

UNIVERSIDAD NACIONAL DE CORDOBA

TITULO

METODOS NO PARAMETRICOS DE CLASIFICACION CON  
VARIABLES CONTINUAS. CASO DE APLICACIÓN EN UNA  
MUESTRA DE EMPRESAS QUE OPERAN BAJO LA FORMA DE  
SOCIEDADES ANÓNIMAS EN LA ARGENTINA

Magíster en Estadística Aplicada

Autor: Norma Patricia Caro

Contadora Pública

Año 2004

## COMISION ASESORA DE TESIS

Director: Dra. Hebe Goldenhersch

Miembros Asesores: Dr. Jose Walter Dorflinger

Dra. Margarita Díaz

Fecha de aprobación de Tesis: 23/06/2004

*A Florencia, Leiza y Manuel que son la razón de mi vida,*

*A Daniel quién me acompañó y ayudó durante este y otros trabajos,*

*A Norma, quien me apoyó desde el principio en todas mis iniciativas, sin ella no sería lo que soy,*

*A mi papá Manuel a quien extraño mucho y sé que siempre está conmigo y*

*A Raúl y Rubén quienes a su manera también están presentes.*



Métodos no paramétricos de clasificación con variables continuas : caso de aplicación en una muestra de empresas que operan bajo la forma de sociedades anónimas en la Argentina por Caro, Norma Patricia se distribuye bajo una [Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).

## RESUMEN

El problema planteado en esta investigación consiste en clasificar empresas constituidas bajo la forma de sociedades anónimas que cotizan sus acciones en la Bolsa de Valores de Buenos Aires. Se trabajó con 50 empresas que fueron seleccionadas conociendo con anticipación su situación de empresas con dificultades económicas-financieras (en crisis) y sin estas dificultades (sanas), a fin de predecir situaciones de crisis a través de la aplicación de metodología no paramétrica de clasificación.

En primer lugar se procedió a realizar un análisis descriptivo multivariado y se probó el cumplimiento de los supuestos que requiere el Discriminante Lineal de Fisher, o el Discriminante Cuadrático. En virtud que éstos no se cumplieron, se aplicó metodología no paramétrica de clasificación: Método Kernel, Vecino más cercano y Árbol de clasificación.

Procesadas las empresas con cada uno de estos métodos, se obtuvieron las siguientes tasas de clasificación errónea: con el método de Kernel: 0,16; con el método del Vecino más cercano: 0,10 y con el árbol de clasificación: 0,22. De donde se deduce que la mejor clasificación fue lograda con el método del vecino mas cercano para  $k=2$  (vecinos mas cercanos).

Por último, se aplican los métodos de clasificación no paramétricos con nuevas observaciones, obteniéndose como conclusión que a partir de fines del año 2001, el escenario económico en donde actúan las empresas ha cambiado notablemente, además de las nuevas normas contables establecidas a partir del año 2002, lo cual sugiere que se definan nuevas variables y que se puedan considerar los estados contables de las empresas en un mediano plazo, luego de superar esta etapa de transición, que hace que los balances no sean comparables.

**PALABRAS CLAVES:** Discriminante no paramétrico, Kernel, Vecino más cercano, Árbol de clasificación, crossvalidation, Estados Contables.

# CLASSIFICATION'S NON PARAMETRIC METHODS WITH CONTINUOUS DATA. RESEARCH IN A SAMPLE OF BUSINESS THAT ARE COMPANIES IN ARGENTINA

## ABSTRACT

The problem expounded in this research is about Company's classification. Companies are corporation that quote their shares at Buenos Aires's share stock market. We researched with 50 companies, which were selected knowing previously their situation, business with economic-financial difficulties and business without their in order to predict economic-financial's crisis through to obtain classification's rules.

The first we proceed to make a descriptive multivariate analysis and we can prove the fulfillment of the Fisher's Discriminant or Cuadratic Discriminant's assumption. By virtue of they didn't prove, we assign non parametric methodology of classification: Kernel Method, Nearest Neighbour Method an Classification Tree.

When we make data processing with this methods we obtain the following crossvalidate error rate associated with the classification: with Kernel Method: 0,16, with Nearest Neighbour Method: 0,10 and with Classification Tree: 0,22. The best classification was Nearest Neighbour Method for  $k=2$  (nearest neighbours).

Lastly, we prove these non parametric classification method with new business, conclusion are about the economic stage has changed since the end of 2001, in addition new accounting norms were established since 2002, it says we have to define new variable and we can take companies's Balance sheet to middle time., after this transition's phase that make Blances sheet don't be comparative .

Key words: non parametric Discriminant, Kernel, Nearest Neighbour, Classification Tree, Crossvalidation, Balance sheet.

## INDICE

Introducción	Pag. 5
Capítulo 1: Métodos no paramétricos de clasificación	Pag. 10
1.1. Discriminación paramétrica y no paramétrica	Pag. 10
1.2. Método de Kernel	Pag. 13
1.2.1. Caso Univariado	Pag. 13
1.2.2. Caso Multivariado	Pag. 17
1.3. Método “El Vecino más cercano”	Pag. 20
1.3.1. Generalidades	Pag. 20
1.3.2. Regla del Vecino más cercano: 1 Vecino más cercano ( $k = 1$ )	Pag. 22
1.3.3. Regla con $k$ vecinos más cercanos ( $k > 1$ )	Pag. 23
1.4. Método del “Árbol de Clasificación”	Pag. 25
1.4.1. Generalidades	Pag. 25
1.4.2. Construcción del Árbol	Pag. 28
1.5. Precisión de los diferentes métodos	Pag. 34
1.5.1. Tasa de error aparente	Pag. 35
1.5.2. Tasa de error por Cross-Validation	Pag. 36
Capítulo 2: Caso de aplicación de los métodos de clasificación no paramétricos	Pag. 39
2.1. Introducción	Pag. 39
2.2. Universo de estudio	Pag. 40
2.3. El Análisis de los Estados Contables como herramienta de diagnóstico	Pag. 41
2.4. Selección del grupo de análisis	Pag. 42
2.5. Variables del Modelo	Pag. 43
2.5.1. Situación financiera a largo plazo	Pag. 44

2.5.2. Situación financiera a corto plazo	Pag. 46	
2.5.3. Situación económica	Pag. 46	
2.6. Procesamiento	Pag. 47	
2.6.1. Supuestos	Pag. 47	
2.6.2. Análisis descriptivo multivariado	Pag. 50	
2.6.3. Definición de modelos	Pag. 57	
2.6.4. Aplicación de metodologías no paramétricas a los datos	Pag. 57	
2.7. Error de Clasificación	Pag. 68	
Capítulo 3: Resultados y clasificación de nuevas observaciones		Pag. 71
3.1. Resultados de los Métodos de Kernel, El Vecino más cercano y Árbol de Clasificación	Pag. 71	
3.2. Aplicación de los métodos de clasificación no paramétricos a nuevas observaciones	Pag. 72	
3.3. Conclusiones	Pag. 74	
Referencias		Pag. 77
Anexo		Pag. 79



## **INTRODUCCION**

### **ANTECEDENTES**

El tema de interés versa sobre la clasificación de empresas según diferentes criterios, con el fin de investigar la situación de las mismas, o bien, construida una regla de clasificación, predecir en qué grupos pueden ubicarse nuevas empresas.

Cuando el supuesto de normalidad multivariada no se cumple y no pueden aplicarse los métodos de clasificación habituales, como la función discriminante de Fisher, en su versión lineal o cuadrática, es necesario trabajar con otros métodos que permitan dicha clasificación. La información que se posee de estas empresas se presenta en sus estados contables y se utiliza para contruir dichas reglas.

Los estados contables publicados son fruto del procesamiento de la información de cada empresa durante un período de tiempo, cuya característica central es la de proporcionar información económica y financiera a distintos usuarios. Esta información esta representada en variables continuas, por lo cual se utilizan métodos no paramétricos de clasificación para este tipo de variables.

### **OBJETIVO GENERAL**

- Estructurar el marco teórico y empírico de los métodos no paramétricos de clasificación con variables continuas a través de la revisión bibliográfica y mediante una aplicación en el campo de las finanzas, formular hipótesis de trabajo, basado en el uso de estas técnicas.

### **OBJETIVOS ESPECIFICOS**

- Analizar el desarrollo estadístico de los métodos de clasificación no paramétricos, como método de Kernel, método del “vecino más cercano” y los árboles de clasificación.

- Resolver problemas de clasificación de las observaciones, cuando las distribuciones de las variables en las poblaciones son desconocidas.
- Clasificar las empresas que operan bajo la forma de sociedad anónima en dos grupos: como empresas sin problemas financieros y económicos por un lado y empresas con dificultades de este tipo por el otro, a los fines de compararlas y formar un criterio que las identifique.
- Asignar nuevas observaciones a la población o clase que corresponda, previamente determinada, a través de estos métodos no paramétricos.
- Permitir una adecuada toma de decisiones para evitar que las empresas que comienzan a tener dificultades, se presenten en concursos preventivos o en quiebra.

## **ACERCA DE LA CLASIFICACION Y PLANTEAMIENTO DEL PROBLEMA**

Una tarea importante en los diversos aspectos del quehacer científico que subyace en las distintas áreas, es la “clasificación” de objetos.

Intentando conceptualizar la naturaleza común de los problemas que pueden presentarse, puede decirse que existen dos tipos de problemas de clasificación. Por un lado tenemos el análisis grupal (cluster analysis) método que impone una estructura de grupo sobre una serie de objetos. Es una técnica multivariada que pretende separar distintos conjuntos de observaciones pertenecientes a distintas poblaciones o grupos y ubicar nuevas observaciones en las clases previamente definidas.

En este trabajo se cuenta con una muestra de elementos de clasificación conocida. De este modo, ya existe una estructura de grupo. El objetivo es crear una regla de clasificación para futuros elementos sobre la base de la muestra de diseño.

Si una muestra fue seleccionada de dos o más poblaciones, es necesario identificar dichas poblaciones o grupos alternativos a partir de un conjunto de variables que se

utilizan para caracterizarlas, en función a los valores que las mismas asumen. Es, entonces el análisis discriminante lineal el que nos permitirá realizar la clasificación de individuos considerando un conjunto de variables que son independientes.

En realidad, el análisis discriminante lineal supone que los grupos son conocidos a priori y su objetivo es obtener combinaciones lineales de las variables independientes de forma tal que sus valores separen lo más apropiadamente posible dichos grupos, mientras que el análisis de clasificación (cluster) no supone a priori tal división en grupos.

En base a esto, la clasificación se realiza con dos objetivos:

- Con fines explicativos: se trata de incluir a cada individuo en una de las poblaciones o grupos establecidos.
- Con fines predictivos: una observación nueva, la cual no fue utilizada para la construcción de la regla de clasificación, se asignará al grupo en el cual tiene mayor probabilidad de pertenecer basándose en sus mediciones.

La estrategia es encontrar combinaciones lineales de las variables originales para discriminar entre los grupos, que minimice la tasa de error o de mala clasificación. No obstante, el resultado de la aplicación está condicionada al cumplimiento de dos fuertes supuestos:

- Que los grupos posean iguales matrices de varianzas y covarianzas.
- Que las muestras obtenidas sean aleatorias e independientes y provenientes de una población normal multivariada.

Si el primer supuesto no se verifica, puede aplicarse el análisis discriminante cuadrático.

Por regla general, para el caso de dos poblaciones, los parámetros  $\mu_1$  y  $\mu_2$  (vectores de medias de la población 1 y 2),  $\Sigma_1$  y  $\Sigma_2$  (matrices de varianza-covarianza de cada una de las poblaciones) son desconocidos y al reemplazarlos por sus estimadores máximo-verosímiles se desvanecen las propiedades de optimalidad de la regla de clasificación dada por el análisis discriminante lineal, o el cuadrático, con el agravante de que si el valor de  $n$  no es sensiblemente mayor que el número de parámetros a

estimar, las matrices de varianza-covarianza de los estimadores se vuelven altamente inestables.

De igual modo, cuando la distribución de las variables en la población es desconocida, todas las observaciones caracterizadas a través de un conjunto de variables continuas que pertenecen a esa población, se utilizan para estimar la función de densidad, que representa la probabilidad que una observación  $\mathbf{x}$  con sus mediciones corresponda a la población  $j$  y poder asignar nuevas observaciones al grupo que corresponde.

Es decir que  $f(j/\mathbf{x})$  debe ser estimada por  $\hat{f}(j/\mathbf{x})$  a fin de cumplir con el objetivo de clasificación.

Ante el incumplimiento de estos supuestos, se hace necesario buscar otras reglas de clasificación que sean adecuadas, como son los métodos no paramétricos que se concentran en la proximidad de los puntos para determinar una clasificación, mediante la estimación de la función de densidad de las distintas poblaciones.

Los métodos no paramétricos utilizados, en este trabajo, son:

- Método de Kernel
- Método del “vecino más cercano”
- Método de árbol de clasificación

En el método de Kernel, a través de la estimación de la densidad de cada clase, se realiza la clasificación de las observaciones utilizando todas las mediciones y permite incluir nuevos individuos en la clase correspondiente.

El método “el vecino más próximo”, similar al anterior, a través de la estimación de la densidad, realiza la clasificación, pero utilizando los puntos de diseño más cercanos a un vector de mediciones  $\mathbf{x}$ .

A su vez, el método del árbol de clasificación, que Hand (1996) considera como recursivo, es una técnica de clasificación no paramétrica que en la partida incluye todos los objetos en un mismo grupo, el cual es subdividido según los distintos valores de las

variables hasta pararse en un punto. Cada grupo terminal será una población. Es posible también que un nuevo objeto sea clasificado de acuerdo a los grupos terminales.

Cualquiera sea el método que se utilice, después de realizar una clasificación, será necesario evaluarlo a través del error de predicción, en otras palabras evaluar la probabilidad de clasificar mal las observaciones, ya que ciertos errores en la clasificación son más serios que otros. Si la severidad relativa de esto puede ser cuantificada en términos de costos, se podría analizar el costo esperado de clasificar mal. Si las funciones de densidad son desconocidas y por lo tanto la tasa de error no pueda obtenerse, se podrá realizar una estimación de esta tasa, a través de la proporción de observaciones mal clasificadas.

Algunos estimadores de la tasa de error, que no dependen de la distribución poblacional, son: la tasa de error aparente, validación cruzada, el método bootstrap, entre otros.

En síntesis, el problema planteado para esta investigación consiste en clasificar empresas constituidas bajo la forma de sociedades anónimas que cotizan sus acciones en la Bolsa de Valores de Buenos Aires, las cuales fueron seleccionadas según dos estados (con dificultades financieras-económicas y sin ellas), a fin de predecir situaciones de crisis en otras empresas, a través del uso de las reglas de clasificación obtenidas.

Para ello, se aplicaron métodos no paramétricos de clasificación, los cuales fueron validados con nuevos casos, midiendo el error de clasificación errónea.

## CAPITULO 1: METODOS NO PARAMETRICOS DE CLASIFICACION

### 1.1. DISCRIMINACION PARAMETRICA Y NO PARAMETRICA.

Dada una muestra de elementos, con una cierta estructura de grupo, el objetivo es determinar una regla de clasificación que permita evaluar las clasificaciones de dichos elementos y clasificar elementos futuros, en base a dicha muestra.

Cada objeto o elemento de análisis se asigna a un grupo mediante el estudio de los valores de las variables (vector de variables). Se desea asignar los objetos a grupos de manera tal que minimice alguna medida de error. Cuando se conoce  $f_j(\mathbf{x})$ , la función de densidad para el grupo  $j$ ,  $\pi_j$  la probabilidad a priori de pertenencia al grupo  $j$  y  $C_{ij}$  el costo de clasificar erróneamente un objeto del grupo  $j$  como del grupo  $i$ , se puede calcular:

$$\sum_j C_{ij} \pi_j f_j(\mathbf{x})$$

para  $j = 1, 2, \dots, g$ , siendo  $g$  el número de grupos y encontrar así la expresión más pequeña.

Dado que usualmente  $\pi_j$  y  $f_j(\mathbf{x})$  no son conocidos, deben de alguna manera ser estimados a través de los elementos de la muestra. La estimación de  $\pi_j$  es directa y la dificultad reside en estimar  $f_j(\mathbf{x})$ .

El método más antiguo de análisis discriminante paramétrico es el método de la función discriminante lineal de Fisher (1936), para  $g = 2$  grupos, quien abordó el problema desde una óptica univariada usando una combinación lineal de las características observadas, el cual se extendió a un enfoque multivariado.

Sea  $\mathbf{x}$  el vector de mediciones sobre un elemento de una población y considerando dos poblaciones  $\pi_1$  y  $\pi_2$ , las funciones de densidad multivariadas asociadas con ambas poblaciones respectivamente serán  $f_1(\mathbf{x})$  y  $f_2(\mathbf{x})$ . Se asume que las variables aleatorias caracterizadas por estas funciones multivariadas tienen vectores medios  $\mu_1$  y  $\mu_2$  y matriz de covarianza común  $\Sigma_1 = \Sigma_2$ . Se considera la combinación lineal  $y = l' \mathbf{x}$ , con parámetros  $\mu_{1y} = l' \mu_1$  y  $\mu_{2y} = l' \mu_2$  y varianza igual a  $V(y) = l' \Sigma l$ .

La idea de Fisher fue maximizar la distancia estadística entre  $\mu_{1y}$  y  $\mu_{2y}$  a través de una selección apropiada del vector de coeficientes de la combinación lineal, la cual se conoce como función lineal discriminante de Fisher.

La importancia del método es que para dos distribuciones normales con matrices de covarianzas idénticas, la superficie de decisión óptima es verdaderamente lineal y si los estimadores de la media y la covarianza son elegidos correctamente el método de Fisher convergerá a esa superficie de decisión óptima.

Cuando las matrices de covarianzas no son idénticas, la función discriminante cuadrática cumple con el objetivo de discriminar los grupos o poblaciones, siempre que exista normalidad multivariada.

Cuando no se verifican los supuestos necesarios para aplicar estos métodos paramétricos se deben utilizar métodos de discriminación no paramétrica, que tienden a estimar la densidad dentro de los grupos.

Se aplican en este trabajo, dos importantes métodos, que se basan en la estimación de la densidad: el “de núcleo” o “Kernel” y el “vecino mas cercano” y un tercer método de clasificación conocido como el “árbol de clasificación”.

En un principio, el método más probado para estimar la densidad es el enfoque del histograma tradicional. Para una variable continua, esto es definido por la partición del rango de los posibles valores de la variable en varios intervalos de igual longitud, graficando para cada intervalo una barra con altura proporcional al número de observaciones en ese intervalo. Tal estimador podrá usarse para producir una clasificación: un nuevo individuo con mediciones  $x$  se clasifica comparando las alturas de los histogramas para cada grupo. Por supuesto, que si la muestra para cada grupo tiene solo pocos casos, entonces las alturas de las barras del histograma tomarán pocos valores, luego, tendría una alta probabilidad estimada ( $\hat{f}(x/i)$ ), lo cual es una desventaja.

Otro obstáculo en el uso del histograma para análisis discriminante está dado por la cantidad de intervalos, ya que si hay más de una variable, la cantidad de intervalos va siendo cada vez más grande, y más todavía si se incrementa el número de variables, dada la estructura fija de los mismos. Lo cual podría evitarse si tenemos intervalos

muy grandes, pero esto, a su vez proporciona otra desventaja, en cuanto a que no refleja con exactitud las variaciones en la densidad.

También este enfoque enfrenta otra desventaja que es el tema de las discontinuidades en las probabilidades estimadas para cada intervalo construido. Al producir un histograma, es necesario decidir que ancho tendrán los intervalos, cuántos serán y donde serán localizados. En esencia, la probabilidad estimada dentro de un intervalo es una estimación del promedio de las probabilidades para los valores de  $x$  en ese intervalo. Si el intervalo es grande, entonces los diferentes valores de la variable corresponderán a diferentes probabilidades, la estimación por el histograma será más sesgada para algunas partes del intervalo. Entonces, intervalos grandes tenderán a tener más observaciones dentro de ellos, por lo tanto la estimación de probabilidad correspondiente tendrá menor varianza. A la inversa, un intervalo menor comprende pocas observaciones que podrán quedarse dentro, por lo tanto, la varianza de la probabilidad estimada para esos intervalos será mayor, aunque al cubrir un pequeño rango de  $x$ , tendrá la ventaja de un menor sesgo. Entonces, habrá que elegir la mejor amplitud del intervalo y su localización, adoptando algún criterio de performance que minimice la relación sesgo-varianza. Aún así no se supera el problema de las discontinuidades.

Otro aspecto a considerar, es la extensión a más de una variable, ya que si la variable está dividida en 10 intervalos y a su vez se cuenta con 10 variables, luego habrá  $10^{10}$  intervalos en el espacio multivariado. No son muchas las muestras que poseen tamaño suficientemente grande al producir probabilidades estimadas exactas de sus intervalos. En efecto, muchos de los intervalos multivariados están vacíos, por lo tanto si la clasificación se basa en la comparación de probabilidades estimadas, muchas de ellas serán cero y no será una buena regla de clasificación.

Los métodos de Kernel y del “vecino más cercano”, como generalización del histograma, solucionan el problema de las discontinuidades, el de localizar los intervalos determinando su extensión y los inconvenientes de índole multivariada, ya que evitan que la estructura de intervalos sea fija. Para ser precisos, la densidad en un punto es estimada desde el intervalo donde está ubicado ese punto.

Las primeras descripciones de los métodos llamados de núcleo son de Fix y Hodgs (1951), Rosenblatt (1956) y Parzen (1962), todos ellos citados por Hand (1999).



En el análisis discriminante no paramétrico las estimaciones de distribución condicional de los grupos y las superficies de decisión son más flexibles que en los métodos tradicionales. No es necesario forzar supuestos y colocar las formas permisibles para distribuciones estimadas y superficies de decisión en una camisa de fuerza. Tampoco se quiere insinuar que los métodos paramétricos no sirven, sino que son apropiados para tipos particulares de problemas.

