Escola Tècnica Superior d'Informàtica Aplicada

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

facultad de informática. valencia

Jenui'07

Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos

R. Alcover, J. Benlloch, P. Blesa, M. A. Calduch, M. Celma, C. Ferri, J. Hernández-Orallo, L. Iniesta, J. Más, M. J. Ramírez-Quintana, A. Robles, J. M. Valiente, M. J. Vicent, L. R. Zúnica

Índice

- Objetivos
- Minería de datos
- Metodología
- Resultados ITIS
- Conclusiones y Trabajo futuro

Motivación

- Creciente atención al rendimiento
  - Factores económicos, políticos y sociales
- Proceso de convergencia Europea
  - Estrategias de adaptación
  - Evaluación y acreditación de títulos
- Mejora de la calidad del proceso educativo
- Propuesta de futuros títulos
  - Indicadores
- Programas UPV
  - Programa de Acción Tutorial
  - Plan de Acciones para la Convergencia Europea (PACE)

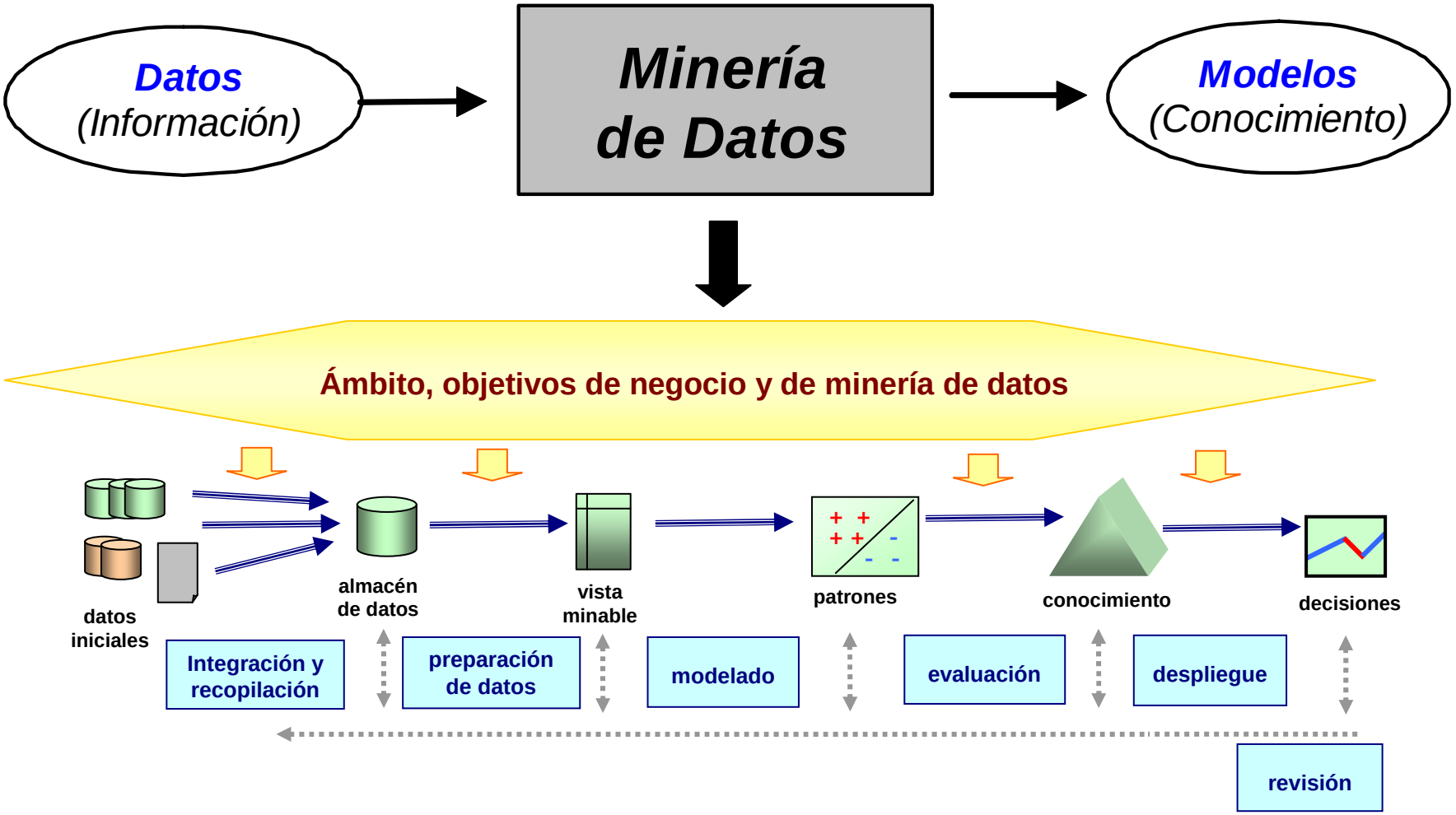
Objetivos

- Analizar la influencia de ciertos parámetros sobre el rendimiento de alumnos de nuevo ingreso en las titulaciones de informática de la UPV
- Predecir el rendimiento a partir de la información en el momento de inicio de los estudios

Minería de datos

- Son herramientas de ayuda a la toma de decisiones
- Incluyen técnicas de análisis de datos
  - Extraer conocimiento
  - “Predecir”
  - Herramientas de minería de datos permiten obtener **información que no está en la base de datos**, pero que se puede inferir de ella.

Minería de Datos



Minería de datos

- Herramienta *SPSS Clementine v.9.0*
  - Árboles de decisión: *árbol C&R*
    - Realiza particiones binarias
    - A cada hoja se le asigna un valor cercano a la media de los elementos que caen en ella
  - Regresión lineal

Metodología

- Definir la población
- Obtener la vista minable
- Elegir tipo de análisis de datos
- Generación y validación de los modelos

Metodología: Población

- Alumnos de nuevo ingreso
- Cursos: 2001,2002 y 2003
- Titulaciones: II, ITIS, ITIG de la UPV
- Nº elementos
  - II: 569 alumnos
  - ITIG: 646 alumnos
  - ITIS: 572 alumnos
- Partición de la población en dos subconjuntos disjuntos
  - Entrenamiento (77%) ⇒ 440
  - Prueba o test (23%) ⇒ 132

Metodología: Vista Minable

- Colección de individuos (alumnos)
  - Vista de la base de datos de la UPV
  - Despersonalización de datos
  - Filtrado de atributos
  - Agrupación de valores de atributos

