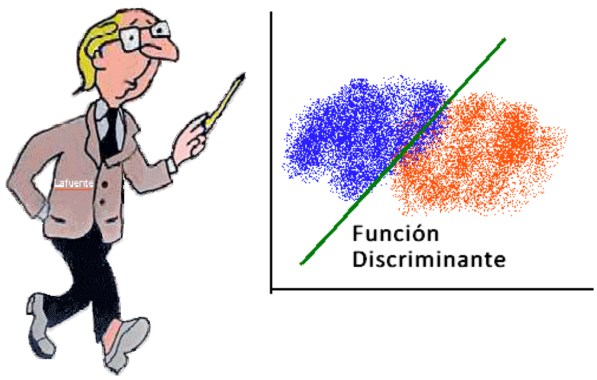
Instrumentos Estadísticos Avanzados
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández

ANÁLISIS DISCRIMINANTE



ANÁLISIS DISCRIMINANTE

El Análisis Discriminante es una técnica estadística multivariante cuya finalidad es analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos para, en el caso de que existan, explicar en qué sentido se dan y facilitar procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados.

- ¿Se puede predecir si una empresa va a entrar en bancarrota?
- ¿Es posible predecir con antelación si un cliente que solicita un préstamo a un banco va a ser un cliente moroso?
- ¿Existe discriminación por razones de sexo o de raza en una empresa o en un colegio?

El Análisis Discriminante se puede considerar como un análisis de regresión donde la variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, mientras que las variables independientes son continuas y determinan a qué grupos pertenecen los objetos.

- *Se pretende encontrar relaciones lineales entre las variables continuas que mejor discriminen en los grupos dados a los objetos.*

- *Construir una regla de decisión que asigne un objeto nuevo con un cierto grado de riesgo, cuya clasificación previa se desconoce, a uno de los grupos prefijados*.

Para efectuar el análisis es necesario considerar una serie de supuestos:

(a) Se tiene una variable categórica y el resto de variables son de intervalo o de razón y son independientes respecto de ella.

(b) Se necesitan al menos dos grupos, y para cada grupo se necesitan dos o más casos.

(c) El número de variables discriminantes debe ser menor que el número de objetos menos 2, es decir, $(x_1, x_2, \cdots, x_p)$ donde $p < (n-2)$ siendo $n \equiv$ número de objetos.

(d) Ninguna variable discriminante puede ser combinación lineal de otras variables discriminantes.

(e) El número máximo de funciones discriminantes es el mínimo [número de variables, número de grupos menos 1] – con q grupos, $(q-1)$ funciones discriminantes -.

(f) Las matrices de covarianzas dentro de cada grupo deben de ser aproximadamente iguales.

(g) Las variables continuas deben seguir una distribución normal multivariante.

MODELO MATEMÁTICO

Partiendo de **q** grupos donde se asignan a una serie de objetos y de **p** variables medidas sobre ellos $(x_1, x_2, \cdots, x_p)$, se trata de obtener para *cada objeto* una *serie de puntuaciones* que indican el grupo al que pertenecen $(y_1, y_2, \cdots, y_m)$, de modo que sean funciones lineales de $(x_1, x_2, \cdots, x_p)$:

$$\begin{cases} y_1 = w_{11}\,x_1 + w_{12}\,x_2 + \cdots + w_{1p}\,x_p + w_{10} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_m = w_{m1}\,x_1 + w_{m2}\,x_2 + \cdots + w_{mp}\,x_p + w_{10} \end{cases} \qquad m = \text{mín}\left[q-1, p\right]$$

tales que discriminen o separen lo máximo posible a los **q** grupos.

Estas combinaciones lineales de las **p** variables deben maximizar la varianza *entre* los grupos y minimizar la varianza *dentro* de los grupos.

DESCOMPOSICIÓN DE LA VARIANZA: La variabilidad total de la muestra se puede descomponer en variabilidad *dentro* de los grupos y *entre* los grupos. Para ello, se parte:

$$\text{Cov}(x_j, x_{j'}) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)\,(x_{ij'} - \overline{x}_{j'})$$

se puede considerar la media de la variable x_j en cada uno de los grupos $(I_1, I_2, \cdots, I_q)$, es decir, $\overline{x}_{kj} = \dfrac{1}{n_k}\displaystyle\sum_{i\in I_k} x_{ij}$ para $k = 1, \cdots, q$.

De esta forma, la media total de la variable x_j se puede expresar como función de las medias dentro de cada grupo: $n_k \overline{x}_{kj} = \displaystyle\sum_{i\in I_k} x_{ij}$

con lo cual, $\overline{x}_j = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_{ij} = \dfrac{1}{n}\sum_{k=1}^{q}\sum_{i\in I_k} x_{ij} = \dfrac{1}{n}\sum_{k=1}^{q} n_k \overline{x}_{kj} = \sum_{k=1}^{q}\dfrac{n_k}{n}\,\overline{x}_{kj}$

Así, $\text{Cov}(x_j, x_{j'}) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{q}\sum_{i\in I_k}(x_{ij} - \overline{x}_j)\,(x_{ij'} - \overline{x}_{j'})$

Poniendo en cada uno de los términos: $\begin{cases} (x_{ij} - \overline{x}_j) = (x_{ij} - \overline{x}_{kj}) + (\overline{x}_{kj} - \overline{x}_j) \\ \\ (x_{ij'} - \overline{x}_{j'}) = (x_{ij'} - \overline{x}_{kj'}) + (\overline{x}_{kj'} - \overline{x}_{j'}) \end{cases}$ se obtiene,

$$\text{Cov}(x_j, x_{j'}) = \frac{1}{n}\sum_{k=1}^{q}\sum_{i\in I_k}(x_{ij} - \overline{x}_j)\,(x_{ij'} - \overline{x}_{j'}) = \frac{1}{n}\sum_{k=1}^{q}\sum_{i\in I_k}(x_{ij} - \overline{x}_{kj})\,(x_{ij'} - \overline{x}_{kj'}) + \sum_{k=1}^{q}\frac{n_k}{n}\,(\overline{x}_{kj} - \overline{x}_j)\,(\overline{x}_{kj'} - \overline{x}_{j'}) =$$

$$= v(x_j, x_{j'}) + f(x_j, x_{j'}) \quad \mapsto \quad \overbrace{t(x_j, x_{j'})}^{\text{covarianza total}} = \overbrace{v(x_j, x_{j'})}^{\text{covarianza dentro grupos}} + \overbrace{f(x_j, x_{j'})}^{\text{covarianza entre grupos}} \quad \Rightarrow \quad \overbrace{T = V + F}^{\text{MATRICIALMENTE}}$$

La covarianza *total* es igual a la covarianza *dentro* de los grupos más la covarianza *entre* grupos.

EXTRACCIÓN FUNCIONES DISCRIMINANTES

La idea básica del Análisis Discriminante consiste en extraer a partir de $(x_1, x_2, \cdots, x_p)$ variables observadas en **k** grupos, **m** funciones $(y_1, y_2, \cdots, y_m)$ de forma que:

$y_i = w_{i1} x_1 + w_{i2} x_2 + \cdots + w_{ip} x_p + w_{i0}$ donde $m = \text{mín}(q-1, p)$, tales que $\text{corre}(y_i, y_j) = 0 \;\; \forall i \neq j$

Si las variables $(x_1, x_2, \cdots, x_p)$ están tipificadas, las funciones $\left(y_i = w_{i1} x_1 + w_{i2} x_2 + \cdots + w_{ip} x_p\right)$ para $(i = 1, \cdots, m)$ se denominan *discriminantes canónicas*.

Las funciones $(y_1, y_2, \cdots, y_m)$ se extraen de modo que:

- y_1 sea la combinación lineal de $(x_1, x_2, \cdots, x_p)$ que proporciona la mayor discriminación posible entre los grupos.

- y_2 sea la combinación lineal de $(x_1, x_2, \cdots, x_p)$ que proporciona la mayor discriminación posible entre los grupos, después de y_1, tal que $\text{corre}(y_1, y_2) = 0$

- En general, y_i es la combinación lineal de $(x_1, x_2, \cdots, x_p)$ que proporciona la mayor discriminación posible entre los grupos, después de y_{i-1}, tal que $\text{corre}(y_i, y_j) = 0$ para $j = 1, \cdots, (i-1)$

MATRICIALMENTE: Se busca una función lineal de $(x_1, x_2, \cdots, x_p)$: $Y = w'X$

Se sabe que La covarianza *total* es igual a la covarianza *dentro* de los grupos más la covarianza *entre* grupos: $\overbrace{T = F + V}^{\text{MATRICIALMENTE}}$.

De modo que, $\text{Var}(y) = w'Tw = w'Fw + w'Vw$

☞ Se maximiza la variabilidad *entre* los grupos para discriminarlos mejor, es decir, se maximiza la varianza *entre* grupos en relación con el total de la varianza: $\text{máx}\left[\dfrac{w'Fw}{w'Tw}\right]$

Considerando la función $f(w) = \dfrac{w'Fw}{w'Tw}$ se observa que es una función *homogénea*, es decir, $f(w) = f(\mu w) \;\; \forall \mu \in R$. El hecho de que sea *homogénea* implica que calcular $\text{máx}\left[\dfrac{w'Fw}{w'Tw}\right]$ equivale a calcular $\text{máx}\left[w'Fw\right]$ tal que $w'Tw = 1$

Como es el esquema habitual de los multiplicadores de Lagrange, se define:

$$L = w'Fw - \lambda\left(w'Tw - 1\right) \;\; \Rightarrow \;\; \frac{\vartheta L}{\vartheta w} = 2Fw - 2\lambda Tw = 0 \;\; \Rightarrow \;\; Fw = \lambda Tw \;\; \Rightarrow \;\; (T^{-1}F)w = \lambda w$$

En consecuencia, el autovector asociado a la primera función discriminante lo es de la matriz $(T^{-1}F)$, que en general no es simétrica.

Como $Fw = \lambda T w$, se tiene $w'Fw = \lambda w'Tw = \lambda$

Por tanto, tomando el vector asociado al máximo autovalor se obtendrá la función que recoge el *máximo poder discriminante*.

El **autovalor asociado** a la función discriminante indica la proporción de varianza total explicada por las **m** funciones discriminantes que recoge la variable y_i

- Para obtener más funciones discriminantes se siguen sacando los autovectores de la matriz $(T^{-1}F)$ asociados a los autovalores elegidos en orden decreciente: $\begin{cases} w_2' \;\Rightarrow\; w_2'X = Y_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ w_m' \;\Rightarrow\; w_m'X = Y_m \end{cases}$

  $m = \text{mín}(q-1, p)$. Estos vectores son linealmente independientes y dan lugar a funciones incorreladas entre sí.

La suma de todos los autovalores $\sum_{i=1}^{m} \lambda_i$ es la proporción de varianza total que queda explicada, o se conserva, al considerar sólo los ejes o funciones discriminantes.

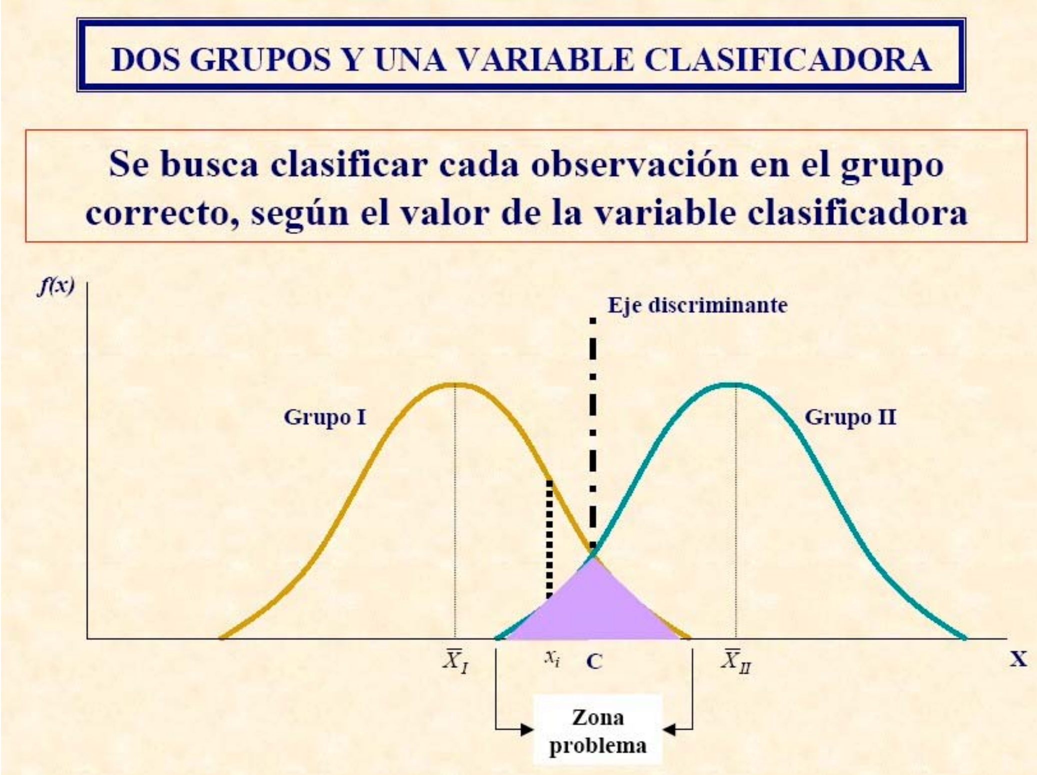
Como consecuencia, el porcentaje explicado por la variable y_i del total de varianza explicada por las funciones $(y_1, y_2, \cdots, y_m)$ es:

$$\frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i} 100\%$$

ANÁLISIS DISCRIMINANTE: OBJETO

Clasificar las observaciones de la muestra en grupos, a partir de la información suministrada por un conjunto de variables.

Un conjunto de variables explicativas o criterio	Una variable categórica señalando los grupos
Variables clasificadoras	Variable dependiente

