

Un modello multilivello per l'analisi della condizione occupazionale dei laureati ¹

Bruno Chiandotto, Silvia Bacci

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

Riassunto. Oggetto del contributo è l'analisi della condizione occupazionale dei laureati dell'Ateneo fiorentino nell'anno solare 2000 a un anno e mezzo/due anni e mezzo dal conseguimento del titolo. Lo scopo perseguito è relativo all'individuazione delle possibili determinanti del fenomeno indagato, sia a livello individuale (quali genere, residenza, diploma di scuola superiore, voto di laurea, ...) che a livello istituzionale (in termini di caratteristiche distintive dei diversi corsi di studio). L'impiego di analisi descrittive è propedeutico per l'applicazione di un *modello di regressione logistica con intercetta casuale a due livelli di aggregazione*, il cui utilizzo è giustificato dalla necessità di tenere in debito conto la struttura gerarchica dei dati oggetto di studio: infatti, come unità di primo livello sono stati considerati i laureati dell'Ateneo fiorentino, mentre come unità di secondo livello i rispettivi corsi di laurea.

Parole chiave: Modelli multilivello, Regressione logistica, Sbocchi occupazionali, Probabilità di occupazione.

1. Introduzione

L'analisi dei profili e degli sbocchi occupazionali dei laureati e dei diplomati, che può essere condotta da tutte le Università che aderiscono al Consorzio *ALMA-*

¹ Il presente lavoro è stato finanziato nell'ambito del PRIN 2002, cofinanziato dal MIUR "Transizioni Università-lavoro e valorizzazione delle competenze professionali dei laureati: modelli e metodi di analisi multidimensionali delle determinanti". 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 S. Bacci.

*LAUREA*², consente di pervenire ad una misura di efficienza (interna) e di efficacia (interna ed esterna) della qualità dei servizi formativi offerti dagli Atenei.

Com'è noto, le rilevazioni effettuate direttamente da AlmaLaurea coinvolgono i laureati della sola sessione estiva ad uno, due e tre anni dal conseguimento del titolo (dall'anno 2003, sempre limitatamente alla sola sessione estiva, vengono intervistati i laureati/diplomati ad uno, tre e cinque anni dal conseguimento del titolo). L'Ateneo Fiorentino, a partire dal 1998, ha deciso di estendere l'indagine all'intero collettivo dei laureati dell'anno solare³ al fine di pervenire ad una migliore comprensione di quella che è la qualità del prodotto finito (i laureati e diplomati) della propria attività formativa, anche attraverso l'acquisizione pressoché completa dei dati a livello di singolo corso di laurea; l'indagine è stata ripetuta per tutti i laureati degli anni 1999, 2000, 2001 e 2002⁴.

I risultati dell'analisi dei dati relativi agli studenti che hanno conseguito un titolo di studio presso l'Università degli Studi di Firenze durante l'anno solare 2000 sono contenuti in un volume di recente pubblicazione (Chiandotto B., Bacci S. e Bertaccini B., 2004).

² 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.

³ La rilevazione effettuata dall'Università di Firenze si caratterizza, rispetto ad AlmaLaurea, per gli aspetti di seguito riportati:

- l'inserimento nella scheda di rilevazione di due ulteriori quesiti sulle motivazioni dell'iscrizione all'università;
- nei quesiti a risposta multipla (per rendere possibili adeguate elaborazioni statistiche) viene richiesta la manifestazione delle priorità;
- al quesito generico sull'utilizzo delle competenze acquisite durante la formazione universitaria della scheda AlmaLaurea è stato aggiunto un quesito molto articolato sulle diverse tipologie di competenze acquisite, sul loro utilizzo e sulle modalità di acquisizione delle stesse (insegnamenti universitari, attività di tirocinio, corsi di specializzazione, corsi di formazione professionale, acquisizioni sul posto di lavoro, ...).

⁴ Chi fosse interessato ad un approfondimento conoscitivo sui risultati delle analisi svolte può consultare: Bulgarelli G. (2002), Bertaccini B. (2000), Chiandotto B. (2002), Chiandotto B. e Bertaccini B. (2003).

In questa nota vengono presentati sinteticamente⁵ alcuni dei risultati conseguiti soffermando l'attenzione sulle conclusioni cui si è pervenuti a seguito di una applicazione di un modello logistico a due livelli (dove i laureati sono le unità di primo livello ed i corsi di studio le unità di secondo livello) nel quale la variabile dipendente di riferimento è la condizione occupazionale dei laureati a un anno e mezzo/due anni e mezzo dal conseguimento del titolo e l'obiettivo che si vuol conseguire è quello dell'individuazione e della misura dell'*effetto netto* delle sue possibili determinanti (fattori esplicativi).

Il secondo paragrafo di questa nota è dedicato alla illustrazione dei risultati dell'analisi, sostanzialmente di natura descrittiva, finalizzata alla individuazione dell'eventuale influenza esercitata sulla condizione occupazionale dei laureati/diplomati presso l'Università di Firenze nell'anno solare 2000 da caratteristiche individuali, quali genere, residenza, diploma di scuola superiore, ecc.

Per pervenire alla misura dell'effetto "netto" eventualmente esercitato da possibili determinanti (sia individuali che istituzionali) della condizione occupazionale si è fatto ricorso ai modelli gerarchici o di regressione multilivello che, com'è noto, hanno la principale caratteristica di tenere in considerazione la struttura gerarchica dei dati oggetto di studio. I risultati delle analisi condotte sono riportati nei paragrafi successivi, mentre alcune considerazioni conclusive completano la nota.

2. Condizione occupazionale dei laureati nell'Ateneo fiorentino nell'anno 2000

Nell'anno solare 2000, hanno concluso gli studi con successo nell'Università degli Studi di Firenze **4846** laureati e **399** diplomati, per un totale di **5245** studenti che costituiscono la popolazione di riferimento delle analisi condotte⁶. Per un quadro d'insieme sulla condizione occupazionale del contingente analizzato, disaggregata per facoltà e per corsi di laurea, si osservino i dati riportati nella Tavola 1.

⁵ La trattazione completa è riportata nel suddetto volume di recente pubblicazione: "I laureati e diplomati dell'Ateneo fiorentino dell'anno 2000: Profilo e sbocchi occupazionali (Chiandotto B., Bacci S. e Bertaccini B., 2004).

⁶ Le indagini sugli Sbocchi Occupazionali sono state condotte nel periodo giugno-luglio 2002 tramite interviste telefoniche ricorrendo alle tecniche *C.A.T.I.* (*Computer Aided Telephone Interviewing*). Il disegno d'indagine ha anche previsto l'invio postale di un questionario opportunamente adattato ai laureati non contattati durante la fase telefonica della rilevazione, fornendo la possibilità di un'eventuale compilazione tramite e-mail. Il contingente dei laureati/diplomati è stato, pertanto, contattato in un arco temporale che varia dai 15 mesi (per i laureati in luglio) ai 30 mesi (per i laureati in gennaio). A livello complessivo si registra un tasso di risposta pari all'82,6% (4330 interviste su un collettivo di 5245 individui); se si escludono le interviste postali, i laureati e diplomati contattati telefonicamente sono stati circa il 76,9% del totale (3736 laureati e 296 diplomati).

Al momento dell'intervista il 76,5% degli intervistati (2889 laureati e 307 diplomati)⁷ si dichiara occupato. Tra i non occupati, soltanto il 9,5% manifesta l'intenzione di cercare lavoro, mentre il 14,0% dichiara di non lavorare e non cercare lavoro. Pertanto, il tasso occupazionale netto, calcolato escludendo quest'ultimo contingente - di fatto costituito da giovani inattivi sul mercato del lavoro e quindi da non considerarsi come disoccupati - sale all'89,0%.

Se si escludono dall'analisi le facoltà di Giurisprudenza e Medicina e Chirurgia, nelle quali, notoriamente, la partecipazione ad attività formative non è generalmente dovuta alla libera scelta del laureato ma è spesso originata da obblighi istituzionali e statuari, i tassi di occupazione generale e netto (cioè calcolato escludendo gli inattivi) salgono rispettivamente all'83,0% e al 90,1% .

Val la pena segnalare l'altissimo livello occupazionale registrato per i diplomati intervistati (94,5%), soprattutto alla luce del fatto che oltre l'83,0% di questo contingente (pari al 87,9% degli occupati) svolge un lavoro iniziato dopo il conseguimento del titolo, mentre i laureati occupati che non proseguono il lavoro iniziato prima del completamento degli studi sono pari soltanto al 60,3% del totale laureati (ovvero l'80,4% degli occupati).

