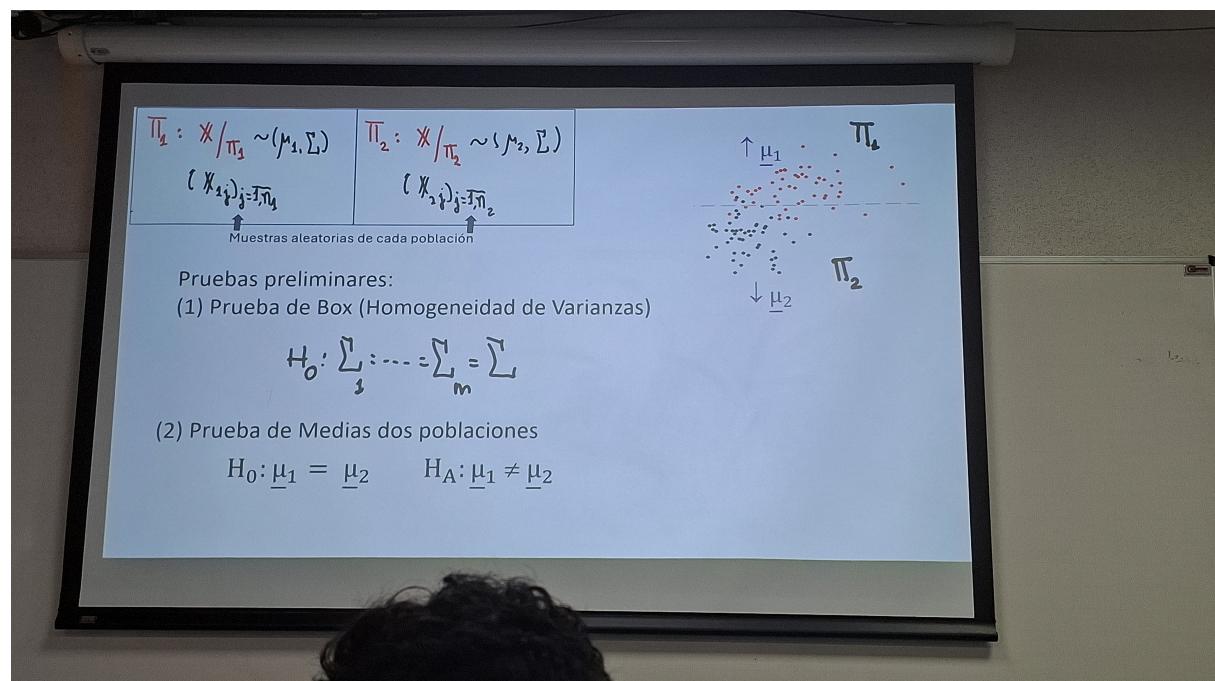
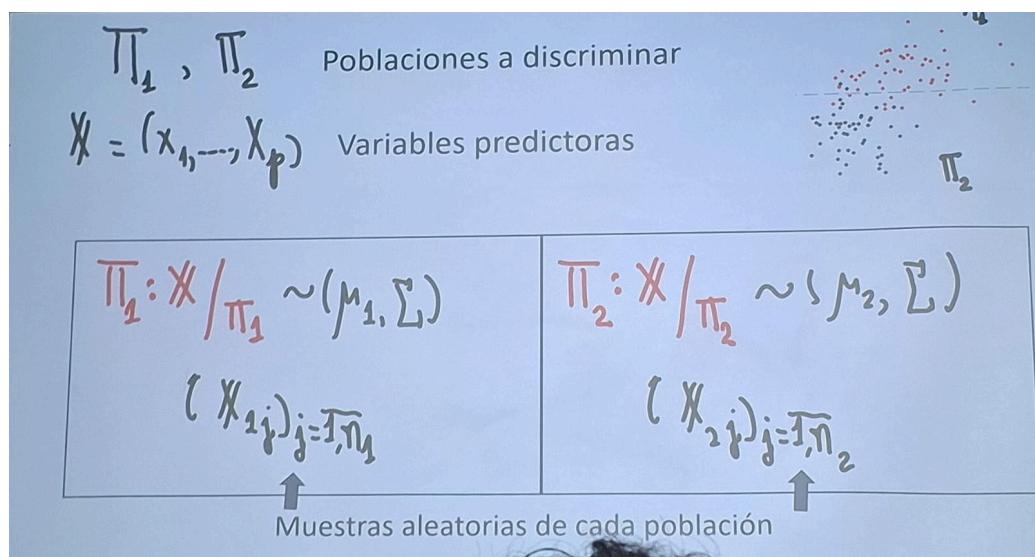
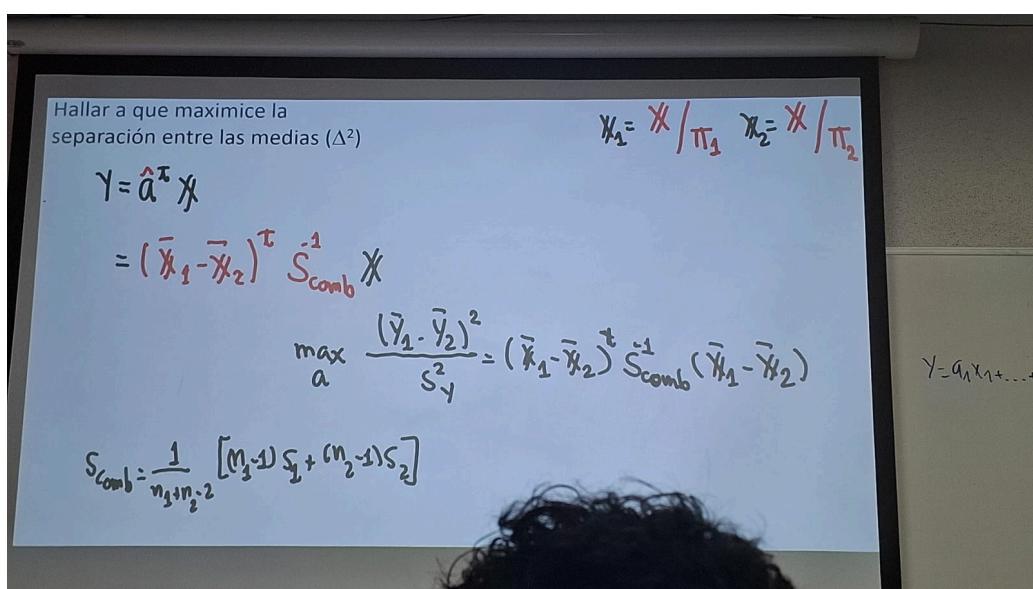
