Cap. VI

Análisis Discriminante

6.1.- PLANTEAMIENTO DEL PROBLEMA

- Una población de $\boldsymbol{n}$ objetos, particionada en $\boldsymbol{q}$ grupos
- Cada grupo G_i tiene n_i elementos, de modo que $n = \sum_{i=1}^{q} n_i$.
- $\boldsymbol{Y} = (Y_1 \ldots Y_p)$ una matriz de p variables numéricas observadas sobre los objetos con el fin de utilizar dicha información para discriminar entre los q grupos anteriores.

Objetivos del Análisis Discriminante:

1) Analizar si existen diferencias entre los grupos en cuanto a su comportamiento con respecto a las variables consideradas y averiguar en qué sentido se dan dichas diferencias.

   ➔ ***Análisis Discriminante Descriptivo***

2) Elaborar procedimientos de clasificación sistemática de individuos de origen desconocido, en uno de los grupos analizados

   ➔ ***Análisis Discriminante Predictivo***

<u>Ejemplo 1:</u> Discriminación con dos grupos

Una empresa está interesada en analizar la opinión de sus clientes con respecto a su labor comercial y de gestión. Para ello realiza una encuesta a una muestra de 100 de ellos en las que le pide que valoren su labor en los siguientes aspectos, haciendo una valoración entre 0 y 10: Velocidad de Entrega (VENTREGA), Nivel de Precios (NIVPREC), Flexibilidad de Precios (FLEXPREC), Imagen de la Empresa (IMGEMPR), Servicio (SERVICIO), Imagen de Ventas (IMGVENTA) y Calidad de Producto (CALIDAD).

Además, tiene clasificados a sus clientes en dos grupos de acuerdo al tamaño de la empresa en la que trabajan: Empresas Pequeñas (TAMAÑO=1) y Empresas Grandes (TAMAÑO=2). El número de clientes pertenecientes a empresas pequeñas es igual a 60 y el de empresas grandes es igual a 40.

El objetivo del estudio es analizar si existen diferencias en cuanto a la percepción de su labor empresarial entre los clientes de un grupo y del otro y, en caso de que existan, analizar en qué sentido se dan dichas diferencias.

En este caso, por lo tanto, existen 7 variables clasificadoras (p=7) y dos grupos a discriminar (q=2). El tamaño de la muestra es n=100 con n_1 = 60 y n_2 = 40.

Ejemplo 2 Discriminación con 6 grupos

Se tiene los datos socioeconómicos de 109 países del mundo del año 1995. Dichos países están clasificados de acuerdo a 6 regiones económicas: OCDE, Europa Oriental, Asia/Pacífico, Africa, Oriente Medio y América Latina. Las variables analizadas son el porcentaje de habitantes en ciudades (URBANA), el aumento de la población (INCR_POB), la tasa de natalidad (TASA_NAT), la tasa de mortalidad (TASA_MOR) y las transformaciones logarítmicas de la población (LOGPOB), la densidad (LOGDENS), la esperanza de vida femenina (LOGESPF) y masculina (LOGESPM), de la tasa de alfabetización (LOGALF), de la tasa de mortalidad infantil (LOGMINF), del cociente nacimientos/muertes (LOGNACDE), de la tasa de fertilidad (LOGFERT) y del PIB percápita (LOGPIBCA).

En este caso se tiene, por lo tanto, que q=6, p=13 y n=109. Además, n_1=21, n_2=14, n_3=17, n_4=19, n_5=17 y n_6=21.

El objetivo del estudio es analizar si existen diferencias entre las diversas regiones socio-económicas y, en caso afirmativo, en qué sentido.

6.2. CÁLCULO DE LAS FUNCIONES DISCRIMINANTES

- La discriminación entre los *q* grupos se realiza mediante el cálculo de unas funciones matemáticas denominadas ***funciones discriminantes***.

- Existen varios procedimientos para calcularlas. Veamos el procedimiento de Fisher.

6.2.1 Procedimiento Discriminante de Fisher

- Considera como *funciones discriminantes*, combinaciones lineales de las funciones clasificadoras, es decir:

$$D = u_1Y_1 + u_2Y_2 + \dots + u_pY_p = \boldsymbol{u'Y}$$

- El vector ***u*** de coeficientes que define una variable discriminante se calcula maximizando:

$$\frac{\text{Variabilidad entre grupos}}{\text{Variabilidad intra grupos}} = \frac{\dfrac{\sum_{g=1}^{q} n_g \left(\bar{d}_g - \bar{d}\right)^2}{q-1}}{\dfrac{\sum_{g=1}^{q}\sum_{k=1}^{n_g} \left(d_{gk} - \bar{d}_g\right)^2}{n-q}} = \frac{\mathbf{u'Bu}}{\mathbf{u'Wu}} \; \frac{n-q}{q-1}$$

donde d_{ik}, $k=1,\dots,n_i$; $i=1,\dots,q$ denota el valor de D en la k-ésima observación del i-ésimo grupo y

- **W =** $\sum_{g=1}^{G} W_g$ = $\sum_{g=1}^{G}\sum_{k=1}^{n_j}\left(\mathbf{y}_{g\mathbf{k}}-\overline{\mathbf{y}}_g\right)\left(\mathbf{y}_{g\mathbf{k}}-\overline{\mathbf{y}}_g\right)'$ =

$$\begin{bmatrix} \sum_{g=1}^{G}\sum_{k=1}^{n_g}\left(y_{1gk}-\overline{y}_{1g}\right)^2 & \cdots & \sum_{g=1}^{G}\sum_{k=1}^{n_g}\left(y_{1gk}-\overline{y}_{1g}\right)\left(y_{Kgk}-\overline{y}_{Kg}\right) \\ \cdots & \cdots & \cdots \\ \sum_{g=1}^{G}\sum_{k=1}^{n_g}\left(y_{Kgk}-\overline{y}_{Kg}\right)\left(y_{1gk}-\overline{y}_{1g}\right) & \cdots & \sum_{g=1}^{G}\sum_{k=1}^{n_g}\left(y_{Kgk}-\overline{y}_{Kg}\right)^2 \end{bmatrix}$$

es *la matriz de suma de cuadrados intra-grupos.*

- **B** = $\sum_{g=1}^{G} n_g\left(\overline{\mathbf{y}}_g-\overline{\mathbf{y}}\right)\left(\overline{\mathbf{y}}_g-\overline{\mathbf{y}}\right)'$ = $\begin{bmatrix} \sum_{g=1}^{G} n_g\left(\overline{y}_{1g}-\overline{y}_1\right)^2 & \cdots & \sum_{g=1}^{G} n_g\left(y_{1g}-\overline{y}_1\right)\left(y_{Kg}-\overline{y}_K\right) \\ \cdots & \cdots & \cdots \\ \sum_{g=1}^{G} n_g\left(y_{Kg}-\overline{y}_K\right)\left(y_{1g}-\overline{y}_1\right) & \cdots & \sum_{g=1}^{G} n_g\left(\overline{y}_{Kg}-\overline{y}_K\right)^2 \end{bmatrix}$

es *la matriz de suma de cuadrados inter-grupos.*

- Se impone, además, la condición de normalización **u'Wu** = **1**
- La solución viene dada por el vector propio $\mathbf{u}_1$ de $\mathbf{W}^{-1}\mathbf{B}$ asociado al mayor valor propio $\boldsymbol{\lambda}_1$ de esta matriz.

OBSERVACIONES:

- Si se quieren calcular r funciones discriminantes con varianza 1, y que sean incorreladas entre sí, es decir, que verifiquen que $\mathbf{u}_i^{'}\mathbf{W}\mathbf{u}_j = \mathbf{I}_{ij}$; i,j=1,…,r, se obtienen como soluciones los r vectores propios de $\mathbf{W}^{-1}\mathbf{B}$ asociados a los r mayores valores propios de esta matriz $\lambda_1 \geq \ldots \geq \lambda_r > 0$. A las funciones $\mathbf{D}_i = \mathbf{u}_i^{'}\mathbf{Y}$ i=1,…,r se les llama ***funciones discriminantes canónicas*** *o* ***funciones discriminantes de Fisher***.