Tavola 1. *Laureati occupati e non occupati al momento dell'intervista per Facoltà e Corso di studi*

	Non lavora	%	Lavora	%	Laureati
AGRARIA	16	18,0	73	82,0	89
Scienze agrarie	2	12,5	14	87,5	16
Scienze forestali	7	20,6	27	79,4	34
Scienze forestali ed ambientali	2	10,0	18	90,0	20
Scienze agrarie tropicali e subtropicali	3	33,3	6	66,7	9
Scienze e tecnologie agrarie	2	20,0	8	80,0	10
ARCHITETTURA	114	17,0	558	83,0	672
Architettura	114	17,0	558	83,0	672
ECONOMIA	84	14,4	498	85,6	582
Economia aziendale	1	5,3	18	94,7	19
Economia e commercio	82	15,5	448	84,5	530
Scienze statistiche ed attuariali	-	-	24	100,0	24
Scienze statistiche ed economiche	1	11,1	8	88,9	9
FARMACIA	6	8,5	65	91,5	71
Chimica e tecnologia farmaceutica	4	12,5	28	87,5	32
Farmacia	2	5,1	37	94,9	39
GIURISPRUDENZA	272	58,6	192	41,4	464
Giurisprudenza	272	58,6	192	41,4	464

⁷ Si segnala che in questo paragrafo si richiamano alcune conclusioni relative all'analisi condotta considerando sia i laureati che i diplomati, mentre per la stima del modello sono stati presi in considerazione soltanto i dati relativi ai laureati.

Tavola 1 (segue)

	Non lavora	%	Lavora	%	Laureati
INGEGNERIA	32	9,3	312	90,7	344
Ingegneria civile	4	4,6	83	95,4	87
Ingegneria elettronica	9	8,9	92	91,1	101
Ingegneria meccanica	5	6,8	69	93,2	74
Ingegneria informatica	6	26,1	17	73,9	23
Ingegneria per l'ambiente e il territorio	5	11,6	38	88,4	43
Ingegneria delle telecomunicazioni	3	18,8	13	81,3	16
LETTERE E FILOSOFIA	139	28,6	347	71,4	486
Filosofia	22	38,6	35	61,4	57
Lettere	77	29,4	185	70,6	262
Lingue e letterature straniere	19	21,6	69	78,4	88
Lingue e letterature straniere moderne	12	23,1	40	76,9	52
Storia	9	33,3	18	66,7	27
MEDICINA E CHIRURGIA	70	48,3	75	51,7	145
Medicina e chirurgia	68	59,6	46	40,4	114
Odontoiatria e protesi dentaria	2	6,5	29	93,5	31
SCIENZE della FORMAZIONE	60	16,2	310	83,8	370
Lingue e letterature straniere	2	11,1	16	88,9	18
Lingue e letterature straniere (europee)	3	18,8	13	81,3	16
Materie letterarie	5	17,9	23	82,1	28
Pedagogia	5	11,1	40	88,9	45
Psicologia	19	29,7	45	70,3	64
Scienze della educazione	26	13,1	173	86,9	199
SCIENZE POLITICHE	73	23,6	236	76,4	309
Scienze politiche	73	23,6	236	76,4	309
SMFN	101	31,2	223	68,8	324
Chimica	14	26,4	39	73,6	53
Fisica	7	30,4	16	69,6	23
Matematica	10	27,8	26	72,2	36
Scienze biologiche	42	36,2	74	63,8	116
Scienze della informazione	-	-	12	100,0	12
Scienze geologiche	14	30,4	32	69,6	46
Scienze naturali	14	36,8	24	63,2	38
TOTALE	967	25,1	2889	74,9	3856

Gli elevati valori (cfr. Tavole 2 e 3) assunti dalla V di Cramer (0,37 per il contingente comprendente tutti i laureati e 0,40 per quello relativo ai soli laureati che non lavoravano al momento della laurea) evidenziano una forte associazione tra corso di laurea frequentato e probabilità che ha il laureato stesso di essere occupato a un anno e mezzo / due anni e mezzo dal conseguimento del titolo.

Oltre la tipologia del titolo di studio conseguito, altri caratteri possono, ovviamente, influenzare la probabilità di occupazione; ad esempio, l'analisi del quadro occupazionale, disaggregato per *genere*, evidenzia differenze rilevanti nei due sessi, soprattutto per quanto riguarda alcune facoltà.

Tavola 2. Misure di associazione relative ai laureati occupati e non occupati al momento dell'intervista

Statistiche d'associazione per	Statistica	Valore	GdL	Prob
Percentuale di laureati che lavorano	Chi-quadro	533,9860	38	< 0,0001
vs				
Corso di laurea in cui è stato conseguito il titolo	V di Cramer	0,3721		

Tavola 3. Misure di associazione relative ai laureati occupati e non occupati al momento dell'intervista che non lavoravano alla laurea

Statistiche d'associazione per	Statistica	Valore	GdL	Prob
Percentuale di laureati che lavorano	Chi-quadro	427,2321	38	< 0,0001
vs				
Corso di laurea in cui è stato conseguito il titolo	V di Cramer	0,4003		

Un'altra possibile determinante della condizione occupazionale al momento dell'intervista è il **tempo di conseguimento del titolo**. Le analisi svolte evidenziano il trend crescente del tasso occupazionale all'aumentare del tempo di conseguimento, da un 72,7% dei laureati e diplomati in tempi brevi per raggiungere un 80,0% di coloro che si laureano in tempi medio lunghi. Questo risultato sembra contraddire la convinzione diffusa che tempi rapidi di acquisizione del titolo conducano ad un rapido inserimento nel mondo del lavoro ed è facilmente spiegabile se si guarda ai tassi di partecipazione ad attività formative post-laurea, che risultano superiori proprio per i laureati in tempi brevi.

La stessa analisi, condotta per classi di **voto al conseguimento del titolo**, mostra una proporzionalità diretta, per quanto riguarda i corsi di laurea, tra la percentuale di occupati ed il giudizio finale conseguito, se si eccettua il calo evidenziato dai laureati che hanno riportato un voto pari a 110/110 e lode: infatti, soltanto il 71,5% dei laureati con il massimo dei voti risulta occupato. Tale valore apparentemente anomalo, peraltro non rilevato per i diplomati, trova però giustificazione nel tasso di partecipazione ad almeno un'attività formativa in corso, che per questa categoria di soggetti risulta il più elevato. È quindi verosimile ritenere che i laureati in tempi brevi e con votazioni elevate siano maggiormente propensi alla prosecuzione degli studi.

Relativamente alla possibile influenza del **titolo di studio dei genitori** non emerge una tendenza particolarmente netta, fatta eccezione per il dato interessante secondo cui i giovani con entrambi i genitori laureati presentano il tasso occupazionale più basso rispetto alle altre categorie (63,8%). Tale fatto, accompagnato dall'elevata percentuale di giovani con entrambi i genitori in possesso di laurea che non hanno mai lavorato (24,2%), è coerente con quanto osservato nello studio svolto a proposito della maggiore propensione a svolgere attività formative post-titolo da parte di questa tipologia di laureati (l'80,6% dichiara di aver già concluso o di stare

ancora svolgendo almeno un'attività di formazione post-laurea contro il 66,7% di coloro i cui genitori sono in possesso al più della licenza elementare).

Un'ulteriore variabile analizzata è la *zona di residenza* degli intervistati: il tasso di occupazione maggiore, pari all'82,8%, si registra nelle regioni del Centro-Nord (eccettuata la Toscana), mentre il tasso più basso, pari al 60,7%, risulta al Sud e nelle Isole. Il Sud Italia e le Isole presentano altresì la percentuale più alta di disoccupati che non hanno mai lavorato (29,1%).

3. Il modello gerarchico a due livelli

Come anticipato nell'introduzione, per procedere ad una più soddisfacente individuazione delle possibili determinanti della condizione occupazionale si è fatto ricorso ai modelli di regressione multilivello⁸.

La finalità che s'intende perseguire attraverso il ricorso ai modelli di regressione multilivello è quella dell'individuazione delle variabili maggiormente esplicative della condizione occupazionale di un laureato a un anno e mezzo / due anni e mezzo dalla laurea (cioè al momento dell'intervista) e di misurarne l'effetto netto. La variabile risposta (indicata con *lavora*) presa in considerazione è, dunque, la condizione occupazionale del laureato al momento dell'intervista; le modalità che essa può assumere sono due (si tratta quindi di una variabile dicotomica): il laureato non lavora (*lavora*=0) oppure il laureato lavora (*lavora*=1)⁹. Trattandosi di una variabile binaria, il modello prescelto è stato il modello di tipo logistico (a due livelli), preferito ad un modello di tipo probit per la maggiore facilità di interpretazione dei risultati attraverso il calcolo degli *odds*¹⁰.