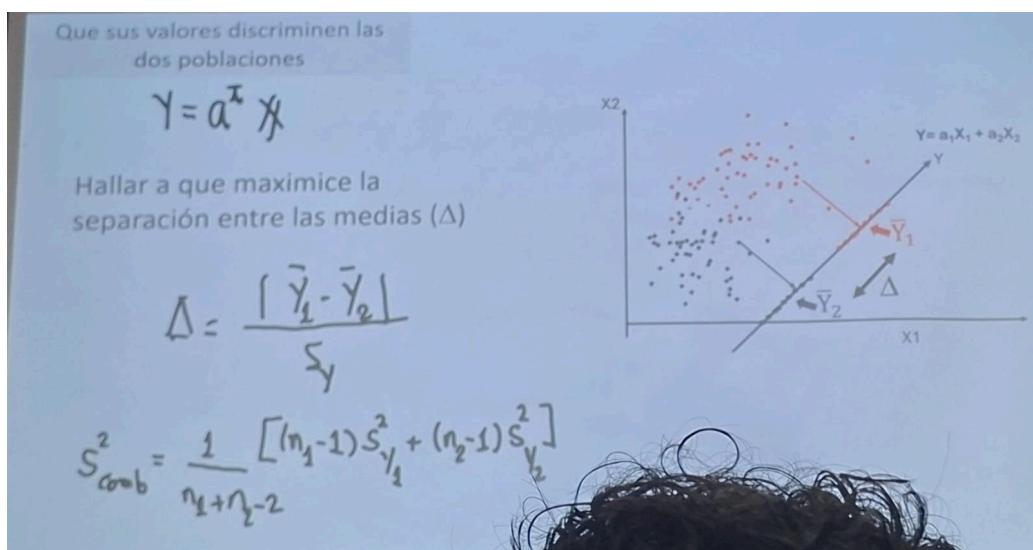
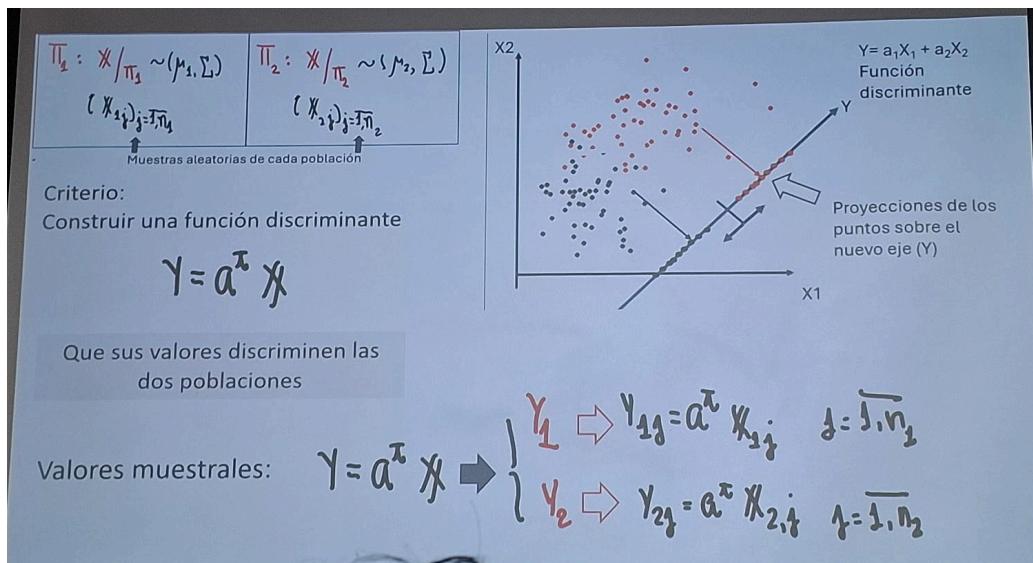