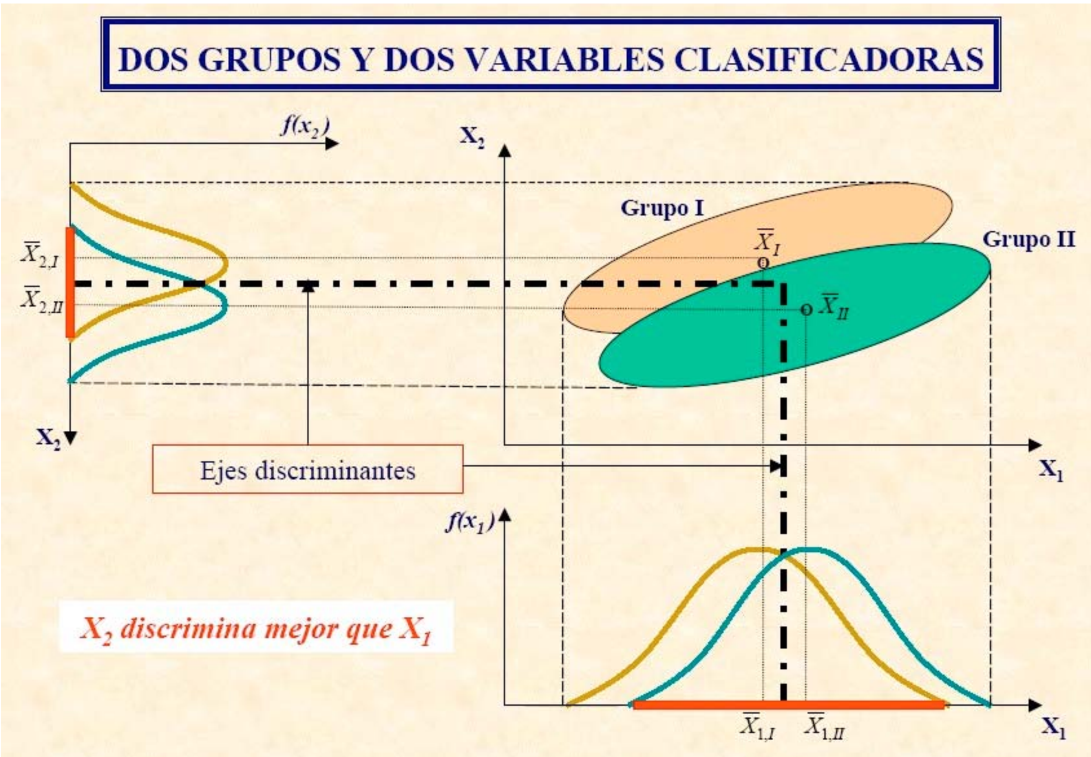


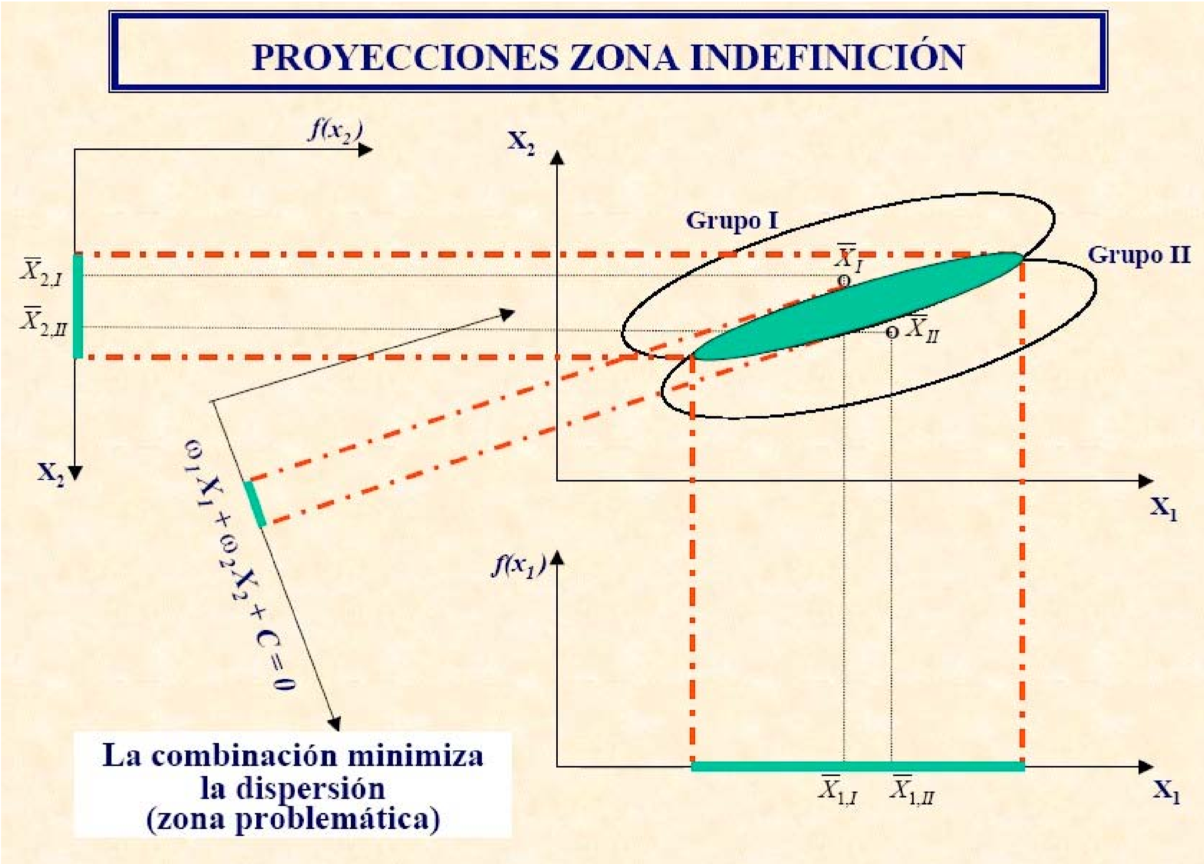
ANÁLISIS DISCRIMINANTE (A.D.): CRITERIO DE CLASIFICACIÓN

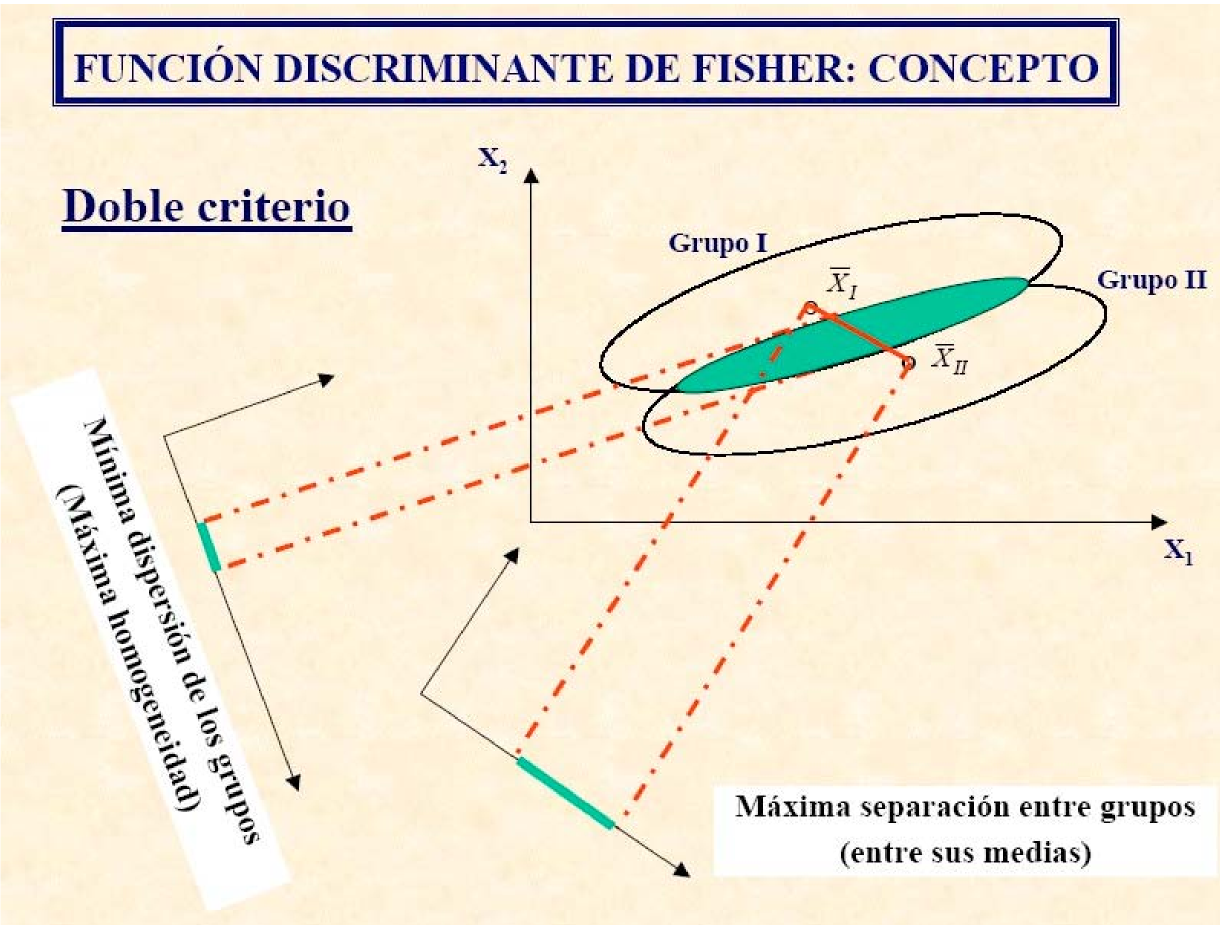
Hipótesis: Las distribuciones sólo se diferencian por su localización (igual forma y varianza)

- Se trata de minimizar los errores de clasificación
- Si $x_i < C$ se clasifica en el grupo I
- Si $x_i > C$ se clasifica en el grupo II

El punto C se denomina *punto de corte discriminante*: $C = \dfrac{\overline{X}_I + \overline{X}_{II}}{2}$




PROYECCIONES ZONA INDEFINICIÓN
$f(x_2)$
X_2
Grupo I
Grupo II
$\overline{X}_{2,I}$
$\overline{X}_{2,II}$
$\overline{X}_I$
$\overline{X}_{II}$
X_2
$\omega_1 X_1 + \omega_2 X_2 + C = 0$
X_1
$f(x_1)$
La combinación minimiza
la dispersión
(zona problemática)
$\overline{X}_{1,I}$
$\overline{X}_{1,II}$
X_1


FUNCIÓN DISCRIMINANTE DE FISHER: CONCEPTO
Doble criterio
X_2
Grupo I
Grupo II
$\overline{X}_I$
$\overline{X}_{II}$
X_1
Mínima dispersión de los grupos
(Máxima homogeneidad)
Máxima separación entre grupos
(entre sus medias)

ENFOQUES DE ANÁLISIS

- Basado en la obtención de funciones discriminantes de cálculo similar a las ecuaciones de regresión lineal múltiple. Consiste en conseguir, a partir de las variables explicativas, unas funciones lineales de éstas con capacidad para clasificar a otros individuos. A cada nuevo caso se aplican dichas ecuaciones y la función de mayor valor define el grupo al que pertenece.

- Basado en técnicas de correlación canónica y de componentes principales (Análisis Factorial) denominado Análisis Discriminante Canónico.

CLASIFICACIÓN EN DOS GRUPOS

- Se estudia la aplicación del Análisis Discriminante (AD) a la clasificación de individuos en el caso de que se puedan asignar solamente a dos grupos a partir de k variables discriminadoras.

- Fisher resuelve el problema mediante su función discriminante: $D = w_1 X_1 + w_2 X_2 + \cdots + w_k X_k$

- Las puntuaciones discriminantes son los valores que se obtienen al dar valores a $(X_1, X_2, \cdots, X_k)$ en la ecuación anterior.

- Se trata de obtener los coeficientes de ponderación w_j

- Si se consideran N observaciones $\rightarrow$ La función discriminante $D_i = w_1 X_{1i} + w_2 X_{2i} + \cdots + w_k X_{ki}$ para $\forall\ i = 1, \cdots, N$.

  (D_i) es la puntuación discriminante correspondiente a la observación i-ésima.

- La función discriminante en forma matricial: $\begin{pmatrix} D_1 \\ D_2 \\ \vdots \\ D_N \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \\ X_{1N} & X_{2N} & \cdots & X_{kN} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{pmatrix}$

- Expresando el modelo en función de las desviaciones a la media, resulta:

$$\begin{pmatrix} D_1 - \overline{d}_1 \\ D_2 - \overline{d}_2 \\ \vdots \\ D_N - \overline{d}_N \end{pmatrix} = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \\ X_{1N} & X_{2N} & \cdots & X_{kN} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{pmatrix} \quad \text{es decir,}$$

$$\boxed{d = X\,w \quad \text{(función discriminante en diferencias)}}$$

- La variabilidad de la función discriminante (suma de cuadrados de las desviaciones de las variables discriminantes con respecto a su media) se expresa:

  Suma de cuadrados explicada por esta función: $\boxed{d'\,d = w'\,X'\,X\,w}$

  $X'\,X$ es una matriz simétrica que expresa las desviaciones cuadráticas con respecto a la media de las variables (*suma de cuadrados total*).

Se puede descomponer en suma de cuadrados *entre* grupos F y suma de cuadrados *dentro* de los grupos V:

$T = X'\,X$ (matriz de suma de cuadrados y productos cruzados (varianzas-covarianzas) para el conjunto de observaciones.

$$\boxed{T = X'\,X = F + V}$$

con lo cual, $\boxed{d'\,d = w'\,X'\,X\,w = w'\,(F+V)\,w = w'\,F\,w + w'\,V\,w}$

- Los ejes discriminantes vienen dados por los vectores propios asociados a los valores propios de la matriz $(V^{-1}\,F)$ ordenados de mayor a menor.

- Las puntuaciones discriminantes se corresponden con los valores obtenidos al proyectar cada punto del espacio k-dimensional de las variables originales sobre el eje discriminante.

- Los coeficientes w se obtienen: $\text{Máx}\,\lambda = \dfrac{w'\,F\,w}{w'\,V\,w} = \dfrac{\text{separación entre grupos}}{\text{separación dentro grupos}}$

CLASIFICACIÓN

- Se obtienen las puntuaciones discriminantes d_i para cada observación, introduciendo los correspondientes valores de las k variables en la función discriminante.

- Se aplica el criterio de clasificación:

  $d_i < C$ $(d_i - C < 0)$ $\rightarrow$ pertenece al grupo I
  $d_i > C$ $(d_i - C > 0)$ $\rightarrow$ pertenece al grupo II

- Otro camino: funciones discriminantes para cada grupo $\rightarrow$ se clasifica la observación en el grupo en que la función correspondiente arroja mayor valor.

HIPÓTESIS

- Las variables son independientes y se distribuyen normalmente $\rightarrow$ problemas en la estimación.

- Las matrices de las varianzas y covarianzas son iguales en todos los grupos $\rightarrow$ afecta a la clasificación.

- No multicolinealidad entre las variables clasificadoras.

- Las relaciones son lineales.

- No existen valores anómalos (*outliers*).

CENTROIDES PARA CADA GRUPO (GRUPO I, GRUPO II)

$$\overline{X}_I = \begin{pmatrix} \overline{X}_{1I} \\ \overline{X}_{2I} \\ \vdots \\ \overline{X}_{kI} \end{pmatrix} \qquad \overline{X}_{II} = \begin{pmatrix} \overline{X}_{1II} \\ \overline{X}_{2II} \\ \vdots \\ \overline{X}_{kII} \end{pmatrix}$$

Los subíndices I y II indican a qué grupo pertenece la variable.

PARA CADA GRUPO

$$\begin{cases} \overline{D}_I = w_1\,\overline{X}_{1I} + w_2\,\overline{X}_{2I} + \cdots + w_k\,\overline{X}_{kI} \\ \\ \overline{D}_{II} = w_1\,\overline{X}_{1II} + w_2\,\overline{X}_{2II} + \cdots + w_k\,\overline{X}_{kII} \end{cases}$$

CRITERIO PARA CLASIFICAR A UN INDIVIDUO

- Si $D_i < C$ se clasifica al individuo i en el grupo I
- Si $D_i > C$ se clasifica al individuo i en el grupo II

C: punto de corte discriminante $C = \dfrac{\overline{D}_I + \overline{D}_{II}}{2}$

EN GENERAL:

$\left\{ D - C = w_1\,\overline{X}_1 + w_2\,\overline{X}_2 + \cdots + w_k\,\overline{X}_K - C \right\}$ se clasifica dependiendo si $(D - C)$ es positivo o negativo.

INFERENCIAS Y CÁLCULO DE PROBABILIDADES

La obtención de la función discriminante la realizó Fisher aplicando un enfoque puramente descriptivo. Cuando en el análisis discriminante se desean abordar cuestiones de carácter inferencial y otros relativos al modelo poblacional se requiere la formulación previa de hipótesis estadísticas.

Las cuestiones de tipo inferencial se refieren a diversos contrastes de significación sobre el modelo, así como contrastes utilizados en el proceso de selección de variables cuando el número de éstas es muy grande y no se conoce *a priori* las variables que son relevantes en el análisis.
Por otra parte, el cálculo de probabilidad de pertenencia a un grupo requiere que previamente se haya postulado algún modelo probabilístico de la población.

Las hipótesis estadísticas que se adoptan, análogas a las postuladas en el análisis multivariante de la varianza, se refieren tanto a la población como al proceso de obtención de la muestra.

☞ *Las hipótesis estadísticas sobre la población:*

(a) La matriz de covarianzas de todos los grupos es igual a Σ (hipótesis de homocedasticidad).

(b) Cada uno de los grupos tiene una distribución normal multivariante.

Las hipótesis implican que $x_g \approx N(\mu_g, \Sigma)$

☞ *Las hipótesis sobre el proceso de obtención de la muestra:* Facilitan la realización del proceso de inferencia a partir de la información disponible: <<Se supone que se ha extraído una muestra aleatoria multivariante independiente en cada uno de los G grupos>>.

Bajo las hipótesis citadas, la función discriminante obtenida por Fisher es óptima. La hipótesis $x_g \approx N(\mu_g, \Sigma)$ exige que las variables clasificadoras sigan una distribución normal. Sin embargo, no sería razonable postular est hipótesis respecto a variables categóricas, utilizadas frecuentemente en el análisis discriminante como variables clasificadoras. Señalar que, cuando se utilizan variables de este tipo, la función discriminante lineal de Fisher no tiene el carácter de óptima.

Contrastes de significación y evaluación de la bondad de ajuste

Con los contrastes de significación que se realizan en el análisis discriminante con dos grupos se trata de dar respuesta a tres tipos de cuestiones diferentes:

(a) ¿Se cumple la hipótesis de homocedasticidad del modelo?
(b) ¿Se cumplen las hipótesis de normalidad?
(c) ¿Difieren significativamente las medias poblacionales de los dos grupos?

Para el contraste de homocedasticidad (si la matriz de covarianzas es la misma para los distintos grupos) se utiliza el estadístico de Barlett-Box:

$$\boxed{M = \frac{\prod_{g=1}^{K} |S_g|^{(n_g-1)/2}}{|\overline{S}|^{(n-K)/2}}}$$

- En el numerador aparecen los determinantes de las estimaciones de la matriz de covarianzas para cada grupo.
- En el denominador, el determinante de la estimación global de la matriz de covarianzas.

Cuando el numerador sea muy superior al denominador, será indicativo de que existe heteroscedasticidad (no existe homogeneidad entre las matrices de covarianzas de cada grupo).

donde: $S_g = \dfrac{V_g}{n_g - 1}$ $\qquad \overline{S} = \dfrac{\sum_{g=1}^{G} V_g}{n-G} = \dfrac{\sum_{g=1}^{G}(n_g-1)S_g}{n-G}$ $\qquad K \equiv$ variables

La matriz S_g es una estimación de la matriz de covarianzas correspondiente a la celda g-ésima Σ_g, $\overline{S}$ es una estimación de la matriz de covarianzas global Σ.

☞ La respuesta a la pregunta *¿Difieren significativamente las medias poblacionales de los dos grupos?* es decisiva para la realización del análisis discriminante. En caso de que la respuesta fuera negativa carecería de interés continuar con el análisis, ya que significaría que las variables introducidas no tienen capacidad discriminante significativa.

