



UNIVERSITAT DE BARCELONA

Modelització de la durada dels estudis universitaris: una aplicació a la Universitat de Barcelona

Ramon Alemany Leira

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

Modelització de la Durada dels Estudis Universitaris: Una Aplicació a la Universitat de Barcelona

Ramon Alemany Leira

Tesi dirigida pel Dr. Manuel Artís Ortuño en el marc del programa de doctorat "Mètodes Quantitatius en Economia" de la Universitat de Barcelona.

Departament d'Econometria,
Estadística i Economia Espanyola.

Barcelona, Novembre de 1993

Arribat el moment de posar punt i final al present treball resulta inevitable la mirada enrera i la constatació dels molts agraïments que haig d'expressar a totes aquelles persones que, d'una manera o altra, m'han ajudat en la seva elaboració.

Agraeixo al Dr. Manuel Artís, director d'aquesta tesi, l'haver-me incorporat a una línia de recerca tan profitosa i interessant, de la qual aquest treball no és més que una primera etapa, i la seva ajuda, especialment en la superació dels entrebancs que ens anàvem trobant i en haver posat a la meva disposició tots els mitjans que he necessitat.

Al Sr. Puig Salellas, President del Consell Social de la Universitat de Barcelona, el qual ens va donar peu a endinsar-nos en la problemàtica que aquí s'aborda, i a Jordi Camós, Secretari Tècnic, per la seva col·laboració.

A Gaspar Rosselló i Toni Forés del Servei de Gestió Acadèmica de la Universitat de Barcelona, l'haver-me facilitat tota la informació dels estudiants sense la qual aquest treball no s'hauria pas pogut realitzar. En aquest sentit, la col·laboració de Montse Casas i Fermí Osuna de la Secretaria de la Facultat d'Econòmiques va ser també molt important.

A Victor Cano i Francis Martín, del Departament d'Economia Aplicada de la Universidad de la Laguna, per les hores que van dedicar a la recaptat de referències bibliogràfiques que m'han estat de moltíssima utilitat.

Als meus companys del Departament d'Econometria, Estadística i Economia Espanyola, que sempre m'han fet costat i, especialment, a Montse Guillén i Eli Torrelles, la intervenció de les quals ha estat de molta ajuda en les darreres fases d'aquest treball.

Als meus pares, pel seu esforç, sacrifici i exemple que han estat fonamentals en la meva formació personal i professional. Als meus germans i amics que m'han animat en tot moment.

A la Marga, per la seva paciència, el seu respecte i el seu carinyo, i a la Mireia, que sempre té a punt un somriure gratificant, per les moltes hores que els hi he robat. Amb totes dues i la perspectiva del/la que vindrà, el camí ha estat molt més planer.

Barcelona, 18 de novembre de 1993

Index

1 Introducció	1
 Primera Part: Metodologia	11
 2 Antecedents i consideracions metodològiques	13
 3 Dades de durada o Supervivència: Introducció	21
3.1 Conceptes teòrics bàsics	24
3.1.1 Variable Temps contínua	26
3.1.2 Variable Temps discreta	27
3.2 Censuració	29
 4 Modelització de Poblacions Homogènies	35
4.1 Mètodes No Paramètrics en Poblacions Homogènies	35
4.1.1 L'estimador del Producte Límit de Kaplan-Meier	35
4.1.2 L'estimador de Taula de Vida	40
4.1.3 Comparació de Funcions de Supervivència	42
4.1.3.1 Contrast de <i>log-rank</i>	42
4.1.3.2 Contrast de Wilcoxon generalitzat	45

4.2 Mètodes Paramètrics en Poblacions Homogènies	46
4.2.1 La Distribució Exponencial	47
4.2.2 La Distribució Weibull	48
4.2.3 La Distribució Lognormal	49
4.2.4 La Distribució Log-Logística	50
4.2.5 La Distribució Gamma	51
4.2.6 La Distribució Gamma generalitzada	52
4.2.7 Criteris de selecció de les distribucions	52
4.2.8 Inferència en els models paramètrics per a poblacions homogènies	54
4.2.8.1 La funció de Versemblança: el cas continu	55
4.2.8.2 La funció de Versemblança: el cas discret	56
4.2.8.3 Estimació de paràmetres i contrast d'hipòtesis	57
5 Modelització de Poblacions Heterogènies	61
5.1 Models de Durada Accelerada (AFT)	64
5.2 Models d'Atzar Proporcional (PH)	71
5.2.1 Estimació paramètrica dels models d'atzar proporcional	73
5.2.2 Estimació semiparamètrica contínua del model d'atzar proporcional	75
5.2.3 Estimació semiparamètrica discreta del model d'atzar proporcional	86
5.2.3.1 Mètode de Prentice i Gloeckler (1978)	87
5.2.3.2 Mètode de Cox (1972)	92
5.3 Variables explicatives canviants en el temps	97
5.4 Múltiples causes de finalització de la durada	102

Segona Part: Aplicació	107
6 La base de dades	109
6.1 Estructura i fonts d'informació	110
6.2 Etapes en la construcció de la base de dades	112
6.2.1 Mètodes d'elecció de la població a estudiar	112
6.2.2 La Cohort del curs 1985-86	114
6.3 Descripció de les variables	116
6.3.1 Variables Personals	117
6.3.1.1 La variable SEXE	117
6.3.1.2 La variable EDAT	118
6.3.1.3 La variable ESTAT CIVIL	119
6.3.1.4 La variable NACIONALITAT	120
6.3.1.5 La variable DOMICILI	120
6.3.1.6 La variable TIPUS DE RESIDÈNCIA	120
6.3.1.7 La variable SITUACIÓ LABORAL	121
6.3.2 Variables de Gestió Acadèmica	122
6.3.2.1 La variable DURADA DELS ESTUDIS	122
6.3.2.2 La variable SITUACIÓ ACADÈMICA DE L'ALUMNE	123
6.3.2.3 La variable FONT DE FINANÇAMENT DELS ESTUDIS	125
6.3.2.4 La variable CLASSE DE MATRÍCULA	126
6.3.2.5 La variable CENTRE	127
6.3.2.6 La variable HORARI DE CLASSES	127
6.3.3 Variables de Rendiment Acadèmic	128

6.3.4	Variables Acadèmiques prèvies	141
6.3.4.1	La variable TÍTOL D'INGRÉS	141
6.3.4.2	La variable RESULTATS DE LES PROVES D'ACCÉS (PAAU)	141
6.4	Transformacions sobre les variables	144
6.5	Subpoblacions de referència	147
7	Resultats de les modelitzacions estàtiques	151
7.1	Model de probabilitat lineal pels abandonaments	154
7.2	Model de probabilitat lineal pels llicenciats	157
7.3	Relació amb la variable "durada dels estudis"	158
7.4	Una altra alternativa clàssica: l'anàlisi discriminant	165
8	Resultats previs de l'anàlisi de dades de supervivència	171
8.1	Anàlisi Prèvies	173
8.1.1	Subpoblació d'Abandonaments	174
8.1.1.1	Variables Personals	174
8.1.1.2	Variables de Gestió Acadèmica	188
8.1.1.3	Variables Acadèmiques Prèvies	196
8.1.1.4	Conclusions Prèvies	200
8.1.2	Subpoblació de Llicenciats	204
8.1.2.1	Variables Personals	204
8.1.2.2	Variables de Gestió Acadèmica	218
8.1.2.3	Variables Acadèmiques Prèvies	226
8.1.2.4	Conclusions Prèvies	230

9 Resultats de la modelització paramètrica	235
9.1 Sub població d'abandonaments	238
9.2 Sub població de llicenciats	248
9.3 Conclusions prèvies	254
10 Resultats de la modelització semiparamètrica contínua	257
10.1 Sub població d'abandonaments	261
10.2 Sub població de llicenciats	287
11 Resultats de la modelització semiparamètrica discreta	301
11.1 Sub població d'abandonaments	303
11.2 Sub població de llicenciats	326
12 Conclusions	337
Bibliografia	345

1. Introducció

Capítol 1

Introducció

Les Administracions Públiques amb competències en matèria d'ensenyament universitari, els responsables de la presa de decisions de política universitària, tant dins com fora de les institucions d'ensenyament superior, tenen un gran interès a conèixer les causes de l'abandonament dels estudis universitaris i, en general, els factors que determinen la finalització dels estudis, bé sia en forma d'èxit per l'obtenció del títol, bé sia en forma de fracàs per l'abandonament dels mateixos. El coneixement d'aquests factors, així com de la situació concreta en cada moment en termes de taxes de finalització dels estudis, fracàs o èxit, per universitats i ensenyaments, els serà d'utilitat en el disseny de la política universitària i en la planificació educativa.

El tractament d'aquesta problemàtica és objecte d'anàlisi tant pels especialistes en educació com pels investigadors en economia. Naturalment, l'abandonament dels estudis té una repercussió en termes de cost econòmic i representa un desaprofitament dels recursos d'una comunitat. Es tracta doncs, des d'aquest punt de vista, d'avaluar en quina mesura la Universitat satisfi les exigències de la societat que inverteix recursos en ella.

D'entre els estudis més rellevants en la descripció i l'anàlisi dels costos econòmics associats al procés educatiu universitari destaquen els treballs de Levin (1972) i Caterall (1987). En un entorn més proper a nosaltres cal destacar l'aportació de Moltó i Oroval (1984).

És habitual creure que aquelles persones que abandonen els estudis, no només desaprofiten l'oportunitat de beneficiar-se del producte universitari completament, sinó que no és estrany pensar que tindran unes possibilitats professionals de rang inferior respecte a aquelles altres que aconseguen llicenciar-se, malgrat que aquestes últimes hagin tingut un rendiment acadèmic per sota de la mitjana. En l'anterior plantejament es presenta la dualitat deguda al fet que existeixen dos col·lectius, l'anàlisi dels quals té un interès especial i mereix un estudi detallat. Es tracta per una banda dels estudiants que havent iniciat una carrera universitària, deixen els estudis de manera definitiva i per altra d'aquells estudiants que en un termini raonablement breu finalitzen els cursos acadèmics que condueixen a l'obtenció d'una titulació oficial.

Degut a la presència dels dos aspectes acabats d'assenyalar, utilitzarem el concepte global de durada dels estudis o permanència. La durada dels estudis fa referència al temps d'estada en el centre universitari, i és un concepte que a més inclou informació sobre la forma en que ha finalitzat l'estada, ja sigui deguda a l'acabament dels cursos o bé a l'abandonament voluntari dels mateixos.

Resulta d'utilitat emprar també el concepte de rendiment acadèmic en termes generals ja que és el que ens proporciona una analogia amb la idea del cost/benefici, i per tant ens dona la possibilitat d'obtenir una ratio que relacioni la quantitat d'input i d'output obtinguda en el procés d'ensenyament universitari.

Sovint aquest input es mesurarà en la quantitat d'assignatures cursades per l'estudiant, quantitat fàcilment assimilable al cost que suposen els seus estudis. L'output mesurarà el nivell d'èxit assolit per l'estudiant, la qualificació obtinguda en finalitzar el curs acadèmic.

Per a poder assolir un coneixement en profunditat dels aspectes econòmics dels estudis universitaris, és necessària la realització d'una anàlisi exhaustiva de la permanència a la Universitat. Això vol dir buscar quines són les causes que determinen l'èxit o l'abandonament dels estudis, en quina mesura és possible determinar aquestes causes i, en definitiva, quins són els principals indicadors que determinen la permanència en els estudis. Finalment, es preveu la possibilitat de poder extreure altres conclusions derivades com ara estratègies que permetin elaborar sistemes educatius amb una major rendibilitat.

La metodologia utilitzada per a tractar les dades que informaran sobre el concepte de permanència o durada dels estudis haurà d'ésser eminentment estadística. Es farà èmfasi en l'utilització de tècniques novedoses, fent les adaptacions més convenientes per a la seva utilització en aquest context. Degut a la temporalitat de la variable durada dels estudis, les tècniques més avançades van encaminades a l'utilització de l'anàlisi de supervivència. Aquests mètodes van ser desenvolupats inicialment en l'àmbit de la bioestadística, per a modelitzar el temps de vida (Cox, 1959, 1972, Crowley i Hu, 1977). Posteriorment la metodologia es va anar utilitzant en altres camps, com l'economia (Lancaster, 1979, Nickell, 1979 i Kiefer, 1988), la sociologia (Allison, 1982) i l'enginyeria (Lawless, 1982). Actualment les seves aplicacions s'adrecen també a l'anàlisi dels processos educatius com ara a les trajectòries professionals del professorat i la durada dels estudis (Willet i Singer, 1991).

Aquesta anàlisi de la durada dels estudis mitjançant la metodologia estadística de les dades de supervivència haurà de permetre l'obtenció d'importants resultats adreçats al disseny de la política universitària i la planificació educativa, d'entre els que destaquen, el coneixement dels mecanismes que regeixen la permanència, l'anàlisi econòmica cost-benefici de l'ensenyament superior i l'elaboració de normatives de permanència, així com el disseny de polítiques encaminades a millorar el rendiment acadèmic, l'elaboració de programes de prevenció de l'abandonament i la detecció de col·lectius d'alt risc.

Les polítiques de permanència troben justificació des de l'anàlisi cost-benefici de l'ensenyament superior. La productivitat dels recursos que les universitats proporcionen a l'entorn social és una qüestió difícil de concretar (Alemany et al., 1990). En primer lloc es fa necessària la definició i quantificació dels factors productius i del producte. D'una banda tenim l'òptica sota la qual es considera que els factors productius únicament fan referència als elements que formen part del pressupost de la Universitat: professorat, inversions, despeses corrents, etc. El producte es defineix com el nombre d'hores lectives per estudiant que imparteix la institució. En aquest context es obvi que un grup de 400 estudiants amb un únic professor que tingui un nivell de retribució mínima seria el paradigma d'una alta productivitat.

Alternativament es pot mesurar la productivitat mitjançant els recursos socials emprats per la Universitat. Des d'aquest enfocament, els agents que participen en l'activitat són uns altres. El producte anteriorment definit com hores de classe impartides, únicament es pot considerar un output intermedi, mentre que el producte final és ara el titulat universitari completament format. En aquesta perspectiva, l'abandonament dels estudis suposaria un cost individual i social que

està directament relacionat amb la pèrdua d'oportunitat d'acabar la carrera, el primer, i en el malbaratament dels recursos socials destinats a la Universitat, el segon.

Així les coses, les polítiques de permanència tindran com a objectiu la minimització d'aquests costos i com a instrument l'elaboració de normatives que permetin, d'una banda, avançar l'abandonament i, de l'altra, aconseguir un increment en el nombre de titulats amb una formació de qualitat en terminis raonables de temps.

L'origen del present treball es situa en l'encàrrec del Consell Social de la Universitat de Barcelona d'un estudi sobre el Rendiment Acadèmic dels estudiants per tal d'elaborar una normativa que reguli la permanència dels mateixos. L'elaboració i aprovació de l'esmentada normativa, tal com disposen la LRU i els Estatuts de la U.B., és competència del Consell Social i es fa necessària atès el grau de massificació de la U.B., entre d'altres raons (Alemany et al., 1990).

Qualsevol estudi del rendiment acadèmic dels estudiants amb l'objectiu d'orientar l'elaboració d'una normativa de permanència, s'ha de fonamentar en l'anàlisi de l'abandonament dels estudis. Partint de la base que tota normativa de permanència representarà uns costos, s'hauria de seleccionar/elaborar aquella que els minimitzés. Així, aquesta minimització vindrà donada per la capacitat de la normativa d'avançar l'abandonament dels estudis. D'altra banda, no s'ha de perdre de vista el fet que la normativa de permanència també generarà beneficis, fonamentalment derivats de la seva capacitat de motivar un major aprofitament de l'estudi i de generar unes millors condicions per aquest aprofitament conseqüència d'una menor massificació.

La comparació de les característiques personals, econòmiques i socials dels estudiants proporcionarà informació sobre la col·lectivitat universitària amb un rendiment acadèmic baix i que abandona els estudis, així com les causes que el provoquen. D'aquesta forma, podrà ser analitzat el cost des d'una perspectiva individual a la vegada que també globalment per tal de detectar la població que inclou el fracàs escolar, i així poder endegar els efectes correctors oportuns.

L'abandonament dels estudis, també dit fracàs escolar, s'hauria d'estudiar des de dos punts de vista. Per un costat, la vessant acadèmica, és a dir, tractant de conèixer i posar de manifest quin és el perfil acadèmic de l'estudiant que abandona en cadascun dels moments de la seva vida acadèmica. Això hauria de permetre l'establiment d'uns mínims en el rendiment acadèmic per sota dels quals l'alumne hauria de ser exclòs i que estiguessin fonamentats en el fet de què avancen l'abandonament. Per altra banda, l'anàlisi de la influència de les característiques personals sobre l'abandonament. Qualsevol valoració/predicció de l'abandonament a partir del rendiment acadèmic ha de tenir present el control de les característiques personals: sexe, edat, estat civil, situació laboral, etc.

En l'esmentat treball inicial l'objectiu de modelitzar la relació entre l'abandonament i les components acadèmiques i personals per tal de fonamentar una normativa de permanència que avancés/predis l'abandonament, va ser parcialment assolit. Existien, però, dues mancances importants: l'absència de la informació de les característiques personals dels estudiants que abandonaven els estudis i la inadequació/el replantejament de la tècnica estadística emprada. El primer problema ha estat parcialment resolt i serà detallat en el capítol sisè corresponent a la descripció de la base de dades utilitzada així com de les fonts d'informació disponibles.

El segon problema, la inadequació de la tècnica, passa pel replantejament de la mateixa amb uns instruments més depurats i adients als objectius de l'anàlisi. En el treball inicial es van emprar models dicotòmics successius (Logits en diferents moments del temps) per tal de mesurar la probabilitat d'abandonar a partir dels valors d'algunes variables acadèmiques. L'anàlisi dels resultats en diferents moments del temps ens permetia concloure quin indicador del rendiment acadèmic predia millor l'abandonament dels estudis però no la manera en que aquest indicador afectava a la decisió d'abandonar i encara menys la relació amb el temps que es trigava a prendre aquella decisió. Aquesta forma de procedir s'emmarca dins d'una metodologia estàtica clàssica la qual es tractada amb més detall en el capítol segon.

Però és precisament la necessitat d'analitzar els resultats en diferents moments del temps per tal de treure conclusions el que ens porta a fer un canvi en el plantejament del problema. Així, més que analitzar la decisió d'abandonar/no abandonar en diferents moments del temps, el que realment ens interessa és el temps que es triga a prendre la decisió d'abandonar, si es que es pren mai. Plantejat d'aquesta forma el problema, la tècnica a utilitzar ha de ser la de l'anàlisi de la supervivència. En els capítols tercer, quart i cinquè del present treball es realitza una descripció d'aquest mètodes associats a l'anàlisi de la supervivència. En el capítol tercer es presenten els conceptes bàsics sobre els quals es fonamenta la metodologia, mentre que en els capítols quart i cinquè es descriuen les alternatives de modelització per a poblacions homogènies (capítol quart) i heterogènies (capítol cinquè), és a dir, quan s'incorporen variables explicatives caracteritzadores de l'heterogeneïtat. En aquest punt del treball finalitzarà la primera part corresponent a la metodologia utilitzada.

La possibilitat de disposar de la informació sobre les característiques personals, socio-econòmiques i acadèmiques dels estudiants, permet una anàlisi estadística rigorosa de les dades de permanència a la Universitat. Mitjançant les eines estadístiques i econòmiques més adients es poden establir models de decisió en aquest context. Els nous desenvolupaments en aquest àmbit conjuntament amb la disponibilitat de les dades ens ha portat a determinar la metodologia de tractament més adequada i a realitzar una anàlisi de la durada dels estudis concreta, a més de donar les directrius interpretatives dels resultats. L'exposició que es durà a terme és de tipus general, fàcilment extrapolable a contextos similars. En la segona part del present treball s'efectua l'aplicació de la metodologia descrita a una cohort d'estudiants de la Universitat de Barcelona. Primerament, el capítol sisè presenta les fonts de la informació i les etapes en la construcció de la base de dades corresponent a la cohort seleccionada, així com una anàlisi descriptiva prèvia de cadascuna de les variables a utilitzar. El capítol setè mostra com les tècniques clàssiques amb un plantejament estàtic són insuficients per a assolir els objectius de caracterització dels factors que determinen finalització dels estudis en forma d'abandonament o d'obtenció del títol oficial. Els capítols següents, des del vuitè fins a l'onzè, tracten de la presentació dels resultats de la utilització dels mètodes de l'anàlisi de la supervivència a la cohort d'estudiants en la què es fonamenta l'aplicació.

Mitjançant la utilització de les tècniques estadístiques dinàmiques de l'anàlisi de la supervivència i la seva aplicació a una cohort d'estudiants concreta, volem detectar els factors que influeixen positivament en l'obtenció de l'output, el títol de llicenciat, i aquells que interrompen el procés productiu, l'abandonament. Aquesta detecció es realitzarà des del punt de vista derivat de l'existència de l'expedient de l'estudiant a més del coneixement d'algunes

característiques personals i socio-econòmiques i no pas a partir d'altres característiques externes (esdeveniments econòmics, socials, particularitats del professorat, etc.) o internes (hàbits personals i altres variables psico-pedagògiques).

Existeixen tres preguntes clau que motiven l'estudi:

1. Es cert que els estudiants tenen major probabilitat d'abandonar els estudis en certs moments de la seva trajectòria acadèmica?
2. La probabilitat d'abandonar els estudis, és possible que tingui una component diferent segons les característiques de l'estudiant?
3. Quina relació podem establir entre la probabilitat o risc d'abandonar i el concepte contrari de finalitzar els estudis ràpidament? Quina és la complementarietat que hi ha entre aquests dos conceptes?

El tractament d'aquestes qüestions ajudarà a la deducció de conclusions i recomanacions dirigides a la detecció de col·lectius d'alt risc per a la possible implantació de mecanismes de prevenció. La localització dels moments de la trajectòria acadèmica que requereixen més recursos pot ajudar a concentrar esforços per a evitar abandonaments innecessaris i incrementar el percentatge de graduats en un temps inferior.

Un estudi de les característiques del treball que es presenta a continuació podrà ser una eina de decisió important en la política educativa si la seva continuïtat en una segona etapa facilita informació sobre l'impacte dels canvis de plans d'estudi i de la entrada en vigor de la normativa de permanència.

El darrer capítol del present treball servirà per a jutjar els resultats de l'aplicació de la metodologia presentada, especialment pel que fa a les anteriors qüestions plantejades, així com la seva possible generalització a altres situacions i els posteriors desenvolupaments de la línia de recerca iniciada.

Primera Part:
Metodologia

2. Antecedents i consideracions metodològiques

Capítol 2

Antecedents i consideracions metodològiques

Generalment, el mètode utilitzat per a l'anàlisi de l'èxit o el fracàs en els estudis és el seguiment d'una cohort. Aquesta es defineix com el conjunt d'alumnes matriculats per primera vegada en el primer curs d'un ensenyament i en un determinat curs acadèmic. Així, són nombrosos els treballs publicats sobre el rendiment acadèmic, l'èxit o abandonament dels estudis realitzats en altres universitats espanyoles i estrangeres, i també a la Universitat de Barcelona. A Espanya els primers treballs es van dur a terme en els anys setanta a càrrec de Rubio (1968), el qual va analitzar l'abandonament de les cohorts des del 1960 fins el 1966 dels alumnes de les Escoles Tècniques Superiors de Madrid, i Benedito, Vicens i Altamira (1970), que van analitzar l'abandonament de la cohort de 1968 dels estudiants de la Universitat de Barcelona. Els posteriors treballs de García Aretio (1987) i de Salvador i García-Valcárcel (1989) realitzen una revisió exhaustiva de les investigacions realitzades en aquest àmbit a les universitats espanyoles citant, entre d'altres, a Herrero i Infestas (1980), Muñagorri i Martínez (1984), Moncada (1986) i les comunicacions presentades per González Tirados,

Latiesa, Oroval i altres a les *Jornadas Internacionales sobre Demanda de Educación Superior y Rendimiento Académico en la Universidad*, organitzades pel CIDE a Madrid l'any 1986. Posteriorment s'han publicat altres importants aportacions com les de García Llamas (1986) i Jiménez Fernández (1987). Pel que respecte als treballs publicats sobre la base dels estudiants d'universitats estrangeres, Willet i Singer (1991) realitzen també una important revisió sistemàtica de les diferents metodologies utilitzades en l'àmbit nordamericà i europeu. Entre molts d'altres citen els treballs de Spady (1970), Tinto (1975) i Terenzini i Pascarella (1980).

Els primers treballs (Rubio, 1968, Benedito et al., 1970) es centraven exclusivament en el càlcul de taxes d'abandonament per a diferents cohorts d'estudiants i constitueixen el bloc que Willet i Singer (1991) anomenen d'estudis retrospectius. Els principal inconvenient de la metodologia emprada és que ignoren el moment del temps en el qual l'estudiant abandona. Aquesta informació és prou important ja que permetria identificar aquells cursos acadèmics de l'estada de l'estudiant a la universitat en els quals la política universitària podria actuar amb l'objectiu de disminuir l'abandonament.

Amb posterioritat, s'utilitzen mètodes estadístics que tracten de determinar els factors que expliquen l'èxit o el fracàs en els estudis. En uns casos la variable a explicar és únicament l'abandonament dels estudis (Spady, 1970, Tinto, 1975, García Aretio, 1987) i, en altres, l'objectiu va més enllà i recull en general la situació acadèmica de l'alumne després de transcorregut un període suficientment llarg com per a que sigui possible trobar estudiants que hagin obtingut el títol, que hagin abandonat, que hagin repetit cursos, etc. (González Tirados, 1986, Latiesa, 1986, Salvador i García-Valcárcel (1989).

Els mètodes estadístics utilitzats són fonamentalment contrastos estadístics d'associació entre variables, com ara el contrast χ^2 (García Aretio, 1987, Salvador i García-Valcárcel, 1989) i l'anàlisi discriminant (García Llamas, 1986, Jiménez Fernández, 1987, Ott, 1988 i Stage, 1988, 1989). Totes aquestes aproximacions tenen en comú la recerca de les característiques individuals que determinen l'èxit o fracàs en els estudis. Uns autors posen l'èmfasi en els atributs dels estudiants com són, a més dels estrictament personals (sexe, edat, estat civil, etc.) altres de caire psico-pedagògic (estils d'aprenentatge, actituds, personalitat, hàbits de treball i estudi, etc.). Alguns autors que destaquen en aquest aspecte són García Llamas (1986), González Tirados (1986), Bryk i Thum (1989), Rumberger, Ghatak, Poulos, Ritter i Dornbusch, (1990). Altres autors com Jiménez Fernández (1987) i Murdock (1987) inclouen també variables relatives a les qualificacions obtingudes prèviament a l'accés a la universitat i al llarg dels anys d'estada en aquesta, a més de mesures de la integració en el món acadèmic i de la recepció de beques o ajudes financeres a l'estudi. Aquests darrers treballs que combinen dades acadèmiques amb d'altres de caire estadístic personal inclouen també factors de control que relacionen el rendiment acadèmic amb l'estatus socioeconòmic, l'estructura familiar i altres factors externs.

Els treballs referenciats fins ara són del tipus anomenat estàtics ja que aborden el problema des de la perspectiva de respondre únicament a la pregunta *abandona l'alumne els estudis?* o, alternativament, *obté el títol oficial?* però no inclouen en l'anàlisi el factor temporal de *quan?* es produeix l'abandonament o es finalitza amb èxit. Alguns autors, però, han reconegut la dimensió dinàmica del problema establint a priori diversos moments del temps o cursos acadèmics i comparant posteriorment els percentatges d'abandonament que es produïen en aquests instants diferents (Schlechty i Vance, 1981) o analitzant els factors que

expliquen l'abandonament mitjançant una anàlisi discriminant (Jiménez Fernández, 1987) o una regressió logística (Alemany et al, 1990) per a successius cursos acadèmics. En aquests treballs sempre es fa referència al fet que la probabilitat d'abandonament no és constant al llarg de la vida acadèmica. Aquesta idea apareix relacionada amb conceptes semblants en altres àmbits que tenen a veure amb l'anàlisi del temps esperat de vida d'un individu. El caràcter no constant de l'esperança de vida és un fet totalment reconegut.

Totes aquestes modelitzacions estàtiques, fins i tot aquelles que tracten d'incorporar una certa dinamicitat en l'anàlisi, reben moltes crítiques. En primer lloc cal que l'objectiu de l'anàlisi vagi més enllà de la predicció de l'abandonament o èxit en els estudis i determinació de les seves causes, entenen aquells únicament de forma estàtica. La variable d'interès ha de ser per força la permanència o durada dels estudis ja que aquesta inclou, a més de la informació de si l'estudiant abandona, continua o s'ha llicenciat, el moment en el qual s'ha produït aquest esdeveniment.

El problema principal que s'ha de tractar en analitzar la variable permanència o durada dels estudis a la Universitat té a veure amb el tipus de dades disponibles. Si l'estudi està limitat a un període temporal no excessivament extens és impossible conèixer si aquells individus que encara continuen els seus estudis en el moment de finalitzar la recollida de dades, els abandonaran posteriorment o bé finalitzaran la Llicenciatura amb èxit. El fet d'ignorar aquesta característica de censurament de les dades, donat que es desconixerà el comportament futur dels individus, ha estat el que ha caracteritzat els principals treballs de recerca pioners en aquest àmbit i que conformen el grup dels mètodes estàtics.

La manca de tractament del problema de la censura pot donar lloc a l'obtenció de conclusions en certa forma contradictòries. Per exemple, s'obtidran resultats diferents depenent de la definició dels grups a discriminar. Algunes variables poden aparèixer com significatives si es considera com variable d'interès l'abandonament en acabar el primer curs i aquestes variables poden no ser rellevants si el moment d'interès és la finalització del tercer any.

El més habitual en estudis de permanència o durada es disposar d'informació censurada. Davant d'aquesta circumstància existeixen bàsicament dues possibilitats de tractament estadístic. La primera i més radical, consistiria en l'eliminació d'aquells individus que encara no han finalitzat els estudis. La implementació d'aquest procediment porta a l'obtenció d'una mostra explícitament esbiaixada, en la qual els individus amb un rendiment acadèmic baix en termes d'un nombre reduït d'assignatures superades no quedarien representats en la mostra final. Només restaria la informació relativa als estudiants que abandonen en els primers anys de la Llicenciatura i aquells que la finalitzen en el temps previst. Per tant si es decidís optar per aquesta solució caldria emprar mètodes estadístics que aconseguissin corregir el biaix per selecció (Shaw, 1988).

La segona alternativa consisteix en reciclar la informació dels individus que apareixen censurats a la mostra. El més directe és realitzar imputacions sobre les condicions i durada dels seus estudis, però la forma de realitzar aquestes imputacions és sempre discutible, ja que no existeix un únic criteri. En un context similar aquesta és la solució que proposen Frank i Keith (1984).

Els mètodes estadístics basats en la discriminació entre dos grups (Jiménez Fernández, 1987, Ott, 1988, Stage, 1988, 1989, Alemany et al., 1990), aquells que

compleixen una determinada condició en un moment del temps i la resta i que tracten d'incorporar la dinamicitat, analitzen anàlogament els estudiants que han abandonat abans, per exemple, del tercer any, distingint-los d'aquells que continuen després del tercer any, però sense considerar la seva condició final. Així s'intenta eliminar l'inconvenient del censurament. Naturalment, com a conseqüència del mètode és impossible captar les diferències existents entre els riscos d'abandonament en altres instants del temps.

Essencialment aquesta és la crítica que reben els tractaments anomenats estàtics en front dels dinàmics. Aquests darrers utilitzen l'anàlisi de supervivència per estudiar els riscos d'abandonament i la seva evolució al llarg del temps. De la mateixa forma es realitza l'anàlisi per a la probabilitat d'obtenir la llicenciatura i els seus canvis al llarg de la vida acadèmica de l'estudiant. Si no es té en compte l'evolució de les probabilitats de permanència al llarg del temps la influència de la durada queda emmascarada i això pot donar lloc a prediccions incorrectes sobre el comportament dels individus. Aquest problema s'ha presentat a la literatura com especialment important en el cas que el període de vida acadèmica en l'anàlisi del comportament dels individus sigui diferent per cadascun d'ells. Si el grup d'interès és una cohort i, per tant, el moment de censurament és comú a tota la mostra les conseqüències de la modelització són menys greus.

Finalment, un altre argument en contra de models basats en l'anàlisi discriminant, en models de regressió o en models Logit, tots ells estàtics, és la poca freqüència en que s'utilitzen variables explicatives que varien al llarg del temps. Naturalment, la utilització d'indicadors que varien al llarg del temps produeix un problema de colinealitat i autocorrelació dels regressors, i per tant es fa necessària la correcció d'aquestes problemes.

Per tant el mètode més adequat per poder realitzar una anàlisi de la permanència o durada dels estudis ha de tenir tres característiques fonamentals:

1. Ha de tenir en compte que la durada dels estudis, és a dir, la temporalitat i la variació de les circumstàncies al llarg del temps hauran de ser captades pel model.
2. Ha de poder tenir mecanismes que facilitin el tractament de la censuració.
3. Ha de permetre la inclusió de variables explicatives per tal de poder contrastar la seva significació en la influència de la variable durada i, posteriorment, servir per configurar una segmentació del col·lectiu d'estudis.
4. Ha de permetre la inclusió de variables explicatives tant constants com canviants en el temps.

La metodologia que s'adapta especialment a aquestes exigències és la de l'anàlisi de dades de supervivència. Tot seguit, en els següents capítols d'aquesta primera part metodològica es realitza una descripció d'aquests mètodes i de les modelitzacions alternatives possibles.

3. Dades de Durada o Supervivència: Introducció

Capítol 3

Dades de Durada o Supervivència: Introducció

L'anàlisi estadística de la supervivència compta amb una llarga història, fonamentada en aplicacions biomèdiques i industrials. En els darrers anys, però, s'han extés les aplicacions al camp de l'economia per l'estudi de la durada de l'atur, de les vagues, etc., i s'han iniciat en l'estudi dels processos educatius.

En l'anàlisi de dades de durada o de supervivència, l'interès es centra sobre un grup o grups d'individus per a cadascun dels quals es defineix un succés puntual (mort, falla, abandonament, etc.) el qual s'esdevé després d'un període de temps, anomenat durada o temps de supervivència.

El primer objectiu de l'anàlisi rau en l'estimació de les funcions de supervivència, inicialment per a poblacions homogènies, tenint en compte el problema de la censuració, és a dir, que per a alguns individus no es disposi de la informació exacta de la durada o temps de supervivència. Per tal d'assolir aquest primer objectiu s'han desenvolupat mètodes paramètrics i no paramètrics.

Els mètodes no paramètrics es fonamenten en la construcció de taules de vida, nom que reben en general les estimacions no paramètriques de la funció de supervivència, i en l'estimador de Producte Límit de Kaplan-Meier. Les referències més importants són Berkson i Gage (1952), Cutler i Ederer (1958), Kaplan i Meier (1958), Chiang (1960), Efron (1967), Gehan (1969), Breslow i Crowley (1974), i Meier (1977).

Pel que fa als mètodes paramètrics, consisteixen en l'especificació de la funció de supervivència atenent a alguna funció de probabilitat coneguda per tal d'estimar els seus paràmetres tenint en compte la possible censuració; així, podem treballar amb models exponencial, Weibull, Gamma, Lognormal, etc. Com a referències bàsiques d'aquests mètodes paramètrics podem esmentar Cox (1972), Mann et al. (1974), i Gross i Clark (1975).

El segon objectiu fonamental de l'anàlisi de supervivència va molt més enllà que el primer. Donat que sovint no treballarem amb poblacions homogènies, calen models que quantifiquin la relació entre la durada o temps de supervivència i un conjunt de variables explicatives que caracteritzen l'heterogeneïtat. Per a la modelització en aquestes circumstàncies existeixen dues famílies de models: els Models de Durada Accelerada (Accelerated Failure Time Models, AFT) i els Models d'Atzar Proporcional (Proportional Hazards Models, PH).

En els de Durada Accelerada (AFT) l'efecte de les variables explicatives és directe sobre el temps de durada o supervivència, mentre que en els d'Atzar Proporcional (PH) aquest efecte es produeix sobre la funció d'atzar, és a dir sobre la probabilitat d'abandonar (morir) en un moment determinat del temps condicionada a la supervivència fins a aquell moment. Les referències bàsiques

d'aquests models són Cox (1972), Kalbfleisch i Prentice (1973), i Prentice (1978).

Els models d'atzar proporcional (PH) tenen una característica que els fa més atractius en algunes ocasions com és que s'han desenvolupat mètodes d'estimació eficients sense necessitat d'especificar la forma de la funció de supervivència (exponencial, Weibull, etc). És per aquesta raó que són anomenats semiparamètrics, donat que només es realitza inferència sobre els paràmetres que acompanyen a les variables explicatives però no sobre els paràmetres de la funció de supervivència.

De la mateixa forma, encara que amb procediments numèrics més complexes que els anteriors, per als models de durada accelerada (AFT) també s'han proposat mètodes d'inferència semiparamètrics basats en contrastos de rangs que obvien l'especificació de la forma de la funció de supervivència.

Finalment, cal assenyalar que aquests models de durada poden ser ampliat amb la incorporació d'algunes qüestions com: introducció de variables explicatives els valors de les quals va canviant en el temps ("time-dependent covariates"), possibilitat de més d'un succés alternatiu que posa punt final a la durada ("multiple risks"), validació dels models (residus, bondat de l'ajust, correcta especificació), etc.

A continuació desenvoluparem els conceptes bàsics en què es fonamenta l'anàlisi estadística de la supervivència i més endavant entrarem en detall en l'especificació i l'estimació dels models enumerats anteriorment.

3.1. Conceptes teòrics bàsics

Inicialment considerarem un conjunt d'individus que formen part d'una població homogènia a on cadascun dels quals té un temps de durada o de supervivència. Així, disposem d'una variable aleatòria no negativa T que representa la durada dels individus.

Cal que aquesta durada estigui definida de forma precisa, la qual cosa es produirà si es compleixen els següents requisits. En primer lloc, s'ha de definir de forma concreta i no ambigua un origen temporal. Això no vol dir que calgui que l'origen temporal hagi de ser el mateix temps de calendari per a cadascun dels individus sinó que la situació inicial, l'estat inicial, sigui homogeni per a tots els individus. En segon lloc, cal disposar d'una escala de mesura del temps, la qual generalment serà el temps real, malgrat que pot ser utilitzada qualsevol altra amb una única limitació: que sigui no negativa. Finalment, el significat del succés puntual que determina la durada ha de ser totalment clar, és a dir, no poden existir ambigüitats que puguin fer dubtar sobre la longitud de la durada.

La distribució de probabilitat de T pot ser especificada de diverses formes, però n'existeixen tres que són especialment útils en aquest context: la funció d'atzar ($\lambda(t)$), la funció de supervivència ($S(t)$) i la funció de densitat ($f(t)$).

La Funció d'Atzar és una caracterització més especialitzada de la distribució d'una variable aleatòria particularment útil en la modelització de les dades de supervivència o durada perquè expressa probabilitats condicionades, és a dir, probabilitats d'abandonar (morir) en un moment determinat del temps condicionades al fet d'haver sobreviscut fins aleshores. No és el mateix la

consideració de la probabilitat de, per exemple, viure 10 anys, que la probabilitat de morir als 10 anys condicionada al fet d'haver viscut fins els 9 anys. Aquesta última idea reflecteix el risc immediat de que es produeixi l'abandonament (mort) per a un individu del qual se sap que ha sobreviscut fins a un moment determinat.

A més, en molts camps d'aplicació (medicina, biologia, economia, etc) es disposa d'informació sobre com canvia la taxa d'abandonament (mortalitat) a mesura que avança el temps. Aquesta informació pot ser utilitzada per a la modelització de la funció d'atzar. Per exemple, en la modelització de l'edat de mort de les persones és evident que $\lambda(t)$ per valors petits de t és elevada degut a la mortalitat infantil

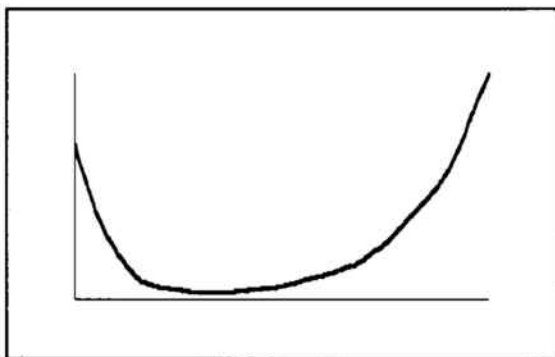


Figura 1

i a les malalties de la infantesa. Més endavant es passa per un període d'una relativa mortalitat inferior, després de la qual les taxes de mortalitat augmenten molt ràpidament (Figura 1). Aquesta informació qualitativa sobre la funció d'atzar sovint és útil en la selecció d'un model de probabilitat per a T .

La *Funció de Supervivència* es defineix, tant en el cas discret com en el continu, com la probabilitat de que T sigui més gran o igual que un valor t , és a dir:

$$S(t) = P(T \geq t), \quad 0 < t < \infty$$

Per a l'especificació de les altres funcions resulta més adient la separació entre el cas continu i el cas discret.

3.1.1 Variable Temps Contínua

La *Funció de Densitat* de T és:

$$f(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = \frac{-dS(t)}{dt}$$

Alternativament tenim que:

$$S(t) = \int_t^{\infty} f(s) ds \quad i \quad f(t) \geq 0 \quad \text{amb} \quad \int_0^{\infty} f(t) dt = 1.$$

El recorregut de T és $[0, \infty)$ i hauria de ser entès com a domini de definició de les funcions de t .

La *Funció d'Atzar* especifica la taxa instantània d'abandonar (morir, fallar) en $T = t$ condicionada a la supervivència fins a t . Es defineix com:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}.$$

A partir d'aquesta es poden deduir les anteriors, la qual cosa suposa que disposant de la *Funció d'Atzar* tenim caracteritzada la distribució de probabilitat de T .

Així, de l'anterior:

$$\lambda(t) = \frac{-d \log S(t)}{dt}$$

de forma que integrant i fent servir el resultat $S(0) = 1$, tenim:

$$S(t) = \exp \left(\int_0^t \lambda(u) d(u) \right).$$

De la mateixa forma, la *Funció de Densitat* quedaria:

$$f(t) = \lambda(t) \exp \left(\int_0^t \lambda(u) d(u) \right).$$

La Funció d'Atzar possibilita una convenient definició de la dependència de la durada. Tindrem *dependència de la durada positiva* en el punt t^* si $d\lambda(t)/dt > 0$ en $t=t^*$ la qual cosa significa, en el nostre cas, que la probabilitat de que un estudiant abandoni és més gran a mesura que les durades són més llargues o la seva supervivència superior. Tindrem *dependència de la durada negativa* en el punt t^* si $d\lambda(t)/dt < 0$ en $t=t^*$ la qual cosa significa, que la probabilitat de què un estudiant abandoni és més petita a mesura que les durades són més llargues.

Existeix una altra funció deduïda de l'anterior que també resulta útil en el treball aplicat. Així es defineix com *Atzar Integrat* a la funció:

$$\Lambda(t) = \int_0^t \lambda(u) du$$

la qual, malgrat no tenir una interpretació directa, és una peça clau en alguns contrastos d'especificació i la seva representació gràfica és un instrument molt útil en la determinació de la distribució concreta de la supervivència.

3.1.2 Variable Temps Discreta

Si T és una variable aleatòria discreta que pren els valors $0 \leq x_0 < x_1 < x_2 < x_3 < \dots$ amb funció de probabilitat associada:

$$f(x_i) = P(T=x_i) \quad i=0,1,2,\dots$$

aleshores la *Funció de Supervivència* és:

$$S(t) = \sum_{j|x_j \geq t} f(x_j)$$

i expressa la probabilitat de què $T \geq t$.

La Funció d'Atzar en x_j és:

$$\lambda_j = P(T=x_j | T \geq x_j) = \frac{f(x_j)}{S(x_j)} \quad j=0,1,2,\dots$$

i expressa la probabilitat de que l'individu abandoni en x_j condicionada al fet d'haver sobreviscut fins a x_j .

Ara podem escriure:

$$\begin{aligned} 1 - \lambda_j &= 1 - \frac{f(x_j)}{S(x_j)} = \frac{S(x_j) - f(x_j)}{S(x_j)} = \\ &= \frac{f(x_j) + f(x_{j+1}) + \dots - f(x_j)}{S(x_j)} = \\ &= \frac{f(x_{j+1}) + f(x_{j+2}) + \dots}{S(x_j)} = \frac{S(x_{j+1})}{S(x_j)} \end{aligned}$$

de manera que:

$$S(x_{j+1}) = (1 - \lambda_j) S(x_j)$$

Recursivament i sabent que $S(x_0)=1$:

$$S(x_1) = (1 - \lambda_0) S(x_0) = (1 - \lambda_0)$$

$$S(x_2) = (1 - \lambda_1) S(x_1) = (1 - \lambda_1)(1 - \lambda_0)$$

$$S(x_3) = (1 - \lambda_2) S(x_2) = (1 - \lambda_2)(1 - \lambda_1)(1 - \lambda_0)$$

del qual es dedueix que la Funció de Supervivència és igual a:

$$S(x_j) = \prod_{i=0}^{j-1} (1 - \lambda_i)$$

la qual cosa expressa que, per tal que la durada sigui $T \geq t$, és necessari i suficient amb sobreviure en tots els moments del temps anteriors a t .

Les expressions de la *Funció de Densitat* i de l'*Atzar Integrat* són:

$$f(x_j) = \lambda_j \prod_{i=1}^{j-1} (1 - \lambda_i)$$

$$\Lambda(x_j) = \sum_{i=0}^{j-1} \lambda_i$$

3.2 Censuració

Una qüestió diferencial en l'anàlisi de supervivència és la possibilitat que alguna de les durades observades es trobi censurada, és a dir, que per a un conjunt d'individus no podem observar el temps de durada total o complet. Un individu, el qual és observat sense abandonar (morir) al llarg de 10 dies després dels quals no es disposa de cap informació sobre el mateix, tindrà una durada que excedirà els 10 dies. Aquesta observació incompleta de la durada s'anomena censurada. Cal notar però, que la censuració, de la mateixa forma que l'abandonament (mort), és un succés puntual i que el període d'observació pels individus censurats cal que sigui enregistrat.

Realitzem les següents hipòtesis inicials. En primer lloc, suposem que en absència de censuració, l' i -èsim individu en una mostra de n té una durada de T_i , una variable aleatòria. En segon lloc, que establim que les n variables aleatòries són independents, idènticament distribuïdes (iid) i cadascuna amb funció de distribució F . En darrer lloc, que existeix un període d'observació c_i tal que el seguiment de l'individu finalitza a c_i si l'abandonament (mort) no s'ha esdevingut fins aleshores. Aleshores el que observem és $Y_i = \min(T_i, c_i)$, a la vegada que l'indicador $\delta_i = 1$ si $T_i \leq c_i$ (observació no censurada), $\delta_i = 0$ si $T_i > c_i$ (observació censurada).

Considerarem diferents tipus de censuració:

- Censuració Aleatòria:

Siguin c_i , $i=1, \dots, n$ variables aleatòries, iid, cadascuna amb funció de distribució G . c_i és el temps de censuració associat a T_i . Així doncs, només observem $(Y_1, \delta_1), \dots, (Y_n, \delta_n)$, a on Y_i i δ_i estan definides com abans. En el cas de la censuració aleatòria hem de fer la següent i decisiva hipòtesi: T_i i c_i són independents. Sense aquesta hipòtesi pocs resultats són accessibles. El seu compliment sembla justificat amb entrades aleatòries en l'estudi i pèrdues en el seguiment també aleatòries. Però si la raó de l'eliminació de l'individu de l'experiment està relacionada amb el curs del propi experiment, podria existir dependència entre T_i i c_i .

- Censuració Tipus I:

Sigui c qualsevol número fix predeterminat el qual anomenarem temps de censuració fix. Aleshores totes les c_i són iguals a $c_i = c$, el temps de censuració fix que està controlat per l'investigador. Si aturem l'experiment o recollida d'informació un cop transcorreguts c períodes de temps sense que l'individu hagi abandonat (mort), tenim Censuració Tipus I. Aquesta Censuració Tipus I és un cas particular de la censuració aleatòria.

- Censuració Tipus II:

Sigui $d < n$ fix. L'observació finalitza després d'un predeterminat número d d'abandonaments (morts). Així, si aturem l'experiment quan una fracció predeterminada d/n d'individus hagi abandonat (mort), tenim Censuració Tipus II. En aquest cas, malgrat pugui existir dependència entre T_i i c_i , aquesta serà sempre respecte la història passada del procés però no en el futur del mateix. Aquesta situació també permet l'anàlisi acurada de les dades de durada.

- Quan els individus entren en l'estudi aleatòriament al llarg del temps i la finalització de l'experiment es produeix en una data predeterminada, aleshores és possible que, en el moment de la finalització de l'enregistrament de durades, alguns individus no hagin abandonat (mort) de forma que l'observació estarà censurada. Aquest també és un cas particular de censuració aleatòria.

En general, la hipòtesi d'independència es mantindrà en el cas que el procediment de censuració depengui arbitràriament de les durades prèvies, dels temps de censuració previs o dels valors de les variables independents incloses en el model. En cas contrari, si per exemple l'individu passa a ser censurat quan es troba davant d'un imminent risc d'abandonar, no es verifica la hipòtesi d'independència.

La censuració aleatòria es produeix en moltes aplicacions. Els individus entren en l'estudi en moments diferents del temps, s'observa el temps de supervivència, però la censuració s'esdevè en alguna de les següents circumstàncies:

- 1.- Pèrdua del seguiment.
- 2.- Eliminació de l'individu.
- 3.- Finalització de l'estudi.

A la Figura 2, l'individu 1 entra en l'experiment posteriorment al moment $t=0$ i li esdevé la mort (abandona) en el moment T_1 , donant lloc a una observació no censurada. L'individu 2 entra en l'experiment i en el moment de la finalització d'aquest encara és viu donant lloc a una observació censurada T_2^+ . L'individu 3 entra en l'experiment i es perd el seguiment abans de la finalització del mateix donant lloc a una altra observació censurada T_3^+ .

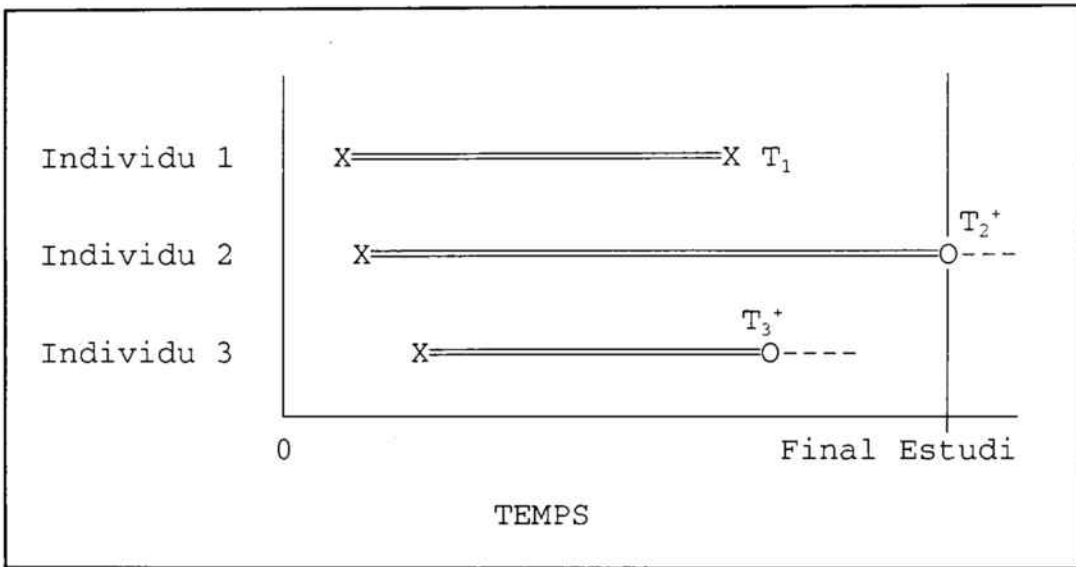


Figura 2

- Altres tipus de Censuració:

Existeixen altres tipus de censuració que apareixen en la literatura. Els anteriors cauen en el grup de la Censuració per la Dreta (Right Censoring): la referència a una durada excessiva no la podem observar, tan sols podem donar una acotació de la mateixa.

Existeix, però, també la Censuració per l'Esquerra. Així, en la censuració aleatòria per l'esquerra només observem Y_i i ϵ_i , a on $Y_i = \max(T_i, c_i)$ mentre que $\epsilon_i = I(c_i \leq T_i)$. Aquesta situació es produeix quan en el moment d'iniciar l'estudi, l'enregistrament de les durades, existeixen individus la durada dels quals fa alguns períodes que ja ha començat però és impossible saber quan.

Quan es produeix tant la censuració per l'esquerra com per la dreta tenim Censuració Intervalar, coneguda en la literatura com "Double Censoring", la qual és d'aplicació en diferents àmbits (Gómez i Lagakos, 1993).

4. Modelització de Poblacions Homogènies

Capítol 4

Modelització de Poblacions Homogènies

4.1 Mètodes No Paramètrics en Poblacions Homogènies

Els mètodes no paramètrics possibiliten alternatives flexibles a les tècniques paramètriques d'estimació de les funcions de supervivència, les quals requereixen l'especificació de la forma funcional de la distribució del temps de durada en absència de censuració, com veurem més endavant. A més, són especialment útils com a instruments gràfics per a la comprovació de la bondat de l'ajust en models més complexes. També han estat desenvolupats contrastos no paramètrics per tal de comparar funcions de supervivència de dues o més mostres.

4.1.1 L'estimador del Producte Límit de Kaplan-Meier

En l'anàlisi de supervivència és especialment útil la construcció de la funció de supervivència mostral per tal d'estudiar la forma de la distribució, la seva dispersió i altres característiques de la mateixa, així com per contrastar l'ajust de les dades a la llei de probabilitat que pot haver generat la mostra de la que disposem. Aquesta estimació de la funció de supervivència i els contrastos no

paramètrics requereixen uns mètodes específics degut a la presència de la censuració.

Suposem que la distribució de la variable aleatòria T_i és discreta, la qual pren valors $x_1 < x_2 < x_3 < \dots$ i que la seva funció de probabilitat associada és:

$$f(x_i) = P(T=x_i) \quad i=1,2,\dots$$

Com hem vist en l'apartat 3.1.2., la funció de supervivència $S(t)$ pot expressar-se en termes de la funció d'atzar discreta de la forma:

$$S(t) = \prod_{j|x_j < t} (1 - \lambda_j)$$

a la vegada que:

$$f(x_j) = \lambda_j \prod_{i=1}^{j-1} (1 - \lambda_i)$$

Una estimació no paramètrica de la funció de supervivència serà doncs:

$$\hat{S}(t) = \prod_{j|x_j < t} (1 - \hat{\lambda}_j)$$

on $\hat{\lambda}_j$ és l'estimador màxim versemblant de λ_j .

Per tal de deduir la versemblança d'una mostra de n individus, en primer lloc seleccionem tots els termes corresponents al moment x_j . Suposem que existeixen d_j abandonaments (morts) i m_j observacions censurades entre els $n_j = n(x_j)$ individus presents en el moment x_j , de tal forma que $n_j = d_j + m_j + n_{j+1}$. La contribució a la versemblança és, pels individus que han abandonat (mort), $f(x_j)$, mentre que pels censurats suposarem que és possible que una observació enregistrada com a censurada en l'instant t hagi abandonat en aquest mateix moment i , per tant, definirem en aquest cas $S(x_j)$ com la probabilitat de sobreviure fins a $T > x_j$. Com a conseqüència la funció de supervivència serà:

$$S(x_j) = \prod_{i=1}^j (1 - \lambda_i)$$

En aquesta situació la funció de versemblança, és:

$$\begin{aligned} L &= \prod_{j=1}^k \left[f(x_j)^{d_j} S(x_j)^{m_j} \right] = \\ &= \prod_{j=1}^k \left[\lambda_j \prod_{i=1}^{j-1} (1-\lambda_i) \right]^{d_j} \left[\prod_{i=1}^j (1-\lambda_i) \right]^{m_j} \end{aligned}$$

de manera que:

$$L = \prod_{j=1}^k \left[\lambda_j^{d_j} \prod_{i=1}^{j-1} (1-\lambda_i)^{d_j} \prod_{i=1}^j (1-\lambda_i)^{m_j} \right]$$

Descomposant els termes pels períodes $j = 1, 2, 3$ tenim que:

$$\begin{aligned} j = 1 & \quad \lambda_1^{d_1} (1-\lambda_1)^{m_1} \\ j = 2 & \quad \lambda_2^{d_2} (1-\lambda_1)^{d_2} (1-\lambda_2)^{m_2} \\ j = 3 & \quad \lambda_3^{d_3} (1-\lambda_1)^{d_3} (1-\lambda_2)^{d_3} (1-\lambda_1)^{m_3} (1-\lambda_2)^{m_3} (1-\lambda_3)^{m_3} \end{aligned}$$

i generalitzant:

$$L = \lambda_1^{d_1} \lambda_2^{d_2} \dots \lambda_k^{d_k} (1-\lambda_1)^{m_1+n_2} (1-\lambda_2)^{m_2+n_3} \dots (1-\lambda_k)^{m_k+n_{k+1}}$$

amb la qual cosa i sabent que $m_j + n_{j+1} = n_j - d_j$, podem reescriure la funció i el logaritme de la funció com:

$$L = \prod_{j=1}^k \left[\lambda_j^{d_j} (1-\lambda_j)^{n_j-d_j} \right]$$

$$\log L = \sum_{j=1}^k \left[d_j \log \lambda_j + (n_j - d_j) \log(1-\lambda_j) \right]$$

Aquesta expressió de la funció de versemblança és una generalització del concepte utilitzat en els models paramètrics donat que aquí maximitzem respecte la funció d'atzar λ_j en comptes de fer-ho respecte un paràmetre.

L'obtenció de l'estimador màxim-versemblant $\hat{\lambda}_j$ es realitza a partir de:

$$\frac{\partial \log L}{\partial \lambda_j} = \frac{d_j}{\lambda_j} - \frac{n_j - d_j}{1 - \lambda_j} = 0$$

i pren el valor:

$$\hat{\lambda}_j = \frac{d_j}{n_j}$$

Substituint aquest estimador màxim-versemblant de la funció d'atzar en l'expressió de l'estimador no paramètric de la funció de supervivència, obtenim l'*estimador de producte límit* o *estimador de Kaplan-Meier* (Kaplan i Meier, 1958, Johansen, 1978):

$$\hat{S}(t) = \prod_{j|x_j < t} \left(1 - \frac{d_j}{n_j} \right)$$

Qualsevol terme en el producte pel qual $d_j=0$ pot ser omès sense afectar l'expressió de l'estimador. Aquest és formalment independent de la selecció de punts x_j pels quals el número observat d'abandonaments (morts) és zero. Per tant, l'estimador de Kaplan-Meier és únicament una funció de les dades.

Com en tot estimador, resulta necessària l'obtenció d'una mesura de la seva precisió, és a dir, una estimació de la variança del mateix, per tal de valorar la fiabilitat dels resultats. Així, en el cas que les possibles durades temporals

x_1, x_2, x_3, \dots siguin fixes i que el mecanisme de censuració permeti que el número d'abandonaments (morts) d_j en cada moment del temps x_j vagi creixent al mateix ritme que ho fa tot la mida de la mostra n , aleshores es poden aplicar les propietats asimptòtiques pels estimadors màxim versemblants (Breslow i Crowley, 1974, Meier, 1975). En algunes aplicacions, però, la mida de la mostra n vindrà determinat pel nombre d'abandonaments d_j de manera que no es pot suposar que ambdós valors creixen en la mateixa taxa i, en conseqüència, no es podrà emprar la teoria asimptòtica sobre l'estimador. Tanmateix, en les condicions habituals pels mecanismes de censuració es pot fer inferència sobre λ_j i sobre les funcions del mateix com $S(t)$.

Assimptòticament, $\sqrt{n}(\hat{\lambda}_j - \lambda_j)$ tindrà una distribució normal multivariant amb mitjana zero i matriu de variàncies i covariàncies igual a la inversa de la matriu d'informació observada. En aquest cas la matriu d'informació és:

$$\left[\frac{\partial^2 \log L}{\partial \lambda_j \partial \lambda_k} \right]_{\hat{\lambda}} = \begin{cases} \frac{-n_j}{\hat{\lambda}_j(1-\hat{\lambda}_j)} & (j=k) \\ 0 & (j \neq k) \end{cases}$$

de manera que:

$$\text{Var}(\hat{\lambda}_j) = \frac{d_j(n_j - d_j)}{n_j^3}$$

Mitjançant la transformació:

$$\log \hat{S}(t) = \sum_{j|x_j < t} \log(1 - \hat{\lambda}_j)$$

obtenim l'expressió de la variança asimptòtica de la transformació¹, usant el resultat $\hat{\lambda}_j = d_j/n_j$:

$$\begin{aligned} \widehat{\text{Var}} [\log \hat{S}(t)] &\approx \sum_{j|x_j < t} (1 - \hat{\lambda}_j)^{-2} \widehat{\text{Var}}(1 - \hat{\lambda}_j) \\ &\approx \sum_{j|x_j < t} \frac{d_j}{n_j(n_j - d_j)} \end{aligned}$$

per a finalment obtenir la de l'estimador de Kaplan-Meier:

$$\widehat{\text{Var}} [\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j|x_j < t} \frac{d_j}{n_j(n_j - d_j)}$$

Aquesta expressió de la variança asimptòtica de l'estimador de Kaplan-Meier rep el nom de *formula de Greenwood* (Greenwood, 1926), la qual va ser deduïda en primer lloc per a l'estimador clàssic de taules de vida.

4.1.2 L'estimador de Taula de Vida

La taula de vida és una forma de sintetitzar les dades de supervivència agrupades en intervals adients. Si, per exemple, les dades venen agrupades en els intervals I_1, \dots, I_k , tals que $I_j = [b_0 + b_1 + \dots + b_{j-1}, b_0 + b_1 + \dots + b_j]$ és d'amplada b_j amb $b_0 = 0$ i $b_k = \infty$, la taula de vida presenta el número d'abandonaments (morts) i observacions censurades de cadascun dels intervals.

¹ L'aproximació s'obté a partir del resultat de la variança de la transformació lineal d'una variable aleatòria segons la qual $\text{Var}[f(u)] = (\partial f/\partial u) \text{Var}(u) (\partial f/\partial u)'$.

Suposem que es produeixen m_j observacions censurades i d_j abandonaments (morts) en l'interval I_j , i que $n_j = \sum_{t \geq j} (d_t + m_t)$ és el número d'individus en risc en l'inici de l'interval j -èsim. L'estimador de taula de vida de la probabilitat condicional d'abandonar (morir) en I_j és:

$$\hat{q}_j = \begin{cases} 0 & \text{si } n_j=0 \\ \frac{d_j}{n_j - \frac{1}{2}m_j} & \text{en altre cas} \end{cases}$$

El terme $m_j/2$ del denominador s'utilitza en un intent d'ajustar-se al fet de que no tots els individus n_j es troben en risc al llarg de tot l'interval I_j .

L'estimador de taula de vida per a la funció de supervivència al final de l'interval I_j és:

$$\hat{S}(b_1 + \dots + b_j) = \prod_{i=1}^j (1 - \hat{q}_i) = \prod_{i=1}^j \left(1 - \frac{d_i}{n_i - \frac{1}{2}m_i} \right)$$

La fórmula de Greenwood amb la modificació de n_j per $(n_j - m_j/2)$ ens proporciona un estimador de la variança de $\hat{S}(t)$:

$$\widehat{Var} [\hat{S}(t)] = [\hat{S}(t)]^2 \sum_{j: t_j \leq t} \frac{d_j}{n'_j (n'_j - d_j)}$$

$$i \quad n'_j = n_j - \frac{1}{2}m_j$$

El mètode de la taula de vida està principalment dissenyat per aquelles situacions en les quals es desconeixen els moments exactes dels abandonaments i de les censuracions, disposant només del número d'abandonaments (morts) i d'observacions censurades en cadascun dels intervals.

4.1.3 Comparació de Funcions de Supervivència

Sovint és interessant la determinació de si dues o més mostres han estat generades per una mateixa funció de supervivència. Una aproximació podria ser la utilització dels resultats asimptòtics per a $\hat{S}(t)$ obtinguts anteriorment per a la construcció d'un contrast d'igualtat de les funcions de supervivència en un moment determinat del temps. Aquesta metodologia, però, no seria del tot eficient donat que es centraria en un únic moment del temps. En aquest sentit han estat dissenyat contrastos d'igualtat de funcions de supervivència per a tot el període mostral de què es disposi. Aquests contrastos són generalitzacions per al cas de treballar amb dades censurades dels contrastos de rangs de Wilcoxon (Gilbert, 1962, Gehan, 1965, Miller, 1981) i de Mantel-Haenszel (o log-rank test) (Mantel i Haenszel, 1959, Mantel, 1966, Peto i Peto, 1972).

4.1.3.1 Contrast de *log-rank*

Aquest contrast és una generalització del contrast de Savage (1956) en presència de censura el qual va ser desenvolupat en una sèrie d'articles de Mantel i els seus col·laboradors (Mantel i Haenszel, 1959, Mantel, 1963 i Mantel 1966). Està particularment indicat quan el quocient de les funcions d'atzar en les poblacions a comparar és aproximadament constant. A més la seva interpretació és senzilla donat que es pot presentar com la diferència entre el número observat

d'abandonaments (morts) a cada grup i el que equivaldria al corresponent número esperat dels mateixos, sota la hipòtesi nul·la d'igualtat de poblacions o de procés de generació de les dades.

Suposem que volem contrastar la igualtat de les funcions de supervivència $S_1(t), \dots, S_r(t)$, a partir de mostres de cadascuna de les r poblacions. Siguin $t_1 < t_2 < \dots < t_k$ els temps d'abandonament (mort) per a la mostra formada pel conjunt dels individus de les r mostres, d_j són els abandonaments (morts) esdevinguts en t_j i n_j són els individus en risc instants abans de t_j ($j=1, \dots, k$). D'igual forma tenim d_{ij} i n_{ij} per a la corresponent mostra i ($i=1, \dots, r$). Les dades en el moment t_j es troben en la forma d'una taula de contingència $r \times 2$ amb d_{ij} abandonaments i $n_{ij} - d_{ij}$ supervivents a la i -èsima fila ($i=1, \dots, r$).

Amb la informació disponible fins el moment t_j , la distribució de d_{1j}, \dots, d_{rj} és simplement el producte de r distribucions binomials:

$$\prod_{i=1}^r \binom{n_{ij}}{d_{ij}} \lambda_j^{d_{ij}} (1-\lambda_j)^{n_{ij}-d_{ij}}$$

a on λ_j és la probabilitat condicional d'abandonar en t_j , la qual, sota la hipòtesi nul·la, és comuna per a cadascuna de les r mostres.

La distribució condicional per a d_{1j}, \dots, d_{rj} donat d_j és, aleshores, la distribució hipergeomètrica:

$$\frac{\prod_{i=1}^r \binom{n_{ij}}{d_{ij}}}{\binom{n_j}{d_j}}$$

Això és així donat que el fet de condicionar la suma de les $d_{ij} + \dots + d_{rj}$ al valor de d_j implica el no reemplaçament del mostratge i, per tant, la distribució hipergeomètrica.

La mitjana de d_{ij} i les covariances entre d_{ij} i d_{lj} , a partir de la funció de distribució són:

$$e_{ij} = \frac{n_{ij} d_j}{n_j}$$

$$(V_j)_{il} = \frac{(n_j n_{ij} \delta_{il} - n_{ij} n_{lj}) d_j (n_j - d_j)}{n_j^2 (n_j - 1)} \quad i \quad \delta_{il} = \begin{cases} 1 & \text{si } i=l \\ 0 & \text{si } i \neq l \end{cases}$$

Així, l'estadístic "log-rank" és:

$$\mathbf{v}' = (v_1, \dots, v_r) = \left(\sum_{j=1}^k (d_{1j} - e_{1j}), \dots, \sum_{j=1}^k (d_{rj} - e_{rj}) \right)$$

el qual tindrà mitjana zero i matriu de variàncies i covariances V . Aquesta serà igual a $V = V_1 + \dots + V_k$ si les k taules de contingència són independents. L'estadístic es pot interpretar com el vector del número d'abandonaments observat en cada població o estrat menys el corresponent vector del número d'abandonaments esperat.

El contrast de la igualtat de les r funcions de supervivència pot fonamentar-se en una distribució assintòtica txi-quadrat amb $r-1$ graus de llibertat:

$$\mathbf{v}' V^{-1} \mathbf{v} \sim \chi_{r-1}^2$$

Els graus de llibertat són $r-1$ donat que els r elements de \mathbf{v} sumen zero. En conseqüència, V^{-1} ha de ser la inversa generalitzada de V o, més fàcilment, l'estadístic χ^2_{r-1} pot construir-se utilitzant $r-1$ elements qualsevols de \mathbf{v} i la corresponent submatriu $(r-1) \times (r-1)$ de V .

4.1.3.2 Contrast de Wilcoxon Generalitzat

Aquest contrast, proposat en aquest context de les dades de supervivència per Gehan (1965) i generalitzat per Breslow (1970), és una variació del contrast de "log-rank" com varen mostrar Tarone i Ware (1977). Així podem definir l'estadístic com:

$$\mathbf{v}' = (v_1, \dots, v_r) = \left(\sum_{j=1}^k w_j (d_{1j} - e_{1j}), \dots, \sum_{j=1}^k w_j (d_{rj} - e_{rj}) \right)$$

el qual tindrà mitjana zero i matriu de variàncies i covariàncies:

$$V_{ii} = \sum_{j=1}^k w_j^2 \frac{(n_j n_{ij} \delta_{ii} - n_{ij} n_{ij}) d_j (n_j - d_j)}{n_j^2 (n_j - 1)} \quad i \quad \delta_{ii} = \begin{cases} 1 & \text{si } i=l \\ 0 & \text{si } i \neq l \end{cases}$$

Doncs bé, en el cas que la ponderació w_j sigui igual a 1 ens trobem davant de l'anterior contrast de "log-rank". Si, per contra, la ponderació w_j pren el valor n_j aleshores tenim el contrast de Wilcoxon. Aquest contrast també es distribueix asimptòticament segons una χ^2 amb un nombre de graus de llibertat iguals al rang de la matriu V , és a dir $r-1$.

4.2 Mètodes Paramètrics en Poblacions Homogènies

Malgrat que resulta més interessant la modelització de les relacions entre el temps de durada i algunes variables explicatives, és necessari l'estudi previ dels models paramètrics en poblacions homogènies. Estudiarem doncs algunes distribucions i les seves característiques pensant en l'esmentada modelització del temps de supervivència. La referència bàsica per a l'estudi de les distribucions de probabilitat que a continuació s'exposen és la de Johnson i Kotz (1970). Alternativament, algunes d'elles es troben estudiades en detall a Cox (1972), Mann et al. (1974) i Gross i Clark (1975).

Definirem $T > 0$ una variable aleatòria que representa el temps de durada i sigui t un punt genèric del seu rang de valors. Modelitzarem el seu comportament probabilístic a partir d'alguna de distribucions de probabilitat que es descriuen a continuació. Aquesta modelització serà especificada principalment mitjançant la funció d'atzar a més de considerar les altres funcions associades, supervivència i densitat. L'especificació d'un model de probabilitat pel temps de supervivència ens permetrà conèixer els paràmetres que caracteritzen la distribució i les seves propietats, així com la realització d'inferència sobre els mateixos.

Utilitzarem igualment, en ocasions, la notació $Y = \log T$ per a caracteritzar algunes de les distribucions, les quals seran en tots els casos contínues. Aquesta transformació logarítmica resulta convenient en el sentit que donarà lloc a models loglineals del temps de durada quan s'incorporin variables explicatives en la modelització.

4.2.1 La Distribució Exponencial

Si la variable aleatòria, T , temps de supervivència segueix una distribució exponencial, la qual depèn únicament d'un paràmetre, la *Funció d'Atzar* serà constant al llarg del temps:

$$\lambda(t) = \lambda > 0$$

el que vol dir que la probabilitat condicional d'abandonar (morir) és independent del moment del temps en el qual és mesurada i, per tant, independent del temps que l'individu porti estudiant, en l'aplicació que veurem posteriorment. Aquesta característica de la distribució exponencial tot i facilitar les especificacions de les funcions i els càlculs conseqüents, resulta molt restrictiva en molts àmbits d'aplicació puix que l'atzar constant implicaria, en el cas de modelitzar la vida d'un individu, que la probabilitat condicional de morir seria la mateixa per a un infant de 5 anys que per a un home de 90. Les altres funcions que caracteritzen la distribució són:

$$S(t) = e^{-\lambda t}, \quad f(t) = \lambda e^{-\lambda t}, \quad \Lambda(t) = \lambda t$$

Si tenim en compte ara $Y = \log T$, la seva funció de densitat serà:

$$f(y) = \exp(y - \alpha - e^{y-\alpha}) \quad -\infty < y < \infty$$

on $\alpha = -\log \lambda$.

Aleshores podem escriure el model $Y = \alpha + W$, on la funció de densitat de W és:

$$f(w) = \exp(w - e^w) \quad -\infty < w < \infty$$

la qual rep el nom de *Distribució de Valor Extrem Tipus I*. Així doncs, tenim un model loglineal per a la variable aleatòria T amb una distribució coneguda per al terme W .

4.2.2 La Distribució Weibull

Generalitzant la distribució exponencial per tal de permetre que la funció d'atzar variï en el temps, obtenim la distribució Weibull, la qual té dos paràmetres i la seva funció d'atzar és:

$$\lambda(t) = \lambda p (\lambda t)^{p-1} \quad \lambda, p > 0$$

Aquesta funció d'atzar és monòtona decreixent per a $p < 1$, monòtona creixent per a $p > 1$ i es redueix a la de la distribució exponencial si $p = 1$. Així, una possible via de discriminació entre la distribució exponencial i la Weibull serà la d'efectuar un contrast estadístic sobre el valor estimat del paràmetre p per tal de verificar la hipòtesi $p = 1$. Aquest instrument estarà molt més justificat que no pas els procediments gràfics i de tabulació proposats posteriorment.

Les altres funcions que la caracteritzen són:

$$S(t) = \exp[-(\lambda t)^p], \quad f(t) = \lambda p (\lambda t)^{p-1} \exp[-(\lambda t)^p], \quad \Lambda(t) = (\lambda t)^p$$

Si considerem $Y = \log T$, en aquest cas la seva funció de densitat és:

$$f(y) = \sigma^{-1} \exp\left(\frac{y - \alpha}{\sigma} - e^{(y - \alpha)/\sigma}\right) \quad -\infty < y < \infty$$

on $\sigma = p^{-1}$ i $\alpha = -\log \lambda$. És a dir, que podem escriure el model log-lineal $Y = \alpha + \sigma W$ on W presenta la mateixa distribució de valor extrem Tipus I que en l'apartat anterior. Per tant, la forma de la funció de densitat per a Y és igual tant en el cas que T tingui distribució exponencial com distribució Weibull. En aquest darrer cas, els paràmetres p i λ només determinen la situació i l'escala de la distribució de Y .

4.2.3 La Distribució Lognormal

Una altra forma de derivar possibles distribucions per a T és mitjançant l'especificació de qualsevol família de distribucions contínues conegudes per a la variable $\log T$. D'entre totes elles, la més habitual seria la distribució normal. Així, el model per a $Y = \log T$ és novament de la forma $Y = \alpha + \sigma W$, a on W es distribueix segons una normal estàndard de densitat:

$$f(w) = \phi(w) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right).$$

La funció de densitat per a T serà, a partir de l'anterior i realitzant un canvi de variable:

$$f(t) = \frac{1}{\sqrt{2\pi}} p t^{-1} \exp\left(\frac{-p^2 (\log \lambda t)^2}{2}\right)$$

on, com en el cas de la Weibull $\sigma = p^{-1}$ i $\alpha = -\log \lambda$.

Les altres funcions que caracteritzen la distribució de probabilitat no tenen una forma explícita i depenen de la funció de distribució d'una normal estàndard (Φ). La funció de supervivència és:

$$S(t) = 1 - \Phi(p \log \lambda t), \quad \Phi(w) = \int_{-\infty}^w \phi(u) du$$

i la funció d'atzar serà $f(t)/S(t)$. Aquesta última pren el valor 0 per a $t = 0$, augmenta fins a un màxim per després anar disminuint aproximant-se cap el zero a mesura que t va creixent.

La distribució exponencial no és un cas particular de la distribució lognormal (com succeïa en les anteriors distribucions) malgrat que és necessària una mostra d'una grandària considerable per tal de poder discriminar empíricament entre una exponencial i una lognormal amb $\sigma = 0.8$.

4.2.4 La Distribució Log-Logística

Una altra distribució coneguda aplicable a $\log T$ és la logística. La distribució log-normal anterior és recomanable quan no hi han observacions censurades però quan la censuració és present, els càlculs i complexitats de les estimacions la fan poc recomanable.

Com a alternativa tenim l'aproximació logística a la normal a on, pel model $Y = \log T = \alpha + \sigma W$, tenim que la distribució de W serà:

$$f(w) = \frac{e^{-w}}{(1 + e^{-w})^2}$$

I, efectuant un canvi de variable, la densitat per a T és:

$$f(t) = \frac{\lambda p (\lambda t)^{p-1}}{[1 + (\lambda t)^p]^2}$$

on novament $p = \sigma^{-1}$ i $\lambda = e^{-\alpha}$.

En aquest cas, les funcions de supervivència i d'atzar sí que tenen formes explícites i són:

$$S(t) = \frac{1}{1 + (\lambda t)^p} \quad \lambda(t) = \frac{\lambda p (\lambda t)^{p-1}}{1 + (\lambda t)^p} .$$

La funció d'atzar és idèntica a la de la distribució Weibull excepte pel que fa al denominador i la seva forma serà monòtona decreixent des de ∞ si $p < 1$ i monòtona decreixent des de λ si $p = 1$. Si $p > 1$, la semblança és amb la log-normal donat que va creixent des de zero fins a un màxim en el punt $t = (p-1)^{1/p} / \lambda$ i decreix a partir d'aleshores.

4.2.5 La Distribució Gamma

Una altra generalització de la distribució exponencial amb dos paràmetres és la distribució Gamma, la funció de densitat de la qual és:

$$f(t) = \frac{\lambda(\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)}$$

on $k, \lambda > 0$ i:

$$\Gamma(k) = \int_0^{\infty} x^{k-1} e^{-x} dx$$

és la *funció gamma*.

Les altres funcions que caracteritzen la distribució Gamma no tenen formes explícites, sinó que depenen de la funció gamma incompleta:

$$I_k(s) = \frac{\int_0^s x^{k-1} e^{-x} dx}{\Gamma(k)}$$

Així doncs, tenim que les funcions de Supervivència i d'Atzar són:

$$S(t) = 1 - I_k(\lambda t) \quad \lambda(t) = \frac{\lambda(\lambda t)^{k-1} \exp(-\lambda t) \Gamma(k)^{-1}}{1 - I_k(\lambda t)}$$

La funció d'Atzar és monòtona creixent des de 0 si $k > 1$ i monòtona decreixent si $k < 1$, en ambdós casos tendint cap el valor de λ a mesura que augmenta t .

En aquest cas, el model per a $Y = \log T$ es pot escriure $Y = \alpha + W$, a on la funció de densitat de W és:

$$f(w) = \frac{\exp(kw - e^w)}{\Gamma(k)} .$$

4.2.6 La Distribució Gamma generalitzada

Aquesta és la generalització de l'anterior amb la incorporació d'un paràmetre d'escala σ en el model per a $Y = \log T$ donant lloc a $Y = \alpha + \sigma W$, on W tindrà la mateixa funció de densitat que en el cas de la Gamma simple. Aquesta generalització amb tres paràmetres va ser introduïda per Stacy (1962) i resulta interessant vist que inclou com a casos particulars algunes de les distribucions comentades anteriorment. Així, si $\sigma = k = 1$ ens trobarem davant de l'exponencial, si $k = 1$ davant la Weibull, si $\sigma = 1$ davant la Gamma simple i en el cas que $k \rightarrow \infty$ la distribució tendirà a la lognormal.

La funció de densitat per a T serà en aquest cas:

$$f(t) = \frac{\lambda p (\lambda t)^{pk-1} \exp[-(\lambda t)^p]}{\Gamma(k)}$$

per a $t > 0$ i on $p = \sigma^{-1}$ i $\lambda = e^{-\alpha}$.

4.2.7 Criteris de selecció de les distribucions

Com assenyalen Cox i Oakes (1984), existeixen moltes formes de comparar famílies de distribucions de probabilitat ja sigui per a constatar les seves diferències o com a base per a una anàlisi empírica. En la pràctica, quan disposem d'una mostra d'individus pels quals s'han enregistrat uns temps de durada en algun procés, no coneixem a priori quina distribució teòrica ha generat aquelles dades. Així, per a la comparació entre distribucions tenim instruments gràfics i de tabulació. Les representacions gràfiques de les funcions d'atzar o d'atzar integrat respecte el temps, tant en valors originals com en logaritmes, ens mostren les diferències en la forma de les distribucions, mentre que els moments de segon i tercer ordre són informatius pel que fa a la dispersió (en comparació a la distribució exponencial) i a l'asimetria en les cues de la distribució.