Si è proceduto alla stima del modello prendendo in considerazione come unità di primo livello tutti i 3856 laureati che hanno partecipato all'indagine¹¹, di cui il

⁸ Al riguardo si possono consultare, tra gli altri, i volumi di Goldstein H. (2003) e Snijders A.B., Bosker R. J. (1999).

⁹ In un primo momento si era pensato di usare una variabile risposta politomica (non ordinale) che tenesse conto del fatto che il laureato non occupato fosse in cerca di lavoro (quindi le possibili modalità sarebbero state: il laureato lavora, il laureato non lavora e non cerca lavoro, il laureato non lavora e cerca lavoro). Purtroppo, i software al momento disponibili non hanno consentito lo sviluppo di modelli multilivello con variabili risposta politomiche non ordinali.

¹⁰ Gli *odds* sono dati dal rapporto tra la probabilità che un laureato lavori rispetto alla probabilità che lo stesso laureato non lavori al momento dell'intervista.

¹¹ Considerato che lo scopo finale dell'analisi è capire che cosa influenza il tasso di neo-occupazione dei laureati (piuttosto che il tasso generale di occupazione), nel volume Chiandotto B., Bacci S. e Bertaccini B. (2004) viene presentato il modello che assume come unità di primo livello soltanto i 2666 laureati dell'anno 2000 che non lavoravano al momento della laurea.; tra le due applicazioni non sono state riscontrate differenze di particolare rilievo.

74,9% è occupato al momento dell'intervista e il 25,1% risulta non occupato; mentre come unità di secondo livello sono stati scelti i 39 corsi di laurea in cui sono suddivisi i laureati dell'anno 2000¹². La scelta delle unità di secondo livello è ricaduta sui corsi di laurea e non sulle facoltà a ragione delle differenze che di fatto sussistono, in maniera più o meno accentuata, tra corsi di laurea diversi all'interno di una stessa facoltà, differenze che hanno un'evidente ricaduta sui tassi di occupazione come evidenziato dai dati riportati in Tavola 1.

4. Variabili esplicative di primo livello e di secondo livello

Le variabili esplicative prese in considerazione per la stima del modello sono state scelte sulla base delle analisi descrittive svolte, di cui molto sommariamente si è dato conto nel secondo paragrafo, e sulla base della conoscenza del fenomeno. In particolare, come **covariate di primo livello** sono state incluse, almeno in una fase iniziale, le seguenti:

Variabili esplicative di primo livello continue

- **Tempo intercorso tra la laurea e l'inizio della ricerca di un lavoro (*tempolavoro*):** l'unità di misura di tale variabile è il numero di mesi e il valore minimo che essa può assumere è 0, nel caso in cui l'intervistato abbia iniziato a cercare un lavoro immediatamente dopo il conseguimento del titolo.
- **Tempo intercorso tra la laurea e l'intervista (*tempointerv*):** l'unità di misura di tale variabile è il numero di mesi e assume valori compresi tra 15 e 30 mesi.
- **Voto di laurea (*votolau*):** è il voto di laurea espresso in 110-mi e assume valori tra 66 e 113 (corrispondente al 110 e lode). Considerato che le variabili relative al voto di laurea e al voto medio agli esami sono fortemente correlate – risultando, infatti, un indice di correlazione di Pearson pari a 0,84 – si è preferito non inserirle entrambe nel modello per evitare fenomeni di multicollinearità. La scelta è caduta sul voto di laurea, piuttosto che sul voto medio agli esami, in quanto nel relazionarsi col mercato del lavoro (sia privato che pubblico) è il voto finale di laurea il principale elemento distintivo del neo-laureato.

¹² A causa dell'esiguo numero di laureati (2) in Agricoltura tropicale e subtropicale e in Economia Politica, è stato deciso di accorpare questi due corsi di laurea con, rispettivamente, Scienze agrarie tropicali e subtropicali ed Economia e commercio, in modo da non inficiare la significatività delle stime. Anche i due corsi di laurea in Lingue e letterature straniere di Lettere e Filosofia e di Scienze della Formazione sono stati considerati come un unico corso.

- **Età alla laurea (*etalau*):** tale variabile assume valori da un minimo di 21,9 anni ad un massimo di 67,6 anni. Nel modello non è invece stata inclusa la variabile relativa all'indice di durata¹³ sia a causa dell'elevata correlazione rispetto all'età alla laurea (indice di correlazione di Pearson pari a 0,58) sia perché, rispetto al mercato del lavoro così come rispetto alla scelta di proseguire ulteriormente la formazione, ha più importanza l'età a cui il laureato consegue il titolo, piuttosto che il numero di anni che esso ha impiegato per completare gli studi.
- **Voto di maturità (*votodip*):** è il voto di maturità espresso in 60-esimi e varia da un minimo di 36/60 al massimo di 60/60.

Variabili esplicative di primo livello discrete

- **Sesso (*sess*):** è una variabile binaria con modalità “femmina” (*femmina*) e “maschio” (*maschio*). Essendo le femmine laureate il 56,4%, come riferimento è stata assunta la modalità femmina¹⁴.
- **Esperienze lavorative durante gli studi (*lavI*):** è una variabile binaria con modalità “no” (*lavI_no* – il laureato non ha avuto esperienze lavorative durante gli studi) e “sì” (*lavI_si* – il laureato ha avuto esperienze lavorative durante gli studi). In tal caso, benché la maggior parte dei laureati abbia avuto esperienze lavorative durante gli studi, si è deciso di assumere come riferimento la modalità “no”: infatti, trattandosi di una variabile binaria con modalità sì/no risulta di più facile e immediata interpretazione un coefficiente di regressione relativo alla modalità “sì” (indicante la presenza della variabile).
- **Svolgimento di uno stage o tirocinio per il conseguimento del titolo (*tirocin*):** è una variabile binaria con modalità: “tirocinio non svolto” (*tirocin_no*) e “tirocinio svolto” (*tirocin_si*). Oltre l'83% dei laureati ha dichiarato di non aver svolto nessun tirocinio, quindi come riferimento è stata assunta la modalità “tirocinio non svolto”.
- **Frequenza alle lezioni (*frequenza*):** è una variabile binaria che può assumere le modalità: “frequenza elevata” (*elevata*) e “frequenza bassa” (*bassa*). Come riferimento è stata scelta la modalità “frequenza elevata”, essendo risultata leggermente più elevata dell'altra. Si puntualizza che la frequenza elevata caratterizza il laureato che ha frequentato regolarmente tutti o quasi tutti i corsi, mentre la frequenza bassa caratterizza il laureato che ha frequentato soltanto alcuni corsi saltuariamente oppure non ha mai frequentato alcun corso.
- **Tipo di maturità conseguita (*tipmat*):** le modalità assumibili da tale variabile sono: “classica” (*classica*), “scientifica” (*scientifica*), “tecnica” (*tecnica*), “altra

¹³ L'indice di durata è definito come rapporto tra durata effettiva degli studi universitari svolti e durata legale.

¹⁴ Tranne le poche eccezioni segnalate nel testo, come regola generale si è scelto di adottare come riferimento per le variabili esplicative discrete la modalità con la frequenza osservata più elevata.

maturità" (*altramat*). Come riferimento è stata adottata la modalità "scientifica", essendo quella che si presenta con la maggior frequenza.

