

RENDIMIENTO DE LOS ESTUDIANTES EN MATEMÁTICA I DE LA UDONE UTILIZANDO EL ALGORITMO II DE SEGMENTACIÓN

ACADEMIC PERFORMANCE OF STUDENTS WHO TAKE THE COURSE IN MATHEMATICS I AT UDONE, USING A SEGMENTATION ALGORITHM II.

EDUARDO HERNÁNDEZ, NELSON BRACHO Y MICHEER LUNA

Departamento de Estadística, Núcleo de Nueva Esparta, Universidad de Oriente.

RESUMEN

Se tomó una muestra de los estudiantes que ingresaron a la Universidad de Oriente Núcleo de Nueva Esparta (UDONE), en el segundo semestre de 1996 y el primero de 1997, en el Departamento de Control de Estudios de la Universidad, con el fin de estudiar el Rendimiento Académico en Matemática I, a través del Algoritmo II del Análisis de Segmentación Descendente basado en Pruebas de Hipótesis de Independencia Condicional. Se consideró como variable respuesta el Rendimiento de los Estudiantes en Matemática I, y como variables explicativas: Sexo, Promedio de Notas de Bachillerato en Matemática, Tipo de Bachiller, Procedencia y Tipo de Liceo. Resultó que las variables explicativas Sexo, Procedencia y Tipo de Liceo no aportaron información a la variable respuesta, mientras que las variables Tipo de Bachiller y Promedio de Notas en Matemática de Bachillerato sí aportaron al Rendimiento Académico en Matemática I.

PALABRAS CLAVES: Segmentación, árbol, rendimiento, independencia condicional.

ABSTRACT

For this study, the Department of Academic Control gave us a sampling of the students who entered the University of Orient in the state of Nueva Esparta (UDONE), Venezuela, during the second semester of 1996 and the first of 1997, in order to analyze Academic Performance in Mathematics I. We used an Algorithm II of Descending Segmentation Analysis based on Tests of Hypothesis of Conditional Independence. Student Performance in Mathematics I was considered as the variable response, while Sex, Average Mathematics Grades in High School, High School Graduate Type, Origin and School Type were considered as explanatory variables. The results show that Sex, Origin and School Type did not contribute any information to the variable response, but Average Mathematics Grades in High School and High School Graduate Type did.

KEY WORDS: Segmentation, tree, performance, conditional independence.

INTRODUCCIÓN

El Análisis de Segmentación se utiliza con fines exploratorios y descriptivos, con el objetivo básico de clasificar la población en grupos que describan de la mejor manera posible la variable respuesta; este análisis se basa en los métodos de Detección Automática de la Interacción (DAI), propuestos por Morgan y Sonquist (1963), y permite conocer qué variables son útiles para describir la variable respuesta, qué categorías de un predictor (variable explicativa) es homogénea respecto a la variable respuesta y cuál es el efecto conjunto de dos o más predictores.

Esta situación particular hace que en principio, sólo se esté interesado en aquellas hipótesis en las cuales está intrínseca la información relacionada con la variable respuesta, ya que en este tipo de análisis lo que se pretende

es encontrar segmentos homogéneos de la población, que sirvan para describir la relación existente entre las variables explicativas y la variable respuesta.

La significación de cada predictor se obtiene con el valor, p , correspondiente al estadístico de prueba χ^2 (Ji-Cuadrado). Pero se debe considerar, que primero se procede a la agrupación de categorías homogéneas en relación a la variable respuesta; luego el cálculo de la significación de cada predictor debería ser corregido tomando en cuenta el proceso de agrupación previo.

Dado que no es recomendable la utilización de distribuciones marginales, Ávila (1996) propone la utilización de hipótesis de independencia condicionada de la forma siguiente:

$$i \perp j / V \setminus \{j\}$$

Recibido: Junio 2000. Aprobado: Mayo 2001.
Versión final: Mayo 2001.

Donde lo anterior significa, i es independiente de j dado el conjunto V de todos los predictores omitiendo el predictor j .

MATERIALES Y MÉTODOS

Los datos fueron recopilados en el Departamento de Control de Estudios de la Universidad de Oriente Núcleo de Nueva Esparta; con el objeto de obtener información sobre el Rendimiento Académico en Matemática I de los Estudiantes Nuevos. La muestra consta de 847 registros de estudiantes que ingresaron, en el segundo semestre de 1996 y el primero de 1997 a la Universidad.

En el análisis se dispone de 6 factores, todos categóricos, se consideró como variable respuesta el Rendimiento de los Estudiantes en Matemática I, y como variables explicativas: Sexo, Promedio de Notas de Bachillerato en Matemática, Tipo de Bachiller, Procedencia y Tipo de Liceo.