A continuació resumim alguns resultats proposats per Cox i Oakes (1984) sobre les representacions gràfiques i tabulacions que ajudarien a identificar les distribucions anteriors:

- La distribució exponencial:

- 1.- Representació gràfica del logaritme de la funció de supervivència estimada, $\log \hat{S}(t)$, respecte el temps, t la qual presentarà una línia recta de pendent negativa.
- 2.- Alternativament, la representació gràfica de la funció d'atzar integrat estimada $\Lambda(t) = -\log S(t)$, en front del temps t , donarà una línia recta de pendent positiva que passa per l'origen de coordenades.
- 3.- El Coeficient de Variació ($\gamma = \mu/\sigma$) és igual a la unitat.
- 4.- El gràfic del moment de tercer ordre estandaritzat ($\gamma_3 = \mu_3/\sigma^3$) respecte el coeficient de variació correspon al punt de coordenades (1,2).²

- La distribució Weibull:

- 1.- Representació gràfica del logaritme de l'Atzar integrat (o $\log[-\log S(t)]$) respecte el logaritme del temps. Aquest gràfic presentarà una línia recta, la pendent de la qual ens donarà una primera aproximació al valor de p mentre que l'ordenada en l'origen de coordenades ens aproximarà al valor de λ .
- 2.- Si el coeficient de variació és $\gamma < 1$, aleshores pot ser una Weibull amb paràmetre $p > 1$. Per contra, si el coeficient de variació és $\gamma > 1$, aleshores pot ser una Weibull amb paràmetre $p < 1$.

² El moment central de tercer ordre per a la distribució exponencial és $\mu_3 = 2/\lambda^3$, mentre que $\sigma^3 = (1/\lambda)^3$, de manera que el quocient serà $\gamma_3 = \mu_3 / \sigma^3 = 2$.

- Altres distribucions:

1.- Si el coeficient de variació és $\gamma < 1$, aleshores pot ser una Gamma amb paràmetre $k > 1$, una Log-normal amb $\sigma < 0.83$, o una Log-logística amb $\sigma < 0.118$. Per contra, si el coeficient de variació és $\gamma > 1$, pot ser una Gamma amb paràmetre $k < 1$, una Log-normal amb $\sigma > 0.83$ o una Log-logística amb $\sigma > 0.118$.

2.- Per a totes les distribucions la comparació entre el coeficient d'asimetria i el coeficient de variació (mitjançant una gràfica, per exemple) ens permet diferenciar les distribucions, donat que per a una mateixa dispersió, un valor més elevat del coeficient d'asimetria indicarà unes cues de la distribució més llargues.

4.2.8 Inferència en els models paramètrics per a poblacions homogènies.

En els apartats anteriors s'han descrit alguns models paramètrics per tal de representar les durades o supervivències, així com alguns criteris per a la selecció entre ells. Un cop seleccionada una distribució de probabilitat per a les durades caldrà disposar de mètodes d'estimació dels paràmetres de la mateixa. Utilitzant la funció de versemblança i introduint la possibilitat de censura en la mostra, obtindrem estimadors dels paràmetres d'interès i podrem contrastar hipòtesis sobre els mateixos. Aquests contrastos facilitaran una anàlisi estadística confirmatòria de la hipòtesi distribucional assumida en l'especificació del model. Les modelitzacions paramètriques descrites anteriorment es troben lligades per restriccions sobre els paràmetres de manera que realitzant els contrastos de les mateixes es podrà justificar l'elecció de la distribució de probabilitat pel temps de supervivència.

4.2.8.1 La funció de Versemblança: el cas continu

Suposem que es disposa d'una mostra de n individus pels quals observem (t_i, δ_i) $i=1,2,\dots,n$. δ_i és un indicador que pren el valor 0 si el temps de durada és censurat i el valor 1 si no està censurat. El model de durada està especificat sobre un paràmetre θ , de tal forma que la funció de supervivència per a l'individu i -èsim és $S(t, \theta)$ mentre que la densitat és $f(t, \theta)$. La funció de versemblança és la densitat conjunta de les variables aleatòries que formen la mostra, avaluada en els valors que prenen les variables aleatòries (realització mostral), la qual depèn dels paràmetres del model. Suposant que la mostra està formada per observacions independents, la densitat conjunta equival al producte de les funcions de densitat marginals.

En el cas de què l'observació de la durada d'un individu sigui no censurada, l'aportació de la mateixa a la funció de versemblança és la densitat marginal avaluada en el punt: $f(t_i, \theta)$. Per contra, si l'observació es troba censurada, és a dir, només sabem que t ha pres un valor superior a una constant c , aleshores no podem especificar la densitat marginal en el punt, donat que no el coneixem, de tal forma que la contribució a la funció de versemblança en aquest cas serà:

$$\int_c^{\infty} f(t, \theta) dt = S(c, \theta)$$

Si la censuració és aleatòria però independent dels temps de durada, és a dir, els individus no poden estar censurats quan es troben davant d'un elevat risc d'abandonar (morir), o, més generalment, quan la llei de probabilitat de la censuració aleatòria no depèn del paràmetre θ , podem escriure la funció de versemblança com:

$$L(\theta) \propto \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(c_i; \theta)^{1-\delta_i}$$

i el logaritme de la versemblança, llevat de l'addició d'una constant:

$$\log L(\theta) = \sum_{i=1}^n [\delta_i \log f(t_i; \theta) + (1-\delta_i) \log S(c_i; \theta)]$$

4.2.8.2 La funció de Versemblança: el cas discret

Si suposem que la distribució de les durades és discreta, amb probabilitats $f_j(\theta)$ en els punts x_j ($x_1 < x_2 < \dots$), i que un individu censurat en c podria haver abandonat (mort) en c , aleshores la contribució a la versemblança d'un individu que abandona en x_j és $f_j(\theta)$, i la d'un individu censurat en c serà:

$$P(t > c) = S(c; \theta) = 1 - \sum_{j: x_j \leq c} f_j(\theta)$$

En termes de la funció d'atzar $\lambda_j(\theta)$ tenim que:

$$f_j(\theta) = \lambda_j(\theta) \prod_{k < j} [1 - \lambda_k(\theta)]$$

$$S(c; \theta) = \prod_{j: x_j \leq c} [1 - \lambda_j(\theta)]$$

La funció de versemblança serà doncs, en el punt x_j a on d_j individus han abandonat d'entre el total de r_j individus que resten en la mostra en el moment x_j :

$$[\lambda_j(\theta)]^{d_j} [1 - \lambda_j(\theta)]^{r_j - d_j}$$

Per a tot el recorregut de valors possibles x_j , $j=1,2,\dots$, el logaritme de la funció de versemblança serà:

$$\log L(\theta) = \sum_j [d_j \log \lambda_j(\theta) + (r_j - d_j) \log(1 - \lambda_j(\theta))]$$

4.2.8.3 Estimació de paràmetres i contrast d'hipòtesis

Un cop determinada apriorísticament la distribució de probabilitat del temps de supervivència a partir, per exemple, de la informació que proporcionen les representacions gràfiques i els procediments de tabulació comentats anteriorment, podrem especificar de manera explícita la funció de versemblança.

Així, en el cas de la distribució exponencial amb paràmetre λ , les funcions de densitat i de supervivència són:

$$f(t) = \lambda e^{-\lambda t} \quad i \quad S(t) = e^{-\lambda t}$$

i, per tant el logaritme de la funció de versemblança serà:

$$\log L(\lambda) = d \log \lambda - \lambda \sum_{i=1}^n t_i$$

essent d el nombre total d'abandonaments enregistrat.

Amb el logaritme de la funció de versemblança, podem en primer lloc obtenir l'estimació màxim versemblant dels paràmetres igualant a zero el vector de primeres derivades del logaritme de la funció de versemblança respecte cadascú dels paràmetres³.

En el cas de l'exponencial tenim:

$$\frac{\partial \log L(\lambda)}{\partial \lambda} = \frac{d}{\lambda} - \sum_{i=1}^n t_i = 0$$
$$\hat{\lambda} = \frac{d}{\sum t_i}$$

³ La convexitat de la funció de versemblança garanteix l'obtenció d'un màxim de la funció.

En segon lloc, mitjançant la matriu de segones derivades del logaritme de la funció de versemblança respecte els paràmetres obtenim la matriu d'informació observada. La seva inversa avaluada en el vector d'estimadors màxim versemblants i canviada de signe ens proporciona una estimació de la matriu de variàncies i covariàncies dels paràmetres estimats.

Per a l'exponencial:

$$\left[\frac{-\partial^2 \log L}{\partial \lambda^2} \right]^{-1} = \frac{\hat{\lambda}^2}{d} = \text{Var}(\hat{\lambda})$$

Aplicant els resultats segons els quals la distribució asimptòtica de l'estimador màxim versemblant del paràmetre θ és normal multivariant amb vector de mitjanes θ i matriu de variàncies i covariàncies igual a la inversa de la matriu d'informació de Fisher, es poden realitzar contrastos d'hipòtesis sobre els paràmetres i estimació per intervals. Aquest resultat asimptòtic conjuntament amb els de què, l'estimador màxim versemblant és solució única de l'equació que iguala a zero les primeres derivades del logaritme de la funció de versemblança respecte els paràmetres i que aquest estimador serà consistent, es mantindran sempre i quan es verifiquin les habituals condicions de regularitat (Rao, 1973, Cox i Hinkley, 1974).

D'altra banda, podem substituir la matriu d'informació de Fisher per la matriu d'informació observada avaluada en els paràmetres màxim versemblants estimats sense que es perdin les propietats desitjables que ens permetran realitzar inferència (Rao, 1973). Així, la matriu d'informació de Fisher:

$$\mathfrak{I}(\theta) = - \left[E \left(\frac{\partial^2 \log L}{\partial \theta_j \partial \theta_k} \right) \right]$$

es pot estimar mitjançant la matriu d'informació observada:

$$I(\hat{\theta}) = \left(\frac{-\partial^2 \log L(\theta)}{\partial \theta_j \partial \theta_k} \right)_{\theta=\hat{\theta}}$$

i es podrà utilitzar pel contrast d'hipòtesis l'estadístic de Wald segons el qual l'expressió:

$$(\hat{\theta} - \theta_0)' I(\hat{\theta}) (\hat{\theta} - \theta_0)$$

sota la hipòtesi nul·la es distribueix segons una χ^2 amb p graus de llibertat, essent p la dimensió del vector θ .

Un altre resultat asimptòtic d'interès en la contrastació d'hipòtesis té a veure amb l'estadístic de la Raó de Versemblança i la seva distribució. Així, si es mantenen les condicions de regularitat de la teoria de la màxima versemblança aleshores sota la hipòtesi nul·la $\theta = \theta_0$, la distribució asimptòtica de:

$$-2 [\log L(\theta_0) - \log L(\hat{\theta})]$$

és la χ^2 amb p graus de llibertat, essent p la dimensió del vector θ .

5. Modelització de Poblacions Heterogènies

Capítol 5

Modelització de Poblacions Heterogènies

En el capítol anterior han estat presentats mètodes no paramètrics i paramètrics per tal de modelitzar les dades de supervivència en presència de censuració i per a poblacions homogènies. En moltes situacions, però, serà d'interès la comparació de les funcions de supervivència o durada per a dos o més grups d'individus i, d'aquesta forma, captar les possibles diferències entre ells respecte les seves durades mitjanes o medianes. Aquesta anàlisi podrà tenir com a objectiu la verificació dels efectes de determinades situacions personals sobre l'eventualitat d'un canvi d'estat (abandonament, mort,...) o bé les conseqüències de determinades polítiques o actuacions sobre les decisions individuals. L'esmentada comparació de funcions de supervivència per a diferents col·lectius pot realitzar-se mitjançant anàlisis qualitatives basades en instruments gràfics o, més formalment, a partir de contrastos estadístics, com ara els de Wilcoxon i log-rank presentats anteriorment en l'apartat corresponent als mètodes no paramètrics. Tot i això, la necessitat de comparacions més acurades així com l'existència de situacions amb poblacions caracteritzades per un nombre elevat de variables obliguen a disposar de modelitzacions que incorporin els efectes d'aquestes.

Presentarem doncs un conjunt de modelitzacions de les durades o temps de supervivència les quals incorporaran, a més dels paràmetres propis de la distribució de les durades, altres paràmetres que recolliran els efectes de les variables explicatives sobre aquelles. Aquestes variables explicatives poden classificar-se segons els seus valors es mantinguin constants al llarg de tot el període analitzat o bé que depenguin del moment del temps en que es mesurin. Inicialment tindrem les modelitzacions dels temps de supervivència amb variables explicatives els valors de les quals es mantenen constants al llarg del temps i en apartats posteriors del present capítol introduïrem les variables canviant en el temps. La introducció d'aquestes tindrà efectes sobre les hipòtesis en les què es fonamenten les modelitzacions de les durades a més d'incrementar la magnitud dels càlculs en els procediments estadístics.

Existeixen dues alternatives de modelització dels temps de supervivència quan s'inclouen variables explicatives. La primera és la de la Durada Accelerada ("Accelerated Failure Time Models", AFT) i la segona és la dels Atzars Proporcional ("Proportional Hazard Models", PH). La diferència bàsica entre ambdues modelitzacions rau en la forma com es produeix l'efecte dels canvis en les variables explicatives sobre el temps de supervivència. Així, en el cas dels models de durada accelerada (AFT) l'efecte de les variables explicatives es produeix directament sobre el temps de supervivència mentre que en el cas dels models d'atzar proporcional (PH) aquest efecte és sobre l'atzar o risc d'abandonar (morir) i, per tant, indirectament sobre el temps de supervivència.

Una i altra modelitzacions alternatives poden especificar-se de forma plenament paramètrica, és a dir, quan es realitza una hipòtesi sobre la distribució de probabilitat del temps de supervivència a més de sobre l'estructura de la relació

entre les variables explicatives i la durada, de manera que haurem de realitzar inferència sobre els paràmetres de la distribució de probabilitat (λ, p, k, \dots) i sobre els paràmetres que mesuren els efectes de les explicatives sobre la durada (β). Alternativament, l'especificació dels models pot ser semiparamètrica, el que vol dir que només s'efectuarà inferència sobre els paràmetres associats als efectes de les explicatives i no sobre la distribució de probabilitat de la durada.

En el apartats que segueixen exposem cadascuna de les alternatives de modelització així com els procediments d'estimació dels paràmetres tant per a l'especificació paramètrica com per a la semiparamètrica. A més, reflectim les relacions existents entre ambdues modelitzacions les quals són especialment interessants en l'especificació paramètrica quan la distribució de probabilitat de les durades és l'exponencial o la Weibull, vist que aleshores resulten equivalents els resultats de la durada accelerada i els de l'atzar proporcional.

En el cas de la modelització d'atzar proporcional distingirem entre l'especificació i estimació per a temps continu de la de temps discret. Aquesta distinció entre temps continu i temps discret pot estar condicionada per la pròpia definició del problema a analitzar a més de per l'estructura de la base de dades disponible. D'altra banda, als efectes d'interpretació i valoració dels resultats de les estimacions, resultarà de més utilitat l'especificació discreta per la qual cosa serà estudiada amb més cura.

5.1 Models de Durada Accelerada (AFT)

Aquesta alternativa de modelització va ser proposada per Cox (1972) i desenvolupada per Prentice (1978); posteriorment ha estat analitzada amb més detall per nombrosos autors. Suposarem que volem estudiar els efectes sobre el temps de supervivència de la pertinença dels individus a dos col·lectius diferenciats. Disposarem d'una variable dicotòmica z , que prendrà el valor 0 si l'individu pertany a un grup i el valor 1 en cas contrari. Aleshores podem representar la funció de supervivència pels individus del grup 0 mitjançant l'expressió $S_0(t)$. Doncs bé, en el model de durada accelerada existeix una constant (ψ) tal que la funció de supervivència pel grup 1, és a dir quan la variable dicotòmica pren el valor $z = 1$, és $S_1(t) = S_0(\psi t)$, de tal forma que les altres funcions associades a la de supervivència seran:

$$f_1(t) = \psi f_0(\psi t) \quad \lambda_1(t) = \psi \lambda_0(\psi t)$$

Una manera d'interpretar l'anterior especificació és que qualsevol individu amb un temps de supervivència igual a t quan $z = 0$, tindria una durada igual a t/ψ en el cas que $z = 1$, la qual cosa implicaria que la relació entre les variables aleatòries corresponents a cadascun dels dos col·lectius seria $T_1 = T_0/\psi$. Com podem observar, en aquesta formulació els efectes de les variables explicatives es produeixen directament sobre el temps de supervivència.

Generalitzant, si disposem d'un vector de variables explicatives z amb valors constants en el temps, suposarem que existeix una funció de les variables explicatives, $\psi(z)$, tal que les funcions caracteritzadores dels temps de supervivència seran:

$$S(t; z) = S_0(t\psi(z)) \quad f(t; z) = f_0(t\psi(z))\psi(z) \quad \lambda(t; z) = \lambda_0(t\psi(z))\psi(z)$$

on S_0 fa referència a unes condicions estàndard, per exemple quan $z = 0$ i aleshores $\psi(0) = 1$. En termes de les variables aleatòries tindrem $T = T_0/\psi(z)$ on T_0 té funció de supervivència S_0 .

Prenent logaritmes en l'expressió anterior ens queda:

$$\log T = \mu_0 - \log \psi(z) + w$$

on $\mu_0 = E(\log T_0)$ i w és una variable aleatòria amb mitjana zero i amb una funció de densitat igual a $f_0(\exp(-w)) \exp(-w)$ la qual no depèn de z . Si, com és el cas habitualment, disposem de diverses variables explicatives serà convenient especificar una forma paramètrica per a $\psi(z)$ que depengui d'uns paràmetres β . Així, donades les restriccions $\psi(z;\beta) \geq 0$ i $\psi(0;\beta) = 1$, una possible especificació que facilitarà, a més, propers desenvolupaments del model, serà $\psi(z;\beta) = \exp(z;\beta)$, essent z un vector fila i β un vector columna, ambdós de dimensió p . Aleshores tindrem el següent model de regressió loglineal:

$$\log T = \mu_0 - z\beta + w$$

L'expressió loglineal del model facilita la interpretació dels efectes de les variables explicatives sobre el temps de supervivència. Així, tenim que el paràmetre $(\partial \log T / \partial z_j = -\beta_j)$ mesura l'efecte sobre la durada com a conseqüència d'una variació unitària en l'explicativa corresponent.

L'especificació plenament paramètrica del model de durada accelerada es deduirà de la distribució de probabilitat per al temps de supervivència que es seleccioni i, per tant, un cop formulades les funcions S_0 , f_0 i λ_0 . D'aquesta forma coneixerem d'una manera explícita la distribució de probabilitat de la variable aleatòria w en el model loglineal anterior.

En aquest punt, cal tornar la vista a l'apartat anterior on han estat estudiades en profunditat algunes distribucions teòriques d'entre les més utilitzades en l'anàlisi de la supervivència. Per a totes aquelles distribucions de probabilitat analitzades ha estat deduïda una expressió loglineal del temps de supervivència en funció d'una variable aleatòria w , la qual tindria una distribució particular segons s'especifiqués la distribució per a la durada. Així, per a la distribució exponencial teníem la següent expressió loglineal de T i la corresponent densitat de W :

$$Y = \log T = \alpha + z\beta + W$$

$$f(w) = \exp(w - e^w) \quad -\infty < w < \infty$$

mentre que, per a la distribució Weibull és la mateixa expressió anterior però amb la introducció d'un paràmetre d'escala, σ , tal que:

$$Y = \log T = \alpha + z\beta + \sigma w$$

$$f(w) = \exp(w - e^w) \quad -\infty < w < \infty$$

A partir d'aquesta especificació paramètrica del model loglineal podrem formalitzar la funció de versemblança per a una mostra de n individus dels quals es disposa d'informació sobre y_i, δ_i , essent aquests el logaritme del temps de supervivència observat i l'indicador d'observació no censurada per a l'individu i -èsim, i sobre les seves característiques individuals representades pel vector de variables explicatives $z_i = (z_{1i}, \dots, z_{pi})$.

La funció de versemblança serà doncs:

$$L(\beta, \sigma) = \prod_{i=1}^n [\sigma^{-1} f(w_i)]^{\delta_i} [1 - F(w_i)]^{1-\delta_i}$$

on ara $w_i = (y_i - z_i\beta)/\sigma$ i $F(w)$ és la funció de distribució de w . A més, la primera component del vector z és igual a la unitat, $z_1 = 1$, de tal forma que el paràmetre $\beta_1 = \alpha = \mu_0$.

L'anterior expressió de la funció de versemblança és general per a totes les distribucions teòriques estudiades amb anterioritat i es concretarà segons sigui la distribució de probabilitat de w en cada cas. Així, per exemple, en la distribució Weibull, on $f(w) = \exp(w - e^w)$ i $S(w) = 1 - F(w) = \exp(-e^w)$, el logaritme de la funció de versemblança serà:

$$\log L(\beta; \sigma) = \sum_{i=1}^n \left[\delta_i \left(\frac{y_i - z_i \beta}{\sigma} - \log \sigma \right) - \exp \left(\frac{y_i - z_i \beta}{\sigma} \right) \right]$$

Les primeres i segones derivades del logaritme de la funció de versemblança general per a qualsevol distribució teòrica són:

$$\frac{\partial \log L}{\partial \beta_j} = -\sigma^{-1} \sum_{i=1}^n z_{ij} [\delta_i f'(w_i)/f(w_i) - (1-\delta_i)f(w_i)/S(w_i)] \quad j=1, \dots, p$$

$$\frac{\partial \log L}{\partial \sigma} = -\sigma^{-1} \sum_{i=1}^n \{ w_i [\delta_i f'(w_i)/f(w_i) - (1-\delta_i)f(w_i)/S(w_i)] - \delta_i \}$$

i

$$\frac{\partial^2 \log L}{\partial \beta_j \partial \beta_k} = -\sigma^{-2} \sum_{i=1}^n A_i z_{ij} z_{ik} \quad j, k=1, \dots, p$$

$$\frac{\partial^2 \log L}{\partial \beta_j \partial \sigma} = -\sigma^{-2} \sum_{i=1}^n A_i w_i z_{ij} - \sigma^{-1} \left(\frac{\partial \log L}{\partial \beta_j} \right) \quad j=1, \dots, p$$

$$\frac{\partial^2 \log L}{\partial \sigma^2} = -\sigma^{-2} \sum_{i=1}^n [A_i w_i^2 - \delta_i] - 2\sigma^{-1} \left(\frac{\partial \log L}{\partial \sigma} \right)$$

on

$$A_i = -\delta_i \left(\frac{\partial^2 \log f(w_i)}{\partial w_i^2} \right) - (1-\delta_i) \left(\frac{\partial^2 \log S(w_i)}{\partial w_i^2} \right)$$

Si la distribució inclou paràmetres de forma (k en el cas de la Gamma), existiran a més altres derivades primeres i segones respecte aquests paràmetres.

El càlcul de les primeres i segones derivades del logaritme de la funció de versemblança respecte els paràmetres β i σ (i k en el seu cas), ens permetrà, mitjançant la utilització d'algoritmes d'optimització no lineal com ara el de Newton-Raphson, l'obtenció dels estimadors màxim versemblants dels paràmetres així com de la matriu d'informació observada, la inversa de la qual avaluada en les estimacions màxim versemblants anteriors, representarà l'estimació de la matriu de variàncies i covariàncies dels paràmetres estimats.

Per a completar la inferència estadística en aquests models paramètrics de durada accelerada podrà fer-se ús de la teoria assintòtica per a les estimacions màxim-versemblants (Rao, 1973) i aprofitar les seves propietats per a deduir estadístics que permetin la contrastació d'hipòtesis sobre els paràmetres del model, tal i com ja ha estat comentat en apartats anteriors. Així podrem estudiar la significació estadística dels paràmetres que acompanyen a les variables explicatives del model per tal de seleccionar aquelles que efectivament són rellevants en l'explicació de l'heterogeneïtat de les funcions de supervivència de la població, és a dir, contrastar la hipòtesi nul·la $H_0: \beta_j = 0$ mitjançant l'estadístic de Wald:

$$\hat{\beta}_j^2 / \hat{var}(\hat{\beta}_j) \sim \chi_1^2$$

També podrem contrastar la significació conjunta dels paràmetres del model associats a variables explicatives mitjançant l'estadístic de la Raó de Versemblança comparant el model sense variables explicatives amb el que les incorpora totes. La hipòtesi nul·la serà en aquest cas $H_0: \beta_1 = \dots = \beta_s = 0$ i l'estadístic de la Raó de Versemblança:

$$-2[\log L(H_0) - \log L(\hat{\beta})] \sim \chi_p^2$$

D'altra banda i tenint en compte que algunes distribucions de probabilitat inclouen com a casos particulars a unes altres, podrem realitzar implícitament una validació de la hipòtesi efectuada sobre la distribució de probabilitat de les durades. Així, la distribució Gamma generalitzada inclou com a casos particulars a l'exponencial, la Weibull, la Gamma simple i la lognormal segons els valors dels tres paràmetres que la caracteritzen de manera que, si contrastem cadascuna de les restriccions que donen lloc als esmentats casos particulars, estarem en disposició de seleccionar a partir de criteris estadístics entre modelitzacions alternatives. Per exemple, en la distribució Gamma generalitzada, si es verifica que $k = 1$ aleshores ens trobarem davant de la distribució Weibull i, per tant, un cop estimat el model més general contrastarem l'esmentada restricció sobre el paràmetre mitjançant l'estadístic de Wald:

$$(\hat{k} - 1)^2 / \hat{var}(\hat{k}) \sim \chi_1^2$$

o, alternativament, mitjançant el contrast dels multiplicadors de Lagrange a partir de l'estimació del model restringit, el qual contrast és asimptòticament equivalent al de Wald i no necessita de l'estimació màxim versemblant del model més general. Aquest darrer doncs, serà d'utilitat en el cas que no es disposi d'unes eines informàtiques, tant pel que fa al software com al hardware, que permetin l'estimació del model Gamma generalitzat.

Finalment, mitjançant les estimacions màxim versemblants dels paràmetres del model de supervivència i coneguda la distribució de probabilitat de les durades és possible realitzar prediccions del temps de supervivència per a uns valors prefixats de les variables explicatives. Així, del model loglineal inicial podrem escriure:

$$\hat{y}_p = \log \hat{t}_p = z\beta + \sigma w_p$$

on w_p és el percentil p de la distribució de probabilitat base, la qual dependrà a la seva vegada de l'estimació màxim versemblant de qualsevol paràmetre de forma de la distribució (k en la Gamma).

A partir de l'expressió anterior es poden deduir dos resultats de molt interès per a l'anàlisi de les dades de supervivència. En primer lloc, a partir del percentil 50 de la distribució de probabilitat base (la de Valor Extrem Tipus I en el cas de treballar amb l'exponencial o la Weibull, per exemple) es pot calcular la *durada mediana* estimada per a un conjunt d'individus caracteritzat per una combinació determinada de valors de les variables explicatives, z . Aquesta durada mediana és un resultat sintetitzador de la informació dels temps de supervivència d'una població especialment indicada en presència de censuració, com és el cas habitualment.

Un segon resultat, derivat de l'anterior, consistirà en el càlcul del temps de supervivència estimat per a un rang de percentils de la distribució base suficientment ample com per permetre la representació gràfica de la funció de supervivència estimada per a un conjunt d'individus caracteritzats pels seus valors en les variables explicatives z . D'aquesta manera podrem comparar la durada per a subcol·lectius d'individus diferents.

Ambdós resultats seran obtinguts en l'aplicació presentada posteriorment per tal de realitzar una comparació de les durades medianes estimades de dos individus arquetípics diferents així com de la forma de les seves funcions de supervivència estimades corresponents.

5.2 Models d'Atzar Proporcional (PH)

Els models d'atzar proporcional van ser proposats inicialment per Cox (1972) en un article publicat en el *Journal of the Royal Statistical Society* el qual va ser la llavor per a posteriors desenvolupaments de nombrosos autors, el que ha portat a la situació actual d'abundància de literatura referida a aquesta alternativa de modelització del temps de supervivència. Les principals aportacions al seu desenvolupament han estat realitzades per Peto (1972), Kalbfleisch i Prentice (1973), Cox (1975), Efron (1977), Prentice i Gloeckler (1978), i Oakes (1981) entre d'altres.

En la formulació d'aquest model tornem a suposar que disposem d'una mostra de n individus caracteritzats per un vector de variables explicatives constants en el temps z . A més, coneixem el temps de supervivència de cadascun dels individus de la mostra, t_i , i l'indicador d'observació no censurada, δ_i . Inicialment, suposem que el temps de supervivència es distribueix de forma contínua i que no existeixen dos individus amb el mateix valor de t_i . En el model d'atzar proporcional l'efecte de les variables explicatives es produeix de manera multiplicativa sobre la funció d'atzar mitjançant un factor ψ que no depèn del temps de durada, t , de manera que:

$$\lambda(t; z) = \lambda_0(t) \psi(z)$$

i en aquest cas λ_0 és l'atzar per a un individu en les condicions estàndard, $z = 0$, de tal forma que $\psi(0) = 1$. Així, les funcions associades a la d'atzar seran:

$$S(t; z) = [S_0(t)]^{\psi(z)} \quad f(t; z) = \psi(z) [S_0(t)]^{\psi(z)-1} f_0(t)$$

De manera equivalent al model de durada accelerada (AFT), en la present especificació podem reparametritzar convenientment la funció $\psi(z)$ i la forma més habitual de fer-ho és $\psi(z;\beta) = \exp(z\beta)$. Així podem interpretar els paràmetres β_j , tenint en compte que,

$$\frac{\partial \log \lambda(t; z)}{\partial z_j} = \beta_j$$

com l'efecte proporcional de la variable explicativa z_j sobre la probabilitat condicional de finalitzar una durada o període de supervivència. Aquest efecte és constant i no depèn del temps de supervivència en que es mesuri.

Els avantatges d'aquesta modelització del temps de supervivència respecte els models de durada accelerada exposats anteriorment són, en primer lloc, la senzilla interpretació dels efectes de les variables explicatives sobre l'atzar o risc d'abandonar (morir). En segon lloc, la hipòtesi de proporcionalitat dels atzars o riscos per a diferents valors de les variables explicatives ha estat confirmada empíricament en nombrosos camps d'aplicació. En tercer lloc, la incorporació de les problemàtiques de les variables explicatives canviants en el temps i dels diferents motius de finalització de les durades és més senzilla en aquesta formulació. Finalment, en els models d'atzar proporcional és possible l'estimació dels paràmetres que mesuren els efectes de les variables explicatives sobre les durades tant si especifiquem de forma explícita la funció d'atzar base λ_0 , a partir d'alguna de les distribucions de probabilitat teòriques analitzades anteriorment, com si aquella funció és arbitrària.

Analitzarem en primer lloc l'estimació plenament paramètrica dels models d'atzar proporcional i, posteriorment, la semiparamètrica, és a dir, aquella que no necessita de l'especificació de la funció d'atzar base.

5.2.1 Estimació paramètrica dels models d'atzar proporcional

Una vegada especificat el model d'atzar proporcional continu, és possible especificar la funció d'atzar base λ_0 a partir d'alguna de les distribucions de probabilitat de les durades analitzades anteriorment i el model es formularà com:

$$\lambda(t; z) = \lambda_0(t; \alpha) \exp(z\beta)$$

essent α el vector de paràmetres que caracteritzen la distribució del temps.

Ara podrem formalitzar la funció de versemblança per a una mostra de n individus dels quals es disposa d'informació sobre t_i, δ_i , essent aquests el temps de supervivència observat i l'indicador d'observació no censurada per a l'individu i -èsim, i sobre les seves característiques individuals representades pel vector de variables explicatives $z_i = (z_{1i}, \dots, z_{pi})$. La funció de versemblança serà doncs:

$$L(\alpha, \beta; t_i) = \prod_{i=1}^n [f(t_i; z_i; \alpha, \beta)]^{\delta_i} [S(t_i; z_i; \alpha, \beta)]^{1-\delta_i}$$

la qual en termes de la funció d'atzar és:

$$L(\alpha, \beta; t_i) = \prod_{i=1}^n [\lambda(t_i; z_i; \alpha, \beta)]^{\delta_i} [S(t_i; z_i; \alpha, \beta)]$$

Reordenant i substituint segons l'especificació del model, la versemblança i el logaritme de la versemblança són:

$$L(\alpha, \beta; t_i) = \prod_{i=1}^n [\lambda_0(t_i; \alpha) \exp(z_i \beta)]^{\delta_i} [S_0(t_i; \alpha)]^{\exp(z_i \beta)}$$

$$\log L(\alpha, \beta; t_i) = \sum_{i=1}^n \{ \delta_i [\log \lambda_0(t_i; \alpha) + z_i \beta] + \exp(z_i \beta) \log S_0(t_i; \alpha) \}$$

En el cas concret de la distribució Weibull, per exemple, tenim que:

$$\lambda_0(t; p, \lambda) = p \lambda (\lambda t)^{p-1} \quad S_0(t; p, \lambda) = \exp[-(\lambda t)^p]$$

i el logaritme de la versemblança serà:

$$\log L(p, \lambda, \beta; t_i) = \sum_{i=1}^n \left\{ \delta_i [\log p + p \log \lambda + (p-1) \log t_i + z_i \beta] - \exp(z_i \beta) (\lambda t_i)^p \right\}$$

expressió a partir de la qual es podran deduir les primeres i segones derivades respecte cadascun dels paràmetres del model $(p, \lambda, \beta_1, \dots, \beta_s)$ les quals ens permetran l'obtenció dels estimadors màxim versemblants dels paràmetres i de les variàncies i covariàncies d'aquests paràmetres estimats.

Utilitzant els resultats assintòtics aplicables a les estimacions màxim versemblants en les condicions de regularitat habituals serà possible, de manera equivalent a com ha estat exposat anteriorment en el model de durada accelerada, contrastar hipòtesis sobre el valor dels paràmetres del model (els de la distribució de la durada i/o els que mesuren els efectes de les explicatives) i efectuar estimació per interval.

L'especificació i estimació paramètrica del model d'atzar proporcional ha estat utilitzada en aplicacions empíriques, especialment en la seva formulació Weibull (Kennan, 1985, Serrat i Gil, 1991, entre moltes d'altres en l'àmbit econòmic). Aquesta especificació, conjuntament amb l'exponencial i la loglogística són les més utilitzades a resultes de que qualsevol altra distribució de probabilitat complica excessivament el procés de maximització de la funció de versemblança, per la qual cosa perdria utilitat davant, d'una banda, de la mateixa especificació en la modelització de la durada accelerada o, d'altra, de l'estimació semiparamètrica del model d'atzar proporcional. Aquesta alternativa semiparamètrica en l'estimació del model d'atzar proporcional és la que ha estat més utilitzada en els treballs aplicats per la flexibilitat que proporciona davant dels mètodes paramètrics anteriors.

5.2.2 Estimació semiparamètrica contínua del model d'atzar proporcional

Per a l'estimació semiparamètrica contínua del model d'atzar proporcional existeixen dues aproximacions alternatives. La primera proposada per Cox (1972 i 1975) és la de la Versemblança Parcial; la segona és la de la Versemblança Marginal i va ser desenvolupada per Kalbfleisch i Prentice (1973), la qual es troba molt relacionada amb els contrastos de rangs exposats anteriorment. Totes dues, però, presenten resultats equivalents i difereixen únicament en la construcció i interpretació de la versemblança corresponent.

L'objectiu del mètode és l'estimació dels paràmetres β i de λ_0 , quan la forma d'aquesta darrera funció d'atzar base és totalment desconeguda. Suposarem que disposem d'una mostra de n individus dels quals es coneix el temps de supervivència, t_i , i, inicialment, sense observacions censurades ni repetides per a un mateix valor de t . Siguin t_1, t_2, \dots, t_n els temps de supervivència per a cadascun dels individus de la mostra a partir dels quals definirem, en primer lloc, un estadístic d'ordre, $O(t)=(t_{(1)}, t_{(2)}, \dots, t_{(n)})$, corresponent a les durades ordenades de forma creixent (és a dir, $t_{(1)} < t_{(2)} < \dots < t_{(n)}$) i, en segon lloc, un estadístic de rang, $r(t)=((1), (2), \dots, (n))$, que assignarà a cada observació del temps el seu lloc en la ordenació creixent de les mateixes. La informació que proporcionen a la vegada $O(t)$ i $r(t)$ és equivalent a la de les durades originals (sense ordenar) i, a més, una transformació monòtona del temps de supervivència, com per exemple $u = t^2$, alteraria els valors del temps de supervivència però deixaria igual l'estadístic de rang $r(t)$, és a dir, no canviaria la ordenació de les durades. D'altra banda, si apliquem la transformació monòtona anterior al model d'atzar proporcional original, $\lambda(t; z) = \lambda_0(t) \exp(z\lambda)$, i efectuant un canvi de variable tindriem:

$$\lambda(u; z) = \lambda_0(u^{1/2}) \frac{1}{2u^{1/2}} \exp(z\beta) = \lambda_1(u) \exp(z\beta)$$

amb el què el problema de la inferència sobre els paràmetres β no es veuria afectada per la transformació i seria la mateixa tant en aquest cas (u) com en l'anterior (t) sempre i quan es desconegui la forma explícita de λ_0 .

Com a conseqüència d'això, únicament l'ordenació dels temps de supervivència, $r(t)$, aporta informació per a la inferència dels paràmetres que mesuren els efectes de les variables explicatives, β , en el cas que la funció d'atzar base sigui desconeguda o, en altres paraules, si no es disposa d'informació sobre la distribució de probabilitat que segueixen les durades serà suficient per a mesurar els efectes de les variables explicatives sobre el risc o atzar d'abandonar (morir) amb conèixer si un individu amb unes característiques personals determinades presenta un temps de supervivència superior o inferior al d'un altre amb característiques diferents. Caldrà, doncs, atendre únicament a l'ordenació de les durades i a la distribució conjunta, $P[(1),(2),\dots,(n)]$, de les seves permutacions possibles.

A partir de la distribució de probabilitat marginal dels rangs, Kalbfleisch i Prentice (1973) obtenen, fonamentant-se en els resultats de Fraser (1968) i de Kalbfleisch i Sprott (1970), la versemblança marginal dels paràmetres β , $L(\beta)$, la qual serà proporcional a la probabilitat d'observar el vector de rangs mostral:

$$\begin{aligned}
 L(\beta) &\propto P[r = ((1),(2),\dots,(n)) ; \beta] = \int_0^\infty \int_{t_{(1)}}^\infty \dots \int_{t_{(n-1)}}^\infty \prod_{i=1}^n f(t_{(i)}; z_{(i)}) dt_{(n)} \dots dt_{(1)} = \\
 &= \int_0^\infty \int_{t_{(1)}}^\infty \dots \int_{t_{(n-1)}}^\infty \prod_{i=1}^n f_i[t_{(i)}; z_{(i)} | (t_{(1)}; z_{(1)}), \dots, (t_{(i-1)}; z_{(i-1)})] = \\
 &= P[(t_{(1)}; z_{(1)})] P[(t_{(2)}; z_{(2)}) | (t_{(1)}; z_{(1)})] \dots P[(t_{(n)}; z_{(n)}) | (t_{(1)}; z_{(1)}), (t_{(2)}; z_{(2)}), \dots, (t_{(n-1)}; z_{(n-1)})] =
 \end{aligned}$$

$$\begin{aligned}
&= \frac{\lambda(t_{(1)}; z_{(1)})}{\lambda(t_{(1)}; z_{(1)})} \cdot \frac{\lambda(t_{(2)}; z_{(2)})}{\lambda(t_{(2)}; z_{(1)}) + \lambda(t_{(2)}; z_{(2)})} \cdots \frac{\lambda(t_{(n)}; z_{(n)})}{\lambda(t_{(n)}; z_{(1)}) + \lambda(t_{(n)}; z_{(2)}) + \cdots + \lambda(t_{(n)}; z_{(n)})} = \\
&= \frac{\lambda_0(t_{(1)}) \exp(z_{(1)}\beta)}{\lambda_0(t_{(1)}) \exp(z_{(1)}\beta)} \cdot \frac{\lambda_0(t_{(2)}) \exp(z_{(2)}\beta)}{\lambda_0(t_{(2)}) [\exp(z_{(1)}\beta) + \exp(z_{(2)}\beta)]} \cdots \\
&\quad \cdots \frac{\lambda_0(t_{(n)}) \exp(z_{(n)}\beta)}{\lambda_0(t_{(n)}) [\exp(z_{(1)}\beta) + \exp(z_{(2)}\beta) + \cdots + \exp(z_{(n)}\beta)]} = \\
&= \frac{\exp(\sum_{i=1}^n z_i\beta)}{\prod_{i=1}^n \left[\sum_{i \in R(t_{(i)})} \exp(z_i\beta) \right]}
\end{aligned}$$

essent $R(t_{(i)})$ el conjunt d'estadístics d'ordre associats als individus que es troben davant de la possibilitat d'abandonar (morir) en el moment $t_{(i)}$, és a dir, estaria format per $R(t_{(i)}) = \{(i), (i+1), \dots, (n)\}$.

En presència d'observacions censurades, Kalbfleisch i Prentice utilitzen un argument semblant a l'anterior, si més no, sempre que es pugui suposar que les observacions censurades s'esdevenen a continuació dels abandonaments. Mitjançant aquest supòsit, el qual és totalment realista en el cas de la censuració Tipus II i més arriscat en el cas de la censuració aleatòria independent, malgrat que la pèrdua d'informació produïda és insignificant de cara a la inferència dels paràmetres β , pot obviar-se l'obtenció de la versemblança marginal a partir de la suma d'un elevat nombre de termes iguals a la versemblança en absència de censuració. Aquesta suma vindrà com a conseqüència de, per exemple, disposant de la següent informació sobre els temps de supervivència de tres individus, 10, 20* i 30, indicant l'asterisc * una observació censurada, l'estadístic de rang corresponent podria ser qualsevol dels dos següents (1,2,3) ó (1,3,2) pel fet que

és desconeix exactament el temps de supervivència de l'observació censurada en 20 i, per tant, podria ser inferior o superior a 30. La probabilitat marginal, doncs, seria la suma de les probabilitats corresponents a les dues situacions possibles calculades com en el cas d'absència de censuració. Si suposem que les observacions censurades es produeixen sempre després dels abandonaments pot calcular-se la distribució marginal conjunta dels rangs mitjançant el producte de les corresponents probabilitats condicionades seguint la regla de la cadena tal com s'ha fet anteriorment.

Siguin ara, doncs, n individus dels quals d han abandonat i la resta són observacions censurades. Els temps de supervivència observats una vegada ordenats de forma creixent són $t_{(1)} < t_{(2)} < \dots < t_{(d)}$, mentre que q_i individus es troben censurats en l'interval $[t_{(i)}, t_{(i+1)})$ per $i = 0, 1, \dots, d$ i on $t_{(d+1)} = \infty$ i $t_{(0)} = 0$. Si les observacions censurades es produeixen després dels abandonaments tindrem que $t_{(i)} < t_{i1}, \dots, t_{iq_i}$ ($i = 0, 1, \dots, d$), essent t_{i1}, \dots, t_{iq_i} els temps de supervivència no observats associats als individus censurats en l'interval $[t_{(i)}, t_{(i+1)})$ i, aleshores, les observacions censurades en l'esmentat interval formaran part del conjunt de risc $R(t_{(i)})$ i s'hauran de tenir en compte per al càlcul de les probabilitats condicionades anteriors. Així doncs, la versemblança marginal en presència de censuració és:

$$L(\beta) = \frac{\exp\left(\sum_{i=1}^d z_{(i)}\beta\right)}{\prod_{i=1}^d \left[\sum_{l \in R(t_{(i)})} \exp(z_l\beta) \right]}$$

on ara $R(t_{(i)}) = \{(i), i1, \dots, iq_i, \dots, (d), d1, \dots, dq_d\}$ novament és el conjunt d'individus que es troben davant de la possibilitat d'abandonar en $t_{(i)}$ el qual incorporarà les observacions censurades quan correspongui.

Però no només ens trobarem davant de l'eventualitat de la presència d'observacions censurades sinó que, a més, malgrat es treballi sobre la hipòtesi del temps com a variable aleatòria contínua, en la pràctica podrem tenir observacions repetides en cada instant, degut a problemes d'error de mesura en l'enregistrament dels temps de supervivència. En les aplicacions empíriques, quan es mesura el temps es fa habitualment mitjançant anys, trimestres, mesos, setmanes, dies, etc., per la qual cosa no serà difícil trobar dues observacions amb el mateix temps de supervivència per bé que l'esdeveniment que produeix la finalització de la durada s'hagi produït en instants diferents (tot i que propers). Així doncs, cal adaptar la versemblança marginal a la possibilitat de l'existència d'observacions repetides.

Els esmentats autors dedueixen la versemblança marginal utilitzant un argument semblant al de la incorporació de les observacions censurades. Dels n individus considerats, s'han enregistrat com abandonaments en l'instant $t_{(i)}$ ($i=1, \dots, d$) els individus i_1, \dots, i_{m_i} , on $\sum_{i=1}^d m_i = n$. De la mateixa forma que en el cas de les observacions censurades la informació disponible sobre l'estadístic de rangs és únicament parcial vist que dels m_i individus que abandonen en el moment $t_{(i)}$ és conegut que tenen una ordenació o rang inferior als m_j individus que ho fan en $t_{(j)}$ per $j > i$, però l'ordenació dels abandonaments enregistrats en $t_{(i)}$ és desconegut. La distribució marginal conjunta de l'estadístic de rangs s'obtindria com a suma de les $\prod_{i=1}^d m_i!$ probabilitats calculades com en el cas d'absència de censuració i observacions repetides, corresponent cada sumand a la possible permutació dels rangs en cada moment del temps.

Si es té en compte que l'ordenació dels temps de supervivència en el moment $t_{(i)}$ serà independent de l'obtinguda per a l'instant $t_{(j)}$, la suma de les probabilitats anterior quedarà reduïda al producte de d sumes, una per a cada

temps de supervivència enregistrat. Es defineix com Q_i al conjunt de totes les permutacions possibles dels símbols $i1, \dots, im_i$, és a dir, totes les ordenacions possibles pels individus que han abandonat en el moment $t_{(i)}$, i sigui $P=(p_1, \dots, p_{m_i})$ un element qualsevol de Q_i . Com abans, $R(t_{(i)})$ és el conjunt de risc o d'individus que es troben davant de l'eventualitat d'abandonar en el moment $t_{(i)}$, i $R(t_{(i)}, p_r)$ és el conjunt $R(t_{(i)}) - (p_1, \dots, p_{r-1})$, és a dir, el conjunt de risc havent eliminat les observacions anteriors a la r en la permutació corresponent. La funció de versemblança marginal per a β és:

$$L(\beta) = \prod_{i=1}^d \left\{ \exp(s_i \beta) \sum_{P \in Q_i} \prod_{r=1}^{m_i} \left[\sum_{l \in R(t_{(i)}, p_r)} \exp(z_l \beta) \right]^1 \right\}$$

on $s_i = \sum_{j=1}^{m_i} z_{ij}$ és la suma de les variables explicatives dels individus que han abandonat en el moment $t_{(i)}$.

Aquesta expressió de la versemblança marginal per a β és la que posteriorment s'utilitzarà en l'aplicació qualificant-la de versemblança marginal exacta per a distingir-la de les aproximacions a la mateixa que han estat proposades. Aquestes aproximacions es troben justificades pel fet que si el nombre d'observacions repetides en cada moment del temps és molt elevat aleshores el càlcul de la funció de versemblança marginal resultarà extremadament complicat vist el gran nombre de sumes presents en el denominador de l'expressió.

La primera aproximació va ser proposada per Peto (1972) i desenvolupada per Breslow (1974), la qual resulta convenient si el nombre d'individus, m_i , que abandonen en cada moment del temps és reduït si es compara amb el total d'individus en disposició d'abandonar, $R(t_{(i)})$, situació en la qual es tindria que $R(t_{(i)})$ és aproximadament igual a $R(t_{(i)}, p_r)$ en cadascuna de les permutacions i

l'expressió de la versemblança marginal quedaria:

$$L(\beta) = \prod_{i=1}^d \frac{\exp(s_i \beta)}{\left[\sum_{l \in R(t_{i0})} \exp(z_l \beta) \right]^{m_i}}$$

Una altra alternativa proposada per Efron (1977) com a aproximació a la versemblança marginal exacta de Kalbfleisch i Prentice consisteix en canviar el denominador de l'expressió original de manera que:

$$L(\beta) = \prod_{i=1}^d \left(\frac{\exp(s_i \beta)}{\prod_{j=1}^{m_i} \left[\sum_{l \in R(t_{i0})} \exp(z_l \beta) - \frac{(j-1)}{d_i} \sum_{l \in D(t_{i0})} \exp(z_l \beta) \right]} \right)$$

on $D(t_{(i)})$ és el conjunt d'individus que abandonen en $t_{(i)}$.

Respecte les anteriors aproximacions cal dir, en primer lloc, que en absència d'observacions repetides per a un mateix moment del temps ambdues expressions coincideixen amb l'expressió de la versemblança exacta i, a la vegada, amb la deduïda inicialment quan es considerava el cas d'absència de censuració. A més, aquestes expressions només seran d'utilitat en el cas que el nombre d'observacions repetides no sigui massa elevat respecte el nombre d'individus en el conjunt de risc. Si aquest nombre és efectivament molt elevat i el càlcul de la versemblança marginal exacta no és accessible per problemes computacionals caldrà efectuar les estimacions sobre la formulació del model discret que s'estudiarà més endavant.

Un cop obtinguda la funció de versemblança marginal per a β , en absència de censuració i observacions repetides o quan ambdues circumstàncies són presents, mitjançant les primeres i segones derivades del logaritme de la funció de

versemblança marginal respecte els paràmetres podran calcular-se les estimacions màxim versemblants dels paràmetres β i de les variàncies i covariàncies dels mateixos. Novament, utilitzant els resultats asimptòtics aplicables a les estimacions màxim versemblants podrem completar la inferència amb el contrast de les hipòtesis sobre els valors dels paràmetres que siguin més oportunes.

Una vegada finalitzada la inferència sobre els paràmetres β que mesuren els efectes de les variables explicatives sobre la funció d'atzar, resta encara pendent l'estimació de la funció d'atzar base sobre la qual es produeixen aquells efectes. Per tal d'obtenir una estimació no paramètrica de la funció d'atzar base (λ_0) o, alternativament, de la funció de supervivència base (S_0), Kalbfleisch i Prentice formulen el model en termes discrets vist que el desconeixement de la distribució de probabilitat de les durades només permetrà l'estimació de la funció base en aquell moment del temps on s'ha enregistrat un abandonament. L'especificació del model en termes de la funció de supervivència és:

$$S(t; z) = [S_0(t)]^{\exp(z\beta)}$$

mentre que en el cas de temps discret sabem que:

$$S(t_i) = \prod_{j=1}^{i-1} (1 - \lambda_j) \quad f(t_i) = \lambda_i \prod_{j=1}^{i-1} (1 - \lambda_j)$$

de tal forma que en el model d'atzar proporcional podem escriure:

$$S(t_i; z) = \prod_{j=1}^{i-1} (1 - \lambda_0(t_j))^{\exp(z\beta)}$$

mentre que, com a resultat de la formulació discreta tenim:

$$1 - \lambda_i = \frac{S(t_{i+1})}{S(t_i)}$$

i, per tant, la funció d'atzar en el moment t_i donat el vector de variables explicatives z serà:

$$1 - \lambda(t_i; z) = \frac{S(t_{i+1}; z)}{S(t_i; z)} = \frac{\prod_{j=1}^i (1 - \lambda_0(t_j))^{\exp(z\beta)}}{\prod_{j=1}^{i-1} (1 - \lambda_0(t_j))^{\exp(z\beta)}} = (1 - \lambda_0(t_i))^{\exp(z\beta)}$$

$$\lambda(t_i; z) = 1 - (1 - \lambda_0(t_i))^{\exp(z\beta)}$$

A partir d'una mostra de n individus, essent $t_{(1)}, \dots, t_{(d)}$ els temps de supervivència observats pels individus que abandonen i D_i el conjunt d'individus que abandonen en $t_{(i)}$ a la vegada que C_i és el conjunt d'individus censurats en l'interval $[t_{(i)}, t_{(i+1)})$ ($i = 0, \dots, d$) on $t_{(0)} = 0$ i $t_{(d+1)} = \infty$ podem escriure la funció de versemblança del model com:

$$L = \prod_{l=1}^n [f(t; z)]^{\delta_l} [S(t; z)]^{1-\delta_l} = \prod_{l=1}^n [\lambda(t; z)]^{\delta_l} S(t; z)$$

i tenint en compte que, per a cada moment del temps on es produeixen abandonaments, els individus que abandonen contribueixen a la versemblança en un factor igual a $\lambda(t_{(i)}; z) = 1 - (1 - \lambda_0(t_{(i)}; z))^{\exp(z\beta)}$ en el moment $t_{(i)}$ i en un altre igual a $S(t_{(i)}; z) = \prod_{j=1}^{i-1} (1 - \lambda_0(t_{(j)}; z))^{\exp(z\beta)}$ per a tots els moments del temps anteriors a $t_{(i)}$, mentre que els individus censurats contribueixen en un factor $(1 - \lambda_0(t_{(i)}; z))^{\exp(z\beta)}$ només en aquells moments del temps inclosos en l'interval $[t_{(i)}, t_{(i+1)})$, podem reescriure la funció de versemblança com:

$$L = \prod_{i=0}^d \left[\prod_{l \in D_i} (1 - \alpha_i)^{\exp(z\beta)} \prod_{l \in R(t_{(i)}) - D_i} \alpha_i^{\exp(z\beta)} \right]$$

on $\alpha_i = 1 - \lambda_0(t_{(i)})$, $\alpha_0 = 1$ i $R(t_{(i)})$ és el conjunt d'individus en disposició d'abandonar en el moment $t_{(i)}$, el qual incorporarà les observacions censurades corresponents.

Kalbfleisch i Prentice proposen maximitzar el logaritme d'aquesta funció de versemblança respecte $\alpha_1, \dots, \alpha_d$ substituint β pels valors obtinguts en la maximització de la versemblança marginal anterior. Diferenciant doncs respecte α_i s'obindrà l'estimador màxim versemblant de α_i com a solució iterativa de l'equació:

$$\sum_{l \in D_i} \frac{\exp(z_l \beta)}{1 - \hat{\alpha}_i^{\exp(z_l \beta)}} = \sum_{l \in R(t_{(i)})} \exp(z_l \beta)$$

Mitjançant les estimacions màxim versemblants de α_i , es poden obtenir les corresponents expressions màxim versemblants per a les funcions de supervivència i d'atzar bases:

$$\hat{S}_0(t_{(i)}; z) = \prod_{j=1}^{i-1} \hat{\alpha}_j \quad \hat{\lambda}_0(t_{(i)}; z) = 1 - \hat{\alpha}_i$$

Cal notar que ambdues funcions, malgrat haver estimat un model d'atzar proporcional sobre temps continu, són funcions esglaonades vist que només és possible obtenir una estimació de les mateixes en els moments del temps on es produeixen abandonaments. A mesura que el supòsit de temps continu es verifiqui d'una forma més acurada i es disposi d'un nombre d'observacions més elevat, existirà un nombre més gran d'instantos on es produeixen abandonaments de manera que les esmentades funcions base aniran suavitzant-se envers una forma contínua.

De la mateixa forma que en el cas del model de durada accelerada, en la modelització d'atzar proporcional serà convenient la representació gràfica de les funcions de supervivència i d'atzar estimades per a individus arquetípics caracteritzats per una determinada combinació de valors de les variables explicatives, z .

Utilitzant les esmentades representacions es posarà de manifest gràficament el desplaçament cap amunt o cap avall de les funcions d'atzar estimades o, alternativament, els desplaçaments en les funcions de supervivència, en modificar-se els valors de les variables explicatives i segons siguin, d'una banda, els signes dels paràmetres β estimats i, d'altra, els seus valors absoluts.

En el cas dels models d'atzar proporcional, a diferència dels de durada accelerada, en no disposar de la distribució de probabilitat dels temps de supervivència no és possible el càlcul de manera exacta de la durada mediana estimada com a mesura de síntesi de les dades de supervivència analitzades. Una aproximació a la mateixa s'obté utilitzant la tècnica de la interpolació lineal en l'interval de la funció de supervivència estimada que conté a la mediana de la distribució si és que aquesta es troba dins del rang de moments del temps en els quals es produeixen abandonaments. En cas contrari, només podrà afirmar-se que la durada mediana és superior al darrer instant on s'han produït abandonaments i, de la inspecció de la funció de supervivència estimada en el seu conjunt, extreure una idea aproximada de si aquella es troba molt o poc allunyada d'aquest darrer moment del temps analitzat.

5.2.3 Estimació semiparamètrica discreta del model d'atzar proporcional

Novament per a l'estimació semiparamètrica sobre temps discret l'aportació inicial va ser la de Cox (1972) existint, però, una alternativa proposada per Kalbfleisch i Prentice (1973) i desenvolupada per Prentice i Gloeckler (1978). Ambdós mètodes són aproximadament iguals per a valors de la funció d'atzar discreta petits per bé que únicament la segona alternativa pot ser utilitzada en el cas que la modelització discreta vingui com a conseqüència de l'agrupació d'observacions temporals en el model continu.

Prentice i Gloeckler obtenen la versió discreta del model d'atzar proporcional analitzat en l'apartat anterior a partir de l'aplicació de la relació sobre la funció de supervivència en el model, $F(t; z) = F_0(t)^{\exp(z\beta)}$ directament sobre un model discret. Així, si la funció d'atzar corresponent a F_0 té contribució $\lambda_0(t_{(i)})$ en el moment t_i aleshores:

$$F_0(t_i) = \prod_{j=1}^{i-1} (1 - \lambda_0(t_j))$$

i de la relació de l'atzar proporcional anterior:

$$F(t_i; z) = \prod_{j=1}^{i-1} (1 - \lambda_0(t_j))^{\exp(z\beta)}$$

mentre que, com a resultat de la formulació discreta tenim:

$$1 - \lambda_i = S(t_{i+1}) / S(t_i)$$

i, per tant, la funció d'atzar en el moment t_i fixat el vector z d'explicatives serà:

$$\lambda(t_i; z) = 1 - \frac{S(t_{i+1}; z)}{S(t_i; z)} = 1 - \frac{\prod_{j=1}^i (1 - \lambda_0(t_j))^{\exp(z\beta)}}{\prod_{j=1}^{i-1} (1 - \lambda_0(t_j))^{\exp(z\beta)}} = 1 - (1 - \lambda_0(t_i))^{\exp(z\beta)}$$

Habitualment aquesta formulació discreta del model d'atzar proporcional s'utilitzarà quan l'enregistrament del temps de supervivència es fa en intervals, bàsicament per errors de mesura en l'enregistrament del temps, i els mètodes d'estimació de la modelització contínua, tan l'exacte com les aproximacions a aquest, o bé no són accessibles per problemes computacionals degut a un elevat nombre d'observacions censurades o repetides en cada interval, o bé les aproximacions no són prou acurades.

En altres ocasions, però, la mesura del temps és estrictament discreta, els contextes acadèmics, per exemple, acostumen a enregistrar el temps discretament, per la qual cosa serà obligatòria l'estimació del model en aquesta formulació discreta. No és menyspreable la circumstància que porta a estimar un model discret doncs en cas que sigui una agrupament del model continu, únicament serà convenient el mètode d'estimació proposat per Prentice i Gloeckler, mentre que si el temps és estrictament discret podrà ser utilitzat també el mètode inicialment proposat per Cox i desenvolupat per altres autors posteriorment.

5.2.3.1 Mètode de Prentice i Gloeckler (1978)

Siguin (t_i, δ, z) la informació disponible per a cada individu d'una mostra de n , essent cadascun d'ells respectivament el temps de supervivència, l'indicador d'observació no censurada i el vector de variables explicatives. Denotarem per D_i el conjunt d'individus que abandonen en el moment t_i i per R_i el conjunt d'individus censurats en t_i o pels quals s'ha observat que han sobreviscut per sobre de t_i . La divisió del temps continu es realitza mitjançant intervals definits de la forma $[t_{i-1}, t_i)$, $i=1, \dots, d$ amb $t_0 = 0$ i $t_d = \infty$ i on el temps de supervivència inclòs en l'interval s'enregistrarà com t_i .

Amb aquesta notació, la contribució a la versemblança d'un individu que abandona en el moment t_i és la funció de densitat en t_i donats els valors de z :

$$f(t_i; z) = \lambda(t_i; z) \prod_{j=1}^{i-1} (1 - \lambda(t_j; z)) = 1 - (1 - \lambda_0(t_i))^{\exp(z\beta)} \prod_{j=1}^{i-1} (1 - \lambda_0(t_j))^{\exp(z\beta)}$$

mentre que la d'un individu que es troba censurat en el moment t_i , suposant que ha sobreviscut fins al començament de l'interval, és:

$$S(t_i; z) = \prod_{j=1}^{i-1} (1 - \lambda(t_j; z)) = \prod_{j=1}^{i-1} (1 - \lambda_0(t_j))^{\exp(z\beta)}$$

Així, la funció de versemblança dels paràmetres β i α_i , essent aquesta darrera $\alpha_i = 1 - \lambda_0(t_i)$ i t_i el moment del temps en el qual l'individu l abandona o es troba censurat, serà:

$$L = \prod_{l=1}^n \left\{ \left[1 - \alpha_i^{\exp(z_i\beta)} \right]^{\delta_i} \prod_{j=1}^{i-1} \alpha_j^{\exp(z_j\beta)} \right\}$$

la qual reescrita en termes dels $d-1$ moments o intervals de temps en els quals es produeixen abandonaments quedarà:

$$L = \prod_{i=1}^{d-1} \left\{ \prod_{l \in D_i} (1 - \alpha_i^{\exp(z_i\beta)}) \prod_{l \in R_i} \alpha_i^{\exp(z_i\beta)} \right\}$$

Disposant de la funció de versemblança pels paràmetres α_i ($i=1, \dots, d-1$) i β_1, \dots, β_p pot formular-se l'obtenció de les seves estimacions a partir de la maximització de la funció. Posteriorment, i utilitzant els resultats asimptòtics habituals per a les estimacions màxim versemblants, serà possible completar la inferència sobre els paràmetres un cop conegudes les estimacions de les variàncies i covariàncies dels mateixos realitzant els contrastos estadístics de significació individual i conjunta.

Tanmateix, Sprott i Kalbfleisch (1969) mostren com les esmentades aproximacions asimptòtiques dels resultats màxim versemblants no només requereixen d'una mostra suficientment gran sinó que en moltes ocasions necessiten d'una reparametrització convenient per tal que la funció de versemblança sigui aproximadament normal multivariant en la seva forma. En aquest sentit, Kalbfleisch i Gloeckler proposen una reparametrització de la funció de versemblança anterior per a superar la restricció sobre els paràmetres α_i segons la qual $0 < \alpha_i < 1$, ($i=1, \dots, d$) el què pot conduir a que no es verifiquin les condicions òptimes per tal d'aplicar les aproximacions asimptòtiques a les estimacions màxim versemblants. A més, la reparametrització proposada garanteix, o si més no facilita, la convergència dels algorismes d'optimització numèrica, com ara el Newton-Raphson, emprats en l'obtenció dels estimadors màxim versemblants.

Així la transformació proposada és $\gamma_i = \log(-\log \alpha_i)$, de manera que el logaritme de la funció de versemblança anterior queda:

$$\log L = \sum_{i=1}^{d-1} \left\{ \sum_{l \in D_i} \log [1 - \exp\{-\exp(\gamma_i + z_l \beta)\}] - \sum_{l \in R_i} \exp(\gamma_i + z_l \beta) \right\}$$

Les primeres derivades del logaritme de la funció de versemblança respecte els paràmetres, les quals igualades a zero ens permetran el càlcul dels estimadors màxim versemblants, són:

$$\frac{\partial \log L}{\partial \gamma_i} = \sum_{l \in D_i} b_{il} - \sum_{l \in R_i} h_{il} \qquad \frac{\partial \log L}{\partial \beta_j} = \sum_{i=1}^{d-1} \left(\sum_{l \in D_i} z_{jl} b_{il} - \sum_{l \in R_i} z_{jl} h_{il} \right)$$

on $h_{il} = \exp(\gamma_i + z_l \beta)$ i $b_{il} = h_{il} \frac{\exp(-h_{il})}{1 - \exp(-h_{il})}$.

D'altra banda, les segones derivades ens permetran el càlcul de la matriu d'informació observada la qual serà d'utilitat tant pel procediment iteratiu en la recerca del màxim de la funció com, un cop obtingut aquest, per a estimar la matriu de variàncies i covariàncies dels estimadors:

$$\frac{-\partial^2 \log L}{\partial \gamma_i^2} = \sum_{l \in D_i} g_{il} + \sum_{l \in R_i} h_{il}$$

$$\frac{-\partial^2 \log L}{\partial \gamma_i \partial \beta_j} = \sum_{l \in D_i} z_{jl} g_{il} + \sum_{l \in R_i} z_{jl} h_{il}$$

$$\frac{-\partial^2 \log L}{\partial \beta_j \partial \beta_m} = \sum_{i=1}^{d-1} \left(\sum_{l \in D_i} z_{jl} z_{ml} g_{il} + \sum_{l \in R_i} z_{jl} z_{ml} \right)$$

on $g_{il} = b_{il} \frac{[\exp(-h_{il}) + h_{il} - 1]}{1 - \exp(-h_{il})}$.

Per a la maximització del logaritme de la funció de versemblança els autors proposen l'algoritme d'optimització de Newton-Raphson amb uns valors inicials en la primera iteració iguals a $\beta_0 = 0$ i $\gamma_0 = \hat{\gamma}(0)$, essent aquest darrer l'estimador màxim versemblant de γ quan $\beta = 0$ el qual es concreta en:

$$\hat{\gamma}_i(0) = \log \left[-\log \left(1 - \frac{d_i}{n_i} \right) \right]$$

on d_i és el nombre d'individus que abandonen en t_i i n_i el total d'individus en disposició d'abandonar en t_i , és a dir, els que abandonen, els censurats i els que sobreviuen a l'esmentat moment o interval de temps. En aquestes condicions s'ha comprovat en diverses aplicacions com el procediment iteratiu és estable i convergeix sense problemes.

La convergència del procediment iteratiu, per contra, no estarà garantida en el cas que el nombre d'abandonaments en intervals de temps concrets sigui molt reduït, la qual cosa es produirà habitualment quan l'agrupament d'observacions en intervals de temps sigui molt ajustat o quan de la mida de la mostra disponible no s'en derivi un nombre suficient d'observacions repetides en cadascun dels intervals i , per tant, podrà mantenir-se la hipòtesis de temps continu i seran d'aplicació els mètodes d'estimació de l'apartat anterior, bé sia a partir de la versemblança marginal exacta de Kalbfleisch i Prentice o bé mitjançant les aproximacions de Breslow o Efron.

Cal observar com, respecte l'anterior mètode d'estimació en la formulació contínua del temps de supervivència, aquest mètode presenta l'avantatge de que l'estimació de la funció d'atzar base i , en conseqüència, la de la funció de supervivència base discretes, s'obtenen simultàniament amb els paràmetres que mesuren els efectes de les explicatives. Així doncs, per a cada combinació de valors de les explicatives que caracteritzen a grups d'individus concrets, una vegada obtingudes les estimacions màxim versemblants dels paràmetres γ i β , la funció de supervivència estimada serà:

$$\hat{S}(t_i; z) = \prod_{j=1}^{i-1} \exp[-\exp(\hat{\gamma}_j + z\hat{\beta})]$$

mentre que en el cas d'absència de variables explicatives, és a dir, si $\beta = 0$, aleshores l'estimador de la funció de supervivència coincidirà amb el de Kaplan-Meier:

$$\hat{S}(t_i; z) = \prod_{j=1}^{i-1} \left(1 - \frac{d_j}{n_j} \right)$$

5.2.3.2 Mètode de Cox (1972)

Una segona metodologia per a estimar el model d'atzar proporcional discret va ser proposada per Cox (1972) i posteriorment desenvolupada per Brown (1975), Thompson (1977), Laird i Oliver (1981) i Efron (1988). Cox va considerar, vist que les λ_i en el cas discret són probabilitats, la possibilitat d'especificar un model logístic lineal per a l'atzar en cadascun dels moments o intervals de temps en els quals s'enregistren abandonaments. L'expressió del model serà doncs:

$$\frac{\lambda(t_i; z)}{1 - \lambda(t_i; z)} = \frac{\lambda_0(t_i)}{1 - \lambda_0(t_i)} \exp(z\beta)$$

o, alternativament, prenent logaritmes:

$$\log \left[\frac{\lambda(t_i; z)}{1 - \lambda(t_i; z)} \right] = \log \left[\frac{\lambda_0(t_i)}{1 - \lambda_0(t_i)} \right] + z\beta$$

El denominador de l'expressió original del model tendirà cap al valor 1 a mesura que la probabilitat condicional λ tendeixi cap a zero, la qual cosa succeirà en el cas de l'agrupament del temps en intervals molt reduïts i, per tant, quan serà d'aplicació el model d'atzar proporcional continu. Altrament, quan els intervals tinguin una amplada suficient per a que el nombre d'abandonaments enregistrat en cadascun d'ells sigui elevat podrem aplicar aquesta especificació tot i tenint present que només serà una aproximació al model d'atzar proporcional amb observacions del temps continu agrupades de tal forma que l'estimació més adient serà la del mètode proposat per Prentice i Gloeckler. Per contra, si el temps de supervivència és estrictament discret serà de total aplicació la present formulació de Cox. Aquesta alternativa d'estimació del model discret resulta interessant vist que reparametriza la probabilitat condicionada d'abandonament a partir d'una distribució logística, la qual es troba molt estudiada en la literatura i per a l'estimació de la qual existeixen diverses aplicacions informàtiques.

Siguin novament (t_i, δ, z) la informació disponible per a cada individu d'una mostra de n observacions independents, essent respectivament el temps de supervivència, l'indicador d'observació no censurada i el vector de variables explicatives. A diferència del que ha estat considerat en apartats anteriors en aquesta formulació del model estrictament discreta caldrà reformular l'expressió de la funció de supervivència. Així, la contribució a la funció de versemblança d'un individu que abandona en t_i serà:

$$f(t_i; z) = \lambda(t_i; z) \prod_{j=1}^{i-1} (1 - \lambda(t_j; z))$$

mentre que la d'una observació censurada serà, en aquest cas:

$$P(T > t_i) = S(t_i; z) = \prod_{j=1}^i (1 - \lambda(t_j; z))$$

Cal notar, com dèiem, el canvi en la definició de la funció de supervivència, $S(t_i; z) = P(T > t_i)$, respecte a la que ha estat utilitzada fins ara en els apartats anteriors $S(t_i; z) = P(T \geq t_i)$. Aquesta modificació en la definició s'explica per la discrecionalitat del temps de tal forma que ara una observació censurada en t_i estarà indicant que s'ha perdut el seguiment de l'individu a partir de t_i i, per tant, el temps de supervivència serà superior a aquest i , vist que el temps no és continu, obligatòriament haurà de ser estrictament superior a t_i .

En cas d'utilitzar aquesta formulació amb dades temporals contínues agrupades, caldrà modificar la definició dels intervals, els quals seran ara $(t_i, t_{i+1}]$ en comptes de com havien estat definits anteriorment, $[t_i, t_{i+1})$, i si les observacions censurades, per hipòtesi, es produeixen després dels abandonaments hauran d'enregistrar-se en l'interval superior que seria el que inclouria l'instant t_{i+1} .⁴

⁴ Miller (1981) realitza l'anàlisi de dades de supervivència utilitzant la definició $S(t) = P(T > t)$.

En aquestes condicions, la funció de versemblança serà:

$$L = \prod_{l=1}^n \left[\lambda(t_l; z_l) \prod_{j=1}^{i_l-1} (1 - \lambda(t_j; z_l)) \right]^{\delta_l} \left[\prod_{j=1}^{i_l} (1 - \lambda(t_j; z_l)) \right]^{1-\delta_l}$$

essent i_l novament el moment del temps en el qual l'individu l abandona o es troba censurat. Prenent logaritmes i reagrupant:

$$\begin{aligned} \log L &= \sum_{l=1}^n \left[\delta_l \log \lambda(t_l; z_l) + \delta_l \sum_{j=1}^{i_l-1} \log(1 - \lambda(t_j; z_l)) + (1-\delta_l) \sum_{j=1}^{i_l} \log(1 - \lambda(t_j; z_l)) \right] = \\ &= \sum_{l=1}^n \left[\delta_l \log \left(\frac{\lambda(t_l; z_l)}{1 - \lambda(t_l; z_l)} \right) + \sum_{j=1}^{i_l} \log(1 - \lambda(t_j; z_l)) \right] \end{aligned}$$

Seguint a Singer i Willet (1993), introduïm una variable indicadora de la situació de cada individu en tots els moments del temps, y_{li} ($i=1, \dots, d$), tal que prengui el valor 1 si l'individu l abandona en el moment i i el valor 0 en qualsevol altre cas, és a dir, si l'individu l abandona més tard o bé es troba censurat, tindrem que:

$$\sum_{j=1}^{i_l} y_{lj} \log \left(\frac{\lambda(t_j; z_l)}{1 - \lambda(t_j; z_l)} \right) = \begin{cases} \log \left(\frac{\lambda(t_l; z_l)}{1 - \lambda(t_l; z_l)} \right) & \text{si } \delta_l = 1 \\ 0 & \text{si } \delta_l = 0 \end{cases} = \delta_l \log \left(\frac{\lambda(t_l; z_l)}{1 - \lambda(t_l; z_l)} \right)$$

i el logaritme de la versemblança serà:

$$\log L = \sum_{l=1}^n \left[\sum_{j=1}^{i_l} y_{lj} \log \left(\frac{\lambda(t_j; z_l)}{1 - \lambda(t_j; z_l)} \right) + \sum_{j=1}^{i_l} \log(1 - \lambda(t_j; z_l)) \right]$$

Desfent la transformació logarítmica i reagrupant tenim:

$$L = \prod_{i=1}^n \prod_{j=1}^{i_j} [\lambda(t_j; z_j)]^{y_{ij}} [1 - \lambda(t_j; z_j)]^{(1-y_{ij})}$$

Aquesta expressió de la funció de versemblança pel model d'atzar proporcional discret és equivalent a la que podríem obtenir per a una seqüència de $N=(i_1 + i_2 + \dots + i_n)$ experiments de Bernoulli amb paràmetres $\lambda(t_i; z_i)$. Brown (1975), Laird i Oliver (1981) i Allison (1982) demostren com efectivament ambdues funcions de versemblança són equivalents la qual cosa permet tractar els N valors dicotòmics y_{ij} observats com un conjunt de variables aleatòries independents i efectuant una hipòtesi respecte la parametrització sobre les probabilitats dels dos valors possibles (0 i 1), obtenir estimadors màxim versemblants dels paràmetres del model. Així, Cox i els altres autors esmentats anteriorment proposen la parametrització logística i, per tant, s'estimarà un model logit a partir de les N observacions dicotòmiques. Tot i això, és possible també especificar la distribució normal i ens trobarem davant d'un model Probit o, alternativament, utilitzar la transformació loglogarítmica complementaria sobre la probabilitat condicional d'abandonar, $\log\{-\log[1 - \lambda(t_i; z_i)]\}$. Aquesta darrera opció és la que van proposar implícitament Prentice i Gloeckler en la formulació del seu mètode d'estimació per al model discret.

Així doncs, tal com assenyalen Willet i Singer (1993), utilitzant els habituals mètodes d'estimació del model Logit o regressió logística, es podran obtenir les estimacions màxim versemblants dels paràmetres del model d'atzar proporcional discret, tant en la formulació estrictament logística com en les alternatives Probit i loglogarítmica complementaria, vist que les aplicacions informàtiques més habituals incorporen la possibilitat de seleccionar qualsevol de les tres parametritzacions alternatives.

Cal però concretar aquesta equivalència en el procediment amb l'objectiu de mostrar clarament les diferències latents d'una o altra parametrització. El procediment dicotòmic tracta d'obtenir, mitjançant una parametrització convenient, una relació lineal entre la probabilitat de que s'esdevingui un succés i una combinació de variables explicatives. Les dues parametritzacions considerades són la Logística (Logit), $\log\{ \lambda(t;z) / (1-\lambda(t;z)) \}$, i la Loglogarítmica complementaria, $\log\{-\log(1-\lambda(t;z))\}$, de manera que la relació entre les probabilitats i les variables explicatives serà per a la logística:

$$\frac{\lambda(t;z)}{1 - \lambda(t;z)} = \frac{\lambda_0(t)}{1 - \lambda_0(t)} \exp(z\beta)$$

$$\log\left(\frac{\lambda(t;z)}{1 - \lambda(t;z)}\right) = \log\left(\frac{\lambda_0(t)}{1 - \lambda_0(t)}\right) + z\beta = \gamma_t + z\beta$$

i per a la loglogarítmica complementaria:

$$\lambda(t;z) = 1 - (1 - \lambda_0(t))^{\exp(z\beta)}$$

$$\log(1 - \lambda(t;z)) = \log[(1 - \lambda_0(t))^{\exp(z\beta)}] = \exp(z\beta) \log(1 - \lambda_0(t))$$

$$\log\{-\log(1-\lambda(t;z))\} = \log\{-\log(1-\lambda_0(t))\} + z\beta = \gamma_t + z\beta$$

posant-se de manifest la clara diferència per als dos mètodes alternatius en l'estructura de la relació entre les variables explicatives i la probabilitat condicional, i, a més, que únicament la loglogarítmica complementaria correspon a la formulació discreta del model d'atzar proporcional.

5.3 Variables explicatives canviants en el temps

La presència de variables explicatives el valor de les quals és canviant en el temps ha estat desenvolupat abastament en el si del model d'atzar proporcional des de la seva primera formulació de Cox (1972,1975). La inclusió d'aquestes variables en el model complica lleugerament l'estimació especialment en el cas del model continu. Així, quan es presentava la funció de versemblança per a l'estimació semiparamètrica del model d'atzar proporcional continu es van esmentar les dues aproximacions a la mateixa, per una banda la de la versemblança marginal de Kalbfleisch i Prentice i, per altra, la de la versemblança parcial de Cox. En el cas de les variables explicatives depenents del temps, la interpretació de la funció de versemblança en termes de la versemblança marginal dels rangs ja no és possible i, en conseqüència, només resulta adient el raonament en termes de la versemblança parcial.

Considerarem, inicialment el model sense observacions repetides, és a dir, únicament amb observacions censurades. Així, disposant d'una mostra de n individus independents dels quals es coneix el temps de supervivència, t_i , i si es tracta d'una observació censurada o no. Dels n individus, d han abandonat i la resta són observacions censurades. Els temps de supervivència observats s'ordenen de forma creixent, $t_{(1)} < t_{(2)} < \dots < t_{(d)}$ i pels q_i individus que es troben censurats en l'interval $[t_{(i)}, t_{(i+1)})$ $i=0, 1, \dots, d$, i on $t_{(0)} = 0$ i $t_{(d+1)} = \infty$, suposem que la censuració es produeix després dels abandonaments de cada interval, de tal forma que aquestes observacions censurades formaran part del conjunt de risc $R(t_{(i)})$, és a dir, aquell que inclou els individus en disposició d'abandonar en el moment $t_{(i)}$. En aquestes condicions, és possible especificar la versemblança parcial tenint en compte que en cada moment del temps on es produeixen abandonaments cal considerar la informació fins aleshores de les variables canviants en el temps.

La versemblança parcial serà doncs:

$$L(\beta) = \prod_{i=1}^d \left(\frac{\exp[z_{(i)}(t_{(i)}) \beta]}{\sum_{l \in R(t_{(i)})} \exp[z_l(t_{(i)}) \beta]} \right)$$

la qual és equivalent a l'obtinguda pel model d'atzar proporcional continu amb totes les explicatives constants, amb una única diferència en la notació utilitzada per a les explicatives les quals ara dependran del moment del temps en que es mesurin, $z_l(t)$.

La inclusió d'observacions repetides del temps de supervivència degudes a errors de mesura en el seu enregistrament comporta una funció de versemblança lleugerament diferent a la deduïda per Kalbfleisch i Prentice en el model amb variables explicatives constants. Quan les variables explicatives són canviants en el temps, la deducció de la versemblança a partir de la distribució marginal dels rangs no és possible. Aquesta versemblança es deduïa a partir del resultat de què una transformació monòtona sobre el temps de supervivència actua transitivament sobre els valors del temps i deixa inalterats el rangs (la ordenació) associats a cadascuna de les observacions, mentre que ara, en presència de variables canviants en el temps, si es produeix una transformació monòtona sobre el temps de supervivència canviarien els valors associats de les explicatives corresponents i s'hauria modificat la informació sobre la qual estimar el model.

Cox (1972) va proposar que, en presència d'observacions repetides, la versemblança parcial es pot calcular mitjançant l'aplicació d'un model logístic discret en el càlcul de la probabilitat d'observar efectivament els d_i individus que abandonen donats el conjunt d'individus que formen el conjunt de risc en el moment $t_{(i)}$.

La versemblança parcial en presència d'observacions repetides serà:

$$L(\beta) = \prod_{i=1}^d \left(\frac{\exp[s_i(t_{(i)})\beta]}{\sum_{l \in R_d(t_{(i)})} \exp[s_l(t_{(i)})\beta]} \right)$$

on $s_i(t_{(i)})$ és la suma de les variables explicatives $z_l(t_{(i)})$ per a tots els individus que abandonen en $t_{(i)}$, $s_l(t_{(i)}) = \sum_{j=1}^{d_i} z_l(t_{(i)})$, $l=(1, \dots, l_d)$ i $R_d(t_{(i)})$ és el conjunt de tots els subconjunts de d_i elements que es podrien formar amb els individus del conjunt de risc $R(t_{(i)})$ sense reemplaçament. Aquesta funció de versemblança és computacionalment difícil per la qual cosa pot ser aproximada utilitzant en el denominador de l'expressió les propostes de Kalbfleisch i Prentice, Breslow o Efron, estudiades anteriorment en l'estimació amb variables explicatives constants.