- **Classe sociale di appartenenza (*p_socgen*):** le modalità assumibili da tale variabile sono: "borghesia" (*borghesia*), "classe media impiegatizia" (*impiegatizia*), "piccola borghesia" (*picc_borgh*), "classe operaia" (*operaia*). La modalità presa come riferimento è "borghesia".
- **Titolo di studio dei genitori (*tit_stud*):** in una prima fase sono state considerate le modalità "al più licenza elementare" (*elementare*), "licenza media inferiore" (*media*), "diploma di scuola superiore" (*superiori*), "un genitore con laurea" (*unalaui*), "entrambi i genitori laureati" (*duelau*); "diploma di scuola superiore" è stato assunto come riferimento. In una seconda fase, poiché, da una parte, la variabile nel suo complesso è risultata significativa sulla base del test di Wald multivariato, ma, dall'altra, soltanto alcune delle singole modalità sono risultate tali, si è deciso di procedere ad un raggruppamento delle suddette modalità, in modo da facilitare l'interpretazione del risultato finale. In particolare, la variabile *tit_stud* è stata ricondotta ad una variabile binaria con modalità "al più scuola dell'obbligo" (*tit_basso*) e "almeno diploma di scuola superiore" (*tit_alto*) – modalità quest'ultima presa come riferimento .
- **Regione di residenza (*macroreg*):** anche in questo caso, per gli stessi motivi esposti al punto precedente, la variabile è stata trasformata da politomica in dicotomica. Le modalità iniziali – "province di Firenze e Prato" (*Fi_Po*), "altra provincia Toscana" (*altraToscana*), "altra regione del Centro-Nord" (*CentroNord*), "Sud e Isole" (*Sud*), "estero" (*estero*) – sono state raggruppate nelle due modalità "residente al Centro-Nord o all'estero" (*resid_nosud*) – modalità assunta come riferimento - e "residente al Sud" (*resid_sud*).
- **Conoscenza della lingua inglese (*ling1_gb*):** le modalità assumibili da tale variabile sono "nessuna/scarsa" (*ingl_scarso*), "sufficiente" (*ingl_suff*), "buona" (*ingl_buono*), "ottima" (*ingl_ottimo*). La maggior parte degli intervistati dichiara una buona conoscenza dell'inglese, quindi questa modalità è stata assunta come riferimento.
- **Conoscenza di word processor (*info3_wp*):** le modalità considerate sono le stesse della variabile *ling1_gb* (conoscenza della lingua inglese); la modalità più frequente risulta essere quella relativa ad una buona conoscenza di word processor (*wp_buono*).
- **Conoscenza di fogli elettronici (*info4_fe*):** stesse modalità di cui sopra; come riferimento è stata assunta la scarsa o inesistente conoscenza dell'uso di fogli elettronici, risultando questa la più diffusa. La conoscenza di altri strumenti informatici non è stata presa in considerazione, in quanto trattasi di strumenti peculiari soltanto per alcuni gruppi di laureati (per es. l'uso di CAD per gli archi-

tetti o la conoscenza di *Database* per gli ingegneri e gli informatici) e, quindi, richiesti eventualmente soltanto per specifiche tipologie di lavori.

Con riferimento alle covariate discrete, il **laureato – base** è una persona di sesso femminile, non ha avuto esperienze lavorative durante gli studi, ha conseguito la maturità scientifica, proviene da una famiglia appartenente alla classe borghese, i genitori sono in possesso del diploma di scuola superiore, risiede nelle province di Firenze o Prato, ha una buona conoscenza della lingua inglese e dei programmi di scrittura su PC, mentre ha scarse conoscenze dell'uso di fogli elettronici.

Entrando nel merito delle variabili esplicative di secondo livello, ideale sarebbe stato poter disporre di un indicatore diretto delle differenze tra corsi di laurea. Non avendo a disposizione nessuna variabile di questo tipo, si è deciso di creare variabili che, in qualche modo, fossero in grado di evidenziare le diverse caratteristiche dei corsi di laurea dell'Ateneo, ricorrendo all'aggregazione per singoli corsi di alcune variabili di primo livello (livello individuale) ritenute più adatte a tale scopo.

In particolare, le **covariate di secondo livello** impiegate sono:

- **Percentuale di maschi per corso di laurea (*maschi_corsi*)**
- **Percentuale di laureati che hanno svolto il tirocinio per il conseguimento del titolo per corso di laurea (*tiroc_corsi*)**
- **Percentuale di laureati che hanno lavorato durante gli studi per corso di laurea (*lav1_corsi*)**
- **Percentuale di laureati con maturità non liceale per corso di laurea (*tipmat_corsi*)**
- **Voto medio di maturità per corso di laurea (*votodip_corsi*):** per ogni corso è stata calcolata la media dei voti di maturità dei laureati ad esso afferenti
- **Voto agli esami medio per corso di laurea (*votoesami_corsi*):** per ogni corso è stata calcolata la media dei voti conseguiti agli esami dai rispettivi laureati. Mentre tra le variabili di primo livello si è scelto di impiegare il voto di laurea piuttosto che il voto medio agli esami per i motivi suddetti, tra le variabili di secondo livello si è preferito ricorrere al voto agli esami, in quanto presenta, rispetto al voto di laurea, una distribuzione maggiormente diversificata e, quindi, è in grado di fornire una rappresentazione migliore delle differenze tra corsi di laurea¹⁵.

¹⁵ In un secondo momento il modello è stato stimato sostituendo alla variabile *votoesami_corsi* il voto di laurea medio per corso di laurea, che però non è risultato significativo.

5. Stima del modello relativa a tutti i laureati dell'anno 2000

Il modello è stato stimato facendo ricorso alla PROC NLMIXED¹⁶ del software SAS¹⁷ con impiego dell'algoritmo di ottimizzazione Dual Quasi-Newton¹⁸ e del metodo di quadratura Gaussiana (non adattiva).

Per la stima del modello si è seguito quanto suggerito dalla teoria in materia. In primo luogo, si è proceduto alla stima del modello vuoto (cioè privo di covariate), al fine di verificare, attraverso l'analisi della varianza dei residui di secondo livello, la ragionevolezza del ricorso ad un'analisi multilivello. In secondo luogo, è stato stimato il modello a intercetta casuale comprendente tutte le variabili di primo livello e, a questo punto, la selezione del modello è avvenuta eliminando progressivamente quelle covariate risultanti non significative (ad un livello di significatività del 10%) sulla base del test di Wald e confrontando tra loro i diversi modelli ottenuti tramite i risultati del test del rapporto di verosimiglianza e i valori assunti dagli indici AIC e BIC (cfr. nota 22). Infine, seguendo la medesima procedura, sono state selezionate le covariate di secondo livello, pervenendo dunque ad un *modello a intercetta casuale*¹⁹ *finale a due livelli di aggregazione*.

Di seguito verranno riportati e commentati soltanto i risultati relativi alla stima del modello a intercetta casuale comprendente le sole variabili di primo livello (compresi termini quadratici e interazioni) e quelli relativi alla stima finale, cioè, quelli relativi alla stima del *Modello a intercetta casuale* nel quale sono presenti le sole *variabili di primo e secondo livello significative*.

Il modello a intercetta casuale in presenza di covariate (variabili esplicative) assume, per ogni gruppo – corso di laurea - j ($j=1, 2, \dots, 39$), la seguente forma generale:

¹⁶ Si tratta di una procedura specifica per la stima di modelli multilivello (con due soli livelli di aggregazione) con variabili risposta discrete (binarie, ordinali, poissoniane ecc.); in altri termini tale procedura tratta dei modelli GLM (*Generalized Linear Models*) in versione multilivello.

¹⁷ Merita sottolineare che il SAS, durante una elaborazione, elimina automaticamente tutte le risposte mancanti (*missing*), fatto questo che ha comportato una notevole riduzione del contingente considerato. Una possibile soluzione per evitare fenomeni di questo genere consiste nel procedere ad imputazioni, che siano il più logiche possibile, dei dati mancanti. In realtà, tale procedura può rivelarsi piuttosto pericolosa in termini di correttezza di stima di un modello, soprattutto se le imputazioni a cui si procede sono numerose e la variabilità di ciascuna covariata è notevole. Quindi, per evitare di inserire elementi di eccessiva discrezionalità nella stima del modello, si è preferito ignorare i *missing*.

¹⁸ In letteratura sono stati sviluppati numerosi algoritmi di ottimizzazione: la scelta è caduta sull'algoritmo Dual Quasi-Newton in quanto è in grado di creare un bilanciamento appropriato tra velocità di calcolo e stabilità per la maggior parte dei modelli gerarchici non lineari.