El método de Kernel se utiliza tanto para variables categóricas, como para variables de escala de intervalo continuo. A su vez, ha sido aplicado a problemas que involucran combinaciones de tipos de variables.

El método del vecino más cercano, cuyos primeros antecedentes datan de Fix y Hodgs (1951), Hart (1968) y Hand y Batchelor (1978), todos citados por Hand (1999), trabaja con las distancias más cercanas en el espacio multivariado de los individuos.

Por último, se aplicó el método de árbol de clasificación, cuyo uso comenzó con Breiman y Friedman (1973) y que Hand (1996) lo considera un método de partición recursivo, por la forma de trabajo, subdividiendo grupos en subgrupos, sin que estos grupos asuman una distribución de probabilidad. El algoritmo, basado en los valores de las variables, decide cómo se particiona cada una según determinadas características y así se divide en ramas. Cuando la variable está ordenada de alguna manera, si los valores son menores a cierto valor, se define una regla de clasificación y va a una rama determinada, de lo contrario va a otra.

Si la variable no está ordenada, se necesita una partición explícita para asignar cada valor de la variable a una rama particular. Cada rama terminal es un grupo.

## **1.2. METODO DE KERNEL**

### **1.2.1. Caso Univariado**

En primer lugar, se presentan características del método de Kernel en espacios univariados con variables continuas.

Sea  $x_1, x_2, \dots, x_n$  una muestra obtenida de una población cualquiera, con la cual se quiere encontrar una estimación  $\hat{f}_j(x/j)$  de la función de densidad de la población.

Como se indicó en el apartado anterior, el histograma es la forma básica del estimador de la densidad no paramétrica. Rosenblatt (1956) propone que, en vez de

tener una estructura de intervalos fijos, independiente de la posición de  $x$ , centrar un intervalo en  $x$ . Al igual que con un histograma, la probabilidad de que un punto caiga en el intervalo centrado en  $x$  es estimada por la proporción de puntos de la muestra que caen en ese intervalo. La densidad en  $x$  es esta probabilidad total dividida por la longitud del intervalo. Por lo tanto, se trabaja con intervalos centrados en  $x$ , de amplitud  $2h$ , esto es  $[x-h ; x+h]$ .

En el enfoque del histograma, cada punto dentro de la amplitud del intervalo contribuye  $1/n$  a la estimación y cada punto fuera de esta amplitud no contribuye. Con la finalidad de superar estas dificultades, se utiliza una función ponderada con las distancias entre  $x$  y el conjunto de puntos. La forma precisa de esta función es lo que se denomina función Kernel.

La estimación de la densidad debe hacerse para cada grupo. A fin de simplificar la notación, se considera en general de la siguiente manera:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{K_1(x-x_i)}{2h} \quad (1)$$

donde:  $K_j$ : función de núcleo o de Kernel

$(x-x_i)$ : distancia entre el valor de la variable y el conjunto de datos de la muestra

$h$ : cantidad que se suma por defecto y por exceso a  $x$ .

si simbolizamos  $u = x - x_i$

donde:

$$K_j(u) = \begin{cases} 1 & \text{si } |u| < h \\ 0 & \text{si } |u| \geq h \end{cases}$$

con lo cual resulta una estimación representada por la distribución de la muestra. Esta estimación puede ser vista como un promedio de  $n$  valores, correspondiente a cada punto de la muestra.

Puntos cercanos a  $x$ , dentro de la distancia  $h$ , aportan  $1/2h$ , mientras que puntos lejanos a  $x$  aportan un valor 0. Entonces el salto de  $1/2h$  a 0 conduce a las discontinuidades.

Si se reemplaza  $K_1/2h$  por  $K_2$ , donde  $K_2(u)$  decrece en forma monótona a medida que  $(x-x_i)$  se incrementa, la estimación es:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_2(x-x_i)$$

Ahora bien, en regiones con varias  $x_i$  habrá más aportes que con pocas  $x_i$ . Podríamos generalizar este concepto si tomáramos otra función para  $K_2$  que la simple distancia entre  $x$  y  $x_i$ .

También, se puede describir esta función como:

$$\hat{f}(x) = \{F_n(x+h) - F_n(x-h)\} / 2hn$$

donde  $F_n(x)$  es la función de distribución de la muestra, entonces  $\hat{f}(x)$  puede ser una aproximación a la derivada de esta función de distribución.

Para  $n$  fija, si  $h$  es grande, la estimación es uniforme porque todo el conjunto de elementos queda dentro del intervalo; si  $h$  es pequeño, la estimación es altamente irregular.

En el límite cuando  $h \rightarrow 0$  la  $\hat{f}(x)$  obtenida como la derivada de la función de distribución presenta una serie de picos (uno en cada  $x_i$ ) lo cual no es útil para los requerimientos comparativos de clasificación. Por ello, se debe tomar un valor de  $h > 0$  a fin de obtener una función de densidad aproximada más útil para la clasificación.

Para solucionar el problema del histograma en cuanto al tamaño de los intervalos, el método de Kernel impone un tipo de restricción sobre la irregularidad de la función de densidad, que consiste en establecer un parámetro atenuante (smoothing), que indica que puntos cercanos al conjunto bajo análisis deben tener la misma probabilidad de ocurrir.

Este parámetro atenuante, el cual designaremos como  $h$ , deberá ser escogido con la perspectiva de especificar cuanta irregularidad se quiere captar en el modelo.

Por ello, se debe hacer explícita la dependencia de la estimación sobre  $h$  mediante la siguiente formula:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)$$

lo cual considerado para cada grupo, es:

$$\hat{f}_j(x) = \frac{1}{n_j h} \sum_{i=1}^{n_j} K\left(\frac{x-x_i}{h}\right)$$

donde:  $x_i \in$  al  $j$ -ésimo grupo.

Esta función, entonces, depende de dos parámetros:

$K$ : que representa la función Kernel y  $h$ : que es el parámetro atenuante (smoothing).

Si bien suelen utilizarse diferentes Kernel y parámetros atenuantes para cada grupo, usualmente ambos son los mismos.

Si  $K$  es tomada en sí misma como una función con densidad unimodal, cumple con las propiedades de ser función de densidad, de manera tal que:

a)  $\int K(x)dx = 1$

b)  $K(x) \geq 0$  para todo  $x$

por lo tanto,  $\hat{f}(x)$  también es una densidad, esto es que cumple con las mismas propiedades:  $\int \hat{f}(x)dx = 1$  y  $\hat{f}(x) \geq 0$  para todo  $x$ .

Respecto a las propiedades de las estimaciones, Rosenblatt, (1956), probó que no hay estimador no negativo que sea insesgado para todo  $f(x)$  continua y en particular no hay estimador de Kernel con núcleo no negativo que sea insesgado para todo  $f(x)$  continua, lo cual fue confirmado por Yamato en 1972.

Así en general para muestras finitas y con estimaciones de núcleo no negativos, debemos aceptar cierto sesgo en las estimaciones.

Shapiro (1969) prueba que a pesar que el sesgo no es cero, el índice de convergencia tiende a cero, es decir la  $E[\hat{f}(x)]$  tiende a  $f(x)$ , mientras  $n$  crece y  $h$  se reduce, lo cual se prueba dentro de la teoría asintótica.

Se sabe que un estimador insesgado o con un sesgo pequeño no posee suficiente valor práctico si posee una varianza grande. Es decir, que se tendría poca confianza de que  $\hat{f}(x)$  esté cerca de  $f(x)$ .

La varianza del estimador definida como  $V[\hat{f}(x)] = E\{\hat{f}(x) - E[\hat{f}(x)]\}^2$ , en términos de resultados asintóticos, debería combinarse con la medida del sesgo para producir una medida única.

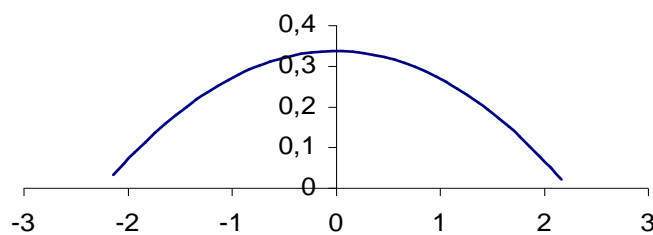
Para usar este método se necesita especificar la forma del Kernel y el tamaño del parámetro atenuante. Se debe hacer una “buena” elección de estos dos parámetros

para proceder a la clasificación. Se considera que la elección del Kernel no es tan crítica como la del  $h$ . Elegir un  $h$  determinado es considerar el mínimo sesgo-varianza para cada  $x$ . La medida que combina ambos aspectos, es aquella que minimice el error cuadrático medio (MSE), donde se pretende determinar el mejor  $h$  para un valor particular  $x$  que ha de ser clasificado. Esta medida del error tiende a 0, cuando  $n$  tiende a ser lo suficientemente grande.

Respecto a la elección del Kernel, este puede ser una función de densidad univariada, tal como la normal, la uniforme, la Epanechnikov, entre otras. Respecto al Kernel Epanechnikov, éste fue encontrado en 1969, como un Kernel óptimo que se obtiene como (Figura 1.1):

$$K_o(x) = \begin{cases} \frac{3}{4\sqrt{5}} - \frac{3x^2}{20\sqrt{5}} & |x| \leq \sqrt{5} \\ 0 & |x| > \sqrt{5} \end{cases}$$

Figura 1.1. Kernel Epanechnikov univariado



Sin duda la elección de ambos parámetros deberá ser aquella combinación que minimice el error de clasificar erróneamente.

### 1.2.2. Caso Multivariado

En el caso univariado, cada punto  $x_i$  de la muestra aportó en la estimación en  $x$ , una cantidad inversamente relacionada a la distancia entre  $x$  y  $x_i$ . La naturaleza de esta relación inversa fue especificada por la forma del núcleo:  $K(x - x_i)$  y por el parámetro

*h.* Todo lo mencionado se extiende al caso multivariado. Nuevamente, cada punto del conjunto aporta a  $\hat{f}(x)$  una cantidad que depende de la distancia entre  $\mathbf{x}$  y  $\mathbf{x}_i$ . Se vuelve a considerar un promedio de estos aportes y la función  $K$  tienen múltiples argumentos.

La estimación multivariada general es:

$$\hat{f}_j(\mathbf{x}) = \frac{1}{n_j} \sum_{i=1}^{n_j} K(\mathbf{x} - \mathbf{x}_i) \quad \text{donde } \mathbf{x}_i \in \text{al } j\text{-ésimo grupo}$$

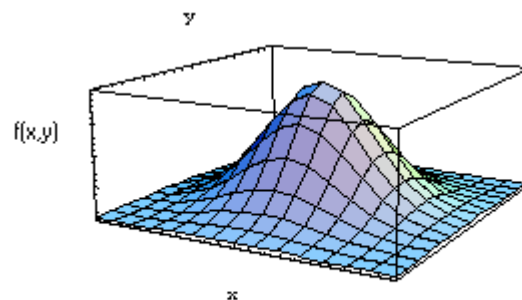
En la función del Kernel está presente el parámetro smoothing, cuya estimación será una matriz atenuante ó smoothing que denotamos con  $\mathbf{H}$ .

Las formas más comunes del Kernel son:

- 1) Kernel Gaussiano (La Figura 1.2. representa la función normal bivariada, a modo ilustrativo de la forma del Kernel en esa dimensión)

$$K(\mathbf{u}) = \frac{1}{(2\pi)^{d/2} |\mathbf{H}|^{1/2}} \exp\{-1/2(\mathbf{u}'\mathbf{H}^{-1}\mathbf{u})\} \quad \text{donde: } \mathbf{H} = \mathbf{r}^d \mathbf{V}_i$$

Figura 1.2. Función Normal Bivariada



Si bien  $\mathbf{V}_i$  puede tener diferentes formas, en las aplicaciones lo más frecuente es utilizar  $\mathbf{V}_i = \mathbf{S}$  (la matriz de covarianzas común ó pooled).

Y  $r$  especifica el valor del radio para estimar la densidad por Kernel y clasificar una observación dentro de uno de los grupos, basándose en la información de la muestra y dentro del radio  $r$  de  $x$ .

El núcleo centrado en  $\mathbf{x}_i$  tiene la misma forma funcional que una distribución normal multivariada  $d$ -dimensional.

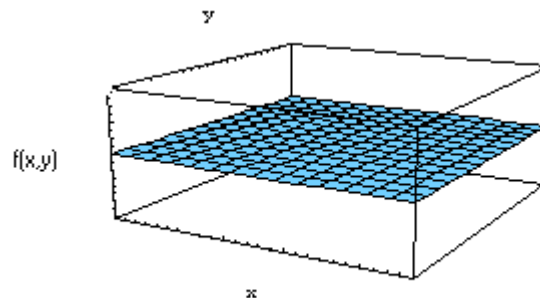
2) Kernel Uniforme (La Figura 1.3. representa en el campo bivariado un Kernel uniforme)

$$K(\mathbf{u}) = \frac{1}{|\mathbf{H}|^{1/2} V_o} \quad \text{para} \quad \mathbf{u}' \mathbf{H}^{-1} \mathbf{u} \leq r^2$$

siendo  $V_o = \frac{\pi^{d/2}}{\Gamma(d/2 + 1)}$  y  $\mathbf{H} = r^d \mathbf{V}_i$

$V_o$  es el volumen de la elipsoide  $d$  dimensional y  $\Gamma$  representa la función Gamma.

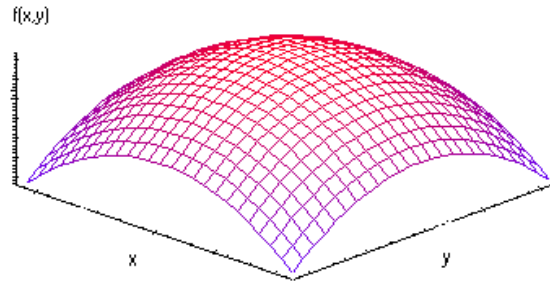
Figura 1.3. Función Uniforme Bivariada



3) Kernel Epanechnikov (La Figura 1.4. representa, a modo ilustrativo, el Kernel de Epanechnikov para dos variables)

$$K(\mathbf{u}) = \frac{1 + d/2}{|\mathbf{H}|^{1/2} V_o} (1 - \mathbf{u}' \mathbf{H}^{-1} \mathbf{u}) \quad \text{para} \quad \mathbf{u}' \mathbf{H}^{-1} \mathbf{u} \leq r^2$$

Figura 1.4. Función Epanechnikov Bivariada



También se utilizan funciones univariadas, que a través del producto de las mismas, pueden extenderse al caso multivariado, lo cual es conocido como producto de Kernel univariados:

$$K(\mathbf{u}) = \frac{1}{h_1 \dots h_d} \prod_{l=1}^d K(\mathbf{u}_l / h_l)$$

K puede ser cualquier forma univariada y en este caso, H resulta ser una matriz diagonal.

Otro caso de producto de Kernel, es:

$$K(\mathbf{u}) = \frac{1}{h^d} \prod_{l=1}^d K(\mathbf{u}_l / h)$$

Esta forma es más común debido a que hay un solo parámetro h que necesita ser estimado. A fin de evitar el inconveniente de tener poblaciones con varianzas diferentes, y que al estimar un solo h, éste atenúe de más o de menos, según el caso. Habbema y Hermans (1977) estimaron un parámetro para cada grupo j y variable i, ( $h_{ij}$ ).

Se extienden las mismas propiedades que para el caso univariado.

### 1.3. METODO “EL VECINO MAS CERCANO”

#### 1.3.1. Generalidades

El método del vecino más cercano trabaja directamente con la probabilidad a posteriori de pertenecer a un grupo. Dado un elemento con su vector de mediciones  $\mathbf{x}$ , el objetivo es estimar la probabilidad condicional que el mismo pertenezca a la clase j. Una estimación de esta probabilidad está dada por la proporción de puntos próximos a  $\mathbf{x}$  que corresponde al grupo i. En particular, si se define una distancia



entre  $\mathbf{x}$  al punto  $k$ -ésimo más cercano de la muestra, es posible estimar la probabilidad por la proporción de casos que pertenecen al grupo  $j$ , entre los  $k$ -ésimos más cercanos a  $\mathbf{x}$ .

Si se trabaja con esta proporción del conjunto entre los  $k$ -vecinos más cercanos, los cuales corresponden al grupo  $j$  como estimación de  $f(j/\mathbf{x})$ , se logra una clasificación a través de la estimación  $\hat{f}(j/\mathbf{x})$ .

Acá  $k$  juega el rol del parámetro suavizante del Kernel, un  $k$  grande tiene menor varianza en la estimación de la probabilidad, pero es posible que introduzca más sesgo en la estimación de  $f(j/\mathbf{x})$ . Por el contrario, un  $k$  pequeño tendrá mayor varianza y menor sesgo. La diferencia entre  $h$  y  $k$  está dada en la cercanía o proximidad identificada por  $k$  acordando la densidad del conjunto de datos.

El parámetro  $k$  indica la distancia a considerar pero no dice cuál es la forma que tiene la misma, para lo cual es necesario elegir una métrica, es decir una unidad de medida para esa distancia. Esto es análogo al problema de la forma del Kernel que se comentó en el apartado anterior.

Para ello, dado dos vectores  $\mathbf{x}$  e  $\mathbf{y}$  representando dos puntos se puede medir la similitud entre ellos, definiendo una función  $d(\mathbf{x},\mathbf{y})$  que mide la distancia entre ambos elementos. Dicha medida de similitud utilizada es la distancia euclidiana al cuadrado:

$$d(\mathbf{x},\mathbf{y}) = (\mathbf{x} - \mathbf{y})' (\mathbf{x} - \mathbf{y}).$$

También es posible utilizar la distancia de Mahalanobis que en vez de considerar la matriz Identidad, considera la inversa de la matriz de covarianzas de cada una de las clases.

Este método, esencialmente estima las probabilidades promedio sobre los vecinos más cercanos. Si todas estas probabilidades son iguales, el vecino más cercano propuesto será una estimación insesgada. Dado que se desconoce  $f(i/\mathbf{x})$ , se puede realizar una estimación no paramétrica de esta probabilidad usando esta metodología. Para ello, se requiere elegir el valor de  $k$ . Su estimación, se realiza usando sólo los  $k$  puntos más cercanos, para lo cual es necesario examinar la muestra completa. En general, una de las formas en que puede hacerse este examen, suponiendo que se quieren encontrar en la muestra elementos cercanos al punto  $\mathbf{x}$  que ha de ser clasificado, es, en primer lugar dividiendo el conjunto de elementos de la muestra en una estructura de subconjuntos, es decir  $c$  grupos semejantes, luego cada uno de los  $c$  grupos será descompuesto en subgrupos, según los puntos más cercanos en el mismo

subgrupo. Se define así el radio como la máxima distancia entre la media del grupo y cualquier punto en ese mismo grupo.

Fix y Hodgs (1951) introducen la estimación de la densidad contando el número de puntos de la muestra ( $m_i(x)$ ) situados cerca de  $\mathbf{x}$ , fijando  $N_i$  (total de elementos de la población cercanos a  $\mathbf{x}$ ) y contando  $m_i(x)$ . Ellos proponen que se fije algún valor  $k$  y que  $N_i$  incluya los  $k$  puntos cercanos a  $\mathbf{x}$ .

Si  $d_{i,k}(x)$  es la distancia euclideana desde  $\mathbf{x}$  al  $k$ -ésimo punto más cercano entre los  $x_{ij}$  ( $j = 1, 2, \dots, n_i$ ), luego la estimación de la densidad por este método, de orden  $k$ , está definida por:

$$\hat{f}_i(x) = kn_i^{-1} / C_i(x) \quad , \quad \text{donde} \quad C_i(x) = V_d [d_{i,k}(x)]^d$$

siendo  $V_d$ , el volumen de la esfera unitaria en  $d$  dimensiones.

1.3.2. Regla con el método del vecino más cercano: 1 Vecino más cercano (k=1)

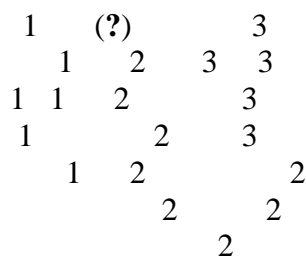
A modo de ejemplo, dado un conjunto de datos que contiene diferentes atributos y diferentes clases, por ejemplo el vector  $\mathbf{x}$  puede estar representado por un dígito y las clases pueden pertenecer al conjunto  $\{0,1,2,\dots,9\}$ ; si se define un conjunto  $D$  con  $P$  puntos,  $D = \{\mathbf{x}^j, c^j\}$  para  $j=1,2,\dots,P$ , la cuestión está dada en ubicar un nuevo  $\mathbf{x}$  en la clase correcta.

Para ello, se calcula la distancia entre  $\mathbf{x}$  (un punto particular) y cada uno de los puntos del conjunto.

$$d^j = d(\mathbf{x}, \mathbf{x}^j)$$

Luego se encuentra otro punto  $\mathbf{x}^{j^*}$ , el más cercano a  $\mathbf{x}$  de tal forma que la  $d^{j^*} < d^j$  para todo  $j=1,2,\dots,P$

Por último se asigna la clase  $c(x) = c^{j^*}$  (la clase del vecino más cercano). La figura siguiente muestra un ejemplo de esta regla:



Donde el nuevo vector representado por (?) es asignado a la clase 2 que es la clase a la cual pertenece el vecino más cercano.

Si hubiese dos o más puntos equidistantes y pertenecientes a distintos niveles, la clase más numerosa es elegida, si esto no sucede puede aplicarse la regla del k-ésimo vecino más cercano ( $k > 1$ ).

