

Alcuni fattori di rischio per l'insuccesso negli studi universitari*

Francesco Delvecchio, Francesco d'Ovidio
Dipartimento di Scienze Statistiche, Università di Bari

Riassunto. L'osservazione dei dati della popolazione universitaria di Bari mostra, fra i diversi corsi di laurea, una certa differenziazione del rischio di insuccesso negli studi, per motivi legati a caratteristiche degli studenti oppure dei corsi stessi. Per verificare se esistono tali differenze ed a quali cause siano dovute, si è fatto ricorso a modelli *logit*, anche con struttura *multilevel*, che hanno consentito di determinare, per gli studenti dell'Ateneo barese, come fattori primari di rischio alcune covariate a livello studente: età all'immatricolazione, tipo di maturità conseguito e voto di maturità, tuttavia con un'influenza differenziata in base al corso di laurea di iscrizione, fornendo così interessanti spunti di interpretazione.

Parole chiave: Insuccesso negli studi universitari, fattori di rischio, modelli multilevel, modelli *logit*.

1. Premessa

Condizione necessaria per l'analisi dei fattori di rischio di abbandono degli studi universitari è, com'è ovvio, la conoscenza delle caratteristiche dell'universo degli studenti iscritti all'Università, onde stabilire chi sia effettivamente a rischio di abbandono e chi non lo sia. A tale scopo, in altro nostro lavoro¹ sono state determinate le categorie fondamentali di esito degli studi a medio-lungo termine, segmentando adeguatamente la popolazione di riferimento (gli studenti immatricolatisi presso ciascun corso di laurea dell'Università di Bari dall'a.a. 1986-87 all'a.a. 1992-93, così come risultanti dagli archivi dell'Ateneo al 28 maggio 2001): *laureati* in corso e fuori corso,

* Il presente lavoro è stato finanziato nell'ambito del progetto "La ricerca di determinanti del rischio mediante analisi di segmentazione di campioni", cofinanziato dal MIUR; il coordinatore nazionale è Luigi Fabbris, coordinatore del gruppo di Bari è Francesco Delvecchio. Del suddetto lavoro, opera congiunta dei due autori, sono stati redatti da F. Delvecchio i paragrafi da 1 a 3 e le Considerazioni conclusive, da F. d'Ovidio i paragrafi 4 e 5.

¹ F. Delvecchio - F. d'Ovidio, 2002.

studenti tuttora iscritti, trasferiti (ad altri corsi di studi dell'Ateneo barese o in altre Università, sia ad altri corsi di studi che al medesimo corso di immatricolazione), *studenti non più iscritti* ma probabilmente prossimi alla laurea, avendo completato l'intero *cursus* accademico, *studenti non più iscritti con un certo rischio di abbandonare gli studi* in futuro (rischio quantificato in lieve, medio e alto, a seconda del numero di anni di non iscrizione e del numero di annualità mancanti al completamento del corso di studi) ed, infine, *rinunciatarî e decaduti*.

Il rischio di abbandono sembra essere, almeno alla semplice osservazione dei dati, alquanto differenziato fra i diversi corsi di laurea, per motivi forse inerenti alle caratteristiche precipue di detti corsi, oltre che degli studenti che vi si sono iscritti.

Tra l'altro, vi è una precisa gerarchia fra le unità statistiche di rilevazione, poiché gli studenti universitari (unità di primo livello) sono raggruppati in corsi di laurea (unità di secondo livello); a causa di tale gerarchia, è possibile che anche le motivazioni individuali relative alla scelta di un corso di studi si annidino entro motivazioni collettive, legate appunto alle caratteristiche dei corsi. Può quindi essere opportuno utilizzare una metodologia di "analisi multilivello", che è in grado di delineare un corretto quadro di indagine per tali aspetti compositi del problema; tuttavia, la complessità del quadro di riferimento implica l'esistenza di problemi di vario tipo².

Per ridurre alcuni problemi computazionali connessi con l'eccessiva complicazione del modello, è stato necessario determinare innanzitutto la più opportuna categorizzazione delle variabili espresse su scala ordinale o intervallare e, per quanto riguarda le variabili qualitative non dicotomiche, l'eventuale aggregazione ottimale delle modalità categoriche troppo numerose. Queste ultime elaborazioni, tra l'altro, sono generalmente utili anche per procedere con l'analisi degli usuali modelli logit³.