- Si r es el número de funciones discriminantes, se tiene que $\mathbf{W}_D = \mathbf{I}_r$ y $\mathbf{B}_D = diag(\lambda_1,\ldots,\lambda_r)$ donde $\mathbf{W}_D$ y $\mathbf{B}_D$ son las matrices **W** y **B** calculadas utilizando las puntuaciones discriminantes.

- Los valores propios λ_i ; i=1,...,r miden el poder de discriminación de la i-ésima variable discriminante de forma que si $\lambda_i = 0$, la variable discriminante no tiene ningún poder discriminante. Dado que el rango de la matriz $\mathbf{W}^{-1}\mathbf{B}$ es a lo más min {q-1,p}, el número máximo de funciones discriminantes que se podrán calcular será igual a min {q-1,p}.

6.2.2 Lambda de Wilks

- Es un estadístico que mide el poder discriminante de un conjunto de variables. Viene dada por

$$\Lambda = \frac{|\mathbf{W}|}{|W + B|} = \frac{1}{\prod_{i=1}^{\min(q-1,p))}(1+\lambda_i)} \quad ,$$

y toma valores entre 0 y 1 de forma que, cuanto más cerca de 0 esté, mayor es el poder discriminante de las variables consideradas y cuanto más cerca de 1, menor es dicho poder.

6.2.3 Correlación Canónica

- La i-ésima **correlación canónica** viene dada por:

$$CR_i = \sqrt{\frac{\lambda_i}{1+\lambda_i}} \qquad i = 1,\ldots,r$$

- Mide, en términos relativos, el poder discriminante de la i-esima variable discriminante ya que es el porcentaje de la variación total en dicha variable que es explicada por las diferencias entre los grupos.

- Toma valores entre 0 y 1 de forma que, cuanto más cerca de 1 esté su valor, mayor es la potencia discriminante de la i-esima variable discriminante.

6.2.4 Determinación del número de funciones discriminantes

- El número de funciones discriminantes significativas se determina mediante un contraste de hipótesis secuencial.

- Si k es el número de funciones discriminantes significativas el proceso comienza con k=0. En el (k+1)-ésimo paso del algoritmo la hipótesis nula a contrastar es

$$H_o: \lambda_{k+1} = \dots = \lambda_{\min\{G-1,p\}} = 0$$

y el estadístico de contraste viene dado por:

$$T = \left(n-1-\frac{p+q}{2}\right)\sum_{j=k+1}^{\min\{q-1,p\}} \log\left(1+\lambda_j\right)$$

el cual se distribuye como una $X^2_{(p-k)(q-k-1)}$ bajo H_o.

Ejemplo1 (continuación)

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	2,046[a]	100,0	100,0	,820

a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	,328	105,244	7	,000

- En las tablas se muestra los valores de λ_1 = 2.046 y de la correlación canónica $\sqrt{\frac{2.046}{1+2.046}}$ = 0.82 obtenidos mediante el programa SPSS. Así mismo, se muestra el resultado obtenido al aplicar el test de hipótesis secuencial utilizado para determinar el número de funciones discriminantes significativas.

- En este caso el número máximo de funciones discriminantes posibles es igual a min{2-1,7} = 1 por lo que sólo será necesario llevar a cabo un test de hipótesis.

- La hipótesis nula será H_o: $\lambda_1 = 0$ y el valor del estadístico T=105.244 correspondiente a una lambda de Wilks igual a 0.328. El p-valor es igual a $P\left[\chi_7^2 \geq 105.244\right]$=0.000 por lo que la función obtenida es significativa y su poder discriminante es alto dado el elevado valor de la correlación canónica.

Ejemplo 2 (continuación)

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,004[a]	65,7	65,7	,926
2	1,182[a]	12,9	78,7	,736
3	,949[a]	10,4	89,1	,698
4	,610[a]	6,7	95,7	,616
5	,390[a]	4,3	100,0	,530

a. Se han empleado las 5 primeras funciones discriminantes canónicas en el análisis.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 5	,015	396,850	65	,000
2 a la 5	,105	212,911	48	,000
3 a la 5	,229	139,173	33	,000
4 a la 5	,447	76,111	20	,000
5	,720	31,101	9	,000

- En este caso el número de funciones discriminantes calculadas es igual a min{6-1,13}=5. En las tablas subsiguientes se muestran los valores propios, la correlación canónica y el porcentaje de varianza de discriminación y el porcentaje acumulado explicados por cada función discriminante, los cuales vienen dados por

$$100\frac{\lambda_i}{\sum_{j=1}^{5}\lambda_j} \quad \text{y} \quad 100\frac{\sum_{j=1}^{i}\lambda_j}{\sum_{j=1}^{5}\lambda_j} \quad \text{respectivamente.}$$

- Se observa, por ejemplo, que las 3 primeras funciones discriminantes explican un 89.1% de la varianza de discriminación. Así mismo, en la siguiente tabla se realiza el contraste para la determinación del número de funciones discriminantes significativas. Así, por ejemplo, en el paso 3 se contrasta la hipótesis nula:

$$H_o: \lambda_3 = \lambda_4 = \lambda_5$$

- En este caso T_{obs} = 139.173 y el p-valor $P\left[\chi^2_{33} \geq 139.173\right]$=0 y se rechazaría la hipótesis nula. Se observa que todas las funciones discriminantes son significativas.

6.3. INTERPRETACIÓN DE LOS RESULTADOS

- Significado de las dimensiones de discriminación entre los grupos proporcionadas por las funciones discriminantes mediante el análisis de la matriz de estructura y de la de los coeficientes estandarizados de las funciones discriminantes.

- Análisis del sentido de la discriminación entre dichos grupos, es decir, averiguar qué grupos separa cada función discriminante y en qué sentido. Este análisis se lleva a cabo mediante representaciones gráficas del espacio de discriminación así como de perfiles multivariantes correspondientes a cada grupo.

6.3.1 Matriz de Estructura

Es una matriz pxr que contiene, por filas, los coeficientes de correlación de las funciones discriminantes con las variables originales. De esta forma es posible interpretar el significado de las mismas utilizando, para cada una de ellas, aquéllas variables con las que está más correlacionada. De cara a facilitar dicha interpretación se suelen realizar rotaciones ortogonales del espacio de discriminación similares a las utilizadas por el Análisis Factorial.

6.3.2 Coeficientes estandarizados de las funciones discriminantes

Vienen dados por la expresión:

$$\mathbf{u^*} = \mathbf{F}^{-1}\mathbf{u}$$

donde **F** = $\mathrm{diag}\left(s_{jj}^{1/2}\right)$ siendo s_{jj} elemento de la diagonal de la matriz $\overline{\mathbf{S}} = \frac{\mathbf{W}}{n-G}$. A partir de ellos se puede deducir la expresión matemática de las funciones discriminantes en términos de las variables originales estandarizadas. Estos coeficientes son poco fiables si existen problemas de multicolinealidad entre las variables clasificadoras.

Ejemplo 1 (continuación)

En las tablas subsiguientes se muestran los coeficientes estandarizados de la función discriminante estimada así como la matriz de estructura. La expresión mátemática de dicha función vendrá dada por:

$$D = 0.466Z_{ventrega} + 0.084Z_{nivprec} + 0.538Z_{flexprec} - 0.068Z_{imgempr}$$
$$-0.093Z_{servicio} + 0.295Z_{imgventa} - 0.6784Z_{calidad}$$

donde Z_i indica la variable i-ésima estandarizada.

