



Universidad  
Tecnológica  
de Pereira

**¡Estamos entre las mejores universidades del país..!**

**BIENVENIDOS**

**Acreditada Institucionalmente de Alta Calidad  
por el Ministerio de Educación Nacional**

**Portal Institucional: [www.utp.edu.co](http://www.utp.edu.co)**



Gestión de la Calidad ISO 9001 : 2008  
Gestión Pública NTC GP 1000 : 2009





Universidad  
Tecnológica  
de Pereira

Portal Institucional: [www.utp.edu.co](http://www.utp.edu.co)

# ANÁLISIS MULTIVARIADO TECNICAS DE SEGMENTACION

JAIRO ALBERTO VILLEGAS FLÓREZ  
SANTIAGO VÁSQUEZ ARTUNDUAGA



Universidad Tecnológica de Pereira  
Acreditada Institucionalmente de Alta Calidad por el Ministerio de Educación Nacional



# TECNICAS DE SEGMENTACION. CONCPETOS Y CLASIFICACION

**CONCEPTOS BASICOS:** La segmentación puede definirse como el proceso de dividir un todo (población, consumidores, etc) en grupos uniformes mas pequeños que tengan características semejantes. Dicho proceso suele realizarse según los valores de denominadas variables que son los que fijan sus características. Los segmentos son grupos homogéneos. Debido a la similitud de los elementos dentro de cada grupo, es probable que respondan de modo similar a determinadas estrategias: marketing, ventas, precios, distribución, tratamientos, etc.

# TECNICAS DE SEGMENTACION. CONCPETOS Y CLASIFICACION

## CLASIFICACION DE LAS TECNICAS DE SEGMENTACION:

1. Técnicas ***Predictivas y Descriptivas***: Las ***predictivas***, en las que las variables que intervienen puede clasificarse inicialmente en dependientes e independientes (similar a técnicas de dependencia o métodos explicativos en análisis multivariante) y las técnicas ***descriptivas***, en las que todas las variables tienen inicialmente el mismo estatus (similar a las técnicas de interdependencia o métodos descriptivos en AM).

# TECNICAS DE SEGMENTACION. CONCPETOS Y CLASIFICACION

Las **técnicas predictivas** especifican el modelo para los datos con base a un conocimiento teórico previo. El modelo supuesto para los datos debe contrastarse después de minería de datos antes de aceptarlo como validos. Fases: **Identificación objetiva, estimación, diagnosis (validación) y predicción.**

**\*Modelos de elección discreta, análisis discriminante; arboles de decisión, redes neuronales, análisis conjunto, etc \***

Estas técnicas también se denominan **Técnicas De Clasificación**, ya que permiten extraer perfiles de comportamiento o clase, cuyo objetivo es crear un modelo que permita clasificar cualquier nuevo individuo.

# TECNICAS DE SEGMENTACION. CONCPETOS Y CLASIFICACION

En las ***técnicas descriptivas*** no se asignan ningún papel predeterminado a las variables (No se supone la existencia de variables dependientes o independientes) ni la existencia de un modelo previo para los datos, los modelos se crean ***automáticamente*** partiendo del reconocimiento de patrones.

***\*Clustering, técnicas de reducción de la dimensionalidad, escalamiento multidimensional, análisis de correspondencia, etc \****

La dos técnicas están enfocadas al descubrimiento del conocimiento debido a los datos



# TECNICAS DE SEGMENTACION. CONCPETOS Y CLASIFICACION

**2. Técnicas *ad-Hoc* y *Post-hoc***, en las segmentación ***ad-hoc***, tanto el numero de segmentos como su tamaño o su descripción de establecen antes de que se inicie el estudio. En primer lugar, el investigador selecciona alguna base a partir de la cual segmentar (beneficio, necesidades, fidelidad de marcar, etc). En segundo lugar, el investigador clasifica a los individuos en segmentos de acuerdo con la base elegida y estudia su relación con otras variables descriptoras de tipo demográfico, socioeconómico, etc.

Ejemplos: Arboles de decisión y análisis discriminante, son técnicas típicas de segmentación ***ad-hoc***.

# TECNICAS DE SEGMENTACION. CONCPETOS Y CLASIFICACION

En la segmentación ***Post-hoc*** no se conocen inicialmente el número de segmentos ni tamaño. Esta situación se da, por ejemplo, cuando se desconocen las relaciones de los individuos ante los valores de las características de segmentación, cuyo caso resulta mas eficaz una segmentación ***Post-hoc***. En este modelo, el numero de segmentos, su tamaño y su descripción se conocen tras el análisis y no antes.