È evidente, poi, che oltre ai problemi computazionali sopra citati, tipici delle metodologie d'analisi iterative, potranno sussistere problemi di interpretazione e gestione dei risultati, in quanto nel modello entrano senz'altro fattori di contesto (ad esempio, le opportunità lavorative), e fattori -assoluti e relazionali- non misurabili, soprattutto a livello di corso di laurea (secondo livello), come le caratteristiche dei

² La *multilevel analysis*, infatti, è una metodologia statistica supportata in modo non del tutto perfetto dalla tecnica computazionale, che ne è nondimeno presupposto fondamentale, essendo detta analisi basata essenzialmente su procedimenti iterativi molto dispendiosi in termini di risorse informatiche, e quindi attuabili solo con sistemi di elaborazione adeguatamente dimensionati; se uno dei problemi di detto tipo di analisi è, infatti, la possibilità della mancata convergenza dei risultati quando la base di dati è poco ampia o le variabili utilizzate sono poco numerose, vi è la contestuale possibilità che il supporto informatico si riveli insufficiente o addirittura si blocchi quando la base di dati è troppo ampia e/o le variabili troppo numerose o troppo informative (difetto, questo, caratteristico del programma da noi utilizzato, *MLwiN* v. 1.2).

³ Allo stato attuale delle conoscenze, tra l'altro, i modelli logit richiedono che sia chiara e univoca l'attribuzione delle variabili utilizzate all'uno o all'altro livello, ma anche che esse siano precisamente definite e determinate, mentre invece accade spesso che tali variabili facciano rilevare diversi aspetti di tipo sfocato, o *fuzzy* (J. J. Hox, 1995), che andrebbero ovviamente affrontati con metodologie adeguate.

docenti, la complicazione dei piani di studio, l'abilità promozionale dei corsi medesimi, ma anche le inclinazioni personali degli studenti nella scelta di un corso anziché di un altro, e così via.

Per questi motivi, abbiamo reputato opportuno far precedere l'analisi da una breve disamina delle variabili espresse al primo livello, allo scopo di determinare il numero e le caratteristiche di quelle effettivamente legate ai fenomeni sotto esame, anche con relazioni di tipo differente per ciascuna delle unità di secondo livello (i corsi di laurea dell'Università di Bari). In seguito, sarà appurato se sia effettivamente necessario il ricorso alla metodologia multilevel, oppure se si possa fare riferimento agli usuali modelli di regressione, che nel caso presente saranno di tipo logistico.

2. I presumibili fattori di rischio rilevati fra i dati disponibili

Nel presente studio si ipotizzerà che, per uno studente universitario, la *probabilità di insuccesso* (ossia di abbandonare gli studi, di essere dichiarato decaduto o di essere classificato “ad alto rischio di abbandono”⁴) sia dipendente da alcuni fattori caratteristici della popolazione studentesca (variabili “di primo livello”) e/o da fattori legati al corso di laurea scelto (variabili “di secondo livello”); la *baseline* rispetto a cui verificare tale ipotesi è costituita dalla condizione opposta e complementare (cioè quella di *non insuccesso*), che non è apertamente di successo, poiché a fianco ai laureati vengono posti sia gli studenti tuttora iscritti fuori corso (al corso di originaria immatricolazione o in un altro corso dell'Ateneo barese), sia un certo numero di studenti non più iscritti, i quali comunque non sembrano avere, nel complesso, le caratteristiche di carente preparazione universitaria fatta rilevare dai rinunciatari (fra i quali, d'ora in poi, si intenderanno sempre compresi gli studenti decaduti dagli studi) e dagli studenti ad alto rischio di abbandono.

Per quanto riguarda i presumibili fattori di rischio, quelli individuati fra i dati disponibili sono:

- il sesso dello studente;
- la sua età all'immatricolazione;
- il tipo di diploma di istruzione secondaria a suo tempo conseguito;
- il voto di maturità;
- la percentuale di annualità superate rispetto al totale di quelle previste per il corso di laurea di appartenenza: variabile, quest'ultima, che può essere utile anche per

⁴ Studenti non più iscritti da oltre tre anni accademici, che non abbiano superato esami negli ultimi cinque anni ed ai quali manchino almeno cinque annualità per completare il corso degli studi. Le caratteristiche del percorso universitario di tale categoria di studenti, in effetti, sono molto simili a quelle degli effettivi rinunciatari (F. Delvecchio – F. d'Ovidio, 2002).

isolare la massa di coloro che si sono iscritti per motivi meramente strumentali, senza sostenere alcun esame e spesso rinunciando agli studi l'anno successivo a quello di immatricolazione.