En l'estimació del model d'atzar proporcional discret per qualsevol de les dues metodologies exposades anteriorment, la inclusió de variables explicatives canviants en el temps no presenta més problemes que els derivats de la complexitat en la construcció de les bases de dades que contenen la informació disponible per a l'estimació del model. Ambdues metodologies, la de Prentice i Gloeckler i la proposada per Cox i posteriorment desenvolupada per altres autors, tenen en comú la seva equivalència a una successió de N variables dicotòmiques, corresponents a l'estat de l'individu (abandonament o censuració) en cadascun dels moments del temps o intervals en els quals es produeixen abandonaments. Així, utilitzant la metodologia de la regressió logística amb la parametrització corresponent a cada cas (logit i loglogarítmica complementaria), les variables explicatives canviants en el temps podran ser introduïdes sense més que pensar en què cal tenir en compte la variable dicotòmica en cada moment del temps i , per tant, podrà distingir-se entre el valor que pren una explicativa en un instant o en un altre.

Una vegada estimats els paràmetres del model d'atzar proporcional, sia en la hipòtesi de temps continu sia en la de temps discret, i en aquest darrer cas per qualsevol dels dos mètodes alternatius, caldrà analitzar la interpretació dels paràmetres que acompanyen a les variables explicatives canviants en el temps.

En aquest sentit cal dir que, malgrat que les variables considerades puguin prendre valors diferents per a cada moment del temps, el seu efecte sobre la funció d'atzar serà constant i vindrà mesurat pel paràmetre corresponent. És per això que les variables canviants en el temps són habitualment utilitzades per a verificar la hipòtesi de proporcionalitat en el model. Aquesta hipòtesi estableix que els efectes sobre l'atzar d'una modificació en els valors d'una variable explicativa hauran de ser independents del moment del temps en que es mesurin, de manera que introduint interaccions entre el temps i qualsevol de les variables explicatives, constants o no, es podrà verificar la hipòtesis de proporcionalitat sense més que contrastar la significació del paràmetre que acompanya la interacció. Si aquest paràmetre és significatiu estadísticament ens trobarem davant l'eventualitat d'una violació de la hipòtesi de proporcionalitat vist que l'efecte sobre l'atzar de variacions en les explicatives serà diferent en cada moment del temps considerat. Contràriament, si el paràmetre de la interacció no és significatiu estadísticament, aleshores es verificarà la hipòtesi i, en el cas de tractar-se d'una variable explicativa el valor de la qual sigui susceptible de variar en el temps, l'efecte de la mateixa sobre la funció d'atzar es mantindrà constant per a tots els moments del temps independentment del valor que hagi pres la variable. Gràficament, la hipòtesi de proporcionalitat estableix que les funcions d'atzar integrat corresponents a dos valors diferents d'una mateixa variable explicativa hauran d'evolucionar de forma paral·lela i, precisament la distancia entre una i altra vindrà determinada pel paràmetre corresponent a l'explicativa.

La contrastació de la hipòtesis de proporcionalitat en el model d'atzar és de vital importància vist que la interpretació dels paràmetres depèn directament de la mateixa i, a més, la seva violació pot donar lloc a conclusions errònies sobre la forma de les funcions d'atzar i supervivència, sobre les durades medianes estimades per interpolació lineal i, en general, sobre els resultats de la inferència estadística del model.

D'altra banda, i tal com s'ha realitzat en la presentació general dels mètodes d'estimació del model d'atzar proporcional continu i discret, una vegada estimats els paràmetres β que mesuren els efectes de les explicatives sobre la funció d'atzar resulta d'utilitat l'estimació de les funcions d'atzar i de supervivència bases, les quals serveixen de referència per a mesurar els efectes de les variables explicatives sobre el temps de supervivència ja que representen les condicions estàndard dels individus analitzats. En presència de variables explicatives canviants en el temps, però, mentre el càlcul de les esmentades funcions base no suposa cap complexitat afegida al procés anteriorment descrit per a les variables explicatives constants, ara la interpretació de les funcions base no és gens simple. Aquesta complexitat es deriva de la pròpia condició de les variables explicatives canviants en el temps, és a dir, si les funcions d'atzar i de supervivència bases representen les condicions estàndard dels individus analitzats resulta molt difícil, si existeixen variables canviants en el temps, fixar les esmentades condicions estàndard ja que la combinació de valors possibles per a l'explicativa corresponent pot ser molt ampla. Així, resultarà d'interès l'anàlisi de dos casos extrems en l'evolució dels valors de la variable explicativa, els quals serviran de límits en torn dels quals fluctuaran la resta de possibles combinacions en els valors de la variable.

5.4 Múltiples causes de finalització de la durada

Fins el moment han estat estudiats mètodes per a l'anàlisi de dades de supervivència amb censuració però tenint en compte un únic tipus d'esdeveniment en la finalització de la durada. Amb tot, en moltes situacions empíriques poden existir més d'una causa per a la finalització de la durada; per exemple, en l'anàlisi de la mortalitat humana les causes de la mort poden ser múltiples i, en l'aplicació que ens ocupa, la finalització dels estudis pot venir motivada per un abandonament dels mateixos, per un trasllat, per l'obtenció del títol corresponent, etc. Cal doncs incorporar en les modelitzacions anteriors la possibilitat de múltiples causes de finalització de la durada.

La problemàtica de les múltiples causes de finalització de les durades ("multiple risks" en la literatura) fa referència a tres qüestions diferenciades. La primera té a veure amb l'estimació de les relacions entre les variables explicatives i la funció d'atzar o de supervivència per a una causa de finalització específica. Aquesta qüestió ha estat la més desenvolupada i el seu objectiu és garantir la validesa dels resultats de l'anàlisi de supervivència quan només es considera una causa específica de finalització.

La segona qüestió s'enmarca en l'estudi de les interrelacions entre diferents causes de finalització de la durada per a unes condicions prefixades. En aquest cas es tracta de verificar si existeixen relacions entre les causes de finalització en el sentit de si són, per exemple, proporcionals els seus atzars específics o bé si les condicions que alteren el risc de que s'esdevingui la finalització per una causa determinada, per exemple, un tractament mèdic per a guarir una malaltia, a la vegada incrementen o disminueixen el risc d'una altra causa, modifiquen la probabilitat condicional de morir pels efectes d'una malaltia diferent.

Finalment, l'estimació de les funcions d'atzar per a una causa específica de finalització havent "eliminat" les altres causes. Aquesta darrera qüestió està relacionada no tant amb aspectes d'inferència estadística com de la pròpia definició del problema a estudiar la qual comporti la possibilitat d'"eliminar" alguna de les causes específiques de finalització, per exemple, en l'anàlisi de la mortalitat humana tindria a veure amb la creació de les condicions exògenes necessàries per a l'erradicació d'una causa de mort i els seus efectes sobre la resta de causes de mortalitat.

Davant de la possible presència de múltiples causes de finalització de les durades és convenient tornar a explicitar els conceptes bàsics de l'anàlisi de supervivència amb la incorporació d'aquelles. Així, suposarem que el temps de supervivència, T , és continu, el qual pot estar censurat i, pel moment, que les variables explicatives, z , prenen valors constants al llarg del temps. A més, quan es produeix la finalització d'una durada pot ser deguda a qualsevol de les m causes diferents possibles, $J = \{1, 2, \dots, m\}$. La taxa instantània de finalització de la durada o funció d'atzar per a totes les causes serà:

$$\lambda(t; z) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t, z)}{\Delta t}$$

De forma equivalent podem definir la funció d'atzar per a una causa específica, j , com:

$$\lambda_j(t; z) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, J=j \mid T \geq t, z)}{\Delta t}$$

per a $j = 1, 2, \dots, m$. Aquesta expressió representa la taxa instantània de finalització de la durada per la causa j en el moment t donades les variables explicatives z i en presència de la resta de causes possibles de finalització.

En el cas que j sigui una única causa possible d'entre el conjunt de m tindrem que:

$$\lambda(t; z) = \sum_{j=1}^m \lambda_j(t; z)$$

i la funció de supervivència serà:

$$S(t; z) = \exp \left[- \int_0^t \lambda(u; z) du \right]$$

mentre que la de densitat per a cada causa específica és:

$$f_j(t; z) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, J=j | z)}{\Delta t} = \lambda_j(t; z) S(t; z) \quad j=1, 2, \dots, m$$

Amb aquestes definicions serà possible explicitar la funció de versemblança en termes de les funcions d'atzar per a cadascuna de les causes específiques. Així, disposant d'una mostra de n individus independents per a cadascun dels quals es coneix $(t_i, \delta_i, j_i, z_i)$, essent t_i el temps de supervivència observat, δ_i l'indicador d'observació censurada, j_i la causa específica de finalització de la durada (la qual no formarà part de la versemblança en el cas que $\delta_i=0$) i z_i el vector de variables explicatives de l'individu, podrem escriure la funció de versemblança com proporcional a:

$$\prod_{i=1}^n \left\{ [\lambda_{j_i}(t_i; z_i)]^{\delta_i} S(t_i; z_i) \right\} = \prod_{i=1}^n \left\{ [\lambda_{j_i}(t_i; z_i)]^{\delta_i} \prod_{j=1}^m \exp \left[- \int_0^{t_i} \lambda_j(u; z_i) du \right] \right\}$$

i reagrupant els termes convenientment la funció de versemblança serà igual al producte dels factors corresponents a cadascuna de les m causes específiques de finalització de la durada. A més, el factor de la versemblança que correspondria a una causa específica j qualsevol serà aquell que obtindriem considerant la resta de causes específiques diferents a la j com a observacions censurades en t_i .

En conseqüència els mètodes estudiats en els apartats anteriors seran vàlids per la inferència sobre les funcions d'atzar específiques sense més que considerar com a observacions censurades totes les causes excepte l'analitzada. Així, per exemple podrem especificar el model d'atzar proporcional continu per a una causa específica com:

$$\lambda_j(t; z) = \lambda_{0j}(t) \exp[z(t) \beta_j] \quad j=1, \dots, m$$

destacant que tant les funcions d'atzar base, λ_{0j} , com els paràmetres associats a cadascuna de les variables explicatives, β_j , poden prendre valors diferents per a cadascuna de les m causes específiques de finalització de la durada.

En el cas que les variables explicatives siguin canviants en el temps també es verifiquen els resultats anteriors i, per tant, es podran utilitzar els mètodes analitzats pel model d'atzar proporcional continu o discret, en cadascuna de les m causes de finalització per separat considerant la resta de causes com a observacions censurades.

Amb els resultats anteriors quedaria resolta la primera de les qüestions plantejades al voltant de la problemàtica de les múltiples causes de finalització de les durades. En aquest sentit, l'aplicació que es desenvoluparà posteriorment inclourà dues causes de finalització de les durades dels estudis com són, d'una banda, l'abandonament i, de l'altra, l'obtenció del títol de llicenciat. Realitzant una anàlisi individual per a cadascuna de les causes de finalització considerant la resta com a observacions censurades s'efectuarà correctament la inferència sobre els paràmetres que relacionen les variables explicatives amb la probabilitat condicional d'abandonar/licenciar-se i, en el seu cas, sobre les funcions d'atzar base discretes.

Segona Part:
Aplicació

6. La base de dades

Capítol 6

La base de dades

La segona part del treball correspon a l'aplicació dels mètodes exposats en la primera part a una cohort d'estudiants en concret. Els objectius d'aquesta aplicació són diversos. En primer lloc, es realitzarà una descripció de la base de dades, de les etapes necessàries per a la seva construcció i de les fonts d'informació. Posteriorment, i mitjançant alguna de les metodologies anomenades estàtiques que han estat utilitzades en l'anàlisi de l'abandonament i l'èxit en els estudis universitaris, tractarem d'evidenciar les mancances que presenten aquestes tècniques "clàssiques" i la necessitat d'utilitzar la metodologia de l'anàlisi de dades de supervivència. Un cop justificat l'ús de l'anàlisi del temps de supervivència a la durada dels estudis es presentaran els resultats de l'aplicació dels mètodes no paramètrics i paramètrics a les durades amb una doble intenció. D'una banda, la de conèixer els factors que determinen les causes de la finalització dels estudis, l'èxit o el fracàs, i de l'altra, analitzar críticament la adequació de la normativa de permanència als objectius que té plantejats, l'avançament en el temps dels abandonaments que fatalment s'han de produir i la col·laboració a la millora de les condicions per tal d'augmentar el rendiment acadèmic dels estudiants via la reducció de la massificació i augmentant la motivació envers l'estudi.

En el present capítol detallem l'estructura i composició de la base de dades utilitzada en la part empírica del treball així com les etapes seguides en la seva construcció i les fonts d'informació. A més, presentem una anàlisi descriptiva prèvia de les variables utilitzades un cop efectuada una tipologia de les mateixes i un repàs de les conseqüències que tindrà el disseny de la base de dades sobre la metodologia estadística emprada.

6.1 Estructura i fonts d'informació

L'objectiu del treball és estudiar i, si és possible, modelitzar la relació entre la durada dels estudis i les components acadèmiques i personals dels estudiants. Per tant la base de dades ha d'incloure aquests dos tipus de variables: les acadèmiques i les personals. Això ens porta a la necessitat d'accedir a dues fonts d'informació paral·leles: d'una banda els expedients acadèmics dels estudiants i d'altra, les dades estadístiques i personals que proporcionen els impresos de matrícula.

El projecte MEGA de Mecanització de la Gestió Acadèmica posat en marxa per la Universitat de Barcelona l'any 1988 ha permès disposar de la base de dades dels expedients acadèmics informatitzada. Malgrat tot, aquesta informació no és completa. Donat que, lògicament, el procés d'introducció dels expedients acadèmics existents en suport magnètic resultava molt costós, es va restringir als expedients "vius" en el curs 1988-89, és a dir, només els d'aquells estudiants que van matricular l'esmentat curs acadèmic. Des del curs 88-89 fins a l'actualitat es disposa de tots els expedients acadèmics, essent la seva gestió informatitzada cada cop més completa i ràpida.

Aquesta base de dades d'expedients acadèmics ens permet conèixer l'any en què l'alumne va iniciar els estudis, les assignatures que ha anat matriculant cada curs acadèmic i el seu rendiment acadèmic concretat en les qualificacions obtingudes. D'altra banda, la recenta aprovació i posada en marxa dels nous plans d'estudis dels ensenyaments possibilita, a partir de les taules d'adaptacions del Pla d'Estudis 1980 als nous plans aprovades pels òrgans competents, la traducció dels expedients dels estudiants de la cohort a la nova situació dels crèdits, la qual cosa facilitarà la introducció en l'anàlisi de la simulació de la posada en marxa de la normativa de permanència aprovada pel Consell Social de la Universitat de Barcelona.

Paral·lelament als expedients acadèmics hem de disposar d'informació de les característiques personals dels estudiants. Part d'aquesta informació la tenim recollida en els impresos de matrícula que emplena l'alumne en iniciar el curs acadèmic. Actualment l'esmentada informació queda enregistrada juntament amb l'expedient acadèmic dins del marc del projecte MEGA. Anys enrera, però, les dades provinents de la matrícula s'enregistraven en una cinta magnètica separada que quedava arxivada un cop finalitzat el curs corresponent.

Aquesta base de dades obtinguda a través dels impresos de matrícula ens dóna a conèixer les característiques personals i socio-econòmiques dels alumnes: edat, sexe, nacionalitat, domicili familiar, situació laboral, etc.

Al costat de les dades personals dels alumnes provinents de la matrícula, per alguns cursos acadèmics concrets existeixen altres arxius que contenen dades estadístiques de l'ingrés de l'estudiant a la universitat com ara els resultats de les Proves d'Accés (PAAU), el centre on cursà l'ensenyament secundari, etc.

6.2 Etapes en la construcció de la base de dades

6.2.1 Mètodes d'elecció de la població a estudiar

Un cop delimitades les fonts de la informació disponibles i les seves restriccions cal decidir sobre la població que serà objecte d'anàlisi. En aquest sentit, existeixen dues alternatives d'elecció: la selecció i seguiment d'una cohort d'estudiants o bé l'estudi i seguiment retrospectiu de tots els alumnes que cursen estudis en un moment del temps determinat, per exemple l'actual curs acadèmic. Totes dues opcions presenten avantatges i inconvenients tant des del punt de vista metodològic com del de la disponibilitat d'informació completa essent, però, molt més adient l'aproximació per les cohorts.

L'elecció d'una cohort, és a dir, del conjunt d'estudiants que van iniciar els estudis en un mateix curs acadèmic, és l'opció més recomanada des del punt de vista metodològic donat que permet fer el seguiment d'un col·lectiu que es troba en les mateixes condicions inicials i que s'enfronta a les mateixes circumstàncies acadèmiques, socials, econòmiques, etc⁵. Aquest fet ens permetrà controlar, en certa manera, la immensa heterogeneïtat existent en el conjunt dels estudiants universitaris que no queda suficientment recollida per les variables de què disposem, la qual cosa facilitarà l'anàlisi de la relació entre la durada dels estudis i el rendiment acadèmic.

L'inconvenient que presenta l'elecció d'una cohort està en haver de trobar el punt d'equilibri entre:

⁵ Terenzini (1982), Hogan (1984).

1.- la necessitat que la durada dels estudis tingui una variabilitat suficient per tal que sigui possible realitzar una anàlisi estadística que capti abastament les diferències existents en les trajectòries acadèmiques dels alumnes, el què ens portaria a seleccionar una cohort allunyada en el temps, i

2.- la disponibilitat de la informació acadèmica i personal dels estudiants seleccionats que, com hem esmentat anteriorment, està restringida pel fet que la informatització dels expedients no es va dur a terme fins el curs 1988-89 i només per aquells estudiants que van matricular aquell any.

L'elecció del conjunt d'estudiants que cursen estudis en un moment determinat del temps tindria l'avantatge, en contraposició a la cohort, de recollir una variabilitat més gran en la durada dels estudis però presenta com a inconvenient que incorporaria una heterogeneïtat en les condicions davant de les què s'enfronten els alumnes que no podrien ser suficientment recollides per les variables de què es disposa i que farien difícil l'anàlisi de les durades dels estudis en relació a les característiques personals i acadèmiques.

A més, aquesta segona opció, si no anés acompanyada de altres mesures en el disseny de la base de dades, portaria a la introducció d'un biaix donat que els estudiants amb un rendiment acadèmic inferior estarien sobrerrepresentats entre el col·lectiu dels que tenen una durada més llarga. Un factor de correcció del biaix estaria en seleccionar dos moments del temps suficientment allunyats i estudiar la trajectòria acadèmica dels estudiants compresos dins del període, el qual s'enfrontaria també a la restricció en la disponibilitat d'informació acadèmica suficientment allunyada en el temps.

6.2.2. La Cohort del curs 1985-86

A la vista de les anteriors consideracions al voltant dels mètodes d'elecció de la població a estudiar, hem considerat més oportuna la selecció d'una cohort. Un cop presa aquesta decisió caldrà fixar quina cohort, és a dir, quin curs acadèmic prendrem com a referència per tal de trobar l'equilibri entre disponibilitat d'informació i variabilitat suficient de la durada dels estudis, del que parlàvem en l'apartat anterior.

Si el més important és la disponibilitat d'informació ens veurem en la obligació de seleccionar la cohort del curs acadèmic 1988-89 que és l'any a partir del qual es realitza la informatització dels expedients acadèmics i, per tant, pels estudiants de la cohort disposaríem de tota la informació acadèmica. Aquesta cohort, però, tindria únicament 5 anys d'història dels quals només pels quatre primers disposaríem efectivament d'informació referent a rendiment acadèmic donat que el darrer any, el curs 1992-93 actualment encara no ha finalitzat. La selecció d'aquesta cohort, malgrat que disposant de tota la informació, seria massa curta fins el punt que cap estudiant no hauria finalitzat els seus estudis (excepte en els casos de les Diplomatures).

Si donem més importància a la variabilitat de la durada dels estudis de tal manera que, per exemple, sigui possible que una part de la cohort hagi finalitzat els seus estudis, ens veurem abocats a seleccionar un curs acadèmic de referència més allunyat en el temps amb la consegüent pèrdua en la disponibilitat d'informació. Així, seleccionant la cohort del curs 1985-86 tindrem 8 anys d'història, dels quals set estarien finalitzats (serà possible trobar llicenciats), però pels cursos acadèmics 1985-86, 1986-87 i 1987-88 tindrem manca d'informació.

Seleccionada la cohort del curs 1985-86 ens trobem que, per aquells individus que van abandonar els estudis abans del curs 1988-89, no disposem d'informació sobre els expedients acadèmics. Aquesta manca d'informació provoca un desconeixement de les circumstàncies personals i acadèmiques dels estudiants que abandonen els estudis en els primers anys, la qual cosa faria perdre validesa a qualsevol anàlisi sobre l'abandonament a realitzar donada la gran importància tan qualitativa com quantitativa de l'abandonament en els primers anys. En aquests primers cursos és quan es produeixen un nombre d'abandonaments més gran i, per tant, qualsevol estudi de l'abandonament que no disposi de l'esmentada informació donarà lloc a uns resultats incomplets i esbiaixats.

Amb la finalitat de resoldre aquesta manca d'informació pels tres primers cursos acadèmics de la cohort en aquells casos en què hagin estat abandonats els estudis, ens veiem obligats a imposar una restricció sobre la població a estudiar. L'objectiu fixat inicialment era, un cop seleccionada la cohort, analitzar la durada dels estudis de tots els estudiants de la Universitat de Barcelona. Tanmateix però, la única forma de completar la informació dels expedients acadèmics per a tots els individus de la cohort és anar directament al suport físic dels mateixos en les secretaries dels centres i atès l'elevat nombre d'estudiants que es troben en l'esmentada situació d'abandonament del estudis en els primers anys, ens veiem abocats a centrar-nos únicament en la Llicenciatura en Ciències Econòmiques i Empresarials.

Així doncs, la població a estudiar quedarà constituïda pels 1351 estudiants que van iniciar els estudis de la Llicenciatura en Ciències Econòmiques i Empresarials el curs 1985-86, per la qual disposàvem, en un primer moment, de les dades de matrícula aquell any i dels expedients acadèmics dels estudiants que

van matricular-se el curs 1988-89. Mitjançant les dades de matrícula es van poder identificar (amb nom i cognoms, i DNI) els estudiants que van causar baixa en el centre al llarg dels cursos 85-86, 86-87 i 87-88. Amb aquesta informació va ser possible localitzar físicament els expedients en els arxius de la secretaria de la Facultat d'Econòmiques i completar la base de dades. Finalment, i per tal d'arrodonir al màxim la informació, van ser també localitzats aquells estudiants pels quals no es disposava en els arxius informàtics dels resultats de les Proves d'Accés a la Universitat. Donat que aquesta dada també queda enregistrada en l'expedient físic de l'alumne, va ser possible completar al màxim la informació referida a la cohort.

6.3 Descripció de les variables

Com s'ha dit anteriorment, la base de dades per a la cohort del curs acadèmic 1985-86, ha estat construïda a partir de la unió de tres fonts d'informació diferents com són, d'una banda, els expedients acadèmics dels estudiants, les dades recollides en els impresos de matrícula i els arxius de dades estadístiques sobre l'ingrés a la universitat. Tota aquesta informació es concreta en els següents blocs:

1.- **Variables personals:** En aquest bloc es situen aquelles variables que tenen a veure amb les característiques pròpies de l'individu com són el SEXE, l'EDAT, l'ESTAT CIVIL, la NACIONALITAT, el DOMICILI, el TIPUS DE RESIDÈNCIA i la SITUACIÓ LABORAL.

2.- **Variables de gestió acadèmica:** Estan incloses en aquest bloc aquelles variables relacionades amb la situació d'entorn acadèmic de l'estudiant, com són la DURADA DELS ESTUDIS, la SITUACIÓ ACADÈMICA DE L'ALUMNE, la FONT DE FINANÇAMENT DELS ESTUDIS, la CLASSE DE MATRÍCULA, el CENTRE i l'HORARI DE CLASSES.

3.- **Variables de rendiment acadèmic:** Aquest bloc el formen el conjunt de variables que recullen el nombre d'assignatures matriculades (MAT85...MAT92), el nombre d'assignatures superades (APR85...APR92) i el nombre d'assignatures no presentades (P85...P92) cada curs acadèmic.

4.- **Variables acadèmiques prèvies:** En aquest bloc s'inclouen aquelles variables que tenen a veure amb l'ingrés de l'alumne a la universitat com són el TÍTOL D'INGRÉS i el RESULTAT DE LES PROVES D'ACCÉS (PAAU).

A continuació es realitza una anàlisi descriptiva més detallada de cadascuna de les variables amb l'objectiu de caracteritzar el col·lectiu d'estudiants que formen la cohort.

6.3.1 Variables Personals

6.3.1.1 La variable SEXE

La variable SEXE presenta la següent distribució de freqüències que mostra com el nombre d'Homes és molt superior al de Dones.

SEXE	Freq. Absoluta	Freq. Relativa
Dones	496	36.7
Homes	855	63.3

6.3.1.2. La Variable EDAT

Aquesta variable recull l'edat dels estudiants en el moment de realitzar la matrícula el curs 1985-86. Les següents mesures descriptives de la variable EDAT ens mostren la concentració existent al voltant dels 18 anys. Així, un 75% dels estudiants tenen 19 o menys anys mentre que només el 1% està per sobre de la frontera dels 30 anys. El primer dels col·lectius, el més jove, és el provinent majoritàriament del COU, mentre que existeix un petit grup de persones amb una edat superior que provenen d'altres estudis o de l'accés a majors de 25 anys.

Com anirà posant-se de manifest al llarg de les pàgines següents, aquesta dicotomia entre els dos col·lectius és molt acusada i serà rellevant en les anàlisis posteriors.

EDAT

Moments

N	1351
Mitjana	18.62842
Variança	4.778866
Desv Std	2.186062

Quantiles(Def=5)

100% Max	50	99%	29
75% Q3	19	95%	21
50% Med	18	90%	20
25% Q1	18	10%	17
0% Min	17	5%	17
		1%	17

Taula de Freqüències

EDAT	Freq. Absoluta	Freq. Relativa	Freq. Abs.Acum.	Freq. Rel.Acum.
17	160	11.8	160	11.8
18	783	58.0	943	69.8
19	239	17.7	1182	87.5
20	78	5.8	1260	93.3
21	29	2.1	1289	95.4
22	10	0.7	1299	96.2
23	14	1.0	1313	97.2
24	4	0.3	1317	97.5
25	5	0.4	1322	97.9
26	6	0.4	1328	98.3
27	3	0.2	1331	98.5
28	3	0.2	1334	98.7
29	4	0.3	1338	99.0
30	2	0.1	1340	99.2
31	3	0.2	1343	99.4
32	2	0.1	1345	99.6
33	1	0.1	1346	99.6
34	1	0.1	1347	99.7
35	1	0.1	1348	99.8
37	2	0.1	1350	99.9
50	1	0.1	1351	100.0

6.3.1.3 La Variable ESTAT CIVIL

La variable ESTAT CIVIL ha estat transformada degut a la poca variabilitat que recollien les categories originals. D'aquesta forma tenim ara dues modalitats, SOLTER i ALTRES, a on aquesta darrera inclou les de CASAT, VIDU, SEPARAT, RELIGIÓS i DIVORCIAT. La distribució de freqüències de la variable és coherent amb la distribució d'edats que hem vist anteriorment:

ESTAT CIVIL	Freq. Absoluta	Freq. Relativa
Solter/a	1329	98.4
Altres	22	1.6

6.3.1.4 La variable NACIONALITAT

La variable NACIONALITAT també presenta una forta concentració com les dues variables anteriors, en aquest cas obviament pel que respecta als de nacionalitat espanyola. La seva distribució de freqüències és:

NACIONALITAT	Freq. Absoluta	Freq. Relativa
Espanyola	1338	99.0
Estrangera	13	1.0

6.3.1.5 La variable DOMICILI

En els impresos de matrícula l'alumne ha d'indicar el codi corresponent a la comarca on es troba el seu domicili. Utilitzant aquests codis hem creat una variable amb les categories BARCELONA i FORA DE BARCELONA. La primera categoria la integren els alumnes que tenen el seu domicili a Barcelona ciutat o a les comarques del seu cinturó industrial: Barcelonès, Baix Llobregat, Vallès Occidental, Vallès Oriental i Maresme.

DOMICILI	Freq. Absoluta	Freq. Relativa
Barcelona	1164	86.2
Fora de Barcelona	187	13.8

6.3.1.6 La variable TIPUS DE RESIDÈNCIA

En els impresos de matrícula la variable TIPUS DE RESIDÈNCIA presentava les següents modalitats: Domicili Patern, Domicili Propi, Domicili Altres Familiars, Col·legi Major o Residència, Hotel-Pensió, Pis d'estudiants i Altres. Aquesta és la seva distribució de freqüències:

TIPUS DE RESIDÈNCIA	Freq. Absoluta	Freq. Relativa
Domicili Patern	1128	83.5
Domicili Propi	31	2.3
Domicili Altres Familiars	44	3.2
Col·legi Major/Residència	75	5.6
Hotel/Pensió	24	1.8
Pis d'estudiants	43	3.2
Altres	6	0.4

Amb la intenció d'agrupar les modalitats més homogènies de manera que quedin reflectides les diferències entre els tipus de residència que suposen un cost econòmic afegit i les altres, hem creat les següents categories: PARES que inclou el domicili patern i el domicili d'altres familiars, PRÒPIA que inclou el domicili propi i ALTRES que recull les altres modalitats, les quals tenen una component de càrrega econòmica important.

TIPUS DE RESIDÈNCIA	Freq. Absoluta	Freq. Relativa
PARES	1172	86.7
PRÒPIA	31	2.3
ALTRES	148	11.0

6.3.1.7 La variable SITUACIÓ LABORAL

En aquesta variable es recull la distinció entre aquells alumnes que treballen dels que no ho fan. Entre els que treballen hi són presents tant els que ho fan a temps parcial com els que ho fan a temps complet. La seva distribució de freqüències és:

SITUACIÓ LABORAL	Freq. Absoluta	Freq. Relativa
Treballa	190	14.1
No Treballa	1161	85.9

6.3.2 Variables de Gestió Acadèmica

6.3.2.1 La variable DURADA DELS ESTUDIS

Aquesta és la variable sobre la qual es fonamenten els objectius del present treball. La durada dels estudis recull el nombre d'anys que l'estudiant ha matriculat des que va iniciar la carrera en el curs acadèmic 1985-86. Definida d'aquesta manera, incorpora la possibilitat que els estudiants deixin de matricular conjunturalment en algun curs acadèmic, la qual cosa es posaria de manifest en el fet que, malgrat que continuï estudiant, la seva estada a la universitat és inferior a 8 anys.

A continuació es presenten algunes mesures descriptives així com la distribució de freqüències de la variable.

DURADA DELS ESTUDIS

Moments

N	1351
Mitjana	5.253886
Variança	6.482902
Desv Std	2.546154

Quantiles(Def=5)

100% Max	8	99%	8
75% Q3	8	95%	8
50% Med	6	90%	8
25% Q1	3	10%	1
0% Min	1	5%	1
		1%	1

Taula de Freqüències

ANYS	Freq. Absoluta	Freq. Relativa	Freq. Abs.Acum.	Freq. Rel.Acum.
1	206	15.2	206	15.2
2	100	7.4	306	22.6
3	76	5.6	382	28.3
4	52	3.8	434	32.1
5	186	13.8	620	45.9
6	159	11.8	779	57.7
7	204	15.1	983	72.8
8	368	27.2	1351	100.0

El creuament d'aquesta variable amb la següent, la SITUACIÓ ACADÈMICA DE L'ALUMNE, ens donarà una primera impressió de quin ha estat el comportament de la cohort en el període: percentatge d'abandonaments, de llicenciats, els que continuen, etc.

6.3.2.2 La variable SITUACIÓ ACADÈMICA DE L'ALUMNE

La variable SITUACIÓ ACADÈMICA ens informa de si l'alumne s'ha llicenciat, ha abandonat els estudis o si encara els continua. Serà de capital importància en les properes anàlisis utilitzant la metodologia pròpia dels models de supervivència, vist que és la que determina si una observació es troba censurada o no, a més d'indicar quin és el motiu que provoca la finalització de la durada dels estudis. La seva distribució de freqüències és:

SITUACIÓ ACADÈMICA	Freq. Absoluta	Freq. Relativa
Continua	392	29.0
Abandonament	645	47.7
Llicenciat	314	23.2

La lectura de la taula anterior ens mostra com el percentatge d'alumnes de la cohort que ha finalitzat els seus estudis no arriba al 25%, mentre que quasi el 50% els ha abandonat. Cal dir però que, dels 645 alumnes que han abandonat els estudis, 113 han deixat de matricular-se només l'actual curs acadèmic 1992-93 per la qual cosa podria ser que un percentatge considerable dels mateixos no hagin abandonat definitivament i tornin a matricular-se en el futur.

La següent taula mostra la distribució de freqüències conjunta de les variables SITUACIÓ ACADÈMICA i DURADA DELS ESTUDIS. En cada casella tenim la freqüència absoluta, la relativa, el percentatge respecte la fila i el percentatge respecte la columna, a més de les freqüències marginals de cadascuna de les variables en les columnes/files de totals.

Freq.Abs. Freq.Rel. % Fila % Colum	DURADA DELS ESTUDIS								Total
	1	2	3	4	5	6	7	8	
S	0	0	1	0	4	3	20	364	392
I Continua	0.00	0.00	0.07	0.00	0.30	0.22	1.48	26.94	29.02
T	0.00	0.00	0.26	0.00	1.02	0.77	5.10	92.86	
U	0.00	0.00	1.32	0.00	2.15	1.89	9.80	98.91	
A									
C	206	100	75	52	55	44	113	0	645
I Abandona	15.25	7.40	5.55	3.85	4.07	3.26	8.36	0.00	47.74
Ó	31.94	15.50	11.63	8.06	8.53	6.82	17.52	0.00	
	100.00	100.00	98.68	100.00	29.57	27.67	55.39	0.00	
A									
C	0	0	0	0	127	112	71	4	314
A Llicenciat	0.00	0.00	0.00	0.00	9.40	8.29	5.26	0.30	23.24
D	0.00	0.00	0.00	0.00	40.45	35.67	22.61	1.27	
È	0.00	0.00	0.00	0.00	68.28	70.44	34.80	1.09	
M									
I Total	206	100	76	52	186	159	204	368	1351
C	15.25	7.40	5.63	3.85	13.77	11.77	15.10	27.24	100.00
A									

De la inspecció de la taula es posa de manifest que:

1.- La circumstància d'abandonar els estudis temporalment per uns anys i tornar a reprendre'ls posteriorment és poc freqüent i en qualsevol cas els que ho han deixat i tornen ho han fet només un curs acadèmic. És a dir, quan un estudiant decideix deixar de matricular un curs acadèmic, molt probablement es tractarà d'un abandonament definitiu, tot i no considerant altres factors com les situacions personals i de rendiment acadèmic.

2.- Un percentatge molt elevat dels abandonaments es produeixen o bé en els dos primers anys o bé en els darrers anys. És a dir, quan s'inicien els estudis o quan la seva estada a la universitat és d'uns 6 ó 7 anys, és el moment més probable per a què es produeixi la decisió d'abandonar els estudis.

3.- Finalment cal aclarir que en l'any 8 no es computen abandonaments donat que fins que no finalitzi l'actual curs acadèmic 1992-93 i es completi el procés de matrícula pel proper curs 1993-94 no s'enregistra l'abandonament. Seguint la mateixa lògica, no haurien d'existir llicenciats l'any 8; els quatre que estan enregistrats ho són per que es va poder constatar, mentre es duia a terme el procés de depuració i perfeccionament de la base de dades, que van finalitzar els estudis en la convocatòria extraordinària de febrer d'aquest curs.

6.3.2.3 La variable FONT DE FINANÇAMENT DELS ESTUDIS

Aquesta variable recull la forma en la qual els estudiants aconseguixen els recursos per tal de pagar la matrícula. Les categories que la formen són BECA, TREBALL PROPI, AJUT DELS PARES i ALTRES, situant-se en aquesta darrera altres fonts de finançament com són el Treball del Cònjuge, l'Ajut d'altra persona i Altres mitjans. La seva distribució de freqüències és:

FONT FINANÇAMENT ESTUDIS	Freq. Absoluta	Freq. Relativa
Beca	100	7.4
Treball Propi	93	6.9
Ajut dels Pares	1130	83.6
Altres	28	2.1

Aquí s'observa com majoritàriament és l'ajut dels pares el que finança els estudis, mentre que es constata que no tots els que declaren treballar⁶ es financen els estudis amb el seu propi treball la qual cosa posa de manifest que molts dels que declaren treballar ho fan a temps parcial o esporàdicament.

6.3.2.4 La Variable CLASSE DE MATRÍCULA

Aquesta variable està molt relacionada amb l'anterior en el sentit que, a més de necessitar d'una font de finançament dels estudis, aquests poden tenir un cost o un altre segons existeixin o no les circumstàncies personals de l'alumne que donen dret a una reducció del preu de la matrícula. Les diferents classes de matrícula són: ORDINÀRIA, FAMÍLIA NOMBROSA DE 1^a, SEMIGRATUÏTA PER ALTRES CAUSES, FAMÍLIA NOMBROSA DE 2^a I D'HONOR, BECARI, GRATUÏTA PER ALTRES CAUSES, FUNCIONARI U.B. i FUNCIONARI M.E.C.

Als nostres efectes, ens interessa més captar les diferències en el preu de la matrícula que no pas les circumstàncies personals de l'alumne (si té família nombrosa o és fill de funcionari del M.E.C.) i és per això que hem recategoritzat la variable en les modalitats MATRÍCULA ORDINÀRIA, MATRÍCULA SEMIGRATUÏTA i MATRÍCULA GRATUÏTA. La distribució de freqüències és:

⁶ Segons es desprèn de la taula de freqüències de la variable SITUACIÓ LABORAL, són 190 els alumnes que declaren treballar.

CLASSE DE MATRÍCULA	Freq. Absoluta	Freq. Relativa
Ordinària	981	72.6
Semigratuïta	189	14.0
Gratuïta	181	13.4

6.3.2.5 La Variable CENTRE

La Llicenciatura en Ciències Econòmiques i Empresariales per la Universitat de Barcelona (com ara les noves Llicenciatures en Economia i en Administració i Direcció d'Empreses) es podia cursar tant a la Facultat d'Econòmiques com en el centre adscrit Abat Oliba, en aquest darrer cas només el primer cicle.

És prou sabut que les circumstàncies que envolten els alumnes que cursen la Llicenciatura en un o altre Centre no són les mateixes, si més no al llarg del primer cicle, especialment si tenim en compte la massificació existent en la Facultat d'Econòmiques que contrasta amb el reduït nombre d'estudiants per classe del Col·legi Abat Oliba. Pot resultar interessant, doncs, la consideració d'aquesta variable per tal d'analitzar la durada dels estudis, l'abandonament i la taxa de Llicenciats. La distribució de freqüències és:

CENTRE	Freq. Absoluta	Freq. Relativa
Facultat Econòmiques	1247	92.3
Abat Oliba	104	7.7

6.3.2.6 La variable HORARI DE CLASSES

Aquesta variable, tot i que està molt associada a d'altres variables considerades anteriorment, com ara la SITUACIÓ LABORAL, pot tenir la seva importància a l'hora d'estudiar la durada dels estudis. És conegut que existeixen

diferències substancials entre els grups de matí i els de tarda-nit, especialment en els cursos superiors de la Llicenciatura, les quals tenen a veure amb les diferents limitacions en la dedicació a l'estudi però també, en moltes ocasions, a la disparitat en els nivells d'atenció, concentració i interès a classe. La distribució segons l'horari de classes és:

HORARI DE CLASSES	Freq. Absoluta	Freq. Relativa
Matí	872	64.5
Tarda/Nit	479	34.5

6.3.3. Variables de Rendiment Acadèmic

Aquest bloc de variables està compost pel gruix de l'expedient acadèmic de l'estudiant i, per tant, disposem de la informació corresponent al nombre d'assignatures matriculades cada curs acadèmic (MAT85...MAT92), del nombre d'assignatures superades per curs (APR85...APR92)⁷, del nombre d'assignatures suspeses (S85...S92) i de les no presentades (P85...P92). Cal notar, com hem anat fent anteriorment, que per l'actual curs 1992-93 que encara no ha finalitzat i pel curs 1991-92 pel qual encara no han estat transferits els resultats als expedients informatitzats, no es disposa de la informació més enllà del nombre d'assignatures matriculades. A continuació detallem les distribucions de freqüències d'aquestes variables de rendiment acadèmic. Per cada curs acadèmic es presenten les variables: nombre d'assignatures matriculades, d'aprovades, de suspeses i de no presentades:

⁷ Aquí s'inclouen les aprovades pròpiament a més de les convalidades i adaptades.

CURS ACADÈMIC 1985-86

ASSIGS	MAT85 Percent		APR85 Percent		S85 Percent		P85 Percent	
0			161	11.9	560	41.5	303	22.4
1			137	10.1	416	30.8	294	21.8
2			121	9.0	223	16.5	315	23.3
3			156	11.5	97	7.2	142	10.5
4			180	13.3	44	3.3	94	7.0
5			257	19.0	10	0.7	66	4.9
6			170	12.6	1	0.1	47	3.5
7	1351	100.0	169	12.5			90	6.7

CURS ACADÈMIC 1986-87

ASSIGS	MAT86 Percent		APR86 Percent		S86 Percent		P86 Percent	
0	222	16.4	366	27.1	507	37.5	455	33.7
1			133	9.8	300	22.2	192	14.2
2	4	0.3	141	10.4	264	19.5	231	17.1
3	12	0.9	148	11.0	169	12.5	173	12.8
4	13	1.0	154	11.4	79	5.8	133	9.8
5	54	4.0	153	11.3	27	2.0	79	5.8
6	84	6.2	129	9.5	4	0.3	36	2.7
7	435	32.2	120	8.9	1	0.1	30	2.2
8	292	21.6	5	0.4			19	1.4
9	171	12.7	2	0.1			2	0.1
10	59	4.4					1	0.1
11	5	0.4						

CURS ACADÈMIC 1987-88

ASSIGS	MAT87 Percent		APR87 Percent		S87 Percent		P87 Percent	
0	324	24.0	438	32.4	695	51.4	549	40.6
1	4	0.3	89	6.6	315	23.3	219	16.2
2	4	0.3	102	7.5	189	14.0	196	14.5
3	9	0.7	112	8.3	100	7.4	137	10.1
4	23	1.7	117	8.7	31	2.3	105	7.8
5	36	2.7	119	8.8	16	1.2	51	3.8
6	92	6.8	143	10.6	5	0.4	36	2.7
7	304	22.5	165	12.2			26	1.9
8	274	20.3	50	3.7			19	1.4
9	180	13.3	15	1.1			9	0.7
10	78	5.8					2	0.1
11	14	1.0	1	0.1			1	0.1
12	4	0.3						
13	4	0.3						
17	1	0.1					1	0.1

CURS ACADÈMIC 1988-89

ASSIGS	MAT88 Percent		APR88 Percent		S88 Percent		P88 Percent	
0	386	28.6	538	39.8	789	58.4	547	40.5
1	1	0.1	73	5.4	285	21.1	182	13.5
2	7	0.5	107	7.9	173	12.8	183	13.5
3	13	1.0	117	8.7	70	5.2	151	11.2
4	20	1.5	150	11.1	26	1.9	93	6.9
5	51	3.8	130	9.6	7	0.5	63	4.7
6	145	10.7	121	9.0	1	0.1	48	3.6
7	304	22.5	84	6.2			57	4.2
8	239	17.7	22	1.6			12	0.9
9	118	8.7	8	0.6			5	0.4
10	45	3.3					7	0.5
11	12	0.9	1	0.1			2	0.1
12	6	0.4						
13	2	0.1						
14	1	0.1						
15	1	0.1					1	0.1

CURS ACADÈMIC 1989-90

ASSIGS	MAT89 Percent		APR89 Percent		S89 Percent		P89 Percent	
0	428	31.7	562	41.6	843	62.4	644	47.7
1	3	0.2	79	5.8	278	20.6	146	10.8
2	6	0.4	101	7.5	131	9.7	173	12.8
3	16	1.2	90	6.7	59	4.4	124	9.2
4	32	2.4	102	7.5	28	2.1	100	7.4
5	59	4.4	92	6.8	8	0.6	69	5.1
6	165	12.2	160	11.8	4	0.3	33	2.4
7	237	17.5	122	9.0			29	2.1
8	177	13.1	31	2.3			18	1.3
9	115	8.5	9	0.7			5	0.4
10	68	5.0	2	0.1			7	0.5
11	27	2.0					1	0.1
12	10	0.7	1	0.1			1	0.1
13	3	0.2						
14	4	0.3					1	0.1
15	1	0.1						

CURS ACADÈMIC 1990-91

ASSIGS	MAT90 Percent		APR90 Percent		S90 Percent		P90 Percent	
0	616	45.6	763	56.5	949	70.2	734	54.3
1	13	1.0	121	9.0	202	15.0	121	9.0
2	45	3.3	100	7.4	132	9.8	114	8.4
3	35	2.6	97	7.2	46	3.4	118	8.7
4	68	5.0	81	6.0	18	1.3	85	6.3
5	53	3.9	79	5.8	2	0.1	75	5.6
6	95	7.0	56	4.1	2	0.1	36	2.7
7	116	8.6	33	2.4			28	2.1
8	136	10.1	10	0.7			21	1.6
9	81	6.0	10	0.7			11	0.8
10	45	3.3	1	0.1			5	0.4
11	27	2.0						
12	11	0.8						
13	3	0.2						
14	2	0.1					2	0.1
15	2	0.1						

CURS ACADÈMIC 1991-92

ASSIGS	MAT91 Percent	
0	766	56.7
1	10	0.7
2	45	3.3
3	23	1.7
4	87	6.4
5	74	5.5
6	75	5.6
7	84	6.2
8	70	5.2
9	53	3.9
10	31	2.3
11	13	1.0
12	11	0.8
13	4	0.3
14	5	0.4

CURS ACADÈMIC 1992-93

ASSIGS	MAT92 Percent	
0	968	71.7
1	15	1.1
2	42	3.1
3	25	1.9
4	77	5.7
5	46	3.4
6	60	4.4
7	40	3.0
8	26	1.9
9	20	1.5
10	10	0.7
11	10	0.7
12	9	0.7
13	3	0.2

Aquest conjunt de variables ens ha de permetre la construcció de diferents indicadors del rendiment acadèmic de l'estudiant i de la seva progressió, els quals seran determinants en l'anàlisi posterior de la durada dels estudis. Als efectes d'utilitzar els resultats posteriors de les estimacions dels models de supervivència per tal d'orientar la política universitària en la seva vessant docent així com contrastar la validesa i optimalitat de la normativa de permanència, recentment aprovada pel Consell Social, són les variables de rendiment acadèmic les fonamentals, restant les altres variables personals i de gestió acadèmica com a controls de l'heterogeneïtat en la funció de supervivència.

A partir de les variables recollides en l'expedient acadèmic, i que tenen a veure amb el Pla d'Estudis 1980 vigent quan va començar la cohort del 85-86, i atesos els nous Plans d'Estudis aprovats i posats en marxa aquests curss, es realitza una traducció d'assignatures a crèdits a partir de les taules d'adaptació que recull el Pla d'Estudis en el seu apartat d'Organització, aprovat per Junta de Govern de la Universitat de Barcelona i homologat pel Consell d'Universitats. La consideració dels crèdits en comptes d'assignatures tindrà rellevància únicament per tal de trobar la connexió entre els resultats de l'anàlisi estadística de la durada dels estudis i la normativa de permanència en vistes a prendre decisions futures sobre la seva idoneïtat. A més, resulta interessant la introducció del concepte de crèdit en l'anàlisi de la durada dels estudis vist que incorpora una ponderació entre les assignatures segons les hores de classe a la setmana de les quals disposen. Aquesta ponderació afegida a la distinció entre assignatures troncal i obligatòries, per una banda, i les optatives i de lliure elecció, per una altra, possibilita un elevat grau de desagregació en l'anàlisi del rendiment acadèmic i de les seves conseqüències sobre la durada dels estudis.

La taula d'adaptacions recollides en els nous Plans d'Estudis és la següent:

Ll.C.Econ. i Emp.	Ll. Economia	Ll. Adm.i Dir.Emp.
Teoria Econòmica I	Microeconomia I (7.5) Macroeconomia I (7.5)	Microeconomia I (7.5) Macroeconomia II (7.5)
Matemàtiques I	Matemàtica Econòmica I (7.5)	Matemàtica Empresarial I (7.5)
Economia Empresa I	Economia Empresa I i II (12)	Economia Empresa I i II (15)
Història Ec.Mundial	Història Ec. Mundial (7.5)	Història Ec. Mundial (7.5)
Sociologia I	Sociologia (Op) (6)	Sociologia del Tr.i Emp. (Op) (6)
Tª de l'Estat i Dret Const.	Sistema Polític i Constituc. (Op) (6)	Sistema Polític i Constituc. (Op) (6)
Dret Civil	Introducció al Dret (7.5)	Dret Privat (6)
Teoria Econòmica II	Microeconomia II (7.5)	Microeconomia II (7.5)
Matemàtiques II	Matemàtica Econòmica II (7.5) Optimització Econòmica (Op) (6)	Matemàtica Empresarial II (7.5) Introducció a l'Optimització (Op) (6)
Economia Empresa II		Administració Empresarial (Op) (6)
Història Ec.Espanya	Història Ec. Espanya (7.5)	Història Ec. Espanya (Op) (6)
Estructura Ec. Mundial	Economia Mundial (7.5)	Economia Mundial (7.5)
Dret Mercantil	Institucions Mercantils (Op) (6)	Dret Empresarial (6)
Tª Comptabilitat I	Introducció a la Comptabilitat (7.5)	Introducció a la Comptabilitat (7.5)
Teoria Econòmica III	Macroeconomia II (7.5)	Macroeconomia II (7.5)
Política Econòmica I	Política Econòmica (7.5)	Política Econòmica (7.5)
Estadística Ec.i Emp.	Estadística Econòmica I i II (15)	Estadística Empresarial I i II (15)
Estructura Ec.Espanya	Economia Espanyola (7.5)	Economia Espanyola (7.5)
Tª Comptabilitat II	Compt. Financera i Analítica (7.5)	Comptabilitat Financera (7.5)
Dret Administratiu	Ordenament Administratiu (Op) (6)	Ordenament Administratiu (Op) (6)
Dret del Treball	Relacions Laborals (Op) (6)	Dret Laboral (Op) (6)
Especialitat Economia General		
Teoria Econòmica IV	Macroeconomia III (6) Microeconomia III (6)	Tª Econòmica de l'Emp. i la Ind. (6)
Política Econòmica II	Política Econòmica i Internacional (6)	Política Econòmica (2on.Cicle) (6)
Econometria	Econometria I i II (12)	Econometria Emp. I i II (9)
Hisenda Pública	Hisenda Pública I i II (9)	Fiscalitat Empresarial I (6)
Organització Econòmica Internacional	Organització Ec.Internacional (6)	Integració Europea (6)
Política Econòmica III	Política Econ. Espanya i la CE (6)	
Hª Doctrines Econòmiques	Hª Doctrines Econòmiques (6)	
Sistema Fiscal	Sistema Fiscal I i II (9)	Fiscalitat Empresarial II (6)
O.C.P.Hisenda i Ent.Públiques	Pressupostos i Compt. Pública (Op) (6)	
Models Econòmics	Econometria III (6) Micro o Macroeconometria (Op) (6)	Models de Programació (6)
Economia de Catalunya	Economia de Catalunya (Op) (6)	
Metodologia de les CC.SS.	Metodologia de les CC.SS. (Op) (6)	
Economia Regional i Urbana	Economia Regional i Urbana (Op) (6)	
Hisendes Locals	Hisenda Autònoma i Local (6)	

Especialitat Economia de l'Empresa		
Compt. Emp. i Est. Costos		Comptabilitat de Costos (6)
Matemàtica de les Op. Financeres		Matemàtica de les Op. Financeres (6)
Teoria Econòmica IV	Microeconomia III (6)	Tª Econòmica de la Emp. i la Ind. (6)
Política Econòmica II	Política Econòmica i Internacional (6)	Política Econòmica (2on.Cicle) (6)
Econometria de l'Empresa	Econometria I i II (12)	Econometria Emp. I i II (9)
Economia de l'Empresa III (Finan)		Gestió Financera (6)
Economia de l'Empresa III (Marq)		Fonaments del Màrqueting (6)
Verificació Compt., An. i Cons. Bal.		Anàlisi Comptable (6)
Política Econòmica de l'Empresa		Direcció Gral. i Estr. de l'Emp. (4.5)
Hisenda Pública i Reg. Fiscal Emp.		Fiscalitat Empresarial I i II (12)
Investigació Operativa i A.Sist.Emp.		Tècniques Operatives de Gestió (6) Models de Programació (6)
Política Econòmica III	Política Econòmica d'Espanya i CE (6)	
Economia de l'Empresa IV (Inv)		Anàlisi Financer (6)
Economia de l'Empresa IV (Mercats)		Gestió i Investigació de Mercats (6)
Especialitat Actuarial i de l'Empresa Financera		
Estadística Actuarial		
Teoria General de l'Assegurança		
Economia de l'Empresa Financera		
Matemàtica de les Op. Financeres		Matemàtica de les Op. Financeres (6)
Teoria Econòmica IV	Microeconomia III (6)	Tª Econòmica de l'Emp. i la Ind. (6)
Política Econòmica II	Política Econòmica i Internacional (6)	Política Econòmica (2on.Cicle) (6)
Econometria de l'Empresa	Econometria I i II (12)	Econometria Emp. I i II (9)
Teoria Matemàtica de l'Assegurança		
Dret de l'Assegurança i Ass. Socials		
Economia de l'Empresa Asseguradora		
Matemàtica de la Inversió		
Política Econòmica III	Política Econòmica d'Espanya i CE (6)	
Dret Bancari i Borsari		
Tª de la Població i Demografia		
Hisenda Pública i Reg. Fiscal Emp.	Hisenda Pública I (6) Sistema Fiscal I (3)	Fiscalitat Empresarial I i II (12)
Especialitat d'Economia Política i Sociologia		
Sociologia II	Sociologia Apl. a l'Economia (Op) (6)	
Teoria de l'Estat II	Sistema Polític i Constituc. (Op) (6)	
Mètodes Estadístics per a les CC.SS.	Econometria I i II (12)	
Teoria Econòmica IV	Microeconomia III i IV (12)	Tª Econòmica de l'Emp. i la Ind. (6) Macroeconomia (Op) (6)
Política Econòmica II	Política Econòmica Internacional (6)	Política Econòmica i Internacional (6)
Població i Estructura Social	Sociologia Apl. a l'Economia (Op) (6)	
Hisenda Pública	Hisenda Pública I i II (9)	Fiscalitat Empresarial I (6)
Organització Econòmica Internacional	Organització Econ. Internacional (6)	Integració Europea (6)
Metodologia de les Ciències Socials	Metodologia Ciències Socials (Op) (6)	

Sociologia III	Sociologia Apl. a l'Economia (6)	
Inst. Polít. i Admtves. de Catalunya	Sistema Polític i Constituc. (Op) (6)	
Economia Regional i Urbana	Economia Regional i Urbana (Op) (6)	
Política Econòmica III	Política Econòmica d'Espanya i CE (6)	
Història dels Moviments Socials		
Sociologia del Treball i l'Empresa	Sociologia Apl. a l'Economia (6)	

Entre parèntesis per a les assignatures dels nous plans d'estudis hi figuren el nombre de crèdits que els corresponen. Les assignatures del Pla d'Estudis 1980 a les quals no s'assigna una adaptació específica podran convalidar-se per crèdits de lliure elecció o optatius fins a un màxim de 30 i 18 crèdits respectivament.

Mitjançant la taula d'adaptacions anterior han estat traduïts els expedients dels alumnes des de les assignatures cap als crèdits. En cas que un estudiant no hagi iniciat el segon cicle o que havent-lo iniciat no correspongui ni a l'especialitat en Economia General ni a la d'Economia de l'Empresa, la traducció s'ha realitzat tenint en compte la taula d'adaptacions per a la llicenciatura on l'alumne obtindria un major nombre de crèdits superats.

Amb la informació provinent dels expedients acadèmics de l'alumne és possible construir variables indicadores del rendiment acadèmic. D'entre les mesures alternatives del rendiment acadèmic ha estat seleccionada la definida com Index de Progressió Acadèmica (Alemany et al., 1990)⁸ vist que resulta ser la que presenta un comportament més acurat per a la predicció dels abandonaments, tal com va ser comprovat empíricament en el treball esmentat anteriorment. Aquest Index de Progressió Acadèmica ha estat construït atenent, d'una banda, a les assignatures superades i, d'altra, als crèdits superats, i es calcula per a cada curs

⁸ Escudero (1986) proposa diverses mesures alternatives d'indicadors de progressió acadèmica com ara el que anomena index de despesa en anys de permanència definit com el quocient entre el nombre de cursos acadèmics diferents matriculats per l'estudiant i el nombre d'anys acadèmics emprats.

acadèmic com el nombre d'assignatures superades/crèdits superats acumulats fins a l'any en qüestió dividit pel nombre total d'assignatures/crèdits necessaris per a obtenir el títol de llicenciat, expressat en tant per cent. A continuació es detallen unes taules resum dels valors dels dos indicadors (IPAS per assignatures i IPAC per crèdits) en cada curs acadèmic, excepte pels darrers cursos acadèmics 1991-92 i 1992-93, vist que la informació d'assignatures/crèdits superats no es troba disponible.

CURS ACADÈMIC 1985-86

IPAS85	Alumnes	Percent	IPAC85	Alumnes	Percent
0%	161	11.9	0%	161	11.9
0 - 6.06%	258	19.1	0 - 6%	215	15.9
6 - 12.12%	336	24.8	6 - 12%	391	28.9
12 - 18.18%	427	21.6	12 - 18%	347	25.7
18 - 21.21%	169	12.5	18 - 21%	237	17.5

CURS ACADÈMIC 1986-87

IPAS86	Alumnes	Percent	IPAC86	Alumnes	Percent
0 - 9.09%	378	28.0	0 - 9%	367	27.2
9 - 18.18%	295	21.8	9 - 18%	289	21.4
18 - 27.27%	275	20.4	18 - 27%	311	23.0
27 - 36.36%	258	19.0	27 - 36%	242	17.9
36 - 42.42%	145	10.7	36 - 42%	142	10.5

CURS ACADÈMIC 1987-88

IPAS87	Alumnes	Percent	IPAC87	Alumnes	Percent
0 - 15.15%	449	33.6	0 - 15%	454	33.6
15 - 30.30%	272	20.1	15 - 30%	294	21.8
30 - 45.45%	289	21.4	30 - 45%	311	23.0
45 - 60.60%	254	18.8	45 - 60%	292	21.6
60 - 63.63%	87	6.4			

CURS ACADÈMIC 1988-89

IPAS88	Alumnes	Percent	IPAC88	Alumnes	Percent
0 - 21.21%	490	36.3	0 - 21%	501	37.1
21 - 42.42%	274	20.3	21 - 42%	313	23.2
42 - 63.63%	342	25.3	42 - 63%	354	26.2
63 - 84.84%	242	17.9	63 - 80%	183	13.5
80 - 90.90%	3	0.2			

CURS ACADÈMIC 1989-90

IPAS89	Alumnes	Percent	IPAC89	Alumnes	Percent
0 - 27.27%	531	39.3	0 - 27%	535	39.6
27 - 54.54%	275	20.4	27 - 54%	325	24.0
54 - 75.75%	234	17.3	54 - 75%	246	18.2
75 - 96.96%	225	16.7	75 - 100%	245	18.1
96 - 100%	86	6.4			

CURS ACADÈMIC 1990-91

IPAS90	Alumnes	Percent	IPAC90	Alumnes	Percent
0 - 27.27%	514	38.0	0 - 27%	524	38.8
27 - 54.54%	191	14.1	27 - 54%	238	17.6
54 - 75.75%	201	14.9	54 - 75%	226	16.7
75 - 96.96%	251	18.6	75 - 100%	363	26.9
96 - 100%	194	14.4			

A la vista de les taules anteriors no s'observen grans diferències entre els indicadors de progressió acadèmica per assignatures o per crèdits. Les petites diferències s'expliquen pel fet que els crèdits introdueixen una ponderació sobre les assignatures superades de tal forma que a igual nombre d'assignatures superades els corresponents valors de l'índex en crèdits poden ser diferents segons quines siguin en concret les assignatures superades. Comparant qualitativament ambdues mesures del rendiment acadèmic cal dir que pels estudiants de la cohort analitzada no era d'aplicació el sistema de crèdits introduït en els nous Plans d'Estudis per la qual cosa les seves decisions fonamentades en el seu rendiment acadèmic ho eren sempre pensant en assignatures i no tant atribuint un pes específic diferent a cadascuna d'elles tal com està implícitament recollit amb la incorporació dels crèdits.

En aquest sentit, la conveniència de la traducció de les assignatures a crèdits segons la taula d'adaptacions anteriorment exposada rau en la possibilitat de simular els efectes que la Normativa de Permanència recentment aprovada pel Consell Social de la Universitat de Barcelona hauria tingut sobre la cohort analitzada. La normativa estableix unes trajectòries mínimes a superar per l'alumne atenent al nombre de crèdits matriculats de forma acumulada al llarg de la seva estada a la universitat amb un factor de repetició fixat en el 1.75, la qual cosa

suposa que gairebé es disposa de dos cursos acadèmics per a aprovar una assignatura matriculada. A més, el primer any d'ingrés a la universitat, amb l'objectiu de compensar els efectes del procés d'adaptació de l'alumne al nou entorn universitari, la normativa únicament exigeix la superació d'un nombre reduït de crèdits independent dels que hagin estat matriculats. En aquestes condicions, els efectes de la normativa sobre la cohort estudiada haurien estat els que recull la següent taula:

	Exclosos	%	No Exclosos	%
Curs 1985-86	326	29.3	955	70.7
Curs 1986-87	179	18.7	776	81.3
Curs 1987-88	124	16.0	652	84.0
Curs 1988-89	92	14.1	560	85.9
Curs 1989-90	61	10.9	499	89.1
Curs 1990-91	40	8.0	459	92.0
Curs 1991-92	70	15.3	389	84.7
Curs 1992-93	48	12.3	341	87.7

La taula anterior mostra com en el primer curs acadèmic és quan s'hauria produït un percentatge més elevat d'alumnes exclosos, malgrat ser la normativa més flexible en el primer any d'estada a la universitat, i a partir d'aleshores el percentatge va disminuint progressivament. En els cursos 91-92 i 92-93, atès que la informació de les assignatures superades no es troba disponible, la simulació del nombre d'alumnes exclosos seria sota el supòsit que els alumnes no haurien superat cap assignatura i, per tant, la normativa només afectaria a aquells amb una trajectòria acumulada més ajustada i en conseqüència més necessitats de la superació d'assignatures any darrera any. Cal notar com, en el cas que un alumne tingui un rendiment acadèmic acceptablement alt al llarg d'uns quants anys, la normativa li permetria tenir un any dolent sense ser exclòs.

6.3.4 Variables acadèmiques prèvies

6.3.4.1 La variable TÍTOL D'INGRÉS

Aquesta variable ens informa del tipus d'accés a la Universitat que han tingut els estudiants de la cohort. Hem volgut distingir entre aquells alumnes que provenen del COU dels que ho fan via l'accés a Majors de 25 anys o que provenen de altres titulacions de grau mitjà o superior. La distribució de freqüències ens mostra com majoritàriament els alumnes de la cohort provenen del COU:

TÍTOL D'INGRÉS	Freq. Absoluta	Freq. Relativa
COU	1305	96.6
Altres	48	3.4

6.3.4.2 La Variable RESULTATS DE LES PROVES D'ACCÉS (PAAU)

Aquí recollim la nota final de les proves d'accés a la Universitat que és la mitjana de les proves d'accés pròpiament dites i de les notes mitjanes al llarg del BUP i del COU. Pel què fa a aquells estudiants provinents de l'accés als Majors de 25 anys només tenim constància de la qualificació d'Apte, la qual cosa ens obliga a qualificar-los numèricament, per tal d'homogeneitzar-los amb la resta d'alumnes i hem convingut assignar-los un 5.00.

Les Proves d'Accés a la Universitat tenen avui en dia una gran importància donat que sobre el resultat de les mateixes es fonamenta el sistema de *numerus clausus* vigent per a l'accés als ensenyaments en tots els centres i universitats. Per a la cohort sobre la qual treballem no van tenir la mateixa repercussió que actualment però si que el seu valor pot ser una aproximació al nivell de formació

i de capacitat i aprofitament de l'estudi per part de l'alumne, els quals poden influir molt directament en el seu rendiment acadèmic posterior a la universitat. A continuació presentem algunes mesures descriptives de la variable així com la seva distribució de freqüències:

RESULTATS DE LES P.A.A.U

Moments

N	1351
Mitjana	5.972605
Variança	0.3685622
Desv Std	0.6070933

Quantiles(Def=5)

100% Max	8.21	99%	7.87
75% Q3	6.30	95%	7.10
50% Med	5.90	90%	6.80
25% Q1	5.50	10%	5.30
0% Min	5.00	5%	5.10
		1%	5.00

Taula de Freqüències

PAAU	Freq. Absoluta	Freq. Relativa	Freq. Abs.Acum.	Freq. Rel.Acum.
5.0	51	3.8	51	3.8
5.1	26	1.9	77	5.7
5.2	45	3.3	122	9.0
5.3	61	4.5	183	13.5
5.4	88	6.5	271	20.0
5.5	91	6.7	362	26.7
5.6	93	6.9	455	33.6
5.7	91	6.7	546	40.3

Taula de Freqüències (Cont.)

PAAU	Freq. Absoluta	Freq. Relativa	Freq. Abs.Acum.	Freq. Rel.Acum.
5.8	95	7.0	641	47.3
5.9	83	6.1	724	53.4
6.0	91	6.7	815	60.1
6.1	72	5.3	887	65.4
6.2	75	5.6	962	71.0
6.3	74	5.5	1036	76.5
6.4	48	3.6	1084	80.1
6.5	44	3.3	1128	83.4
6.6	46	3.4	1174	86.8
6.7	36	2.7	1210	89.5
6.8	24	1.8	1234	91.3
6.9	26	1.9	1260	93.2
7.0	15	1.1	1275	94.3
7.1	11	0.8	1286	95.1
7.2	17	1.3	1303	96.4
7.3	7	0.5	1310	96.9
7.4	10	0.7	1320	97.6
7.5	7	0.5	1327	98.1
7.6	3	0.2	1330	98.3
7.7	3	0.2	1333	98.5
7.8	4	0.3	1337	98.8
7.9	5	0.4	1342	99.2
8.0	5	0.4	1347	99.6
8.1	1	0.1	1348	99.7
8.2	3	0.2	1351	100.0

Sobre les variables descrites en el present capítol hem realitzat algunes transformacions, fonamentalment hem obtingut variables indicadores per a les categòriques amb més de dues modalitats. A més, ha estat necessari dividir la població en dues subpoblacions segons els estudiants hagin abandonat o s'hagin llicenciat per tal d'estudiar de forma independent les probabilitats de finalització de la durada dels estudis per cadascun dels dos motius citats i com els hi afecten les variables personals i de rendiment acadèmic.

6.4 Transformacions sobre les variables

La realització dels contrastos i estimació dels models propis de l'anàlisi de la supervivència requereix treballar amb variables quantitatives. D'entre les variables explicatives que hem seleccionat només dues són estrictament quantitatives, l'EDAT i els RESULTATS DE LES PROVES D'ACCÉS A LA UNIVERSITAT, a més de les variables de rendiment acadèmic. Aquestes no ha calgut transformar-les, malgrat que s'han creat variables categòriques a partir de les originals amb efectes exclusivament de la confecció de gràfics il·lustratius de l'associació entre la variable i la durada dels estudis. Per a la resta de variables, les quals són qualitatives, s'han creat variables "dummy" o fictícies.

Les variables tal com han estat utilitzades en les anàlisis posteriors són les següents:

Variables Personals:

- **SEXE:**

$$SEXE = \begin{cases} 1 & \text{si Home} \\ 0 & \text{si Dona} \end{cases}$$

- **EDAT:**

Les categories creades pels gràfics són:

$$EDATR = \begin{cases} 1 & \text{si } 17 \leq EDAT \leq 20 \\ 2 & \text{si } 21 \leq EDAT \leq 24 \\ 3 & \text{si } EDAT \geq 25 \end{cases}$$

- ESTAT CIVIL:

$$ESCI = \begin{cases} 1 & \text{si Solter} \\ 0 & \text{si Casat/Separat/Vidu/} \\ & \text{Divorciat/Religiós} \end{cases}$$

- NACIONALITAT:

$$NAC = \begin{cases} 1 & \text{si Espanyola} \\ 0 & \text{si Estrangera} \end{cases}$$

- DOMICILI:

$$DOMIC = \begin{cases} 1 & \text{si Barcelona} \\ 0 & \text{si Fora Barcelona} \end{cases}$$

- TIPUS DE RESIDÈNCIA:

El tipus de residència està recollit per dues variables fictícies, restant la categoria Domicili dels Pares (la qual inclou Domicili d'altres familiars) com a categoria de referència.

$$RESID1 = \begin{cases} 1 & \text{si Domicili Propi} \\ 0 & \text{si Altre cas} \end{cases}$$

$$RESID2 = \begin{cases} 1 & \text{si Altres Residències} \\ 0 & \text{si Altre cas} \end{cases}$$

- SITUACIÓ LABORAL:

$$LABOR = \begin{cases} 1 & \text{si Treballa} \\ 0 & \text{si No treballa} \end{cases}$$

Variables de Gestió Acadèmica:

- FONT DE FINANÇAMENT DELS ESTUDIS:

La Font de Finançament dels estudis està recollida per tres variables fictícies, restant la categoria Ajut dels Pares com a categoria de referència.

$$FONF1 = \begin{cases} 1 & \text{si Beca} \\ 0 & \text{si Altre cas} \end{cases}$$

$$FONF2 = \begin{cases} 1 & \text{si Treball Propi} \\ 0 & \text{si Altre cas} \end{cases}$$

$$FONF3 = \begin{cases} 1 & \text{si Altres Fonts} \\ 0 & \text{si Altre cas} \end{cases}$$

- CLASSE DE MATRÍCULA:

La Classe de Matrícula està recollida per dues variables fictícies, restant la categoria Matrícula Ordinària com a categoria de referència.

$$CLMA1 = \begin{cases} 1 & \text{si Matrícula Semigratuïta} \\ 0 & \text{si Altre cas} \end{cases}$$

$$CLMA2 = \begin{cases} 1 & \text{si Matrícula Gratuïta} \\ 0 & \text{si Altre cas} \end{cases}$$

- CENTRE:

$$CENTRE = \begin{cases} 1 & \text{si Fac. Econòmiques} \\ 0 & \text{si Abat Oliba} \end{cases}$$

- HORARI:

$$HORARI = \begin{cases} 1 & \text{si Matf} \\ 0 & \text{si Tarda/Nit} \end{cases}$$

Variables Acadèmiques Prèvies:

- TÍTOL D'INGRÉS:

$$TITOL = \begin{cases} 1 & \text{si } C.O.U. \\ 0 & \text{si } \textit{Altres Títols} \end{cases}$$

- RESULTATS P.A.A.U.:

Les categories creades pels gràfics són:

$$PAAUR = \begin{cases} 1 & \text{si } 5.50 \geq PAAU \\ 2 & \text{si } 5.50 < PAAU \leq 5.90 \\ 3 & \text{si } 5.90 < PAAU \leq 6.30 \\ 4 & \text{si } PAAU > 6.30 \end{cases}$$

6.5 Subpoblacions de referència

En el nostre treball, la variable temps de supervivència és la durada dels estudis la qual vindrà determinada per dos successos puntuals com són, d'una banda l'abandonament dels estudis i de l'altra la finalització dels estudis amb èxit, és a dir, quan l'alumne es llicencia. Qualsevol d'aquestes dues circumstàncies marquen el final de la durada. Aquesta situació es coneix en la literatura de l'anàlisi de la supervivència com a *riscos en competència* o *múltiples causes de finalització de la durada* ("Competing risks")⁹. L'existència d'aquests dos tipus d'esdeveniments diferents que determinen la durada dels estudis ens obliga a realitzar una anàlisi separada per a ambdues subpoblacions.

⁹ Gail (1975), Seal (1977), Prentice, Kalbfleisch et al. (1978), Elandt-Johnson i Johnson (1980).

En la primera, que anomenarem dels Abandonaments, inclourem tots els estudiants de la cohort però prenent com a observacions censurades, a més d'aquelles que ho són pròpiament, és a dir, les dels estudiants la durada dels quals encara no ha finalitzat donat que continuen els estudis, les d'aquells alumnes que s'han llicenciat.

A la segona subpoblació, que anomenarem dels Llicenciats, inclourem tots els estudiants de la cohort però prenent com a observacions censurades, a més d'aquelles que ho són pròpiament, és a dir, les dels estudiants la durada dels quals encara no ha finalitzat donat que continuen els estudis, les d'aquells alumnes que han abandonat. En aquesta segona subpoblació, el fet que cap estudiant s'hagi llicenciat abans dels 5 anys portarà a limitar la durada des del cinquè any en endavant, donat que les observacions censurades pels anys anteriors no es tindran en compte en l'anàlisi.

Un cop efectuada l'anàlisi per a cadascuna de les subpoblacions i estimats els models de durada corresponents, podrem unificar ambdós tipus de resultats per a un estudi complet de la durada dels estudis. Les funcions d'atzar resultants seran la suma de les corresponents a cadascuna de les subpoblacions, tal com ha s'ha posat de manifest en l'apartat 2.8. La metodologia, doncs, permet analitzar els efectes de les variables explicatives de caire personal i les de rendiment acadèmic sobre la finalització de les durades de manera independent per a cadascuna de les dues subpoblacions a estudiar.

7. Resultats de les modelitzacions estàtiques

Capítol 7

Resultats de les modelitzacions estàtiques

Com s'ha comentat en el capítol segon, el problema de l'anàlisi de la durada dels estudis pot tractar-se des del punt de vista d'una situació de classificació. En efecte, un estudiant es classificarà o es situarà en un estat possible d'entre tres: "continua els seus estudis", "ja ha finalitzat amb èxit els seus estudis" o bé "ha abandonat els estudis". Habitualment, ens interessarem per situacions on establirem una dualitat entre dos possibles estats. Per exemple, considerarem primerament que volem veure el comportament dels estudiants que han abandonat en front de la resta, aquells que segueixen estudiant o bé que ja han acabat. Alternativament, podem interessar-nos pel conjunt d'estudiants que s'han llicenciat davant el col·lectiu complementari, que en qualsevol dels casos no ha aconseguit finalitzar completament els seus estudis. Com ja sabem, la disponibilitat d'informació mostral ens proporciona un coneixement dels individus d'estudi, i concretament trobem una sèrie de mesures o indicadors de les característiques individuals que, per comoditat, hem convingut a transformar en variable dicotòmiques en la majoria dels casos, per tal de captar l'existència o

l'absència d'alguna característica. Aquestes variables podem considerar que estan associades al comportament individual relatiu a les circumstàncies personals i acadèmiques.

L'aproximació a l'anàlisi de la permanència o durada dels estudis pot tractar-se no tan sols des de la perspectiva dels models de supervivència, on el factor clau és la modelització de la variable que indica la durada o temps que l'individu s'ha mantingut en un estat, sinó que podem centrar-nos des d'un punt de vista més bàsic en un estudi de classificació de l'individu en qualsevol dels possibles estats que s'han definit en el començament. Aquesta aproximació estarà basada en el tractament de la informació mostral de què es disposa i en la utilització de la metodologia estadística que requereix el problema.

La utilització de la metodologia de classificació és anterior a l'ús de les tècniques de modelització en aquest context. El treball de Willett i Singer (1991) aporta una revisió de la bibliografia existent i, particularment, fa referència als inconvenients de la utilització de models discriminants en el tractament de problemes de permanència. En qualsevol cas, encara que des d'un punt de vista estadístic ja es justifica sobradament la utilització de l'anàlisi de supervivència, convé veure com efectivament el fet d'emprar una metodologia més clàssica basada en l'anàlisi discriminant pot conduir a conclusions que poden ésser congruents amb una anàlisi més avançada, però que són força inestables. De fet, amb els resultats que es presentaran a continuació, es pretén mostrar com la variable "durada dels estudis", és a dir, el factor de temporalitat, constitueix una característica que té una influència essencial en el comportament de permanència (en la supervivència dels individus) i que per tant s'ha d'incloure en l'estudi.

Per tant, els resultats mostraran com la qüestió pot ésser considerada des d'una altra perspectiva, alternativa a l'anàlisi més avançada com és l'anàlisi de supervivència. Si volem establir un mètode de discriminació entre els diferents grups d'alumnes, l'aspecte clau esdevé la construcció una suma ponderada de les variables explicatives quantitatives i explicatives indicadores (dicotòmiques) de manera que aquesta suma permeti distingir entre dos casos o individus en un dels dos estats contraposats. La suma ponderada es pot considerar o transformar en una puntuació, de manera que el procediment de predicció de pertinença a un grup, és a dir la classificació sigui un pas immediat. D'alguna forma podríem dir que busquem una modelització de la probabilitat d'abandonar els estudis, sense tenir en compte la component indicativa del moment en què es produeix l'abandonament. Anàlogament es realitza el raonament sobre la probabilitat de finalitzar amb èxit els estudis.

La metodologia existent per al tractament que s'acaba d'exposar engloba des de l'anàlisi de regressió múltiple, i el model de probabilitat lineal, fins l'anàlisi discriminant o bé les tècniques iteratives que es basen en la maximització d'alguna funció d'"informació". L'inconvenient principal de tots aquests mètodes apareix en la inconcreció que es produeix sobre el tractament que han de rebre les dades censurades.

Naturalment, un dels aspectes fonamentals de l'estudi global que estem realitzant, va encaminat a obtenir una metodologia que proporcioni una certa habilitat en la predicció de quins són els individus que tenen més propensió (major probabilitat) a l'abandonament, possiblement sense tenir en compte el temps que han estat estudiant. Com ja hem esmentat abans, la inclusió d'aquest factor és l'aspecte bàsic que caracteritza l'anàlisi de supervivència.

Una possibilitat molt senzilla que permetria veure la possibilitat de classificació dels individus en els estats que podrien anomenar, el primer "satisfactori" (continua estudiant o ja és llicenciat), o bé el segon, "insatisfactori" (s'ha produït l'abandonament)¹⁰ seria l'establiment d'un model de regressió lineal múltiple i la construcció d'una taula de puntuacions. En aquest cas, la variable dependent seria una variable dicotòmica. Aquesta variable pren un dels dos possibles valors en el cas que l'individu es trobi en l'estat "satisfactori", i l'altre alternatiu en el cas que l'individu estigui en el cas de l'estat "insatisfactori". En realitat, aquest procediment és el més senzill, i per tant el que permet una més fàcil implementació i presentació als efectes il·lustratius que aquí es pretenen. El què estem construint és un model de probabilitat lineal, en una situació on la utilització de models adequats a la modelització de variables dicotòmiques potser és recomanable. De tota manera, per la seva simplicitat s'ha optat per a presentar inicialment els resultats que s'obtenen mitjançant la utilització del model de regressió. En el model plantejat, establim que com explicatives de la variable dependent prendrem les variables que han estat emprades en els models de supervivència.

7.1 Model de probabilitat lineal pels abandonaments

Definim Y la variable que pren el valor 1 si l'estudiant ha abandonat els seus estudis, i el valor 0 si l'estudiant no els ha abandonat. Notem com la principal crítica es produeix precisament pel fet de disposar d'informació censurada, ja que possiblement alguns dels estudiants que no han abandonat els

¹⁰ Anàlogament es tractaria el cas en què l'estat "satisfactori" correspongués al grup d'estudiants que s'han llicenciat i l'estat "insatisfactori" agrupés els estudiants que, o bé continuen els estudis o bé els han abandonat.

estudis, és possible que els abandonin en un futur posterior a la data de l'estudi, i per tant la variable indicadora que intentem modelitzar no prendria el valor correcte per a aquest grup d'individus.

Els resultats es presenten a continuació:

ESTIMACIONS DEL MODEL DE PROBABILITAT LINEAL SUBPOBLACIÓ D'ABANDONAMENTS

Anàlisi de la Variança:					
Font	DF	Suma de Quadrats	SQ / DF	Estad F	Prob>F
Model	8	42.89872	5.36234	24.464	0.0001
Error	1342	294.16272	0.21920		
C Total	1350	337.06144			
R ²	0.1273				
Adj R ²	0.1221				
Variable	Paràmetre Estimat	Error Estàndard	T per H0: Paràmetre=0	Prob > T	
INTERCEP	1.288304	0.26105916	4.935	0.0001	
RESPAN	-0.001714	0.00022244	-7.707	0.0001	
CENTRE	0.239069	0.05380443	4.443	0.0001	
NAC	-0.388044	0.13105443	-2.961	0.0031	
EDAT	0.023554	0.00805304	2.925	0.0035	
HORA	-0.116197	0.03173579	-3.661	0.0003	
RESID1	-0.271047	0.10772345	-2.516	0.0120	
FONF2	0.211098	0.06382780	3.307	0.0010	
FONF3	0.220914	0.09212994	2.398	0.0166	

El model és globalment significatiu i igualment ho són cadascun dels paràmetres coeficients de l'equació. Encara que el coeficient de determinació indiqui una bondat de l'ajust baixa, aquest fet es deu a la naturalesa dicotòmica de la variable dependent, i només reflecteix la inadequació d'utilitzar una combinació de regressors per a ajustar-la. Aquests resultats s'han de prendre com una senzilla aplicació de la metodologia clàssica. La interpretació dels resultats obtinguts és molt directa i de la mateixa destaquen dues característiques importants: d'una banda totes les variables que hem inclòs en el model són significatives i de l'altra el signe dels coeficients que acompanyen a aquestes és directament interpretable i indica el tipus d'influència de les característiques individuals en el fet que es produeixi o no l'abandonament dels estudis. En aquest sentit, és directe constatar que els individus amb nacionalitat espanyola, que estudien en horari de matí resideixen en domicili propi (RESID1=1) i han obtingut una qualificació més elevada en les proves d'accés, tindran menys propensió a l'abandonament. Contràriament, les característiques relatives a no estudiar en el centre adscrit, tenir font de finançament dels estudis Treball propi o altres fonts (FONF2=1 i FONF3=1), o bé tenir una edat més elevada, són aquelles que produeixen més inclinació a l'abandonament.

