

¿Se puede construir una tipología de estudiantes según su trayectoria de avance en la carrera?

Mara Dutto
Juan Jose Goyeneche

Agosto 2011

Serie DT (11/04)
ISSN : 1688-6453

¿SE PUEDE CONSTRUIR UNA TIPOLOGÍA DE ESTUDIANTES SEGÚN SU TRAYECTORIA DE AVANCE EN LA CARRERA?

María Dutto

Juan José Goyeneche *

RESUMEN

El objetivo de este trabajo es investigar si es posible construir una tipología de estudiantes según su trayectoria de avance en la carrera, que pueda servir como insumo para la mejora de la gestión universitaria. Como es un estudio exploratorio, a modo de ejemplo se toman únicamente los estudiantes de algunas generaciones

Se utiliza la información generada por el Sistema de Gestión de Bedelías de la Universidad de la República, que contiene los registros de las actividades de los estudiantes y algunas de sus características sociodemográficas.

Se emplea la técnica de análisis de cluster utilizando distintas formas de conformar los grupos de trayectorias. Finalmente, se analiza la conformación de los grupos según algunas variables sociodemográficas y se plantean líneas para seguir investigando.

Palabras clave: *educación, gestión, análisis de cluster, desempeño escolar, graduación*

1. Introducción

Este trabajo se realizó en el marco del grupo de estudio sobre el Plan 90 que lleva adelante el Instituto de Estadística. El grupo está conformado por Ignacio Álvarez, Joaquín Amiel, Sebastián Castro, Natalia Da Silva, María Dutto, Juan José Goyeneche, Gabriela Mathieu y Guillermo Zoppolo.

La información proviene del Sistema de Gestión de Bedelías de la Universidad de la República, que contiene los registros de las actividades de los estudiantes, y algunas de sus características sociodemográficas. Las variables utilizadas se observan en las Tablas 1 y 2.

*Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, UDELAR

Variables sobre las actividades		
Nombre	Tipo	Descripción
estci	categórica	cédula de identidad sin dígito verificador
mat	categórica	código de la materia
fecha	fecha	fecha de la actividad
tact	categórica	tipo de actividad: C (cursos), E (exámenes), N (cursos invalidados), D (cursos caducados)
nota	numérica	tipo de registro: N (normal), R (reválida), A (automático) y C (cambio de plan)

Tabla 1: Variables sobre las actividades

Variables sociodemográficas		
Nombre	Tipo	Descripción
lugar	categórica	departamento o país de nacimiento
nacido	fecha	fecha de nacimiento
sexo	categórica	sexo
inst	categórica	Instituto del que proviene
tipinst	categórica	Tipo de instituto: extranjero, público secundario, UTU, público-otros, privado laico y privado religioso

Tabla 2: Variables sociodemográficas

Se generaron nuevas variables que podrían explicar la trayectoria escolar (Tabla 3), especialmente referidas al desempeño de los estudiantes en sus primeros años.¹

Antes de pensar en encontrar tipologías de trayectorias fue necesario definir una única trayectoria por estudiante. Muchos estudiantes están anotados a más de un ciclo, pero no necesariamente los siguen con la misma intensidad. Si hiciéramos el análisis por carrera-ciclo (por ejemplo para la 4-1, que es Contador Público) solamente en base a las inscripciones surgirían algunos problemas. Por un lado, al no incluir a los del ciclo básico que no llegaron a optar por una carrera, se está subestimando la deserción en los primeros años. Por otro lado, se sabe que existen muchos estudiantes que se inscriben a más de un ciclo e incluso hacen materias de todos ellos pero finalmente optan por uno, por lo que no sería adecuado considerarlos como desertores de uno de los ciclos.

Desde el punto de vista práctico lo que se hizo fue lo siguiente:

- Se eliminaron los estudiantes cuya única carrera o la carrera donde acumulaban más semestres fuera una de las opciones de la Licenciatura en Estadística, ya que nuestro objeto de estudio es el Plan 90.