En otras palabras, esta regla está basada en la estimación de la densidad. Suponga que  $f_i(x)$  es estimada a través de la densidad Kernel,  $\hat{f}_1(x)$  donde  $K$  es la densidad uniforme sobre los  $N_1$  valores más cercanos, siendo estos los suficientes para contener un número  $k$  de puntos de la muestra ( $x_{ij}$ , siendo  $j = 1,2$  y  $i = 1,2, \dots, n_j$ ). Así, algunos  $N_2$  más cercanos son usados para estimar  $\hat{f}_2(x)$  de  $f_2(x)$ , Luego esto puede ser visto como una regla basada en el tamaño relativo de  $\hat{f}_1(x)$  y  $\hat{f}_2(x)$ , lo cual es equivalente al tamaño relativo de  $\frac{k_1}{n_1}$  y  $\frac{k_2}{n_2}$  donde  $k_i$  es el número de puntos de la muestra combinada en los  $N_i$  más cercanos de  $x$  desde cada grupo  $G_j$  ( $j = 1,2$ ).

Para  $k = 1$ , la regla basada en el tamaño relativo de  $\frac{k_1}{n_1}$  y  $\frac{k_2}{n_2}$  es precisamente el objeto-más cercano, la cual asigna un elemento con características del vector  $\mathbf{x}$  al grupo de los vecinos más cercanos en el conjunto.

Una de las desventajas de esta metodología es el hecho de que existan valores alejados, por lo cual considerar el  $K$  vecino más cercano es más robusto para clasificar estos puntos.

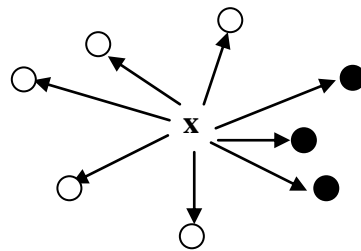
### 1.3.3. Regla con k vecinos más cercanos ( $k > 1$ )

En esta regla, se quiere incluir más de un vecino en la decisión sobre la clase a la que pertenece un nuevo punto  $\mathbf{x}$ . Para ello, considerando la distancia euclidiana y la hiperesfera centrada en  $\mathbf{x}$  con radio  $r$ , se incrementa este radio hasta que la hiperesfera contenga exactamente  $K$  puntos. La clase  $c(x)$  está dada a su vez por la clase mas numerosa dentro de la hiperesfera.

Para elegir  $K$ , habrá que buscar aquel que de la mejor performance, ya que si es muy grande, todas las observaciones se clasifican en la clase más numerosa y queda prácticamente una sola clase.

Más allá de todas las pruebas que se realizaron, dada la importancia en la elección de  $K$ , se propone  $n^{3/8}$  o  $n^{2/8}$  dependiendo de la similitud o las diferencias entre las estructuras de las matrices de covarianzas subyacentes a los grupos.

El valor de  $K$  determina el grado de irregularidad en la estimación de la función de densidad, por ello es un parámetro de suavizado. Cuando es grande, la densidad estimada es más suavizada que cuando  $K$  es pequeño. En la figura siguiente puede observarse el siguiente ejemplo:



○ Clase A

● Clase B

Se está ante un cierto espacio con distancias definidas, el estimador de la densidad por el método del  $k$ -ésimo vecino más cercano puede ser usado para estimar la confianza de la muestra que se está analizando. Usando el 8 vecino más cercano, la confianza de que el nuevo elemento pertenezca a la clase A es  $5/8$  (proporción de elementos a posteriori de medir las distancias), mayor a  $3/8$ , por lo tanto pertenece a la clase A.

En síntesis, el método no depende de que los datos estén distribuidos normalmente, sino de la distancia entre las parejas detectores de observaciones.

La idea básica de este análisis discriminante es que para una nueva observación que deba clasificarse, en primer lugar se encuentra otra observación en el conjunto de datos que esté más cerca a esa nueva observación (es decir, la mínima distancia). Así se asigna la nueva observación al grupo del que proviene la observación vecina más cercana a ella.

Si hubiera un empate, es decir, si las distancias entre la nueva observación y dos o más de las demás observaciones son idénticas, perteneciendo a diferentes grupos, se le asigna la clase que posee mayor cantidad de elementos en ese radio, de lo contrario el procedimiento busca su siguiente vecina más cercana, a menos que las observaciones empatadas pertenezcan al mismo grupo. Si esto sucede, la nueva

observación se asigna a ese grupo. Si las observaciones no pertenecen al mismo grupo y la siguiente vecina mas cercana se acopla a uno de estos grupos, entonces el procedimiento busca la siguiente vecina más cercana y así sucesivamente.

#### 1.4. METODO DEL “ARBOL DE CLASIFICACION”

##### 1.4.1. Generalidades

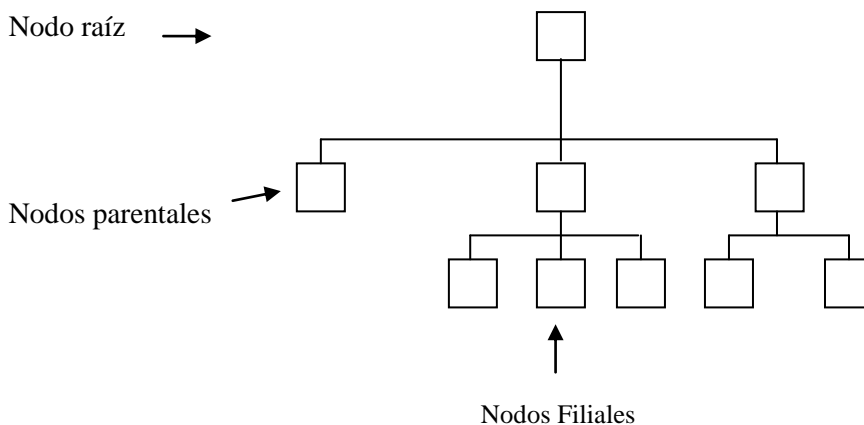
Actualmente ha aumentado en popularidad el uso de árboles de clasificación y regresión, dada la versatilidad y fácil aplicabilidad en problemas donde se trata en particionar conjuntos de datos e identificar sus estructuras subyacentes. Se han comprobado sus ventajas tanto para pequeños como para grandes conjuntos de datos, como también cuando se trata de clasificar con relativa exactitud casos ya conocidos o bien otros casos que se presentan por primera vez en el modelo.

Los árboles de clasificación (classification trees) trabajan con variables dependientes (variables de criterio) categóricas, mientras que el método de árboles de regresión (regression tree) lo hace con variables de criterio continuas.

Estamos ante un método recursivo que se usa para predicción y que ha surgido como una alternativa a otros modelos paramétricos o no paramétricos con finalidades similares.

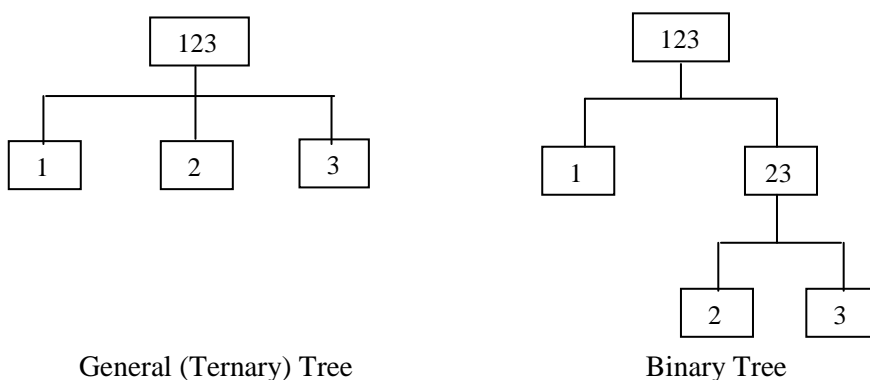
El árbol presenta un nodo raíz con ramificaciones, según se presenta en la figura 1.5. Dentro de las ramificaciones, llamamos nodos parentales a los nodos ubicados en el mismo nivel y nodos filiales respecto a los que se ubican en forma descendente respecto a otro nodo superior.

Figura 1.5. Árbol de Clasificación



El total de elementos de la muestra está subdividido en subconjuntos o grupos. Si el total de datos se divide en dos grupos, se está en presencia de un árbol binario. En la práctica, suele ser necesario clasificar los datos en más de dos grupos (general tree), lo que no significa que el hecho de dividir en muchos grupos apunte a una mayor precisión, ya que cada nodo es una permutación de otros, lo cual puede observarse en la figura 1.6.

Figura 1.6. Tipos de árboles



Es decir, que ambos árboles presentados tienen el mismo número de parámetros (puntos de división) y cualquier estadística asociada al árbol general (de la izquierda) puede ser convertida trivialmente para estimar uno del árbol binario. A este respecto, Hand (1996) lo considera como un método de partición recursivo, ya que sobre la base de los valores de la variable decide como particionar al conjunto de datos.

A modo de ejemplo, se supone el caso que se presenta en la figura 1.7.(a), donde se muestra un árbol simple, con dos variables continuas. En primer lugar, si la variable  $x$  de un nuevo objeto que pretende ser clasificado es  $x < 1$ , entonces se examina la variable  $y$ . Si  $y < 1$  el objeto es asignado a la clase 1, si sucede lo contrario ( $y > 1$ ) el objeto se asigna a la clase 2. Por otro lado, si  $x > 1$ , se sigue con la misma variable para valores ahora mayores a 1,5; donde si esto sucede el objeto va a la clase 1, de lo contrario va a la clase 2, independiente del valor de  $y$  para esta rama. Esto mismo puede representarse con otra estructura (figura 1.7.(b))

Figura 1.7. (a) Ejemplo de árbol de clasificación con dos variables continuas

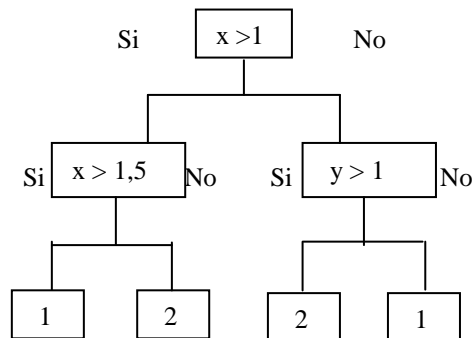
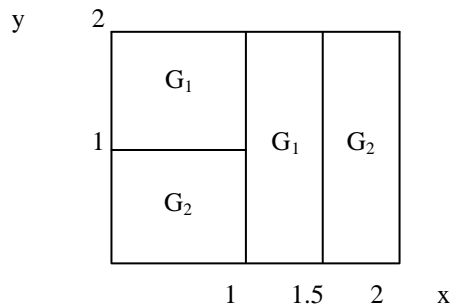


Figura 1.7. (b) Ejemplo de clasificación en dos grupos según Fig. 2.3. (a)



Una conclusión importante, que se deduce, es que una variable puede ser considerada en más de una ocasión, según su rango de valores, para clasificar objetos.

Si bien en este ejemplo se introduce una idea de este método, la cuestión radica en como se construye el mismo. Qué variables se deben usar? Cómo se particiona la variable? La partición produce un nodo, el cual está basado en más de una variable, en una combinación, y en tal caso ¿en cuál de ellas?, ¿cómo se caracterizan los grupos dadas las diferentes particiones de las variables?, etc.

Desde otra perspectiva de análisis, algunos autores lo consideran a este método como no paramétrico, porque en los grupos no se presupone ninguna distribución de probabilidad, lo cual lo hace muy flexible y de fácil interacción entre las variables. Otro mérito es su simplicidad conceptual, tiene como ventaja el hecho de trabajar con vectores de mediciones de distinta dimensionalidad de objetos.

El árbol binario es la forma más común y básicamente consiste en tomar los valores de la variable y particionarlos en brazos, que previamente se han definido. Cuando la variable está ordenada de alguna manera, se define una regla de clasificación, si los valores son menores que lo que indica la regla va al brazo 1; y si son mayores van al otro brazo. En cambio, si la variable no está ordenada, se necesita una partición explícita para asignar cada categoría de variable a un brazo particular. En definitiva, cuando se construye un árbol, es preciso decidir cómo son en particular los nodos considerados y sobre qué variable se hará la división, al igual que la propia naturaleza de esta división.

#### 1.4.2. Construcción del árbol

Una clasificación es una partición de  $X$  (espacio de mediciones, es decir, aquél que contiene todos los posibles vectores de mediciones) en  $J$  subconjuntos disjuntos:  $A_1, \dots, A_j$ , siendo entonces  $X = \bigcup_j A_j$  para todo  $x \in A_j$ , siendo  $j$ , la cantidad de grupos o poblaciones,  $j = 1, 2, \dots, G$ . Para construir una regla de clasificación necesitamos contar con una muestra. Esta muestra estará formada por los vectores de mediciones que caracterizan a los diferentes elementos.

El propósito del análisis de clasificación, depende del problema, puede ser por un lado un propósito de producir una clasificación precisa o bien determinar su estructura predictiva. Si se aspira a éste último se debe tener idea de las variables y de las interacciones, es decir, conocer las características del fenómeno, esto determina cuando un objeto está en un grupo y no en otro. A este respecto, muchas de las técnicas estadísticas disponibles fueron diseñadas para conjuntos pequeños de datos que cuentan con estructura estándar, con todas las variables del mismo tipo, es decir con el supuesto fundamental que el fenómeno es homogéneo. Esto es, que la relación entre variables está sujeta a todo el espacio de mediciones, lo cual conduce a modelos donde sólo pocos parámetros fueron necesarios para trazar los efectos que envuelven varios factores. Por el contrario, cuando la muestra es grande y a su vez se definen muchas variables, pueden diseñarse otras estructuras y habrá que tener en cuenta otros aspectos, tales como la alta dimensionalidad, tipos de datos mixtos,



estructura no estándar de los mismos y lo más difícil: la no homogeneidad, esto es, relaciones entre variables en diferentes partes del espacio de mediciones.

Para construir el árbol de clasificación, el primer problema es cómo usar la muestra para determinar la división binaria del espacio de mediciones  $X$  en partes.

Esta idea será implementada en el siguiente sentido:

Primero, será conveniente comenzar por definir la proporción de los nodos  $p(j/t)$ , para  $j = 1, 2$  como la proporción de casos  $\mathbf{x} \in t$  (nodo) correspondiente a la clase  $j$ , por lo tanto:  $p(1/t) + p(2/t) = 1$

En segundo lugar, se define una medida  $r(t)$  de impureza de  $t$ , como una función no negativa  $\phi$  de dichas proporciones, resultando máxima la impureza si las dos clases están igualmente mezcladas en cada nodo y será pequeña cuando éstos contengan una sola clase.

Tercero, definir el conjunto de alternativas  $S$  de divisiones binarias  $s$  de cada nodo. Esta asociación de divisiones  $s$  envía todos los casos  $\mathbf{x}$  en  $t$  que conteste “sí” a  $t_L$  y todo  $\mathbf{x}$  que conteste “no” a  $t_R$ . Para pasar de un nivel a otro, habrá que hacer preguntas. Al terminar el árbol, se ha definido una regla heurística. Cuando un nodo  $t$  no llega a distinguir la impureza, entonces  $t$  no se divide más y ese nodo es terminal. Específicamente, si:

$$p(J_0/t) = \max_j p(J/t)$$

entonces  $t$  ha designado a la clase  $J_0$  como el nodo terminal.

Previamente, es necesario definir conceptualmente las variables con las que se va a trabajar. Ellas son:

*Variable criterio:* es aquella cuyo resultado se desea predecir a partir de otras variables, también se conoce con el nombre de variable dependiente o explicativa.

*Variables predictoras:* son aquellas que predicen el patrón de comportamiento de la variable criterio, también llamadas variables independientes.

Existen cuatro métodos para desarrollar árboles, que se sintetizan en la Tabla 1.1.

Tabla 1.1. Métodos de árboles de decisión

MÉTODO	CARACTERÍSTICAS
CHAID (Detector automático de interacciones mediante pruebas Chi-cuadrado) (Kass 1980)	Genera árboles no binarios y la variable criterio puede ser categórica o numérica continua. El árbol se detiene cuando se encuentran categorías estadísticamente heterogéneas. Utiliza estadísticos chi cuadrado para identificar divisiones óptimas.
CHAID exhaustivo (Biggs, de Ville y Suen 1991)	Mejora el método de Kass cuando las categorías que han quedado son heterogéneas, ya que sigue fundiendo categorías, analizando luego cuales son las que proporcionan una mayor asociación con la variable criterio calculando para ello un valor p corregido por esa asociación.
CART (Árboles de regresión y clasificación) (Breiman, Friedman, Olshen y Stone 1984)	Genera árboles binarios, por lo cual puede utilizar la misma variable predictora más de una vez. Se trata de un proceso recursivo que se repite hasta alcanzar el criterio de homogeneidad, minimizando medidas de impureza.
QUEST (Árbol estadístico eficiente, insesgado y rápido) (Loh y Shih 1997)	Genera árboles binarios, tratando por separado la selección de variables y los puntos de división. Si todas la variables predictoras son igualmente informativas respecto a la variable criterio, se selecciona cualquiera de ellas con la misma probabilidad.

El CHAID, (detector automático de interacciones mediante chi – cuadrado) es una técnica estadística muy eficiente para generar árboles, presentada por Kass (1980), que utiliza como criterio, la significación de un contraste estadístico, evaluando todos los valores de una variable predictora potencial. Reúne los valores evaluados como estadísticamente homogéneos respecto de la variable criterio y conserva inalterados todos los valores que resulten distintos (no homogéneos). Luego, selecciona la mejor variable predictora para formar la primera rama del árbol de decisión, de forma que cada nodo esté compuesto por un grupo de valores homogéneos de la variable seleccionada y así hasta desarrollar por completo, el árbol.

La prueba estadística que se utiliza cuando la variable criterio es categórica es la chi-cuadrada, mientras que si la variable criterio es continua, se usa la prueba F.

Este método puede generar más de dos categorías en cualquier nivel del árbol, es decir, que no es binario.

El CHAID exhaustivo es una modificación del método anterior, presentado por Biggs, de Ville y Suen (1991) y tiende a resolver problemas que presenta el CHAID. Particularmente, cuando no se encuentra la división óptima de una variable, éste método permite encontrar la mejor división para cada predictor y elegir el

mismo que se debe dividir. En muchas ocasiones, dependiendo de los datos utilizados, es posible que no haya diferencia entre los resultados obtenidos con ambos métodos.

El CART (árboles de regresión y clasificación) es un algoritmo binario para generar árboles, creado por Breiman, Friedman, Olshen y Stone (1984). Esta metodología consiste en dividir los datos en dos subconjuntos, de modo que los datos comprendidos en cada uno de los subconjuntos sean más homogéneos que en el subconjunto anterior. Se trata de un proceso recursivo, que se repite hasta alcanzar el criterio de homogeneidad o hasta llegar al criterio de detención definido: Por ello, puede utilizarse la misma variable predictora varias veces en distintos niveles de árbol.

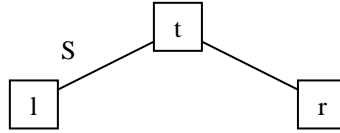
Por último, el QUEST (árbol estadístico eficiente, insesgado y rápido) es también un algoritmo binario, creado por Loh y Shih (1997). Este método trata por separado la selección de variables y la selección del punto de división. Si todas las variables predictoras son igualmente informativas respecto a la variable criterio, QUEST selecciona cualquiera de estas variables con la misma probabilidad.

En esta oportunidad se ha aplicado al conjunto de datos el método CART. Su *algoritmo* funciona eligiendo en cada nodo una división, de modo que cada nodo filial sea más puro que su nodo parental. La pureza se refiere a los valores de la variable criterio, de tal forma que en un nodo completamente puro, todos los casos tienen el mismo valor para la variable criterio. Para ello, el método mide la impureza de la división de un nodo dado, definiendo una *medida de impureza*. El índice de impureza mide en un nodo, las diferencias entre las probabilidades correspondientes a cada clase.

Se supone, por ejemplo: que un nodo posee 10 objetos, 5 son de la clase 0 y 5 de la clase 1. Si los elementos de la clase 0 corresponden a valores de la variable predictora  $x$ , menores que 6, los de la clase 1, serán, por el contrario, mayores que 6. Este nodo, entonces, es relativamente impuro, tienen igual número de elementos de cada una de las clases. Un nuevo objeto tiene el 0,50 de chance de pertenecer a la clase 0 y una proporción igual para la clase 1.

En la figura 1.8., un nodo puede dividirse en dos divisiones, según un criterio.

Figura 1.8. Ejemplo de división de un nodo



Donde:

t: un nodo raíz,

S: criterio de división,

l y r: nodos filiales de t

$i(t)$ : medida de impureza del nodo t.

$\pi(l)$ : proporción de casos (l) que tiene el nodo t, según la división S.

$\pi(r)$ : proporción de casos (r) que tiene el nodo t, según la división S.

Una medida del cambio de impureza, la cual puede ser producida por la división S del nodo t, esta dada por:

$$\Delta i(S, t) = i(t) - \pi(l)i(l) - \pi(r)i(r)$$

mientras mayor sea  $\Delta i$ , es mejor la división, es un criterio de partición.

En otro orden de ideas, una medida global de impureza es:

$$I(A) = \sum_{t \in T} i(t)\theta(t)$$

siendo T: el conjunto de nodos terminales del árbol y  $\theta(t)$  es la proporción de elementos que caen en el nodo t.

En el caso que existan dos árboles A y A' (formado por una división S del nodo terminal t de A), se definen ambas medidas de impureza:

$$I(A) = \sum_{t \in T_A} i(t)\theta(t)$$

y

$$I(A') = \sum_{t \in T_{A-t}} i(t)\theta(t) + i(l)\theta(l) + i(r)\theta(r)$$

donde l y r representa las ramas izquierda y derecha del nodo t.

$$\theta(l) = \theta(t)\pi(l)$$

$$\theta(r) = \theta(t)\pi(r)$$

Si el cambio en la impureza es:

$$\Delta i = I(A) - I(A')$$

reemplazando cada una de los términos por sus expresiones equivalentes, queda:

$$\Delta i = \Delta i(S, t)\theta(t) \quad (1)$$

esto es la reducción de la impureza del árbol causado por la división del nodo  $t$ , que es proporcional a la reducción de la impureza del nodo, donde el factor de proporcionalidad es el tamaño de ese nodo.