Naturalment, una anàlisi com l'anterior, realitzada en el present estudi de magnitud més àmplia és únicament exploratori i confirmatori dels resultats. Les prediccions sobre la variable Y que es poden construir a partir de les anteriors estimacions, ens ajudaran a determinar la crucialitat de la variable "durada dels estudis", i en tot cas, en la selecció de les variables que utilitzarem com a regressors en els models més avançats.

7.2 Model de probabilitat lineal pels llicenciats

Si ara la variable dependent Y2 correspon a un indicador que determina si un estudiant ja s'ha llicenciat abans del moment de finalitzar el present estudi, obtindrem a través d'un model de probabilitat lineal els resultats que es presenten a continuació.

ESTIMACIONS DEL MODEL DE PROBABILITAT LINEAL SUBPOBLACIÓ DE LLICENCIATS

Anàlisi de la Variança:					
Font	DF	Suma de Quadrats	SQ / DF	Estad F	Prob>F
Model	8	31.86181	3.98273	25.554	0.0001
Error	1342	209.15817	0.15586		
C Total	1350	241.01999			
R ²	0.1322				
Adj R ²	0.1270				
Variable	Paràmetre Estimat	Error Estàndard	T per H0: Paràmetre=0	Prob > T	
INTERCEP	-1.101061	0.22013168	-5.002	0.0001	
RESPAN	0.002166	0.00018757	11.547	0.0001	
CENTRE	-0.135214	0.04536926	-2.980	0.0029	
NAC	0.186502	0.11050840	1.688	0.0917	
EDAT	-0.002963	0.00679053	-0.436	0.6626	
HORA	0.066227	0.02676042	2.475	0.0135	
RESID1	0.060101	0.09083513	0.662	0.5083	
FONF2	-0.112821	0.05382121	-2.096	0.0362	
FONF3	-0.057392	0.07768629	-0.739	0.4602	

Com abans, el model té com a principal defecte que aquells individus que en un futur sí que obtindran la seva llicenciatura, en aquest moment són inclosos en el grup dels que continuen estudiant o bé han abandonat. De fet, el model té utilitat en termes de determinar el grup d'individus que aconseguix finalitzar els estudis en un període no inferior a la durada que estableix el moment en què finalitza la recopilació de dades per al present estudi.

Els comentaris respecte als resultats d'aquest model són semblants als assenyalats anteriorment per al cas de modelitzar els abandonaments. Destaca únicament en aquest cas la significació de la variable relativa a la qualificació en les proves d'accés i la pèrdua de significació en un bon nombre de variables, de manera que sembla evident que algunes característiques personals i acadèmiques ja no semblen determinar una major propensió a acabar els estudis en un termini breu.

7.3 Relació amb la variable "durada dels estudis"

El següent pas ha consistit en fer un creuament de la informació que s'obté de l'estimació dels anteriors models amb la variable "durada" dels estudis, per tal de mostrar l'essencialitat de tenir en compte aquest factor. De fet el raonament es basa en suposar que el factor temporal no es té en compte, per obtenir després resultats que no són estables en diferents talls temporals. Per dir-ho en altres paraules, es pot veure com les propensions a pertànyer als diferents estats, la probabilitat d'abandonar o d'acabar els estudis, no es manté constant al llarg dels diferents moments de la trajectòria acadèmica. D'aquí la necessitat de dur a terme una modelització que admeti aquestes variacions.

Les taules següents són il·lustratives. Suposem que utilitzem el primer model estimat per a realitzar una predicció per tots els individus de la mostra. En realitat, si la predicció s'acostés a 1, diríem que l'individu prediríem que abandona, i respectivament si s'acosta a 0 que no abandona. Establím com a punt de tall el 0.5. Ara determinem diferents nivells de durada dels estudis: d'un a tres anys, de 4 a 5 anys, 6, 7 o 8 anys de durada. Ara, com que estem utilitzant la informació de la mostra i per cada individu sabem en quin estat es troba, podem fer les classificacions següents:

CLASSIFICACIÓ SEGONS EL MODEL DE PROBABILITAT LINEAL
SUBPOBLACIÓ D'ABANDONAMENTS
Estudiants que continuen els seus estudis

Anys	1 a 3	4 a 5	6	7	8
Predicció $Y < 0.5$	1	3	2	10	227
Predicció $Y \geq 0.5$	0	1	1	10	137
Total	1	4	3	20	364

Interpretem l'anterior taula: tenim per exemple 20 estudiants que durant 7 anys estant realitzant la seva carrera i no l'ha finalitzada, però no han abandonat (únicament ho han pogut fer temporalment, tornant més tard a matricular-se). El model hauria de classificar tots els 20 estudiants amb un predicció propera a 0, donat que no abandonen. Però el model classifica incorrectament el 50% dels casos. Per al cas dels estudiants amb una durada de 8 anys, el model funciona més acceptablement classificant correctament el percentatge corresponent a la freqüència relativa 227/364. No parlem dels altres casos donat que hi ha molt pocs individus.

Passem ara a revisar la taula més interessant; com que el model estava dedicat a predicció de l'abandonament, veiem què succeeix per a aquest col·lectiu d'individus de què disposem en la mostra.

CLASSIFICACIÓ SEGONS EL MODEL DE PROBABILITAT LINEAL

SUBPOBLACIÓ D'ABANDONAMENTS

Estudiants que abandonen els seus estudis

Anys	1 a 3	4 a 5	6	7	8
Predicció $Y < 0.5$	147	51	21	63	0
Predicció $Y \geq 0.5$	234	56	23	50	0
Total	381	107	44	113	0

Fixem-nos com el model aconsegueix un bon percentatge de classificacions correctes únicament si es tracta d'abandonaments que es produeixen en els primers anys dels estudis. De fet, dels 107 estudiants que abandonen en el quart o cinquè any d'estudis el model en classifica 51 com si no abandonessin. I el percentatge de mala classificació obtinguda per les prediccions del model és encara pitjor si tenim en compte el grup que abandona en el setè any d'estudis, situant-se en aquest cas per sobre del 50% de mala classificació. Una vegada més veiem com una aproximació com l'anterior no proporciona resultats acceptables, al menys per certes franges, és a dir per certes durades. Essencialment el resultat clau és que els percentatges de classificació correcta, és a dir, el comportament del model, són excessivament inestables i podrien variar moltíssim depenent del moment del temps de vida acadèmica en què decidim efectuar la predicció.

Finalment, per al grup d'individus que són llicenciats i amb diferents durades dels estudis, veiem quines són les freqüències absolutes que s'obtenen en realitzar una predicció pel model de regressió.

**CLASSIFICACIÓ SEGONS EL MODEL DE PROBABILITAT LINEAL
SUBPOBLACIÓ D'ABANDONAMENTS**

Estudiants que finalitzen amb èxit els seus estudis

Anys	1 a 3	4 a 5	6	7	8
Predicció $Y < 0.5$	0	123	91	56	4
Predicció $Y \geq 0.5$	0	4	21	15	0
Total	0	127	112	71	4

Veiem com en aquest cas el model classifica correctament un elevat percentatge dels casos, encara que únicament ens diria que no abandonen. Malgrat això hi ha una forta discrepància dels resultats obtinguts per a la durada de 5 anys i les durades superiors. Aquests resultats no haurien de sorprendre donat que justament aquest grup d'individus que es llicencia, és el que hauria d'estar més clarament diferenciat del grup d'individus que abandonen. Els que continuen els seus estudis, potser es trobarien en una distància intermitja i per tant serien més difícils de classificar o detectar.

Ara podem dur a terme la mateixa anàlisi amb el model de regressió que es presentava en segon lloc i que servia per modelitzar la variable dicotòmica que indica si un estudiant es llicencia en el termini de temps breu durant el qual es va estudiar aquest cohort. Les taules de classificació obtingudes amb les prediccions de l'anterior model són les següents:

**CLASSIFICACIÓ SEGONS EL MODEL DE PROBABILITAT LINEAL
SUBPOBLACIÓ DE LLICENCIATS**

Estudiants que continuen els seus estudis

Anys	1 a 3	4 a 5	6	7	8
Predicció $Y < 0.5$	1	4	3	18	353
Predicció $Y \geq 0.5$	0	0	0	2	11
Total	1	4	3	20	364

Estudiants que abandonen els seus estudis

Anys	1 a 3	4 a 5	6	7	8
Predicció $Y < 0.5$	375	104	42	112	0
Predicció $Y \geq 0.5$	6	3	2	1	0
Total	381	107	44	113	0

Estudiants que finalitzen amb èxit els seus estudis

Anys	1 a 3	4 a 5	6	7	8
Predicció $Y < 0.5$	0	103	102	66	4
Predicció $Y \geq 0.5$	0	25	10	5	0
Total	0	127	112	71	4

El principal defecte de la utilització del model de probabilitat, sovint com en aquest cas, està en la inadequació del punt del criteri de classificació basat en la predicció. En la situació anterior, per exemple, i fixant-nos en la darrera taula, veiem com el model no detecta a la majoria dels estudiants que es llicencien, especialment si la durada es superior al mínim de 5 anys. Encara que de manera molt lleugera, també es mostra com la incorrecta classificació té diferents percentatges segons els moment temporal que es prengui. Potser en aquest cas l'estudi seria més evident si es construís una malla de valors per a la predicció més detallada.

En les taules següents, amb una categorització més detallada de les prediccions de la variable que indica quins són els estudiants que s'han llicenciat, queda encara més clar com el comportament de les prediccions, que estan relacionades amb el funcionament del model per classificar correctament nous estudiants, és de naturalesa molt diferent segons quina sigui la durada dels estudis. Fixem-nos en particular en la darrera taula i veurem com l'histograma de les prediccions de la variable Y2 seria diferent si es fes per durades de 5, 6, 7 o 8 anys. Per a les altres dues taules l'únic que es veu és que, en el cas dels alumnes que continuen els seus estudis com que la majoria encara segueixen, la major part es troba en l'interval de durada de 8 anys. Pels estudiants que han abandonat, el model dona prediccions de Y2 sempre al voltant de 0.2, la qual cosa és correcta, però la dispersió d'aquestes prediccions és força diferent segons el moment en què es considera l'abandonament. En particular, el moment on hi ha menys dispersió és en els abandonaments després de poca durada dels estudis.

CLASSIFICACIÓ SEGONS EL MODEL DE PROBABILITAT LINEAL**SUBPOBLACIÓ DE LLICENCIATS****Estudiants que continuen els seus estudis**

Anys	1 a 3	4 a 5	6	7	8
< 0	0	0	0	2	6
[0, 0.2)	0	1	1	9	146
[0.2, 0.4)	1	3	2	6	175
[0.4, 0.6)	0	0	0	1	34
[0.6, 0.8)	0	0	0	2	2
[0.8, 1.0)	0	0	0	0	1
[1.0, 1.2)	0	0	0	0	0
> 1.2	0	0	0	0	0
Total	1	4	3	20	364

Estudiants que abandonen els seus estudis

Anys	1 a 3	4 a 5	6	7	8
< 0	42	10	0	3	0
[0, 0.2)	197	51	21	53	0
[0.2, 0.4)	127	34	19	51	0
[0.4, 0.6)	14	10	4	6	0
[0.6, 0.8)	1	2	0	0	0
[0.8, 1.0)	0	0	0	0	0
[1.0, 1.2)	0	0	0	0	0
> 1.2	0	0	0	0	0
Total	381	107	44	113	0

Estudiants que finalitzen amb èxit els seus estudis

Anys	1 a 3	4 a 5	6	7	8
< 0	0	0	1	2	0
[0, 0.2)	0	8	25	18	0
[0.2, 0.4)	0	68	63	42	4
[0.4, 0.6)	0	38	19	7	0
[0.6, 0.8)	0	12	4	2	0
[0.8, 1.0)	0	1	0	0	0
[1.0, 1.2)	0	0	0	0	0
> 1.2	0	0	0	0	0
Total	0	127	112	71	4

7.4 Una altra alternativa clàssica: l'anàlisi discriminant

Una segona aproximació clàssica al problema d'estudiar el fet que un individu se situï en un estat entre dos alternatius consisteix en efectuar una anàlisi discriminant. Aquesta anàlisi calcula la distància estadística de cada individu al "centre" de cadascun dels grups que es volen discriminar i per els quals es té també informació estadística. Cal tenir en compte però, que tota la inferència estadística que es deriva de l'anterior anàlisi està basada en el supòsit de distribució normal multivariant de les variables explicatives que s'utilitzen en el procés. En el cas que ens ocupa, amb un elevat nombre de variables indicadores, aquesta hipòtesi no es pot assumir, i per tant, els contrastos que es realitzen sobre els paràmetres estimats no es podran donar com a definitius¹¹.

¹¹ Existeixen altres aproximacions no paramètriques que eviten la hipòtesi de normalitat i segurament serien més apropiades en aquest cas.

Ens centrarem en la variable que defineix el grup d'abandonaments i discriminarem entre aquest col·lectiu i el seu complementari. Els resultats de la funció lineal discriminant són:

**RESULTATS ANÀLISI DISCRIMINANT
SUBPOBLACIÓ D'ABANDONAMENTS**

	0	1
CONSTANT	-212.56298	-208.85755
RESPAN	0.23602	0.22816
CENTRE	14.49580	15.59215
NAC	113.13063	111.35110
EDAT	8.68763	8.79565
HORA	-2.99193	-3.52479
RESID1	-40.16926	-41.41225
FONF2	-30.31438	-29.34631
FONF3	-3.15841	-2.14532

I els resultats de la reclassificació del individus de la mostra:

Nombre d'observacions i Percentatge classificat en Y:

Des de Y	0	1	Total
0	486	220	706
	68.84	31.16	100.00
1	244	401	645
	37.83	62.17	100.00
Total	730	621	1351
Percentatge	54.03	45.97	100.00

Els percentatges de mala classificació:

	0	1	Total
Percentatge	0.3116	0.3783	0.3450
Prob Priori	0.5000	0.5000	

Com a comentari final direm que el mètode proporciona percentatges de classificació per sobre del 60%, que es poden considerar acceptables, però la principal crítica és precisament que no podem assegurar que el criteri de classificació hagi de ser el mateix al llarg dels diferents períodes de durada. La segona crítica, fonamental, igual que l'anterior fa referència a què el tractament òbvia el censurament present a les dades i no es pot fer cap reflexió sobre les conseqüències d'ignorar aquest fet. Per concloure aquest apartat es pot efectuar una anàlisi de la funció lineal discriminant, no ja pel que fa referència a la significació, però si a la interpretació. Aquelles característiques que fan inclinar l'individu a no abandonar són: més elevats resultats en les proves d'accés, ser espanyol, tenir horari de matí i residència en domicili propi. Aquelles que fan augmentar la tendència a abandonar són el centre UB, una edat més elevada i financen els estudis amb el treball propi o altres fonts. Aquesta interpretació, fixem-nos com és totalment congruent amb el què s'observava per al corresponent model de probabilitat lineal.

Si ara fem un mateix anàlisi discriminant per a la variable Y2, obtenim uns resultats molt similars als anteriors,

RESULTATS ANÀLISI DISCRIMINANT
SUBPOBLACIÓ DE LLICENCIATS

	0	1
CONSTANT	-212.97184	-221.80050
RESPAN	0.24127	0.25524
CENTRE	14.94801	14.07593
NAC	112.16250	113.36537
EDAT	8.80575	8.78663
HORA	-3.20811	-2.78097
RESID1	-41.34380	-40.95617
FONF2	-29.86840	-30.59606
FONF3	-2.26081	-2.63096

I els resultats de la reclassificació del individus de la mostra:

Nombre d'observacions i Percentatge classificat en Y:

Des de Y	0	1	Total
0	725	312	1037
	69.91	30.09	100.00
1	113	201	314
	35.99	64.01	100.00
Total	838	513	1351
Percentatge	62.03	37.97	100.00

Els percentatges de mala classificació:

	0	1	Total
Percentatge	0.3009	0.3599	0.3304
Prob Priori	0.5000	0.5000	

En aquest cas els resultats de classificació són encara millors que els anteriors, el fet es deu potser a la major distància del col·lectiu d'estudiants que es llicencien. No s'inclouen probabilitats a priori per als dos grups.

Aquest capítol s'ha dedicat a explorar, amb els models més senzills possibles, els resultats que s'obtindrien en les nostres dades si féssim servir els models i tècniques més clàssiques. L'objectiu ha estat confirmar les interpretacions de les influències de les diferents variables en un augment de la probabilitat d'abandonar o bé de llicenciar-se. En aquest context les interpretacions són molt senzilles i el signe dels paràmetres ja aporta informació sobre la influència de les variables que acompanyen respectivament. En segon lloc, hem mostrat com l'eix temporal aporta un element d'informació que és necessari incorporar en l'anàlisi, donat que algunes conclusions variarien segons quin moment de la durada fixéssim i, per tant, no es pot fer un judici global, sense tenir en compte aquest factor.

Spady (1970), Tinto (1975) i Trenzini i Pascarella(1980) ja han tractat la utilització de mètodes clàssics en contextes similars al present i, com en el cas que s'acaba de presentar, els procediments que han emprat són febles donat que depenen del moment de la trajectòria acadèmica en què s'interromp l'estudi i no es realitza un tractament explícit de la censuració de les dades encara que com en el cas que ens ocupa el fet de tractar una cohort fa que aquest efecte sigui menys rellevant (com esmenten Willett i Singer, 1991). La darrera observació que s'ha de fer és que els models com els anteriors, a fi d'evitar correlacions entre els regressors, no inclouen característiques que puguin ésser variables al llarg del temps o que s'hi considerin i en aquest sentit són explicacions estàtiques per un procés que és intrínsecament dinàmic.

8. Resultats previs de l'anàlisi de dades de supervivència

Capítol 8

Resultats previs de l'anàlisi de dades de supervivència

L'aplicació de la metodologia de l'anàlisi de la supervivència descrita en en els capítols 3 a 5 al cas concret dels estudiants de la cohort de 1985-86 en la Llicenciatura en Ciències Econòmiques i Empresariales de la Universitat de Barcelona ens permetrà analitzar la durada dels estudis en relació a la informació que proporciona la base de dades sobre les característiques personals i acadèmiques de l'alumne de tal manera que serà possible concretar quins són els factors que, d'una banda, precipiten els abandonaments i, d'altra, faciliten el camí envers l'obtenció del títol de llicenciat. L'anàlisi, però, anirà molt més enllà atès que, no tan sols mostrarà els factors que determinen l'eventualitat de l'acabament del estudis per un o altre motiu, sinó que a més els hi afegirà una component temporal, és a dir, evidenciarà en quin moment del temps serà més o menys probable l'acabament de la durada. Aquest perfil temporal de les funcions d'atzar és el que caracteritza l'anàlisi de les durades i sobre el qual es fonamenta la diferència bàsica amb els mètodes clàssics del capítol anterior.

En una primera fase prèvia, dins del present capítol, estudiarem l'associació entre les variables seleccionades i la durada dels estudis utilitzant els mètodes no paramètrics clàssics. Aquests seran bàsicament contrastos estadístics de rangs i instruments gràfics amb els quals serà possible determinar l'existència de relacions entre els factors de caire personal i acadèmic i el nombre d'anys que constitueix l'estada dels individus a la universitat. Aquesta anàlisi s'efectuarà tant de forma individual per a cadascuna de les variables com per a totes elles considerades en el seu conjunt. Aquesta darrera, consistent en un contrast conjunt per etapes, és de gran utilitat per tal de posar de manifest les interrelacions existents entre les esmentades variables explicatives i, per tant, de la informació redundant recollida en les mateixes.

En els capítols següents, amb els resultats previs disponibles, modelitzarem la durada dels estudis per tal de captar i mesurar els efectes dels factors personals i acadèmics que expliquen l'heterogeneïtat existent en la població. Amb aquest objectiu, s'utilitzaran diferents alternatives de modelització com ara les paramètriques i les semiparamètriques, i dins d'aquestes les que consideren el temps continu o discret, comparant els resultats obtinguts en cadascuna d'elles. També serà analitzada especialment la introducció de variables explicatives canviants en el temps, concretades en el nostre context en les variables de rendiment acadèmic mesurat mitjançant els indicadors de progressió acadèmica. Finalment, realitzarem una interpretació dels paràmetres estimats i obtindrem les funcions d'atzar i supervivència estimades.

8.1 Anàlisi Prèvia

Habitualment es realitza una anàlisi prèvia de les variables explicatives mitjançant instruments gràfics i contrastos estadístics no paramètrics basats en els rangs. L'objectiu és identificar aquelles variables més relacionades amb la durada per tal d'incloure-les en les modelitzacions alternatives posteriors. A més, permeten també captar les interrelacions existents entre les variables, les quals poden donar lloc a incloure interaccions en els models.

Per a cadascuna de les variables es realitza el gràfic de la funció de supervivència estimada per mètodes no paramètrics, la funció d'atzar i les transformades de la de supervivència $\log[S(t)]$ i $\log\{-\log[S(t)]\}$. Les tres primeres en front de la durada en anys i la darrera en front el logaritme dels anys. Cadascuna d'aquestes funcions es calcula per als diferents estrats que es deriven dels valors possibles de la variable explicativa considerada. D'aquesta forma es posa de manifest visualment si existeix una diferència significativa entre les funcions de supervivència (i les deduïdes d'aquesta) per a diferents valors possibles de la variable explicativa. Aquests gràfics també ens han de permetre d'una banda, plantejar hipòtesis sobre la distribució de probabilitat que pot seguir la durada per tal d'estimar models paramètrics i d'altra banda, verificar les hipòtesis que s'han de donar per a poder estimar els models d'atzar proporcional.

Al costat dels mètodes gràfics, es realitzen contrastos estadístics per a verificar la igualtat entre les funcions de supervivència: el contrast del log-rank i el de Wilcoxon generalitzat, els quals ens mesuren l'associació existent entre la variable explicativa i la durada dels estudis, per a cadascuna de les variables per separat i per a totes conjuntament.

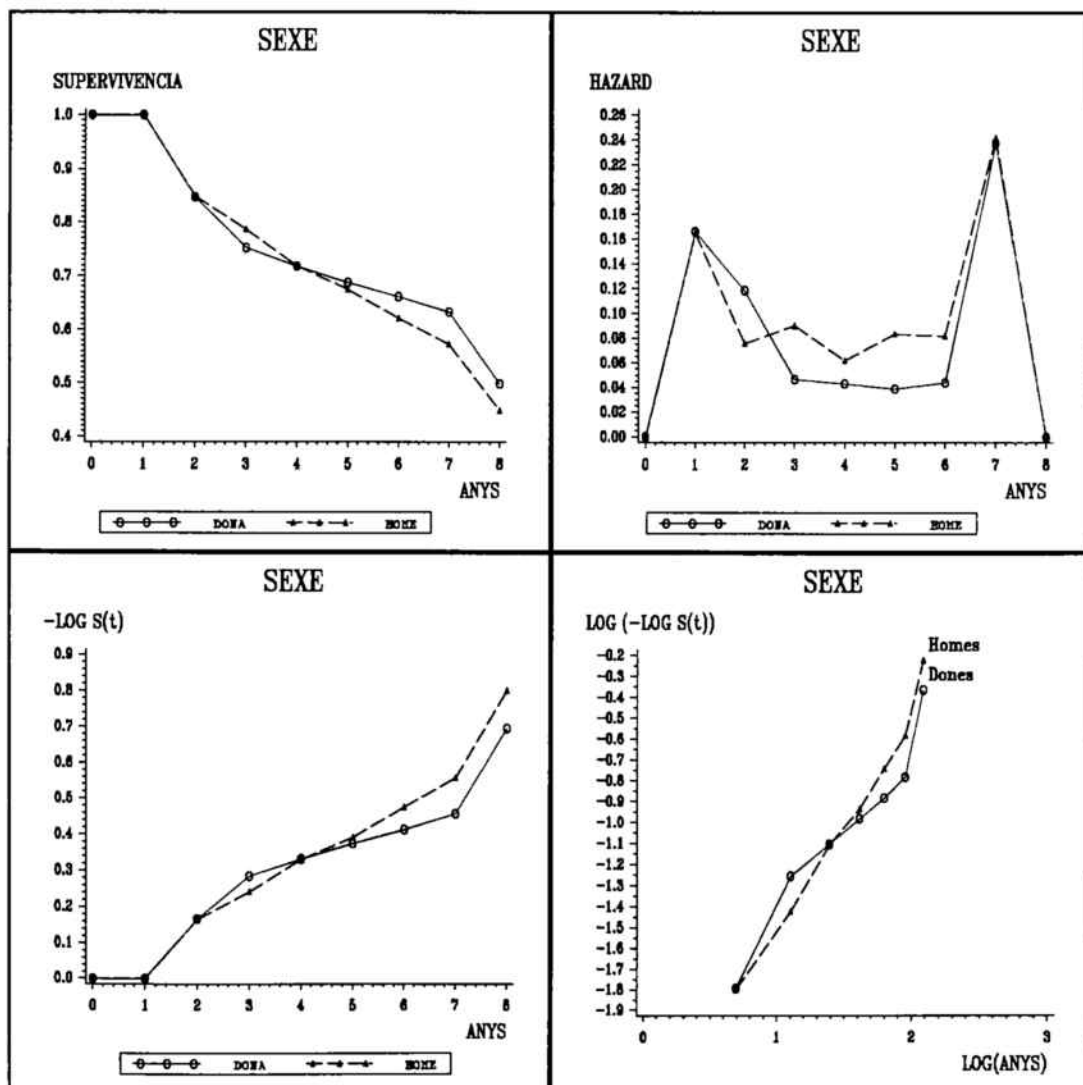
8.1.1 Sub població d'Abandonaments

8.1.1.1 Variables Personals

Per a la variable SEXE tant els gràfics com els contrastos estadístics mostren la no existència d'heterogeneïtat entre Homes i Dones pel que fa a la funció de supervivència. En concret, els contrastos estadístics de Wilcoxon i Log rank es distribueixen, sota la hipòtesi nul·la d'igualtat en les funcions de supervivència, segons una χ^2 amb un grau de llibertat i, essent 3.84 el valor crític de la mateixa al nivell de confiança del 5% a una sola cua, es verificaria el no rebuig de la hipòtesi d'igualtat de les dues funcions de supervivència corresponents a cadascuna de les modalitats de la variable SEXE, Homes i Dones. En conseqüència, estadísticament no existeix evidència empírica d'associació entre el SEXE i la durada dels estudis i, per tant, no existeix una tendència a l'abandonament diferenciada pels homes respecte les dones, o viceversa.

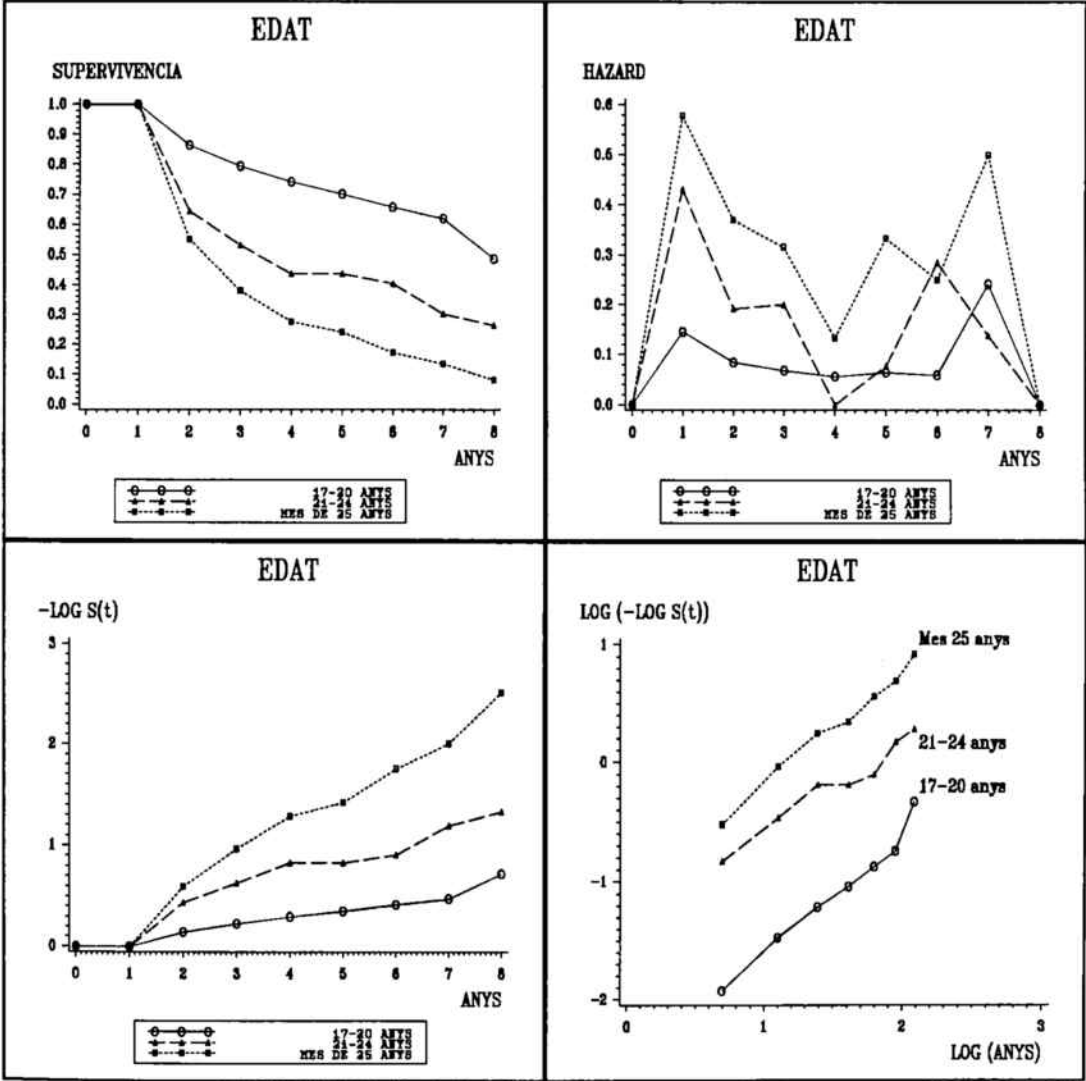
Malgrat tot, gràficament es posa de manifest com el col·lectiu de les dones en els primers anys d'estada a la universitat tenen una risc d'abandonar lleugerament més elevat que els homes per bé que posteriorment són els homes els que presenten unes probabilitats condicionals d'abandonar més elevades i en conseqüència durades més curtes fins a l'abandonament. Aquesta situació queda reflectida clarament en el gràfic de la funció d'atzar estimada ("Hazard") on s'observa com entre els anys tercer i sisè les corresponents funcions es mantenen al voltant d'un valor constant, el qual és inferior per les dones que pels homes. Cal tornar a destacar que l'actual anàlisi es realitza individualment per a cada variable sense tenir en compte les possibles interrelacions amb d'altres, les quals podrien posar de manifest associacions estadísticament significatives per a la variable sexe en combinació amb altres característiques personals o acadèmiques.

SEXE



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-11.1008	9.3690	1.4038	0.2361
LOG RANK	-17.9169	12.2625	2.1349	0.1440

EDAT

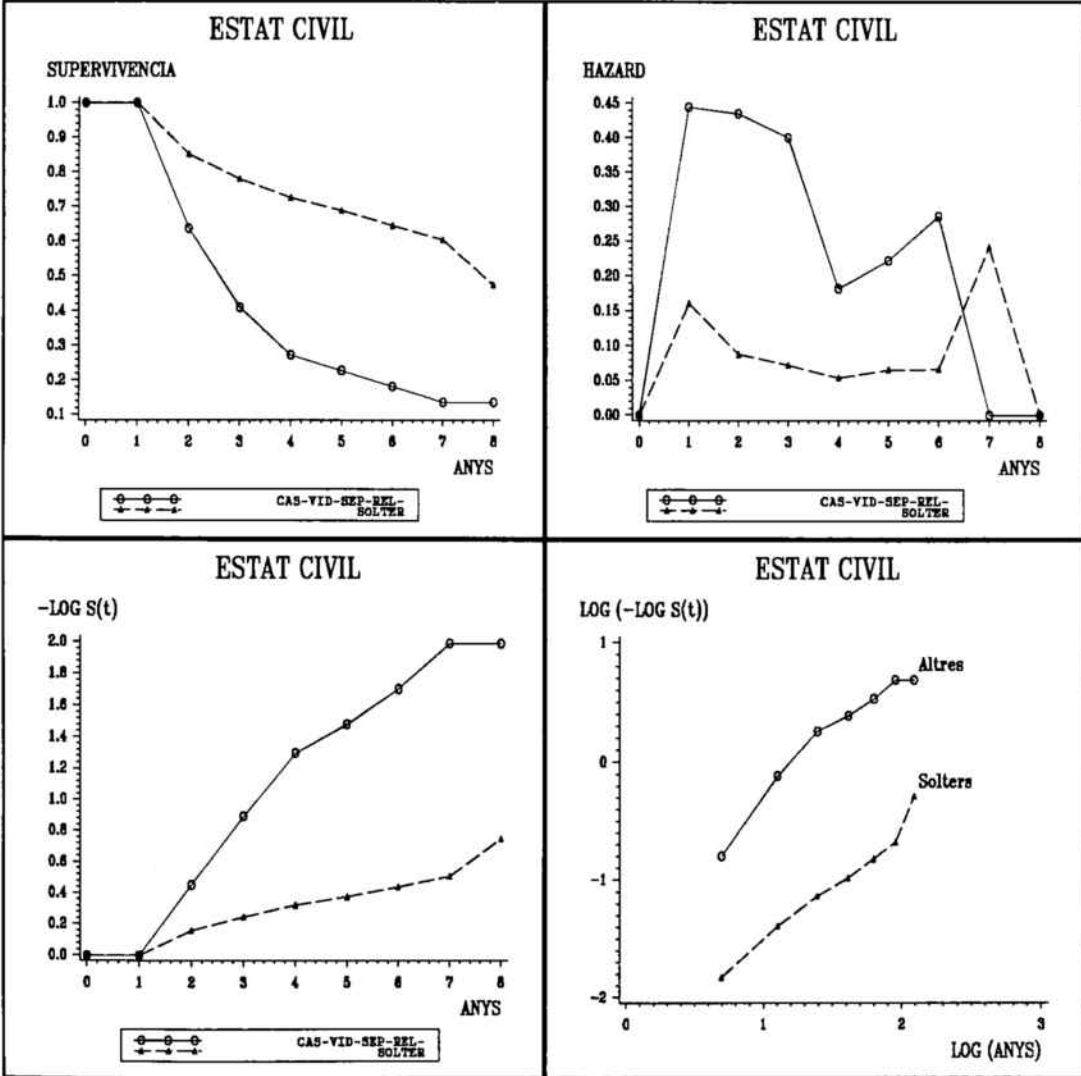


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-383.9	39.9376	92.4076	0.0001
LOG RANK	-435.7	43.5048	100.3	0.0001

En contraposició a l'anterior, per a l'EDAT els contrastos estadístics de Wilcoxon i Log rank permeten rebutjar estadísticament i de forma contundent la hipòtesi nul·la d'igualtat entre les funcions de supervivència. Així, els estadístics de contrast en ambdós casos són molt superiors al valor crític corresponent al nivell de confiança del 5% per a la distribució χ^2 amb un grau de llibertat, essent la probabilitat superior al valor de l'estadístic més petita que 0.0001.

Per a la construcció dels gràfics de les funcions de supervivència (i relacionades) per estrats segons els valors de la variable EDAT, aquesta ha estat categoritzada en les modalitats 17-20 ANYS, 21-24 ANYS i MÉS DE 25 ANYS. Efectivament en concordància amb els resultats dels contrastos estadístics, les funcions de supervivència són prou diferents per a cadascuna de les modalitats, essent el grup dels majors de 25 anys els que tenen una supervivència inferior i, per tant, durades més curtes fins a l'abandonament, mentre que el col·lectiu dels més joves, aquells en edats compreses entre els 17 i 20 anys els que abandonen més tard. Les probabilitats condicionals d'abandonar que mesuren les funcions d'atzar, en consonància amb les funcions de supervivència, també mostren, globalment, uns valors més baixos pels més joves que pels altres grups d'edats, tot i que puntualment en els anys 4 i 5 els valors corresponents als alumnes entre 21 i 24 anys d'edat són baixos i semblants als dels compresos entre 17 i 20. D'altra banda, la inspecció del gràfic de la funció d'atzar integrat en escala logarítmica, $\log\{-\log[S(t)]\}$ versus $(\log t)$, ens permetria acceptar la hipòtesi del model d'atzar proporcional segons la què els efectes de les variables sobre la funció d'atzar es mantenen constants al llarg del temps, la qual cosa en termes gràfics suposarà una evolució aproximadament paral·lela dels logaritmes dels atzars integrats per als diferents estrats.

ESTAT CIVIL

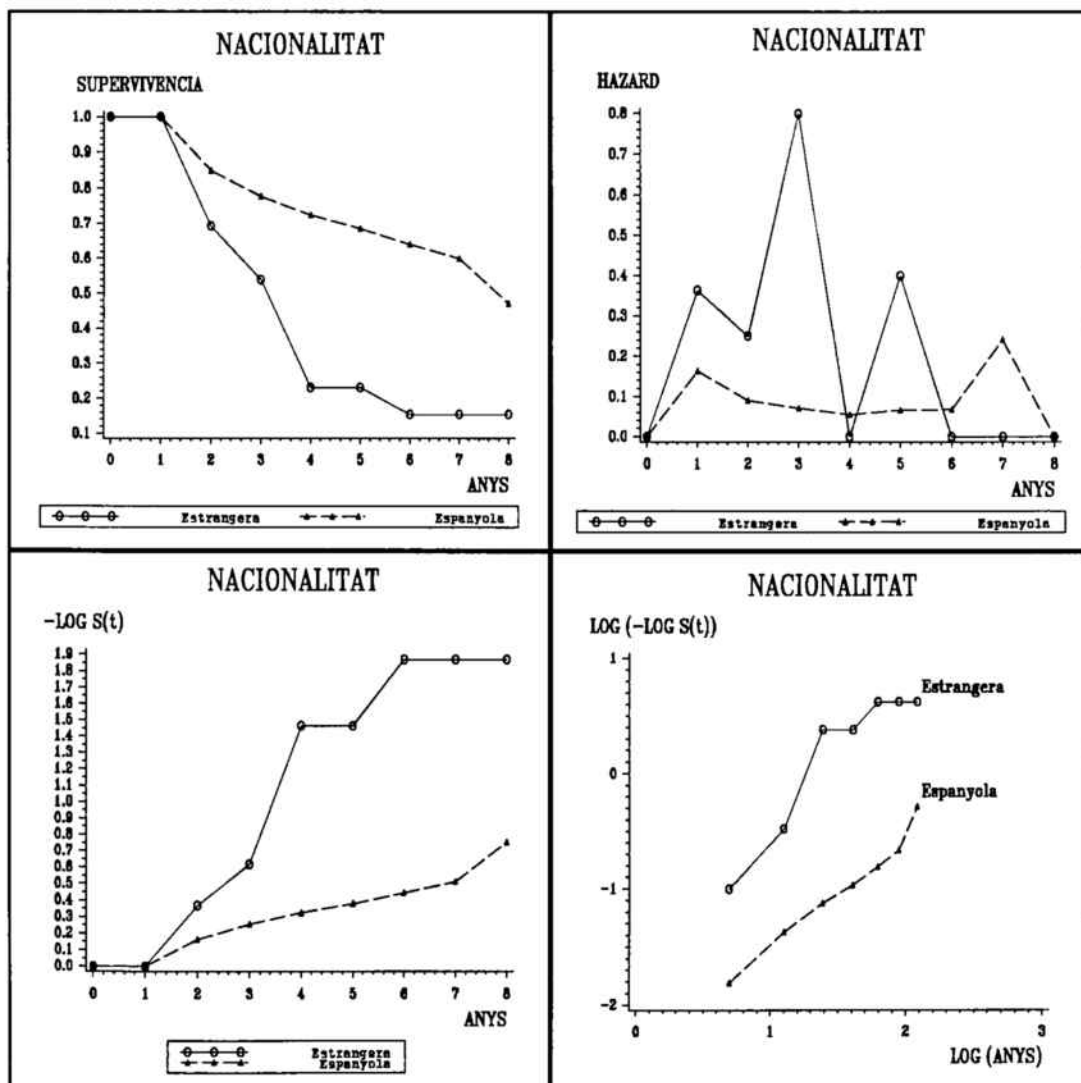


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	10.7515	2.3972	20.1153	0.0001
LOG RANK	12.3215	2.5681	23.0194	0.0001

Per a la variable ESTAT CIVIL s'observa un comportament clarament diferenciat en les funcions de supervivència per estrats. El col·lectiu dels casats, separats, divorciats o vidus presenta una supervivència molt inferior a la dels solters, essent les variacions de l'una a l'altra més elevades en els primers anys d'estada a la universitat. Aquesta darrera circumstància es fa evident en el gràfic de l'atzar integrat, $-\log[S(t)]$, envers el temps, pel qual la funció corresponent als no solters creix amb una pendent molt més accentuada que no pas la de l'estrat dels solters. En conseqüència, les durades fins a l'abandonament són extremadament més curtes pels no solters, els quals tindrien una durada mediana entre els 2-3 anys mentre que pels solters aquella es situaria per sobre dels 8 anys, és a dir, pel col·lectiu dels no solters més del 50% abandonen els estudis en els tres primers anys d'estada a la universitat. La durada mediana podria estimar-se per interpolació lineal mitjançant el càlcul del temps al qual li correspondria un valor del 50% en la funció de supervivència.

Les conclusions anteriors deduïdes de la inspecció dels gràfics es veuen reafirmades pels resultats dels contrastos d'associació entre la variable ESTAT CIVIL i la durada dels estudis, segons els quals es posa de manifest el rebuig de la hipòtesi nul·la d'igualtat en les funcions de supervivència al nivell de confiança del 5% i, per tant, la variable ESTAT CIVIL recull una part de l'heterogeneïtat existent en la supervivència de la cohort analitzada. A més, la variable podrà ser incorporada com a explicativa en els models d'atzar proporcional vist que el gràfic del logaritme de l'atzar integrat mostra una separació aproximadament constant entre les funcions pels dos estrats considerats.

NACIONALITAT

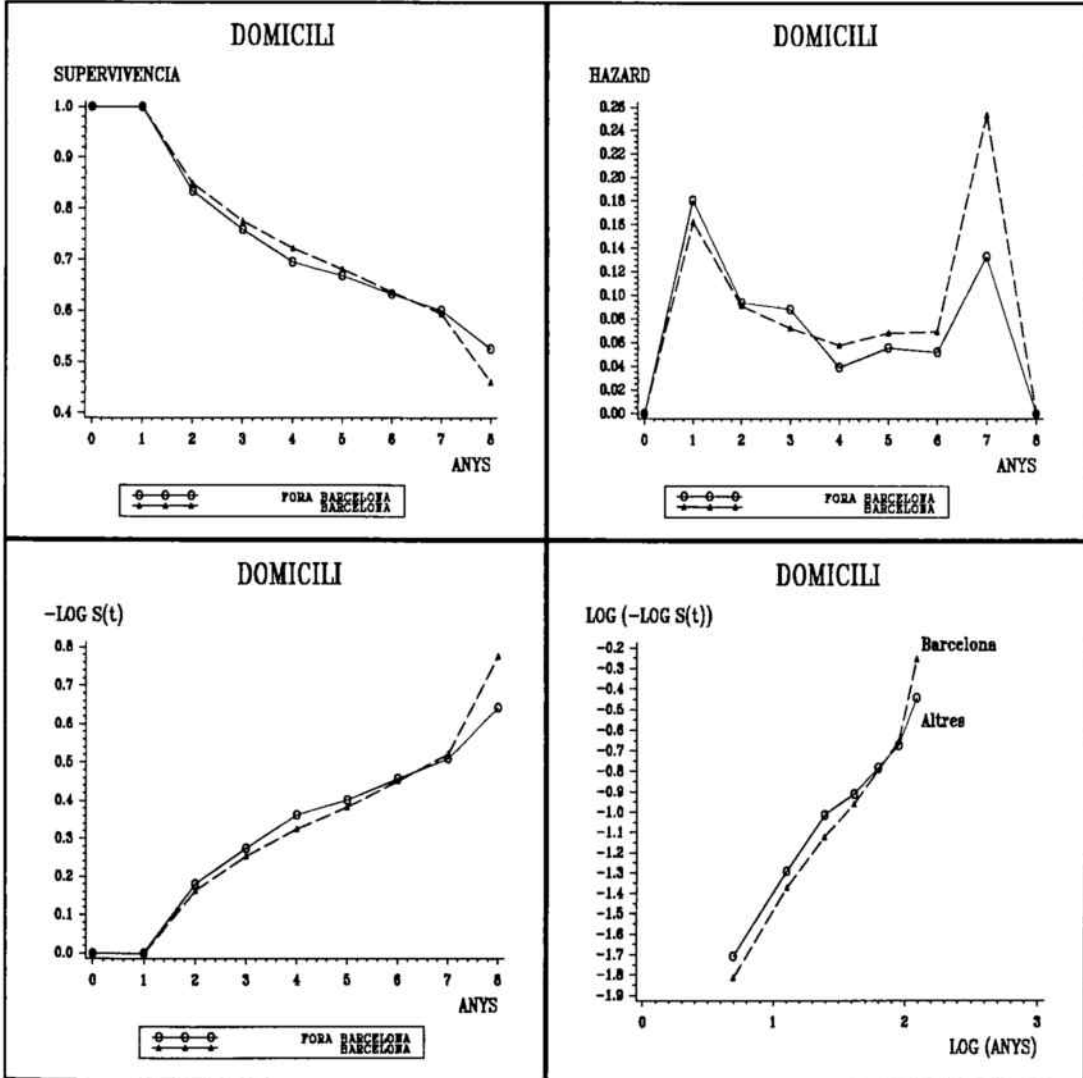


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	5.8754	1.8928	9.6353	0.0019
LOG RANK	6.8729	2.0238	11.5335	0.0007

Amb la variable NACIONALITAT es posa de manifest com els estrangers abandonen de manera més clara i prematura els estudis, estant les funcions de supervivència molt allunyades per a cadascun dels estrats. En aquest cas però, el salt més important es produeix en el tercer any, moment en el qual la funció d'atzar corresponent als estrangers pren un valor molt elevat. Malgrat tot, aquesta funció d'atzar presenta un comportament poc estable en el temps degut principalment al reduït nombre d'observacions incloses en aquest estrat la qual cosa fa que, en el moment que algun estudiant del col·lectiu abandona, l'efecte en termes relatius és molt gran i augmenta la probabilitat condicional d'abandonar. Pel col·lectiu dels alumnes de nacionalitat espanyola la funció d'atzar estimada per mètodes no paramètrics presenta un comportament molt semblant al del conjunt de la població amb un decreixement lent a partir dels segon any i una alça en el darrer any analitzat deguda a la informació incompleta disponible pel curs 1992-93.

Els resultats dels contrastos estadístics són també clars respecte la necessitat d'incloure la variable NACIONALITAT per tal de caracteritzar o recollir part de l'heterogeneïtat de les funcions de supervivència en la cohort de referència. La hipòtesi nul·la d'absència d'associació és rebutjada al nivell de confiança del 5% tot i que no de forma tan contundent com en el cas de l'EDAT o l'ESTAT CIVIL. Novament aquest fet estaria relacionat amb la poca variabilitat existent per a la característica personal NACIONALITAT doncs només l'1% del total d'estudiants pertanyen al grup dels estrangers. Tot i això, els resultats són prou significatius com per justificar la inclusió de la variable en l'anàlisi i, a més, podrà ser incorporada en un model d'atzar proporcional vist que el logaritme de les funcions d'atzar integrat evolucionen aproximadament en paral·lel.

DOMICILI



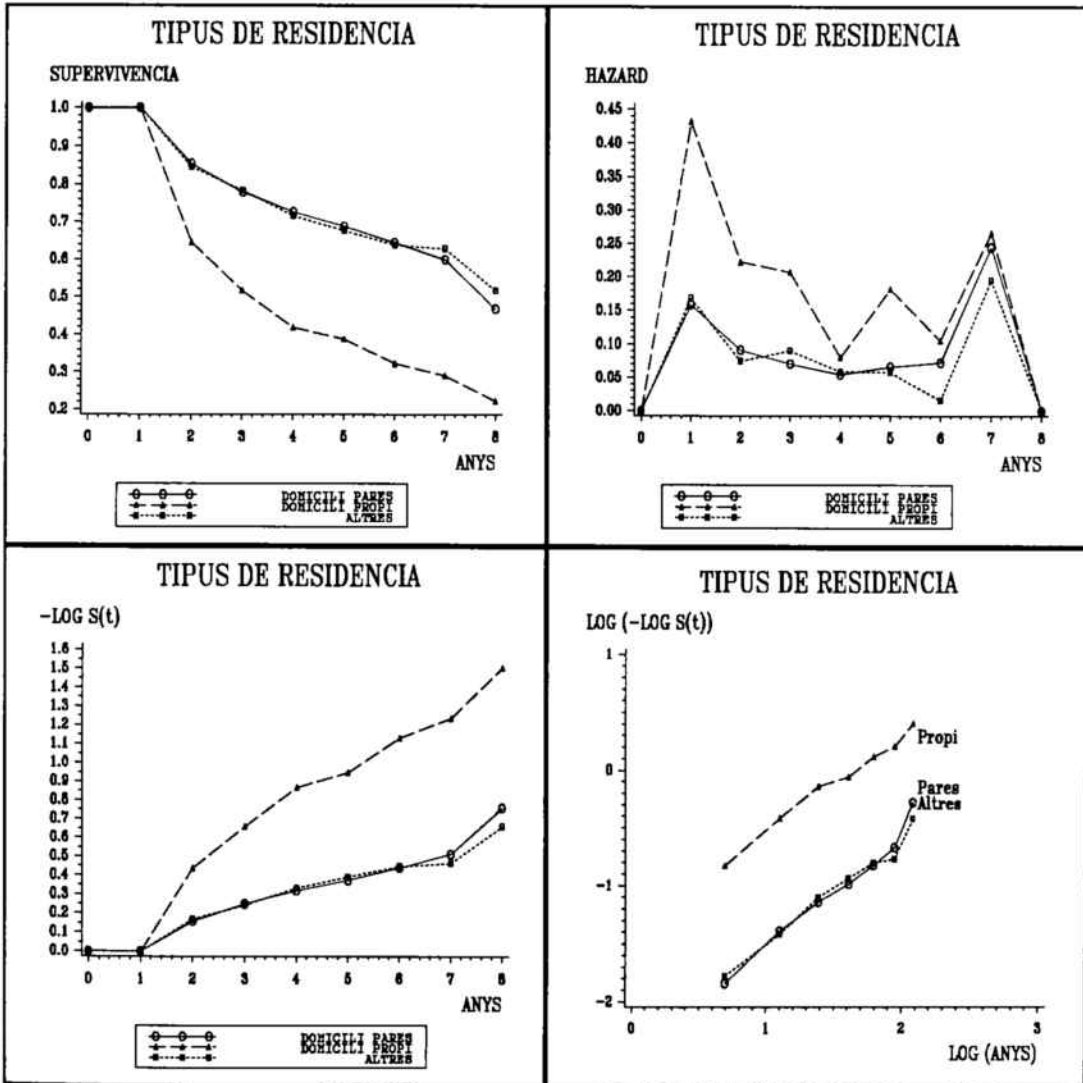
Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-2.4032	6.5573	0.1343	0.7140
LOG RANK	-5.6661	8.5722	0.4369	0.5086

En la variable DOMICILI, la qual recull per una banda els individus amb domicili a Barcelona ciutat i el seu cinturó industrial i de l'altra la resta d'alumnes amb un domicili més llunyà, s'observa un comportament semblant al de la variable SEXE analitzada anteriorment. Així, tant les representacions gràfiques de les funcions de supervivència per estrats i de les funcions derivades de la mateixa com els contrastos estadístics d'associació, permeten afirmar que no existeixen diferències significatives entre ambdós grups d'estudiants.

Els contrastos estadístics permeten no rebutjar la hipòtesi d'igualtat en les funcions de supervivència al nivell de confiança del 5%, prenent l'estadístic de contrast χ^2 uns valors molt propers a zero tant pel contrast de Wilcoxon (0.1343) com pel del log rank (0.4369). Clarament aquests valors de l'estadístic es troben molt per sota del valor crític de la distribució amb un grau de llibertat que és 3.84.

Sobre aquesta variable cal dir que pot presentar petites distorsions en la informació continguda com a conseqüència de la resposta de l'alumne en l'imprès de matrícula. Així, aquells estudiants amb un domicili familiar lluny de l'entorn de Barcelona ciutat però que hi resideixen durant el curs en aquesta darrera no sempre omplen de forma homogènia el codi corresponent al domicili. Per aquest motiu la variable, que en principi hauria de recollir part de l'heterogeneïtat de la població pel que fa a les seves durades fins a l'abandonament, no sembla que sigui prou rellevant, si més no individualment, com per a incloure-la en les modelitzacions posteriors. Caldrà, però, atendre a les possibles interaccions amb altres variables abans d'eliminar-la totalment de l'anàlisi.

TIPUS DE RESIDÈNCIA

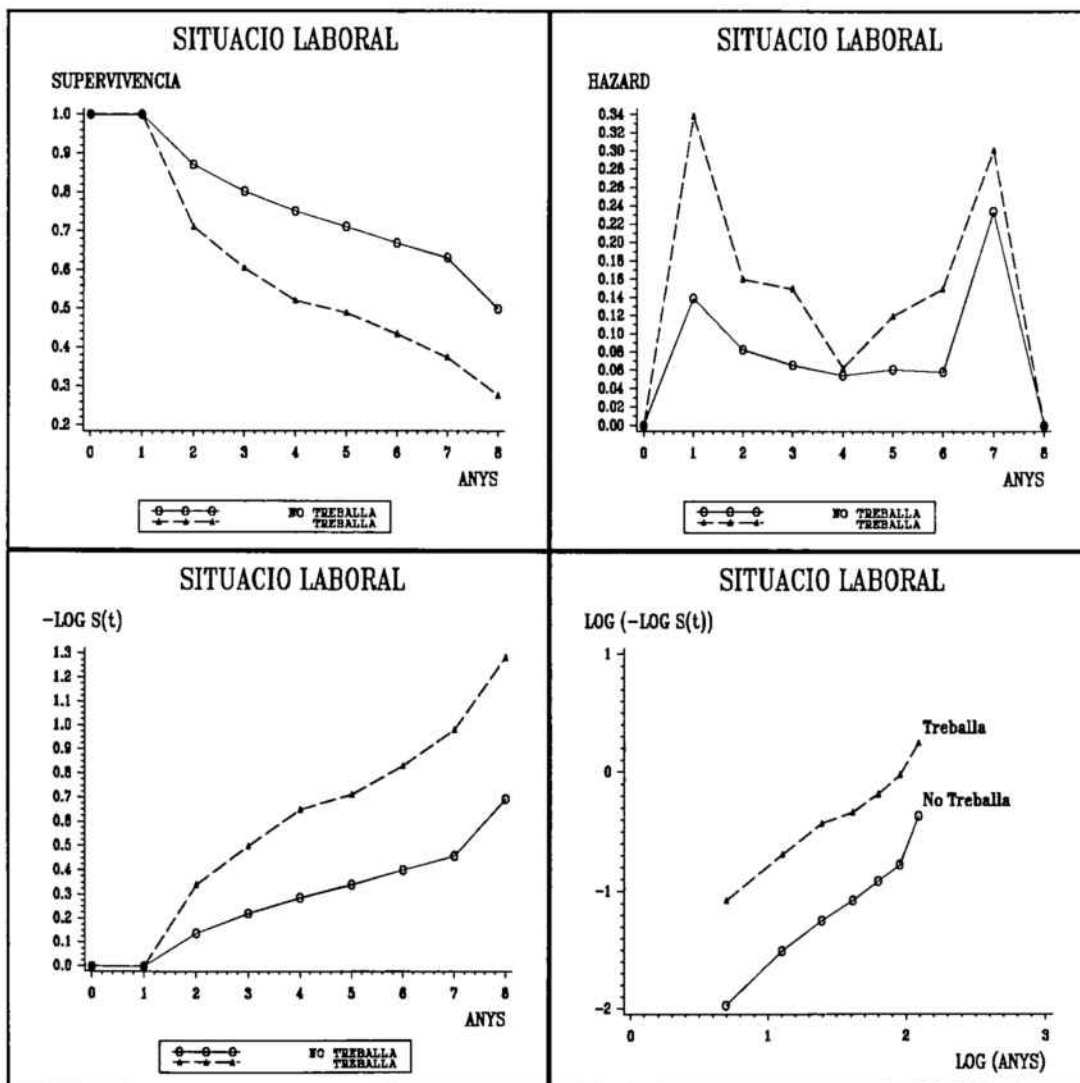


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON			15.1792	0.0004
Domicili Pares	7.2101	6.5018	1.2297	0.2675
Domicili Propi	-11.0813	2.8633	14.9783	0.0001
Altres Domicilis	3.8712	5.9865	0.4182	0.5179
LOG RANK			16.5342	0.0003
Domicili Pares	6.4308	8.3950	0.5868	0.4437
Domicili Propi	-12.5808	3.3478	14.1223	0.0002
Altres Domicilis	6.1500	7.8565	0.6128	0.4337

La variable TIPUS DE RESIDÈNCIA ha estat caracteritzada a partir de tres modalitats alternatives: DOMICILI PARES, la qual inclou també aquells alumnes que resideixen en el domicili d'altres familiars, DOMICILI PROPI i ALTRES (pisos d'estudiants, residències universitàries, etc). Les funcions de supervivència per a les modalitats DOMICILI PARES i ALTRES es superposen de tal forma que no és possible distingir comportaments diferents per aquest dos grups exceptuant els darrers anys (a partir del sisè) pels quals el grup dels que resideixen amb els seus pares o altres familiars tenen una probabilitat condicional d'abandonar o valor de la funció d'atzar lleugerament més elevada. El tercer grup restant, els alumnes que resideixen en domicili propi, presenten una supervivència molt més reduïda que els anteriors amb funció d'atzar més elevada, especialment els primers anys d'estada a la universitat, i atzar integrat amb una pendent molt més accentuada.

Els contrastos estadístics són ara dobles. En primer lloc es realitza un contrast global per a verificar l'associació de la variable TIPUS DE RESIDÈNCIA globalment considerada amb la durada dels estudis, el qual resulta clarament significatiu al nivell de confiança del 5%, és a dir, es rebutjaria la hipòtesi nul·la d'igualtat en les funcions de supervivència per estrats. En segon lloc el contrast s'efectua per a cadascuna de les tres modalitats de la variable per separat, resultant únicament significativa la del domicili propi vist que el seu comportament es troba molt diferenciat de les altres dues modalitats. Per contra, les categories Domicili Propi i Altres, considerades individualment, malgrat ser prou diferenciades cadascuna d'elles de la categoria Domicili Propi, pel fet de tenir comportaments molt homogenis entre elles es derivaria un no rebuig de la hipòtesi nul·la del contrast d'igualtat en les funcions de supervivència.

SITUACIÓ LABORAL



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-47.3511	6.7354	49.4229	0.0001
LOG RANK	-54.9243	8.1810	45.0727	0.0001

La variable SITUACIÓ LABORAL, tal com era previsible, és rellevant en la caracterització de l'heterogeneïtat de la cohort d'estudiants analitzada pel que fa a la durada dels estudis fins a l'abandonament. Els alumnes que declaren treballar, ja sigui a temps complet com parcial, tenen unes durades més curtes que els que declaren no treballar. Els valors de la funció d'atzar són més elevats en tots els anys acadèmics, excepte el quart, pel col·lectiu d'estudiants integrats en el mercat laboral, així com també la funció d'atzar integrat, $-\log[S(t)]$, presenta una tendència creixent molt més accentuada per aquests darrers. D'altra banda, la variable podrà ser incorporada en la modelització proporcional de l'atzar vist que l'evolució del logaritme de les funcions d'atzar integrat pels dos subcol·lectius és clarament paral·lela al llarg del temps, si més no en els sis primers anys estudiats.

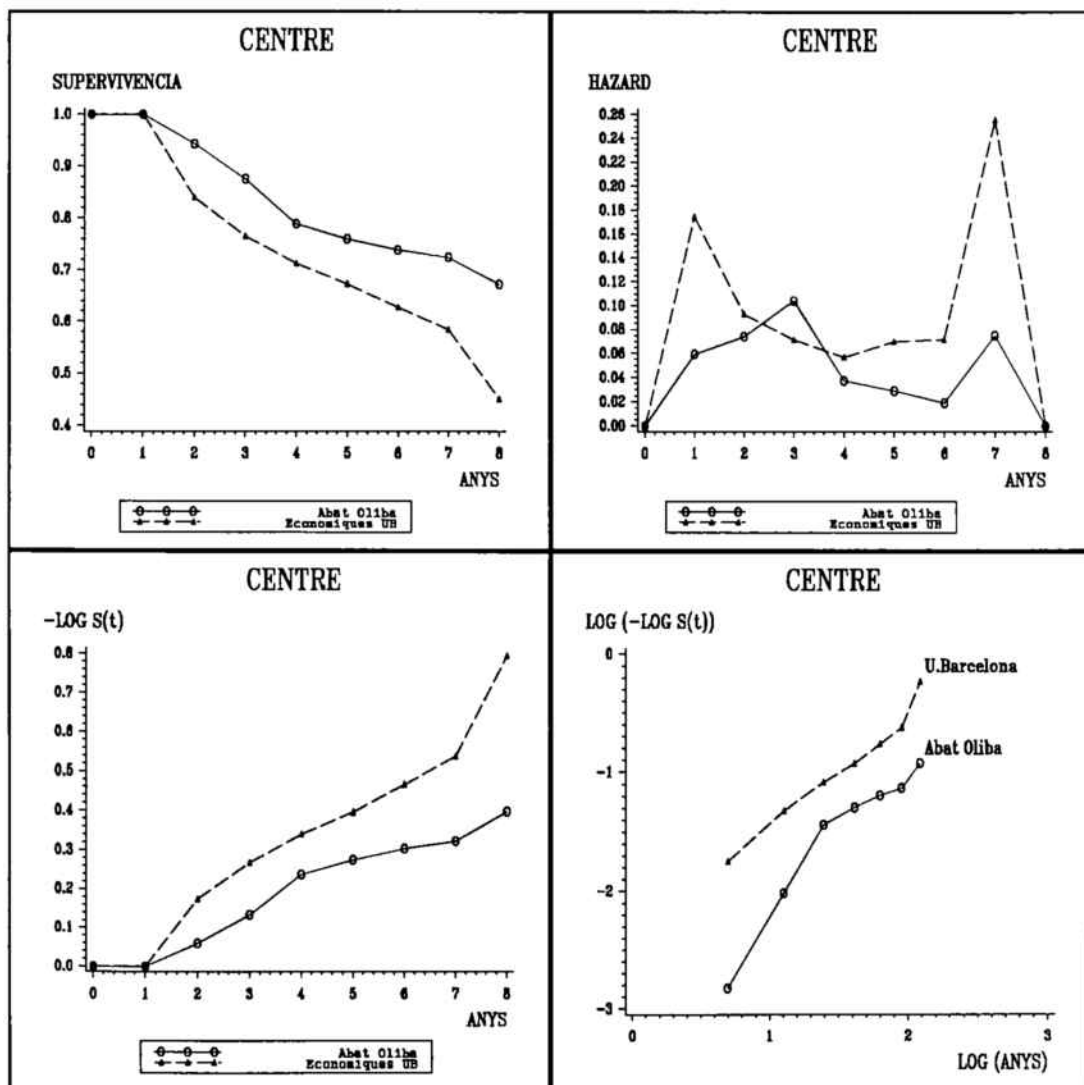
Els contrastos estadístics de Wilcoxon i Log rank permeten rebutjar la hipòtesi nul·la d'igualtat en les funcions de supervivència al nivell de significació del 5% i, com a conseqüència, queda garantida estadísticament l'associació entre la situació laboral considerada individualment i la durada dels estudis. Malgrat tot, la variable situació laboral es troba molt lligada a d'altres variables personals i de situació acadèmica de l'alumne com ara l'Edat, l'Estat Civil, la Font de Finançament dels estudis, l'Horari de classes, etc. per la qual cosa en un contrast conjunt amb aquestes caldrà verificar si la redundància en la informació proporcionada per les mateixes és prou gran com per que la variable no hagi d'intervenir en l'anàlisi. En qualsevol cas, es posa de manifest com els alumnes que duen a terme una activitat laboral a banda dels estudis universitaris tenen una tendència o risc superior d'abandonament.

8.1.1.2. Variables de Gestió Acadèmica

La variable CENTRE divideix la cohort en dos grups segons els alumnes hagin cursat el primer cicle de la llicenciatura en el centre adscrit Abat Oliba o bé a la Facultat d'Econòmiques. En aquest cas, el nombre d'alumnes del centre adscrit és reduït i no arriba al 8% del total, la qual cosa provoca determinades inestabilitats en els valors estimats per a les funcions de supervivència, atzar, etc. Tot i això, la representació gràfica de les funcions de supervivència mostren unes durades més llargues fins a l'abandonament entre els estudiants del centre adscrit Abat Oliba per bé que els valors de la funció d'atzar d'aquest col·lectiu no han estat sempre per sota dels corresponents als alumnes de la Facultat. En aquest sentit, cal destacar com, pels estudiants de la Facultat, en el primer any d'estada a la universitat l'abandonament és molt elevat i posteriorment es va estabilitzant fins arribar novament a valors alts en el setè any. Contràriament, entre els estudiants del centre adscrit Abat Oliba, l'abandonament va creixent des del primer any fins el tercer i és un cop superat aquest curs quan la funció d'atzar comença a experimentar una baixada. Aquest comportament contradictori és encara més destacable pel fet que en el quart curs acadèmic el percentatge d'alumnes que han abandonat gairebé seria el mateix pels dos centres.

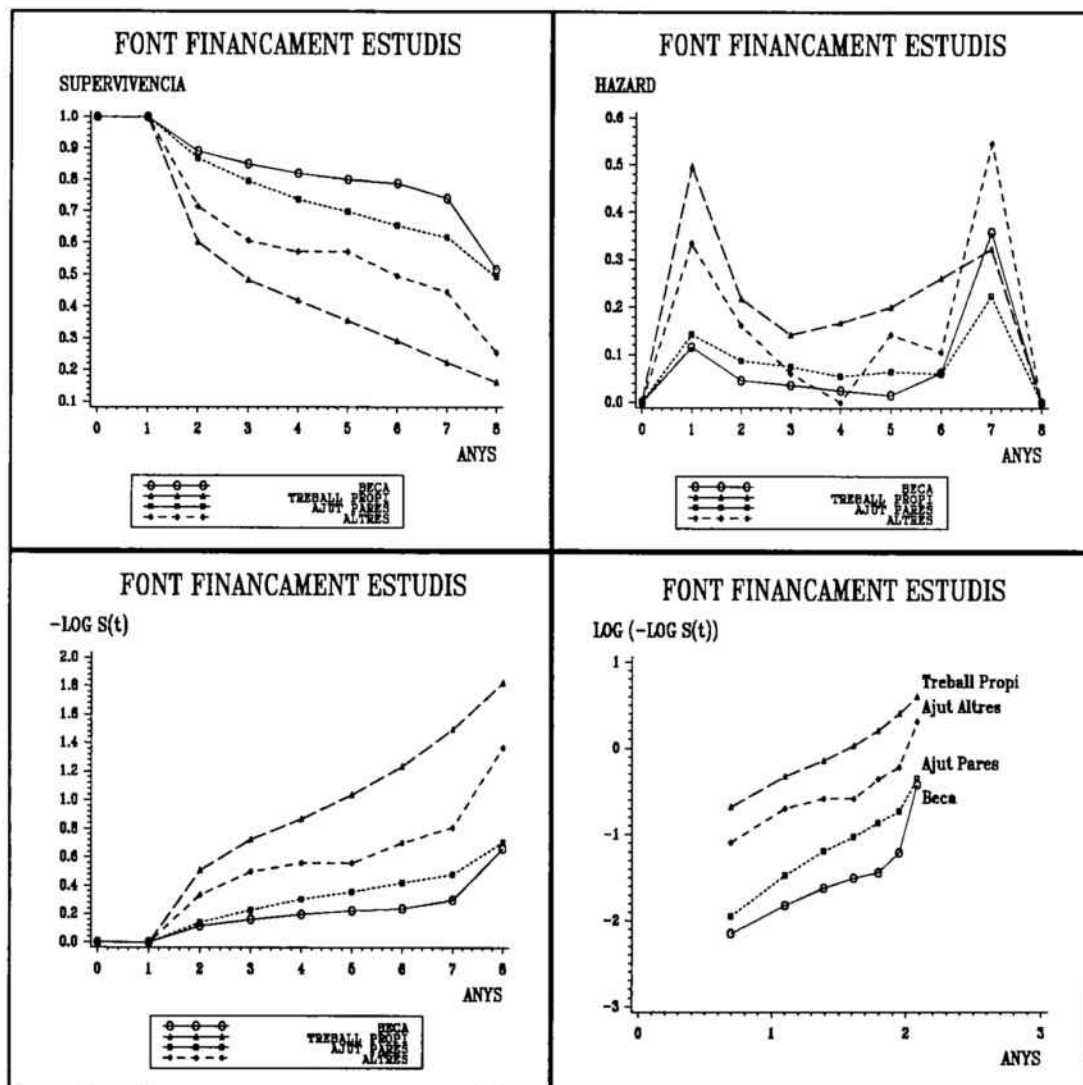
Els contrastos estadístics permeten rebutjar la hipòtesi d'igualtat en les funcions de supervivència i, en conseqüència, determinarien una associació entre la durada dels estudis i el centre on es cursa el primer cicle. Caldrà doncs incorporar la variable en les modelitzacions posteriors tot i que ara la hipòtesi de proporcionalitat de l'atzar no es verificaria de forma clara disposant únicament dels gràfics del logaritmes de les funcions d'atzar integrat vist que la distància que separa les corresponents a ambdós grups d'estudiants no es manté constant.

CENTRE



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-17.4039	5.1134	11.5845	0.0007
LOG RANK	-21.7253	6.9574	9.7508	0.0018

FONT DE FINANÇAMENT DELS ESTUDIS

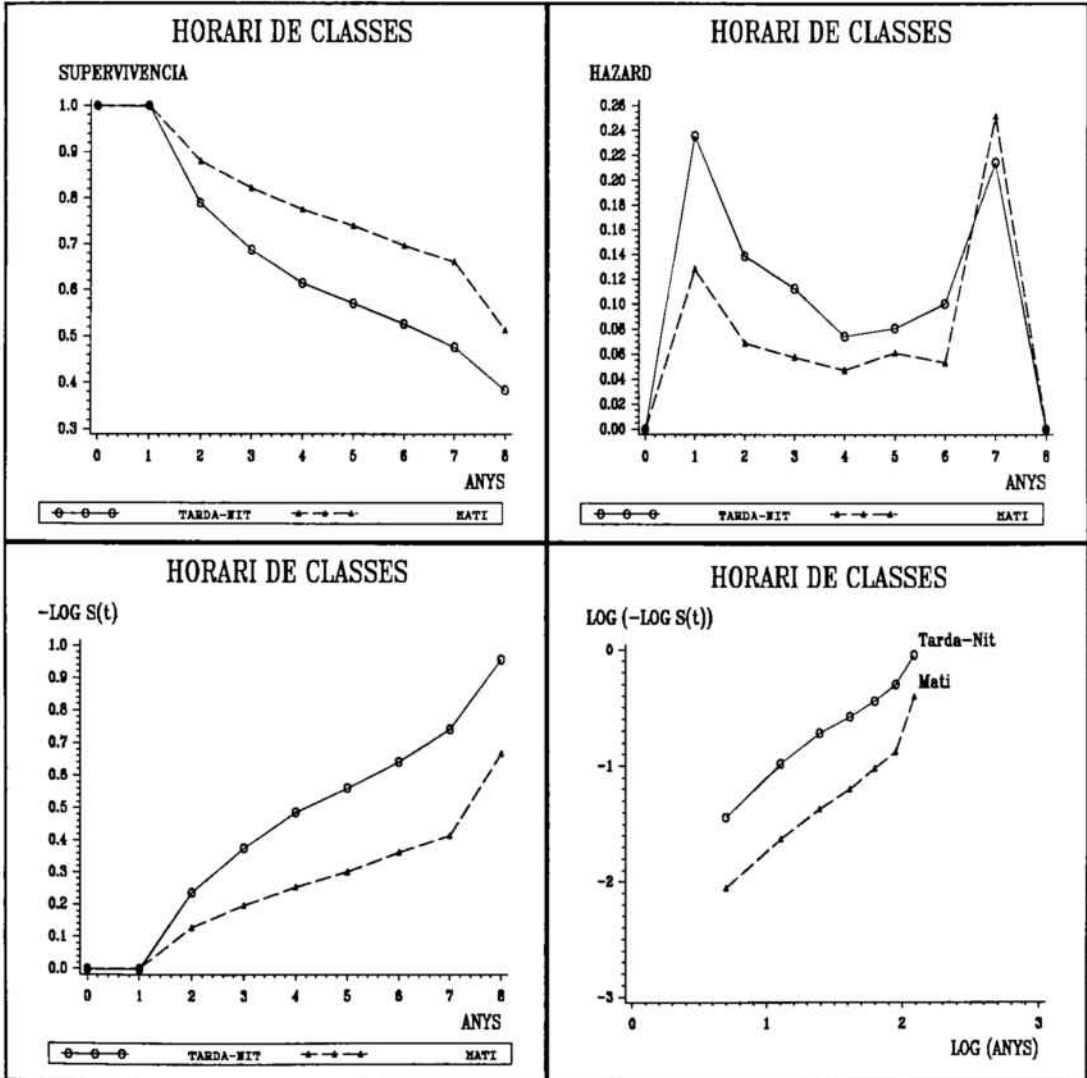


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON			74.6130	0.0001
Ajut Pares	32.8498	7.1425	21.1526	0.0001
Beca	12.1959	5.0852	5.7518	0.0165
Treball Propi	-39.1895	4.8372	65.6364	0.0001
Altres	-5.8561	2.7398	4.5686	0.0326
LOG RANK			85.9983	0.0001
Ajut Pares	38.0233	8.9114	18.2042	0.0001
Beca	14.0651	6.7946	4.2851	0.0384
Treball Propi	-45.2481	5.4825	68.1161	0.0001
Altres	-6.8403	3.3108	4.2687	0.0388

Les representacions gràfiques de les funcions de supervivència per a cadascun dels quatre estrats en que la cohort d'estudiants analitzada ha estat dividida segons els valors de la variable FONT DE FINANÇAMENT DELS ESTUDIS mostren com aquesta es troba associada a la durada dels estudis. Així, Beca, Ajut Pares, Ajut Altres i Treball Propi tenen respectivament durades cada cop més curtes fins a l'abandonament. D'altra banda, les modalitats Beca i Ajut dels Pares d'una banda, i d'altra Ajut Altres amb Treball Propi es troben lleugerament interrelacionades, és a dir, tenen comportament semblants entre les mateixes i, per exemple, les dues primeres en la funció d'atzar mostren un valor elevat en el primer curs acadèmic que posteriorment va reduint-se fins al setè any en que torna a créixer, mentre que les altres dues presenten evolucions més inestables.

Altrament, dels contrastos estadístics s'obtenen uns resultats coherents amb els anteriors i tant els dos contrastos globals com els marginals per a cadascuna de les modalitats permeten rebutjar la hipòtesi d'igualtat en les funcions de supervivència. Els estadístics de contrast per a totes les modalitats conjuntament evidencien una forta associació entre la variable Font de Finançament i la durada dels estudis. En els contrastos individuals per a cadascuna de les modalitats, els corresponents estadístics de Beca i Ajut Altres tenen nivells de significació que, tot i ser inferiors al 5%, posarien de manifest com aquestes dues modalitats estan interrelacionades amb les altres dues. Malgrat tot, cal notar les diferències existents entre el finançament via beques, el qual va lligat a un bon rendiment acadèmic i en conseqüència a una menor tendència a l'abandonament, del finançament via ajut dels pares en el qual queden inclosos un col·lectiu d'estudiants amb connotacions diferents front la decisió d'abandonar.

HORARI



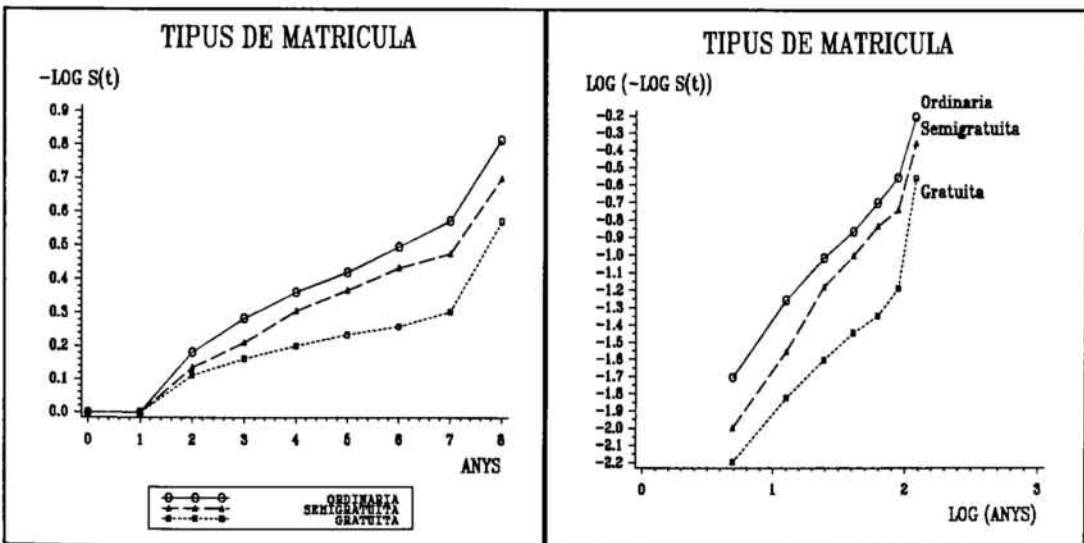
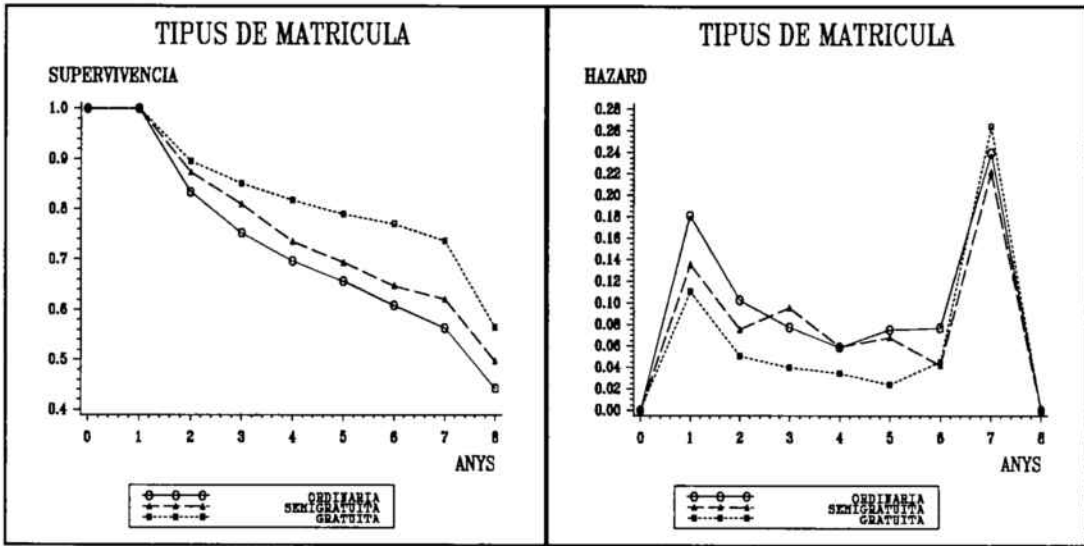
Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	58.5089	9.3131	39.4687	0.0001
LOG RANK	67.1310	11.8857	31.9003	0.0001

Per a la variable HORARI DE CLASSES tenim un comportament semblant al d'altres variables analitzades anteriorment. Els gràfics de les funcions de supervivència estimades per mètodes no paramètrics en cadascun dels dos grups, Matí i Tarda-Nit, mostren com els alumnes d'aquest darrer grup tenen durades dels estudis més curtes i una major tendència l'abandonament. En aquest cas les funcions d'atzar de cadascun dels dos grups tenen formes molt semblants amb valors alts en el primer curs acadèmic i després una evolució al voltant d'un valor més baix que finalment creix en el darrer any acadèmic. La forma de les funcions d'atzar és semblant però els valors concrets són força més elevats pel grup d'horari Tarda-Nit.

