

Unidad 3

Análisis Discriminante

• PROFESOR: ARTURO ZUÑIGA

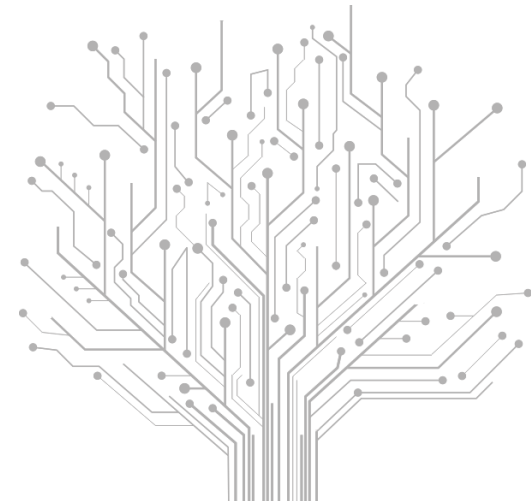
Contenido

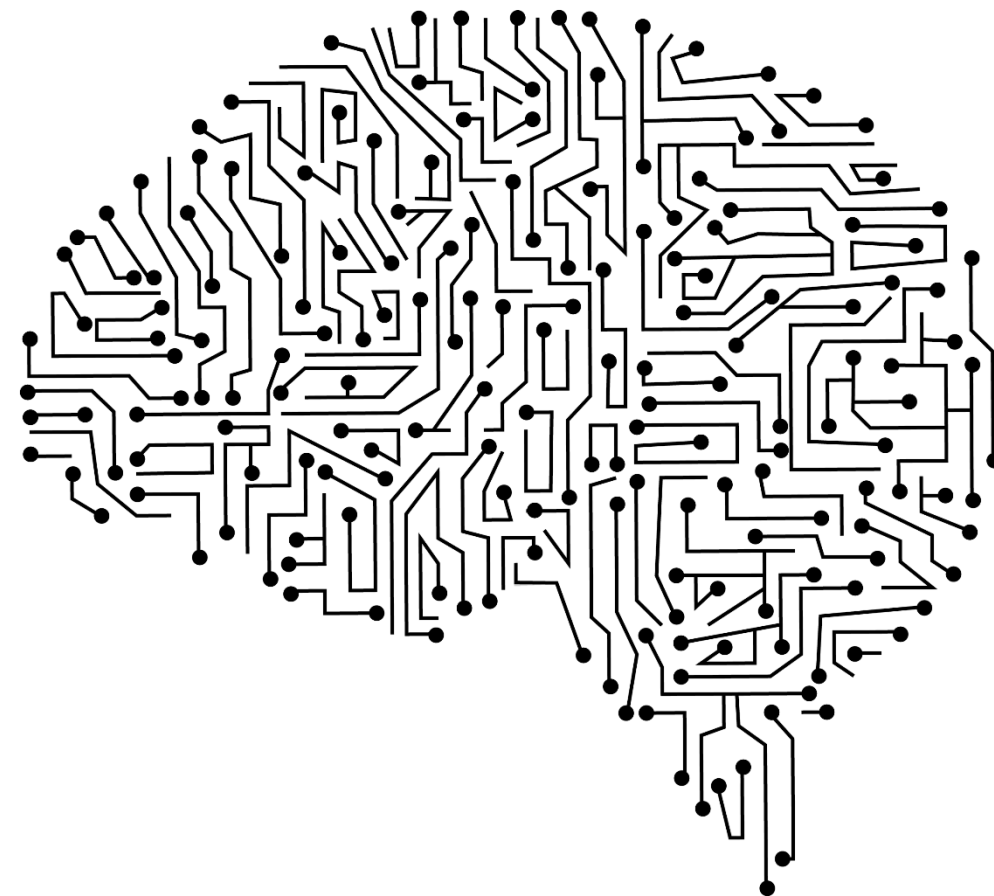
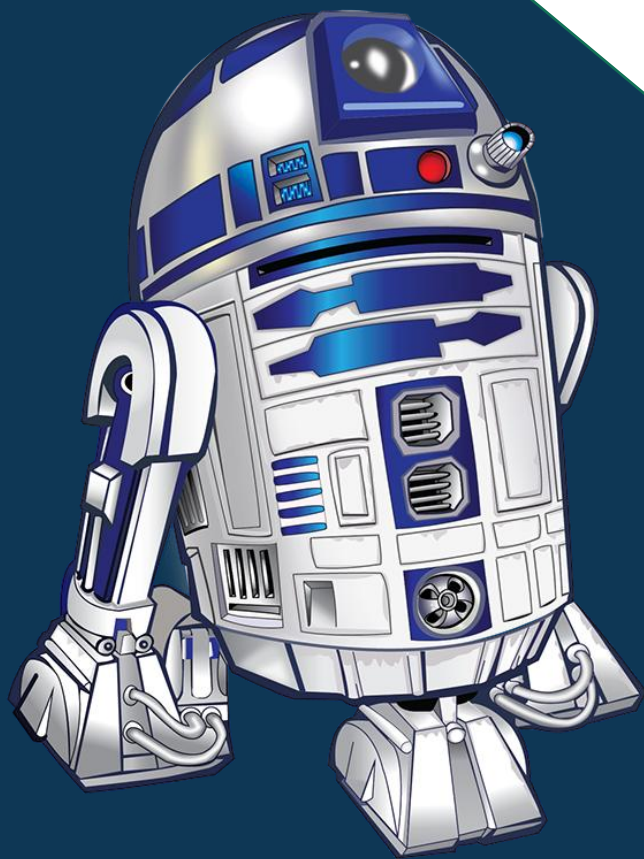
1. Análisis discriminante



contenidos

1. Introducción
2. Metodología Estadística
 - 2.1 Pruebas estadísticas preliminares
 - 2.2 Métodos de discriminación
 - Función discriminante lineal de Fisher
 - Función discriminante de Máxima verosimilitud
 - Análisis de regresión discriminante
 - Distancia de Mahalanobis
 - 2.3 Evaluación de la función discriminante
 - 2.4 Evaluación de la clasificación
 - 2.5 Clasificación de nuevas observaciones
 - 2.6 Análisis discriminante con más de dos grupos
3. Ejemplo de aplicación