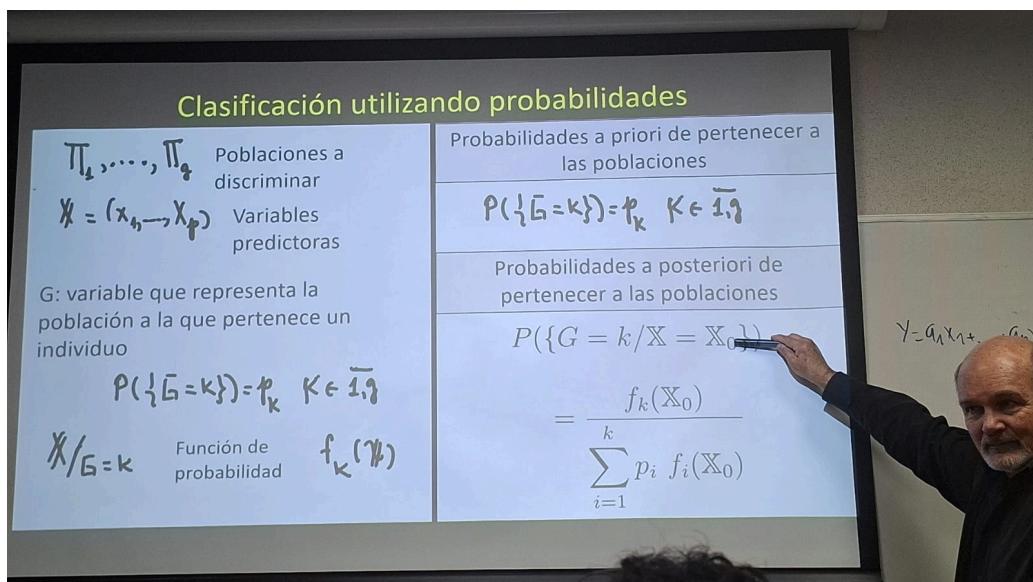
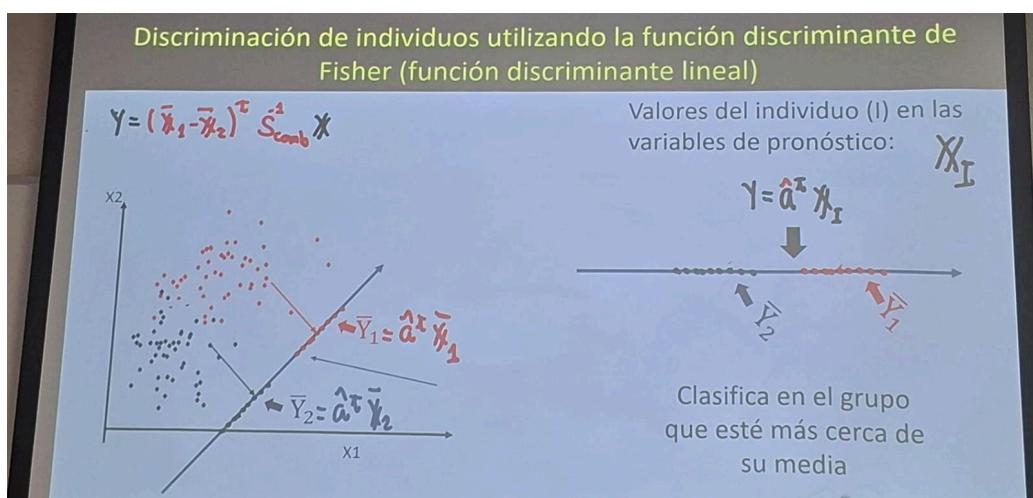
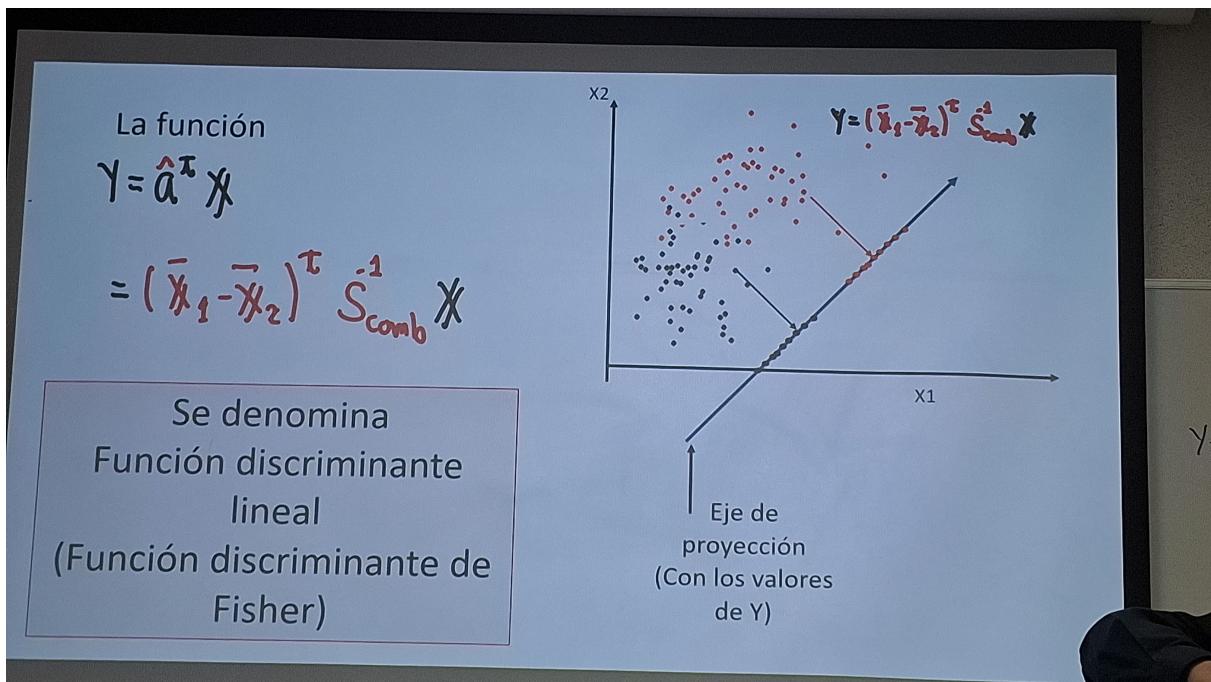
# Analisis discriminante