¹⁹ Si puntualizza che sono stati testati anche modelli con coefficienti casuali che, però, non hanno dato risultati degni di attenzione, in quanto, nella maggior parte dei casi, o l'algoritmo di massimizzazione della funzione di verosimiglianza non ha raggiunto la convergenza oppure i coefficienti stimati non sono risultati significativi.

$$Y_{ij} = P_{ij} + \varepsilon_{ij},$$

dove Y_{ij} è il valore assunto dalla variabile risposta per l' i -esimo individuo del j -esimo gruppo, P_{ij} indica la probabilità di essere occupato per un laureato estratto casualmente da un *generico* corso di laurea, mentre ε_{ij} rappresenta la componente residua di I livello avente, per ipotesi, media pari a zero e varianza strettamente dipendente dalla probabilità media di successo per ogni gruppo j . Usualmente, il modello viene presentato in funzione di una opportuna trasformazione di P_{ij} , detta trasformazione logit:

$$\log it(P_{ij}) = \log\left(\frac{P_{ij}}{1 - P_{ij}}\right) = \beta_{0j} + \sum_{k=1}^m \beta_k \cdot x_{kij} + \sum_{t=1}^s \beta_t \cdot z_{tj}.$$

Più precisamente, essendo l'intercetta β_{0j} variabile in funzione del gruppo j , cioè

$$\beta_{0j} = \gamma_0 + U_{0j},$$

il modello assume la seguente forma:

$$\log it(P_{ij}) = \log\left(\frac{P_{ij}}{1 - P_{ij}}\right) = \gamma_0 + \sum_{k=1}^m \beta_k \cdot x_{kij} + \sum_{t=1}^s \beta_t \cdot z_{tj} + U_{0j},$$

dove si è indicato con γ_0 la parte fissa dell'intercetta, con x_k la k -esima variabile esplicativa di I livello, con z_t la t -esima variabile esplicativa di II livello e con β_k e β_t i corrispondenti coefficienti di regressione²⁰, mentre U_{0j} rappresenta la componente residua di II livello (in altri termini, la parte variabile dell'intercetta). La ragionevolezza dell'assunzione di tale modello risiede nell'ipotesi di distribuzione normale dei residui di II livello con media pari a 0 e varianza costante²¹.

La forma esplicita assunta dal modello, considerando in primo luogo soltanto le covariate di I livello, è la seguente:

²⁰ Si ricorda che nel modello logistico ciascun coefficiente di regressione è uguale al logaritmo dell'*odds ratio* associato ad un incremento pari ad 1 della corrispondente variabile esplicativa. In particolare, gli *odds ratio* non sono altro che il rapporto tra gli *odds* di due diverse categorie di soggetti; per esempio, possono rappresentare la probabilità di occupazione rispetto alla probabilità di non occupazione per i maschi rapportata alla probabilità di occupazione rispetto alla probabilità di non occupazione per le femmine.

²¹ Si fa notare che una condizione fondamentale per l'adozione di un modello multilivello è la significatività della componente di varianza di II livello, condizione questa che è risultata sempre verificata nel corso delle elaborazioni svolte.

$$\begin{aligned} \text{logit}(P_{ij}) = & \gamma_0 + \beta_1 \cdot \text{tempolavoro}_{ij} + \beta_2 \cdot \text{tempointerv}_{ij} + \beta_3 \cdot \text{votolau}_{ij} + \beta_4 \cdot \text{etalau}_{ij} + \\ & + \beta_5 \cdot \text{votodip}_{ij} + \beta_6 \cdot \text{sesso}_{ij} + \beta_7 \cdot \text{lav1}_{ij} + \beta_8 \cdot \text{tirocina}_{ij} + \beta_9 \cdot \text{frequenza}_{ij} + \beta_{10} \cdot \text{classica}_{ij} + \\ & + \beta_{11} \cdot \text{tecnica}_{ij} + \beta_{12} \cdot \text{altramat}_{ij} + \beta_{13} \cdot \text{impiegatizia}_{ij} + \beta_{14} \cdot \text{picc_borgh}_{ij} + \beta_{15} \cdot \text{operaia}_{ij} + \\ & + \beta_{16} \cdot \text{tit_basso}_{ij} + \beta_{17} \cdot \text{resid_sud}_{ij} + \beta_{18} \cdot \text{ingl_scarso}_{ij} + \beta_{19} \cdot \text{ingl_suff}_{ij} + \\ & + \beta_{20} \cdot \text{ingl_ottimo}_{ij} + \beta_{21} \cdot \text{wp_scarso}_{ij} + \beta_{22} \cdot \text{wp_suff}_{ij} + \beta_{23} \cdot \text{wp_ottimo}_{ij} + \\ & + \beta_{24} \cdot \text{fe_suff}_{ij} + \beta_{25} \cdot \text{fe_buono}_{ij} + \beta_{26} \cdot \text{fe_ottimo}_{ij} + U_{0j} \end{aligned}$$

Nella Tavola 4 sono riportati i risultati della stima del modello.

Le variabili più significative - con p-value inferiore a 0,0001 - sono *tempolavoro* e *resid_sud*. Riguardo a *tempolavoro*, l'effetto sulla probabilità di occupazione è molto tenue, ma comunque negativo (variazione negli *odds ratio* pari 0,98): ciò indica che quanto maggiore è il tempo che un laureato lascia intercorrere tra la laurea e l'inizio della ricerca di un lavoro, tanto inferiore è la probabilità di risultare occupato al momento dell'intervista. Riguardo a *resid_sud*, l'effetto esercitato sulla probabilità di occupazione è, invece, molto più forte: per i laureati residenti al Sud o nelle Isole la probabilità di occupazione rispetto a quella di non occupazione è soltanto il 30% circa della probabilità di occupazione rispetto a quella di non occupazione che si rileva per i laureati residenti al Centro Nord (o all'estero).

Con un p-value pari a 0,0075, il titolo di studio dei genitori (*tit_stud*) risulta una delle variabili più significative del modello; la variazione negli *odds ratio* pari a 1,64 conferma quanto già rilevato in sede di analisi descrittiva: i laureati i cui genitori hanno al più la licenza di scuola dell'obbligo mostrano una probabilità di occupazione di oltre 1,6 volte rispetto ai colleghi con genitori almeno diplomati. Seguono, con livelli di significatività intorno all'1,5%, le covariate *sesso* (p-value=1,54%) e *votodip* (p-value=1,79%). Riguardo al sesso, la probabilità di occupazione per i laureati maschi rispetto alla probabilità di non occupazione è di circa il 43,0% superiore a quella delle laureate; relativamente a *votodip*, l'effetto stimato risulta debole, anche se positivo, indicando che a voti di maturità più elevati corrispondono tendenzialmente maggiori probabilità di occupazione.

Altre due variabili risultate significative sono *votolau* (p-value=5,59%) e *lav1* (p-value=8,85%); l'effetto del voto di laurea è molto debole, ma negativo - al contrario del voto di maturità - indicando che a voti di laurea più alti corrisponde tendenzialmente una minore probabilità di occupazione. L'effetto stimato per *lav1* è, invece, più consistente, risultando una variazione negli *odds ratio* pari a 1,27, esplicitiva di una maggiore probabilità di occupazione per i laureati che hanno già avuto esperienze lavorative durante gli studi rispetto a quelli che non ne hanno mai avute. Nessuna delle altre variabili inserite nel modello risulta significativa.

Tavola 4. Modello a intercetta casuale con tutte le variabili esplicative di primo livello

	Riferimento	Stima	Errore Standard	Valore t	p-value	Variazione odds ratio
Effetti fissi						
Intercetta	--	2,56660	1,51810	1,69	0,0991	--
Tempolavoro	--	-0,02611	0,00195	-13,42	<,0001	0,9742
Tempointerv	--	0,01387	0,01260	1,1	0,278	1,0140
Votolau	--	-0,02287	0,01159	-1,97	0,0559	0,9774
Etalau	--	-0,02345	0,02813	-0,83	0,4095	0,9768
Votodip	--	0,02486	0,01004	2,48	0,0179	1,0252
Sesso	femmina	0,36100	0,14220	2,54	0,0154	1,4348
lav1	lav1_no	0,23730	0,13570	1,75	0,0885	1,2678
Tirocin	tirocin_no	0,19550	0,21040	0,93	0,3586	1,2159
Frequenza	elevata	0,04945	0,14380	0,34	0,7327	1,0507
Classica	scientifica	-0,15750	0,17250	-0,91	0,3671	0,8543
Tecnica	scientifica	0,19660	0,18250	1,08	0,2881	1,2173
Altramat	scientifica	0,19980	0,20180	0,99	0,3285	1,2212
Impiegatzia	borghesia	0,15910	0,16030	0,99	0,3270	1,1725
picc_borgh	borghesia	-0,27100	0,20200	-1,34	0,1877	0,7626
Operaia	borghesia	-0,35760	0,24340	-1,47	0,1499	0,6994
tit_basso	tit_alto	0,49170	0,17400	2,83	0,0075	1,6351
resid_sud	resid_nosud	-1,15230	0,24910	-4,63	<,0001	0,3159
ingl_scarso	ingl_buono	-0,06570	0,22190	-0,3	0,7688	0,9364
ingl_suff	ingl_buono	0,15130	0,15030	1,01	0,3204	1,1633
ingl_ottimo	ingl_buono	0,15510	0,19610	0,79	0,4339	1,1678
wp_scarso	wp_buono	0,18560	0,21020	0,88	0,3829	1,2039
wp_suff	wp_buono	0,04030	0,17770	0,23	0,8219	1,0411
wp_ottimo	wp_buono	0,11530	0,18770	0,61	0,5429	1,1222
fe_suff	fe_scarso	0,28500	0,17630	1,62	0,1143	1,3298
fe_buono	fe_scarso	0,34960	0,20310	1,72	0,0934	1,4185
fe_ottimo	fe_scarso	0,37750	0,27470	1,37	0,1773	1,4586
Effetti casuali						
Var(U_{0j})	--	0,26200	0,07699	3,4	0,0016	
Test di Wald multivariato		GdL	Valore F di Fisher		p-value	
Tipmat		38	1,26		0,3015	
p_socgen		38	1,92		0,1432	
ling1_gb		38	0,61		0,6143	
info3_wp		38	0,34		0,7958	
info4_fe		38	1,43		0,2491	