¹Las variables prim1, prim2, anu1 y anu2 son categóricas porque valen -1 si el estudiante no rindió ningún examen en el período de tiempo considerado y -2 si no se anotó a ningún curso

VARIABLES GENERADAS A PARTIR DE LAS ORIGINALES		
Nombre	Tipo	Descripción
edad.ing	numérica	edad al ingreso a la facultad
fecing_f	fecha	fecha de ingreso a la facultad
gen	numérica	generación (año de ingreso a la facultad)
lugar.inst	categoría	departamento o país del instituto de procedencia
prim1	categoría	cantidad de materias de primero aprobadas en su primer año en facultad
prim2	categoría	cantidad de materias de primero aprobadas en sus dos primeros años en facultad
anu1	categoría	cantidad de materias anuales de primero aprobadas en su primer año en facultad
anu2	categoría	cantidad de materias anuales de primero aprobadas en sus dos primeros años en facultad
esc1	numérica	escolaridad en las materias de primero
esc_1ano	numérica	escolaridad en su primer año en facultad

Tabla 3: Variables generadas a partir de las originales

- Se unieron las carrera/ciclo 4-1, 4-4 y 4-5, recodificándolas como “40” (Administrativo-Contable). La razón es que en el caso de un estudiante que hace las carreras de Contador Público y Licenciado en Administración - Contador (sector privado o público), es difícil elegir una, ya que tienen muy pocas materias de diferencia. Por lo tanto, se optó por unir las en una sola carrera: la “40”. Para determinar la trayectoria de avance de estos estudiantes se toma en cada momento del tiempo el mayor avance entre las carreras 4-1, 4-4 y 4-5.
- Se eligió una única carrera/ciclo para cada estudiante inscripto en más de una: aquella en la que hayan avanzado más para los que llegaron al 50 % de la carrera y “49” que representa ciclo básico o aún no definidos por una carrera para el resto.

Luego de este procesamiento los estudiantes de carreras del Plan 90 quedan clasificados como: “40” (15046), “42” (2120) y “49”(10418).

Nota: el avance se mide en semestres aprobados sobre el total de semestres necesarios para el egreso. No se mide en materias aprobadas porque se considera que las materias anuales “pesan” más que las semestrales con relación al avance. Para no distorsionar la comparación del avance en las carreras 4-1, 4-4 y 4-5 de los estudiantes que aprobaron solo materias comunes a las tres carreras, se optó por usar el mismo denominador para todas, aunque las carreras de Licenciado en Administración tiene más semestres que la de Contador Público.

En la Figura 1 se puede ver el gráfico de las trayectorias en sus primeros 12 años de los estudiantes para las generaciones 1990 a 1998. En el primer caso están coloreadas por la

cantidad de materias de primero aprobadas en su primer año en facultad y en el segundo por el status que tienen en abril de 2010. Nótese que a los efectos del cálculo de la variable status, los estudiantes de la generación 1990 se los observa por una ventana de 20 años y a los de la generación 1998 sólo por 12.

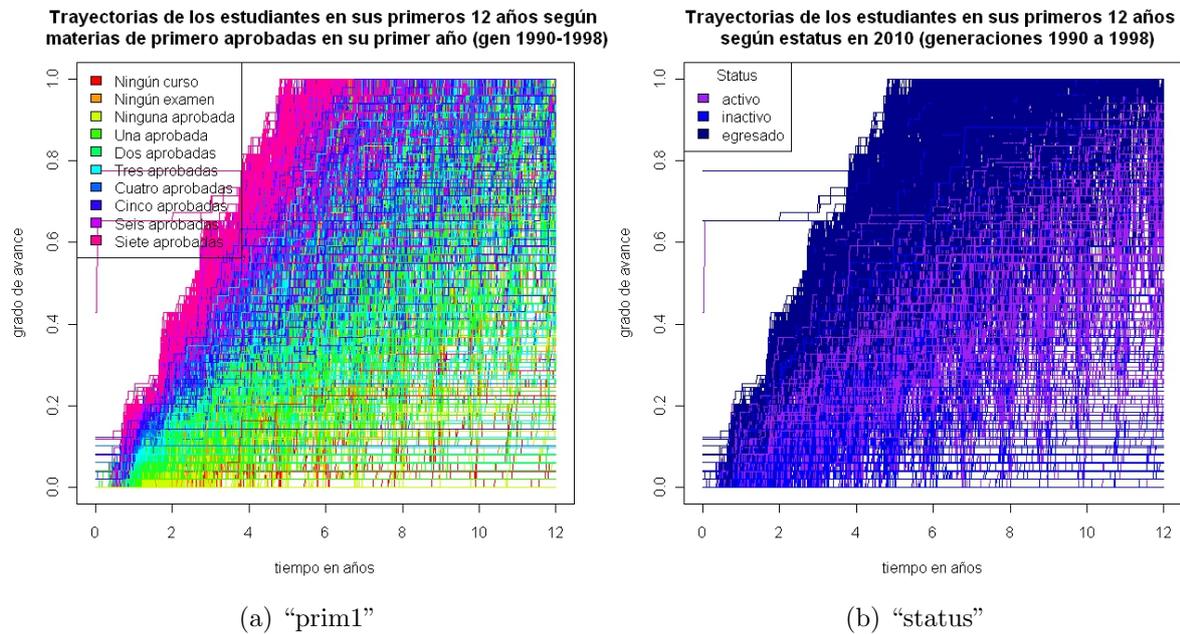


Figura 1: Trayectorias por "prim1" y "status"

En la Figura 1 puede verse que a grandes razgos tanto "prim1" como "status" parecerían contribuir a diferenciar las trayectorias de los estudiantes. Sin embargo, no es una división clara.

