

Un modello multilivello per l'analisi della durata degli studi universitari¹

Bruno Chiandotto, Roberta Varriale

Dipartimento di Statistica "G. Parenti" - Università degli Studi di Firenze

Riassunto. Nel lavoro si analizza il fenomeno dei tempi di conseguimento della laurea, una delle maggiori criticità del sistema universitario italiano. Per cercare di individuare le possibili determinanti è stata svolta un'analisi sia sui dati di archivio relativi agli studenti immatricolatisi presso l'Ateneo fiorentino nel ventennio 1980-2000, sia sui dati (di archivio e raccolti nell'ambito del progetto AlmaLaurea) relativi ai laureati nell'anno solare 2000. Su questi ultimi dati, avendo come finalità la misura dell'effetto "netto" esercitato dai fattori individuali e da fattori istituzionali (variabili specifiche dei corsi di studio) sui tempi di conseguimento del titolo, è stato introdotto un *modello lineare gerarchico a due livelli*; tale modello tiene conto del fatto che gli studenti (unità di primo livello) risultano naturalmente aggregati nei Corsi di laurea (unità di secondo livello).

Parole chiave: Tempi di conseguimento del titolo, Modelli multilivello, Regressione lineare gerarchica.

1. Introduzione

Tra gli aspetti negativi che hanno caratterizzato e caratterizzano ancora oggi il sistema universitario italiano assumono particolare rilevanza gli abbandoni e la durata delle carriere: la percentuale di studenti che abbandonano gli studi in Italia è eccessiva e, per coloro che invece riescono a conseguire il titolo universitario, il tempo impiegato per concludere il percorso di studi è troppo elevato.

¹ Il presente lavoro è stato finanziato nell'ambito del progetto "Transizioni Università-lavoro e valorizzazione delle competenze professionali dei laureati: modelli e metodi di analisi multidimensionali delle determinanti", cofinanziato dal MIUR; coordinatore nazionale è Luigi Fabbris, coordinatore del gruppo di Firenze è Bruno Chiandotto (titolo del progetto dell'unità di ricerca locale "Valutazione del processo formativo universitario, sbocchi professionali e pianificazione dei percorsi formativi: modelli e metodi"). L'idea iniziale, la struttura e l'impostazione del lavoro sono dovuti al contributo di entrambi gli autori, mentre le elaborazioni e l'implementazione del modello vanno attribuite a R. Varriale.

Il problema dell'eccessiva durata delle carriere universitarie, tipico del sistema universitario italiano, appare ancora più accentuato se si analizza la situazione dell'Ateneo fiorentino (Chiandotto B. e Bertaccini B., 2003), il che induce a presumere che, su questo fenomeno, le conclusioni di un approfondimento conoscitivo utilizzando i dati fiorentini possano essere ragionevolmente estese anche a gran parte degli altri Atenei italiani. L'individuazione delle possibili determinanti del fenomeno dei tempi di conseguimento del titolo eccessivamente lunghi dovrebbe suggerire interventi finalizzati alla eliminazione di una tale criticità².

Il secondo paragrafo di questa nota è dedicato ad una sintetica illustrazione dei risultati dell'analisi svolta sugli immatricolati presso l'Università di Firenze negli anni accademici dal 1980/81 al 2000/01, finalizzata all'individuazione dell'eventuale influenza esercitata sulla durata degli studi sia dal corso di laurea sia da caratteristiche individuali (quali genere, residenza, diploma di scuola superiore, ecc.)³.

Nel terzo paragrafo vengono riassunti, altrettanto sinteticamente, i risultati dell'analisi relativa agli studenti che, essendosi immatricolati presso l'Università di Firenze dall'anno accademico 1980/81 all'anno accademico 2000/01, e non avendo mai effettuato un passaggio di corso di studi, hanno conseguito la laurea presso l'Università di Firenze durante l'anno solare 2000⁴.

Successivamente, facendo sempre riferimento ai laureati dell'anno 2000, per pervenire alla misura dell'effetto "netto" eventualmente esercitato da possibili determinanti (sia individuali che istituzionali) sulla durata degli studi, sono stati introdotti i modelli gerarchici o di regressione multilivello; il ricorso a tali modelli è stato suggerito dalla struttura dei dati che è di tipo gerarchico a due livelli: le unità di primo livello sono gli studenti, quelle di secondo livello sono i corsi di laurea. I risultati delle analisi condotte sono riportati nel quarto paragrafo; alcune conclusioni completano la nota.

² Sul problema della valutazione dei processi formativi finalizzata alla eliminazione di eventuali criticità presenti nel sistema universitario si veda Chiandotto B. (2002).

³ Una trattazione più dettagliata si trova in Varriale R. (2004), un altro significativo contributo sull'argomento è stato fornito da Bulgarelli G. (2002).

⁴ Anche in questo caso si tratta di un'esposizione estremamente sintetica, maggiori dettagli si trovano in Varriale R. (2004), sullo stesso argomento si può utilmente consultare Chiandotto B., Bacci S. e Bertaccini B. (2004).

2. Esito degli studi universitari degli immatricolati nell'Ateneo fiorentino nel periodo 1980-2000

Secondo la definizione utilizzata dall'Istat⁵, sono stati considerati immatricolati gli studenti "iscritti per la prima volta al primo anno di un corso di laurea o di diploma universitario"; in particolare, sono stati esaminati gli immatricolati ai soli Corsi di laurea.

Ai fini dell'analisi, è stato adottato l'approccio longitudinale, scegliendo come evento di comune origine l'immatricolazione presso l'Università di Firenze in un determinato anno accademico; all'interno della popolazione oggetto di studio sono state, pertanto, individuate 21 coorti. Ogni coorte è stata osservata per 10 anni⁶, trascorso tale periodo lo studente può: aver abbandonato gli studi (*abbandono*), essersi già laureato (*laureato*), essere ancora iscritto (*iscritto*). Per valutare il fenomeno dei tempi di conseguimento del titolo sono state, pertanto, considerate 13 coorti.

I laureati presso l'Ateneo fiorentino tra il 1980 e il 31 luglio 2003, immatricolatisi nel periodo intercorso tra l'a.a. 1980/81 e l'a.a. 1992/93, sono stati 32636; a livello di Ateneo il tasso medio di laurea è del 30.4%; mentre i tassi registrati per le diverse Facoltà variano da un minimo di 22.9% per Scienze della formazione ad un massimo di 40.6% per Medicina e Chirurgia.

Solo il 3.2% del totale dei laureati dell'Ateneo completa il ciclo di studi in corso, mentre più dell'80% lo fa con almeno 2 anni di ritardo. Il tempo che gli studenti impiegano per concludere gli studi universitari ha un ovvio riflesso sull'età che i laureati stessi hanno al conseguimento del titolo: l'età media di Ateneo è di 26.8 anni, leggermente più elevata per i maschi (27 anni) rispetto alle femmine (26.7 anni), abbastanza simile per tutte le Facoltà, tranne che per la Facoltà di Scienze della Formazione nella quale i laureati hanno un'età media di quasi 28 anni.

⁵ Gli studenti rientranti nella suddetta definizione di immatricolati sono stati classificati in base al Corso di laurea di prima iscrizione; per tali studenti si dispone di informazioni classificabili in "variabili d'ingresso" (principalmente dati anagrafici e relativi agli studi pre-universitari), "di soggiorno" (per esempio informazioni su eventuali passaggi di corso, rinunce) e "d'uscita" (esito finale degli studi). Le variabili d'ingresso e quelle "in itinere" rappresentano i fattori individuali, o variabili esplicative, che si suppone possano influenzare l'esito e la durata degli studi. Tali informazioni risultano aggiornate, per ciascuna delle unità di analisi, al 31 luglio 2003.

⁶ Nella scelta di tale periodo si è tenuto conto che la durata media degli studi risulta pari a 7-8 anni e che entro 8 anni dall'immatricolazione si registra più del 70% del totale delle lauree osservate per ogni generazione. Inoltre, è stato rilevato che il tasso marginale di crescita del tasso di laurea tende nel tempo ad attestarsi su valori abbastanza costanti e che scegliendo un periodo di osservazione pari a 15 anni gli studenti che avrebbero fatto parte del collettivo di riferimento per le successive analisi sarebbero stati solo 69560 (anziché 107267) sui 174072 studenti immatricolatisi tra l'a.a. 1980/81 e l'a.a. 2000/01.

Ovviamente, anche l'analisi basata sull'indice di durata⁷ porta a conclusioni non confortanti. A livello di Ateneo, infatti, l'indice di durata medio assume il valore di 1.7: questo vuol dire che gli studenti impiegano più di una volta e mezzo del tempo ritenuto necessario per legge a terminare gli studi universitari.

La Facoltà che presenta l'indice di durata più basso è Medicina e Chirurgia (1.24), mentre la Facoltà con l'indice più alto è Economia (1.9), seguita da Lettere e Filosofia (1.86) e Giurisprudenza (1.83). Come prevedibile, sono gli studenti con un diploma di tipo liceale e coloro che hanno riportato votazioni più elevate all'esame di maturità a presentare valori più bassi dell'indice di durata.