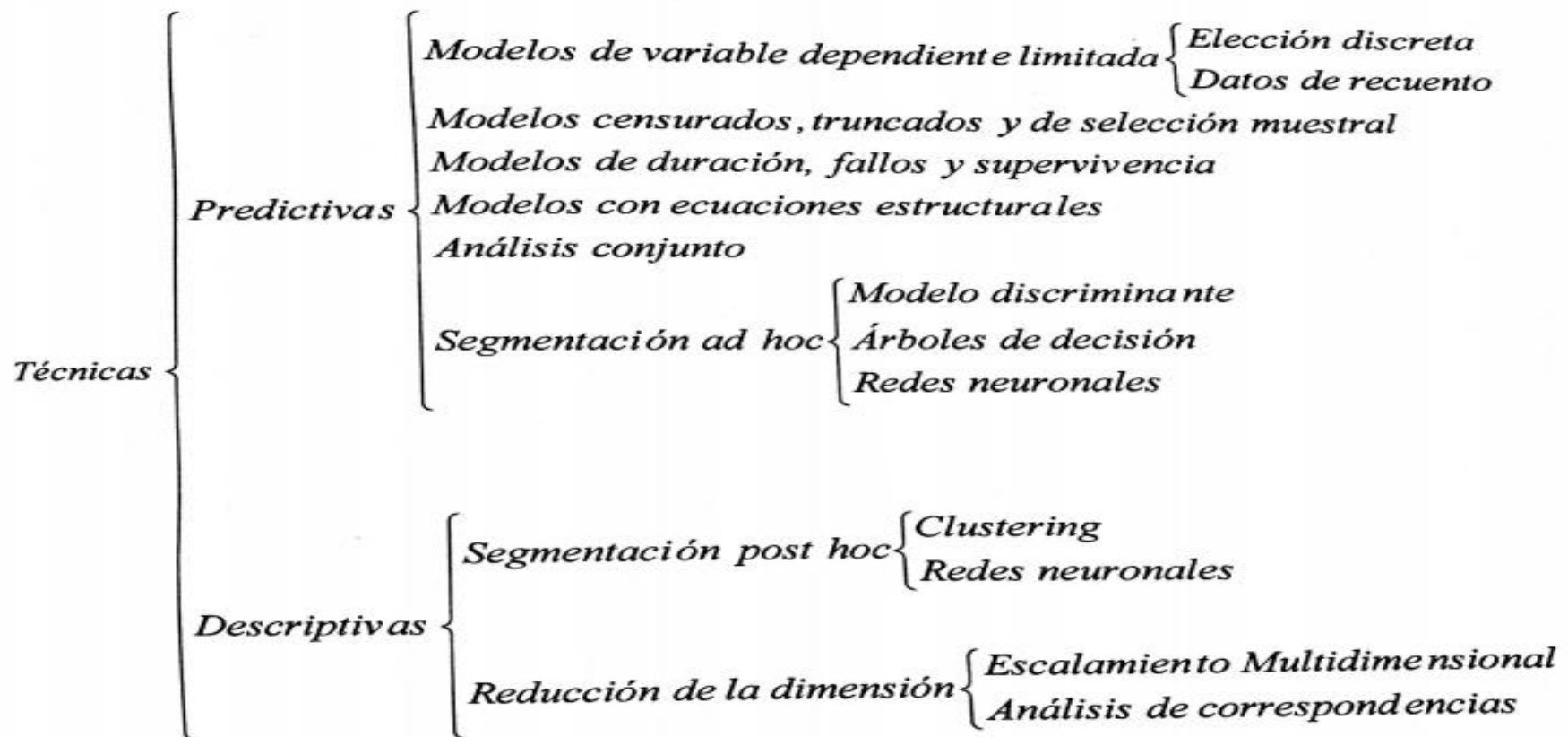
Inicialmente se realiza un exploración de la población y a continuación un análisis conglomerados que agrupa a los sujetos de acuerdo con la similitud de sus perfiles respecto a algunas variables de segmentación



# TECNICAS DE SEGMENTACION. CONCPETOS Y CLASIFICACION

Preestablecidas (comportamiento de compras, actitudes, beneficios buscados, etc)

A continuación se presenta una *clasificación de las técnicas de segmentación*:





Universidad  
Tecnológica  
de Pereira

Portal Institucional: [www.utp.edu.co](http://www.utp.edu.co)

# ANÁLISIS MULTIVARIADO SEGMENTACION AD-HOC ARBOLES DE DECISION

## ARBOLES DE DECISION COMO TECNICA PREDICTIVA DE SEGMENTACION



Vicerrectoría Académica - Centro de Registro y Control Académico

Universidad Tecnológica de Pereira

Acreditada Institucionalmente de Alta Calidad por el Ministerio de Educación Nacional



# ARBOLES DE DECISION

Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial. Dada una base de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.