2. Análisis de grupos

El análisis de grupos o de conglomerados tiene como objetivo conformar grupos homogéneos de acuerdo a ciertas características de interés y realizar tipologías de individuos. Se trata de una técnica descriptiva, exploratoria. El resultado va a depender tanto de las variables elegidas como de la definición de distancia o disimilaridad entre individuos y grupos de individuos.

Existen métodos jerárquicos y no jerárquicos. Los primeros generan particiones encajadas y se pueden clasificar en dos: agregativos y divisivos. Los agregativos parten de I grupos, cada uno conformado por un individuo y al final del proceso están todos en un solo grupo. En cada paso se unen solamente dos grupos, por lo que el procedimiento tiene I pasos.

Si dos individuos se unen en un momento del proceso, siguen juntos hasta el final. En los divisivos el proceso es inverso. En los métodos no jerárquicos por el contrario, se define un número de grupos deseado, y dos individuos que estaban en un mismo grupo pueden separarse durante el proceso.

Método de Ward

El método de Ward es un método jerárquico agregativo. Para explicarlo, es necesario recordar que la varianza total se puede descomponer en varianza dentro de los grupos y varianza entre los grupos, o en otras palabras, la suma de cuadrados totales (SCT) es igual a la suma de cuadrados dentro (SCD) más la suma de cuadrados entre (SCE)

$$SCT = SCD + SCE$$

$$\sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_j)^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} \sum_{j=1}^J (x_{ij(k)} - \bar{x}_{kj})^2 + \sum_{k=1}^K n_k (\bar{x}_{kj} - \bar{x}_j)^2$$

siendo I el número de individuos, J el número de variables, K el número de grupos, x_{ij} el valor del individuo i en la j -ésima variable, \bar{x}_j el promedio de todos los individuos en la variable j , \bar{x}_{kj} el promedio de todos los individuos del grupo k en la variable j y n_k la cantidad de individuos que tiene el grupo k .

En el método de Ward en cada paso, de todas las formas posibles de unir dos grupos, se elige la que produce mínima varianza dentro de los grupos. Forma grupos más compactos que otros métodos como vecino más lejano y más cercano, con cantidad de elementos similares, es decir, es menos sensible a la formación de grupos con pocos individuos “atípicos”.

Elección del número de grupos

La elección de la cantidad de grupos es subjetiva, pero existen algunas ayudas. El dendrograma permite ver la estructura de particiones encajadas y la distancia a la que se fueron juntando los grupos. También existen algunas reglas que según Blanco (2006) funcionan muy bien cuando los grupos son esféricos y diferenciados.

El

$$R^2 = \frac{SCE}{SCT}$$

es la proporción de variación explicada por un determinado agrupamiento, cuando cada individuo es un grupo vale 1, y va cayendo, llegando a 0 cuando todos están en un mismo grupo. En general uno quiere quedarse con un número razonablemente pequeño de grupos. Cuando de una partición con k grupos se pasa a otra con $k - 1$ grupos, el R^2 disminuye (o queda igual). Se elige un número k de grupos que implique un balance entre un R^2 “alto” y un número de grupos “pequeño”. Si se grafica el R^2 en función del número de

grupos, uno podría quedarse con el número de grupos del “final de la meseta”, antes de que el R^2 caiga abruptamente.

La regla de Calinski o

$$pseudo.F = \frac{R^2/(K-1)}{(1-R^2)/(I-K)}$$

donde K es el número de grupos e I es el número de individuos. Si el indicador tiene un máximo relativo, se prefiere esa estructura de grupos. Si crece o decrece monótonamente, no puede determinarse una estructura de grupos.

El

$$pseudo.t^2 = \frac{tr(W_{G \cup L}) - (tr(W_G) + tr(W_L))}{(tr(W_G) + tr(W_L))/(n_G + n_L - 2)}$$

donde W_G y W_L son las matrices de variación dentro de los grupos G y L respectivamente y $W_{G \cup L}$ la variación dentro del grupo que surge de la unión entre G y L . Si hay una caída importante del indicador al pasar de k grupos a $k+1$, se recomienda quedarse con $k+1$.