3. Tempi di conseguimento del titolo dei laureati nell'anno solare 2000

In questa parte del lavoro vengono analizzati i dati forniti dal Consorzio Interuniversitario AlmaLaurea⁸ relativi agli studenti che, appartenendo alla popolazione esaminata nel paragrafo precedente, hanno conseguito la laurea presso l'Università degli studi di Firenze durante l'anno solare 2000; i dati utilizzati sono stati forniti dal Consorzio Interuniversitario AlmaLaurea e derivano sia da indagini predisposte nell'ambito del Progetto AlmaLaurea sia da fonti di tipo amministrativo. Il tipo di strumento utilizzato per le prime è il questionario strutturato compilato dai laureandi (tasso di risposta pari al 91%) al conseguimento del titolo, composto da domande chiuse a risposta unica e suddiviso in 6 aree tematiche che riguardano: notizie anagrafiche, curriculum scolastico e formativo, notizie sull'esperienza universitaria appena conclusa, situazione lavorativa, notizie sulla famiglia, intenzioni e prospettive future.

I dati AlmaLaurea sono stati successivamente integrati con quelli forniti dall'Ufficio Servizi Statistici e Controllo di Gestione dell'Università di Firenze uti-

⁷ L'indice di durata (I_d) delle singole Facoltà e CdL è costruito rapportando la durata effettiva degli studi alla durata legale del rispettivo corso; tale indice rende possibile il confronto tra laureati delle diverse Facoltà e diversi CdL, cresce al crescere del ritardo e assume valori maggiori o uguali a 1.

⁸ Il consorzio interuniversitario *ALMALAUREA* nasce nel 1994 per iniziativa dell'Osservatorio Statistico dell'Università di Bologna ed attualmente è gestito dalle Università aderenti con il sostegno del Ministero dell'Istruzione, dell'Università e della Ricerca.

I principali obiettivi dei servizi offerti da *ALMALAUREA* sono, da una parte, quelli di assicurare agli organi di governo degli atenei appartenenti al consorzio, ai nuclei di valutazione, alle commissioni impegnate nella didattica e nell'orientamento, attendibili e tempestive basi documentarie e di verifica, volte a favorire i processi decisionali e la programmazione delle attività; dall'altra di creare una sempre più stretta collaborazione tra università e mondo produttivo, facilitando, attraverso la propria banca dati, l'accesso dei giovani al mercato del lavoro italiano ed internazionale.

Per ulteriori informazioni, si può consultare il sito Internet: www.almalaurea.it

lizzati per le analisi a cui si è riferito nel paragrafo precedente, in modo da consentire il confronto dei risultati conseguiti nelle due diverse analisi. Nel procedere all'integrazione tra i due insiemi di dati, però, alcuni records relativi a studenti laureati presso l'Ateneo fiorentino nell'anno solare 2000 non sono stati utilizzati⁹; per tale motivo il collettivo di riferimento oggetto delle successive analisi è risultato composto da 4382 studenti. Inoltre, sono stati esclusi dalle indagini quei laureati che non hanno compiuto l'intero ciclo di studi nell'Ateneo fiorentino e nello stesso corso di laurea, ottenendo una popolazione di riferimento **3978** unità.

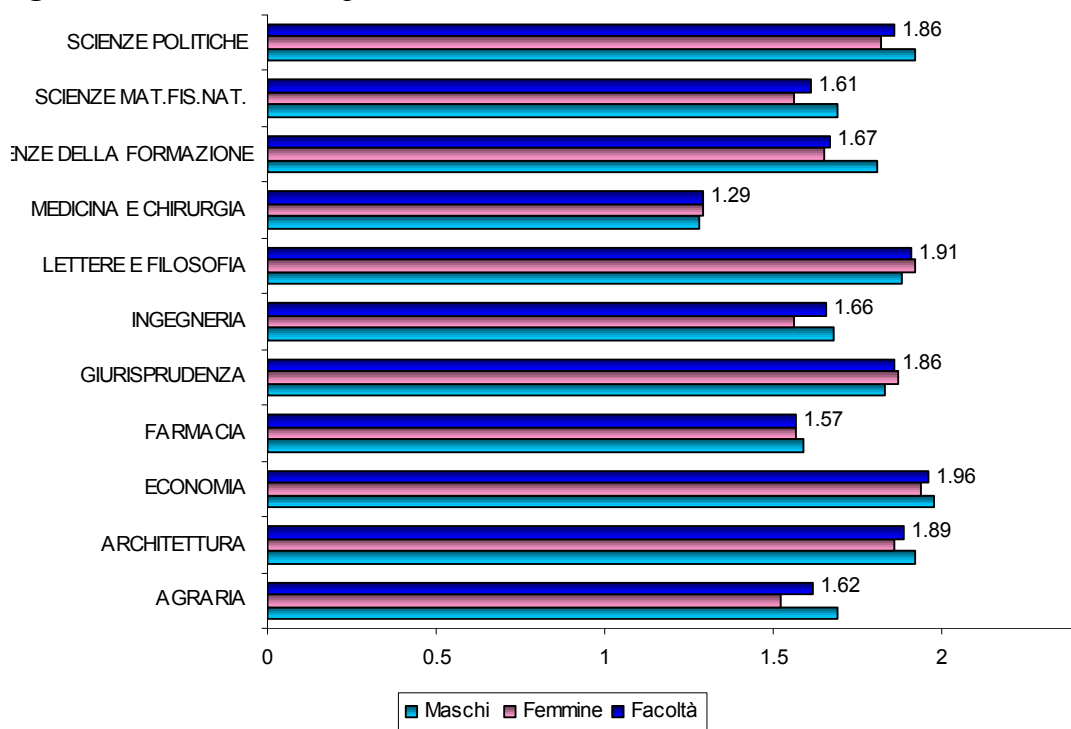
Come già sottolineato, uno degli aspetti più negativi che caratterizza la figura del laureato "tipo" è l'età molto elevata al conseguimento del titolo, età elevata che è il diretto riflesso dell'eccessiva durata degli studi; infatti, solo l'11% dei laureati consegue il titolo ad un'età inferiore ai 24 anni, mentre quasi il 30% termina gli studi universitari ad un'età superiore ai 28 anni. A livello di Facoltà, i laureati in Architettura presentano un'età media al conseguimento del titolo più elevata (29.6), mentre i laureati in Scienze Matematiche Fisiche e Naturali sono i più giovani con un'età media di 26.9 anni.

Naturalmente, la variabile ora analizzata serve solo a fornire una indicazione, seppure interessante, della durata degli studi universitari; infatti, molteplici sono i fattori che possono determinare l'innalzamento dell'età al raggiungimento del termine degli studi, e non tutti possono essere imputabili alla durata del piano di studi del corso prescelto. Si è preferito analizzare, pertanto, la durata degli studi universitari sia attraverso un approccio diretto, ovvero attraverso la sua misurazione in anni, sia indirettamente facendo ricorso all'indice di durata.

Dall'analisi della variabile *durata degli studi* per il contingente dei laureati dell'anno solare 2000, il risultato è tutt'altro che confortante. A livello di Ateneo, confrontando le durate legali delle varie Facoltà con quelle medie effettivamente impiegate dagli studenti, si può notare come gli studenti impiegano mediamente da 1.5 a 4.4 anni in più rispetto al tempo considerato necessario per legge al raggiungimento del termine degli studi. Inoltre, mentre solo il 5% degli studenti analizzati si laurea in corso, ben il 50% di questi si laurea dopo il quarto anno fuori corso.

Facendo riferimento all'indice di durata, all'interno dell'intero Ateneo si registra un valore medio pari a 1.8: questo significa che gli studenti impiegano quasi il doppio del tempo ritenuto necessario per legge a completare gli studi universitari. I valori assunti da questo indice a livello di Facoltà sono riportati nella Figura 1.

⁹ Degli originari 4846 laureati: 195 studenti si sono immatricolati in altri Atenei, 53 studenti si sono immatricolati prima dell'a.a. 1980/81, 41 studenti erano già in possesso di altro titolo, 4 studenti si sono immatricolati ad un diploma, 140 studenti si sono già immatricolati in precedenza senza arrivare al conseguimento del titolo, 25 studenti si sono immatricolati ad anni successivi al primo, 6 studenti si sono immatricolati oltre il 31 Luglio di ogni anno.

Figura 1. Indice di durata per Facoltà e sesso

Successivamente, si è cercato di individuare le possibili relazioni esistenti tra i caratteri di interesse e la durata degli studi e, a tal fine, oltre ad effettuare un'analisi di tipo descrittivo, si è proceduto al computo di due statistiche di associazione, il Chi-quadro di Pearson e la V di Cramer¹⁰.