Existen cuatro medidas de impureza distintas, que dependen del tipo de variable criterio.

*Gini*: el índice Gini en el nodo  $t$ , se define como:

$$g(t) = \sum_{j \neq i} p(j/t)p(i/t)$$

donde  $i$  y  $j$  son categorías de la variable criterio.

$$g(t) = 1 - \sum_j p^2(j/t)$$

cuando los casos de un nodo están distribuidos uniformemente entre las categorías, el índice Gini toma su valor máximo de  $1 - (1/k)$ , donde  $k$  es el número de categorías de la variable criterio. Cuando todos los casos del nodo pertenecen a la misma categoría, el índice Gini es igual a 0. La expresión (1) ya presentada, es el criterio Gini y es el valor que se informa en el árbol como “mejora”.

*Binario*: Este índice se basa en la división de las categorías criterio en dos grupos, y después busca la mejor división de la variable predictora según estos dos grupos. La función del criterio binario para la división  $S$  en el nodo  $t$  es una medida de impureza. La división  $S$  es la que maximiza este criterio.

*Binario Ordenado*: es una modificación del índice anterior, para variables criterio ordinales. La diferencia consiste en que, con el criterio binario ordenado, sólo las categorías contiguas se pueden combinar para formar los grupos.

*Desviación mínimo-cuadrática (LSD)*: para variables criterio continuas se utiliza la medida de impureza LSD, la cual no es otra cosa que la varianza ponderada intranodo para el nodo  $t$ , y es igual a la estimación del riesgo por resustitución para dicho nodo.

Para llevar a cabo el análisis CART, su algoritmo realiza las siguientes etapas:

1. comenzando por el nodo raíz  $t=1$ , se debe buscar la división  $s^*$ , entre todos los candidatos posibles  $S$ , que de lugar a la mayor reducción de la impureza:

$$\phi(s^*,1) = \max_{s \in S} \phi(s,1)$$

luego dividir el nodo 1 ( $t=1$ ) en dos nodos,  $t=2$  y  $t=3$ , utilizando la división  $s^*$ .

2. Repetir el proceso de búsqueda de divisiones para uno de los nodos  $t=2$  y  $t=3$ , y así sucesivamente

3. Continuar con el proceso de desarrollo del árbol hasta alcanzar al menos una de las reglas de parada

Dentro de las *reglas de parada*, se cuentan dos tipos:

a) reglas debida a los datos:

\* Todos los casos de un nodo tienen valores idénticos en todos los predictores.

\* El nodo se vuelve puro, ya que todos los casos tienen el mismo valor en la variable criterio.

b) reglas debido al soft utilizado:

\* La profundidad del árbol ha alcanzado el valor máximo preestablecido

\* El número de casos que constituyen el nodo es menor que el tamaño mínimo preestablecido para un nodo parental (de igual nivel que otro)

\* La división del nodo tiene como resultado un nodo filial (de nivel inferior que otro, del cual desciende) cuyo número de casos es menor que el tamaño mínimo preestablecido para un nodo filial.

Este último grupo de reglas puede configurarse en el soft utilizado, mientras que el primer grupo es el que realmente nos interesa a los fines estadísticos.

### 1.5. PRECISION DE LOS DIFERENTES METODOS

La precisión de los diferentes métodos es medida a través de la tasa de error. La importancia de esta tasa está en evaluar una regla de clasificación.

Para clasificar un nuevo elemento cuyas mediciones están representadas en un vector  $\mathbf{x}$ , en uno de los  $G$  grupos ( $G_1, \dots, G_g$ ), se utiliza una regla discriminante,  $r(x,t)$  formada por los datos de la muestra:  $t' = (y_1, y_2, \dots, y_n)$ ,

donde  $y_j = (x'_j, z'_j)'$  y  $(z_j)_i$  será 1 si  $x_j$  corresponde a  $G_i$  y asume el valor 0 si no pertenece a dicho grupo. Para  $i = 1, 2, \dots, g$  ;  $j = 1, 2, \dots, n$

Con el objetivo de evaluar la clasificación, nuestro interés se centra en la tasa de error, al clasificar un elemento:

$$ec_{ij}(F_i, t) = \Pr\{r(x, t) = j / x \in G_i, t\}$$

$$i, j = 1, 2, \dots, g$$

Es decir, es la probabilidad condicional sobre  $t$ , que un elemento seleccionado aleatoriamente de  $G_i$  sea asignado a  $G_j$  por la regla. Siendo,  $F_i$  la función de distribución de  $x$ , en el grupo  $G_i$ . La tasa esperada de error es la tasa de clasificación errónea no condicional de la regla.

La tasa de error específica de un grupo es la probabilidad condicional de clasificación errónea de elementos elegidos aleatoriamente de  $G_i$ . Lo cual, puede denotarse:

$$ec_i(F_i, t) = \sum_{j \neq i}^g ec_{ij}(F_i; t)$$

Por otro lado, la tasa de error condicional para un individuo seleccionado aleatoriamente de un subconjunto de Grupos con probabilidades a priori diferentes, es decir con proporciones  $\pi_1, \dots, \pi_g$ , es:

$$ec(F, t) = \sum_{i=1}^g \pi_i ec_i(F_i; t)$$

A fin de estimar las tasas de error, veremos a continuación, dos posibles estimadores:

### 1.5.1. Tasa de error aparente

En primer lugar, consideraremos la estimación de la tasa de error para el primer grupo. Un estimador no paramétrico común de la tasa de error condicional  $ec_1(F_1; t)$  es la tasa de error aparente  $A_1(t)$ , sugerida por Smith (1947) que se obtiene de clasificar la regla de clasificación a los datos de la muestra de  $G_1$ . Es decir,  $A_1(t)$

es la proporción de observaciones de este grupo mal clasificadas por la regla, que puede denotarse como:

$$A_1(t) = \frac{1}{n_1} \sum_{j=1}^n z_{1j} Q[1, r(\mathbf{x}_j; \mathbf{t})]$$

donde  $n_1$  es el número de observaciones del grupo  $G_1$ . Definimos  $Q$  como una función indicadora que será igual a 1 si el objeto  $j$ -ésimo es mal clasificado con la regla establecida y será igual a 0, si el objeto es clasificado correctamente.

Como esta tasa es obtenida aplicando la regla al mismo conjunto de datos con los cuales fue formada, ésta provee una evaluación optimista de la verdadera tasa de error condicional. La sobredimensión del error puede ser peligrosa y provocar un elevado sesgo.

La proporción de observaciones en cada grupo, mal clasificadas por aplicación de la regla, permite obtener la tasa de error, como la probabilidad condicional que  $\mathbf{x}_j$  pertenece a un grupo en particular.

### 1.5.2. Tasa de error por Cross-Validation

Una manera de evitar el sesgo de la tasa de error aparente como consecuencia de que la regla ha sido testeada con el mismo conjunto de datos, puede estimarse la tasa de error a través de otra metodología, como la aflicción del método Holdout, en el cual los datos disponibles son divididos en dos subconjuntos: uno que será la muestra de entrenamiento y el otro, como muestra de validación. La regla discriminatnea formada por los datos de la muestra de entrenamiento y luego validada con la otra muestra. El método es insesgado pero no es aplicable cuando la muestra es de tamaño reducido. Un caso particular de éste es el método de “dejar uno afuera”, técnica descrita por Lachenbruch y Mickey (1968) o Crossvalidation trabajado por Stone (1974) y Geisser (1975).

En este trabajo se validaron los resultados con la tasa de validación cruzada (crossvalidation). La proporción de observaciones mal clasificadas está dada por:

$$A_1^{cv} = \frac{1}{n_1} \sum_{j=1}^n z_{1j} Q[1, r(\mathbf{x}_j; \mathbf{t}_{(-j)})]$$



donde:  $t_{(j)}$  denota el conjunto con el punto  $y_j$  que es excluido ( $j= 1, \dots, n$ );  
 $n_1$  es el número de observaciones del grupo 1.

$$Q = 0, \text{ si } 1 = r(\mathbf{x}_j ; \mathbf{t}) \text{ y } Q = 1 \text{ si } 1 \neq r(\mathbf{x}_j ; \mathbf{t})$$

Por lo tanto, antes de que la regla sea aplicada, se excluye un elemento y la regla es recalculada sobre la base de los elementos restantes. Es decir, se toman  $n-1$  en vez de  $n$  observaciones como un conjunto y determina con esas  $n-1$  una regla que luego es aplicada para clasificar la observación que ha quedado excluida. Así para cada una de las observaciones. El ratio de clasificación errónea para cada grupo es la proporción de observaciones muestrales en ese grupo, en el cual son mal clasificadas. Los datos fueron procesados con SAS y se calculó la tasa de error por crossvalidation.

Generalmente, el ratio de error es estimado a través del promedio de las estimaciones de los ratios de error de los grupos específicos individuales, donde las probabilidades a priori son usadas en las ponderaciones.

A fin de evaluar las reglas de clasificación de los métodos no paramétricos utilizados, se midió su precisión para cada uno de ellos, según la tasa de error por crossvalidation.

La precisión en el método del Árbol de Clasificación indica hasta qué punto el comportamiento del árbol es satisfactorio respecto a la predicción del resultado deseado o la clasificación de los individuos. El proceso inverso a la validez predictiva se denomina riesgo y las tablas de clasificación errónea indican la precisión del árbol final.

El riesgo estimado de clasificación errónea se calcula como la proporción de casos de la muestra clasificados incorrectamente por el árbol.

Podar un árbol suele ser útil, a fin de evitar que el tamaño sea excesivo, ya que en un árbol voluminoso, los nodos terminales pueden no proporcionar información útil. El motivo principal de la poda es evitar el sobreajuste, es decir que el árbol recoja no solo los patrones reales presentes en los datos, sino también parte del “ruido” o error específico de la muestra.

El hecho de eliminar una rama del árbol debe provocar un riesgo de clasificación errónea no mucho mayor que el encontrado en los árboles posibles y debe minimizar el costo por complejidad, ambas situaciones pueden definirse como:

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}|$$

donde:

$R(T)$ : riesgo de clasificación errónea para el árbol  $T$ .

$|\tilde{T}|$ : número de nodos terminales del árbol  $T$

Esta medida es una combinación lineal del riesgo para el árbol  $T$  y su complejidad.

Si  $\alpha$  es el costo de complejidad por nodo terminal, entonces  $R_\alpha(T)$  es la suma del riesgo para el árbol  $T$  y este costo.

Cualquier árbol que se pueda generar tiene un tamaño máximo (un solo caso en cada nodo terminal), si  $\alpha=0$ , el árbol máximo tiene un riesgo mínimo, ya que todos los casos están perfectamente pronosticados. A mayor  $\alpha$ , menor será el número de nodos terminales, eliminándose las divisiones más débiles.

En esta metodología se dividió la muestra en una serie de submuestras más pequeñas. Posteriormente se generan los árboles, excluyendo los datos de cada submuestra. El primer árbol se basa en todos los casos excepto aquellos de la primera submuestra, el segundo se basa en todos los casos excepto los de la segunda submuestra y así sucesivamente. Para cada árbol se calcula el riesgo de clasificación errónea aplicando el árbol a la submuestra que se excluyó cuando se generó. La estimación del riesgo mediante validación cruzada para todo el árbol se calcula como el promedio de los riesgos para todos los árboles.

## **CAPITULO 2: CASO DE APLICACIÓN DE LOS METODOS DE CLASIFICACION NO PARAMETRICOS**

### **2.1 INTRODUCCION**

Desde hace algún tiempo las empresas tienen como objetivo predecir los resultados futuros del gerenciamiento empresarial. En 1961 se propuso el modelo “Z score” como herramienta de clasificación de riesgo de empresas manufactureras con cotización pública de acciones en los EEUU (Altman 1961). Este modelo surge de la suma de ciertos índices de balances, ponderados por coeficientes que representan su importancia relativa; si bien es un método sencillo, trajo algunos inconvenientes, ya que se requiere de un profundo análisis de la experiencia del cumplimiento e incumplimiento de las empresas. Además no era fácil trasladar el modelo a otras actividades, por lo cual recibió algunas críticas.

En 1977, Altman, Haldeman y Narayanan, desarrollaron el modelo Zeta, con la intención de mejorar el viejo modelo score, considerando empresas de mayor tamaño y con perfil financiero. Este modelo utilizó muestras de firmas insolventes donde el tamaño del activo dos años antes de la falencia tenía determinados niveles, mayores a los que se utilizaron en el modelo score.

Además se consideró la importancia del tiempo en los datos y se incorporaron otras ramas de actividad, no necesariamente industriales. Se le agregaron objetivos predictivos y no sólo de clasificación.

Como conclusión, llegaron a que, tomando un año antes de la falencia, se clasificaron con éxito el 90 % de las empresas que conformaron la muestra y con un horizonte de 5 años, el 70 %.

En esta oportunidad se pretende analizar la calidad de los métodos no paramétricos en la predicción a mediano plazo de los procesos de gestación e instalación de estados de vulnerabilidad financiera en empresas privadas del medio.

La Convención de Basilea, realizada en el año 1995 recomendó a los países miembros la adopción generalizada de estas técnicas de predicción, sobre todo en el sistema de la banca oficial y privada, como medio idóneo para disminuir los encajes

bancarios, acrecentar la oferta de préstamos, aumentar la rentabilidad del negocio bancario y por esta vía reducir las tasas de interés activas.

## 2.2 UNIVERSO DE ESTUDIO.

Se elige un conjunto de empresas que están en estado de falencia o de insolvencia en un cierto período (que según las definiciones del párrafo siguiente, son empresas que cotizan sus acciones en ronda reducida) y simultáneamente otro grupo de empresas similares (tamaño y rama de actividad) pero que exhiben en el período un comportamiento financiero regular o bueno.

Estas empresas se seleccionan de los registros que posee la Bolsa de Comercio de Buenos Aires que reúne alrededor de 250 empresas (sociedades anónimas) que cotizan acciones en Bolsa y para permanecer en la ronda de empresas cotizantes las entidades deben cumplir con los requisitos establecidos en el Capítulo XIV del Reglamento. En este sentido, el art. 38 inciso a), b) y c) establecen *“La bolsa debe transferir la negociación de las acciones a “ronda reducida” cuando la sociedad emisora se encuentre en alguna de las siguientes situaciones: a) solicite su concurso preventivo, b) surjan de un estado contable o de información suministrado por la sociedad, resultados no asignados negativos que insumen las reservas y el 50 % del capital ajustado, c) surjan de un estado contable o de información suministrado por la sociedad, resultados no asignados negativos por un monto superior al 75 % del patrimonio neto. El monto del patrimonio neto a considerar será el resultando de excluir el resultado no asignado...”*.

A su vez el art. 41 inciso a) establece también *“la transferencia de negociación a ronda reducida cuando la emisora incluya en su información contable, conceptos, que a juicio de la Bolsa, se aparten de las normas contables profesionales y ello afecte significativamente su situación patrimonial y los resultados”*.

Si bien en un grado menor al concurso preventivo o a la quiebra, no cabe duda que la transferencia a ronda reducida constituye un signo ostensible de vulnerabilidad financiera, razón por la cual en el presente trabajo se califica como **empresa en crisis** aquella que durante el período de análisis entró por primera vez a ronda reducida.

### **2.3. EL ANÁLISIS DE LOS ESTADOS CONTABLES COMO HERRAMIENTA DE DIAGNOSTICO.**

La crisis empresarial es un estado caracterizado por estar afectada la condición de viabilidad, en la cual se produce un desequilibrio generalizado cuya principal manifestación es la disminución de la capacidad de regenerar ciclos de actividad (Escandell, José). Se afecta a la estructura de rentabilidad y a la estructura de financiación.

El análisis de estados contables es el sistema de información que investiga, a partir de la información contable, cuál es la situación de la empresa, para determinar sus causas y sugerir los cursos de acción más adecuados en las circunstancias, según sea la finalidad perseguida.

Es conveniente separar la investigación en sus aspectos económicos de los aspectos financieros, subdividiendo este último en las circunstancias del corto plazo y la de largo plazo. La situación financiera se refiere a la suficiencia del patrimonio para cumplir los fines del ente. Este concepto implica la capacidad para proveer los recursos necesarios para el funcionamiento del ente en adecuadas condiciones, así como para afrontar las obligaciones que surgen de los compromisos que asume por la realización de su actividad.

**El análisis de la situación financiera a largo plazo** tiene por objeto el estudio de la suficiencia del patrimonio neto para el funcionamiento del ente en adecuadas condiciones, así como para afrontar las obligaciones que surgen de los compromisos que asume por la realización de su actividad y comprende el estudio de la estructura de financiación, de la política de inversión, de las políticas financieras y de la capacidad de autofinanciación, en un período de al menos varios ejercicios anuales.

**La situación financiera a corto plazo** comprende el instrumental que mide la capacidad de la entidad para hacer frente a los compromisos y obligaciones del corto plazo. El capital de corto plazo es aquel que es necesario para cumplimentar el funcionamiento normal del ciclo operativo, o sea el capital de trabajo o capital corriente (Activo Corriente – Pasivo Corriente), es decir, se pretende estudiar la suficiencia del capital de trabajo para afrontar las obligaciones que surgen de los compromisos que asume por la realización de su actividad.

Por último, **el análisis de la situación económica** estudia los resultados de la empresa con el objeto de determinar su capacidad de generación de utilidades y su finalidad es verificar la eficiencia de la administración.

Estas estructuras pueden ser utilizadas para establecer un marco de referencia para el proceso de diagnóstico en las organizaciones, acorde con su realidad y con las políticas dadas por la misma.

El diagnóstico empresarial constituye el nexo entre el estudio - investigación y la programación de actividades mediante la detección de anomalías para generar en la organización, una situación que escape a la crisis.

A fin de entender en que consisten los estados contables y cómo se produce su análisis, en el cuadro 2.1. se esquematizan sus principales partes:

- a) Activo, que comprende el Activo Corriente o de corto plazo (AC) y el Activo No Corriente o de largo plazo (ANC).
- b) Pasivo, que comprende el Pasivo Corriente o de corto plazo (PC) y el Pasivo No Corriente o de largo plazo (PNC).
- c) Patrimonio Neto o capital propio, es todo lo que posee la empresa deducidas las deudas. Esta formado, a su vez, por el capital aportado por los dueños de la empresa (socios) y los resultados obtenidos en los diferentes períodos).

Cuadro 2.1 Esquema de los Estados Contables



#### 2.4 SELECCIÓN DEL GRUPO DE ANALISIS

En el período investigado, Julio de 1995 a Diciembre de 2000, se seleccionaron 25 empresas que pasaron a ronda reducida, a las cuales se les pudo asignar una empresa como par en el grupo de control.

El año de ingreso a ronda reducida se consideró año base y a los fines específicos de las estimaciones efectuadas en el presente trabajo, se tomaron en cuenta los balances cerrados dos años antes del pase a ronda reducida. Es decir, que el horizonte de pronóstico oscila en dos años aproximadamente.

La conformación de ambos grupos por rama de actividad y por tamaño (volumen de Activo e Ingresos Totales) se muestran en los Cuadros 2.2 y 2.3.

Cuadro 2.2. Empresas en ambos grupos por rama de actividad.

RAMA DE ACTIVIDAD	EN RONDA REDUCIDA (CANTIDAD)	EN RONDA NORMAL (CANTIDAD)
Comercio	2	5
Siderurgia	2	1
Generación de electricidad	1	1
Metalmecánica	7	9
Alimenticia	5	4
Textil	2	2
Papel	2	1
Química	4	2
Totales	25	25

Cuadro 2.3. Empresas en ambos grupos según Activos e Ingresos Totales.

Ingresos Totales			Activos Totales		
Intervalos (miles de pesos)	En ronda reducida (cantidad)	En ronda normal (cantidad)	Intervalos (miles de pesos)	En ronda reducida (cantidad)	En ronda normal (cantidad)
550 – 20500	12	12	550 – 50550	13	13
20500 – 40500	4	3	50550 – 100500	5	6
40500 – 60500	3	5	100500 – 600550	4	4
60500 – 653500	5	3	600550 – 858050	1	0
653500 – 1246500	1	2	858050 – 1115500	2	2
Totales	25	25	Totales	25	25
Promedio	60.672	116.208	Promedio	175.029	140.689

## 2.5. VARIABLES DEL MODELO

Al igual que la mayoría de los modelos empleados en la investigación estadística, el presente consta de variables dependientes e independientes. La

variable dependiente o criterio es una sola y es de carácter nominal, está constituida por una escala bipolar en la cual el código 1 representa empresas en crisis por haber ingresado a ronda reducida y el código 2 identifica empresas en ronda normal, todo esto conforme a las definiciones expresadas en el apartado 2.2.

Las variables independientes o predictoras consideradas, en principio fueron 12 indicadores, bajo la forma de ratios, calculados en base a la información contenida en los balances anuales que las sociedades anónimas presentan a la Bolsa de Comercio.

Conceptualmente, cada uno de los indicadores fue obtenido conforme a las siguientes definiciones:

#### 2.5.1. Situación Financiera a largo plazo

##### **1. Solvencia = Activo / Pasivo**

El índice de solvencia compara el activo con los fondos prestados por terceros para financiar los activos que la empresa necesita para su actividad. Cuando el índice es mayor que 1(unos) ello significa que hay activos por valores superiores al monto del pasivo y viceversa.

