INSTITUTO DE ESTADISTICA
Universidad de la República
Facultad de Ciencias Económicas
y de Administración

¿Son útiles los modelos multiestado para analizar la trayectoria escolar?

Mara Dutto
Gabriela Mathieu

Agosto 2011

Serie DT (11/03)
ISSN : 1688-6453

Eduardo Acevedo 1139
Montevideo Uruguay
CP 11200

Tel: 2410 25 64 - 2418 73 81
Fax: 2410 46 34
e-mail: secretaria@iesta.edu.uy

¿SON ÚTILES LOS MODELOS MULTIESTADO PARA ANALIZAR LA TRAYECTORIA ESCOLAR?

María Dutto Gabriela Mathieu *

RESUMEN

Este trabajo se realizó en el marco del grupo de estudio sobre el Plan 90 de las carreras de Licenciatura en Economía, Contador Público y Licenciatura en Administración de la Facultad de Ciencias Económicas y de Administración. [1]

El objetivo es modelar las trayectorias de los estudiantes con modelos multiestado. Para esto se prueba el uso de dos paquetes del programa R, que trabajan con este tipo de modelos pero con enfoques diferentes: *mstate* y *msm*.

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. Como es un estudio exploratorio, a modo de ejemplo se toman únicamente los estudiantes de la generación 90 de la carrera de Contador Público.

Palabras clave: *educación, modelos multiestado, trayectoria escolar, análisis de supervivencia, R*

1. Introducción

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 originales utilizadas se muestran en la tabla 1

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

[1]El equipo está integrado por Ignacio Álvarez, Joaquín Amiel, Sebastián Castro, Natalia Da Silva, María Dutto, Juan José Goyeneche, Gabriela Mathieu y Guillermo Zoppolo, del Instituto de Estadística.

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)
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 1: Variables utilizadas

Además, a partir de los datos se generaron nuevas variables que podrían llegar a explicar la trayectoria escolar, que se muestran en la tabla 2. [2] Estas variables refieren al desempeño de los estudiantes en sus primeros años y con ellas se quiere explicar la trayectoria futura.

[2]Las 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.

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	categórica	departamento o país del instituto de procedencia
prim1	categórica	cantidad de materias de primero aprobadas en su primer año en facultad
prim2	categórica	cantidad de materias de primero aprobadas en sus dos primeros años en facultad
anu1	categórica	cantidad de materias anuales de primero aprobadas en su primer año en facultad
anu2	categórica	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
escol	numérica	escolaridad a marzo de 2010
acum	numérica	créditos (horas) acumuladas a marzo de 2010

Tabla 2: Variables generadas

2. Modelos multiestado

Los modelos multiestado buscan analizar la trayectoria de individuos, que pueden pasar por diferentes estados. No todas las transiciones entre estados son posibles, y definir cuáles lo son es una parte central de la modelación. Un mismo problema se puede modelar con estructuras de estados distintas. En las trayectorias educativas, por ejemplo, se podría definir como estados posibles la condición de "activo", "inactivo" o "egresado", asumiendo que se puede pasar de "activo" a "inactivo" y viceversa y de cualquiera de ellos a "egresado". Otra forma sería definir como estados: "activo", "inactivo por primera vez", "activo por segunda vez", "inactivo por segunda vez", "egresado",etc, redefiniendo las transiciones posibles (de "inactivo por primera vez" se puede pasar solamente a "activo por segunda vez" o a "egresado").

