

Desempeño educativo de los estudiantes de la Licenciatura en Economía

Daniel Czarniewicz & Mauro Gopar & Romina Quagliotti

2018

Índice

Abstract	1
Introducción	2
Los datos	2
Estadísticas descriptivas de la población	4
Metodología	9
Análisis de cluster	9
Análisis discriminante	10
Resultados obtenidos	11
Análisis de cluster	11
Selección del método	13
Caracterización de los grupos	15
Análisis discriminante	17
Conclusiones	19
Anexo	20
Complete Linkage	20
Average Linkage	20
k-medoides	21
Ward	22
Referencias	24

Abstract

En el presente trabajo se buscó estudiar el desempeño educativo de los estudiantes de la Licenciatura en Economía mediante la aplicación de técnicas de análisis de clusters y de análisis discriminante. Para ello se utilizaron datos provenientes del Sistema de Gestión de Bedelías y el Formulario Estadístico de ingreso a la Universidad correspondientes a las generaciones de ingreso 2012 y 2013. Se pudo confirmar la existencia de una estructura de tres grupos: estudiantes con desempeño bajo, medio, y alto. Así mismo se encontró que la conformación de dichos grupos se encuentra explicada, en parte, por la edad de los estudiantes, su situación laboral, la procedencia de los mismos de instituciones de educación secundaria privada, y la cantidad de carreras a la que están inscriptos. No se encontró evidencia de que el nivel educativo alcanzado por los padres contribuya al buen desempeño de sus hijos en la carrera.

Introducción

Este trabajo tiene como objetivo caracterizar el desempeño de los estudiantes de la Licenciatura en Economía de la Facultad de Ciencias Económicas y Administración (FCEA) al cabo de sus primeros cuatro años. Se utilizan técnicas de análisis de cluster para formar grupos de individuos según indicadores de su avance, y su escolaridad, y técnicas de análisis discriminante para determinar cuáles variables sociodemográficas contribuyen a explicar la pertenencia de cada individuo a su grupo. Para ellos trabajamos con la información de las actividades de los estudiantes y los formularios estadísticos de los mismos.

El trabajo consta de las siguientes partes. Luego de esta introducción se describen los datos utilizados para el análisis y se brinda una caracterización de toda la población. A continuación se explican las metodologías utilizadas para la conformación de los grupos. Seguido de esto, se presentan los principales resultados obtenidos tanto del análisis de clusters como del análisis dicriminante. Por última se ofrecen algunas conclusiones.

Los datos

La muestra está compuesta por 595 estudiantes que se inscribieron en la Licenciatura en Economía e ingresaron a FCEA en los años 2012 y 2013. Esta elección se debe a que contamos con datos hasta el año 2016, por lo cual, únicamente disponemos de datos para completar un período de cuatro años para estas dos generaciones. Cabe aclarar que la elección de la duración del análisis no es aleatoria ni subjetiva, sino que se desprende de la duración teórica de la carrera bajo estudio. No se anaizan generaciones anteriores dado que en 2012 comenzó a implementarse el nuevo plan de estudios para la carrera de economía, por lo cual, los resultados entre estudiantes de distintos planes no serían comparables. Adicionalmente, el interés planteado por los autores consiste en analizar el desempeño de los estudiantes bajo el plan actual.

Se utilizaron dos bases de datos. La primera contiene todas las actividades (reválidas, e inscripción y resultados de cursos y exámenes) realizadas por los estudiantes entre los años 2012 y 2016. Esta se obtuvo del Sistema de Gestión de Bedelías de la Universidad de la República (UdelaR). La segunda base de datos con que se trabajó corresponde a la utilizada en Arim et al. (2016). La misma contiene las siguientes variables que son de nuestro interés:

- **gen**: año en que ingresó a FCEA
- **cred1**: créditos aprobados en el primer año
- **cred1y2**: créditos acumulados en el primer y segundo año
- **cred1.2y3**: créditos acumulados en el primer, segundo y tercer año
- **cred1234**: créditos acumulados en los primeros cuatro años
- **sit_labNT**: situación laboral de los que no trabajaban al momento de ingreso a facultad
- **horas**: cantidad de horas que trabajaba al momento de ingreso a facultad
- **sexto**: toma los valores “Montevideo privada”, “Montevideo pública”, “Interior privada”, “Interior pública”, “UTU Montevideo”, “UTU Interior”, e indica en qué tipo de institución el estudiante realizó su último año de secundaria
- **edad** al momento de ingreso a la facultad
- **Xfem**: binaria que toma el valor 1 si es mujer

- **Xhijos**: binaria que toma el valor 1 si tiene hijos
- **Xocup**: binaria que toma el valor 1 si trabaja
- **XeduP**: variable categórica que refleja el máximo nivel educativo del padre: bajo, medio y alto. El nivel educativo bajo incluye desde primaria incompleta hasta educación media incompleta (inclusive), el nivel educativo medio incluye hasta educación terciaria/universitaria incompleta (inclusive), y el nivel educativo alto incluye educación terciaria/universitaria completa o más
- **XeduM**: ídem **XeduP** pero relativo a la madre del estudiante
- **carr11**: binaria que toma el valor 1 si el individuo está inscripto a la carrera de Contador Público (además de en la Licenciatura en Economía)
- **carr14**: binaria que toma el valor 1 si el individuo está inscripto a la carrera de Licenciatura en Administración (además de en la Licenciatura en Economía)
- **EDAoMCEDA**: binaria que toma el valor 1 si el individuo en algún momento de su vida académica se inscribió en algún curso de la Escuela de Administración (EDA) o cursó la asignatura métodos cuantitativos (Cálculo I) a través de la EDA

A partir de la base de actividades se crearon las siguientes variables para medir el desempeño de los estudiantes:

- **esco**: promedio de notas de los estudiantes en sus primeros cuatro años
- **esco.ec**: promedio de notas en materias del área economía de los estudiantes en sus primeros cuatro años
- **esco.mc**: ídem **esco.ec** pero relativo al área de métodos cuantitativos
- **prop.aprob.exam**: proporción de las aprobaciones en FCEA (sin reválidas) que el estudiante rindió por examen

A partir de la base de información sociodemográfica se crearon las siguientes variables:

- **cred2**: cantidad de créditos aprobados en el segundo año
- **cred3**: cantidad de créditos aprobados en el tercer año
- **cred4**: cantidad de créditos aprobados en el cuarto año
- **sexto.priv**: binaria que toma el valor 1 si el estudiante cursó sexto en una institución privada
- **sexto.mvd**: binaria que toma el valor 1 si el estudiante cursó sexto en una institución de Montevideo
- **busca.trabajo**: binaria que toma el valor 1 si el estudiante buscaba trabajo al momento de su ingreso a FCEA (a partir de `sit_labNT`)

A su vez, ya que se busca estudiar el desempeño de estudiantes que están cursando la Licenciatura en Economía, se descartaron aquellos estudiantes que se encontraban inscriptos a la Licenciatura en Economía y a alguna otra carrera de FCEA y que, por sus actividades, podía asumirse que están realizando la otra carrera y no la Licenciatura en Economía. Para ello, se definió que un estudiante estaba cursando la carrera de Contador Público y no la Licenciatura en Economía si: a) estaba inscripto en Contador Público (**carr11** = 1), y b) en los cuatro años analizados había aprobado menos de 9 materias obligatorias de la Licenciatura en Economía y más de 4 del área de contabilidad (estos indicadores se calcularon a partir de la base de actividades). Por otro lado, se definió que un estudiante se encontraba cursando la Licenciatura en Administración y no la de Economía si: a) estaba inscripto en esa carrera (**carr14** = 1), y b) había aprobado, en los cuatro años,

menos de 12 cursos obligatorios de la Licenciatura en Economía y más de 6 del área de administración. Estos límites se definieron en base a a cantidad de materias que tienen en común las carreras, y la cantidad de cursos opcionales de estas áreas que hay en la Licenciatura en Economía.

Estadísticas descriptivas de la población

Se espera que se desprenda del análisis una estructura de grupos que presente una diferenciación clara entre estudiantes con desempeño bajo y alto, y un tercer grupo heterogeneo con desempeños que llamaremos medios. A su vez, se espera también que el estudiantes provenientes de liceos privados de Montevideo se encuentren concentrados dentro del grupo de estudiantes con alto rendimiento. Adicionalmente, se espera verificar que el desempeño de los estudiantes esté vinculado con el nivel educativo alcanzado por sus padres. Por último, se espera que no haya conclusiones claras respecto a la relación entre la situación laboral de los estudiantes tal como fuera relevada en el Formulario Estadístico y el desempeño de los mismos. Esto se debe a que dicha información corresponde únicamente al momento de ingreso a facultad, mientras que el desempeño se evalúa a lo largo de cuatro años.