① Creadas	@13 de mayo de 2025 11:21
② Asignatura	Analisis Multivariado

Lineal con FISHER







LA funcion discriminante lineal(de fisher) casos multiples poblaciones ¿Como clasificar individuos?

$\pi_1, \dots, \pi_g$  Poblaciones a discriminar

$\mathbf{x} = (x_1, \dots, x_p)$  Variables predictoras

$\pi_i: \mathbf{x} / \pi_i \sim (\mu_i, \Sigma)$

$x_i = \mathbf{x} / \pi_i$

$(x_{ij})_{j=1, n_i}$

↑

Muestras aleatorias de cada población

Pruebas preliminares:

- (1) Prueba de Box (Homogeneidad de Varianzas)

$H_0: \sum_j \dots \sum_m = \sum$

(2) Prueba de Medias p poblaciones

$H_0: \mu_1 = \dots = \mu_p$

$H_A: \text{alguna (s) diferentes}$

Para estimar el valor de los coeficientes (cargas de la función discriminante) primero hay que hacer una descomposición de la matriz de varianzas y covarianzas

Muestras aleatorias de cada población

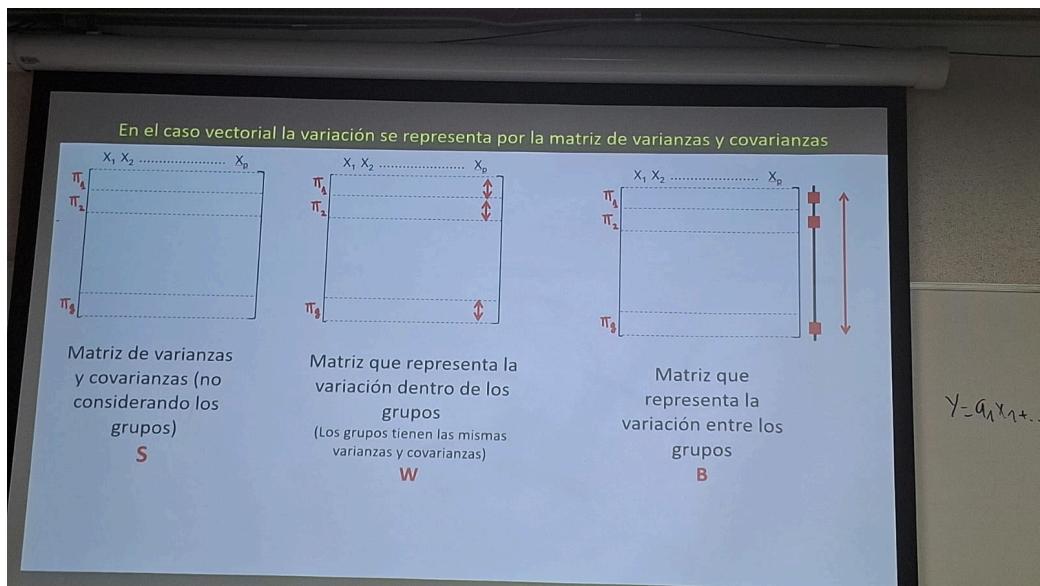
$i = 1, g$

$x_i = \mathbf{x} / \pi_i$

$(x_{ij})_{j=1, n_i}$

Matriz de datos particionada por filas

	$x_1$	$x_2$	.....	$x_p$
$\pi_1$				
$\pi_2$				
$\vdots$				
$\pi_g$				



$$S = W + B$$

$W$  es mejor ariacion pequeña

$B$  es mejor variacion grande

Construir una función discriminante  $\gamma = \alpha^T \chi$

$S = W + B$

$\text{cov}(\chi) = S$

$\hat{\text{var}}(\gamma) = \alpha^T S \alpha$

$\text{Var}_x(\gamma) = \alpha^T W \alpha + \alpha^T B \alpha$

Variación Total Variación dentro ↓ Variación entre ↑

Criterio para determinar  $A$

$\hat{\alpha} = \alpha \text{ s.t. } \max_{\alpha} \frac{\alpha^T B \alpha}{\alpha^T W \alpha}$