Coeficientes estandarizados de las funciones discriminantes canónicas

	Función
	1
Velocidad de Entrega	,466
Nivel de Precios	,084
Flexibilidad de Precios	,538
Imagen de la Empresa	-,068
Servicio	-,093
Imagen de las Ventas	,295
Calidad del Producto	-,684

Matriz de estructura

	Función
	1
Calidad del Producto	-,656
Flexibilidad de Precios	,592
Velocidad de Entrega	,568
Nivel de Precios	-,332
Servicio	,147
Imagen de las Ventas	,030
Imagen de la Empresa	-,020

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas
Variables ordenadas por el tamaño de la correlación con la función.

Analizando la matriz de estructura de la función discriminante se observa que dicha función realiza un contraste entre la Velocidad de Entrega y la Flexibilidad de Precios, por un lado, y la Calidad del Producto y el Nivel de Precios, por el otro, de forma que clientes con un valor de D positivo serán clientes con una tendencia a valorar por encima de la media a la labor de la empresa en aspectos más específicos como rapidez y flexibilidad y a valorar por debajo aspectos más genéricos como son la calidad del producto y el nivel de precios. Lo contrario ocurre con clientes con valores de D negativos.

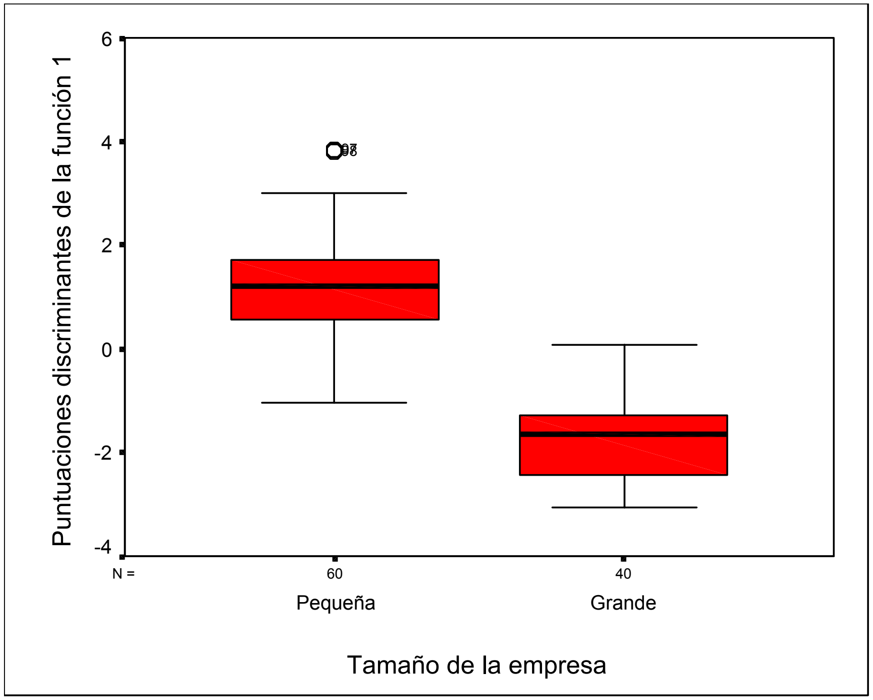
La siguiente tabla contiene las puntuaciones medias $\overline{d}_i$; i=1,2 para cada grupo.

Funciones en los centroides de los grupos

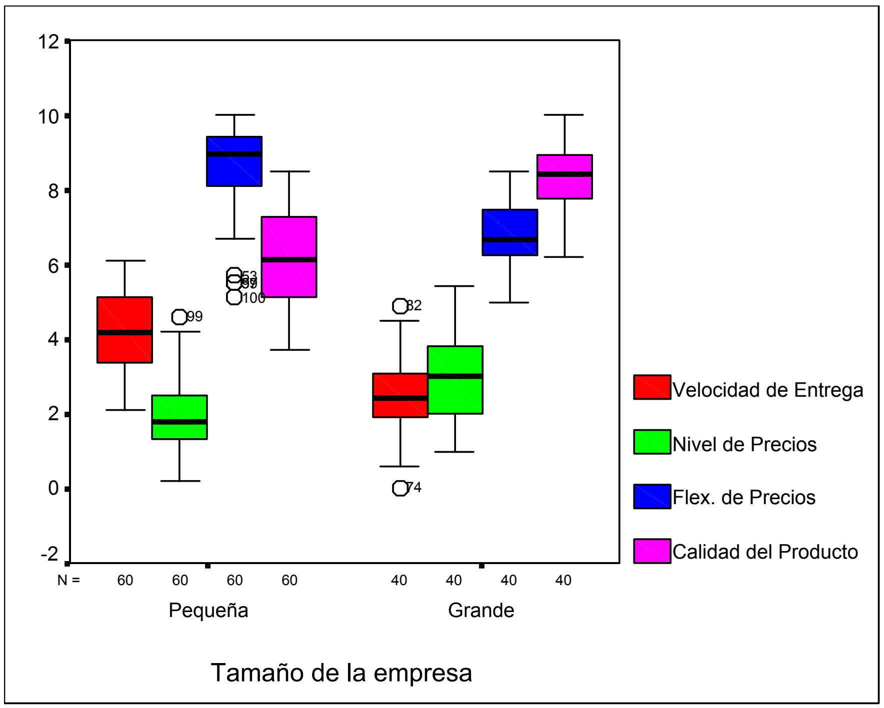
	Función
Tamaño de la empresa	1
Pequeña	1,156
Grande	-1,734

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

y el gráfico subsiguiente los diagramas de caja de dichas puntuaciones



Se observa que, las empresas pequeñas, tienden a valorar mejor a la empresa en aspectos más específicos como son la velocidad de entrega y flexibilidad de precios y, por el contrario, las empresas grandes tienden a valorar mejor los aspectos más generales como son el nivel de precios y la calidad del producto ofrecido. Estos resultados se confirman al comparar los diagramas de caja de cada una de las variables en los dos grupos como se muestra en el gráfico siguiente



Ejemplo 2 (continuación)

En la tabla subsiguiente se muestra la matriz de estructura de las funciones discriminantes tras aplicar una rotación Varimax. Se observa que la primera función discrimina en función del equilibrio demográfico existente en cada país, la segunda tiene que ver con aspectos relacionados con el crecimiento demográfico del mismo, la tercera con su calidad de vida y la quinta con su tamaño demográfico. La cuarta función, cuyo poder discriminante no es muy alto, no ofrece una interpretación tan clara aunque parece estar relacionado con el nivel de desarrollo económico-cultural del país debido a su mayor correlación con PIBPCA, Habitantes en ciudades y tasa de alfabetización.