¹⁰ Questi indici sono calcolati attraverso il confronto tra le frequenze osservate e le frequenze teoriche nell'ipotesi di indipendenza stocastica tra i caratteri considerati e, mentre il primo è espresso in termini assoluti, l'indice di Cramer varia tra 0 e 1. A ragione del contesto di analisi, valori superiori a 0.10 di tale indice inducono a concludere a favore della presenza di un livello di dipendenza abbastanza elevato tra i caratteri analizzati.

La statistica *Chi-quadro* è solitamente utilizzata per l'analisi di distribuzioni discrete, ma può essere calcolata anche per distribuzioni continue raggruppando i dati in classi di modalità; considerazioni analoghe valgono per l'indice V di Cramer. Per questo motivo si è dovuto procedere alla suddivisione in classi dell'indice di durata e, dato che questa variabile misura il tempo impiegato dallo studente per conseguire il titolo e quindi cresce all'aumentare degli anni di iscrizione "fuori corso", nel farlo si è cercato di creare una corrispondenza tra classe dell'indice e il numero degli anni fuori corso; le classi così ottenute sono 6. Classe 1 – (0-1.14) – 0; Classe 2 – (1.1401-1.37) – 1; Classe 3 – (1.3701-1.62) – 2; Classe 4 – (1.6201-1.87) – 3; Classe 5 – (1.8701-2.12) – 4; Classe 6 – (2.1201-7) – Più di 4 anni, dove, rispettivamente, si riporta (in parentesi) l'intervallo dell'indice di durata ed il numero di anni fuori corso.

Se si procede al confronto tra il valore assunto dall'indice di durata per i due sessi si riscontra una lievissima differenza a livello complessivo (0.04) a favore del genere femminile, differenza questa che non presenta grande variabilità anche a livello di singole facoltà. Se si fa, invece, riferimento alle statistiche di associazione, si rileva la presenza di un legame tra i due caratteri essendo abbastanza elevato il valore assunto (0.10) dalla V di Cramer.

Un risultato inatteso è quello concernente la relazione esistente fra *residenza* degli studenti e *durata* degli studi universitari; infatti, la residenza degli studenti non sembra incidere in maniera così netta sul valore dell'indice di durata: tale valore rimane identico per gli studenti provenienti da Firenze e dalle province di Prato, Pistoia e Arezzo e lievemente più alto per coloro con residenza nelle altre province della Toscana; un valore più alto dell'indice, invece, si rileva per i giovani con residenza fuori dalla Toscana. Anche l'analisi delle statistiche di associazione (che rileva come la V di Cramer assume un valore pari a 0.08) sembra confermare la mancanza di un forte legame tra la residenza dello studente e la durata dei suoi studi universitari, ma è da ricordare ancora una volta che in realtà la variabile d'interesse sarebbe il *domicilio* degli studenti e non la loro residenza.

Una certa incidenza sulla *durata* degli studi ha, invece, il *titolo di studio dei genitori* dei laureati. Tale fatto è confermato sia dal valore della V di Cramer (0.12) sia dalla semplice lettura dei dati: da una situazione in cui entrambi i genitori sono laureati e l'indice assume un valore medio di 1.58 si passa a situazioni in cui in famiglia vi è al più una licenza elementare dove si registra un indice medio di 1.97. Piuttosto basso è invece il valore assunto dall'indice V (0.07) quando si considera la *classe sociale*¹¹ della famiglia di appartenenza.

Esaminando l'indice di durata in funzione della carriera preuniversitaria si osservano risultati del tutto prevedibili: i giovani con una carriera preuniversitaria regolare presentano un valore dell'indice di durata minore rispetto a chi ha affrontato l'esame di maturità con uno o più anni di ritardo; si registra una relazione inversa tra *voto alla maturità* e indice di durata (al crescere della votazione al diploma del laureato diminuisce il valore assunto dall'indice di durata); gli studenti che hanno concluso in tempi più contenuti la carriera universitaria sono quelli provenienti dal liceo scientifico e classico, mentre quelli che vi hanno impiegato più tempo provengono da altri tipi di maturità e scuole secondarie di tipo tecnico.

¹¹ Per la classificazione della variabile *classe sociale* si è adottato lo schema proposto da A. Cobalti e A. Schizzerotto, *La mobilità sociale in Italia*, Bologna, Il Mulino, 1994, adottato anche da *ALMALAUREA*. La posizione socio-economica può assumere le modalità borghesia, classe media impiegatizia, piccola borghesia e classe operaia. In proposito si segnala che gli imprenditori, i libero professionisti e i dirigenti, appartengono alla *borghesia* indipendentemente dal titolo; gli impiegati o intermedi con laurea sono nella *classe media impiegatizia*; i lavoratori in proprio, i soci di cooperative e i coadiuvanti appartengono alla *piccola borghesia*; gli impiegati con un titolo di studio della scuola dell'obbligo, gli operai ed i lavoratori a domicilio sono nella classe operaia

Le statistiche di associazione mostrano come vi sia un forte legame tra carriera preuniversitaria dello studente e tempo impiegato per conseguire la laurea: la significatività delle statistiche Chi-quadro di Pearson è sempre molto elevata e la V di Cramer è sempre superiore a 0.10. Da notare, è che la relazione più intensa si ha tra voto alla maturità e indice di durata (la V di Cramer è quasi pari a 0.13) a conferma dell'influenza esercitata da questo fattore sull'esito della carriera universitaria.

Per quanto riguarda le variabili relative alla carriera universitaria dello studente, è possibile innanzitutto osservare come, al crescere del *ritardo dell'immatricolazione* all'Università, cresca anche il valore dell'indice di durata. Ancora, è possibile osservare che coloro che hanno ottenuto risultati migliori sia a livello di *voto medio* agli esami che alla laurea impiegano meno tempo per conseguire il titolo. Il fatto che esista una forte relazione tra i risultati ottenuti dagli studenti sia per quanto riguarda la votazione conseguita sia per quanto riguarda il tempo di conseguimento del titolo è confermato, inoltre, dal calcolo delle statistiche di associazione: la significatività del Chi-quadro è sempre molto elevata e la V di Cramer supera in entrambi i casi il valore di 0.14.

Riguardo il modo di vivere l'esperienza universitaria, l'elevato valore della V di Cramer (0.21) mostra come vi sia un legame molto forte tra tempi di conseguimento del titolo e la *frequenza* alle lezioni; infatti, chi frequenta con regolarità tutti o quasi tutti i corsi impiega meno tempo ($I_d = 1.75$) rispetto a chi frequenta saltuariamente, al più, alcuni corsi ($I_d = 2$).

Il fenomeno della frequenza alle lezioni è sicuramente collegato a quello delle esperienze lavorative durante la carriera universitaria: chi non lavora conclude gli studi nettamente prima (il valore dell'indice di durata è di 1.64 contro 1.86 per coloro che lavorano) e tra chi lavora ha più difficoltà a mantenere un buon ritmo di studi chi ha un contratto di lavoro di tipo stabile rispetto a chi ha rapporti di lavoro di tipo occasionale.

Interessanti sono i risultati che si ottengono sia dal calcolo dell'indice di durata in relazione alla necessità o meno di svolgere attività di *stage o tirocinio* per il completamento degli studi sia dal calcolo delle statistiche di associazione tra queste variabili: queste attività sembrano incidere positivamente sulla durata della carriera universitaria (il valore dell'indice è di 1.54 per coloro che sono stati coinvolti in tali attività e 1.84 per gli altri) ed il loro legame con i tempi di conseguimento del titolo sembra molto forte (il valore della V di Cramer è ben 0.26).

In relazione al rapporto che può esistere tra tempi di laurea e posizione nei confronti degli *obblighi di leva*, si registra un valore molto alto dell'indice di durata per gli studenti che hanno già svolto il servizio militare o civile ($I_d = 2.02$), mentre una situazione migliore si rileva per chi si trova nella condizione di non dover svolgere il servizio militare ($I_d = 1.83$). Valori molto bassi dell'indice si osservano per gli studenti che stanno adempiendo agli obblighi di leva, o per coloro che stanno aspettando di farlo, probabilmente dovuto al fatto che, consapevoli dei propri obblighi, tali

studenti hanno organizzato in maniera migliore i propri piani di studio (la forte relazione tra questa variabile e i tempi di durata è rilevata dalla V di Cramer pari a 0.37).

Per quanto riguarda la *soddisfazione* sull'esperienza universitaria appena conclusa, coloro che ne danno un giudizio ottimo sono coloro che hanno impiegato meno tempo a concludere gli studi ($I_d = 1.57$), mentre coloro che impiegano più tempo non sono coloro che ne danno un giudizio pessimo, ma mediocre. Anche in questo caso, comunque, la V di Cramer assume un valore abbastanza alto (0.14).

Relativamente all'ipotesi di *reiscrizione* all'Università, inoltre, sono coloro che vorrebbero reinscrivere allo stesso corso di laurea ad aver impiegato meno tempo alla conclusione degli studi ($I_d = 1.76$), mentre gli studenti che non vorrebbero reinscrivere all'Università sono coloro per cui si osserva il valore dell'indice di durata più elevato (1.96).