L'impianto della ricerca qui effettuata sarà di tipo campionario, a partire dalla base di dati fornita dal Centro Informativo dell'Università di Bari, sia a causa della natura essenzialmente inferenziale delle analisi statistiche che ne sono alla base, e sia per trascendere il limite computazionale accennato in premessa⁵.

L'analisi cercherà di individuare, innanzitutto, le relazioni che possono sussistere, per gli immatricolati ad ognuno dei corsi di laurea presenti nell'Ateneo barese, fra l'esito degli studi e le caratteristiche degli studenti e del corso di laurea, con particolare attenzione all'abbandono effettivo o implicito. Dal complesso degli immatricolati sono stati esclusi, ovviamente, coloro per i quali non si ha alcuna notizia sull'esito degli studi, ossia i trasferiti ad altro Ateneo; dall'universo risultante è stato quindi estratto, con procedura proporzionale stratificata per corso di laurea e per esito degli studi, un campione casuale al 6% della numerosità complessiva dei corsi (pari a 76.364 studenti), con il vincolo di un numero minimo di unità da campionare, pari a 50 per corso di laurea⁶ (allo scopo di rendere più affidabili le stime relative ai subcampioni estratti dai c.d.l. numericamente esigui), cosicché il campione ammonterà, complessivamente, a 4.750 unità.

Le numerosità dei subcampioni estratti dai vari strati sono riportate nella Tab.1. Come appare evidente, la quota di rinunciatari, decaduti e "rinunciatari potenziali" è molto elevata nel campione, ammontando ad oltre il 54%; si osserva, inoltre, una certa variabilità, posto che la quota di studenti che abbandonano o sono considerati "a forte rischio di abbandono" va dal minimo del 25,1%, per Medicina e Chirurgia, ai massimi livelli rilevati per Scienze Politiche (75,5%) e Scienze Ambientali (78,0%).

Data la sua elevata numerosità, tale campione è presumibilmente ben rappresentativo della realtà universitaria barese del periodo in esame.

A questo punto occorre definire come saranno strutturate le covariate da utilizzare nel modello. In altri termini, è necessario sceglierne la classificazione più adeguata, così da ottenere il massimo dettaglio informativo con il minore utilizzo di risorse.

Non vi sono problemi per quanto riguarda le variabili categoriche: il sesso o l'aver sostenuto esami sono, infatti, variabili dicotomiche, mentre il tipo di scuola frequentato, in base a quanto risulta da altri studi (ad es., F. Delvecchio -N. Tedesco,

⁵ Infatti, per quanto si possano semplificare i modelli, riducendo numero e modalità delle variabili in essi comprese, gli algoritmi iterativi alla base dell'analisi, lavorando con decine di migliaia di dati, potrebbero saturare la memoria dell'elaboratore e, di conseguenza, mandare in stallo il programma.

⁶ Detto vincolo di numerosità sufficiente ha operato per i corsi di laurea in Scienze Forestali ed Ambientali, Scienze Statistiche ed Economiche, Chimica e Tecnologie Farmaceutiche, Chimica, Fisica, Scienze Ambientali, Scienze Naturali. La distorsione introdotta da tale accorgimento nell'efficienza delle stime può essere ritenuta irrilevante, dato l'esiguo numero di unità aggiunte rispetto alla numerosità globale.

Tab. 1 – Numerosità dei subcampioni estratti dagli strati. Percentuali di insuccesso nei vari corsi di laurea e nei subcampioni estratti dagli stessi.

