Análisis discriminante.

INTRODUCCIÓN

Las técnicas de análisis discriminante se proponen la determinación de un criterio que nos permita decidir a qué grupo pertenece un cierto individuo, a partir de la información disponible sobre él cifrada en términos de los valores que toman ciertas variables consideradas.

A diferencia del análisis cluster, no pretendemos clasificar a los individuos en grupos, procediendo nosotros mismos a su agrupación, sino que, aquí, los grupos ya están constituidos y lo que buscamos es descubrir qué tiene de específico cada grupo para ser capaces de asignar de manera correcta los individuos a los grupos. Se trata, en definitiva, de discriminar a qué grupo pertenece, o debería pertenecer cada individuo.

Se dispone, por tanto, de un conjunto de datos relativos a unos elementos o individuos, pertenecientes a dos o más grupos definidos " a priori", de manera precisa. El objetivo inicial es encontrar si ciertas variables, de entre las que se dispone, nos pueden permitir asignar correctamente a los individuos a las clases a las que pertenecen. El fin último es ser capaces de asignar un "individuo anónimo " (del que se desconoce su afiliación) a la clase o grupo al que pertenece.

En consecuencia, podemos considerar que la discriminación se plantea dos objetivos distintos aunque íntimamente relacionados; y, en ese sentido, cabe decir que admite dos enfoques: Una discriminación entendida como decisión de adscripción a un grupo de un individuo ; y una discriminación entendida como una descripción de la estructura diferenciadora de los grupos.

Cuando se habla de discriminación suele hacerse referencia al problema de la identificación y se la considera como una técnica de toma de decisiones. Éste fue, de hecho, el principal enfoque que se dio a los primeros trabajos de Fisher, Wald, Anderson, Welch, von Mises etc.

En estas páginas dedicaremos un breve espacio al *problema de la identificación* (discriminación como decisión) y nos centraremos luego en el enfoque descriptivo, haciendo especial hincapié en el *análisis factorial discriminante*.

ANÁLISIS DISCRIMINANTE

SITUACIÓN DE APLICACIÓN:

SE DISPONE DE N INDIVIDUOS CLASIFICADOS EN g GRUPOS
SE HA RECABADO INFORMACIÓN SOBRE n VARIABLES (CUYOS VALORES CARACTERIZAN A CADA UNO DE LOS INDIVIDUOS)

PROBLEMA A RESOLVER: LA DISCRIMINACIÓN DE LOS GRUPOS

A) OBTENER UN CRITERIO PARA ASIGNAR UN INDIVIDUO ANÓNIMO A UNO U OTRO GRUPO

B) DETERMINAR LA ESTRUCTURA DIFERENCIADORA (DISCRMINANTE) DE LOS GRUPOS

QUÉ VBLES. DISCRIMINAN MEJOR
QUÉ COMB. LINEAL. DE ESAS VBLES. DISCRIMINA ÓPTIMAMENTE

A) DISCRIMINACIÓN COMO "DECISIÓN":

ESTUDIAR LA SIMILARIDAD DE INDIVIDUOS DEL MISMO GRUPO Y LA DISIMILARIDAD DE INDIVIDUOS DE DISTINTOS GRUPOS. (DUAL DE A.CLUSTER)
MINIMIZAR ERROR DE CLASIFICACIÓN /REGLA BAYES -MINIMIZAR F.PERDIDA.

B) DISCRIMINACIÓN COMO "DESCRIPCIÓN"(A.F.D.)

1) UTILIZAR EL MANOVA COMO CRITERIO DE SELECCIÓN ALGORITMICA DE LAS VARIABLES MÁS DISCRIMINANTES
2) FACTORIZAR LA SOLUCIÓN ANTERIOR PARA DETERMINAR LOS EJES DEL ESPACIO DE LAS VARIABLES QUE MEJOR DISCRIMINAN LOS GRUPOS

HABITUALMENTE LOS PAQUETES ESTADÍSTICOS COMBINAN LOS DOS ESQUEMAS DE ACTUACIÓN SEGÚN EL SIGUIENTE PROCESO: DETERMINACIÓN DE VARIABLES MÁS DISCRIMINANTES/ FACTORIZACIÓN/ DETERMINACIÓN DE DISCRIMINADORES/RECLASIFICACIÓN.