A continuación se presentan algunas estadísticas descriptivas y gráficos vinculados a las variables de análisis. Se trabajó con 316 estudiantes de la generación 2012 y 279 estudiantes de la generación 2013. Por otra parte, 316 observaciones corresponden a hombres, mientras que 279 corresponden a mujeres.

Cuadro 1: Sexo según generación de ingreso

	2012	2013
Hombres	163	153
Mujeres	153	126

En la Figura 1 puede apreciarse la distribución de créditos acumulados al cabo de los primeros cuatro años de carrera, distinguiendo a los estudiantes según su liceo de procedencia. Vemos que existe una alta proporción de estudiantes con una baja cantidad de créditos al final del período de análisis. A su vez, el gráfico presenta indicios de que la distribución de estudiantes provenientes de liceos privados de Montevideo, presentan una cola superior más pesada que la distribución de los demás grupos.

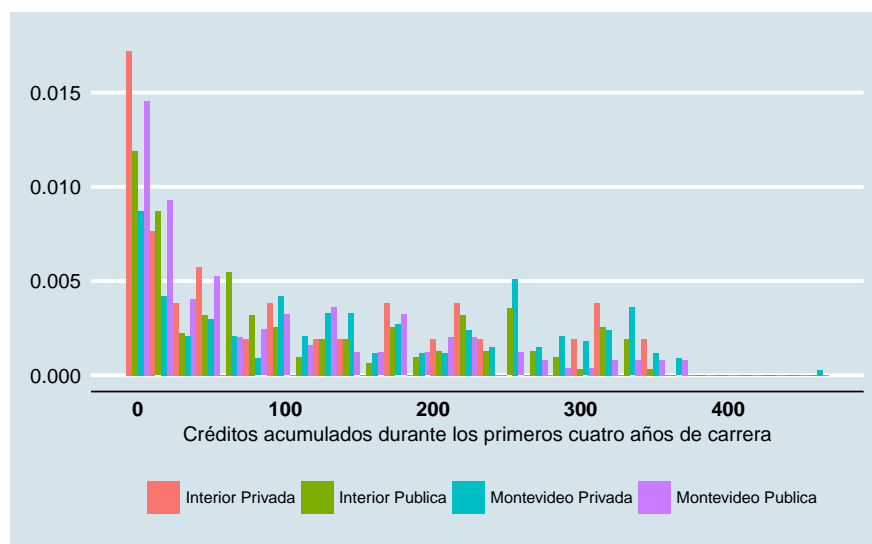


Figura 1: Créditos acumulados al cabo de cuatro años según sexto.

De igual forma, en la Figura 2 puede apreciarse que la distribución de estudiantes provenientes de liceos privados de Montevideo presenta una asimetría izquierda sugiriendo entonces que estos se concentran en las notas más altas.

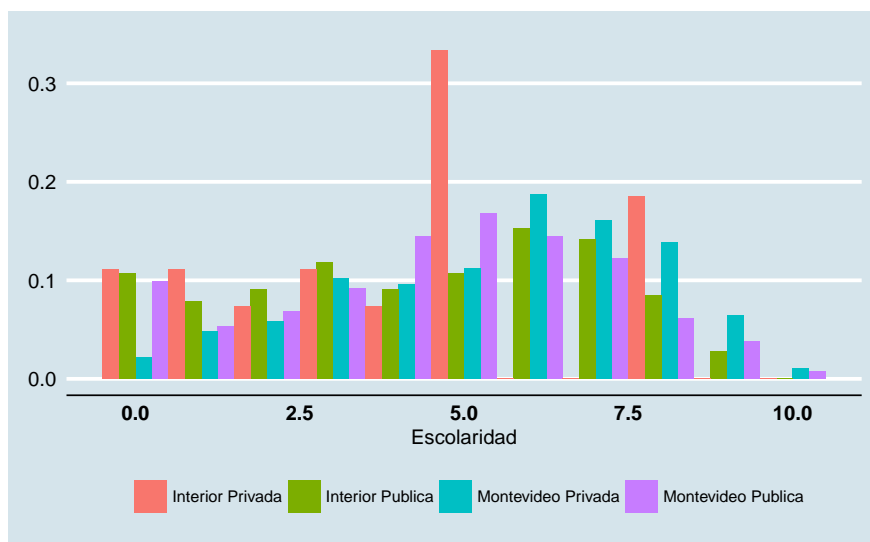


Figura 2: Escolaridad según sexto.

Por su parte, la Figura 3 muestra la relación entre el avance (medido en créditos) y la escolaridad para los estudiantes de mayor avance mientras que no existe una relación clara para los estudiantes de menor avance.

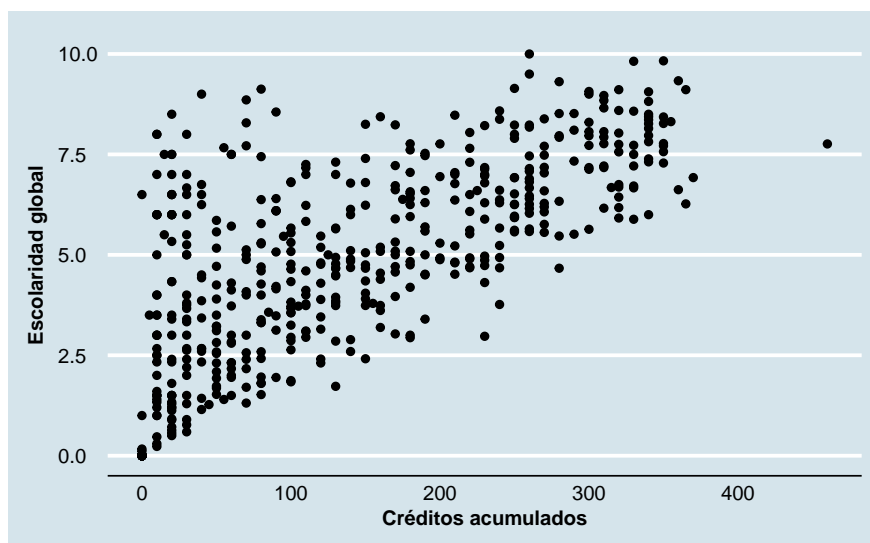


Figura 3: Escolaridad y créditos acumulados.

En las Figuras 4 y 5 puede apreciarse que los estudiantes que optan por rendir sus materias mediante examen (en lugar de curso) son quienes presentan, al mismo tiempo, menor avance y peor escolaridad.

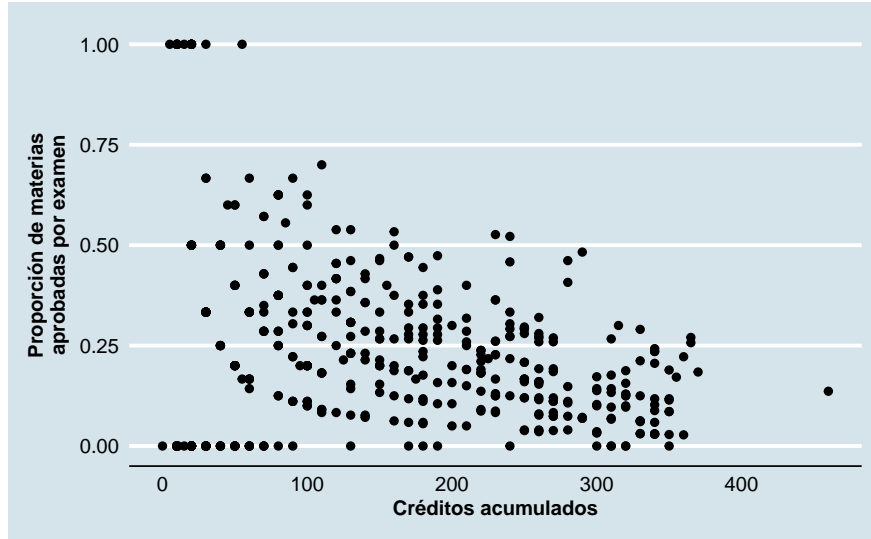


Figura 4: Proporción de actividades aprobadas por examen y créditos acumulados.