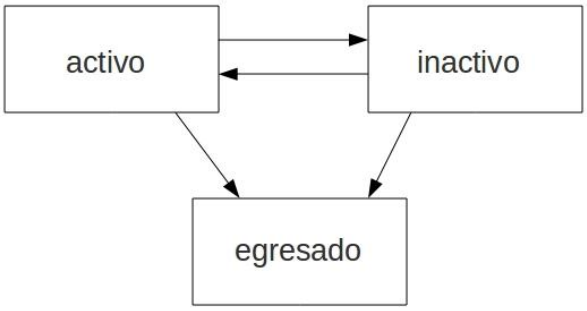

Figura 1: Ejemplo de grafo

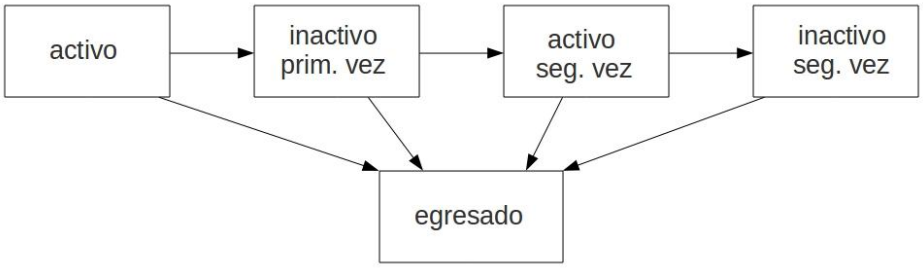

Figura 2: Segundo ejemplo de grafo

La estructura de estados se puede representar mediante un grafo, como puede verse en las figuras 1 y 2. Las flechas muestran las transiciones posibles entre los estados.

No siempre se puede ver todo el proceso. En algunos casos las observaciones están censuradas. [3] Quizás la censura más común y a la que nos enfrentaremos al estudiar las trayectorias de los estudiantes es la censura por la derecha, es decir, cuando un proceso se observa hasta determinado momento del tiempo T (salvo que en ese tiempo el proceso esté en un estado absorbente. [4])

2.1. Modelos markovianos

Detrás de los modelos multiestado hay un proceso estocástico en tiempo continuo. Este se define como una familia de variables aleatorias $\{X_t\}$ definidas en un espacio de probabilidad común $(\Omega, \mathrm{A}, \mathbf{P})$ cuyo índice t, el tiempo, toma valores en el conjunto $[0, \infty)$. Si fijamos $\omega \in$

[3] No debe confundirse con una observación truncada, que es cuando hay una exclusión sistemática de tiempos de supervivencia y la exclusión depende del tiempo mismo (por ejemplo, se quitan todos los tiempos superiores a un umbral).

[4] Un estado es absorbente si una vez que el proceso entra en él no puede salir (en nuestro caso "egreso" es absorbente).

Ω, obtenemos la función $\{X_t(\omega)\}$ al variar t; cada una de esas funciones es una trayectoria del proceso (Petrov y Mordecki, 2008). Vamos a trabajar con procesos estocáticos en tiempo continuo, pero con espacio de estados S discreto. Al número de estados le llamaremos n.

La ley del proceso puede determinarse por las matrices de transición de dimensión $n \times n$ cuyos elementos son las probabilidades de transición entre un estado y otro en cada momento t o por las intensidades de transición.

Las probabilidades de transición son:

$$\mathbf{P}_{mk}(v,t) = \mathbf{P}(X_t = k | X_v = m, F_{t-})$$

donde k es el estado en el que está el proceso al tiempo t y F_{v-} es la σ-álgebra que cubre la información en $[0,v]$, es decir, la historia del proceso. La probabilidad está condicionada a la trayectoria de proceso hasta el tiempo v.

La intensidad de la transición del estado m al k (función *hazard*) se define como:

$$\lambda_{mk}(t|X_u, u \in [0,t]) = \lim_{\Delta t \to 0} \mathbf{P} \frac{(X_{t+\Delta t} = l | F_{t-}, X_t = m)}{\Delta t}$$

Representa el riesgo instantáneo de moverse de un estado a otro (Jackson, 2011). Dado un tiempo t y una historia del proceso, las intensidades de transición forman una matriz Q, de dimensiones $n \times n$, cuyas filas suman 0 (los elementos de la diagonal quedan determinados por los otros elementos de la fila).

Se obtiene un modelo markoviano si asumimos que las probabilidades de transición no dependen de la historia del proceso, es decir, el comportamiento futuro depende del estado presente pero no de cómo se llegó hasta ahí. En términos matemáticos esto quiere decir que:

$$\mathbf{P}_{mk}(v,t) = \mathbf{P}(X_t = k | X_v = m)$$

Además, un proceso markoviano es homogéneo si suponemos que las intensidades de transición son constantes como funciones del tiempo (en ese caso las notamos simplemente λ_{mk}). Es intuitivo ver que si el proceso es homogéneo $\mathbf{P}_{mk}(v,t) = \mathbf{P}_{mk}(0, t-v)$. En este contexto, el tiempo en que el proceso se encuentra en un determinado estado r tiene una distribución exponencial, con tasa λ_{rr} (el r-ésimo elemento de la diagonal de la matriz Q), por lo que su esperanza es $1/\lambda_{rr}$.