DISCRIMINACIÓN COMO DECISIÓN

DISCRIMINACIÓN COMO IDENTIFICACIÓN (PERSPECTIVA DECISIONAL)

En esta perspectiva la discriminación se plantea como el problema de identificar (adscribir) un individuo anónimo.
La situación de nuestro problema de decisión es la siguiente:
Tenemos un individuo, **w**, que puede pertenecer a una de varias posibles poblaciones H_1 ,H_2 ,...,H_r. Pero no sabemos a cuál de ellas pertenece de hecho.

- Disponemos de un conjunto de n variables X_1 ,X_2 ,..X_n.

- Nuestro individuo nos viene caracterizado por los valores que toman para él las n variables. En consecuencia, podemos caracterizarlo por el vector de observaciones:

$$X(w)=\begin{pmatrix} x_1(w) \\ x_2(w) \\ . \\ . \\ x_n(w) \end{pmatrix}$$

- Nos planteamos la obtención de una regla de decisión que nos permita asignar al individuo w a una de las r poblaciones o grupos.

La ejecución de esta regla de decisión es lo que constituye la identificación o discriminación.

La solución del problema suele acometerse construyendo ciertas funciones (funciones = f (X)), llamadas discriminantes, con la esperanza de que estas funciones nos definan sobre R^n una partición { R_1,R_2 ..,R_r } ,de modo que podamos adoptar el criterio de decisión de que si $X(w)\in R_i$, entonces $w\in H_i$.

Básicamente hay tres posiciones, (que bajo ciertas condiciones, presentan resultados similares) de afrontar el problema de la discriminación y la determinación de las funciones discriminantes :

A. Apoyarse para la construcción de las funciones discriminantes en la distancia entre el individuo w y los distintos grupos.
B. Apoyarse en el criterio de minimizar la probabilidad de error de clasificación (con o sin información inicial) o maximizar la verosimilitud de pertenencia al grupo.
C. Plantear la identificación como un típico problema de decisión cuya solución pasa por la minimización de la función de pérdida asociada a la "mala clasificación " (o misclassification).

Utilizando un criterio de discriminación geométrico, resulta razonable asignar el individuo anónimo a aquel grupo del que se encuentre más cercano. Lo habitual es considerar la distancia al centro de gravedad del grupo y utilizar una medida de la distancia que no esté sujeta ni a la escala de medida de las variables ni a su covariación o correlación , y, en este sentido, la distancia de Mahalanobis es la más adecuada.

Así pues si consideramos la distancia entre el individuo w (representado por el vector de observaciones X) y el grupo H_i, como la distancia de Mahalanobis entre X y el centro de gravedad del grupo i-simo M_i:

$$D(X,H_i) = (X - M_i)' V^{-1} (X - M_i)$$

El criterio de discriminación será asignar w al grupo H_i si:

$$D(X,H_i) = \min_s\{D(X,H_s)\}$$

Es fácil obtener una función discriminante que discrimine entre los grupos H_i y H_j:

$$W_{ij}(X) = (M_i - M_j)'V^{-1} X - 1/2\,(M_i - M_j)' V^{-1} (M_i + M_j)$$

que es una función lineal debida a Fisher y, posteriormente desarrollada, para el caso de disponer únicamente de información muestral ,por Wald y Anderson

De esta forma, si para un cierto i se cumple que :

$W_{ij}(X) > 0 \quad \forall\, i \neq j$ asignaremos el individuo anónimo al grupo H_i.

Otro criterio de discriminación basado en la distribución de las variables en los distintos grupos será el que asigne el individuo anónimo a aquel grupo en el que se maximice la función de verosimilitud.

Según este criterio asignaremos el individuo w al grupo H_i si :

$L_i(X) = \max_s\{L_s(X)\}$ (donde L es la función de verosimilitud).

En este caso la función discriminante para cada par de grupos H_i y H_j sería :

$V_{ij}(X) = \log L_i(X) - \log L_j(X)$

De forma que si para cierto i se cumple que:

$V_{ij}(X) > 0 \quad \forall\, i \neq j$ asignaremos el individuo anónimo al grupo H_i.

Si las poblaciones son normales multivariantes y las matrices de varianzas coinciden, el criterio máximo verosímil y el geométrico coinciden y el discriminador máximo verosímil asume la misma expresión que el discriminador lineal de Wald-Anderson.