Análisis discriminante

Introducción



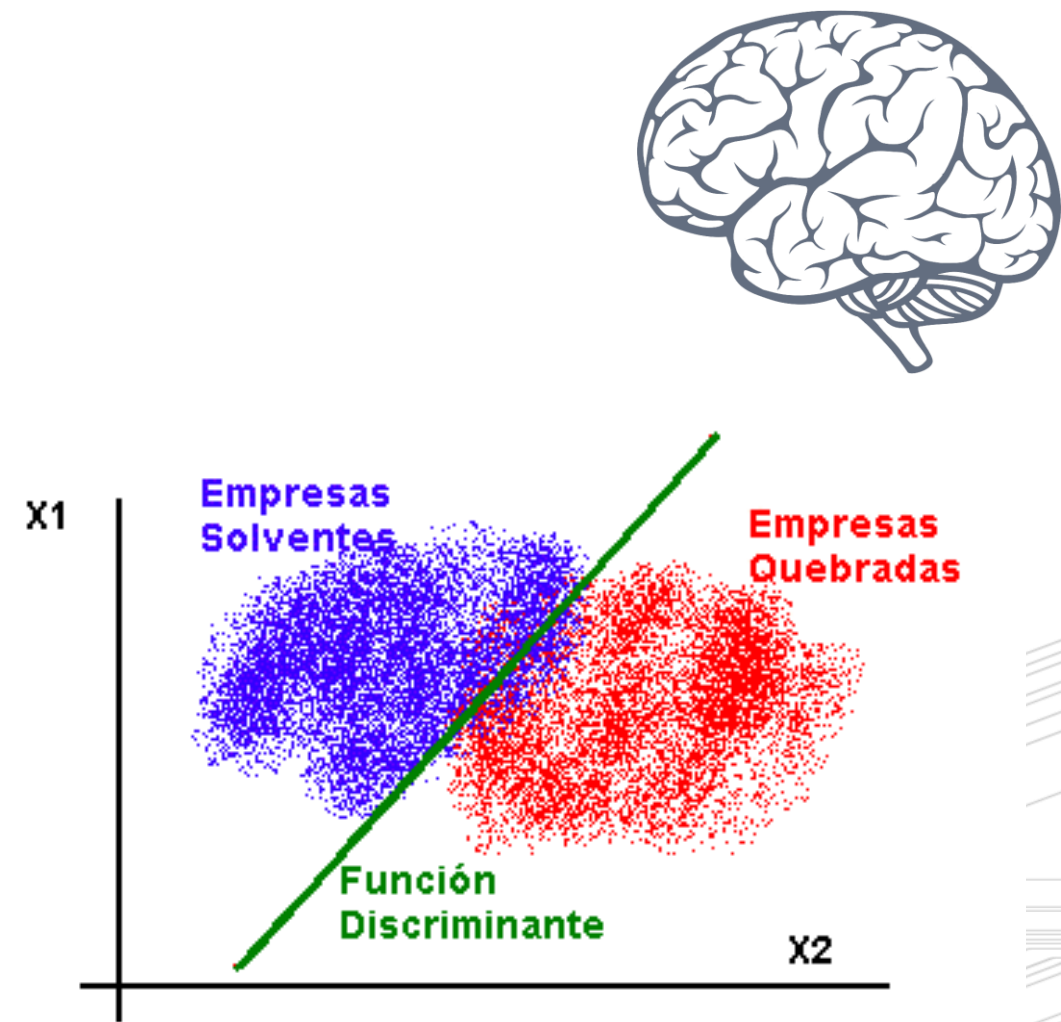
- ❑ Es un método multivariado cuyo **propósito es el de clasificar nuevos individuos u objetos en grupos o poblaciones previamente definidas**, para lo cual se desarrollan reglas o funciones discriminantes que son construidas en base del conjunto de variables que son medidas sobre los individuos u objetos.
- ❑ El AD, es considerado un **método de dependencia**; las variables que permiten clasificar son llamadas clasificadoras, predictoras o explicativas (variables independientes), mientras que se usa una variable categórica (variable dependiente) para identificar al grupo que pertenece el individuo

Introducción

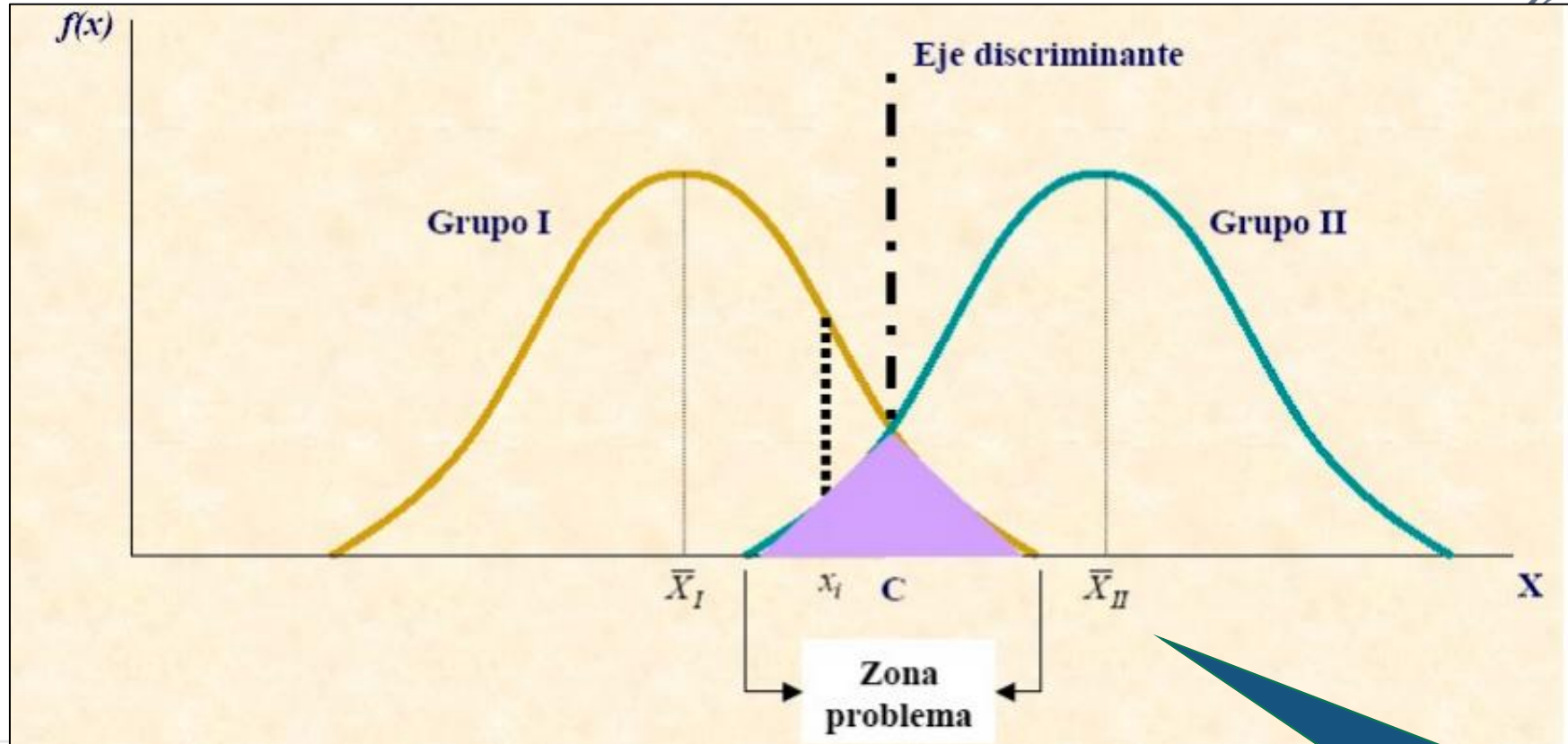
El AD, se usa para fines explicativos (identificar variables que mejor discriminen entre los grupos), predictivos (seleccionar variables) y reclasificación (nuevas observaciones).