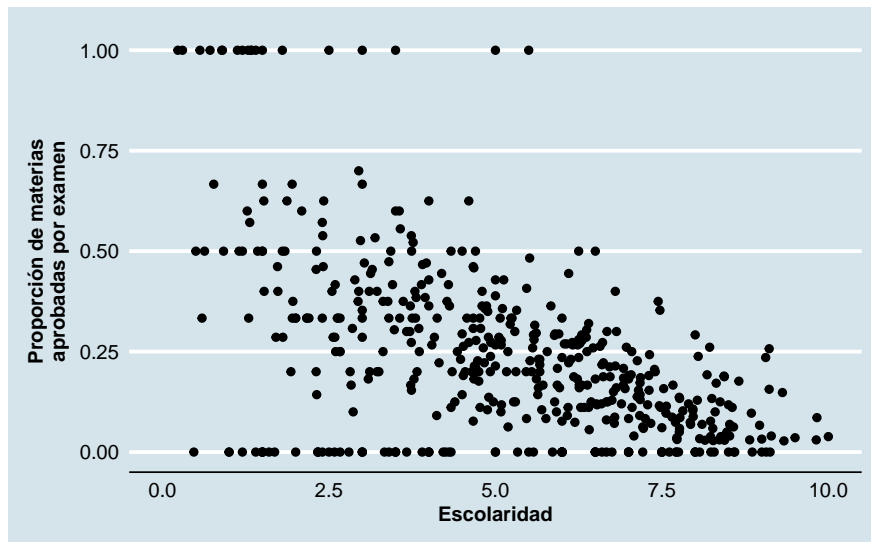


Figura 5: Proporción de actividades aprobadas por examen y escolaridad.

En el cuadro 2, se puede observar que la mayoría de los estudiantes tienen padres con similar nivel educativo, siendo Bajo-Bajo la combinación mayoritaria. Cuando los niveles difieren, en la mayoría de los casos la madre alcanzó un nivel educativo mayor que el padre.

Cuadro 2: Educación del padre (por filas) y de la madre (por columnas)

	Bajo	Medio	Alto
Bajo	153	72	37
Medio	37	103	48
Alto	20	30	76

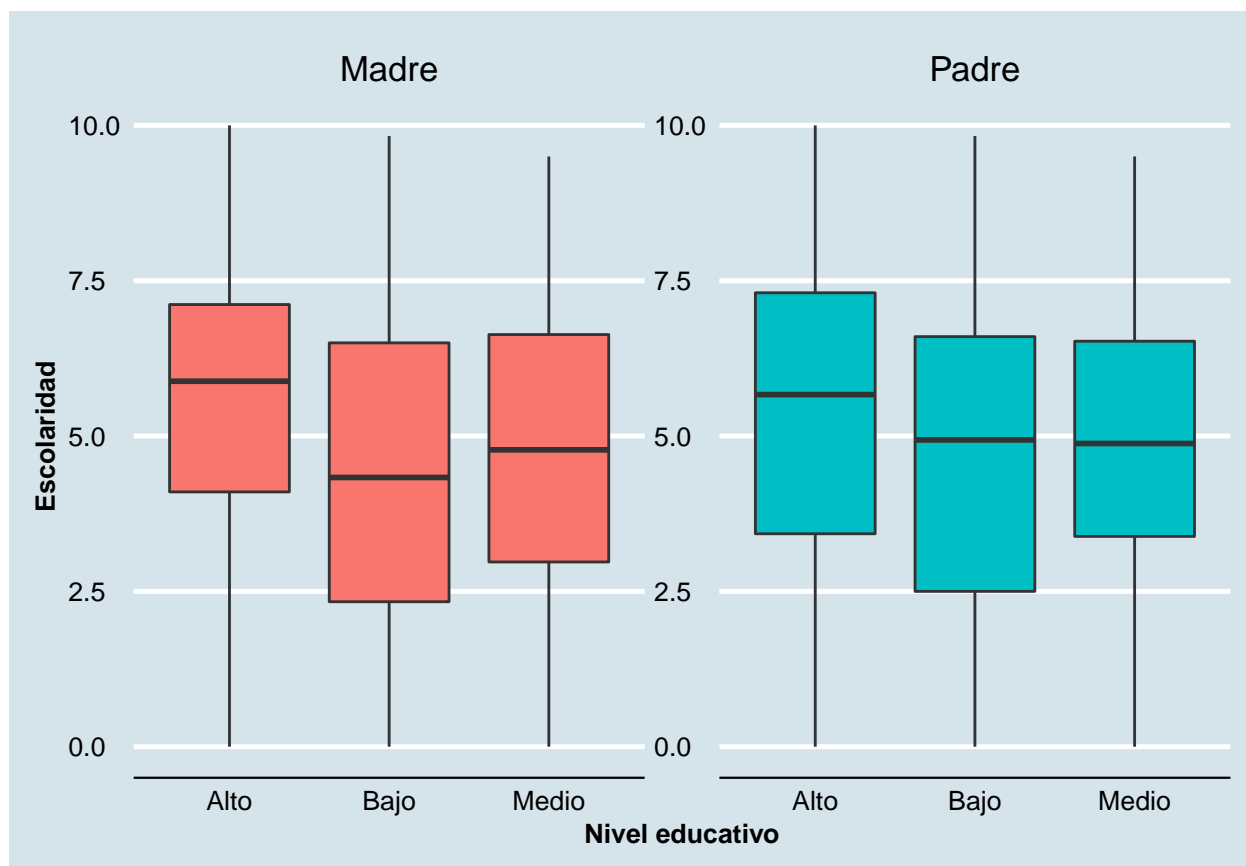


Figura 6: Boxplot de escolaridad según nivel educativo de la madre y el padre.

En las Figuras 6 y 7 se presenta la dispersión de las escolaridades y de los créditos acumulados, respectivamente, según los distintos niveles educativos del padre y la madre. Se observa que, tanto para el padre como la madre, cuanto mayor su nivel educativo mayor la escolaridad del estudiante así como también la cantidad de créditos acumulados. Dicha diferencia está más marcada para el caso del nivel educativo de la madre.

En el Cuadro 3 se observa que la gran mayoría de los estudiantes son jóvenes y no trabajan. Debido a esta característica de la población bajo estudio se resolvió trabajar con la variable **Xocup** en lugar de la variable **horas**.

Cuadro 3: Horas trabajadas según edad de ingreso

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	32	33	35	39	44
No corresponde	19	307	76	31	10	8	3	2	1	0	0	0	1	2	0	0	1	0	0
Menos de 10 horas	0	7	7	1	2	0	2	0	0	2	1	1	0	0	0	0	0	0	0
Entre 10 y 20 horas	0	13	4	1	2	2	0	2	0	0	0	0	0	0	0	1	0	0	0
Entre 20 y 30 horas	2	3	1	6	2	1	2	1	2	1	1	0	0	0	0	0	0	1	0
Entre 30 y 40 horas	0	6	5	4	2	3	0	1	2	3	0	0	0	1	1	0	0	1	0
Más de 40 horas	0	3	2	3	1	2	2	2	0	2	4	4	2	2	0	3	0	1	1

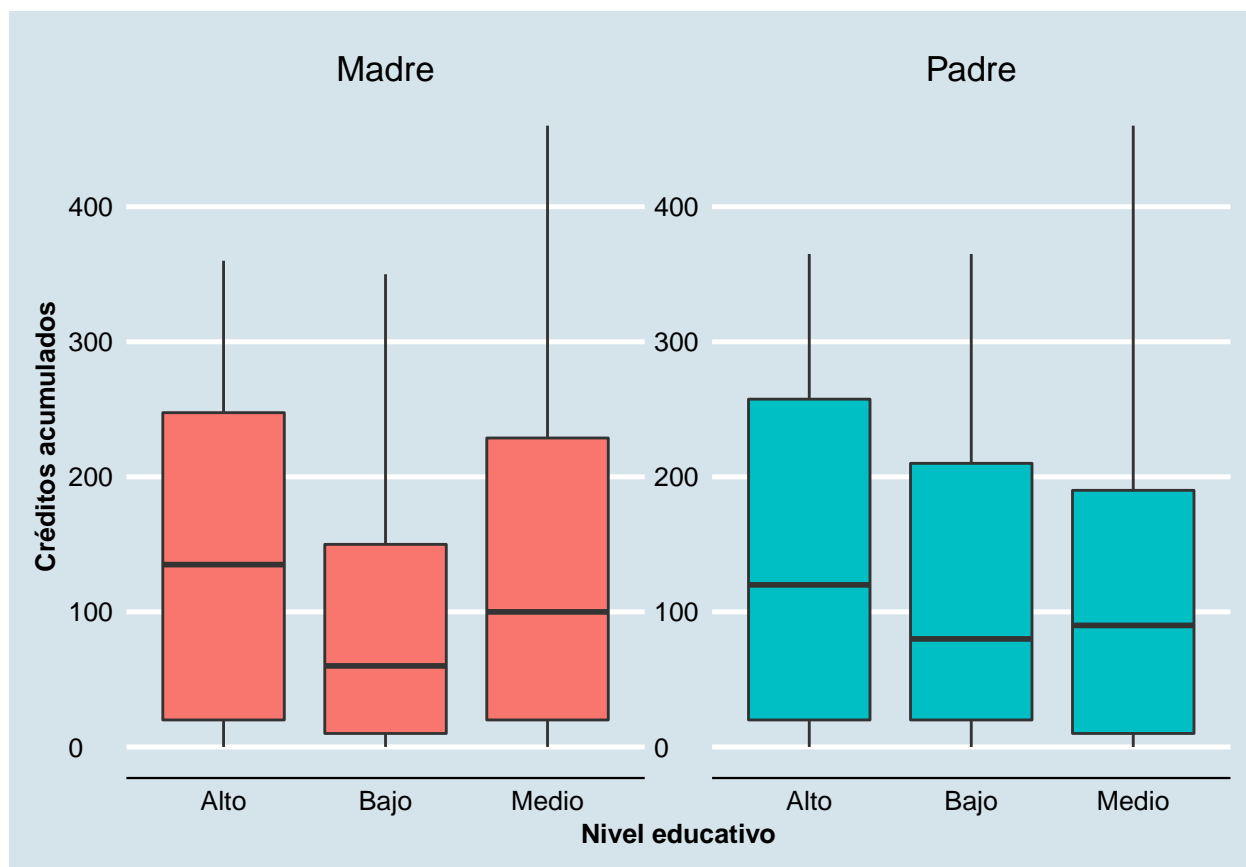


Figura 7: Boxplot de créditos acumulados según nivel educativo de la madre y el padre.