##### **2. Financiación propia del Activo = Patrimonio Neto/Activo**

Compara los fondos propios con el total de los recursos invertidos en la empresa. Esta ratio se vincula con el endeudamiento, con el cual mantiene una relación inversamente proporcional. El índice representa qué porcentaje del activo está financiado por recursos propios, es decir, si su valor es 0.50, ello indicaría que la mitad de los activos están financiados por recursos propios y la otra mitad por fondos de terceros.

##### **3. Costos Marg. de Financiamiento = Utilidad Neta/Intereses Pagados**

Debido a su alta performance en otros países, primordialmente en U.S.A. (Altman, 1989), se incluye este indicador a los efectos de medir su valor predictivo en el conjunto de empresas investigadas. Una relación baja acusa costos elevados en la contratación de fondos de terceros, lo que indicaría que la mala situación financiera



de la empresa la ha llevado a buscar fuentes de financiamiento fuera del sistema bancario, es decir, en mercados informales donde las tasas son más elevadas.

También se lo conoce como la cantidad de veces que se ganan los intereses, indica la capacidad de la empresa para generar utilidades, para hacer frente al costo de los pasivos y consecuentemente, el margen de seguridad de los acreedores con relación al cobro de los intereses.

#### **4. Inmovilización del activo = Activo No Corriente / Activo**

El coeficiente que se obtiene indica que proporción del activo está inmovilizado, es decir la importancia relativa de los activos fijos en el total del activo. Se supone, en una empresa en marcha, que los bienes que se destinan al cumplimiento de los fines de la misma seguirán haciéndolo indefinidamente. Para un acreedor, indica la dificultad de realización de los recursos invertidos en caso que deba recurrirse a ellos para recuperar la acreencia.

#### **5. Inmovilización del Patrimonio Neto = Activo No Corriente / Patrim. Neto**

Expresa la política de financiación de los activos no corrientes. Los fundamentos que sustentan la relación responden al criterio de que la empresa debe recurrir a fuentes de fondos que se ajusten a la naturaleza y duración de los bienes que van a ser financiados con ellos. Dado el reducido grado de liquidez de los activos no corrientes por la longitud del ciclo de vida que los caracteriza, es conveniente que sean financiados con fondos de baja exigibilidad, destacándose como más idónea fuente de fondos el patrimonio neto y las deudas a largo plazo.

#### **6. Autofinanciación = Utilidades no Distribuidas / Activo**

Es una ratio que pone de relieve la capacidad de la empresa para financiar parte de sus inversiones con los ahorros genuinos generados a partir del funcionamiento regular de sus actividades.

#### **7. Endeudamiento a largo plazo = Pasivo No corriente / Activo**

Indica la proporción del pasivo a largo plazo respecto del activo total y constituye una medida del endeudamiento a largo plazo Es un índice que informa sobre si el

activo total alcanza o no para cubrir las deudas a largo plazo.

#### 2.5.2. Situación Financiera a corto plazo

##### **8. Liquidez ácida = Activo corriente – Bienes de Cambio / Pasivo Corriente**

Por medio de este índice se compara el activo rápido con el pasivo corriente. El piso para este índice es 0,50, lo que surge de considerar que los bienes de cambio representan el 50 % del activo corriente. Mediante su empleo se procura resaltar a través de una medición cuantitativa una característica cualitativa, cual es la diferente liquidez de los diferentes activos corrientes.

##### **9. Movilidad del Activo = Capital de Trabajo/ Activo Total**

Este coeficiente expresa el grado de fluidez de los activos e indica en qué medida la inmovilización del activo total resulta neutralizada por el capital de trabajo, es decir, que proporción de capital de trabajo en el total de activo tiene la empresa. Una de las características de las empresas en crisis es tener muchos activos inmovilizados o de largo plazo y poco capital de trabajo, por lo cual no puede hacer frente a las deudas a corto plazo.

##### **10. Rotación del Activo = Ingresos Total/ Activo Total**

Es el coeficiente que mide la densidad del activo, esto es, cuantas veces el activo total gira en las ventas totales. Se pretende evaluar si el producido de las ventas de la empresa cubre el total del Activo.

#### 2.5.3. Situación Económica

##### **11. Rentabilidad Económica = Utilidad antes de impuesto / Activo Total**

El índice de rentabilidad económica es una medida del resultado del negocio, de la administración de todos los recursos confiados a la administración, con total prescindencia de si los fondos fueron aportados por los propietarios o por terceros.

##### **12. Margen Bruto = Utilidad Bruta / Ventas**

A través de este índice se cuantifica la proporción que representa la utilidad bruta

con respecto a las ventas brutas totales, sin deducciones ni en el numerador ni en el denominador.

En este sentido, los ratios utilizados a modo de predictores se consignan en el cuadro 2.4.

Cuadro 2.4. Variables independientes

Variable	Indicador económico-financiero
X <sub>1</sub>	Solvencia
X <sub>2</sub>	Propiedad del Activo
X <sub>3</sub>	Utilidad Neta/Intereses
X <sub>4</sub>	Inmovilización del Activo
X <sub>5</sub>	Inmovilidad del Patrimonio Neto
X <sub>6</sub>	Autofinanciación
X <sub>7</sub>	Pasivo No Cte/Activo
X <sub>8</sub>	Liquidez Acida
X <sub>9</sub>	Movilidad del Activo
X <sub>10</sub>	Rotación del Activo
X <sub>11</sub>	Rentabilidad Económica
X <sub>12</sub>	Margen Bruto

## 2.6. PROCESAMIENTO

### 2.6.1. Supuestos

En primer lugar se probaron los supuestos que deberían cumplirse para la aplicación de métodos paramétricos, según lo manifestado por Johnson, R y Wichern, D (1992):

- a) Igualdad de matrices de covarianzas
- b) Normalidad multivariada

Respecto al primer supuesto, se realizó la prueba de homogeneidad de matrices de covarianzas, con el test Box's M, citado por Díaz, M (2001), utilizando el paquete estadístico SPSS donde la hipótesis nula establece que las matrices de covarianzas son iguales.

Cuadro 2.5. Test Box's M de Igualdad de Matrices de Covarianzas

<b>Box's M</b>		218,589
<b>F</b>	Approx.	4,983
	df1	36
	df2	7752,623
	Sig.	,000

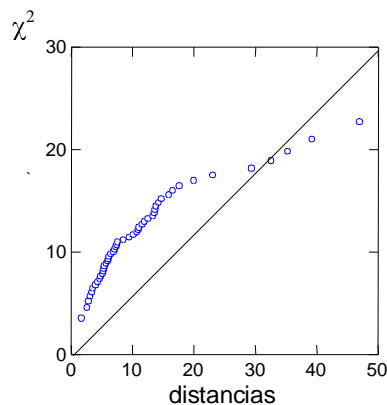
Con lo cual se rechaza la hipótesis nula, si bien no es necesario este supuesto para aplicar discriminante cuadrático, es necesario cumplir con el segundo supuesto, de la distribución normal multivariada.

Para este supuesto, se realizó el gráfico de probabilidad “chi cuadrado” (Q-Q plot). Para la construcción del mismo, se ordenaron las distancias al cuadrado ( $d_j^2$ ), de menor a mayor.

$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \quad \text{para } j= 1,2,\dots,n$$

y se graficaron los pares  $(d_j^2; \chi_p^2((j-1/2)/n))$ , siendo la segunda componente el percentil de la distribución chi cuadrado con  $p$  grados de libertad (Figura 2.1.).

Figura 2.1. Chi cuadrado



Dicho gráfico nos muestra que los puntos no están alineados sobre la línea de 45°. Hasta un valor de  $d_j^2$  cercano a 30 están agrupadas la mayor cantidad de las observaciones y por arriba de dicha línea. En general, puede afirmarse que las variables no se comportan según una distribución normal multivariada.

La normalidad multivariante (la combinación de dos o más variables) implica que las variables individuales sean normales en un sentido univariado y que sus combinaciones también sean normales. Por lo tanto si una variable es normal multivariante, también es normal univariante. Sin embargo, lo contrario no es necesariamente cierto (dos o más variables normales univariantes no son necesariamente normal multivariante). Por tanto, una situación en la que todas las variables exhiben normalidad univariante ayudará a obtener normalidad multivariante, aunque no lo garantiza. Por otro lado, si las variables normales

univariantes no son normales, es una prueba que al considerar todas las variables simultáneamente, no tendrán distribución normal multivariante.

Por ello, se evaluará la normalidad de cada una de las variables incluidas en el análisis (Figura 2.2.).

Salvo las variables “PROPACT” (Propiedad del Activo) e “INMACT” (inmovilización del Activo) que presentan un comportamiento aproximadamente normal, las restantes variables no tienen esta distribución, por lo cual podemos confirmar que las variables en su conjunto no poseen el supuesto de la distribución para aplicar un método de clasificación paramétrico.

Debido a que ambos supuestos no se cumplen, se procedió a aplicar métodos no paramétricos.

Figura 2.2. Gráficos Q – Q

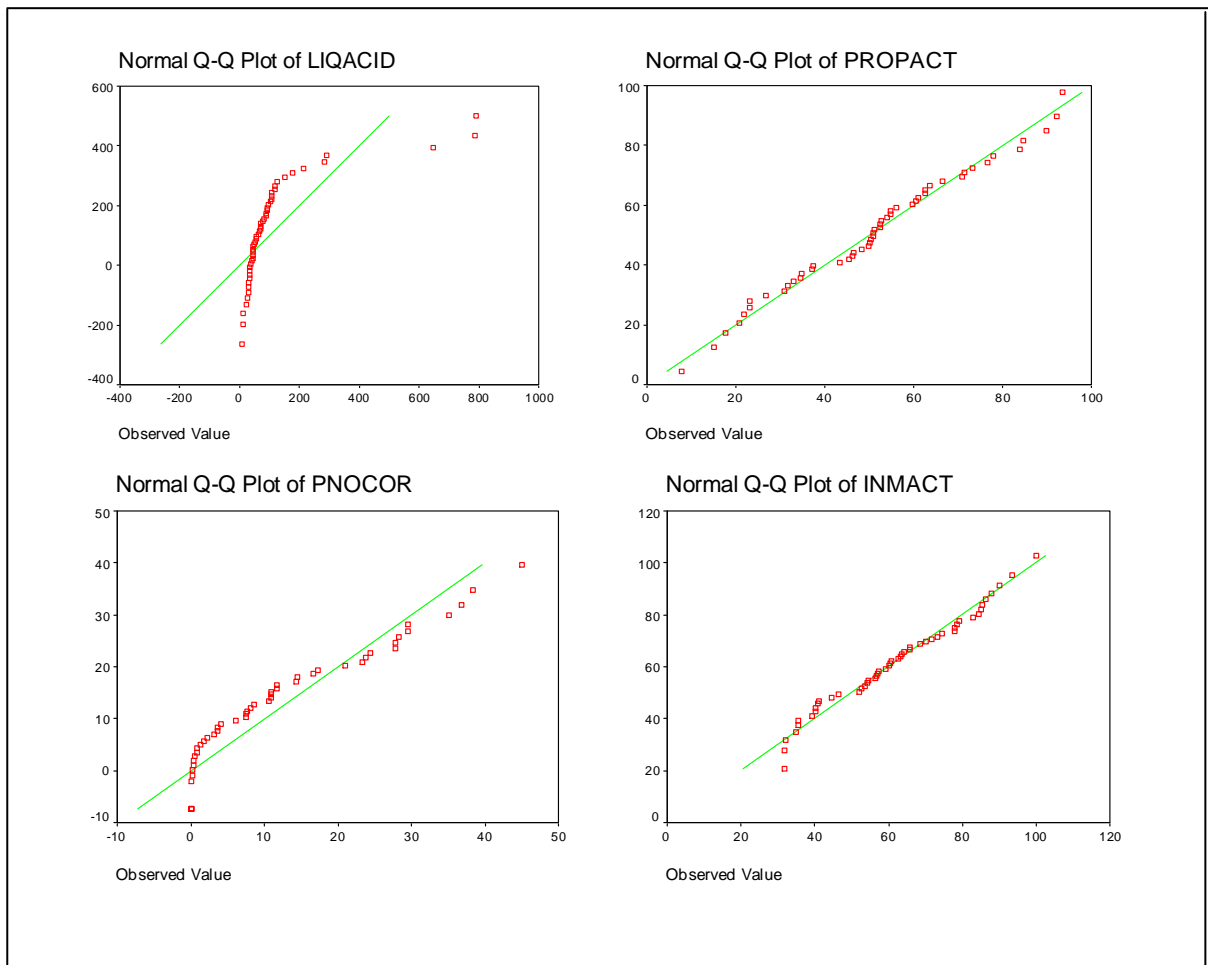
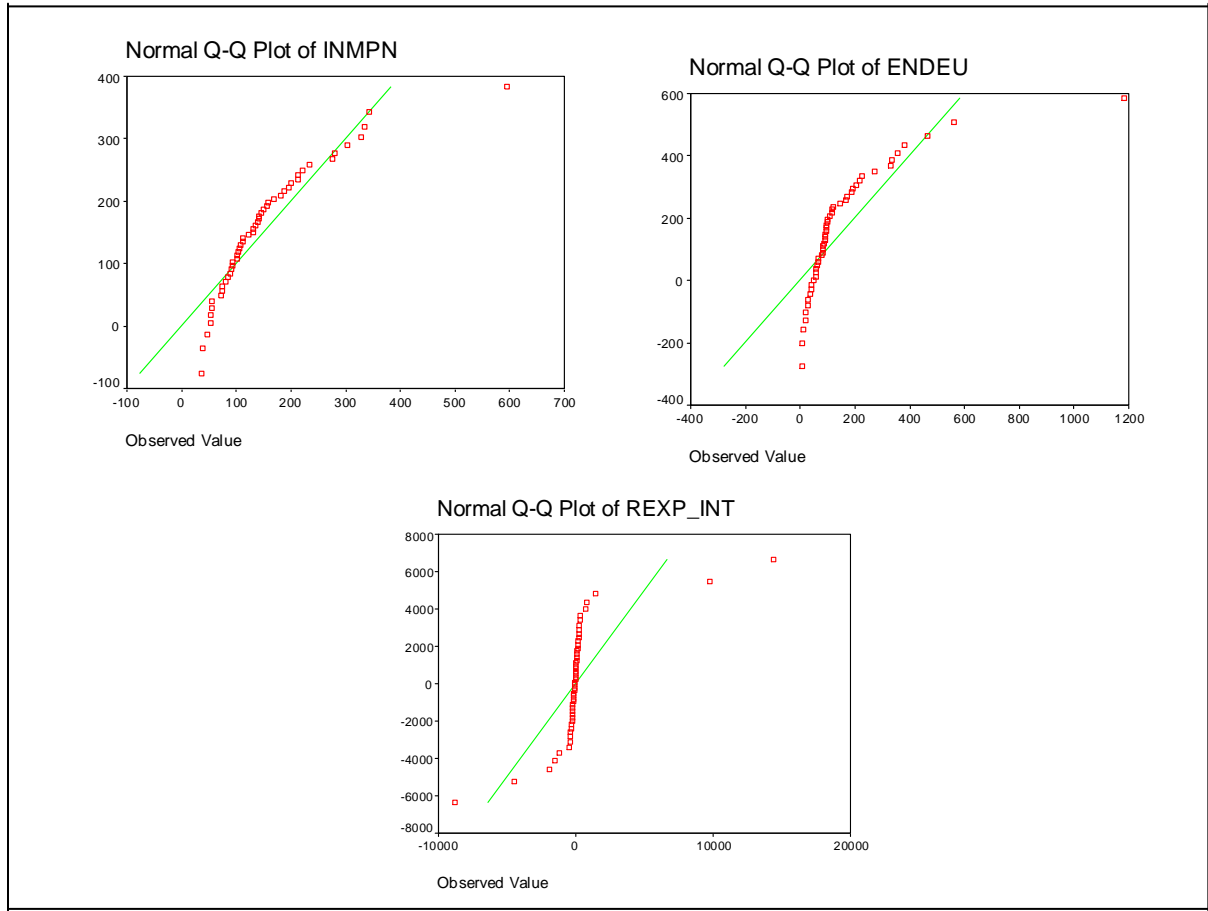


Figura 2.2. (Continuación) Gráficos Q – Q



### 2.6.2. Análisis descriptivo multivariado

Con el paquete estadístico Infostat se realizó el cálculo de algunos estadísticos descriptivos multivariados, entre ellos:

- a) El vector de medias total: por cada variable de la matriz de datos se calcula la media muestral. En este caso se cuenta con 12 medias muestrales ( $p= 12$ ), cada una obtenida como:

$$\bar{x}_j = \sum_{i=1}^n \frac{x_{ij}}{n} \quad \text{para } j=1,2,\dots,p$$

Si no consideramos el criterio de clasificación, el vector de medias total está conformado con las medias de las  $p$  variables calculadas a partir de las observaciones, lo cual se muestra en la segunda columna del cuadro 2.6.

Si se considera el criterio de clasificación se podrán obtener las medias de cada grupo, según se muestra en las columnas 3 (empresas en crisis) y 4 (empresas sin dificultades) del cuadro 2.6

Cuadro 2.6. Vector de medias total y por grupos

Variables	Medias	Medias del grupo 1	Medias del Grupo 2
LIQACID	118.18	89.69	146.68
SOLVENC	293.52	214.28	372.76
PROPACT	51.14	42.5	59.78
PNOCOR	11.80	12.19	11.41
AUTOFIN	-2.44	-9.05	4.18
INMACT	61.61	65.31	57.91
INMPN	153.25	189.97	116.52
RENTECO	-4.30	-8.85	0.24
MAREXP	-5.57	-12.09	0.95
CTRAACT	2.69	-7.65	13.03
REXPINT	141.20	-149.77	432.17
ROTACT	0.79	0.65	0.92

De donde puede observarse que las empresas que están en crisis poseen liquidez, solvencia y rotación del activo, en promedio, menor que en las empresas con comportamiento normal. Es decir, la Liquidez promedio de las empresas en crisis (89,69) es menor a la de las empresas sanas (146,68) lo cual refleja que las primeras poseen, en términos relativos, menos activos rápidos para cancelar el pasivo corriente. En cuanto a la Solvencia, en las empresas con falencia (214,28) el total del activo necesario para cubrir el pasivo es menor que en las empresas sanas (372,76).

Por último, respecto a la Rotación del Activo, en las empresas en crisis (0,65), el producido de las ventas no cubre el total del activo de la misma forma que en las empresas sanas (0,92).

Respecto a la rentabilidad y a los márgenes de utilidad, estos son negativos en un gran número de ellas, por lo cual su promedio es menor a cero. En las empresas en crisis una rentabilidad negativa (-8,85) indica la existencia de quebrantos en el ejercicio contable y que los márgenes también lo sean (-12,09), refleja la existencia de quebrantos brutos como margen respecto del nivel de ventas.

En cuanto a la inmovilización del activo y del patrimonio neto, en las empresas en crisis (65,31 y 189,97), estas son mayores que en el grupo de empresas que no lo están (57,91 y 116,52), lo cual significa que los activos no corrientes (inmovilizados)

son mayores que los activos totales ó que el patrimonio neto, lo cual es un indicador que suele estar presente en las empresas en crisis, ya que al tener activos inmovilizados, comienzan a tener dificultades en cuanto al cumplimiento de los compromisos con el resto de los activos.

Las empresas en crisis no se autofinancian (-9,05) y sus pasivos no corrientes en promedio son mayores que las empresas del segundo grupo.

b) Matriz de varianza-covarianza: la varianza muestral, calculada a partir de las  $n$  mediciones sobre cada variable, será denotada por  $S_j^2$ . Para una matriz de datos de dimensión  $n \times p$  habrá  $p$  varianzas muestrales, cada una obtenida a partir de la expresión:

$$S_j^2 = c \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad \text{para } j=1, \dots, p$$

donde la constante  $c$  es  $1/(n-1)$ .

La covarianza muestral mide la asociación lineal entre dos variables, es decir mide cómo varían dos variables conjuntamente. La Covarianza entre la variable  $j$ -ésima y la variable  $k$ -ésima es obtenida por:

$$S_{jk} = c \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad \text{para } j, k = 1, \dots, p$$

Esta matriz posee  $p$  varianzas y  $[p(p-1)]/2$  covarianzas, lo cual se refleja en el Cuadro 2.7.



Cuadro 2.7. Matriz de covarianza común (insesgada)

	Liq. Acid.	Solv.	Prop. Activ	Pas. Noct e/Activ	Auto Fin	Inmv Act.	Inmv. PN	Rent Econ	Marg. Brut..	Ctrab /Acti	Ut. neta/ Inter.	Rot. Activ.
Liq. acid	29218	44832	2268	-344	674	-703	-6541	239	667	2524	78509	-12.6
Solve ncia		79418	4409	-1112	1184	-1275	-13027	528	797	4459	32652	-6.63
Prop Act			434	-96	1130	-44	-1710	94	118	383	10378	0.09
Pa nc/ act.				154	-75	120	463	-4	14	-73	-3455	3.4
Auto fin					517	-45	-559	160	158	121	21572	2.46
Inm Act						334	565	-14	-93	-198	-465	-4.6
Inm PN							10460	-494	672	1742	28372	-7.1
Rent econ								155	148	104	23384	0.01
Marg Bruto									504	217	13921	2.4
Cap tra/Ac										528	6751	1.01
Ut. neta/ Inter.											8408781	-91
Rot. Activ												0.36

c) Matriz de correlación total: es una matriz de orden  $p$  que posee el valor 1 sobre la diagonal principal y los coeficientes de correlación de Pearson entre cada par de variables son los restantes elementos. El coeficiente de correlación de Pearson es una medida de la magnitud de la asociación lineal entre dos variables que no depende de las unidades de medida de las variables originales. Para las variables  $j$ -ésima y  $k$ -ésima se define como:

$$r_{jk} = \frac{S_{jk}}{\sqrt{S_j^2 S_k^2}}$$

El coeficiente de correlación muestral representa la covarianza de los valores muestrales estandarizados.