L'eliminazione delle covariate non significative e l'analisi dei possibili termini quadratici nonché dei plausibili effetti d'interazione hanno portato alla individuazione del seguente modello:

$$\logit(P_{ij}) = \gamma_0 + \beta_1 \cdot \text{tempolavoro}_{ij} + \beta_2 \cdot \text{votodip}_{ij} + \beta_3 \cdot \text{sesso}_{ij} + \beta_4 \cdot \text{lav1}_{ij} + \\ + \beta_5 \cdot \text{tit_basso}_{ij} + \beta_6 \cdot \text{resid_sud}_{ij} + \beta_7 \text{lav1} \times \text{tempolavoro} + U_{0j}$$

I risultati conseguiti attraverso la procedura di stima sono riportati nella Tavola 5.

Rispetto al modello iniziale, la differenza più rilevante risiede nell'esclusione del voto di laurea dalle covariate significative: in pratica ciò è indice della sostanziale perdita di potere discriminatorio da parte del giudizio finale, a causa essenzialmente dell'eccessiva concentrazione dello stesso intorno ai valori più elevati.

Per quanto riguarda le variabili esplicative risultate significative, gli effetti sulla probabilità di occupazione esercitati sia dal tempo intercorso tra la laurea e l'inizio della ricerca di un lavoro (*tempolavoro*) che dal voto conseguito alla maturità (*votodip*) risultano sostanzialmente immutati rispetto al modello iniziale. Per contro, per le rimanenti covariate presenti in entrambi i modelli l'effetto stimato mostra variazioni sensibili, benché la "direzione" di tali variazioni rimanga immutata; l'unica variabile aggiuntiva, rispetto al modello iniziale, risultata significativa è l'interazione tra *lav1* e *tempolavoro*.

Per tener conto in maniera adeguata della natura gerarchica dei dati si è infine proceduto all'inserimento delle variabili esplicative di II livello citate precedentemente, così da pervenire ad un modello definitivo. In realtà, a seguito di alcune semplici considerazioni che verranno espone nelle righe successive, si è preferito concentrare l'attenzione su due possibili modelli alternativi, che differiscono per una sola variabile di secondo livello: la variabile *votoesami_corsi* è stata sostituita dalla variabile *maschi_corsi*.

Tavola 5. Modello a intercetta casuale con le variabili esplicative di primo livello significative e un effetto di interazione

	Riferimento	Stima	Errore Standard	Valore t	p-value	Variazione odds ratio
Effetti fissi						
intercetta	--	0,86910	0,39200	2,22	0,0327	--
<i>tempolavoro</i>	--	-0,03310	0,00210	-11,41	<0,0001	0,9675
<i>votodip</i>	--	0,01580	0,00762	2,07	0,0451	1,0159
<i>sesso</i>	<i>femmina</i>	0,32410	0,11590	2,80	0,0081	1,3828
<i>lav1</i>	<i>lav1_no</i>	0,26210	0,12410	2,11	0,0413	1,2997
<i>tit_basso</i>	<i>tit_alto</i>	0,32810	0,11220	2,92	0,0058	1,3883
<i>resid_sud</i>	<i>resid_nosud</i>	-0,79200	0,19660	-4,03	0,0003	0,4529
<i>lav1*tempolavoro</i>	--	0,01157	0,00353	3,28	0,0022	1,0116
Effetti casuali						
<i>Var(U_{0j})</i>	--	0,41290	0,08024	5,15	<0,0001	

Modello A (Modello B)

$$\log it(P_{ij}) = \gamma_0 + \beta_1 \cdot tempolavoro_{ij} + \beta_2 \cdot votodip_{ij} + \beta_3 \cdot sesso_{ij} + \beta_4 \cdot lav1_{ij} + \beta_5 \cdot tit_basso_{ij} + \beta_6 \cdot resid_sud_{ij} + \beta_7 (lav1 \times tempolavoro)_{ij} + \beta_8 \cdot votodip_corsi_j + \beta_9 \cdot (votoesami_corsi)_j \Leftarrow (maschi_corsi)_j + U_{0j}$$

Tavola 6. Modello A: modello a intercetta casuale con variabili esplicative di primo e secondo livello significative e un effetto di interazione

	Riferimento	Stima	Errore Standard	Valore t	p-value	Variazione odds ratio
Effetti fissi						
intercetta	--	-0,01130	2,58150	0,00	0,9965	--
tempolavoro	--	-0,03261	0,00294	-11,08	<0,0001	0,9679
votodip	--	0,01468	0,00788	1,86	0,0701	1,0148
sesso	Femmina	0,29750	0,11830	2,52	0,0162	1,3465
lav1	lav1_no	0,24420	0,12490	1,95	0,0580	1,2766
tit_basso	tit_alto	0,33650	0,11290	2,98	0,0050	1,4000
resid_sud	resid_nosud	-0,89950	0,19960	-4,51	<0,0001	0,4068
lav1*tempolavoro	--	0,01170	0,00358	3,27	0,0023	1,0118
votodip_corsi	--	0,14200	0,04092	3,47	0,0013	1,1526
votoesami_corsi	--	-0,23090	0,06306	-3,66	0,0008	0,7938
Effetti casuali						
Var(U _{0j})	--	0,42750	0,07939	5,39	<0,0001	

Tavola 7. Modello B: modello a intercetta casuale con variabili esplicative di primo e secondo livello significative e un effetto di interazione (maschi_corsi al posto di votoesami_corsi)

	Riferimento	Stima	Errore Standard	Valore t	p-value	Variazione odds ratio
Effetti fissi						
intercetta	--	-5,65490	1,93010	-2,93	0,0057	--
tempolavoro	--	-0,03260	0,00294	-11,09	<0,0001	0,9679
votodip	--	0,01458	0,00787	1,85	0,0717	1,0147
sesso	Femmina	0,29410	0,12080	2,44	0,0197	1,3419
lav1	lav1_no	0,23800	0,12450	1,91	0,0635	1,2687
tit_basso	tit_alto	0,32530	0,11250	2,89	0,0063	1,3844
resid_sud	resid_nosud	-0,91690	0,19950	-4,60	<0,0001	0,3998
lav1*tempolavoro	--	0,01203	0,00357	3,37	0,0017	1,0121
votodip_corsi	--	0,12360	0,04187	2,95	0,0054	1,1316
maschi_corsi	--	0,84440	0,35030	2,41	0,0209	2,3266
Effetti casuali						
Var(U _{0j})	--	0,31660	0,06388	4,96	<0,0001	

In un confronto tra i due modelli, gli indici AIC e BIC²², assumono valori inferiori per il Modello A (AIC = 2324 e BIC = 2343) rispetto al Modello B (AIC = 2332 e BIC=2350); ciò induce a concludere per una migliore bontà di adattamento del Modello A a scapito del Modello B. D'altra parte, osservando i valori della varianza della componente residua di secondo livello ci si accorge che per il Modello A si ottiene un valore pari a 0,43 che, da una parte, risulta sostanzialmente uguale all'analogo valore ottenuto per il modello privo di variabili di secondo livello e, dall'altra, è comunque superiore alla varianza degli errori di secondo livello del Modello B, pari a 0,32. In base a tali risultati sembra, dunque, che l'inserimento della variabile *maschi_corsi* al posto di *votoesami_corsi* riesca a spiegare una quota maggiore della variabilità del fenomeno, benché l'adattamento del modello risulti leggermente peggiore.