# ARBOLES DE DECISION

Un árbol de decisión tiene unas entradas las cuales pueden ser un objeto o una situación descrita por medio de un conjunto de atributos y a partir de esto devuelve una respuesta la cual en últimas es una decisión que es tomada a partir de las entradas. Los valores que pueden tomar las entradas y las salidas pueden ser valores discretos o continuos. Se utilizan más los valores discretos por simplicidad, cuando se utilizan valores discretos en las funciones de una aplicación se denomina clasificación y cuando se utilizan los continuos se denomina regresión.





Universidad  
Tecnológica  
de Pereira

Portal Institucional: [www.utp.edu.co](http://www.utp.edu.co)

# GENERALIDADES PARA IBM SPSS Y LOS ARBOLES DE DECISION



Vicerrectoría Académica - Centro de Registro y Control Académico

Universidad Tecnológica de Pereira

Acreditada Institucionalmente de Alta Calidad por el Ministerio de Educación Nacional



# GENERALIDADES PARA IBM SPSS Y LOS ARBOLES DE DECISION

El procedimiento ***Árbol de Clasificación*** crea un modelo de clasificación basado en arboles, y ***clasifica casos en grupos o pronostica valores de una variable*** (criterio) dependiente basada en valores de variables independientes (predictores). El procedimiento proporciona herramientas de validación para análisis de clasificación exploratorio y confirmatorio, y puede utilizarse en múltiples técnicas que se describen a continuación:







# GENERALIDADES PARA IBM SPSS Y LOS ARBOLES DE DECISION

**Segmentación:** Identifica individuos que pueden ser miembros de un grupo específico.

**Estratificación:** Asigna los casos a una categoría de entre varias, por ejemplo, grupos de alto riesgo, bajo riesgo y riesgo intermedio.

**Predicción:** Crea reglas y las utiliza para predecir eventos futuros, como la verosimilitud de que una persona cause mora en un crédito o el valor de reventa potencial de un vehículo o una casa.

**Reducción de datos y clasificación de variables:** Selecciona un subconjunto útil de predictores para utilizarlo en la creación de un modelo paramétrico formal.





# GENERALIDADES PARA IBM SPSS Y LOS ARBOLES DE DECISION

**Identificación de Iteración:** Identifica las relaciones que pertenecen solo a subgrupos específicos y las especifica en un modelo paramétrico formal.

**Fusión de categorías y Distretizacion de variables continuas:** Recodifica las categorías de grupo de los predictores y las variables continuas, con una pérdida mínima de información.





Universidad  
Tecnológica  
de Pereira

Portal Institucional: [www.utp.edu.co](http://www.utp.edu.co)

# ANÁLISIS MULTIVARIADO

## TIPOS DE ARBOLES DE DECISION



Vicerrectoría Académica - Centro de Registro y Control Académico

Universidad Tecnológica de Pereira

Acreditada Institucionalmente de Alta Calidad por el Ministerio de Educación Nacional



# TIPOS DE ARBOLES DE DECISION

## ARBOLES CHAID.

El método CHAID (Chi-square Automatic Interaction Detector) es en conclusión una serie de métodos basados en el Detector Automático de Iteraciones (AID) de Morgan y Sonquist. Se trata de un método exploratorio de análisis de datos, útil para identificar variables importantes, y sus interacciones enfocadas a la segmentación y a los análisis descriptivos, que suelen ser pasos previos a otros análisis posteriores. La variable dependiente puede ser *cualitativa (nominal u ordinal)* o *cuantitativa*. Para variables cualitativas, el análisis lleva a cabo una serie de análisis  $\chi^2$  entre las variable dependiente y la predictora.



# TIPOS DE ARBOLES DE DECISION

En el caso de variables dependientes cuantitativas, se recorre a métodos de análisis de varianza, en los que los intervalos (divisiones) se determinan óptimamente para las variables independientes, de forma que maximice la capacidad para explicar la varianza de la medida dependiente. Se divide cada nodo localizando el par de categorías permisibles del predictor con menor valor de  $\chi^2$ . Si el nivel de significación es menor que un cierto nivel crítico, se unen ambas categorías y se repite el proceso. Si es mayor, se convierten en dos candidatas a la división de la variable.