Corso di laurea di immatricolazione	Numerosità dei subcampioni		Percentuali di insuccesso	
	Insuccesso	Non insuccesso	Sub- popolazioni	Sub- campioni
Scienze Agrarie	44	25	63,2	63,8
Scienze Forestali e Ambientali	32	18	64,0	64,0
Economia e Commercio	471	419	52,9	52,9
Scienze Statistiche ed Economiche	27	23	53,6	54,0
Farmacia	32	40	44,9	44,4
Chimica e Tecnologie farmaceutiche	22	28	44,0	44,0
Giurisprudenza	673	636	51,4	51,4
Lettere	106	138	43,5	43,4
Filosofia	28	25	52,3	52,8
Lingue e Letterature straniere	140	145	49,0	49,1
Medicina e Chirurgia	42	125	25,3	25,1
Medicina Veterinaria	22	28	44,1	44,0
Pedagogia (Sc. Formaz. Primaria)	106	67	61,4	61,3
Magistero (Scienze Educazione)	63	44	58,7	58,9
Chimica	22	28	44,4	44,0
Fisica	23	27	45,2	46,0
Informatica	130	92	58,5	58,6
Matematica	27	27	50,6	50,0
Scienze Ambientali	39	11	78,1	78,0
Scienze Biologiche	84	71	54,4	54,2
Scienze Geologiche	39	18	68,4	68,4
Scienze Naturali	35	15	69,5	70,0
Scienze Politiche	372	121	75,5	75,5
Intero campione	2579	2171	54,1	54,3

1997), è comunque un carattere da mantenere nel massimo dettaglio possibile: ragioni per cui si è effettuata (anche per uniformità con quanto fatto dal C.S.I. dell'Ateneo nei riguardi della Maturità Professionale) la sola aggregazione fra Maturità Tecnica Industriale e Maturità Tecnica Agraria, il cui contributo all'ampiezza dell'universo è comunque minoritario (circa il 2%); si aggiunga, inoltre, che circa la metà degli studenti in possesso di quest'ultimo titolo di studio si concentra, com'era prevedibile, nei corsi di laurea in Scienze Agrarie ed in Scienze Forestali.

Per quanto riguarda le variabili discrete o continue prese in esame (età all'immatricolazione, voto di maturità, percentuale di esami di profitto sostenuti), dopo una breve analisi esplorativa diretta alla massimizzazione delle differenze fra percentuali di insuccesso dei dati dell'universo, si è ritenuto opportuno definire i seguenti

intervalli: per l'età all'immatricolazione, meno di 20 anni compiuti e da 20 anni in poi; per il voto di maturità, fra 36 e 44 sessantesimi, fra 45 e 53 sessantesimi e fra 54 e 60 sessantesimi; per quanto riguarda la percentuale degli esami di profitto, infine, si è posto: nessun esame sostenuto, da un esame fino al 10% delle annualità previste dal corso di studi, da più del 10% al 30%, da più del 30% al 50%, oltre il 50%.

3. Definizione del modello

Per verificare l'ipotesi di indipendenza del fenomeno dell'insuccesso universitario dai vari fattori su definiti, va innanzitutto ricordato che la variabile risposta è una dicotomica (in cui si è posto il *non insuccesso* come *baseline*=0 e l'*insuccesso*=1); per cui, se si indica con p_{ij} la probabilità che abbia insuccesso un individuo che si trovi *i.ma* condizione di primo livello e nella *j.ma* condizione di secondo livello, con n_{ij} il numero delle unità campionate (senza ripetizione) che si trovano nella condizione *ij* e con f_{ij} la proporzione di dette unità che presentano insuccesso, è noto che f_{ij} si distribuisce secondo una ipergeometrica; è noto altresì che, allorché la numerosità campionaria è inferiore alla decima parte dell'universo (come nel caso presente), l'ipergeometrica si approssima ad una binomiale di parametri p_{ij} ed n_{ij} , con valore atteso e varianza, rispettivamente, pari a

$$E(f_{ij}) = p_{ij} \quad \text{e} \quad \text{Var}(f_{ij}) = p_{ij}(1-p_{ij})/n_{ij}.$$

Poiché si dimostra⁷ che il *logit* $f_{ij} = \ln[f_{ij}/(1-f_{ij})]$ è distribuito approssimativamente secondo una normale, è senz'altro possibile, fatte salve le altre condizioni alla base del GLM, utilizzare un'equazione di regressione lineare per spiegare e prevedere *logit* p_{ij} . Va sottolineato, al proposito, che i modelli *logit* sono perfettamente idonei a verificare se ed in quale misura le covariate (e, nel caso presente, anche l'appartenenza ai diversi gruppi, ossia l'essersi immatricolati ai diversi corsi di laurea) influenzino il comportamento dello studente in tema di abbandono degli studi: ciò perché, com'è evidente, la quantità $p_{ij}/(1-p_{ij})$ esprime il rapporto fra la probabilità che uno studente che si trova nella condizione *i* di primo livello e *j* di secondo livello abbandoni gli studi (o sia "ad alto rischio di abbandono") e la probabilità che si trovi in condizione diversa (laureato, ancora iscritto, non più iscritto ma con rischio di abbandono lieve o al più medio).

