

# TEMA 8. ANÁLISIS DISCRIMINANTE

## Análisis de Datos Multivariantes

2017/18

# 1 Introducción

- Supongamos un conjunto de variables  $\underline{X} = (X_1, \dots, X_p)^t$  medidas sobre varias poblaciones (grupos)  $G_1, \dots, G_k$ .
  - Supongamos que se dispone de una muestra de cada uno de los grupos

$$\underline{x}_{i1}, \dots, \underline{x}_{in_i}, \quad i = 1, \dots, k$$

- **Discriminar:** Determinar funciones (lineales) de las variables que permitan separar (diferenciar) entre los grupos.
  - Determinar la contribución de cada variable a la separación entre los grupos.
  - Determinar el plano óptimo sobre el que proyectar los puntos para obtener la máxima separación entre los grupos

- **Clasificar:** Predecir o asignar una observación de origen incierto

$$\underline{x}_0 = (x_{01}, \dots, x_{0p})^t$$

a uno de los grupos.

- Los problemas se solapan, una regla discriminante se puede usar para clasificar.

- Ejemplos:

- Clasificar el nivel de riesgo de un crédito en función del: nivel de ingresos, edad, personas a su cargo, tipo de empleo.
- Detección de e-mail spam
  - \* Definir un criterio de clasificación automática de correo spam en función de las palabras del mismo.
- Pacientes que han padecido un ataque cardiaco.
  - \* Predecir si el paciente va a sufrir un segundo ataque en función de variables clínicas y otros factores (dieta, edad,...)

- Reconocimiento digital de escritura manual.
- Clasificación (asignación a un determinado autor) de una obra literaria en función de las figuras gramaticales de la misma.
- Clasificación (asignación a una especie) de restos óseos en función de las medidas antropométricas.

## 2 Discriminación de dos poblaciones con distribuciones conocidas

- Sea una población  $P$  formada por dos subpoblaciones o grupos  $G_1$  y  $G_2$ , con proporciones  $\pi_1$  y  $\pi_2$  respectivamente ( $\pi_1 + \pi_2 = 1$ ).
- Sea  $\underline{X} = (X_1, \dots, X_p)^t$  con función de densidad:  $f_i(\underline{x})$  en  $G_i$ ,  $i = 1, 2$ .
- **Una regla discriminante** es una partición del espacio muestral  $\mathcal{R}$  en dos subconjuntos  $\mathcal{R}_1$  y  $\mathcal{R}_2$

$$\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \quad ; \quad \mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset ,$$

de forma que dado un individuo con valores  $\underline{x}_0 = (x_{01}, \dots, x_{0p})^t$  se considera el siguiente criterio de asignación:

$$\text{Asignar } \underline{x}_0 \text{ a } G_i \quad \text{si} \quad \underline{x}_0 \in \mathcal{R}_i, \quad i = 1, 2.$$

## 2.1 Regla de clasificación óptima. Errores de clasificación

- Probabilidades de los errores de clasificación

$$P[\text{asignar a } G_2 \mid \text{siendo de } G_1] = P(2/1) = \int_{\mathcal{R}_2} f_1(\underline{x}) d\underline{x}$$

$$P[\text{asignar a } G_1 \mid \text{siendo de } G_2] = P(1/2) = \int_{\mathcal{R}_1} f_2(\underline{x}) d\underline{x}$$

- Probabilidad total de error de clasificación de una regla discriminante

$$P[\mathcal{R}; f] = P(1/2)\pi_2 + P(2/1)\pi_1$$

- Objetivo: Determinar la regla que minimice  $P[\mathcal{R}; f]$ .

**Lema 1** Sea  $g : \mathbb{R}^p \longrightarrow \mathbb{R}$ . Entonces

$$\min_{\mathcal{R}_1 \subset \mathbb{R}^p} \int_{\mathcal{R}_1} g(\underline{x}) d\underline{x} = \int_{\mathcal{R}_g^*} g(\underline{x}) d\underline{x} \quad \text{donde} \quad \mathcal{R}_g^* = \{\underline{x} \in \mathbb{R}^p : g(\underline{x}) < 0\}$$

**Nota.**  $\mathcal{R}_g^*$  no queda determinada de forma única.

**Lema 2** *La regla discriminante que minimiza  $P[\mathcal{R}; f]$  viene dada por:*

*Asignar  $\underline{x}_0$  a  $G_1$  si  $\pi_1 f_1(\underline{x}_0) > \pi_2 f_2(\underline{x}_0)$*

*Asignar  $\underline{x}_0$  a  $G_2$  si  $\pi_1 f_1(\underline{x}_0) < \pi_2 f_2(\underline{x}_0)$*