Sin embargo, en el caso en el que cada población tenga una matriz de varianzas propia el discriminador adopta una expresión cuadrática. De forma que para cada dos grupos la función discriminante queda como:

$$Q_{ij}(X) = \frac{1}{2}X'(V_j^{-1} - V_i^{-1})X + X'(V_j^{-1}M_i - V_i^{-1}M_j)X + \frac{1}{2}M_j'V_j^{-1}M_j + \frac{1}{2}M_i'V_i^{-1}M_i + \frac{1}{2}\log|V_j| + \frac{1}{2}\log|V_i|$$

Welch fue el primero en proponer la discriminación siguiendo un criterio de minimizar la probabilidad de error de calsificación o bien, maximizar la verosimilitud de pertenencia a un grupo. Estudió también, la posibilidad de disponer de conocimiento a priori acerca de la pertenencia de un individuo a cada grupo y el modo de utilizarlo en la discriminación a través del uso del Teorema de Bayes.

Con información inicial el discriminador asociado era, entonces:

$B_{ij}(X) = \log L_i(X) - \log L_j(X) + \log(q_i/q_j)$

donde q_i y q_j son las probabilidades iniciales de que w pertenezca a H_i y H_j , respectivamente.

Otros muchos autores han hecho hincapié en las distintas consecuencias que pueden tener distintos errores de clasificación. En muchas situaciones prácticas distintos errores de clasificación pueden tener asociados costes muy distintos. En ese caso, no es el error de clasificación lo que se debe minimizar sino el coste esperado o decantarse por aquella alternativa que nos minimice el máximo coste u optar por algún otro criterio decisional adecuado.

DISCRIMINACIÓN COMO DESCRIPCIÓN-ANÁLISIS FACTORIAL-DISCRIMINANTE-

OBJETIVOS DEL ANÁLISIS FACTORIAL DISCRIMINANTE

Así pues, los objetivos que perseguimos con el análisis factorial discriminante son básicamente:

1. Seleccionar qué variables resultan más discriminantes de entre las disponibles.
2. Obtener los llamados factores o funciones discriminantes, que no son otra cosa que nuevas variables que serán combinaciones lineales de las variables discriminantes originales con un poder de discriminación óptimo.
3. Interpretar los factores discriminantes; lo que podrá hacerse a partir de las correlaciones de éstos con las variables originales, análogamente a como interpretábamos la estructura factorial en el Análisis Factorial. Aquí se trata, obviamente, de interpretar la estructura de diferenciación o estructura discriminante.
4. Ser capaces, por último, de clasificar a los individuos en base a los factores discriminantes obtenidos.

ANÁLISIS MULTIVARIANTE DE LA VARIANZA (MANOVA)

Si el análisis de la varianza univariante pretende contrastar hipótesis lineales sobre la influencia de los distintos niveles de uno o varios factores en el comportamiento de una variable (unidimensional), el análisis multivariante de la varianza (MANOVA) tiene la misma pretensión, pero considerando un vector (multidimensional) de variables.

La aplicación paradigmática del análisis de la varianza es determinar si existen diferencias significativas entre los distintos niveles o grupos de un factor (categórico), a través del contraste de igualdad de medias. Porque es, fundamentalmente, ésta la aplicación que nos interesa, aquí nos detendremos un poco en ella, para el caso multivariante del MANOVA, para luego poder extender sus resultados y consideraciones al análisis factorial discriminante.

Los supuestos del análisis serán los mismos en el caso del MANOVA que en el del análisis factorial discriminante y, en consecuencia, los mantendremos ya desde aquí:

a. Consideramos un vector aleatorio Y de dimensión n sobre el cuál obtenemos g muestras correspondientes a los g niveles, categorías o **grupos** considerados.
b. Suponemos que Y sigue, en cada una de las poblaciones de los g grupos una distribución Normal n-variante con vector de medias M (i= 1,2,...g),eventualmente distinto para cada grupo y matriz de covarianzas V, la misma para todas las poblaciones.