2.2. Modelos Markov extendidos

Un caso más general es el de los llamados modelos Markov extendidos, que asumen que las intensidades de transición λ_{mk} dependen del tiempo en el que el proceso permaneció en el estado m ($t - T$, donde T es el momento en el que ingresó al estado m) y del tiempo t. En

particular, cuando las intensidades dependen de $t - T$ pero no de t se habla de procesos semi-Markov. (Hougaard, 1999; Commenges, 1999)

La heterogeneidad de las intensidades de transición para distintos individuos puede ser explicada en parte con la inclusión de algunas variables en el modelo. Suponemos que las características de un determinado individuo pueden ser resumidas en un vector de variables $Z(t)$ que en principio pueden variar con el tiempo.

Un aspecto atractivo de los modelos multiestado es que se pueden aplicar los modelos de análisis de supervivencia univariados basados en la función *hazard* en tiempo continuo (Andersen y Phoar Perme, 2008) como Weibull, Exponencial, Gomperz, Log-logístico, Lognormal, Gamma Generalizado, de riesgos proporcionales (o de Cox) y de tiempo de fallo acelerado (Jenkins, 2005).

El modelo "semiparamétrico" de Cox, uno de los más usados, es para una intensidad genérica:

$$\lambda(t|Z_i(t)) = \lambda_0(t)\exp(\beta' Z_i(t)) = \lambda_0(t)\alpha_i(t)$$

donde β es el vector de parámetros con el que se combinan linealmente las variables explicativas[5], $\lambda_0(t)$ es la línea de base de la función *hazard* (cuando todas las variables son 0) y $\alpha_i(t)$ es específica para cada individuo. Se puede ver que:

$$\beta_k = \frac{d\log\lambda(t|Z(t))}{dZ_k(t)}$$

se interpreta como la elasticidad de la intensidad respecto a la variable $Z_k(t)$, o sea, resume el efecto proporcional en la función *hazard* de un cambio en $Z_k(t)$.

3. Implementación

En el programa estadístico R, existen varios paquetes que permiten trabajar con modelos multiestado. En este trabajo usamos el *msm* y el *mstate*.

3.1. Paquete *msm*

El paquete *msm* permite ajustar modelos multiestado markovianos en tiempo continuo, cuando se observan exactamente los tiempos de transición o incluso cuando el proceso se observa en tiempos arbitrarios (es decir, cuando se desconoce que pasó entre observaciones). También permite trabajar con datos censurados e incluir estados absorbentes.

Admite la incorporación de variables explicativas para las intensidades de transición, dependientes o no del tiempo. En el caso de las dependientes del tiempo se asume que

[5]Los β_k también pueden asumirse dependientes del tiempo.

son constantes a tramos, pudiendo cambiar solamente en los tiempos observados (o de censura). Dentro de este paquete, la función *msm*() implementa la estimación máximo verosímil de los parámetros del modelo.

La matriz de datos debe contener al menos el identificador del sujeto, el tiempo de las observaciones y el estado observado del proceso. En la tabla 3 se puede ver la estructura de los datos utilizados [6].

estci	tiempo	status	acum	escol	sexo	edad .ing	tip inst	prim1	prim2	anu1	anu2
1	0,00	1	0,00	0,00	F	28	Rel	6	7	3	3
1	8,76	2	0,61	3,51	F	28	Rel	6	7	3	3
1	9,73	1	0,65	3,50	F	28	Rel	6	7	3	3
1	15,67	3	1,00	4,00	F	28	Rel	6	7	3	3
2	0,00	1	0,00	0,00	M	36	Pco	7	7	3	3
2	5,92	3	1,00	7,59	M	36	Pco	7	7	3	3
3	0,00	1	0,00	0,00	M	27	Pco	1	1	0	0
4	0,00	1	0,00	0,00	M	25	Pco	3	4	1	2
4	7,90	2	0,31	5,91	M	25	Pco	3	4	1	2
4	7,92	1	0,32	5,75	M	25	Pco	3	4	1	2
4	9,92	2	0,32	5,75	M	25	Pco	3	4	1	2