Las hipótesis nula y alternativa para dar respuesta a la cuestión, en el caso de dos grupos $\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$

El contraste de la hipótesis se puede realizar específicamente mediante el estadístico T^2 de Hotelling:

$$T^2 = (\overline{y}_1 - \overline{y}_2)' \, \overline{S}^{-1} (\overline{y}_1 - \overline{y}_2) \left(\frac{n_1 n_2}{n_1 + n_2} \right) \quad \text{donde} \quad \overline{S} = \frac{V_1 + V_2}{n_1 + n_2 - 2}$$

La matriz $\overline{S}$ es un estimador insesgado de la matriz de covarianzas poblacional Σ, obtenido bajo el supuesto de que la matriz de covarianzas poblacional es la misma en los dos grupos.

Bajo la hipótesis nula, el estadístico T^2 de Hotelling se distribuye:

$$\left(\frac{n_1+n_2-K-1}{K}\right)\frac{T^2}{n_1+n_2-2} \approx F_{K,\,n_1+n_2-K-1}$$

Existen otros estadísticos para realizar el contraste, diseñados para el caso general de G grupos, tales como el estadístico de Rao o el estadístico V de Barlett (estos dos últimos estadísticos están construidos a partir de la Λ de Wilks).

En el caso de que se rechace la hipótesis nula $H_0: \mu_1 = \mu_2$, se puede aplicar el análisis univariante de la varianza para contrastar la hipótesis de igualdad de medias para cada una de las variables clasificadoras por separado.

Como medida de evaluación de la bondad de ajuste se utiliza el coeficiente *eta cuadrado* (η^2), que es el coeficiente de determinación obtenido al realizar la regresión entre la variable dicotómica, que indica la pertenencia al grupo, y las puntuaciones discriminantes. A la raíz cuadrado de este coeficiente se le denomina correlación canónica.

$$\eta = \sqrt{\frac{\lambda}{1+\lambda}} \quad \text{(correlación canónica)}$$

$\lambda \equiv$ ratio que se obtiene al maximizar $\text{Máx}\,\lambda = \dfrac{w_1^{'}\,F\,w_1}{w_1^{'}\,V\,w_1} = \dfrac{\text{separación entre grupos}}{\text{separación dentro grupos}}$

Cálculo de probabilidades de pertenencia a una población

Las funciones discriminantes del tipo $\begin{Bmatrix} D = w_1 X_1 + w_2 X_2 + \cdots + w_k X_k \\ D - C = w_1 X_1 + w_2 X_2 + \cdots + w_k X_k - C \end{Bmatrix}$ clasifican a los diferentes individuos en uno u otro grupo, pero no ofrecen más información acerca de los individuos investigados.

En muchas ocasiones es conveniente tener información complementaria a las puntuaciones discriminantes. Si bien con estas puntuaciones se puede clasificar a cada individuo, también es interesante disponer de información sobre la probabilidad de su pertenencia a cada grupo, pues con este dato se puede realizar análisis más matizados, e incluir otras informaciones tales como la información *a priori* o los costes que implica una información errónea.

Para realizar este tipo de cálculos se suelen asumir las hipótesis estadísticas sobre la población:

(c) La matriz de covarianzas de todos los grupos es igual a Σ (hipótesis de homocedasticidad).

(d) Cada uno de los grupos tiene una distribución normal multivariante.

Las hipótesis implican que $x_g \approx N(\mu_g, \Sigma)$, considerando además que se conocen los parámetros poblacionales.

El cálculo de probabilidades se realiza en el contexto de la teoría de la decisión, que permite tener en cuenta la probabilidad de pertenencia a un grupo, como los costes de una clasificación errónea.

La clasificación de los individuos se realiza utilizando el teorema de Bayes. La aplicación del teorema de Bayes permite el cálculo de las probabilidades *a posteriori* a partir de estas probabilidades *a priori* y de la información muestral contenida en las puntuaciones discriminantes.

En el caso general de G grupos, el teorema de Bayes establece que la probabilidad *a posteriori* de pertenencia a un grupo g con una puntuación discriminante D, con probabilidades *a priori* π_g es:

$$\boxed{\text{Prob}(g/D) = \frac{\pi_g\ \text{Prob}(D/g)}{\sum_{i=1}^{G} \pi_i\ \text{Prob}(D/i)}}$$

La probabilidad condicionada $\text{Prob}(D/g)$ se obtiene calculando la probabilidad de la puntuación observada suponiendo la pertenencia a un grupo g.

Dado que el denominador $\sum_{i=1}^{G} \pi_i\ \text{Prob}(D/i)$ es una constante, se utiliza también la forma equivalente:

$$\boxed{\text{Prob}(g/D) \propto \pi_g\ \text{Prob}(D/g)} \quad \propto \equiv \text{proporcionalidad}$$

La clasificación de cada individuo se puede realizar mediante la comparación de las probabilidades *a posteriori*. Así, se asignará un individuo al grupo para el cual sea mayor su probabilidad a *posteriori*.

Se presenta el cálculo de probabilidades en el caso de dos grupos, de forma que sea fácilmente generalizable al caso de G grupos.

El cálculo de probabilidades se realiza bajo tres supuestos diferentes: (a) Cálculo de probabilidades sin información a priori. (b) Cálculo de probabilidades con información a priori. (c) Cálculo de probabilidades con información a priori considerando los costes.

☞ ***Cálculo de probabilidades a posteriori sin información a priori***

En el cálculo de estas probabilidades se considera que no existe conocimiento previo de las probabilidades de pertenencia a cada grupo. Cuando no existe dicha información, se adopta el supuesto de que la probabilidad de pertenencia a ambos grupos es la misma, es decir, se adopta el supuesto de que $\pi_I = \pi_{II}$. Esto implica que estas probabilidades a priori no afectan a los cálculos de las probabilidades a posteriori.

Bajo las hipótesis estadísticas sobre la población, la probabilidad de pertenencia a cada grupo, dada la puntuación discriminante obtenida, viene dada por la expresión:

$$\boxed{\text{Prob}(g/D) = \frac{e^{F_g}}{e^{F_I} + e^{F_{II}}} \quad g = I, II} \quad F_I \text{ y } F_{II} \text{ son las funciones definidas.}$$

Un individuo se clasifica en el grupo para el que la probabilidad sea mayor. Este criterio implica que un individuo se clasificará en el grupo I si $\boxed{F_I > F_{II}}$

Aplicando la fórmula de probabilidad a posteriori se llega a los mismos resultados que aplicando la fórmula discriminante de Fisher. Esto implica que el punto de corte C es el mismo: $\boxed{C = \frac{\overline{D}_I + \overline{D}_{II}}{2}}$.

☞ ***Cálculo de probabilidades a posteriori con información a priori***

En ocasiones se dispone de información de la probabilidad *a priori* sobre pertenencia de un individuo a cada uno de los grupos. Por ejemplo, se puede tener información de que los préstamos fallidos suponen un 10% del total de los préstamos concedidos a lo largo de cinco años. Para tener en cuenta este tipo de información se introducen probabilidades a priori en el análisis.

Cuando se utilizan probabilidades a priori los individuos se clasifican en el grupo para el que la probabilidad a posteriori sea mayor.

$$\boxed{\text{Prob}(g/D) = \frac{\pi_I\, e^{F_g}}{\pi_I\, e^{F_I} + \pi_{II}\, e^{F_{II}}} \quad g = I, II}$$ F_I y F_{II} son las funciones definidas.

Con este criterio, un individuo se clasifica en el grupo I si: $\boxed{F_I \ln\pi_I > F_{II} \ln\pi_{II}}$.

La aplicación implica que el punto de corte discriminante C vendrá dado por la expresión:

$$\boxed{C_p = \frac{\overline{D}_I + \overline{D}_{II}}{2} - \ln\frac{\pi_{II}}{\pi_I}}$$

La ratio de probabilidades *a priori* debe establecerse de forma que el punto de corte se desplace hacia el grupo con menor probabilidad *a priori*. Al desplazarse el punto de corte de esta forma, se tenderá a clasificar una proporción menor de individuos en el grupo con menor probabilidad *a priori*.

☞ ***Cálculo de probabilidades a posteriori con información a priori y considerando costes***

Hasta ahora no se ha considerado el coste que una clasificación errónea puede tener. En muchas ocasiones el coste de clasificación errónea puede diferir para cada uno de los grupos. Por ejemplo, en la concesión de préstamos, clasificar como fallido a un cliente cumplidor y clasificar como cumplidor a un fallido, no es lo mismo para la entidad bancaria. En la primera de las posibilidades, el coste para el banco es dejar de percibir los intereses del préstamo y la posible pérdida de un cliente que en realidad es cumplidor. Por el contrario, en la segunda posibilidad el coste para el banco es la pérdida de la cantidad prestada, ya que el cliente clasificado como cumplidor es realmente fallido. En principio, y bajo el criterio de una prudente administración financiera, parece que el segundo tipo de coste es superior al primero.

Cuando se introducen costes de clasificación no puede hablarse ya de cálculo de probabilidades a *posteriori*. No obstante se puede obtener un criterio para clasificar minimizando el coste total de clasificación errónea. Este total viene dado por la expresión:

$$\boxed{\pi_I\ \text{Prob}(II/I)\ \text{Coste}(II/I) + \pi_{II}\ \text{Prob}(I/II)\ \text{Coste}(I/II)}$$

Cada probabilidad se encuentra multiplicada por el coste en que se incurre. Al minimizar la expresión, bajo las hipótesis estadísticas sobre la población, el punto de corte discriminante $C_{p,c}$ se obtiene con la expresión:

$$\boxed{C_{p,c} = \frac{\overline{D}_I + \overline{D}_{II}}{2} - \ln\frac{\pi_{II}\ \text{Coste}(I/II)}{\pi_I\ \text{Coste}(II/I)}}$$

En los desarrollos anteriores se ha supuesto que las probabilidades son conocidas. En la práctica, se utilizan estadísticos muestrales en su lugar. El empleo de estadísticos muestrales tiene como consecuencia que se subestime la probabilidad de clasificación errónea, cometiéndose por lo tanto sesgos sistemáticos en la clasificación. Para disminuir estos sesgos se han propuesto, entre otros, dos procedimientos alternativos.

- Un procedimiento consiste en dividir la muestra total en dos submuestras, utilizando la primera muestra para estimar la función discriminante, mientras que la segunda se utiliza para su validación. Así, la potencia discriminante de la función vendrá determinada por el porcentaje de individuos clasificados en esta segunda muestra.

- El segundo procedimiento consiste en excluir un individuo del grupo I, calcular la función discriminante, y clasificar después al individuo que se ha excluido. Haciendo lo mismo con el resto de individuos del grupo I, se estima la Prob(II/I) con el porcentaje de individuos que han sido clasificados en el grupo II. Procediendo análogamente con los individuos del grupo II, se estima la Prob(I/II). A este segundo procedimiento se le conoce con la denominación *jacknife*.

Se adjunta una tabla resumen del Ejercicio 1, donde se acompaña las puntuaciones discriminantes para los 16 clientes.

Cliente	Grupo pertenencia	Patrimonio Neto	Deuda Pendiente	Puntuación discriminante	Grupo clasificado
1	I	1,3	4,1	-5,9957	I
2	I	3,7	6,9	-6,1213	I
3	I	5	3	-1,141	I
4	I	5,9	6,5	-3,4715	I
5	I	7,1	5,4	-1,2043	I
6	I	4	2,7	-1,8964	I
7	I	7,9	7,6	-2,4267	I
8	I	5,1	3,8	-1,7831	I
9	II	5,2	1	0,93	II
10	II	9,8	4,2	2,7086	II
11	II	9	4,8	1,3214	II
12	II	12	2	7,036	II
13	II	6,3	5,2	-1,8459	I
14	II	8,7	1,1	4,4593	II
15	II	11,1	4,1	4,1473	II
16	II	9,9	1,6	5,2353	II

En la tabla siguiente (resultados de la clasificación) se refleja el resumen de la clasificación de la tabla de arriba. A veces se utiliza en el análisis discriminante la expresión de *matriz de confusión* para referirse a la tabla siguiente:

Resultados de la clasificación[a]

		Préstamo	Grupo de pertenencia pronosticado: Fallidos	Grupo de pertenencia pronosticado: No Fallidos	Total
Original	Recuento	Fallidos	8	0	8
		No Fallidos	1	7	8
	%	Fallidos	100,0	,0	100,0
		No Fallidos	12,5	87,5	100,0

a. Clasificados correctamente el 93,8% de los casos agrupados originales.

◊ En la tabla que sigue se han calculado las probabilidades *a posteriori* (sin incorporar información a priori ni considerar gastos) de pertenencia a cada grupo utilizando la fórmula:

$$\text{Prob}(g/D) = \frac{e^{F_g}}{e^{F_I} + e^{F_{II}}} \quad g = I, II$$

Como puede observarse, las probabilidades de pertenencia al propio grupo son elevadas, excepto en el cliente cumplidor 13 que se clasifica erróneamente en el grupo de fallidos y que por añadidura tiene una probabilidad muy baja (0,1367) de pertenencia al grupo de los cumplidores.

Fallidos			*No Fallidos*		
Cliente	Prob(I/D)	Prob(II/D)	Cliente	Prob(I/D)	Prob(II/D)
1	0,9975	0,0025	9	0,2826	0,7174
2	0,9978	0,0022	10	0,0622	0,9378
3	0,7575	0,2425	11	0,2100	0,7900
4	0,9698	0,0302	12	0,0009	0,9991
5	0,7687	0,2313	13	0,1367	0,8633
6	0,8693	0,1307	14	0,0114	0,9886
7	0,9185	0,0815	15	0,0155	0,9845
8	0,8558	0,1442	16	0,0053	0,9947

◊ Como segunda aplicación, se realiza la clasificación incorporando información *a priori*.

Para clasificar a los clientes se va a utilizar el punto de corte $C_p = \frac{\overline{D}_I + \overline{D}_{II}}{2} - \ln \frac{\pi_{II}}{\pi_I}$

Si se establece que $\pi_1 = 0{,}10$ y $\pi_2 = 0{,}90$, el valor que se obtiene:

$$C_p = \frac{0{,}518 + 6{,}522}{2} - \ln\frac{0{,}9}{0{,}1} = 3{,}520 - 2{,}1972 = 1{,}323$$

con lo que la función discriminante de Fisher será:

$$D - C = 1{,}035.\text{Patrimonio\_Neto} - 0{,}932.\text{Deuda\_Pendiente} - 1{,}323$$

resultando:

Cliente	Grupo pertenencia	Patrimonio Neto	Deuda Pendiente	Puntuación discriminante	Grupo clasificado
1	I	1,3	4,1	-3,7987	I
2	I	3,7	6,9	-3,9243	I
3	I	5	3	1,056	I
4	I	5,9	6,5	-1,2745	I
5	I	7,1	5,4	0,9927	I
6	I	4	2,7	0,3006	I
7	I	7,9	7,6	-0,2297	I
8	I	5,1	3,8	0,4139	I
9	II	5,2	1	3,127	II
10	II	9,8	4,2	4,9056	II
11	II	9	4,8	3,5184	II
12	II	12	2	9,233	II
13	II	6,3	5,2	0,3511	I
14	II	8,7	1,1	6,6563	II
15	II	11,1	4,1	6,3443	II
16	II	9,9	1,6	7,4323	II

Los clientes 3, 5, 6 y 8, que antes estaban clasificados como fallidos, se clasifican ahora como cumplidores, ya que su puntuación discriminante ha pasado de negativa a positiva. Lo mismo ocurre con el cliente 13 que anteriormente estaba clasificado erróneamente como fallido cuando era cumplidor.

◊ Ahora se calcula el punto de corte teniendo en cuenta la información a priori e incorporando también los costes de la clasificación errónea. Como respecto al coste, se adopta el criterio de clasificar como cumplidor a un cliente fallido es 20 veces superior al coste de clasificar como fallido a un cliente cumplidor. Es decir, se establece que, la ratio: $\dfrac{\text{Coste}(\text{II}/\text{I})}{\text{Coste}(\text{I}/\text{II})} = 20$

El punto de corte discriminante será:

$$C_{p,c} = \frac{\overline{D}_I + \overline{D}_{II}}{2} - \ln \frac{\pi_{II}\ \text{Coste}(\text{I}/\text{II})}{\pi_I\ \text{Coste}(\text{II}/\text{I})} = \frac{0{,}518 + 6{,}522}{2} - \ln \frac{0{,}9}{0{,}1.20} = 4{,}319$$

La incorporación de los costes ha determinado que el nuevo punto de corte discriminante $C_{p,c}$ esté situado a la derecha del punto C, a diferencia de lo que ocurría cuando solamente se tenían en cuenta las probabilidades *a priori*.

con lo que la función discriminante de Fisher será:

$$D - C = 1{,}035.\text{Patrimonio\_Neto} - 0{,}932.\text{Deuda\_Pendiente} - 4{,}319$$

resultando:

Cliente	Grupo pertenencia	Patrimonio Neto	Deuda Pendiente	Puntuación discriminante	Grupo clasificado
1	I	1,3	4,1	-6,7947	I
2	I	3,7	6,9	-6,9203	I
3	I	5	3	-1,94	I
4	I	5,9	6,5	-4,2705	I
5	I	7,1	5,4	-2,0033	I
6	I	4	2,7	-2,6954	I
7	I	7,9	7,6	-3,2257	I
8	I	5,1	3,8	-2,5821	I
9	II	5,2	1	0,131	II
10	II	9,8	4,2	1,9096	II
11	II	9	4,8	0,5224	II
12	II	12	2	6,237	II
13	II	6,3	5,2	-2,6449	I
14	II	8,7	1,1	3,6603	II
15	II	11,1	4,1	3,3483	II
16	II	9,9	1,6	4,4363	II

Se comprueba que no altera la clasificación de ningún cliente respecto a la utilización del punto de corte inicial C. Es decir, la incorporación de los costes de clasificación errónea ha compensado, más o menos, la menor probabilidad *a priori* de ser un cliente fallido.