Bajo estos supuestos, consideraremos, también, que cada observación n-dimensional para cada grupo, i, puede expresarse de acuerdo con el siguiente modelo:

$$Y_i = M + A_i + E_i$$

Donde:
M es el vector de medias *general*.
A_i es un vector n-dimensional que nos indica el efecto propio del nivel o grupo i-simo.
E_i es un vector aleatorio que nos indica la desviación errática de las observaciones y se supone que sigue una distribución normal n-dimensional con vector de medias el vector nulo y matriz de varianzas V, la misma para todos los grupos (i=1,2,...g)
En estas circunstancias es fácil comprobar cómo el vector Y_i tendrá, en cada grupo, i, una distribución:

$$Y_i \rightarrow N\, [\, (M + A_i\,);\, V\,]$$

Sobre este modelo nos plantemos contrastar la hipótesis nula de que todos los vectores A sean nulos:

$$H_0\!: A_1 = A_2 = ... = A_g = 0$$

Esta hipótesis equivale a considerar que **no hay diferencias en los vectores de medias de Y en cada uno de los grupos** o que las medias en cada grupo son las mismas y coinciden el vector M .

Para la realización del contraste, partimos, como en el caso univariante, de la descomposición de la varianza total; en este caso de la matriz de varianzas y covarianzas total.

La matriz de varianzas muestrales T puede verse como la suma de otras dos matrices de varianzas:

$$T = B + W$$

Donde B es la matriz de varianzas "entre-grupos" (Between-groups) y W es la matriz de varianza "intra- grupos" (Within-groups).

B expresa las varianzas y covarianzas, considerando los centroides de los grupos como observaciones.

W, en cambio, expresa la suma para todos los grupos de las varianzas y covarianzas de las observaciones de cada grupo.

Pues bien, la matriz NB, donde N es el número total de observaciones muestrales, puede probarse que sigue una distribución de Wishart con parámetros n, g-1 , V (lo que se expresa como W_n(g-1, V)).

La distribución puede considerarse como una generalización de la distribución χ^2 de Pearson, que puede definirse de acuerdo con el siguiente esquema general:

Si tenemos una matriz de n columnas y m filas, Z; donde cada columna está formada por un vector aleatorio m-dimensional que tiene una distribución normal m-variante con vector de medias el vector nulo y matriz de varianzas V, la misma para todas las columnas de la matriz; entonces la matriz A = Z'Z sigue una distribución de Wishart de parámetros n, m y V [lo que puede expresarse como: W_n (m,V)]

Una propiedad importante de esta distribución es que si realizamos un muestreo aleatorio de tamaño N sobre una población normal multivariante N [M,V],la matriz formada por el producto del escalar N y la matriz de varianzas muestral,S, sigue una distribución de Wishart de parámetros n, N-1, V: NS $\rightarrow W_n$ (N-1, V)

Es, precisamente, a partir de esta propiedad como puede probarse el resultado de que :

$$NB \rightarrow W_n (g-1,V)$$

Igualmente puede probarse también que si la hipótesis nula: H_0: $A_1 = A_2 = ... = A_g = 0$ es cierta, entonces la matriz NW seguirá, también una distribución de Wishart de parámetros n, N-g, V y será independiente de la distribución de NB.

Obviamente también, considerando esa misma propiedad, NS (siendo S la matriz de varianzas totales muestral) seguirá también una distribución W_n (N-1,V)

Teniendo en cuenta ésto, el contraste de la hipótesis nula: H_0: $A_1 = A_2 = ... = A_g = 0$ se

lleva a cabo evaluando el valor del estadístico Λ (lambda de Wilks): $\Lambda = |W| \,/\, |T|$

Estadístico que sigue una distribución Λ de Wilks de parámetros n, N-g ,g-1 .

Es, precisamente este estadístico el que nos conducirá a determinar si los vectores de medias de los grupos son significativamente diferentes o no; es decir, si la hipótesis nula es rechazable o no:

Para un nivel de significación α:

- Aceptaremos la hipótesis nula si $\Lambda > \Lambda_\alpha$
- Rechazaremos la hipótesis nula si $\Lambda < \Lambda_\alpha$

Siendo Λ el valo crítico que verifica P ($\Lambda > \Lambda_\alpha$) = α en una distribución Λ (n, N-g,g-1).