Tabla 3: Estructura de los datos utilizados (*msm*)

Nos quedamos solo con los estudiantes inscriptos a la carrera 4-1 (Contador Público) de la generación 1990 y las materias que corresponden a esa carrera. Luego de procesar los datos dejamos solo los tiempos en los que se dan las transiciones, ya que los conocemos exactamente. Por lo tanto, el estudiante que tiene una observación en el tiempo cero y ninguna otra es porque sigue activo a marzo de 2010 (fecha de corte de la base).

La función *statetable.msm*() cuenta el número de veces que los individuos experimentaron cada una de las transiciones posibles. El resultado para nuestros datos puede verse en la tabla 4, en base a los 901 estudiantes de la generación 1990 de la carrera de Contador Público.

statetable.msm	1	2	3
1	0	668	341
2	161	0	17

Tabla 4: Resultado de *statetable.msm*() aplicada a los datos

A priori no parecería razonable asumir un modelo markoviano en nuestro caso porque por

[6] Las cédulas de los estudiantes (variable "estci") se cambiaron para preservar su identidad

ejemplo la probabilidad de pasar de "activo" a "egresado", no debe ser igual si el estudiante estuvo siempre en "activo" o pasó una o varias veces por "inactivo". Sin embargo, hicimos la prueba de estimar el modelo.

Es necesario crear una matriz Q inicial, que sea coherente con los estados y las transiciones posibles del modelo (en este caso representado en el grafo de la figura 1). En la diagonal se ponen todos ceros, ya que como vimos, los elementos de la diagonal se obtienen por diferencia para que la suma de la fila de cero. También se ponen ceros en las transiciones que no son posibles.

Luego, con la función *crudeinits.msm*() se calcula la semilla de la matriz Q para usar en la estimación del modelo por máxima verosimilitud, usando la Q inicial y los datos. Es necesario explicitar cuál es la variable de la base que indica el estado en el que está el proceso y cuál el tiempo.

El modelo se estima con la función *msm*(). Como parámetros se incluyen las covariables y se explicita que observamos los tiempos exactos en los que se dan las transiciones. Las variables a incluir no pueden tener datos faltantes, por lo tanto optamos por no usar "tipinst" y "lugar.inst" [7]. Otro parámetro de la función *msm*() que se puede ajustar es *pci*, que indica los tiempos de corte en los que las intensidades son constantes. Incluimos las variables "sexo", "edad.ing", "prim1", "prim2", "anu1", "anu2", "escol" y "acum".

Al estimar el modelo en nuestro caso devuelve una advertencia por los estudiantes en los que no se observó ninguna transición (porque siguen activos). Devuelve también la matriz de intensidades de transición estimada (que no depende del tiempo) evaluada en los valores medios de las covariables. También se pueden obtener las matrices de probabilidades de transición para cada tiempo. Devuelve además una estimación del tiempo medio de permanencia en el estado 1 ("activo") y en el estado 2 ("inactivo") [8] (ver tabla 5). Se probó con distintas combinaciones de covariables y esto no mejora.

	estimates	SE	L	U
State 1	5,267354	0,3442681	4,634031	5,987231
State 2	1,287592	0,1566944	1,014356	1,634428

Tabla 5: Tiempo medio estimado de permanencia

Para hacer el diagnóstico de la adecuación del modelo se puede graficar la prevalencia observada (el porcentaje de individuos que está en cada estado en cada momento) contra la estimada por el modelo (figura 3). En nuestro caso se nota claramente que las dos curvas están muy alejadas para todos los estados, por lo que el modelo no es adecuado.

[7] En el caso de "tipinst" podríamos haber imputado, ya que eran solo dos datos faltantes, en el caso de "lugar.inst" eran la mayoría, ya que esta información se empezó a relevar más adelante

[8] No incluye al estado 3 ("egresado") porque es absorbente.