El AD, esta afectado por las unidades de medida de las variables explicativas.

En el AD, se puede distinguir el **análisis discriminante simple**, cuando la variable dependiente considera dos categorías (dos grupos: $G=2$) y el **análisis discriminante múltiple**, cuando existen más de dos categorías (más de dos grupos: $G>2$).



Dos grupos una variable clasificadora



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

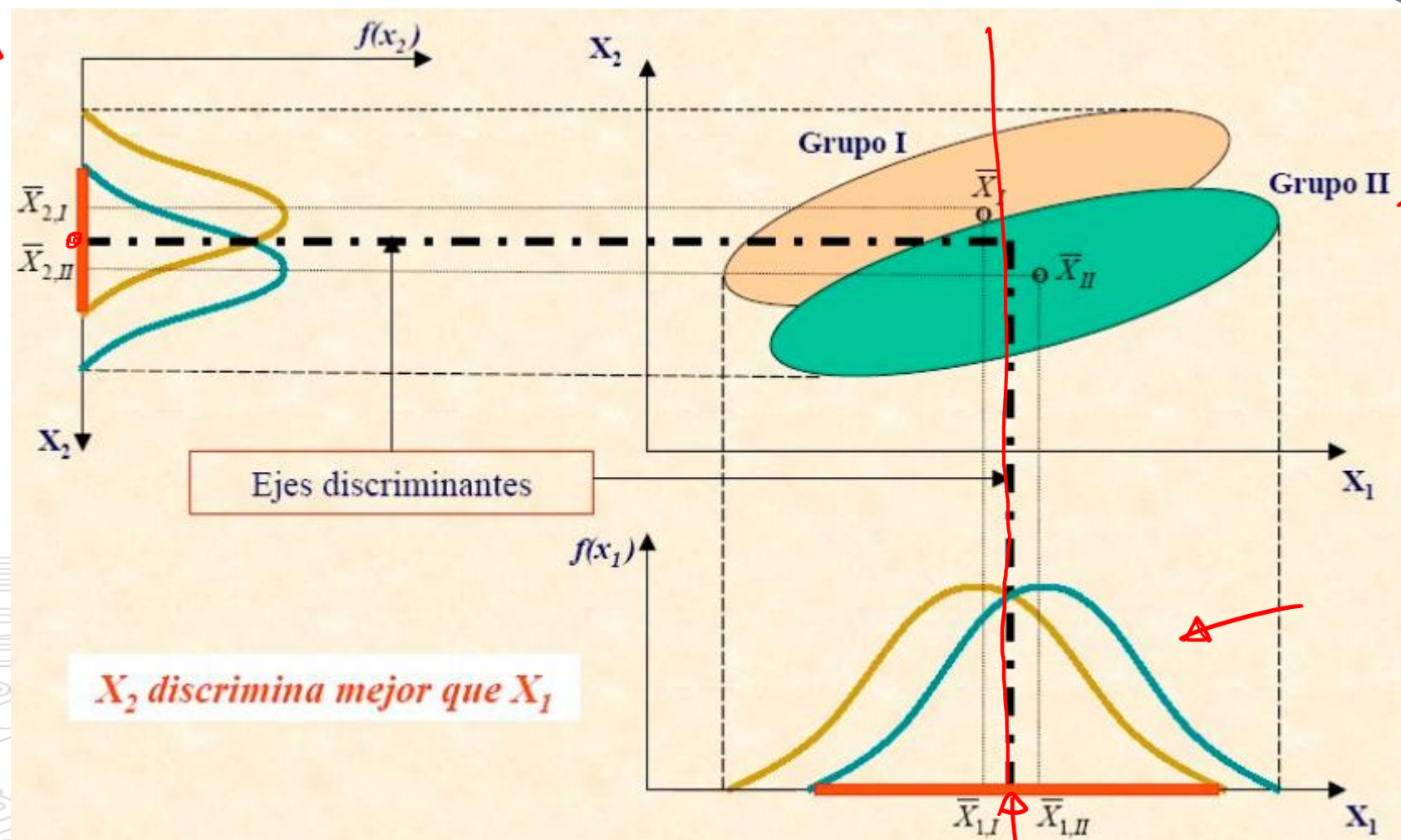
Hipótesis: Las distribuciones sólo se diferencian por su localización (igual forma y varianza)