En la práctica el contraste se realiza después de una transformación previa del estadístico en una F o una $\chi 2$

ANÁLISIS DE LA VARIANZA (UNIVARIANTE Y MULTIVARIANTE)

	VARIABLES INDIVIDUOS GRUPOS	H_o	INFORMACIÓN MUESTRAL (DESC. VARIAN.)	MODELO/ DISTRIBU. MUESTRALES	ESTADÍSTICO/ CRITERIO DE DECISIÓN
ANOVA	VBLES=1 IND=N= ΣN_i GRUPOS=g	$\mu_1=\mu_2=...=\mu_g$ (MEDIAS IGUALES, EN TODOS LOS GRUPOS)	VARIANZAS: -INTRA-GRUPO S^2_W ENTRE-GRUPOS S^2_B TOTAL S^2 $S^2 = S^2_W + S^2_B$	MODELO: $X \to N[\mu_i, \sigma^2]$ (σ^2 CONSTANTE EN TODOS LOS GRUPOS) DIS.MUEST. : $NS^2_W \to \chi^2_{N-g}$ $NS^2_B \to \chi^2_{g-1}$ $NS^2 \to \chi^2_{N-1}$	$F = \frac{NS^2_B}{NS^2_W} \frac{(N-g)}{(N-1)}$ $F \to F_{(g-1),(N-g)}$ SI: $F > F_\alpha$ RECHAZAMOS H_o (EXISTEN DIFERENCIAS SIGNIFICATIVAS)
MANOVA	VBLES= n IND=N= ΣN_i GRUPOS=g	$M_1 = M_2 = ... = M_g$ (VECTORES DE MEDIAS IGUALES , EN TODOS LOS GRUPOS)	MATRICES DE VARIANZAS: INTRA-GRUPO W ENTRE-GRUPOS B TOTAL T $T = W + B$	MODELO: $X \to N[M_i, V]$ (V CONSTANTE EN TODOS LOS GRUPOS) DIS.MUEST. : $NW \to W_n(N-g,V)$ $NB \to W_n(g-1,V)$ $NT \to W_n(N-1,V)$	$\Lambda = \frac{\lvert W \rvert}{\lvert T \rvert}$ $\Lambda \to \Lambda_{n,(N-g),(g-1)}$ SI: $\Lambda < \Lambda_\alpha$ RECHAZAMOS H_o (EXISTEN DIFERENCIAS SIGNIFICATIVAS)

DETERMINACION DE LAS VARIABLES MAS DISCRIMINANTES

Como se ha comentado, uno de los objetivos de la discriminación descriptiva es determinar cuáles son , de entre las variables disponibles, aquellas que tienen un mayor poder discriminativo. Es evidente que este objetivo es tanto más importante si el número de variables de las que se dispone información es muy elevado o si este número excede al número máximo de variables que pueden utilizarse en una factorización posterior.

Determinar qué variable es más discriminante y las que tienen, sucesivamente, mayor poder discriminativo puede hacerse a partir de un análisis de la varianza (univariante) para cada variable; obteniendo que una variable es tanto más discriminativa cuanto mayor sea su estadístico F.