Nel presente lavoro, è opportuno verificare per prima cosa se, fra i valori rilevati nei diversi corsi di laurea, vi siano differenze sistematiche o se le variazioni possono essere considerate casuali. Allo scopo, si farà riferimento ad un *test di eterogeneità* (D. Commenges - H. Jacqmin, 1994), che per modalità dicotomiche è formulato come

⁷ Cfr. P. Mc Cullagh - J. A. Nelder, 1989.

nella formula seguente⁸:

$$C = \frac{\sum^J [n_j^2 (f_j - \hat{p})^2] - n\hat{p}(1-\hat{p})}{\hat{p}(1-\hat{p})\sqrt{2\sum^J n_j(n_j-1)}}$$

in cui J è il numero di gruppi, n_j è la numerosità del j -mo gruppo, f_j è la frequenza degli insuccessi nel j -mo gruppo (prescindendo dalla condizione in cui i soggetti si trovano nel primo livello), n è la numerosità campionaria complessiva e \hat{p} è la frequenza degli insuccessi nell'intero campione, assunta come stima della frequenza degli insuccessi nell'universo. Il valore empirico del test può essere confrontato, per grandi campioni come quello qui esaminato, con il valore soglia fornito, al prefissato livello di significatività α , dalla v.c. normale standardizzata.

Nel caso in cui non risulti significativa la differenza fra i vari corsi di laurea, per valutare l'influenza delle covariate di primo livello sulla probabilità di abbandono degli studi, si potrà utilizzare un semplice modello *logit* a risposta dicotomica con variabili esplicative categoriche, la cui formulazione generale, com'è noto, è⁹:

$$\text{logit } p_{ij} = \beta'X,$$

dove β rappresenta il vettore dei coefficienti delle variabili introdotte nel modello, comprese le eventuali interazioni e le variabili *dummy* inserite.

Il test di eterogeneità suddetto, nel caso presente, fornisce un valore di 24,782, che è di gran lunga superiore al valore soglia $z_{0,005}=2,57583$ al livello $\alpha=0,01$ (per test bilaterale): ciò equivale ad affermare che è verosimile che sull'esito degli studi vi sia una cospicua influenza del corso di laurea di prima iscrizione. Il valore empirico del test decresce, poi, man mano che vengono esclusi dal campione i corsi di laurea con le percentuali di insuccesso estreme: da un lato Medicina e Chirurgia, con la più bassa quota di insuccessi, dall'altro Scienze Ambientali e Scienze Politiche, con le percentuali di insuccesso più elevate. Escludendo detti corsi di laurea, si ottiene un valore empirico inferiore al valore soglia: per i rimanenti corsi di laurea si può, quindi, accettare l'ipotesi di base (nessun effetto gruppo). Ciò induce ad ipotizzare che si sia soltanto in presenza di un'influenza dovuta all'azione degli opposti *outliers*.

Allo scopo di valutare comunque l'importanza dell'effetto gruppo qui individuato, si esaminerà innanzitutto un modello "a componenti di varianza", con sola intercetta¹⁰. Detto modello, nel caso di funzione *logit*, assume la semplice forma

⁸ È qui parzialmente ripresa la notazione utilizzata in T. Snijders - R. Bosker, 1999, anziché la formulazione originale, più generale ma meno immediata.

⁹ Cfr., ad es., F. Delvecchio, 2002.

¹⁰ Il concetto di modello "a componenti di varianza", invero, deriva direttamente dal modello a sola intercetta, che nella sua forma generale è $y_{ij} = \gamma_{00} + \eta_{0j} + \varepsilon_{ij}$: infatti, essendo γ_{00} costante, da tale forma si ricava che $\text{Var}(Y_{ij}) = \sigma^2(\varepsilon) + \sigma^2(\eta)$, ossia la varianza della variabile risposta si scompone nella somma delle varianze degli errori di primo e secondo livello.

$$\text{logit } p_{ij} = \gamma_{00} + \eta_{0j},$$

in cui i valori η_{0j} rappresentano l'errore casuale relativo all'effetto di gruppo¹¹, mentre γ_{00} è una costante che indica, appunto, l'effetto medio dovuto al corso di laurea di prima iscrizione.