Da segnalare, infine, il forte grado di associazione tra indice di durata e *Facoltà* (V di Cramer pari a 0.26) che risulta ancora più elevato quando si misura l'associazione tra durata e *Corso di studi* (V di Cramer pari a 0.36).

4. Il modello gerarchico a due livelli

Come anticipato nell'introduzione, per procedere ad una più soddisfacente individuazione delle possibili determinanti dei tempi di conseguimento del titolo si è fatto ricorso ai modelli di regressione multilivello¹².

Nei due paragrafi precedenti è stato analizzato il fenomeno della durata degli studi nell'Ateneo fiorentino, sia relativamente al periodo 1980-2000 che ai laureati dell'anno solare 2000, facendo ricorso a tecniche statistiche sostanzialmente di tipo descrittivo, tecniche che non consentono di pervenire ad una misura dell'effetto netto delle determinanti (fattori esplicativi) del fenomeno di interesse che nel caso specifico è rappresentato dall'indice di durata. Tale possibilità è, invece, offerta, dai modelli di regressione e, in particolare, dai modelli di regressione di tipo multilivello. In proposito, vale la pena ricordare che il ricorso ad un modello di regressione multilivello è consigliabile ogni volta che le unità (dette unità di primo livello) su cui si rileva il fenomeno oggetto di studio risultano naturalmente aggregate in gruppi differenti (le unità di secondo livello), che a loro volta possono essere aggregate in unità di terzo livello e così via: in tali casi è ragionevole ritenere che la variabilità del fenomeno dipenda non solo da variabili esplicative individuali (o di primo livello), ma altresì dal fatto che un certo individuo appartenga ad un determinato gruppo avente caratte-

¹² Al riguardo si possono consultare, tra gli altri, i volumi di Goldstein H. (2003) e Snijders A.B., Bosker R. J. (1999).

ristiche peculiari che lo contraddistinguono dagli altri gruppi; nel caso specifico trattato, gli studenti (unità di primo livello) risultano naturalmente aggregati in Corsi di laurea (unità di secondo livello).

Come già sottolineato, la variabile risposta considerata è l'indice di durata. Il "vantaggio" principale che proviene dall'utilizzo di questo indicatore è che attraverso di esso è possibile effettuare un confronto tra tempi di laurea osservati in diverse Facoltà e diversi CdL; unico "svantaggio" nell'utilizzo della variabile continua *ind_durata* è che la sua distribuzione ha un andamento di tipo normale, ma troncato a sinistra (il valore minimo osservato è 1)¹³.

I dati cui si farà riferimento sono quelli relativi ai laureati dell'anno solare 2000 (unità di primo livello), che si sono immatricolati per la prima volta nell'Ateneo fiorentino a partire dall'a.a. 1980/81 e che non hanno cambiato corso di studi. Come precedentemente illustrato, sono stati esclusi dall'analisi quei laureati che non hanno compiuto l'intero ciclo di studi nell'Ateneo fiorentino e nello stesso corso di laurea, ottenendo così una popolazione di riferimento di 3978 unità; l'insieme di dati così ottenuto è risultato, però, composto da numerosi records con dati mancanti relativi ad alcune variabili e, dato che il software utilizzato per l'applicazione del modello di regressione multilivello esclude questi records dall'analisi e volendo evitare di ricorrere a complicate tecniche di imputazione, si è proceduto a cancellare tali records, ottenendo così una popolazione di riferimento composta da **1896** osservazioni.

Come unità di secondo livello sono stati scelti i Corsi di laurea in cui gli studenti hanno conseguito il titolo di studio e non le rispettive Facoltà di appartenenza in quanto si è ritenuto che solo dall'analisi di questi si potessero ottenere informazioni utili alla spiegazione del fenomeno dei tempi di laurea all'interno dell'Ateneo fiorentino. Infatti, è possibile osservare come spesso le Facoltà sono suddivise al loro interno in CdL con caratteristiche sostanzialmente differenti tra loro. Inoltre, come prevedibile, anche dall'analisi dei tempi medi di laurea è possibile osservare come i diversi CdL all'interno delle stesse Facoltà differiscono notevolmente tra loro; solo per fare un esempio relativo alla Facoltà di Economia, il CdL in Scienze Statistiche ed Attuariali ha un indice di durata pari a 1.60 mentre per Economia e Commercio si osserva un valore dell'indice addirittura pari a 1.97 (quasi il doppio della durata legale!).

Un'ultima osservazione che conferma quanto appena descritto riguarda il già segnalato (cfr. par. 2) alto grado di associazione riscontrato sia tra i tempi di laurea degli studenti e le Facoltà frequentate, sia tra i tempi di laurea degli studenti ed i loro CdL: associazione molto forte nel primo caso, ma notevolmente più forte nel secondo.

¹³ Il fatto che la variabile *ind_durata* sia troncata a sinistra è una delle possibili cause della non normalità dei residui. Quest'ultimo aspetto ha comunque delle conseguenze soprattutto sul valore degli errori standard delle stime dei coefficienti di regressione e delle componenti di varianza (e di conseguenza sulla validità dei test utilizzati), e non su quello delle stime puntuali di tali parametri.

Coerentemente a quanto suggerito dalla teoria i passi della procedura di stima impiegata sono stati:

- stima del **modello nullo**, al fine di verificare, attraverso la stima della varianza dei residui di secondo livello, la ragionevolezza del ricorso ad un'analisi multilivello e di scomporre la variabilità del fenomeno tra variabilità *entro* i gruppi e variabilità *tra* gruppi.
- stima del **modello a intercetta casuale** comprendente tutte le variabili di primo livello (compresi termini quadratici e interazioni) risultate significative.
- stima del **modello a intercetta casuale finale**¹⁴ (due livelli) costituito dalle variabili di primo e secondo livello (compresi termini quadratici e interazioni) risultate significative.

Di seguito verranno riportati e commentati soltanto i risultati relativi alla stima dei due Modelli a intercetta casuale comprendenti, rispettivamente, le sole variabili di primo livello e quelle di primo e secondo livello relative alla stima finale.

Il modello utilizzato è un modello multilivello ad intercetta casuale del tipo:

$$Y_{ij} = (\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j) + (u_{0j} + r_{ij})$$

dove:

$$r_{ij} \sim iid N(0, \sigma^2) \quad , \quad u_{0j} \sim iid N(0, \tau_{00}) \quad ;$$

i residui del modello, r_{ij} e u_{0j} , sono indipendenti tra loro, j è l'indice utilizzato per descrivere i gruppi (corsi di laurea - $j = 1, 2, \dots, 37$), mentre i è l'indice che descrive le unità (laureati all'interno di ogni gruppo - $i = 1, 2, \dots, n_j$); la variabile risposta Y_{ij} (misurata al livello individuale) è rappresentata dall'indice di durata degli studi ed ha distribuzione di tipo normale troncata nella coda di sinistra (il valore minimo osservato è 1); X_{ij} indicano le variabili esplicative, di primo livello mentre Z_j rappresentano le variabili esplicative di secondo livello. I termini r_{ij} e u_{0j} rappresentano gli errori residui del modello (ovvero quella parte di variabilità di Y_{ij} che non è catturata dalle variabili esplicative) rispettivamente a livello individuale ed a livello di gruppo

Per l'applicazione del modello lineare gerarchico è stata utilizzata la PROC MIXED del software SAS¹⁵.

¹⁴ Si puntualizza che sono stati testati anche modelli con coefficienti casuali; in questo caso, però, l'algoritmo utilizzato non ha mai raggiunto la convergenza.

¹⁵ La procedura PROC MIXED permette di scegliere i metodi di stima dei parametri. Nel contesto dei modelli multilivello gli stimatori maggiormente impiegati sono quelli di *Massima Verosimiglianza* (Maximum Likelihood, ML) e quelli della *Massima Verosimiglianza residua* (Residual Maximum Likelihood, REML). Questi due metodi danno risultati molto simili per quanto riguarda la stima dei coefficienti di regressione mentre differiscono maggiormente nella stima delle componenti della varianza; inoltre, la devianza calcolata attraverso il metodo REML può essere utilizzata nei *test del rapporto di verosimiglianza* solo se i due modelli comparati sono composti dalla stessa parte fissa e differente parte casuale. Per tale motivo in questa applicazione è stato scelto di calcolare le stime dei coefficienti di regressione attraverso il metodo della *Massima Verosimiglianza* (specificando l'argomento METHOD=ML nella procedura PROC MIXED).

4.1 Il modello stimato ed i risultati ottenuti

Come già segnalato, il primo passo della procedura prevede la stima di un modello di regressione a due livelli senza variabili esplicative, ossia del cosiddetto modello nullo.