Els contrastos estadístics permeten no rebutjar la hipòtesi d'absència d'associació entre la durada dels estudis i l'horari de classes al nivell de confiança del 5%, essent els valors dels estadístics de contrast molt superiors al valor crític en taules per a la distribució χ^2 amb un grau de llibertat, el qual al 5% de significació és 3.84. Caldrà doncs incorporar la variable horari en les modelitzacions posteriors, verificant-se a més la hipòtesi de proporcionalitat de l'atzar si més no a la vista de la informació que proporciona el gràfic del logaritme de la funció d'atzar integrat.

La variable HORARI també es troba molt relacionada amb altres variables personals i de gestió acadèmica com ara l'Edat, la Situació Laboral, l'Estat Civil, etc., donat que els alumnes de més edat que treballen i no són solters integraran majoritàriament el grup de Tarda-Nit, per la qual cosa caldrà comprovar la seva significació en contrastos d'associació realitzats conjuntament amb altres variables.

CLASSE DE MATRÍCULA



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON			15.8794	0.0004
Ordinària	-30.4880	8.6941	12.2973	0.0005
Semigratuïta	6.3101	6.8220	0.8556	0.3550
Gratuïta	24.1779	6.5647	13.5645	0.0002
LOG RANK			14.5141	0.0007
Ordinària	-36.3085	11.4543	10.0508	0.0015
Semigratuïta	7.1709	8.9273	0.6452	0.4218
Gratuïta	29.1376	8.8031	10.9557	0.0009

La variable CLASSE DE MATRÍCULA està formada per les categories Matrícula Ordinària, Matrícula Semigratuïta i Matrícula Gratuïta les quals en el seu conjunt i segons els resultats dels contrastos estadístics es troben associades a la durada dels estudis. Com a conseqüència dels contrastos de Wilcoxon i Log rank conjunts es verifica la hipòtesi d'associació a un nivell de significació del 5%. Malgrat tot, individualment considerades no totes les modalitats són significatives estadísticament sinó que per la Matrícula Semigratuïta obtenim un estadístic de contrast de valor molt proper a zero. Aquesta situació és deguda, com pot observar-se en els gràfics de les funcions de supervivència i derivades de la mateixa, a que la categoria Semigratuïta té un comportament molt semblant i proper al de la Ordinària tot i que no són exactament iguals.

Alternativament, els alumnes amb Matrícula Gratuïta tenen una funció de supervivència estimada superior a la de les altres dues modalitats i, per tant, unes durades més curtes i menor tendència a abandonar els estudis. Cal dir que dins del col·lectiu de matrícula gratuïta es troben inclosos aquells alumnes que gaudeixen d'una beca a més dels Funcionaris i fills d'aquests o membres de famílies nombroses, per la qual cosa la variable pot trobar-se interrelacionada amb la que recollia la Font de Finançament dels estudis, el que caldrà comprovar mitjançant la utilització dels contrastos d'associació conjunts.

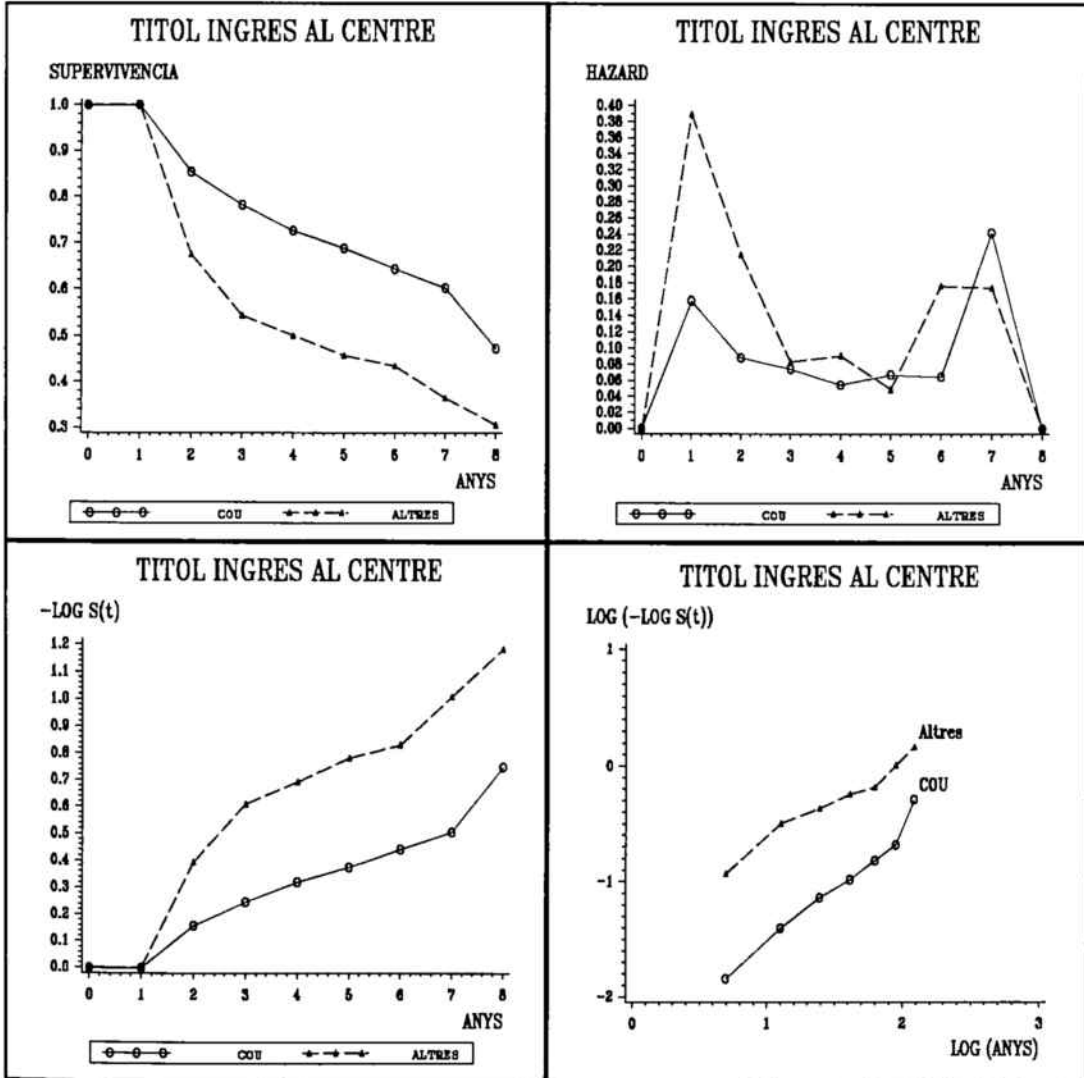
La representació gràfica del logaritme de la funció d'atzar integrat mostra com de la consideració conjunta de la matrícula ordinària amb la semigratuïta es derivaria que evolucionen aproximadament en paral·lel respecte la matrícula gratuïta, per la qual cosa serà d'aplicació el model d'atzar proporcional incorporant aquesta variable.

8.1.1.3 Variables Acadèmiques Prèvies

La variable TÍTOL D'ACCÉS es troba molt relacionada amb la dicotomia existent entre dos grans grups d'alumnes en la cohort, d'una banda els més joves provinents del COU i d'altra els d'edats més avançades els quals han accedit per altra via: Majors de 25 anys, altres titulacions, etc. En conseqüència, el comportament de les dues modalitats d'accés considerades en termes de les funcions de supervivència és molt diferenciat tal com s'havia produït en el cas, per exemple, de l'Edat o l'Estat Civil. Així, observem com els alumnes provinents del COU tenen unes durades més llargues fins a l'abandonament amb una durada mediana estimada superior als set anys, mentre la resta d'estudiants abandonen abans amb una durada mediana entre 3-4 anys. Cal destacar com els valors de la funció d'atzar pels dos col·lectius són aproximadament iguals en els anys centrals de la durada (del 3er al 6è), essent els abandonaments més probables en els dos primers anys pels alumnes amb accés diferent del COU.

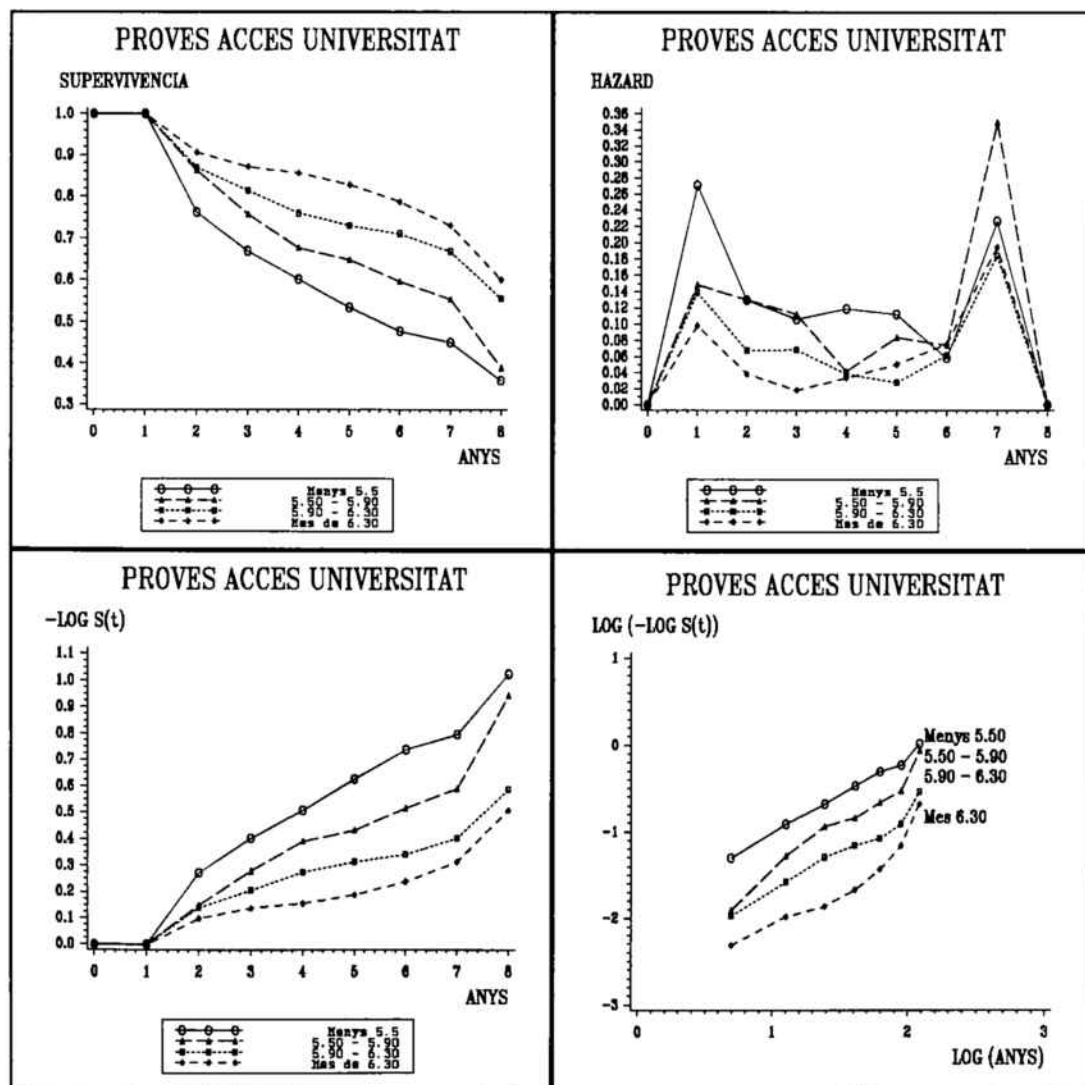
Mitjançant els contrastos estadístics es verifica la hipòtesis de forta dependència o associació entre les durades dels estudis i la titulació d'accés, trobant-se l'estadístic de contrast en ambdós casos per sobre del valor crític per a la distribució amb un grau de llibertat. La introducció de la variable dictòmica en les posteriors modelitzacions es troba doncs justificada tot i que cal tenir en compte les possibles interrelacions amb altres variables tal com ha estat comentat anteriorment i, per tant, caldrà comprovar els resultats dels contrastos conjunts. De la representació gràfica del logaritme de la funció d'atzar integrat pot deduir-se la conveniència de la modelització mitjançant la hipòtesis de l'atzar proporcional vist que la distància que separa ambdues funcions és aproximadament constant al llarg de tot el període analitzat, exceptuant els dos darrers anys.

TÍTOL D'ACCÉS



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-11.9877	3.4150	12.3225	0.0004
LOG RANK	-13.0209	4.1792	9.7074	0.0018

PROVES D'ACCÉS A LA UNIVERSITAT



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	10541.6	1167.6	81.5118	0.0001
LOG RANK	12476.3	1509.7	68.2972	0.0001

Finalment, la variable PROVES D'ACCÉS A LA UNIVERSITAT, la qual és quantitativa donat que incorpora la qualificació obtinguda en el global de les proves, ha estat categoritzada per tal de representar gràficament diferents nivells i verificar l'existència d'associació amb la durada dels estudis. Les categories són MENYS DE 5.50, ENTRE 5.50 i 5.90, ENTRE 5.90 i 6.30, i MÉS DE 6.30. Amb les representacions gràfiques constatem l'associació de la variable i la durada dels estudis vist que la supervivència és menor pels estudiants amb qualificacions més baixes en les proves d'accés i va pujant de nivell a mesura que obtenen unes qualificacions més elevades. Pel que fa a les funcions d'atzar també s'observa un desplaçament a l'alça des del nivell més alt en les qualificacions, pel qual correspondrien valors inferiors en la probabilitat condicional d'abandonar, cap amunt en les diferents categories seleccionades.

Els resultats dels contrastos estadístics d'associació també són prou contundents, prenent els estadístics de contrast uns valors molt elevats i clarament diferents de zero del què es conclou la importància de la variable en la caracterització de l'heterogeneïtat de la població en termes de la seva supervivència fins a l'abandonament. A més, la incorporació de la mateixa en les modelitzacions posteriors podrà efectuar-se utilitzant la hipòtesi de proporcionalitat en l'atzar vist que els logaritmes de les funcions d'atzar integrat en cadascun dels estrats evolucionen aproximadament en paral·lel. Aquesta variable serà de molta importància en les anàlisis posteriors, especialment quan no estiguin incorporades les variables de rendiment acadèmic de l'estudiant, donat que serà una aproximació a aquest rendiment, és a dir, els estudiants amb resultats satisfactoris en les proves d'accés tindran, a priori, un rendiment acadèmic més elevat, essent aquest l'element clau en l'explicació de les durades dels estudis.

8.1.1.4. Conclusions Prèvies

Les pàgines anteriors ens mostren els resultats dels contrastos d'associació entre les variables explicatives i la durada dels estudis així com els gràfics de la funció de supervivència i relacionades amb aquesta, segons els valors de les variables explicatives. Els contrastos i els gràfics són individuals, és a dir, tractant cadascuna de les variables explicatives marginalment i separada de les altres.

Observem com totes les variables explicatives estan fortament associades a la durada dels estudis per a la subpoblació dels abandonaments excepte en els casos de les variables SEXE i DOMICILI, mentre que per a les variables TIPUS DE RESIDÈNCIA i CLASSE DE MATRÍCULA, malgrat que en el seu conjunt són significatives, en alguna de les seves modalitats no ho són. En el cas de la variable TIPUS DE RESIDÈNCIA, les modalitats PARES i ALTRES no són significatives individualment donat que tenen un comportament en termes de supervivència molt semblant. Per altra banda, en la variable CLASSE DE MATRÍCULA la modalitat SEMIGRATUÏTA presenta un comportament molt semblant al de la modalitat ORDINÀRIA en front a la de GRATUÏTA. Aquests resultats ens portarien a recodificar aquestes dues variables dicotomitant-les per tal d'unificar les modalitats més homogènies.

Conjuntament amb l'anàlisi individual per a cadascuna de les variables, resulta interessant la realització del contrast d'associació conjunt per a totes les variables. D'aquesta forma serà possible detectar les associacions existents entre les variables explicatives, les quals eren intuïdes en el descriptiu inicial i tindran rellevància en el moment de dur a terme les modelitzacions i estimacions posteriors.

A continuació es presenten dues taules amb el contrast conjunt per a totes les variables explicatives realitzat per etapes, és a dir, incloent una a una cadascuna de les variables explicatives tenint en compte que seran incloses segons sigui la seva associació amb la durada dels estudis, començant per aquella més associada. El procediment consisteix en seleccionar la variable individualment més associada amb la durada, en aquest cas l'EDAT. Posteriorment es realitzarien els contrastos conjunts dos a dos de l'EDAT amb cadascuna de les altres variables i es seleccionaria aquella que aportés un increment en l'estadístic χ^2 més elevat, en aquest cas la variable RESPAN (PROVES D'ACCÉS A LA UNIVERSITAT). En una tercera etapa es realitzarien tots els contrastos conjunts tres a tres de l'EDAT, RESPAN i cadascuna de les altres variables per a seleccionar la que donés lloc a un increment més gran en l'estadístic, ara la variable FONF2 (FONT DE FINANÇAMENT DEL ESTUDIS Treball Propi), i així succesivament.

Ambdós contrastos d'associació Wilcoxon i Log Rank tenen resultats coherents i mostren com només les variables EDAT, RESULTATS P.A.A.U., FONT DE FINANÇAMENT Treball Propi, NACIONALITAT, CENTRE, HORARI, TIPUS DE RESIDÈNCIA Domicili Propi i FONT DE FINANÇAMENT Altres Fonts, aporten significació a l'associació conjunta. La introducció de la resta de variables explicatives no afegeix un increment en l'associació amb la durada dels estudis que sigui significatiu. El nivell de confiança prefixat per a valorar la significació estadística de les variables ha estat del 10%. L'equivalència en els resultats d'ambdós contrastos garanteixen l'estabilitat en les associacions de les variables explicatives amb la durada dels estudis fins a l'abandonament fruit de l'extensa mostra disponible la qual permet mantenir les propietats asimptòtiques dels esmentats estadístics.

Contrast de WILCOXON

Variable	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment	
EDAT	92.4076	0.0001	92.4076	0.0001	EDAT
RESPAN	141.5	0.0001	49.0665	0.0001	P.A.A.U.
FONF2	160.0	0.0001	18.5305	0.0001	FINANC.Treball
NAC	170.9	0.0001	10.8563	0.0010	NACIONALITAT
CENT	177.5	0.0001	6.6474	0.0099	CENTRE
HORA	193.3	0.0001	15.7653	0.0001	HORARI
RESID1	198.6	0.0001	5.3434	0.0208	RESID.Pròpia
FONF3	204.8	0.0001	6.1726	0.0130	FINANC.Altres
CLMA2	207.2	0.0001	2.3991	0.1214	Matr.Gratuïta
LABOR	208.3	0.0001	1.1072	0.2927	SIT.LABORAL
DOMIC	209.1	0.0001	0.8453	0.3579	DOMICILI
RESID3	209.9	0.0001	0.7433	0.3886	RESID.Pares
SEXE	210.6	0.0001	0.7378	0.3904	SEXE
TITOL	211.3	0.0001	0.6688	0.4135	TITOL ACCÈS
ESCI	211.6	0.0001	0.3148	0.5748	ESTAT CIVIL
FONF1	211.8	0.0001	0.1891	0.6636	FINANC.Beca
CLMA3	212.0	0.0001	0.1601	0.6890	Matr.Ordinària

Els resultats indiquen com algunes variables molt associades individualment amb la durada dels estudis per a la sub població d'abandonaments, com ara la SITUACIÓ LABORAL, la CLASSE DE MATRÍCULA, l'ESTAT CIVIL i el TÍTOL D'ACCÉS, quan són tractades conjuntament amb la resta de variables no afegixen gaire a l'associació conjunta. Aquest fet s'explicaria per les interrelacions existents entre les variables, per exemple, l'EDAT amb l'ESTAT CIVIL i el TÍTOL D'ACCÉS, la RESIDÈNCIA en Domicili Propi amb la SITUACIÓ LABORAL, etc. les quals tenen a veure amb l'existència de dos grups d'estudiants clarament diferenciats: d'una banda els més joves, solters, provinents del COU que no treballen i financen amb l'ajut dels pares i, d'altra, els de més edat, no solters, que treballen i financen els estudis amb el seu propi treball.

Contrast de LOG RANK

Variable	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment	
EDAT	100.3	0.0001	100.3	0.0001	EDAT
RESPAN	139.9	0.0001	39.5901	0.0001	P.A.A.U.
FONF2	156.6	0.0001	16.7509	0.0001	FINANC.Treball
NAC	168.0	0.0001	11.3520	0.0008	NACIONALITAT
CENT	173.9	0.0001	5.9400	0.0148	CENTRE
HORA	186.8	0.0001	12.8798	0.0003	HORARI
RESID1	192.4	0.0001	5.5919	0.0180	RESID.Propia
FONF3	198.9	0.0001	6.4486	0.0111	FINANC.Altres
CLMA2	200.8	0.0001	1.9777	0.1596	Matr.Gratuita
LABOR	201.9	0.0001	1.0935	0.2957	SIT.LABORAL
SEXE	202.3	0.0001	0.3336	0.5635	SEXE
TITOL	202.5	0.0001	0.2809	0.5961	TITOL ACCÈS
FONF1	202.8	0.0001	0.2981	0.5850	FINANC.Beca
CLMA3	203.1	0.0001	0.2377	0.6258	Matr.Ordinària
RESID3	203.2	0.0001	0.1412	0.7071	RESID.Pares
DOMIC	203.8	0.0001	0.5825	0.4453	DOMICILI
ESCI	203.8	0.0001	0.00671	0.9347	ESTAT CIVIL

Les vuit variables que conjuntament determinen de forma significativa l'associació amb la durada dels estudis són les que tindran un paper fonamental en la construcció dels models de durada posteriorment estimats. Per a la resta de variables caldrà tenir en compte les interrelacions existents i analitzar la possibilitat d'incloure en el model alguns termes d'interacció que afegixin informació rellevant. En aquest sentit cal dir que, fins el moment, no han estat considerades en aquestes anàlisis prèvies per a la subpoblació d'abandonaments les variables del rendiment acadèmic dels estudiants les quals previsiblement tindran uns efectes determinants sobre les durades dels estudis i estaran per força connectades amb les característiques personals i de gestió acadèmica dels alumnes. La incorporació, doncs, del rendiment acadèmic pot alterar els resultats obtinguts.

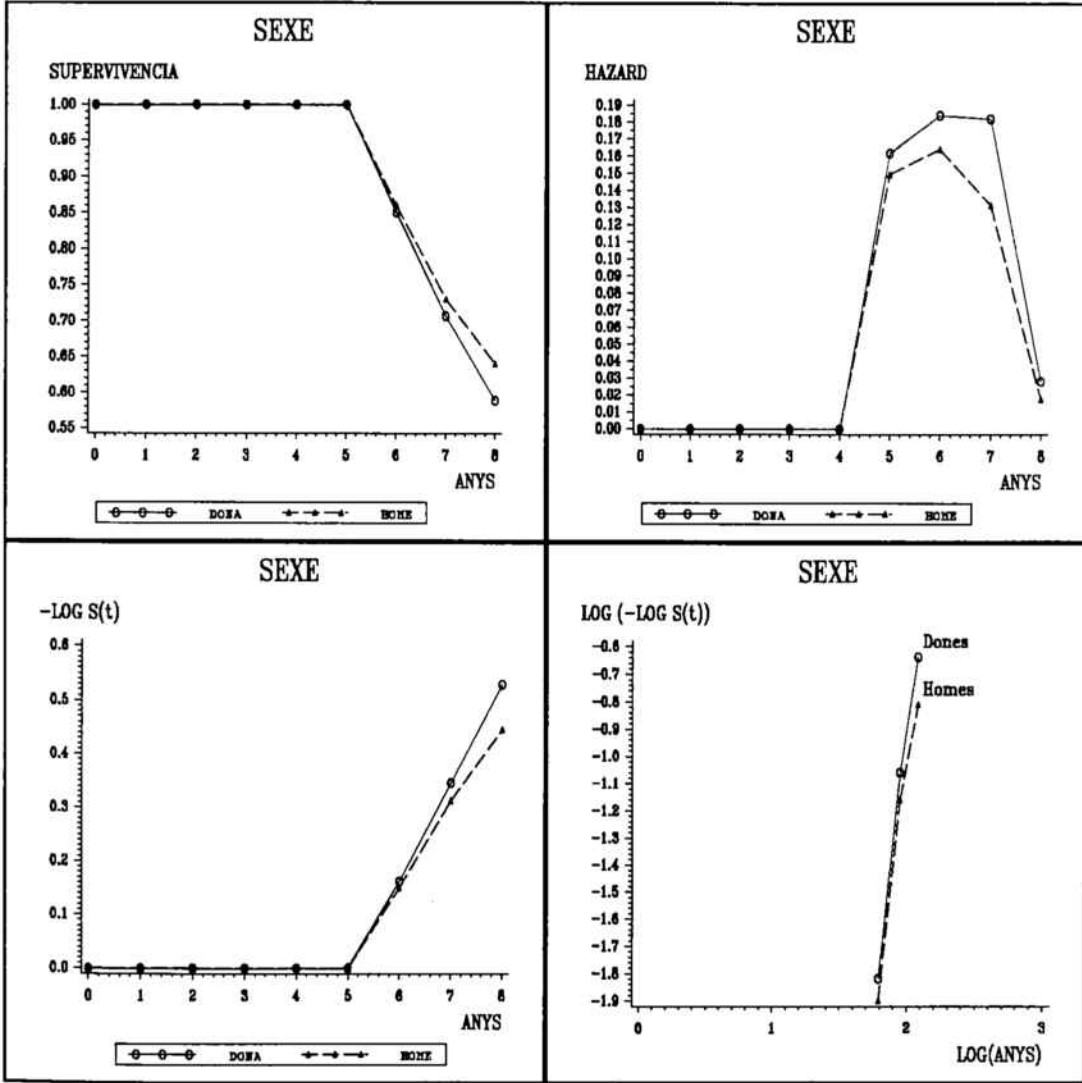
8.1.2 Sub població de Llicenciats

8.1.2.1 Variables Personals

En la sub població dels llicenciats per a la variable SEXE s'obtenen unes conclusions equivalents a les de la sub població d'abandonaments fruit de la inspecció de les representacions gràfiques de la funció de supervivència i la resta de funcions relacionades així com dels contrastos estadístics d'associació o igualtat en les funcions per estrats. En els gràfics cal destacar el que serà una constant per aquesta sub població com és que les funcions es mantenen constants fins el cinquè any acadèmic degut a que fins aleshores no s'han enregistrat llicenciats. En conseqüència, no es disposa de un període massa llarg per a analitzar la finalització per motiu de l'obtenció del títol. Tot i això, s'observa una evolució de les funcions de supervivència i relacionades amb aquesta molt semblant tant pels homes com per les dones, essent remarcable una petita tendència superior per part de les dones a finalitzar els estudis abans que els homes. Aquesta més gran probabilitat condicional a llicenciar-se per a les dones es posa de manifest en el gràfic de la funció d'atzar, els valors de la qual pels homes són inferiors als de l'altra col·lectiu. En aquest sentit s'ha de dir que, pel darrer any acadèmic considerat, únicament han estat enregistrats com a llicenciats aquells alumnes que ho varen fer en la convocatòria extraordinària de febrer per la qual cosa el valor de la funció d'atzar és molt més baix que en els anys anteriors com a conseqüència de la manca d'enregistrament dels estudiants que s'han llicenciat en les convocatòries de juny i setembre de 1993.

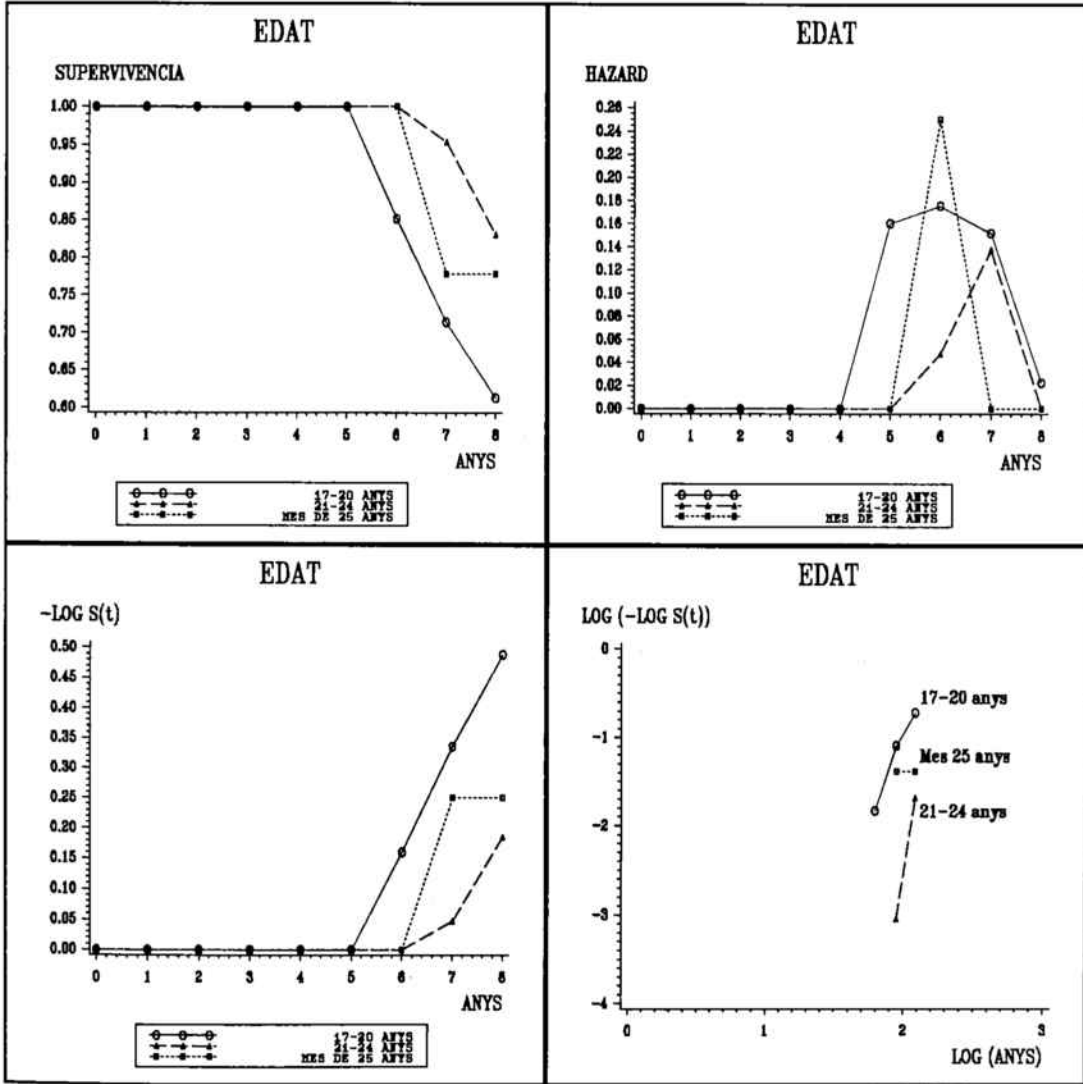
D'altra banda, els contrastos estadístics de Wilcoxon i Log rank permeten validar la hipòtesi d'igualtat en les funcions de supervivència pels dos sexes amb un nivell de confiança del 5%.

SEXE



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	10.1942	7.1478	2.0341	0.1538
LOG RANK	12.3535	8.4909	2.0678	0.1504

EDAT

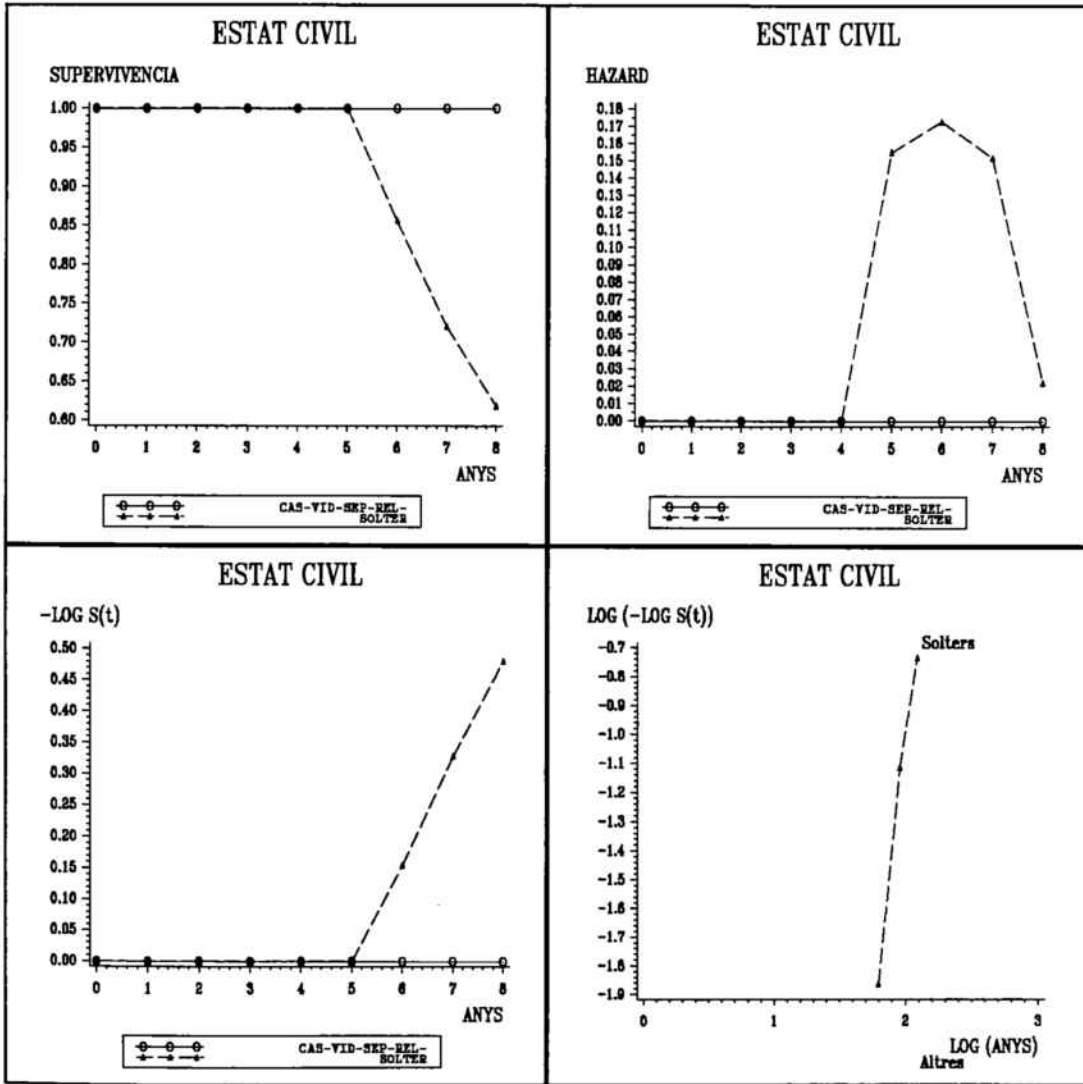


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	69.8533	18.0210	15.0250	0.0001
LOG RANK	75.1900	22.4187	11.2487	0.0008

La variable EDAT també es troba associada a la durada dels estudis fins a l'obtenció del títol de llicenciat com succeïa en la subpoblació d'abandonaments. Aquest resultat és clar a partir de les representacions gràfiques de les funcions de supervivència per a les diferents categories creades a partir de les edats dels estudiants. Així, s'observa com els estudiants més joves, aquells amb edats en iniciar els estudis situades entre els 17 i els 20 anys són els que tenen durades més curtes i, per tant, els que es llicencien abans, mentre que els que van començar la carrera amb edats superiors triguen més temps a obtenir el títol. Tot i això, el grup d'estudiants amb edats superiors als 25 anys, el qual no és massa nombrós, té un valor en la funció d'atzar pel sisè any acadèmic molt més elevat que el dels altres dos grups d'edats trencant l'esquema de a major edat en iniciar els estudis més llargs resulten aquests fins a l'obtenció del títol. D'altra banda i tal com s'anirà repetint contínuament en aquesta subpoblació, en el darrer any acadèmic no estan enregistrats la totalitat dels estudiants llicenciats puix que el curs encara no ha finalitzat i els expedients acadèmics no han estat actualitzats. Aquest fet provoca una caiguda en la funció d'atzar per a la darrera observació la qual no correspondrà a la realitat.

Els contrastos estadístics d'associació, tot i els comentaris anteriors, indiquen la desigualtat en les funcions de supervivència i, en conseqüència una associació entre la durada dels estudis i l'edat considerada com a variable quantitativa. La hipòtesi nul·la d'igualtat en les distribucions es rebutja amb claredat per ambdós contrastos vist que en cas que aquella hipòtesi nul·la fos correcta la probabilitat d'obtenir un estadístic χ^2 més gran al resultat mostral actual és inferior a una mil·lèsima.

ESTAT CIVIL



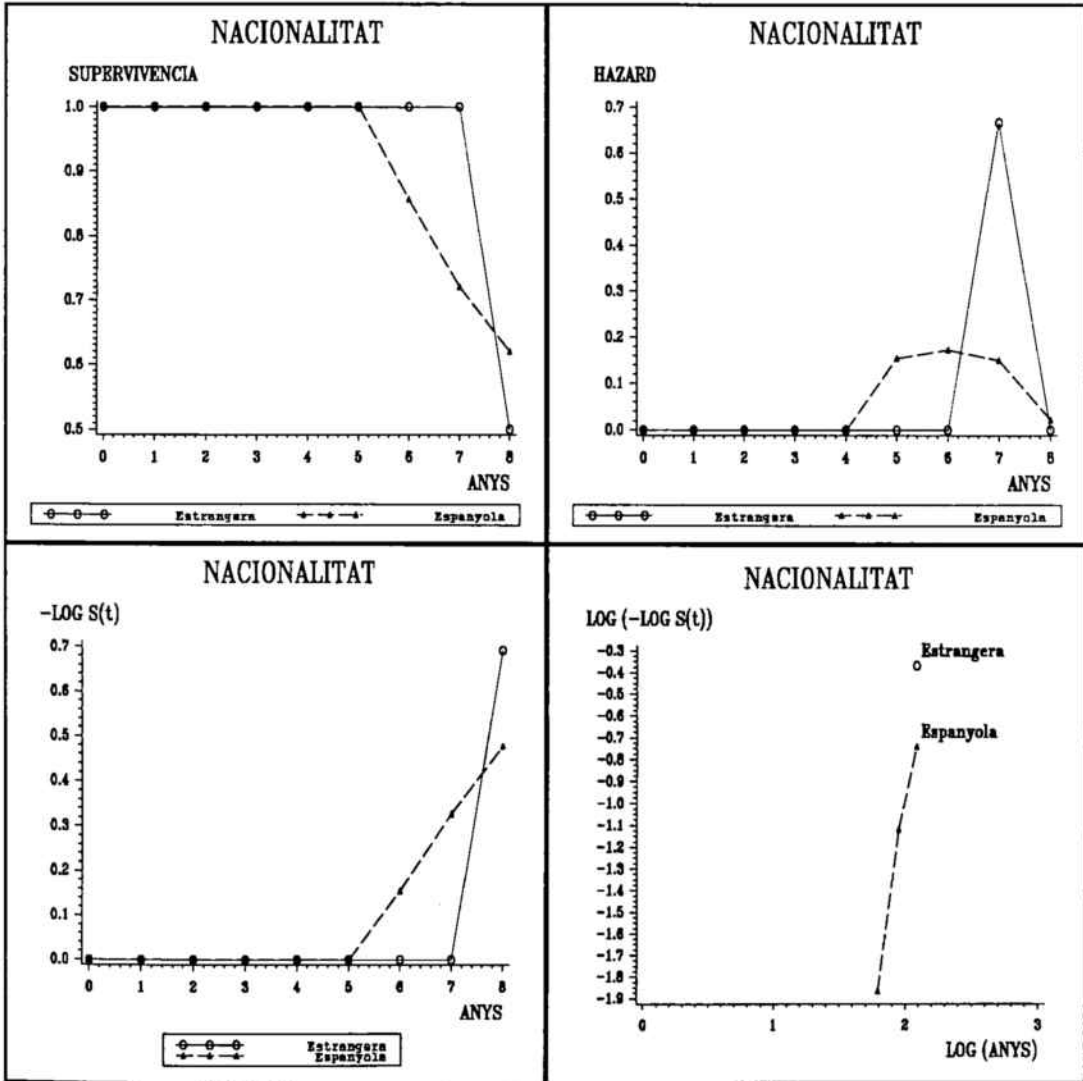
Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-1.5042	1.0029	2.2494	0.1337
LOG RANK	-1.6995	1.3001	1.7087	0.1912

Per a la variable ESTAT CIVIL en la subpoblació dels llicenciats ens trobem davant la situació en la qual cap dels estudiants del grup dels no solters, és a dir, dels casats, separats, divorciats, vidus o religiosos, no ha obtingut el títol fins a la darrera informació disponible. Aquest fet s'expressa en que la funció de supervivència per aquest estrat pren el valor 1 al llarg de tot el període analitzat mentre que pel grup dels solters va decreixent indicant la progressiva obtenció del títol per part d'aquests estudiants.

Com a conseqüència de l'anterior s'extreu la conclusió de l'existència de relació entre l'estat civil i la durada dels estudis fins a l'abandonament per bé que els contrastos estadístics d'associació no indicarien el mateix. Aquests darrers impliquen el no rebuig de la hipòtesi nul·la d'igualtat en les funcions de supervivència al nivell de confiança del 5%, essent aquest resultat incoherent degut a que en l'estrat corresponent als alumnes no solters totes les observacions es troben censurades. Una única observació no censurada per aquest grup hauria permès l'obtenció d'uns resultats més adreçats a la violació de la hipòtesi nul·la anterior i, per tant, a la conclusió d'existència d'associació entre durada dels estudis fins a l'obtenció del títol i estat civil.

De la mateixa forma que en la subpoblació d'abandonaments, però, la variable considerada es trobarà molt interrelacionada amb altres variables fruit de la ja comentada dicotomia existent en les característiques personals i de gestió acadèmica dels alumnes segons la qual per una banda tenim els estudiants joves, solters, procedents del COU i que financen els estudis amb ajut dels pares i, de l'altra, els de més edat, no solters, amb domicili propi i que financen els estudis mitjançant el seu propi treball.

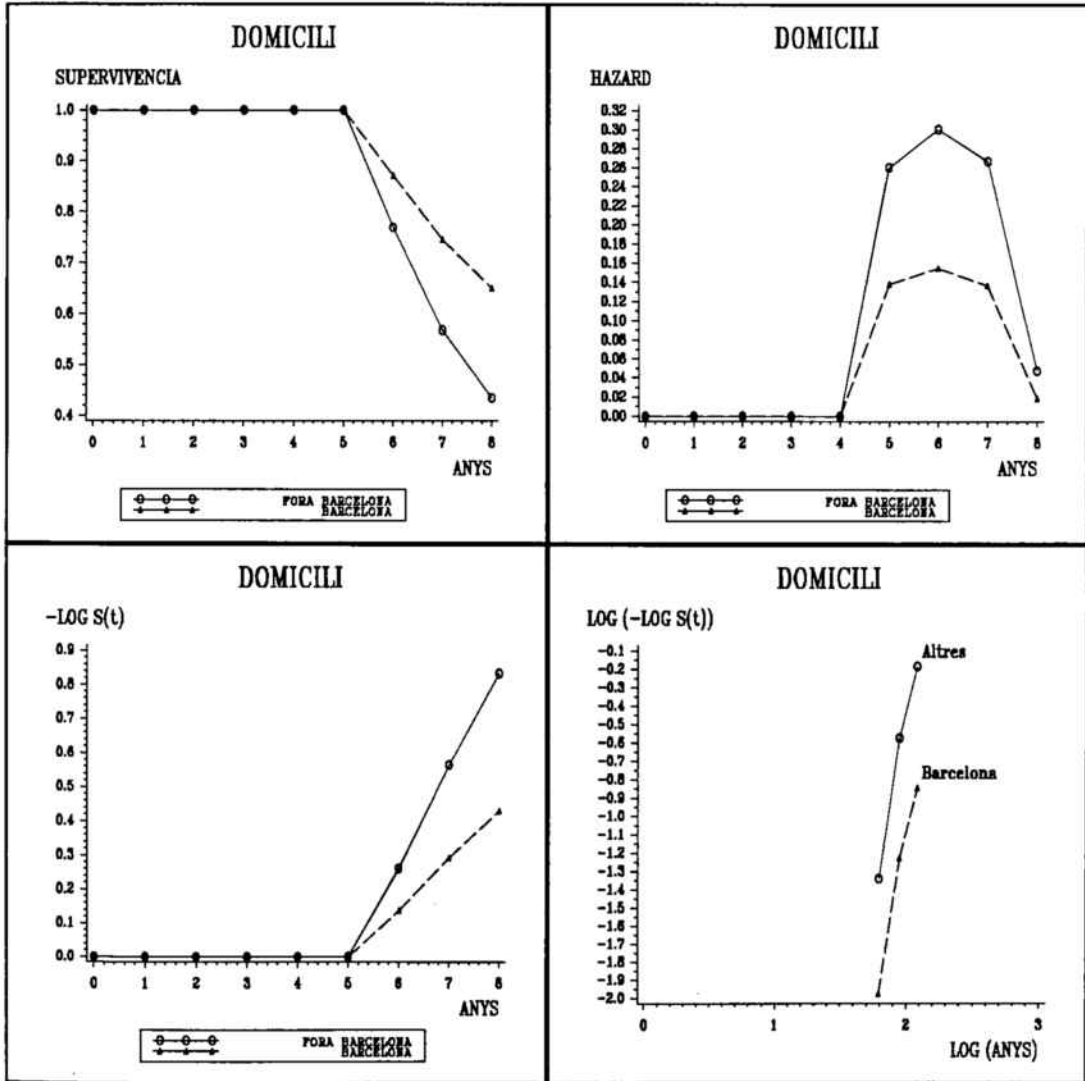
NACIONALITAT



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-0.1380	0.8825	0.0244	0.8758
LOG RANK	0.0190	0.9889	0.000368	0.9847

En la NACIONALITAT es recull molt poca variabilitat per quant només un 1% de la població total d'estudiants és de nacionalitat estrangera. Aquesta situació queda reflectida en els gràfics de les funcions de supervivència i deduïdes de la mateixa en les quals s'observa com, mentre que els de nacionalitat espanyola tenen una evolució progressiva cap a l'obtenció del títol amb una funció d'atzar que va creixent a partir del cinquè any acadèmic i que posteriorment baixa com a conseqüència de l'enregistrament incomplet dels llicenciats en el darrer any, pel col·lectiu d'estrangers la funció de supervivència experimenta un descens molt pronunciat a partir del setè any indicatiu d'una situació en la què un percentatge molt elevat del grup ha obtingut el títol. Això mateix es fa palès en la funció d'atzar on per aquests estudiants pren un valor molt elevat en el penúltim any acadèmic. Així doncs, com en el cas de la variable ESTAT CIVIL anteriorment comentada, novament ens trobem davant d'una situació complexa d'interpretació vist que d'una banda les representacions gràfiques mostren un comportament clarament diferenciat pels dos estrats considerats i, de l'altra, els contrastos estadístics, condicionats per l'estructura de les dades, portarien a no rebutjar la hipòtesi d'igualtat en les funcions de supervivència. Caldrà doncs anar amb compte en la incorporació d'aquesta variable en les modelitzacions posteriors en les quals la variable no serà de gran interès com a conseqüència de la manca de variabilitat en la informació que recull si més no en la subpoblació dels llicenciats. En la d'abandonaments, malgrat que la informació continguda en la variable és la mateixa, el percentatge d'abandonaments al llarg dels anys analitzats permet extraure unes conclusions més acurades les quals vindran reflectides en els resultats de les estimacions dels models de durada alternatius.

DOMICILI

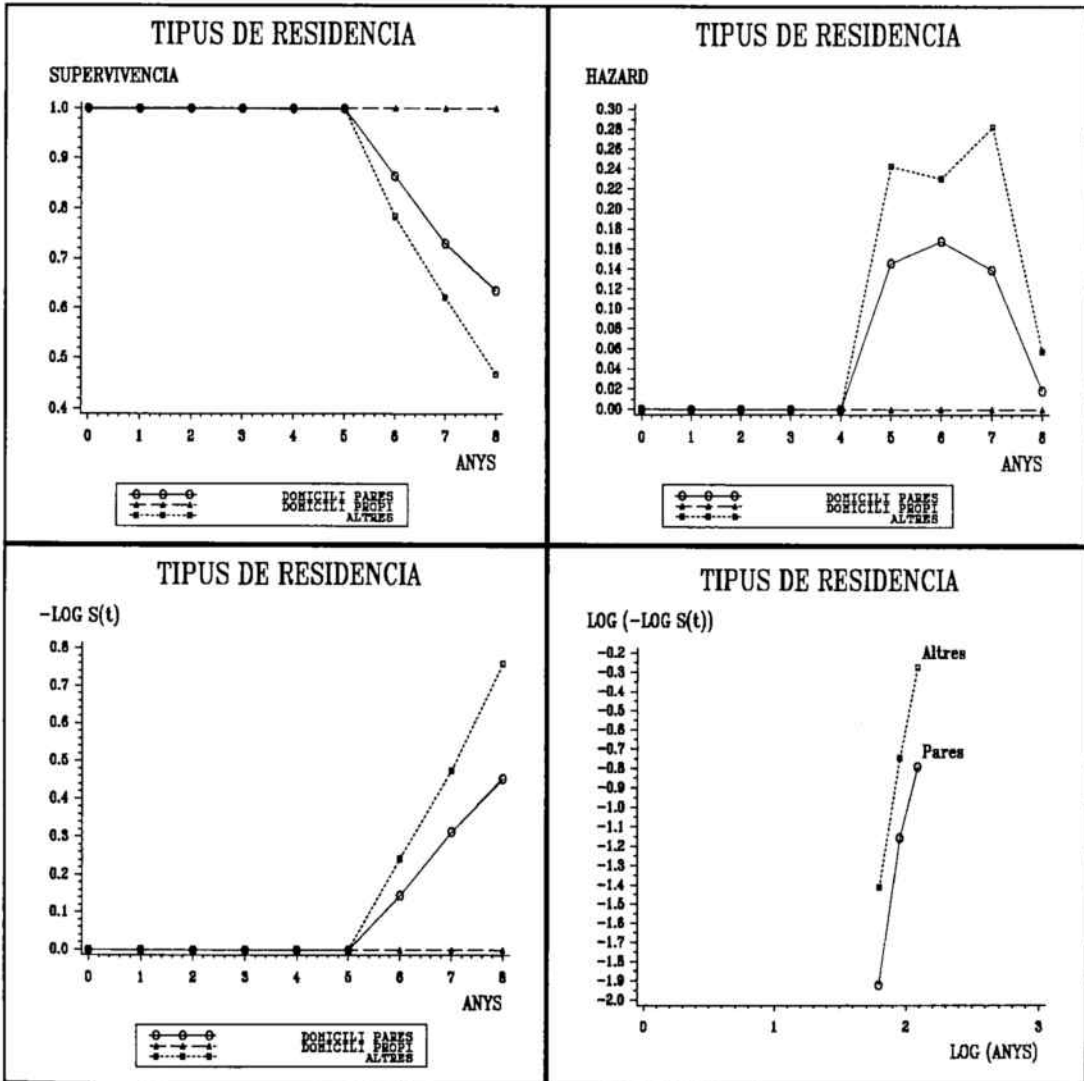


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	23.4511	5.0905	21.2225	0.0001
LOG RANK	26.6223	5.8657	20.5991	0.0001

A diferència del que ha estat posat de manifest per a la subpoblació dels abandonaments, la variable DOMICILI es troba associada amb la durada dels estudis fins a l'obtenció de la llicenciatura. Els gràfics de les funcions de supervivència per a les dues categories de la variable, els de domicili a Barcelona i el seu entorn industrial d'una banda i, els d'altre domicili més allunyat de l'altra, mostren clarament com els estudiants residents fora de Barcelona tenen unes durades inferiors a les d'aquells que tenen més a prop el centre on cursen els estudis. Les funcions d'atzar també posen de manifest aquesta situació amb valors més elevats i, per tant, probabilitats condicionals de llicenciar-se més grans, pels alumnes residents en poblacions allunyades de Barcelona. Això pot ser explicat pel cost addicional que tenen els estudiants que han de desplaçar-se fins a Barcelona per a cursar els estudis el qual serveix de revulsiu per a finalitzar els estudis amb una major celeritat. En certa manera, també caldria tenir present que només aquells estudiants amb uns rendiments acadèmics més satisfactoris al llarg de l'ensenyament secundari es plantegen, ells o els seus pares, l'accés a un estudis superiors lluny del seu lloc de residència.

Els contrastos estadístics ens permeten validar estadísticament les impressions obtingudes amb les representacions gràfiques anteriors evidenciant una clara associació entre la variable DOMICILI i la durada dels estudis fins a la seva finalització amb èxit. Els estadístics de contrast prenen uns valors prou alts de tal forma que, en cas de verificar-se la hipòtesi nul·la d'igualtat en les distribucions pels dos estrats, la probabilitat d'un valor superior a l'obtingut per l'estadístic és inferior a una deu-mil·lèsima.

TIPUS DE RESIDÈNCIA

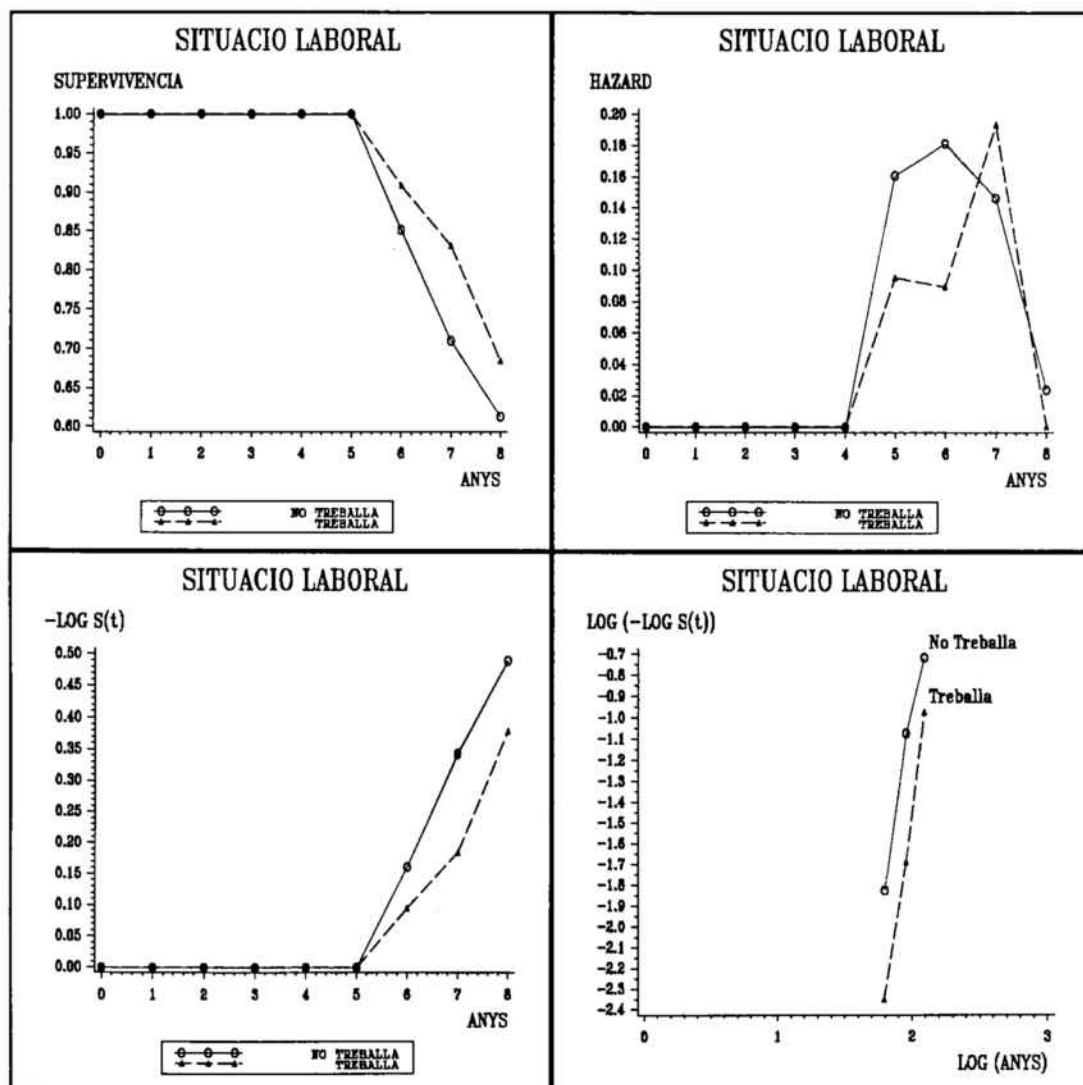


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON			15.9085	0.0004
Domicili Pares	11.3033	4.8553	5.4197	0.0199
Domicili Propi	3.8340	1.5776	5.9064	0.0151
Altres Domicilis	-15.1373	4.6527	10.5846	0.0011
LOG RANK			16.7418	0.0002
Domicili Pares	13.1338	5.7038	5.3022	0.0213
Domicili Propi	4.3764	2.0773	4.4387	0.0351
Altres Domicilis	-17.5102	5.3964	10.5286	0.0012

En la variable TIPUS DE RESIDÈNCIA ens trobem la mateixa situació que en alguna de les variables comentades anteriorment per a la subpoblació dels llicenciats consistent en què una de les categories en les quals ha estat estratificada la variable, concretament la d'aquells alumnes que resideixen en Domicili Propi, no presenta cap observació corresponent a la finalització amb èxit dels estudis. Tanmateix, per a les altres dues categories si s'enregistren alumnes llicenciats i, a més, la tendència a obtenir el títol és més elevada pels residents en Altres domicilis que no pas pels que viuen a casa dels seus pares. Les funcions d'atzar mostren clarament el que s'acaba de dir. En relació als comentaris realitzats per a la variable DOMICILI, cal esperar que els alumnes que resideixen en pisos d'estudiants, residències universitàries, pensions, etc. tinguin un cost econòmic més gran que la resta d'estudiants per la qual cosa es veuran més motivats per finalitzar la carrera el més aviat possible a més de l'explicació en termes de un rendiment acadèmic superior en l'ensenyament secundari d'aquells que acaben decidint cursar estudis superiors lluny de la seva residència habitual.

Els contrastos estadístics no contravenen els resultats anteriors verificant-se, globalment i individualment per a cadascuna de les tres categories que incorpora la variable, l'associació existent amb la durada dels estudis. Globalment la significació estadística dels contrastos és molt elevada mentre que individualment i per a les categories Domicili Pares i Domicili Propi el nivell de confiança que permetria rebutjar la hipòtesi nul·la d'igualtat en les funcions de supervivència es trobaria vorejant el 5%, especialment en el contrast del Log rank. Cal considerar, però, l'elevat nombre d'observacions censurades per a la modalitat Domicili Propi ja que cap estudiant del grup s'ha llicenciat fins el moment, si més no amb la informació disponible.

SITUACIÓ LABORAL



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	8.6139	4.4046	3.8247	0.0505
LOG RANK	9.0641	5.3656	2.8538	0.0912

Contràriament al que seria d'esperar, la variable SITUACIÓ LABORAL no presenta les mateixes circumstàncies que l'ESTAT CIVIL i el TIPUS DE RESIDÈNCIA en el sentit que la categoria Treballa, molt relacionada amb els no solters que resideixen en domicili propi, sí que disposa d'observacions en situació d'haver obtingut el títol de llicenciat. Això cal interpretar-ho com una demostració de què la inclusió entre els que treballen dels que ho fan a temps complet, a temps parcial o esporàdicament, desvirtua lleugerament la informació de la variable en termes de caracteritzar l'heterogeneïtat de la subpoblació pel que fa a la durada dels estudis. Tot i això, les representacions gràfiques de les funcions de supervivència pels dos col·lectius mostren com aquells que no treballen tenen unes durades inferiors i, per tant, es llicencien abans. Pels que treballen, la funció d'atzar pren un valor, en el setè any acadèmic, per sobre del de l'altra col·lectiu antagònic la qual cosa provoca inestabilitats en els resultats estadístics.

Els contrastos d'associació permeten sancionar estadísticament els comentaris anteriors vist que es rebutjaria la hipòtesi d'igualtat de les funcions de supervivència malgrat que no amb tanta claredat com ha estat fet per a d'altres variables. En aquest cas, el nivell de confiança amb el qual es pot rebutjar la hipòtesi nul·la es trobaria al voltant del 10%. Les funcions logaritme de l'atzar integrat per a cadascun dels dos grups d'estudiants evolucionen aproximadament en paral·lel al llarg dels anys considerats malgrat que a una distància petita, la qual cosa servirà per a permetre la incorporació de la variable en un model d'atzar proporcional en el si del qual podran ser analitzades les interrelacions de la variable amb altres explicatives d'interès.

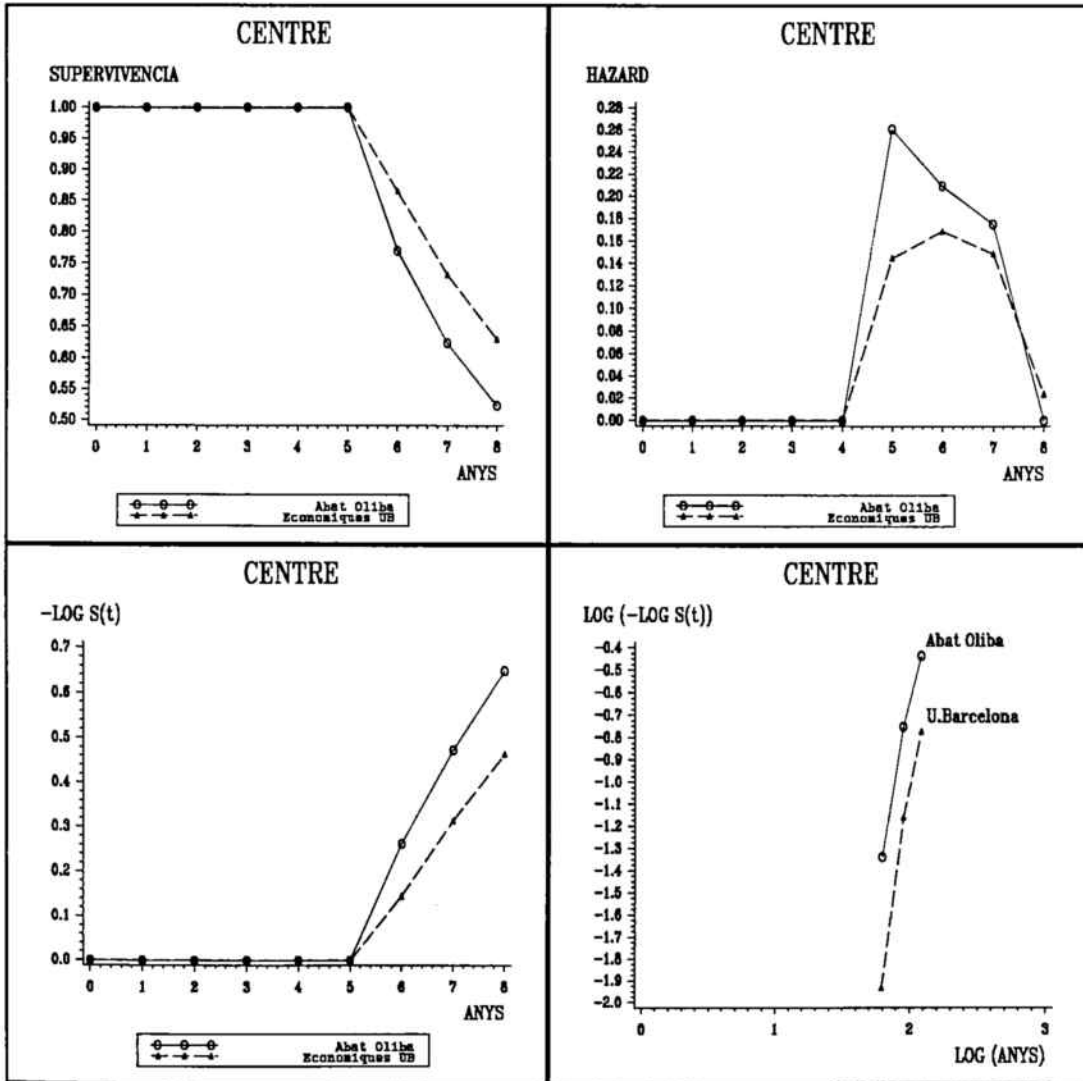
8.1.2.2 Variables de Gestió Acadèmica

Per a la variable CENTRE, la qual diferencia entre els alumnes que cursen el primer cicle de la llicenciatura a la Facultat d'Econòmiques i els que ho fan al centre adscrit Abat Oliba, bé sia mitjançant els gràfics de les funcions de supervivència, bé sia pels resultats dels contrastos estadístics, es deriva l'existència d'associació amb la durada dels estudis fins a l'obtenció del títol.

En els gràfics de les funcions de supervivència s'observa com els alumnes que realitzen el primer cicle en el Col·legi Abat Oliba tenen unes durades inferiors i, per tant, es llicencien abans que els de la Facultat d'Econòmiques. Tanmateix, de l'anàlisi de la funció d'atzar es conclou que la probabilitat condicional de llicenciar-se pels alumnes de l'Abat Oliba únicament és molt més gran que la corresponent als alumnes de la Facultat en el cinquè any, és a dir, pel primer curs acadèmic en el qual és factible haver finalitzat els estudis. Més enllà d'aquest any ambdues funcions d'atzar tendeixen cap a valors propers amb el descens produït en el darrer curs analitzat degut a que la informació pel mateix és incompleta. Això s'interpreta com que, un cop integrats tots els estudiants en el segon cicle de la Facultat, es van esvaïnt les possibles diferències entre uns i altres.

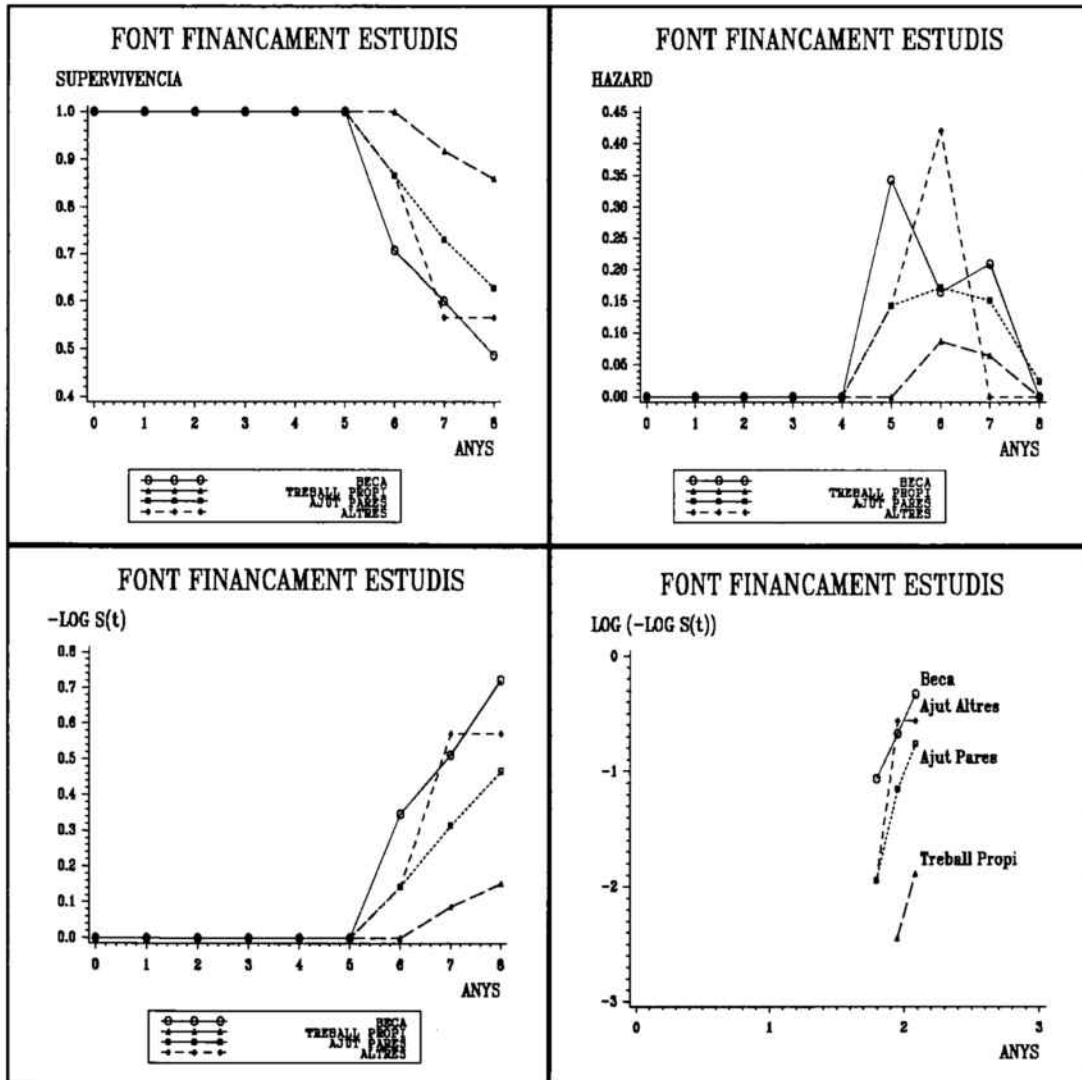
Pel que fa als contrastos estadístics, dels resultats dels mateixos es rebutjaria la hipòtesi nul·la d'igualtat en les funcions de supervivència i, en conseqüència es verificaria estadísticament l'associació de la variable amb la durada dels estudis. En ambdós contrastos la hipòtesi nul·la es rebutja amb un nivell de confiança del 5%. A més la variable podrà ser incorporada en la modelització d'atzar proporcional vist que el logaritme de les funcions d'atzar integrat pels dos col·lectius mantenen una distància constant en el període analitzat.

CENTRE



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	9.4423	4.1131	5.2701	0.0217
LOG RANK	9.9399	4.8882	4.1349	0.0420

FONT DE FINANÇAMENT DELS ESTUDIS

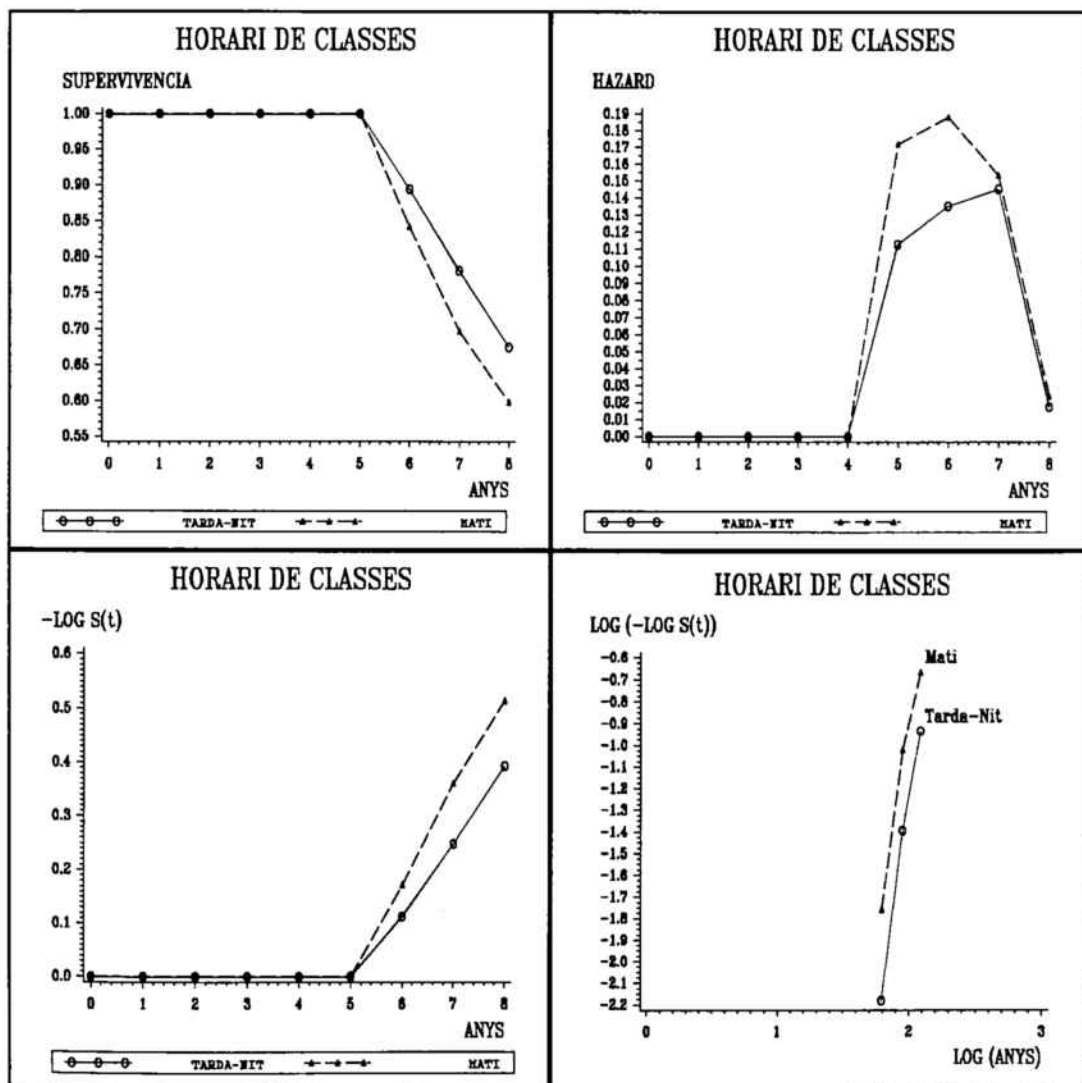


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON			17.4198	0.0005
Ajut Pares	6.2235	4.9733	1.5660	0.2108
Beca	-12.7169	4.0373	9.9217	0.0016
Treball Propi	7.4768	2.6272	8.0996	0.0044
Altres	-0.9834	1.8802	0.2736	0.6010
LOG RANK			15.2788	0.0016
Ajut Pares	5.8857	5.9756	0.9701	0.3246
Beca	-13.0485	4.7912	7.4171	0.0065
Treball Propi	8.1960	3.2857	6.2220	0.0126
Altres	-1.0331	2.2106	0.2184	0.6402

De les representacions gràfiques corresponents a les funcions de supervivència de cadascuna de les modalitats de la variable FONT DE FINANÇAMENT DELS ESTUDIS es dedueix l'associació amb la durada dels estudis fins a la finalització amb èxit. D'una banda, els alumnes que gaudeixen d'una beca són els que obtenen el títol més ràpidament, tot seguit vindrien els que disposen de l'ajut dels pares i, en darrer lloc, els que financen els estudis mitjançant els seu propi treball són els més tardans en finalitzar la carrera. El grup dels que disposen d'altres fonts de finançament, com ara l'ajut del cònjuge o d'altra persona, mostren una situació inestable tant per la funció de supervivència com per la d'atzar deguda, principalment, al reduït nombre d'observacions en aquesta modalitat la qual cosa fa que uns pocs llicenciats distorsionin els resultats en el càlcul de la probabilitat condicional de finalització dels estudis. Així les coses, aquesta modalitat d'Altres fonts de finançament es troba molt relacionada amb la modalitat d'ajut dels pares.

Els contrastos estadístics mostren que globalment es pot rebutjar la hipòtesi d'igualtat en les funcions de supervivència per bé que en algunes modalitats, com ara la d'Ajut dels pares i la d'Altres fonts, el contrast individual seria contradictori amb el global. Tal com s'acaba de comentar aquestes dues modalitats tindrien una estructura de supervivència semblant i, a més, la darrera es trobaria molt condicionada pel reduït nombre d'efectius dels què disposa. Fixant l'atenció en les tres altres fonts de finançament, Ajut Pares, Treball Propi i Beca, els logaritmes de les funcions d'atzar integrat evolucionen aproximadament en paral·lel i, en conseqüència, la variable podrà ser incorporada en el model d'atzar proporcional sense que es vegi afectada la hipòtesi de l'efecte proporcional sobre l'atzar.

HORARI

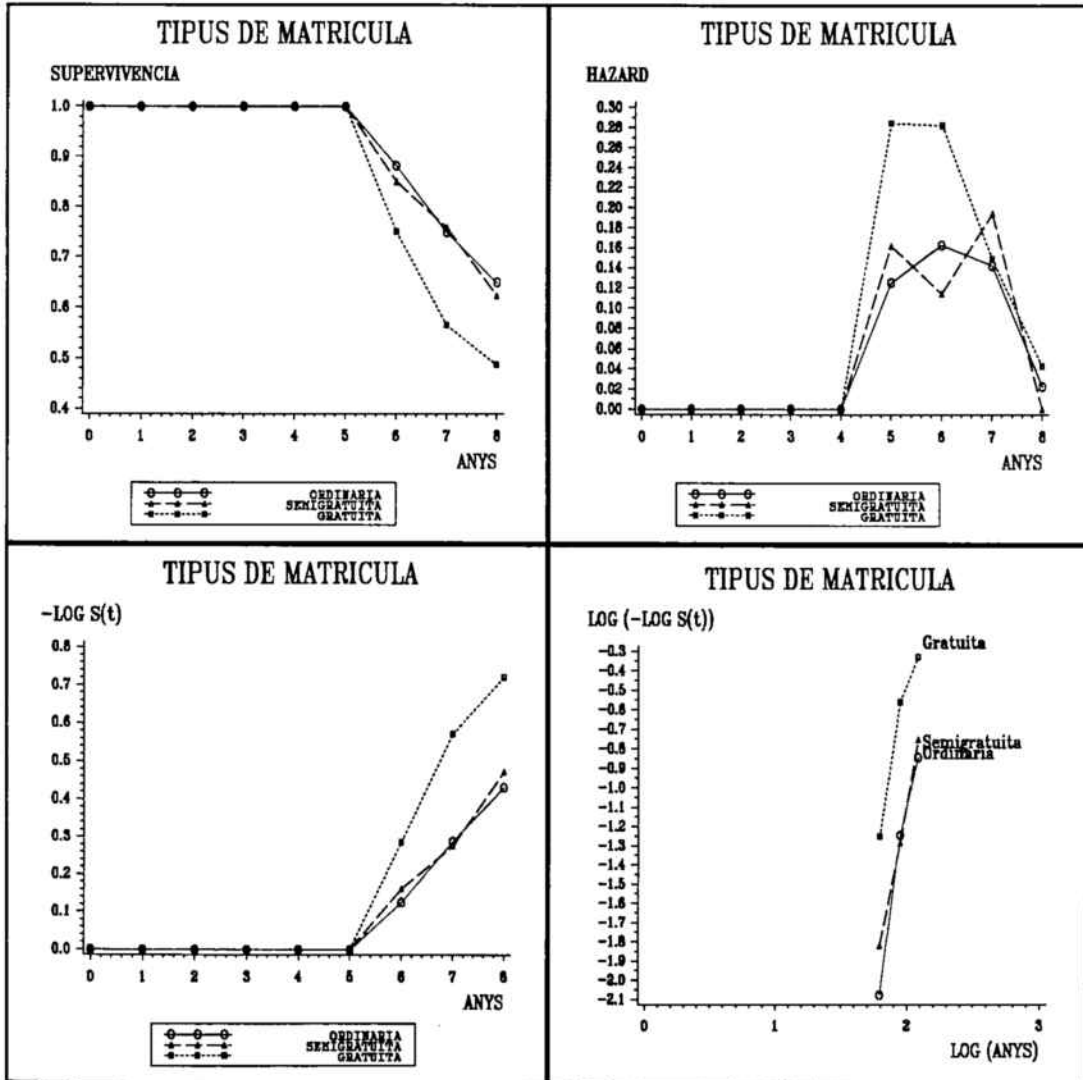


Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-16.7758	6.7100	6.2506	0.0124
LOG RANK	-18.0904	8.1417	4.9370	0.0263

L'HORARI de classes és una variable que explica part de l'heterogeneïtat en la supervivència dels estudiants de la cohort analitzada. Els gràfics de les funcions de supervivència evidencien una durada més llarga fins a la finalització de la carrera dels estudiants que segueixen l'horari de Tarda-Nit. Atenent a les funcions d'atzar estimades per a cadascun dels dos grups d'estudiants, s'observa com la tendència a obtenir el títol de llicenciat en cada moment del temps, és a dir, la probabilitat condicional de finalitzar els estudis, és més elevada pels de l'horari de Matí especialment en el primer any en el qual és possible llicenciar-se. A partir d'aleshores ambdues funcions convergeixen cap a valors més propers. Novament cal insistir en les interrelacions previsibles entre els variables i, en aquest cas, l'Horari es troba fortament associat amb la Situació Laboral, l'Edat, l'Estat Civil, etc. per la qual cosa caldrà esperar a la realització del contrast d'associació conjunt amb aquestes i altres variables per a constatar la importància de la variable en la caracterització de l'heterogeneïtat en la supervivència de la població.

Dels contrastos estadístics individuals es dedueix que l'associació entre l'Horari i la durada dels estudis és significativa al nivell de confiança del 5%. Altrament, la variable podrà ser inclosa en un model d'atzar proporcional com a explicativa sense haver d'esperar un incompliment de la hipòtesi d'efectes proporcionals sobre l'atzar per modificacions en les modalitats considerades. Aquesta conclusió es deriva de la inspecció del gràfic dels logaritmes de les funcions d'atzar integrat els quals es troben separats per una distància aproximadament constant pels diversos cursos acadèmics dels quals es disposa.