Variación entre ↑ Variación dentro ↓

Por ese motivo el criterio para determinar A es

$$\hat{a} = \underset{a}{\operatorname{Arg\,max}} \frac{a^T B a}{a^T W a}$$

Para  $c > 1$

$$a^T X < \underbrace{c a^T}_{b^T} y$$

$\hat{a} = \underset{a}{\operatorname{Arg\,max}} (a^T B a)$

Condición  $a^T W a = 1$

Solución

Vector propio (e) asociado al mayor valor propio ( $\lambda$ ) de  $W^{-1} B$

$\Pi_1, \dots, \Pi_g$  Poblaciones a discriminar

$X = (x_1, \dots, x_p)$  Variables predictoras

$W^{-1} B$

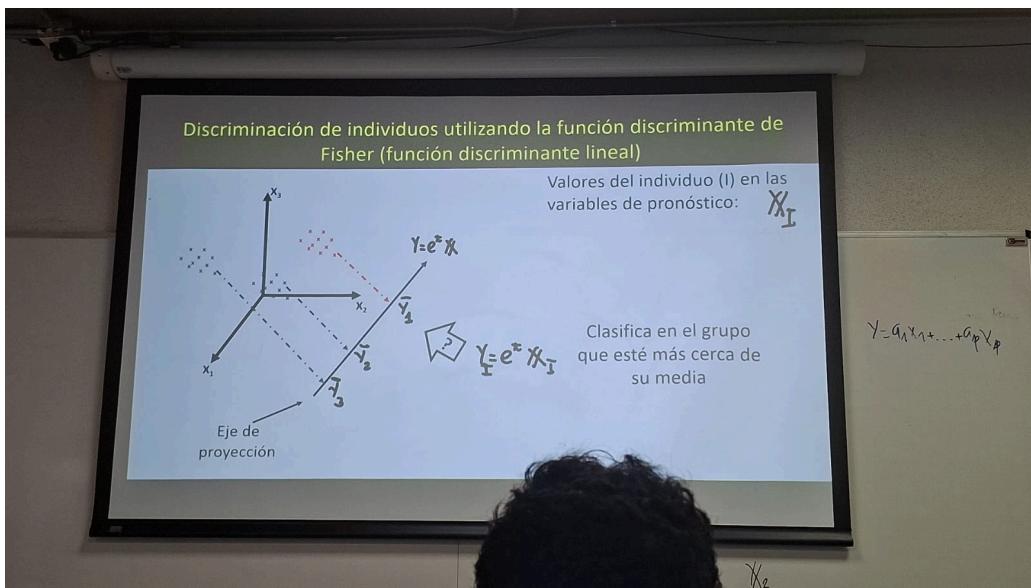
$(\lambda, e)$

Mayor valor propio

Vector valor propio asociado

Función discriminante de Fisher

$$Y = e^T X$$



Siguiendo la lógica del ACP

$$\hat{a} = \underset{a}{\operatorname{argmax}} (a^T B a)$$

Condición

$$a^T W a = 1$$

Se pueden considerar varios ejes de proyección

$$(\lambda_i, e_i) \quad i=1, \dots, s = \min(q-1, p)$$

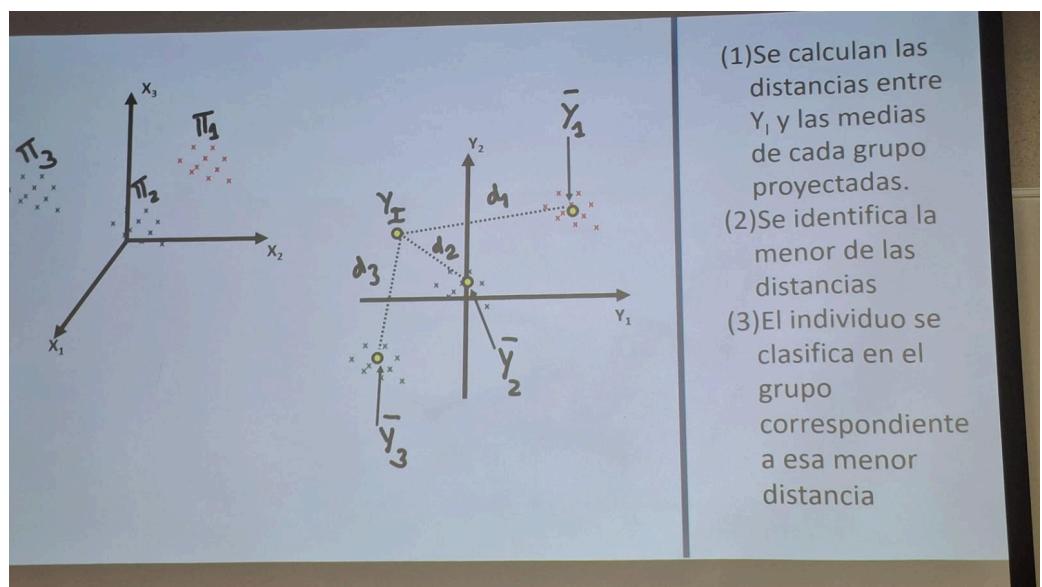
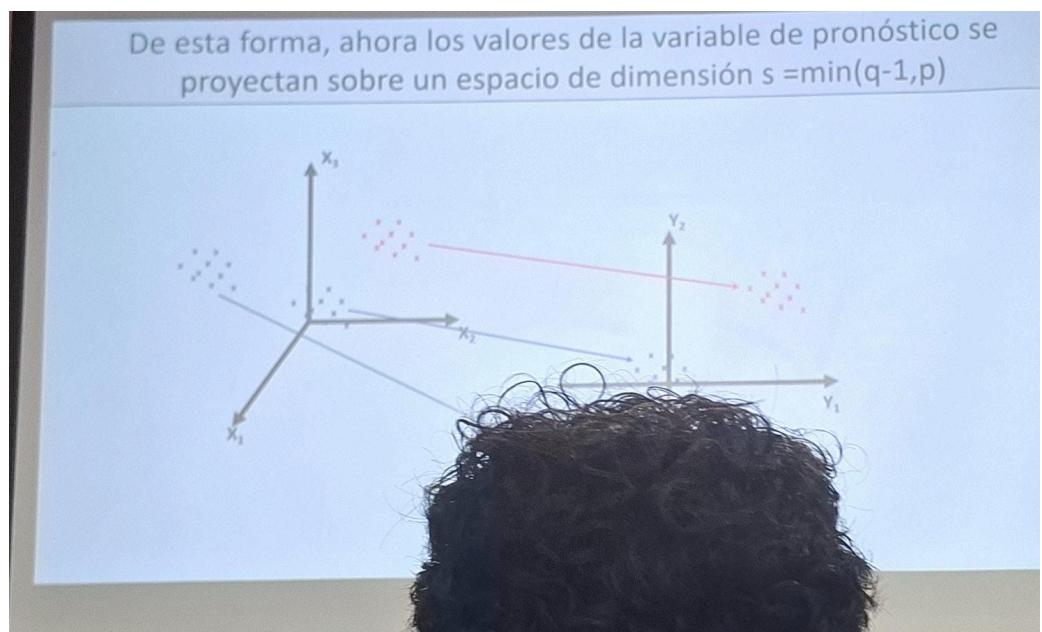
$$Y_i = e_i^T X$$

De esta forma se tiene una familia de proyecciones cada una de las cuales (de forma descendente) introduce la mayor separación entre los grupos

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_s \end{bmatrix} = \begin{bmatrix} e_1^T X \\ \vdots \\ e_s^T X \end{bmatrix} = \begin{bmatrix} e_1^T \\ \vdots \\ e_s^T \end{bmatrix} X = A X$$

De esta forma, ahora los valores de la variable de pronóstico se proyectan sobre un espacio de dimensión  $s = \min(q-1, p)$

$q=2, s=1$



Calidad (importancia) de la separación entre grupos que se logra con cada función discriminante

Para la función discriminante  $i$ , utilizando la clasificación de los individuos que con ella se lograría, se calcula la **separación entre grupos** que se lograría con ella

$$\text{Separ}_i = \boxed{\quad}$$

Para cuantificar la importancia de la función discriminante  $i$ , relativa a las demás se calcula la proporción **traza**

$$\frac{\text{Separ}_i}{\sum_i \text{Separ}_i} (*100)$$

Mientras sea mayor la separación que se logra con la función discriminante más importante es la función

### EJEMPLO: Cangrejos

probabilidad que un individuo aparesca en un grupo prior proba

Numero de funciones discriminantes LD; se determina  $S = \min(q-1, p)$  en este caso 4 - 1 , 5