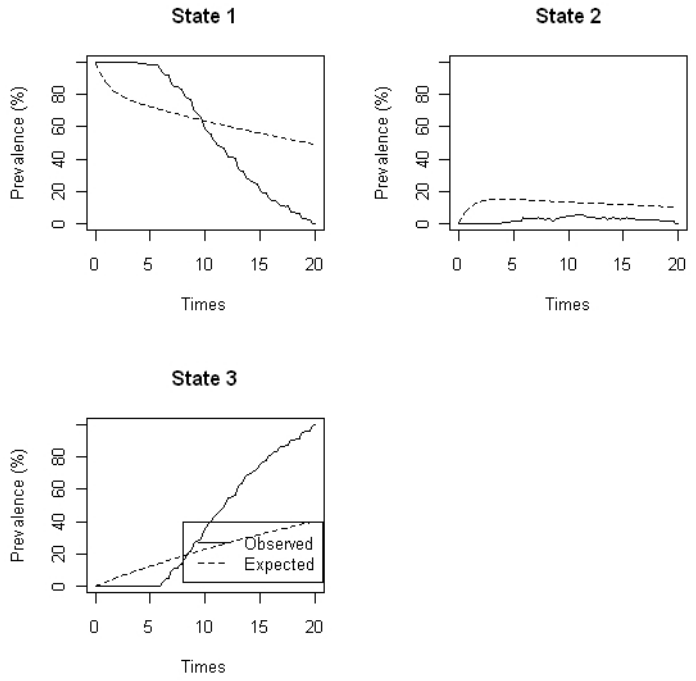

Figura 3: Gráfico diagnóstico para el modelo

Por lo tanto, buscamos que las intensidades fueran constantes solo a tramos: $[0, 5)$,$[5, 10)$ y más de 10, ya que entre otras cosas nadie se puede recibir antes de los 5 años, por lo que la función *hazard* de "inactivo" o "activo" a "egresado" tendría que ser nula antes de los 5 años. Esto se realiza agregando $pci = c(5, 10)$ (o solo $pci = 5$, según se quiera) como parámetro de la función $msm()$. El problema es que no se puede estimar (hay que seguir investigando el motivo); devuelve el error *numerical overflow in calculating likelihood*.

3.2. Paquete *mstate*

El paquete *mstate* también se puede utilizar para estimar modelos multiestado. Por un lado, es más amplio que el *msm* porque no se restringe a modelos markovianos, pero por otro, no permite incluir transiciones bidireccionales entre estados (el grafo de la figura 1 no se puede estimar con este paquete). Por lo tanto, para este trabajo se optó ajustar dos modelos: uno similar al grafo de la figura 2 pero con ocho estados porque hay estudiantes que han tenido hasta tres períodos de inactividad intermedios y otro con tres estados: "siempre activo", "alguna vez inactivo" y "egresado". Este modelo es similar al del grafo de la figura 1, pero sin la posibilidad de pasar de "inactivo" a "activo" (todos comienzan en el estado "activo").

En primer lugar debemos organizar los datos de tal manera de que puedan ser usados

como insumo para la función *msm*(). Uno de los formatos admisibles se muestra en la tabla 6; cada fila corresponde a una transición posible para el estudiante. En las columnas se indica la trayectoria de los estudiantes (los tiempos en años y los estados efectivamente visitados), por ejemplo, "$recib_t$" es el tiempo en el que el estudiante se recibió o el tiempo de corte de la base en caso de que no se haya recibido, y "$recib_s$" vale 1 si se recibió y 0 en otro caso.

estci	$recib_t$	$recib_s$	$inact1_t$	$inact1_s$	$inact2_t$	$inact2_s$	$inact3_t$	$inact3_s$
1	15,7	1	8,8	1	15,7	0	15,7	0
2	5,9	1	5,9	0	5,9	0	5,9	0
3	20,0	0	20,0	0	20,0	0	20,0	0
4	20,0	0	7,9	1	9,9	1	20,0	0
5	10,3	1	10,3	0	10,3	0	10,3	0
6	20,0	0	6,3	1	20,0	0	20,0	0
$react1_t$	$react1_s$	$react2_t$	$react2_s$	$react3_t$	$react3_s$	sexo	...	tipinst
9,7	1	15,7	0	15,7	0	F	...	Rel
5,9	0	5,9	0	5,9	0	M	...	Pco
20,0	0	20,0	0	20,0	0	M	...	Pco
7,9	1	20,0	0	20,0	0	M	...	Pco
10,3	0	10,3	0	10,3	0	M	...	Pco
20,0	0	20,0	0	20,0	0	M	...	Pri