Cluster de curvas

Queremos hacer análisis de *cluster* de trayectorias temporales (en nuestro caso el avance en la carrera). Una de las razones para hacer esto es que si se tienen muchos individuos la representación gráfica de las trayectorias es muy difícil (queda un gráfico lleno de líneas en el que nada se distingue) y una opción sería graficar las trayectorias típicas (D’Urso, 2000). Para esto es necesario hacer un análisis de cluster.

La clave está en definir una medida de disimilaridad entre las curvas, que dependerá del objetivo del trabajo. En nuestro caso se discretizó cada curva en años y se usaron esos valores como insumo para el análisis de *cluster*. Se probó trabajar con los incrementos relativos año a año pero el resultado fue mucho más pobre, con tipologías menos claras.

3. Resultados

En esta sección trataremos de aplicar el análisis de cluster a las trayectorias estudiantiles de avance en la carrera. En primer lugar se calcula la variable tiempo en años, que es la resta de la fecha de la actividad con la fecha de ingreso del estudiante a la facultad. Cada trayectoria es una función a tramos que aumenta cada vez que el estudiante aprueba una materia. Si la trayectoria llega a 1 es porque el estudiante se recibió.

Se decidió trabajar con las generaciones 1990 a 1998 porque se los observa por una ventana de al menos 12 años, tiempo razonable para observar la mayoría de los egresos. Se trabajaron todas esas generaciones juntas, haciendo un análisis de *cluster* por el método

de Ward con el avance de cada estudiante en los años 1 a 12. El análisis se hizo en el programa R, usando la función *agnes()* de la librería *cluster*. Tomando en cuenta los indicadores ya explicados (R^2 , pseudo-F y pseudo-t) y lo que surge del dendrograma (Figura 2) se decidió quedarse con 6 grupos.

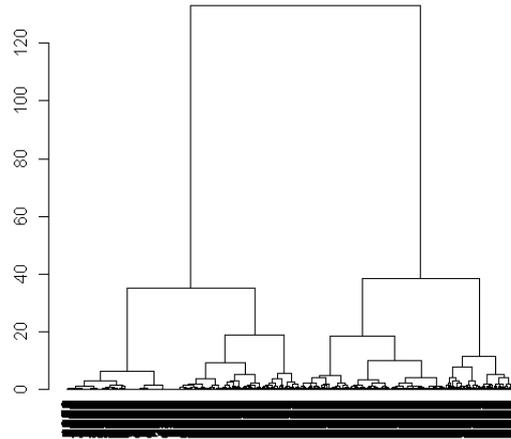


Figura 2: Dendrograma

En la Tabla 4 se muestra que las proporciones de los grupos se mantienen estables para las distintas generaciones. Los “primeros” son aproximadamente el 20 %, los “segundos” un 10 %, etc. En los nueve años estudiados hay un grupo de “primeros”, un grupo de “segundos”, etc., y las proporciones de estos grupos se mantienen a través de las generaciones.

tablita	primeros	segundos	terceros	cuartos	quintos	sextos	Sum
1990	20.3 %	11.0 %	15.2 %	10.4 %	18.0 %	25.0 %	100.0 %
1991	21.1 %	10.1 %	15.8 %	8.4 %	19.7 %	24.9 %	100.0 %
1992	22.4 %	11.1 %	15.0 %	8.3 %	17.0 %	26.1 %	100.0 %
1993	21.2 %	10.0 %	15.6 %	9.0 %	18.6 %	25.6 %	100.0 %
1994	19.9 %	13.6 %	15.2 %	7.0 %	20.4 %	23.9 %	100.0 %
1995	23.0 %	11.6 %	14.9 %	8.3 %	19.6 %	22.6 %	100.0 %
1996	22.0 %	11.4 %	15.2 %	9.4 %	19.2 %	22.8 %	100.0 %
1997	19.5 %	11.2 %	15.0 %	7.6 %	17.9 %	28.8 %	100.0 %
1998	23.7 %	11.3 %	13.9 %	8.0 %	19.0 %	24.1 %	100.0 %
1990-98	21.4 %	11.3 %	15.1 %	8.4 %	18.8 %	24.9 %	100.0 %

Tabla 4: Distribución porcentual de estudiantes por grupo según generación

En la Figura 3 se ve el gráfico de las trayectorias coloreadas por grupo. La división de grupos parece bastante clara, y por lo tanto se nombran como “primeros”, “segundos”, etc, con relación a su rapidez para avanzar en la carrera. Los “primeros” son los que avanzan más rápido, los “segundos” son los que les siguen, y así sucesivamente. A grandes rasgos se puede ver que los “primeros” se reciben en su mayoría entre los 5 y 6 años, los “segundos” entre los 6 y los 12 y el resto no se recibe en los primeros 12 años de actividad en la Facultad.