CLASIFICACIÓN EN MÁS DE DOS GRUPOS: ANÁLISIS DISCRIMINANTE MÚLTIPLE

- Número máximo de ejes discriminantes $\text{mín}(G-1, k)$, donde G es el número de categorías. Se obtienen $(G-1)$ ejes discriminantes si el número de variables explicativas es mayor o igual que $(G-1)$ - generalmente, este hecho suele ser cierto -.
- Cada una de las funciones discriminantes D_i se obtiene como función lineal de las k variables explicativas: $D_i = w_{i1} X_1 + w_{i2} X_2 + \cdots + w_{ik} X_k \quad i = 1, \cdots, G-1$
- Los $(G-1)$ ejes vienen definidos respectivamente por los vectores $(w_1, w_2, \cdots, w_{G-1})$

$$w_1 = \begin{pmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{1k} \end{pmatrix} \quad , \quad w_2 = \begin{pmatrix} w_{21} \\ w_{22} \\ \vdots \\ w_{2k} \end{pmatrix} \quad , \quad \cdots\cdots \quad , \quad w_{G-1} = \begin{pmatrix} w_{G-1,1} \\ w_{G-1,2} \\ \vdots \\ w_{G-1,k} \end{pmatrix}$$

Para la obtención del primer eje discriminante se maximiza la ratio variabilidad entre grupos entre variabilidad dentro grupos, es decir:

$$\boxed{\text{Máx } \lambda_1 = \frac{w_1^{'} F w_1}{w_1^{'} V w_1} = \frac{\text{separación entre grupos}}{\text{separación dentro grupos}} \quad \textit{(criterio obtención del primer eje discriminante)}}$$

Derivando la ratio e igualando a cero: $\dfrac{\vartheta \lambda_1}{\vartheta w_1} = 0$, con lo cual:

$$\frac{\vartheta \lambda_1}{\vartheta w_1} = \frac{2F w_1 (w_1^{'} V w_1) - 2V w_1 (w_1^{'} F w_1)}{(w_1^{'} V w_1)^2} = 0 \quad \Rightarrow \quad 2F w_1 (w_1^{'} V w_1) - 2V w_1 (w_1^{'} F w_1) = 0$$

operando con la expresión, resulta: $\dfrac{2F w_1}{2V w_1} =_1 \dfrac{(w_1^{'} F w_1)}{(w_1^{'} V w_1)} = \lambda_1 \quad \mapsto \quad F w_1 = V w_1 \lambda_1$

siendo, por tanto, $\boxed{\lambda_1 w_1 = V^{-1} F w_1}$

La obtención del vector w_1 resulta un problema de cálculo de un vector característico asociado a la matriz no simétrica $(V^{-1} F)$. De las raíces características que se obtienen al resolver la ecuación $\left[\lambda_1 w_1 = V^{-1} F w_1\right]$ se retiene la mayor, ya que λ_1 es la ratio que se pretende maximizar y w_1 es el vector característico asociado a dicha raíz característica.

Como λ_1 es la ratio $\left[\dfrac{w_1^{'} F w_1}{w_1^{'} V w_1}\right]$ medirá el poder discriminante del primer eje discriminante. El resto de los ejes discriminantes son otros vectores característicos de la matriz $(V^{-1} F)$, ordenados según el orden decreciente de las raíces características. Así, el segundo eje discriminante tendrá menor poder discriminante que el primero, pero más que cualquiera de los restantes.

Puesto que la matriz $(V^{-1} F)$ no es simétrica, en general, esto implicará que los ejes discriminantes no serán ortogonales, es decir, no serán perpendiculares entre sí.

Contrastes de significación

En el análisis discriminante múltiple se plantean contrastes específicos para determinar si cada uno de los valores λ_i es estadísticamente significativo, es decir, para determinar si cada uno de los valores λ_i contribuye o no a la discriminación entre los diferentes grupos.

Este tipo de contrastes se realiza a partir del estadístico V de Barlett. El estadístico V es una función de la Λ de Wilks y se aproxima a una chi-cuadrado, tiene interés en el análisis discriminante por su descomponibilidad.

Estadístico V de Barlett: $\boxed{V = \left[n - 1 - \frac{K+G}{2}\right] (\ln \Lambda)}$ $\quad V \approx \chi^2_{K(G-1)}$ $\quad \begin{cases} K \equiv \text{variables categóricas} \\ G \equiv \text{grupos} \end{cases}$

Este estadístico se utiliza en el análisis multivariante para contrastar las hipótesis $\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_G \\ H_1 : \text{No todas las } \mu_g \text{ son iguales} \end{cases}$

☞ En el análisis multivariante de la varianza con un factor se contrasta esta hipótesis para determinar si el factor (variable categórica con G grupos) explica la variabilidad del vector de variables dependientes de forma significativa.

☞ En el análisis discriminante múltiple la hipótesis a contrastar sigue siendo la misma, aunque los papeles se han invertido. Ahora se realiza el contraste para tratar de dar respuesta a la pregunta: *¿Las K variables clasificadoras contribuyen significativamente a discriminar entre los G grupos?*

Si no se rechaza la hipótesis nula citada, no se debería continuar el análisis, puesto que las variables clasificadoras utilizadas en la investigación no tienen ningún poder discriminante significativo.

- Para examinar el poder discriminante de cada uno de los ejes que se construyen en el análisis discriminante, se descompone el estadístico V en productos a partir de la descomposición de la Λ de Wilks. De acuerdo con su definición, el recíproco de Λ se puede descomponer:

$$\frac{1}{\Lambda} = \frac{|T|}{|V|} = |V|^{-1}|T| = |V^{-1}|\,|T| = |V^{-1}T| = |V^{-1}(F+V)| = |I + V^{-1}F|$$

teniendo en cuenta que el determinante de una matriz es igual al producto de sus raíces características, se obtiene que:

$$\frac{1}{\Lambda} = |I + V^{-1}F| = (1+\lambda_1)(1+\lambda_2) \;\cdots\; (1+\lambda_{G-1})$$

sustituyendo en el estadístico V de Barlett, se obtiene la expresión alternativa del estadístico:

$$\boxed{\text{Estadístico V de Barlett: } V = \left[n - 1 - \frac{K+G}{2}\right] \sum_{g=1}^{G-1} \ln(1+\lambda_g)}$$

Si se rechaza la hipótesis nula, significa que al menor uno de los ejes discriminantes es estadísticamente significativo. Esto implica a su vez que el primer eje discriminante es estadísticamente significativo, debido a que es precisamente el que tiene mayor poder discriminante.

En caso de que se acepte la hipótesis de que el primer eje discriminante es significativo, se pasa a contrastar la significación conjunta del resto de los ejes discriminantes, utilizando el estadístico:

$$\boxed{V = \left[n - 1 - \frac{K+G}{2} \right] \sum_{g=2}^{G-1} \ln(1+\lambda_g)}$$

De forma general, se puede establecer la expresión de contrastación secuencial mediante el estadístico:

Estadístico V de Barlett: $$\boxed{V_j = \left[n - 1 - \frac{K+G}{2} \right] \sum_{g=j+1}^{G-1} \ln(1+\lambda_g) \quad \text{donde} \quad j = 0, 1, 2, \cdots, G-2}$$

Así, en el proceso secuencial se van eliminando del estadístico V las raíces características que van resultando significativas, deteniendo el proceso cuando se acepte la hipótesis nula de no significatividad de los ejes discriminantes que queden por contrastar.

PRÉSTAMOS-RIESGO

Cuando una entidad financiera concede un préstamo personal a un cliente se enfrenta a la doble posibilidad de que sea reintegrado o de que no lo sea. En este último caso el préstamo será finalmente clasificado como fallido. Obviamente, si la entidad financiera conociera de antemano que una persona va a resultar fallida no le concedería el préstamo en ningún caso. En esta línea, puede utilizar la información existente en la entidad sobre préstamos concedidos en el pasado para la concesión de préstamos futuros de forma que se evite, o al menos, se reduzca la posibilidad de conceder préstamos que después fueran fallidos.

En los archivos de la entidad financiera existe información de las características de las personas a las que se les ha concedido un préstamo, ya que el cliente en el momento de solicitar el préstamo ha facilitado datos acerca de cuestiones tales como ingresos, edad, sexo, situación familiar, antigüedad en el puesto de trabajo, régimen de tenencia de la vivienda, etc. Es muy posible que los clientes cumplidores tengan unas características distintas a los clientes fallidos.

Utilizando estas características se trata de *establecer unas funciones* que clasifiquen lo más correctamente posible a los clientes a los que se les ha concedido un préstamo en cumplidores y fallidos (finalidad explicativa). Posteriormente, estas funciones se emplearán, en el caso de que se haya realizado adecuadamente dicha clasificación, para determinar si se conceden o no los préstamos futuros a futuros solicitantes (finalidad predictiva).

ANÁLISIS DISCRIMINANTE CON SPSS

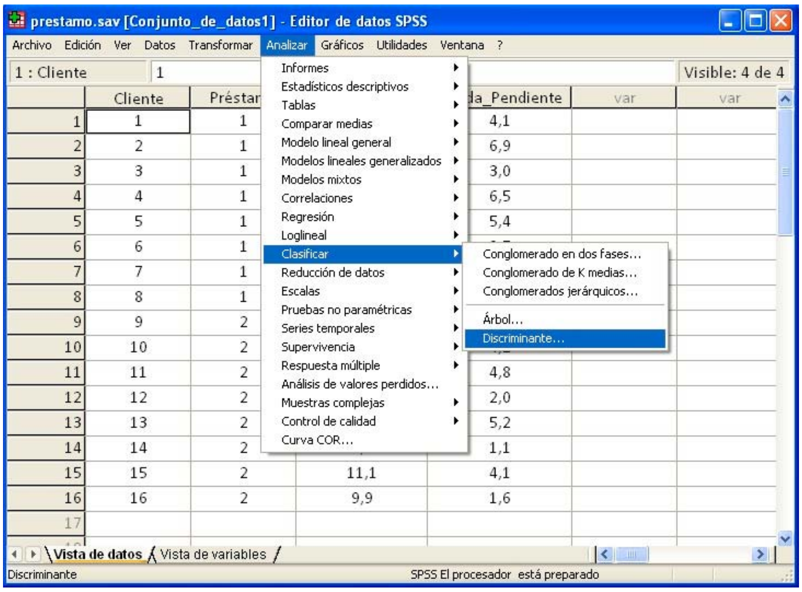
La tabla adjunta contiene información de 16 clientes de una entidad financiera a los que se les concedió un préstamo. Pasados 3 años desde la concesión del préstamo, de los 16 clientes, había 8 que fueron clasificados como fallidos (*grupo 1*) mientras que los otros 8 clientes fueron cumplidores (*grupo 2*), ya que reintegraron el préstamo.
Para cada uno de los 16 clientes se dispone de información sobre X_1 = 'su patrimonio neto' y X_2 ='sus deudas pendientes', en el momento de la solicitud. Con esta información se pretende construir una función discriminante que separe/diferencie lo más posible a los dos grupos y que permita clasificar, con los menores errores posibles, a los distintos clientes en los dos grupos.

Cliente	Préstamo	Patrimonio Neto	Deuda Pendiente
1	1	1,3	4,1
2	1	3,7	6,9
3	1	5,0	3,0
4	1	5,9	6,5
5	1	7,1	5,4
6	1	4,0	2,7
7	1	7,9	7,6
8	1	5,1	3,8
9	2	5,2	1,0
10	2	9,8	4,2
11	2	9,0	4,8
12	2	12,0	2,0
13	2	6,3	5,2
14	2	8,7	1,1
15	2	11,1	4,1
16	2	9,9	1,6

El director de la entidad financiera tiene dos nuevas solicitudes de un préstamo instantáneo. El primer solicitante dispone de un patrimonio neto de 10,1, con unas deudas pendientes de 6,8. Para el segundo solicitante los valores de estas variables son 9,7 y 2,2 respectivamente. ¿Qué decisión debe tomar?
(Nota.- Las unidades monetarias se expresan en 100.000 euros)

- Para hacer un Análisis Discriminante, se selecciona sucesivamente del menú principal:

Analizar → Clasificar → Discriminante



En primer lugar, hay que elegir cuál es la **Variable de Agrupación**, es decir, qué variable juega el papel de variable categórica dependiente cuyas categorías definen los posibles grupos de pertenencia de los individuos. En este caso, la variable es *Préstamo*. Además, en el botón con el nombre *Definir Rango*, es necesario especificar cuáles son los valores Mínimo y Máximo de esta variable. Se introducen los valores correspondientes: *Mínimo*: 1 y *Máximo*: 2.

Las otras dos variables, X_1 = 'Patrimonio_ Neto' y X_2 ='Deuda_Pendiente', se eligen como variables independientes, cuyos valores se utilizan para construir la *función discriminante*. Estas variables pueden introducirse en el modelo simultáneamente o por etapas

SPSS ofrece en los distintos botones activados del cuadro de diálogo: 'Seleccionar', 'Estadísticos', 'Clasificar', 'Guardar'. El botón 'Método' sólo se activa si previamente se ha elegido Introducir las variables con un **Método por pasos**.

Seleccionar: Permite reducir el análisis a un subgrupo de la muestra total, subgrupo que vendrá definido por una variable de selección. Este no es el caso, no se elige esta opción.

ESTADÍSTICOS UTILIZADOS:

- **F de Snedecor:** Se compara para cada variable las desviaciones de las medias de cada uno de los grupos a la media total, entre las desviaciones a la media *dentro* de cada grupo.

  - Si F es grande para cada variable, entonces las medias de cada grupo están muy separadas y la variable discrimina bien.

  - Si F es pequeña para cada variable, la variable discrimina poco, ya que habrá poca homogeneidad en los grupos y éstos estarán muy próximos.

- **λ de Wilks:** Se consideran las variables de modo individual, la λ es el cociente entre la suma de cuadrados *dentro* de los grupos y la suma de cuadrados total (sin distinguir grupos). Esto equivale a las desviaciones a la media dentro de cada grupo, entre las desviaciones a la media total sin distinguir grupos.

  - Si λ es pequeño la variable discrimina mucho: la variabilidad total se debe a las diferencias *entre* grupos, no a las diferencias *dentro* de grupos.

VARIABLES ORIGINALES QUE SE CONSIDERAN: La idea del análisis discriminante es construir funciones lineales de las variables originales que discriminen entre los distintos grupos. Sin embargo, no todas las variables discriminan de la misma forma o tienen los mismos valores de la F de Snedecor o de la λ de Wilks. Por ello, a la hora de construir las funciones lineales, no es necesario incluir a todas las variables iniciales en la función.

Como criterio general para seleccionar una variable se emplea la selección del valor de la λ de Wilks o, de modo equivalente, del valor de su F asociada.

Se utilizan fundamentalmente dos métodos de selección de variables: el método directo (Introducir independientes juntas) y el método stepwise (Usar método de selección por pasos). En el método directo se consideran todas las variables originales que verifiquen un criterio de selección.

El método stepwise funciona con varios pasos:

(a) Se incluye en el análisis la variable que tenga el mayor valor real aceptable para el criterio de selección o de *entrada*.

(b) Se evalúa el criterio de selección para las variables no seleccionadas. La variable que presenta el valor más alto para el criterio se selecciona (siempre que se encuentre dentro de un límite).

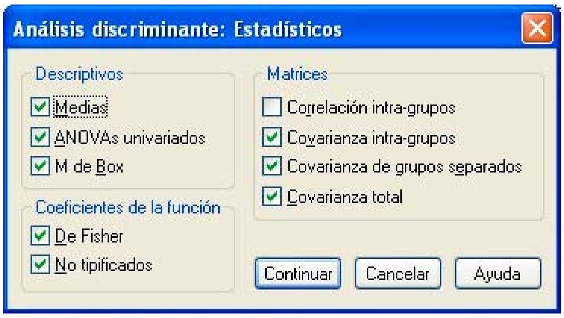
(c) Se examinan las variables seleccionadas según un criterio de *salida* y se examinan también las variables no seleccionadas, para ver si cumplen el criterio de *entrada*. Se excluyen o se incluyen variables según cumplan los criterios de entrada y salida.

(d) Se repite el proceso © hasta que ninguna variable más pueda ser seleccionada o eliminada.

Además de todo lo expuesto, en el SPSS se considera un número máximo de pasos, dado que una variable puede ser incluida y eliminada en más de una ocasión. Se toma el doble del número de variables originales como número máximo de pasos del método stepwise.