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

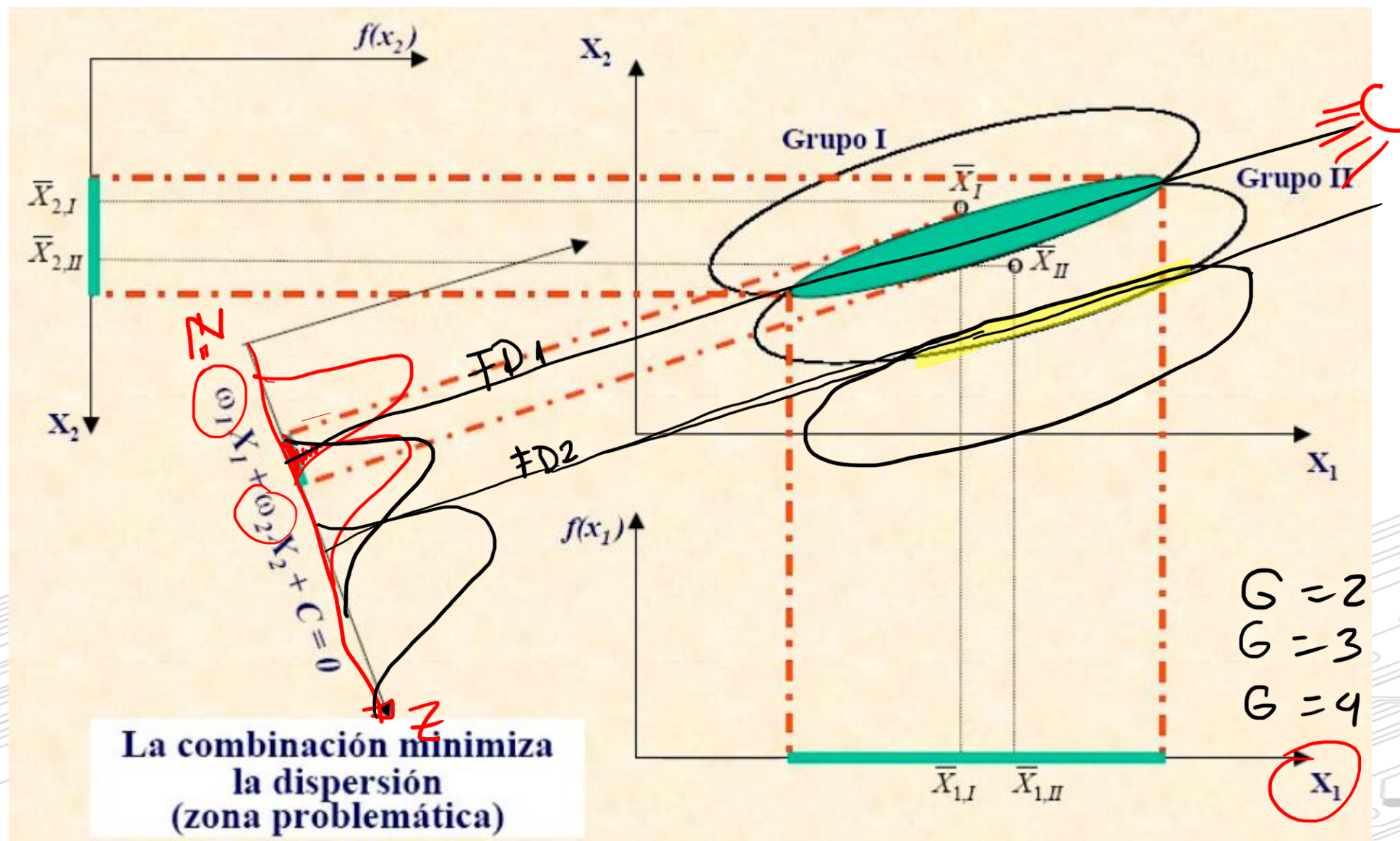
El punto C se denomina *punto de corte discriminante*:
$$C = \frac{\bar{X}_I + \bar{X}_{II}}{2}$$

Se busca clasificar cada observación en el grupo correcto, según el valor de la variable clasificadora

**Dos
grupos
de
variables
clasifica
doras**



11°C



Características

La diferencia principal entre ambos (ADS Y ADM), es el número de funciones discriminantes que se pueden obtener. En el ADS se obtiene sólo una función discriminante; mientras que el ADM se pueden obtener el *Mínimo*($G-1, p$) funciones discriminantes, donde $p = N^{\circ}$ de variables indep.

En el AD es posible retener o seleccionar un grupo del conjunto de variables independientes para obtener la función discriminante. Se emplean técnicas de selección de variables que son aplicables en el análisis de regresión lineal múltiple: selección hacia adelante (forward), hacia atrás (backward) y por pasos (stepwise).



Pruebas estadísticas preliminares I

1) Prueba de homogeneidad de matrices de variancias-covariancias.

Formulación de las hipótesis:

$$H_0 : \Sigma_1 = \Sigma_2 = \dots \Sigma_g = \Sigma$$

H_1 : Al menos una matriz diferente

Prueba estadística. Se utiliza la M de Bartlett-Box:

$$M = \frac{\prod_{g=1}^K |S_g|^{\frac{n_g - 1}{2}}}{|\bar{S}|^{\frac{n - K}{2}}}$$

$$S_g = \hat{\Sigma}_g = \frac{W_g}{n_g - 1}$$

$$\bar{S} = \hat{\Sigma} = \frac{\sum_{g=1}^G W_g}{n - G} = \frac{\sum_{g=1}^G (n_g - 1)}{n - G}$$

Decisión estadística. Se han desarrollados aproximaciones a la Chi-cuadrado y F.



Pruebas estadísticas preliminares II

- 2) Probar a priori el poder discriminante de cada variable.

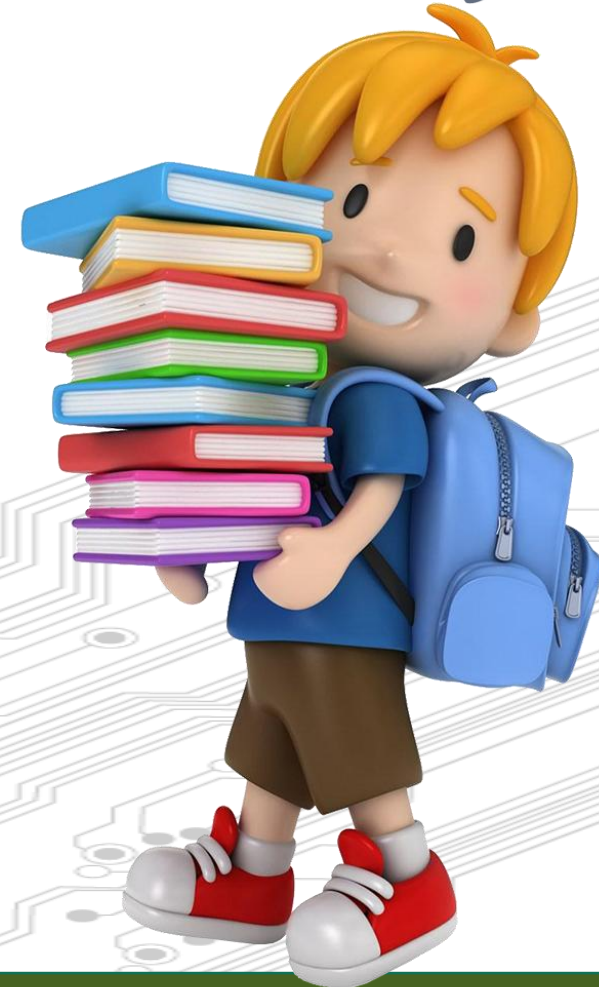
Formulación de las hipótesis:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_g$$

H1 : Al menos una media diferente

Prueba estadística. Se utiliza la estadística de Lambda de Willks o la F.