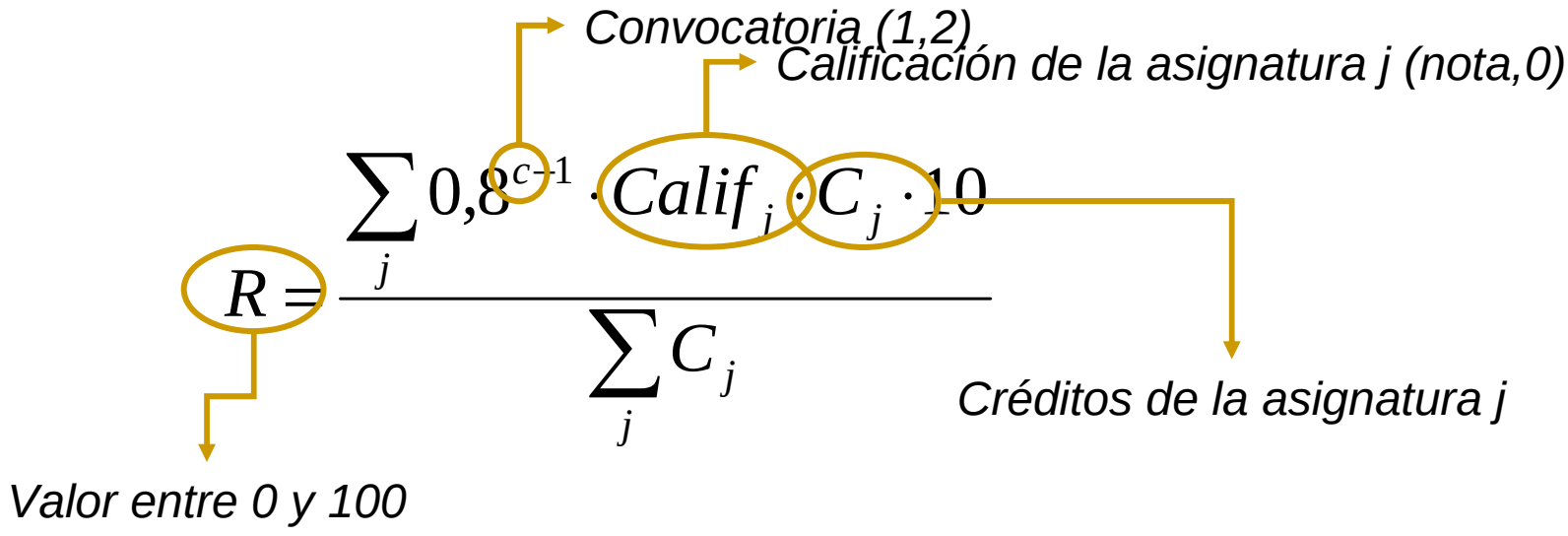
Metodología: Vista Minable

- Atributos:
  - *Ocupacio P*: Ocupación del padre
  - *Ocupacio M*: Ocupación de la madre
  - *Ocupacio A*: Ocupación del alumno
  - *Ing Nota*: Nota de acceso del alumno
  - *Ing Est*: Estudios con los que accede
  - *D_Altr Estud*: Otros estudios universitarios
  - *D_Estudis P*: Estudios del padre.
  - *D_Estudis M*: Estudios de la madre.
  - *Dpaises*: Derivado del país de nacimiento
  - *Residencia Alumno*
  - *Residencia Familia Alumno*
  - *Edad Ingreso*

Atributos derivados