Al considerar las relaciones de interdependencia entre los predictores se pueden utilizar dos tipos de matrices: la matriz de correlación total y la combinada (pooled matrix), esta última obtenida promediando las correlaciones calculadas dentro de

cada uno de los dos grupos (en crisis y sin dificultades). En el Cuadro 2.8 se consignan las correlaciones promedios y sus significaciones estadísticas dentro de los grupos (pooled matrix)

Cuadro 2.8. Matriz de Correlaciones de Predictores

	Liq. acid.	Solv.	Prop. Activ	Pas. Nocte/Activ	Auto fin	Inmv Act.	Inmv. PN	Rent Econ	Marg. Brut..	Ctrab /Acti	Ut. neta/ Inter.	Rot. Activ .
Liq. Acid		.93	.63	-.16	.17	-.23	-.37	.11	.17	.64	-.16	-.12
Solven cia			.75	-.32	.18	-.25	-.45	.15	.13	.69	-.04	-.04
Prop Act				-.37	.24	-.11	-.80	.36	.25	.80	.17	-.01
Pa nc/ act.					-.26	.53	.36	-.03	.05	-.25	-.10	-.46
Auto Fin						-.11	-.24	.56	.31	.23	.33	.18
Inm Act							.30	-.06	-.23	-.47	-.01	-.42
Inm PN								-.39	-.29	-.74	-.10	-.12
Rent Econ									.53	.36	.65	-.000
Marg Bruto										.48	.21	.18
Cap tra/Ac											.10	.07
Ut. neta/ Inter.												-.05
Rot. Activ												

A partir del cuadro anterior se ordenaron las principales correlaciones, las cuales prescindiendo del signo varían desde .930 hasta .530. A tal efecto, el Cuadro 2.9 refleja el ordenamiento de las correlaciones en el conjunto de las doce variables independientes consideradas. Cabe destacar que si bien algunos valores son obvios, en razón de su propia constitución, los restantes reflejan los resultados finales del conjunto de balances concretamente investigado.

Cuadro 2.9. Principales Correlaciones

Binomio de Variables	Coef.
1)Solvencia – Liquidez Acida	.930
2)Propiedad del Activo – Inmovilización del PN	-.800
3)Propiedad del Activo – Movilidad del Activo	.800
4)Propiedad del Activo – Solvencia	.750
5)Cap. Trabajo – Inmovilización del PN	-.740
6) Rentabilidad Económica – Autofinanciación	.560
7) Pasivo No Cte/Activo – Inmovilización del Act	.530
8)Rentabilidad Económica- Utili. Neta/ Intereses	.530

La discusión de los signos que preceden los coeficientes del Cuadro 2.9 constituye sin lugar a dudas una etapa importante en la convalidación de la consistencia interna de las cifras de los balances y la pertinencia y razonabilidad de las fórmulas utilizadas en el cálculo de los diferentes indicadores. A tal fin, es conveniente desglosar las fórmulas empleadas y expresar cada uno de los indicadores en función de sus pares constitutivos.

1) Par SOLVENCIA – LIQUIDEZ ACIDA.

Poseen el mismo signo, ya que su estructura de composición es semejante relaciona el activo total (en caso de solvencia) o partes componentes del mismo, con el Pasivo total o el pasivo corriente.

2) Par PROPIEDAD DEL ACTIVO – INMOVILIZACION DEL PATRIMONIO NETO.

Existe una realación inversa entre ambos indicadores:

$$\begin{aligned} \text{PROPACT} &= \text{PN} / \text{ACTIVO} \\ (\text{PROPACT})^{-1} &= \text{ACTIVO} / \text{PN} \\ &= (\text{ACTIVO NO CTE} + \text{ACT.CTE}) / \text{PN} \\ &= \text{INMOV. PN} + (\text{ACT. CTE} / \text{PN}) \end{aligned}$$

A mayor nivel de Inmovilización del Patrimonio Neto, menor valor para el indicador que mide la Propiedad del Activo, justificando el signo negativo.

3) Par PROPIEDAD DEL ACTIVO – CAPITAL DE TRABAJO

$$\begin{aligned} \text{PROPACT} &= \text{PN} / \text{ACTIVO} \\ \text{MOVILIDAD DEL ACTIVO} &= \text{CAPITAL DE TRABAJO} / \text{ACTIVO} \\ \text{PROPACT} &= [\text{CAPITAL DE TRABAJO} + (\text{ACTCTE} - \text{PASCTE})] / \text{ACTIVO} \\ &= \text{MOVILIDAD DEL ACTIVO} + (\text{ACTCTE} - \text{PASCTE}) / \text{ACTIVO} \end{aligned}$$

Siendo la movilidad del activo uno de los componentes de la propiedad del activo, ante un incremento en uno de ellos, incrementa el otro, es decir que ambos índices

miden una relación respecto del Activo, por ello su signo positivo.

4) Par PROPIEDAD DEL ACTIVO – SOLVENCIA

La relación entre ambos indicadores es positiva:

$$\begin{aligned} \text{PROP. ACT} &= \text{PN} / \text{ACTIVO} \\ \text{SOLVENCIA} &= \text{ACTIVO} / \text{PASIVO} \\ \text{PROPACT} &= (\text{PN} / \text{ACTIVO}) = (\text{ACT} - \text{PAS}) / \text{ACTIVO} \\ &= 1 - (\text{SOLVEN.})^{-1} \end{aligned}$$

Si Solvencia aumenta, la Propiedad del Activo también crece; por lo tanto, el signo es positivo.

5) Par CAPITAL DE TRABAJO – INMOVILIDAD DEL PATRIMONIO NETO

Analizando la estructura de cada uno de estos índices:

$$\begin{aligned} \text{CAPTRAB} &= (\text{ACTCE} - \text{PASCTE}) / \text{ACTIVO} \\ \text{INMOV. PN} &= \text{ACT NOCTE} / \text{PN} \\ (\text{CAPTRAB})^{-1} &= (\text{ACT} / (\text{ACTCTE} - \text{PASCTE})) = (\text{PAS} + \text{PN} / \text{ACTCTE} - \text{PASCTE}) \\ &= \text{PN} (\text{PAS} / \text{PN} + 1) / (\text{ACT} - \text{ACTNOCTE} - \text{PASCTE}) \\ &= ((\text{PAS} / \text{PN}) + 1) / ((\text{ACT} / \text{PN}) - (\text{ACTNOCTE} / \text{PN}) - (\text{PASCTE} / \text{PN})) \\ &= (\text{ENDEUDAMIENTO} + 1) / ((\text{PROPACT})^{-1} - \text{INMPN} - (\text{PASCTE} / \text{PN})) \end{aligned}$$

Si la Inmovilidad del Patrimonio Neto aumenta, crece también (Captrab)<sup>-1</sup> y por consiguiente decrece Capital de Trabajo, el signo es por consiguiente negativo.

6) Par RENTABILIDAD ECONÓMICA – AUTOFINANCIACIÓN.

$$\begin{aligned} \text{RENTABECON} &= \text{UTILID} / \text{ACTIVO} = \text{UTILID DISTRIB} + \text{UTILID NO DISTRIB} / \text{ACT} \\ &= (\text{UTILID DISTRIB} / \text{ACT}) + (\text{UTILID NO DISTRIB} / \text{ACT}) \\ \text{RENTAB.ECON} &= \text{UTIL.DISTRIB}/\text{ACT} + \text{AUTOFINANCIAMIENTO} \end{aligned}$$

Por tal razón, el coeficiente de correlación es positivo.

7) Par PASNOCTE / ACTIVO – INMOVILIZACION DEL ACTIVO

Desarrollando cada uno de los índices y viendo cual es su relación:

$$\begin{aligned} &\text{PASNOCTE} / \text{ACTIVO} \\ \text{INMOVACT} &= \text{ACTNOCTE} / \text{ACTIVO} \\ &= (\text{ACT} - \text{ACTCTE}) / \text{ACT.} \\ &= (\text{PASCTE} + \text{PASNOCTE} + \text{PN} - \text{ACTCTE}) / \text{ACT} \\ &= (\text{PASCTE} / \text{ACTIVO}) + (\text{PASNOCTE} / \text{ACTIVO}) + \text{PROPACT} - \text{ACTCTE} / \text{ACT} \end{aligned}$$

Luego, Inmovilización del Activo tiene el mismo signo que Pasivo No corriente/Activo y por lo tanto el coeficiente de correlación es positivo

8) Par RENTABILIDAD ECONOMICA – UTIL.NETA / INTERESES

Considerando los componentes de ambos índices:

$$\text{RENTAB.ECON} = \frac{\text{UTILIDAD (antes de impuestos)}}{\text{ACTIVO}} \\ \frac{\text{UTILNETA}}{\text{INTERESES}}$$

Ambos índices dependen directamente de la utilidad Bruta y tienen, obviamente, el signo positivo.

### 2.6.3. Definición de modelos

Trabajar con todas las variables (Modelo A) es una de las posibilidades. Otra es trabajar con 10 variables, luego de eliminar Propiedad del Activo y Solvencia (Modelo B), que se seleccionaron del total de 12 variables por la alta correlación que presentan con otros indicadores.

Por último, otro modelo alternativo es el que considera ocho variables (Cuadro 2.10), las mismas resultan de un proceso de selección de variables paso a paso, en el que se fueron removiendo variables (Modelo C).

Cuadro 2.10 Variables independientes (Modelo C)

Variable	Indicador económico-financiero
X <sub>1</sub>	Liquidez ácida
X <sub>2</sub>	Propiedad del Activo
X <sub>3</sub>	Pasivo No Corriente/Activo
X <sub>4</sub>	Autofinanciación
X <sub>5</sub>	Inmovilización del Activo
X <sub>6</sub>	Inmovilidad del Patrimonio Neto
X <sub>7</sub>	Utilidad sobre Interese
X <sub>8</sub>	Endeudamiento

Se aplicaron metodologías no paramétricas a estos modelos.

### 2.6.4. Aplicación de metodologías no paramétricas a los datos

#### A) Kernel y vecino más cercano

Para estimar las densidades de grupos específicos y producir una regla de clasificación, mediante el método de Kernel o el vecino más cercano, el procesamiento fue realizado con el paquete estadístico SAS.

La distancia al cuadrado entre dos vectores de observaciones  $\mathbf{x}$  e  $\mathbf{y}$ , en el grupo  $i$  es tomado como:

$$d_i^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})' \mathbf{V}_i^{-1} (\mathbf{x} - \mathbf{y})$$

donde  $\mathbf{V}_i$ , puede tomar diferentes formas. En este caso se trabajó con la matriz pooled.

La clasificación de un vector de observaciones  $\mathbf{x}$  se basa en la estimación de la densidad del grupo especificado para el conjunto que se analiza. Se evalúan las probabilidades posteriores, es decir una observación  $\mathbf{x}$  es clasificada en el grupo  $u$ , si  $i = u$  produce un valor de  $p(i/\mathbf{x})$ . Si esta probabilidad es menor que el valor especificado,  $\mathbf{x}$  está clasificado en otro grupo.

El método Kernel usa un radio fijo y un Kernel especificado a fin de estimar la densidad del grupo  $i$  para cada observación  $\mathbf{x}$ . Si  $\mathbf{z}$  es un vector  $p$ -dimensional, entonces el volumen de la esfera unitaria  $p$ -dimensional, donde  $\mathbf{z}'\mathbf{z} = 1$ , es:

$$V_0 = \frac{\pi^{p/2}}{\Gamma(p/2 + 1)}$$

Por ello, en el grupo  $i$ , el volumen de la elipsoide  $p$ -dimensional para  $\{\mathbf{z} / \mathbf{z}' \mathbf{V}_i^{-1} \mathbf{z} = r^2\}$  es:

$$V_r(i) = r^p |\mathbf{V}_i|^{1/2} V_0$$

Se usan diferentes tipos de Kernel para estimar la densidad, entre ellos se trabajó con Kernel uniforme, normal y Epanechnikov. En relación con el criterio de proximidad se puede trabajar tanto con la distancia de Mahalanobis como la Euclidiana.

Dentro de los diferentes Kernel, con los que se puede trabajar, el  $r$  debe ser elegido de manera que optimice el criterio tomado. Uno de ellos es el que minimiza el MISE aproximado de la densidad estimada. El valor óptimo, entonces de  $r$  depende de la función del Kernel.

El valor óptimo de  $r$  está dado por:

$$[A(K_i) / n_i]^{1/(p+4)}$$

donde la constante optima  $A(K_i)$  depende del kernel (Epanechnikov, 1969).

Por ejemplo, para:

$$\text{Kernel uniforme: } A(K_i) = \frac{2^{p+1}(p+2)\Gamma(p/2)}{p}$$

$$\text{Kernel normal: } A(K_i) = \frac{4}{2p+1}$$

$$\text{Kernel Epanechnikov: } A(K_i) = \frac{2^{p+2} p(p+2)(p+4)\Gamma(p/2)}{2p+1}$$

El método del Vecino más cercano fija el número  $k$  de un conjunto de puntos para cada observación  $\mathbf{x}$ . Este método encuentra el radio  $r_k(\mathbf{x})$  que es la distancia desde  $\mathbf{x}$  al  $k$ -ésimo conjunto de puntos cercanos en la métrica  $V_i^{-1}$ . Este  $K$  es usualmente relativo, una aproximación práctica es probar diferentes valores de suavizado dentro del contexto de una aplicación particular y elegir uno que tome la mejor tasa crossvalidada del error.

El cuadro 2.11 muestra el resumen del procesamiento de los diferentes modelos con ambos métodos.

Cuadro 2.11 Resumen del procesamiento con SAS

Cantidad de variables	Método	R	Tasa crossvalidada
12	Kernel Normal	0.5318	0.18
	K. Epanechnikov	10.0182	0.16
	Kernel Uniforme	3.826	0.30
12	Vecino mas cercano	K = 1	0.18
		K = 2	0.06
10	Kernel Normal	0.4982	0.26
	K. Epanechnikov	8.4771	0.32
	Kernel Uniforme	3.033	0.38
10	Vecino mas cercano	K = 1	0.26
		K = 2	0.12
8	Kernel Normal	0.4595	0.22
	<b>K. Epanechnikov</b>	<b>6.9327</b>	<b>0.16</b>
	Kernel Uniforme	2.3533	0.36
8	<b>Vecino mas cercano</b>	K = 1	0.24
		<b>K = 2</b>	<b>0.10</b>

Se descarta la posibilidad de considerar el modelo de 12 variables, debido a la conveniencia de seleccionar aquel modelo que tenga menos cantidad de variables, a fin de hacerlo comparable con otros métodos de clasificación.

Del análisis de la tabla se observa que el modelo con 8 variables es el que ofrece a través de los diferentes métodos, la mejor tasa crossvalidada, con un Kernel Epanechnikov, la tasa de clasificar mal es de 0.16 (Situación 1) y con el método del vecino más cercano con  $K = 2$ , es de 0.10 (Situación 2).

Considerando ambas situaciones, puede observarse que:

Para la Situación 1, los cuadros 2.12, 2.13 y 2.14 representan las salidas que indican: Cuál es la probabilidad a posteriori de pertenecer a un grupo o a otro, la tasa de clasificación errónea y cuáles fueron las empresas mal clasificadas.

PROCESAMIENTO CON SAS (KERNEL EPANECHNIKOV)

Cuadro 2.12 Probabilidades a posteriori

Posterior Probability of Membership in CONDICIO Classified					
Obs	CONDICIO	CONDICIO	From 1	into 2	
7	1	2 *	0.4846	0.5154	
13	1	2 *	0.4919	0.5081	
25	1	2 *	0.4936	0.5064	
27	1	2 *	0.4578	0.5422	
33	1	2 *	0.4656	0.5344	
43	1	2 *	0.4707	0.5293	
44	2	1 *	0.5256	0.4744	
48	2	1 *	0.5048	0.4952	

\* Misclassified observation

Cuadro 2.13. Tasa de error crossvalidada

Error Count Estimates for CONDICIO			
	1	2	Total
Rate	0.2400	0.0800	<b>0.1600</b>
Priors	0.5000	0.5000	

Cuadro 2.14. Empresas mal clasificadas

From CONDICIO	1	2
1	19 0.5724	6 0.5226
2	2 0.5152	23 0.5481
Total	21 0.5669	29 0.5428



Del total de empresas, el 16 % se clasifica erróneamente. En este caso 6 empresas en crisis se clasifican como buenas y 2 empresas sanas como en crisis.

En la situación 2, con el método del vecino más cercano y con tasa crossvalidada, se obtienen los resultados de los cuadros 2.15, 2.16 y 2.17:

PROCESAMIENTO CON SAS (VECINO MAS CERCANO, K = 2)

Cuadro 2.15. Probabilidades a posteriori

Posterior Probability of Membership in CONDICIO Classified				
Obs	CONDICIO	CONDICIO	From 1	into 2
7	1	2 *	0.0000	1.0000
27	1	2 *	0.0000	1.0000
43	1	2 *	0.0000	1.0000
44	2	1 *	1.0000	0.0000
48	2	1 *	1.0000	0.0000

\* Misclassified observation

Cuadro 2.16. Tasa de error crossvalidada

Error Count Estimates for CONDICIO			
	1	2	Total
Rate	0.1200	0.0800	<b>0.1000</b>
Priors	0.5000	0.5000	

Cuadro 2.17. Empresas mal clasificadas

From CONDICIO	1	2
1	22 0.8219	3 1.0000
2	2 1.0000	23 0.7657
Total	24 0.8367	26 0.7927

Del análisis de estos cuadros, se desprende sólo 5 empresas fueron mal clasificadas, tres que estando en crisis se clasificaron como sanas y dos en otro sentido. Por ello la tasa de mal clasificación es de 0,10.

#### b) Árbol de clasificación

En la aplicación del método del árbol de clasificación, los datos se procesaron con Infostat y SPSS. Los modelos basados en árboles constituyen una alternativa a los modelos lineales aditivos para los problemas de clasificación, ya que se aplican cuando no es posible detectar interacciones entre variables.

El árbol de clasificación, en su versión CART es un conjunto de muchas reglas determinadas por un procedimiento de ajuste por particiones binarias recursivas, donde un conjunto de datos es sucesivamente particionado. El grupo se separa en dos subgrupos a partir de una de las variables regresoras de manera tal que la heterogeneidad a nivel de la variable dependiente, sea mínima de acuerdo a la medida de heterogeneidad seleccionada. Los dos subgrupos formados se separan nuevamente si hay suficiente heterogeneidad y el tamaño del nodo es superior al mínimo establecido para continuar el algoritmo. El proceso se detiene cuando no se cumple una de estas condiciones. En cada instancia de separación el algoritmo analiza todas las variables regresoras y selecciona, para realizar la partición, aquella que permite conformar grupos homogéneos.

El paquete Infostat provee dos medidas de heterogeneidad dentro de nodos, la Deviance y la Suma de cuadrados. Debido a que la variable dependiente es una variable de clasificación se utilizó la primera.

La expresión de la Deviance calculada es:

$$D(\mu_i; y_i) = -2 \sum_k n_{ik} \log(p_{ik})$$

También se determina un mínimo tamaño de nodo para continuar la partición y el umbral de heterogeneidad dentro del grupo para terminar.

Se ingresan las variables independientes ordenadas, tal cual surge de pruebas F de análisis de la varianza, y así ordenados los estadísticos de mayor a menor, es como se introducen dichas variables.

La salida de Infostat, en el cuadro 2.18 muestra como han sido particionadas las variables y la formación de los nodos. Las variables que han incidido en la clasificación son: \* Costo marginal de financiamiento

\* Inmovilidad del Activo

\* Inmovilidad del Patrimonio Neto.

La figura 2.3. (salida de SPSS) esquematiza el árbol de clasificación.