In particolare, attraverso il modello nullo è stato possibile esplicitare il coefficiente di correlazione intra-classe ρ , che misura il grado di omogeneità tra osservazioni appartenenti allo stesso gruppo: in questo caso, quasi il 40% della varianza totale dell'indice di durata è dovuta all'effetto del Corso di Laurea. Inoltre, è stato calcolato il valore della devianza - ovvero il grado di "non adattabilità" del modello (Hox J.J., 2002) - che è stato successivamente utilizzato come benchmark per il confronto di differenti modelli.

Una volta verificata l'effettiva esistenza di un'organizzazione dei dati su due livelli di analisi, lo studio è proseguito con la stima del modello multilivello (completo) contenente le variabili esplicative.

Relativamente alle variabili categoriche introdotte nel modello, si è reso necessario individuare per ciascuna di esse una modalità base o di riferimento nei confronti della quale valutare l'effetto di tutti gli altri livelli esistenti¹⁶, come in una normale regressione. Relativamente a tale aspetto si è scelto di considerare come modalità base di ciascuna covariata la caratteristica più diffusa nella popolazione di studio, secondo i risultati ottenuti attraverso l'analisi descrittiva preliminare; l'individuo che possiede tutte queste caratteristiche verrà denominato *individuo-base*.

Per quanto riguarda le variabili continue, per semplificare l'interpretazione dei risultati, si è scelto di centrare le variabili continue di primo livello rispetto alla loro media di gruppo e quelle di secondo livello rispetto alla loro media totale¹⁷. Si è scelto dunque di utilizzare l'approccio *group mean centering*, che consente di considerare il cosiddetto "frog-pond effect" (Hox J.J., 2002). Nel caso qui considerato, ad esempio, scegliendo come variabile esplicativa il voto al diploma riportato dagli studenti (*voto*), studiare questo effetto vuol dire analizzare come varia la relazione tra tempo di conseguimento del titolo e la variabile voto, in dipendenza dal voto medio al diploma osservabile nello specifico corso di laurea.

Le variabili esplicative utilizzate per la selezione del modello sono state scelte in base ai suggerimenti forniti dalla conoscenza del fenomeno e alle conclusioni risultanti dall'analisi descrittiva svolta.

¹⁶ Dovendo utilizzare queste variabili ai fini dell'applicazione ai dati del modello multilivello, si è scelto talvolta di adottare classificazioni delle variabili meno dettagliate rispetto a quelle presentate durante l'analisi descrittiva. Naturalmente, un'analisi che tenga presente classificazioni più dettagliate delle variabili potrà essere spunto per approfondimenti successivi.

¹⁷ Il calcolo della media di gruppo e della media totale delle medie di gruppo, come suggerito da Snijders e Bosker (1999), è stato effettuato su tutti i valori individuali a disposizione per la determinata variabile analizzata prima della procedura di cancellazione dei records aventi dati mancanti relativi ad altre variabili.

Le variabili di primo livello, riportate in Tabella 1, possono essere classificate in tre gruppi, ognuno dei quali riguarda un differente aspetto della vita dello studente:

- **variabili legate ai caratteri strutturali** (variabili 1-4)
- **variabili legate alla preparazione preuniversitaria** (variabili 5-7)
- **variabili legate all'esperienza universitaria** (variabili 8-14).

Tabella 1. Variabili esplicative di primo livello

| n. | Nome variabile | Descrizione | Modalità di risposta | Categoria di riferimento |
|----|-------------------|---|--|--|
| 1 | <i>sexso</i> | Sesso | 1= maschi 2= femmine | Femmine |
| 2 | <i>residenza</i> | Residenza | 1= fuori Toscana 2= altra provincia Toscana 3= Po - Pt - Ar 4= Firenze | Firenze |
| 3 | <i>tit_gen</i> | Titolo di studio dei genitori ¹⁸ | 1= al più un diploma inferiore 2= almeno un genitore con diploma superiore 3= almeno un genitore con laurea | Almeno un genitore con laurea |
| 4 | <i>cl_sociale</i> | Classe sociale della famiglia di origine | 1= borghesia 2= classe operaia 3= classe media impiegatizia o piccola borghesia | Classe media impiegatizia o piccola borghesia |
| 5 | <i>diploma</i> | Tipo di diploma di scuola superiore | 1= altro 2= tecnico 3= liceale | Liceale |
| 6 | <i>voto_dipl</i> | Voto al diploma di scuola superiore (in sessantesimi) | | Variabile centrata rispetto alla media di gruppo |
| 7 | <i>eta_dipl</i> | Età al diploma di scuola superiore | 1= maggiore di 19 anni (percorso di studi non regolare) 2= minore o uguale a 19 anni (percorso di studi regolare) | Percorso di studi regolare |
| 8 | <i>rit_iscr</i> | Tempo tra l'esame di maturità e l'iscrizione all'Università | 1= maggiore o uguale a 1 anno 2= 0 anni | 0 anni |

¹⁸ Per questa variabile è stata scelta come categoria di riferimento non la modalità più frequente (che sarebbe stata "almeno un genitore con diploma superiore"), ma quella ritenuta più significativa per studiare il contributo del livello di istruzione presente in famiglia sui tempi di laurea dello studente.

| n. | Nome variabile | Descrizione | Modalità di risposta | Categoria di riferimento |
|----|-------------------|--|---|--|
| 9 | <i>frequenz</i> | Frequenza alle lezioni | 1= non regolare 2= regolare ad almeno alcuni corsi | Frequenza regolare |
| 10 | <i>esp_lav</i> | Tipo di esperienze lavorative durante gli studi universitari ¹⁹ | 1= stabile 2= non stabile | Non stabile |
| 11 | <i>tiroc</i> | Tirocinio o stage svolto per il completamento degli studi | 1= sì 2= no | No |
| 12 | <i>voto_30</i> | Votazione media riportata agli esami (in trentesimi) | | Variabile centrata rispetto alla media di gruppo |
| 13 | <i>tempo_tesi</i> | Tempo impiegato per la stesura della tesi (in mesi) | | Variabile centrata rispetto alla media di gruppo |
| 14 | <i>militare</i> | Servizio militare o civile svolto durante gli studi | 1= svolto durante gli studi universitari 2= non svolto durante gli studi universitari. | Non svolto durante gli studi universitari |

Nella tabella, per ogni variabile sono stati indicati: un nome convenzionale, una breve descrizione del suo significato, le modalità di risposta ricodificate e la categoria di riferimento (categoria riferita all'individuo-base).

Da quanto indicato nella Tabella 1 emerge il profilo dell'individuo-base che risulta essere:

- **femmina**
- **residente a Firenze**
- **almeno un genitore con laurea**
- **appartenenza alla classe media impiegatizia o alla piccola borghesia**
- **diploma di tipo liceale**
- **voto al diploma medio all'interno del proprio CdL**
- **percorso di studi pre-universitari regolare**
- **iscritto subito all'Università**
- **frequenza regolare alle lezioni universitarie**

¹⁹ La scelta di inserire come variabile di primo livello il tipo di esperienze lavorative avute durante gli studi universitari e non il fatto di aver avuto o meno tali esperienze è dovuta al fatto che tutti i laureati appartenenti alla popolazione analizzata hanno indicato di aver avuto almeno un'esperienza di tipo lavorativo durante gli studi.

- **esperienza di lavoro non stabile**
- **nessuna attività di tirocinio o stage per il completamento degli studi**
- **votazione media agli esami pari alla media nel proprio CdL**
- **tempo medio all'interno del proprio CdL per la stesura della tesi**

Le variabili di secondo livello (Tabella 2) utilizzate sono di tipo contestuale, ovvero variabili che si riferiscono a caratteristiche proprie di ogni corso di laurea, espresse attraverso il valore della media di gruppo delle variabili di primo livello²⁰.

Tabella 2. Variabili esplicative di secondo livello

| n. | Nome variabile | Descrizione | Modalità di risposta | Valore di riferimento |
|----|-----------------------|---|----------------------|---|
| 1 | <i>maschi_CdL</i> | Percentuale di maschi nel CdL | | |
| 2 | <i>noliceali_CdL</i> | Percentuale di studenti con un diploma superiore diverso dal diploma liceale | | |
| 3 | <i>voto_dipl_CdL</i> | Voto medio nel CdL riportato dagli studenti al diploma di scuola superiore (in sessantesimi) | | Variabile centrata rispetto alla media totale |
| 4 | <i>frequenz_CdL</i> | Percentuale di studenti che non frequentano regolarmente tutte o quasi tutte lezioni | | |
| 5 | <i>tiroc_CdL</i> | Percentuale di studenti che hanno svolto attività di tirocinio o stage per il completamento degli studi | | |
| 6 | <i>voto_30_CdL</i> | Votazione media nel CdL riportata agli esami (in trentesimi) | | Variabile centrata rispetto alla media totale |
| 7 | <i>tempo_tesi_CdL</i> | Tempo medio nel CdL impiegato per la stesura della tesi (in mesi) | | Variabile centrata rispetto alla media totale |
| 8 | <i>numchiuso</i> | Presenza nel CdL di limitazioni all'accesso | 1= sì 2= no | No |

²⁰ Nel caso in cui le variabili di primo di livello sono categoriche, le rispettive variabili di secondo livello sono espresse dalla percentuale di studenti in ogni CdL per cui si osserva una modalità di risposta diversa da quella base.