Metodología

Análisis de cluster

Para este trabajo, en primer lugar, se realizaron clusters con métodos jerárquicos agregativos y no jerárquicos. Dentro de los jerárquicos agregativos, se ensayaron distintos algoritmos: *single linkage*, *complete linkage*, *average linkage* y *Ward*. En cuanto a los no jerárquicos, se recurrió a los algoritmos *k-medoides* y *fuzzy sets*.

Un método jerárquico agregativo parte de considerar a cada individuo como un grupo y los une progresivamente hasta llegar a un solo grupo. El que sea jerárquico implica que dos individuos que se unen a un nivel de distancia permanecen en el mismo grupo durante todo el proceso. Para clasificar se debe definir una métrica de distancia entre los individuos, y el algoritmo mediante el cual se unirán individuos con grupos, y grupos con grupos. Los individuos que se unen siempre son los que tienen menor distancia. Todos los algoritmos unen a los grupos con menor distancia, pero difieren en cómo definen la distancia entre dos grupos.

Single linkage: considera que la distancia entre dos clusters es la mínima entre las distancias dos a dos de las observaciones pertenecientes a cada cluster.

$$d_{I,J} = \min\{d_{ij} \mid \forall i \in I \forall j \in J\}$$

Complete Linkage: toma como distancia entre dos clusters la máxima de las distancias dos a dos entre las unidades de cada cluster.

$$d_{I,J} = \max\{d_{ij} \mid \forall i \in I \forall j \in J\}$$

Average Linkage: toma como medida de distancia la media de las distancias entre pares de individuos pertenecientes a cada grupo.

$$d_{I,J} = \frac{1}{n_I n_J} \sum_{i=1}^{n_I} \sum_{j=1}^{n_J} d_{i,j}$$

Método de Ward: busca minimizar el incremento de varianza interna al pasar de k a $k - a$ grupos.

Los métodos jerárquicos indican la forma en que se unen los individuos y los grupos, debiéndose luego definir la estructura de grupos adecuadas. Para ellos se debe recurrir a las *stopping rules*.

Análisis visual del dendograma: dado que las barras verticales del dendograma indican el nivel al que se unen los individuos/grupos, analizar visualmente la longitud de dichas barras da una idea de la estructura de grupos que presentan los datos. Este método es muy informal y solo permite una primera aproximación. En caso de querer obtener una cantidad previamente especificada de grupos, esto determinará el nivel en el cual cortar el dendograma.

Criterio del R^2 : este indicador relaciona la variación explicada y la variación total, donde la variación explicada está representada por la estructura de grupos que se encuentra en cada nivel. Por lo tanto, el valor del R^2 se encuentra en el intervalo $[0, 1]$, valiendo 0 cuando todas las observaciones se encuentran en un mismo grupo (es decir, la variación explicada por la estructura de grupos es 0 ya que todos los individuos están en un mismo grupo), y valiendo 1 cuando cada individuo es un grupo (es decir, cuando la variación explicada es igual a la variación total). En términos matemáticos:

$$R^2 = 1 - \frac{\sum_{k=1}^g \sum_{i=1}^{n_k} \sum_{j=1}^p (x_{ijk} - \bar{x}_{jk})^2}{\sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2}$$

Criterio pseudo-F: relaciona la suma de variaciones entre los grupos (variación explicada) con la suma de variaciones en los grupos (variación residual):

$$Pseudo F = \frac{\text{Variación entre grupos}/(k-1)}{\text{Variación en los grupos}/(n-k)}$$

La cual puede calcularse también como:

$$Pseudo F = \frac{tr(B)/(k-1)}{tr(W)/(n-k)}$$

O como cocientes del R^2 :

$$Pseudo F = \frac{tr(B)/(k-1)}{tr(W)/(n-k)}$$

Reglas empíricas de utilización:

- Si el indicador crece de forma monótona al crecer el número de grupos, no se puede determinar una estructura de grupos.
- Si disminuye de forma monótona al crecer el número de grupos, no se puede determinar una estructura de grupos, pero se puede decir que existe jerarquía.
- Cuando se halla un máximo, la población presenta una estructura de grupos en ese máximo.

Criterio pseudo- t^2 : este indicador busca determinar la significación de juntar dos grupos. Busca determinar en cada paso, si la disminución de la SCR (variación intra-grupos) como resultado de pasar de k a $k+1$ grupos es significativa o no. Algebraicamente, este indicador también se basa en las tazas de las matrices de varianza en el grupo (W), antes y después de unir dos grupos:

$$Pseudo-t^2 = \frac{tr(W_{GL}) - [tr(W_G) + tr(W_L)]}{[tr(W_G) + tr(W_L)] / (n_G + n_L - 2)}$$

donde:

- n_i es la cantidad de observaciones en el grupo $i = G, L$.
- $tr(W_i)$ es la traza en el grupo $i = G, L, GL$.

Si se analiza el vector de valores de pseudo- t^2 desde abajo (1 grupo con n observaciones) hacia arriba (n grupos con 1 observación cada uno), si en $k+1$ grupos presenta una caída “fuerte” respecto de k , entonces nos quedamos con $k+1$ grupos.

Por otro lado, los métodos no jerárquicos generan particiones no encajadas, donde las observaciones se pueden reclasificar durante el proceso, y es necesario definir la cantidad de grupos previamente. Existen diferentes algoritmos para clasificar a los individuos en los grupos, en este trabajo se utilizó fuzzy sets y k-medoides.

Fuzzy sets: calcula el coeficiente de pertenencia de cada individuo a cada grupo.

K-medoides: construye los grupos asignando a cada individuo considerando la menor distancia de este con el centroide de cada grupo. En una primera instancia los centroides deben ser definidos, luego se calculan los centroides de los grupos formados y se repite este proceso un número predefinido de veces o hasta que converja a una estructura.

Análisis discriminante

En cuanto al análisis de discriminante, este consiste en una técnica de clasificación supervisada que tiene tres objetivos: analizar si existen diferencias entre los grupos respecto a un conjunto de covariables, generar procedimientos sistemáticos para clasificar observaciones de origen desconocido en alguno de los grupos considerados, y reclasificar las observaciones en los grupos.

Hay distintos tipos de análisis discriminante: lineal, cuadrático y logístico. Para definir cuál utilizar, se realizan dos tests. Por un lado, se contrasta la hipótesis nula de que los grupos se distribuyen normal multivariante, este contrasta compara la asimetría y la kurtosis de cada distribución con la de una normal. Por el otro, se

contrasta la homogeneidad de las varianzas entre los grupos. Si no se rechaza la hipótesis nula en ninguno de los dos tests da indicios de que es adecuado realizar un análisis discriminante lineal. Si no se rechaza la hipótesis nula en el constaste de normalidad multivariada y sí se rechaza en el de homogeneidad de varianzas, es preferible realizar un análisis discriminante cuadrático. Finalmente, si se rechaza la hipótesis nula en ambos contrastes se debe realizar un análisis discriminante logístico.

Luego de realizar los tests, se plantea un modelo de la pertenencia a cada grupo explicada por las covariables. A continuación se procede a realizar el diagnóstico del modelo. Se evalua la significación global con una prueba stepwise, tanto forward como backward, utilizando el criterio de información de Akaike, y la significación individual de cada covariable realizando el test de Wald. Con el modelo ajustado, se realiza validación cruzada para medir la sensibilidad de los parámetros a la muestra seleccionada.

Resultados obtenidos

Análisis de cluster

Como se mencionó en la metodología, se realizaron agrupaciones con distintos métodos, jerárquicos y no jerárquicos. A continuación se presentan los resultados relevantes. En el anexo se encuentran las salidas correspondientes a los métodos que fueron descartados.