Cuadro 2.18 Árbol de Clasificación

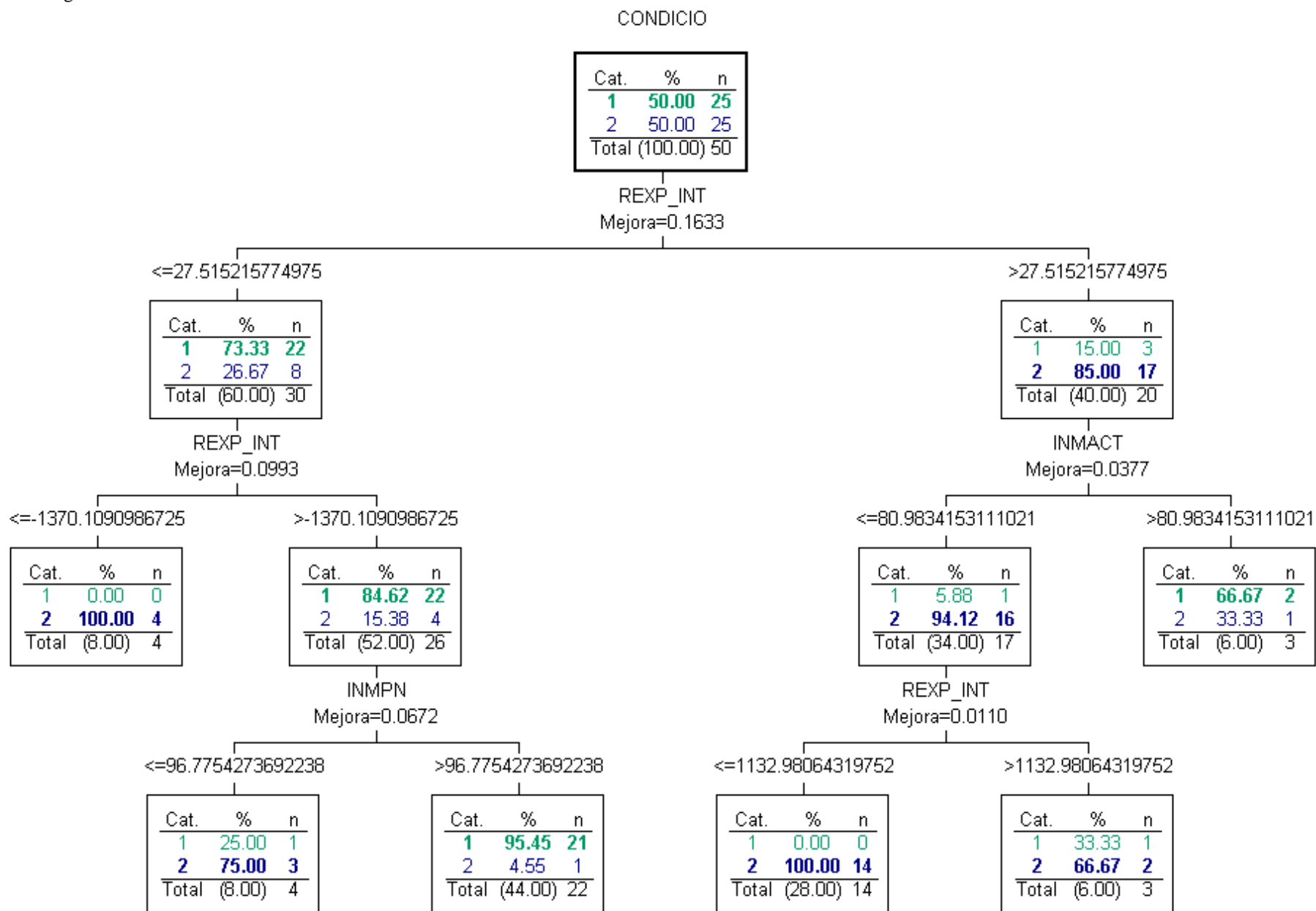
Árboles de clasificación/regresión									
H= Deviance (suma (ni*ln(pi))									
Nodo	Formación	H	Predicción	n	Media	Varianza	Mínimo	Máximo	
Raíz				1.50	50	1.50	0.26	1.00	2.00
1	REXP_INT(<=27.52)	34.79	1.27	30	1.27	0.20	1.00	2.00	
1.1	REXP_INT(<=-1370.11)	0.00	2.00	4	2.00	0.00	2.00	2.00	
1.2	REXP_INT(>-1370.11)	22.32	1.15	26	1.15	0.14	1.00	2.00	
1.2.1	INMPN(<=96.78)	4.50	1.75	4	1.75	0.25	1.00	2.00	
1.2.2	INMPN(>96.78)	8.14	1.05	22	1.05	0.05	1.00	2.00	
2	REXP_INT(>27.52)	16.91	1.85	20	1.85	0.13	1.00	2.00	
2.1	INMACT(<=80.98)	7.61	1.94	17	1.94	0.06	1.00	2.00	
2.2	INMACT(>80.98)	3.82	1.33	3	1.33	0.33	1.00	2.00	

La primera variable que clasifica es la que mide el costo marginal de financiamiento, es decir, el cociente entre la Utilidad Neta y los Intereses pagados. Una relación baja acusa costos elevados en la contratación de fondos de terceros, lo que indicaría que la mala situación financiera de la empresa la ha llevado a buscar fuentes de financiamiento fuera del sistema bancario, es decir, en mercados informales donde las tasas son más elevadas.

En el primer nivel de clasificación se distinguen dos grupos:

- a) 30 empresas con los niveles más bajos de esta variable (menores a 27,51), donde el 73,33 % de este total son empresas de la categoría 1, o sea de 25 empresas en crisis, 22 poseen niveles bajos de esta variable. En otras palabras son empresas que tienen altos costos en la contratación de fondos de terceros y poca capacidad para generar utilidades.

Figura 2.3. Árbol de Clasificación



- b) 20 empresas con niveles superiores de esta variable (mayores a 27,51) de las cuales el 85 % son empresas sin dificultades, las cuales han podido generar utilidades para hacer frente a sus obligaciones, si es que las tienen.

En el segundo nivel de clasificación, por un lado las empresas del grupo a) se clasifican según la misma variable, el costo de financiamiento, en:

- c) 4 empresas con niveles muy bajos de esta variable (menores a -1370,10), lo cual puede significar altos intereses a pagar a los acreedores, es decir, que aunque sean empresas sin dificultades, deben enfrentar altos costos. Se está ante la presencia de un nodo terminal, cuyas empresas se asignan al **grupo 2**.
- d) 26 empresas con niveles de costo comprendidos entre -1370,10 y 27,51, de las cuales el 84,62 % pertenecen al grupo de empresas en crisis, lo cual implica que también tienen costos de altos intereses.

Las empresas del grupo b) se clasifican, según la variable inmovilidad del activo, a su vez en:

- e) 17 empresas con menores niveles de inmovilización, de las cuales 16 (94,12 %) son empresas sin dificultades.
- f) 3 empresas con mayores niveles de inmovilización de activos, donde 2 de ellas son empresas en crisis. Este constituye un nodo terminal, son empresas que pueden asignarse al **grupo 1**.

Siguiendo con esta clasificación, las empresas del grupo d) vuelven a clasificarse según la variable inmovilización del Patrimonio Neto, en los siguientes grupos:

- g) 4 empresas poseen menores niveles de inmovilización del patrimonio neto, de las cuales 3 son empresas normales (75 %). Este constituye un nodo terminal de empresas, que se puede asignar al **grupo 2**.

- h) 22 empresas con mayores nivel de inmovilización del patrimonio neto, de donde 21 (95,45 %) son empresas con dificultades, constituye también un nodo terminal, que se asigna al **grupo 1**.

Por último las empresas del grupo e) se clasifican según el costo de financiamiento nuevamente en:

- i) 14 empresas sin dificultades presentan costos elevados (entre 27,51 y 1132,98) por lo cual puede decirse que son empresas con capacidad para generar utilidades y poder cubrir sus costos de financiamiento. Este nodo terminal está formado, entonces, por empresas sin dificultades (**grupo 2**).
- j) 3 empresas con costos más elevados todavía por lo cual este nodo terminal también está formado en su mayoría (2) por empresas sanas. Este nodo terminal también pertenece al **grupo 2**.

En definitiva, el grupo 1 está constituido por 25 empresas y el grupo 2, por las 25 empresas restantes.

A continuación se presenta el resumen de ganancias que produce el procesamiento del árbol.

Cuadro 2.19. Resumen de Ganancias

Variable criterio: CONDICIO N° 2						
Nodo	Nodo: n	Nodo: %	Resp: n	Resp: %	Ganancia %	Indice %
3	4	8.00	4	16.00	100.00	200.00
9	14	28.00	14	56.00	100.00	200.00
7	4	8.00	3	12.00	75.00	150.00
10	3	6.00	2	8.00	66.67	133.33
6	3	6.00	1	4.00	33.33	66.67
8	22	44.00	1	4.00	4.55	9.09

En el resumen de ganancias para los nodos terminales, para la variable criterio: Condición 2 (empresas sin dificultades), pueden leerse los estadísticos que representan:

*Nodo: n* indica que cada nodo terminal ha capturado *n* casos por ejemplo el nodo 3 tiene 4 empresas del total de 50.

*Nodo: %* indica el porcentaje de empresas en ese nodo, por ejemplo en el nodo 3 se han captado el 8 % del total de empresas.

*Resp: n* indica que del total de empresas captadas en el nodo, *n* son empresas según la variable criterio 2, es decir empresas sin dificultades. Siguiendo con el ejemplo del nodo 3, del total de 4 empresas, las 4 son empresas sin dificultades, lo que es equivalente decir que el nodo 3 no tiene empresas en crisis.

*Resp: %* indica el porcentaje de empresas sin dificultades del total de empresas, en el caso del nodo 3, las empresas sin dificultades representan el 16 % del total.

*Ganancia %* muestra el porcentaje de casos del nodo que presentan el valor criterio en la variable criterio, es decir que para el nodo 3, dividimos 4 por 4 y obtenemos el 100 % de empresas sin dificultades, lo que equivale a decir que no hay empresas en crisis.

*Indice %* indica la composición del nodo respecto a las empresas sin dificultades comparada con la composición de toda la muestra. Este valor se obtiene como la tasa de ganancia % respecto a la proporción de respuestas de la categoría criterio de toda la muestra (porcentaje de empresas sin dificultades en el nodo raíz: 50 %).

En definitiva, este cuadro muestra cuáles nodos tienen la mayor y la menor proporción de categorías criterio. En este caso, lo interesante es saber que subgrupo de empresas (nodos) tienen mayores posibilidades de ser clasificados como buenas (sin dificultades). Para este problema tanto el nodo 3 y el 9 no presentan clasificaciones erróneas por lo que ambos muestran un valor de ganancia del 100 %. Dado que ello duplica el porcentaje de los casos con clasificación buena de toda la muestra (50 %), el índice de ganancia es del 200 %. Claramente estos son los casos que deben darse. Por el contrario, el nodo 8 presenta la menor proporción de empresas sin dificultades. Por lo tanto las empresas con altos intereses en la contratación de terceros y poca capacidad para generar utilidades, acompañado de una alta inmovilización del Patrimonio Neto son las que más condiciones tienen de estar en crisis.

Entonces, los grupos que se detectaron son los siguientes:

Clase 1	Empresas con dificultades (en crisis): a) con altas tasas de interés en la contratación de fondos de terceros b) con altos niveles de inmovilización del activo y del Patrimonio Neto.
Clase 2	Empresas sin dificultades: a) con bajas tasas de interés en la contratación de fondos de terceros b) con bajos niveles de inmovilización del activo y del Patrimonio Neto.

Siendo la matriz de clasificación errónea la siguiente:

		Categoría real		
		1. En crisis	2. Sin dificultades	Total
Categoría Estimada	1. En crisis	23	2	25
	2. Sin dificultades	2	23	25
	Total	25	25	50

Esta es una tabla de recuento de los valores para las categorías real y estimada. En las casillas diagonales de la tabla se muestran los valores de clasificación correcta. Estos elementos representan los acuerdos entre el valor real y el valor estimado, medido en términos de aciertos. En los elementos no diagonales de la matriz se encuentra el recuento de las clasificaciones erróneas, en este caso en particular 2 empresas sin dificultades fueron clasificadas como en crisis y dos empresas en crisis fueron clasificadas como sin dificultades, situación que es mas riesgosa que la anterior.

La tasa de error obtenida

	Reestimación	Validación cruzada
Estimación del riesgo	0.08	<b>0.22</b>

Es decir, que el árbol ha clasificado erróneamente el 8 % de las veces, considerando la tasa de error aparente, mientras que la tasa de error crossvalidada con fines de predicción es del 22 %.

## 2.7. ERROR DE CLASIFICACION

En este trabajo de investigación, la muestra se ha seleccionado según un muestreo separado, que es el que generalmente se utiliza en análisis discriminante y consiste en extraer una muestra de tamaño fijo  $n_i$  de cada clase  $G_i$  ( $i = 1, 2, \dots, g$ ). Sobre la base de la muestra de entrenamiento se establece una regla de clasificación que se denota  $r_i$  (Díaz, M. 2001).

Un criterio de clasificación puede ser evaluado según la performance en la clasificación de una futura observación.

Cuando se trabaja con métodos no paramétricos, la tasa de error puede estimarse según distintas metodologías. En el procesamiento de los datos se utilizaron:



- a) Tasa de error aparente: que establece la proporción de objetos de la muestra que fueron mal clasificados, esta tasa subestima considerablemente la verdadera tasa de error condicional, ya que reclasifica los elementos de la muestra con la regla a partir de ellos.
- b) Crossvalidation: que permite omitir observaciones del conjunto de datos, recalculan la regla de clasificación sobre la base de las restantes observaciones y luego usar la nueva regla para clasificar las observaciones omitidas; por último se calcula la proporción de errores que se cometen con el procedimiento.

Para los métodos de Kernel y vecino más cercano, SAS usa dos tipos de ratios de error de estimación al evaluar el criterio de clasificación basado en parámetros estimados para una muestra:

- estimación de la cantidad de error (ratio de error aparente)
- estimación de la probabilidad posterior del error.

La estimación de la cantidad de error es calculada por aplicación del criterio de clasificación derivado de una muestra o un conjunto de pruebas y luego contar los números de observaciones mal clasificadas. La cantidad estimada es la proporción de observaciones mal clasificadas en el grupo. Cuando el conjunto de prueba es independiente de la muestra, la estimación es insesgada.

De cualquier modo, esta puede tener una varianza alta, especialmente si el conjunto de prueba es pequeño.

Otro ratio es el de la crossvalidation (validación cruzada), que toma  $n-1$  fuera de  $n$  observaciones como un conjunto y determina con esas  $n-1$  una función discriminante y luego las aplica en clasificar la observación que ha quedado afuera. Así para cada una de las observaciones. El ratio de clasificación errónea para cada grupo es la proporción de observaciones muestrales en ese grupo, en el cual son mal clasificadas.

Las probabilidades posteriores de pertenencia a cada grupo se basan en las probabilidades posteriores de las observaciones clasificadas en algún grupo. La estimación de las probabilidades posteriores son buenas para el ratio de error cuando estas ocurren.

Generalmente, el ratio de error es estimado a través del promedio de las estimaciones de los ratios de error de los grupos específicos individuales, donde las probabilidades a priori son usadas en las ponderaciones.

Para nuestro caso, las tasas de error fueron:

Método de Kernel	0,16
Método de Vecino más cercano	0,10

Para el árbol de clasificación se midió la precisión a través de la tasa de error por crossvalidation, lo cual implica dividir la muestra en una serie de muestras más pequeñas. Posteriormente se generan los árboles, excluyendo los datos de cada submuestra. Por ejemplo, con una validación cruzada de diez veces, los datos se dividen en diez submuestras y luego se generan diez árboles. El primer árbol se basa en todos los casos excepto aquellos del primer número de submuestras; el segundo árbol se basa en todos los casos excepto los del segundo número de submuestras, etc. Para cada árbol se calcula el riesgo de clasificación errónea aplicando el árbol a la submuestra que se excluyó cuando se lo generó. La estimación del riesgo mediante validación cruzada para todo el árbol se calcula como promedio de los riesgos de todos los árboles.

Método de Árbol de Clasificación	0,22
----------------------------------	------

Siendo este método el menos satisfactorio para nuestros datos.

## CAPITULO 3:

## RESULTADOS Y VALIDACION

### 3.1. RESULTADOS DE LOS METODOS DE KERNEL, EL VECINO MAS CERCANO Y ARBOL DE CLASIFICACION.

Luego del procesamiento con las 50 empresas clasificadas en dos grupos y fijando como estrategia la elección de un modelo que muestre el mejor comportamiento en lo relativo a tasas de error, la performance del criterio de clasificación fue considerar aquel que minimiza la tasa crossvalidada.

Se consideró un modelo que contiene ocho variables que resultaron significativas, lo cual se muestra en el cuadro 3.1.

Cuadro 3.1. Variables del modelo

X <sub>1</sub>	Liquidez ácida	Compara el activo rápido con las deudas de corto plazo.
X <sub>2</sub>	Propiedad del Activo	Compara los fondos propios con el total de los recursos de la empresa.
X <sub>3</sub>	Pasivo No corriente s/activo	Mide si las deudas a largo plazo pueden ser cubiertas por el activo.
X <sub>4</sub>	Autofinanciación	Mide la capacidad de financiar parte de sus inversiones con los ahorros operativos generados.
X <sub>5</sub>	Inmovilización del Activo	Indica la proporción del Activo que está inmovilizado
X <sub>6</sub>	Inmovilización del Pat. Neto	Expresa la política de financiación de los activos no corrientes.
X <sub>7</sub>	Utilidad sobre intereses	Mide la capacidad de la empresa para generar utilidades y hacer frente al costo de los pasivos.
X <sub>8</sub>	Endeudamiento	Mide la cantidad de pesos prestados por terceros respecto al capital propio.

Luego del procesamiento con el método de Kernel, el del vecino más cercano y el árbol de clasificación, los resultados se muestran en el cuadro 3.2, en donde se observa que la menor tasa de error (crossvalidation) se obtuvo con el método del vecino mas cercano, considerando  $k = 2$  observaciones más cercanas a cada uno de los elementos que se consideró en el análisis (0,10).

Cuadro 3.2. Métodos y tasas de error en el modelo

Modelos	Tasa de error crossvalidada
Kernel (Epanechnikov)	0.16
Vecino más cercano ( $k = 2$ )	0.10
Arbol de clasificación	0,22

### 3.2. APLICACIÓN DE LOS METODOS DE CLASIFICACION NO PARAMETRICOS CON NUEVAS OBSERVACIONES.

En una primera etapa se consideraron empresas que desde la segunda mitad del año 2000 en adelante ingresaron en ronda reducida. Se seleccionaron cinco de ellas con sus correspondientes pares (empresas que no están en situación de crisis), según los totales de activo y los ingresos por ventas al cierre de ejercicio, dos años antes al ingreso a rueda reducida. Así se conformaron cinco pares de empresas con la finalidad de predecir la crisis financiera, aplicando esta metodología no paramétrica de clasificación..

El Cuadro 3.3 muestra las nuevas empresas, según su actividad.

Cuadro 3.3. Empresas por rama de actividad

Rama de Actividad	En ronda reducida (cantidad)	En ronda normal (cantidad)
Inversiones	1	1
Molinos harineros	1	1
Industrias textiles	1	1
Explot. Agropecuaria	1	1
Servicios de energía	1	1
<b>Totales</b>	<b>5</b>	<b>5</b>

El Cuadro 3.4. muestra los Ingresos totales y el nivel de Activos para formar los nuevos pares que fueron clasificados con los diferentes métodos.

Cuadro 3.4. Empresas en ambos grupos según Activos e Ingresos Totales

Ingresos Totales			Activos Totales		
Intervalos (en miles de pesos)	En ronda reducida (cant.)	En ronda normal (cant.)	Intervalos (en miles de pesos)	En ronda reducida (cant.)	En ronda normal (cant.)
10500-40500	3	2	0 – 100500	2	2
40500-60500	0	2	100500-600550	2	2
60500-856500	2	1	600550-858050	1	1
Totales	5	5	Totales	5	5

Se procesaron todas las empresas (60) con el paquete SAS, utilizando el método de Kernel y del vecino más cercano, como así también se construyó el árbol de clasificación con el paquete SPSS y se obtuvieron las siguientes conclusiones:

- A) Es fundamental distinguir las empresas según la fecha en la cual ingresan en ronda reducida. Para las empresas que lo hicieron con anterioridad a fines del año 2001, los balances tienen un comportamiento similar al observado en la base de datos con la cual se realizó la clasificación (capítulo 2), mientras que para las empresas que ingresaron a ronda reducida con posterioridad a fines del año 2001, la clasificación no tuvo capacidad para distinguir estas empresas en crisis de las empresas sanas y por lo tanto fueron mal clasificadas. Con lo cual puede concluirse que los índices definidos a través de los estados contables no son los adecuados, lo que hace suponer que si las empresas ingresan a cotizar en ronda reducida es porque existen otras variables económicas, o bien, otros índices contables a definir que implicarían proseguir con el estudio de esta metodología definiendo un nuevo universo de variables para el presente escenario, en donde las empresas deben actuar.
- B) Por otro lado, no podemos dejar de mencionar el hecho de que la Comisión Nacional de Valores ha impuesto nuevas resoluciones respecto a la presentación de los estados contables, entre ellas la exigencia del ajuste por inflación a partir de mediados del año 2002 y luego su suspensión a partir de Marzo de 2003, lo cual hace que en algunos períodos haya estados contables no comparables con otros. Con relación a ello, la Resolución N° 415 de la Comisión Nacional de Valores, dispone que los balances presentados por las sociedades cotizantes a partir del 31 de julio de 2002 deben estar ajustados por inflación. La Resolución N° 441/03 de dicho Organismo dispuso que a partir del 1° de marzo de 2003 las entidades sujetas a la fiscalización de la Comisión, discontinuarán, a todos los efectos, la aplicación del método de reexpresión de los estados contables en moneda homogénea. También son de aplicación las Nuevas Normas Contables, por lo cual la Resolución N° 434/03 de la Comisión Nacional de Valores dispuso que serán de aplicación para los ejercicios iniciados a partir del 1° de enero de 2003, las Resoluciones Técnicas N° 16, 17, 18, 19 y 20 de la Federación

Argentina de Consejos Profesionales de Ciencias Económicas, admitiéndose su aplicación anticipada.

C) Se está ante una etapa de transición. A fin de reelaborar el estudio, es necesario que transcurra un considerable período de tiempo para contar con empresas, cuyos estados contables sean comparables. En estos momentos tenemos empresas que aplican normas contables anteriores y otras que están aplicando las nuevas resoluciones.

En una segunda etapa, se procedió a excluir a aquellas empresas que ingresaron en ronda reducida con posterioridad a fines del año 2001.

Se trabajó con los siguientes pares de empresas (Cuadro 3.5):

Cuadro 3.5. Empresas por rama de actividad

<b>Rama de Actividad</b>	<b>En ronda reducida (cantidad)</b>	<b>En ronda normal (cantidad)</b>
Explot. Agropecuaria	1	1
Servicios de energía	1	1
<b>Totales</b>	<b>2</b>	<b>2</b>

Procesando estas nuevas empresas la tasa de error fue similar a la encontrada con las 50 empresas que se utilizaron para definir la regla.

Por último, en una última etapa, se consideró uno sólo de estos pares, obteniéndose resultados similares, es decir que las tasas de error fueron parecidas a las que se obtuvieron en el capítulo 2.

Se presentan como anexo las salidas computacionales del procesamiento de los datos.