Se espera que se rechace H_0 para cada variable, indicando que hay diferencias significativas entre grupos (las variables permitirán una discriminación)



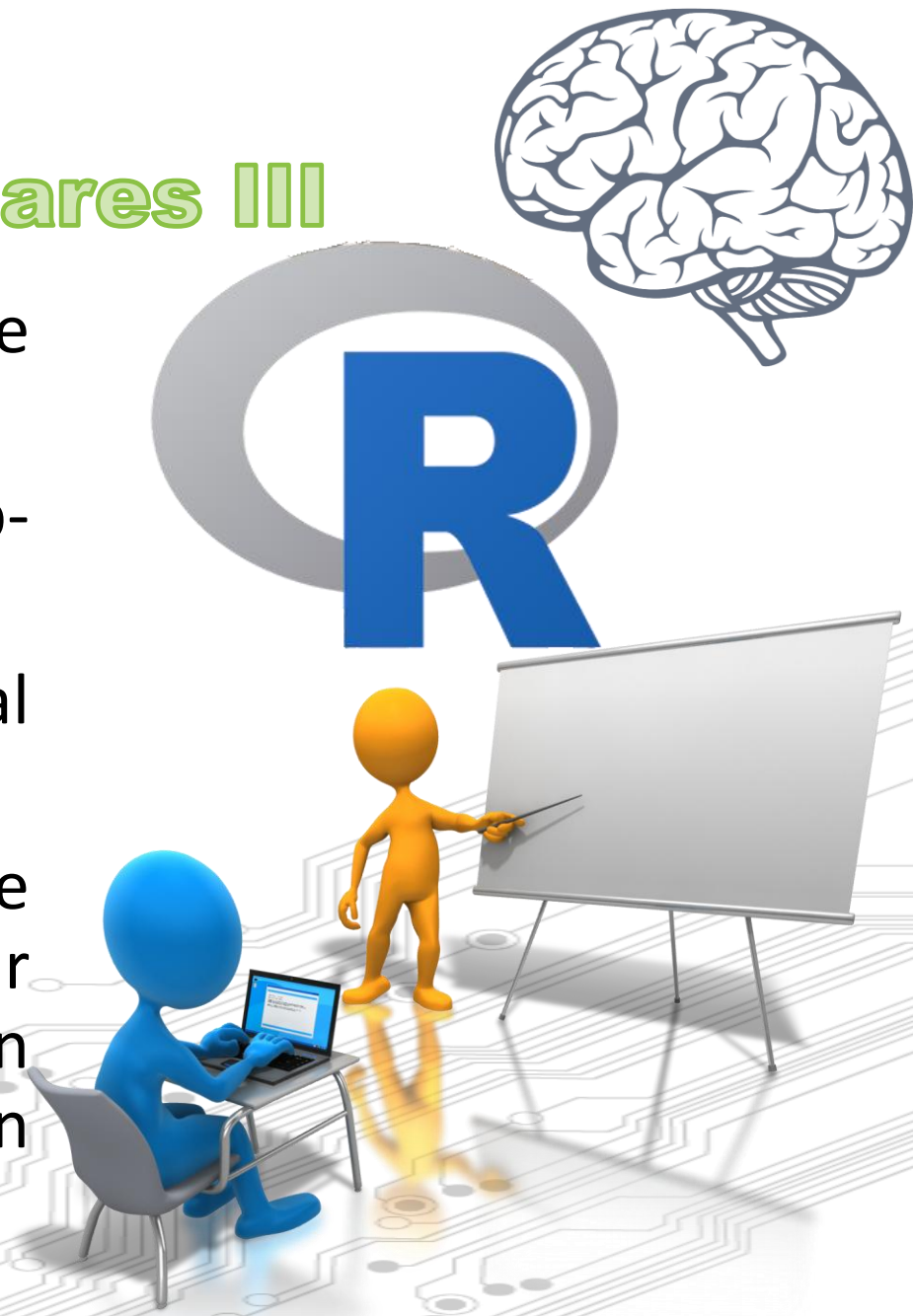
Pruebas estadísticas preliminares III

3. Con el Software R se hará la prueba de normalidad p-variada de Shapiro.

H0: Las p variables tienen distribución normal p-variada.

Ha: Las p variables no tienen distribución normal p-variada.

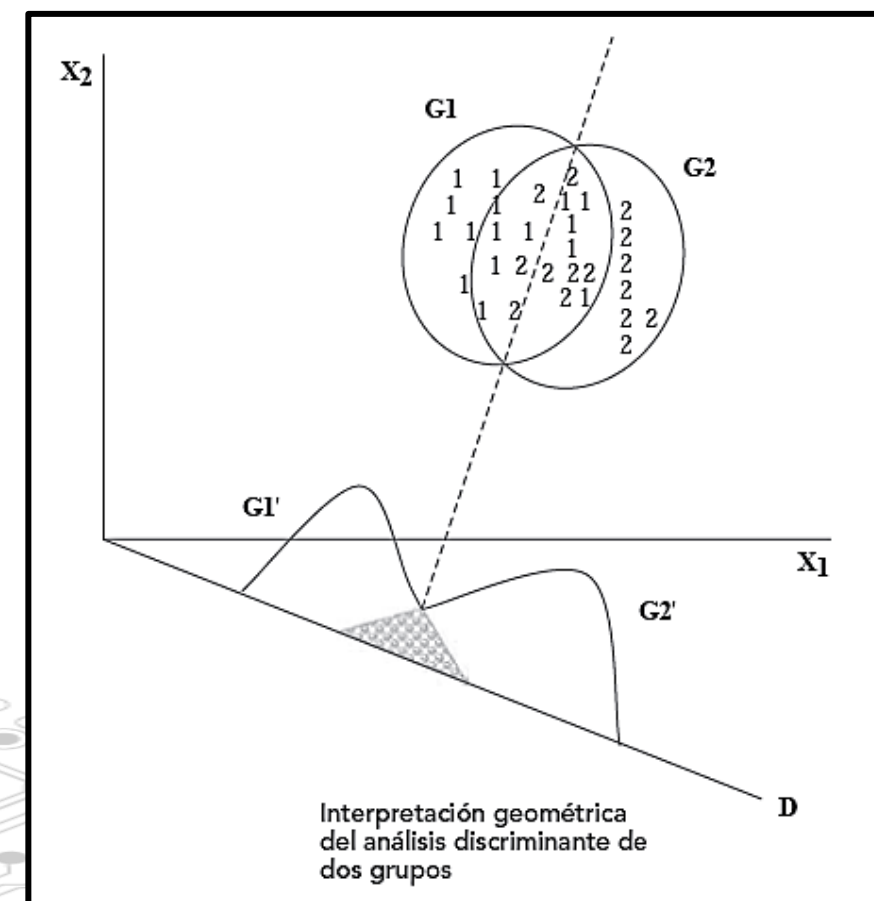
NOTA: Los datos que no cumplan el supuesto de normalidad multivariante pueden causar problemas en la estimación de la función discriminante. De ser posible usar la regresión logística como alternativa.