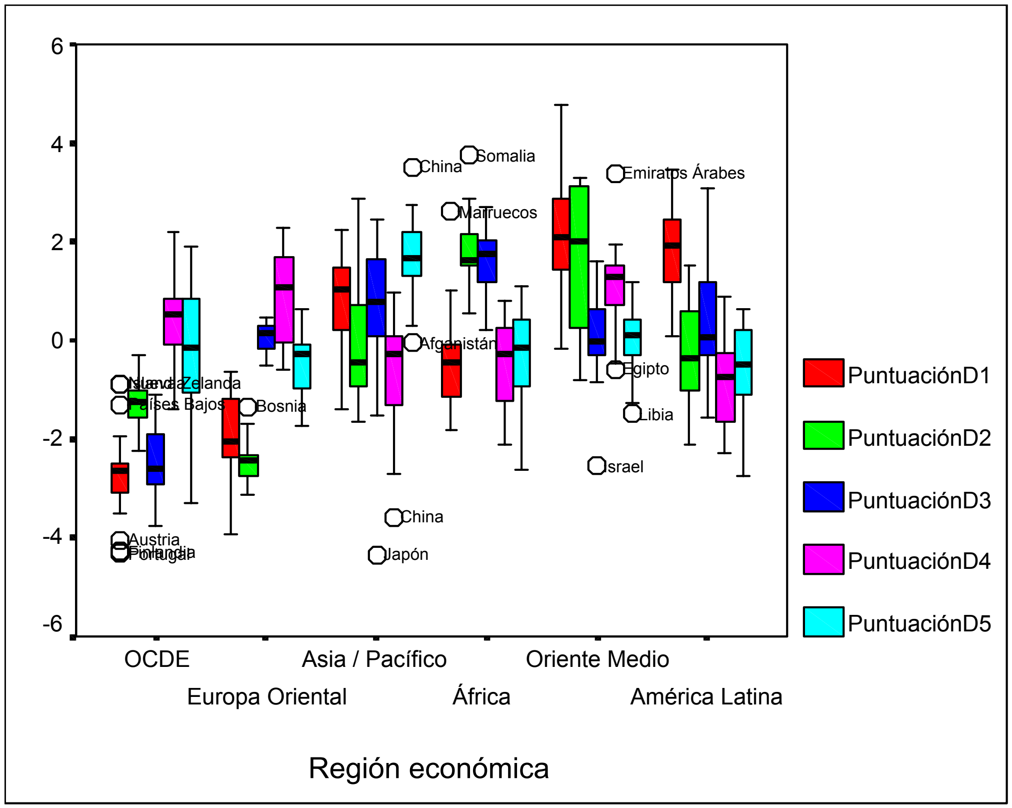
Matriz de estructura rotada

	Función				
	1	2	3	4	5
LOGNACDE	,644*	,410	-,088	-,035	-,030
Tasa de mortalidad (por 1.000 habitantes)	-,566*	,251	,480	-,154	-,112
Tasa de natalidad (por 1.000 habitantes)	-,084	,745*	,454	-,197	-,137
Aumento de la población (% anual)	,190	,738*	,124	,046	-,115
LOGFERT	-,038	,733*	,473	-,012	-,230
LOGALF	-,008	,602*	,420	-,322	,051
LOGESPF	-,128	,291	,890*	-,021	,032
LOGESPM	-,197	,158	,846*	,035	-,014
LOGMINF	-,036	,337	,786*	-,082	-,009
LOGPIBCA	,040	-,146	-,709*	,447	-,097
Habitantes en ciudades (%)	,320	-,261	-,545*	,359	-,256
LOGDENS	,053	-,112	,024	,085	,616*
LOGPOB	-,040	-,049	,014	-,113	,554*

Correlaciones intra-grupo combinadas y rotadas entre las variables discriminantes y las funciones canónicas discriminantes estandarizadas.
Variables ordenadas por el tamaño de la correlación con la función.

*. Mayor correlación absoluta entre cada variable y cualquier función discriminante.

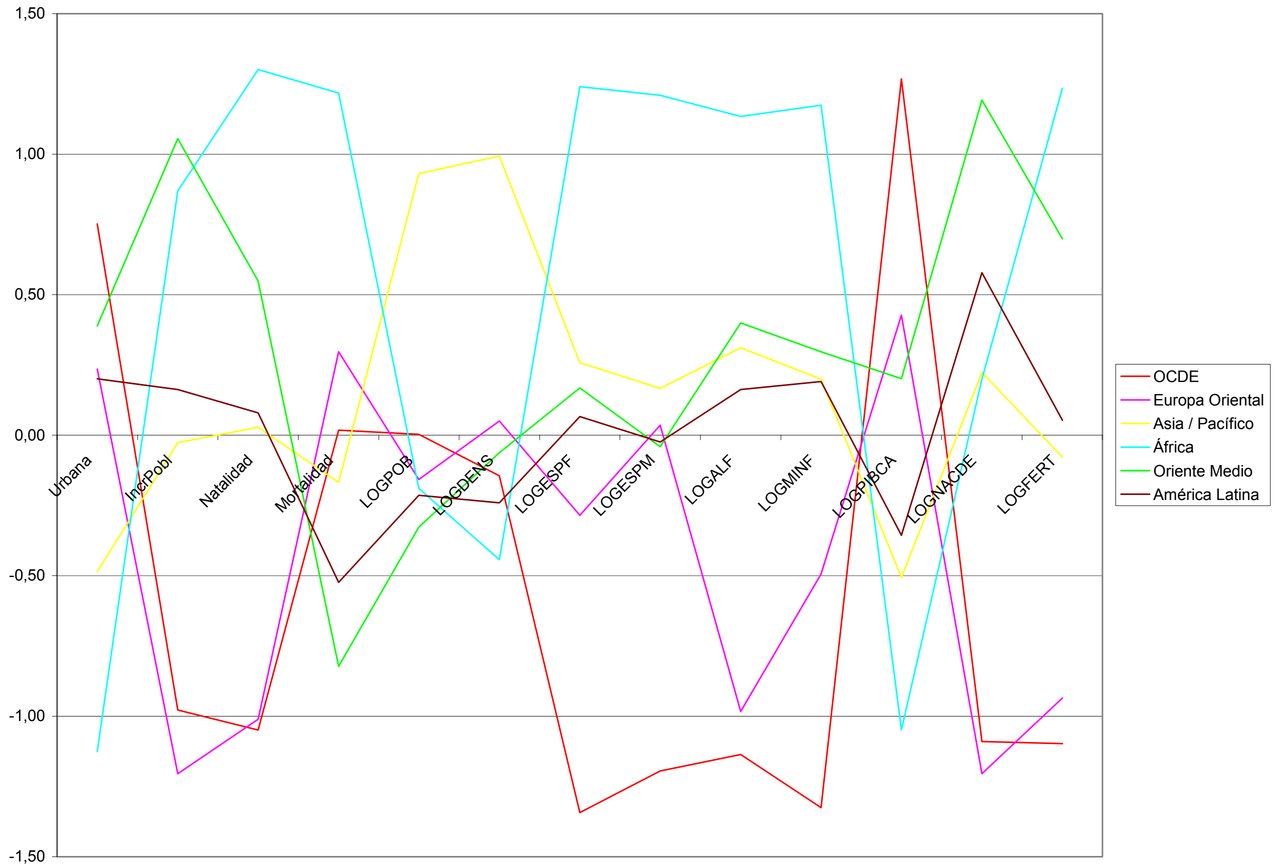
En el gráfico siguiente se muestra el diagrama de cajas de las puntuaciones discriminantes estimadas para cada uno de los países clasificados por región económica.



Se observa que las dos primeras funciones discriminantes separan, esencialmente, a los países de la OCDE y de la Europa Oriental del resto debido al mayor equilibrio demográfico existente en las dos regiones anteriores por su baja natalidad y su baja mortalidad. La tercera función separa a los países de la OCDE debido a su mayor nivel de vida que se traduce en una mayor esperanza de vida, un mayor PIB per cápita y un mayor porcentaje de hombres y mujeres viviendo en ciudades. La cuarta función discrimina, esencialmente a las regiones más pobres y menos desarrolladas (Asia/Pacífico, Africa y América Latina) frente a las más ricas y menos desarrolladas (OCDE, Europa Oriental y Oriente Medio)

La quinta función separa a los países asiáticos del resto debido a su mayor población y su mayor densidad. Respecto a la cuarta no se ve un patrón claro de separación.

La siguiente figura muestra los perfiles multivariantes de cada región los cuales corroboran las interpretaciones anteriores.



6.4.- SELECCIÓN DE VARIABLES CLASIFICADORAS

El problema de selección de variables intenta responder a la pregunta ¿Son necesarias todas las variables clasificadoras para discriminar?

Para responderla existen, esencialmente, tres tipos de algoritmos: algoritmos de selección de variables hacia adelante, eliminación hacia atrás y de regresión por pasos.

Los *algoritmos de selección hacia adelante* comienzan eligiendo la variable que más discrimina entre los q grupos. A continuación seleccionan la segunda más discriminante y así sucesivamente. Si de las variables que quedan por elegir ninguna discrimina de forma significativa entre los grupos analizados el algoritmo finaliza.