Clusters jerárquicos agregativos

Se observa en la figura 8, correspondiente a un Dendrograma elaborado a partir del algoritmo Single Linkage, que en la base estudiada hay una gran presencia de observaciones atípicas. Esto se aprecia en la forma del gráfico, en particular, que muchas observaciones se unen a cortas distancias y tienden a agregarse de a una a un gran grupo.

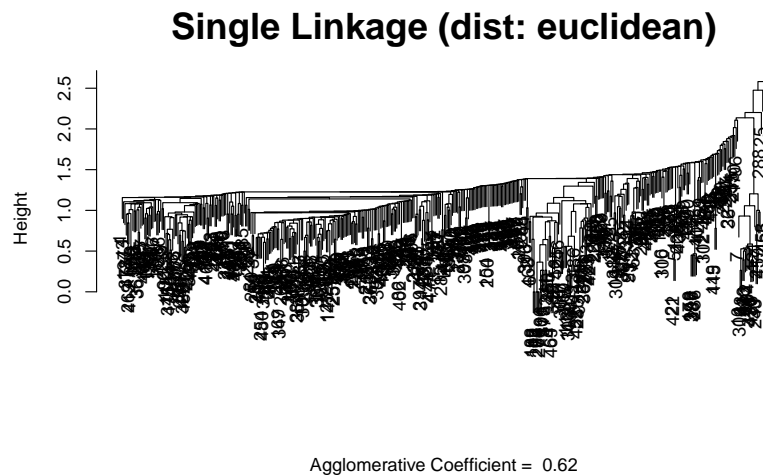


Figura 8: Dendrograma cluster jerárquicos agregativos. Algoritmo: Single Linkage.

En la Figura 9, se observa el dendrograma correspondiente al algoritmo de Ward y en la Figura 10 se grafican los valores de los distintas medidas de bondad de ajuste (R^2 , $Pseudo - F$ y $Pseudo - t^2$). Debido a la forma de constriuir los grupos que tiene este algoritmo, el resultado muestra estructuras más definidas. Parecería una opción razonable, dado los niveles de las uniones finales, elegir una estructura de tres grupos (señalado en

el gráfico con línea verde punteada). Se verifica esta conclusión analizando los valores de las medidas, ya que en una estructura el R^2 se estabiliza, el $Pseudo - F$ presenta un máximo local y el $Pseudo - t^2$ tiene una fuerte caída con respecto a dos grupos. Dados estos resultados, se decide continuar el análisis considerando tres grupos.

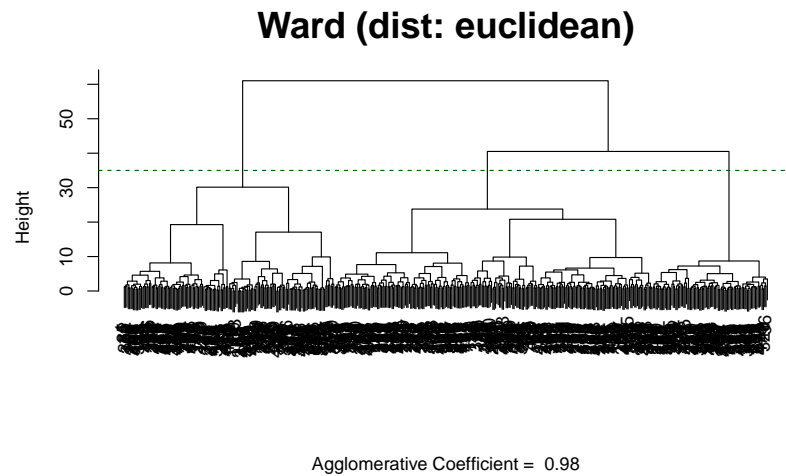


Figura 9: Dendrograma cluster jerárquicos agregativos. Algoritmo: método de Ward.

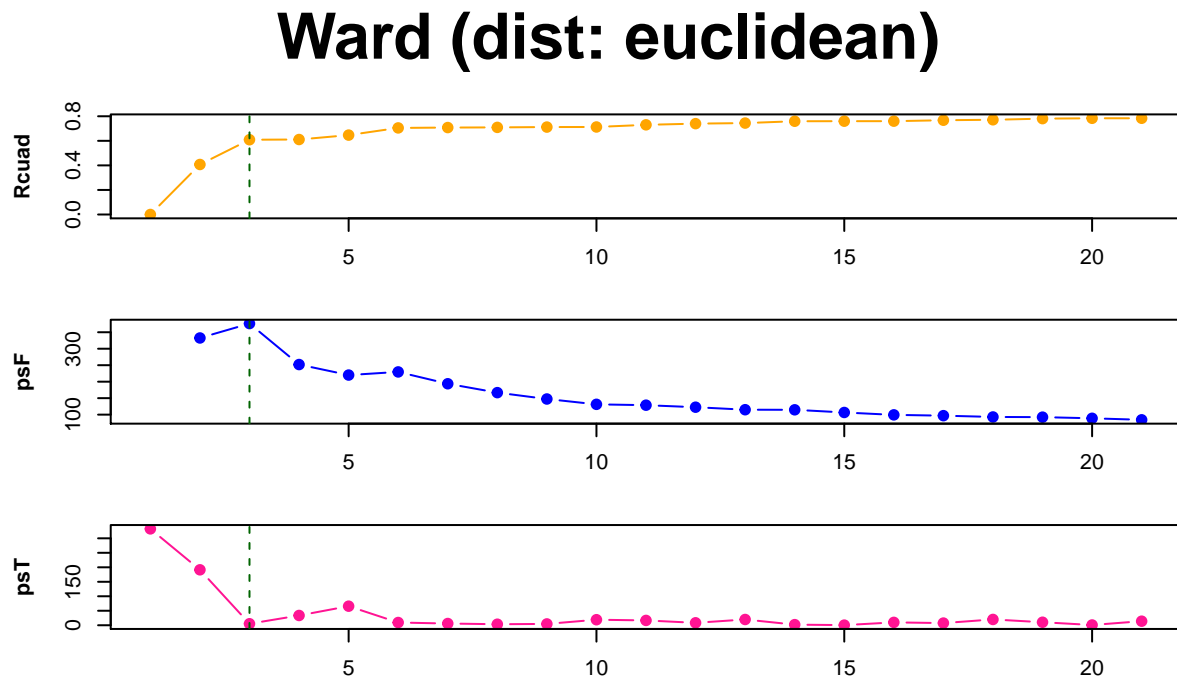


Figura 10: Indicadores. Algoritmo: método de Ward.

Clusters no jerárquicos

Para realizar esta clasificación, se define una estructura de tres grupos, cantidad que se constató como adecuada ya que fue la que se desprendió del análisis jerárquico.

Fuzzy sets

Se construyeron los grupos asignando cada individuo al grupo donde su coeficiente de pertenencia era mayor. El resultado de esta clasificación se aprecia en la figura 11, que muestra la estructura de siluetas. Existen indicios de que algunas observaciones no están bien clasificadas ya que las siluetas de los grupos 1 y 2 presentan valores negativos. No obstante, la gran mayoría está clasificada de forma adecuada, siguiendo este criterio.

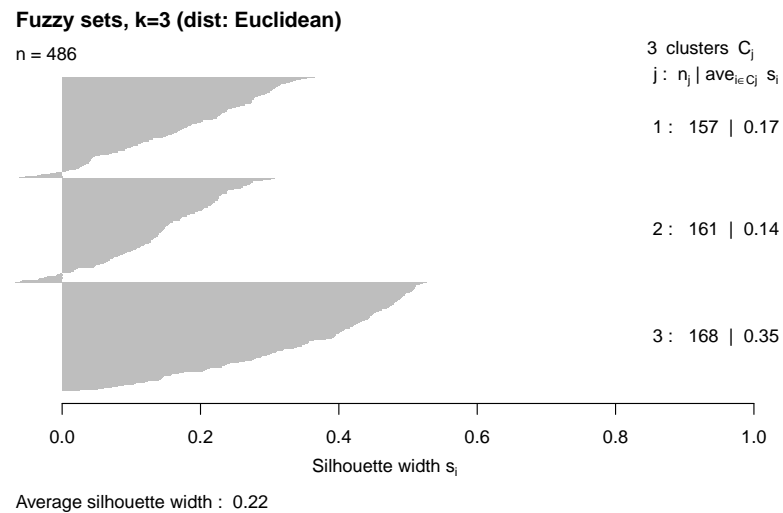


Figura 11: Sihlouette de tres grupos, Algoritmo: Fuzzy sets.

Selección del método

Tal como se desprende de las Figuras 12 y 13, ambos métodos definen un grupo integrado por estudiantes de alto desempeño (alta escolaridad y mayor avance). Sin embargo, la composición presenta diferencias para ambos métodos. El algoritmo de Ward logra restringir mejor el acceso a dicho grupo. Es decir, no clasifica como estudiantes de alto rendimiento a aquellos que presentan rendimientos medios. Esto se debe a la naturaleza del método.