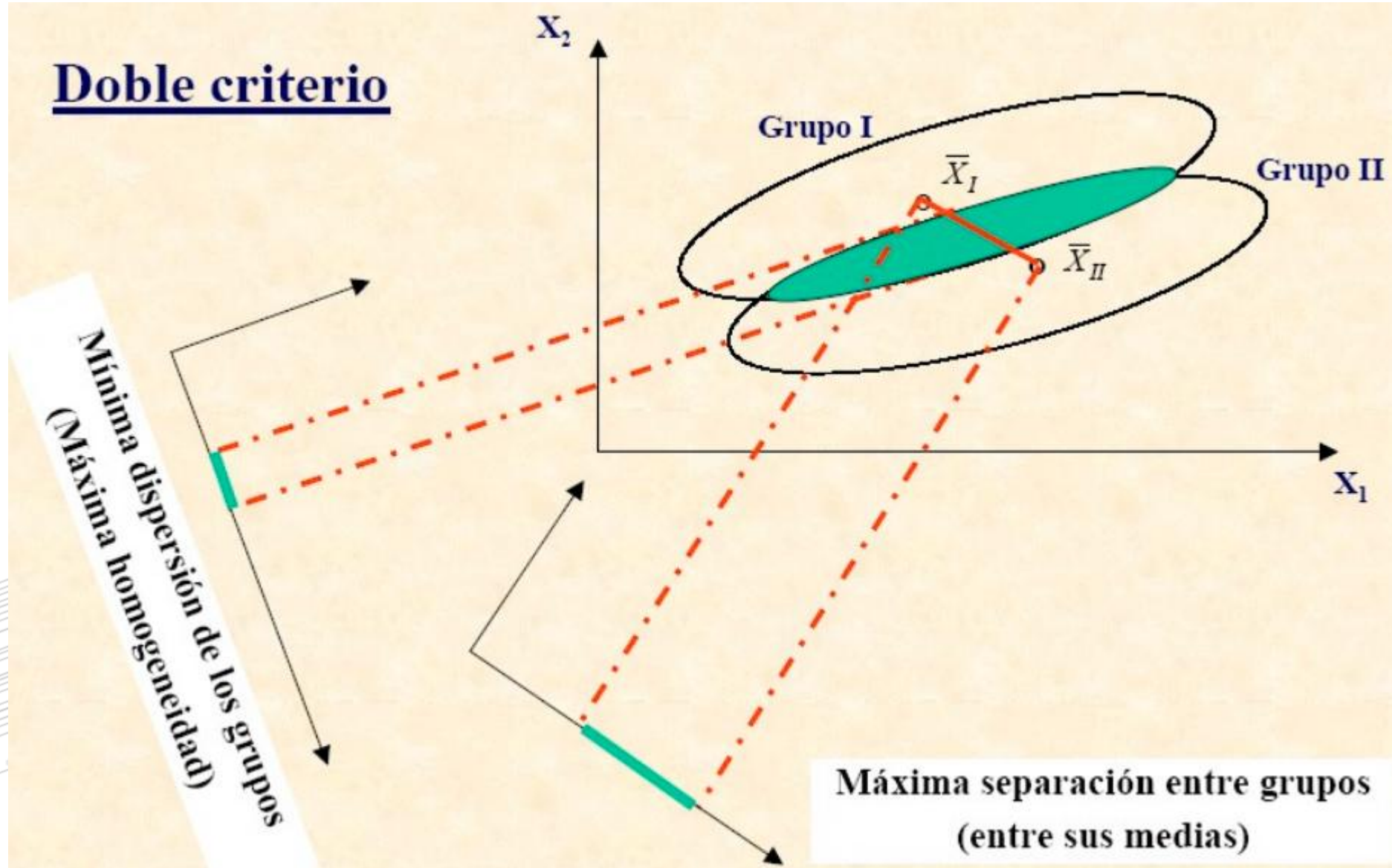


Métodos de discriminación

- 1.- Función discriminante lineal de Fisher
- 2.- Función discriminante de Máxima Verosimilitud
- 3.- Análisis de Regresión Discriminante
- 4.- Distancia de Mahalanobis



Función discriminante de Fisher





2.2 Métodos de discriminación

1) Función discriminante lineal de Fisher

La regla de clasificación se basa en una función lineal de las p variables explicativas. Se define las puntuaciones discriminantes:

$$D_i = a_1 X_{1i} + a_2 X_{2i} + a_3 X_{3i} + \dots + a_p X_{pi} \quad i = 1, 2, \dots, n$$

Para dos grupo ($g=2$), los a_i corresponde al autovector asociado a λ_1 que es el mayor autovalor de la matriz $W^{-1}B$

La regla de discriminación:

Si $D_i \leq C$, se clasifica al individuo i en el grupo I

Si $D_i > C$, se clasifica al individuo i en el grupo II

Punto de corte discriminante:

$$C = \frac{D_I + D_{II}}{2}$$

2.2 Métodos de discriminación

2) Función discriminante de Máxima Verosimilitud

La regla de discriminación asigna una observación (vector de variables \underline{x}) a una población Π_j con función de densidad $f_i(\underline{x})$ para la cual se obtiene la mayor probabilidad.

$$L_j(\underline{x}) = \max_i f_i(\underline{x})$$

Para el caso de distribución multivariante:

$$\Pi_i \approx N_p(\mu_i, \Sigma) \quad i = 1, 2, \dots, g$$

Se asigna \underline{x} a Π_j si para el índice i , se minimiza la suma de cuadrados de la distancia de Mahalanobis:

$$(\underline{X} - \underline{\mu}_i)' \Sigma^{-1} (\underline{X} - \underline{\mu}_i)$$