Los *algoritmos de eliminación hacia detrás* proceden de forma inversa a los anteriores. Se comienza suponiendo que todas las variables son necesarias para discriminar y se elimina la menos discriminante entre los grupos analizados y así sucesivamente. Si las variables no eliminadas discriminan significativamente entre los grupos analizados el algoritmo finaliza.

Los *algoritmos de regresión por pasos* utilizan una combinación de los dos algoritmos anteriores permitiendo la posibilidad de arrepentirse de decisiones tomadas con precipitación bien sea eliminando del conjunto seleccionado una variable introducida en el conjunto de discriminación en un paso anterior del algoritmo, bien sea introduciendo en dicho conjunto una variable eliminada con anterioridad.

Para determinar qué variables entran y salen en cada paso de este tipo de algoritmos se utilizan diversos criterios de entrada y salida. Uno de los más utilizados es el del Lambda de Wilks que es el que exponemos, a continuación. Otros criterios pueden verse, por ejemplo, en el manual del SPSS.

6.4.1 Criterio de la lambda de Wilks

Utiliza la lambda de Wilks para medir la potencia discriminante ganada/perdida al introducir/sacar una variable del conjunto de discriminación.

Sea Λ_q la lambda de Wilks basada en las *q* primeras variables.

Para ver si es necesario incluir la variable $\mathbf{Y_{q+1}}$ en el conjunto de discriminación se utiliza el estadístico

$$F = \frac{n-G-q}{G-1}\left(\frac{\Lambda_q}{\Lambda_{q+1}}-1\right) \qquad \sim \mathbf{F}_{G-1,\, n-G-q}$$

si la variable $\mathbf{Y_{q+1}}$ no aporta información relevante al proceso de discriminación entre los grupos. Un valor alto/bajo de F indica una pérdida significativa/no significativa de información si la variable Y_{q+1} no es incluida/es incluida en el conjunto de discriminación.

Utilizando dicha variable es posible, por ejemplo, proporcionar un p-valor de entrada y otro de salida de forma que si el p-valor obtenido al introducir una variable en el conjunto de discriminación, no es inferior al p-valor de entrada, la variable considerada no entra en dicho conjunto y si el p-valor obtenido al eliminarla del conjunto de discriminación no es superior al de salida, la variable considerada no sale de dicho conjunto.

Ejemplo 1 (continuación)

En las tablas subsiguientes se muestran los resultados obtenidos al aplicar el algoritmo de selección de variables utilizado por SPSS tomando como criterio de entrada un p-valor igual a 0.05 y como criterio de salida un p-valor igual a 0.10. Las variables seleccionadas son (por orden de selección) la calidad del producto, la flexibilidad de precios y la velocidad de entrega no siendo eliminada del conjunto de discriminación, ninguna de las variables seleccionadas.

Estadísticos por pasos

Variables introducidas/eliminadas[a,b,c,d]

Paso	Introducidas	Lambda de Wilks							
						F exacta			
		Estadístico	gl1	gl2	gl3	Estadístico	gl1	gl2	Sig.
1	Calidad del Producto	,532	1	1	98,000	86,200	1	98,000	,000
2	Flexibilidad de Precios	,388	2	1	98,000	76,552	2	97,000	,000
3	Velocidad de Entrega	,341	3	1	98,000	61,879	3	96,000	,000

En cada paso se introduce la variable que minimiza la lambda de Wilks global.

a. El número máximo de pasos es 14.

b. La significación máxima de F para entrar es .05.

c. La significación mínima de F para salir es .10.

d. El nivel de F, la tolerancia o el VIN son insuficientes para continuar los cálculos.

Variables en el análisis

Paso		Tolerancia	Sig. de F que eliminar	Lambda de Wilks
1	Calidad del Producto	1,000	,000	
2	Calidad del Producto	1,000	,000	,583
	Flexibilidad de Precios	1,000	,000	,532
3	Calidad del Producto	,992	,000	,460
	Flexibilidad de Precios	,970	,000	,414
	Velocidad de Entrega	,963	,000	,388

Variables no incluidas en el análisis

Paso		Tolerancia	Tolerancia mín.	Sig. de F que introducir	Lambda de Wilks
0	Velocidad de Entrega	1,000	1,000	,000	,602
	Nivel de Precios	1,000	1,000	,000	,816
	Flexibilidad de Precios	1,000	1,000	,000	,583
	Imagen de la Empresa	1,000	1,000	,779	,999
	Servicio	1,000	1,000	,040	,958
	Imagen de las Ventas	1,000	1,000	,674	,998
	Calidad del Producto	1,000	1,000	,000	,532
1	Velocidad de Entrega	,992	,992	,000	,414
	Nivel de Precios	,928	,928	,099	,517
	Flexibilidad de Precios	1,000	1,000	,000	,388
	Imagen de la Empresa	,948	,948	,172	,522
	Servicio	,977	,977	,012	,498
	Imagen de las Ventas	,920	,920	,023	,504
2	Velocidad de Entrega	,963	,963	,000	,341
	Nivel de Precios	,836	,836	,842	,388
	Imagen de la Empresa	,936	,936	,085	,376
	Servicio	,969	,969	,009	,361
	Imagen de las Ventas	,914	,914	,019	,366
3	Nivel de Precios	,835	,835	,758	,341
	Imagen de la Empresa	,920	,920	,241	,336
	Servicio	,550	,547	,742	,340
	Imagen de las Ventas	,903	,903	,066	,329

Las tablas subsiguientes muestran los resultados obtenidos utilizando las variables seleccionadas. Se observa que los resultados obtenidos son esencialmente los mismos que los obtenidos utilizando todas las variables.

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	1,934[a]	100,0	100,0	,812

a. Se han empleado las 1 primeras funciones discriminantes canónicas en el análisis.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	,341	103,860	3	,000

Coeficientes estandarizados de las funciones discriminantes canónicas

	Función
	1
Velocidad de Entrega	,437
Flexibilidad de Precios	,526
Calidad del Producto	-,629

Matriz de estructura

	Función
	1
Calidad del Producto	-,674
Flexibilidad de Precios	,609
Velocidad de Entrega	,584
Nivel de Precios[a]	-,378
Imagen de las Ventas[a]	-,193
Imagen de la Empresa[a]	-,163
Servicio[a]	,120

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas
Variables ordenadas por el tamaño de la correlación con la función.

a. Esta variable no se emplea en el análisis.

6.4.2 Inconvenientes de los procedimientos de selección de variables

Conviene destacar los siguientes (ver Huberty -1989- para más detalles):

1) No tienen por qué llegar a la solución óptima
2) Utilizan como criterios de selección, criterios de separación de grupos y no de clasificación
3) El nivel de significación global es superior al establecido para entrar y sacar variables debido a la realización simultánea de varios test de hipótesis.

6.5. PROCEDIMIENTOS DE CLASIFICACIÓN

Existen varios métodos de clasificación dependiendo del número de grupos a clasificar (dos o más grupos), de las hipótesis hechas acerca del comportamiento de las variables en cada grupo (normalidad conjunta, homocedasticidad) así como del criterio utilizado para llevar a cabo dicha clasificación.

Uno de los criterios más utilizados es el *criterio Bayes* que es el que expondremos, a continuación, distinguiendo entre el caso de dos y más de dos grupos, si la discriminación se lleva a cabo bajo hipótesis de normalidad o no normalidad y/o bajo hipótesis de homo y heterocedasticidad.

6.5.1 Discriminación de dos poblaciones normales homocedásticas

Suponga que $\mathbf{Y} = \mathbf{Y}_i \sim \mathbf{N}_p(\mu_i, \Sigma)$, i=1,2 en cada uno de los grupos.

Sea **y** el valor de las variables de clasificación de una nueva observación cuya pertenencia a uno de los dos grupos se desconoce.