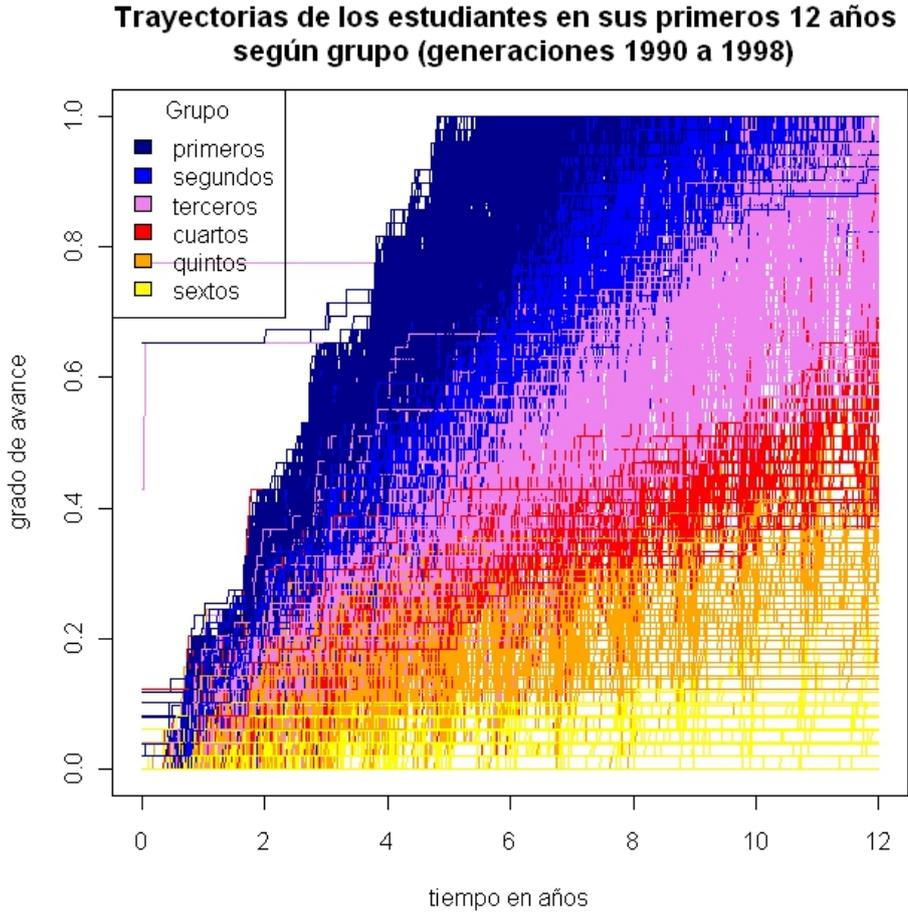


Figura 3: Los grupos de trayectorias

Luego de la obtención de los grupos se procedió a caracterizarlos según las variables sociodemográficas de las que disponemos información.

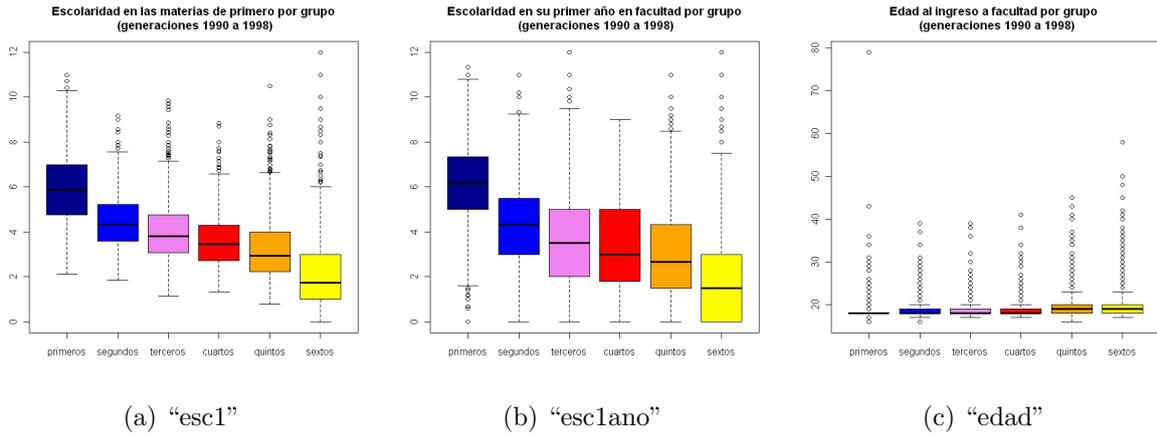


Figura 4: Diagramas de caja de edad y escolaridad por grupo

En los diagramas de caja (Figura 4) se ve que hay una relación entre la rapidez del avance y la escolaridad (tanto en primer año como en las materias de primero). Los “primeros” tienen escolaridad mucho mayor que el resto, mientras que los “sextos” son los que tienen peores valores. Entre los cuatro grupos intermedios no se nota una diferencia clara. Con respecto a la edad de ingreso a Facultad se puede observar que hay mayor dispersión en los grupos más lentos.