Alla luce di queste considerazioni si è ritenuto più opportuno presentare entrambi i modelli. Il motivo che ha indotto a non inserire le due variabili *maschi_corsi* e *votoesami_corsi* in un unico modello è da ricercarsi nella loro elevata correlazione ($-0,47$)²³, l'inserimento di entrambe le variabili nello stesso modello avrebbe comportato (come infatti è successo) la selezione di una sola delle due.

In entrambi i modelli, delle sei variabili di secondo livello prese in considerazione inizialmente, soltanto due risultano significative: il voto medio agli esami o la percentuale di maschi, appunto, e il voto medio di maturità per corso di laurea. Ri-

²² Il criterio d'informazione di Akaike (Akaike's Information Criterion – AIC) e il criterio d'informazione bayesiano di Schwarz (Schwarz's Bayesian Information Criterion – BIC) sono due grandezze impiegate ogni volta che si è interessati ad un confronto tra modelli contenenti *diverse* variabili esplicative. Per un generico modello M i due indici vengono così definiti:

$$AIC = Dev_M + 2k$$

$$BIC = Dev_M + k \ln(N)$$

dove Dev_M è il doppio prodotto della differenza tra la log-verosimiglianza del c.d. *modello saturo* – vale a dire il modello contenente un parametro per ciascuna osservazione e, quindi, dotato della massima capacità di adattamento ai dati osservati (e della minima capacità di sintesi delle informazioni) – e la log-verosimiglianza del modello M, k è il numero complessivo di parametri stimati e N è il numero di unità di secondo livello. Dal momento che quanto più piccola è la devianza del modello tanto migliore è la sua bontà di adattamento ai dati osservati (in tal caso infatti si riduce la “distanza” tra il modello M ed il modello saturo) e dal momento che al diminuire della devianza anche i valori dei due indici diminuiscono (a parità di k), in un confronto tra due o più modelli diversi possiamo ritenere migliore, in termini di bontà di adattamento, quello che presenta il valore AIC o BIC più basso di tutti gli altri.

Si osservi, però, che entrambi gli indici aumentano all'aumentare di k , cioè del numero di parametri inseriti nel modello, anche se questi potrebbero non essere significativi: è, quindi, necessario ricorrere con cautela all'impiego di questi due indicatori, accertandosi in via preventiva che i parametri inseriti nei modelli posti a confronto raggiungano il livello di significatività desiderato.

²³ I corsi di laurea dove prevale la presenza maschile (es. i corsi della facoltà di Ingegneria) sono caratterizzati da votazione media più bassa, mentre nei corsi di laurea a prevalenza femminile (es. i corsi della facoltà di Lettere e Filosofia) i voti medi agli esami sono più alti.

guardo al voto medio agli esami per corso di laurea si osserva un effetto negativo sulla probabilità di occupazione, risultando una variazione negli *odds ratio* pari a 0,79: in altri termini, un laureato proveniente da un corso di laurea caratterizzato da voti medi elevati ha circa il 20% in meno di possibilità di occupazione rispetto ad un laureato proveniente da un corso di laurea caratterizzato da voti medi più bassi. Per comprendere meglio si pensi ai corsi di laurea attivati a Ingegneria e a quelli attivati a Lettere e Filosofia: le analisi svolte hanno evidenziato come i voti medi agli esami di Ingegneria siano di gran lunga inferiori a quelli di Lettere e Filosofia, mentre i tassi di occupazione sono sicuramente più elevati nel primo caso che non nel secondo. Coerentemente, il Modello B mostra che per un laureato proveniente da un corso di laurea a prevalenza maschile la probabilità di occupazione rispetto a quella di non occupazione è di oltre 2,3 volte la stessa probabilità per un laureato proveniente da un corso di laurea a prevalenza femminile. Riguardo al voto medio di maturità per corso di laurea (*votodip_corsi*), si rileva un effetto positivo: quanto più sono “bravi” gli studenti di un certo corso di laurea e tanto maggiore è la probabilità che gli stessi, una volta laureati, risultino occupati nel breve termine; più precisamente, all'aumentare del voto medio di maturità per corso di laurea, la probabilità di occupazione rispetto a quella di non occupazione aumenta di 1,13-1,15 volte (rispettivamente Modello B e Modello A).

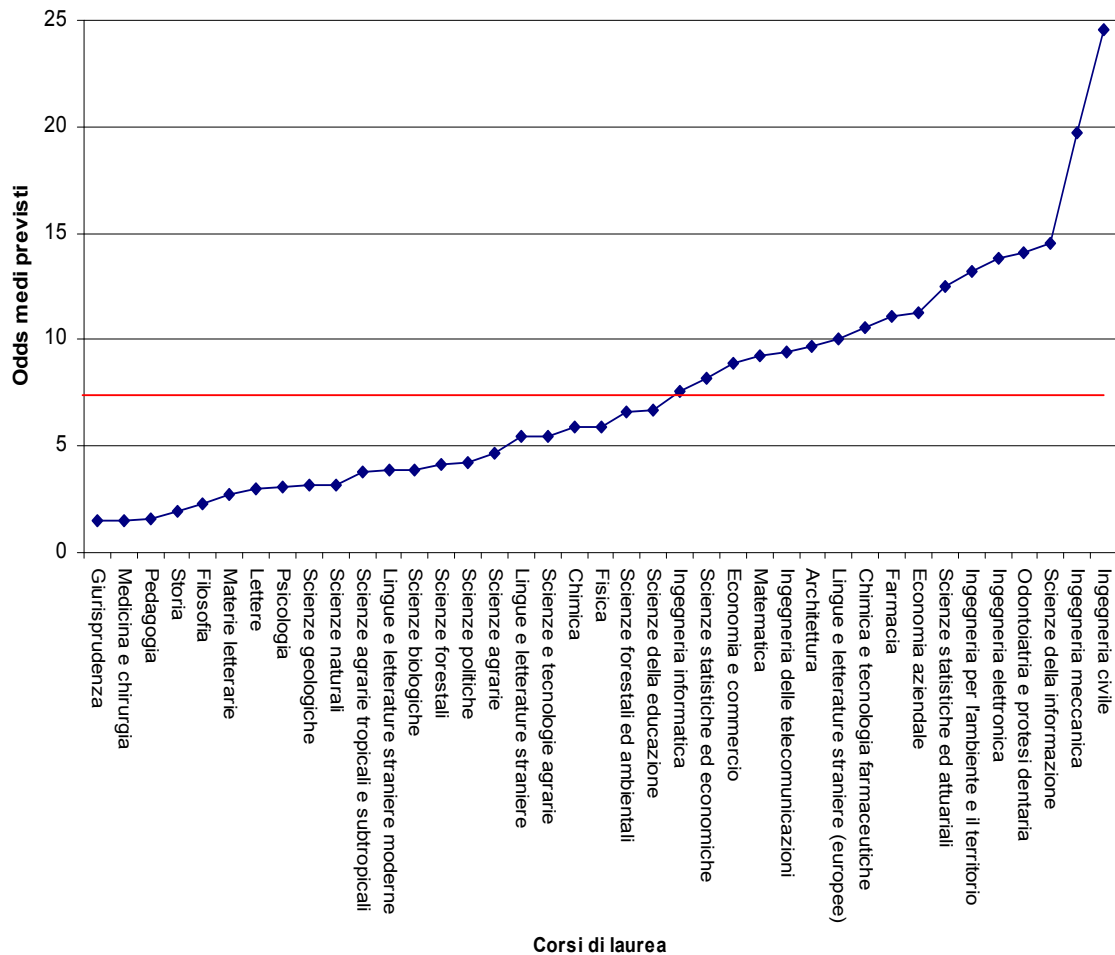
Riguardo, infine, alle stime dei coefficienti delle variabili di primo livello comprese nei due modelli, non si rilevano particolari differenze né tra il Modello A e il Modello B (anche in termini di significatività delle variabili), né tra ciascuno di questi due rispetto al modello privo di variabili di secondo livello.

A completamento dell'analisi, per ognuno dei 39 corsi di laurea sono stati stimati i relativi *odds* sulla base del modello a intercetta casuale con variabili esplicative di primo e secondo livello²⁴. Dal momento che gli *odds* vengono stimati tenendo conto delle caratteristiche individuali e di contesto di ciascun corso di laurea, a livello interpretativo il valore assunto, per un particolare corso di laurea, dal relativo *odd* esprime la probabilità di occupazione rispetto alla probabilità attesa di non occupazione a un anno e mezzo / due anni e mezzo dalla laurea per il laureato medio proveniente da quel corso. Quella che si ottiene è una graduatoria esplicativa della diversa capacità di ciascun corso di creare occupazione nel breve termine (cfr. Figura 1).