El **criterio Bayes** utiliza el teorema de Bayes para determinar a qué grupo pertenece.

Para ello considera $\{\pi_i = P[G_i], i = 1,2\}$ las probabilidades a priori de que la observación considerada pertenezca a cada grupo. Se suelen tomar $\pi_i = 0.5$, i = 1,2 si no se dispone de información previa o $\pi_i = {}^{n_i}\!/_{n}$, i=1,2 si los tamaños muestrales de cada grupo reflejan la composición de la población analizada.

Aplicando el teorema de Bayes se tiene que:

$$P\left[G_i \mid \mathbf{y}\right] = \frac{\pi_i f_i(\mathbf{y})}{\pi_1 f_1(\mathbf{y}) + \pi_2 f_2(\mathbf{y})} \qquad \text{i=1,2}$$

donde $\mathbf{f_i}(\mathbf{y}) \propto \exp[-0.5(y-\mu_i)'\Sigma^{-1}(y-\mu_i)]$, i=1,2 son las funciones de densidad de **Y** en cada uno de los grupos.

La observación **y** se asignará al grupo G_1 si:

$$P[G_1 \mid \mathbf{y}] > P[G_2 \mid \mathbf{y}] \Leftrightarrow \pi_1 f_1(\mathbf{y}) > \pi_2 f_2(\mathbf{y})$$

$$\Leftrightarrow (\mathbf{y}-\mu_1)'\Sigma^{-1}(\mathbf{y}-\mu_1) < (\mathbf{y}-\mu_2)'\Sigma^{-1}(\mathbf{y}-\mu_2) - \log \frac{\pi_2}{\pi_1}$$

$$\Leftrightarrow \mathbf{y}'\Sigma^{-1}(\mu_2-\mu_1) < 0.5(\mu_1+\mu_2)'\Sigma^{-1}(\mu_2-\mu_1) - \log \frac{\pi_2}{\pi_1}$$

Observaciones

1) Si $\pi_1 = \pi_2$ el criterio Bayes asignará la observación **y** al grupo cuya media, μ_i, esté a menor distancia de Mahalanobis, la cual viene dada por $d(\mathbf{y},\mu_i) = (\mathbf{y}-\mu_i)'\Sigma^{-1}(\mathbf{y}-\mu_i)$. Esta distancia también se utiliza para examinar la existencia de atípicos.

2) Para evaluar si un punto es atípico, se calcula el p-valor dado por:

$$P\left[\chi^2_{k-1} \geq D_{obs}\right]$$

donde se utiliza el hecho de que, bajo hipótesis de normalidad, $D_{obs}=\left(d-\bar{d}_i\right)' S_D^{-1}(d-\bar{d}_i) \sim \chi^2_{k-1}$ donde $d = (d_1,...,d_k)'$ son las puntuaciones en las k funciones discriminantes de cada individuo y S_D es su matriz de varianzas-covarianzas.

3) El criterio Bayes utiliza como función de clasificación, la función lineal dada por $\mathbf{y}'\Sigma^{-1}(\mu_2-\mu_1)$ y establece como punto de corte entre los dos grupos

$$\mathbf{0.5}\ (\mu_1+\mu_2)'\Sigma^{-1}(\mu_2-\mu_1) - \log \frac{\pi_2}{\pi_1}$$

4) Geométricamente, el espacio p-dimensional de los objetos queda dividido en dos regiones separadas por el hiperplano $\mathbf{y'\Sigma^{-1}(\mu_2-\mu_1) = 0.5(\mu_1+\mu_2)'\Sigma^{-1}(\mu_2-\mu_1)} - \log\frac{\pi_2}{\pi_1}$

5) Si existe un coste asociado diferente a la asignación incorrecta a cada uno de los grupos, de forma que la matriz de pérdidas viene dada por:

Asignado\Verdadero	G_1	G_2
G_1	0	c_{12}
G_2	c_{21}	0

se calculan las pérdidas esperadas medias a posteriori:

$$L(\text{Asignar a } G_1/\mathbf{y}) = c_{12}P[G_2 | \mathbf{y}]$$

$$L(\text{Asignar a } G_2/\mathbf{y}) = c_{21}P[G_1 | \mathbf{y}]$$

y se asigna la observación $\mathbf{y}$ al grupo G_1 si:

$$L(\text{Asignar a } G_1/\mathbf{y}) < L(\text{Asignar a } G_2/\mathbf{y})$$

$$\Leftrightarrow \mathbf{y'\Sigma^{-1}(\mu_2-\mu_1) < 0.5(\mu_1+\mu_2)'\Sigma^{-1}(\mu_2-\mu_1)} - \log\frac{c_{21}\pi_2}{c_{12}\pi_1}$$

6.5.2 Discriminación de dos poblaciones normales heterocedásticas

Si $\mathbf{Y}=\mathbf{Y}_i \sim \mathbf{N}_p(\mu_i,\Sigma_i)$ i=1,2 en cada uno de los grupos con $\Sigma_1 \neq \Sigma_2$ entonces las funciones de densidad de **Y** vendrán dadas por:

$$f_i(\mathbf{y}) \propto |\Sigma_i|^{-1/2} \exp[-0.5(\mathbf{y}-\mu_i)'\Sigma_i^{-1}(\mathbf{y}-\mu_i)] \qquad i=1,2$$

y se tendrá que:

$$P[G_1/\mathbf{y}] > P[G_2/\mathbf{y}]$$

$$\Leftrightarrow (\mathbf{y}-\mu_1)'\Sigma_1^{-1}(\mathbf{y}-\mu_1) - (\mathbf{y}-\mu_2)'\Sigma_2^{-1}(\mathbf{y}-\mu_2) < \log \frac{|\Sigma_2|\pi_1}{|\Sigma_1|\pi_2}$$

La función discriminante será, por lo tanto, la forma cuadrática

$$(\mathbf{y}-\mu_1)'\Sigma_1^{-1}(\mathbf{y}-\mu_1) - (\mathbf{y}-\mu_2)'\Sigma_2^{-1}(\mathbf{y}-\mu_2)$$

en lugar de ser una función lineal como en el caso anterior.

Conviene hacer notar, sin embargo, que el criterio lineal especificado anteriormente es más robusto que el criterio cuadrático a la hipótesis de normalidad y es el que se suele utilizar habitualmente.

6.5.3 Discriminación de q grupos

Los criterios vistos con dos grupos se generalizan a más de dos grupos de forma trivial. Así, por ejemplo, suponga que $\mathbf{Y} \sim \mathbf{N}_p(\boldsymbol{\mu}_i, \Sigma)$ i=1,…,q .

Las funciones de densidad de **Y** vendrán dadas por:

$$f_i(\mathbf{y}) \propto \exp[-0.5(\mathbf{y}-\boldsymbol{\mu}_i)'\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu}_i)] \qquad i=1,\ldots,q$$

El criterio Bayes clasifica la observación **y** en el grupo g si:

$$P[G_g/\mathbf{y}] = \max_{k=1,\ldots,q} P[G_k/\mathbf{y}]$$

$$\Leftrightarrow \quad \mathbf{y}'\Sigma^{-1}\boldsymbol{\mu}_g - 0.5\boldsymbol{\mu}_g\Sigma^{-1}\boldsymbol{\mu}_g + \log \pi_g = \max_{k=1,\ldots,q} \{ \mathbf{y}'\Sigma^{-1}\boldsymbol{\mu}_k - 0.5\boldsymbol{\mu}_k\Sigma^{-1}\boldsymbol{\mu}_k + \log \pi_k \}$$

Las funciones de clasificación son lineales y vienen dados por:

$$\mathbf{y}'\Sigma^{-1}\boldsymbol{\mu}_g - 0.5\boldsymbol{\mu}_g\Sigma^{-1}\boldsymbol{\mu}_g + \log \pi_g \qquad g = 1,\ldots,q$$

Ejemplo 1 (continuación)

La siguiente tabla contiene los resultados obtenidos, para 10 clientes de la empresa, en el proceso de clasificación aplicando el criterio Bayes con probabilidades a priori iguales para cada grupo y bajo hipótesis de homocedasticidad y normalidad. SPSS (como muchos otros programas) calcula las probabilidades a posteriori de cada grupo para cada caso, así como la distancia de Mahalanobis.