Tabla 6: Estructura de los datos construida a partir del SGB

Con la función *msprep*() se transforman los datos que estaban en el formato anterior para poder ser usados por *mstate*. Para esto se debe construir la matriz T [9] que indica qué transiciones son posibles y las numera.

$$T = \begin{pmatrix} 1 & & & & & & 2 \\ & 3 & & & & & 4 \\ & & 5 & & & & 6 \\ & & & 7 & & & 8 \\ & & & & 9 & & 10 \\ & & & & & 11 & 12 \\ & & & & & & 13 \\ & & & & & & \end{pmatrix}$$

El resultado se muestra en la tabla 7. Cada estudiante tiene seguro dos filas con las transiciones posibles (de 1 a 2 y de 1 a 8); si pasa al estado 2 se le abren otras dos posibilidades: pasar del 2 al 3 y del 2 al 8. Si efectivamente llega al estado 3, de ahí puede pasar al estado 4 ó al estado 8, como ocurre con el estudiante 1 ($id = 1$). En cada

[9] Los casilleros vacíos son NA en el R

transición posible (se numeran mediante "trans"), se miden el tiempo al inicio ("Tstart") de esa transición y al final ("Tstop"), siendo la variable "time", el tiempo transcurrido. Finalmente, la variable "status" da cuenta si la transición en cuestión se verificó (vale 1) o no (vale 0).

id	from	to	trans	Tstart	Tstop	time	status	sexo	...
1	1	2	1	0,0	8,8	8,8	1	F	...
1	1	8	2	0,0	8,8	8,8	0	F	...
1	2	3	3	8,8	9,7	0,9	1	F	...
1	2	8	4	8,8	9,7	0,9	0	F	...
1	3	4	5	9,7	15,7	6,0	0	F	...
1	3	8	6	9,7	15,7	6,0	1	F	...
2	1	2	1	0,0	5,9	5,9	0	M	...
2	1	8	2	0,0	5,9	5,9	1	M	...
3	1	2	1	0,0	20,0	20,0	0	M	...
3	1	8	2	0,0	20,0	20,0	0	M	...

Tabla 7: Tabla de datos en el formato propio de *mstate*

La función *events*() resume la información contenida en la base de datos en cuanto a las transiciones. Las tablas 8 y 9 (en porcentajes) se leen por fila, por ejemplo, de los que entraron estaban activos, 61 % pasaron a "inactivos por primera vez", 37 % egresaron y el 3 % seguía activo hasta el 2010. Como puede verse, la fila de egreso es nula, ya que una vez que el estudiante egresa no puede pasar a ningún estado. Es interesante ver que de los estudiantes que alguna vez fueron inactivos, solo el 5 % se había recibido luego de 20 años de observación.

	act	inact1	react1	inact2	react2	inact3	react3	egr	no event
act	0	61	0	0	0	0	0	37	3
inact1	0	0	26	0	0	0	0	3	71
react1	0	0	0	78	0	0	0	9	14
inact2	0	0	0	0	17	0	0	1	82
react2	0	0	0	0	0	63	0	0	37
inact3	0	0	0	0	0	0	17	0	83
react3	0	0	0	0	0	0	0	0	100
egr									

Tabla 8: Salida de la función *events*() para el caso de ocho estados

	act	inact1	egr	no event
act	0	61	37	3
inact1	0	0	5	95
egr				

Tabla 9: Salida de la función *events*() para el caso de tres estados

La función *msfit*() permite estimar los parámetros y las intensidades de transición (cuando todas las covariables son 0), sus varianzas y covarianzas. Devuelve tres objetos: *Haz* (la estimación de la función *hazard* acumulada para cada transición), *varHaz* (las varianzas y covarianzas de las funciones anteriores), y *trans* (la matriz de transición).

Posteriormente utilizando las estimaciones de *msfit*() y la función *probtrans*() se pueden obtener las probabilidades de transición para un estudiante en particular.

La función *expand.covs*() permite expandir las covariables de manera que los coeficientes β_k puedan ser diferentes para cada transición.