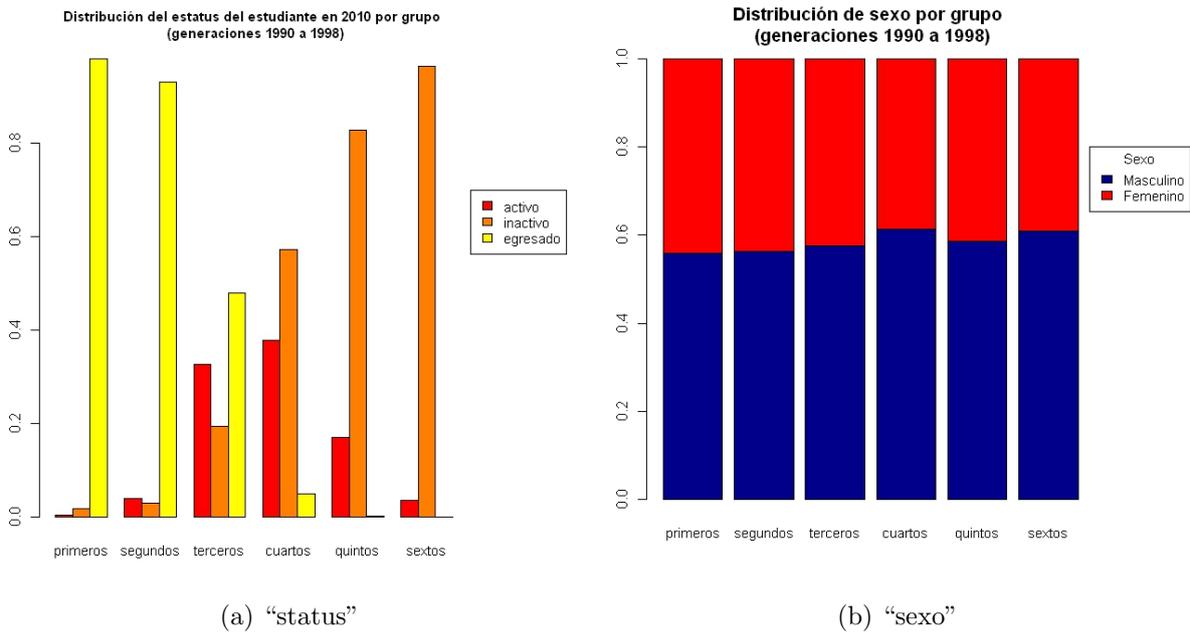


Figura 5: Distribución de status y sexo por grupo

En las Figuras 5 puede verse que los dos grupos más rápidos tienen un porcentaje muy alto de egreso (si se los observa hasta el 2010), mientras que los dos últimos casi no tienen egresados. Por lo tanto, la tipología encontrada parecería ser adecuada para explicar el egreso. Con relación a la distribución por sexo no se observan diferencias al interior de los grupos.