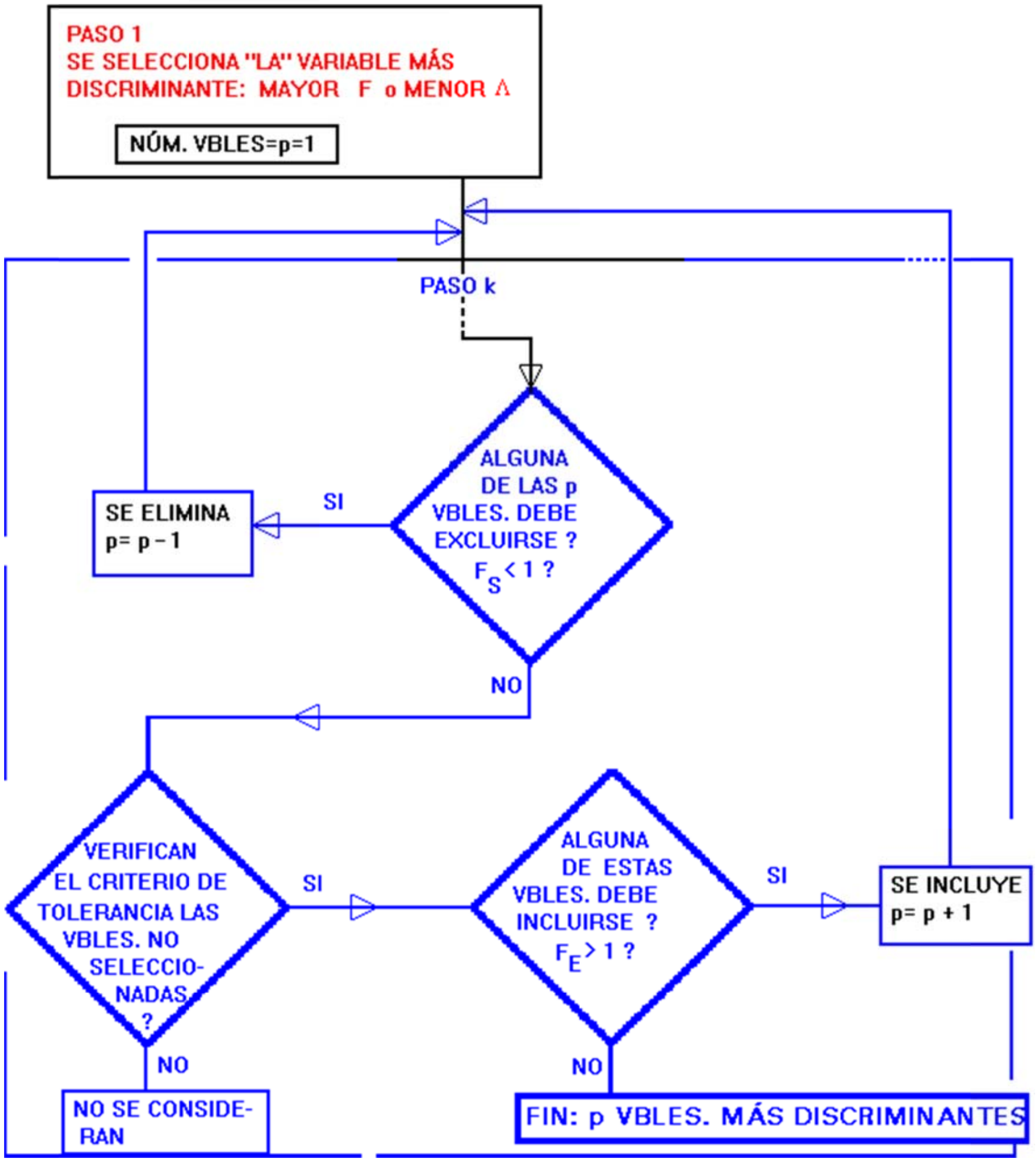
Si queremos determinar qué grupos de variables de un determinado tamaño es el más discriminativo, podemos considerar todos los posibles grupos de ese tamaño y realizar un análisis MANOVA , decantándonos por el grupo de variables que presente un menor valor de su estadístico Λ de Wilks.

Procediendo de esta manera, podemos, desgraciadamente, acabar seleccionando un conjunto de variables que resulten estar altamente correlacionadas entre sí, proporcionándonos, en consecuencia, una discriminación redundante. Para evitar este problema, los distintos algoritmos de discriminación "paso-a-paso " utilizan "criterios de tolerancia" para la selección de las variables como variables discriminantes

De todas manera, la determinación del conjunto de variables a considerar como variables discriminantes es, en la práctica, algo que se realiza a la vez que la determinación de los factores o ejes discriminantes utilizando los ya mencionados procedimientos paso a paso.

Especialmente nos centraremos en el procedimiento de selección y discriminación basado en la mínimización del estadístico Λ de Wilks, que es el que utilizaremos aquí.

ALGORITMO DE SELECCIÓN DE LAS (p ≪ n) VARIABLES MÁS DISCRIMINANTES



CRITERIOS DE SELECCION Y EXCLUSIÓN

SELECCIÓN: ENTRADA DE UNA NUEVA VARIABLE

TOLERANCIA + MEJOR DISCRIMINACIÓN

TOLERANCIA: $T_i = 1 - R_i^2$ (POR DEFECTO SPSS: $T > 0.001$)

SIENDO R_i^2 EL CUADRADO DEL COEF. CORR. MULT. X / VBLES. YA SELECC.