In generale, gli *odds* che si ottengono assumono valori molto elevati, elemento questo che indica come i laureati dell'Ateneo fiorentino non incontrino particolari difficoltà di inserimento nel mondo del lavoro (l'*odd* stimato per l'intero ateneo risulta pari a 7,3); tuttavia, le differenze tra i singoli corsi di laurea risultano essere

²⁴ Per non appesantire troppo la lettura, i risultati commentati in queste note conclusive si riferiscono soltanto al Modello A, cioè al modello contenente la variabile di secondo livello *votoesami_corsi*; si puntualizza, comunque, che le stesse analisi svolte per il Modello B non hanno posto in evidenza differenze particolari e, dunque, non avrebbero condotto a conclusioni sostanzialmente diverse.

Figura 1. Odds medi previsti per ogni corso di laurea, stimati sulla base del Modello A (tutti i laureati)



piuttosto consistenti. A tal proposito si individuano quattro gruppi principali: un primo gruppo comprende quei corsi con *odds* inferiori a 5; un secondo gruppo comprende corsi con *odds* superiori a 5 e inferiori a 10; al terzo gruppo appartengono i corsi con *odds* superiori a 10 e inferiori a 15; infine, dell'ultimo gruppo fanno parte i corsi con *odds* superiori a 15.

In particolare, al primo gruppo appartengono, tra gli altri, i corsi di laurea in Giurisprudenza e Medicina e chirurgia, che con gli *odds* in assoluto più bassi (pari a 1,5) confermano ampiamente quanto già sottolineato riguardo all'alto numero di laureati in tali discipline impegnati in attività di formazione e specializzazione post-laurea.

Per contro, i due corsi di laurea che fanno registrare *odds* superiori a 15 appartengono alla Facoltà di Ingegneria (Ingegneria meccanica e Ingegneria civile con *odds* rispettivamente pari a 19,7 e 24,6); facoltà questa che presenta, per tutti i corsi di laurea, valori superiori alla media di Ateneo, essendo infatti Ingegneria informatica il corso con *odd* più basso (lievemente superiore a 7,3).

6. Considerazioni conclusive

I risultati delle analisi svolte consentono di concludere che la probabilità di occupazione rispetto alla probabilità di non occupazione per un laureato risulta maggiore per i maschi che non per le femmine, per coloro che hanno avuto esperienze lavorative durante gli studi piuttosto che per chi non ne ha mai avute, per coloro che provengono da famiglie con basso profilo culturale rispetto ai laureati i cui genitori sono in possesso per lo meno di un diploma di scuola superiore, per coloro che risiedono nelle regioni del Centro-Nord piuttosto che per i residenti nel Sud Italia. Inoltre, la probabilità di occupazione è tanto maggiore quanto più elevato è il voto di maturità e quanto più breve è il tempo che il laureato lascia intercorrere tra il conseguimento del titolo e l'inizio della ricerca di un lavoro, variabile questa il cui effetto dipende altresì dal fatto che il laureato abbia avuto precedenti esperienze di lavoro durante gli studi. Infine, a parità di valori assunti dalle variabili individuali, la probabilità di occupazione è maggiore per i laureati che provengono da corsi di laurea caratterizzati da elevate percentuali di maschi, che si distinguono per voti medi agli esami relativamente bassi e che attirano gli studenti più bravi, cioè con voti di maturità più elevati.

I risultati delle analisi svolte, molto sommariamente richiamati in questa nota, giustificano, a nostro parere, ampiamente il ricorso ai *modelli multilivello* quando si procede all'analisi di dati che riguardano gli studenti universitari²⁵; infatti, è del tutto

²⁵ 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 fiorentino. Testimoniano, tra l'altro, tale interesse altri due lavori presentati in questa sede: "L'abbandono degli studi universitari" (Chiandotto B. e Giusti C.); "Un modello multilivello per l'analisi della durata degli studi universitari" (Chiandotto B. e Varriale R.), il già citato volume "I laureati e diplomati dell'Ateneo fiorentino dell'anno 2000: profilo e sbocchi occupazionali" (Chiandotto B., Bacci S. e Bertaccini B.; 2004) che, oltre a contenere una versione molto più dettagliata dell'analisi svolta ed i cui risultati sono stati qui riassunti, tratta anche dell'applicazione del modello di regressione logistica ordinale a due livelli dove la variabile di interesse di cui si vogliono individuare le possibili determinanti è rappresentata dall'utilizzo sul posto di lavoro delle competenze acquisite

evidente la natura gerarchica dei dati: le unità di primo livello sono i laureati o i 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

BERTACCINI B. (2000) *I laureati dell'Ateneo Fiorentino dell'anno 1998 – Profilo e sbocchi occupazionali*, Università degli Studi di Firenze, Firenze.

BIGGERI L., BINI M. (1999) A multilevel logistic model for the analysis of the Italian universities effectiveness, *Proceedings of the Annual Meeting of the American Statistical Association*, Baltimore (agosto 1999).

BIGGERI L., GRILLI L., BINI M. (2001) The transition from university to work: a multilevel approach to the analysis of the time to obtain the first job, *Journal of the Royal Statistical Society - Series A*, **162(2)**: 293-305.

BINI M., PRATESI M. (2001) Un modello multi-livello per stimare l'efficacia esterna della formazione universitaria con un disegno di campionamento complesso, *Atti del Convegno "Processi e Metodi Statistici di Valutazione"*, Roma, 4-6 giugno 2001, Società Italiana di Statistica.

BINI M. (1999) Valutazione della Efficacia dell'Istruzione Universitaria rispetto al Mercato del Lavoro, RdR 03/99, *Osservatorio per la Valutazione del Sistema Universitario - MURST* – Roma, consultabile anche sul sito www.cnvsu.it/publidoc/comitato/default.asp.

BULGARELLI G. (2002) *I laureati dell'Ateneo Fiorentino dell'anno 1997 – Profilo e sbocchi occupazionali*, Università degli Studi di Firenze, Firenze.

CHIANDOTTO B. (2002) *Profilo e condizione occupazionale dei laureati dell'Ateneo Fiorentino ad uno, due e tre anni dal conseguimento del titolo*, Università degli Studi di Firenze, Firenze.

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, 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, Firenze.

FIELDING A., YANG M., GOLDSTEIN H. (2003) Multilevel ordinal models for examination grades, *Statistical modelling*, vol. 3, n. 2: 127 - 153.

all'università. Nello stesso contesto si colloca il volume di Bini M. (1999): "Valutazione della Efficacia dell'Istruzione Universitaria rispetto al Mercato del Lavoro".

- FOTOUHI A. R. (2004) *Comparisons of estimation procedures for nonlinear multilevel models*, <http://www.jstatsoft.org/v08/i09/Paper.pdf>.
- GOLDSTEIN H. (2003) *Multilevel statistical models*, Arnold Publishers, London.
- HECK R. H., THOMAS S.L. (2000) *An Introduction to Multilevel Modelling Techniques*, LEA, London.
- HEDEKER D. (2004) *Multilevel Models for Ordinal and Nominal Variables*, <http://tigger.uic.edu/~hedeker/ml.html>.
- SINGER J.D. (1998) Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models, *Journal of educational and Behavioral Statistics*, vol.24, n. 2: 323 - 355.
- SNIJDER A.B., BOSKER R. J. (1999) *Multilevel analysis. An introduction to basic and advanced multilevel modelling*, Sage Publications, London.

A multilevel model for analysing the graduate employment

Summary. *The subject of this contribution is an analysis of the employment state of graduates from the University of Florence during the year of 2000, up to one and a half/ two years and a half from the date of graduation. The relative aim is pursued by detecting the possible determining factors of the phenomenon under study at both an individual (such as sex, place of residence, high school diploma, final marks towards the degree) and an institutional (in terms of the distinctive characteristics of the various courses of studies) level. The use of descriptive analyses is preparatory for the application of a model of logistic regression with random intercept having two levels of aggregation, recourse to which is justified by the need to make due allowances for the hierarchical structure of the data that are the subject of the study. In fact, graduates of the University of Florence were considered as a first-level unit, while the respective degree courses were considered as a second-level unit.*

Keywords: *Multilevel models, Logistic regression, Employment opportunities, Probability of employment.*