Seguendo la strategia di selezione²¹ sopra descritta si è ottenuto il modello:

$$\begin{aligned} ind_dur_{ij} = & \gamma_{00} + \gamma_{10j} sesso_{ij} + \gamma_{20j} tit_gen_{ij} + \gamma_{30j} diploma_{ij} + \gamma_{40j} voto_dipl_{ij} + \\ & + \gamma_{50j} frequenz_{ij} + \gamma_{60j} esp_lav_{ij} + \gamma_{70j} tiroc_{ij} + \gamma_{80j} voto_30_{ij} + \gamma_{90j} tempo_tesi_{ij} + \\ & + \gamma_{10,0j} militare_{ij} + \gamma_{11,0j} sesso * tempo_tesi_{ij} + \gamma_{12,0j} sesso * voto_30_{ij} + \\ & + \gamma_{13,0j} diploma * tit_gen_{ij} + \gamma_{14,0j} voto_dipl * tempo_tesi_{ij} + (u_{0j} + r_{ij}) \end{aligned}$$

I risultati ottenuti sono riportati nella Tabella 3; i coefficienti riportati in quarta colonna indicano quanto cambia il tempo di laurea di uno studente iscritto ad un corso di laurea di durata quadriennale al variare di un'unità delle variabili esplicative corrispondenti.

L'intercetta $\gamma_{00} = 1.65$ indica il valore dell'indice di durata (che per un corso di laurea di durata quadriennale corrisponde a circa 6 anni e 7 mesi) quando le variabili esplicative di primo livello assumono un valore pari a 0 e gli errori di primo e di secondo livello sono nulli, ovvero quando è osservato il tempo di conseguimento del titolo del cosiddetto individuo-base all'interno di un CdL-base.

Tra le variabili di primo livello che non sono risultate significative (*residenza*, *cl_sociale*, *eta_dipl*, *rit_iscr*) vi è la residenza dello studente. Tale risultato conferma le osservazioni svolte in precedenza: la residenza degli studenti non sembra incidere in maniera così netta sul valore dell'indice di durata anche se, come già segnalato, la variabile da considerare dovrebbe essere il domicilio e non la residenza degli studenti.

Le stime dei coefficienti di regressione che hanno un valore positivo sono quelle riguardanti la relazione tra indice di durata e le variabili *tit_gen*, *diploma*, *frequenz*, *esp_lav*, *tempo_tesi*, *militare*: questo significa che "allontanandosi" dal profilo base si ha un incremento dell'indice di durata e quindi un aumento dei tempi di laurea.

La variabile *diploma* è significativa se presa nel suo insieme ma, dall'analisi dei risultati riportati in Tabella 3, il passaggio da una situazione in cui lo studente ha una preparazione preuniversitaria di tipo liceale ad una situazione in cui il tipo di scuola superiore frequentata ricade nella categoria "altro" (né liceale, né tecnica) non sembra avere influenza sui tempi di laurea degli studenti, mentre significativo è avere una preparazione tecnica rispetto a quella liceale²².

²¹ Si è provveduto a migliorare di volta in volta il modello inserendovi differenti variabili esplicative e le loro interazioni e togliendo quelle covariate non risultate significative attraverso l'analisi del *test di Wald* al livello di significatività del 5%. Per un confronto tra modelli ottenuti attraverso l'inserimento di parametri aggiuntivi ci si è basati sul *test della devianza* al livello di significatività del 5%; nonostante il software proceda in automatico al calcolo della statistica della *devianza*, il test ad esso relativo è stato calcolato manualmente. Inoltre, per confrontare modelli con differenti parametri, è stato utilizzato l'indice di adattamento *AIC* (Akaike's Information Criterion).

²² Si segnala che sono state utilizzate anche altre classificazioni interne della variabile *diploma*, ma nessuna di queste ha apportato dei miglioramenti complessivi al modello.

Tabella 3. Effetti fissi: coefficienti di regressione

| Effetto | Categorie | Stima | STIMA (espressa in mesi) | Errore standard | Valore t | Pr > t |
|----------------------|-----------|--------|--------------------------------|--------------------|----------|---------|
| Intercetta | | 1.647 | 79.08 | 0.0567 | 29.05 | <.0001 |
| sex | 1 | -0.102 | -4.891 | 0.0233 | -4.38 | 0.0001 |
| sex | 2 | 0 | 0 | . | . | . |
| tit_gen | 1 | 0.188 | 9.010 | 0.0294 | 6.39 | <.0001 |
| tit_gen | 2 | 0.109 | 5.251 | 0.0281 | 3.9 | 0.0002 |
| tit_gen | 3 | 0 | 0 | . | . | . |
| diploma | 1 | 0.0243 | 1.164 | 0.0577 | 0.42 | 0.6759 |
| diploma | 2 | 0.180 | 8.664 | 0.0571 | 3.16 | 0.0027 |
| diploma | 3 | 0 | 0 | . | . | . |
| voto_dipl | | -0.011 | -0.513 | 0.0014 | -7.65 | <.0001 |
| frequenz | 1 | 0.083 | 4.00 | 0.0352 | 2.37 | 0.0319 |
| frequenz | 2 | 0 | 0 | . | . | . |
| esp_lav | 1 | 0.132 | 6.350 | 0.0224 | 5.91 | <.0001 |
| esp_lav | 2 | 0 | 0 | . | . | . |
| tiroc | 1 | -0.140 | -6.730 | 0.0323 | -4.34 | 0.0002 |
| tiroc | 2 | 0 | 0 | . | . | . |
| voto_30 | | -0.011 | -0.544 | 0.0095 | -1.19 | 0.2338 |
| tempo_tesi | | 0.013 | 0.647 | 0.0022 | 5.98 | <.0001 |
| militare | 1 | 0.228 | 10.944 | 0.0265 | 8.6 | <.0001 |
| militare | 2 | 0 | 0 | . | . | . |
| sex*tempo_tesi | 1 | 0.009 | 0.419 | 0.0033 | 2.65 | 0.0082 |
| sex*tempo_tesi | 2 | 0 | 0 | . | . | . |
| sex*voto_30 | 1 | -0.032 | -1.561 | 0.0128 | -2.53 | 0.0114 |
| sex*voto_30 | 2 | 0 | 0 | . | . | . |
| diploma*tit_gen | 1*1 | 0.056 | 2.683 | 0.0662 | 0.84 | 0.4026 |
| diploma*tit_gen | 1*2 | 0.057 | 2.720 | 0.0705 | 0.8 | 0.4257 |
| diploma*tit_gen | 1*3 | 0 | 0 | . | . | . |
| diploma*tit_gen | 2*1 | -0.181 | -8.688 | 0.0638 | -2.84 | 0.0068 |
| diploma*tit_gen | 2*2 | -0.190 | -9.106 | 0.0666 | -2.85 | 0.0065 |
| diploma*tit_gen | 2*3 | 0 | 0 | . | . | . |
| diploma*tit_gen | 3*1 | 0 | 0 | . | . | . |
| diploma*tit_gen | 3*2 | 0 | 0 | . | . | . |
| diploma*tit_gen | 3*3 | 0 | 0 | . | . | . |
| voto_dipl*tempo_tesi | | 0.001 | 0.030 | 0.0003 | 2.37 | 0.0177 |

Molto interessante, inoltre, è osservare che, assunto come casuale l'effetto del corso di laurea, essere maschio ha un effetto positivo sui tempi di laurea. Qualora le altre variabili esplicative abbiano valore pari a 0, si avrà:

$$ind_dur_j(maschio) = 1,65 - 0,10(maschio = 1)_{ij} + (u_{0j} + r_{ij})$$

e

$$ind_dur_j(femm) = 1,65 - 0,10(femm = 0)_{ij} + (u_{0j} + r_{ij}),$$

ossia un valore dell'indice di durata di circa 1.55 (che per un corso di laurea di durata quadriennale corrisponde a circa 6 anni e 2 mesi) per i maschi e 1.65 per le femmine (corrispondente a circa 6 anni e 7 mesi).