*Asignar  $\underline{x}_0$  aleatoriamente si  $\pi_1 f_1(\underline{x}_0) = \pi_2 f_2(\underline{x}_0)$*

### **Demostración**

$$\begin{aligned} P[\mathcal{R}; f] &= P(1/2)\pi_2 + P(2/1)\pi_1 = \pi_2 \int_{\mathcal{R}_1} f_2(\underline{x}) d\underline{x} + \pi_1 \int_{\mathcal{R}_2} f_1(\underline{x}) d\underline{x} = \\ &= \pi_2 \int_{\mathcal{R}_1} f_2(\underline{x}) d\underline{x} + \pi_1 \left[ 1 - \int_{\mathcal{R}_1} f_1(\underline{x}) d\underline{x} \right] \\ &= \int_{\mathcal{R}_1} [\pi_2 f_2(\underline{x}) - \pi_1 f_1(\underline{x})] d\underline{x} + \pi_1 \end{aligned}$$

**NOTA.** Caso unidimensional y  $\pi_1 = \pi_2$

**NOTA:** Se pueden construir reglas discriminantes en función de otros criterios:

- Criterio de máxima verosimilitud
- Asignando costes a los errores de clasificación,  $c(1/2)$  y  $c(2/1)$  y minimizando el coste total de error de clasificación.
- Minimizar la probabilidad total de error de clasificación a posteriori.
- Criterio minimax :  $\min \{ \max (P(1/2), P(2/1)) \}$



## 2.2 Poblaciones normales con idéntica matriz de covarianzas

- $\underline{X} \sim N_p(\underline{\mu}_1, \Sigma)$  en  $G_1$  y  $\underline{X} \sim N_p(\underline{\mu}_2, \Sigma)$  en  $G_2$

$$f_i(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\underline{x} - \underline{\mu}_i)^t \Sigma^{-1} (\underline{x} - \underline{\mu}_i) \right\} \quad i = 1, 2$$

- La regla discriminante que minimiza  $P[\mathcal{R}; f]$  dada por

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \pi_1 f_1(\underline{x}_0) > \pi_2 f_2(\underline{x}_0)$$

se puede expresar de distintas formas

- **Expresión 1. (En términos de la distancia de Mahalanobis)**

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \Delta^2(\underline{x}_0; \mu_2) - \Delta^2(\underline{x}_0; \mu_1) > 2 \ln \left( \frac{\pi_2}{\pi_1} \right)$$

con

$$\Delta^2(\underline{x}_0; \mu_i) = (\underline{x}_0 - \mu_i)^t \Sigma^{-1} (\underline{x}_0 - \mu_i)$$

- **Expresión 2. (Como una función lineal de las componentes de  $\underline{x}$ )**

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \lambda^t \underline{x}_0 - \frac{1}{2} \lambda^t (\mu_1 + \mu_2) > \ln \left( \frac{\pi_2}{\pi_1} \right)$$

con

$$\text{con } \lambda^t = (\mu_1 - \mu_2)^t \Sigma^{-1}$$

- **Expresión 3. Supuesto  $\pi_1 = \pi_2$**

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \lambda^t \underline{x}_0 > \frac{1}{2} \lambda^t (\mu_1 + \mu_2)$$

## 2.3 Poblaciones normales con distintas matrices de covarianzas

- $\underline{X} \sim N_p(\underline{\mu}_1, \underline{\Sigma}_1)$  en  $G_1$  y  $\underline{X} \sim N_p(\underline{\mu}_2, \underline{\Sigma}_2)$  en  $G_2$

$$f_i(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\underline{\Sigma}_i|^{1/2}} \exp \left\{ -\frac{1}{2}(\underline{x} - \underline{\mu}_i)^t \underline{\Sigma}_i^{-1} (\underline{x} - \underline{\mu}_i) \right\} \quad i = 1, 2$$

- La regla discriminante que minimiza  $P[\mathcal{R}; f]$  viene dada por

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } Q(\underline{x}_0) > \ln\left(\frac{\pi_2}{\pi_1}\right)$$

donde

$$\begin{aligned} Q(\underline{x}) &= \frac{1}{2} \ln \left[ \frac{|\underline{\Sigma}_2|}{|\underline{\Sigma}_1|} \right] - \frac{1}{2} (\underline{\mu}_1^t \underline{\Sigma}_1^{-1} \underline{\mu}_1 - \underline{\mu}_2^t \underline{\Sigma}_2^{-1} \underline{\mu}_2) \quad (\text{constante}) \\ &\quad + (\underline{\mu}_1^t \underline{\Sigma}_1^{-1} - \underline{\mu}_2^t \underline{\Sigma}_2^{-1}) \underline{x} \quad (\text{lineal}) \\ &\quad - \frac{1}{2} \underline{x}^t (\underline{\Sigma}_1^{-1} - \underline{\Sigma}_2^{-1}) \underline{x} \quad (\text{cuadrático}) \end{aligned}$$

- $Q(\underline{x})$  es una **función cuadrática** de las componentes de  $\underline{x}$ .