2.2 Métodos de discriminación

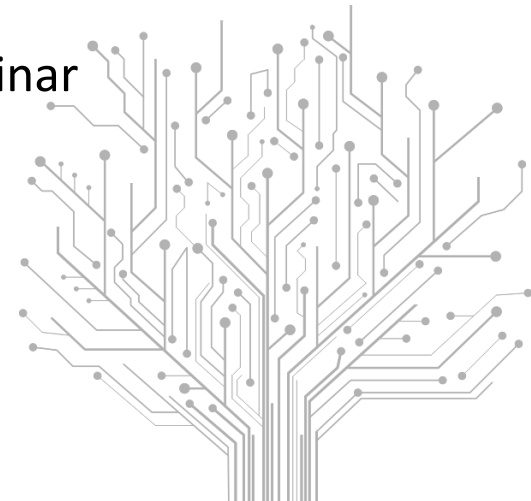
3) Análisis de Regresión Discriminante

Este método se basa en las técnicas de selección de variables utilizadas en el análisis de regresión lineal múltiple, para la formación de la función de discriminación. El método selecciona las variables independientes que mejor actúan como variables clasificadoras.

- Selección hacia adelante (Forward).
- Selección hacia atrás (Backward).
- Selección por pasos (Stepwise).

Estos métodos de selección utilizan como criterios para adicionar y eliminar variables:

- La lambda de Wills
- La razón F entre grupos
- La V de Rao



2.2 Métodos de discriminación

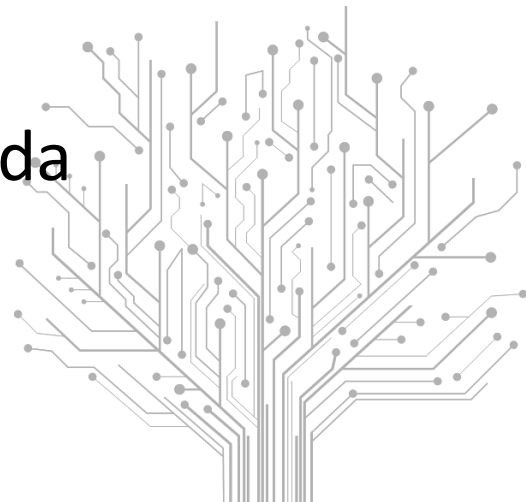
4) Distancia de Mahalanobis

Considerando \underline{x}_I y \underline{x}_{II} los centroides o centros de gravedad (promedios) de cada grupo, entonces se define la distancia de Mahalanobis de un punto i a cada uno de los dos centroides como:

$$d_{iI}^2 = (\underline{x}_i - \underline{x}_I)' S^{-1} (\underline{x}_i - \underline{x}_I)$$

$$d_{iII}^2 = (\underline{x}_i - \underline{x}_{II})' S^{-1} (\underline{x}_i - \underline{x}_{II})$$

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



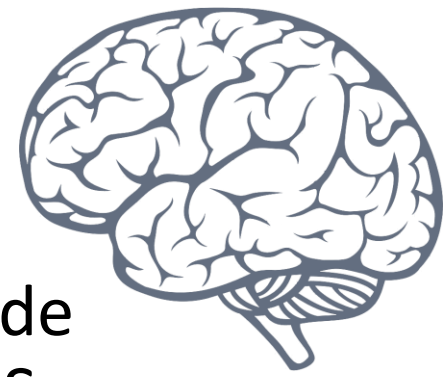
Evaluación de funciones discriminantes



Evaluar la significación de la capacidad de discriminación entre grupos que tendrá las funciones discriminantes.

- ☐ Prueba de bondad de ajuste
- ☐ Prueba de que los $(G-1)$ ejes discriminantes son significativos
- ☐ Correlación de cada variable y la función discriminante
- ☐ Índice de potencia.

2.3 Evaluación de las funciones discriminantes



- 1) **Prueba de bondad de ajuste.** Medir la proporción de variabilidad explicada por cada eje o función discriminante. Se utiliza el coeficiente Eta (correlación canónica).

$$\eta_i = \sqrt{\frac{\lambda_i}{1 + \lambda_i}}$$

Cuando η_i se aproxima a 1, indicará que la función i-ésima muestra gran capacidad explicativa para la función discriminante. Eta (correlación canónica) i al cuadrado representa la proporción de la variancia total de la variable a explicar que es explicada por la función discriminante i.

2.3 Evaluación de las funciones discriminantes

- 2) Probar si los $(g-1)$ ejes de discriminación son significativos. La determinación del número de funciones discriminantes es una prueba secuencial.

Formulación de las hipótesis:

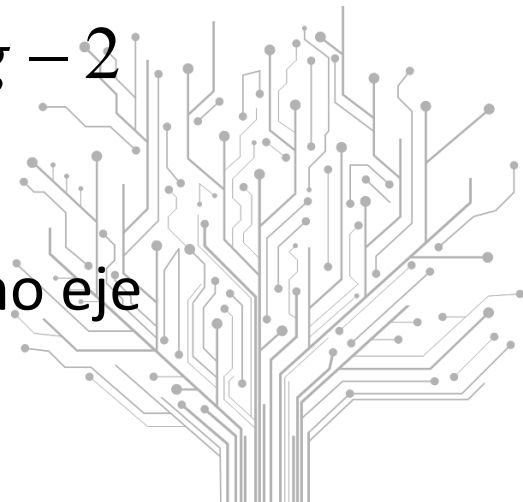
$$H_0 : \lambda_{k+1} = \dots = \lambda_{\min(G-1, p)} = 0$$

H1 : Al menos una diferente

Prueba estadística. Se utiliza la V de Bartlett:

$$V_j = \left\{ n - 1 - \frac{p + g}{2} \right\} \sum_{i=j+1}^{g-1} \ln(1 + \lambda_i) \quad j = 0, 1, 2, \dots, g - 2$$

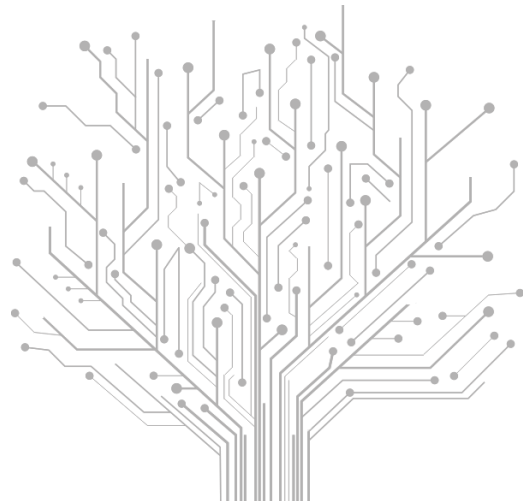
Se tiene $V_j \approx X^2_{(p-g)(g-j-1)}$ y λ_i es el autovalor generado por el i -ésimo eje discriminante.