Se estimaron varios modelos usando la función *coxph*() usando *strata(trans)* lo que sería equivalente a estimar un modelo de cox para cada transición con la misma función. Se hizo tanto para el caso de los ocho estados como para el de tres, con distintas combinaciones de variables. La prueba de razón de verosimilitud para el modelo en su conjunto en todos los casos da significativa al 5 %, pero sin embargo el R^2 es muy bajo [10].

Quedan por explorar muchas herramientas del paquete *mstate*, lo que se expuso en este trabajo es solo una parte de sus potencialidades. Por razones de tiempo no se pudo profundizar en sus funciones para el diagnóstico de la bondad del modelo ni para la predicción.

4. Conclusiones

Del análisis realizado vemos que asumir que los datos provienen de un proceso markoviano no da un buen ajuste del modelo, según se desprende de aplicar el paquete *msm*.

Más allá de esto, sería interesante incluir otras covariables que podrían explicar la trayectoria escolar como si trabaja y cuántas horas. Sería bueno explotar la información proveniente de los censos estudiantiles, de la que aún no disponemos.

Otras variables que podrían incorporar y con las que contamos son: si el estudiante realiza otra carrera de la facultad (además de la que se tomó como principal) y el origen del liceo del que proviene (Montevideo o Interior).

[10] Resta profundizar en el alcance del R^2 en este caso.

No parece conveniente la inclusión de "escol" y "acum" al punto de corte de la base, ya que estamos viendo "el final de la película" (estamos haciendo depender una intensidad de transición de información futura).

Como se dijo al principio, este trabajo busca explorar los modelos multiestado para analizar las trayectorias de los estudiantes. No pretende dar por cerrada la respuesta sino que más bien busca abrir una línea de investigación, muy incipiente aún. Resta mucho por profundizar, tanto desde el punto de vista teórico como instrumental. Lo que se presentó es solo un pequeño avance en esta dirección, el puntapié inicial para poner en orden las ideas.

5. Bibliografía

Referencias

Andersen, P.K. y Pohar Perme, M. (2008): "Inference for outcome probabilities in multi-state models". En: *Lifetime Data Analysis* No.14, Springer Science+Business Media, pp. $405 - 431$. DOI: $10{,}1007/s10985 - 008 - 9097 - x$

Commenges, D. (1999): "Multi-state Models in Epidemiology" En: *Lifetime Data Analysis* No.5, Boston: Kluwer Academic Publishers. pp. $315 - 327$

Guttorp, P. (1995): "Chapter 3. Continuous Time Markov Chains" En: *Stochastic Modeling of Scientific Data* Londres: Chapman & Hall. pp. $125 - 188$

Hougaard, P. (1999): "Multi-state Models: A Review" En: *Lifetime Data Analysis* No.5, Boston: Kluwer Academic Publishers. pp. $239 - 264$

Jackson, C.H. (2011): "Multi-State Models for Panel Data: The msm Package for R" En: *Journal of Statistical Software* Volumen 38, No.8, Enero de 2011. Disponible en: `http://www.jstatsoft.org` [Consulta: 10/3/2011]

Jackson, C.H. (2007): *Multi-state modelling with R: the msm package* Versión 0.74. Disponible en `rss.acs.unt.edu/Rdoc/library/msm/doc/msm-manual.pdf` [Consulta: 28/7/2011]

Jenkins, S. (2005): *Survival Analysis* Notas de curso no publicadas. Institute for Social and Economic Research, University of Essex. Disponible en: `http://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf` [Consulta: 28/7/2011]

Petrov, V. y Mordecki, E. (2008): *Teoría de la probabilidad* Montevideo: DIRAC. Segunda edición.

Putter, H. (2011): *Tutorial in biostatistics: Competing risks and multi-state models. Anal-*

yses using the mstate package Disponible en: cran.r-project.org/web/packages/mstate/vignettes/Tutorial.pdf[Consulta 1/7/2011]

Putter, H.; Fiocco, M. y Geskus, R.B. (2007): "Tutorial in biostatistics: Competing risks and multi-state models" En: *Statistics in Medicine* No.26, pp. 23892430. Disponible en: www.interscience.wiley.com DOI:10.1002/sim.2712