### 3 Discriminación en dos poblaciones: caso paramétrico

- Suponemos conocida la forma funcional de la distribución de  $\underline{X}$  excepto un número finito de sus parámetros

$$\underline{X} \sim f_i(\underline{x}, \theta_i), \quad \theta_i \in \mathbb{R}^q, \quad \text{en } G_i, \quad i = 1, 2$$

- Consideramos una muestra de cada subpoblación

$$\underline{x}_{11}, \dots, \underline{x}_{1n_1} \text{ m.a. de } G_1 \quad \underline{x}_{21}, \dots, \underline{x}_{2n_2} \text{ m.a. de } G_2$$

- Sean  $\hat{\theta}_1$  y  $\hat{\theta}_2$  estimadores eficientes de  $\theta_1$  y  $\theta_2$ , respectivamente
- Regla discriminante:

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \pi_1 f_1(\underline{x}_0; \hat{\theta}_1) > \pi_2 f_2(\underline{x}_0; \hat{\theta}_2)$$

### 3.1 Poblaciones normales con idéntica matriz de varianzas

- Suponemos  $\underline{X} \sim N_p(\mu_i, \Sigma)$  en  $G_i$ ,  $i = 1, 2$
- Consideramos una muestra aleatoria de cada subpoblación

$\underline{x}_{11}, \dots, \underline{x}_{1n_1}$  m.a. de  $G_1$

$\underline{x}_{21}, \dots, \underline{x}_{2n_2}$  m.a. de  $G_2$

- Estimadores de los parámetros:

$$\hat{\mu}_i = \bar{\underline{x}}_i, \quad i = 1, 2;$$

$$\hat{\Sigma} = \frac{1}{n_1 + n_2 - 2} [\mathbf{S}_1 + \mathbf{S}_2] \text{ con } \mathbf{S}_i = \sum_{j=1}^{n_i} (\underline{x}_{ij} - \bar{\underline{x}}_i)(\underline{x}_{ij} - \bar{\underline{x}}_i)^t, \quad i = 1, 2$$

- **Expresión 1. (En términos de la distancia de Mahalanobis)**

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \widehat{\Delta}^2(\underline{x}_0; \underline{\bar{x}}_2) - \widehat{\Delta}^2(\underline{x}_0; \underline{\bar{x}}_1) > 2 \ln \left( \frac{\pi_2}{\pi_1} \right)$$

con

$$\widehat{\Delta}^2(\underline{x}_0; \underline{\bar{x}}_i) = (\underline{x}_0 - \underline{\bar{x}}_i)^t \widehat{\Sigma}^{-1} (\underline{x}_0 - \underline{\bar{x}}_i)$$

**Expresión 2. (Como una función lineal de las componentes de  $\underline{x}$ )**

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \widehat{\lambda}^t \underline{x}_0 - \frac{1}{2} \widehat{\lambda}^t (\underline{\bar{x}}_1 + \underline{\bar{x}}_2) > \ln \left( \frac{\pi_2}{\pi_1} \right)$$

con

$$\widehat{\lambda}^t = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)^t \widehat{\Sigma}^{-1}$$

**Expresión 3. Supuesto  $\pi_1 = \pi_2$**

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \underbrace{\widehat{\lambda}^t \underline{x}_0}_{\widehat{y}_0} > \underbrace{\frac{1}{2} \widehat{\lambda}^t (\underline{\bar{x}}_1 + \underline{\bar{x}}_2)}_{\widehat{m}}$$

– **Función lineal discriminante de Fisher:**  $\widehat{y} = \widehat{\lambda}^t \underline{x}$ .

## 3.2 Poblaciones normales con distintas matrices de varianzas:

- Suponemos

$$\underline{X} \sim N_p(\mu_i, \Sigma_i) \text{ en } G_i, \quad i = 1, 2$$

- Consideramos una muestra aleatoria de cada subpoblación

$$\underline{x}_{11}, \dots, \underline{x}_{1n_1} \text{ de } G_1$$

$$\underline{x}_{21}, \dots, \underline{x}_{2n_2} \text{ de } G_2$$

- Se consideran como estimadores de los parámetros:

$$\hat{\mu}_i = \bar{\underline{x}}_i \quad ; \quad \hat{\Sigma}_i = \frac{1}{n_i - 1} \mathbf{S}_i, \quad i = 1, 2$$

- Regla discriminante

Asignar  $\underline{x}_0$  a  $G_1$  si  $\hat{Q}(\underline{x}_0) > \ln \left( \frac{\pi_2}{\pi_1} \right)$

donde

$$\begin{aligned} \hat{Q}(\underline{x}) = & \frac{1}{2} \ln \frac{|\hat{\Sigma}_2|}{|\hat{\Sigma}_1|} - \frac{1}{2} (\bar{\underline{x}}_1^t \hat{\Sigma}_1^{-1} \bar{\underline{x}}_1 - \bar{\underline{x}}_2^t \hat{\Sigma}_2^{-1} \bar{\underline{x}}_2) \\ & + (\bar{\underline{x}}_1^t \hat{\Sigma}_1^{-1} - \bar{\underline{x}}_2^t \hat{\Sigma}_2^{-1}) \underline{x} - \frac{1}{2} \underline{x}^t (\hat{\Sigma}_1^{-1} - \hat{\Sigma}_2^{-1}) \underline{x} \end{aligned}$$

–  $\hat{Q}(\underline{x})$  es una función cuadrática en las observaciones.



## 4 Discriminación en el caso de $k$ subpoblaciones

- Sea una población  $P$  formada por  $k$  subpoblaciones o grupos  $G_i$  con proporciones  $\pi_i$ ,  $i = 1, \dots, k$  ( $\sum_{i=1}^k \pi_i = 1$ ).
- Sea  $\underline{X}$  un vector aleatorio  $p$ -dimensional con función de densidad:  $f_i(\underline{x})$  en  $G_i$ .
- Una regla discriminante es un criterio que permite realizar una partición del espacio muestral  $\mathcal{R}$  en  $k$  subconjuntos  $\mathcal{R}_1, \dots, \mathcal{R}_k$  de forma que dado un individuo con valores  $\underline{x}_0$  se considera el siguiente criterio de asignación:

Asignar  $\underline{x}_0$  a  $G_i$  si  $\underline{x}_0 \in \mathcal{R}_i$

## 4.1 Errores de clasificación. Regla discriminante óptima

- Probabilidades de error de clasificación asociadas a la partición  $\mathcal{R}_1, \dots, \mathcal{R}_k$

$$P \left[ \text{asignar a } G_j \mid \text{siendo de } G_i \right] = P(j / i) = \int_{\mathcal{R}_j} f_i(\underline{x}) d\mathbf{x}$$

$$P(i) = P[\text{asignación errónea} / G_i] = \sum_{j \neq i} P(j / i) = 1 - P(i / i)$$

- Probabilidad total de clasificación errónea

$$P[\mathcal{R}, f] = \sum_i P(i) \pi_i = 1 - \sum_i P(i / i) \pi_i$$

- Objetivo: Determinar la regla que minimice  $P[\mathcal{R}; f]$ .

### Teorema 3 (véase Seber [1])

$P[R, f]$  se minimiza para la partición  $\{\mathcal{R}_1^*, \dots, \mathcal{R}_k^*\}$  con

$$\mathcal{R}_i^* = \left\{ \underline{x} : \pi_i f_i(\underline{x}) \geq \pi_j f_j(\underline{x}) , \quad j = 1, 2, \dots, k \right\}$$

- En consecuencia, la regla discriminate óptima es:

$$\text{Asignar } \underline{x}_0 \text{ a } G_i \text{ si } \max_j \{ \pi_j f_j(\underline{x}_0) \} = \pi_i f_i(\underline{x}_0)$$

- En el caso de parámetros desconocidos, el razonamiento es semejante al caso de dos grupos: estimar los parámetros y trabajar con las funciones de densidad o probabilidad estimadas.
- Se pueden definir las mismas tasas de error y sus correspondientes estimaciones.