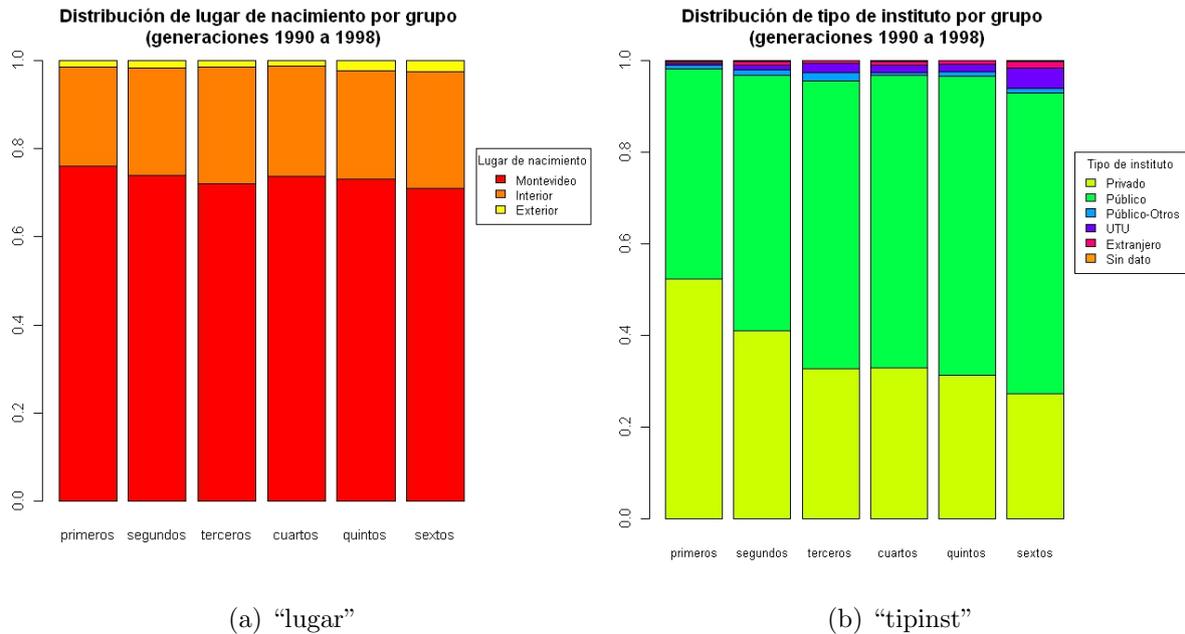
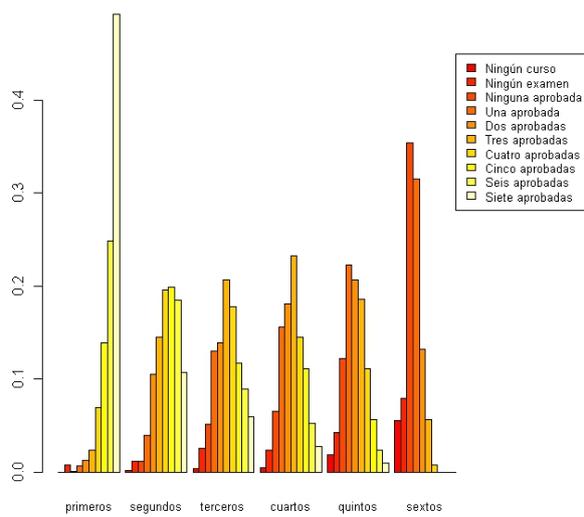


Figura 6: Distribución de lugar de nacimiento y tipo de instituto por grupo

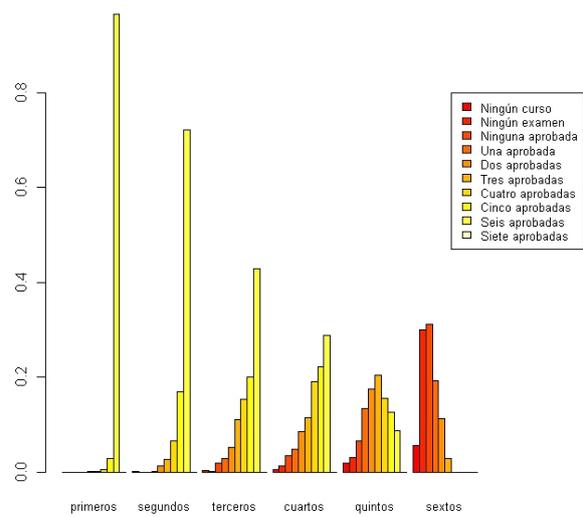
En los gráficos de barras (Figura) se observa que no hay mucha diferencia al interior de los grupos en cuanto al lugar de procedencia (Montevideo, interior o exterior), pero sí en cuanto al tipo de instituto de donde proviene, ya que los “primeros” y “segundos” tienen mayor proporción de estudiantes provenientes de liceos privados que el resto, y los “sextos”, mayor proporción de estudiantes de la UTU.

Distribución de materias de primero aprobadas en su primer año por grupo (generaciones 1990 a 1998)



(a) "prim1"

Distribución de materias de primero aprobadas en dos dos primeros años por grupo (generaciones 1990 a 1998)



(b) "prim2"

Figura 7: Distribución de la aprobación de las materias de primero por grupo

En las Figuras y se ve que los grupos se diferencian bastante por su comportamiento en "prim1", "prim2", "anu1" y "anu2". Los "primeros" en general aprueban todas las materias de primero en sus dos primeros años en facultad, mientras que de los "sextos" casi ninguno lo hace.