Metodología: Vista Minable

- Rendimiento:



EJEMPLO. Rendimiento de un alumno en una asignatura única según calificación y convocatoria.

Calificación	Convocatoria Ordinaria	Convocatoria Extraordinaria
N.P. ó 0	0	0
2	20	16
5	50	40
9	90	72

EJEMPLO. Curso completo. 5 asignaturas en cada semestre del mismo número de créditos.

Asignatura	Febrero	Junio	Septiembre	Rendimiento
1A	NP	3	-	24
2A	2	NP	-	20
3A	7		-	70
4A	4	7	-	56
5A	NP	5	-	40
1B	-	1	NP	10
2B	-	6		60
3B	-	NP	4	32
4B	-	5		50
5B	-	NP	7	56

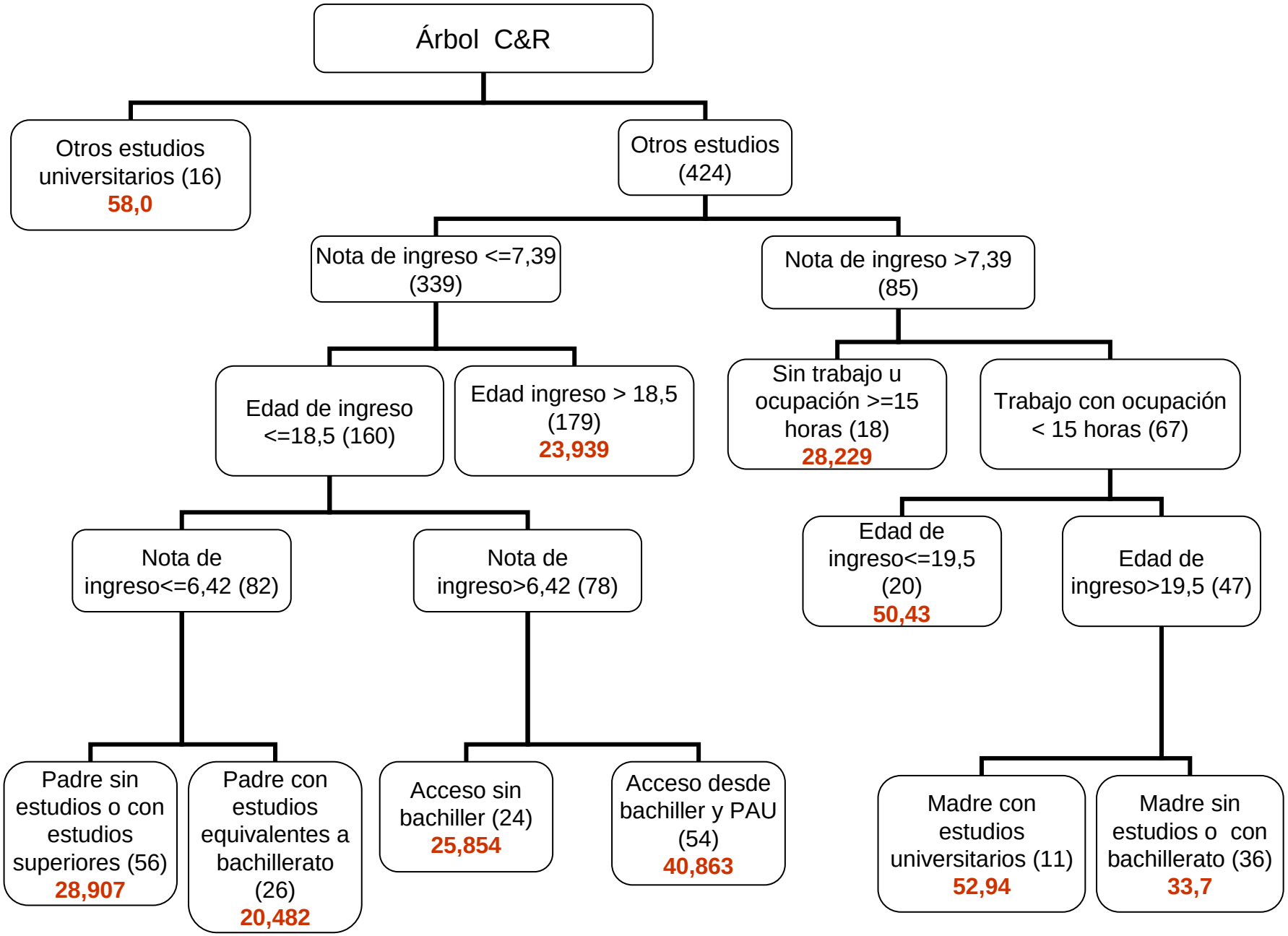
Rendimiento del alumno ese curso: **41,8**