### **3.3. CONCLUSIONES.**

En la selección de las empresas se recurrió a los registros de la Bolsa de Comercio de Buenos Aires, que reúne de manera periódica los balances de aproximadamente 250 sociedades anónimas. Para cotizar en Bolsa, las empresas deben sujetarse a

determinados requisitos dispuestos en el Reglamento Interno de esta institución, también allí se establece que si no se cumplen con determinadas prescripciones, la empresa pasa a cotizar en la llamada “ronda reducida”. Así fue que se tuvo en cuenta aquellas sociedades que estuviesen en crisis económica-financiera en el período Julio de 1995 a Mayo de 2000, lo que se ve reflejado en su ingreso a ronda reducida, es decir el nivel de operaciones es menor al que debería tener una empresa normal, cuyas acciones son requeridas en el mercado. Es obvio que si una empresa comienza a tener dificultades, sus acciones no serán cotizadas en alza, es más, las operaciones irán disminuyendo hasta tanto la empresa muestre posibilidades de reactivación.

A su vez, se vio la necesidad de seleccionar otras empresas que acompañaran a las anteriores, de tamaño y rubro similar. Si una empresa ingresa a ronda reducida, se elige para esta unidad de análisis el último balance cerrado entre uno y dos años antes y lo mismo para la empresa que actúa como control, tratando de tomar la misma fecha de cierre o lo más cercana posible. Si ambas empresas son similares en cuanto a tamaño y rama de actividad y sus balances son comparables en los mismos momentos del tiempo, es altamente probable que las diferencias entre estar en quiebra y no estarlo puedan ser explicadas en términos de las diferencias existentes en la composición del vector económico-financiero representativo de los balances cerrados en un año dado. El vector económico-financiero está formado por los ratios del balance, con los cuales se definieron las variables del modelo.

Con las 50 empresas seleccionadas, con las variables definidas con los datos de los balances (variables continuas), a través de métodos de clasificación no paramétricos se pretendió discernir si la firma entrará o no en estado de falencia al cabo de dos años aproximadamente.

Con el método de Kernel, la tasa de error crossvalidada fue de 0,16, con el método del vecino más cercano, con  $k = 2$  (dos observaciones más cercanas), la tasa de error se mejoró siendo de 0,10 y por último con el árbol de clasificación la tasa de error crossvalidada fue de 0,22, siendo éste el método que mostró resultados menos satisfactorios.

Luego se aplicaron estos métodos de clasificación con nuevas observaciones, considerándose empresas que habían ingresado en ronda reducida en el período que va desde mediados del año 2000 en adelante, hasta mediados del año 2002 y sus correspondientes pares de empresas en situación económica – financiera normal,

donde los siguientes hechos: el cambio de escenario económico producido a partir de fines del año 2001, la aplicación de nuevas normas contables, como así también las nuevas resoluciones de la Comisión Nacional de Valores, en materia de ajuste por inflación, nos indican que se está en una etapa de transición, donde se deberá plantear un nuevo universo de variables que hagan que los estados contables sean comparables y se pueda discriminar entre empresas, siempre con el objetivo de predecir las crisis económica-financiera. Por otro lado, considerando empresas que ingresaron a ronda reducida con anterioridad a finales del año 2001, las tasas de error crossvalidada para los métodos de Kernel y del Vecino más cercano fueron similares a las encontradas, como así también con el árbol de clasificación.

Por último, es necesario tener en cuenta, la importancia en la definición de la base de datos para la aplicación de esta metodología no paramétrica de clasificación a fin de lograr conclusiones consistentes. Respecto a los métodos utilizados, en otras aplicaciones, el método del vecino más cercano fue el que obtuvo resultados más satisfactorios.



## REFERENCIAS

1. Aitchison, J y Aitken, C (1976) *Multivariate binary discrimination by the kernel method*. Biometrika 63, 413-420
2. Altman, E. (1988). *The prediction of Corporate Bankruptcy*. Garland Publishing, Inc. N York. London.
3. Anderson, T.W. (1984, 2da.edic.). *An Introduction to Multivariate Statistical Analysis*. 1ra edic. (1958). New York, Wiley.
4. As.Vs. (1989). "Discriminant Analysis and Clustering" *Statistical Science*, Vol. 4, , N°1, 34-69.
5. Barber, David (2001) *Learning from Data I, Nearest Neighbour Classification* Curso en Internet <http://anc.ed.ac.uk>.
6. Breiman, L.Friedman, J. Olshen, R. y Stone, C. (1998) *Classification and regression Trees*. Ed. Chapman & Hall.
7. Diaz, Margarita (2001) *Performance del Análisis Discriminante Regularizado en la predicción de Crisis Financieras* . Tesis Maestría en Estadística Aplicada U.N. de Córdoba.
8. Hall, P. y Matthew, W. (1988) *On Nonparametric Discrimination Using Density Differences* Biometrika vol 75 N° 3 pp 541-547.
9. Hand, D.J. (1982) *Kernel discriminant analysis*. Wiley.
10. Hand, D.J. (1981) *Discrimination and Classification*. New York, Wiley.
11. Hand, D.J. (1999). *Construction and Assessment of Classification Rules*.. Chichester, Wiley.
12. Izenman, Alan (1991) *Recent Developments in Nonparametric Density Estimation* JASA vol. 86 N° 413 pp 205-224.
13. Johnson, R.A. y Wichern, D.W. (1992, 3 edic.). *Applied Multivariate Statistical Analysis*. 1ra. edic.(1982). New York, Prentice-Hall.
14. Littell, R. Milliken, G, Stroup, W y Wolfinger R. (1996) *SAS System for Proc Discrim*. SAS Institute Inc. Cary, North Carolina.
15. Saunders, A. (1997) *Financial Institutions Management, a modern perspective*. Mc.Graw Hill, 2da. edic.
16. SPSS Inc. Manual del Usuario "Answer Tree 2.0." (1998).

17. Tutz, G. (1986) *An alternative choice of smoothing for kernel-based density estimates in discrete discriminant analysis*. *Biometrika* 73, 405-411.
18. Wilkinson, Leland (1992) *Tree Structured Data Analysis: AID, CHAID and CART* SPSS Inc. 233. South Wacker, Chicago, paper presentado en Joint Software Conference.

**ANEXO**

**PROCESAMIENTO APLICANDO LOS METODOS NO PARAMETRICOS  
CON NUEVAS OBSERVACIONES**

En un primer momento, se procesaron 60 empresas con el paquete SAS, considerando que algunas ingresaron en ronda reducida con anterioridad a fines del 2001 y otras con posterioridad.

Método de Kernel (60)

Number of Observations and Percent Classified into CONDICIO					
From CONDICIO		1	2	Total	
1	1	19	11	30	
		63.33	36.67	100.00	
2	2	5	25	30	
		16.67	83.33	100.00	
Total		24	36	60	
		40.00	60.00	100.00	
Priors		0.5	0.5		

Error Count Estimates for CONDICIO			
	1	2	Total
Rate	0.3667	0.1667	0.2667
Priors	0.5000	0.5000	

Posterior Probability of Membership in CONDICIO					
Obs	From CONDICIO	Classified into CONDICIO		1	2
7	1	2 *		0.4859	0.5141
8	2	1 *		0.5087	0.4913
9	1	2 *		0.4955	0.5045
13	1	2 *		0.4943	0.5057
14	2	1 *		0.5029	0.4971
25	1	2 *		0.4989	0.5011
27	1	2 *		0.4684	0.5316
33	1	2 *		0.4871	0.5129
43	1	2 *		0.4851	0.5149
44	2	1 *		0.5129	0.4871
48	2	1 *		0.5016	0.4984
51	1	2 *		0.4758	0.5242
52	2	1 *		0.5799	0.4201
55	1	2 *		0.4728	0.5272
57	1	2 *		0.4885	0.5115
59	1	2 *		0.4877	0.5123

\* Misclassified observation

Método de Vecino más Cercano (60)

Number of Observations and Percent Classified into CONDICIO

From CONDICIO	1	2	Total
1	23 76.67	7 23.33	30 100.00
2	4 13.33	26 86.67	30 100.00
Total	27 45.00	33 55.00	60 100.00
Priors	0.5	0.5	

Error Count Estimates for CONDICIO

	1	2	Total
Rate	0.2333	0.1333	0.1833
Priors	0.5000	0.5000	

Posterior Probability of Membership in CONDICIO

Obs	From CONDICIO	Classified into CONDICIO	
		1	2
8	2	1 *	1.0000
21	1	2 *	0.0000
27	1	2 *	0.0000
43	1	2 *	0.0000
44	2	1 *	1.0000
48	2	1 *	1.0000
51	1	2 *	0.0000
52	2	1 *	1.0000
55	1	2 *	0.0000
57	1	2 *	0.0000
59	1	2 *	0.0000

\* Misclassified observation

En un segundo momento, se procesaron 54 empresas con el paquete SAS, donde se excluyeron aquellas empresas que ingresaron a ronda reducida con posterioridad a fines del año 2001.

Método de Kernel (54)

Number of Observations and Percent Classified into CONDICIO				
From CONDICIO	1	2	Total	
1	19	8	27	
	70.37	29.63	100.00	
2	4	23	27	
	16.67	83.33	100.00	
Total	23	31	54	
	40.00	60.00	100.00	
Priors	0.5	0.5		
Error Count Estimates for CONDICIO				
	1	2	Total	
Rate	0.2963	0.1481	0.2222	
Priors	0.5000	0.5000		
Posterior Probability of Membership in CONDICIO				
		Classified		
	From	into		
Obs	CONDICIO	CONDICIO	1	2
7	1	2 *	0.4857	0.5143
9	1	2 *	0.4928	0.5072
13	1	2 *	0.4951	0.5049
14	2	1 *	0.5032	0.4968
17	1	2 *	0.4960	0.5040
27	1	2 *	0.4651	0.5349
33	1	2 *	0.4807	0.5193
38	2	1 *	0.5002	0.4998
43	1	2 *	0.4798	0.5202
44	2	1 *	0.5158	0.4842
53	1	2 *	0.4612	0.5388
54	2	1 *	0.5154	0.4846
* Misclassified observation				

Método de Vecino más Cercano (54)

Number of Observations and Percent Classified into CONDICIO				
From CONDICIO		1	2	Total
1		22	5	27
		81.48	18.52	100.00
2		2	25	27
		7.41	92.59	100.00
Total		24	30	54
		44.44	55.56	100.00
Priors		0.5	0.5	

Error Count Estimates for CONDICIO			
	1	2	Total
Rate	0.1852	0.0741	0.1296
Priors	0.5000	0.5000	

Posterior Probability of Membership in CONDICIO				
Obs	From CONDICIO	Classified into CONDICIO	1	2
7	1	2 *	0.0000	1.0000
21	1	2 *	0.0000	1.0000
27	1	2 *	0.0000	1.0000
43	1	2 *	0.0000	1.0000
44	2	1 *	1.0000	0.0000
48	2	1 *	1.0000	0.0000
53	1	2 *	0.0000	1.0000

\* Misclassified observation

Por último, se agregó un par de empresas a las mismas empresas con que se realizó el procesamiento anterior.

Método de Kernel (agregando un par de empresas)

Number of Observations and Percent Classified into CONDICIO				
	From CONDICIO	1	2	Total
	1	17	9	26
		65.38	34.62	100.00
	2	2	24	26
		7.69	92.31	100.00
	Total	19	33	52
		36.54	63.46	100.00
	Priors	0.5	0.5	

Error Count Estimates for CONDICIO				
		1	2	Total
	Rate	0.3462	0.0769	0.2115
	Priors	0.5000	0.5000	

Posterior Probability of Membership in CONDICIO				
Obs	From CONDICIO	Classified into CONDICIO	1	2
7	1	2 *	0.4850	0.5150
9	1	2 *	0.4946	0.5054
13	1	2 *	0.4926	0.5074
14	2	1 *	0.5003	0.4997
17	1	2 *	0.4961	0.5039
25	1	2 *	0.4957	0.5043
27	1	2 *	0.4565	0.5435
33	1	2 *	0.4705	0.5295
43	1	2 *	0.4734	0.5266
44	2	1 *	0.5206	0.4794
51	1	2 *	0.4931	0.5069

\* Misclassified observation

Método de Vecino más Cercano (agregando un par de empresas)

Number of Observations and Percent Classified into CONDICIO				
	From CONDICIO	1	2	Total
	1	22	4	26
		84.62	15.38	100.00
	2	2	24	26
		7.69	92.31	100.00
	Total	24	30	54
		46.15	53.85	100.00
	Priors	0.5	0.5	
Error Count Estimates for CONDICIO				
		1	2	Total
	Rate	0.1538	0.0769	0.1154
	Priors	0.5000	0.5000	
Posterior Probability of Membership in CONDICIO				
Obs	From CONDICIO	Classified into CONDICIO	1	2
7	1	2 *	0.0000	1.0000
27	1	2 *	0.0000	1.0000
43	1	2 *	0.0000	1.0000
44	2	1 *	1.0000	0.0000
48	2	1 *	1.0000	0.0000
51	1	2 *	0.0000	1.0000
* Misclassified observation				

A continuación se presentan los Árboles de Clasificación, para cada uno de los casos. En primer lugar con 60 empresas: de las diez nuevas empresas agregadas, tres de las que estaban en la categoría de empresas en crisis fueron mal clasificadas, lo cual coincide con lo obtenido con los otros métodos.



Luego con cuatro nuevas empresas y por último agregando un par de empresas. A continuación se expresan las tasas de error encontradas por validación cruzada.

Con diez nuevas empresas (Figura 3.1.):

Matriz de clasificación errónea				
Categoría estimada	Categoría real			Total
	1	2	3	
1	25	2	27	
2	5	28	33	
Total	30	30	60	

	Reestimación	Validación cruzada
Estimación de riesgo	0,116667	0,3
ET de la estimación del riesgo	0,0414438	0,0591608

Con cuatro nuevas empresas (Figura 3.2.):

Matriz de clasificación errónea				
Categoría estimada	Categoría real			Total
	1	2	3	
1	24	3	27	
2	3	24	27	
Total	27	27	54	

	Reestimación	Validación cruzada
Estimación de riesgo	0,111111	0,37037
ET de la estimación del riesgo	0,0427667	0,0657149

Con un par de empresas (Figura 3.3.):

Matriz de clasificación errónea				
Categoría estimada	Categoría real			Total
	1	2	3	
1	23	2	25	
2	2	23	25	
Total	25	25	50	

	Reestimación	Validación cruzada
Estimación de riesgo	0,08	0,2
ET de la estimación del riesgo	0,0383667	0,0565685

Figura 3.1.

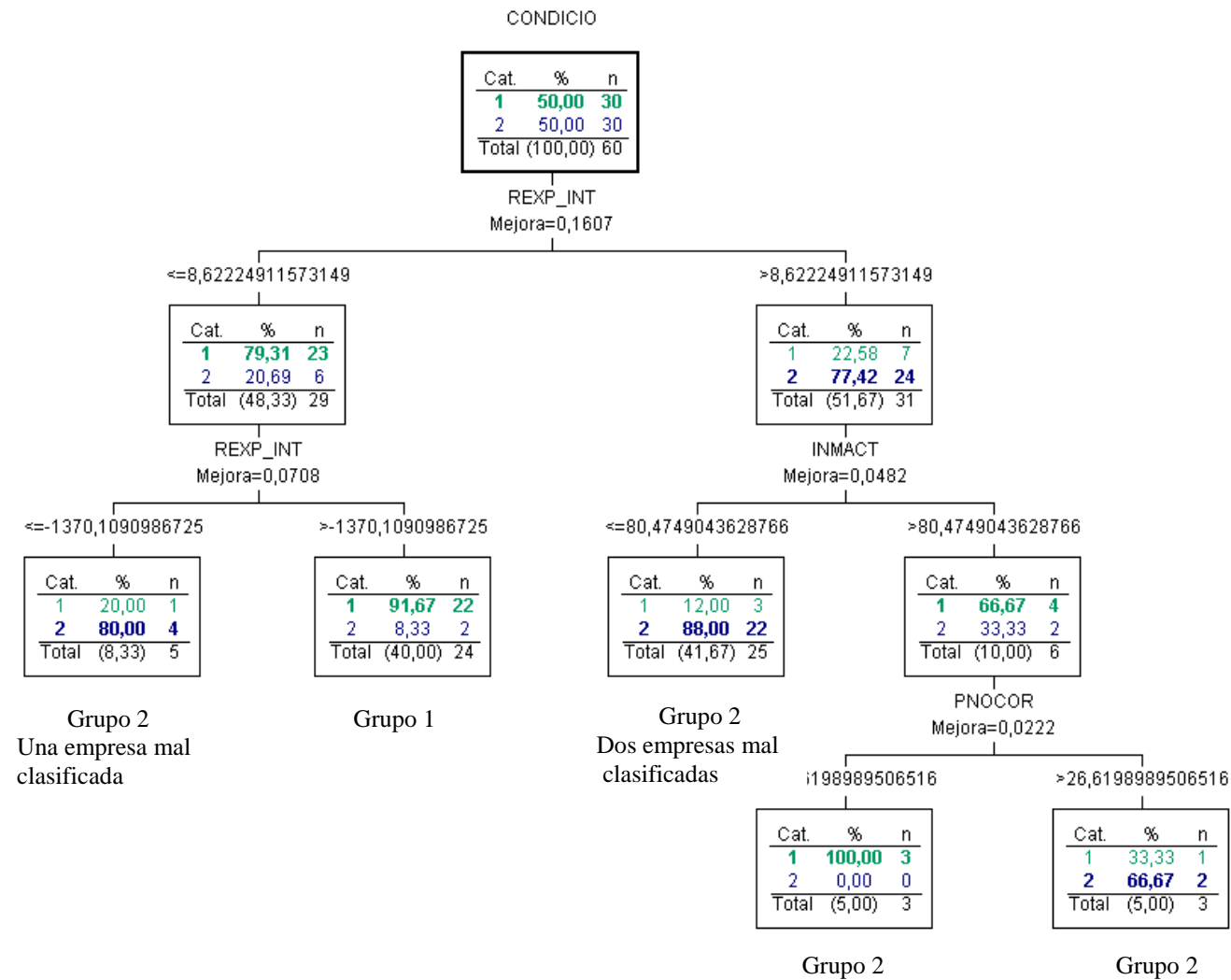


Figura 3.2.

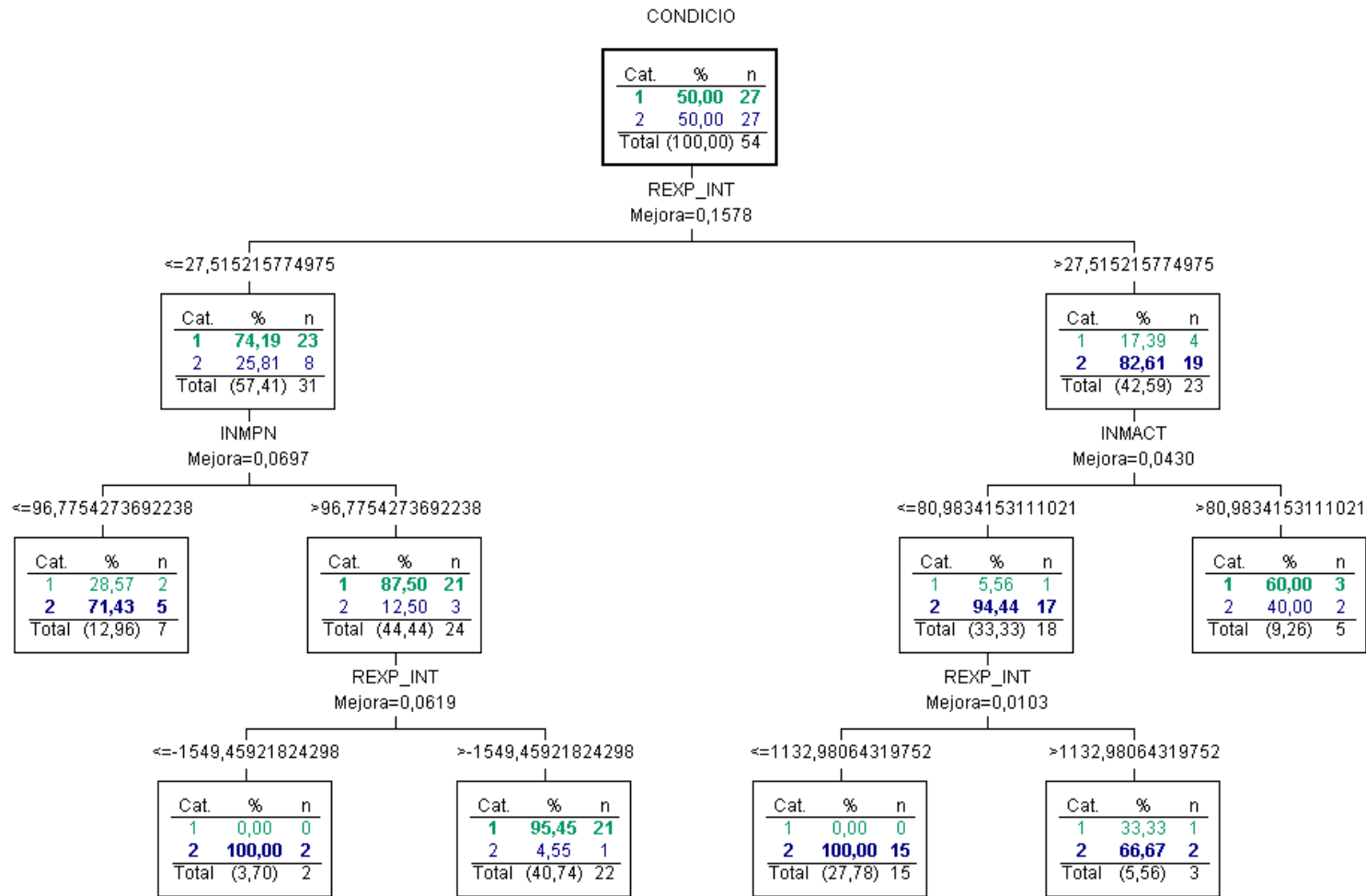


Figura 3.3.