# TIPOS DE ARBOLES DE DECISION

## **ARBOLES CART.**

El método CART (classification And Regression Tree) o *C&RT* es una alternativa al método CHAID exhaustivo para árboles de clasificación (variables dependientes categóricas). Este método surgió para intentar superar algunas dificultades y deficiencias que por entonces mostraba la formulación original del CHAID, que estaba limitado inicialmente variables dependientes nominales y variables independientes (predictores) categóricas hasta la aparición de su versión exhaustiva (estaba claro que se necesitaba utilizar predictores de cualquier nivel de medida).





# TIPOS DE ARBOLES DE DECISION

Además, *CART* tiene una estructura estadística más fuerte que CHAID, lo que le llevo a ser utilizado en campos de la investigación como la medicina.

El método comienza dividiendo la muestra en subconjuntos y evaluando cada predictor cuantitativo para encontrar el mejor punto de corte o cada predictor categórico y para encontrar las mejores agrupaciones de categorías. A continuación se comparan también los predictores, seleccionando el predictor y la división que produce la mayor bondad de ajuste.



# TIPOS DE ARBOLES DE DECISION

Para *predictores cuantitativos* suele utilizarse la minimización del error cuadrático o de la desviación absoluta media con respecto a la mediana. Para *predictores cualitativos* suele utilizarse el *coeficiente Gini* para evaluar la probabilidad de una mala clasificación (valor cero para la clasificación perfecta y valor uno para una mala clasificación).

*CART se utiliza para arboles de clasificación con variable dependiente cualitativa y para arboles de regresión con variable dependiente cuantitativa, y genera arboles binarios (solo produce divisiones binarias).*





# TIPOS DE ARBOLES DE DECISION

## **ARBOLES QUEST.**

Los árboles QUEST (Quick, Unbiased, Efficient Statistical Tree) - Árbol Estadístico Rápido, Insesgado y Eficiente – consiste en un algoritmo de clasificación absorbente creado específicamente para solventar dos de los principales problemas que presentaron métodos como CHAID exhaustivo y CART, a la hora de dividir un grupo de sujetos en función de una variable independiente. Este tipo de árboles mitigan la complejidad computacional y los sesgos de selección de variables. Se trata de evitar que se seleccionen aquellas variables que cuentan con un mayor número de categorías.







# TIPOS DE ARBOLES DE DECISION

QUEST intenta seleccionar el mejor predictor y su mejor punto de corte como tareas separadas, calculando en cada nodo la asociación entre cada predictor y la variable dependiente mediante el estadístico F del ANOVA de Levene para predictores continuos y ordinales o mediante una  $\chi^2$  de Pearson para predictores nominales. Se consiguen divisiones binarias de la variable dependiente a través de la creación de dos superclases en el predictor, aplicando un algoritmo conglomerativo.





# TIPOS DE ARBOLES DE DECISION

Por último, para eliminar el sesgo en la elección de variables, se elige el predictor que tiene mayor asociación con la variable dependiente. Posteriormente, para hallar el mejor punto de corte se recurre a un análisis discriminante cuadrático, repitiéndose el proceso recursivamente hasta que lo permitan las reglas de parada establecidas en el algoritmo. Se esta forma, se eliminan sesgos de respuesta y se simplifica el cálculo.





# TIPOS DE ARBOLES DE DECISION

En cuanto a la valoración de los métodos de construcción de árboles, podría establecerse un orden de jerarquía (nunca absoluto) que sitúe el método QUEST como superior a CART y este último método superior al CHAID. No olvidemos que QUEST admite métodos de validación mediante poda y permite utilizar combinaciones lineales de variables. Pero debe quedar claro que esta evaluación solo es válida en líneas generales







Universidad  
Tecnológica  
de Pereira

Portal Institucional: [www.utp.edu.co](http://www.utp.edu.co)

# ANÁLISIS MULTIVARIADO

## TECNICAS DE SEGMENTACION ARBOLES DE DECISION

### PRACTICAS EN IBM SPSS v20

[ARBOL CHAID](#)

[ARBOL CART](#)

[ARBOL QUEST](#)



Vicerrectoría Académica - Centro de Registro y Control Académico

Universidad Tecnológica de Pereira  
Acreditada Institucionalmente de Alta Calidad por el Ministerio de Educación Nacional