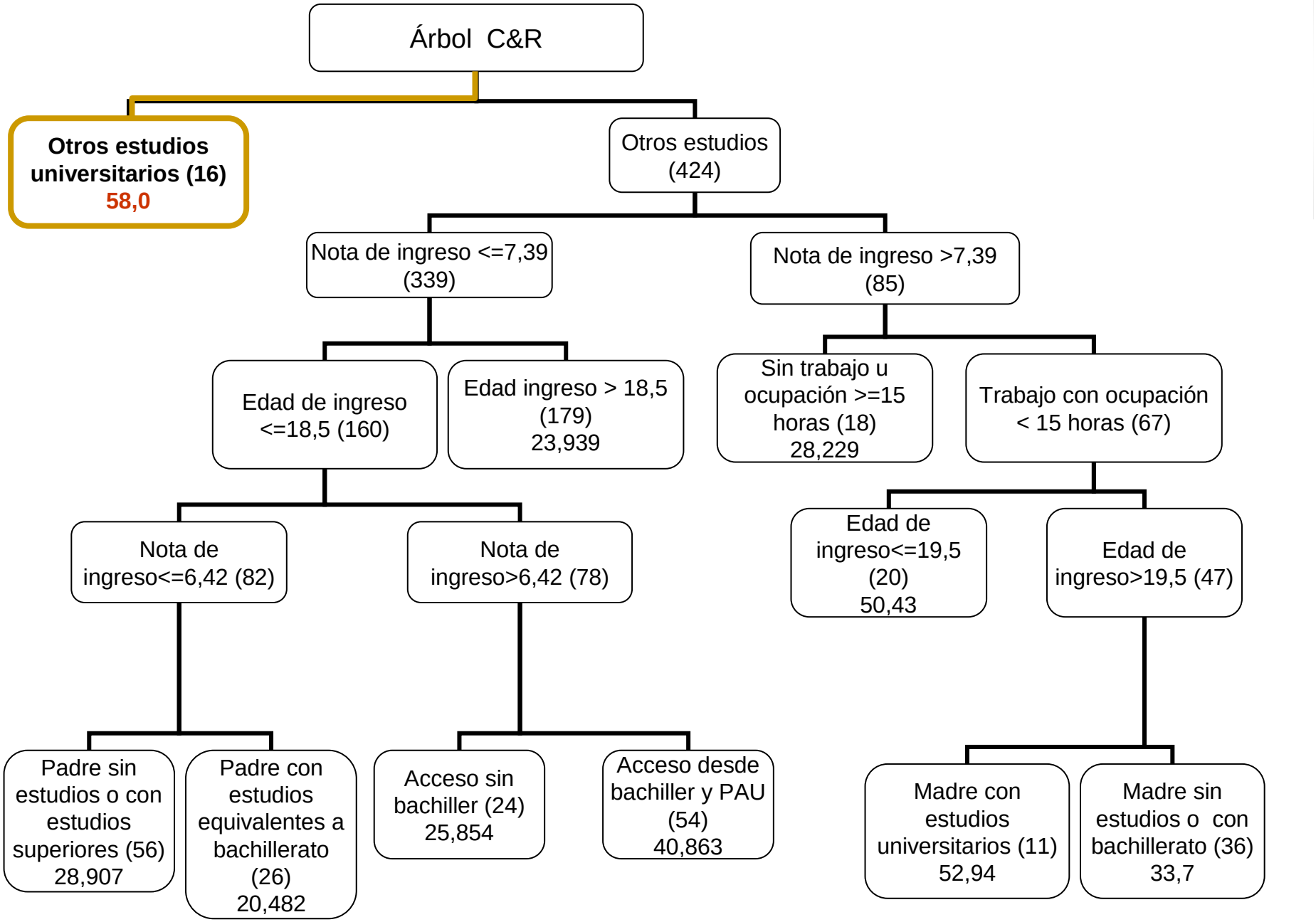
Metodología: Tipo de análisis

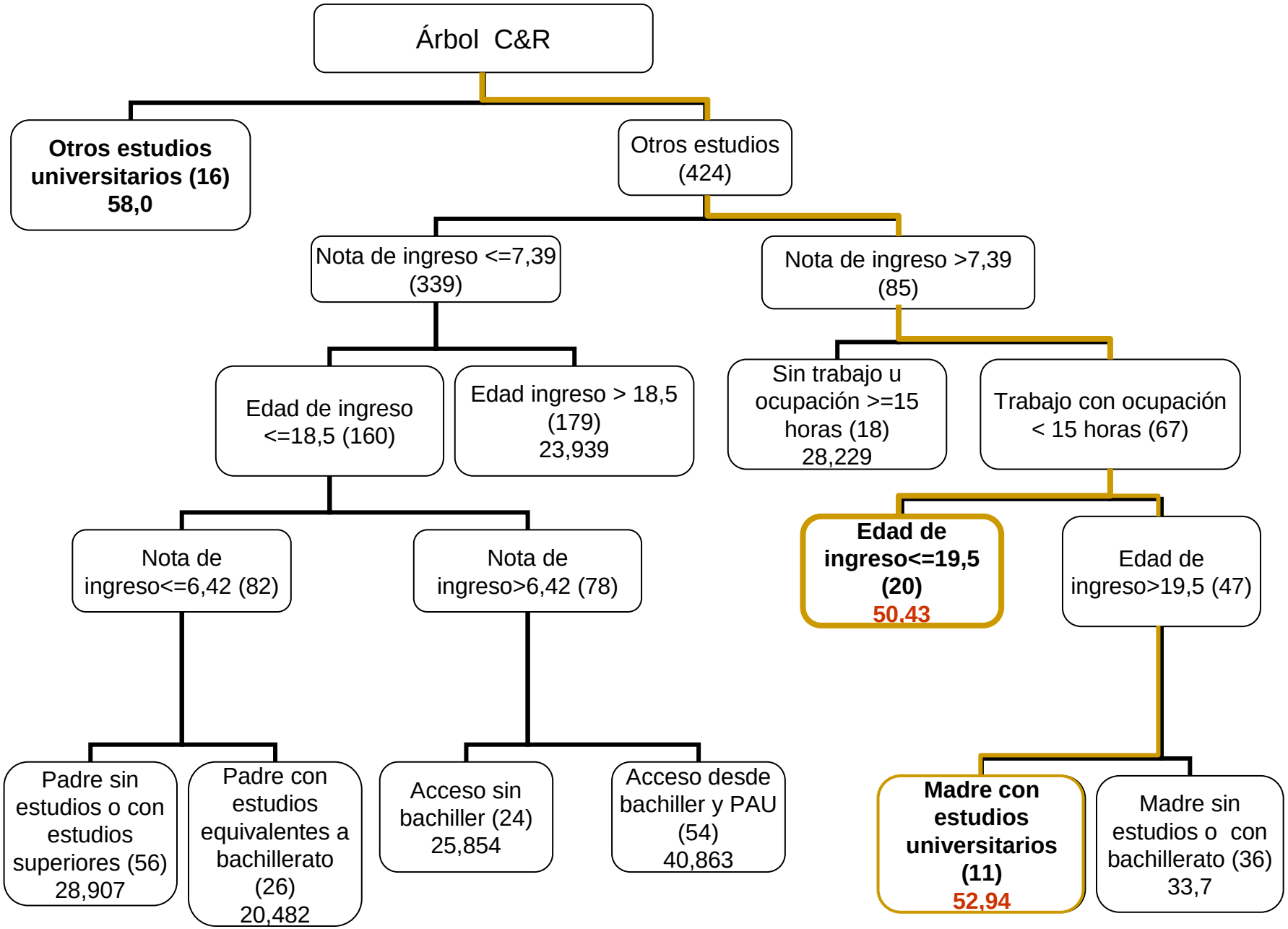
- Árboles de decisión
  - Decisiones o condiciones organizadas de forma jerárquica
  - Cada nodo representa una condición o test sobre un atributo
  - Cada rama que parte de ese nodo corresponde a un posible valor para ese atributo
  - Las hojas representan el valor de la variable predicha