CLASSE DE MATRÍCULA



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON			19.6550	0.0001
Ordinària	21.7520	6.7106	10.5070	0.0012
Semigratuïta	1.5361	5.1560	0.0888	0.7658
Gratuïta	-23.2882	5.2833	19.4294	0.0001
LOG RANK			18.5126	0.0001
Ordinària	23.1089	8.0351	8.2714	0.0040
Semigratuïta	1.6090	6.2432	0.0664	0.7966
Gratuïta	-24.7179	6.2223	15.7807	0.0001

L'estructura de la supervivència dels estudiants de la cohort fins a la finalització dels estudis amb èxit atenent a les modalitats de la variable CLASSE DE MATRÍCULA és de les mateixes característiques que per a la subpoblació dels abandonaments. En concret, els alumnes que gaudeixen de matrícula gratuïta tenen durades més curtes que la resta i cal tenir en compte que, entre aquests es troben inclosos els alumnes amb beca els quals tenen en general un rendiment acadèmic més elevat i, en conseqüència, acaben la carrera més ràpidament. Les altres dues modalitats, per la seva banda, tenen comportaments molt semblants en termes de supervivència per bé que en la funció d'atzar s'enregistren petites diferències en els valors pel sisè curs acadèmic. La equivalència entre les funcions de supervivència de les modalitats Matrícula Ordinària i Matrícula Semigratuïta queda reflectida en els resultats dels contrastos estadístics individuals.

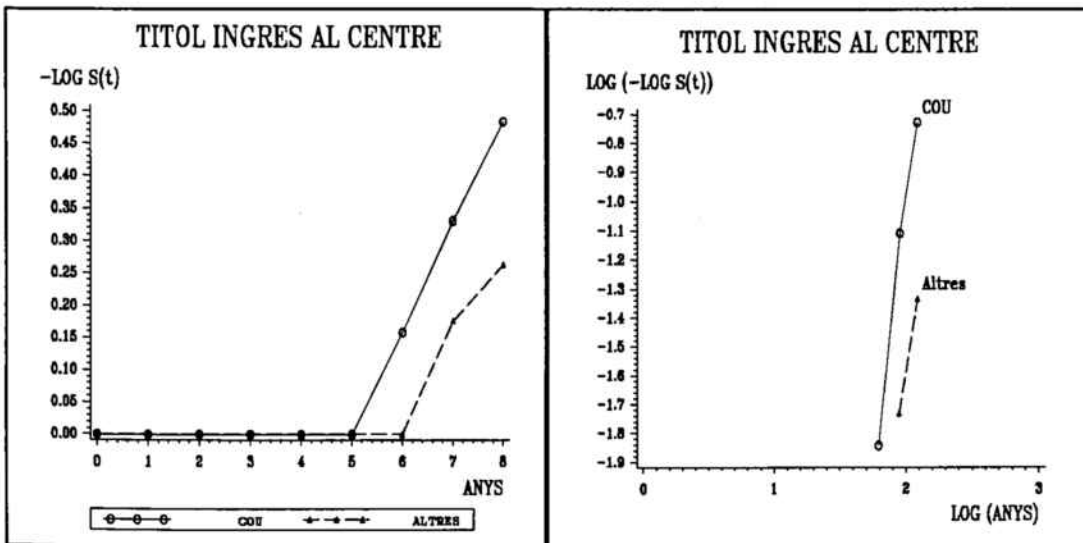
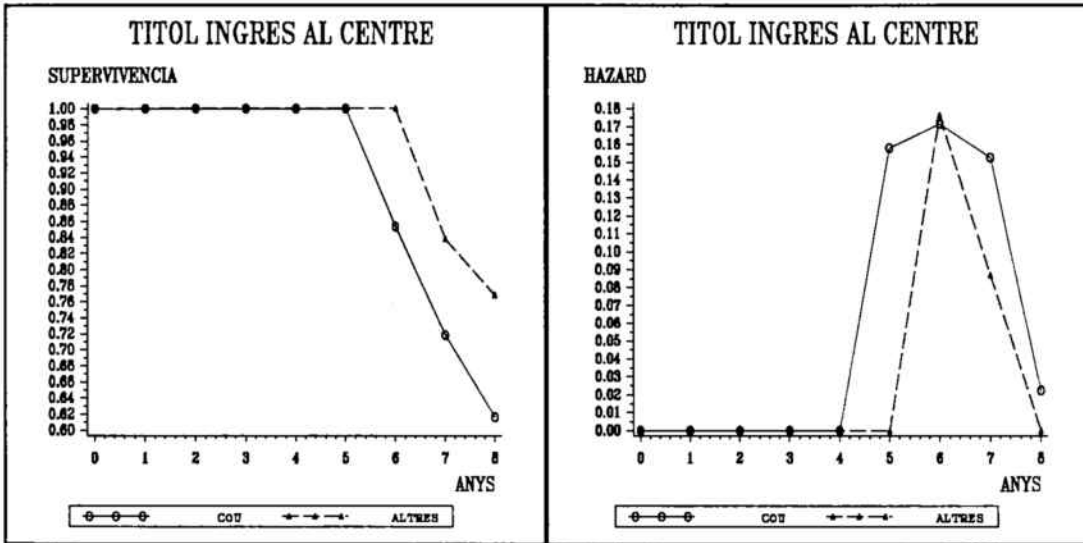
Globalment, per a la variable CLASSE DE MATRÍCULA es rebutja clarament la hipòtesi d'igualtat en les funcions de supervivència i, per tant, l'associació amb la durada dels estudis és significativa estadísticament. Així, la probabilitat d'obtenir un valor superior a l'estadístic mostral en cas que la hipòtesi nul·la fos la correcta és inferior a una deu-mil·lèsima. Tanmateix, per a la modalitat Matrícula Semigratuïta i en consonància amb els comentaris efectuats anteriorment, el contrast no permetria rebutjar la igualtat en les funcions de supervivència degut a que els alumnes de l'esmentada modalitat tindrien una estructura de supervivència molt semblant als de la Matrícula Ordinària. En cas de considerar el grup dels de matrícula gratuïta en front dels de la ordinària més els de la semigratuïta la variable podrà ser incorporada en la modelització d'atzar proporcional vist que els logaritmes de la funció d'atzar integrat mantenen una distància aproximadament constant en el període analitzat.

8.1.2.3 Variables Acadèmiques Prèvies

La variable TÍTOL D'ACCÉS caracteritza els estudiants de la cohort segons si han accedit a la universitat per la via del COU d'una banda o, de l'altra per les proves d'accés a majors de 25 anys o altres titulacions de grau mig o superior. L'anàlisi en termes de la durada dels estudis fins a l'obtenció del títol de llicenciat està condicionada pel reduït nombre d'observacions incloses en la segona modalitat de tal forma que pocs alumnes del grup llicenciats provoquen valors de la funció d'atzar elevats. Les funcions de supervivència mostren com els alumnes provinents del COU es llicencien més ràpidament que no pas els altres i que en aquest darrer grup no comencen a finalitzar la carrera fins el sisè any. Les funcions d'atzar, tal com s'ha dit, no són del tot comparables degut als resultats inestables en el grup dels que accedeixen per via diferent del COU.

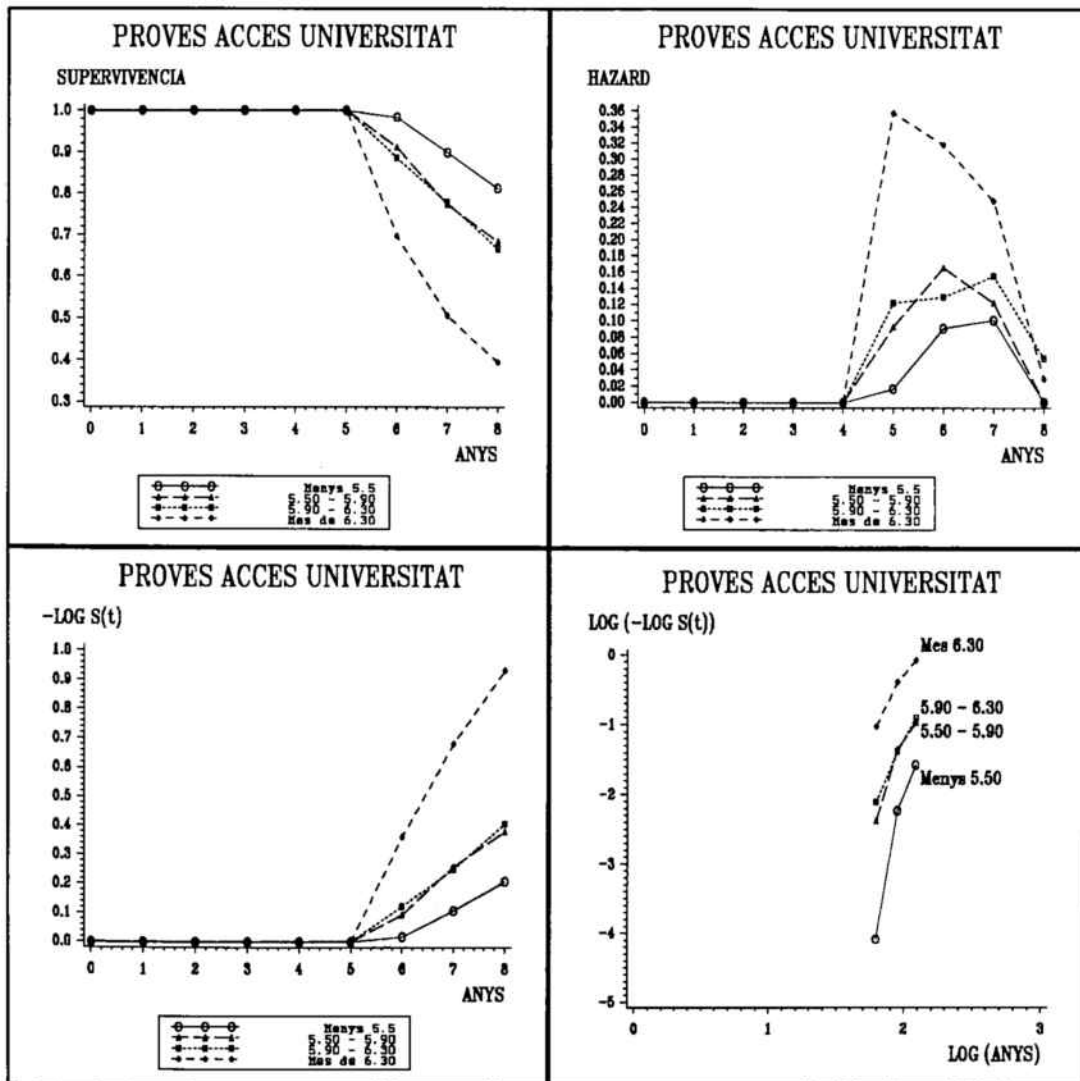
Els contrastos estadístics, malgrat les diferències existents entre les funcions de supervivència, si més no gràficament, no permeten el rebuig de la hipòtesi nul·la d'igualtat en aquelles al nivell de confiança del 5%, tot i que relaxant la significació al 10% s'obté un resultat dubtós en el contrast de Wilcoxon. Aquesta manca d'associació, estadísticament parlant, no coincideix amb el que s'observa en els gràfics es troba fortament condicionada per la inestabilitat generada en la modalitat d'Altres procedències diferents del COU. Malgrat tot, les interrelacions existents entre les variables personals i de gestió acadèmica seleccionades garantitzen que l'heterogeneïtat que recull la variable, la qual té a veure obviament amb l'Edat, la Situació laboral, etc., quedarà prou de manifest en les modelitzacions posteriors.

TÍTOL D'ACCÉS



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	3.5604	2.2351	2.5374	0.1112
LOG RANK	3.8083	2.7592	1.9051	0.1675

PROVES D'ACCÉS A LA UNIVERSITAT



Contrast	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
WILCOXON	-9531.8	866.2	121.1	0.0001
LOG RANK	-10273.7	1020.7	101.3	0.0001

Per acabar, la variable RESULTATS DE LES PROVES D'ACCÉS A LA UNIVERSITAT què és quantitativa doncs recull la qualificació global obtinguda en les proves, torna a estar clarament associada a la durada dels estudis fins a l'obtenció del títol de llicenciat tal com ha succeït en la subpoblació dels abandonaments. Malgrat tot, les representacions gràfiques de les funcions de supervivència i deduïdes d'aquesta per als diferents estrats analitzats mostren com ara les diferències més grans en termes de durades es produeixen entre els dos grups més extrems, és a dir, entre aquells que van superar la qualificació global de 6.30 punts i els que van obtenir menys de 5.50 punts. Les qualificacions situades entre aquests dos extrems no mostren diferències massa pronunciades en les durades ni en les funcions d'atzar.

Els contrastos estadístics de Wilcoxon i Log rank no contravenen els comentaris anteriors vist que, malgrat efectuar els contrastos considerant la variable com a quantitativa, l'associació amb la durada dels estudis es significativa estadísticament al nivell de confiança del 5%. No només en aquest nivell de significació sinó que la probabilitat d'obtenir un valor superior a l'estadístic de contrast mostrat en el cas que no existís associació entre les variables és inferior a una deu-mil·lèsima. Igual que en la subpoblació anterior, la gràfica dels logaritmes de les funcions d'atzar integrat garanteixen la inclusió de la variable Proves d'accés en un model d'atzar proporcional sense violar cap hipòtesi en la seva especificació. En la subpoblació dels llicenciats també és vàlid el comentari sobre l'especial importància d'aquesta variable en tant que representativa del rendiment acadèmic de l'estudiant, el qual és determinant en la progressió fins a l'obtenció del títol de llicenciat.

8.1.2.4. Conclusions Prèvies

De la mateixa forma que s'ha fet per a la subpoblació d'abandonaments, per a la de llicenciats s'han explicat els resultats dels contrastos d'associació entre les variables explicatives i la durada dels estudis així com els gràfics de la funció de supervivència i relacionades amb aquesta segons els valors de les variables explicatives. Els contrastos i els gràfics són individuals, és a dir, tractant cadascuna de les variables explicatives marginalment i separada de les altres.

Observem ara com només les variables explicatives que estan fortament associades a la durada dels estudis per a la subpoblació dels llicenciats són l'EDAT, el DOMICILI, el TIPUS DE RESIDÈNCIA, CENTRE, FONT DE FINANÇAMENT DELS ESTUDIS, HORARI, CLASSE DE MATRÍCULA i RESULTATS P.A.A.U., mentre que la SITUACIÓ LABORAL i el TÍTOL D'ACCÉS tenen una associació més feble, amb nivells de significació al voltant del 10%. Per a les variables FONT DE FINANÇAMENT DELS ESTUDIS i CLASSE DE MATRÍCULA, malgrat que en el seu conjunt són significatives, en alguna de les seves modalitats no ho són. En el cas de la variable FONT DE FINANÇAMENT DELS ESTUDIS, les modalitats AJUT PARES i ALTRES no són significatives individualment donat que tenen un comportament en termes de supervivència molt semblant. Per altra banda, en la variable CLASSE DE MATRÍCULA, de la mateixa forma que per a la subpoblació d'abandonaments, la modalitat SEMIGRATUÏTA presenta un comportament equivalent al de la modalitat ORDINÀRIA en front a la de GRATUÏTA. Aquests resultats ens portarien a recodificar aquestes dues variables dicotomitant-les per tal d'unificar les modalitats més homogènies.

Conjuntament amb l'anàlisi individual per a cadascuna de les variables, és d'interès la realització del contrast d'associació conjunt per a totes elles. D'aquesta forma serà possible detectar les associacions existents entre les variables explicatives i reconèixer la informació redundat que contenen.

A continuació es presenten dues taules amb el contrast conjunt per a totes les variables explicatives realitzat per etapes, és a dir, incloent una a una cadascuna de les variables explicatives tenint en compte que seran incloses segons sigui la seva associació amb la durada dels estudis, començant per aquella més associada.

Contrast de WILCOXON					
Variable	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment	
RESPAN	121.1	0.0001	121.1	0.0001	P.A.A.U.
DOMIC	139.5	0.0001	18.3783	0.0001	DOMICILI
CENT	145.5	0.0001	6.0594	0.0138	CENTRE
FONF2	150.1	0.0001	4.6080	0.0318	FINANC.Treball
FONF1	152.3	0.0001	2.1511	0.1425	FINANC.Beca
HORA	153.6	0.0001	1.3281	0.2491	HORARI
EDAT	154.6	0.0001	0.9641	0.3262	EDAT
SEXE	155.4	0.0001	0.8184	0.3656	SEXE
RESID1	155.8	0.0001	0.3877	0.5335	RESID.Propia
ESCI	156.6	0.0001	0.8446	0.3581	ESTAT CIVIL
TITOL	157.0	0.0001	0.3397	0.5600	TITOL ACCÈS
CLMA2	157.2	0.0001	0.2115	0.6456	Matr.Gratuïta
FONF4	157.3	0.0001	0.1349	0.7134	FINANC.Pares
LABOR	157.4	0.0001	0.1122	0.7376	SIT.LABORAL
RESID2	157.4	0.0001	0.00885	0.9251	RESID.Altres
NAC	157.4	0.0001	0.00790	0.9292	NACIONALITAT
CLMA3	157.4	0.0001	0.000057	0.9940	Matr.Ordinària

Novament, ambdós contrastos d'associació WILCOXON i LOG RANK tenen resultats coherents i mostren com només les variables RESULTATS P.A.A.U., DOMICILI, CENTRE i FONT DE FINANCIACIÓ Treball Propi, aporten significació a l'associació conjunta. Aquestes dues darreres entren en diferent ordre en el contrast conjunt i la significació en el contrast del LOG RANK per a la última en entrar, CENTRE, és del 8%.

La introducció de la resta de variables explicatives no afegeix un increment en l'associació amb la durada dels estudis que sigui significatiu (tots es troben per sobre del 10% de significació).

Contrast de LOG RANK					
Variable	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment	
RESPAN	101.3	0.0001	101.3	0.0001	P.A.A.U.
DOMIC	120.5	0.0001	19.1510	0.0001	DOMICILI
FONF2	124.5	0.0001	4.0587	0.0439	FINANC.Treball
CENT	127.6	0.0001	3.0250	0.0820	CENTRE
CLMA2	128.8	0.0001	1.2290	0.2676	Matr.Gratuïta
EDAT	129.9	0.0001	1.1264	0.2885	EDAT
HORA	130.5	0.0001	0.6338	0.4260	HORARI
ESCI	130.8	0.0001	0.3008	0.5834	ESTAT CIVIL
RESID1	131.7	0.0001	0.8854	0.3467	RESID.Pròpia
FONF4	132.1	0.0001	0.3842	0.5354	FINANC.Pares
SEXE	132.4	0.0001	0.2784	0.5977	SEXE
TITOL	132.6	0.0001	0.2123	0.6450	TITOL ACCÈS
LABOR	132.8	0.0001	0.1644	0.6851	SIT.LABORAL
FONF1	132.8	0.0001	0.0547	0.8152	FINANC.Beca
CLMA1	132.9	0.0001	0.0342	0.8533	Matr.Semigratuïta
RESID2	132.9	0.0001	0.0296	0.8635	RESID.Altres
NAC	132.9	0.0001	0.0140	0.9057	NACIONALITAT

Els resultats indiquen com algunes variables molt associades amb la durada dels estudis per a la subpoblació dels llicenciats (EDAT, HORARI, TIPUS DE RESIDÈNCIA, CLASSE DE MATRÍCULA), quan són tractades conjuntament amb la resta de variables no afegeixen gaire a l'associació conjunta. Aquest fet estaria relacionat amb les interrelacions existents entre les variables.

Les variables que han resultat ser significatives en els contrastos d'associació efectuats conjuntament són les que primerament s'inclouran en les modelitzacions posteriors. La resta de variables també seran incorporades per tal de confirmar les interrelacions existents sense descartar la possibilitat de, en el moment d'introduir en l'anàlisi les variables de rendiment acadèmic de l'alumne, puguin modificar-se les conclusions obtingudes sobre les variables més rellevants en la caracterització de l'heterogeneïtat de la subpoblació pel que fa a la durada dels estudis fins a l'obtenció del títol.

9. Resultats de la modelització paramètrica

Capítol 9

Resultats de la modelització paramètrica

Un cop efectuades les anàlisis prèvies descriptiva i d'associació de les variables explicatives amb la durada dels estudis per a les dues subpoblacions de referència, abandonaments i llicenciats, és el moment d'iniciar l'especificació i estimació dels models de supervivència. En aquest capítol presentem les diferents especificacions realitzades i els resultats de les estimacions corresponents així com una comparació entre elles. Les especificacions alternatives són: la paramètrica i les semiparamètriques contínua i discreta.

L'especificació paramètrica requereix establir la hipòtesi sobre la distribució de probabilitat que seguirà la durada dels estudis. Els resultats prèvis conjuntament amb el potencial de càlcul que proporciona el software estadístic disponible han permès la comparació entre especificacions funcionals alternatives i la selecció de la més acurada. L'especificació semiparamètrica dels models d'atzar proporcional (Cox,1972) és més flexible. Aquesta no requereix de cap hipòtesi sobre la distribució de la durada dels estudis i només realitza inferència sobre els paràmetres que acompanyen les variables explicatives.

D'altra banda, el context acadèmic en el qual es situa el treball fa necessària una especial atenció a la consideració del temps com a variable contínua o discreta, tant per les seves conseqüències sobre els mètodes d'estimació com per la diferent interpretació dels resultats obtinguts. És per això que s'han especificat i estimat els models d'atzar proporcional sota una i altra hipòtesis.

Finalment, en les especificacions dels models semiparamètrics d'atzar proporcional han estat incloses variables explicatives els valors de les quals es mantenen constants en el temps i variables explicatives els valors de les quals van canviant amb el temps. La possibilitat d'introduir aquest darrer tipus de variables resulta especialment important ja que permet incorporar l'evolució en el rendiment acadèmic de l'estudiant com a factor explicatiu de la durada dels estudis.

Tal com ha estat comentat en pàgines anteriors l'especificació paramètrica d'un model de durada requereix establir la hipòtesi sobre la distribució de probabilitat de la variable temps. Feta aquesta hipòtesi, podran estimar-se els paràmetres de la distribució de probabilitat i aquells que acompanyen les variables explicatives, de tal forma que podrem caracteritzar, d'una banda, l'escala i la forma de la distribució de probabilitat de la durada i, d'altra, l'efecte que sobre la mateixa tenen les característiques individuals.

Per a l'elecció de la distribució de probabilitat de la variable temps disposem bàsicament d'instruments empírics com ara el gràfic de la transformació de la funció de supervivència, estimada mitjançant mètodes no paramètrics (Kaplan-Meier o Taula de Vida). Així, la representació gràfica del $-\log(S(t))$ en front del temps t , i del $\log(-\log(S(t)))$ en front del logaritme del temps ($\log t$) ens poden donar una idea aproximada de quina és la distribució de la durada.

En el cas que el gràfic del $-\log(S(t))$ en front del temps sigui aproximadament una línia recta que passa per l'origen de coordenades ens trobarem davant d'una distribució exponencial. Si el gràfic del $\log(-\log(S(t)))$ en front del logaritme del temps segueix aproximadament una recta que no passa per les coordenades (0,0) tindrem una distribució Weibull, on la pendent i l'ordenada en l'origen de la recta són aproximacions als paràmetres de la distribució.

De les representacions gràfiques presentades en el capítol anterior per a cadascuna de les variables explicatives, amb funcions separades per estrats segons els valors possibles de les variables explicatives, es podria derivar la hipòtesi de què la durada dels estudis segueix una distribució Weibull. La representació del $\log(-\log(S(t)))$ en front del $(\log t)$ mostren per a les dues subpoblacions i per a totes les variables explicatives uns núvols de punts que poden considerar-se aproximacions a rectes.

Alternativament a les especificacions exponencial i Weibull s'utilitzen també altres distribucions com la Gamma, la lognormal i la loglogística. Aquestes altres especificacions són més complexes en termes formals però poden recollir més acuradament el comportament de la variable temps modelitzada, bàsicament pel fet que incorporen un nombre més gran de paràmetres. Els instruments gràfics esmentats anteriorment, però, no són gaire informatius en el cas que la durada segueixi alguna d'aquestes distribucions.

Totes aquestes circumstàncies ens porten a la necessitat de, disposant d'una eina informàtica prou potent, estimar totes les especificacions alternatives per tal de seleccionar aquella que tingui un ajust més elevat, a més d'utilitzar la informació teòrica segons la qual la distribució Gamma inclou, com a casos

particulars, a les distribucions Weibull i lognormal, mentre que la Weibull inclou a l'exponencial. Així, la disponibilitat d'un software estadístic com és el SAS (versió 6) amb el PROC LIFEREG el qual incorpora tots aquests models paramètrics ens ha permès la seva estimació i un cop feta la corresponent tasca de comparació entre models, seleccionar l'especificació més correcta mitjançant contrastos estadístics sense haver de recórrer exclusivament a la informació proporcionada pels gràfics.

Els resultats de les estimacions es presenten en les taules següents separant les anàlisis corresponents a les subpoblacions dels abandonaments i dels llicenciat. Per a les estimacions dels diferents models paramètrics han estat seleccionades aquelles variables explicatives que conjuntament tenien una associació més significativa amb la durada dels estudis, en cadascuna de les dues subpoblacions de referència.

9.1 Subpoblació d'abandonaments

Els resultats de les estimacions dels models paramètrics per a la subpoblació d'abandonaments són clars respecte a la selecció del model més adient per a l'explicació de la durada dels estudis quan intervenen les variables explicatives més associades amb la durada i sense incloure les variables de rendiment acadèmic canviants en el temps.

ESTIMACIONS DELS MODELS PARAMÈTRICS SUBPOBLACIÓ D'ABANDONAMENTS

Variable Dependent =Log(ANYS) DURADA DELS ESTUDIS
 Variable de Censura= SITUACIÓ ACADÈMICA ALUMNE
 Valor(s) de Censura= 1 (Continua) 3 (Llicenciat)
 Observacions No censurades= 645
 Observacions Censurades per la Dreta= 706

	EXPONENCIAL	WEIBULL	GAMMA	LNORMAL	LLOGISTIC
INTERCP	0.1984218 (0.678844)	0.14816704 (0.570528)	0.08475622 (0.770226)	0.03296843 (0.722702)	0.07325451 (0.7427)
EDAT	-0.0587108 (0.014527)	-0.0554447 (0.011996)	-0.0833178 (0.024169)	-0.0777913 (0.021053)	-0.0846495 (0.022694)
NAC	1.22292271 (0.305663)	1.10237045 (0.256977)	1.00208719 (0.358692)	1.06863232 (0.337827)	1.09449817 (0.317253)
RESID1	0.58468421 (0.24603)	0.51042695 (0.205927)	0.89985470 (0.305761)	0.82265053 (0.282996)	0.86786035 (0.285134)
FONF2	-0.5883751 (0.156765)	-0.514145 (0.131621)	-0.5252389 (0.174151)	-0.5487032 (0.165969)	-0.5669101 (0.163623)
FONF3	-0.6513076 (0.252838)	-0.5663815 (0.21313)	-0.6518906 (0.255916)	-0.6590359 (0.249755)	-0.670354 (0.251022)
CENT	-0.7183923 (0.19635)	-0.6135814 (0.165649)	-0.6428038 (0.161305)	-0.6524072 (0.163109)	-0.6589978 (0.166126)
HORA	0.33788537 (0.091732)	0.29302353 (0.077258)	0.33570418 (0.089314)	0.34309311 (0.087063)	0.36265841 (0.086579)
RESPAN	0.50519026 (0.078152)	0.42712014 (0.067217)	0.4593180 (0.068222)	0.46704676 (0.068466)	0.47560091 (0.069333)
SCALE	1	0.83727161 (0.029241)	1.25803297 (0.050103)	1.18021194 (0.036247)	0.68657083 (0.023152)
SHAPE			-0.3646726 (0.195427)		
LOG L	-1490.278742	-1478.211691	-1448.997955	-1450.891312	-1464.887936

SELECCIÓ DE MODELS

Weibull vs. Exponencial:

$$H_0: \text{SCALE} = 1$$

<i>Estadístic</i>	<i>ChiSquare</i>	<i>Pr>ChiSq</i>
$\frac{(0.83727161 - 1)^2}{(0.029241)^2}$	= 30.970076	0.0001

Gamma vs. Weibull:

$$H_0: \text{SHAPE} = 1$$

<i>Estadístic</i>	<i>ChiSquare</i>	<i>Pr>ChiSq</i>
$\frac{(-0.3646726 - 1)^2}{(0.195427)^2}$	= 48.762709	0.0001

Gamma vs. Lognormal:

$$H_0: \text{SHAPE} = 0$$

<i>Estadístic</i>	<i>ChiSquare</i>	<i>Pr>ChiSq</i>
$\frac{(-0.3646726 - 0)^2}{(0.195427)^2}$	= 3.482067	0.05

D'una banda, tenint en compte que el model Weibull inclou com a cas particular l'Exponencial mitjançant el paràmetre d'escala ($SCALE=1$) i un cop estimats els dos models, efectuem un contrast de Wald del que es deriva la no acceptació de la hipòtesi nul·la i, per tant, la selecció del model Weibull en front de l'Exponencial. Si tot seguit considerem el model Gamma el qual, amb les oportunes restriccions sobre el paràmetre de forma de la distribució, inclou els models Weibull ($SHAPE=1$) i Lognormal ($SHAPE=0$) com a casos particulars i efectuem els corresponents contrastos de Wald sobre el valor del paràmetre, seleccionarem el model Gamma en front de qualsevol de les altres especificacions, tot i que davant del model lognormal el contrast està vorejant la significació.

D'altra banda, si a més dels contrastos de restriccions sobre els paràmetres tenim en compte altres mesures de la bondat de l'ajust del model, constatem que pel que fa al logaritme de la versemblança ($\log L$) el model que presenta un valor absolut més petit (i per tant una versemblança més elevada) correspon també a l'especificació Gamma.

Finalment, cal destacar que malgrat haver seleccionat l'especificació Gamma, una anàlisi comparativa entre els models ens mostra la gran estabilitat tant del signe com dels valors puntuals dels paràmetres estimats per a les diferents especificacions. Aquest fet posa de manifest com, independentment de la hipòtesi efectuada sobre la distribució de probabilitat de la durada dels estudis per a la subpoblació d'abandonaments, el conjunt de variables explicatives seleccionat presenta una associació molt estable i definida amb la durada.

Pel que fa referència als paràmetres estimats, tots ells són significatius al nivell de confiança del 5% excepte en el cas del terme independent ($INTERCP$)

afegit a la combinació lineal de variables explicatives el qual no és significatiu, la qual cosa és coherent amb els resultats que s'havien obtingut mitjançant els contrastos d'associació de Wilcoxon i Log Rank.

Els paràmetres estimats es poden interpretar directament com la variació en el logaritme de la durada quan s'incrementa en una unitat el valor de la variable explicativa corresponent si aquesta és quantitativa o quan es produeix un canvi en la modalitat de les variables qualitatives. A més, permeten realitzar la predicció de la durada dels estudis per a la subpoblació d'abandonaments. La predicció del logaritme del temps serà:

$$(\log t)_p = \hat{\alpha} + z\hat{\beta} + \hat{\sigma}w_p$$

a on w_p és el percentil p de la distribució de l'error.

Habitualment es realitzen les prediccions per a la mediana de la distribució de l'error com a valor de referència, és a dir, per aquell valor que es situaria en el centre de la distribució de supervivència i que tindria el 50% d'individus amb una durada inferior o superior. A més, el càlcul de les prediccions per a un nombre suficientment gran de percentils de la distribució del terme d'error, facilita la representació gràfica de les funcions de supervivència estimades.

Com a il·lustració dels resultats i de la interpretació de les estimacions obtingudes, seleccionem dos individus arquetípics i realitzem la predicció de la durada. D'aquesta forma constatarem el significat del signe i del valor del paràmetre estimat. Els dos individus de referència tenen les següents característiques:

Individu A:

18 anys d'EDAT, de NACIONALITAT Espanyola (NAC=1), amb RESIDÈNCIA al domicili patern (RESID1=0), FINANÇAMENT DELS ESTUDIS mitjançant l'ajut dels pares o amb beca (FONF2=0 i FONF3=0), el CENTRE on cursa la llicenciatura és la Facultat d'Econòmiques (CENT=1) i en HORARI de matí (HORA=1).

Individu B:

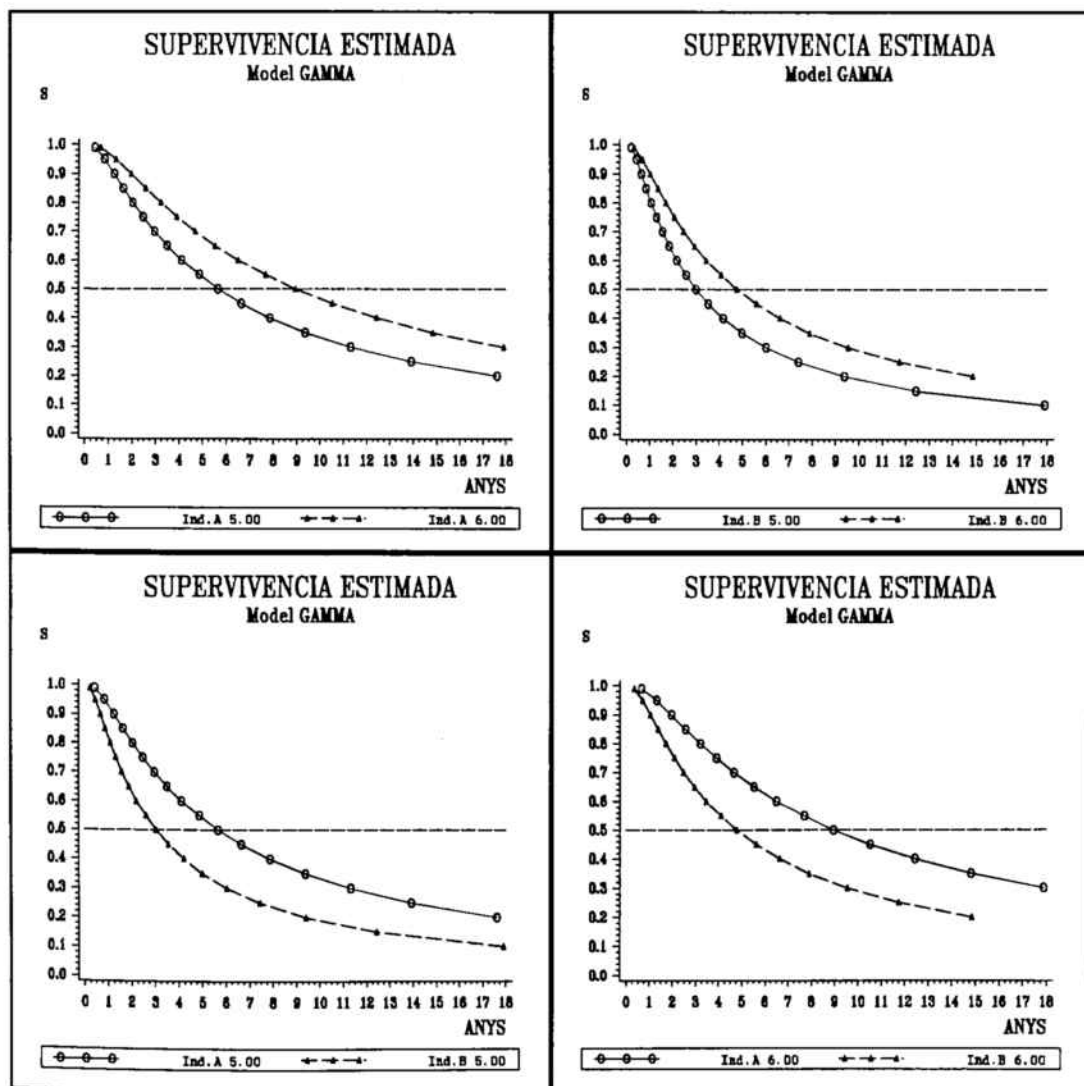
25 anys d'EDAT, de NACIONALITAT Espanyola (NAC=1), amb RESIDÈNCIA en el domicili propi (RESID1=1), FINANÇAMENT DELS ESTUDIS mitjançant el seu propi treball (FONF2=1 i FONF3=0), el CENTRE on cursa la llicenciatura és la Facultat d'Econòmiques (CENT=1) i en HORARI de tarda/nit (HORA=0).

Per ambdós individus considerarem dos resultats alternatius en les proves d'accés P.A.A.U., 5.00 i 6.00.

Les prediccions de les durades dels estudis fins a l'abandonament dels mateixos són:

	Individu A		Individu B	
	PAAU=5.00	PAAU=6.00	PAAU=5.00	PAAU=6.00
log (DURADA)	1.7317287	2.1910467	1.1874157	1.6467337
DURADA	5.65 anys	8.94 anys	3.28 anys	5.19 anys

Tot seguit presentem les representacions gràfiques de les funcions de supervivència estimades. Comparem dos a dos els individus, en primer lloc cadascun dels dos individus segons sigui la qualificació obtinguda en el global de les proves P.A.A.U., i en segon lloc, per a una mateixa qualificació en les P.A.A.U., cadascun dels dos individus A i B.



Com pot observar-se en el quadre anterior, per a un mateix individu i per tant mantenint invariants la resta de característiques personals i de gestió acadèmica, l'increment d'un punt en el resultat de les proves P.A.A.U. suposa un increment en el logaritme de la durada dels estudis igual al valor del paràmetre estimat per a la variable RESPAN, 0.4593180.

Constatem doncs per una banda, que el paràmetre estimat representa l'increment/disminució en el logaritme de la durada dels estudis en incrementar-se en una unitat la variable explicativa corresponent (si aquesta és quantitativa) i mantenint constants la resta de variables (*ceteris paribus*). En el cas de variables explicatives qualitatives pel fet d'haver estat introduïdes en el model com a dicotòmiques, el paràmetre representa la variació en el logaritme de la durada en passar d'una modalitat a una altra de l'atribut que representa. D'altra banda, el fet que es produeixi un increment o disminució en el logaritme de la durada dependrà del signe del paràmetre.

Especialment interessant resulta el càlcul de l'antilogaritme del paràmetre estimat ja que ens permet la interpretació en termes de les variacions de la durada directament en anys. La predicció de la durada mediana en anys per l'Individu A passaria de 5.65 anys a 8.94 anys pel fet d'haver obtingut un punt més en les P.A.A.U., la qual cosa representa que la durada mediana en anys s'hauria de multiplicar per l'antilogaritme del paràmetre estimat ($\exp(0.4593180)=1.583$) per a obtenir la durada mediana en anys corresponent a l'obtenció d'un punt més en els resultats de les P.A.A.U.

En conseqüència, el càlcul de l'antilogaritme del paràmetre estimat ens permet identificar aquelles variables que més "acceleren/desacceleren" la durada dels estudis. Aquelles variables amb un paràmetre estimat amb signe negatiu tindran un antilogaritme inferior a la unitat i, per tant, desacceleraran la durada dels estudis, mentre que les que tenen un paràmetre estimat de signe positiu accelereraran la durada i aquesta acceleració dependrà del valor del paràmetre. Això es traduirà gràficament en els desplaçaments de les funcions de supervivència estimades, cap a la dreta (signe positiu del paràmetre, acceleració) o cap a l'esquerra (signe negatiu, desacceleració). Pel model Gamma seleccionat els antilogaritmes dels paràmetres són:

	β	$\exp(\beta)$
EDAT	-0.0833178	0.920059
NAC	1.00208719	2.72396
RESID1	0.89985470	2.45925
FONF2	-0.5252389	0.591414
FONF3	-0.6518906	0.52106
CENT	-0.6428038	0.525816
HORA	0.33570418	1.39892
RESPAN	0.4593180	1.58299

En el model Gamma seleccionat presenten signe positiu i, per tant, acceleren la durada dels estudis fins a l'abandonament, les següents variables:

- NACIONALITAT:

Els individus de nacionalitat Espanyola allarguen més la decisió d'abandonar, en el cas de que aquesta es produeixi, que no pas els de nacionalitat Estrangera que abandonen abans.

- RESIDÈNCIA EN DOMICILI PROPI:

Els individus amb residència al domicili propi tenen una propensió a posposar la decisió d'abandonar més elevada que els que tenen altres tipus de residència (Pares o Altres).

- HORARI:

Els individus que segueixen el curs en horari de matí prenen la decisió d'abandonar més tard que els d'horari tarda/nit.

- RESULTATS P.A.A.U.:

A mesura que augmenta la qualificació obtinguda en el global de les proves d'accés a la universitat més tardana és la decisió d'abandonar els estudis.

Les variables que presenten signe negatiu, és a dir, les que desacceleren la durada prèvia a la decisió d'abandonar són:

- FONT DE FINANÇAMENT DELS ESTUDIS:

Aquells alumnes que financien els estudis mitjançant el seu treball personal o amb altres fonts diferents de l'ajut dels pares o les beques, tenen una tendència a abandonar més ràpidament els estudis.

- EDAT:

A mesura que augmenta l'edat dels individus en iniciar els estudis més curta serà l'estada a la universitat abans de l'abandonament. És a dir, els que comencen la carrera amb edats més avançades abandonen abans que els que la inicien més joves.

- CENTRE:

Els alumnes que cursen els estudis a la Facultat d'Econòmiques abandonen els estudis amb més celeritat que els que ho fan al centre adscrit Abat Oliba.

9.2 Sub població de llicenciats

ESTIMACIONS DELS MODELS PARAMÈTRICS SUBPOBLACIÓ DE LLICENCIATS

Variable Dependent =Log(ANYS) DURADA DELS ESTUDIS
Variable de Censura= SITUACIÓ ACADÈMICA ALUMNE
Valor(s) de Censura= 1 (Continua) 2 (Abandona)
Observacions No censurades= 314
Observacions Censurades per la Dreta= 1037

	EXPONENCIAL	WEIBULL	GAMMA	LNORMAL	LLOGISTC
INTERCP	7.03399512 (0.561082)	3.34646996 (0.142622)	2.40303722 (0.170029)	3.21825901 (0.140265)	3.28242333 (0.141511)
RESPAN	-0.7531981 (0.079918)	-0.2139074 (0.020559)	-0.1189517 (0.025517)	-0.2150970 (0.020954)	-0.2283628 (0.021005)
FONF2	1.42864858 (0.580586)	0.25961419 (0.13367)	0.12269574 (0.05587)	0.21058161 (0.096878)	0.24230094 (0.11483)
CENT	0.27697042* (0.178909)	0.04540422 * (0.041334)	0.06987968 (0.035618)	0.09422553 (0.041954)	0.09797137 (0.044437)
DOMIC	0.54146524 (0.139429)	0.1593116 (0.032505)	0.06832126 (0.029763)	0.1526294 (0.034351)	0.1616863 (0.034524)
SCALE	1	0.22887092 (0.010888)	0.26800055 (0.026325)	0.31091091 (0.013852)	0.18346067 (0.008643)
SHAPE			-2.7774205 (0.446895)		
LogL	-685.8884871	-401.1707266	-326.1502002	-364.3212712	-379.8643763

L'asterisc * indica significació al voltant del 10%.

SELECCIÓ DE MODELS

Weibull vs. Exponencial:

$$H_0: SCALE = 1$$

<i>Estadístic</i>	<i>ChiSquare</i>	<i>Pr>ChiSq</i>
$\frac{(0.22887092 - 1)^2}{(0.010888)^2}$	= 5016.0	0.0001

Gamma vs. Weibull:

$$H_0: SHAPE = 1$$

<i>Estadístic</i>	<i>ChiSquare</i>	<i>Pr>ChiSq</i>
$\frac{(-2.7774205 - 1)^2}{(0.446895)^2}$	= 71.446290	0.0001

Gamma vs. Lognormal:

$$H_0: SHAPE = 0$$

<i>Estadístic</i>	<i>ChiSquare</i>	<i>Pr>ChiSq</i>
$\frac{(-2.7774205 - 0)^2}{(0.446895)^2}$	= 38.625338	0.0001

Els resultats de les estimacions dels models paramètrics per a la subpoblació dels llicenciats també són clars, com en el cas de la subpoblació d'abandonaments, respecte a la selecció del model més adient per a l'explicació de la durada dels estudis quan intervenen les variables explicatives més associades amb la durada i sense incloure les variables de rendiment acadèmic canviants en el temps.

Els contrastos sobre els paràmetres d'escala i de forma de la distribució mostren com es selecciona el model Gamma en front de qualsevol de les altres especificacions donat que la hipòtesi nul·la dels tres contrastos és rebutjada sobradament. Com en el cas de la subpoblació d'abandonaments, si tenim en compte el logaritme de la versemblança observem que el valor absolut més petit (i per tant la versemblança més elevada) correspon també al model Gamma.

Pel que fa referència als paràmetres estimats, tots ells són significatius al nivell de confiança del 5%, excepte en el cas de la variable CENTRE en la qual per a les especificacions Exponencial i Weibull el paràmetre és significatiu al nivell de confiança del 10%. Aquesta mateixa variable ja presentava una associació més feble en els contrastos de Wilcoxon i Log rank realitzats anteriorment.

A diferència del que succeïa en la subpoblació d'abandonaments, ara les estimacions obtingudes per a les diferents especificacions alternatives no presenten una gran estabilitat. Particularment, per a les especificacions exponencial i Weibull s'obtenen unes estimacions puntuals molt allunyades de les de la resta d'especificacions. Així, malgrat que el signe dels paràmetres estimats és el mateix per a tots els models, els valors puntuals van, per exemple en el cas de la variable FONF2, dicotòmica que recull la modalitat FINANÇAMENT DELS ESTUDIS pel Treball Propi, des del valor 1.4286 en el cas del model exponencial fins el de 0.1226 pel model Gamma. Aquesta inestabilitat és deguda al fet que en la subpoblació dels llicenciats existeix poca variabilitat en la durada dels estudis doncs del total de vuit anys acadèmics sobre els que es disposa d'informació, en els cas dels llicenciats queda restringit a la meitat donada la impossibilitat de llicenciar-se en un temps inferior als 5 anys.

La interpretació dels paràmetres obviament serà la mateixa que per a la subpoblació anterior, tan pel que fa a la directa interpretació del paràmetre sobre el logaritme de la durada com a la interpretació de l'antilogaritme del paràmetre sobre la durada en anys. Donat que les variables explicatives incloses en el model són diferents de les anteriors, seleccionarem ara una altra parella d'individus arquetípics de la subpoblació dels llicenciats i realitzarem la predicció de la durada per tal d'interpretar les estimacions.

Els dos individus tenen les següents característiques:

Individu A:

DOMICILI a Barcelona i el seu cinturó industrial (DOMIC=1), FINANÇAMENT DELS ESTUDIS mitjançant el treball propi (FONF2=1) i el CENTRE on cursa la llicenciatura és la Facultat d'Econòmiques (CENT=1).

Individu B:

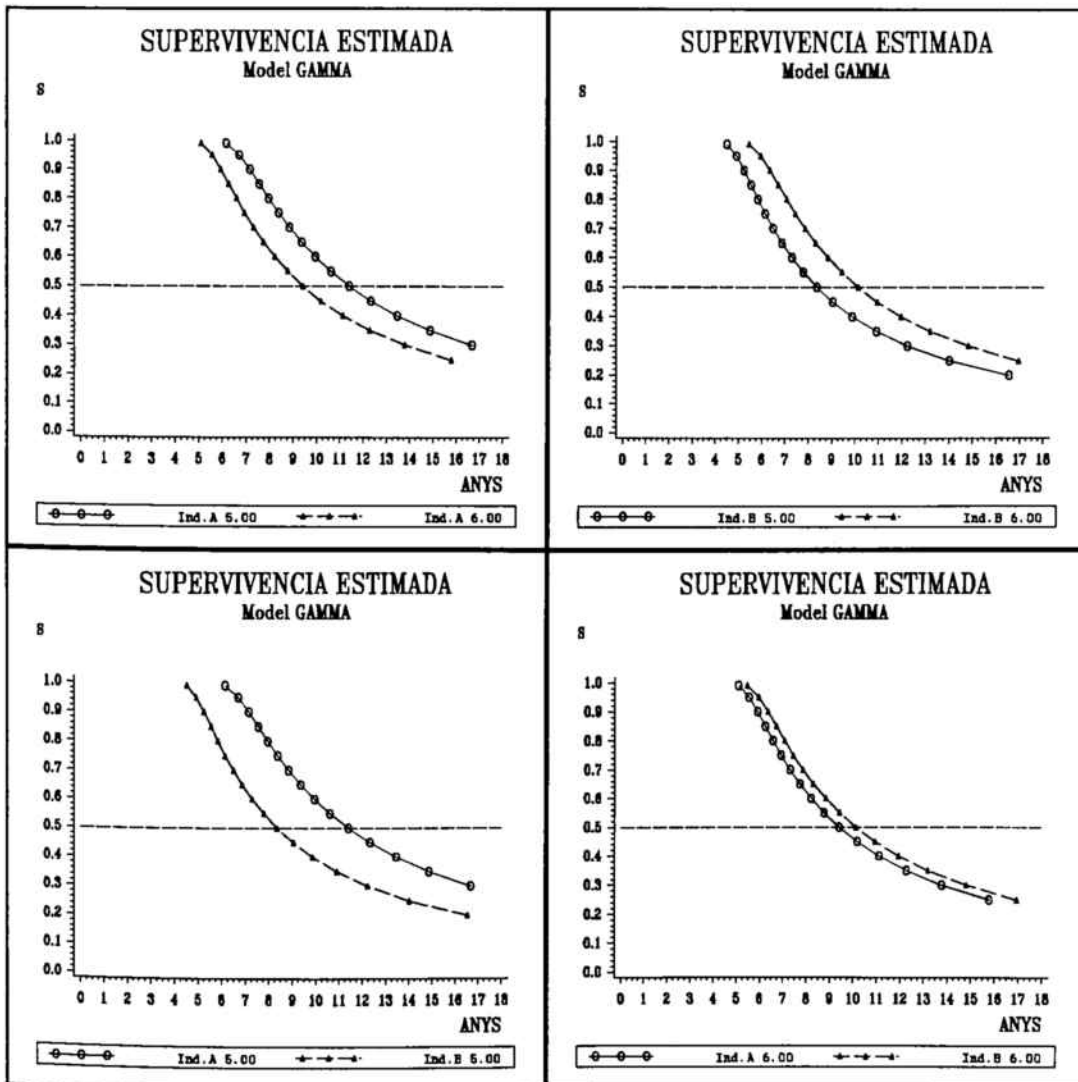
DOMICILI fora de Barcelona (DOMIC=0), FINANÇAMENT DELS ESTUDIS mitjançant qualsevol font diferent del seu propi treball (FONF2=0) i el CENTRE on cursa la llicenciatura és la Facultat d'Econòmiques (CENT=1).

Per ambdós individus considerarem dos resultats alternatius en les proves d'accés P.A.A.U., 5.00 i 6.00, com s'ha fet anteriorment.

Les prediccions de les durades dels estudis fins a llicenciar-se són:

	Individu A		Individu B	
	PAAU=5.00	PAAU=6.00	PAAU=5.00	PAAU=6.00
log (DURADA)	2.4337	2.31481	2.24273	2.12382
DURADA	11.40 anys	10.12 anys	9.42 anys	8.36 anys

Els gràfics de les funcions de supervivència estimades són:



Com pot observar-se en el quadre anterior, per a un mateix individu i per tant mantenint invariants la resta de característiques personals i de gestió acadèmica, l'increment d'un punt en el resultat de les proves P.A.A.U. suposa una disminució en el logaritme de la durada dels estudis igual al valor del paràmetre estimat per a la variable RESPAN, -0.1189517. En aquest cas, la durada mediana dels estudis per l'Individu A és, suposant una puntuació de 5.00 en les proves P.A.A.U., de 11.40 anys, la qual seria desaccelerada en el cas d'haver obtingut un punt més per un factor igual a 0.88785, restant en una durada mediana de 10.12 anys. Gràficament, l'anterior interpretació es traduiria en un desplaçament cap a l'esquerra de la funció de supervivència estimada quan comparem un mateix individu característic amb les dues qualificacions en les P.A.A.U. alternatives.

Per a la sub població dels llicenciats, els antilogaritmes dels paràmetres estimats són:

	β	$\exp(\beta)$
RESPAN	-0.1189517	0.88785
FONF2	0.12269574	1.13054
CENT	0.06987968	1.07238
DOMIC	0.06832126	1.07071

En el model Gamma seleccionat presenten signe positiu i, per tant, augmenten la durada dels estudis fins a llicenciar-se, les següents variables:

- FONT DE FINANÇAMENT DELS ESTUDIS:

Els individus que financien els estudis mitjançant el seu propi treball necessiten més anys per a llicenciar-se que els alumnes amb altres fonts de finançament.

- CENTRE:

Els alumnes que cursen els estudis a la Facultat d'Econòmiques els finalitzen més lentament que els que ho fan al centre adscrit Abat Oliba.

- DOMICILI:

Els alumnes el domicili dels quals es troba a Barcelona ciutat o en el seu cinturó industrial tenen una durada més llarga per a llicenciar-se que aquells que viuen fora de Barcelona.

La variable que presenta signe negatiu, en contraposició al que succeïa per a la subpoblació d'abandonaments és:

- RESULTATS P.A.A.U.:

A mesura que augmenta la qualificació obtinguda en el global de les proves d'accés a la universitat, més ràpidament l'alumne finalitzarà els seus estudis amb èxit.

9.3 Conclusions prèvies

En primer lloc, cal destacar la selecció del model Gamma com a especificació paramètrica més adient per a representar la durada dels estudis en totes dues subpoblacions. L'elecció d'aquesta especificació es fonamenta en els resultats de les mesures de bondat de l'ajust i en els contrastos estadístics sobre els paràmetres de la distribució, i vindria justificada per la introducció d'un paràmetre de forma de la distribució el qual afecta a l'assimetria negativa de la distribució, la presència del qual permet ajustar d'una manera més acurada les

durades més llargues. En aquest sentit, l'especificació lognormal alternativa, la qual també presenta assimetria cap a l'esquerra, proporcionava com en el cas de la subpoblació d'abandonaments, un ajust prou remarcable.

En segon lloc, respecte al conjunt de variables seleccionades que han resultat significatives en la modelització paramètrica cal fer dos aclariments. Primer, la seva significació s'ha d'interpretar com a fruit de ser les característiques personals que d'una manera més satisfactòria recullen l'heterogeneïtat existent en les durades, tant pels que abandonen com pels que es llicencien, i no pas com a factors explicatius de les durades en un sentit causal. Segon, la no incorporació de la resta de característiques personals, pel fet de no resultar significatives en la modelització paramètrica, no s'ha d'entendre com una conseqüència de la seva poca importància per a representar l'heterogeneïtat en les durades degut al fet que en molts casos les variables excloses són redundants respecte de les incloses en la modelització. Així, per exemple, la informació que proporciona la variable SITUACIÓ LABORAL vindria incorporada en les variables FONT DE FINANÇAMENT DELS ESTUDIS, en la modalitat Treball Propi, i en la variable TIPUS DE RESIDÈNCIA modalitat Domicili Propi.

10. Resultats de la modelització semiparamètrica contínua

Capítol 10

Resultats de la modelització semiparamètrica contínua

La modelització paramètrica realitzada anteriorment requereix establir la hipòtesi prèvia sobre la distribució de probabilitat de les durades, la qual cosa introdueix un cert nivell d'incertesa sobre els resultats, bàsicament derivat de les possibles conseqüències de l'error en l'especificació de la funció de distribució. A més, l'estat actual en el desenvolupament de les eines estadístiques informàtiques, possiblement conseqüència de la manca en l'estudi teòric i metodològic de la introducció de variables explicatives el valor de les quals és canviant en el temps, no ens ha permès la consideració del rendiment acadèmic com a factor explicatiu de les durades dels estudis.

La flexibilitat respecte de la hipòtesi distribucional dels models d'atzar proporcional (Cox, 1972) ha traslladat tots els esforços de desenvolupaments teòrics i de recerca aplicada envers l'estimació semiparamètrica dels models de durada. En aquest sentit, ha estat molt estudiada i s'ha implementat informàticament la possibilitat de la introducció de variables explicatives canviants

en el temps. Aquesta darrera possibilitat, conjuntament amb la flexibilitat distribucional d'aquesta modelització, ens obre moltes perspectives en la present anàlisi donat que podrem incorporar les variables de rendiment acadèmic en la modelització de la durada dels estudis. Aquestes variables són determinants ja que totes les decisions de l'alumne sobre la seva continuïtat en les llicenciatures o sobre la intensificació del seu aprofitament acadèmic estaran condicionades per quin ha estat el seu rendiment en els anys anteriors, és a dir, per quina és la seva situació objectiva per tal d'arribar a obtenir el títol de llicenciat finalitzant els estudis amb èxit.

La modelització semiparamètrica d'atzar proporcional es fonamenta en dues hipòtesis bàsiques com són, d'una banda, si la variable temps de durada a modelitzar és contínua o discreta, i d'altra banda, la hipòtesi de la proporcionalitat en els atzars o riscos quan es modifiquen els valors de les explicatives.

La consideració del temps de durada com a continu o discret vindrà en part donada per les característiques pròpies de l'anàlisi efectuada, però també tindrà molta repercusió el disseny i grandària de la base de dades disponible. Així, en la nostra aplicació sobre els estudiants universitaris, el fet de situar-nos en un entorn acadèmic significa que totes les decisions acadèmiques es produeixen en intervals discrets, generalment cada any en finalitzar el curs acadèmic, properament en acabar els semestres lectius, per la qual cosa serà més natural treballar sobre la hipòtesi del temps de durada com a variable discreta. En altres ocasions, malgrat que les decisions es prenguin en qualsevol moment del temps, segons sigui l'escala de mesura d'aquest ens podem trobar davant d'una discretització de la variable i com a conseqüència disposar de moltes observacions per a un mateix moment del temps. Segons intervingui el temps en el model d'atzar proporcional

tindrem una modelització contínua o discreta, existint mètodes d'estimació alternatius en cada cas. En aquest treball es realitzen les dues hipòtesis per tal de poder efectuar comparació de resultats i treure conseqüències per a possibles aplicacions posteriors de la metodologia.

D'altra banda, la hipòtesi de la proporcionalitat en l'atzar o risc implica que els efectes sobre el risc d'abandonar o llicenciar-se per modificacions en els valors de les explicatives haurà de ser constant al llarg del temps. És a dir, s'exigeix que variacions en el valor dels factors explicatius de la durada produeixin uns efectes sobre el risc d'abandonar o llicenciar-se, segons sigui el cas, que hauran de ser independents del moment del temps en que es mesurin, iguals en acabar el primer any que en acabar el tercer any, per exemple.

La verificació de la hipòtesi de proporcionalitat de l'atzar es realitza, habitualment, utilitzant instruments gràfics. La representació gràfica utilitzada en capítols anteriors del $\log(-\log(S(t)))$ en front del logaritme del temps ens haurà de mostrar, per a diferents valors de la variable explicativa, unes línies rectes aproximadament paral·leles o, si més no, la distància entre els núvols de punts haurà de ser constant per a tots els moments del temps. La inspecció de les esmentades gràfiques, presentades en el capítol 4, ens mostren com la hipòtesi pot mantenir-se per a les variables personals més rellevants. Més endavant es verificarà la hipòtesi d'una manera més formal.

La presència, però, de les variables de rendiment acadèmic, els valors de les quals són canviants en el temps ens col·locarà en una situació on l'esmentada hipòtesi de proporcionalitat no es mantindrà. Així, és poc raonable pensar que el rendiment acadèmic produeix uns efectes sobre el risc d'abandonar o llicenciar-se,

que són invariants respecte el moment del temps en que es mesuren: no tindrà els mateixos efectes haver superat un 10% de les assignatures després d'un any a la universitat que un cop transcorreguts 8 cursos acadèmics. Tindrem doncs, amb la introducció d'aquestes darreres variables, una doble problemàtica a resoldre: els seus valors són canviants en el temps i violen la hipòtesi de proporcionalitat del model.

A continuació presentem les estimacions del model d'atzar proporcional continu, en primer lloc incorporant únicament les variables que han estat considerades en la modelització paramètrica. D'aquesta forma realitzarem una primera anàlisi comparativa entre els models paramètrics i semiparamètrics. En segon lloc introduïrem les variables de rendiment acadèmic com a canviants en el temps i analitzarem els resultats, a més de verificar la violació de la hipòtesi de proporcionalitat.

Per a l'estimació del model d'atzar proporcional sota la hipòtesi de que el temps és continu utilitzarem el PROC PHREG que té implementat el software estadístic SAS (versió 6). L'esmentat procediment realitza l'estimació a partir de la maximització de la funció de versemblança marginal exacta (Fraser, 1968, Kalbfleisch i Sprott, 1970, i Kalbfleisch i Prentice, 1980), malgrat ser extremadament difícil en presència de moltes observacions repetides per a un mateix moment del temps, com és el nostre cas.

Tot seguit, es detallen per a cadascuna de les dues subpoblacions considerades, els resultats de les estimacions conjuntament amb les mesures habituals de bondat de l'ajust. Així, al costat del valor del paràmetre estimat es presenten el seu error estàndard, el contrast de Wald, la significació d'aquest i la

ratio del risc. Aquesta darrera informació correspon a calcular l'antilogaritme del paràmetre estimat, $\exp(\beta)$, i s'interpreta com el desplaçament en paral·lel de la funció d'atzar quan s'incrementa en una unitat la variable explicativa quantitativa o quan es produeix un canvi en les modalitats de les qualitatives.

10.1 Sub població d'abandonaments

Com es pot observar en la taula següent, tots els paràmetres són significatius al nivell de confiança del 5%. Aquest resultat és coherent amb els que havíem obtingut amb els contrastos d'associació de Wilcoxon i Log-rank, i en els models de durada paramètrics. Les variables seleccionades són les que aporten un grau més elevat de significació per tal de representar l'heterogeneïtat en les funcions de supervivència dels individus.

La interpretació dels paràmetres varia respecte l'efectuada en els models de durada paramètrics. En aquell cas l'antilogaritme del paràmetre representava l'efecte sobre la durada en anys d'una variació en l'explicativa corresponent i, per tant, els efectes de les variables explicatives es produïa directament sobre el temps de supervivència, accelerant-lo o desaccelerant-lo. En el cas dels models d'atzar proporcional, per la seva banda, l'antilogaritme del paràmetre representa l'efecte sobre la funció d'atzar base d'una variació en l'explicativa corresponent, és a dir, ara les modificacions en els valors de les explicatives no tenen efectes directament sobre el temps de durada dels estudis sinó sobre el risc de finalitzar els estudis, ja sigui per l'abandonament dels mateixos o per assolir la llicenciatura.

ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC CONTINU
SUBPOBLACIÓ D'ABANDONAMENTS
 (Sense el Rendiment Acadèmic)

Variable Dependent: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIO ACADEMICA ALUMNE
 Valor(s) de Censura: 1 (Continua) 3 (Llicenciat)
 Mètode Estimació: EXACT Màx.Versemblança Marginal

Total Observacions: 1351
 Abandonaments: 645
 Obs.Censurades: 706 % Censura: 52.26

Contrast Global Hipòtesi Nul·la: BETA=0

Criteri	Sense Covariants	Amb Covariants	Chi-Square del Model
-2 LOG L	4051.732	3867.689	184.043 amb 8 DF (p=0.0001)
Score	.	.	242.343 amb 8 DF (p=0.0001)
Wald	.	.	217.572 amb 8 DF (p=0.0001)

Variable	Paràmetre Estimat	Error Standard	Wald Chi-Square	Pr > Chi-Square	Risc Ratio
EDAT	0.059496	0.01474	16.28719	0.0001	1.061
NAC	-1.198805	0.30781	15.16766	0.0001	0.302
RESID1	-0.607051	0.24770	6.00608	0.0143	0.545
FONF2	0.629521	0.15770	15.93519	0.0001	1.877
FONF3	0.699358	0.25368	7.59998	0.0058	2.012
CENT	0.726580	0.19650	13.67299	0.0002	2.068
HORA	-0.338943	0.09199	13.57464	0.0002	0.713
RESPAN	-0.508447	0.07855	41.89814	0.0001	0.601

Així doncs, en el model d'atzar proporcional la predicció del temps de supervivència es realitza atenent a les variacions sobre la funció d'atzar degudes als canvis en les variables explicatives, la qual cosa té la seva equivalència en termes de la funció de supervivència:

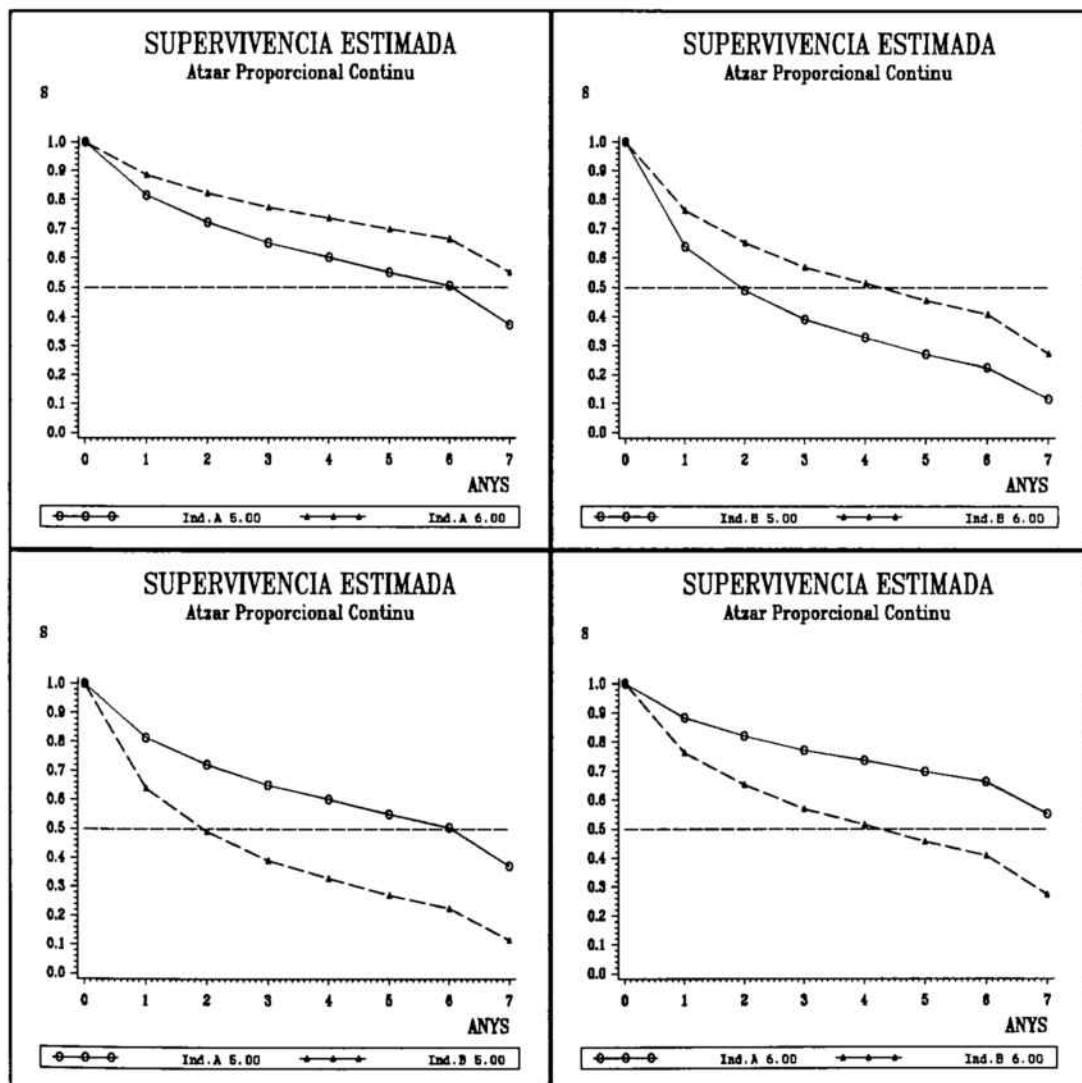
$$\hat{\lambda}(t; z) = \hat{\lambda}_0(t) \exp(z\beta) \qquad \hat{S}(t) = [\hat{S}_0(t)]^{\exp(z\beta)}$$

a on $\hat{\lambda}_0(t)$ és una funció d'atzar base arbitrària i $\hat{S}_0(t)$ és la seva corresponent funció de supervivència. Les estimacions d'aquestes funcions base s'obtenen a partir de la solució iterativa a la maximització de la funció de versemblança completa per a les durades, condicionada als valors mostrals de les durades i a la substitució dels paràmetres β per les seves estimacions obtingudes a partir de la maximització de la versemblança marginal. En el cas d'estimar el model sense variables explicatives, $z = \mathbf{0}$, les funcions base corresponents coincidiran amb les estimacions no paramètriques de Kaplan-Meier.

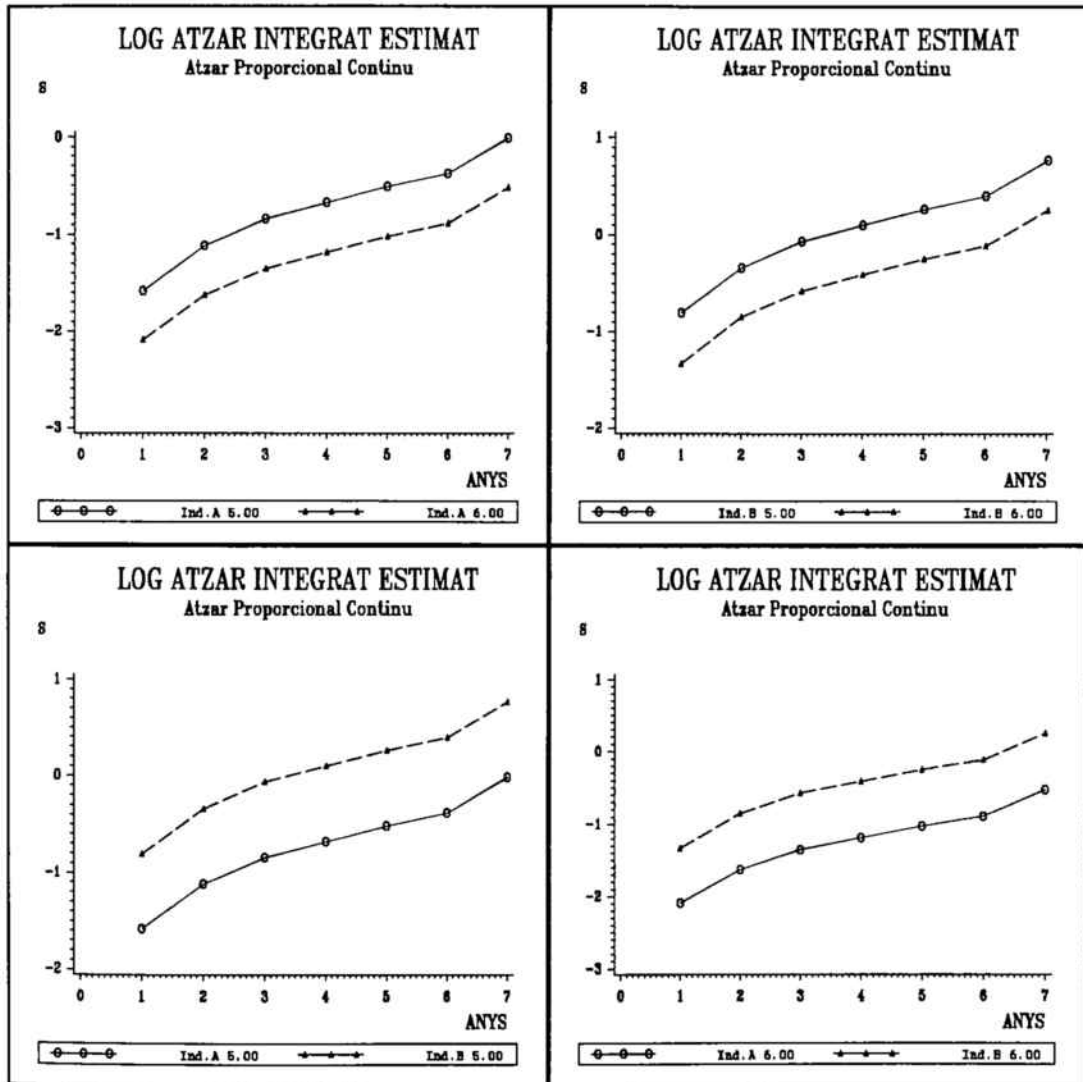
Utilitzant els mateixos individus de referència que en l'apartat anterior corresponent als models de durada paramètrics, les prediccions de les durades medianes corresponents a cadascun dels individus A i B, segons hagi estat la qualificació obtinguda en les proves P.A.A.U. es pot obtenir ara a partir de la funció de supervivència estimada mitjançant interpolació lineal. D'aquesta manera tenim:

	Individu A		Individu B	
	PAAU=5.00	PAAU=6.00	PAAU=5.00	PAAU=6.00
DURADA	6.03 anys	> 7 anys	1.93 anys	4.25 anys

Les funcions de supervivència estimades són:



A més de la representació de la funció de supervivència estimada, podem representar la funció d'atzar integrat, $\Lambda(t) = -\log S(t)$, la qual representa la funció d'atzar acumulada des del moment 0 fins el moment t, o encara més il·lustrativa és la representació gràfica del logaritme de l'atzar integrat donat que serà lineal en els paràmetres del model: $\log \Lambda(t) = \log[-\log S(t)] = \log \Lambda_0(t) + z\beta$.



Aquesta presentació gràfica mostra el desplaçament paral·lel del logaritme de l'atzar integrat, és a dir la proporcionalitat del model davant de variacions en els valors de les explicatives. Un desplaçament cap avall estarà associat a un paràmetre negatiu del model (antilogaritme inferior a la unitat) mentre que un desplaçament cap amunt estarà associat a un paràmetre positiu (antilogaritme superior a la unitat). En el següents gràfics s'observa com per a un mateix individu, A o B, un increment en la qualificació obtinguda en les proves P.A.A.U.

desplaça el logaritme de l'atzar integrat cap a baix donat que el paràmetre que acompanya la variable RESPAN pren el valor -0.508447. Això vol dir que a mesura que augmenta la qualificació en les proves P.A.A.U. es fa més petit el risc d'abandonar, la funció d'atzar es redueix en un factor proporcional al valor del paràmetre. Aquesta disminució de la funció d'atzar es tradueix en un desplaçament cap amunt de la funció de supervivència i, per tant, unes durades més llargues.

D'aquesta lectura dels resultats s'en deriva, en comparació als resultats obtinguts per a la modelització paramètrica, una interpretació dels signes dels paràmetres contrària a la que fèiem en aquella situació.

Així doncs pel model d'atzar proporcional continu tenim que presenten signe positiu i, per tant, augmenten el risc d'abandonar (funció d'atzar) i redueixen la supervivència (durada dels estudis fins a l'abandonament), les següents variables:

- FONT DE FINANÇAMENT DELS ESTUDIS:

Els alumnes que financen els seus estudis mitjançant el treball personal o d'altres fonts diferents de l'ajut dels pares o les beques, tenen un risc d'abandonament més elevat que la resta i, per tant, unes durades més curtes abans de l'abandonament.

- EDAT:

Els alumnes que inicien els estudis universitaris amb una edat més elevada tenen un risc superior d'abandonar i durades dels estudis més curtes. És a dir, aquells que comencen més joves els estudis i que prenen la decisió d'abandonar ho fan més tard que els que han iniciat els estudis amb una edat superior.

- CENTRE:

Els alumnes que cursen els estudis a la Facultat d'Econòmiques abandonen els estudis amb un risc més elevat que els que ho fan al centre adscrit Abat Oliba.

Les variables que presenten signe negatiu, és a dir, les que redueixen el risc d'abandonar i augmenten la durada prèvia a la decisió d'abandonar són:

- NACIONALITAT:

Els individus estrangers tenen un risc d'abandonar en cada moment del temps més elevat que no pas aquells alumnes de nacionalitat espanyola.

- RESIDÈNCIA EN DOMICILI PROPI:

Els alumnes que resideixen al seu propi domicili tenen una risc d'abandonament més petit que la resta i, per tant, unes durades dels estudis més llargues.

- HORARI:

Aquells alumnes amb horari de matí tenen durades dels estudis més llargues fruit d'una funció d'atzar inferior.

- RESULTATS P.A.A.U.:

En augmentar la qualificació obtinguda en el global de les proves d'accés a la universitat es produeix una reducció en la funció d'atzar i, per tant, més llargues són les durades, més improbable la decisió d'abandonar.

La interpretació dels paràmetres, doncs, malgrat ser diferent de la que es va efectuar en el cas de la modelització paramètrica de les durades, és coherent amb aquells resultats i amb els que es van obtenir dels contrastos d'associació de Wilcoxon i Log-rank. Cal destacar com el signe dels paràmetres estimats és el contrari en els models paramètrics al dels paràmetres del model d'atzar proporcional, bàsicament degut a la diferent interpretació dels efectes de les variacions en les explicatives, influint unes directament sobre el temps de durada i les altres sobre el risc de que finalitzi la durada.

D'ara en endavant, aprofitarem les possibilitats que obre la modelització de l'atzar proporcional amb la introducció de variables canviants en el temps i, en particular, el rendiment acadèmic.

La variable rendiment acadèmic, tal com s'ha comentat en el capítol corresponent a la base de dades, pot ser mesurada de formes alternatives. Així tenim, per una banda, les assignatures matriculades, superades, suspeses i no presentades. Aquestes assignatures, amb la informació proporcionada pels expedients acadèmics informatitzats, poden ser traduïdes a crèdits segons les taules d'adaptació dels plans d'estudis anteriors als plans recentment aprovats. Utilitzant les assignatures o els crèdits que representen es poden construir indicadors del rendiment acadèmic. Particularment interessant és l'Indicador de Progressió Acadèmica, IPA, (Alemany, et al, 1990) tant en la seva versió d'assignatures com la de crèdits. Aquest indicador mesura el rendiment acadèmic mitjançant el nombre d'assignatures aprovades (crèdits superats) en relació al total d'assignatures (crèdits) necessari per a llicenciar-se. Altres indicadors del rendiment acadèmic tenen en compte el nombre d'assignatures (crèdits) aprovats o superats respecte de les assignatures (crèdits) matriculades.

En aquesta darrera via es situa el criteri sobre el qual es fonamenta la normativa de permanència aprovada recentment pel Consell Social de la Universitat. Aquesta normativa regula la permanència a partir de la superació d'un nombre de crèdits acumulats superior a un percentatge del nombre de crèdits matriculats acumulats des del primer any d'inici dels estudis. Aquest és el principi general que estableix la normativa i que presenta algunes excepcions, com ara el primer any acadèmic i altres situacions excepcionals (servei militar, segons cicles, etc.). L'existència de la normativa de permanència ens permet la creació d'una variable dicotòmica per tal d'indicar si l'alumne es veuria afectat per la normativa de permanència o no, i que vindria a ser també una mesura del rendiment acadèmic de l'estudiant al llarg dels anys en el quals segueix els estudis. En les estimacions que segueixen han estat introduïdes, alternativament, totes aquestes mesures del rendiment acadèmic de l'estudiant.

La particularitat que presenta la introducció d'alguna variable de rendiment acadèmic en la modelització de les durades dels estudis està, com ja s'ha dit, en la variabilitat del seu valor en diferents moments del temps. Així, l'indicador de progressió acadèmica mostrarà com va evolucionant el rendiment acadèmic de l'estudiant any darrera any en l'objectiu final d'obtenir la llicenciatura. En aquest sentit, la introducció del concepte de variables canviants en el temps ens obliga a fer algunes puntualitzacions respecte la resta de característiques personals i de gestió acadèmica fins ara utilitzades. Entre les variables incorporades en les modelitzacions alternatives de la durada dels estudis algunes, o millor, gaire bé totes, podrien també ser interpretades com a variables els valors de les quals són canviants en el temps. Així, per exemple, les variables HORARI, SITUACIÓ LABORAL, FONT DE FINANÇAMENT DELS ESTUDIS, TIPUS DE RESIDÈNCIA, etc. podran prendre en cada curs acadèmic un valor diferent.

Malgrat tot, per a aquestes característiques personals i de gestió acadèmica només disposem d'informació sobre el seu valor en el primer any d'inici dels estudis, donat que no s'ha tingut accés a les dades de matrícula en successius cursos acadèmics i que els expedients acadèmics s'han completat a partir de diversos orígens, el que ha tingut com a conseqüència que variables de gestió acadèmica, com ara l'HORARI, en el cas d'ésser considerades com a canviants en el temps inclourien informació que no seria homogènia per a tots els individus. Per tant, en la nostra aplicació, la única variable canviant en el temps ha estat la del rendiment acadèmic tot i que gairebé totes les variables d'interès, estrictament, també ho són i, com a conseqüència, les característiques personals i de gestió acadèmica s'han d'interpretar amb el matís de que fan referència al primer curs acadèmic, és a dir, a l'any d'inici dels estudis.

A continuació presentem les estimacions tot introduint el rendiment acadèmic en les seves diverses mesures alternatives i, en primer lloc, les variables personals i de gestió acadèmica utilitzades en els models anteriors.

Com podem observar, la variable de rendiment acadèmic introduïda com a canviant en el temps és la més significativa en qualsevol de les mesures alternatives en que ha estat incorporada: IPAS és l'indicador de progressió acadèmica per assignatures (assignatures superades acumulades / total assignatures per a llicenciar-se) expressat en tant per cent. IPAC és el mateix indicador de progressió acadèmica però per crèdits utilitzant les taules d'adaptacions dels plans d'estudis anteriors als nous plans i aprovades per Junta de Govern de la Universitat de Barcelona. NORM és la variable dicotòmica que pren en consideració si l'alumne seria exclòs o no per l'aplicació de la normativa de permanència aprovada pel Consell Social.