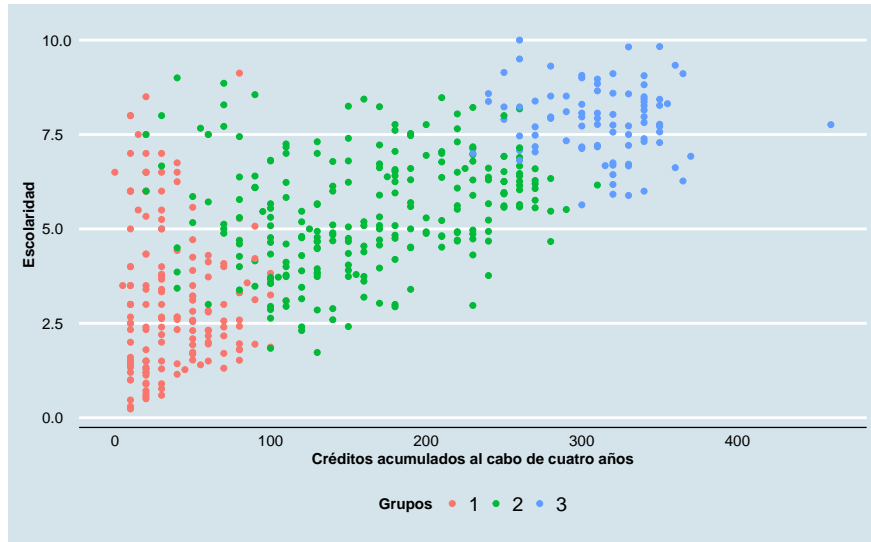


Figura 12: Escolaridad y Créditos acumulados según grupos (método de Ward).

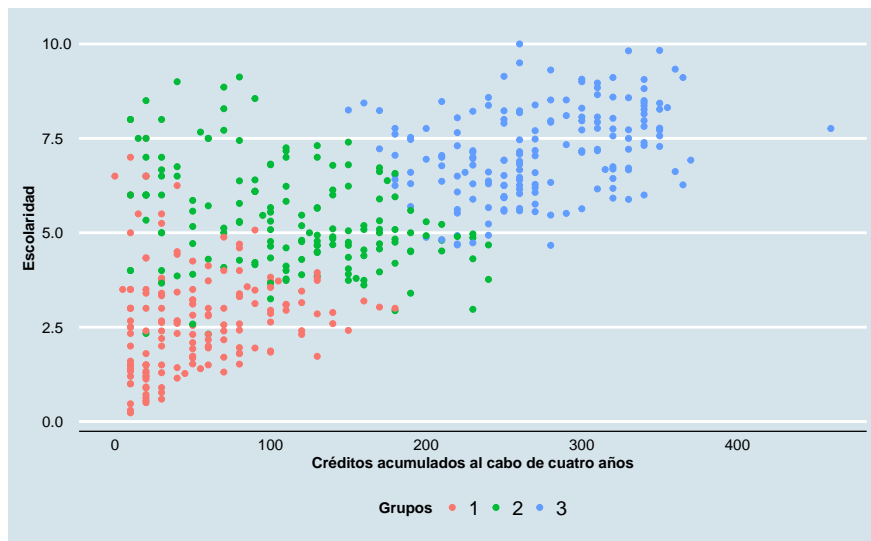


Figura 13: Escolaridad y Créditos acumulados según grupos (Fuzzy sets).

Considerando la asignación de grupos por los métodos de Ward y Fuzzy, se resuelve continuar el análisis con los grupos conformados mediante el método jerárquico agregativo de Ward, ya que es de particular interés para este trabajo que la pertenencia al grupo de alto desempeño sea restrictiva. Esto asegura que todas las unidades que pertenecen a dicho grupo efectivamente tengan un alto desempeño. Por otra parte, la pertenencia al grupo de desempeño medio se esperaba fuera más abarcativa.

		Grupos Fuzzy Sets		
		1	2	3
Grupos	1	122	37	0
	2	35	124	81
	3	0	0	87

En la Figura 14 se observa que los individuos pertenecientes al grupo de mejor desempeño son los que presentan una menor proporción de materias aprobadas por examen. En los otros dos grupos es clara la

diferencia si se observa la cantidad de créditos acumulados, pero se encuentran más dispersos si se considera la proporción de aprobaciones rendidas por examen. Observando la Figura 15, la tendencia es la misma que en el caso anterior, pero no es tan clara la separación de los grupos ya que los individuos de cada uno se encuentran más dispersos.

Cuadro 4: Medias de variables de desempeño por grupo

Grupos	Esco.	Esco. EC	Esco. MC	Créd. 1°	Créd. 2°	Créd. 3°	Créd. 4°	Prop. de exámenes
1	3.27	3.25	0.97	18.93	7.86	5.60	4.84	0.30
2	5.42	5.97	4.31	54.54	41.10	34.98	31.19	0.25
3	7.84	8.09	7.02	87.13	77.99	74.89	71.61	0.11

En resumen, se encuentran tres grupos de desempeño. Uno compuesto por individuos con muy poco avance en la carrera en los cuatro años y baja escolaridad, el cual llamaremos grupo de bajo desempeño, otro compuesto por individuos con la mayor escolaridad y gran cantidad de créditos aprobados por año, al que llamaremos de alto rendimiento. Por último, un grupo residual, al que llamaremos de desempeño intermedio.

Caracterización de los grupos

En esta sección se describen los grupos de desempeño de acuerdo a las características sociodemográficas de los individuos que los componen. En primer lugar, en base al Cuadro 5, se observa que la proporción de individuos que tienen hijos, si bien es baja para todos los grupos, es más alta en el grupo de bajo desempeño. A su vez, la proporción de estudiantes que trabajaban al momento de su ingreso a Facultad es considerablemente mayor en el mismo grupo. Por otra parte, considerando el cuadro 6, se observa que para los grupos de medio y alto desempeño la mayoría de los estudiantes provienen de liceos privados de Montevideo, aunque esta proporción en el grupo de mejor desempeño corresponde a más de un 50 % mientras que en el de medio constituye un 39,2 %. Por otro lado, en el grupo de desempeño más bajo la mayoría de los estudiantes (38,4 %) provienen de liceos públicos del Interior. Por último, en cuanto al nivel educativo del padre, el cual puede observarse en el Cuadro 7, los estudiantes de los tres grupos tienen en su mayoría padres con nivel educativo bajo, superando el 50 % en el grupo de bajo desempeño. En cambio, al considerarse el nivel educativo de la madre no se observa lo mismo. En este caso, los estudiantes del grupo de bajo desempeño son en su mayoría hijos de mujeres con nivel educativo bajo, mientras que en los otros dos grupos en promedio la madre tiene nivel educativo medio. Los individuos sin clasificar (NA) son aquellos que no rindieron ningún examen ni aprobaron ningún curso en los cuatro años estudiados.

Cuadro 5: Medias de variables sociodemográficas por grupo

Grupos	Tamaño	Mujer	Hijos	Trabaja	Carr. 11	Carr. 14
1	159	0.49	0.04	0.29	0.48	0.16
2	240	0.42	0.01	0.13	0.46	0.21
3	87	0.60	0.00	0.10	0.34	0.15
NA	109	0.44	0.06	0.44	0.28	0.06

Cuadro 6: Procedencia por grupo (en %)

	Interior Privada	Interior Publica	Montevideo Privada	Montevideo Publica
1	6.3	38.4	25.2	30.2
2	4.2	32.5	39.2	24.2
3	4.6	24.1	55.2	16.1
NA	8.3	33.0	25.7	33.0

Cuadro 7: Nivel educativo del padre por grupo (en %)

	Bajo	Medio	Alto	NA
1	51.6	23.9	20.1	4.4
2	42.5	35.0	20.8	1.7
3	36.8	29.9	32.2	1.1
NA	42.2	37.6	14.7	5.5

Cuadro 8: Nivel educativo de la madre por grupo (en %)

	Bajo	Medio	Alto	NA
1	45.9	31.4	22.0	0.6
2	32.5	36.7	30.0	0.8
3	24.1	37.9	36.8	1.1
NA	39.4	35.8	24.8	0.0

Análisis discriminante

Es de interés para este trabajo utilizar análisis de discriminante para dos fines distintos: reclasificación y descripción. En primer lugar, se realizó un análisis discriminante para evaluar la necesidad de una reclasificación de los individuos. En segundo lugar, se construyó otra función discriminante para identificar a las variables sociodemográficas que contribuyen a explicar la pertenencia de los individuos a cada grupo.

Reclasificación

Con el fin de determinar qué tipo de discriminante fuera más apropiado a los datos se realizó el test de normalidad multivariada de Mardia. En el Cuadro 9 se puede observar la salida de dicho test. Se constata que se rechazó la hipótesis de normalidad de los grupos con un *p-valor* de 0, por lo que se procedió con un análisis discriminante logístico.

