

# TEMA 8

## TÉCNICAS ESTADÍSTICAS PARA EL ANÁLISIS DE LA INFORMACIÓN CONTABLE

8.1.La inferencia estadística: del análisis univariante al análisis multivariante.

8.2.La reducción de la dimensionalidad: el análisis factorial.

8.3.Los sistemas de clasificación.

8.3.1.El análisis discriminante.

8.3.2.Los modelos de variable respuesta cualitativa.

## 8.1. LA INFERENCIA ESTADÍSTICA: DEL ANÁLISIS UNIVARIANTE AL ANÁLISIS MULTIVARIANTE

- Para analizar la situación de una empresa es necesario conocer un conjunto de indicadores, no solamente un ratio determinado.
- Recurrir al análisis multivariante, que es un conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos relativos a diversas variables.
- Para cada empresa se miden varias variables (datos multivariantes).
- Algunos métodos (los sistemas de clasificación) son útiles porque orientan la toma de decisiones.
- A partir de un conjunto de datos para una empresa, permitirán pronosticar la pertenencia futura de la empresa a un grupo determinado (empresas insolventes, etc.).

## 8.2. LA REDUCCIÓN DE LA DIMENSIONALIDAD: EL ANÁLISIS FACTORIAL

- Para una determinada empresa, pueden calcularse muchos indicadores (ej. a partir de las cuentas anuales pueden obtenerse muchos ratios).
- La introducción de demasiados indicadores en un modelo de decisión puede plantear problemas.
- Una solución es recurrir a técnicas de reducción de la dimensionalidad.
- A partir de un conjunto de variables ( $x_1, x_2, \dots, x_n$ ), permiten obtener un conjunto de factores ( $F_1, F_2, \dots, F_K$ ),  $k < n$ .
- Los factores contendrán un elevado porcentaje de la información contenida en las variables originales.

• **La reducción de la dimensionalidad es posible debido a que algunas variables miden fenómenos comunes.**

• **Por ejemplo, los ratios de liquidez:**

- **Liquidez a corto plazo ( $l_1$ ):  $AC/PC$**

- **Acid-test ( $l_2$ ):  $(AC-Existencias)/PC$**

- **Ratio de tesorería ( $l_3$ ):  $(Tesorería+IFT)/PC$**

• **Las técnicas posibilitan la obtención de un factor:  $F=c_1 \times l_1 + c_2 \times l_2 + c_3 \times l_3$**

• **A ese factor podemos llamarlo genéricamente “liquidez”.**

• **Los dos métodos de reducción de la dimensionalidad más empleados son:**

- **Análisis factorial clásico.**

- **Extracción de componentes principales.**

- Ambos métodos se aplican a través de paquetes estadísticos de uso habitual (SPSS, R, etc.).
- En la práctica, proporcionan una matriz que relaciona las variables originales con los factores obtenidos:

**VARIABLES**

|       |          |          |     |          |
|-------|----------|----------|-----|----------|
|       | $V_1$    | $V_2$    | ... | $V_n$    |
| $F_1$ | $C_{11}$ | $C_{12}$ | ... | $C_{13}$ |
| $F_2$ | $C_{21}$ | $C_{22}$ | ... | $C_{23}$ |
| ...   | ...      | ...      | ... | ...      |
| $F_k$ | $C_{k1}$ | $C_{k2}$ | ... | $C_{k3}$ |

**COEFICIENTES**

- Para saber que fenómeno mide un factor se determinará con qué variables guarda elevada correlación (coeficientes c altos).

## 8.3. SISTEMAS DE CLASIFICACIÓN

- Son herramientas estadísticas que a partir de un conjunto de variables independientes ( $x_1, x_2, \dots, x_n$ ) permiten predecir el valor de una variable dependiente categórica ( $Y$ ).
- $Y$  indica una serie de categorías a las cuales puede pertenecer la empresa.
- En su forma más básica  $Y$  es una variable binaria (0/1, empresa solvente/empresa insolvente, etc.).
- El sistema permite construir una función  $f$  de tal manera que  $Y = f(x_1, x_2, \dots, x_n)$ .
- La función  $f$  se infiere a partir de un conjunto de ejemplos para los cuales conocemos tanto  $Y$  como las variables  $x_i$ .



•Posteriormente, la función será aplicada a un conjunto de empresas para las cuales conoceremos las  $x_i$  pero se desea estimar el valor de  $Y$  ya que se desconoce.

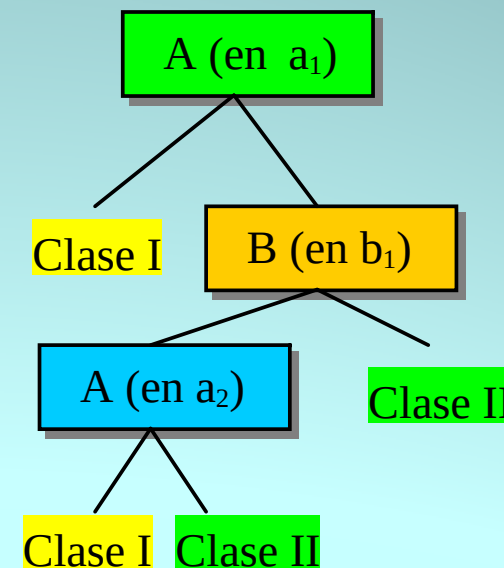
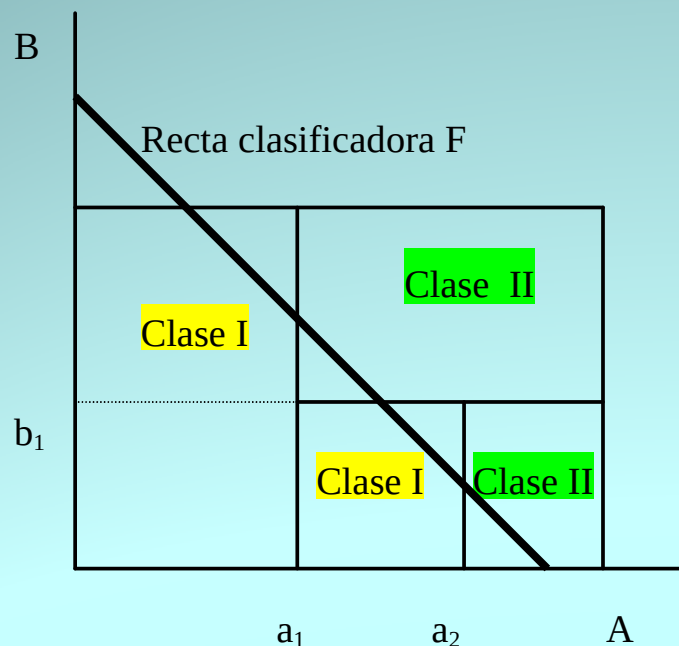
•De esta manera, los sistemas de clasificación se usan como sistemas de ayuda a la decisión.

•Ejemplo típico de modelo de decisión construido a partir de un sistema de clasificación:

- Las  $x_i$  son los ratios que definen la situación económico-financiera de la empresa.
- $Y$  es una variable binaria que vale 0 si la empresa será insolvente en el futuro y 1 en caso contrario.
- Estimamos  $f$  a partir de un conjunto de ejemplos conocidos y la utilizamos para decidir sobre la concesión de prestamos.

## Tipos de sistemas clasificadores

- Los sistemas clasificadores pueden operar a través de:
- Definir una única superficie separadora.
- Realizar particiones sucesivas en el espacio que forman las variables.





## Tipos de sistemas clasificadores

| Familia                                               | Técnica revisada                    | Procedente de ...       | Clasifica mediante...          |
|-------------------------------------------------------|-------------------------------------|-------------------------|--------------------------------|
| Análisis discriminante                                | Análisis discriminante lineal       | Estadística             | Hipersuperficies               |
| Mod. de variable respuesta cualitativa                | Análisis <i>logit</i>               | Estadística             | Hipersuperficies               |
|                                                       | Análisis <i>probit</i>              | Estadística             | Hipersuperficies               |
| Sistemas expertos                                     | Sistemas expertos basados en reglas | Inteligencia Artificial | Particiones                    |
|                                                       | Sistemas CBR                        | Inteligencia Artificial | Particiones / hipersuperficies |
| Redes neuronales artificiales                         | Redes <i>perceptron</i>             | Inteligencia Artificial | Hipersuperficies               |
|                                                       | Mapas autoorganizados               | Inteligencia Artificial | Hipersuperficies               |
| Sistemas de inducción de reglas y árboles de decisión | Algoritmos de Quinlan               | Inteligencia Artificial | Particiones                    |
|                                                       | Algoritmos genéticos                | Inteligencia Artificial | Particiones / hipersuperficies |

TEMA 8

TEMA 9

## El análisis discriminante

- En su forma más básica (Análisis discriminante lineal con dos categorías) consiste en la inferencia de una función discriminante  $Z = c_0 + c_1 \times v_1 + c_2 \times v_2 + \dots + c_n \times v_n$
- Aquellas empresas que alcancen un valor  $Z > 0$  serán clasificadas en una categoría y las que tengan un valor  $Z < 0$  en la otra.

Deuda ( $V_1$ )

**REGIÓN DE  
INSOLVENCIA  
( $Z > 0$ )**

**REGIÓN DE  
SOLVENCIA  
( $Z < 0$ )**

**Recta discriminante:**

$$0 = C_0 + c_1 \times v_1 + c_2 \times v_2 + \dots + c_n \times v_n$$

Patrimonio ( $V_2$ )

## El análisis discriminante

- Los coeficientes  $c_i$  se determinan resolviendo un sistema de ecuaciones diferenciales en las cuales se parte de las siguientes hipótesis.
  - Homoscedasticidad: Igualdad en las matrices de varianzas y covarianzas de los grupos.
  - Normalidad multivariante.
- Si no se cumplen estas hipótesis el sistema no alcanza resultados óptimos.
- Pueden utilizarse métodos para reducir el número de variables.
- El análisis discriminante es aplicable a través de paquetes estándar (SPSS, R, etc.).
- Si se trabaja con más de dos categorías (K categorías) el sistema infiere K-1 rectas discriminantes.

## Ejemplo de análisis discriminante con tres categorías

Deuda

**EMPRESAS  
INSOLVENTES**

Recta discriminante  $r_1$

Recta discriminante  $r_2$

**EMPRESAS  
DUDOSAS**

**EMPRESAS  
SOLVENTES**

Patrimonio

## Los modelos de variable respuesta cualitativa

- Parten de hipótesis menos restrictivas y son una alternativa al análisis discriminante.
- Para cada empresa, se calcula su probabilidad de pertenencia a cada uno de los grupos.
- La técnica más utilizada dentro de este grupo es el análisis logit o regresión logística.
- Para dos categorías, en primer lugar se calcula  $\beta$ , que es una combinación lineal de las variables independientes:

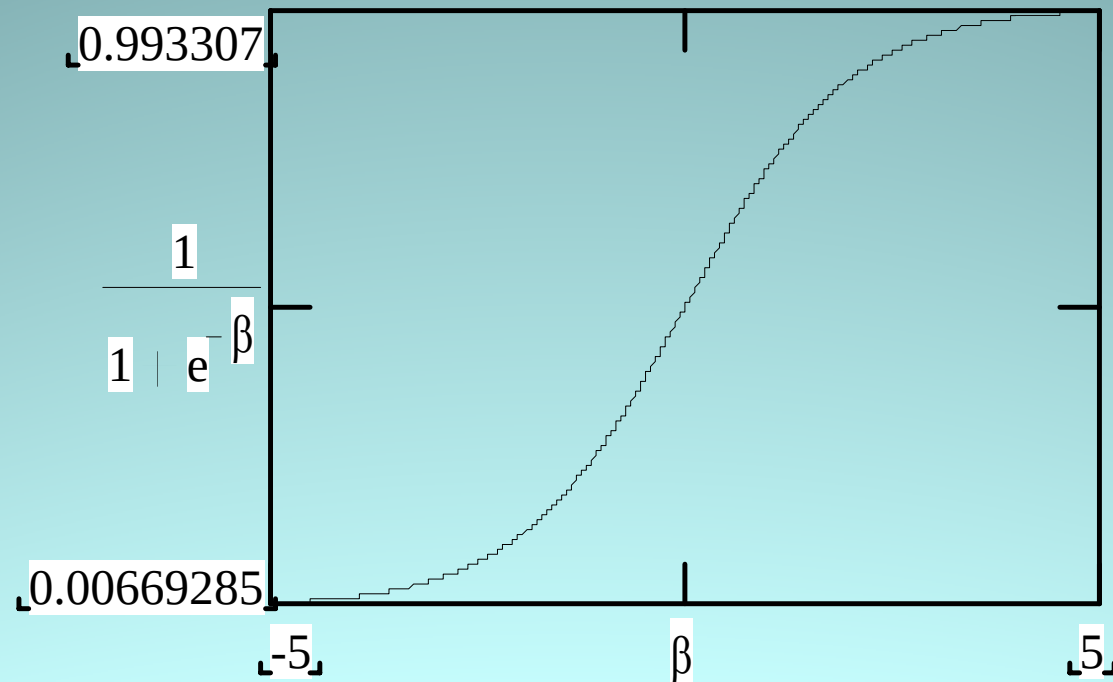
$$\beta = \alpha + \sum_{i=1}^n \alpha_i x_i$$

- En segundo lugar se calcula la probabilidad de pertenencia a cada grupo a través de la función logística:

$$p = \frac{e^{\beta}}{1 + e^{\beta}} = \frac{1}{1 + e^{-\beta}}$$

## Los modelos de variable respuesta cualitativa

- La función logística toma la siguiente forma:



- Los coeficientes  $\alpha_i$  se determinan a través de un método de estimación por máxima verosimilitud.



## Los modelos de variable respuesta cualitativa

- El otro modelo de variable respuesta cualitativa más utilizado es el análisis probit.
- El probit se diferencia del logit en que en lugar de la función logística se utiliza la función de distribución de la normal estándar:

$$p = \phi(\beta)$$

- Al igual que para el análisis discriminante, los modelos de variable respuesta cualitativa:
  - Permiten la utilización de métodos para reducir el número de variables.
  - Si se trabaja con más de dos categorías (K categorías) el sistema infiere K-1 funciones.
  - Pueden estimarse a través de paquetes estándar (SPSS, R, etc.).