ESTIMACIONES DEL MODELO SEMIPARAMÉTRIC CONTINU

SUBPOBLACIÓ D'ABANDONAMENTS

(Amb el Rendiment Acadèmic - Variables seleccionades)

Variable Dependents: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIO ACADEMICA ALUMNE
 Valor(s) de Censura: 1 (Continua) 3 (Licenciada)
 Mètode Estimació: EXACT MÀX. VARIÀNCIA MARGINAL

Total Observacions: 1351

Abandonaments: 645

Obs.Censurades: 706 % Censura: 52.26

Contrast Global Hipòtesi Nulla: BETA=0

Criteri	Sense Covariants		Amb Covariants		Model Chi-Square	Criteri	Sense Covariants		Amb Covariants		Model Chi-Square		
	4051.732	3071.604	980.128	(p=0.0001)			4051.732	3091.277	960.455	(p=0.0001)			
-2 LOG L	4051.732	3071.604	980.128	(p=0.0001)	4051.732	3091.277	960.455	(p=0.0001)	-2 LOG L	4051.732	3270.966	780.766	(p=0.0001)
Score	.	.	1008.385	(p=0.0001)	Score	.	990.341	(p=0.0001)	Score	.	825.192	(p=0.0001)	
Wald	.	.	689.124	(p=0.0001)	Wald	.	685.996	(p=0.0001)	Wald	.	530.918	(p=0.0001)	

Variable	Paràmetre Estimats		Error Standard		Wald Chi-Square	Variable	Paràmetre Estimats		Error Standard		Wald Chi-Square			
	-1.151237	0.31337	13.49608	(p=0.0002)			-1.158149	0.31343	13.65388	(p=0.0002)				
NAC	-1.151237	0.31337	13.49608	(p=0.0002)	NAC	-1.158149	0.31343	13.65388	(p=0.0002)	NAC	-1.034814	0.31219	10.98743	(p=0.0009)
RESID1	-0.156466	0.25160	0.38675	(p=0.540)	RESID1	-0.160583	0.25208	0.40581	(p=0.5241)	RESID1	-0.119626	0.24983	0.2927	(p=0.6321)
FONF2	0.422638	0.15722	7.22660	(p=0.0072)	FONF2	0.424003	0.15697	7.29654	(p=0.0069)	FONF2	0.384970	0.15589	6.09836	(p=0.0135)
FONF3	0.705068	0.25796	7.47070	(p=0.0063)	FONF3	0.679745	0.25711	6.98940	(p=0.0082)	FONF3	0.486076	0.25672	3.58502	(p=0.0583)
CENT	-0.208060	0.20254	1.05529	(p=0.3043)	CENT	-0.131899	0.20175	0.42741	(p=0.5133)	CENT	0.232714	0.19932	1.36322	(p=0.2430)
HORA	-0.005918	0.09438	0.00393	(p=0.9500)	HORA	-0.006082	0.09416	0.00417	(p=0.9485)	HORA	-0.101791	0.09205	1.22284	(p=0.2688)
EDAT	0.040789	0.01754	5.40706	(p=0.0201)	EDAT	0.040745	0.01735	5.51245	(p=0.0189)	EDAT	0.044723	0.01477	9.17288	(p=0.0025)
RESPAN	0.029270	0.07910	0.13694	(p=0.7113)	RESPAN	0.049916	0.07982	0.39105	(p=0.5317)	RESPAN	-0.079801	0.07755	1.05883	(p=0.3035)
IPAS	-0.081645	0.00346	555.90488	(p=0.0001)	IPAC	-0.088808	0.00377	555.04531	(p=0.0001)	NORM	-2.437254	0.12104	405.48428	(p=0.0001)

L'elevada aportació del rendiment acadèmic en l'explicació de la durada dels estudis és un resultat esperat donat que qualsevol decisió relacionada amb la finalització dels estudis estarà lligada al rendiment acadèmic, és a dir, a la situació objectiva de progressió acadèmica. Aquell alumne amb uns resultats acadèmics molt baixos obtinguts de manera sistemàtica i reiterada (un nombre reduït d'assignatures/crèdits superades/ts després de molts anys a la universitat i d'haver matriculat nombroses vegades les assignatures sense aprovar-les) obviament estarà més predisposat a abandonar els estudis que no pas aquell alumne amb un rendiment acadèmic elevat al qual resten un nombre reduït d'assignatures (crèdits) per a llicenciar-se. A més aquest factor de rendiment acadèmic estarà molt per sobre de qualsevols altres circumstàncies personals que puguin envoltar l'alumne.

D'altra banda, però, observem com la significació de les altres característiques personals i de gestió acadèmica s'ha modificat si la comparem amb els resultats obtinguts en les modelitzacions anteriors. Ara les variables TIPUS DE RESIDÈNCIA en la modalitat Domicili Propi (RESID1), CENTRE, HORARI i RESULTAT PROVES ACCÈS A LA UNIVERSITAT (RESPAN), han deixat de ser significatives la qual cosa ens obliga a fer algunes consideracions. En primer lloc la introducció de la variable corresponent al Rendiment Acadèmic pot haver modificat l'estructura de la informació redundant que contenen més d'una variable al mateix temps. D'altra banda, el propi Rendiment Acadèmic estarà associat a les característiques personals i de gestió acadèmica de tal forma que les interrelacions entre les variables del model de durada han canviat considerablement respecte les que existien en les especificacions inicials.

Com a conseqüència de l'anterior realitzem l'estimació del model d'atzar proporcional continu amb totes les variables explicatives, personals, de gestió acadèmica i de rendiment acadèmic, en aquest darrer cas amb cadascuna de les mesures alternatives del mateix. Els resultats de les estimacions venen a continuació i mostren com efectivament algunes de les variables personals o de gestió acadèmica que anteriorment no eren significatives, quan considerem el rendiment acadèmic passen a tenir rellevància. Per exemple, el SEXE, el DOMICILI, i el TIPUS DE RESIDÈNCIA ara en la modalitat Altres (diferent de Domicili Patern i Domicili Propi) són variables significatives per la sub població d'abandonaments.

Els RESULTATS P.A.A.U. (RESPAN), que en les modelitzacions anteriors (sense el Rendiment Acadèmic) tenien especial rellevància donat que eren els que aportaven un nivell d'associació més elevat amb la durada, no és significativa. Aquesta situació es pot explicar pel fet que els Resultats de les proves P.A.A.U. estava actuant com a "proxy" del Rendiment Acadèmic en el sentit que els alumnes amb uns resultats en les P.A.A.U. més elevats podem esperar que tindran un rendiment acadèmic més elevat. És per això que quan intervé en la modelització la variable Rendiment Acadèmic, aleshores la informació de la variable Resultats P.A.A.U. passa a ser redundant i la seva significació estadística desapareix.

Un comportament semblant al de la variable RESULTATS P.A.A.U. s'observa en les variables CENTRE i HORARI les quals inclouen informació redundant amb la pròpia del Rendiment Acadèmic, és a dir, existeix una associació elevada entre aquelles variables i el Rendiment, mesurat de qualsevol de les formes alternatives proposades. Bàsicament aquesta associació vindrà explicada per

ESTIMACIONES DEL MODELO SEMIPARAMÉTRIC CONTINU SUBPOBLACIÓ D'ABANDONAMENTS (Amb el Rendiment Acadèmic - Totes les Variables)

Variable Dependents: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIO ACADEMICA ALUMNE
 Valor(s) de Censura: 1 (Continua) 3 (Llicenciat)
 Mètode d'estimació: EXACT MÀX. VARIÀNCIA MARGINAL

Total Observacions: 1351
 Abandonaments: 645
 Obs.Censurades: 706 % Censura: 52.26
 Contrast Global Hipòtesi Nulla: BETA=0

Sense Covariants			Amb Covariants			Sense Covariants			Amb Covariants		
Variable	Paràmetre Estimant	Error Standard	Wald Chi-Square	Variable	Paràmetre Estimant	Error Standard	Wald Chi-Square	Variable	Paràmetre Estimant	Error Standard	Wald Chi-Square
NAC	-1.185830	0.31948	13.77723 (p=0.0002)	NAC	-1.191001	0.31927	13.91584 (p=0.0002)	NAC	-1.163183	0.31764	13.40975 (p=0.0003)
SEXE	-0.148503	0.08694	2.91788 (p=0.0876)	SEXE	-0.138703	0.08682	2.52210 (p=0.1101)	SEXE	-0.093526	0.08508	1.20831 (p=0.2717)
ESCI	-0.064765	0.44855	0.02085 (p=0.8852)	ESCI	-0.081632	0.44658	0.03341 (p=0.8550)	ESCI	-0.466053	0.40091	1.35137 (p=0.2450)
DOMIC	-0.670449	0.16156	17.22035 (p=0.0001)	DOMIC	-0.645609	0.16103	16.07391 (p=0.0001)	DOMIC	-0.478936	0.16154	8.79040 (p=0.0030)
RESID1	-0.341294	0.39365	0.75168 (p=0.3859)	RESID1	-0.350615	0.39334	0.79454 (p=0.3727)	RESID1	-0.512252	0.35381	2.09617 (p=0.1477)
RESID2	-0.431834	0.18038	5.73127 (p=0.0167)	RESID2	-0.396447	0.17968	4.86848 (p=0.0274)	RESID2	-0.318308	0.18147	3.47532 (p=0.0623)
FONF1	0.032501	0.22357	0.02113 (p=0.8844)	FONF1	0.021131	0.22356	0.00893 (p=0.9247)	FONF1	0.178875	0.21705	0.67920 (p=0.4099)
FONF2	0.327104	0.20826	2.46700 (p=0.1163)	FONF2	0.322464	0.20745	2.41615 (p=0.1201)	FONF2	0.396936	0.20214	3.85590 (p=0.0496)
FONF3	0.799776	0.26488	9.11679 (p=0.0025)	FONF3	0.768059	0.26347	8.49819 (p=0.0036)	FONF3	0.596153	0.26240	5.16172 (p=0.0231)
LABOR	0.062522	0.15091	0.17165 (p=0.6787)	LABOR	0.080811	0.15040	0.28868 (p=0.5911)	LABOR	0.053116	0.14695	0.13065 (p=0.7178)
CENT	-0.219708	0.20441	1.15527 (p=0.2824)	CENT	-0.141492	0.20367	0.48264 (p=0.4872)	CENT	0.287918	0.20122	2.04732 (p=0.1525)
HORA	-0.021060	0.09586	0.04827 (p=0.8261)	HORA	-0.018267	0.09558	0.03652 (p=0.8484)	HORA	-0.109552	0.09307	1.38555 (p=0.2392)
CLMA1	-0.077409	0.12172	0.40446 (p=0.5248)	CLMA1	-0.059247	0.12166	0.23717 (p=0.6263)	CLMA1	-0.065312	0.11963	0.29807 (p=0.5851)
CLMA2	0.011366	0.17961	0.00400 (p=0.9495)	CLMA2	-0.007797	0.17917	0.00189 (p=0.9653)	CLMA2	-0.145781	0.17539	0.69089 (p=0.4059)
TITOL	0.148795	0.22112	0.45281 (p=0.5010)	TITOL	0.132159	0.22077	0.35836 (p=0.5494)	TITOL	-0.178675	0.22738	0.61750 (p=0.4320)
EDAT	0.046264	0.01928	5.76002 (p=0.0164)	EDAT	0.046127	0.01900	5.89163 (p=0.0152)	EDAT	0.053650	0.01693	10.03623 (p=0.0015)
RESPAN	0.018778	0.08128	0.05338 (p=0.8173)	RESPAN	0.042032	0.08193	0.26322 (p=0.6079)	RESPAN	-0.098506	0.07881	1.56212 (p=0.2114)
IPAS	-0.083680	0.00354	557.54769 (p=0.0001)	IPAC	-0.090801	0.00385	557.15669 (p=0.0001)	NORM	-2.384756	0.12212	381.35788 (p=0.0001)

un Rendiment Acadèmic més elevat entre els alumnes que cursen els estudis al centre adscrit Abat Oliba i entre aquells que segueixen l'horari de matí.

Pel que fa a la variable TIPUS DE RESIDÈNCIA es produeix un canvi en la modalitat de la mateixa que és significativa. Sense el Rendiment Acadèmic la modalitat significativa era la del Domicili Propi mentre que ara ho és la modalitat Altres (diferents del domicili patern o propi). Això s'explica perquè els estudiants que resideixen en el Domicili Propi són majoritàriament individus que han iniciat els estudis amb més edat, que treballen i es financen els estudis mitjançant el seu treball, els quals s'associen a rendiments acadèmics baixos. Per contra, aquells que resideixen en altres domicilis diferents del patern o propi són individus que viuen lluny del centre on cursen els estudis i, per tant, el seu rendiment acadèmic repercutirà de forma clara sobre la decisió envers la continuïtat dels estudis, avançant l'abandonament en el cas d'uns resultats insatisfactoris.

Per a la variable SEXE cal fer un comentari especial. Aquesta variable era clarament no significativa tant en les modelitzacions sense el rendiment acadèmic com en els contrastos d'associació de Wilcoxon i Log-rank. Fins i tot, en el gràfic de les funcions de supervivència per homes i dones estimades per mitjans no paramètrics, tampoc no es posava de manifest una diferència significativa entre les mateixes. En canvi, mitjançant la introducció de les variables de rendiment acadèmic es controla aquest aspecte de l'heterogeneïtat de les durades i com a conseqüència la variable SEXE passa a ser significativa. És a dir, fixat un determinat nivell en el rendiment acadèmic, existeix evidència d'una diferent durada dels estudis fins a l'abandonament pels Homes que per les Dones. Una explicació es troba en que les Dones tenen un rendiment acadèmic més extremat (molt alt o molt baix) el que porta a diferències en les durades per raó del sexe.

El mateix que s'ha dit per a la variable SEXE és extrapolable a la variable DOMICILI de l'alumne. Aquesta variable no presentava associació amb la durada dels estudis en el cas de la subpoblació dels abandonaments mentre que amb la incorporació del rendiment acadèmic en l'anàlisi esdevé significativa. Així, fixat un determinat nivell en el rendiment acadèmic, la tendència a prendre decisions respecte la finalització prematura dels estudis és diferent per aquells alumnes que viuen a Barcelona ciutat o en el seu cinturó industrial de la que prenen els que viuen fora de Barcelona. Així, aquells alumnes que tenen el domicili en comarques allunyades de l'entorn de Barcelona, pel cost que per ells suposa el desplaçament fins al centre on cursen els estudis, tindran una més gran tendència a abandonar en el cas de que el seu rendiment acadèmic no sigui del tot satisfactori. La qual cosa és coherent amb el comentari realitzat per al canvi de modalitat significativa en la variable TIPUS DE RESIDÈNCIA.

Realitzades les anteriors consideracions sobre les variacions experimentades en el conjunt de variables significatives en la modelització de la durada dels estudis és el moment de presentar els resultats de les noves estimacions i realitzar les oportunes interpretacions dels mateixos.

En primer lloc haurem de seleccionar una mesura del rendiment acadèmic d'entre les tres alternatives amb les què hem estimat els models. Les variables IPAS i IPAC són mesures del rendiment acadèmic gairebé equivalents vist que el principi que les informa és el mateix: assignatures/crèdits acumulades/ts respecte el total d'assignatures/crèdits necessaris per a llicenciar-se. Els resultats de l'estimació dels models alternatius incorporant una i altra mesura ens mostren com efectivament són equivalents puix que tots els paràmetres estimats tenen valors semblants.

ESTIMACIONES DEL MODELO SEMIPARAMÉTRIC CONTINU

SUBPOBLACIÓ D'ABANDONAMENTS

(Amb el Rendiment Acadèmic - Noves variables seleccionades)

Variable Dependent: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
 Valor(s) de Censura: 1 (Continua) 3 (Llicenciat)
 Mètode Estimació: EXACT Màx. Versemblança Marginal

Total Observacions: 1351
 Abandonaments: 645
 Obs.Censurades: 706 % Censurat: 52.26

Contrast Global Hipòtesi Null: $\beta = 0$

Criteria	Sense Covariants	Amb Covariants	Model Chi-Square	Criteria	Sense Covariants	Amb Covariants	Model Chi-Square
-2 LOG L	4051.732	3056.051	995.681 (p=0.0001)	-2 LOG L	4051.732	3076.562	975.170 (p=0.0001)
Score	.	.	1006.418 (p=0.0001)	Score	.	.	987.956 (p=0.0001)
Wald	.	.	691.431 (p=0.0001)	Wald	.	.	686.605 (p=0.0001)

Variable	Paràmetre Estimant	Error Standard	Wald Chi-Square	Variable	Paràmetre Estimant	Error Standard	Wald Chi-Square
NAC	-1.256781	0.31563	15.85491 (p=0.0001)	NAC	-1.249565	0.31540	15.69614 (p=0.0001)
SEXE	-0.143204	0.08631	2.75312 (p=0.0971)	SEXE	-0.135014	0.08620	2.45312 (p=0.1173)
DOMIC	-0.634998	0.15975	15.80082 (p=0.0001)	DOMIC	-0.609570	0.15900	14.69838 (p=0.0001)
RESID2	-0.395608	0.17693	4.99963 (p=0.0254)	RESID2	-0.363004	0.17606	4.25098 (p=0.0392)
FONF2	0.370837	0.14709	6.35597 (p=0.0117)	FONF2	0.380891	0.14639	6.76941 (p=0.0093)
FONF3	0.704345	0.25062	7.89829 (p=0.0049)	FONF3	0.672479	0.24883	7.30356 (p=0.0069)
EDAT	0.042806	0.01716	6.22234 (p=0.0126)	EDAT	0.041399	0.01701	5.92040 (p=0.0150)
IPAS	-0.083145	0.00339	602.40444 (p=0.0001)	IPAC	-0.090352	0.00369	599.45514 (p=0.0001)

I no només les estimacions puntuals dels paràmetres són estables sinó també les estimacions de les variàncies dels estimadors, els errors estàndards, no varien significativament d'un model a l'altre. Malgrat aquesta estabilitat en els paràmetres estimats, no són exactament iguals donat que la traducció de les assignatures a crèdits incorpora una determinada ponderació entre les assignatures. És a dir, en la variable IPAS de progressió acadèmica segons assignatures totes elles tenen el mateix pes o importància en la formació del valor de l'indicador, en concret el seu pes serà igual a l'invers del total d'assignatures necessàries per a llicenciar-se. Per contra, en la variable IPAC de progressió acadèmica segons crèdits no totes les assignatures tenen el mateix pes o importància en la formació de l'indicador per bé que existiran assignatures de més crèdits que unes altres.

Pel que fa als resultats globals de l'ajust del model incorporant com a rendiment acadèmic les variables IPAS i IPAC observem com l'ajust del model és més elevat en el cas de l'Indicador de Progressió Acadèmica per assignatures. També la significació individual de cadascuna de les variables en els models corresponents és més elevada per a la progressió en termes d'assignatures que no pas en termes de crèdits. Aquests resultats eren els esperats donat que les decisions dels alumnes pertanyents a la cohort que ha servit per a realitzar la present aplicació encara no podien estar fonamentades en els crèdits, corresponents als nous plans d'estudis aprovats molt recentment, i conseqüentment, és el nombre d'assignatures superades respecte el total d'assignatures necessàries per a finalitzar els estudis sobre el què l'alumne es basa per a abandonar els estudis.

D'altra banda, si fixem l'atenció sobre la tercera mesura del rendiment acadèmic alternativa que hem anomenat NORM i que representa la implementació de la normativa de permanència en l'anàlisi de la durada dels estudis, constatem

uns resultats particularment diferents als anteriors. Segons els resultats obtinguts la variable NORM no representaria tan acuradament el rendiment acadèmic com ho feien les variables IPAS i IPAC anteriorment analitzades. L'estabilitat en les estimacions puntuals dels paràmetres i dels seus errors estàndards contrasta ara amb uns valors molt diferents i amb significacions estadístiques contraries. Així per exemple, la variable SEXE que resultava significativa estadísticament en els models anteriors passa a ser irrellevant quan considerem la variable NORM indicadora de l'exclusió o no de la universitat pel rendiment acadèmic insatisfactori. D'igual forma, la bondat de l'ajust per aquest darrer model és molt més baixa que les que s'obtenien en les especificacions anteriors.

Els resultats semblen indicar que les variables construïdes per a ser representatives del rendiment acadèmic de l'alumne tenen comportaments diferents segons es tracti de les de progressió acadèmica (tant en termes d'assignatures com de crèdits) o bé de la superació de la normativa de permanència.

El diferent comportament estadístic de les variables ens fa pensar en la possibilitat que cadascun del dos tipus de mesures està captant informació diferenciada. Per una banda tindríem una mesura del rendiment acadèmic la qual és directament interioritzada per l'estudiant i que tindrà en compte en el moment de prendre les decisions sobre la continuïtat dels estudis. D'altra banda, una mesura del rendiment acadèmic sobre la qual la universitat i, en concret, el Consell Social, han fixat uns mínims per sota dels quals l'estudiant no pot continuar, les trajectòries mínimes. Obviament la variable NORM indicadora de si l'alumne supera o no la normativa de permanència al llarg dels anys en els quals estudia a la universitat és una "proxy" del rendiment acadèmic però amb connotacions diferents als indicadors de progressió atès que incorpora una frontera

arbitrària prefixada del què s'entén per rendiment acadèmic acceptable i inacceptable.

Aquesta dualitat en la interpretació dels indicadors del rendiment acadèmic, que hem anat incorporant en la modelització de la durada dels estudis per a la subpoblació dels abandonaments, fa que sigui possible incorporar ambdues mesures en un mateix model sense haver d'esperar que la informació redundat existent en les variables de rendiment ocasioni inestabilitats sobre les estimacions puntuals i els seus errors estàndards. Introduint ambdós tipus de variables recollirem d'una banda una variable de rendiment acadèmic més subjectiva, sobre la qual l'individu fonamenta les seves decisions respecte la finalització prematura dels estudis, a la vegada que tindrem una variable de rendiment acadèmic més objectiva què comportaria una finalització normativa dels estudis. Així, podrem comparar les durades, mitjançant les funcions de supervivència estimades, pels estudiants que haurien estat exclosos respecte els altres tot i tenint en compte altres factors com les característiques personals i de gestió acadèmica.

La taula següent mostra els resultats de l'estimació dels paràmetres en el model d'atzar proporcional continu amb la presència de variables canviants en el temps com ara el rendiment acadèmic i la normativa de permanència a la vegada. Com a expressió del rendiment acadèmic que hem convingut en qualificar de subjectiu ha estat seleccionat l'indicador de progressió acadèmica en termes de crèdits malgrat que la mesura alternativa en termes d'assignatures era estadísticament més significativa i proporcionava una bondat de l'ajust més elevada. Aquesta selecció no ha estat pas arbitrària sinó que es basa en què la implantació dels nous plans d'estudis farà que aviat els estudiants realitzin els seus càlculs sobre el que els hi resta per a llicenciar-se en termes de crèdits (particular-

**ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC CONTINU
SUBPOBLACIÓ D'ABANDONAMENTS
(Rendiment Acadèmic i Normativa de Permanència)**

Variable Dependent: ANYS DURADA DELS ESTUDIS
Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
Valor(s) de Censura: 1 (Continua) 3 (Llicenciat)
Mètode Estimació: EXACT Màx.Versemblança Marginal

Total Observacions: 1351
Abandonaments: 645
Obs.Censurades: 706 % Censura: 52.26

Contrast Global Hipòtesi Nul·la: BETA=0

Criteria	Sense Covariants	Amb Covariants	Chi-Square del Model
-2 LOG L Score	4051.732	2973.208	1078.524 amb 9 DF (p=0.0001)
Wald	.	.	1022.320 amb 9 DF (p=0.0001)
	.	.	702.559 amb 9 DF (p=0.0001)

Variable	Paràmetre Estimat	Error Standard	Wald Chi-Square	Pr > Chi-Square	Risc Ratio
NAC	-1.235310	0.31961	14.93898	0.0001	0.291
SEXE	-0.152284	0.08588	3.14403	0.0762	0.859
DOMIC	-0.695229	0.16095	18.65866	0.0001	0.499
RESID2	-0.415181	0.17742	5.47612	0.0193	0.660
FONF2	0.351550	0.14558	5.83172	0.0157	1.421
FONF3	0.664031	0.24889	7.11800	0.0076	1.943
EDAT	0.037810	0.01653	5.23302	0.0222	1.039
IPAC	-0.070693	0.00408	300.28488	0.0001	0.932
NORM	-1.316874	0.13682	92.63449	0.0001	0.268

ment en termes de crèdits troncalis o obligatoris, optatius i de lliure elecció) més que no pas en termes d'assignatures per a llicenciar-se, sobre tot si tenim en compte la flexibilitat que obre el nou sistema en quant al propi disseny curricular del estudis per l'alumne.

Les perspectives de futur i la voluntat de què el present treball, més que representar una situació actual concreta en un centre determinat, vol ser una eina que faciliti a les autoritats acadèmiques, d'una banda, el coneixement en tot moment de l'evolució i l'estat dels estudiants i, per altra, el control sobre la normativa de permanència i els seus paràmetres, és el que ha derivat en la selecció de la variable IPAC per davant de la IPAS. Tot i això, els resultats d'ambdues opcions no són pas massa diferents com per a fer de la selecció d'una o altra variable una decisió molt compromesa. La importància de seguir una o altra via queda restringida al terreny de la interpretació dels resultats.

Centrant-nos en aquesta darrera qüestió, la interpretació dels paràmetres estimats en aquest model és la mateixa que havíem realitzat pel model d'atzar proporcional sense el rendiment acadèmic, és a dir, l'antilogaritme del paràmetre estimat representa el desplaçament cap amunt (o cap avall) de la funció d'atzar base en incrementar-se en una unitat la variable corresponent, si aquesta és quantitativa. La direcció del desplaçament dependrà del signe del paràmetre essent cap amunt si el paràmetre és positiu (antilogaritme més gran que la unitat) i cap avall si és negatiu (antilogaritme inferior a la unitat). Vist que el model es formula amb variacions sobre la funció d'atzar quan es modifiquen els valors de les variables explicatives, la interpretació respecte la durada dels estudis serà la inversa i, per tant, un paràmetre positiu indicarà una disminució en la durada dels estudis (desplaçament cap amunt de la funció d'atzar, més risc d'abandonar). Cal

matissar que els desplaçaments de les funcions d'atzar són els mateixos per a qualsevol moment del temps en que es mesurin (són invariants respecte el temps) en virtut de la hipòtesi de proporcionalitat.

Així les coses, les variables que suposen un desplaçament cap amunt de la funció d'atzar i , en conseqüència, una disminució de la durada dels estudis són:

- FONT DE FINANÇAMENT DELS ESTUDIS:

Els alumnes que financen els seus estudis mitjançant el treball propi o amb altres fonts diferents de l'ajut dels pares o les beques, tenen un risc d'abandonament més elevat que la resta d'alumnes i , per tant, unes durades més curtes prèvies a l'abandonament.

- EDAT:

En augmentar l'edat d'inici dels estudis universitaris també augmenta el risc d'abandonar i les durades dels estudis són més curtes.

Per altra banda, les característiques que disminueixen el risc d'abandonar són:

- NACIONALITAT:

Els estudiants de nacionalitat estrangera tenen un risc d'abandonar superior al dels estudiants de nacionalitat espanyola, mantenint constants la resta de variables explicatives del model.

- SEXE:

Les dones tenen un risc més gran d'abandonar que no pas els homes. Les durades dels estudis seran doncs més curtes per les dones. Aquesta

interpretació, com les anteriors, està subjecte a la fixació de la resta de variables introduïdes en el model en uns valors determinats (ceteris paribus).

- DOMICILI:

Aquells alumnes que tenen el domicili a Barcelona ciutat o en el seu cinturó industrial tenen durades dels estudis més llargues i, per tant, un risc més petit d'abandonament. Els estudiants que viuen lluny de Barcelona es veuen més condicionats per la distància en la decisió d'abandonar, mantenint invariants la resta de característiques personals, de gestió acadèmica i el rendiment acadèmic.

- TIPUS DE RESIDÈNCIA:

El tipus de residència diferent al domicili patern o propi, és a dir, els col·legis majors, pisos d'estudiants, etc. comporta un risc d'abandonament més petit i durades més llargues.

Comentari a part mereixen les variables de rendiment acadèmic i la normativa de permanència. Totes dues variables presenten un paràmetre estimat amb signe negatiu el que suposa un antilogaritme del paràmetre inferior a la unitat i, per tant, un desplaçament cap avall en augmentar en una unitat la variable explicativa quantitativa o en passar de la modalitat de referència a l'altra modalitat en el cas de les qualitatives. Aquesta reducció en el risc d'abandonar implica unes durades més llargues. En conseqüència, en augmentar l'Indicador de Progressió Acadèmica mesurat en crèdits (el mateix s'hagués obtingut amb les assignatures) es redueix el risc d'abandonar i les durades dels estudis són més llargues. D'altra banda, tenint en compte que la modalitat de referència (NORM=0) per a la

variable Normativa de Permanència és l'exclusió per no superar la normativa, es comprova com aquells alumnes que superen la normativa tenen un risc d'abandonar inferior i unes durades més llargues mantenint constants la resta de característiques individuals.

La interpretació d'aquestes dues variables, però, ha de tenir present que són variables canviants en el temps. L'indicador de progressió acadèmica va prenent valors diferents en acabar cada curs acadèmic segons hagin estat els resultats acadèmics de l'alumne i evolucionant cap al total de crèdits necessaris per a llicenciar-se. Al mateix temps, la normativa de permanència haurà de ser superada any darrera any per a poder continuar els estudis, és a dir, un cop no s'hagi superat en un curs acadèmic determinat el mínim de crèdits establert per la normativa atenent al nombre total de crèdits matriculats ponderat pel factor de repetició, la variable prendrà el valor zero en els successius anys acadèmics indicant que hauria estat exclòs. Malgrat prendre valors diferents en moments diferents del temps el paràmetre estimat per a la variable corresponent és únic i mesura l'efecte sobre el risc o la durada de manera invariant en el temps. D'aquesta forma, encara que el rendiment acadèmic d'un estudiant sigui molt erràtic i canviant en el temps, l'efecte sobre l'atzar o la durada dels estudis que mesura el paràmetre estimat és un efecte constant i independent del temps. Aquesta hipòtesi de proporcionalitat restringeix en gran mesura els efectes del rendiment acadèmic diguem-ne "subjectiu" per bé que hauria de ser possible que els efectes sobre la durada dels estudis o sobre la decisió d'abandonar fruit d'una variació en el rendiment (augmentar/disminuir el nombre de crèdits superats) fossin diferents, per exemple, en el primer curs acadèmic que després de cinc cursos acadèmics. En el primer curs acadèmic, l'efecte psicològic de no superar cap assignatura o pocs crèdits faria que els alumnes en aquesta situació

abandonéssin en un percentatge molt elevat mentre que després d'uns anys a la universitat els resultats dolents en un curs acadèmic concret (pocs crèdits superats) no haurien de tenir el mateix efecte que en un primer curs.

Per les raons anteriorment exposades, a més de per altres raons purament de rigor metodològic, hauríem de verificar la hipòtesi de la proporcionalitat en l'atzar davant de variacions en les explicatives. A més d'efectuar aquesta verificació de la hipòtesi també restaria pendent la tasca de càlcul de les prediccions i, en conseqüència, de les funcions de supervivència i d'atzar integrat estimades, tal com s'ha realitzat anteriorment i vist que proporciona una lectura més clara dels resultats de les estimacions. Especialment útil serà en el nostre cas seleccionar uns individus representatius (com en anàlisis anteriors) i representar les seves funcions estimades de manera que podem calcular de forma aproximada unes durades medianes, mesures que caracteritzaran els diferents grups d'individus. Aquests individus representatius haurien de ser seleccionats de tal forma que resultés factible la comparació entre les durades medianes d'aquells individus que superen la normativa de permanència i les d'aquells altres que serien exclosos.

Totes dues tasques a fer, verificació de la hipòtesi de proporcionalitat i estimació de les funcions de supervivència i d'atzar integrat, són especialment complexes en el cas de treballar amb el model d'atzar proporcional amb el temps continu o, més ben dit, en el cas d'especificar el model d'atzar proporcional discret els objectius anteriorment formulats són metodològicament més accessibles a la vegada que els resultats més intel·ligibles. És per això que posposem les esmentades anàlisis fins a la presentació del model discret i comentarem a partir d'aquí els resultats del model d'atzar proporcional continu per a la subpoblació del llicenciats.

10.2 Subpoblació de llicenciats

Tal com s'ha realitzat per a la subpoblació d'abandonaments, en primer lloc presentarem i comentarem els resultats de la modelització d'atzar proporcional contínua. Posteriorment introduïrem les variables de rendiment acadèmic i analitzarem les modificacions produïdes respecte els resultats anteriors.

En la taula següent observem com tots els paràmetres estimats són significatius a un nivell de confiança del 5% excepte en el cas de la variable CENTRE la qual seria significativa al 8%. Aquests resultats són coherents amb els que havíem obtingut inicialment amb els contrastos d'associació de Wilcoxon i Log-rank i amb la modelització paramètrica de les durades dels estudis, fins i tot pel que respecta a la variable CENTRE ja que en aquells casos també era menys significativa que les altres variables. En el seu conjunt els paràmetres també són clarament significatius.

La interpretació dels paràmetres és la mateixa que la realitzada en el cas de la subpoblació d'abandonaments, és a dir, l'antilogaritme del paràmetre mesura l'efecte sobre la funció d'atzar base d'una modificació en el valor de la variable (segons sigui quantitativa o qualitativa) explicativa corresponent.

Realitzant les prediccions sobre un conjunt d'individus característics podem representar gràficament les funcions de supervivència estimades per a cadascun d'ells i calcular les durades medianes.

ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC CONTINU
SUBPOBLACIÓ DE LLICENCIATS
(Sense el Rendiment Acadèmic)

Variable Dependent: ANYS DURADA DELS ESTUDIS
Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
Valor(s) de Censura: 1 (Continua) 2 (Abandona)
Mètode Estimació: EXACT Màx.Versemblança Marginal

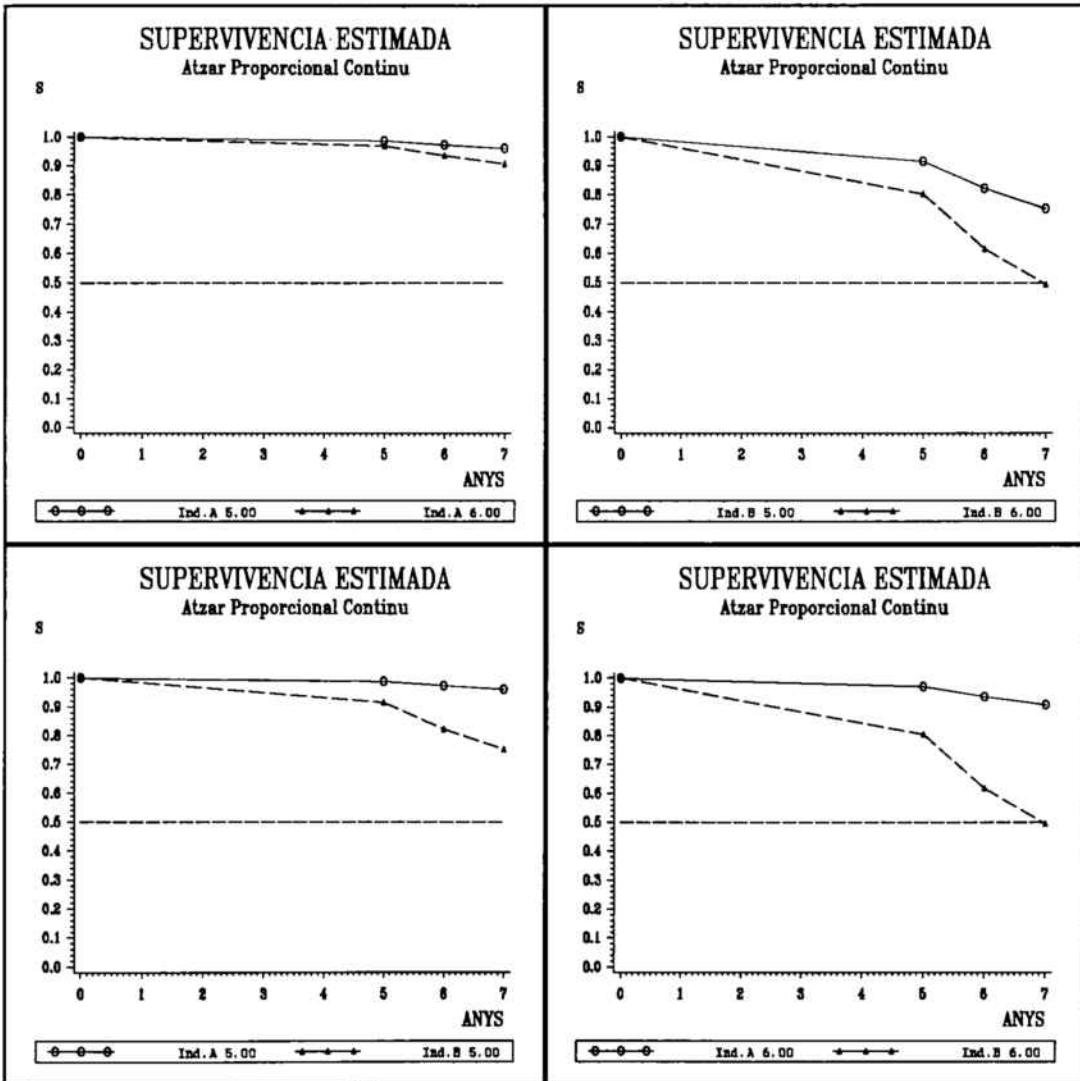
Total Observacions: 1351
Llicenciats: 314
Obs.Censurades: 1037 % Censura: 76.76

Contrast Global Hipòtesi Nul·la: BETA=0

Criteri	Sense Covariants	Amb Covariants	Chi-Square del Model
-2 LOG L Score	1814.825	1675.393	139.431 amb 4 DF (p=0.0001)
Wald	.	.	154.810 amb 4 DF (p=0.0001)
			149.435 amb 4 DF (p=0.0001)

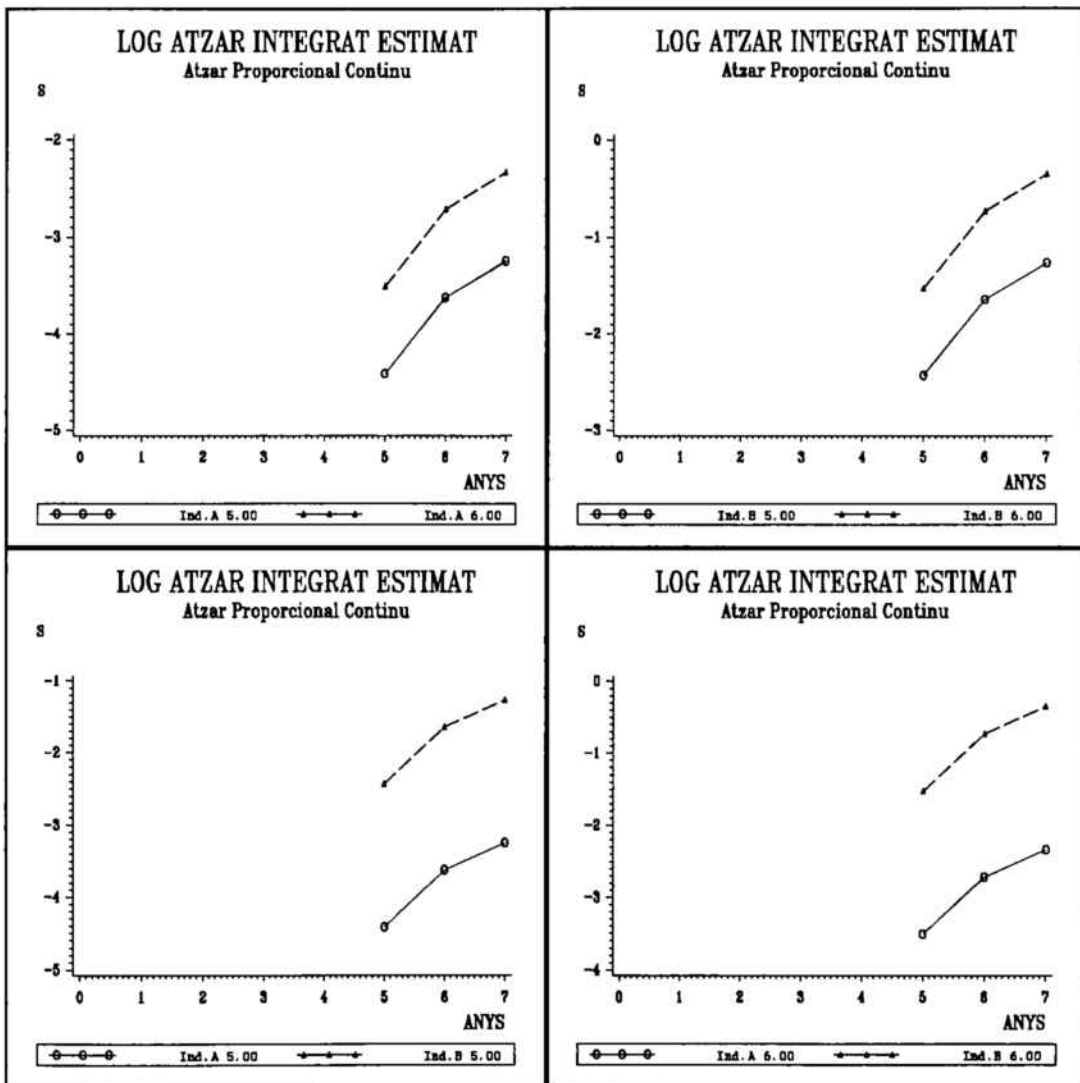
Variable	Paràmetre Estimat	Error Standard	Wald Chi-Square	Pr > Chi-Square	Risc Ratio
FONF2	-1.280817	0.58068	4.86519	0.0274	0.278
RESPAN	0.907839	0.08421	116.21991	0.0001	2.479
CENT	-0.311341	0.18018	2.98573	0.0840	0.732
DOMIC	-0.700956	0.14062	24.84686	0.0001	0.496

Pels mateixos individus característics seleccionats en la modelització paramètrica i segons les qualificacions alternatives (5.00 i 6.00) obtingudes en les Proves d'Accés a la Universitat, les funcions de supervivència estimades són les següents:



De les representacions gràfiques de les funcions de supervivència estimades es deriva que les durades medianes són totes superiors als 8 anys excepte en el cas de l'individu B en el cas d'haver obtingut una qualificació de 6.00 en les proves P.A.A.U. pel qual la durada mediana estimada per interpolació és de 6.96 anys.

D'altra banda, mitjançant la representació gràfica del logaritme de la funció d'atzar integrat estimat podrem visualitzar els efectes de les variacions en les característiques individuals:



Pot observar-se com en els dos primers gràfics el desplaçament de les funcions és de la mateixa magnitud i igual al valor del paràmetre associat a la variable RESPAN, Proves Accés a la Universitat, (0.907839), doncs els primers gràfics corresponen a un mateix individu amb dues qualificacions alternatives en les P.A.A.U. (5.00 i 6.00). Per la seva banda, els dos següents gràfics també presenten el mateix desplaçament en paral·lel de les funcions, el qual és equivalent a la combinació dels paràmetres estimats corresponent a les característiques

individuals que diferencien els individus A i B què són la FONT DE FINANÇAMENT DELS ESTUDIS (FONF2) i el DOMICILI (DOMIC) i per tant el desplaçament serà de $(-1.280817 - 0.700956) = -1.98178$. De la mateixa manera que en el cas de la sub població d'abandonaments, el signe del paràmetre és el que determina si el desplaçament del logaritme de la funció d'atzar integrat és cap amunt o cap avall o, el que és el mateix, si la funció d'atzar augmenta o disminueix.

En la sub població de llicenciats i sense incorporar el rendiment acadèmic, les variables que tenen un paràmetre estimat amb signe negatiu i, per tant, desplacen el logaritme de l'atzar integrat cap avall respecte la funció base són la FONT DE FINANÇAMENT DELS ESTUDIS en la modalitat Treball Propi, CENTRE i DOMICILI, mentre que la variable RESULTATS PROVES P.A.A.U. presenta un signe positiu i, per tant, desplaça la funció cap amunt. És a dir, aquells alumnes amb unes qualificacions en les P.A.A.U. més elevades tenen un risc més gran de finalitzar amb èxit la carrera, mentre que els que tenen el Domicili a Barcelona i el seu cinturó industrial, els que cursen els estudis en la Facultat d'Econòmiques i els que es financen els estudis mitjançant el seu propi treball tenen un risc menor de finalitzar amb èxit els estudis.

De la mateixa manera que en les anàlisis anteriors, aprofitarem els avantatges de l'especificació del model d'atzar proporcional per a introduir les variables explicatives canviants en el temps i, en particular, les variables de rendiment acadèmic. S'han utilitzat les mateixes variables indicadores del rendiment acadèmic que en les anàlisis anteriors així com la normativa de permanència.

D'igual forma que en el cas de la subpoblació d'abandonaments, la introducció de les variables de rendiment acadèmic ha produït variacions respecte de la rellevància de les variables personals i de gestió acadèmica seleccionades, essent en aquest cas encara més radical el canvi per bé que cap de les variables presents en el model, a excepció de la de rendiment acadèmic és significativa en absolut. Anteriorment s'han comentat en detall les interrelacions existents entre les variables personals i de gestió acadèmica, entre elles mateixes i amb les variables de rendiment acadèmic. Aquestes interrelacions porten a situacions com l'anterior a on variables els paràmetres de les quals eren estadísticament diferents de zero, en incorporar el rendiment acadèmic deixen de ser significatives. La informació que proporcionen aquestes variables ja està incorporada en el rendiment acadèmic i la redundància es manifesta en la irrellevància de les variables.

L'estimació dels models amb totes les variables personals i de gestió acadèmica amb cadascuna de les mesures del rendiment acadèmic ens mostren uns resultats dels quals es poden destacar diverses qüestions. En primer lloc, a diferència del que s'havia posat de manifest en la subpoblació d'abandonaments, en aquest cas les mesures del rendiment acadèmic IPAS (Indicador de Progressió Acadèmica per assignatures) i IPAC (Indicador de Progressió Acadèmica per crèdits) no presenten comportaments homogenis. Les estimacions puntuals dels paràmetres així com els seus corresponents errors estàndards no es mantenen entorn dels mateixos valors i donen lloc a nivells de significació totalment oposats per a una mateixa variable. A més, les mesures de la bondat de l'ajust mostren uns resultats qualitativament diferents als obtinguts en l'anàlisi dels abandonaments. Pel que fa a la variable NORM, la qual recull la superació de la normativa de permanència, també els resultats són poc estables i, en aquest sentit, el seu comportament és semblant a l'observat en la subpoblació d'abandonaments.

ESTIMACIONES DEL MODELO SEMIPARAMÉTRIC CONTINU SUBPOBLACIÓ DE LLICENCIATS (Amb el Rendiment Acadèmic - Variables seleccionades)

Variable Dependent: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
 Valor(s) de Censura: 1 (Continua) 2 (Abandona)
 Mètode Estimació: EXACT MÀX. VARIÀNCIA MARGINAL

Total Observacions: 1351

Llicenciats: 314

Obs.Censurades: 1037 % Censura: 76.76

Contrast Global Hipòtesi: Nul·la: BETA=0

Criteri	Sense			Amb			Criteri	Sense			Amb		
	-2 LOG L	Score	Wald	-2 LOG L	Score	Wald		-2 LOG L	Score	Wald	-2 LOG L	Score	Wald
	1814.825	791.809	1023.016 (p=0.0001)	1814.825	792.609	1022.215 (p=0.0001)		1814.825	1309.858	504.967 (p=0.0001)	1814.825	1309.858	504.967 (p=0.0001)
	.	.	607.927 (p=0.0001)	.	.	624.714 (p=0.0001)	Score	.	.	445.422 (p=0.0001)	.	.	445.422 (p=0.0001)
	.	.	263.489 (p=0.0001)	.	.	374.376 (p=0.0001)	Wald	.	.	186.402 (p=0.0001)	.	.	186.402 (p=0.0001)
Variable	Paràmetre	Error	Wald	Paràmetre	Error	Wald	Variable	Paràmetre	Error	Wald	Paràmetre	Error	Wald
	Estimat	Standard	Chi-Square	Estimat	Standard	Chi-Square		Estimat	Standard	Chi-Square	Estimat	Standard	Chi-Square
FONF2	0.070756	0.61382	0.01329 (p=0.9082)	-0.018870	0.61477	0.0009421 (p=0.9755)	FONF2	-0.436126	0.58341	0.55883 (p=0.4547)	-0.436126	0.58341	0.55883 (p=0.4547)
RESPAN	0.121759	0.09695	1.57718 (p=0.2092)	0.015546	0.09646	0.02597 (p=0.8720)	RESPAN	0.585017	0.09605	37.09583 (p=0.0001)	0.585017	0.09605	37.09583 (p=0.0001)
CENT	0.048823	0.19292	0.06405 (p=0.8002)	-0.026736	0.19246	0.01930 (p=0.8895)	CENT	-0.175289	0.18125	0.93529 (p=0.3335)	-0.175289	0.18125	0.93529 (p=0.3335)
DOMIC	0.047960	0.15281	0.09850 (p=0.7536)	0.189825	0.15587	1.48310 (p=0.2233)	DOMIC	-0.442887	0.14271	9.63096 (p=0.0019)	-0.442887	0.14271	9.63096 (p=0.0019)
IPAS	0.204845	0.01305	246.30516 (p=0.0001)	0.200082	0.01104	328.56979 (p=0.0001)	NORM	3.577777	0.32563	122.21397 (p=0.0001)	3.577777	0.32563	122.21397 (p=0.0001)

ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC CONTINU SUBPOBLACIÓ DE LLICENCIATS (Amb el Rendiment Acadèmic - Totes les Variables)

Variable Dependent: ANYS DURADA DELS ESTUDIS
Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
Valor(s) de Censura: 1 (Continua) 2 (Abandona)
Mètode d'estimació: EXACT MÀX. VERSEMBLANÇA MARGINAL

Total Observacions: 1351

Llicenciats: 314

Obs.Censurades: 1037 % Censura: 76.76

Contrast Global Hipòtesi Nulla: BETA=0

Criteri	Sense Covariants		Amb Covariants		Model Chi-Square	Criteri	Sense Covariants		Amb Covariants		Model Chi-Square	
	-2 LOG L	Score	Wald	-2 LOG L			Score	Wald	-2 LOG L	Score		Wald
-2 LOG L	1814.825	785.991	1028.827	1814.825	1037.842	-2 LOG L	1814.825	776.982	1037.842	1037.842	(p=0.0001)	
Score	617.483	617.483	617.483	635.129	635.129	Score	635.129	635.129	635.129	635.129	(p=0.0001)	
Wald	261.846	261.846	261.846	339.722	339.722	Wald	339.722	339.722	339.722	339.722	(p=0.0001)	
Variable	Paràmetre	Error	Wald	Paràmetre	Error	Wald	Paràmetre	Error	Wald	Paràmetre	Error	Wald
Estimat	Estimat	Standard	Estimat	Estimat	Standard	Estimat	Estimat	Standard	Estimat	Estimat	Standard	Estimat
Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square	Chi-Square
NAC	-0.407577	1.06759	0.14575	-1.757575	1.06900	2.70314	NAC	-0.445172	1.01735	0.19148	0.19148	(p=0.6617)
SEXE	0.122822	0.13279	0.85550	0.205834	0.13490	2.32823	SEXE	0.236307	0.12240	3.72703	3.72703	(p=0.0535)
ESCI	-4.267237	3.085	1.91359E-6	-3.460973	3.115	1.23424E-6	ESCI	0.537205	2933	3.35397E-8	3.35397E-8	(p=0.9999)
DOMIC	-0.081971	0.20360	0.16209	-0.154525	0.20289	0.58004	DOMIC	-0.554444	0.20292	7.46583	7.46583	(p=0.0063)
RESID1	-13.686945	1877	0.0000532	-13.613871	1978	0.0000473	RESID1	-14.882604	1947	0.0000584	0.0000584	(p=0.9939)
RESID2	-0.147710	0.22032	0.44949	-0.565748	0.22827	6.14260	RESID2	-0.047283	0.22059	0.04595	0.04595	(p=0.8303)
FONF1	-0.054259	0.25118	0.04666	0.128787	0.24993	0.26552	FONF1	-0.104732	0.23213	0.20357	0.20357	(p=0.6519)
FONF2	0.066367	0.83327	0.00634	-0.118920	0.82434	0.02081	FONF2	-0.323120	0.67556	0.22877	0.22877	(p=0.6324)
FONF3	0.180593	0.50288	0.12896	0.235312	0.49586	0.22520	FONF3	0.209934	0.46354	0.20511	0.20511	(p=0.6506)
LABOR	0.213364	0.26588	0.64399	0.133937	0.26457	0.25627	LABOR	0.052689	0.24769	0.04525	0.04525	(p=0.8315)
CENT	0.026866	0.25621	0.01100	0.003572	0.25656	0.0001939	CENT	-0.068346	0.24137	0.08018	0.08018	(p=0.7771)
HORA	0.050403	0.19289	0.06828	-0.051371	0.19420	0.06997	HORA	-0.143387	0.18239	0.61806	0.61806	(p=0.4318)
CLMA1	0.176921	0.18421	0.92244	0.130473	0.18409	0.50233	CLMA1	-0.044757	0.17008	0.06925	0.06925	(p=0.7924)
CLMA2	-0.065596	0.21586	0.09235	-0.006527	0.21143	0.0009531	CLMA2	0.037266	0.19600	0.03615	0.03615	(p=0.8492)
TITOL	-0.610146	0.52700	1.34043	-0.574271	0.52325	1.20450	TITOL	-0.235900	0.51979	0.20597	0.20597	(p=0.6499)
EDAT	-0.028631	0.10235	0.07825	-0.020793	0.10143	0.04202	EDAT	0.303729	0.08558	0.12892	0.12892	(p=0.7196)
RESPAN	0.148565	0.11202	1.75875	0.081583	0.11254	0.52553	RESPAN	0.630136	0.10769	34.23739	34.23739	(p=0.0001)
IPAS	0.206568	0.01323	243.71460	0.210366	0.01213	300.93447	IPAC	3.616537	0.32488	123.91874	123.91874	(p=0.0001)

Quan tenim en compte la subpoblació dels llicenciats, les mesures del rendiment acadèmic tenen interpretacions diferents i capten diferents aspectes del rendiment acadèmic. Així, la variable IPAS per assignatures vist que pondera per igual totes les assignatures, està valorant de la mateixa manera dos individus que aproven el mateix nombre d'assignatures encara que siguin diferents conjunts de les mateixes. Per tant, als efectes de finalitzar els estudis amb èxit l'indicador IPAS no serà el més adient donat que la superació d'un nombre d'assignatures determinat no garanteix una bona direcció cap a la llicenciatura si aquelles no són assignatures troncales de la carrera.

Per la seva banda, la variable IPAC per crèdits pondera de forma diferent les assignatures segons la seva importància en termes dels crèdits i, per tant, de cara a la finalització dels estudis amb èxit seria un indicador més acurat. Així, dos individus amb el mateix valor per a l'indicador IPAS poden representar situacions objectives de rendiment acadèmic contraposades segons sigui el nombre de crèdits que correspon a cadascuna de les assignatures superades i, en conseqüència, uns valors per a la variable IPAC extremadament oposats.

Aquestes consideracions sobre les variables IPAS i IPAC són aclaridores del diferent nivell de significació individual de les mateixes segons sigui la subpoblació analitzada. En el cas de la d'abandonaments per bé que els estudiants prenen les decisions en funció de la percepció subjectiva del rendiment acadèmic, la qual es realitza sobre el nombre d'assignatures superat, és la variable IPAS la que gaudeix d'una significació més elevada. Per a la subpoblació dels llicenciats, la finalització dels estudis amb èxit no està directament associada a una decisió individual sinó que serà una situació objectiva segons sigui el rendiment acadèmic de l'estudiant per la qual cosa la variable més significativa és l'indicador IPAC.

La variable NORM, per la seva part, com a indicadora de si l'estudiant ha superat la normativa de permanència, està actuant com aproximació al rendiment acadèmic i lògicament resulta molt significativa en raó a que els alumnes que finalitzen els estudis amb èxit superen sobradament el nombre de crèdits que la trajectòria mínima exigeix. Malgrat tot, els resultats de les estimacions puntuals per a la resta de paràmetres són molt diferents dels obtinguts amb els dos altres indicadors del rendiment donat que aquesta variable no incorpora la totalitat de la informació de l'evolució en la progressió acadèmica, és a dir, dos individus que superin la normativa de permanència poden tenir rendiments acadèmics molt diferenciats. Aquest fet també es manifesta pel nivell de la bondat de l'ajust del model el qual es troba molt més allunyat dels valors obtinguts en les modelitzacions alternatives del que reflexava la subpoblació d'abandonaments i la significació individual de la variable NORM és molt més reduïda que anteriorment tot indicant que no és la mesura més oportuna per tal d'explicar o caracteritzar la finalització amb èxit dels estudis.

Una altra constatació feta a partir de l'anàlisi dels resultats i derivada dels comentaris anteriors són els canvis en les variables significatives respecte de la modelització sense el rendiment acadèmic. Si ens concentrem en l'estudi del model que incorpora l'indicador IPAC, per ser el que millor representa el rendiment acadèmic als efectes de la finalització amb èxit dels estudis, observem com ara les variables significatives són el TIPUS DE RESIDÈNCIA (modalitat Altres domicilis) i, en menor mesura, la NACIONALITAT. Així, ja no són rellevants per a caracteritzar l'heterogeneïtat de les durades dels estudis les variables RESULTATS P.A.A.U., DOMICILI, CENTRE i FONT DE FINANÇAMENT (modalitat Treball Propi) perquè la informació que incorporen està recollida en el rendiment acadèmic, com ja s'ha comentat anteriorment.

**ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC CONTINU
SUBPOBLACIÓ DE LLICENCIATS
(Rendiment Acadèmic i Variables Significatives)**

Variable Dependent: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
 Valor(s) de Censura: 1 (Continua) 2 (Abandona)
 Mètode Estimació: EXACT Màx.Versemblança Marginal

Total Observacions: 1351
 Llicenciats: 314
 Obs.Censurades: 1037 % Censura: 76.76

Contrast Global Hipòtesi Nul·la: BETA=0

Criteria	Sense Covariants	Amb Covariants	Chi-Square del Model
-2 LOG L	1814.825	784.436	1030.389 amb 3 DF (p=0.0001)
Score	.	.	606.071 amb 3 DF (p=0.0001)
Wald	.	.	346.078 amb 3 DF (p=0.0001)

Variable	Paràmetre Estimats	Error Standard	Wald Chi-Square	Pr > Chi-Square	Risc Ratio
NAC	-1.712445	1.06134	2.60329	0.1066	0.180
RESID2	-0.485546	0.17649	7.56857	0.0059	0.615
IPAC	0.209498	0.01159	326.70506	0.0001	1.233

L'estimació del model introduint a la vegada la variable de rendiment acadèmic IPAC i l'indicador de la superació de la normativa de permanència conjuntament amb les característiques personals i de gestió acadèmica proporciona uns resultats avinents amb els comentaris realitzats. La variable de superació de

la normativa de permanència no és significativa en presència del rendiment acadèmic a la vista de que la única informació que afegeix no és important en l'explicació de les durades dels estudis que finalitzen amb èxit. Així doncs, els resultats per a la subpoblació de llicenciats, eliminant totes aquelles variables que no són significatives estadísticament, són els de la taula anterior. Per a la interpretació dels paràmetres estimats haurem de tenir present que els efectes dels canvis en les variables explicatives es produeixen directament sobre la funció d'atzar base i inversament sobre la durada dels estudis.

Les variables NACIONALITAT i TIPUS DE RESIDÈNCIA en la modalitat Altres domicilis presenten un paràmetre de signe negatiu de tal forma que aquells estudiants de nacionalitat espanyola o que viuen en residències diferents del domicili patern o propi tenen (mantenint constants les altres variables) un risc inferior de llicenciar-se i, per tant, necessiten més anys per a finalitzar els estudis amb èxit.

Per contra, la variable de rendiment acadèmic IPAC presenta un paràmetre amb signe positiu de manera que la millora en el rendiment acadèmic suposarà un augment en la funció d'atzar el que equivaldrà a un nombre d'anys inferior fins a llicenciar-se. Malgrat ser els valors de la variable IPAC canviant en el temps, l'efecte sobre la durada és independent del moment del temps en que es mesuri per raó de la hipòtesi de proporcionalitat del model. Caldrà doncs contrastar l'esmentada hipòtesi del model tot i que en aquesta subpoblació de llicenciats és més versemblant que es mantingui donat que la possibilitat de llicenciar-se està en funció de la superació de les assignatures incloses en el pla d'estudis corresponent i no pas del nombre d'anys que l'estudiant hagi matriculat.

L'anàlisi de les funcions de supervivència i d'atzar integrat estimades, així com els càlculs de les durades medianes i el contrast de la hipòtesi de proporcionalitat del model, tal com s'ha dit en l'apartat corresponent als abandonaments, es realitzarà d'una forma més senzilla i adient mitjançant el model d'atzar proporcional discret.

11. Resultats de la modelització semiparamètrica discreta

Capítol 11

Resultats de la modelització semiparamètrica discreta

La consideració del temps com a variable discreta facilita la formalització i interpretació del model d'atzar proporcional. La diferència fonamental del model discret respecte del continu està en la interpretació de la funció d'atzar, essent en el primer cas una probabilitat condicionada mentre que en el segon és una taxa instantània de sortida del procés. Així, si l'atzar o risc és una probabilitat condicionada podrem reparametritzar-la per tal que segueixi alguna distribució de probabilitat senzilla i d'aquesta forma estimar els paràmetres dels quals depèn. Existeixen dues especificacions alternatives del model d'atzar proporcional discret. La primera es deriva de l'aplicació directa del model sobre el temps discret de la qual obtenim que la funció d'atzar estarà relacionada amb les variables explicatives de la següent forma:

$$\lambda(x_i; z) = 1 - (1 - \lambda_i)^{\exp(z\beta)}$$

Aquesta especificació implica que la reparametrització loglogarítmica complementaria, $\log[-\log(1 - \lambda_i)]$, resulta ser una combinació lineal de les variables explicatives.

La segona especificació del model d'atzar discret, proposada per Cox (1972) és la de l'aplicació d'un model logístic discret segons el qual la relació entre la funció d'atzar i les variables explicatives seria de la forma:

$$\frac{\lambda(t;z)}{1 - \lambda(t;z)} = \frac{\lambda_0(t)}{1 - \lambda_0(t)} \exp(z\beta)$$

i la seva transformació logarítmica, $\log [\lambda_i / (1 - \lambda_i)]$, també és lineal en els paràmetres.

Aquestes dues especificacions de l'atzar discret permeten l'estimació conjunta dels paràmetres que acompanyen a les variables explicatives, els quals mesuren els efectes sobre l'atzar de variacions en els seus valors, i dels paràmetres que determinen la funció d'atzar base. En el cas de treballar amb valors petits en les probabilitats condicionades λ_i , ambdues especificacions són aproximadament iguals. En cas contrari cal tenir en compte que la transformació logística és només una aproximació, essent la loglogarítmica complementaria la més adequada.

La possibilitat d'estimar el model d'atzar continu pel mètode de la maximització de la versemblança marginal està condicionada en la pràctica per l'estructura de les dades temporals. Així, en el cas de treballar sobre la base d'un temps continu però enregistrat en forma d'interval o, alternativament, quan la formulació del temps és plenament discreta, la presència de moltes observacions repetides per a un mateix moment del temps (o interval) provoca que aquella estimació sigui computacionalment impracticable. Tot i que la nostra aplicació es desenvolupa de fet en un entorn discret ha estat possible l'estimació per aquest mètode, els resultats de la qual s'han presentat en l'apartat anterior. Malgrat tot, s'ha efectuat l'especificació i estimació del model discret utilitzant les dues alternatives comentades anteriorment, tant per la subpoblació d'abandonaments com per la de llicenciats.

Per a l'estimació del model d'atzar proporcional discret utilitzant ambdues especificacions alternatives s'ha utilitzat el PROC LOGISTIC de SAS (versió 6) el qual disposa de les dues opcions reparametritzadores, LOGIT i CLOGLOG, de la probabilitat condicionada que representa l'atzar, que ens proposem estimar. Prèviament, però, cal fer una reestructuració de la base de dades sobre la qual es desenvolupa l'aplicació. Aquesta reestructuració consisteix en la creació d'un registre per a cada individu de la mostra en cadascun dels anys en els quals ha estat matriculat, registre que incorporarà una nova variable dicotòmica indicadora de si l'individu ha abandonat (s'ha llicenciat) o si pel contrari continua. A més la nova estructura de la base de dades també inclourà un conjunt de variables dicotòmiques, una per a cada moment del temps (o interval) al llarg del període analitzat, i que serviran per a indicar si el registre de l'individu pertany a un moment del temps o a un altre. Aquesta forma de reorganitzar la informació permet que les variables els valors de la qual són canviants en el temps puguin estar representades per una única variable que prendrà un valor diferent en cadascun dels registres associat a un interval determinat.

11.1 Sub població d'abandonaments

La taula següent posa de manifest, en primer lloc, que la hipòtesi sobre el temps, continu o discret, en l'aplicació que ens ocupa i donada l'estructura de les dades en termes de nombre d'abandonaments en cada interval respecte el total d'individus subjectes a la possibilitat d'abandonar, no és determinant en els resultats. Les estimacions del model d'atzar proporcional continu i les del discret amb la reparametrització loglogarítmica complementaria són iguals, existint només

ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC SUBPOBLACIÓ D'ABANDONAMENTS (Especificacions Contínua i Discretes)

Variable Dependent: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
 Valor(s) de Censura: 1 (Continus) 3 (Llicenciat)
 Mètode d'estimació: Màxima Versemblança

Total Observacions: 1351
 Abandonaments: 645
 Obs.Censurades: 706 % Censura: 52.26

Contrast Global Hipòtesis Nul·la: BETA=0

MODEL CONTINU

Criteria	Sense Covariants	Amb Covariants	Model Chi-Square
-2 LOG L	4051.732	2973.208	1078.524 (p=0.0001)
Score			1022.320 (p=0.0001)
Wald			702.559 (p=0.0001)

MODEL DISCRET CLOGLOG

Criteria	Sense Covariants	Amb Covariants	Model Chi-Square
-2 LOG L	12761.691	3014.310	9747.381 (p=0.0001)
Score			8701.473 (p=0.0001)
AIC	12761.691	3046.310	
SC	12761.691	3155.339	

MODEL DISCRET LOGIT

Criteria	Sense Covariants	Amb Covariants	Model Chi-Square
-2 LOG L	9329.761	3021.799	6307.962 (p=0.0001)
Score			4749.052 (p=0.0001)
AIC	9329.761	3053.799	
SC	9329.761	3162.829	

Variable

Variable	Paràmetre Estimant	Error Standard	Wald Chi-Square
NAC	-1.235310	0.31961	14.93898 (p=0.0001)
SEXE	-0.152284	0.08588	3.14403 (p=0.0762)
DOMIC	-0.695229	0.16095	18.65866 (p=0.0001)
RESID2	-0.415181	0.17742	5.47612 (p=0.0193)
FONF2	0.351550	0.14558	5.83172 (p=0.0157)
FONF3	0.664031	0.24889	7.11800 (p=0.0076)
EDAT	0.037810	0.01653	5.23302 (p=0.0222)
NORM	-1.316874	0.13682	92.63449 (p=0.0001)
IPAC	-0.070693	0.00408	300.28488 (p=0.0001)

Variable

Variable	Paràmetre Estimant	Error Standard	Wald Chi-Square
D01	-2.2217	0.0876	643.2582 (p=0.0001)
D02	-2.8318	0.1131	626.7689 (p=0.0001)
D03	-3.0932	0.1284	579.9267 (p=0.0001)
D04	-3.3280	0.1500	492.1774 (p=0.0001)
D05	-3.2513	0.1513	461.7913 (p=0.0001)
D06	-3.2428	0.1661	381.0410 (p=0.0001)
D07	-1.7191	0.1128	232.2584 (p=0.0001)
NAC	-1.2361	0.3347	13.6356 (p=0.0002)
SEXE	-0.1523	0.0855	3.1721 (p=0.0749)
DOMIC	-0.6962	0.1577	19.4997 (p=0.0001)
RESID2	-0.4159	0.1764	5.5397 (p=0.0184)
FONF2	0.3517	0.1471	5.7152 (p=0.0168)
FONF3	0.6650	0.2419	7.5550 (p=0.0060)
EDAT	0.0378	0.0170	4.9320 (p=0.0264)
NORM	-1.3173	0.1377	91.4948 (p=0.0001)
IPAC	-0.0708	0.00422	281.6497 (p=0.0001)

Variable

Variable	Paràmetre Estimant	Error Standard	Wald Chi-Square
D01	-2.1554	0.0948	516.9671 (p=0.0001)
D02	-2.8081	0.1257	498.9741 (p=0.0001)
D03	-3.0965	0.1439	463.2051 (p=0.0001)
D04	-3.3634	0.1675	403.3918 (p=0.0001)
D05	-3.2556	0.1718	359.1983 (p=0.0001)
D06	-3.2669	0.1913	291.5814 (p=0.0001)
D07	-1.2723	0.1385	84.3313 (p=0.0001)
NAC	-1.6078	0.4198	14.6676 (p=0.0001)
SEXE	-0.1950	0.1012	3.7154 (p=0.0539)
DOMIC	-0.7250	0.1916	14.3182 (p=0.0002)
RESID2	-0.3882	0.2128	3.3280 (p=0.0681)
FONF2	0.4402	0.1810	5.9159 (p=0.0150)
FONF3	0.6797	0.3025	5.0501 (p=0.0246)
EDAT	0.0571	0.0235	5.9060 (p=0.0151)
NORM	-1.2541	0.1528	67.3402 (p=0.0001)
IPAC	-0.0832	0.00516	259.8235 (p=0.0001)

petites diferències en els errors estàndards que, per altra banda, no són en absolut sistemàtiques com per a concloure que unes estimacions són més eficients en un o altre cas.

Essent els resultats d'aquestes dues especificacions equivalents, cal assenyalar els avantatges de la modelització discreta sobre la contínua. Per començar, la modelització discreta incorpora les variables dicotòmiques D01 fins a la D07 associades als diferents moments del temps analitzats, essent els paràmetres que les acompanyen els atzars base que serveixen de referència per a mesurar les variacions en les explicatives. Així doncs, la modelització discreta ens proporciona directament estimacions de la funció d'atzar base discreta a partir de la qual podem construir la de supervivència base i mesurar les durades medianes per a diferents subconjunts d'individus en funció de les seves característiques.

D'altra banda, la reorganització de la base de dades sobre la que es realitza l'estimació dels models discrets obre la possibilitat d'estimar funcions de supervivència per a grups d'individus tenint en compte la presència de variables explicatives canviants en el temps, mentre que en la modelització contínua no podem anar més enllà de l'anàlisi dels resultats de les estimacions dels paràmetres i de l'ajust global del model sense la possibilitat, per exemple, de realitzar una anàlisi de les durades segons l'evolució en el rendiment acadèmic d'un estudiant.

A més, una altra de les tasques pendents de resoldre en la modelització contínua era la de verificar la hipòtesi de proporcionalitat del model, és a dir, que els efectes de variacions en les característiques individuals sobre l'atzar siguin independents del moment del temps en que es mesurin. Mitjançant la modelització discreta, la qual incorpora les dicotòmiques associades a cada moment del temps

analitzat, podem introduir interaccions entre aquestes dicotòmiques i qualsevol variable explicativa de tal forma que si la interacció és significativa no es verificarà la hipòtesi de proporcionalitat vist que l'efecte serà diferent en un moment o altre del temps.

Un cop feta la comparació entre la modelització contínua i la discreta reparametritzada amb la funció loglogarítmica complementària, atès que l'especificació discreta alternativa, reparametritzada mitjançant la funció logística, tindrà també els avantatges d'interpretació que acabem de detallar, és obligat comparar el resultats d'ambdues modelitzacions discretes.

En aquest sentit cal remarcar que la formulació logística és només una aproximació al model d'atzar proporcional, la qual tendirà als valors de la loglogarítmica complementària a mesura que els intervals en que es divideix el temps esdevinguin més petits i, per tant, les probabilitats condicionades d'abandonament, o atzars λ_i , per a cada moment del temps seran més petites per existir un nombre reduït d'individus que abandonen. Tot i això, la seva estimació és convenient per bé que quan es produeix la situació d'atzars petits deguda a l'existència de pocs abandonaments en cada interval de temps i, per tant, quan l'aproximació és més acurada, ens trobem davant d'instabilitats en l'algoritme de Newton-Raphson per a la maximització de la funció de versemblança i la convergència del mateix no estarà garantida. Així, malgrat que aquesta situació no és produïda en la nostra aplicació concreta, ha estat estimat el model discret amb la reparametrització logística.

Els resultats de la modelització logística confirmen que són una aproximació vist que les estimacions puntuals es troben allunyades de les

obtingudes amb l'especificació loglogarítmica complementària, especialment en els casos de les variables NACIONALITAT, SEXE, FONT DE FINANÇAMENT DELS ESTUDIS (modalitat Treball Propi) i EDAT per a les quals les diferències relatives entre uns i altres paràmetres estimats superen el 25% del valor de l'estimació correcta.

A la vista d'aquest estudi comparatiu, les posteriors anàlisis per a la interpretació dels resultats sobre individus representatius i la verificació de la hipòtesi de proporcionalitat s'efectuaran sobre la modelització discreta loglogarítmica complementària. Primerament, caracteritzarem dos individus representatius a partir de les característiques personals i de gestió acadèmica sobre els quals farem diverses hipòtesis en el seu rendiment acadèmic i, per tant, en la seva situació respecte de la normativa de permanència. Per aquests dos individus en les diverses situacions possibles estimarem les funcions de supervivència i calcularem les durades medianes. Els individus són:

Individu A: Home (SEXE=1), de 25 anys d'EDAT, amb nacionalitat espanyola (NAC=1), domicili a Barcelona ciutat (DOMIC=1) i que resideix al domicili propi (RESID2=0), el qual finança els estudis amb el seu propi treball (FONF2=1 i FONF3=0).

Individu B: Dona (SEXE=0), de 18 anys d'EDAT, amb nacionalitat espanyola (NAC=1), domicili a comarques allunyades de Barcelona (DOMIC=0) que resideix en un pis d'estudiants (RESID2=1) i que finança els estudis mitjançant l'ajut del seus pares (FONF2=0 i FONF3=0).

Per a cadascun d'aquests individus, considerarem les següents trajectòries acadèmiques en funció del seu rendiment acadèmic mesurat en crèdits:

Trajectòria 1:

Curs	1	2	3	4	5	6	7	8
C.Matriculats	50	50	50	50	50	50	50	50
C.Matr.Acum.	0	50	100	150	200	250	300	350
C.Superats	15	15	15	24	15	18	24	24
C.Sup.Acum.	15	30	45	69	84	102	126	150
Normativa	1	0	0	0	0	0	0	0

Trajectòria 2:

Curs	1	2	3	4	5	6	7	8
C.Matriculats	50	50	50	50	50	50	50	50
C.Matr.Acum.	0	50	100	150	200	250	300	350
C.Superats	21	21	21	21	21	21	21	21
C.Sup.Acum.	21	42	63	84	105	126	147	168
Normativa	1	1	0	0	0	0	0	0

Trajectòria 3:

Curs	1	2	3	4	5	6	7	8
C.Matriculats	50	50	50	50	50	50	50	50
C.Matr.Acum.	0	50	100	150	200	250	300	350
C.Superats	22.5	22.5	22.5	22.5	22.5	25.5	24	24
C.Sup.Acum.	22.5	45	67.5	90	112.5	138	162	186
Normativa	1	1	1	0	0	0	0	0

Trajectòria 4:

Curs	1	2	3	4	5	6	7	8
C.Matriculats	50	50	50	50	50	50	50	50
C.Matr.Acum.	0	50	100	150	200	250	300	350
C.Superats	27	27	21	22.5	22.5	22.5	19.5	18
C.Sup.Acum.	27	54	75	97.5	120	142.5	162	180
Normativa	1	1	1	1	0	0	0	0

Trajectòria 5:

Curs	1	2	3	4	5	6	7	8
C.Matriculats	50	50	50	50	50	50	50	50
C.Matr.Acum.	0	50	100	150	200	250	300	350
C.Superats	30	30	30	30	24	24	24	24
C.Sup.Acum.	30	60	90	120	144	168	192	216
Normativa	1	1	1	1	1	1	1	1

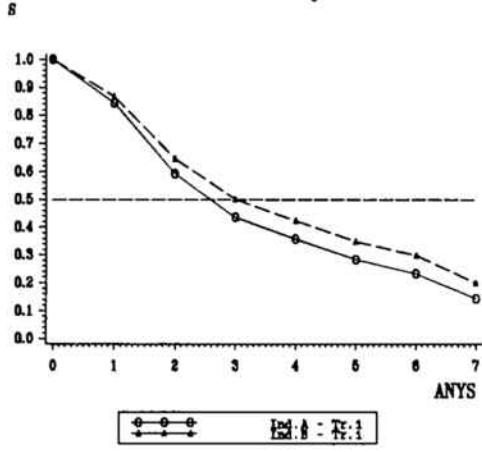
Amb les anteriors caracteritzacions d'individus i de trajectòries i els resultats de l'estimació del model discret CLOGLOG, realitzem l'anàlisi comparativa de les funcions de supervivència estimades i de les corresponents durades medianes. Recordarem aquí que l'indicador del rendiment acadèmic IPAC es computa de manera acumulada i en percentatge respecte el total de crèdits necessaris per a finalitzar els estudis mitjançant l'obtenció del títol. A més, l'estimació del model d'atzar proporcional discret s'ha realitzat sobre els valors de les explicatives expressats en desviacions respecte la seva mitjana per tal que les funcions d'atzar base serveixin de referència a variacions per sobre o per sota de la mitjana.

La representació gràfica de les funcions de supervivència estimades per als dos individus en cadascuna de les trajectòries alternatives en el rendiment acadèmic mostren com les diferències en les durades dels estudis fins a l'abandonament tenen com a principal factor explicatiu el rendiment acadèmic. Així, per a qualsevol de les trajectòries proposades les funcions de supervivència gaire bé són iguals, mentre que si les analitzem comparant una trajectòria amb una altra observem com a mesura que el rendiment acadèmic és més elevat també més elevada serà la durada dels estudis fins a l'abandonament. Aquest resultat és de molta importància i implica que, per a un mateix rendiment acadèmic, els efectes sobre la durada comparant els individus A i B són menys importants que els que produeixen diferències en el rendiment per a un mateix individu.

Cal tenir present, però, la forta interdependència entre aquelles característiques personals i el rendiment acadèmic, com s'ha posat de manifest en els primer resultats de la modelització de la durada dels estudis quan es comparava la significació estadística de les variables explicatives personals amb la introducció en el model del rendiment acadèmic i sense aquest. Aquest fet permet constatar com la recentment aprovada normativa de permanència ha estat dissenyada correctament a la vista de que la mateixa determina les trajectòries mínimes en funció del nombre de crèdits que l'alumne matricula cada any (amb un mínim de 50) i, per tant, deixa en mans de l'estudiant la fixació de l'itinerari o ritme que vol seguir en la realització dels estudis, el qual haurà de ser conscient de la seva situació personal i actuar en conseqüència: si les seves característiques personals l'obliguen a dur un ritme lent en la progressió acadèmica caldrà que matriculi un nombre de crèdits suficientment baix com per a garantir un rendiment acadèmic relatiu que sigui acceptable, la qual cosa repercutirà positivament sobre la superació de la normativa de permanència.

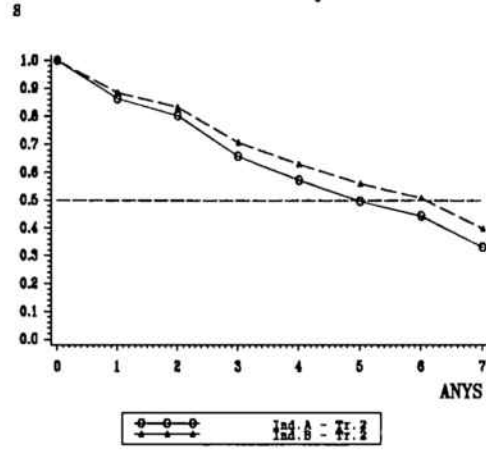
SUPERVIVENCIA ESTIMADA

Individus A i B - Trajectoria 1



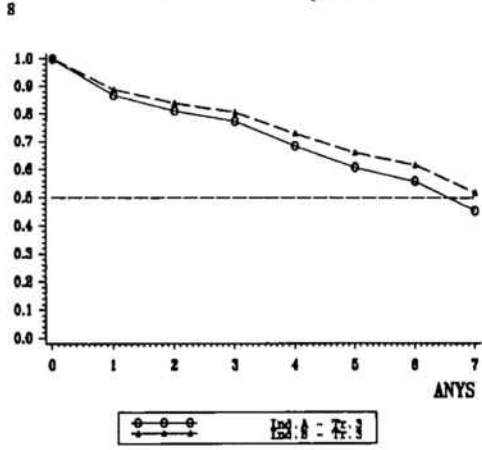
SUPERVIVENCIA ESTIMADA

Individus A i B - Trajectoria 2



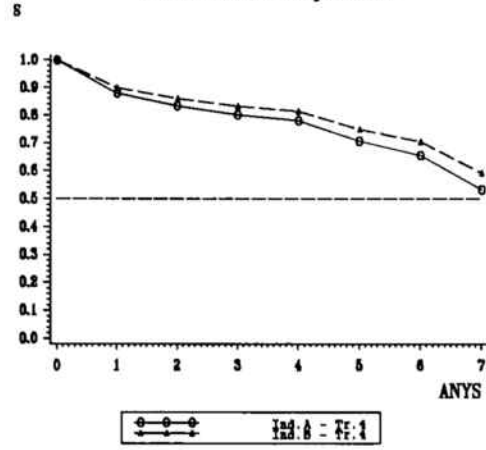
SUPERVIVENCIA ESTIMADA

Individus A i B - Trajectoria 3



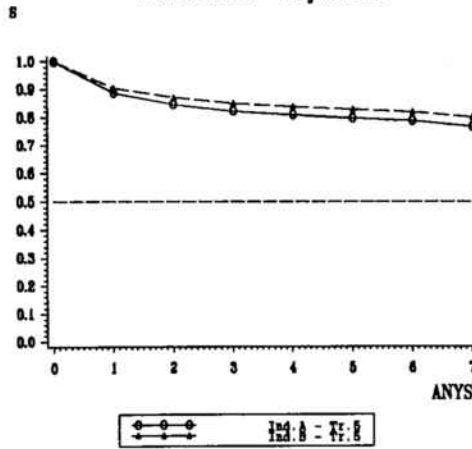
SUPERVIVENCIA ESTIMADA

Individus A i B - Trajectoria 4



SUPERVIVENCIA ESTIMADA

Individus A i B - Trajectoria 5



Més il·lustratiu encara són els resultats si els anàlitzem des de la perspectiva de les durades medianes estimades a partir de les funcions de supervivència per interpolació lineal.