En SPSS se considera también para cada variable la *tolerancia asociada*: Se define para un conjunto de p variables, $R_i \equiv$ coeficiente de correlación múltiple, que expresa el porcentaje de variabilidad de la variable $(x_i \quad i=1,\cdots,p)$ recogida por el resto de $(p-1)$ variables. $R_i^2 \equiv$ coeficiente de determinación. La *tolerancia* se define como $(1-R_i^2)$. *Cuanto mayor sea la tolerancia de una variable, más información independiente del resto de variables recogerá*.

De este modo, si en una iteración dada del procedimiento stepwise la variable seleccionada verifica que su tolerancia con respecto a las variables ya incluidas en la función discriminante es muy pequeña entonces la variable no se incluye en dicha etapa. Así, se evita la redundancia de información.



La opción [*Estadísticos*] se encuentra dividida en tres grandes áreas: Descriptivos, Coeficientes de la función y Matrices.

- DESCRIPTIVOS:

Medias: Proporciona el vector de medias (los centroides) y desviaciones típicas de cada variable para cada grupo.

Univariante ANOVA: Contrasta igualdad de medias entre los grupos para cada variable.

M de Box: Contrasta la hipótesis nula de que las matrices de varianzas-covarianzas poblacionales son iguales en los distintos grupos.

- COEFICIENTES DE LA FUNCIÓN:

De Fisher: Coeficientes de la función de clasificación bajo Normalidad

No tipificados: Coeficientes de la función discriminante canónica de Fisher 'centrados'

- MATRICES:

Covarianza de grupos separados: Proporciona la matriz de varianzas y covarianzas de cada grupo, es decir, las matrices S_1 y S_2 , donde:

$$S_k = \begin{bmatrix} \sum_{i=1}^{n_k}\left(X_{i1}^{(k)} - \overline{X}_1^{(k)}\right)^2 & \sum_{i=1}^{n_k}\left(X_{i1}^{(k)} - \overline{X}_1^{(k)}\right)\left(X_{i2}^{(k)} - \overline{X}_2^{(k)}\right) \\ \sum_{i=1}^{n_k}\left(X_{i1}^{(k)} - \overline{X}_1^{(k)}\right)\left(X_{i2}^{(k)} - \overline{X}_2^{(k)}\right) & \sum_{i=1}^{n_k}\left(X_{i2}^{(k)} - \overline{X}_2^{(k)}\right)^2 \end{bmatrix} \quad k = 1,2$$

Covarianza intra-grupos: Proporciona la matriz de varianzas y covarianzas 'combinada', obtenida como media ponderada de las dos anteriores, es decir:

$$S = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

Covarianza Total: Proporciona la matriz de varianzas y covarianzas de (X_1, X_2) para todos los $n_1 + n_2 = 16$ individuos de la población, sin distinción de grupo.

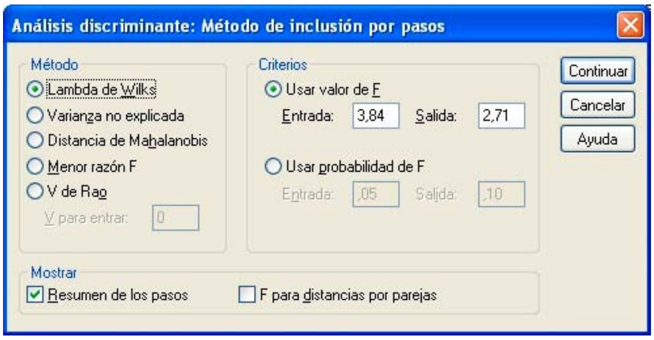
COMPROBACIÓN SUPUESTOS PARAMÉTRICOS: La función discriminante minimiza la probabilidad de equivocarse al clasificar a los individuos en cada grupo. Para ello, las variables originales se deben distribuir como una normal multivariante y las matrices de covarianzas deben de ser iguales en todos los grupos. En la práctica es una técnica robusta y funciona bien aunque las dos restricciones anteriores no se verifiquen.

- Si un conjunto de variables se distribuye como una normal multivariante, entonces cualquier combinación lineal de ellas se distribuye como una normal multivariante. Por ello, si alguna de las variables originales no se distribuye como una normal, entonces es seguro que todas las variables conjuntamente no se distribuirán como una normal multivariante.

- La segunda restricción se ocupa de la igualdad entre las matrices de covarianzas de los grupos. Para comprobar esto, se puede utilizar la *Prueba M de Box*, que tiene como hipótesis nula que las matrices de covarianzas son iguales. Se basa en el cálculo de los determinantes de las matrices de covarianzas de cada grupo. El valor obtenido se aproxima por una F de Snedecor. Si el p_valor < 0,05 se rechaza la igualdad entre las matrices de covarianzas.

  El *test de M de Box* es sensible a la falta de normalidad multivariante, es decir, matrices iguales pueden aparecer como significativamente diferentes si no existe normalidad. Por otra parte, si las muestras son grandes, pierde efectividad (es más fácil rechazar la hipótesis nula).

En este caso, se dejan las opciones que vienen por defecto en SPSS.



- Lambda de Wilks: Estadístico que mide el poder discriminante de un conjunto de variables

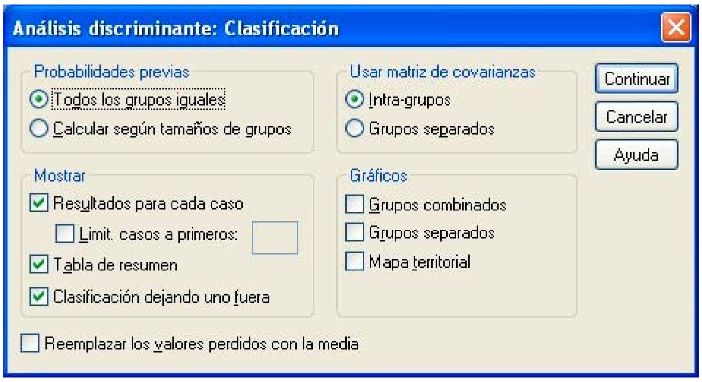
$$\Lambda = \frac{|V|}{|T|} = \frac{|V|}{|V+F|} = \frac{1}{\prod_{i=1}^{\min(q-1,\,p)} (1+\lambda_i)} \qquad (0 \le \Lambda \le 1)$$

  Cuanto más cerca de 0 mayor es el poder discriminante de las variables consideradas, y cuanto más cerca de 1 menor es el poder discriminante.

- Estadísticos asociados: F de Rao; χ^2 de Barlett (tests sobre las diferencias de medias en ambos grupos)

- La i-ésima correlación canónica viene dada por: $CR_i = \sqrt{\dfrac{\lambda_i}{1+\lambda_i}}$

  Mide, en términos relativos, el poder discriminante de la i-ésima función discriminante, ya que es el porcentaje de la variación total en dicha función que es explicada por la diferencia entre los grupos, $0 \le CR_i \le 1$, cuanto más cerca de 1 esté su valor, mayor es la potencia discriminante de la i-ésima función discriminante.



Una opción interesante en la opción [*Clasificación*] es la de '*Reemplazar los valores perdidos con la media*'. En más de una investigación, por algún motivo en la base de datos hay valores perdidos, y para que estos no afecten los resultados finales, existe ésta opción de reemplazo, que se recomienda utilizar.

- PROBABILIDADES PREVIAS*:*

Son las probabilidades a priori para cada grupo. En este caso serían p_1 = p(pertenecer al grupo 1), p_2 = p(pertenecer al grupo 2). Estos valores se utilizan, por ejemplo, en la regla de clasificación de la máxima verosimilitud bajo el supuesto de normalidad.

Todos los grupos iguales: p_1 = p(pertenecer al grupo 1) = p_2 = p(pertenecer al grupo 2) = ½

- USAR MATRIZ DE COVARIANZA:

Intra-grupos: De esta manera se especifica que cuando se obtengan los autovectores de la matriz $(V^{-1}F)$, que son precisamente los coeficientes de las distintas funciones discriminantes, se utilice la restricción **a'Sa=1**, utilizando la matriz de varianzas *entre* grupos 'combinada' **S**.

- MOSTRAR:

Resultados para cada caso: Muestra el proceso de clasificación paso a paso para cada uno de los 16 individuos de la población, con las probabilidades a posteriori para cada uno de ellos, calculadas a partir de las puntuaciones discriminantes.

Tabla de resumen: Proporciona la *matriz de confusión*, es decir la matriz de clasificación para los propios 16 individuos de la muestra para los que conocemos de antemano su adscripción.

Clasificación dejando uno fuera: Proporciona la matriz de clasificación pero obtenida con el *método Jacknife*, que obtiene, en general una estimación de la proporción de clasificaciones erróneas más fiable.

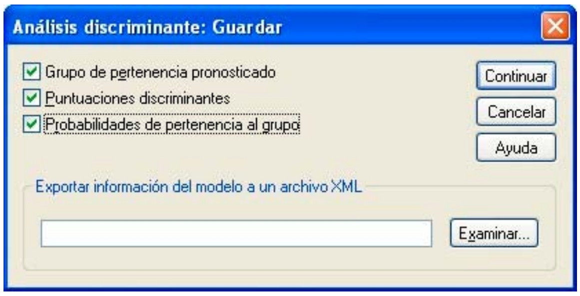
- GRÁFICOS:

Grupos combinados: Representa las puntuaciones discriminantes o valores de la(s) funcion(es) discriminante(s), para los 16 individuos de la muestra (8 de cada grupo) todos juntos en un gráfico, junto con sus centroides.
Como sólo hay una función discriminante este gráfico *no se hace* (si se selecciona, luego no aparece).

Grupos separados: Representa un gráfico como el anterior pero para cada grupo.
En este caso, representaría en el primer gráfico únicamente los 8 individuos del grupo 1 y en el segundo sólo los 8 del grupo 2.

Mapa territorial: Con una única función discriminante no lo hace .

Si se desea que el análisis sea 'Guardado' se procede a dar un clic en el botón de la opción [*Análisis discriminante*].



El *Visor de resultados* de SPSS muestra:

Resumen del procesamiento para el análisis de casos

Casos no ponderados		N	Porcentaje
Válidos		16	100,0
Excluidos	Códigos de grupo para perdidos o fuera de rango	0	,0
	Perdida al menos una variable discriminante	0	,0
	Perdidos o fuera de rango ambos, el código de grupo y al menos una de las variables discriminantes.	0	,0
	Total excluidos	0	,0
Casos Totales		16	100,0

Estadísticos de grupo

Préstamo		Media	Desv. típ.	N válido (según lista) No ponderados	Ponderados
Fallidos	Patrimonio_Neto	5,000	2,0709	8	8,000
	Deuda_Pendiente	5,000	1,8639	8	8,000
No Fallidos	Patrimonio_Neto	9,000	2,2891	8	8,000
	Deuda_Pendiente	3,000	1,7444	8	8,000
Total	Patrimonio_Neto	7,000	2,9518	16	16,000
	Deuda_Pendiente	4,000	2,0268	16	16,000

Se muestran los estadísticos descriptivos: media y desviación típica total de (X_1, X_2) sobre los $n = n_1 + n_2 = 16$ individuos y para los dos grupos: Media y desviación típica de (X_1, X_2) para los $n_1 = 8$ clientes del grupo 1, y media y desviación típica de (X_1, X_2) para los $n_2 = 8$ clientes del grupo 2.

Se observa que el punto de corte discriminante de los dos grupos para la variable X_1 = 'Patrimonio_Neto' se encuentra en el valor 7:

$$\overline{X}_{1,I} = 5 \qquad \overline{X}_{1,II} = 9 \qquad C_1 = \frac{\overline{X}_{1,I} + \overline{X}_{1,II}}{2} = \frac{5+9}{2} = 7$$

El punto de corte se toma como referencia para clasificar a un individuo en uno u otro grupo (fallido, cumplidores): Si el Patrimonio_Neto es menor que 7 se clasifica al cliente como fallido (grupo 1), mientras que se clasifica como cumplidor (grupo 2) si el Patrimonio_Neto es mayor que esa cifra.

Por otra parte, el punto de corte discriminante de los dos grupos para la variable X2 = 'Deuda_Pendiente' de los dos grupos será:

$$\overline{X}_{2,I} = 5 \qquad \overline{X}_{2,II} = 3 \qquad C_1 = \frac{\overline{X}_{2,I} + \overline{X}_{2,II}}{2} = \frac{5+3}{2} = 4$$

Si las deudas pendientes son mayores que 4 se clasifica al cliente como fallido (grupo 1), mientras que se clasifica como cumplidor (grupo 2) si las deudas pendientes son menores que esa cifra.

Pruebas de igualdad de las medias de los grupos

	Lambda de Wilks	F	gl1	gl2	Sig.
Patrimonio_Neto	,510	13,433	1	14	,003
Deuda_Pendiente	,740	4,910	1	14	,044

Los contrastes de igualdad de medias entre los dos grupos para cada variable (en ambos casos se rechaza la hipótesis nula, p_valor < 0,05, es decir, los dos grupos, en media son diferentes).

La información de esta tabla de ANOVAs univariados suele utilizarse como prueba preliminar para detectar si los grupos difieren en las variables de clasificación seleccionadas; sin embargo, hay que considerar que una variable no significativa a nivel univariante podría aportar información discriminativa a nivel multivariante.

Matrices intra-grupo combinadas[a]

		Patrimonio_ Neto	Deuda_ Pendiente
Covarianza	Patrimonio_Neto	4,764	1,001
	Deuda_Pendiente	1,001	3,259

a. La matriz de covarianzas tiene 14 grados de libertad

Matrices de covarianzas[a]

Préstamo		Patrimonio_ Neto	Deuda_ Pendiente
Fallidos	Patrimonio_Neto	4,289	1,824
	Deuda_Pendiente	1,824	3,474
No Fallidos	Patrimonio_Neto	5,240	,177
	Deuda_Pendiente	,177	3,043
Total	Patrimonio_Neto	8,713	-1,199
	Deuda_Pendiente	-1,199	4,108

a. La matriz de covarianzas total presenta 15 grados de libertad.

La salida de la matriz de covarianzas proporciona:

$$S_1 = \begin{bmatrix} 4,289 & 1,824 \\ 1,824 & 3,474 \end{bmatrix} \;,\; S_2 = \begin{bmatrix} 5,240 & 0,177 \\ 0,177 & 3,043 \end{bmatrix} \;,\; S_{total} = \begin{bmatrix} 8,713 & -1,199 \\ -1,199 & 4,108 \end{bmatrix}$$

Por otra parte, la media ponderada de S_1 y S_2 debe de coincidir con la matriz 'intra-grupos combinada', denominada S. Es decir, debe verificarse que:

$$S = \begin{bmatrix} 4,764 & 1,001 \\ 1,001 & 3,259 \end{bmatrix} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2} = \frac{7}{14}\begin{bmatrix} 4,289 & 1,824 \\ 1,824 & 3,474 \end{bmatrix} + \frac{7}{14}\begin{bmatrix} 5,240 & 0,177 \\ 0,177 & 3,043 \end{bmatrix}$$

Aparece después la *Prueba de Box* para el contraste de la hipótesis nula de igualdad de las matrices de varianzas-covarianzas poblacionales. Uno de los supuestos del análisis discriminante es que todos los grupos proceden de la misma población y, más concretamente, que las matrices de varianzas-covarianzas poblacionales correspondientes a cada grupo son iguales entre sí.

Análisis 1

Prueba de Box sobre la igualdad de las matrices de covarianza

Logaritmo de los determinantes

Préstamo	Rango	Logaritmo del determinante
Fallidos	2	2,449
No Fallidos	2	2,767
Intra-grupos combinada	2	2,676

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

Resultados de la prueba

M de Box		,951
F	Aprox.	,268
	gl1	3
	gl2	35280,000
	Sig.	,849

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

El *estadístico M de Box* toma la forma: $M = (n-g)\,\log|S| - \sum_{j=1}^{g}(n_j-1)\,\log|S_j|$

Donde S es la matriz de varianzas-covarianzas combinada, S_j es la matriz de varianzas-covarianzas del grupo j-ésimo, n es el número total de casos y g el número de grupos. El estadístico M carece de distribución muestral conocida, pero puede transformarse en un estadístico F e interpretarse como tal (muchos investigadores critican este estadístico por ser demasiado sensible a pequeñas desviaciones de la normalidad multivariante y a tamaños muestrales grandes, tendiendo a ser conservador).

Se observa que la primera tabla ofrece los logaritmos de los determinantes de todas las matrices utilizadas en el cálculo del estadístico M. Dado que el estadístico es multivariante, la tabla permite comprobar qué grupos (cuando hay más de dos) difieren más.

La tabla (Resultados de la prueba) ofrece la prueba *M de Box* y su transformación en un estadístico F. El resultado de la prueba hace que no se rechace la igualdad de matrices de varianzas-covarianzas (Sig=0,849 > 0,05), concluyendo que los dos grupos tienen la misma matriz de varianzas-covarianzas (no hay un grupo más variable que otro).