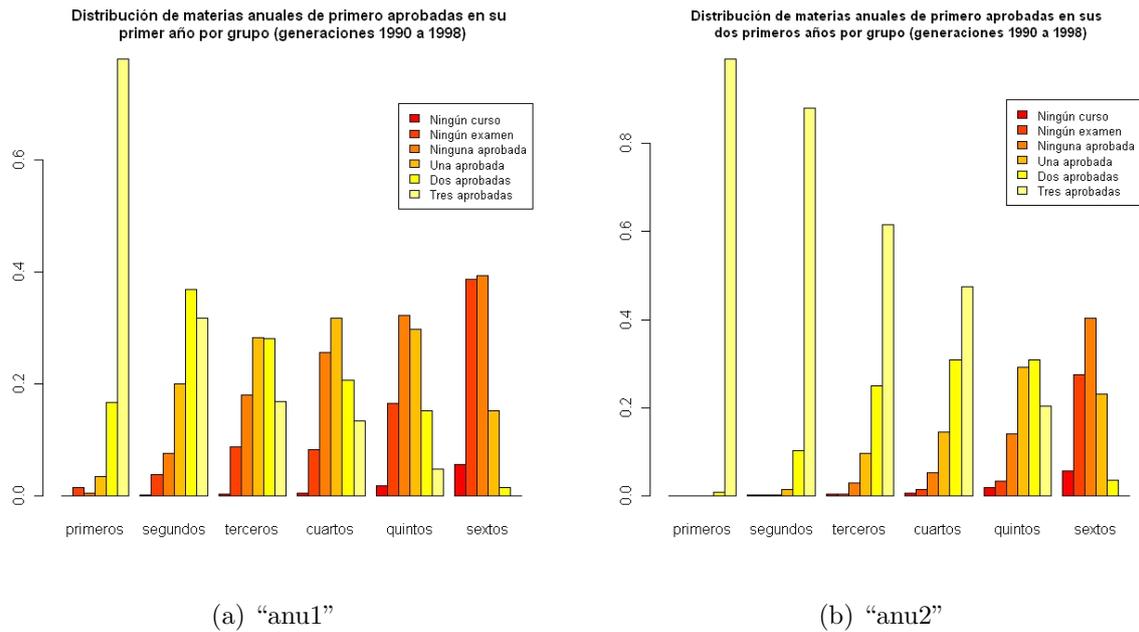


Figura 8: Distribución de la aprobación de las materias anuales de primero por grupo

4. Discusión

En el artículo se hace un análisis exploratorio de las trayectorias de los estudiantes en su carrera “principal”, entendiendo como tal aquella en que se inscribieron, o aquella donde tienen mayor avance en el caso de haberse inscrito a más de una carrera/ciclo.

La actuación durante los primeros dos años en Facultad, tanto en materias aprobadas como en escolaridad en esas materias, es decisiva para la trayectoria de los estudiantes. Los estudiantes que no aprueban materias y los que aprueban una o dos materias en su primer año, o en sus dos primeros años, son un grupo de alto riesgo respecto a no terminar sus estudios en carreras del Plan 90. Por otro lado, en los distintos grupos conformados por el análisis de *cluster* no hay mayores diferencias entre alumnos por lugar de procedencia o por sexo.

Un punto a destacar como positivo del análisis es que con variables que pueden calcularse fácil y rápidamente (en el primer año de actividad) es posible identificar con bastante precisión la actuación futura del estudiante. Estas variables son función de información que está disponible en las bases de Sistema de Gestión de Bedelías una vez finalizado el año académico (luego del período de marzo del año siguiente al ingreso).

Consideramos que hay variables que podrían contribuir de manera importante a la calidad de exploración y descripción: las variables que se encuentran en el Formulario Estadístico

que se llena al ingreso a Facultad. En este formulario hay información sobre educación de los padres, situación laboral, composición del hogar, etc. Si bien esta información es estática al momento de ingreso, es contemporánea con el primer año académico del estudiante y estimamos que sea de gran valor en futuros trabajos.

Otro tipo de variables, como grado de motivación del estudiante, vocación por la carrera, etc. posiblemente sean mejores o más completas como explicativas de fenómenos como deserción, avance rápido en la carrera, etc. Esa información debería ser estudiada por equipos interdisciplinarios pero sería más costosa en tiempo y recursos que la información mencionada en este artículo, la cual se encuentra disponible en bases de datos de la Universidad de la República.

5. Bibliografía

Referencias

- Blanco, Jorge (2006): *Introducción al análisis multivariado*. IESTA, UDELAR, Montevideo.
- D'Urso, Pierpaolo (2000): *Classificazione fuzzy per matrici a tre vie temporali* Tesi di Dottorato di Ricerca in Statistica Metodologica, XII Ciclo. Dipartimento di Statistica, Probabilità e Statistiche Applicate. Università di Roma "La Sapienza", Roma.