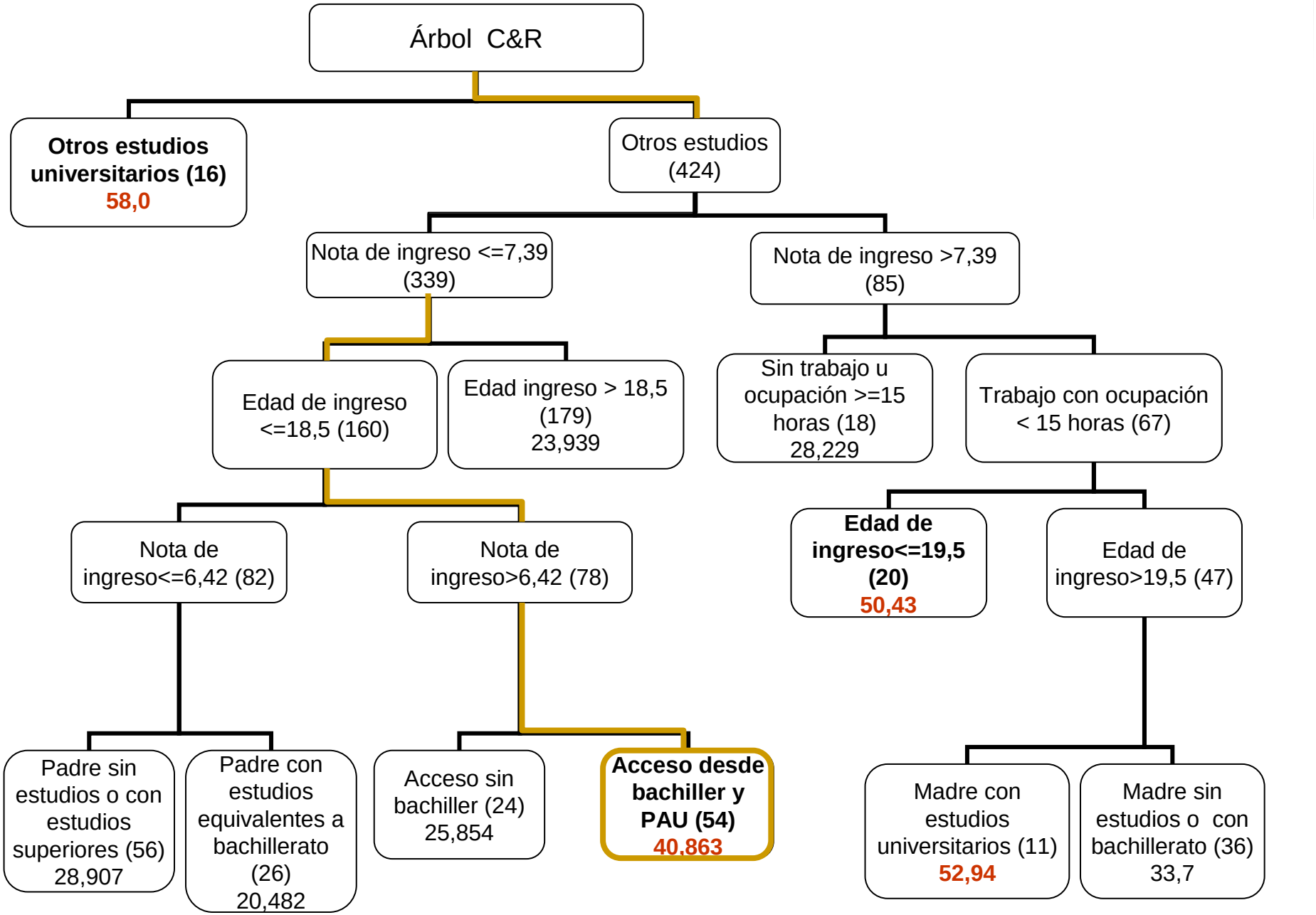
Resultados: Árbol C&R

D_Altr Estud in [1] => 58,0 (16)
D_Altr Estud in [2] (424)
Ing Nota <= 7,390 (339)
Edad Ingreso <= 18,500 (160)
Ing Nota <= 6,420 (82)
D_Estudis P in [2] => 20,482 (26)
D_Estudis P in [1 3] => 28,907 (56)
Ing Nota > 6,420 (78)
Ing Est in [2] => 25,854 (24)
Ing Est in [10] => 40,863 (54)
Edad Ingreso > 18,500 => 23,939 (179)
Ing Nota > 7,390 (85)
Ocupacio A in [1] => 28,229 (18)
Ocupacio A in [2 3] (67)
Edad Ingreso <= 19,500 => 50,43 (20)
Edad Ingreso > 19,500 (47)
D_Estudis_ M in [1 2] => 33,7 (36)
D_Estudis_ M in [3] => 52,94 (11)


Árbol C&R
Otros estudios universitarios (16)
58,0
Otros estudios (424)
Nota de ingreso <=7,39 (339)
Nota de ingreso >7,39 (85)
Edad de ingreso <=18,5 (160)
Edad ingreso > 18,5 (179)
23,939
Sin trabajo u ocupación >=15 horas (18)
28,229
Trabajo con ocupación < 15 horas (67)
Nota de ingreso<=6,42 (82)
Nota de ingreso>6,42 (78)
Edad de ingreso<=19,5 (20)
50,43
Edad de ingreso>19,5 (47)
Padre sin estudios o con estudios superiores (56)
28,907
Padre con estudios equivalentes a bachillerato (26)
20,482
Acceso sin bachiller (24)
25,854
Acceso desde bachiller y PAU (54)
40,863
Madre con estudios universitarios (11)
52,94
Madre sin estudios o con bachillerato (36)
33,7


Árbol C&R
Otros estudios universitarios (16)