Una volta inserite nel modello le variabili esplicative di primo livello e le loro interazioni, sono state aggiunte anche le variabili esplicative di secondo livello. Il modello di regressione specificato separatamente per i singoli gruppi risulta, pertanto, espresso dall'equazione:

$$Y_{ij} = \left(\beta_{0j} + \sum_p \beta_{pj} X_{p ij} \right) + r_{ij}$$

dove:

$$\beta_{0j} = \gamma_{00} + \sum_q \gamma_{0q} Z_{qj} + u_{0j}$$

Seguendo la strategia di selezione del modello sopra descritta, per il coefficiente β_{0j} si è ottenuta l'equazione:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} voto_dipl_CdL + \gamma_{02} numchiuso + u_{0j}$$

Quindi, il modello completo assume la forma:

$$\begin{aligned} ind_dur_{ij} = & \gamma_{00} + \gamma_{10j} sesso_{ij} + \gamma_{20j} tit_gen_{ij} + \gamma_{30j} diploma_{ij} + \gamma_{40j} voto_dipl_{ij} + \\ & + \gamma_{50j} frequenz_{ij} + \gamma_{60j} esp_lav_{ij} + \gamma_{70j} tiroc_{ij} + \gamma_{80j} voto_30_{ij} + \gamma_{90j} tempo_tesi_{ij} + \\ & + \gamma_{10,0j} militare_{ij} + \gamma_{11,0j} sesso * tempo_tesi_{ij} + \gamma_{12,0j} sesso * voto_30_{ij} + \\ & + \gamma_{13,0j} diploma * tit_gen_{ij} + \gamma_{14,0j} voto_dipl * tempo_tesi_{ij} + \\ & + \gamma_{01} voto_dipl_CdL + \gamma_{02} numchiuso + (u_{0j} + r_{ij}) \end{aligned}$$

Nella Tabella 4 sono riportati i risultati ottenuti.

L'intercetta $\gamma_{00} = 1.69$ indica il valore dell'indice di durata (che per un corso di laurea di durata quadriennale corrisponde a circa 6 anni e 9 mesi) quando tutte le variabili esplicative hanno un valore pari a 0 e gli errori di primo e di secondo livello sono nulli, ovvero quando si osserva il tempo di conseguimento del titolo del cosiddetto individuo-base all'interno di un CdL-base.

Tabella 4. Effetti fissi: coefficienti di regressione

| Effetto | Categorie | Stima | STIMA (espressa in mesi) | Errore standard | Valore T | Pr > t |
|------------------------------|-----------|--------|--------------------------------|--------------------|----------|---------|
| Intercetta | | 1.689 | 81.058 | 0.0500 | 33.78 | <.0001 |
| sesso | 1 | -0.096 | -4.612 | 0.0229 | -4.19 | 0.0002 |
| sesso | 2 | 0 | 0 | . | . | . |
| tit_gen | 1 | 0.168 | 8.050 | 0.0291 | 5.77 | <.0001 |
| tit_gen | 2 | 0.089 | 4.291 | 0.0278 | 3.21 | 0.002 |
| tit_gen | 3 | 0 | 0 | . | . | . |
| diploma | 1 | -0.002 | -0.109 | 0.0570 | -0.04 | 0.9685 |
| diploma | 2 | 0.170 | 8.150 | 0.0563 | 3.01 | 0.004 |
| diploma | 3 | 0 | 0 | . | . | . |
| voto_dipl | | -0.011 | -0.513 | 0.0014 | -7.75 | <.0001 |
| frequenz | 1 | 0.085 | 4.068 | 0.0347 | 2.44 | 0.0275 |
| frequenz | 2 | 0 | 0 | . | . | . |
| esp_lav | 1 | 0.127 | 6.086 | 0.0221 | 5.75 | <.0001 |
| esp_lav | 2 | 0 | 0 | . | . | . |
| tiroc | 1 | -0.149 | -7.162 | 0.0317 | -4.7 | <.0001 |
| tiroc | 2 | 0 | 0 | . | . | . |
| voto_30 | | -0.010 | -0.479 | 0.0094 | -1.06 | 0.288 |
| tempo_tesi | | 0.013 | 0.625 | 0.0022 | 5.85 | <.0001 |
| militare | 1 | 0.209 | 10.042 | 0.0262 | 7.97 | <.0001 |
| militare | 2 | 0 | 0 | . | . | . |
| sesso*tempo_tesi | 1 | -0.035 | -1.674 | 0.0127 | -2.76 | 0.0059 |
| sesso*tempo_tesi | 2 | 0 | 0 | . | . | . |
| sesso*voto_30 | 1 | 0.009 | 0.442 | 0.0032 | 2.83 | 0.0047 |
| sesso*voto_30 | 2 | 0 | 0 | . | . | . |
| diploma*tit_gen | 1*1 | 0.070 | 3.342 | 0.0652 | 1.07 | 0.2915 |
| diploma*tit_gen | 1*2 | 0.073 | 3.527 | 0.0695 | 1.06 | 0.2961 |
| diploma*tit_gen | 1*3 | 0 | 0 | . | . | . |
| diploma*tit_gen | 2*1 | -0.174 | -8.338 | 0.0629 | -2.76 | 0.0083 |
| diploma*tit_gen | 2*2 | -0.176 | -8.462 | 0.0657 | -2.68 | 0.0101 |
| diploma*tit_gen | 2*3 | 0 | 0 | . | . | . |
| diploma*tit_gen | 3*1 | 0 | 0 | . | . | . |
| diploma*tit_gen | 3*2 | 0 | 0 | . | . | . |
| diploma*tit_gen | 3*3 | 0 | 0 | . | . | . |
| voto_dipl*tempo_tesi | | 0.001 | 0.029 | 0.0003 | 2.29 | 0.0221 |
| voto_dipl_CdL | | -0.050 | -2.418 | 0.0155 | -3.26 | 0.0025 |
| numchiuso | 1 | -0.378 | -18.154 | 0.0521 | -7.26 | . |
| numchiuso | 2 | 0 | 0 | . | . | . |

I coefficienti di regressione delle covariate *voto_dipl_CdL* e *numchiuso* esprimono l'effetto di queste due variabili di secondo livello sulla media tra gruppi dell'indice di durata. Il fatto che il coefficiente γ_{01} sia negativo indica che all'aumentare del voto medio che gli studenti hanno ottenuto al diploma migliorano i tempi di laurea medi all'interno del gruppo.

Ancora più interessante, inoltre, è analizzare il coefficiente di regressione della variabile *numchiuso*. Quando $u_{0j} = 0$, si ha:

$$\beta_{0j} = 1.69 - 0.37(\text{numchiuso})$$

quindi:

$$\beta_{0j}(\text{numchiuso} = 1) = 1.32$$

$$\beta_{0j}(\text{numchiuso} = 0) = 1.69$$

Questo vuol dire che il valore medio dell'indice di durata risulta inferiore di 0.37 (corrispondente a circa 1 anno e quasi 6 mesi per un corso di laurea di durata quadriennale) per quei CdL in cui vi è il cosiddetto numero chiuso rispetto a quelli in cui non esiste nessuna limitazione all'accesso per le immatricolazioni.

Le variabili di secondo livello che singolarmente sono risultate significative, ma che successivamente, attraverso l'applicazione del test della devianza e il calcolo dell'indice di adattamento AIC sono state escluse dal modello, sono state: *noticeali_CdL*, *frequenz_CdL*, *tempo_tesi_CdL*. Questo indica che vi è una relazione tra il tipo di preparazione preuniversitaria degli studenti che si iscrivono in un determinato CdL, la loro frequenza media e il tempo medio richiesto per la stesura della tesi e i tempi medi di laurea osservati all'interno dello specifico CdL, ma che le variabili *voto_dipl_CdL* e *numchiuso* spiegano una maggior variabilità del fenomeno risposta.

Passando all'analisi dei coefficienti di regressione delle variabili di primo livello e delle loro interazioni si rileva come questi sostanzialmente non siano cambiati rispetto al modello precedentemente presentato. Inoltre, ancora una volta, le stime dei coefficienti di regressione che hanno un valore positivo sono quelle riguardanti la relazione tra indice di durata e le variabili *tit_gen*, *diploma*, *frequenz*, *esp_lav*, *tempo_tesi*, *militare*.

Infine, è stata svolta un'analisi dei residui sia di primo che di secondo livello per verificare alcune ipotesi poste alla base del modello.

In particolare, è risultato che il valore della varianza di entrambe le componenti residue sono inferiori rispetto a quelli stimati attraverso il modello nullo: parte della variabilità della variabile risposta dovuta sia all'effetto individuale che a quello di gruppo è stata spiegata attraverso l'inserimento delle variabili esplicative. Inoltre, è stato calcolato il coefficiente residuo di correlazione intra-classe; comparando tale valore con quello calcolato attraverso il modello vuoto, si osserva come attraverso l'inserimento delle variabili esplicative vi sia stata una diminuzione della percentuale della varianza totale dell'indice di durata dovuta all'effetto del Corso di Laurea.

L'analisi dei residui di secondo livello relativi al modello nullo ed al modello finale ha permesso, tra le altre cose, di ottenere interessanti informazioni per quanto riguarda il cosiddetto "effetto corso di laurea" sui tempi di conseguimento del titolo degli studenti; infatti, è stato possibile costruire una sorta di "graduatoria" dei corsi di laurea in termini di efficacia dovuta esclusivamente alle loro peculiarità. In particolare, è stato rilevato che parte della variabilità dei tempi di laurea degli studenti che attraverso una prima analisi di natura descrittiva sembrava dovuta all'effetto proprio dei diversi corsi di laurea è in realtà spiegabile altrimenti.