2.3 Evaluación de las funciones discriminantes

- 3) **Correlación entre las variables y las funciones discriminantes.** Son coeficientes que permiten identificar las variables independientes que mayor influyen en el poder discriminante de cada función. Miden el poder discriminante que tiene cada variable en la clasificación de cada individuo.
- 4) **Indice de potencia.** Permite evaluar la contribución de cada función discriminante respecto al número de funciones significativas.

$$I_i = \frac{\lambda_i}{\sum \lambda_i}$$

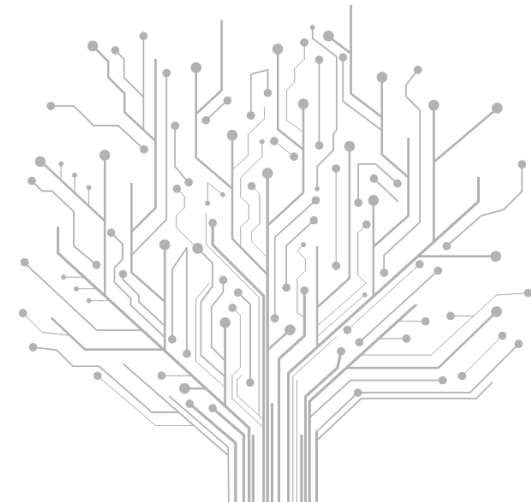


2.4 Evaluación de la clasificación

Calcular las puntuaciones discriminantes que permiten clasificar a un individuo en algún grupo y poder obtener probabilidades de su pertenencia a cada grupo.

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

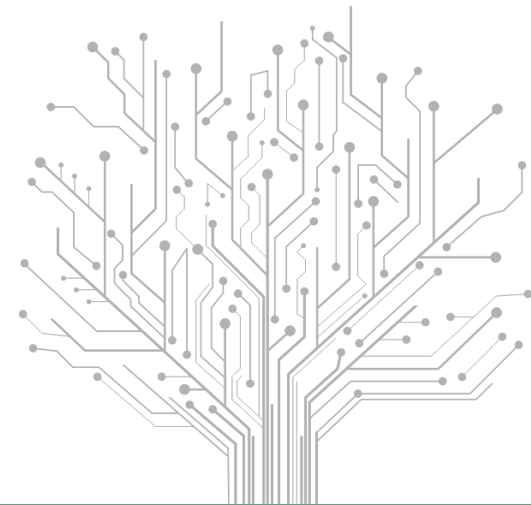
Se tiene las probabilidades a priori y las probabilidades condicionales se obtienen calculando la probabilidad de las puntuaciones observadas suponiendo la pertenencia a un grupo g .



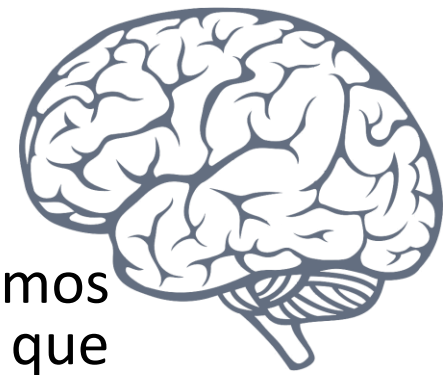
2.6 Análisis discriminante con más de dos grupos

Análisis Discriminante Múltiple, el número máximo de ejes discriminantes que se pueden obtener viene dado por $\min(G-1, p)$. Por lo tanto se pueden obtener hasta $G-1$ ejes discriminantes, siendo casi siempre cierto que el número de variables explicativas p es mayor o igual que $G-1$.

Todo el procedimiento estadístico para dos grupos, son generalizados para el caso del análisis discriminante múltiple.



Ejemplo: Análisis discriminante simple.



Una entidad financiera cuenta con algunas características de préstamos concedidos a sus clientes. Se trata de construir una función discriminante que clasifique con los menores errores posibles a los clientes en dos grupos: fallidos y no fallidos. Si se obtienen buenos resultados de clasificación, en etapas posteriores se utilizará la función discriminante construida para determinar si se concede el préstamo a nuevos solicitantes. De esta forma, si a un nuevo solicitante se le clasifica a priori como fallido, no se le concederá el préstamo solicitado.

Variable Dependiente

Tipo de cliente: 1=Fallido 2=No Fallido

Variables Independientes

mon_pre = Monto de préstamo (miles soles)

pat_net = Patrimonio neto (miles soles)

deu_pen = Deuda pendiente (miles soles)



Ejemplo 2 de aplicación

Análisis discriminante múltiple. El objetivo del estudio es tratar de determinar cuáles son las variables que realmente sirven para discriminar entre las empresas que tienen resultados monetarios clasificados como Bajos, Medios y Altos.

- **Variable Dependiente**

Resultados monetarios de las empresas: 1=Bajos, 2=Medios y 3=Altos

- **Variables Independientes**

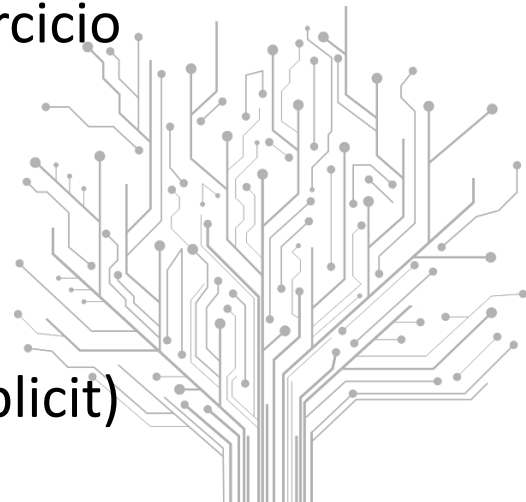
Incremento (%) en el presupuesto publicitario con respecto al periodo anterior (inc_publ)

Incremento (%) en el presupuesto promocional con respecto al ejercicio anterior (inc_prom)

Inversión de patrocinio (patrocin).

Duración en días de las actividades promocionales de la empresa (promocio)

Duración en días de las actividades publicitarias de la empresa (publicit)





Muchas gracias

Producto peruano