La técnica que se aplicó es el Algoritmo II de Segmentación Descendente, basado en la Hipótesis de Independencia Condicionada, propuesto por Avila (1996); este se divide en 5 etapas:

1. Se parte de un árbol inicial en el cual se consideran todas las variables explicativas con todas sus categorías en el estudio.
2. Búsqueda de las variables a colapsar. Se colapsa el árbol en todas aquellas variables para las cuales se acepta la hipótesis de independencia condicionada. Supóngase que se acepta la hipótesis de independencia condicionada para las variables $i \in m / V \setminus \{n\}$, se colapsa el árbol en la variable m y se conservan las variables $V \setminus \{m, n\}$ como predictoras.
3. Se repite la etapa 2 con las variables restantes hasta que el árbol no se pueda colapsar más.
4. Búsqueda de la variable para colapsar en subgrupo:
 - a) Se agrupa la información de aquellas ramas del árbol en las que se presentan rasgos similares.
 - b) Se toma la variable que aun siendo significativa proporciona el valor p más grande, ya que es probable encontrar subgrupos en los cuales se puede colapsar.
 - c) Si se acepta la hipótesis de independencia

condicionada, se colapsa el árbol en este subgrupo.

- d) Se repite el paso anterior hasta que no queden subgrupos a colapsar, momento en el cual el análisis se da por finalizado.
5. Se hace la representación del análisis, a través de un mosaico visual y se sacan las conclusiones pertinentes.

RESULTADOS Y DISCUSIÓN

Siguiendo las etapas del Algoritmo II de Segmentación Descendente, se procede a hacer el análisis para el Rendimiento Académico en Matemática I. En la primera etapa se obtuvo el Árbol General de Clasificación mostrado en la figura 1.

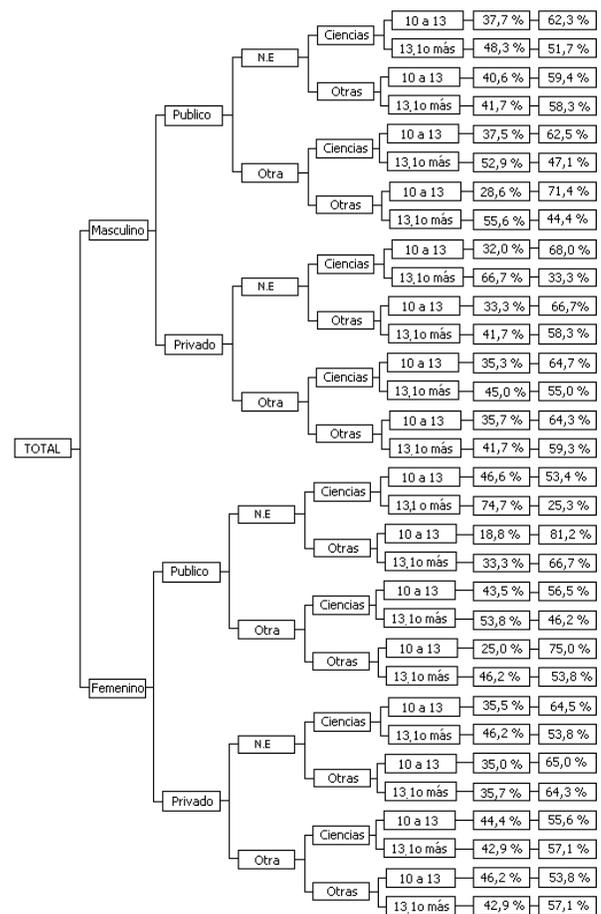


Fig. 1: Árbol General con todas las variables que entran en el estudio, con sus respectivos porcentajes de Aprobados y Reprobados.

Las variables respuesta y explicativas utilizadas en el estudio se muestran en la tabla 1, con sus respectivas categorías.

Tabla 1: Clasificación de las variables que se tomaron en cuenta para el estudio.

Tipo de Variable	Variable	Notación	Categorías
Respuesta	• Rendimiento en Matemática I	i	Aprobados Reprobados
Explicativa	• Sexo	j	Masculino Femenino
Explicativa	• Tipo de Liceo	k	Público Privado
Explicativa	• Lugar de Procedencia	l	Nueva Esparta Otra Región
Explicativa	• Tipo de Bachiller	m	Ciencias Otra Mención
Explicativa	• Promedio de Notas en Matemáticas de Bachillerato.	n	10 a 13 13,1 o más

Fuente: Elaboración propia.

El Promedio de Notas de Matemáticas de Bachillerato se clasificó en dos categorías, debido a que el análisis de segmentación utiliza tablas de contingencias, y no es recomendable que las celdas de estas tablas al hacer los

contrastes individuales estén vacías, ya que de ser así pueden arrojar resultados erróneos.