Solo per far un esempio, presupponendo come casuale l'effetto proprio del gruppo, i corsi di laurea in Medicina e Chirurgia ed in Odontoiatria e protesi dentaria sono risultati molto efficienti in termini di tempi di conseguimento del titolo degli studenti mentre dopo l'inserimento nel modello delle variabili esplicative tale effetto positivo è in qualche modo "scomparso". Questo può essere spiegato dal fatto che i due CdL considerati godono degli effetti positivi nei confronti dei tempi di laurea esercitati dalla presenza delle limitazioni all'accesso delle immatricolazioni, effetti che sono stati controllati inserendo nel modello la variabile esplicativa relativa alla presenza del numero chiuso all'interno del corso di studi. Inoltre, si rileva che sia presupponendo la totale casualità dell'effetto del corso di laurea sia cercando di spiegarlo attraverso l'utilizzo delle variabili esplicative, gli effetti "estremamente" positivi e quelli "estremamente" negativi dei CdL rimangono gli stessi: solo per fare un esempio, i CdL ad avere un effetto migliore sono Psicologia, Scienze Forestali e Scienze dell'Educazione, mentre quelli ad avere effetti peggiori sono Materie Letterarie e Lingue e Letterature Straniere Moderne.

5. Conclusioni

In questa nota sono stati descritti molto sinteticamente i risultati di un'analisi dei tempi di conseguimento del titolo dei laureati dell'Ateneo fiorentino soffermando l'attenzione sui laureati dell'anno solare 2000; su tali dati si è anche proceduto alla stima di un modello multilivello. Obiettivo dell'applicazione è stato quello di esaminare i comportamenti individuali degli studenti universitari per quanto riguarda i tempi di laurea in funzione di variabili esplicative di primo e secondo livello, dove le unità di primo livello sono gli studenti e quelle di secondo livello sono i Corsi di laurea. I risultati dell'analisi possono essere utili ai fini di una miglior comprensione di un fenomeno considerato unanimemente una criticità assoluta del sistema universitario italiano.

Il modello a cui si è giunti è un modello lineare gerarchico ad intercetta casuale, in cui si suppone un effetto costante tra gruppi delle variabili esplicative sulla variabile risposta (Y_{ij}).

Le covariate di primo livello che sono risultate significative nello spiegare i tempi di laurea degli studenti sono state sia variabili legate ai loro caratteri strutturali (genere, titolo di studio dei genitori), sia variabili legate alla loro preparazione preuniversitaria (il tipo di scuola superiore frequentata, la votazione riportata all'esame di maturità), sia variabili legate alla loro carriera universitaria (frequenza alle lezioni, il tipo di esperienze lavorative avute durante gli studi, lo svolgimento o meno di un tirocinio, la votazione riportata agli esami, il tempo impiegato per la stesura della tesi, lo svolgimento o meno del servizio militare o civile durante gli studi); inoltre, è stato possibile rilevare come il fenomeno "tempi di laurea" è spiegato anche da alcune interazioni tra le variabili di primo livello. A livello di corso di laurea, i tempi impiegati dagli studenti per il conseguimento del titolo dipendono sia dal voto medio riportato dagli stessi all'esame di maturità, sia dalla presenza o meno del cosiddetto "numero chiuso" all'interno del CdL.

Naturalmente, le politiche universitarie d'intervento che dovrebbero essere messe in atto ai fini della risoluzione del problema dell'eccessiva durata degli studi potranno riguardare solo variabili legate alla vita universitaria degli studenti che in qualche modo "agiscono" sui tempi di laurea degli stessi. Dall'analisi di tali variabili è possibile rilevare come queste si trovano in relazione con l'organizzazione interna dei corsi di studi: il tipo di frequenza richiesta, il fatto di lasciar tempo o meno per diversi tipi di esperienze lavorative (stabili o non stabili), il tempo richiesto per la stesura tesi, l'obbligatorietà o meno di svolgere attività di tirocinio o stage, possono essere considerati indici di una "buona o cattiva" organizzazione interna della corso di studi.

Anche il fatto che dall'applicazione del modello sia risultato significativo il cosiddetto "numero chiuso" nello spiegare le differenze tra CdL rileva come una miglior organizzazione del corso incida sui tempi di laurea degli studenti. Infatti, generalmente, la limitazione all'accesso delle immatricolazioni, convogliando all'interno dei Corsi di laurea solo un circoscritto numero di studenti solitamente molto motivati, ha degli effetti positivi sull'organizzazione della didattica, sulla gestione dei servizi agli studenti, sul numero dei docenti per studente, ecc. Anche l'inserimento nei curricula di attività di tirocinio o stage può avere degli effetti positivi sui tempi di laurea degli studenti agendo positivamente sull'organizzazione interna dei piani di studi dei CdL.

I risultati delle analisi svolte, molto sommariamente richiamati in questa nota, giustificano ampiamente, a nostro parere, il ricorso ai modelli multilivello quando si procede all'analisi di dati che riguardano gli studenti universitari²³; infatti, è del tutto

²³ In tale direzione si sta muovendo da tempo il gruppo *VALMON* (*Valutazione e Monitoraggio*). Il gruppo, coordinato da B. Chiandotto e costituito da laureandi, dottorandi e docenti del Dipartimento di Statistica dell'Università degli Studi di Firenze, da diversi anni svolge attività di studio e ricerca nel contesto della valutazione e del monitoraggio dei processi formativi che si svolgono nell'Ateneo fio-

evidente la natura gerarchica dei dati: le unità di primo livello sono gli studenti o i laureati/diplomati, mentre le unità di secondo livello sono i corsi di studio. Ovviamente la gerarchizzazione può essere estesa ad un numero di livelli più elevato: ad esempio le Facoltà possono rappresentare il terzo livello e gli Atenei il quarto livello.

Riferimenti bibliografici

- BULGARELLI G. (2002) *Esito degli studi degli immatricolati dell'Ateneo Fiorentino dal 1980/81 al 1997/9*, Università degli Studi di Firenze, consultabile anche sul sito www.unifi.it/aut_dida/indexval.html.
- CHIANDOTTO B. (2002) *Valutazione dei processi formativi: cosa, come e perché, in Valutazione della Didattica e dei Servizi nel Sistema Università*. In D'ESPOSITO M.R. (a cura di) *Valutazione della Didattica e dei Servizi nel Sistema Università*. CUSL, Salerno 2002.
- CHIANDOTTO B., BACCI S., BERTACCINI B. (2004) *I laureati e diplomati dell'Ateneo Fiorentino dell'anno 2000: profilo e sbocchi professionali*, Università degli Studi di Firenze.
- CHIANDOTTO B., BERTACCINI B. (2003) *I laureati e diplomati dell'Ateneo Fiorentino dell'anno 1999: profilo e sbocchi professionali*, Università degli Studi di Firenze.
- COBALTI A., SCHIZZEROTTO A. (1994) *La mobilità sociale in Italia*, Il Mulino, Bologna.
- GOLDSTEIN H. (2003) *Multilevel Statistical Models*, Edward Arnold, London.
- HOX J.J. (2002) *Multilevel Analysis: Techniques and Applications*, LAWRENCE ERLBAUM ASSOCIATES, Mahwah (New Jersey), London.
- SAS INSTITUTE INC. (1999) *SAS/STAT® User's Guide, Version 8*, SAS Institute Inc., Cary NC.
- SNIJDERS T., BOSKER R. (1999) *An Introduction to Basic and Advanced Multilevel Modeling*, Sage, London.
- VARRIALE R. (2004) *Tempi di conseguimento del titolo nell'Università degli Studi di Firenze nel periodo 1980-2000 e applicazione di un modello lineare gerarchico ai laureati nell'anno solare 2000*, Tesi di laurea, Università degli Studi di Firenze.

rentino. Tale interesse è testimoniato, tra l'altro, da altri due lavori presentati in questa sede: "Un modello multilivello per l'analisi della condizione occupazionale dei laureati" (Chiandotto B. e Bacci S.); "L'abbandono degli studi universitari" (Chiandotto B. e Giusti C.).

A multilevel model for the analysis of university students' time to degree

Summary. *This paper deals with the analysis of the time that students need to graduate, since this topic represents one of the most critical aspects of the Italian University System. The analysis is conducted using data concerning the students enrolled at the University of Florence in the academic years from 1980 to 2000, focusing on the data concerning the graduates during the calendar year 2000. The aim of the work is to find out the predictors that explain the time that students need to graduate. The initial analysis is conducted by descriptive statistical methods, then the work is accomplished by applying a hierarchical linear model on students' graduation time. The use of a multilevel analysis was suggested by the two-level data structure: the first level units are identified by the students; the second level units are the degree courses in which these students graduated.*

The level-1 predictors that result in being significant to explain the time that students need to graduate are variables regarding students' structural characteristics, their educational background and their university career. At level-2 (degree courses), the time that students need to graduate depends on the average high school final grade and the presence of limits of the maximum number of students allowed in that specific degree course.

Keywords: *Time needed to graduate, Multilevel models, Hierarchical linear regression.*