Cuadro 9: Test de normalidad multivariada de Mardia

Test	Statistic	p value	Result
Mardia Skewness	1205.3560355978	1.54145864746748e-178	NO
Mardia Kurtosis	8.62857783047672	0	NO
MVN	NA	NA	NO

Dado que el objetivo en esta primera instancia fue de reclasificación, la especificación del primer modelo fue la siguiente:

$$y_i = f(esco, escomc, escoec, cred1, cred2, cred3, cred4, propaprobexam) \quad \forall i = 1, \dots, n$$

Luego de estimado el modelo, se realizó una prueba de Stepwise (tanto forward como backward), a partir de la cual se descartó la variable **esco**, y se estimó un nuevo modelo con las demás variables. El Cuadro 10 muestra la estimación de dicho modelo. Por su parte, el Cuadro 11 muestra los p-valores asociados al contraste de significación individual de Wald para el modelo bajo análisis. Se encontró que, salvo **carr11** para el grupo 2, y **Xfem** y **carr14** para el grupo 3, todas las variables resultaron significativas para explicar la pertenencia de los individuos a ambos grupos.

Cuadro 10: Estimación del modelo logístico multinomial - Variable dependiente: pertenencia a cada grupo

	(Intercept)	esco.mc	esco.ec	cred1	cred2	cred3	cred4	prop.aprob.exam
2	-14.36	0.88	0.92	0.07	0.10	0.10	0.07	4.66
3	-67.40	3.20	1.93	0.23	0.21	0.26	0.17	-18.50

Cuadro 11: p-valores

	(Intercept)	esco.mc	esco.ec	cred1	cred2	cred3	cred4	prop.aprob.exam
2	0	0	0.00	0	0	0	0	0.01
3	0	0	0.02	0	0	0	0	0.03

En la siguiente tabla, se muestran los valores predichos por este modelo, comparados con las asignaciones de grupos a priori. Los resultados muestran una gran coincidencia entre ambas, sugiriendo que no es necesaria una reclasificación de los individuos.

		Grupos Discriminante		
		1	2	3
Grupos	1	150	9	0
Cluster	2	8	228	4
	3	0	4	83

Descripción

En segundo lugar, se realizó un análisis de discriminante con el objetivo de indagar acerca de qué covariables eran significativas para explicar la pertenencia de los individuos a los distintos grupos. Al igual que en la parte anterior, se realizó el test de normalidad multivariada, cuyos resultados se encuentran en el Cuadro 12 donde puede apreciarse que se rechazó que los grupos se distribuyeran normal multivariada, por lo cual se realizó un análisis de discriminante logístico.

Cuadro 12: Test de normalidad multivariada de Mardia

Test	Statistic	p value	Result
Mardia Skewness	5701.00611044267	0	NO
Mardia Kurtosis	35.4773392996285	0	NO
MVN	NA	NA	NO

Se define el siguiente modelo de la pertenencia a los grupos explicada por las variables sociodemográficas:

Cuadro 13: Estimación modelo logístico multinomial - Variable dependiente: pertenencia a los grupos

	2	3
(Intercept)	-796.83	62.81
gen	0.40	-0.03
edad	-0.12	-0.39
Xfem	-0.45	0.23
Xhijos	0.14	-6273.89
Xocup	-0.82	-0.70
XeduPMedio	0.46	0.19
XeduPAlto	-0.09	0.23
XeduMMedio	0.14	0.38
XeduMAlto	0.37	0.79
carr11	-0.16	-0.51
carr14	0.52	-0.02
busca.trabajo	-0.48	-1.64
sexto.mvd	0.35	0.36
sexto.priv	0.15	0.63

A continuación, se realizó la prueba de Stepwise, y a raíz de esta se eliminaron del modelo las variables `Xhijos`, `XeduP`, `XeduM` y `sexto.mvd`, y el modelo resultante fue:

Cuadro 14: Estimación modelo logístico multinomial II - Variable dependiente: pertenencia a los grupos

	2	3
(Intercept)	-794.18	-691.99
gen	0.40	0.35
edad	-0.10	-0.38
Xfem	-0.45	0.23
Xocup	-0.80	-0.66
carr11	-0.20	-0.59
carr14	0.44	-0.03
busca.trabajo	-0.47	-1.70
sexto.priv	0.42	1.00

Cuadro 15: p-valores

	(Intercept)	esco.mc	esco.ec	cred1	cred2	cred3	cred4	prop.aprob.exam
2	0	0	0.00	0	0	0	0	0.01
3	0	0	0.02	0	0	0	0	0.03

En este caso, las variables `gen`, `edad`, `Xocup`, `busca.trabajo` y `sexto.priv` son significativas a un nivel del 5% para explicar la formación de los grupos 2 y 3. Por otro lado, `Xfem` y `carr14` no resultaron significativas para explicar la pertenencia al grupo 3, pero sí lo fueron para explicar la pertenencia al grupo 2. En cambio, `carr11` contribuye a explicar la pertenencia al grupo 3, pero no al grupo 2.

Conclusiones

En primer lugar, se encontró una estructura de tres grupos definida en función del desempeño de los estudiantes. Uno de estos grupos estuvo conformado por estudiantes de alto desempeño, otro por desempeños bajos, y un tercero residual, más heterogeneo, de desempeño intermedio. Por otra parte, se encontró que los estudiantes provenientes de liceos privados de Montevideo no se encuentran concentrados en el grupo de estudiantes de alto rendimiento, (contrariamente a como se esperaba de forma a priori), sino que se encuentran presentes en todos los grupos. No obstante, el grupo de alto rendimiento sí presenta un alto porcentaje de estudiantes provenientes de liceos privados de Montevideo. A su vez, no se encontró evidencia de que la educación de los padres sea relevante para la conformación de grupos de rendimiento.

Por último, contrariamente a lo que se esperaba, se encontró que la situación laboral del estudiante al momento de su inscripción a la carera sí resultó significativa para la conformación de los grupos. Esto probablemente se deba a que aquellos estudiantes que estaban trabajando al inicio de la misma hayan continuado haciendolo en los años subsiguientes. Asimismo, se encontró que el hecho de que el estudiante buscara trabajo en el momento de su ingreso a FCEA resultó relevante para explicar la pertenencia de los individuos a distintos grupos de desempeño, lo que podría deberse a que estos estudiantes hayan comenzado a trabajar antes que los que no estaban interesados en hacerlo cuando comenzaron la carrera.

Anexo

Complete Linkage

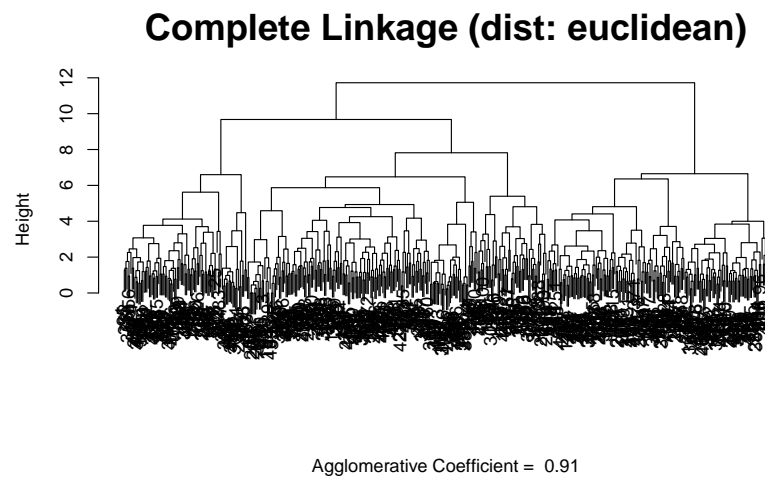


Figura 14: Dendrograma cluster jerárquicos agregativos. Algoritmo: Complete Linkage.