A continuación aparecen los resultados del análisis discriminante (estadísticos por pasos):

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

		Lambda de Wilks							
						F exacta			
Paso	Introducidas	Estadístico	gl1	gl2	gl3	Estadístico	gl1	gl2	Sig.
1	Patrimonio_ Neto	,510	1	1	14,000	13,433	1	14,000	,003
2	Deuda_ Pendiente	,368	2	1	14,000	11,151	2	13,000	,002

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

a. El número máximo de pasos es 4.

b. La F parcial mínima para entrar es 3.84.

c. La F parcial máxima para salir es 2.71

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

Las variables son introducidas/eliminadas del modelo en la medida en que tengan asociado un menor valor del estadístico Λ de Wilks.

Variables en el análisis

Paso		Tolerancia	F para salir	Lambda de Wilks
1	Patrimonio_Neto	1,000	13,433	
2	Patrimonio_Neto	,935	13,136	,740
	Deuda_Pendiente	,935	5,016	,510

Variables no incluidas en el análisis

Paso		Tolerancia	Tolerancia mín.	F para entrar	Lambda de Wilks
0	Patrimonio_Neto	1,000	1,000	13,433	,510
	Deuda_Pendiente	1,000	1,000	4,910	,740
1	Deuda_Pendiente	,935	,935	5,016	,368

Lambda de Wilks

						F exacta			
Paso	Número de variables	Lambda	gl1	gl2	gl3	Estadístico	gl1	gl2	Sig.
1	1	,510	1	1	14	13,433	1	14,000	,003
2	2	,368	2	1	14	11,151	2	13,000	,002

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	1,716[a]	100,0	100,0	,795

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

Como hay *g*=2 grupos y *p*=2 variables, sólo hay q=min (k, g-1)=1 función discriminante, o equivalentemente, la matriz $(V^{-1}F)$ tiene rango q=min (k, g-1)=1 y sólo hay un autovalor distinto de cero, λ_1=1,716, que es el que aparece en la tabla.

El *autovalor de una función* se interpreta como la parte de variabilidad total de la nube de puntos proyectada sobre el conjunto de todas las funciones atribuible a la función. Si su valor es grande, la función discriminará mucho.

Además, se refleja el coeficiente *eta* o *correlación canónica*: $\eta = \sqrt{\frac{\lambda_1}{1+\lambda_1}} = \sqrt{\frac{1,716}{1+1,716}} = 0,795$

Las *correlaciones canónicas*, miden las desviaciones de las puntuaciones discriminantes *entre* grupos respecto a las desviaciones totales sin distinguir grupos. Si su valor es grande (próximo a 1) la dispersión será debida a las diferencias *entre* grupos, y en consecuencia, la función discriminará mucho.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	,368	12,987	2	,002

El estadístico del contraste de significación global Lambda de Wilks: $\Lambda = \frac{1}{1+\lambda_1} = \frac{1}{1+1,716} = 0,368$

que conduce a rechazar la hipótesis nula de igualdad de medias [p-valor = 0,02 < 0,05], lo que indica la conveniencia de extraer una (la única posible) función discriminante, o lo que es lo mismo, que dicha función sea significativa.

Interpretación de las funciones discriminantes: a la vista de los valores de $\rho(X_1,y)$, y $\rho(X_2,y)$, parece que la variable que más contribuye a la discriminación es X_1 ='Patrimonio_Neto'

Coeficientes estandarizados de las funciones discriminantes canónicas

	Función
	1
Patrimonio_Neto	,922
Deuda_Pendiente	-,686

Matriz de estructura

	Función
	1
Patrimonio_Neto	,748
Deuda_Pendiente	-,452

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.

COEFICIENTES ESTANDARIZADOS: Aparecen los coeficientes de la *función discriminante* canónica estandarizados, estos coeficientes aparecen cuando se tipifican o estandarizan cada una de las variables clasificadoras para que tengan media 0 y desviación típica 1. De esta forma se evitan los problemas de escala que pudieran existir entre las variables y, consecuentemente, la magnitud de los coeficientes estandarizados son un indicador de la importancia que tiene cada variable en el cálculo de la función discriminante. En esta línea, se observa que la variable Patrimonio_Neto (X_1) tiene una influencia que es casi un 50% superior a la ejercida por la variable Deuda_Pendiente (X_2).

MATRIZ DE ESTRUCTURA: Es conveniente conocer cuáles son las variables que tienen *mayor poder discriminante* en orden a clasificar a un individuo en uno de los grupos (fallidos, cumplidores). Una forma de medir ese poder discriminante es calculando el coeficiente de correlación entre cada una de las variables y la función discriminante. Esta es precisamente la información que se da en la tabla (Matriz de estructura), en este caso, la correlación de la función discriminante con la variable Patrimonio_Neto (0,748) es mayor en valor absoluto que con la variable Deuda_Pendiente (0,452). Las comparaciones deben hacerse siempre en valor absoluto. En el programa SPSS las variables aparecen ordenadas de acuerdo con el valor absoluto de los coeficientes de correlación.

Los Coeficientes de las funciones discriminantes canónicas de Fisher son:

Coeficientes de las funciones canónicas discriminante:

	Función
	1
Patrimonio_Neto	,422
Deuda_Pendiente	-,380
(Constante)	-1,437

Coeficientes no tipificados

En la tabla aparece información de los coeficientes de la función discriminante canónica no estandarizados. Los coeficientes de esta función son estrictamente proporcionales a los coeficientes de la función discriminante de Fisher $(D-C)$. En este caso, el factor de proporcionalidad es 0,408; esto es, cada coeficiente es igual a 0,408 multiplicado por el coeficiente de la función discriminante de Fisher. Estos coeficientes no estandarizados se obtienen utilizando la regla de normalización de $w'\,V\,w=1$, así pues, se toma como norma el denominador de la variación dentro de los grupos:

Los coeficientes w se obtienen: $\text{Máx}\ \lambda = \dfrac{w'\,F\,w}{w'\,V\,w} = \dfrac{\text{variación entre grupos}}{\text{variación dentro grupos}}$

Centroides de cada grupo (media de la función discriminante en cada grupo):

Funciones en los centroides de los grupos

	Función
Préstamo	1
1	-1,225
2	1,225

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

Con los resultados obtenidos, el punto de corte discriminante será el punto medio de las funciones en los centroides de los grupos: $C = \dfrac{D_1 + D_2}{2} = \dfrac{-1,225 + 1,225}{2} = 0$

Estadísticos de clasificación:

Probabilidades a priori de pertenencia a los grupos (se supone $p_1 = p_2 = 1/2$)

Probabilidades previas para los grupos

		Casos utilizados en el análisis	
Préstamo	Previas	No ponderados	Ponderados
1	,500	8	8,000
2	,500	8	8,000
Total	1,000	16	16,000

Coeficientes de la función de clasificación: Aquí se muestran los coeficientes de las funciones de clasificación que se obtendrían bajo el supuesto de Normalidad bivariante para (X_1, X_2) en ambas poblaciones, utilizando el criterio de la máxima verosimilitud y probabilidades $(p_1 = p_2 = 1/2)$ a priori iguales.

Coeficientes de la función de clasificación

	Préstamo	
	Fallidos	No Fallidos
Patrimonio_Neto	,777	1,813
Deuda_Pendiente	1,296	,364
(Constante)	-5,876	-9,396

Funciones discriminantes lineales de Fisher

Las funciones de clasificación son:

$$\begin{cases} F_I = 0,777.\text{Patrimonio\_Neto} + 1,296.\text{Deuda\_Pendiente} - 5,876 \\ \\ F_{II} = 1,813.\text{Patrimonio\_Neto} + 0,364.\text{Deuda\_Pendiente} - 9,396 \end{cases}$$

Para el grupo 1, la función de clasificación es de la forma: $\hat{d}_I(x) = \bar{x}_1' \; S^{-1} \; x - \frac{1}{2} \, \bar{x}_1' \; S^{-1} \; \bar{x}_1 + \ln(p_1)$

Estadísticos de grupo

Préstamo		Media
Fallidos	Patrimonio_Neto	5,000
	Deuda_Pendiente	5,000
No Fallidos	Patrimonio_Neto	9,000
	Deuda_Pendiente	3,000
Total	Patrimonio_Neto	7,000
	Deuda_Pendiente	4,000

Los centros de gravedad o centroides de los dos grupos serán:

$$\bar{x}_I = \begin{bmatrix} \bar{X}_{1,I} \\ \bar{X}_{2,I} \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \qquad \bar{x}_{II} = \begin{bmatrix} \bar{X}_{1,II} \\ \bar{X}_{2,II} \end{bmatrix} = \begin{bmatrix} 9 \\ 3 \end{bmatrix}$$

Matriz intra-grupo combinada:

$$S = \begin{bmatrix} 4,764 & 1,001 \\ 1,001 & 3,259 \end{bmatrix}$$

$$\hat{d}_I(x) = \begin{bmatrix} 5 & 5 \end{bmatrix} \begin{bmatrix} 4,764 & 1,001 \\ 1,001 & 3,259 \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 5 & 5 \end{bmatrix} \begin{bmatrix} 4,764 & 1,001 \\ 1,001 & 3,259 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 5 \end{bmatrix} + \ln(0,5) =$$

$$= 0,777.\overbrace{Patrimonio\_Neto}^{x_1} + 1,296.\overbrace{Deuda\_Pendiente}^{x_2} - 5,876$$

Para el grupo 2, la función de clasificación es de la forma: $\hat{d}_{II}(x) = \bar{x}_2' \; S^{-1} \; x - \frac{1}{2} \, \bar{x}_2' \; S^{-1} \; \bar{x}_2 + \ln(p_2)$

$$\hat{d}_{II}(x) = \begin{bmatrix} 9 & 3 \end{bmatrix} \begin{bmatrix} 4,764 & 1,001 \\ 1,001 & 3,259 \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 9 & 3 \end{bmatrix} \begin{bmatrix} 4,764 & 1,001 \\ 1,001 & 3,259 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 3 \end{bmatrix} + \ln(0,5) =$$

$$= 1,813.\overbrace{Patrimonio\_Neto}^{x_1} + 0,364.\overbrace{Deuda\_Pendiente}^{x_2} - 9,396$$

Cada sujeto será asignado al grupo en el que obtenga un mayor valor de estas funciones.

La *función discriminante de Fisher* $\left[D - C = F_{II} - F_I\right]$:

$$\boxed{D - C = 1,035.Patrimonio\_Neto - 0,932.Deuda\_Pendiente - 3,520}$$

El programa SPSS no ofrece la función discriminante de Fisher.

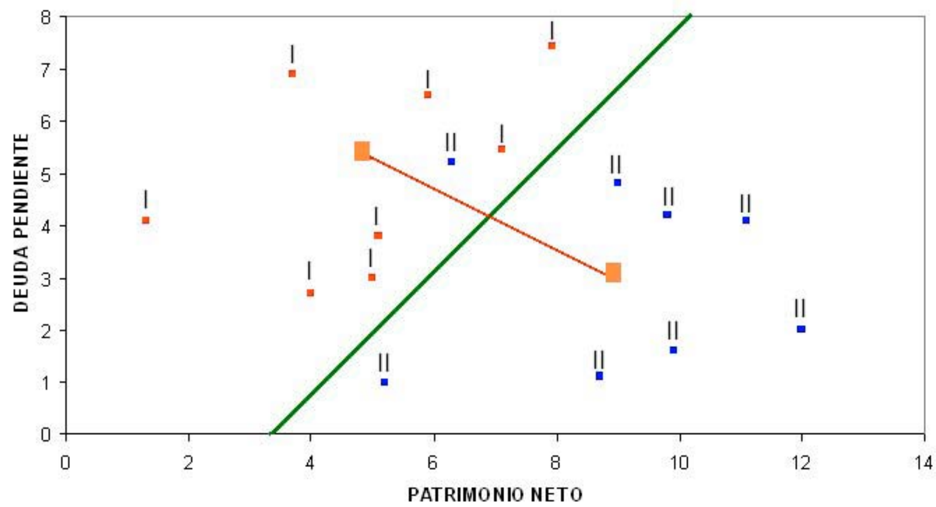
- *Estadísticos por casos:* Para cada caso, se muestran las puntuaciones discriminantes, las distancias de Mahalanobis de dichas puntuaciones al centroide de cada grupo y las probabilidades a posteriori obtenidas a partir de esas distancias.

Estadísticos por casos

	Número de caso	Grupo real	Grupo mayor: Grupo pronosticado	P(D>d | G=g): p	P(D>d | G=g): gl	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Segundo grupo mayor: Grupo	P(G=g | D=d)	Distancia de Mahalanobis al cuadrado hasta el centroide	Puntuaciones discriminantes: Función 1
Original	1	1	1	,222	1	,998	1,491	2	,002	13,479	-2,446
	2	1	1	,203	1	,998	1,617	2	,002	13,854	-2,497
	3	1	1	,447	1	,757	,578	2	,243	2,856	-,465
	4	1	1	,849	1	,970	,036	2	,030	6,972	-1,415
	5	1	1	,462	1	,769	,540	2	,231	2,942	-,490
	6	1	1	,651	1	,869	,204	2	,131	3,994	-,773
	7	1	1	,813	1	,919	,056	2	,081	4,901	-,989
	8	1	1	,618	1	,856	,249	2	,144	3,810	-,727
	9	2	2	,398	1	,717	,714	1	,283	2,577	,380
	10	2	2	,906	1	,938	,014	1	,062	5,439	1,107
	11	2	2	,494	1	,790	,468	1	,210	3,119	,541
	12	2	2	,099	1	,999	2,715	1	,001	16,795	2,873
	13	2	1**	,636	1	,863	,224	2	,137	3,909	-,752
	14	2	2	,551	1	,989	,355	1	,011	9,279	1,821
	15	2	2	,639	1	,985	,220	1	,015	8,523	1,694
	16	2	2	,361	1	,995	,833	1	,005	11,310	2,138

**. Caso mal clasificado

En este caso solo se ha encontrado un caso mal clasificado según la función lineal discriminante, se trata del grupo 2 (caso 13 en la tabla de estadísticos de clasificación) que ha sido incluido erróneamente dentro del grupo 1.



Como puede verse los dos centros de gravedad equidistan de la recta delimitadora.

📖 El director de la entidad financiera clasifica a las dos solicitudes de préstamos. Para ello, basta sustituir, en la *función discriminante de Fisher*, los valores de Patrimonio_Neto y Deuda_Pendiente:

$$D - C = 1,035.\text{Patrimonio\_Neto} - 0,932.\text{Deuda\_Pendiente} - 3,520$$

Primer solicitante: $D - C = 1,035.(10,1) - 0,932.(6,8) - 3,520 = 0,5959$

Segundo solicitante: $D - C = 1,035.(9,7) - 0,932.(2,2) - 3,520 = 4,469$

Como la puntuación es positiva en ambos casos, se clasifican a los dos solicitantes en el grupo de los cumplidores, si bien hay que hacer notar que el segundo solicitante tiene una puntuación discriminante mucho más elevada.

☞ CRITERIOS ALTERNATIVOS DE CLASIFICACIÓN: Existen otros muchos criterios de clasificación. Entre ellos, destacar el análisis de regresión y la aplicación de la distancia de Mahalanobis. A continuación se indican sus rasgos básicos, así como su relación con el análisis discriminante de Fisher.

- ANÁLISIS DE REGRESIÓN: La relación entre el análisis discriminante y el análisis de regresión es muy estrecha. Si se realiza una ajuste por mínimos cuadrados, tomando como variable dependiente la variable dicotómica que define la pertenencia a uno u otro grupo y como variables explicativas a las variables clasificadoras, se obtienen unos coeficientes que tienen una estricta proporcionalidad con los coeficientes de la función discriminante de Fisher,

  A partir del coeficiente de determinación, que se calcula en el análisis de regresión, se puede pasar con facilidad a la distancia de Mahalanobis entre los dos centroides de los dos grupos.

- DISTANCIA DE MAHALANOBIS (1936): Es una generalización de la distancia euclídea, que tiene en cuenta la matriz de covarianzas intra-grupos. El cuadrado de la distancia de Mahalanobis (DM_{ij}^2) entre los grupos i y j en un espacio de p dimensiones, siendo (V_w) la matriz de covarianzas intra-grupos, viene definida de forma: $DM_{ij}^2 = (x_i - x_j)' \, V_w^{-1} \, (x_i - x_j)$

  donde los vectores x_i y x_j representan dos puntos en el espacio p dimensional. En la terminología usual para designar esta distancia se prescinde de la M (introducida para evitar confusiones con las puntuaciones discriminantes a las que se ha designado por D).

  El cuadrado de la distancia euclídea d_{ij}^2 entre los puntos (i, j) viene dado por la expresión:

$$d_{ij}^2 = (x_i - x_j)\,(x_i - x_j) = \sum_{h=1}^{p} (X_{ih} - X_{jh})^2$$

  La distancia euclídea es el caso particular de la distancia de Mahalanobis en la que $(V_w = I)$. Es decir, la distancia euclídea no tiene en cuenta la dispersión de las variables y las relaciones existentes entre ellas, mientras que en la distancia de Mahalanobis sí que se descuentan estos factores al introducir en la expresión $DM_{ij}^2 = (x_i - x_j)' \, V_w^{-1} \, (x_i - x_j)$ la inversa de la matriz de covarianzas intra-grupos.