Así, por ejemplo, para el caso 1, perteneciente al grupo 1, el grupo pronosticado utilizando todos los casos del análisis es el grupo 1 debido a que P(G=1/D=d) = 0.934 frente a P(G=2/D=d) = 0.066.

La distancia de Mahalanobis al centroide de este grupo es igual a 0.279 y el p-valor $P\left[\chi_1^2 \geq 0.279\right]$ =0.597 por lo que dicho caso no es sospechoso de ser atípico.

Estadísticos de clasificación

Estadísticos por casos

	Número de casos	Grupo real	Grupo mayor					Segundo grupo mayor			Puntuaciones discriminantes
			Grupo pronosticado	P(D>d | G=g)		P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Grupo	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Función 1
				p	gl						
Original	1	1	1	,597	1	,934	,279	2	,066	5,580	,628
	2	1	1	,746	1	,994	,105	2	,006	10,330	1,480
	3	1	1	,872	1	,990	,026	2	,010	9,309	1,317
	4	1	1	,706	1	,956	,142	2	,044	6,314	,779
	5	1	1	,230	1	,669	1,444	2	,331	2,852	-,045
	6	1	1	,815	1	,971	,055	2	,029	7,057	,922
	7	1	1	,772	1	,993	,084	2	,007	10,109	1,445
	8	1	1	,557	1	,997	,344	2	,003	12,088	1,743
	9	1	1	,625	1	,941	,239	2	,059	5,767	,667
	10	1	1	,336	1	,999	,925	2	,001	14,837	2,118
Validación cruzada[a]	1	1	1	,355	7	,916	7,754	2	,084	12,533	
	2	1	1	,266	7	,993	8,818	2	,007	18,822	
	3	1	1	,516	7	,989	6,201	2	,011	15,245	
	4	1	1	,736	7	,950	4,371	2	,050	10,252	
	5	1	1	,639	7	,622	5,176	2	,378	6,168	
	6	1	1	,787	7	,967	3,935	2	,033	10,688	
	7	1	1	,266	7	,993	8,819	2	,007	18,606	
	8	1	1	,116	7	,997	11,559	2	,003	23,228	
	9	1	1	,525	7	,928	6,124	2	,072	11,241	
	10	1	1	,582	7	,999	5,640	2	,001	19,622	

Para los datos originales, la distancia de Mahalanobis al cuadrado se basa en funciones canónicas.
Para los datos validados mediante validación cruzada, la distancia de Mahalanobis al cuadrado se basa en observaciones.

a. La validación cruzada sólo se aplica a los casos del análisis. En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.

6.5.4 Homocedasticidad

La homocedasticidad es una hipótesis que se utiliza en algunas de las técnicas multivariantes (ANOVA, MANOVA, Análisis Discriminante) y se refiere a suponer la igualdad de las matrices de varianzas y covarianzas de las variables analizadas en diversos grupos.

El propósito de los test de homocedasticidad es contrastar la existencia de esta igualdad que, en muchas ocasiones, va ligada a una falta de normalidad de las variables analizadas. Para ello se suele utilizar el test M de Box. Este test toma como hipótesis nula la de homocedasticidad y como alternativa la de heterocedasticidad, es decir:

$$H_o: \Sigma_1 = \ldots = \Sigma_G \quad \text{vs.} \quad H_1: \text{No todas } \Sigma_g \text{ son iguales}$$

El estadístico del test está construido a partir del estadístico: $M = \dfrac{\prod_{g=1}^{q} |S_g|^{\frac{n_g - 1}{2}}}{|\overline{S}|^{\frac{n-q}{2}}}$

donde $S_g = \dfrac{W_g}{n_g - 1}$;g = 1,…,q y $\overline{S} = \dfrac{\sum_{i=1}^{q} W_i}{n - q}$

Observaciones

1) La hipótesis de normalidad es necesaria para los test de significación de las funciones discriminantes. El efecto de la falta de normalidad sobre la regla de clasificación es menor. Si no hay normalidad se aconseja utilizar otros procedimientos como, por ejemplo, la regresión logística

2) La hipótesis de homocedasticidad afecta a la validez de los test de significación y de la regla de clasificación. La violación de esta hipótesis puede producir graves desajustes si hay diferencias grandes entre el tamaño de los grupos y si el número de variables es elevado.
   Si hay normalidad conviene utilizar la regla de clasificación cuadrática especialmente si el tamaño muestral es grande

3) Una posible solución a los problemas de la falta de normalidad y homocedasticidad es llevar a cabo transformaciones de las variables. Las transformaciones más utilizadas son las de Box-Cox que vienen dadas por (X+C)p con C, p constantes reales $p \neq 0$ o log(X+C) si p=0. En general si la distribución es muy asimétrica hacia la derecha se pueden intentar transformaciones del tipo anterior con $p < 1$ (las más utilizadas son con $p = 0.5$ y la transformación logarítmica). Si lo es hacia la izquierda se aplica la transformación a - X. Si la distribución de los datos es muy leptocúrtica (curtosis muy grande) se suelen utilizar valores de $p < 0$ (el más utilizado es $p = -1$). Si es platicúrtica entonces conviene utilizar valores de $p > 1$.

4) Una forma empirica de determinar el valor de p más apropiado son los **gráficos nivel-dispersión (Spread-versus-level plot).** Dichos gráficos representan en abscisas un estimador robusto del logaritmo del nivel medio por grupos (en SPSS el logaritmo de la mediana) y en ordenadas un estimador robusto de la dispersión (en SPSS el logaritmo del rango intercuartílico) y estiman el coeficiente de regresión β mediante regresión lineal. A partir de β es posible deducir cuál es el valor de p más apropiado.

6.5.5 Discriminación no paramétrica

Si no hay normalidad conjunta existen varias opciones posibles:

- Transformar las variables para conseguir normalidad
- Llevar a cabo el análisis con los rangos
- Utilizar estimadores no paramétricos de $f_i(\mathbf{y})$

Si algunas de las variables clasificadoras no son cuantitativas, se intenta transformarlas a cuantitativas. La forma de llevar a cabo este paso depende del tipo de variable:

- Las variables binarias se transforman a 0-1
- Las variables ordinales se transforman en rangos
- Las variables nominales utilizan transformaciones basadas en sus distribuciones de frecuencias como, por ejemplo, la de Lancaster-Fisher descrita en Huberty (1994), Capítulo 10.

6.6.- EVALUACIÓN DEL PROCEDIMIENTO DE CLASIFICACIÓN

Se evalúan tres aspectos del mismo: su eficiencia, su significación estadística y su significación práctica

6.6.1 Evaluación de la eficiencia

Para evaluar su eficiencia se construye la *tabla de confusión* que es una tabla de frecuencias cruzadas que refleja los resultados de aplicar dicho procedimiento a los casos observados. Así, en el caso de la discriminación de dos grupos dicha tabla sería de la forma:

		Grupo Predicho	
		1	2
Grupo Real	1	n_{11}	n_{12}
	2	n_{21}	n_{22}

donde n_{ij} es el número de casos pertenecientes al grupo i y para los cuales el mecanismo de clasificación ha predicho que pertenecen al grupo j.

La proporción de bien clasificados vendrá dada por $100\left(\frac{n_{11}+n_{22}}{n}\right)\%$

El proceso de evaluación se puede llevar a cabo de varias formas. Tres de las más utilizadas son las siguientes:

- Con los casos utilizados en el análisis
- Dividiendo la muestra en dos partes: una para estimar las funciones discriminantes y otra para evaluarla
- Utilizando, para cada caso, las funciones discriminantes estimadas mediante el resto de los casos

El primer procedimiento no es muy aconsejable puesto que tiende a sobrevalorar el proceso de clasificación. Suele funcionar bien si $\min_g n_g > 5p$. El segundo procedimiento es aconsejable si n es suficientemente grande y funciona bien si $\min_g n_g > 3p$ tomando en torno a un 35% de la muestra para validar. En el resto de los casos se aconseja el tercer procedimiento. Otros procedimientos para evaluar el mecanismo de predicción pueden verse en Huberty (1994) capítulo 6.

6.6.2 Significación estadística

Se evalúa comparando los resultados obtenidos con los que se obtendrían aplicando un mecanismo aleatorio.

Los dos mecanismos más utilizados son *el criterio de aleatoriedad proporcional*, que clasifica de acuerdo a la distribución $\frac{n_g}{n}$ g=1,...,q y el de máxima aleatoriedad que clasifica todas las observaciones asignándolas al grupo de mayor tamaño.

Para comparar los resultados se utilizan estadísticos con distribución aproximadamente normal, bajo la hipótesis de que no existen diferencias.

Así, en el caso de que el criterio utilizado sea el del mecanismo aleatorio

$$Z_g = \frac{(o_g - e_g)\sqrt{n_g}}{\sqrt{e_g(n_g - e_g)}}$$

para evaluar los resultados en cada grupo y

$$Z = \frac{(o - e)\sqrt{n}}{\sqrt{e(n - e)}}$$

con para evaluar el proceso globalmente, siendo:

$o_g = n_{gg}$ número de clasificaciones correctas en el grupo g

$e_g = \frac{n_g^2}{n}$ el número esperado de dichas clasificaciones

$o = \sum_{g=1}^{G} o_g$ número de clasificaciones correctas

$e = \sum_{g=1}^{G} e_g$ el número de clasificaciones correctas esperadas

6.6.3 Significación práctica

Aún cuando un procedimiento sea significativamente mejor que un mecanismo aleatorio desde un punto estadístico, no tiene por qué ser mucho mejor desde un punto de vista práctico. Debido a esto es necesario medir el grado de mejoría de la regla propuesta con respecto a la clasificación debida al azar.

Para ello se utiliza el índice **I** cuya expresión viene dada por

$$I = \frac{\frac{o}{n} - \frac{e}{n}}{1 - \frac{e}{n}} \times 100$$

si se evalúa al proceso globalmente, o

$$I_g = \frac{\frac{o_g}{n_g} - \frac{e_g}{n_g}}{1 - \frac{e_g}{n_g}} \times 100$$

si se evalúa al proceso en el grupo g.

Este índice mide el porcentaje de reducción en el error que resultaría si se utilizara la regla propuesta por el Análisis Discriminante.

Ejemplo 1 (continuación)

La tabla subsiguiente muestra la tabla de confusión obtenida utilizando todos los casos del análisis y el procedimiento de validación cruzada. Se observa, en particular, que el procedimiento de clasificación ha funcionado correctamente en un 89% = $100\frac{51+38}{100}$ de los casos originales y un 87%= $100\frac{50+37}{100}$ si el procedimiento seguido en la evaluación de la eficiencia, ha sido el de validación cruzada.

Resultados de la clasificación[b,c]

		Tamaño de la empresa	Grupo de pertenencia pronosticado		Total
			Pequeña	Grande	
Original	Recuento	Pequeña	51	9	60
		Grande	2	38	40
	%	Pequeña	85,0	15,0	100,0
		Grande	5,0	95,0	100,0
Validación cruzada[a]	Recuento	Pequeña	50	10	60
		Grande	3	37	40
	%	Pequeña	83,3	16,7	100,0
		Grande	7,5	92,5	100,0

a. La validación cruzada sólo se aplica a los casos del análisis. En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.

b. Clasificados correctamente el 89,0% de los casos agrupados originales.

c. Clasificados correctamente el 87,0% de los casos agrupados validados mediante validación cruzada.

En la siguiente tabla se evalúa la significación estadística y la significación práctica de los resultados obtenidos comparando el procedimiento de clasificación con el mecanismo aleatorio proporcional.

Grupo	e_g	Z_g	p-valor	I_g
Pequeñas	36	3.69	0.00	41.67
Grandes	16	6.78	0.00	12.50
Global	52	7.01	0.00	27.08

Así, por ejemplo, e_1 = $60\frac{60}{100}$ = 36 es el número esperado de éxitos obtenidos en el grupo de empresas pequeñas mediante el mecanismo aleatorio proporcional y Z_1 = $\frac{(50-36)\sqrt{60}}{\sqrt{36x24}}$ =3.69 y el p-valor es $P[Z\geq 3.69]$=0.00. La significación práctica será igual a I_1 = $\frac{60-50}{60-36}100$ = 41.67 por lo que nuestro mecanismo mejora al azar en un 41.67% en las empresas pequeñas. Se observa que todos los resultados son significativos aunque la mejora práctica, en cada uno de ellos no es excesivamente alta.

Resumen

El Análisis Discriminante es una técnica estadística multivariante con una finalidad doble:

1) Un fin descriptivo consistente en analizar si existen diferencias entre una serie de grupos en los que se divide una población, con respecto a un conjunto de variables y, en caso afirmativo, averiguar a qué se deben.

2) Un fin predictivo consistente en proporcionar procedimientos sistemáticos de clasificación de nuevas observaciones de origen desconocido en algunos de los grupos considerados.

Para llevar a cabo un análisis de este tipo se deben los siguientes pasos:

1) Plantear el problema a resolver

2) Analizar si existen diferencias significativas entre los grupos

3) Establecer el número y composición de las dimensiones de discriminación entre los grupos analizados

4) Evaluar los resultados obtenidos desde un punto de vista predictivo analizando la significación estadística y práctica del procedo de discriminación

Conviene hacer notar, finalmente, que el Análisis Discriminante no es la única técnica estadística implicada en el proceso de clasificación de observaciones en grupos previamente fijados por el analista. Otra alternativa interesante viene dada por los modelos de regresión con variable dependiente cualitativa (de los que el Análisis Discriminante podría considerarse un caso particular) como son, por ejemplo, los modelos de regresión logit y probit.

Bibliografía

- Como libro de consulta dedicado al Análisis Discriminante:

  HUBERTY, C.J. (1994). *Applied Discriminant Analysis*. Wiley. Interscience

- Libros de Análisis Multivariantes que contienen buenos capítulos acerca del Análisis Discriminante:

  A) Desde un punto de vista más práctico,

  AFIFI, A.A. and CLARK, V. (1996) *Computer-Aided Multivariate Analysis*. Third Edition. Texts in Statistical Science. Chapman and Hall.

  EVERITT, B. And GRAHAM, D. (1991). *Applied Multivariate Data Analysis*. Arnold.

  HAIR, J., ANDERSON, R., TATHAM, R. y BLACK, W. (1999). *Análisis Multivariante*. 5ª Edición. Prentice Hall.

  SHARMA, S. (1998). *Applied Multivariate Techiques*. John Wiley and Sons.

  URIEL, E. (1995). *Análisis de Datos: Series temporales y Análisis Multivariante*. Colección Plan Nuevo. Editorial AC.

  B) Desde un punto de vista más matemático:

  JOBSON, J.D. (1992) *Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods*. Springer-Verlag.

  MARDIA, K.V., KENT, J.T. y BIBBY, J.M. (1994). *Multivariate Analysis*. Academic Press.

- Enfocados hacia SPSS:

  FERRAN, M.(1997). *SPSS para WINDOWS. Programación y Análisis Estadístico*. Mc.Graw Hill.

  VISAUTA, B. (1998) *Análisis Estadístico con SPSS para WINDOWS (Vol II. Análisis Multivariante)*. Mc-Graw Hill.