides, es la que tiene la mayor correlación con la función 1. En la función 2, la variable alcohol es la que tiene la correlación más alta con esta función.

Cuadro 6. Matriz de estructura. Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas.

	Función	
	1	2
Flavenoides	.529*	-.068
OD280/OD315 de los Vinos	.476*	-.214
Total de Fenoles	.339*	.026
Tonalidad	.313*	-.305
Alcalinidad de la Ceniza	-.207*	-.116
Proantocianinas	.197*	-.012
Fenoles No Flavenoides	-.196*	.006
Alcohol	.152	.579*
Intensidad del Color	-.164	.556*
Prolina	.361	.531*
Ceniza	-.002	.173*
Magnesio	.066	.149*

*.Variables ordenadas por el tamaño de la correlación con la función.

A partir de las puntuaciones discriminantes en las dos funciones (columnas "puntuaciones discriminantes") se obtiene la máxima probabilidad *a posteriori* de pertenecer a un grupo (columna "P(G=g│A=a)"), el grupo en el cual se ha obtenido dicha probabilidad está indicado en la columna "Grupo pronosticado". El grupo pronosticado se compara con el grupo real al que pertenece el individuo (columna "Grupo real"). Si uno de los vinos ha sido clasificado en el mismo grupo al cual pertenecía inicialmente entonces las funciones discriminantes lo han clasificado correctamente. Las clasificaciones erróneas son mostradas con un doble asterisco (**) en la columna que contiene la identificación del grupo real, en nuestro caso todos los vinos fueron clasificados correctamente. Se puede notar que todas las probabilidades *a posteriori* obtenidas para cada uno de los vinos son altas. Las observaciones marcadas con (u) son aquellas que no fueron seleccionadas para generar las funciones discriminantes.

Cuadro 7. Estadísticas por casos

Estadísticos							
Individuo	**Grupo real**	**Grupo mayor**				**Puntuaciones discriminantes**	
		Grupo pronosticado	**P(A>a / G=g)**		**P(G=g | A=a)**	**F1**	**F2**
			p	gl			
1	1	1	.435	2	1.000	4.782	4.782
2	1	1	.571	2	1.000	4.520	1.046
3	1	1	.941	2	1.000	3.506	1.165
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
14	1	1	.120	2	1.000	5.587	1.911
15	1	1	.020	2	1.000	5.746	3.268
16[u]	1	1	.552	2	1.000	3.544	2.600
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
86	2	2	.671	2	1.000	.819	-2.186
87[u]	2	2	.369	2	1.000	-1.021	-3.386
88	2	2	.652	2	1.000	-.456	-3.183
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
163[u]	3	3	.348	2	1.000	-3.243	.492
164	3	3	.363	2	1.000	-2.795	1.340
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
176	3	3	.440	2	1.000	-4.566	2.814
177	3	3	.605	2	1.000	-4.174	2.589
178	3	3	.215	2	1.000	-5.279	2.965

Como último paso está probar la calidad de la clasificación de los vinos por medio de la matriz de clasificación. Los resultados se muestran en el cuadro 8 resultados de la clasificación. La clasificación en base a la muestra de análisis (casos seleccionados) indica que por cada grupo el 100% de los vinos fueron clasificados correctamente. Cuando se ha hecho la clasificación en base a la muestra de validación (casos no seleccionados) el 100% de los vinos fueron clasificados correctamente. Con esto se concluye que las funciones obtenidas tiene una buena capacidad predictora.

Cuadro 8. Resultados de la clasificación

				Grupo de pronosticado			
			Variedad	1	2	3	Total
Casos seleccionados	Original	Recuento	1	54	0	0	54
			2	0	65	0	65
			3	0	0	45	45
		%	1	100.0	.0	.0	100.0
			2	.0	100.0	.0	100.0
			3	.0	.0	100.0	100.0
Casos no seleccionados	Original	Recuento	1	5	0	0	5
			2	0	6	0	6
			3	0	0	3	3
		%	1	100.0	.0	.0	100.0
			2	.0	100.0	.0	100.0
			3	.0	.0	100.0	100.0

CONCLUSIONES

Los valores cercanos a cero (0) del Lambda de Wilks obtenidos en la prueba de significancia de las funciones discriminantes canónicas generadas nos demuestra que ambas funciones tienen un alto poder discriminante.

Por medio del mapa territorial se puede visualizar claramente el poder discriminatorio de ambas funciones discriminantes canónicos al observar que los centroides de los grupos están alejados uno del otro.

Mediante la matriz de estructura podemos observar que las variable Flavenoides y OD280/OD315 de los vinos diluidos poseen mayor correlación con la función 1; y, las variables alcohol, intensidad de color y prolina son las que poseen mayor correlación con la función 2.

Mediante, la matriz de clasificación se muestra que el 100% de los vinos fueron clasificadas correctamente por medio de las funciones discriminantes canónicas, por lo que se comprueba que estas poseen un alto poder discriminante y pueden ser utilizadas para futuras pruebas de clasificación de individuos nuevos.

REFERENCIAS

Ferrán, A., M. 2001. *Spss para windows – Análisis estadístico*. España, McGraw-Hill/Interamericana.

Gondar, N., J. E. 2001. *Análisis discriminante* [Internet] Disponible en: <http://www.estadistico.com/arts.html?20011112>.

Johnson, D., E. 2000. *Métodos multivariados aplicados al análisis de datos*. Edición en español. C. V., International Thomson Editores, S. A., pp 240.

Manzano, A. J. 2001. *El análisis discriminante* [Internet] Universidad de Valencia. Disponible en: : <http://www.uv.es>

Mardia, K. V., et al. Kent, J. T. e J. M. Bibby. 1979. *Multivariate analysis*. 6ta Edición. London, Academic Press.

Rodríguez, C. 1999. *El análisis discriminante y su aplicación.* Vicerrectoría de Investigación y Postgrado, Universidad de Panamá.

Salvador, F., M. 2000. *Análisis discriminante* [Internet] Disponible en: <http://www.5campus.com/leccion/discri>

Recibido noviembre de 2003, aceptado junio de 2004.