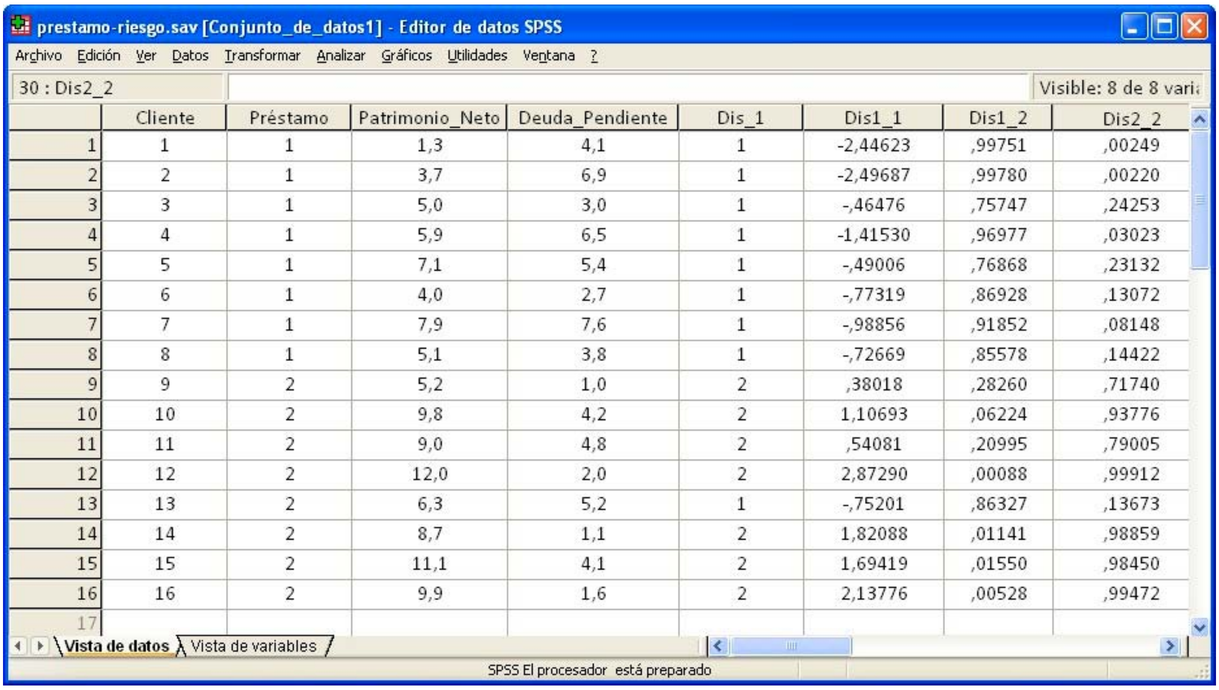
  Con el criterio de Mahalanobis, aplicando $DM_{ij}^2 = (x_i - x_j)' \, V_w^{-1} \, (x_i - x_j)$, se calcula la distancia entre cada punto y los dos centroides.

  Así, para el punto i-ésimo se obtienen estas dos distancias: $\begin{cases} DM_{i,I}^2 = (x_i - x_I)' \, V_w^{-1} \, (x_i - x_I) \\ DM_{i,II}^2 = (x_i - x_{II})' \, V_w^{-1} \, (x_i - x_{II}) \end{cases}$

  La aplicación de este criterio consiste en asignar cada individuo al grupo para el que la distancia de Mahalanobis es menor.

  La distancia de Mahalanobis clasifica a los individuos exactamente igual que lo hace la función discriminante de Fisher. La diferencia entre uno y otro tipo de procedimiento es que, mientras la distancia de Mahalanobis se calcula en el espacio de las variables originales, en el criterio de Fisher se sintetizan todas las variables en la función discriminante, que es la utilizada para realizar la clasificación.

☞ En el fichero *(prestamo-riesgo.sav)* se han guardado las columnas: Dis_1 (*Grupo pronosticado para el análisis 1*), Dis1_1 (*Puntuación discriminante de la función 1 para el análisis 1*), Dis1_2 (*Probabilidades de pertenencia al grupo 1 para el análisis 1*) y Dis2_2 (*Probabilidades de pertenencia al grupo 2 para el análisis 1*)



	Cliente	Préstamo	Patrimonio_Neto	Deuda_Pendiente	Dis_1	Dis1_1	Dis1_2	Dis2_2
1	1	1	1,3	4,1	1	-2,44623	,99751	,00249
2	2	1	3,7	6,9	1	-2,49687	,99780	,00220
3	3	1	5,0	3,0	1	-,46476	,75747	,24253
4	4	1	5,9	6,5	1	-1,41530	,96977	,03023
5	5	1	7,1	5,4	1	-,49006	,76868	,23132
6	6	1	4,0	2,7	1	-,77319	,86928	,13072
7	7	1	7,9	7,6	1	-,98856	,91852	,08148
8	8	1	5,1	3,8	1	-,72669	,85578	,14422
9	9	2	5,2	1,0	2	,38018	,28260	,71740
10	10	2	9,8	4,2	2	1,10693	,06224	,93776
11	11	2	9,0	4,8	2	,54081	,20995	,79005
12	12	2	12,0	2,0	2	2,87290	,00088	,99912
13	13	2	6,3	5,2	1	-,75201	,86327	,13673
14	14	2	8,7	1,1	2	1,82088	,01141	,98859
15	15	2	11,1	4,1	2	1,69419	,01550	,98450
16	16	2	9,9	1,6	2	2,13776	,00528	,99472
17								

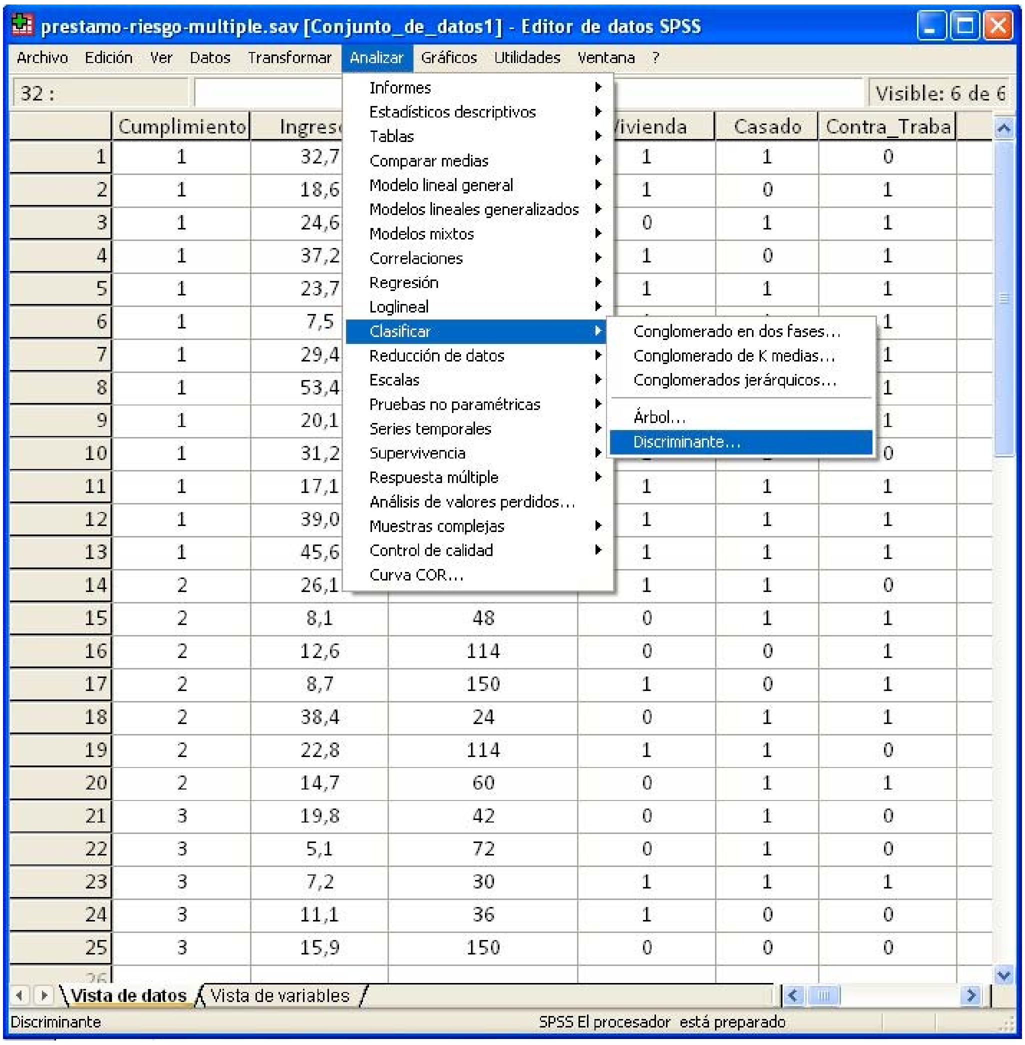
CONCESIÓN PRÉSTAMOS - RIESGO

Un banco realiza un estudio con el objetivo de identificar con la mayor precisión posible aquellas solicitudes de préstamos que probablemente puedan llegar a convertirse en morosos o fallidos en el caso que se concedieran. Para ello, dispone de la información reflejada en la tabla adjunta, relativa a 25 clientes y a las variables que se analizan:

- ☞ *Cumplimiento:* Grado de cumplimiento del cliente en el reintegro del préstamo. Toma el valor 1 si el cliente es cumplidor, 2 si es moroso y 3 si es fallido.
- ☞ *Ingresos:* Ingresos anuales del cliente, en miles de euros.
- ☞ *Patrimonio Neto:* Patrimonio neto del cliente en miles de euros.
- ☞ *Vivienda:* Variable dicotómica que toma el valor 1 si el cliente es propietario; 0 en caso contrario.
- ☞ *Casado:* Variable dicotómica que toma el valor 1 si está casado; 0 en otro caso.
- ☞ *Contrato Trabajo:* Variable dicotómica que toma el valor 1 si el cliente es asalariado con contrato fijo; 0 en otro caso.

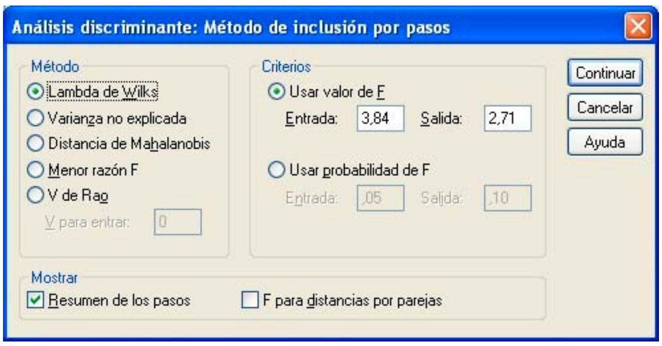
Cliente	Cumplimiento	Ingresos	Patrimonio neto	Vivienda	Casado	Contrato trabajo
1	1	32,7	336	1	1	0
2	1	18,6	204	1	0	1
3	1	24,6	138	0	1	1
4	1	37,2	270	1	0	1
5	1	23,7	114	1	1	1
6	1	7,5	132	1	1	1
7	1	29,4	90	0	1	1
8	1	53,4	228	1	1	1
9	1	20,1	324	0	1	1
10	1	31,2	480	1	1	0
11	1	17,1	108	1	1	1
12	1	39	132	1	1	1
13	1	45,6	216	1	1	1
14	2	26,1	234	1	1	0
15	2	8,1	48	0	1	1
16	2	12,6	114	0	0	1
17	2	8,7	150	1	0	1
18	2	38,4	24	0	1	1
19	2	22,8	114	1	1	0
20	2	14,7	60	0	1	1
21	3	19,8	42	0	1	0
22	3	5,1	72	0	1	0
23	3	7,2	30	1	1	1
24	3	11,1	36	1	0	0
25	3	15,9	150	0	0	0

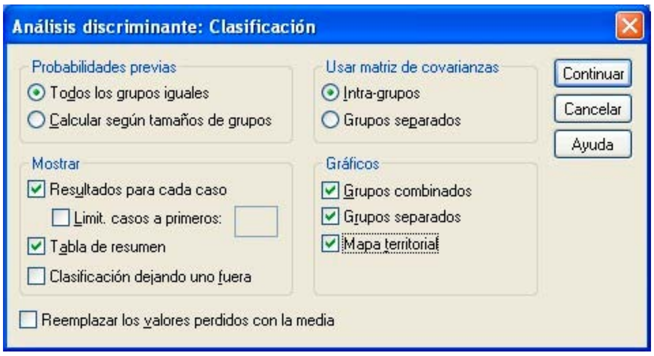
Se trata de un Análisis discriminante múltiple, ya que el banco ha clasificado a los clientes en *tres grandes grupos*, habrá que construir funciones discriminantes que permitan clasificar, con los menores errores posibles, a los clientes en los diferentes grupos. Si se obtienen buenos resultados, estas funciones discriminantes se podrán utilizar para analizar si se concede un préstamo o no a un futuro cliente peticionario.



Se selecciona *Cumplimiento* como variable de agrupación (cuyo rango es 1 y 3) y las otras cinco variables como independientes.

El método de inclusión por pasos.





El Visor de resultados de SPSS muestra:

Estadísticos de grupo

Cumplimiento		Media	Desv. típ.	N válido (según lista) No ponderados	Ponderados
Cliente Cumplidor	Ingresos	29,238	12,5658	13	13,000
	Patrimonio_Neto	213,231	114,3840	13	13,000
	Vivienda	,769	,4385	13	13,000
	Casado	,846	,3755	13	13,000
	Contra_Traba	,846	,3755	13	13,000
Cliente moroso	Ingresos	18,771	10,9939	7	7,000
	Patrimonio_Neto	106,286	71,4743	7	7,000
	Vivienda	,429	,5345	7	7,000
	Casado	,714	,4880	7	7,000
	Contra_Traba	,714	,4880	7	7,000
Cliente fallido	Ingresos	11,820	6,0693	5	5,000
	Patrimonio_Neto	66,000	49,6588	5	5,000
	Vivienda	,400	,5477	5	5,000
	Casado	,600	,5477	5	5,000
	Contra_Traba	,200	,4472	5	5,000
Total	Ingresos	22,824	12,9464	25	25,000
	Patrimonio_Neto	153,840	111,3896	25	25,000
	Vivienda	,600	,5000	25	25,000
	Casado	,760	,4359	25	25,000
	Contra_Traba	,680	,4761	25	25,000

Las medias de las cinco variables introducidas como independientes en el análisis son mayores en la categoría de cumplidores que en las otras categorías.
Así, los clientes cumplidores, en relación con los otros dos grupos (morosos, fallidos), tienen mayores ingresos, un mayor patrimonio, son propietarios de la vivienda que habitan están casados y son asalariados con contrato fijo.

Las ANOVAs indican que no se observan diferencias significativas entre los cumplidores, morosos y fallidos, en cuanto al hecho de ser propietario o no de la vivienda (Vivienda) y el estar casado o no (Casado).

Pruebas de igualdad de las medias de los grupos

	Lambda de Wilks	F	gl1	gl2	Sig.
Ingresos	,688	4,990	2	22	,016
Patrimonio_Neto	,663	5,584	2	22	,011
Vivienda	,870	1,639	2	22	,217
Casado	,948	,609	2	22	,553
Contra_Traba	,721	4,262	2	22	,027

En consecuencia, las variables (Vivienda) y (Casado) no deberían tener una gran influencia a la hora de clasificar a los clientes en uno u otro grupo. Obsérvese que en ambos casos, p_valor > 0,05, se acepta la hipótesis nula, es decir, los grupos en media son iguales.

En la siguiente tabla se observa el contraste de la *Prueba de Box* para determinar si es aceptable o no la hipótesis de homocedasticidad. Primero aparece el logaritmo del determinante de las matrices de covarianzas de los residuos de cada celda, calculadas según la expresión $S_g = \frac{V_g}{n_g - 1}$ (la matriz S_g es una estimación de la matriz de covarianzas correspondiente a la celda g-ésima Σ_g), y la matriz de covarianzas global, calculada según la expresión $\overline{S} = \dfrac{\sum_{g=1}^{G} V_g}{n-G} = \dfrac{\sum_{g=1}^{G}(n_g-1)S_g}{n-G}$ (donde $\overline{S}$ es una estimación de la matriz de covarianzas global Σ), así como el rango de cada una de estas matrices.

Las matrices son de orden 5x5, ya que existen cinco variables clasificadoras.

Si las matrices son no singulares (tienen inversa) su rango debe de ser 5. Se observa, en este caso, que la matriz correspondiente al grupo 3 (cliente fallido) no se calcula porque existen muy pocos casos para ser no singular, en efecto se puede observar que el número de individuos que pertenecen al grupo 3 (clientes fallidos) es justamente 5 y con este tamaño la matriz de covarianzas de los residuos es necesariamente singular.

Análisis 1

Prueba de Box sobre la igualdad de las matrices de covarianza

Logaritmo de los determinantes

Cumplimiento	Rango	Logaritmo del determinante
Cliente Cumplidor	5	7,623
Cliente moroso	5	6,076
Cliente fallido	.[a]	.[b]
Intra-grupos combinada	5	8,345

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

a. Rango < 5

b. Muy pocos casos para ser no-singular

Resultados de la prueba[a]

M de Box		40,339
F	Aprox.	1,693
	gl1	15
	gl2	604,127
	Sig.	,048

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

a. Algunas matrices de covarianzas son singulares y el procedimiento ordinario no es válido. Los grupos no singulares se compararán con sus propias matrices de covarianzas intra-grupo combinadas. El logaritmo de su determinante es 9,348.