Average Linkage

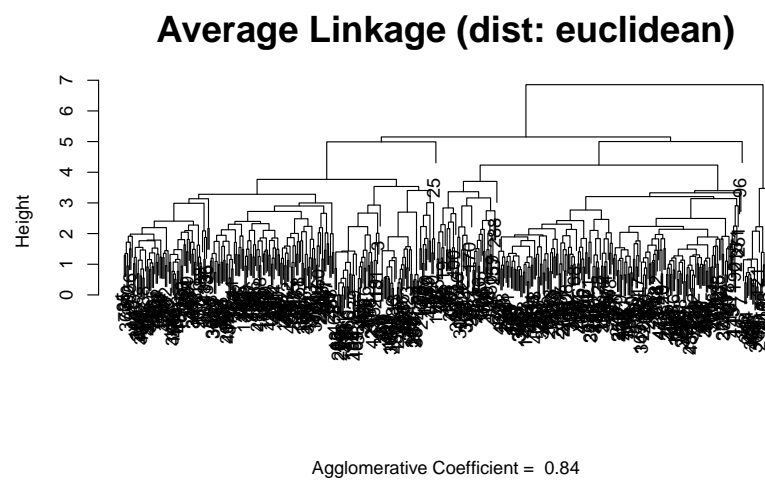


Figura 15: Dendrograma cluster jerárquicos agregativos. Algoritmo: Average Linkage.

k-medoides

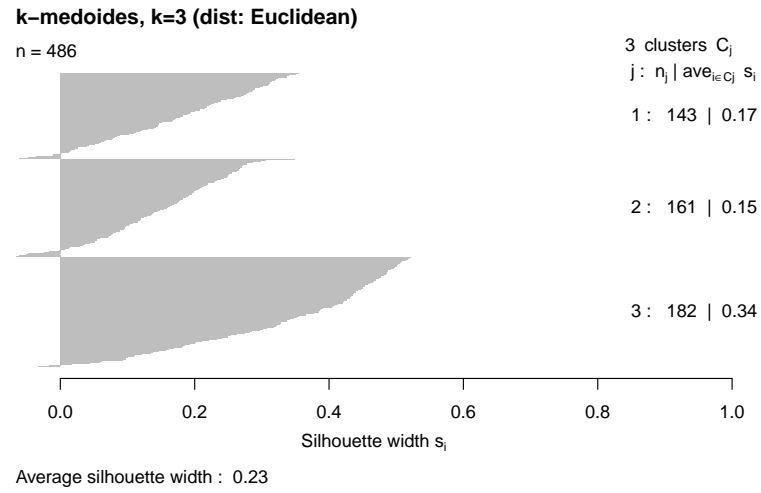


Figura 16: Silhouette cluster no jerárquicos. Algoritmo: k-medoides.

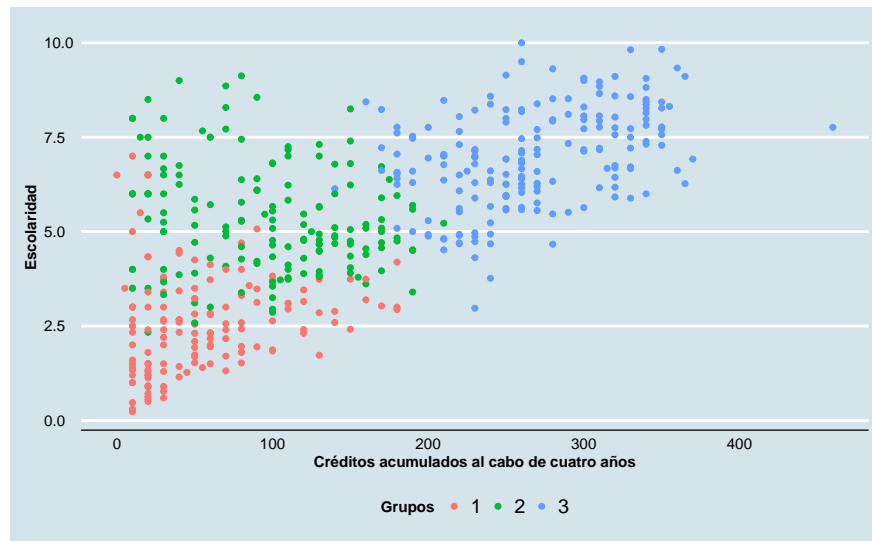


Figura 17: Caracterización de los grupos. Algoritmo: k-medoides.

Ward

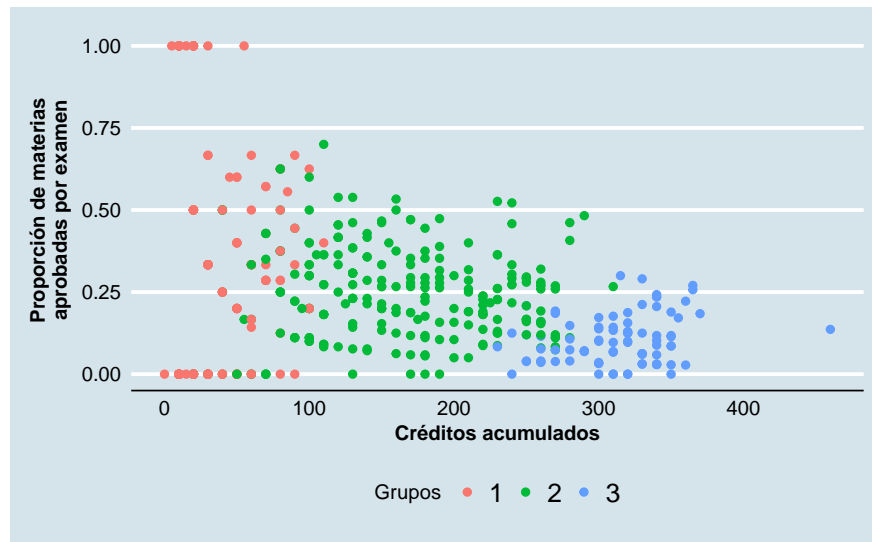


Figura 18: Proporción de actividades aprobadas por examen y créditos acumulados. Algoritmo: Ward

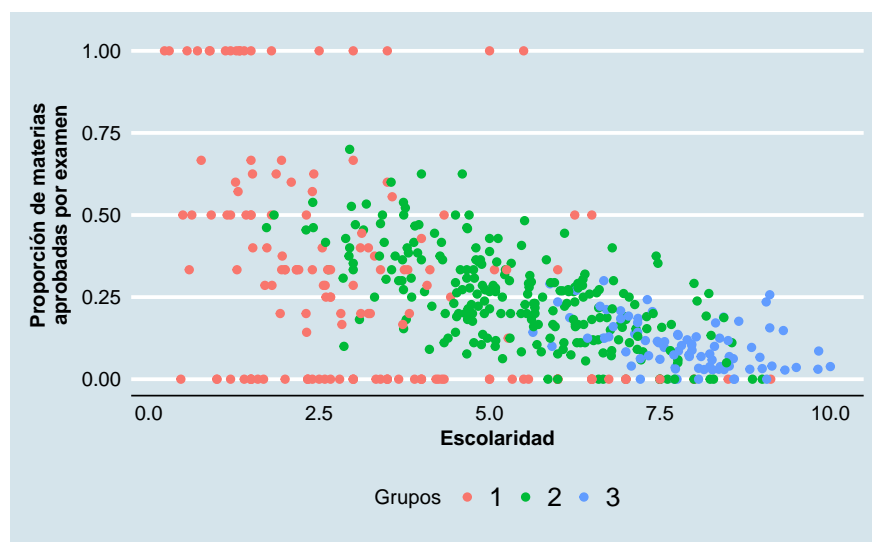


Figura 19: Proporción de actividades aprobadas por examen y créditos acumulados. Algoritmo: Ward

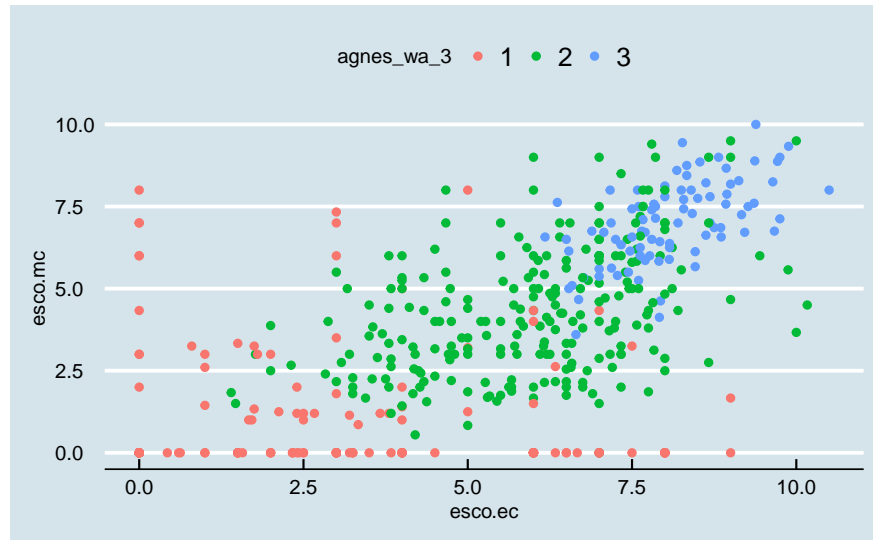


Figura 20: Proporción de actividades aprobadas por examen y créditos acumulados. Algoritmo Ward

Referencias

Arim, Rodrigo, Juan José Goyeneche, Noemi Katzkowicz, Gabriela Sicilia, Elena Vernazza, y Guillermo Zoppolo. 2016. «Evaluación del impacto del plan de estudios 2012 sobre los resultados académicos de los estudiantes».