En la Etapa 2 se procedió a contrastar las hipótesis de independencia condicionada:

Tabla 2: Contrastes de Independencia Condicional utilizando el Algoritmo II de Segmentación Descendente.

Etapa	Contrastes	χ^2	GL	p
2	Rendimiento \perp Sexo / Tipo de Liceo, Procedencia, Tipo de Bachiller, Promedio de Notas de Bachillerato en Matemáticas.	28,95	16	0,0243
	Rendimiento \perp Tipo de Liceo / Sexo, Procedencia, Tipo de Bachiller, Promedio de Notas de Bachillerato en Matemáticas.	16,62	16	0,4103
	Rendimiento \perp Tipo de Bachiller / Sexo, Tipo de Liceo, Procedencia, Promedio de Notas de Bachillerato en Matemáticas.	29,95	16	0,0183
	Rendimiento \perp Procedencia / Sexo, Tipo de Liceo, Tipo de Bachiller, Promedio de Notas de Bachillerato en Matemáticas.	9,10	16	0,9226
	Rendimiento \perp Promedio de Notas de Bachillerato en Matemáticas / Sexo, Tipo de Liceo, Procedencia, Tipo de Bachiller.	28,62	16	0,0266
3.1	Rendimiento \perp Promedio de Notas de Bachillerato en Matemáticas / (Sexo, Tipo de Bachiller).	19,86	4	0,0005
	Rendimiento \perp Sexo / (Tipo de Bachiller, Promedio de Notas de Bachillerato en Matemáticas).	6,43	4	0,1693
	Rendimiento \perp Tipo de Bachiller / (Sexo, Promedio de Notas de Bachillerato en Matemáticas).	17,90	4	0,0001
3.2	Rendimiento \perp Tipo de Bachiller / Promedio de Notas.	12,00	2	0,0025
	Rendimiento \perp Promedio de Notas / Tipo de Bachiller	20,89	2	0,0000
4	Promedio de Notas	21,36	1	0,0245
	Tipo de Bachiller	12,29	1	0,0005

Fuente: Elaboración propia.

$i \downarrow j / V \setminus \{ j \} = \text{Rendimiento} \downarrow \text{Sexo} / (\text{Tipo de Liceo, Procedencia, Tipo de Bachiller, Promedio de Notas de Bachillerato en Matemáticas}).$

$i \downarrow k / V \setminus \{ k \} = \text{Rendimiento} \downarrow \text{Tipo de Liceo} / (\text{Sexo, Procedencia, Tipo de Bachiller, Promedio de Notas de Bachillerato en Matemáticas}).$

$i \downarrow l / V \setminus \{ l \} = \text{Rendimiento} \downarrow \text{Procedencia} / (\text{Sexo, Tipo de Liceo, Tipo de Bachiller, Promedio de Notas de Bachillerato en Matemáticas}).$

$i \downarrow m / V \setminus \{ m \} = \text{Rendimiento} \downarrow \text{Tipo de Bachiller} / (\text{Sexo, Tipo de Liceo, Procedencia, Promedio de Notas de Bachillerato en Matemáticas}).$

$i \downarrow n / V \setminus \{ n \} = \text{Rendimiento} \downarrow \text{Promedio de Notas de Bachillerato en Matemáticas} / (\text{Sexo, Tipo de Liceo, Procedencia, Tipo de Bachiller}).$

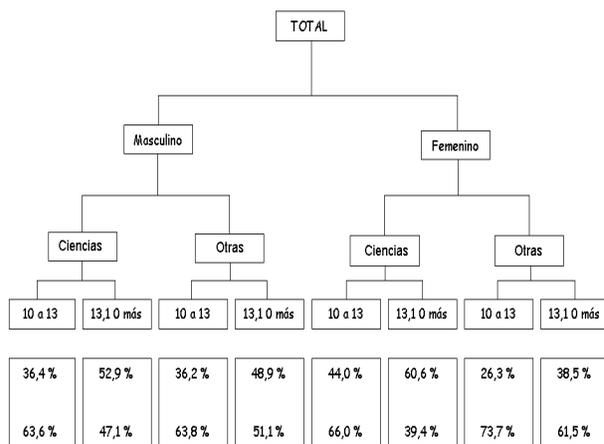


Fig. 2: Árbol al segmentar las variables Tipo de Liceo y Procedencia (% Aprobados y Reprobados).

El resumen de los contrastes de independencia condicionada de esta etapa y los correspondientes a las etapas 3 y 4 aparecen en la tabla 2.

En la etapa 2 se colapsaron las variables Tipo de Liceo y Procedencia, ya que son las variables que tienen el contraste con el $p > 0,05$, esto quiere decir que el Rendimiento Académico en Matemática I no depende del Tipo de Liceo y la Procedencia del Estudiante, dando como resultado el árbol de clasificación mostrado en la figura 2.

En la etapa 3.1, se repitió la etapa 2 al árbol de la figura 2. En este caso el contraste para la variable Sexo tiene un valor $p > 0,05$, lo cual significa que el Rendimiento es independiente de la variable Sexo, por lo que se procedió a

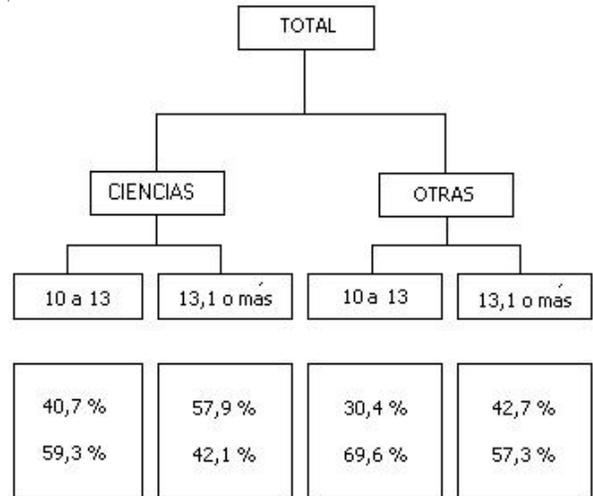


Figura 3: Árbol al Segmentar la Variable Sexo. (% de Aprobados y Reprobados).

agrupar el árbol en esta variable, dando como resultado el árbol de la figura 3.

En la etapa 3.2 se repitió la etapa 2 al árbol de la figura 3, los contrastes correspondientes resultaron altamente significativos ($p < 0,01$), esto implica que no se puede colapsar el árbol en estas variables, debido a que el Rendimiento es dependiente de las variables Tipo de Bachiller y Promedio de Notas, permaneciendo el mismo árbol de la figura 3.

En la etapa 4, se buscan las variables que aun siendo significativas es probable encontrar categorías dentro de

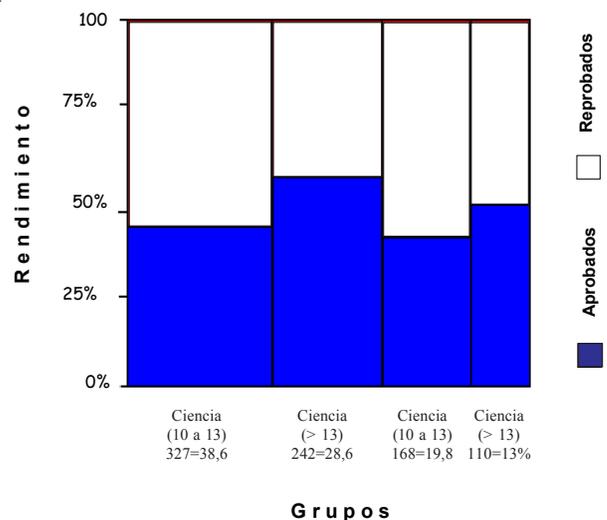


Fig. 4: Mosaico resultante del análisis de segmentación.

ellas que se puedan agrupar. Los contrastes individuales de las variables Tipo de Bachiller y Promedio de Notas resultaron significativos ($p < 0,03$), en consecuencia no se pudo agrupar las categorías de estas variables y el árbol final del Análisis de Segmentación es el mismo de la figura 3.

La representación del Mosaico Visual resultante del Análisis de Segmentación se muestra en la figura 4.

CONCLUSIONES

El porcentaje general de estudiantes que reprueban Matemática I es mayor con un 57,1% que los que aprueban con un 42,9%. Las variables Procedencia, Tipo de Liceo y Sexo no influyen en el Rendimiento, mientras que Tipo de Bachiller y Promedio de Notas de Bachillerato si lo hacen. El grupo de estudiantes con el mayor porcentaje de

aprobación (57,9%), es el de bachilleres en ciencias con un promedio de notas de bachillerato en Matemáticas mayor de 13,1 puntos; y el grupo con el menor porcentaje (30,4%) de aprobación es el de otras menciones con promedio menor de 13,1 puntos.

REFERENCIAS BIBLIOGRÁFICAS

- ÁVILA, C. 1996. Una Alternativa al Análisis de Segmentación Basada en el Análisis de Hipótesis de Independencia Condicionada. Tesis Doctoral del Departamento de Estadística y Matemática Aplicada. Universidad de Salamanca, España.
- MORGAN, J. AND SONQUIST J. 1963. Problems in the analysis of survey data and a proposal. J. Amer. Statistical Assoc. 58: 415-434.