Debido a que la matriz del grupo 3 (fallidos) es singular, SPSS contrasta la igualdad de las matrices de covarianzas poblacionales en los grupos 1 y 2, respectivamente, cliente cumplidores y morosos, estimando la matriz de covarianzas global con los datos de estos dos grupos. El nivel de significación crítico que se obtiene en este contraste es 0,048, con lo que se acepta la hipótesis nula para un nivel de significación del 1% (0,048 > 0,01), pero no para un nivel del 5% (0,048 < 0,05, rechazándose la hipótesis nula).

Resumen de las funciones canónicas discriminantes

Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	2,264[a]	98,1	98,1	,833
2	,043[a]	1,9	100,0	,203

a. Se han empleado las 2 primeras funciones discriminantes canónicas en el análisis.

Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 2	,294	26,343	4	,000
2	,959	,909	1	,340

En la tabla de Lambda de Wilks se aplica el contraste de significación para el conjunto de los dos ejes discriminantes. El contraste V de Barlett que se aplica es:

$$V_j = \left[n-1-\frac{K+G}{2}\right] \sum_{g=j+1}^{G-1} \ln(1+\lambda_g) \quad \text{donde} \quad j=0,1$$

$$V_0 = \left[n-1-\frac{K+G}{2}\right] \left[\ln(1+\lambda_1)+\ln(1+\lambda_2)\right] = \left[25-1-\frac{2+3}{2}\right] \left[\ln(1+2{,}264)+\ln(1+0{,}043)\right] = 26{,}343$$

Los grados de libertad de la chi-cuadrado son $K(G-1)=2(3-1)=4$ y el nivel de significación crítico es 0,000 < 0,05 rechazando, por tanto, la hipótesis nula, lo que significa que al menos uno de los ejes discriminantes es significativo, es decir, el primer eje discriminante es significativo (es el que tiene mayor poder discriminante). Adviértase que si no se rechaza la hipótesis nula no debería continuar el análisis.

Obsérvese que se cumple la relación entre la landa de Wilks y las raíces características (autovalores):

$$\Lambda = \frac{1}{(1+\lambda_1)\,(1+\lambda_2)} = \frac{1}{(1+2{,}264)\,(1+0{,}043)} = 0{,}294$$

Una vez determinada la significatividad del primer eje discriminante, se contrasta la significatividad de los restantes, en este caso, del segundo eje discriminante. El contraste a aplicar es el siguiente:

$$V_1 = \left[n-1-\frac{K+G}{2}\right]\left[\ln(1+\lambda_2)\right] = \left[25-1-\frac{2+3}{2}\right]\left[\ln(1+0{,}043)\right] = 0{,}909$$

Los grados de libertad de la chi-cuadrado son $(K-1)(G-1-1)=(2-1)(3-1-1)=1$ (en el análisis no entran 3 variables clasificadoras) y el nivel de significación crítico es 0,340 > 0,05, aceptando la hipótesis nula, lo que significa que el segundo eje discriminante no es significativamente distinto de 0 para cualquiera de los niveles de significación usuales.

La relación entre la landa de Wilks (obtenida después de excluir la primera función discriminante) y la segunda raíz característica (segundo autovalor) es la siguiente:

$$\Lambda_1 = \frac{1}{(1+\lambda_2)} = \frac{1}{(1+0{,}043)} = 0{,}959$$

Como información complementaria, se calcula la correlación canónica de cada función discriminante con la variable categórica que define los grupos, obteniéndose:

$$\eta_1 = \sqrt{\frac{\lambda_1}{1+\lambda_1}} = \sqrt{\frac{2{,}264}{1+2{,}264}} = 0{,}833 \qquad \eta_2 = \sqrt{\frac{\lambda_1}{1+\lambda_1}} = \sqrt{\frac{0{,}043}{1+0{,}043}} = 0{,}203$$

Los resultados obtenidos confirman que la capacidad explicativa de la segunda función discriminante es muy inferior a la primera. Una confirmación final de esta conclusión es que el porcentaje de varianza explicada con la primera función discriminante es del 98,1%, mientras que la varianza explicada con la segunda función discriminante es del 1,9%. Con lo que a efectos prácticos se podría prescindir de la segunda función discriminante, sin que afectase de forma importante a los resultados de la clasificación.

Coeficientes estandarizados de las funciones discriminantes canónicas

	Función	
	1	2
Patrimonio_Neto	1,193	-,511
Contra_Traba	1,154	,594

Matriz de estructura

	Función	
	1	2
Casado [a]	-,332*	,016
Contra_Traba	,394	,919*
Patrimonio_Neto	,457	-,889*
Vivienda[a]	-,020	-,233*
Ingresos [a]	-,033	-,209*

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.

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

a. Esta variable no se emplea en el análisis.

COEFICIENTES ESTANDARIZADOS: Aparecen los coeficientes de la función discriminante canónica estandarizados (media 0 y desviación típica 1), de esta forma se evitan los problemas de escala que pudieran existir entre las variables y, en consecuencia, la magnitud de los coeficientes estandarizados son un indicador de la importancia que tiene esta variable en el cálculo de la función discriminante.

MATRIZ DE ESTRUCTURA: Conviene conocer cuáles son las variables que tienen mayor poder discriminante en orden a clasificar a un individuo en uno de los grupos (cumplidor, moroso, fallido). Una forma de medir ese poder discriminante es calculando el coeficiente de correlación entre cada una de las variables y la función discriminante. Con un asterisco se indica el coeficiente más grande (en valor absoluto) que tiene cada variable.
Así, la variable *Casado* tienen su mayor coeficiente con la función discriminante 1, mientras que las variables *Contrato_Trabajo* e *Ingresos* lo tienen con la función discriminante 2.

Coeficientes de las funciones canónicas discriminantes

	Función	
	1	2
Patrimonio_Neto	,013	-,005
Contra_Traba	2,734	1,406
(Constante)	-3,796	-,126

Coeficientes no tipificados

Funciones en los centroides de los grupos

	Función	
Cumplimiento	1	2
Cliente Cumplidor	1,202	-,087
Cliente moroso	-,505	,305
Cliente fallido	-2,418	-,201

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

Aparecen las puntuaciones de los centroides de los grupos (*Patrimonio_Neto*, *Contrato_Trabajo*) con respecto a las funciones discriminantes (conviene darse cuenta que en este caso no hay un punto de corte discriminante, pues el conjunto de datos se encuentra separado en tres grupos).

Ahora falta calcular el valor de tres funciones de clasificación, y se clasificará a cada individuo en aquél grupo cuya función discriminante resulte tomar el mayor valor.

Coeficientes de la función de clasificación

	Cumplimiento		
	Cliente Cumplidor	Cliente moroso	Cliente fallido
Patrimonio_Neto	,063	,039	,018
Contra_Traba	13,721	9,604	3,662
(Constante)	-13,590	-6,607	-2,051

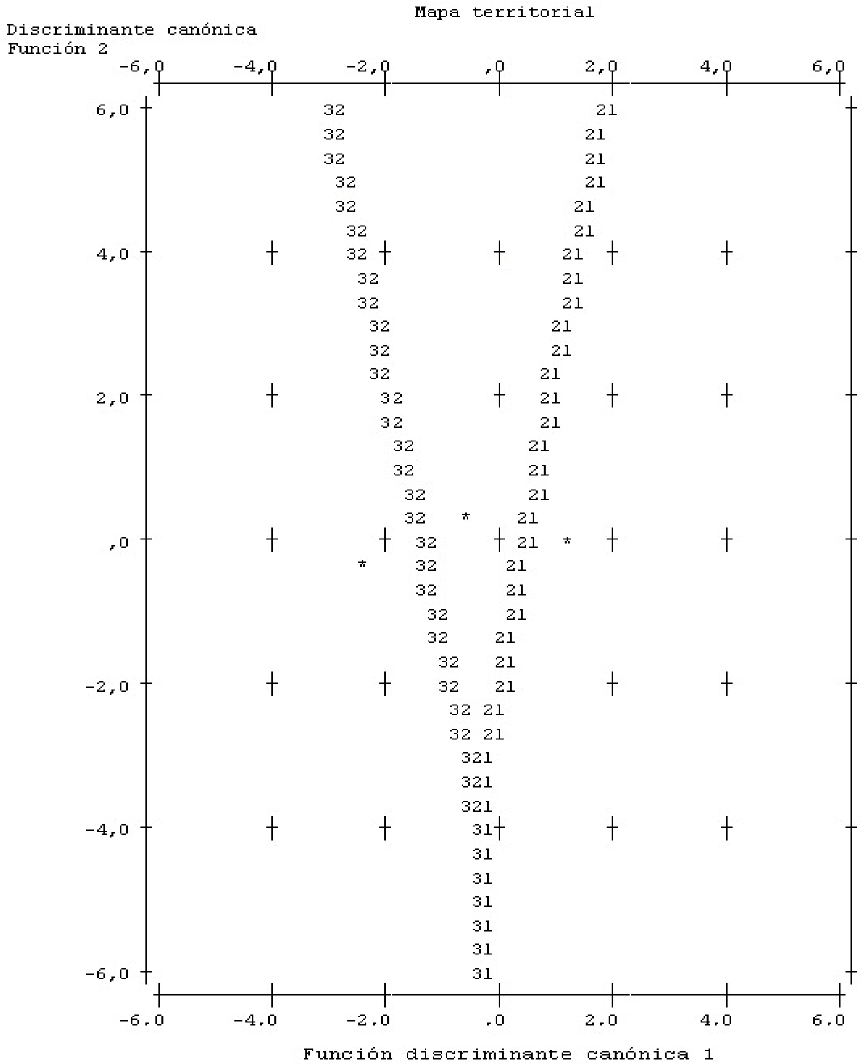
Funciones discriminantes lineales de Fisher

De esta forma, las funciones de clasificación son:

$$\begin{cases} F_I = 0{,}063.\text{Patrimonio\_Neto} + 13{,}721.\text{Contrato\_Trabajo} - 13{,}590 & \text{(cliente cumplidor)} \\ F_{II} = 0{,}039.\text{Patrimonio\_Neto} + 9{,}604.\text{Contrato\_Trabajo} - 6{,}607 & \text{(cliente moroso)} \\ F_{III} = 0{,}018.\text{Patrimonio\_Neto} + 3{,}662.\text{Contrato\_Trabajo} - 2{,}051 & \text{(cliente fallido)} \end{cases}$$

Para su aplicación, se calcula la puntuación de cada individuo en cada uno de los grupos, utilizando las funciones clasificadoras. Finalmente, un individuo se clasifica en el grupo en el que ha alcanzado la puntuación más elevada.

El *mapa territorial* sirve para ver cómo quedan la clasificación en función de las dos funciones lineales discriminantes:




```
Símbolos usados en el mapa territorial

Símbol  Grupo  Etiqu
------  -----  --------------------

   1       1   Cliente Cumplidor
   2       2   Cliente moroso
   3       3   Cliente fallido
   *           Indica un centroide de grupo
```

El mapa territorial delimita, en el plano de las dos funciones discriminantes (no estandarizadas), las áreas que se asignan a cada grupo. El área situada en la parte derecha de la función discriminante 1 es la correspondiente al grupo 1, mientras que el área de la izquierda corresponde al grupo 3. Se clasifican en el grupo 2 los individuos con puntuaciones discriminantes canónicas situadas en el triángulo de la parte central.
La salida de SPSS recoge el cálculo de probabilidades a posteriori, puntuaciones discriminantes y resultados de la clasificación. En este caso, no aparece la columna etiquetada con (*valores faltantes*) donde se refleja casos o individuos para los que no se dispone de información completa. Aparece la columna *Grupo real* de pertenencia y *Grupo pronosticado*, que cuando aparece con un asterisco refleja que el individuo a que corresponda se le clasifica de forma errónea.

Las columnas siguientes son relativas al cálculo de probabilidades. Las probabilidades *a posteriori* P(G/D) se calculan para cada grupo con la fórmula:

$$\mathrm{Prob}(g/D) = \frac{\pi_I \, e^{F_g}}{\pi_I \, e^{F_I} + \pi_{II} \, e^{F_{II}}} \quad g = I, II \quad \begin{cases} g \equiv \text{grupo} \\ \pi_i \equiv \text{probabilidad a priori} \end{cases} \qquad \text{(extendida a tres variables)}$$

Con este criterio se clasifica a un individuo en el grupo I si: $F_I \ln \pi_I > F_{II} \ln \pi_{II}$.
La aplicación de este criterio implica que el punto de corte discriminante C_p viene definido por:

Punto de corte con información a priori: $\boxed{C_p = \dfrac{\overline{D}_I + \overline{D}_{II}}{2} - \ln \dfrac{\pi_{II}}{\pi_I}}$

En la salida del SPSS se indica la probabilidad a posteriori más alta con indicación al grupo a que corresponde y la segunda probabilidad más alta con indicación del grupo. Junto a la probabilidad más alta aparece la probabilidad de la puntuación discriminante P(D/G), que no tiene interés especial en el análisis.

Las dos últimas columnas se refieren a las puntuaciones discriminantes. Cada una de ellas corresponde a una función discriminante. En SPSS estas puntuaciones se calculan utilizando los coeficientes de las funciones discriminantes canónicas no estandarizadas.

Estadísticos por casos

			Grupo mayor					Segundo grupo mayor			Puntuaciones discriminantes	
				P(D>d | G=g)			Distancia de Mahalanobis al cuadrado hasta el centroide			Distancia de Mahalanobis al cuadrado hasta el centroide		
	Número de caso	Grupo real	Grupo pronosticado	p	gl	P(G=g | D=d)		Grupo	P(G=g | D=d)		Función 1	Función 2
Original	1	1	1	,134	2	,706	4,018	2	,274	5,914	,434	-1,938
	2	1	1	,922	2	,875	,163	2	,125	4,062	1,506	,179
	3	1	1	,717	2	,593	,664	2	,401	1,447	,675	,535
	4	1	1	,523	2	,971	1,297	2	,029	8,311	2,337	-,177
	5	1	2**	,637	2	,537	,901	1	,451	1,252	,373	,665
	6	1	1	,673	2	,558	,791	2	,435	1,290	,600	,568
	7	1	2**	,751	2	,661	,572	1	,315	2,055	,071	,794
	8	1	1	,824	2	,925	,387	2	,075	5,418	1,809	,050
	9	1	1	,179	2	,992	3,441	2	,008	13,004	3,017	-,468
	10	1	1	,018	2	,987	8,001	2	,013	16,694	2,248	-2,715
	11	1	2**	,671	2	,570	,798	1	,415	1,432	,298	,697
	12	1	1	,673	2	,558	,791	2	,435	1,290	,600	,568
	13	1	1	,883	2	,903	,248	2	,097	4,713	1,658	,115
	14	2	2	,225	2	,533	2,985	3	,343	3,869	-,850	-1,388
	15	2	2	,773	2	,790	,515	1	,140	3,981	-,458	1,021
	16	2	2	,637	2	,537	,901	1	,451	1,252	,373	,665
	17	2	1**	,798	2	,660	,452	2	,336	1,801	,827	,471
	18	2	2	,677	2	,800	,780	3	,120	4,576	-,760	1,150
	19	2	3**	,863	2	,892	,295	2	,107	4,537	-2,361	-,741
	20	2	2	,793	2	,767	,464	1	,180	3,363	-,307	,956
	21	3	3	,689	2	,975	,744	2	,025	8,063	-3,267	-,353
	22	3	3	,852	2	,953	,320	2	,047	6,357	-2,890	-,515
	23	3	2**	,707	2	,802	,693	3	,106	4,746	-,684	1,118
	24	3	3	,648	2	,978	,869	2	,022	8,444	-3,343	-,320
	25	3	3	,670	2	,789	,800	2	,204	3,504	-1,907	-,935

**. Caso mal clasificado

Estadísticos por caso: Para cada caso, se muestran las puntuaciones discriminantes, las distancias de Mahalanobis de dichas puntuaciones al centroide de cada grupo y las probabilidades a posteriori obtenidas a partir de esas distancias.

Se observa que hay seis casos mal clasificados, comprobándose como las probabilidades de pertenencia son mayores para la pertenencia al grupo mayor, y también que las puntuaciones discriminantes son las que sitúan a cada caso en el mapa territorial.

Los resultados de la investigación son satisfactorios, ya que contiene un porcentaje elevado de clientes clasificados satisfactoriamente (76%), si bien preocupa el caso de un cliente moroso (cliente 17) que ha sido calificado como cumplidor. Este tipo de error de clasificación tiene mucha importancia, el banco se preocupa sobre todo que un cliente moroso o fallido pueda ser considerado como cumplidor, pues el coste de una clasificación errónea de este tipo es elevado para la entidad.

Hoja nº ....

UAM
Universidad Autónoma
de Madrid

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

UNICAMENTE PARA USO ESCOLAR (IMPRESO EN PAPEL RECICLADO)

Hoja nº ....

Universidad Autónoma de Madrid

Asignatura .... Grupo ....

Apellidos .... Nombre ....

Ejercicio del día ....

Instrumentos Estadísticos Avanzados
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández