

Análisis discriminante

March 2024

Introducción

Objetivo

- Cómo **clasificar individuos entre varios grupos** a partir de sus medidas en diversas variables aleatorias.
 - Para ello construiremos **funciones discriminantes** que servirán para decidir en qué población incluimos a cada sujeto.
- Esta técnica se puede aplicar a muy **diferentes situaciones**:
 - diagnóstico de enfermedades,
 - clasificación de individuos de diferentes especies,
 - diagnóstico de autoría en obras de arte,
 - clasificación de perfiles de clientes (por ejemplo en la concesión de créditos), etc.
- Cuando **no se conozcan las características** de las poblaciones en las que se pueden clasificar los individuos,
 - necesitaremos disponer de una **muestra** de las variables en estudio en individuos de cada grupo (al menos dos individuos por cada grupo) y de las medidas de los elementos a clasificar en esas variables.

Criterios

- La clasificación se basará en la **distancia de Mahalanobis** del individuo a cada una de las poblaciones (sus medias).
- La utilización de esta distancia es equivalente bajo normalidad a la utilización del criterio de **máxima verosimilitud**,
 - que clasificará a un individuo en donde sus medidas sean **más probables (verosímiles)**, es decir, donde la función de densidad sea mayor.
- Este segundo criterio permitirá la **extensión** de dicha clasificación **a más de dos poblaciones** con diferentes matrices de covarianzas incluso sin la necesidad de la normalidad de las mismas.

Recordemos...

Distancia de Mahalanobis

- La **distancia de Mahalanobis** del vector \mathbf{x} al vector $\boldsymbol{\mu}$ basada en la matriz V se define como

$$d_V(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})' V^{-1} (\mathbf{x} - \boldsymbol{\mu})}.$$

- Si V es la matriz identidad, obtenemos la **distancia Euclídea**.
- Caso de normalidad**: la función de densidad se expresa como

$$f(\mathbf{x}) = \frac{1}{\sqrt{|V|}(2\pi)^k} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' V^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right),$$

para $\mathbf{x} \in \mathbb{R}^k$, donde $\boldsymbol{\mu}$ es el vector de medias y V es la matriz de covarianzas.

→ Las **circunferencias para la distancia de Mahalanobis** con centro en $\boldsymbol{\mu}$ coincidirán con las **curvas de nivel** de la función de densidad ($f(x) = cte.$).

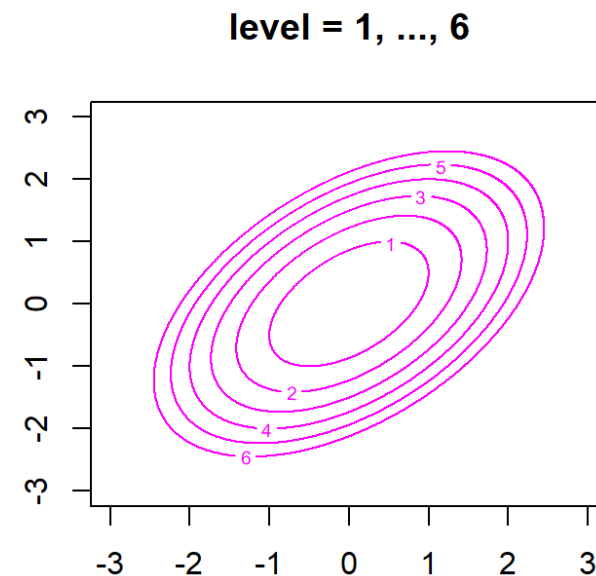
Para una normal bivalente

- Por ejemplo, para una distribución

$$\mathcal{N}_2 \left(\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, V = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} \right)$$

- Las **circunferencias para la distancia de Mahalanobis** con centro en $\boldsymbol{\mu}$ coincidirán con las **curvas de nivel** de la función de densidad.

► Code



Dos poblaciones normales con la misma matriz de covarianzas

Clasificación teórica

- Supongamos que $\mathbf{X} = (X_1, \dots, X_k)'$ e $\mathbf{Y} = (Y_1, \dots, Y_k)'$ son dos v.a. normales k -dimensionales con vectores de medias $\mu_{\mathbf{X}}$ y $\mu_{\mathbf{Y}}$ y matriz de covarianzas común V definida positiva.
- Supongamos que $\mathbf{Z} = (Z_1, \dots, Z_k)'$ representa las medidas obtenidas para el individuo que se quiere clasificar y que \mathbf{Z} proviene de \mathbf{X} o de \mathbf{Y} .
 - Es decir, \mathbf{Z} será un v.a. normal k -dimensional con media igual a $\mu_{\mathbf{X}}$ o $\mu_{\mathbf{Y}}$ y matriz de covarianzas V .
- En la práctica \mathbf{z} será un punto de \mathbb{R}^k que debemos clasificar en \mathbf{X} o en \mathbf{Y} .
- La idea de Fisher es usar una función discriminante D (Función discriminante de Fisher) unidimensional lineal basada en \mathbf{Z} :

$$D = \mathbf{a}'\mathbf{Z} = a_1Z_1 + \dots + a_kZ_k,$$

donde $\mathbf{a} \in \mathbb{R}^k$.

- Si $\mathbf{Z} \rightarrow \mathcal{N}_k(\boldsymbol{\mu}, V)$, entonces

$$D = \mathbf{a}'\mathbf{Z} \rightarrow \mathcal{N}_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'V\mathbf{a})$$

ya que $E[\mathbf{a}'\mathbf{Z}] = \mathbf{a}'E[\mathbf{Z}]$ y

$$Var(\mathbf{a}'\mathbf{Z}) = Cov(\mathbf{a}'\mathbf{Z}) = \mathbf{a}'Cov(\mathbf{Z})\mathbf{a} = \mathbf{a}'V\mathbf{a},$$

donde $\boldsymbol{\mu} = E(\mathbf{Z}) = \boldsymbol{\mu}_X$ ó $\boldsymbol{\mu}_Y$.

- Esta función debe elegirse de forma que discrimine (aleje) a los individuos de \mathbf{X} de los de \mathbf{Y} .

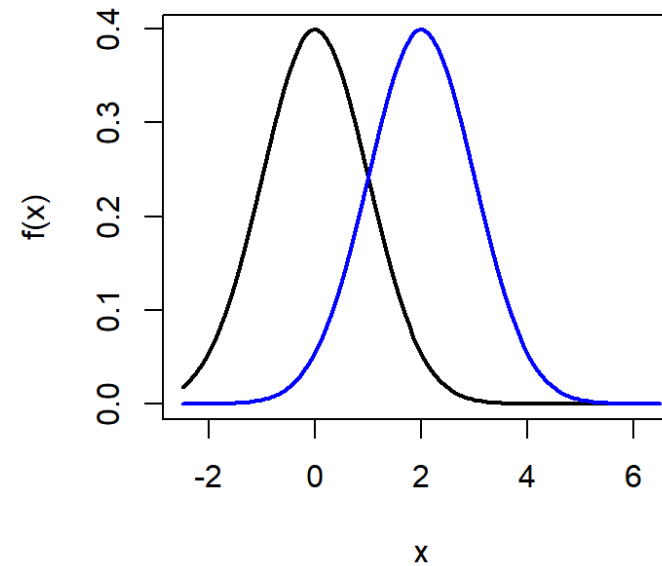
→ Debemos **resolver el problema** siguiente:

$$\max_{\mathbf{a}} \frac{(\mathbf{a}'\boldsymbol{\mu}_X - \mathbf{a}'\boldsymbol{\mu}_Y)^2}{\mathbf{a}'V\mathbf{a}}.$$

→ El objetivo es alejar las **proyecciones** de las medias $\mathbf{a}'\boldsymbol{\mu}_X$ y $\mathbf{a}'\boldsymbol{\mu}_Y$ y disminuir la varianza común $\sigma^2 = \mathbf{a}'V\mathbf{a}$.

- **Ejemplo:** funciones de densidad de las proyecciones en cada grupo

► Code



- La solución se obtiene en el teorema siguiente.

Teorema

- Si V es **definida positiva**, la solución general del problema

$$\max_{\mathbf{a}} \frac{(\mathbf{a}'\boldsymbol{\mu}_{\mathbf{X}} - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{Y}})^2}{\mathbf{a}'V\mathbf{a}}$$

viene dada por

$$\mathbf{a} = \lambda V^{-1}(\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{Y}})$$

para $\lambda \neq 0$, y el máximo vale $d_V^2(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{Y}})$.

Demostración

- La demostración se basa en la desigualdad de Cauchy-Schwarz:

$$(\mathbf{x}'\mathbf{y})^2 \leq (\mathbf{x}'\mathbf{x})(\mathbf{y}'\mathbf{y}),$$

donde se da la igualdad si, y solo si, $\mathbf{x} = \lambda\mathbf{y}$.

- Como V es definida positiva, existe su inversa V^{-1} y $\mathbf{a}'V\mathbf{a} > 0$ para todo vector $\mathbf{a} \neq \mathbf{0}$.
- Entonces, tenemos

$$\begin{aligned}
 \frac{(\mathbf{a}'\boldsymbol{\mu}_X - \mathbf{a}'\boldsymbol{\mu}_Y)^2}{\mathbf{a}'V\mathbf{a}} &= \frac{(\mathbf{a}'V^{1/2}V^{-1/2}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y))^2}{\mathbf{a}'V\mathbf{a}} \\
 &\leq \frac{\mathbf{a}'V\mathbf{a}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)}{\mathbf{a}'V\mathbf{a}} \\
 &= (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y) \\
 &= d_V^2(\boldsymbol{\mu}_X, \boldsymbol{\mu}_Y),
 \end{aligned}$$

donde hemos considerado $\mathbf{x}' = \mathbf{a}'V^{1/2}$ e $\mathbf{y} = V^{-1/2}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)$.

- Además, se verifica la igualdad si, y solo si $\mathbf{x} = \lambda\mathbf{y}$, es decir, si

$$V^{1/2}\mathbf{a} = \lambda V^{-1/2}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y),$$

lo que implica que $\mathbf{a} = \lambda V^{-1}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)$.

Función discriminante de Fisher

- Llamaremos **función discriminante de Fisher** a la v.a.

$$D = L(\mathbf{Z}) = \mathbf{a}'\mathbf{Z} = (\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{Y}})'V^{-1}\mathbf{Z}.$$

- Si las variables \mathbf{X} e \mathbf{Y} son normales, entonces la nueva variable D será normal

$$D \rightarrow \mathcal{N}_1((\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{Y}})'V^{-1}\boldsymbol{\mu}, d_V^2(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{Y}})),$$

donde $\boldsymbol{\mu} = E(\mathbf{Z})$ es igual a $\boldsymbol{\mu}_{\mathbf{X}}$ ó $\boldsymbol{\mu}_{\mathbf{Y}}$.

- Hemos considerado $\lambda = 1$, pero esto no influye en la clasificación ya que podemos tomar cualquier otro λ no nulo.
 → Por ejemplo, si tomamos

$$\lambda = \frac{1}{\|\mathbf{a}\|}$$

obtenemos una proyección en la dirección de \mathbf{a} .

Regla de discriminación

- Consideramos la **función discriminante de Fisher** y $K = L \left(\frac{\mu_{\mathbf{X}} + \mu_{\mathbf{Y}}}{2} \right)$.
- La **regla de discriminación** será:
 - Si $L(\mathbf{Z}) > K$, entonces \mathbf{Z} es clasificado en \mathbf{X} .
 - Si $L(\mathbf{Z}) < K$, entonces \mathbf{Z} es clasificado en \mathbf{Y} .
- En realidad clasificamos a un individuo con características \mathbf{z} según $\mathbf{a}'\mathbf{z}$ esté más cerca de $\mathbf{a}'\mu_{\mathbf{X}}$ o de $\mathbf{a}'\mu_{\mathbf{Y}}$, ya que, como

$$(\mu_{\mathbf{X}} - \mu_{\mathbf{Y}})'V^{-1}(\mu_{\mathbf{X}} - \mu_{\mathbf{Y}}) \geq 0,$$

entonces

$$\mathbf{a}'\mu_{\mathbf{X}} = (\mu_{\mathbf{X}} - \mu_{\mathbf{Y}})'V^{-1}\mu_{\mathbf{X}} \geq (\mu_{\mathbf{X}} - \mu_{\mathbf{Y}})'V^{-1}\mu_{\mathbf{Y}} = \mathbf{a}'\mu_{\mathbf{Y}},$$

- Es decir, con esta función discriminante, la proyección de la media de \mathbf{X} será siempre mayor que la proyección de la media de \mathbf{Y} .

- Ocurrirá lo mismo si tomamos $\lambda > 0$ y lo contrario si tomamos $\lambda < 0$.
- De esta forma, se crean **dos regiones** en el conjunto de posibles valores de \mathbf{Z} ,
 - la región de individuos que serán clasificados en \mathbf{X} :

$$R_{\mathbf{X}} = \{\mathbf{z} \in \mathbb{R}^k : L(\mathbf{z}) > K\}$$

→ la de los que lo serán en \mathbf{Y} :

$$R_{\mathbf{Y}} = \{\mathbf{z} \in \mathbb{R}^k : L(\mathbf{z}) < K\}$$

¿Cómo de buena es la función discriminante de Fisher obtenida?

- La función discriminante de Fisher será mejor cuanto **más alejadas estén las medias $\mathbf{a}'\boldsymbol{\mu}_X$ y $\mathbf{a}'\boldsymbol{\mu}_Y$** , y cuanto **más pequeña sea la varianza $\mathbf{a}'V\mathbf{a}$** .
- Así, el cociente

$$\frac{(\mathbf{a}'\boldsymbol{\mu}_X - \mathbf{a}'\boldsymbol{\mu}_Y)^2}{\mathbf{a}'V\mathbf{a}} = (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y) = d_V^2(\boldsymbol{\mu}_X, \boldsymbol{\mu}_Y)$$

(que no depende de λ) puede servir para comparar una función de discriminación con otra.

Nota

- La discriminación será buena si las **medias poblacionales están alejadas** según la distancia de Mahalanobis asociada a V .

Otro criterio para medir la bondad de un criterio de clasificación

- Podemos calcular las **probabilidades de malas (buenas) clasificaciones**.
- Si llamamos **error tipo 1**, e_1 , al que clasifica a un individuo de la población \mathbf{X} en la población \mathbf{Y} , entonces

$$\begin{aligned}
 \Pr(e_1) &= \Pr(\mathbf{Z} \in R_{\mathbf{Y}} \mid \mathbf{Z} \equiv \mathbf{X}) = \Pr(L(\mathbf{X}) < K) \\
 &= \Pr\left(\mathbf{a}'\mathbf{X} < \mathbf{a}'\frac{\boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\mu}_{\mathbf{Y}}}{2}\right) \\
 &= \Pr\left(\frac{\mathbf{a}'\mathbf{X} - \mathbf{a}'\boldsymbol{\mu}_{\mathbf{X}}}{\sqrt{\mathbf{a}'V\mathbf{a}}} < \frac{\mathbf{a}'(\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{X}})}{2\sqrt{\mathbf{a}'V\mathbf{a}}}\right) \\
 &= \Pr\left(U < \frac{(\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{Y}})'V^{-1}(\boldsymbol{\mu}_{\mathbf{Y}} - \boldsymbol{\mu}_{\mathbf{X}})}{2\sqrt{(\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{Y}})'V^{-1}(\boldsymbol{\mu}_{\mathbf{X}} - \boldsymbol{\mu}_{\mathbf{Y}})}}\right) \\
 &= \Pr\left(U < -\frac{1}{2}d_V(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{Y}})\right),
 \end{aligned}$$

donde $U \rightarrow N_1(0, 1)$.

- De forma análoga, si llamamos **error tipo 2**, e_2 , al que clasifica a un individuo de la población \mathbf{Y} en la población \mathbf{X} , entonces puede comprobarse que

$$\Pr(e_2) = \Pr(\mathbf{Z} \in R_{\mathbf{X}} \mid \mathbf{Z} \equiv \mathbf{Y}) = \Pr\left(U > \frac{1}{2}d_V(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{Y}})\right) = \Pr(e_1).$$

- Por lo tanto, las **probabilidades de clasificaciones erróneas** son **iguales** y solo **dependen de la distancia de Mahalanobis entre las medias de las poblaciones**.
- Lógicamente las **probabilidades de clasificaciones correctas** vienen dadas por:

$$\Pr(c_1) = \Pr(\mathbf{Z} \in R_{\mathbf{X}} \mid \mathbf{Z} \equiv \mathbf{X}) = 1 - \Pr(e_1),$$

$$\Pr(c_2) = \Pr(\mathbf{Z} \in R_{\mathbf{Y}} \mid \mathbf{Z} \equiv \mathbf{Y}) = 1 - \Pr(e_2),$$

y también son **iguales**.

Un caso sencillo

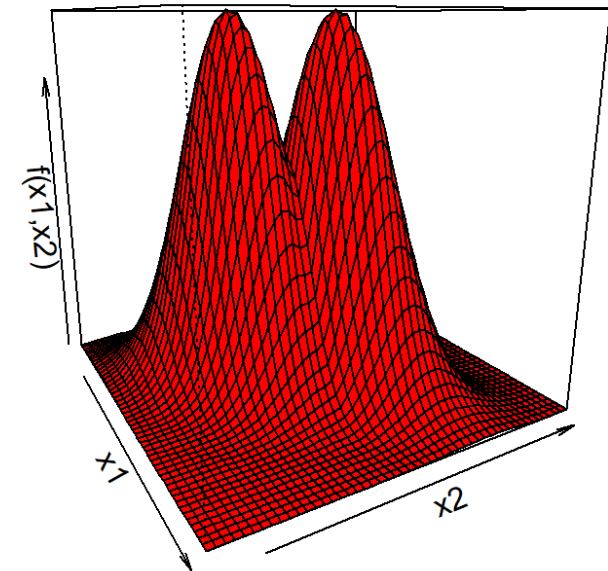
Supongamos que tenemos que decidir si un individuo con medidas

$$\mathbf{z} = (z_1, z_2)' = (2, 0.9)'$$

se clasifica en una población normal bivalente de media $\boldsymbol{\mu}_X = (0, 0)'$ o en una de media $\boldsymbol{\mu}_Y = (1, 2)'$ siendo la matriz de covarianzas común

$$V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

► Code



- La **función discriminante de Fisher** será:

$$\begin{aligned}
 D = L(\mathbf{Z}) &= \mathbf{a}'\mathbf{Z} = (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}\mathbf{Z} = -(1, 2) \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}^{-1} \mathbf{Z} \\
 &= -(1, 2) \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix} \mathbf{Z} = -(0, 2)\mathbf{Z} = -2Z_2,
 \end{aligned}$$

esto es, $L(z_1, z_2) = -2z_2$.

- La **distancia de Mahalanobis al cuadrado** entre las dos poblaciones vale

$$d_V^2(\boldsymbol{\mu}_X, \boldsymbol{\mu}_Y) = (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y) = (0, 2)(1, 2)' = 4.$$

- Un individuo \mathbf{Z} será clasificado en la primera población si

$$-2Z_2 > K = a' \frac{\mu_{\mathbf{X}} + \mu_{\mathbf{Y}}}{2} = -(0, 2)(0.5, 1)' = -2,$$

es decir, si $Z_2 < 1$.

- En este caso, $\mathbf{z} = (1, 0.9)$ será clasificado en \mathbf{X} , con una probabilidad de error global

$$\begin{aligned}
 \Pr(e_2) &= \Pr(\mathbf{Z} \in R_{\mathbf{X}} \mid \mathbf{Z} \equiv \mathbf{Y}) \\
 &= \Pr(U > \frac{1}{2}d_V(\mu_{\mathbf{X}}, \mu_{\mathbf{Y}})) \\
 &= \Pr(U > 1) \\
 &= 1 - F_U(1) \\
 &= 1 - 0.8413 = 0.1587,
 \end{aligned}$$

donde la función de distribución normal estándar $F_U(1)$ se puede calcular con las tablas estadísticas o en R con la instrucción `pnorm(1)`.

- Otra función discriminante equivalente será

$$L^*(z_1, z_2) = z_2,$$

(proyección sobre el eje y) con la que obtendríamos

$$L^*(\mu_{\mathbf{X}}) = L^*(0, 0) = 0,$$

$$L^*(\mu_{\mathbf{Y}}) = L^*(1, 2) = 2,$$

$$K^* = \frac{L^*(\mu_{\mathbf{X}}) + L^*(\mu_{\mathbf{Y}})}{2} = 1$$

y

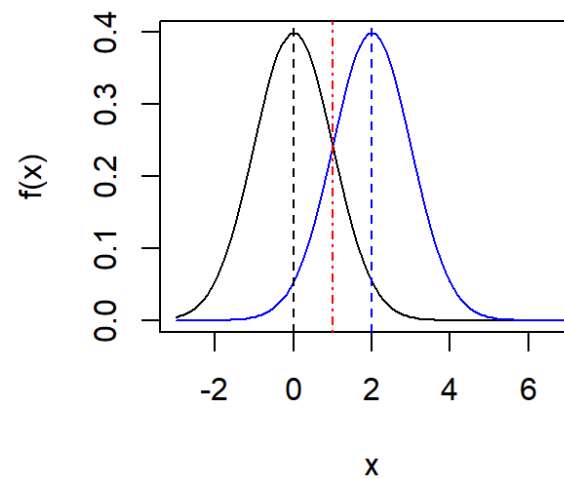
$$L^*(\mathbf{z}) = L^*(1, 0.9) = 0.9,$$

con lo que \mathbf{z} se clasificará en \mathbf{X} .

- Las proyecciones con esta función serán $N(0, 1)$ ($L^*(\mathbf{X})$) y $N(2, 1)$ ($L^*(\mathbf{Y})$).

- Funciones de densidad de las proyecciones sobre el eje y en cada grupo para las poblaciones:

► Code



- La probabilidad de error **0.1587** corresponde a las áreas menores determinadas por la recta vertical en el punto de corte de las densidades en $K = 1$.



- Welch (1939) probó que, si las **poblaciones son normales**, el procedimiento de clasificación mediante la función discriminante de Fisher es máximo verosímil, es decir, que se clasifica a un individuo con características \mathbf{z} en \mathbf{X} si y solo si $f_{\mathbf{X}}(\mathbf{z}) > f_{\mathbf{Y}}(\mathbf{z})$.
- Además, se comprueba que también es equivalente al criterio de clasificación basado en la distancia de Mahalanobis mínima.

Teorema

- Si las variables \mathbf{X} e \mathbf{Y} son **normales multivariantes** con **matriz de covarianzas común** V y función discriminante de Fisher L , entonces equivalen:
 1. $L(\mathbf{z}) > K$.
 2. $d_V(\mathbf{z}, \mu_{\mathbf{X}}) < d_V(\mathbf{z}, \mu_{\mathbf{Y}})$.
 3. $f_{\mathbf{X}}(\mathbf{z}) > f_{\mathbf{Y}}(\mathbf{z})$.

Demostración

- La primera condición $L(\mathbf{z}) > K$ es

$$\mathbf{a}'\mathbf{z} > \mathbf{a}'\frac{\boldsymbol{\mu}_X + \boldsymbol{\mu}_Y}{2},$$

con $\mathbf{a}' = (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}$, es decir,

$$\begin{aligned} (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}\mathbf{z} &> \frac{1}{2}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}(\boldsymbol{\mu}_X + \boldsymbol{\mu}_Y) \\ &= \frac{1}{2}\boldsymbol{\mu}_X'V^{-1}\boldsymbol{\mu}_X - \frac{1}{2}\boldsymbol{\mu}_Y'V^{-1}\boldsymbol{\mu}_Y \end{aligned}$$

lo que equivale a

$$2\boldsymbol{\mu}_X'V^{-1}\mathbf{z} - 2\boldsymbol{\mu}_Y'V^{-1}\mathbf{z} > \boldsymbol{\mu}_X'V^{-1}\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y'V^{-1}\boldsymbol{\mu}_Y.$$

- La segunda condición es equivalente a

$$(\mathbf{z} - \boldsymbol{\mu}_X)'V^{-1}(\mathbf{z} - \boldsymbol{\mu}_X) < (\mathbf{z} - \boldsymbol{\mu}_Y)'V^{-1}(\mathbf{z} - \boldsymbol{\mu}_Y)$$

y, desarrollando, se obtiene

$$\mathbf{z}'V^{-1}\mathbf{z} - 2\boldsymbol{\mu}_X'V^{-1}\mathbf{z} + \boldsymbol{\mu}_X'V^{-1}\boldsymbol{\mu}_X < \mathbf{z}'V^{-1}\mathbf{z} - 2\boldsymbol{\mu}_Y'V^{-1}\mathbf{z} + \boldsymbol{\mu}_Y'V^{-1}\boldsymbol{\mu}_Y.$$

Es decir,

$$2\boldsymbol{\mu}_X'V^{-1}\mathbf{z} - 2\boldsymbol{\mu}_Y'V^{-1}\mathbf{z} > \boldsymbol{\mu}_X'V^{-1}\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y'V^{-1}\boldsymbol{\mu}_Y$$

y, por tanto, las dos primeras condiciones son equivalentes.

- Si las poblaciones son normales, la tercera condición es

$$c \exp \left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{X}})'V^{-1}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{X}}) \right) > c \exp \left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{Y}})'V^{-1}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{Y}}) \right),$$

donde $c = \frac{1}{\sqrt{|V|}(2\pi)^k}$. Es decir,

$$(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{X}})'V^{-1}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{X}}) < (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{Y}})'V^{-1}(\mathbf{z} - \boldsymbol{\mu}_{\mathbf{Y}})$$

lo que es equivalente a la condición segunda.

Observación

- En la demostración anterior la hipótesis de normalidad no es necesaria para demostrar la equivalencia entre las dos primeras condiciones.

Diferente importancia a los dos tipos de errores

- En ocasiones no es conveniente dar la misma importancia a los dos tipos de errores.
- Podemos usar el **criterio** utilizado en los contrastes de hipótesis (Neyman-Pearson):
 - **fijar un máximo para uno de los errores** $\Pr(e_1) \leq \alpha$
 - intentar **reducir la probabilidad del otro error** ($\Pr(e_2)$)
- Usando este criterio sobre la función discriminante de Fisher, cambiará la constante K , que ahora se calculará a partir de la relación

$$\Pr(e_1) = \Pr(\mathbf{Z} \in R_Y | \mathbf{Z} \equiv \mathbf{X}) = \Pr(L(\mathbf{X}) < K_\alpha) = \alpha,$$

donde $L(\mathbf{X}) \rightarrow N_1((\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}\boldsymbol{\mu}_X, \sigma^2)$ y

$$\sigma^2 = d_V^2(\boldsymbol{\mu}_X, \boldsymbol{\mu}_Y) = (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y).$$

- De esta forma, la probabilidad del otro error valdrá

$$\Pr(e_2) = \Pr(L(\mathbf{Y}) > K_\alpha),$$

donde $L(\mathbf{Y}) \rightarrow N_1((\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)'V^{-1}\boldsymbol{\mu}_Y, \sigma^2)$.

Criterio de mínimo coste (probabilidad de error)

- Otro criterio consiste en asignar un coste a cada uno de los posibles errores ($c_1, c_2 > 0$).
- Supongamos que se conocen las probabilidades **a priori** de pertenencia a cada una de las poblaciones:
 - $q_1 = \Pr(\mathbf{Z} \equiv \mathbf{X})$
 - $q_2 = \Pr(\mathbf{Z} \equiv \mathbf{Y})$
- Entonces usando el teorema de la probabilidad total, se tiene

$$\begin{aligned}
 \Pr(error) &= \Pr(\mathbf{Z} \in R_{\mathbf{Y}} \mid \mathbf{Z} \equiv \mathbf{X}) \Pr(\mathbf{Z} \equiv \mathbf{X}) \\
 &\quad + \Pr(\mathbf{Z} \in R_{\mathbf{X}} \mid \mathbf{Z} \equiv \mathbf{Y}) \Pr(\mathbf{Z} \equiv \mathbf{Y}) \\
 &= \Pr(e_1)q_1 + \Pr(e_2)q_2
 \end{aligned}$$

- Y el **coste esperado** asociado para una constante k será

$$c(k) = c_1 \Pr(e_1)q_1 + c_2 \Pr(e_2)q_2,$$

donde

$$\Pr(e_1) = \Pr(L(\mathbf{X}) < k) = G\left(\frac{k - L(\boldsymbol{\mu}_X)}{\sigma}\right),$$

$$\Pr(e_2) = \Pr(L(\mathbf{Y}) > k) = 1 - G\left(\frac{k - L(\boldsymbol{\mu}_Y)}{\sigma}\right),$$

donde G es la función de distribución de la normal estándar $\mathcal{N}(0, 1)$.

- Para minimizar el coste esperado, debe tomarse

$$k = \mathbf{a}' \frac{\boldsymbol{\mu}_X + \boldsymbol{\mu}_Y}{2} + \log\left(\frac{c_2 q_2}{c_1 q_1}\right).$$

(la demostración puede verse en Navarro (2024, p. 157))

Criterio de máxima probabilidad a posteriori

- Supongamos que se conocen las probabilidades **a priori** q_1 y q_2 .
- Entonces se pueden calcular las probabilidades **a posteriori** (es decir, cuando conocemos los valores de \mathbf{Z}) para un individuo con medidas \mathbf{z} mediante el Teorema de Bayes como:

$$\Pr(\mathbf{Z} \equiv \mathbf{X} \mid \mathbf{Z} = \mathbf{z}) = \frac{\Pr(\mathbf{Z} = \mathbf{z} \mid \mathbf{Z} \equiv \mathbf{X}) \Pr(\mathbf{Z} \equiv \mathbf{X})}{\Pr(\mathbf{Z} = \mathbf{z})},$$

$$\Pr(\mathbf{Z} \equiv \mathbf{Y} \mid \mathbf{Z} = \mathbf{z}) = \frac{\Pr(\mathbf{Z} = \mathbf{z} \mid \mathbf{Z} \equiv \mathbf{Y}) \Pr(\mathbf{Z} \equiv \mathbf{Y})}{\Pr(\mathbf{Z} = \mathbf{z})},$$

con

$$\begin{aligned} \Pr(\mathbf{Z} = \mathbf{z}) &= \Pr(\mathbf{Z} = \mathbf{z} \mid \mathbf{Z} \equiv \mathbf{X}) \Pr(\mathbf{Z} \equiv \mathbf{X}) \\ &\quad + \Pr(\mathbf{Z} = \mathbf{z} \mid \mathbf{Z} \equiv \mathbf{Y}) \Pr(\mathbf{Z} \equiv \mathbf{Y}). \end{aligned}$$

- Según este criterio, se clasificaría al individuo \mathbf{z} en la población en la que tenga mayor probabilidad a posteriori.

Varias poblaciones con la misma matriz de covarianza

- Cuando tengamos más de dos poblaciones con **matriz de covarianzas común** V , podemos usar el criterio de mínima distancia de Mahalanobis a las medias de los grupos.
- $\mathbf{X}^{(i)}$ representan las diferentes poblaciones con medias $\boldsymbol{\mu}^{(i)} = E(\mathbf{X}^{(i)})$ y matriz de covarianzas V , para $i = 1, \dots, m$.
- Para un individuo con características \mathbf{z} calcularemos

$$\begin{aligned}
 d_V^2(\mathbf{z}, \boldsymbol{\mu}^{(i)}) &= (\mathbf{z} - \boldsymbol{\mu}^{(i)})' V^{-1} (\mathbf{z} - \boldsymbol{\mu}^{(i)}) \\
 &= \mathbf{z}' V^{-1} \mathbf{z} - 2(\boldsymbol{\mu}^{(i)})' V^{-1} \mathbf{z} + (\boldsymbol{\mu}^{(i)})' V^{-1} \boldsymbol{\mu}^{(i)}.
 \end{aligned}$$

- Como la parte cuadrática $\mathbf{z}'V^{-1}\mathbf{z}$ es común, podemos quedarnos solo con la parte lineal (en realidad, su opuesta) dada por

$$L_i(\mathbf{z}) = (\boldsymbol{\mu}^{(i)})'V^{-1}\mathbf{z} - (\boldsymbol{\mu}^{(i)})'V^{-1}\boldsymbol{\mu}^{(i)} / 2$$

conocida como **función discriminante lineal (FDL)**, clasificándose un individuo con características \mathbf{z} en el grupo en el que tenga un valor máximo dicha función discriminante.

- Como consecuencia se obtiene el teorema siguiente.

Teorema

- Si $\mathbf{X}^{(i)}$ tienen medias $\boldsymbol{\mu}^{(i)} = E(\mathbf{X}^{(i)})$ y **matriz de covarianzas común** V para $i = 1, \dots, m$, entonces equivalen:
 1. $L_i(\mathbf{z}) \geq L_j(\mathbf{z})$ para todo j .
 2. $d_V^2(\mathbf{z}, \boldsymbol{\mu}^{(i)}) \leq d_V^2(\mathbf{z}, \boldsymbol{\mu}^{(j)})$ para todo j .

Corolario

Si solo hay **dos grupos**, este criterio de clasificación es equivalente a usar la función discriminante de Fisher.

Demostración

- La demostración del corolario es inmediata ya que demostramos que el criterio de mínima distancia de Mahalanobis era equivalente a usar la función discriminante de Fisher.
- En este caso también podemos aplicar el criterio de máxima verosimilitud clasificando a \mathbf{Z} en el grupo para el que $f_i(\mathbf{z})$ sea máxima, donde f_i representa la densidad del grupo i .
- Bajo normalidad, tenemos el resultado siguiente.

Teorema

- Si $\mathbf{X}^{(j)} \sim N(\boldsymbol{\mu}^{(j)}, V)$ para $j = 1, \dots, m$, entonces equivalen:
 1. $L_i(\mathbf{z}) \geq L_j(\mathbf{z})$ para todo j .
 2. $d_V^2(\mathbf{z}, \boldsymbol{\mu}^{(i)}) \leq d_V^2(\mathbf{z}, \boldsymbol{\mu}^{(j)})$ para todo j .
 3. $f_i(\mathbf{z}) \geq f_j(\mathbf{z})$ para todo j .

Demostración

- La demostración es inmediata ya que la densidad normal multivariante del grupo i vale

$$f_i(\mathbf{z}) = \frac{1}{\sqrt{|V|} (2\pi)^k} \exp \left(-\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu}^{(i)})' V^{-1} (\mathbf{z} - \boldsymbol{\mu}^{(i)}) \right)$$

y será máxima cuando la distancia de Mahalanobis al cuadrado

$$d_V^2(\mathbf{z}, \boldsymbol{\mu}^{(i)}) = (\mathbf{z} - \boldsymbol{\mu}^{(i)})' V^{-1} (\mathbf{z} - \boldsymbol{\mu}^{(i)})$$

sea mínima.

- Esto no será, en general, cierto si las poblaciones no son normales o si tienen distintas matrices de covarianzas.
- Como consecuencia inmediata, se obtiene el resultado siguiente.

Proposición

- Si todas las poblaciones son **normales con matriz de covarianzas común V** , entonces los criterios de clasificación de **máxima verosimilitud** y de **mínima distancia de Mahalanobis** son equivalentes a aplicar el criterio de **discriminación de Fisher** paso a paso tomando las poblaciones de dos en dos.

- De esta forma, podríamos estudiar primero si **Z** se clasifica en la población 1 o en la 2.
- En el segundo paso discriminaríamos entre la 3 y la ganadora del primer paso y así, sucesivamente.
- Sin embargo, este método no se puede aplicar en la práctica ya que al discriminar entre las poblaciones 1 y 2 solo se utilizarían los individuos de estas poblaciones para estimar **V** .

Ejemplo

Supongamos que tenemos que decidir si un individuo con medidas $\mathbf{z} = (x, y)' = (1, 0.9)'$ se clasifica

- en una población normal bivalente de media $(0, 0)'$,
- en una de media $(1, 2)'$ o
- en una de media $(-1/2, 1)$

siendo la matriz de covarianzas común

$$V = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}.$$

- Entonces las funciones discriminantes lineales (LDF) para $\mathbf{z} = (x, y)'$ serán:

$$L_1(x, y) = (\boldsymbol{\mu}^{(1)})'V^{-1}\mathbf{z} - (\boldsymbol{\mu}^{(1)})'V^{-1}\boldsymbol{\mu}^{(1)}/2 = 0,$$

$$\begin{aligned}
 L_2(x, y) &= (\boldsymbol{\mu}^{(2)})'V^{-1}\mathbf{z} - (\boldsymbol{\mu}^{(2)})'V^{-1}\boldsymbol{\mu}^{(2)}/2 \\
 &= (1, 2) \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} - \frac{1}{2}(1, 2) \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\
 &= (1, 2) \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \frac{1}{2}(1, 2) \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \\
 &= 2y - 2
 \end{aligned}$$

$$\begin{aligned}
 L_3(x, y) &= (\boldsymbol{\mu}^{(3)})' V^{-1} \mathbf{z} - \frac{1}{2} (\boldsymbol{\mu}^{(3)})' V^{-1} \boldsymbol{\mu}^{(3)} \\
 &= (-1/2, 1) \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
 &\quad - \frac{1}{2} (-1/2, 1) \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix} \begin{pmatrix} -1/2 \\ 1 \end{pmatrix} \\
 &= -\frac{4}{3}x + \frac{5}{3}y - \frac{7}{6}.
 \end{aligned}$$

- En particular, para $\mathbf{z} = (1, 0.9)'$ obtenemos

$$L_1(1, 0.9) = 0$$

$$L_2(1, 0.9) = 2(0.9) - 2 = -0.2$$

$$L_3(1, 0.9) = -\frac{4}{3} + \frac{5}{3}0.9 - \frac{7}{6} = -1,$$

- Por lo que \mathbf{z} se clasificará en la primera población (donde L_i es máxima).

- Para calcular las regiones de clasificación para los grupos 1 y 2 haremos $L_1 = L_2$, es decir,

$$2y - 2 = 0$$

obteniendo $y = 1$ (como ya vimos anteriormente usando la función discriminante de Fisher).

- Para los grupos 1 y 3, obtenemos

$$-\frac{4}{3}x + \frac{5}{3}y - \frac{7}{6} = 0,$$

es decir, $y = \frac{4}{5}x + \frac{7}{10}$.

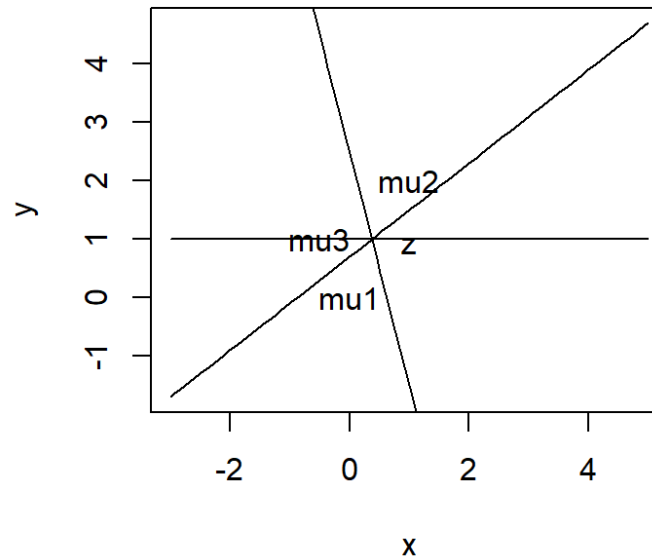
- Para los grupos 2 y 3, obtenemos

$$2y - 2 = -\frac{4}{3}x + \frac{5}{3}y - \frac{7}{6},$$

es decir, $y = -4x + \frac{5}{2}$.

- Regiones de clasificación para las poblaciones:

► Code



- Las tres rectas se cortan en el punto $(3/8, 1)$.

Varias poblaciones con distintas matrices de covarianza

Criterios de clasificación

- Los criterios de clasificación por **máxima verosimilitud** o por **mínima distancia de Mahalanobis a las medias de los grupos** pueden utilizarse aunque las poblaciones no tengan la misma matriz de covarianzas.
- No es necesario que las poblaciones sean normales, pudiéndose aplicar incluso a poblaciones de tipo discreto (siempre que se conozcan las densidades o las funciones puntuales de probabilidad).
- Cuando las poblaciones sean **normales** suele hablarse de **Análisis Discriminante Cuadrático** (ADC o QDA) ya que las funciones que determinan las regiones de clasificación son polinomios de grado 2.
 - Sin embargo, en este caso, las funciones discriminantes para mínima distancia o máxima verosimilitud no coinciden.

- Bajo la hipótesis de normalidad, el criterio de **máxima verosimilitud** buscará el máximo de

$$f_i(\mathbf{z}) = \frac{1}{\sqrt{|V_i|} (2\pi)^k} \exp \left(-\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu}^{(i)})' V_i^{-1} (\mathbf{z} - \boldsymbol{\mu}^{(i)}) \right)$$

o, equivalentemente, el **mínimo** de

$$Q_i(\mathbf{z}) = c - 2 \log f_i(\mathbf{z}) = (\mathbf{z} - \boldsymbol{\mu}^{(i)})' V_i^{-1} (\mathbf{z} - \boldsymbol{\mu}^{(i)}) + \log |V_i|,$$

con término constante $c = -k \log(2\pi)$, conocida como **función discriminante cuadrática (QDF)** para $i = 1, \dots, m$.

- Un individuo con medidas \mathbf{z} se clasificará en el grupo donde la función discriminante cuadrática sea mínima (máxima verosimilitud).

- El criterio basado en la **distancia de Mahalanobis** usará la función discriminante cuadrática

$$Q_i^*(\mathbf{z}) = d_{V_i}^2(\mathbf{z}, \boldsymbol{\mu}^{(i)}) = (\mathbf{z} - \boldsymbol{\mu}^{(i)})' V_i^{-1} (\mathbf{z} - \boldsymbol{\mu}^{(i)}).$$

Observaciones

- Los resultados pueden ser diferentes (cuando los determinantes de las matrices de covarianzas sean diferentes).
- En general, las funciones discriminantes cuadráticas son muy **sensibles** cuando las **poblaciones no son normales**, por lo que no es muy recomendable su uso en este caso, siendo preferible usar funciones discriminantes lineales.

Un ejemplo

- Sean dos poblaciones normales bidimensionales con medias $\boldsymbol{\mu}_1 = (2, 0)'$ y $\boldsymbol{\mu}_2 = (0, 0)'$, y matrices de covarianzas

$$V_1 = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}, \quad y \quad V_2 = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

- Se pide:
 - Calcular las funciones discriminantes cuadráticas,
 - Proporcionar el criterio de clasificación,
 - Dibujar las regiones de clasificación en \mathbb{R}^2 , y
 - Clasificar a $\mathbf{z} = (1, 1)'$.

- Como

$$V_1^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

la primera QDF es

$$\begin{aligned}
 Q_1(x, y) &= (x - 2 \quad y) V_1^{-1} \begin{pmatrix} x - 2 \\ y \end{pmatrix} \\
 &= (x - 2 + y)(x - 2) + (x - 2 + 2y)y \\
 &= x^2 - 4x + 4 + 2yx - 4y + 2y^2.
 \end{aligned}$$

- Análogamente, para el segundo grupo tenemos:

$$\begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix}^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$$

con lo que la segunda QDF será

$$Q_2(x, y) = (x \quad y)V_2^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = 0.5x^2 + 2y^2.$$

- Entonces $z = (x, y)$ se clasifica en 1 si y solo si $Q_1(x, y) < Q_2(x, y)$, es decir, si

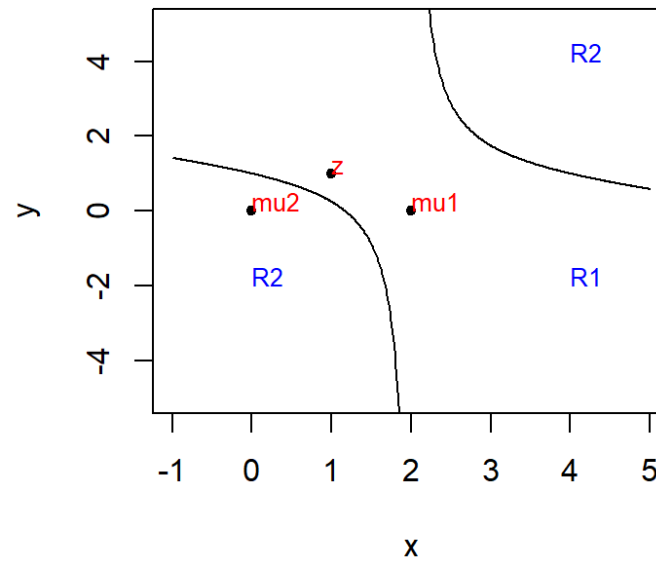
$$(2x - 4)y < 4x - 4 - x^2/2.$$

Como $Q_1(1, 1) = 1$ y $Q_2(1, 1) = 5/2$, se clasifica en 1.

- Para calcular las regiones de clasificación notamos que, en la ecuación anterior, tenemos tres casos:
 1. Si $x > 2$, $Q_1(x, y) < Q_2(x, y)$, si y solo si, $y < (4x - 4 - x^2/2)/(2x - 4)$.
 2. Si $x < 2$, $Q_1(x, y) < Q_2(x, y)$, si y solo si, $y > (4x - 4 - x^2/2)/(2x - 4)$.
 3. Si $x = 2$, $0 < 8 - 4 - 4/2 = 2$, por lo que $Q_1(x, y) < Q_2(x, y)$ y se clasifica en 1.

- Las regiones de clasificación:

► Code



Clasificación a partir de una muestra

- En la práctica, los valores de las **medias** y las **matrices de covarianzas teóricos** usados en los criterios de clasificación serán **desconocidos**, por lo que tendrán que ser **estimados**.
- Estas estimaciones dependerán de las hipótesis de partida (normalidad, igualdad de matrices de covarianzas, etc...), hipótesis que en muchos casos deberán ser corroboradas mediante algún procedimiento cuando no se está muy seguro de su validez.
- En general, si estudiamos k variables numéricas (Z_1, \dots, Z_k) en m poblaciones distintas indicadas por una variable discreta G (grupo) para una muestra de n individuos (muestra de entrenamiento), tendremos una tabla de datos de la forma

	Z_1	\dots	Z_k	G
ω_1	$z_{1,1}$	\dots	$z_{1,k}$	g_1
\dots	\dots	\dots	\dots	\dots
ω_n	$z_{n,1}$	\dots	$z_{n,k}$	g_n

donde $g_i \in \{1, \dots, m\}$ para todo i .

- Para cada valor de $G = j$, esta tabla proporcionará una m.a.s. de la variable aleatoria k -dimensional $\mathbf{Z} = (Z_1, \dots, Z_k)'$ condicionada por $G = j$ que, en muchas ocasiones, podremos suponer normal.
 - En realidad tendremos m muestras de m poblaciones normales k -dimensionales $\mathcal{N}_k(\boldsymbol{\mu}^{(j)}, V_j)$.
- Así, en la práctica, tendremos m medias teóricas y m matrices de covarianzas desconocidas,
 - por lo que tendremos que estimarlas.

- Para dichas **estimaciones** usaremos:

$$\hat{\boldsymbol{\mu}}^{(j)} = \frac{1}{n_j} \sum_{i=1}^n \boldsymbol{\omega}_i \mathbf{1}(G(\boldsymbol{\omega}_i) = j)$$

$$\hat{V}_j = \frac{1}{n_j - 1} \sum_{i=1}^n \mathbf{1}(G(\boldsymbol{\omega}_i) = j) (\boldsymbol{\omega}_i - \hat{\boldsymbol{\mu}}^{(j)}) (\boldsymbol{\omega}_i - \hat{\boldsymbol{\mu}}^{(j)})'$$

$$\boldsymbol{\omega}_i = (z_{i,1}, \dots, z_{i,k})'$$

$$n_j = \sum_{i=1}^n \mathbf{1}(G(\boldsymbol{\omega}_i) = j),$$

donde $n = n_1 + \dots + n_m$, y $\mathbf{1}(G(\boldsymbol{\omega}_i) = j)$ es la función indicador que vale 1 si el individuo i pertenece a la clase j -ésima y vale 0 en caso contrario. Y, por lo tanto, n_j es el número de individuos de la muestra pertenecientes a la población j -ésima.

- Necesitamos $n_j > 1$ para todo j .

- Si $(\mathbf{Z}|G = j)$ es normal, $\hat{\mathbf{V}}_j$ es insesgado para \mathbf{V}_j , teniendo $(n_j - 1)\hat{\mathbf{V}}_j$ una distribución (en el muestreo) Wishart $\mathbf{W}_k(n_j - 1, \mathbf{V}_j)$.
- Si suponemos que las **matrices de covarianzas teóricas son todas iguales** ($\mathbf{V}_1 = \dots = \mathbf{V}_m = \mathbf{V}$), entonces la matriz de covarianzas común \mathbf{V} se aproximará mediante la matriz de covarianzas ponderada (**pooled**)

$$\hat{\mathbf{V}} = \frac{1}{n - m} \sum_{j=1}^m (n_j - 1) \hat{\mathbf{V}}_j$$

que será un estimador insesgado para \mathbf{V} .

- A partir de estos estimadores se pueden obtener **estimaciones de las distintas funciones discriminantes** y con ellas, obtener **clasificaciones (empíricas) para nuevos individuos**.
- Como los estimadores se aproximan a los verdaderos valores de los parámetros, las clasificaciones se parecerán (si n es grande) a las que se obtendrían usando los verdaderos parámetros.

- Por ejemplo, para el caso de dos poblaciones con la misma matriz de covarianzas, la **función discriminante de Fisher se estimará** mediante

$$\hat{D} = \hat{L}(\mathbf{Z}) = \hat{a}'\mathbf{Z} = (\hat{\mu}_X - \hat{\mu}_Y)'\hat{V}^{-1}\mathbf{Z}.$$

→ De esta forma, la **probabilidad del error tipo 1 se estimará** mediante

$$\Pr(e_1) = \Pr\left(U < -\frac{1}{2}\hat{d}_V\right),$$

donde $U \equiv N_1(0, 1)$ y

$$\hat{d}_V = \sqrt{(\hat{\mu}_X - \hat{\mu}_Y)'\hat{V}^{-1}(\hat{\mu}_X - \hat{\mu}_Y)}$$

es la **distancia de Mahalanobis muestral entre la medias (muestrales) de los grupos**.

- Bajo la **hipótesis de normalidad**, estos estimadores son asintóticamente insesgados.

- En Srivastava y Carter (1983, pag. 238) pueden verse otros estimadores basados en las distribuciones obtenidas para los distintos errores a partir de la hipótesis de normalidad.
- Se procederá de forma similar en los otros casos.
- Por ejemplo, las **Funciones Discriminantes Lineales (FDL) muestrales** serán

$$\hat{L}_i(\mathbf{z}) = (\hat{\boldsymbol{\mu}}^{(i)})' \hat{\mathbf{V}}^{-1} \mathbf{z} - (\hat{\boldsymbol{\mu}}^{(i)})' \hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\mu}}^{(i)} / 2,$$

clasificándose \mathbf{Z} en G_i si $\hat{L}_i(\mathbf{z}) \geq \hat{L}_j(\mathbf{z})$ para todo j .

- Análogamente, las **funciones discriminantes cuadráticas (FDC o QDF) muestrales** serán

$$\hat{Q}_i(\mathbf{z}) = c - 2 \log \hat{f}_i(\mathbf{z}) = (\mathbf{z} - \hat{\boldsymbol{\mu}}^{(i)})' \hat{\mathbf{V}}_i^{-1} (\mathbf{z} - \hat{\boldsymbol{\mu}}^{(i)}) + \log | \hat{\mathbf{V}}_i |,$$

clasificándose \mathbf{Z} en G_i si $\hat{Q}_i(\mathbf{z}) \leq \hat{Q}_j(\mathbf{z})$ para todo j .