Il modello con sola intercetta assume, a partire dai dati del campione precedentemente descritto, la forma

$$\text{logit } p_{ij} = 0,184 + \eta_{0j},$$

con una deviazione standard della stima dell'intercetta $\hat{\sigma}(\hat{\gamma}_{00})$ pari a 0,109: si può quindi affermare che l'effetto medio del modello è abbastanza ridotto.

La stima della varianza dell'errore di 2° livello è $\hat{\sigma}^2(\eta) = 0,224$, con deviazione standard pari a 0,080. La verifica statistica della significatività di detta stima potrebbe essere effettuata tramite il test di Wald¹², che però non è adeguato nel caso di stime di varianze, la cui distribuzione campionaria è notoriamente asimmetrica (cfr., ad es., A. S. Bryk - S. W. Raudenbush, 1992); nondimeno, tale argomento è superabile se il programma utilizzato per l'elaborazione dei dati fornisce stime robuste degli errori standardizzati, come, ad esempio, la modifica "sandwich" proposta da G. Verbeke - E. Lesaffre (1997). In effetti, molti dei più recenti pacchetti statistici dedicati all'analisi multilevel (come il programma *MLwiN*, utilizzato per il presente studio) sono appunto strutturati in modo da calcolare stime robuste anche in assenza di normalità delle distribuzioni. Sulla base di tale assunto, la stima della varianza dell'errore di 2° livello risulta significativa.

Il coefficiente di correlazione intraclasse, costruito rapportando detta stima alla varianza d'errore complessiva

$$\hat{\rho}_w = \frac{\hat{\sigma}^2(\eta)}{\hat{\sigma}^2(\varepsilon) + \hat{\sigma}^2(\eta)},$$

¹¹ Nel modello, che considera la *probabilità* p_{ij} di abbandonare gli studi e non la *frequenza relativa osservata* f_{ij} , non sono ovviamente riportati gli errori ε_{ij} di primo livello che, invece, sono parte integrante del modello multilivello generale (cfr., ad es., H. Goldstein, 1995).

¹² Il test di Wald, com'è noto, è fornito da $Z = (\text{stima}) / (\text{deviazione standard della stima})$, dove il valore empirico Z viene comparato con il valore soglia, al livello α , della distribuzione normale standardizzata. Tuttavia, è stato osservato da A. S. Bryk e S. W. Raudenbush (1992) che, in caso di regressori fissi di primo livello, è più opportuno comparare tale rapporto con il valore soglia di una distribuzione T con $J-K-1$ gradi di libertà, ove J è il numero di unità di secondo livello campionate e K il numero di covariate relative a tali unità (nel caso qui esaminato si ha $K=0$, per cui $g=J-1$). In fase esplorativa, ad ogni buon conto, si può utilizzare un criterio più immediato, interpretando come "statisticamente significative" le stime superiori, in valore assoluto, a circa il doppio del loro errore standard (essendo $z_{0,025} = 1,96$); tale criterio è giustificato dall'osservazione che detti errori standard sono asintotici, per cui nella pratica dell'indagine campionaria, se la numerosità è elevata (come nel caso presente), essi risultano delle buone approssimazioni ai valori della v.c. normale standardizzata (cfr. J. J. Hox, 1995).

che esprime quanta parte della varianza totale è spiegata dall'effetto gruppo¹³, ha comunque un valore non elevato ($\hat{\rho}_w = 0,183$).

Sintetizzando le precedenti osservazioni, si può affermare che, verosimilmente, sull'insuccesso negli studi universitari è poco rilevante l'effetto medio dell'essersi immatricolato ai diversi corsi di laurea, ma che vi è una significativa variabilità fra un corso e l'altro, in quanto circa il 18% della variabilità complessiva della distribuzione può essere associata all'*effetto gruppo*. È possibile, tra l'altro, che alcune variabili "a livello studente" abbiano un certo effetto in alcuni corsi ed un effetto diverso in altri.