MEJORA DE LA DISCRIMINACIÓN: A PARTIR DE:
MANOVA SOBRE LAS p VBLES YA INCLUIDAS $\longrightarrow \Lambda_P$
MANOVA SOBRE LAS p VBLES YA INCLUIDAS + LA VBLE CANDIDATO $\longrightarrow \Lambda_{P+1}$

$$F_E = \left[\frac{N-g-p}{g-1}\right]\left[\frac{1-(\Lambda_{P+1}/\Lambda_P)}{(\Lambda_{P+1}/\Lambda_P)}\right]$$

(POR DEFECTO SPSS: $F_E > 1$)

EXCLUSIÓN: SALIDA DE UNA VARIABLE YA CONSIDERADA

MEJORA DE LA DISCRIMINACIÓN

MEJORA DE LA DISCRIMINACIÓN: A PARTIR DE:
MANOVA SOBRE LAS p VBLES YA INCLUIDAS $\longrightarrow \Lambda_P$
MANOVA SOBRE p - 1 VARIABLES (EXCLUYENDO LA CANDIDATA) $\longrightarrow \Lambda_{P-1}$

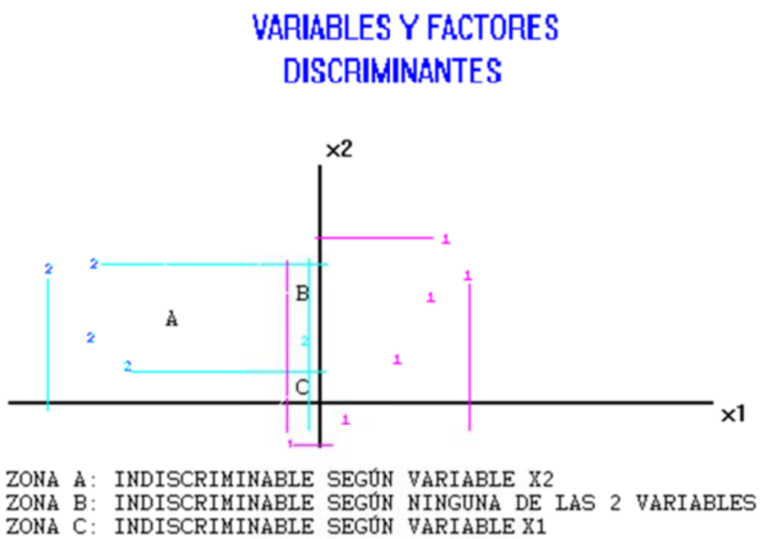
$$F_S = \left[\frac{N-g-p-1}{g-1}\right]\left[\frac{1-(\Lambda_P/\Lambda_{P-1})}{(\Lambda_P/\Lambda_{P-1})}\right]$$

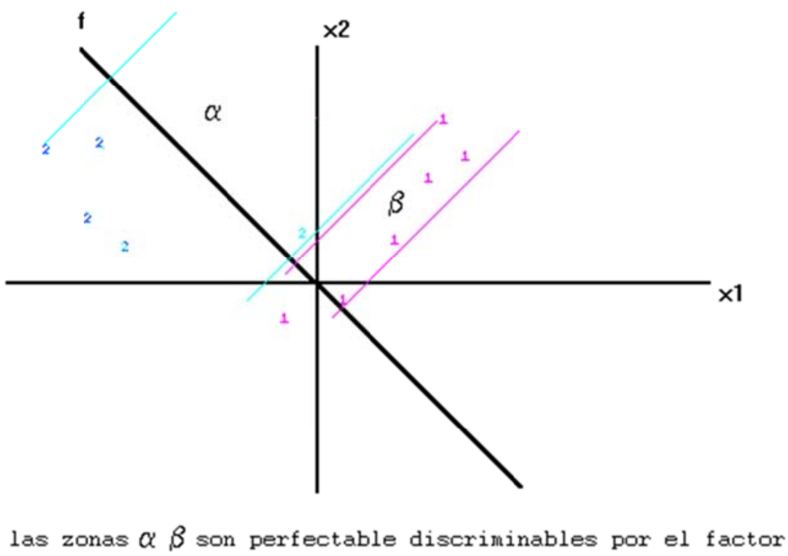
(POR DEFECTO SPSS: $F_S < 1$)

Determinación de los factores o funciones (o ejes) discriminantes

<u>Obtención de los factores discriminantes</u>

Una vez obtenidas el conjunto de variables más discriminantes, la solución puede mejorarse obteniendo un conjunto alternativo de nuevas variables discriminantes, llamadas factores o funciones discriminantes, capaces de discriminar de una forma más contundente y que se obtienen como combinación lineal de la solución obtenida. En el siguiente gráfico se pone de manifiesto, lo que queremos decir:





FACTORES (FUNCIONES/ EJES) DISCRIMINANTES

UNA VEZ SELECCIONADAS LAS p VBLES. MÁS DISCRIMINANTES QUEREMOS DETERMINAR LAS ***DIRECCIONES DE*** R^p, *DEFINIDAS POR VECTORES DEL ESPACIO DE LAS VARIABLES*, *PARA LAS CUALES LA PROYECCIÓN DE LA NUBE DE PUNTOS SEA TAL QUE LOS GRUPOS ESTÉN LO MÁS SEPARADOS POSIBLE Y LOS INDIVIDUOS DEL MISMO GRUPO LO MÁS PRÓXIMOS POSIBLE.*

OBTENCIÓN:

CADA FACTOR F_r ES UNA COMBINACIÓN LINEAL DE LAS $X_1, X_2, \ldots X_p$

$$F_r = \sum_{j=1}^{p} u_{jr} X_j \qquad F_r = u'_r X$$

(BUSCAMOS DETERMINAR u)

SI V ES LA MATRIZ DE VARIANZAS DEL VECTOR DE VARIABLES INICIALES (X), LA VARIANZA DEL FACTOR F SERÁ:

$$V(F_r) = u'_r V u_r = \underbrace{u'_r W u_r}_{\text{INTERNA}} + \underbrace{u'_r B u_r}_{\text{EXTERNA}}$$

PRETENDEMOS DETERMINAR LOS FACTORES (SUS COEFICIENTES) QUE HAGAN MÁXIMA LA VARIANZA EXTERNA Y MINIMICEN LA VARIANZA INTERNA .POR TANTO EL PRIMER FACTOR DISCRIMINANTE SERÁ ÁQUEL CUYO VECTOR DE COEFICIENTES u

$$\max \frac{u'Bu}{u'Wu} \equiv \max \frac{u'Bu}{u'V u}$$

PUEDE PROBARSE QUE LOS SUCESIVOS FACTORES DISCRIMINANTES RESULTAN SER ENTONCES LAS COMBINACIONES LINEALES DADAS POR LOS SUCESIVOS *VECTORES PROPIOS* DE LA MATRIZ $V^{-1}B$ (ASOCIADOS A LOS VALORES PROPIOS CONSIDERADOS DE MAYOR A MENOR)

NÚMERO DE FACTORES:

=RANGO ($V^{-1}B$) =(en general) g - 1

INTERPRETACIÓN:

CADA FACTOR F_r PUEDE INTERPRETARSE A PARTIR DE LOS COEFICIENTES DE LA COMBINACIÓN LINEAL u_r (COEFICIENTES DE LA *MATRIZ PATRÓN*) Y DE LAS CORRELACIONES CON LAS VARIABLES ORIGINALES (COEFICIENTES DE LA *MATRIZ DE ESTRUCTURA DISCRIMINANTE*)

EVALUACIÓN:

UNA VEZ DETERMINADOS LOS FACTORES, PODREMOS EVALUAR LAS LLAMADAS *PUNTUACIONES DISCRIMINANTES* DE CADA INDIVIDUO Y LOS CENTROS DE GRAVEDAD DE CADA GRUPO, LO QUE PERMITIRÁ LA RECLASIFICACIÓN Y UNA MEJOR INTERPRETACIÓN, RESPECTIVAMENTE.

RECLASIFICACIÓN

EVALUADAS LAS PUNTUACIONES DISCRIMINANTES DE UN INDIVIDUO j
(VECTOR F_j)

PRESUPONIENDO QUE EN CADA GRUPO G_i LOS FACTORES DISCRIMINANTES SIGUEN UNA DISTRIBUCIÓN MULTINORMAL

PUEDE DETERMINARSE LA VEROSIMILITUD DE QUE $j \in G_i$, COMO: $f(F_j/G_i)$

PARTIENDO DE UNA PROBABILIDAD INICIAL $P(j \in G_i) = P(G_i)$
(QUE SERÁ 1/g , POR DEFECTO)

APLICANDO EL T. DE BAYES PODREMOS OBTENER LA PROBABILIDAD FINAL DE QUE EL INDIVIDUO j PERTENEZCA AL GRUPO G :

$$P(G_i / F_j) = \frac{P(G_i) \,.\, f(F_j/G_i)}{\sum_{\forall i} P(G_i) \,.\, f(F_j/G_i)}$$

FINALMENTE ASIGNAREMOS EL INDIVIDUO j AL GRUPO QUE HAGA MÁXIMA $P(G_i / F_j)$