## 4.2 Poblaciones normales

- Suponemos

$$\underline{X} \sim N_p(\mu_i, \Sigma_i) \text{ en } G_i, \quad i = 1, \dots, k$$

- Consideramos una muestra aleatoria de cada subpoblación

$$\underline{x}_{i1}, \dots, \underline{x}_{in_i} \text{ de } G_i \quad i = 1, \dots, k$$

- En caso de **igualdad de matrices de varianzas y covarianzas**

$$\text{Asignara } \underline{x}_0 \text{ a } G_i \text{ si } \hat{L}_i(\underline{x}_0) = \max_{j=1, \dots, k} \hat{L}_j(\underline{x}_0)$$

donde

$$\hat{L}_i(\underline{x}_0) = \underline{\bar{x}}_i^t \hat{\Sigma}^{-1} \underline{x}_0 - \frac{1}{2} \underline{\bar{x}}_i^t \hat{\Sigma}^{-1} \underline{\bar{x}}_i + \ln \pi_i \quad i = 1, \dots, k$$

siendo:

$$\hat{\Sigma} = \frac{1}{n - k} [\mathbf{S}_1 + \dots + \mathbf{S}_k] \quad \text{y} \quad \mathbf{S}_i = \sum_{j=1}^{n_i} (\underline{x}_{ij} - \underline{\bar{x}}_i)(\underline{x}_{ij} - \underline{\bar{x}}_i)^t \quad i = 1, \dots, k$$

- En caso de **matrices de varianzas y covarianzas distintas**, se obtiene:

$$\text{Asignar } \underline{x}_0 \text{ a } G_i \text{ si } T_i(\underline{x}_0) = \max_{j=1,\dots,k} T_j(\underline{x}_0)$$

donde

$$T_i(\underline{x}_0) = \left[ \ln \pi_i - \frac{1}{2} \underline{x}_i^t \hat{\Sigma}_i^{-1} \underline{x}_i - \frac{1}{2} \ln |\hat{\Sigma}_i| \right] + \left[ \underline{x}_i^t \hat{\Sigma}_i^{-1} \underline{x}_0 \right] - \frac{1}{2} \underline{x}_0^t \hat{\Sigma}_i^{-1} \underline{x}_0$$

siendo

$$\hat{\Sigma}_i = \frac{1}{n_i - 1} \mathbf{S}_i$$

**Nota.** Al igual que en el caso de la regresión, existen técnicas de discriminación paso a paso (stepwise) para seleccionar el conjunto de variables más adecuado para una discriminación óptima, disminuyendo la dimensión del espacio utilizado para la discriminación.

## 5 Coordenadas discriminantes canónicas

- Sea una población  $P$  formada por  $k$  subpoblaciones o grupos  $G_i$  con proporciones  $\pi_i$ ,  $i = 1, \dots, k$  ( $\sum_{i=1}^k \pi_i = 1$ ).
- Sea  $\underline{X}$  un vector aleatorio  $p$ -dimensional con función de densidad:  $f_i(\underline{x})$  en  $G_i$ .

**Objetivo:** *Determinar las combinaciones lineales de las variables*

$$Y = \underline{\mathbf{c}}^t \underline{X}, \quad \underline{\mathbf{c}} \in \mathbb{R}^p$$

*que "más discriminan" (más diferencian) entre las subpoblaciones.*

## 5.1 Problema original (Fisher) para dos poblaciones $k = 2$

- Supongamos dos poblaciones con  $\Sigma_1 = \Sigma_2$ .
- Consideramos una muestra aleatoria de cada subpoblación

$$\underline{x}_{i1}, \dots, \underline{x}_{in_i} \text{ de } G_i \quad i = 1, 2$$

- Plantea determinar la combinación lineal de las variables que

$$\max_{\underline{c} \in \mathbb{R}^p} \frac{(\underline{c}^t \bar{\underline{x}}_1 - \underline{c}^t \bar{\underline{x}}_2)^2}{\underline{c}^t \hat{\Sigma} \underline{c}} = \max_{\underline{c} \in \mathbb{R}^p} \frac{(\text{Distancia entre las medias})^2}{\text{varianza muestral de } \underline{c}^t \underline{x}} \quad (1)$$

- Regla discriminante:

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \left| \underline{c}^t \underline{x}_0 - \underline{c}^t \bar{\underline{x}}_1 \right| \leq \left| \underline{c}^t \underline{x}_0 - \underline{c}^t \bar{\underline{x}}_2 \right|$$

- El problema (1) es equivalente a

$$\max_{\underline{\mathbf{c}} \in \mathbb{R}^p} \left( \underline{\mathbf{c}}^t \underline{\bar{\mathbf{x}}}_1 - \underline{\mathbf{c}}^t \underline{\bar{\mathbf{x}}}_2 \right)^2 \quad \equiv \quad \max_{\underline{\mathbf{c}} \in \mathbb{R}^p} \underline{\mathbf{c}}^t (\underline{\bar{\mathbf{x}}}_1 - \underline{\bar{\mathbf{x}}}_2) (\underline{\bar{\mathbf{x}}}_1 - \underline{\bar{\mathbf{x}}}_2)^t \underline{\mathbf{c}} \\ s.a. \quad \underline{\mathbf{c}}^t \hat{\underline{\Sigma}} \underline{\mathbf{c}} = 1 \quad s.a. \quad \underline{\mathbf{c}}^t \hat{\underline{\Sigma}} \underline{\mathbf{c}} = 1$$

- Solución: (único) Autovalor-vector de  $\hat{\underline{\Sigma}}^{-1} (\underline{\bar{\mathbf{x}}}_1 - \underline{\bar{\mathbf{x}}}_2) (\underline{\bar{\mathbf{x}}}_1 - \underline{\bar{\mathbf{x}}}_2)^t (= \hat{\underline{\Sigma}}^{-1} \underline{\mathbf{d}} \underline{\mathbf{d}}^t)$

\* Autovalor:  $\underline{\mathbf{d}}^t \hat{\underline{\Sigma}}^{-1} \underline{\mathbf{d}}$ ,      Autovector:  $\underline{\mathbf{c}} = \hat{\underline{\Sigma}}^{-1} \underline{\mathbf{d}}$

Autovector- normalizado

$$\underline{\mathbf{e}} = \frac{\hat{\underline{\Sigma}}^{-1} \underline{\mathbf{d}}}{\left( \underline{\mathbf{d}}^t \hat{\underline{\Sigma}}^{-1} \underline{\mathbf{d}} \right)^{1/2}}$$

- **Regla discriminante:**      Asignar  $\underline{x}_0$  a  $G_1$  si  $|\hat{\underline{\mathbf{e}}}^t \underline{x}_0 - \hat{\underline{\mathbf{e}}}^t \underline{\bar{x}}_1| < |\hat{\underline{\mathbf{e}}}^t \underline{x}_0 - \hat{\underline{\mathbf{e}}}^t \underline{\bar{x}}_2|$

- Operando se tiene

$$\text{Asignar } \underline{x}_0 \text{ a } G_1 \text{ si } \underline{\mathbf{c}}^t \underline{x}_0 \geq \frac{1}{2} \underline{\mathbf{c}}^t (\underline{\bar{x}}_1 + \underline{\bar{x}}_2)$$

(coincide con Regla Discriminante Lineal de Fisher)



## 5.2 k-poblaciones normales con idéntica matriz de varianzas

- El problema (1) para dos poblaciones es equivalente a

$$\max_{\underline{\mathbf{c}} \in \mathbb{R}^p} T_c^2$$

con  $T_c$  el estadístico asociado al contraste  $H_0 : \underline{\mathbf{c}}^t \mu_1 = \underline{\mathbf{c}}^t \mu_2$  en poblaciones normales con iguales matrices de covarianzas.

- Maximizar  $T_c^2$  equivale a minimizar el  $p_c$ -valor.

- Para  $k$  poblaciones se plantea:

- Suponemos  $\underline{X} \sim N_p(\mu_i, \Sigma)$  en  $G_i$ ,  $i = 1, \dots, k$
- Consideramos una muestra aleatoria de cada subpoblación:  $\underline{x}_{i1}, \dots, \underline{x}_{in_i}$  de  $G_i$   $i = 1, \dots, k$
- El objetivo es determinar  $\underline{c}$  de forma que

$$\max_{\underline{c} \in \mathbb{R}^p} F_c$$

con  $F_c$  el estadístico asociado al contraste

$$H_0 : \underline{c}^t \underline{\mu}_1 = \underline{c}^t \underline{\mu}_2 = \dots = \underline{c}^t \underline{\mu}_k$$

- \* Maximizar  $F_c$  equivale a minimizar el  $p_c$ -valor.
- \* Dado  $\underline{c} \in \mathbb{R}^p$ , en cada población  $G_i$ ,  $i = 1, \dots, k$  se tiene que

$$\underline{c}^t \underline{X} \sim N_1(\gamma_i, \sigma_c^2), \text{ con } \gamma_i = \underline{c}^t \underline{\mu}_i \text{ y } \sigma_c^2 = \underline{c}^t \Sigma \underline{c},$$

- El estadístico para contrastar  $H_0$  viene dado

$$F_c = \frac{n - k}{k - 1} \frac{\mathbf{c}^t \mathbf{B} \mathbf{c}}{\mathbf{c}^t \mathbf{W} \mathbf{c}} \quad \left( n = \sum_{i=1}^k n_i \right)$$

$$\mathbf{W} = \sum_{i=1}^k \mathbf{S}_i \quad \text{matriz s.c. dentro de los grupos}$$

$$\mathbf{B} = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^t \quad \text{matriz s.c. entre de los grupos}$$

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (\text{media en } G_i) \quad \text{y} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} \quad (\text{media conjunta})$$

- La región crítica para un test de tamaño  $\alpha$  viene dada por

$$R_c : F_c \geq F_{k-1, n-k, 1-\alpha}$$

- **COORDENADAS DISCRIMINANTES: Definición**

- **1ª coordenada discriminante:**  $Y_1 = \underline{e}_1^t \underline{X}$  con  $\underline{e}_1$  solución de

$$\sup_{\underline{c} \neq 0} \frac{\underline{c}^t \underline{B} \underline{c}}{\underline{c}^t \underline{W} \underline{c}} \quad \text{equivalentemente} \quad \sup_{s.a. \quad \underline{c}^t \underline{W} \underline{c} = 1} \underline{c}^t \underline{B} \underline{c}$$

- **2ª coordenada discriminante:**  $Y_2 = \underline{e}_2^t \underline{X}$  con  $\underline{e}_2$  solución de

$$\sup_{\substack{\underline{c} \neq 0: \\ \underline{e}_1^t \underline{W} \underline{c} = 0}} \frac{\underline{c}^t \underline{B} \underline{c}}{\underline{c}^t \underline{W} \underline{c}} \quad \text{equivalentemente} \quad \sup_{s.a. \quad \underline{c}^t \underline{W} \underline{c} = 1, \quad \underline{e}_1^t \underline{W} \underline{c} = 0} \underline{c}^t \underline{B} \underline{c}$$

- **3ª coordenada discriminante...**

## ● COORDENADAS DISCRIMINANTES: Solución

– Por los resultados vistos en el Tema 6 (Análisis de Correspondencias), las coordenadas discriminantes vienen determinadas por los autovectores asociados a la matriz  $\mathbf{W}^{-1}\mathbf{B}$ .

\* El número de autovalores no nulos es  $r = \min \{p, k - 1\}$

– **1ª coordenada discriminante:**  $Y_1 = \underline{\mathbf{e}}_1^t \underline{\mathbf{X}}$  con  $\underline{\mathbf{e}}_1$  autovector asociado al mayor autovalor de  $\mathbf{W}^{-1}\mathbf{B}$ .

– **2ª coordenada discriminante:**  $Y_2 = \underline{\mathbf{e}}_2^t \underline{\mathbf{X}}$  con  $\underline{\mathbf{e}}_2$  autovector asociado al segundo mayor autovalor de  $\mathbf{W}^{-1}\mathbf{B}$ .

– ...

## 5.3 Regla discriminante

- Sea

$$\mathbf{C}_r = \begin{bmatrix} \underline{\mathbf{e}}_1^t \\ \vdots \\ \underline{\mathbf{e}}_r^t \end{bmatrix}$$

Las coordenadas discriminante de  $\underline{x}_0$  :

$$\mathbf{C}_r \underline{x}_0$$

Las coordenadas discriminantes de las medias de cada grupo:  $\mathbf{C}_r \bar{\underline{x}}_i, i = 1, \dots, k$

- Regla:

$$\text{Asignar } \underline{x}_0 \text{ a } G_i \text{ si } d(\mathbf{C}_r \underline{x}_0, \mathbf{C}_r \bar{\underline{x}}_i) = \min_{s=1, \dots, k} d(\mathbf{C}_r \underline{x}_0, \mathbf{C}_r \bar{\underline{x}}_s)$$

- Nota:  $\mathbf{C}_r \underline{x}$  son las coordenadas de  $\underline{x}$  respecto de la base  $\hat{\Sigma} \underline{\mathbf{e}}_1, \dots, \hat{\Sigma} \underline{\mathbf{e}}_p$ , verificándose

$$d^2(\mathbf{C}_r \underline{x}, \mathbf{C}_r \underline{y}) = (\underline{x} - \underline{y})^t \hat{\Sigma}^{-1} (\underline{x} - \underline{y})$$

- Generalmente se usan sólo las dos primeras coordenadas para construir la regla discriminante,

**Asignar**  $\underline{x}_0$  a  $G_i$  si  $d(\mathbf{C}_2 \underline{x}_0; \mathbf{C}_2 \bar{\underline{x}}_i) = \min_{s=1, \dots, k} d(\mathbf{C}_2 \underline{x}_0; \mathbf{C}_2 \bar{\underline{x}}_s)$

con

$$\mathbf{C}_2 = \begin{bmatrix} \underline{\mathbf{e}}_1^t \\ \underline{\mathbf{e}}_2^t \end{bmatrix}$$

- \* (Los puntos bidimensionales  $\mathbf{C}_2 \bar{\underline{x}}_s = \bar{\mathbf{y}}_s$  se denominan centroides).
- \* Representación gráfica:.....

- **Consideraciones**

- En la práctica se suele realizar la traslación

$$\underline{Z} = [\underline{X} - \underline{\bar{x}}]$$

para que el centroide global  $\underline{\bar{x}}$  se transforme en el origen de coordenadas.

Las coordenadas de los datos trasladados vienen dadas por

$$\mathbf{C}_2 \underline{x} - \mathbf{C}_2 \underline{\bar{x}}.$$

- A veces es interesante construir las coordenadas discriminantes sobre las variables tipificadas (considerando  $\hat{\underline{\Sigma}} = \frac{1}{(n-k)} \mathbf{W}$ )

$$\underline{Z} = \frac{1}{(n-k)^{1/2}} \text{diag} \left\{ \frac{1}{w_{11}^{1/2}}, \dots, \frac{1}{w_{pp}^{1/2}} \right\} [\underline{X} - \underline{\bar{x}}] = \mathbf{M} [\underline{X} - \underline{\bar{x}}]$$

Planteando el problema de optimización sobre las nuevas variables  $\underline{Z}$ , y teniendo en cuenta que  $\mathbf{B}_z = \mathbf{M} \mathbf{B}_x \mathbf{M}^t$  y  $\mathbf{W}_z = \mathbf{M} \mathbf{W}_x \mathbf{M}^t$ , se obtiene que los autovalores  $\underline{\mathbf{u}}_j$  asociados a las variables  $\underline{Z}$  vienen dados por

$$\underline{\mathbf{u}}_j = \mathbf{M}^{-1} \underline{\mathbf{e}}_j = (n-k)^{1/2} \text{diag} \left\{ w_{11}^{1/2}, \dots, w_{pp}^{1/2} \right\} \underline{\mathbf{e}}_j$$



- Las componentes de  $\underline{u}_j$  identifican qué variables influyen más en cada función discriminante canónica.
- Al estar tipificadas las variables originales, todas tienen idéntica "variabilidad". Los coeficientes representan "los pesos" de cada variable en la definición de la variable discriminante.
- La relación entre las variables originales y las variables discriminantes también puede evaluarse a través del coeficiente de correlación lineal entre ambas.

## 6 Estimación de la tasa de error

- Probabilidades de los errores de clasificación

$$P[\text{asignar a } G_2 \mid \text{siendo de } G_1] = P(2/1)$$

$$P[\text{asignar a } G_1 \mid \text{siendo de } G_2] = P(1/2)$$

- Probabilidad total de error de clasificación de una regla discriminante

$$P[\mathcal{R}; f] = P(1/2)\pi_2 + P(2/1)\pi_1$$

### a) Tasas de error aparente

$$e_{i,app} = \frac{m_i}{n_i}, \quad i = 1, 2$$

$m_i$  = número de observaciones de la muestra de  $G_i$  erróneamente clasificadas

$$e_{app} = \pi_1 e_{1,app} + \pi_2 e_{2,app}$$

En caso de que las proporciones  $\pi_i$  sean desconocidas,

$$\hat{\pi}_i = \frac{n_i}{n_1 + n_2} \Rightarrow e_{app} = \frac{m_1 + m_2}{n_1 + n_2}$$

### b) Método Holdout (Entrenamiento - Test)

- Este método utiliza un subconjunto de la muestra para la obtención de la regla discriminante y otro subconjunto (generalmente más pequeño) para su validación y estimación de errores. Evita el carácter optimistamente sesgado de otras técnicas.

### c) Método de cross-validación (jackknife)

- Se determina la regla discriminante usando toda la muestra salvo un individuo y se aplica la regla resultante al individuo omitido.
- Se realiza este procedimiento para cada una de las observaciones muestrales

$$e_{i,c} = \frac{a_i}{n_i} \quad i = 1, 2,$$

$a_i$  = número de observaciones erróneamente clasificadas de la muestra de  $G_i$

$$e_c = \pi_1 e_{1,c} + \pi_2 e_{2,c}$$

- En caso de que las proporciones  $\pi_i$  sean desconocidas,

$$e_c = \frac{a_1 + a_2}{n_1 + n_2}$$

# Referencias bibliográficas

- [1] Seber, G.A.F. "Multivariate observations". John Wiley & Sons
- [2] Dillon, W. y Goldstein, M. "Multivariate Analysis. Methods and Applications". John Wiley & Sons.
- [3] Johnson, R.A. y Wichern, D.W. "Applied Multivariate Statistical Analysis". Prentice-Hall, Inc.