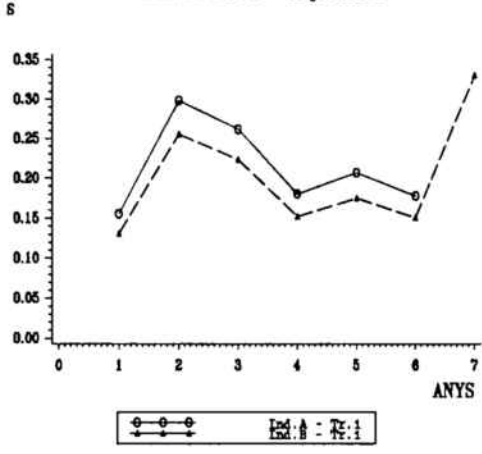
	Individu A	Individu B
Trajectòria 1	2.59 anys	3.03 anys
Trajectòria 2	4.96 anys	6.09 anys
Trajectòria 3	6.57 anys	≈ 7 anys
Trajectòria 4	> 7 anys	> 7 anys
Trajectòria 5	>> 7 anys	>> 7 anys

En aquesta taula veiem com l'individu A, l'home de 25 anys que viu a Barcelona en domicili propi i que es finança els estudis amb el seu propi treball, té una durada mediana més curta que l'individu B, la dona de 18 anys que viu lluny de Barcelona en un pis d'estudiants i que finança els estudis amb l'ajut dels pares, tot i que la diferència entre una i altra durada està al voltant d'un any més per l'individu B. Si la inspecció de la taula la fem en vertical veurem com la reducció en les durades medianes estimades és molt més important quan passem de rendiments acadèmics acceptables (Traj.5) cap a uns més insuficients (Traj.1).

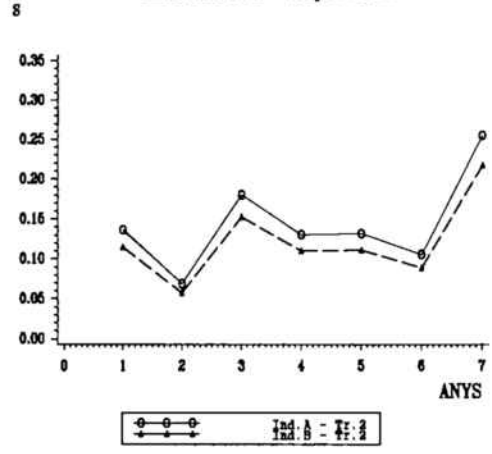
Pel que fa referència a aquestes trajectòries representatives caldria realitzar el següent comentari per tal de situar perfectament els resultats anteriors. Aquestes trajectòries van des de la número 5, la qual és la més acceptable de totes, tot i que representa una progressió acadèmica segons la qual l'estudiant aconseguiria l'objectiu de finalitzar els estudis amb èxit després d'una estada d'aproximadament 11 cursos acadèmics, fins a la més insuficient, la trajectòria 1. Aquesta darrera suposa una progressió acadèmica segons la qual després de 8 anys a la universitat només s'hauria superat un 50 per cent dels crèdits necessaris per a obtenir el títol.

Aquestes trajectòries voldrien representar un ventall de possibilitats de progressió acadèmica, des d'un nivell baix fins a un més acceptable, les quals tinguin una traducció en termes de la normativa de permanència que permeti estimar la durada mediana dels estudiants que no superarien la normativa en cadascun dels cursos acadèmics. D'aquesta manera, la trajectòria 1 es tradueix en que l'alumne seria exclòs un cop finalitzat el segon curs acadèmic, en la trajectòria 2 després del tercer any, en la 3 després del quart any i en la 4 un cop acabat el cinquè curs acadèmic, mentre que el que seguís la trajectòria 5 superaria la normativa al llarg dels set anys considerats. La representació gràfica de les funcions d'atzar estimades per a cadascuna de les trajectòries ens mostren clarament la repercusió de la normativa de permanència sobre la probabilitat d'abandonar els estudis en un moment determinat del temps, condicionada a haver-los seguit fins aleshores. Així, el darrer gràfic dels que segueixen, corresponent a la Trajectòria 5 segons la qual l'estudiant mai no seria exclòs de la universitat, representa la situació límit més favorable pel que fa a la permanència. Observem com l'atzar es va reduint any darrera any, excepte en el setè curs acadèmic on experimenta una petita alça, el que significa que essent el rendiment acadèmic acceptable la probabilitat d'abandonar els estudis condicionada a haver-los seguit fins aquell moment es fa cada cop més petita, l'abandonament és cada cop més improbable i la durada dels estudis més llarga. Si ens fixem ara en la gràfica corresponent a la trajectòria 4 veurem com la funció d'atzar experimenta un desplaçament a l'alça en el cinquè curs acadèmic, precisament l'any en el qual seria exclòs l'alumne amb aquest nivell en el rendiment acadèmic. Aquest desplaçament a l'alça es va produint un any abans a mesura que passem de la trajectòria 4 cap a la més insatisfactòria, la 1. A més s'observa com les diferències entre el individu A i B es van fent més petites a mesura que el rendiment acadèmic és més elevat.

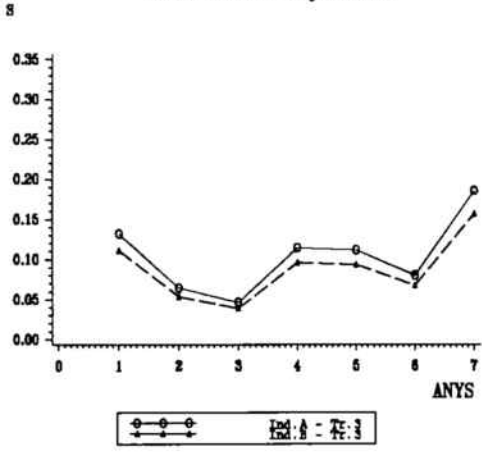
ATZAR DISCRET ESTIMAT
Individus A i B - Trajectoria 1



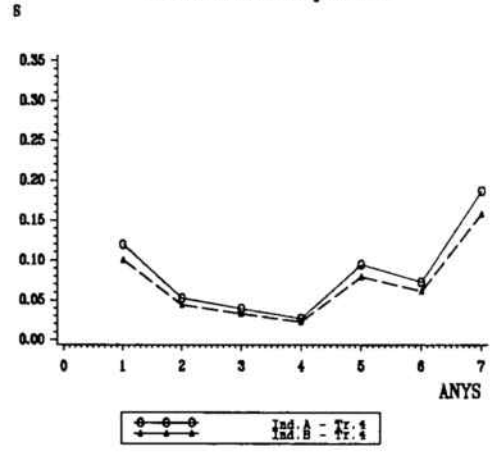
ATZAR DISCRET ESTIMAT
Individus A i B - Trajectoria 2



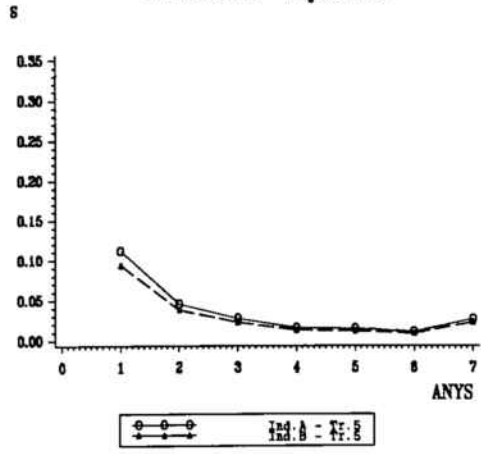
ATZAR DISCRET ESTIMAT
Individus A i B - Trajectoria 3



ATZAR DISCRET ESTIMAT
Individus A i B - Trajectoria 4



ATZAR DISCRET ESTIMAT
Individus A i B - Trajectoria 5



En conseqüència, a més de la interpretació sobre el rendiment acadèmic mesurat per l'indicador IPAC en termes dels crèdits, la introducció de la situació en que es trobaria l'alumne davant de la hipotètica posada en marxa de la normativa de permanència mostra com la probabilitat d'abandonar els estudis és més elevada per aquells que no superarien la normativa. En particular, centrant l'anàlisi en la trajectòria 1, la durada mediana estimada és de 2.89 anys per l'individu A, la qual cosa s'ha d'interpretar en el sentit que el 50% dels alumnes en aquestes circumstàncies haurien abandonat els estudis abans dels tres primers cursos acadèmics i, de la funció de supervivència estimada podem concloure que un cop transcorreguts 7 anys haurien abandonat els estudis un 80% dels individus A amb una trajectòria tipus 1. La normativa de permanència, doncs, hauria avançat una decisió de l'estudiant que es produirà amb tota seguretat, amb el corresponent estalvi de temps i diners tant per l'estudiant com per a la universitat. Queda, doncs, garantit l'esperit de la normativa de permanència en la seva pretensió d'avançar una decisió que s'ha de produir fatalment.

Aquests objectius de la normativa de permanència estan més dubtosos en el cas de la trajectòria 4. Per aquesta, la durada mediana estimada es situa per sobre dels 7 anys, la qual cosa significa que un cop transcorregut aquest període menys d'un 50% dels alumnes en aquesta situació haurien abandonat els estudis mentre que la normativa els hauria exclòs un cop finalitzat el cinquè curs acadèmic. En aquest sentit, cal fer el següent raonament. La present aplicació s'ha realitzat sobre una cohort d'estudiants integrats en un pla d'estudis estructurat en assignatures i amb una normativa de permanència que es fonamentava en un nombre màxim de convocatòries d'examen per assignatura, per la qual cosa les seves decisions respecte les assignatures a matricular i l'esforç necessari per a superar-les no estava condicionat per la normativa de permanència. Així, en el cas

de que l'actual normativa de permanència hagués estat vigent, amb tota seguretat els estudiants haurien pres decisions diferents en el procés de matrícula les quals comportarien la superació de la normativa. És a dir, les estimacions dels paràmetres que mesuren els efectes sobre la durada dels estudis de variacions en el rendiment acadèmic via els indicadors IPAC i NORM, serien diferents en el cas de que la cohort sobre la qual s'ha treballat hagués tingut la normativa vigent, essent d'esperar que aquesta produeixi uns efectes positius sobre el rendiment acadèmic dels estudiants, bàsicament com a conseqüència de la racionalització de les decisions de l'alumne.

Tota l'anàlisi de resultats realitzada fins el moment pot ser encara més clarificadora pel que fa als efectes del rendiment acadèmic sobre la durada dels estudis fins a l'abandonament un cop verificada la hipòtesi de que els esmentats efectes no depenen del moment del temps en que es mesuren, és a dir, la hipòtesi de proporcionalitat del model d'atzar. L'esmentada hipòtesi pot ser contrastada mitjançant la incorporació al model de les corresponents interaccions entre el temps i les variables sobre les quals existeixin raons per pensar que no verifiquen l'esmentat comportament.

Totes les característiques personals i de gestió acadèmica seleccionades, malgrat ser la majoria d'elles estrictament canviants en el temps, han estat considerades invariants prenent els valors del primer any i mantenint-los constants per a tot el període. Aquest fet, conjuntament amb l'anàlisi de cadascuna d'elles, implica que els seus efectes sobre la durada dels estudis no és d'esperar que canviï en moment diferents del temps, la qual cosa s'ha comprovat mitjançant la introducció de la interacció de cadascuna de les variables amb el temps, no presentant significació estadística els paràmetres corresponents.

Per la seva part, el rendiment acadèmic mesurat, per exemple, per l'indicador IPAC, tindrà un efecte diferent sobre les decisions d'abandonament o continuació dels estudis segons el moment del temps en que el mesurem. Així, un rendiment acadèmic molt insuficient comportarà decisions més enèrgiques respecte l'abandonament en el primer curs acadèmic que no pas transcorreguts cinc anys i un cop acumulat un nombre de crèdits important. A continuació presentem el resultat de l'estimació del model discret incorporant les interaccions entre l'indicador IPAC i el temps, les quals vindran representades per les variables IPAD01 fins a la IPAD07, mesurant cadascuna d'elles l'efecte del rendiment acadèmic acumulat sobre la durada dels estudis en cadascun dels set períodes considerats.

Aquests resultats ens mostren com, mitjançant la introducció de les interaccions entre IPAC i temps, es produeixen canvis en la significació estadística de les variables seleccionades si els comparem amb les del model anterior. En particular, la variable EDAT, la qual era significativa quan l'efecte del rendiment acadèmic sobre la durada era el mateix per a qualsevol moment del temps, passa a ser irrellevant en la nova modelització. L'explicació a aquesta circumstància es troba en que la variable EDAT està molt associada al rendiment acadèmic de tal forma que aquells individus que van iniciar els estudis amb edats superiors tenen un rendiment acadèmic més baix, especialment si ens referim als individus d'edats superiors als 25 anys. Quan l'efecte sobre la durada del rendiment acadèmic era invariant respecte el temps una part de l'heterogeneïtat en les durades que no arribava a explicar la variable IPAC era deguda a l'EDAT en iniciar els estudis i la variable resultava significativa. Un cop ampliat el model de manera que els efectes del rendiment acadèmic sobre les durades no són constants, aquella heterogeneïtat que recollia l'EDAT queda incorporada en la variable IPAC.

ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC SUBPOBLACIÓ D'ABANDONAMENTS (Especificació Discreta amb Interaccions)

Variable Dependent: ANYS DURADA DELS ESTUDIS
Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
Valor(s) de Censura: 1 (Continua) 3 (Llicenciat)
Mètode Estimació: Màxima Versemblança (CLOGLOG)

Total Observacions: 1351
Abandonaments: 645
Obs.Censurades: 706 % Censura: 52.26

Contrast Global Hipòtesi Nulla: BETA=0

Criteri	Sense		Amb		Chi-Square del Model	Criteri	Sense		Amb		Chi-Square del Model
	Covariants	Covariants	Covariants	Covariants			Covariants	Covariants			
-2 LOG L	12761.691	2718.172	10043.519		(p=0.0001)	-2 LOG L	12761.691	2717.446	10044.245		(p=0.0001)
Score	.	.	8865.310		(p=0.0001)	Score	.	.	8868.408		(p=0.0001)
AIC	12761.691	2760.172	.			AIC	12761.691	2761.446	.		
SC	12761.691	2903.273	.			SC	12761.691	2911.361	.		

Variable	Paràmetre		Error		Wald		Chi-Square	Variable	Paràmetre		Error		Wald		Chi-Square
	Estimat	Standard	Standard	Standard	Chi-Square	(p=)			Estimat	Standard	Standard	Standard	Chi-Square	(p=)	
D01	-3.9807	0.2319	294.7679	(p=0.0001)	D01	-3.9769	0.2318	294.3400	(p=0.0001)						
D02	-3.2800	0.1966	278.4679	(p=0.0001)	D02	-3.2780	0.1965	278.3629	(p=0.0001)						
D03	-2.8693	0.1650	302.3381	(p=0.0001)	D03	-2.8680	0.1650	302.1202	(p=0.0001)						
D04	-3.2032	0.2201	211.9000	(p=0.0001)	D04	-3.2019	0.2200	211.8797	(p=0.0001)						
D05	-2.6089	0.1553	282.1293	(p=0.0001)	D05	-2.6078	0.1553	281.8080	(p=0.0001)						
D06	-2.8485	0.2247	160.7351	(p=0.0001)	D06	-2.8476	0.2246	160.7305	(p=0.0001)						
D07	-1.0956	0.1016	116.1762	(p=0.0001)	D07	-1.0956	0.1017	116.1465	(p=0.0001)						
NAC	-1.1690	0.3979	8.6318	(p=0.0033)	NAC	-1.1682	0.3966	8.6737	(p=0.0032)						
SEXE	-0.1654	0.0871	3.6040	(p=0.0576)	SEXE	-0.1704	0.0873	3.8053	(p=0.0511)						
DOMIC	-0.7854	0.1680	21.8556	(p=0.0001)	DOMIC	-0.7942	0.1677	22.4406	(p=0.0001)						
RESID2	-0.4860	0.1844	6.9428	(p=0.0084)	RESID2	-0.5010	0.1847	7.3578	(p=0.0067)						
FONF2	0.4306	0.1306	10.8709	(p=0.0010)	FONF2	0.3658	0.1515	5.8325	(p=0.0157)						
FONF3	0.8688	0.2604	11.1318	(p=0.0008)	FONF3	0.8429	0.2609	10.4410	(p=0.0012)						
					EDAT	0.0158	0.0177	0.7977	(p=0.3718)						
NORM	0.6168	0.1728	12.7358	(p=0.0004)	NORM	0.6120	0.1730	12.5125	(p=0.0004)						
IPAD01	-0.4604	0.0299	236.4102	(p=0.0001)	IPAD01	-0.4588	0.0300	233.8517	(p=0.0001)						
IPAD02	-0.2108	0.0179	138.3149	(p=0.0001)	IPAD02	-0.2101	0.0179	136.9905	(p=0.0001)						
IPAD03	-0.1140	0.0114	100.1101	(p=0.0001)	IPAD03	-0.1137	0.0114	99.5074	(p=0.0001)						
IPAD04	-0.1095	0.0112	95.3156	(p=0.0001)	IPAD04	-0.1092	0.0112	94.8818	(p=0.0001)						
IPAD05	-0.0668	0.00739	81.6942	(p=0.0001)	IPAD05	-0.0666	0.00739	81.2261	(p=0.0001)						
IPAD06	-0.0830	0.00906	84.0420	(p=0.0001)	IPAD06	-0.0829	0.00906	83.7087	(p=0.0001)						
IPAD07	-0.0601	0.00536	125.7940	(p=0.0001)	IPAD07	-0.0599	0.00536	124.8453	(p=0.0001)						

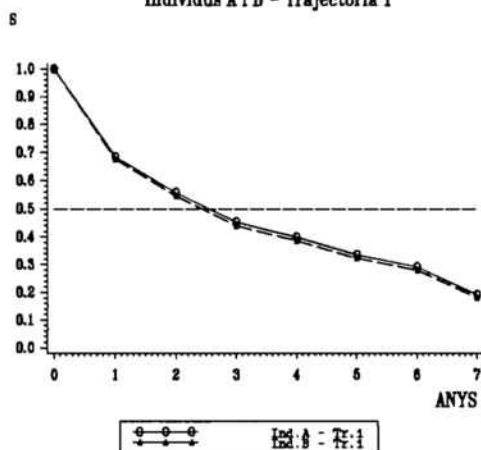
Amb aquests nous resultats tornem a estimar les funcions de supervivència pels individus A i B caracteritzats anteriorment segons les trajectòries 1 a 5. Així mateix, calcularem les durades medianes estimades i les funcions d'atzar. Les durades medianes estimades per a cadascun dels dos individus i segon les diferents trajectòries són les següents:

	Individu A	Individu B
Trajectòria 1	2.56 anys	2.44 anys
Trajectòria 2	6.08 anys	5.96 anys
Trajectòria 3	6.39 anys	6.27 anys
Trajectòria 4	≈ 7 anys	≈ 7 anys
Trajectòria 5	> 7 anys	> 7 anys

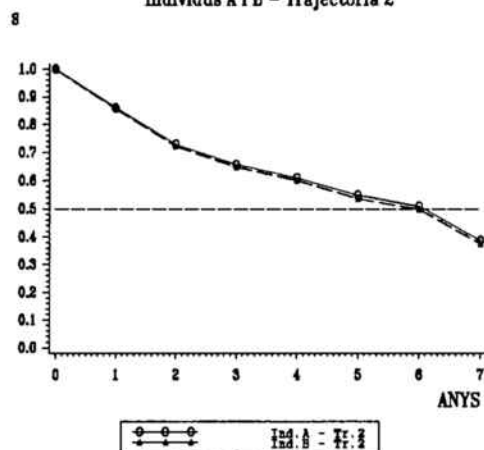
Com podem observar en el quadre, amb la nova modelització incorporant les interaccions entre el rendiment acadèmic i el temps, les durades medianes dels individus A i B són les mateixes tal com ho són també les funcions de supervivència estimades. Comparant aquest resultat amb el de l'anterior modelització ens adonem com el fet de no considerar la possibilitat de que els efectes del rendiment acadèmic sobre la durada dels estudis puguin ser diferents segons el moment del temps en que es mesurin portaria a una estimació incorrecta de la durada mediana de manera que inferiem una durada més curta per l'individu A que pel B mentre que els nous resultats indiquen que les dues durades medianes són aproximadament iguals pels dos individus i en qualsevol trajectòria, essent fins i tot lleugerament més curta l'estada mediana de l'individu B.

De la mateixa forma, s'observa com les durades medianes estimades per a un individu segons sigui la trajectòria acadèmica també són en general més curtes que les obtingudes en l'estimació anterior, particularment per l'individu B.

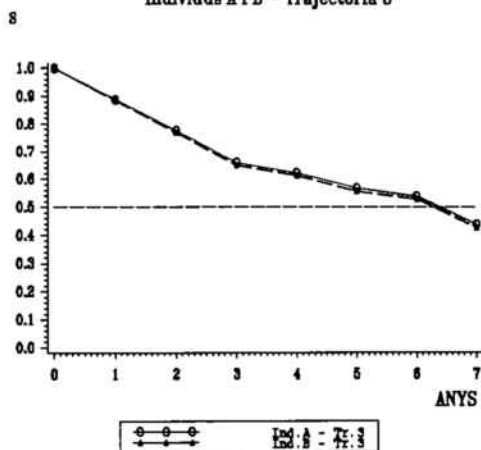
SUPERVIVENCIA ESTIMADA
Individus A i B - Trajectoria 1



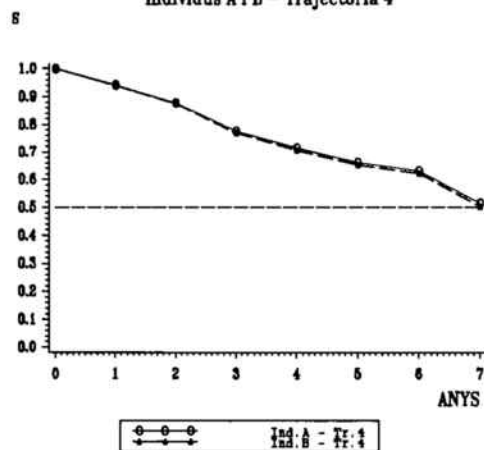
SUPERVIVENCIA ESTIMADA
Individus A i B - Trajectoria 2



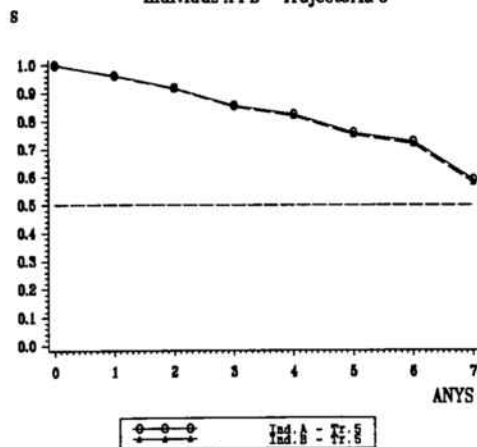
SUPERVIVENCIA ESTIMADA
Individus A i B - Trajectoria 3



SUPERVIVENCIA ESTIMADA
Individus A i B - Trajectoria 4



SUPERVIVENCIA ESTIMADA
Individus A i B - Trajectoria 5

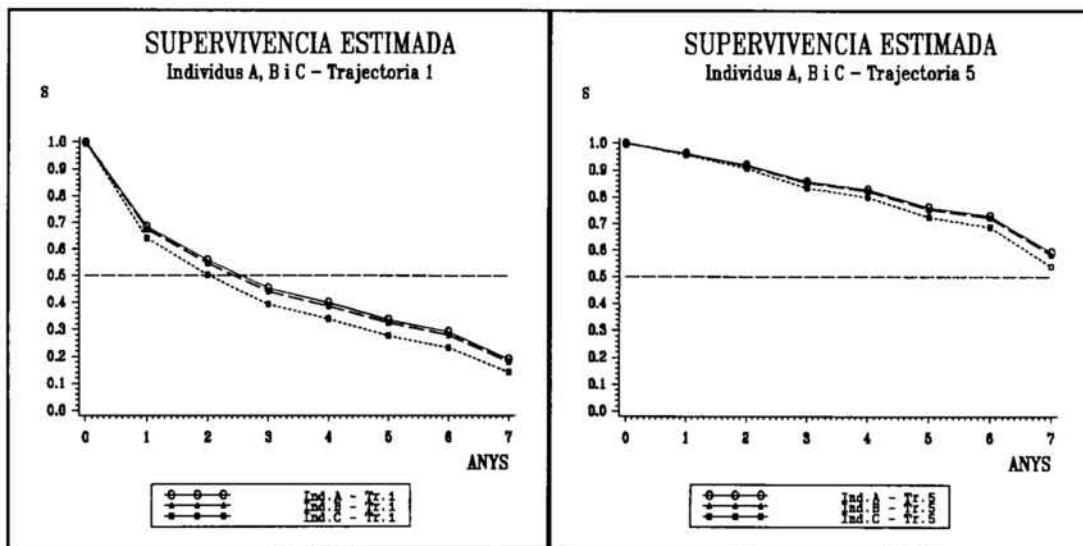


La circumstància que s'ha produït anteriorment segons la qual els dos individus A i B tenen funcions de supervivència aproximadament iguals no ens ha de fer pensar en que les característiques personals i de gestió acadèmica introduïdes en el model siguin innecessàries, la qual cosa només seria correcte en el cas de que haguéssin resultat no significatives estadísticament. De fet, s'ha generat una combinació de característiques individuals que ha portat a una semblança total entre ambdues funcions de supervivència estimades i les seves conseqüències. Pel contrari si incorporem un individu més en l'anàlisi (Individu C) amb les mateixes característiques que l'individu A, exceptuant el SEXE, podem analitzar les diferències en les durades existents entre els homes i les dones, mantenint constants la resta de variables del model. Així ens trobem amb les següents durades medianes estimades:

	Individu A Home	Individu C Dona
Trajectòria 1	2.56 anys	2.02 anys
Trajectòria 2	6.08 anys	4.89 anys
Trajectòria 3	6.39 anys	5.42 anys
Trajectòria 4	= 7 anys	6.65 anys
Trajectòria 5	> 7 anys	> 7 anys

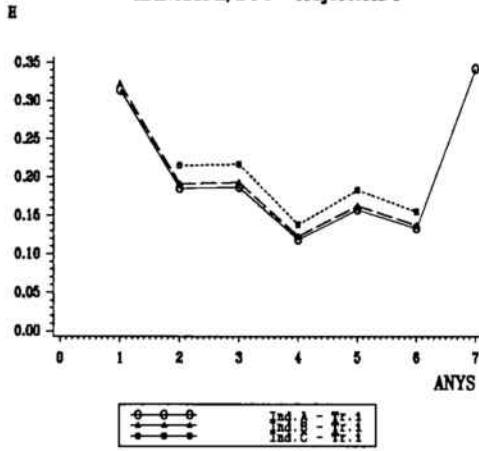
Aquests resultats indiquen que les durades medianes en el col·lectiu de les dones són més curtes, mantenint constants la resta de variables, és a dir per a un mateix rendiment acadèmic. Aquestes diferències entre sexes, però, són mínimes si les comparem amb les que es produeixen per modificacions en el rendiment acadèmic, essent aquest darrer el que més determina el sentit de les decisions envers l'abandonament, tal i com havia quedat en evidència en la modelització anterior. En aquest cas, degut a la presència de les interaccions entre el rendiment i el temps es posa encara més de manifest.

A continuació tenim la representació gràfica de les funcions de supervivència estimades pels tres individus per a les trajectòries 1 i 5, verificant-se, per una banda, com les corresponents als individus A i B es superposen mentre que la del C es troba lleugerament per sota de les altres dues i, per una altra, que el desplaçament important es produeix en modificar la trajectòria acadèmica de l'alumne.

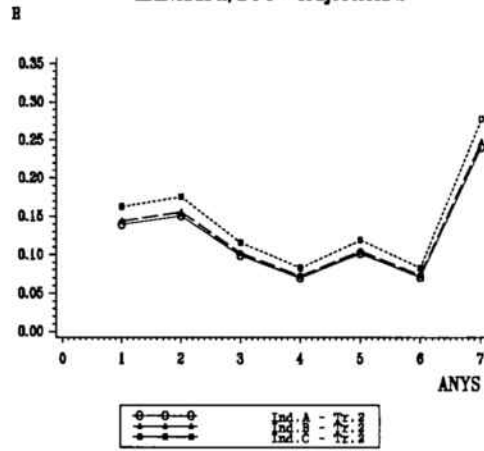


De la mateixa forma que s'ha realitzat per a la modelització sense les interaccions del temps amb el rendiment, és interessant visualitzar les funcions d'atzar estimades en la nova situació segons la qual ja no es verifica la hipòtesi de proporcionalitat. Si en primer lloc comparem les funcions d'atzar estimades pels individus A, B i C podrem observar l'efecte sobre les mateixes d'una variació en el SEXE, el qual, vist que la variable no interacciona amb el temps, haurà de ser aproximadament en paral·lel des d'una modalitat fins a l'altra. Efectivament en els gràfics següents s'observa el lleuger desplaçament paral·lel, és a dir, d'aproximadament la mateixa magnitud per a tots els moments del temps analitzats, que es produeix pel canvi de modalitat en la variable SEXE.

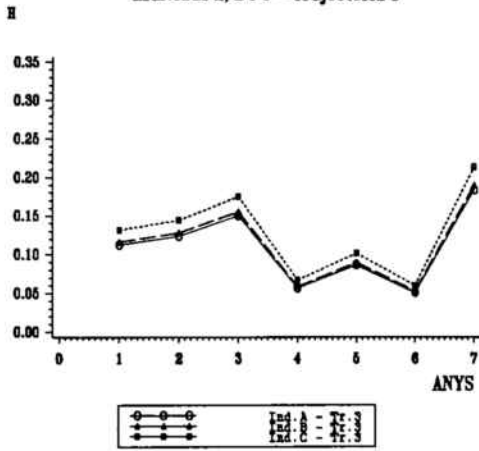
ATZAR DISCRET ESTIMAT
Individus A, B i C - Trajectoria 1



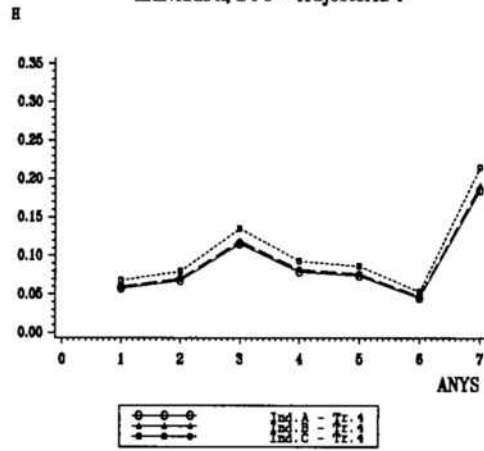
ATZAR DISCRET ESTIMAT
Individus A, B i C - Trajectoria 2



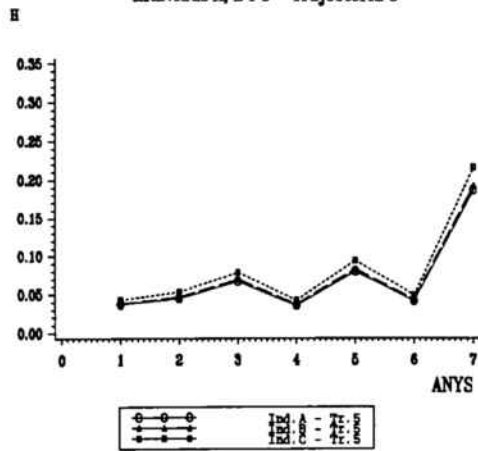
ATZAR DISCRET ESTIMAT
Individus A, B i C - Trajectoria 3



ATZAR DISCRET ESTIMAT
Individus A, B i C - Trajectoria 4



ATZAR DISCRET ESTIMAT
Individus A, B i C - Trajectoria 5



La distància que separa les funcions d'atzar dels individus A i C la qual és aproximadament la mateixa en tots els anys vindrà determinada pel valor del paràmetre estimat corresponent a la variable. El valor d'aquest serà la diferència entre les transformacions loglogarítmiques complementàries dels dos atzars corresponents a cadascuna de les modalitats, el que representem en la següent taula corresponent a la trajectòria 1 i sabent que el paràmetre estimat pel sexe pren un valor de -0.1654.

	Any1		Any4	
	A : HOME	C: DONA	A: HOME	C: DONA
Atzar ($\hat{\lambda}$)	0.31468	0.35972	0.11871	0.13852
$\log(-\log(1-\lambda))$	-0.973208	-0.807773	-2.06855	-1.90312
Diferència	A - C = -0.165435		A - C = -0.16543	

Pel que fa a la forma que presenten les funcions d'atzar estimades un cop introduïdes les interaccions entre el temps i el rendiment acadèmic, s'observen notables diferències respecte la modelització anterior. Així, per exemple, en el cas de la trajectòria més satisfactòria (Traj.5) quan el rendiment tenia un efecte igual per a tots els anys la funció d'atzar mostrava una tendència lleugerament decreixent que s'alterava en el darrer any amb una petita alça, mentre que en la nova modelització observem una tendència contrària, és a dir, creixent a mesura que avança el temps. Aquest comportament reflexa clarament com per aquesta trajectòria tot i ser la més satisfactòria presenta una baixada en el rendiment acadèmic de l'estudiant a partir del cinquè any, la qual cosa té conseqüències (petites) sobre la probabilitat d'abandonar condicionada a haver arribat fins aquells anys.

Si ens centrem en la trajectòria més insuficient (Traj. 1) constatem com en la nova modelització el risc més elevat d'abandonament es produeix en el primer any i un cop passat aquest moment crític comença a baixar fins el quart curs acadèmic moment a partir del qual torna a anar a l'alça. Tot i això, aquestes probabilitats condicionades són molt més elevades que les de la trajectòria 5 en tots els cursos acadèmics.

Tant aquests resultats de la nova modelització com els de la que no incorporava les interaccions entre el temps i el rendiment acadèmic tenen en comú un increment en el risc d'abandonar en el setè any o, el que és el mateix, un descens en la funció de supervivència. Aquesta situació no és del tot coherent amb el que seria d'esperar, especialment si fem referència a la trajectòria 5, la més satisfactòria d'entre les seleccionades. Això s'explica tenint en compte que la base de dades corresponents als expedients acadèmics només està actualitzada fins el sisè any, havent mantingut els valors del rendiment acadèmic constants a partir d'aquest curs acadèmic, és a dir, en el setè any el rendiment és el mateix que en l'any anterior donat que no es disposa de la informació sobre els resultats del curs acadèmic. Aquest fet haurà produït una estimació incorrecta del paràmetre corresponent a l'indicador IPAC en el setè any, concretament una sobreestimació de l'efecte del rendiment acadèmic en aquest curs en raó a que reflexaria una aturada en la progressió dels estudiants. En conseqüència caldria interpretar amb cura els resultats pel que fa al risc d'abandonament, i supervivència associada en el darrer any analitzat, així com de les durades medianes estimades per aquells casos en els quals aquelles es situen per sobre dels 6 anys.

11.2 Sub població de llicenciats

Sobre la sub població dels llicenciats podrien fer-se els mateixos comentaris respecte les modelitzacions alternatives contínua i discretes (cloglog i logística). Podem observar novament com l'estimació exacta del model sobre el temps continu coincideix amb les del model discret reparametritzat amb la funció loglogarítmica complementaria. És destacable en aquest cas la diferència notable existent entre les dues especificacions discretes alternatives, tant pel que fa a les estimacions puntuals com als seus errors estàndards i en conseqüència sobre les significacions estadístiques de les variables presents en el model. Per a la sub població d'abandonaments, doncs, no seria gens aconsellable l'aproximació logística a la modelització de l'atzar discreta.

En la modelització discreta realitzada han estat introduïdes les variables dicotòmiques D05 a D08 associades als diferents moments del temps a analitzar. En la sub població dels llicenciats el període d'estudi va des del cinquè any fins el vuitè mentre que en la d'abandonaments anava des del primer any fins el setè. Això és degut a que en la sub població de llicenciats no es produeix el succés que defineix la durada abans del cinquè curs acadèmic, és a dir no hi ha cap estudiant que s'hagi llicenciat en quatre o menys cursos acadèmics per la qual totes les observacions amb durades inferiors als cinc anys seran observacions censurades vist que no és possible que siguin llicenciats. Per tant, en els primers quatre anys no existeixen llicenciats i l'anàlisi de supervivència comença a partir del cinquè.

D'altra banda i també com a diferència del que havia estat definit en el cas de la sub població d'abandonaments, el període analitzat arriba ara fins el vuitè curs acadèmic mentre que en el cas anterior arribava només fins el setè any.

ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC SUBOBLACIÓ DE LICENCIATS (Especificacions Contínua i Discretes)

Variable Dependent: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
 Valors de Censura: 1 (Continua) 2 (Abandona)
 Mètode d'estimació: Màxima Versemblança

Total Observacions: 1351
 Llicenciats: 314
 Obs.Censurades: 1037 % Censura: 76,76

Contrast Global Hipòtesi Nulla: BETA=0

MODEL CONTINU				MODEL DISCRET CLOGLOG				MODEL DISCRET LOGIT			
Criteria	Sense Covariants	Amb Covariants	Model Chi-Square	Criteria	Sense Covariants	Amb Covariants	Model Chi-Square	Criteria	Sense Covariants	Amb Covariants	Model Chi-Square
-2 LOG L	1814,825	784,436	1030,389 (p=0,0001)	-2 LOG L	5704,048	804,063	4899,985 (p=0,0001)	-2 LOG L	4189,382	792,517	3396,864 (p=0,0001)
Score	.	.	606,071 (p=0,0001)	Score	.	.	3908,135 (p=0,0001)	Score	.	.	2159,063 (p=0,0001)
Wald	.	.	346,078 (p=0,0001)	AIC	5704,048	818,063	.	AIC	4189,382	806,517	.
				SC	5704,048	860,158	.	SC	4189,382	848,613	.
Variable	Parametre Estimat	Error Standard	Wald Chi-Square	Variable	Parametre Estimat	Error Standard	Wald Chi-Square	Variable	Parametre Estimat	Error Standard	Wald Chi-Square
				D05	-9,6257	0,5601	295,3927 (P=0,0001)	D05	-12,2534	0,8019	233,4823 (p=0,0001)
				D06	-9,1860	0,5363	293,4173 (P=0,0001)	D06	-11,9236	0,7840	231,2914 (p=0,0001)
				D07	-8,3774	0,4819	302,1973 (P=0,0001)	D07	-10,6426	0,6926	236,1198 (p=0,0001)
				D08	-10,6668	0,6814	245,0618 (P=0,0001)	D08	-13,3712	0,8881	226,7035 (p=0,0001)
NAC	-1,712445	1,06134	2,60329 (p=0,1066)	NAC	-1,7161	0,9886	3,0133 (P=0,0826)	NAC	-2,1762	1,4627	2,2136 (p=0,1368)
RESID2	-0,485546	0,17649	7,56857 (p=0,0059)	RESID2	-0,4882	0,1864	6,8573 (P=0,0088)	RESID2	0,0201	0,2614	0,0059 (p=0,9388)
IPAC	0,209498	0,01159	326,70506 (p=0,0001)	IPAC	0,2099	0,0124	286,4122 (P=0,0001)	IPAC	0,2837	0,0185	233,9056 (p=0,0001)

L'explicació es troba en que la base de dades disponible, malgrat no tenir actualitzats els expedients pels dos darrers cursos acadèmics (91/92 i 92/93), si que disposa d'alguna informació sobre alumnes que s'han llicenciat en el curs 92/93, els quals, naturalment, ho han fet en la convocatòria extraordinària de Final de Carrera que es realitza al febrer. Així, s'han enregistrat alumnes llicenciats en el vuitè curs acadèmic, tot i que l'esmentat curs encara no ha finalitzat. Per contra, en la subpoblació d'abandonaments, el fet que encara no ha acabat el curs 92/93 (el vuitè any del període analitzat) i que tampoc s'ha realitzat la matriculació pel curs següent, no és possible disposar de la informació dels abandonaments en el mateix, de manera que totes les observacions seran censurades, no existirà cap succés definidor de la durada i, per tant, el període de temps a analitzar anirà del primer fins al setè any.

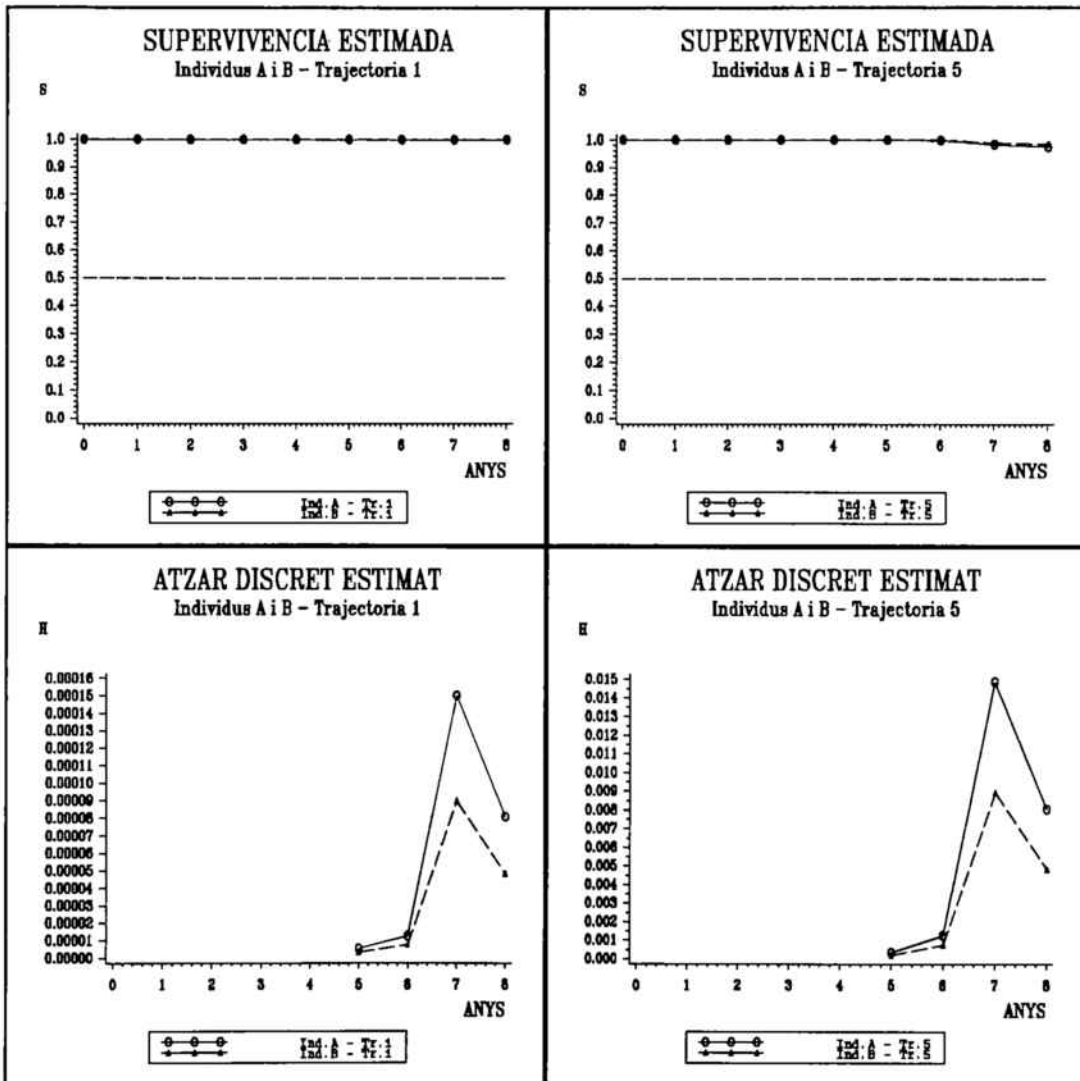
Seleccionada la modelització més adient per a les durades en la subpoblació dels llicenciats, la discreta reparametritzada mitjançant la funció loglogarítmica complementària, estem en disposició de representar les funcions de supervivència i d'atzar estimades, segons els individus que s'han anat caracteritzant al llarg del treball i les trajectòries en el rendiment acadèmic definides anteriorment. Els individus A i B que ens serviran per a realitzar l'anàlisi comparativa són ara, ateses les variables significatives del model:

Individu A: Alumne de nacionalitat espanyola ($NAC=1$) i que resideix en domicili diferent del patern o propi ($RESID2=1$).

Individu B: Alumne de nacionalitat espanyola ($NAC=1$) i que resideix en el domicili patern o propi ($RESID2=0$).

Pel que fa a les trajectòries acadèmiques, analitzarem les que es van definir en la subpoblació d'abandonaments, de tal manera que els individus pels quals havíem estimat una durada mediana i representat les seves funcions de supervivència i d'atzar des de l'òptica de la finalització sense èxit dels seus estudis, podran ara comptar amb una anàlisi de la durada equivalent però des de la perspectiva de la finalització de la mateixa per l'obtenció del títol de llicenciat.

L'elecció d'unes trajectòries més satisfactòries que les definides anteriorment ens portarien a resultats coneguts si tenim en compte, per exemple, que per a un alumne que superi en cada curs acadèmic una cinquena part (60) del total de crèdits necessaris per a l'obtenció del títol de llicenciat (300), enregistrarem una durada de 5 anys amb una probabilitat igual a la unitat. Aquesta situació determinista per a la modelització de la durada en la subpoblació dels llicenciats explicaria la minvada presència de variables personals o de gestió acadèmica prou significatives en el model, la qual cosa s'ha d'interpretar en el sentit que en la durada dels estudis fins a la seva finalització amb èxit l'únic factor realment important és el rendiment acadèmic, el qual actua d'una manera determinista sobre la durada, mentre que les característiques personals de l'alumne en tot cas tindran influència sobre el rendiment acadèmic i, per via d'aquest, indirectament sobre la durada. Pel contrari, en la durada fins a l'abandonament les característiques personals, a més de condicionar el rendiment acadèmic, també algunes d'elles tenen rellevància en l'explicació de l'heterogeneïtat de les durades. En altres paraules, la relació entre rendiment i durada no és determinista en aquest cas sinó que existeix un grau d'aleatorietat o heterogeneïtat no explicada per aquell i que sí explicarien algunes característiques personals de l'alumne. Així, com hem vist abans, en igualtat de condicions personals i de rendiment acadèmic, una dona tindria una tendència més gran a abandonar els estudis.



Les funcions de supervivència estimades pels individus A i B en les trajectòries acadèmiques 1 i 5 ens posen de manifest l'elevada durada dels estudis fins a la finalització amb èxit dels mateixos. Les diferències entre els individus A i B, per a una mateix nivell en el rendiment acadèmic, novament són molt reduïdes. Les trajectòries considerades són molt insuficients, fins i tot la 5, des de la perspectiva de l'obtenció del títol de llicenciat i, per tant, les durades medianes molt superiors als vuit anys analitzats.

Aquests resultats no fan res més que posar de manifest com, amb una progressió acadèmica del tipus de les seleccionades, el nombre d'anys necessari per a llicenciar-se seria molt elevat i que aquesta relació entre rendiment acadèmic i durada és determinista, com s'ha comentat abans, vist que a 30 crèdits per any l'estudiant trigaria 10 anys a llicenciar-se.

En aquest context de la subpoblació dels llicenciats es fa encara més necessària la contrastació de la hipòtesi de proporcionalitat de l'atzar en el model. Així, tot i la relació determinista entre el rendiment acadèmic i la durada dels estudis en el cas de les trajectòries acadèmiques més satisfactòries, és d'esperar que a mesura que l'alumne va progressant en el seu rendiment acadèmic amb l'objectiu d'aconseguir la finalització de la carrera, les probabilitats de llicenciar-se condicionades a haver arribat fins aquell curs acadèmic aniran augmentant any darrera any. La representació gràfica de les funcions d'atzar estimades per a les trajectòries 1 i 5 en els individus A i B no reflexen aquest increment en l'atzar o risc d'abandonar a mesura que va passant el temps i s'evoluciona en el rendiment. Verificarem doncs la hipòtesi de proporcionalitat mitjançant la interacció del temps i l'indicador IPAC. Un cop estimat aquest model amb les interaccions comprovem que efectivament no es verifica la hipòtesi de l'atzar proporcional essent tots els paràmetres de les interaccions significativament diferents de zero i, per tant, l'efecte del rendiment acadèmic sobre la durada dels estudis fins a la finalització amb èxit dels mateixos dependrà del moment del temps en que es mesuri. Aquest efecte és d'esperar que sigui cada cop més important vist que a mesura que la progressió acadèmica assoleix uns nivells elevats restaran a l'alumne un nombre menor de crèdits a superar i, per tant, la probabilitat de finalitzar la carrera, condicionada a haver arribat fins aquell punt, serà cada cop més gran.

ESTIMACIONS DEL MODEL SEMIPARAMÈTRIC
SUBPOBLACIÓ DE LLICENCIATS
(Especificació Discreta amb Interaccions)

Variable Dependent: ANYS DURADA DELS ESTUDIS
 Variable de Censura: SITUACIÓ ACADÈMICA ALUMNE
 Valor(s) de Censura: 1 (Continua) 2 (Abandona)
 Mètode Estimació: Màxima Versemblança (CLOGLOG)

Total Observacions: 1351
 Abandonaments: 314
 Obs.Censurades: 1037 % Censura: 76.76

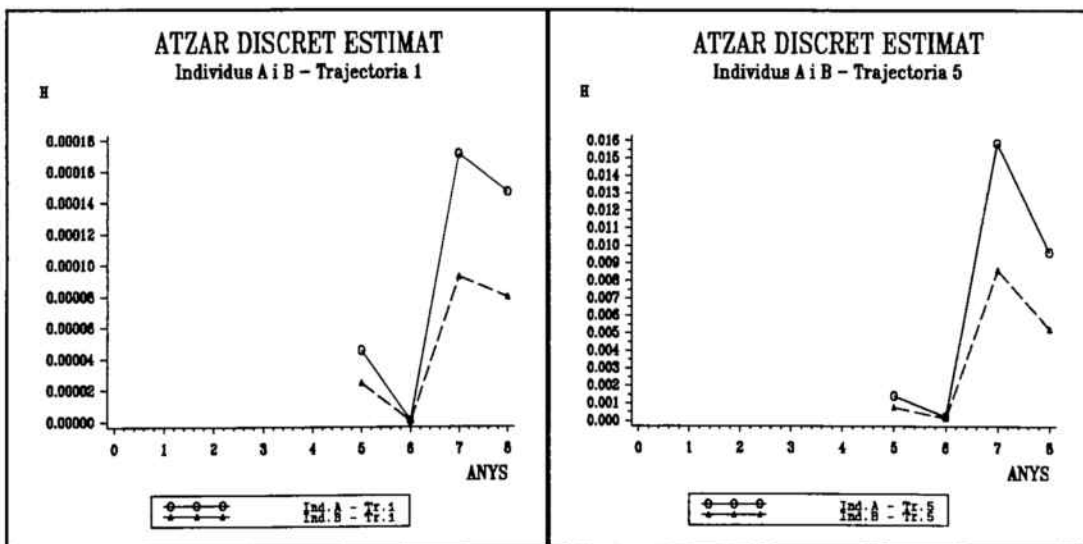
Contrast Global Hipòtesi Nul.la: BETA=0

Criteri	Sense	Amb	Chi-Square del Model
	Covariants	Covariants	
-2 LOG L Score	5704.048	786.595	4917.453 amb 10 DF (p=0.0001) 3938.849 amb 10 DF (p=0.0001)
AIC	5704.048	806.595	.
SC	5704.048	866.732	.

Variable	Paràmetre Estimat	Error Standard	Wald Chi-Square	Pr > Chi-Square
D05	-7.9275	0.6780	136.7272	0.0001
D06	-11.4721	1.2313	86.8148	0.0001
D07	-8.2059	0.8890	85.2066	0.0001
D08	-9.9203	2.9087	11.6322	0.0006
NAC	-1.9258	0.9939	3.7539	0.0527
RESID2	-0.6139	0.1906	10.3764	0.0013
IPAD05	0.1721	0.0153	125.7773	0.0001
IPAD06	0.2627	0.0285	85.2313	0.0001
IPAD07	0.2060	0.0236	76.1661	0.0001
IPAD08	0.1902	0.0793	5.7538	0.0165

Aquest resultat esperat no es reflexa en les estimacions del model obtingudes a on el paràmetre associat al rendiment acadèmic en cada moment del temps augmenta el seu valor de l'any cinc a l'any 6 però a partir del setè any torna a experimentar una baixada. Això s'explica, tal com ha estat comentat anteriorment, pel fet que la base de dades disponible no està actualitzada en els expedients acadèmics pels darrers dos anys del període analitzat, el que suposa haver enregistrat, pels cursos 7 i 8, zero crèdits superats per a tots els alumnes matriculats. Aquesta situació haurà introduït un biaix a la baixa en l'estimació dels paràmetres del rendiment acadèmic pels setè i vuitè anys.

El biaix present en les estimacions dels paràmetres del rendiment acadèmic pels dos darrers anys es veurà reflexat també sobre les funcions d'atzar estimades, les quals haurien de tenir una forma creixent des del cinquè curs acadèmic en endavant. L'evolució creixent de l'atzar es veu alterada per la situació anteriorment comentada i la forma inferida de les estimacions obtingudes és la de creixement en els primers anys i decreixement en els darrers, resultat que no s'ajusta a la realitat.



Les funcions d'atzar estimades pels individus A i B segons les trajectòries 1 i 5, ens mostren com respecte els individus A i B els desplaçaments són aproximadament paral·lels indicant que es verifica la hipòtesi de proporcionalitat en les característiques personals, mentre que si comparem les dues trajectòries acadèmiques els efectes sobre l'atzar són diferents segons el nivell del rendiment de l'estudiant en cada moment del temps analitzat. Per a la correcta lectura dels gràfics cal tenir present com en el cas de la trajectòria 1 l'escala sobre la qual es situa l'atzar estimat oscil·la entre el 0 i el 0.00018 mentre que per a la trajectòria 5 arriba fins el 0.016 i, per tant, la probabilitat de llicenciar-se és molt més gran quan el rendiment acadèmic és més satisfactori.

12. Conclusions

Capítol 12

Conclusions

Aquest capítol final està dedicat a sintetitzar els principals resultats obtinguts al llarg del treball mirant d'establir l'oportuna connexió amb els seus objectius. En el primer capítol es van enunciar les principals qüestions plantejades al voltant de l'anàlisi de la durada dels estudis o permanència, la primera de les quals feia referència a la possibilitat d'afirmar que el risc de finalització dels estudis és canviant per a diferents moments de la trajectòria acadèmica de l'alumne. Aquesta qüestió ha estat tractada des del punt de vista metodològic mitjançant l'avaluació dels resultats que proporcionen les tècniques estadístiques habitualment utilitzades en el tractament de variables dicotòmiques. Així, s'ha posat de manifest com els valors de les prediccions obtingudes mitjançant els esmentats mètodes tenen distribucions diferents segons quina sigui la durada dels estudis, el què voldria dir que la probabilitat condicional o risc d'acabament de la durada depèn del moment del temps en que sigui mesurada. Així les coses, ha calgut emprar una metodologia que tingui en compte aquesta variabilitat en el temps del risc d'abandonar els estudis o de finalitzar-los amb èxit, és a dir, que a més de respondre a la qüestió de *si* un alumne finalitza o no els estudis ens permeti conèixer *quan* és més probable que ho faci.

En el context de l'anàlisi de dades de supervivència serà, doncs, possible respondre a la segona qüestió plantejada en el començament del treball sobre si les característiques individuals de l'alumne tenen influència sobre la probabilitat condicional o risc d'acabament dels estudis. En aquest punt ha estat necessari distingir entre dos conjunts de característiques de l'alumne. D'una banda, les que tenen a veure amb la seva situació personal (sexe, edat, estat civil, situació laboral,...) i acadèmica (classe de matrícula, horari de classes, centre,...) i, d'altra, les referides al seu rendiment acadèmic, essent el tractament d'aquestes darreres més complex ja que incorporen una component de variabilitat en el temps.

Mitjançant mètodes d'anàlisi alternatius, com ara contrastos estadístics no paramètrics, modelitzacions paramètriques que requereixen d'una hipòtesi distribucional prèvia sobre la durada dels estudis, i modelitzacions semiparamètriques que permeten obviar aquesta darrera hipòtesi, s'ha verificat que, efectivament, les característiques personals i acadèmiques de l'alumne, en el cas de considerar *únicament* aquestes, expliquen part de les diferències en les probabilitats condicionals o riscos d'acabament dels estudis, especialment en els abandonaments.

Així, individualment considerades, les variables EDAT, ESTAT CIVIL (Solter vs. altres), NACIONALITAT (Espanyola vs. Estrangera), TIPUS DE RESIDÈNCIA (Pròpia vs. Pares o Altres), SITUACIÓ LABORAL (Treballa vs. No Treballa), CENTRE (Facultat d'Econòmiques vs. Abat Oliba), FONT DE FINANÇAMENT DELS ESTUDIS (Treball vs. Altres vs. Pares o Beca), HORARI (Matí vs. Tarda-Nit), CLASSE DE MATRÍCULA (Gratuïta vs. Ordinària o Semigratuïta), TÍTOL D'ACCÉS A LA UNIVERSITAT (COU vs. altres) i RESULTATS PROVES D'ACCÉS A LA UNIVERSITAT mostren una influència

significativa sobre la durada dels estudis fins a l'abandonament, és a dir, per a cadascuna de les categories de les variables existeixen diferències respecte el risc d'abandonar en diferents moments de la trajectòria acadèmica de l'alumne. Mentre que, per a la durada dels estudis fins a l'obtenció del títol, a les variables anteriors s'afegiria la del DOMICILI (Barcelona i cinturó industrial vs. altres comarques).

Malgrat tot, les anteriors característiques personals i acadèmiques es troben molt interrelacionades entre elles, existint doncs informació redundat, tal com posen de manifest els resultats d'incloure conjuntament totes les variables en els contrastos estadístics i modelitzacions dutes a terme. Així, per exemple, els alumnes no solters són majoritàriament els de més edat, els quals treballen, financen els estudis mitjançant el seu propi treball i resideixen en domicili propi.

Tanmateix, de la incorporació a les modelitzacions del rendiment acadèmic de l'estudiant, en forma d'indicador de la progressió acadèmica mesurat segons el nombre d'assignatures/crèdits superades/ts respecte el total necessari per a l'obtenció del títol, s'obtenen uns resultats que modifiquen substancialment les anteriors conclusions, les quals han estat extretes sobre la base de considerar, únicament i exclusiva, les característiques personals i acadèmiques. Així, variables que es mostraven associades a la durada dels estudis i al risc d'acabament d'aquests quan eren tractades conjuntament totes elles (TIPUS DE RESIDÈNCIA en Domicili Propi, CENTRE, HORARI DE CLASSES, RESULTATS P.A.A.U.), han deixat d'estar-ho mentre que d'altres han esdevingut ara rellevants (SEXE) quan no ho eren ni tan sols individualment. Aquest fet evidencia la forta interdependència existent entre les característiques personals i el rendiment acadèmic.

D'altra banda, el rendiment acadèmic és el factor més important en la determinació de la durada dels estudis, molt per sobre de les altres característiques personals que encara es mostren rellevants. Aquestes darreres són, en el cas de la durada fins a l'abandonament dels estudis, les variables SEXE, DOMICILI, TIPUS DE RESIDÈNCIA Altres, FONT DE FINANÇAMENT DELS ESTUDIS Treball Propi i Altres, i NACIONALITAT. La principal conclusió que s'extreu d'aquests resultats és, en primer lloc, que els efectes del rendiment acadèmic sobre les durades dels estudis són molt més importants que els que es trobarien associats a les característiques personals i de caire acadèmic. En segon lloc, que aquestes darreres característiques individuals produeixen efectes sobre el risc d'acabament dels estudis per via del rendiment acadèmic, és a dir, condicionen el rendiment i aquest la durada. En darrer lloc, les característiques individuals que encara són rellevants en l'explicació de la durada dels estudis al costat del rendiment acadèmic tenen, majoritàriament, un caràcter econòmic la qual cosa implicaria que davant d'un mateix rendiment acadèmic dos estudiants prendran decisions relatives a l'abandonament o continuació dels estudis que seran diferents segons siguin les seves condicions econòmiques recollides indirectament mitjançant les variables DOMICILI (cost del desplaçament fins el centre on es cursen els estudis), TIPUS DE RESIDÈNCIA Altres (cost del pis d'estudiants, residència, etc.) i FONT DE FINANÇAMENT DELS ESTUDIS Treball Propi i Altres (cost dels estudis assumit directament per l'estudiant).

En darrer terme, ha estat incorporada en les modelitzacions una variable dicotòmica que recull la situació de l'estudiant davant de la normativa de permanència recentment aprovada pel Consell Social de la U.B.. Aquesta variable complementa la mesura del rendiment acadèmic ja que en l'aplicació de la normativa intervé com a factor més important el nombre acumulat de crèdits

matriculats per l'estudiant en cada curs acadèmic de tal forma que, darrera d'un mateix rendiment acadèmic mesurat en termes d'assignatures/crèdits superats pot existir un desigual esforç per part de diferents estudiants en el sentit que cadascun d'ells pot haver necessitat matricular un nombre diferent d'assignatures/crèdits per a arribar a aquell nivell en el rendiment acadèmic. En conseqüència, la variable de situació davant la normativa de permanència incorpora en certa mesura la relació èxit/fracàs en la superació de les assignatures/crèdits i resulta ser de molta importància en la determinació de la durada dels estudis fins a l'abandonament, sense alterar substancialment els resultats per a la resta de variables rellevants.

Una altra qüestió incorporada en les modelitzacions de les durades fins a l'acabament dels estudis ha estat la possibilitat de què els efectes sobre el risc o probabilitat condicional de finalització de la durada no siguin iguals en diferents moments del temps. Així, s'ha verificat com l'efecte del rendiment acadèmic, el qual és per ell mateix canviant en el temps, sobre el risc d'abandonar o llicenciar-se no és independent del moment del temps en que es trobi l'alumne, és a dir, que un mateix nombre d'assignatures/crèdits superades/ts no tindran les mateixes repercussions sobre el risc d'abandonar/llicenciar-se en el primer any d'estada a la universitat que després d'uns quants anys.

Dels anteriors resultats es poden formular les següents conclusions que tenen a veure amb les polítiques de permanència. Primerament, s'ha verificat com els alumnes que no superen la normativa de permanència recentment aprovada i que, per tant, haurien estat exclosos en el cas de ser vigent per a la cohort analitzada, tenen unes probabilitats condicionals o riscos d'abandonar molt superiors a les d'aquells alumnes que no es veurien afectats per la normativa, amb la qual cosa s'assegura l'avançament de l'abandonament i es minimitza el cost de la normativa.

En segon lloc, vist que, principalment, el rendiment acadèmic és el factor que determina l'abandonament, mentre que les característiques personals i de caire acadèmic actuen sobre aquest darrer de forma indirecta per via del rendiment, caldrà que la normativa s'instrumentalitzï mitjançant uns nivells mínims en el rendiment acadèmic, considerant aquest darrer de forma acumulada.

En tercer lloc i relacionat amb l'anterior, per tal que la normativa no sigui discriminatòria davant de situacions personals que donen lloc a un rendiment acadèmic baix, caldrà que els mínims en el rendiment per sota dels quals l'alumne seria exclòs tinguin en compte la decisió de l'estudiant respecte el ritme al que vol progressar en els estudis, és a dir, haurà de ser l'estudiant el que, ateses les seves condicions personals, prengui una decisió sobre el ritme en la progressió acadèmica i si aconsegueix mantenir-lo li serà garantida la permanència. Obviament, serà obligada la fixació d'una trajectòria mínima per sota de la qual cap estudiant podria establir el seu ritme de progressió.

Finalment, i responent parcialment a la darrera qüestió plantejada en la introducció d'aquest treball referida a la relació existent entre les dues causes de finalització de les durades dels estudis, abandonament i obtenció del títol, cal dir que la preocupació de la normativa haurà de ser la de la minimització del risc d'excloure un alumne amb una probabilitat condicional o risc elevat de llicenciar-se. En aquest sentit, els resultats obtinguts no són prou concloents degut a la manca d'un període de temps suficientment allunyat per a l'anàlisi de la durada fins a l'acabament amb èxit dels estudis, mancança que es concreta en la disponibilitat de només tres cursos acadèmics en els què ha estat possible llicenciar-se el que fa que no existeixin enregistraments d'alumnes llicenciats amb ritmes de progressió acadèmica més lents i vorejant els mínims de la normativa.

Aquesta darrera qüestió ens permet introduir les possibles línies de desenvolupament posterior del treball. En primer lloc, l'inevitable pas del temps i el cada dia més depurat i complet sistema informàtic de gestió acadèmica a la Universitat de Barcelona permetran disposar en uns quants anys d'una base de dades molt més ampla per a l'anàlisi de la durada dels estudis, als efectes de possibilitar el tractament de diferents cohorts d'estudiants i per a tots els ensenyaments de la universitat. A més l'actualització anual de la base de dades analitzada facilitarà la validació dels resultats obtinguts. L'anàlisi de les cohorts d'estudiants que segueixen els nous plans d'estudis posats en marxa enguany serà de gran interès, especialment pel que fa als efectes de l'aplicació de la normativa de permanència. Les alternatives de modelització presentades en aquest treball garanteixen la possibilitat d'obtenció de resultats adients malgrat que la base de dades disponible sigui cada cop més gran vist que, en el cas extrem de fer computacionalment impossible l'estimació d'alguns dels models estudiats, mitjançant la modelització discreta per via de la reparametrització logística, queda assegurada l'obtenció, relativament senzilla, de bones aproximacions.

D'altra banda i des del punt de vista metodològic, resten obertes dues línies de recerca posteriors com són, primerament, la de la validació de les modelitzacions de les durades aquí presentades, les quals tenen a veure amb la realització de contrastos d'especificació, anàlisi de residus, etc. Altrament, resulta obligada la incorporació en els models d'una component aleatòria que reculli l'heterogeneïtat no observada (equivalent al terme de pertorbació aleatòria en el model de regressió). En aquesta línia es situen els recents desenvolupaments econòmics de la metodologia de l'anàlisi de dades de durada o supervivència, conegut en aquest context, de forma més general, per anàlisi de dades de transició (Lancaster, 1990).

Bibliografia

Bibliografia

- Alemany, R., A. Coduras i A. Costa (1990). *Rendiment Acadèmic i Permanència a la Universitat de Barcelona*. Consell Social. Universitat de Barcelona.
- Allison, P.D. (1982). Discrete-time methods for the analysis of event-histories, en *Sociological Methodology*. San Francisco. Josey-Bass.
- Artís, M. i R. Alemany (1991). Rendimiento Académico y Permanencia de los estudiantes, V Reunión ASEPELT-España, Gran Canaria.
- Benedito, V., A. Vicens i Altamira (1970). *Estudio sobre las Tasas de Abandono en la Universidad de Barcelona*, ICE, Universidad de Barcelona.
- Berkson, J. i R.P. Gage (1952). Survival Curve for Cancer Patients Following Treatment, *Journal of the American Statistical Association*, 47, 501-515.
- Bloosfeld, H.P. (1990). Changes in educational carrers in the Federal Republic of Germany, *Sociology of Education*, 63, 165-177.
- Breslow, N.E. (1970). A Generalized Kruskal-Wallis Test for Comparing K-Samples subject to Unequal Patterns of Censorship, *Biometrika*, 57, 579-594.

- Breslow, N.E. (1974). Covariance Analysis of censored survival data. *Biometrics*, 30, 89-100.
- Breslow, N.E. i J.Crowley (1974). A Large Sample Study of the Life Table and Product Limit Estimates under Random Censorship, *Annals of Statistics*, 2, 437-453.
- Brown, C.C. (1975). On the use of indicator variables for studying the time dependence of parameters in a response-time model. *Biometrics*, 31, 863-872.
- Bryk, A.S. i Y.M. Thum (1989). The effects of high school organization on dropping out: An exploratory investigation. *American Educational Research Journal*, 26, 353-383.
- Buckley, P.J. y G. J. Hooley (1988). The non-completion of doctoral research in Management: Symptoms, Causes and Cures. *Educational Research*, 30, 110-119.
- Caterall, J.S. (1987). On the social costs of dropping out from school. *The High School Journal*, 71, 19-30.
- Chiang, C.L. (1960). A Stochastic Study of Life Table and its Applications: I. Probability distribution of the biometric functions, *Biometrics*, 16, 618-635.
- Cox, D.R. (1959). The analysis of exponentially distributed life-times with two types of failure. *Journal of The Royal Statistical Society*, 21, 411-421.

- Cox, D.R. (1972). Regression Models and Life Tables (with discussion), *Journal of the Royal Statistical Society, B*, 26, 103-110,
- Cox, D.R. (1975). Partial Likelihood. *Biometrika*, 62, 269-276.
- Cox, D.R. i D.V. Hinkley (1974). *Theoretical Statistics*. Chapman-Hall. London.
- Cox, D.R. i D. Oakes (1984). *Analysis of survival data*. Chapman-Hall. London.
- Crowley, J. i M. Hu (1977). Covariance Analysis of Heart Transplant Survival Data, *Journal of The American Statistical Association*, 72, 27-36.
- Cutler, S.J. i F.Ederer (1958). Maximum Utilization of the Life Table in Analizing Survival, *Journal of Chronic Diseases*, 8, 699-712.
- Efron, B. (1967). The Two Sample Problem with Censored Data. *Proc. Fifth Berkeley Symposium in Mathematical Statistics*, IV, New York: Prentice Hall, 831-853.
- Efron, B. (1977). Efficiency of Cox's Likelihood function for censored data. *Journal of the American Statistical Association*, 72, 557-565.
- Efron, B. (1988). Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve, *Journal of the American Statistical Association*, 83, 414-425.
- Elandt-Johnson, R.C. i N.L. Johnson (1980). *Survival Models and Data Analysis*. New York. Wiley.

Escudero, T. (1986). Algunos criterios y evidencias del rendimiento universitario, en M.Latiesa, *Jornadas sobre Educación Superior*.

Frank, A.R. i T.Z. Keith (1984). Academic abilities of persons entering and remaining in special education, *Exceptional Children*, 51, 76-77.

Fraser, D.A.S. (1968). *The Structure of Inference*. New York. Wiley.

Gail, M.H. (1975). A review and critique of some models used in competing risks analysis, *Biometrics*, 31, 209-222.

Garcia Aretio, L. (1987). *Rendimiento Académico y Abandono en la Educación Superior a Distancia*, ICE, U.N.E.D.

Garcia Llamas, J.L. (1986). El Análisis Discriminante y su Utilización en la Predicción del Rendimiento Académico, *Revista de Educación*, 280, 229-254.

Gehan, E. (1965). A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-censored Samples, *Biometrika*, 52, 203-223.

Gehan, E. (1969). Estimating Survivor Functions from the Life Table, *Journal of Chronic Diseases*, 21, 629-644.

Gilbert, J.P. (1962). *Random Censorship*. PhD Dissertation. University of Chicago.

- Gómez, G. i S.W. Lagakos (1993). Estimation of the infection time and latency distribution of AIDS with doubly censored data, *Biometrika*, propera publicació.
- González Tirados, R.M. (1986). Un método para el análisis del fracaso escolar universitario: Estudio piloto. en M.Latiesa, *Jornadas de Educación Superior*.
- Greenwood, M. (1926). The errors of sampling of the survivorship tables. *Reports on Public Health and Statistical Subjects*, 33, London, HMSO, App.1.
- Gross, A.J. i V.A. Clark (1975). *Survival Distributions: Reliability Applications in the Biomedical Sciences*, New York, Wiley.
- Herrero, S. i A. Infestas (1980). *El Rendimiento Académico en la Universidad*, ICE, Universidad de Salamanca.
- Hogan, D.P. (1984). Cohort comparisons in the timing of life events. *Developmental Review*, 4, 289-310.
- Jiménez Fernández, C. (1987). Abandono Estudiantil en la Universidad a Distancia. Un Estudio Empírico sobre su Evolución y Predicción. (I) y (II). *Revista de Educación*, 283 i 284, 285-316 i 317-347.
- Johansen, S (1978). The product limit estimator as maximum likelihood estimator. *Scandinavian Journal of Statistics*, 5, 195-199.

- Johnes, J. i J. Taylor (1989). Undergraduate non-completion rates: differences between U.K. Universities. *Higher Education*, 18, 209-225.
- Johnson, N.L. i S. Kotz (1970). *Distribution in Statistics. Continuous univariate distributions*, (2 vols.), Boston. Houghton Mifflin.
- Kalbfleisch, J.D. i R.L. Prentice (1973). Marginal Likelihoods based on Cox's Regression and Life Model, *Biometrika*, 60, 267-278.
- Kalbfleisch, J.D. i R.L. Prentice (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Kalbfleisch, J.D. i D.A. Sprott (1970). Application of likelihood methods to models involving a large number of parameters (with discussion). *Journal of the Royal Statistical Society*. 32, 175-208.
- Kaplan, E.L. i P.Meier (1958). Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, 53, 457-481.
- Kennan, J. (1985). The duration of contract strikes in U.S. Manufacturing, *Journal of Econometrics*, 28, 5-28.
- Kiefer, N.M. (1988). Economic Duration Data and Hazard Functions, *Journal of Economic Literature*, 26, 646-679.
- Laird, N. i D. Olivier (1981). Covariance analysis of censored survival data using log-linear analysis techniques, *J. Am. Stat. Assoc.*, 76, 231-240.

- Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment, *Econometrica*, 47, 939-956.
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge University Press. Cambridge.
- Latiesa, M. (1986). *Demanda de Educación Superior y Rendimiento Académico en la Universidad*, (Comp.), CIDE, Jornades organitzades pel CIDE i la Secretaria del Consejo de Universidades.
- Lawless, J.F. (1982) *Statistical methods for lifetime data*. Wiley. New York.
- Levin, H.M. (1972) The costs to the nation of inadequate education. en *The effects of dropping out*. U.S. Government Office, Washington, D.C.
- Lynch, M. (1991) The duration of invalidity benefit claims: a proportional hazard model. *Applied Economics*, 23, 1043-1052.
- Mann, N.R., R.E. Schafer i N.D. Singpurwalla (1974). *Methods for Statistical Analysis of Reliability and Life Data*, New York, Wiley.
- Mantel, N. (1963). Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure, *Journal of the American Statistical Association*, 58, 690-700.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration, *Cancer Chemother Rep.*, 50, 163-170.

- Mantel, N. i W. Haenszel (1959). Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of the National Cancer Institute*, 22, 719-748.
- Meier, J. (1977). Estimation of a Distribution Function from Incomplete Observations. *Perspectives in Probability and Statistics (J.Gani Ed.)*, New York, Academic Press, 67-87.
- Miller, R.G. (1981). *Survival Analysis*. New York. Wiley.
- Moltó, T. i E. Oroval (1984). *Costes y Rendimientos en la Enseñanza Superior*, ICE, Universitat de Barcelona.
- Moncada, A. (1986). *Estudio sobre las causas de retraso y abandono en los estudios universitarios*. (Mimeo)
- Murdock, T.A. (1987). It isn't just money: The effects of financial aid on student persistence, *The Review of Higher Education*, 11, 75-101.
- Newby, M. (1985). A further note on the distribution of strike duration. *Journal of the Royal Statistical Society, A*, 148, 350-356.
- Nickell, S.J. (1979). Estimating the Probability of Leaving Unemployment, *Econometrica*, 47, 1249-1266.
- Oakes, D. (1981). Survival times: aspects of partial likelihood (with discussion). *International Statistical Review*, 49, 199-233.

- Oroval, E. (1986). Resultados del análisis individualizado del rendimiento de una promoción de estudiantes en la Universidad de Barcelona, en M. Latiesa, *Jornadas de Educación Superior*.
- Ott, M.D. (1988). An analysis of predictors of early academic dismissal, *Research in Higher Education*, 28, 34-48.
- Peto, R. (1972). Contribution to discussion paper by D.R.Cox. *Journal of the Royal Statistical Society*. 34, 205-207.
- Peto, R. i J. Peto (1972). Assymptotically efficient rank invariant test procedures (with discussion), *Journal of the Royal Statistical Society*, 135, 185-206.
- Prentice, R.L. (1978). Linear Rank Tests with Right Censored Data, *Biometrika*, 65, 167-179.
- Prentice, R.L. i L.A.Gloeckler (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34, 57-67.
- Prentice, R.L., J.D. Kalbfleisch, A.V. Peterson, N. Flournoy, V.T. Farewell i N.E. Breslow (1978). *The Analysis of Failure Times in the Presence of Competing Risks*, *Biometrics*, 34, 541-554.
- Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. New York, John Wiley and sons.

- Rubio, J. (1968). El abandono de los estudios en la Enseñanza Superior, *Revista de Educación*, 197.
- Rumberger, R.W., R. Ghatak, G. Poulos, P.L. Ritter i S.M. Dornbusch (1990). Family influences on dropout behavior in one California High School, *Sociology of Education*, 63, 283-299.
- Salvador, L. i Garcia-Valcárcel, A. (1989). *El Rendimiento Académico en la Universidad de Cantabria: Abandono y Retraso en los Estudios*, CIDE, Ministerio de Educación y Ciencia.
- SAS (1990). *SAS/STAT user's guide*. (Version 6). Cary, N.C. SAS Institute.
- Savage, I.R. (1956). Contributions to the Theory of Rank Order Statistics- The Two Sample Case, *Ann. Math. Stat.*, 27, 590-615.
- Schlechty, P.C. i V.S. Vance (1981). Do academically able teachers leave education? The North Carolina case, *Phi Delta Kappan*, 63, 106-112.
- Seal, H.L. (1977). Studies in the history of probability and statistics, XXXV, Multiple decrements or competing risks, *Biometrika*, 64, 429-439.
- Serrat, A. i F.J. Gil (1991). Movilidad en el mercado de trabajo: Análisis de la duración de los procesos económicos con varias salidas. *V Reunión ASEPELT-España*, Gran Canaria, Junio.

- Shaw, D. (1988). On-site samples regression, *Journal of Econometrics*, 37, 211-233.
- Simpson, C., Baker, K. y G. Mellinger (1980). Conventional failures and unconventional dropouts: comparing different types of university withdrawals. *Sociology of Education*, 53, 203-214.
- Singer, J.D. i J.B.Willet (1993). It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics*, 18, 155-196.
- Spady, W. (1970). Dropouts from higher education: An interdisciplinary review and synthesis, *Interchange*, 1, 109-121.
- Sprott, D.A. i J.D.Kalbfleisch (1969). Examples of likelihoods and comparison with point estimates and large sample approximations. *Journal of the American Statistical Association*, 64, 429-439.
- Stacy, E.W. (1962). A Generalization of the Gamma Distribution. *Ann. Math. Stat.*, 33, 1187-1192.
- Stage, F.K. (1988). University Attrition: LISREL with logistic regression for persistence criterion, *Research in Higher Education*, 29, 343-357.
- Stage, F.K. (1989). Motivation, academic and social integration and the early dropout, *American Educational Research Journal*, 26, 385-402.

- Tarone, R.E. i J. Ware (1977). On Distribution-free tests for equality of survival distributions, *Biometrika*, 64, 156-160.
- Terenzini, P. (1982). Designing attrition studies. *New directions for institutional research: Studying student attrition* (E.Pascarella Ed.), 55-71. San Francisco: Jossey-Bass.
- Terenzini, P. i E. Pascarella (1980). Towards the Validation of Tinto's Model of College Attrition: A review of recent studies, *Research in Higher Education*, 12, 271-282
- Thompson, W.A. (1977). On the Treatment of Grouped Observations in Life Studies, *Biometrics*, 33, 463-470.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research, *Review of Educational Research*, 45, 89-125.
- Watson, C. (1974). *Focus on drop outs*. Ontario Institute for the Study of Education. Toronto.
- Willet, J.B. i J.D. Singer. (1991). From Whether to When: New Methods for Studying Student Dropout and Teacher Attrition, *Review of Educational Research*, 61, 407-450.