58,0
Otros estudios (424)
Nota de ingreso <=7,39 (339)
Nota de ingreso >7,39 (85)
Edad de ingreso <=18,5 (160)
Edad ingreso > 18,5 (179)
23,939
Sin trabajo u ocupación >=15 horas (18)
28,229
Trabajo con ocupación < 15 horas (67)
Nota de ingreso<=6,42 (82)
Nota de ingreso>6,42 (78)
Edad de ingreso<=19,5 (20)
50,43
Edad de ingreso>19,5 (47)
Padre sin estudios o con estudios superiores (56)
28,907
Padre con estudios equivalentes a bachillerato (26)
20,482
Acceso sin bachiller (24)
25,854
Acceso desde bachiller y PAU (54)
40,863
Madre con estudios universitarios (11)
52,94
Madre sin estudios o con bachillerato (36)
33,7


Árbol C&R
Otros estudios universitarios (16) 58,0
Otros estudios (424)
Nota de ingreso <=7,39 (339)
Nota de ingreso >7,39 (85)
Edad de ingreso <=18,5 (160)
Edad ingreso > 18,5 (179) 23,939
Sin trabajo u ocupación >=15 horas (18) 28,229
Trabajo con ocupación < 15 horas (67)
Nota de ingreso<=6,42 (82)
Nota de ingreso>6,42 (78)
Edad de ingreso<=19,5 (20) 50,43
Edad de ingreso>19,5 (47)
Padre sin estudios o con estudios superiores (56) 28,907
Padre con estudios equivalentes a bachillerato (26) 20,482
Acceso sin bachiller (24) 25,854
Acceso desde bachiller y PAU (54) 40,863
Madre con estudios universitarios (11) 52,94
Madre sin estudios o con bachillerato (36) 33,7


Árbol C&R
Otros estudios universitarios (16)
58,0
Otros estudios (424)
Nota de ingreso <=7,39 (339)
Nota de ingreso >7,39 (85)
Edad de ingreso <=18,5 (160)
Edad ingreso > 18,5 (179)
23,939
Sin trabajo u ocupación >=15 horas (18)
28,229
Trabajo con ocupación < 15 horas (67)
Nota de ingreso<=6,42 (82)
Nota de ingreso>6,42 (78)
Edad de ingreso<=19,5 (20)
50,43
Edad de ingreso>19,5 (47)
Padre sin estudios o con estudios superiores (56)
28,907
Padre con estudios equivalentes a bachillerato (26)
20,482
Acceso sin bachiller (24)
25,854
Acceso desde bachiller y PAU (54)
40,863
Madre con estudios universitarios (11)
52,94
Madre sin estudios o con bachillerato (36)
33,7

Resultados: Árbol C&R

Atributo	I	IR	IR2
D_Altr Estud	440	1	0,255
Ing Nota	584	1,327	0,339
Edad Ingreso	406	0,922	0,236
D_Estudis P	82	0,186	0,047
D_Estudis M	47	0,106	0,027
Ocupacio A	85	0,193	0,049
Ing Est	78	0,177	0,045
TOTAL	**1722**	**3,97**	**1**

Importancia Relativa normalizada

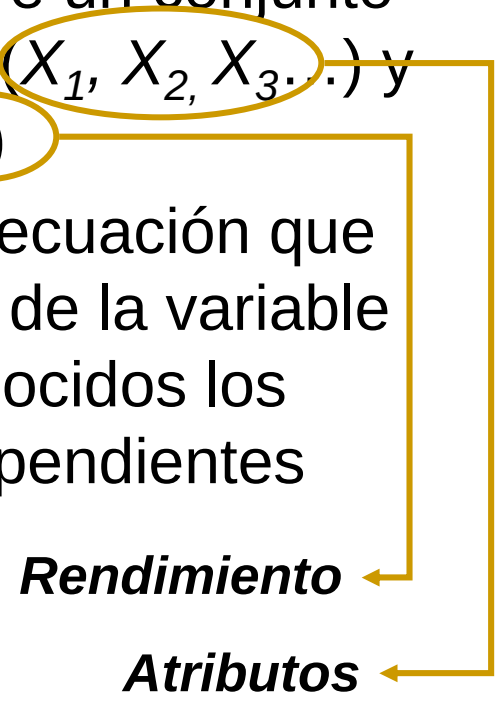
- Utilización en un 5% de las decisiones (IR2>0.05)

Resultados: Árbol C&R

- Mejor rendimiento
  - Alumnos con estudios universitarios previos (58,0)
  - Nota de entrada alta, menor edad y sin trabajo (50,43)
- Peor rendimiento
  - Nota de entrada baja
  - Sin bachiller LOGSE

Metodología: Tipo de análisis

- Regresión lineal multivariante
  - Es un método que nos permite establecer una relación matemática entre un conjunto de variables independientes (X_1, X_2, X_3...) y una variable dependiente (Y)
  - El objetivo es encontrar una ecuación que nos permita predecir el valor de la variable dependiente (Y) una vez conocidos los valores de las variables independientes



Resultados: Regresión lineal

Rendimiento =

D_Altr Estud_1 (21)· 35,55 +
D_Estudis M_3 (140)· 4,063 +
Ing Est_10 (248)· 9,673 +
Ing Est_4 (3)· -36,54 +
Ing Nota (572)· 7,572 +
Ocupacio M_8 (73)· 6,573 +
-27,52

Resultados: Regresión lineal

- Mejor rendimiento
  - Alumnos con estudios universitarios previos (35,6)
  - Bachiller LOGSE con PAU (9,7)
  - Madre con estudios superiores (4,1) y trabajo no remunerado (6,6)

Resultados

$$RECM = \sqrt{ECM} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - R_i)^2}{n}}$$

R_i: Rendimiento real

y_i: Rendimiento estimado

n: Tamaño del conjunto de datos

- RECM Árbol C&R
  - 17,95
  - Datos de test
- RECM Regresión
  - 17,45
  - Datos de test
- Total de alumnos
  - Rendimiento medio: 30,7
  - Desviación típica: 19,6

Conclusiones

- Posibilidad de determinar los factores determinantes del rendimiento de los alumnos de nuevo ingreso
- Posibilidad de predecir el rendimiento

Trabajo futuro

- Extender el análisis a otros cursos
- Utilización de diferentes indicadores:
  - Abandono, duración de los estudios…
- Análisis por asignaturas o áreas
- Ampliación del estudio a otras universidades
  - Proyecto del Ministerio