La proporción de la traza: cuando separa los grupos cada función discriminante

Prior probabilities of groups:

1	2	3	4
0.25	0.25	0.25	0.25

Group means:

	FL	RW	CL	CW	BD
1	14.842	11.718	32.014	36.810	13.350
2	13.270	12.138	28.102	32.624	11.816
3	16.626	12.262	33.688	37.188	15.324
4	17.594	14.836	34.618	39.036	15.632

Coefficients of linear discriminants:

	LD1	LD2	LD3
FL	-1.5543139	-0.1951885	-1.6667377
RW	-0.6247546	-1.5394972	0.4558782
CL	-0.1875489	1.0953923	0.6807588
CW	1.5156077	-0.6435178	-0.6548549
BD	-1.3551090	0.5153193	1.2859743

Proportion of trace:

	LD1	LD2	LD3
0.6861	0.2995	0.0144	

Separación

color sexo grupo

B	M	1
B	F	2
O	M	3
O	F	4

Prior probabilities of groups:

1	2	3	4
0.25	0.25	0.25	0.25

Group means:

	FL	RW	CL	CW	BD
1	14.842	11.718	32.014	36.810	13.350
2	13.270	12.138	28.102	32.624	11.816
3	16.626	12.262	33.688	37.188	15.324
4	17.594	14.836	34.618	39.036	15.632

Coefficients of linear discriminants:

	LD1	LD2	LD3
FL	-1.5543139	-0.1951885	-1.6667377
RW	-0.6247546	-1.5394972	0.4558782
CL	-0.1875489	1.0953923	0.6807588
CW	1.5156077	-0.6435178	-0.6548549
BD	-1.3551090	0.5153193	1.2859743

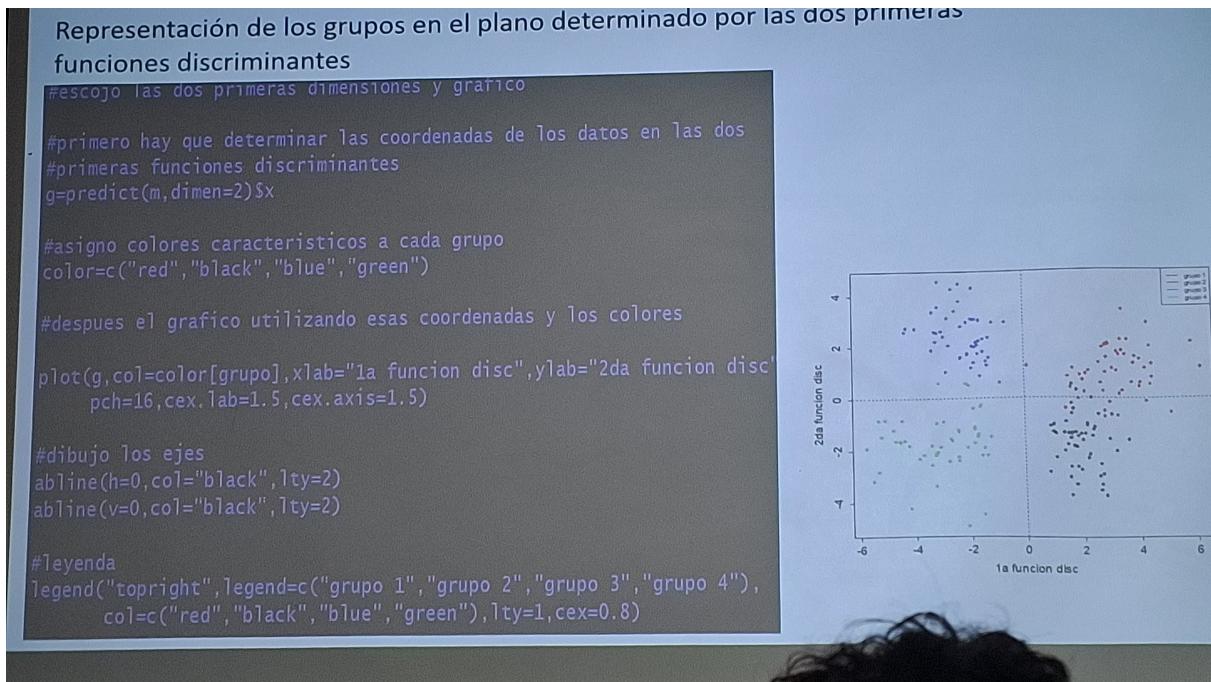
Proportion of trace:

	LD1	LD2	LD3
0.6861	0.2995	0.0144	

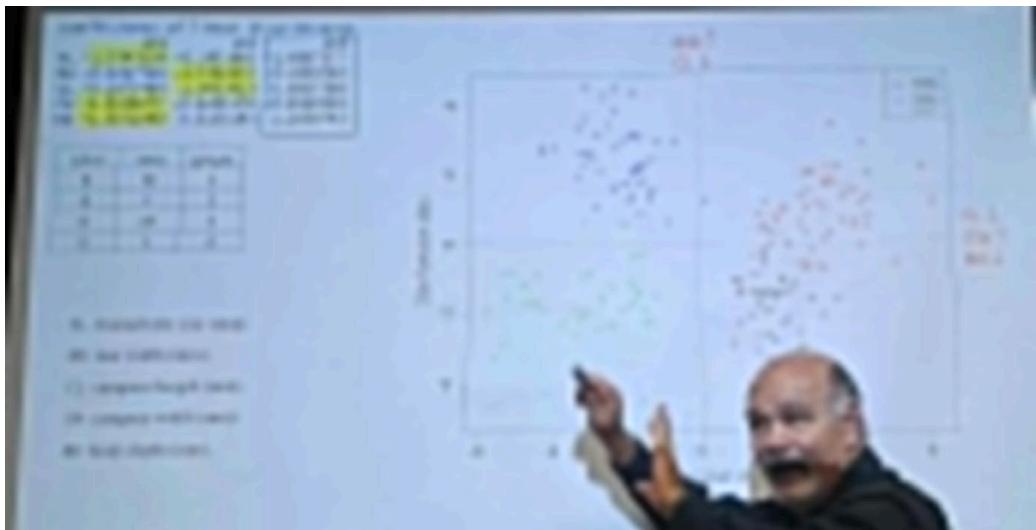
Separación

color sexo grupo

B	M	1
B	F	2
O	M	3
O	F	4



## INTERPRETAR LOS EJES Y LOS GRUPOS



## 📌 Similitudes entre LDA y PCA:

Característica común	Descripción
----------------------	-------------

▼ Reducción de dimensiones	Ambos pueden proyectar los datos a un espacio de menor dimensión.
▼ Transformación lineal	Ambos usan combinaciones lineales de las variables originales.
▼ Preprocesamiento común	Ambos suelen aplicarse después de escalar o centrar los datos.

## 🔍 Diferencias clave:

Característica	PCA	LDA
Tipo de algoritmo	No supervisado	Supervisado
Usa etiquetas	✗ No	✓ Sí
Objetivo	Maximizar la <b>varianza total</b>	Maximizar la <b>separación entre clases</b>
En qué se enfoca	Las direcciones con mayor variabilidad	Las direcciones que <b>mejor separan las clases</b>
Resultado	Componentes que explican la varianza	Componentes que explican la discriminación
Clasifica	✗ No (solo reduce dimensiones)	✓ Sí (puede clasificar nuevos datos)

## 🎯 Analogía simple:

Imagina que tienes datos de dos especies de flores:

- **PCA:** ignora qué especie es cada flor, y solo busca direcciones que expliquen la mayor variación total entre todas las flores.
- **LDA:** **usa las etiquetas de especie** y busca direcciones que **separen mejor una especie de otra**.

## 📈 Visualmente:

- **PCA:**  
Agrupa las variables en componentes principales que capturan la mayor varianza, sin importar de qué clase son los datos.
- **LDA:**

Busca una línea (o plano) que maximice la distancia entre los **centros de clases** y minimice la dispersión **dentro** de cada clase.

---

## ✓ En resumen:

- ◆ PCA → Proyecta datos para explorar estructura general.
- ◆ LDA → Proyecta datos con el fin de **clasificarlos** eficazmente.

Entonces, sí: puedes pensar en LDA como "PCA con conciencia de clases", y además con capacidad de **clasificación**.

Se determina el numero de ejes con  $\min(q-1,p)$  eso lo hace internamente donde q es el numero de poblaciones y p el numero de variables

En cada uno de los problemas

- (1) Determine las funciones discriminantes que serían necesarias para representar las poblaciones consideradas
- (2) Analice el ajuste del modelo discriminante que construyó comparando la predicción de pertenencia a los grupos con la realidad
- (3) Utilice las funciones discriminantes para representar gráficamente los grupos y caracterizarlos

#### Base de datos vino

Type The type of wine, into one of three classes, 1 (59 obs), 2(71 obs), and 3 (48 obs).

Alcohol Alcohol

Malic Malic acid

Ash Ash

Alkalinity Alkalinity of ash

Magnesium Magnesium

Phenols Total phenols

Flavanoids Flavanoids

Nonflavanoids Nonflavaneoid phenols

Proanthocyanins Proanthocyanins

Color Color intensity.

Hue Hue

Dilution D280/OD315 of diluted wines.

Proline Proline

#### (4) Un vino que tenga

Alcohol Alcohol **15**

Malic Malic acid **2**

Ash Ash **2.5**

Alkalinity Alkalinity of ash **15**

Magnesium Magnesium **112**

Phenols Total phenols **4**

Flavanoids Flavanoids **38**

Nonflavanoids Nonflavaneoid phenols **0.5**

Proanthocyanins Proanthocyanins **2**

Color Color intensity. **7**

Hue Hue **1.5**

Dilution D280/OD315 of diluted wines. **3**

Proline Proline **1500**

[En cuál de los tipos clasificaría](#)