In particolare, la precedente verifica con il test di eterogeneità ha fatto notare che si è in presenza di tre valori estremi molto differenziati dalla media: il corso di laurea in Medicina e Chirurgia e quelli in Scienze Ambientali e Scienze Politiche. Tramite le opzioni del programma *MLwiN* è possibile, tuttavia, verificare immediatamente l'influenza di tali *outliers* sulla variabilità complessiva, semplicemente escludendoli da una successiva elaborazione.

Escludendo i tre anzidetti corsi di laurea, il modello assume la forma *logit* $p_{ij} = 0,143 + \eta_{0j}$, con una deviazione standard del parametro intercetta pari a 0,067; la stima della varianza dell'errore di 2° livello è $\hat{\sigma}^2(\eta) = 0,048$, statisticamente non significativa (la deviazione standard di detta stima è, infatti, pari a 0,027). Il coefficiente di correlazione intraclasse, inoltre, è $\hat{\rho}_w = 0,046$, il che significa che l'influenza dell'effetto gruppo sulla variabilità complessiva, escludendo i due *outlier*, è quasi impercettibile.

Di conseguenza, una ipotesi di lavoro validamente alternativa all'analisi multi-level, per superare eventuali difficoltà computazionali o per chiarire alcuni aspetti particolari rilevati con la stessa, potrebbe essere quella di utilizzare semplicemente modelli *logit* diversi: uno per l'*outlier* positivo (Medicina e Chirurgia), uno per gli *outliers* negativi (Scienze Ambientali e Scienze Politiche) ed uno per l'insieme degli altri corsi di laurea del campione.

Proseguendo nell'analisi multilevel, a questo punto, si è ritenuto opportuno fare riferimento ad un semplice modello con H regressori fissi di primo livello e nessuno di secondo livello (definibile anch'esso "a componenti di varianza"):

$$\text{logit } p_{ij} = \gamma_{00} + \sum_{h=1}^H \gamma_{h0} X_{hij} + \eta_{0j} .$$

¹³ Si fa presente che, trattandosi di una distribuzione (binomiale) che nella relazione fra i valori osservati e quelli attesi $f_{ij} = p_{ij} + \varepsilon_{ij}$ comprende gli errori di primo livello, affinché sia effettivamente $\text{Var}(f_{ij} | p_{ij}) = p_{ij}(1-p_{ij})/n_{ij}$ si include la variazione di 1° livello nella forma $\varepsilon_{ij} \sqrt{p_{ij}(1-p_{ij})/n_{ij}}$ e si fissa $\sigma^2(\varepsilon) = 1$, per cui la stima del coefficiente di correlazione intraclasse diviene $\hat{\rho}_w = \hat{\sigma}^2(\eta) / (1 + \hat{\sigma}^2(\eta))$.

In seguito, sulla base dei risultati ottenuti da detto modello, si potrà scegliere se passare o no ad un modello con regressori a coefficienti casuali, individuando quali delle modalità significative possano variare il proprio effetto nei diversi corsi di laurea.

4. Risultati dell'analisi multilevel

Per rendere evidenti le categorie che fanno rilevare livelli meno elevati di rinunce o di situazioni "ad alto rischio", ossia quelle da considerare *baseline*, riportiamo nella Tab. 2 la distribuzione degli insuccessi secondo le modalità dei possibili fattori di rischio.

Il primo modello multilevel a regressori fissi utilizza quindi, come regressori di primo livello, tutte le covariate definite a livello di studente (in forma dicotomica: 0=no, 1=sì): x_1 =*sex*(M), x_2 =*età*(≥ 20 anni), x_3 =*voto maturità* [36-44], x_4 =*voto maturità* [45-53], x_5 =*nessun esame superato*, x_6 =*% annualità superate* $\in]0-10]$,

Tab. 2 – Numero e percentuali di insuccessi nel campione, secondo le modalità dei possibili fattori di rischio.

<i>Caratteristiche</i>		Insuccesso	Non insuccesso	% di insuccesso
<i>Sesso</i>	Femmina	1269	1312	49,2
	Maschio	1310	859	60,4
<i>Età di immatricolaz.</i>	Meno di 20 anni	1180	1789	39,7
	Da 20 anni in poi	1399	382	78,6
<i>Tipo di maturità</i>	Classica	244	519	32,0
	Scientifica	407	628	39,3
	Linguistica	74	77	49,0
	Magistrale	276	190	59,2
	Tecnica commerciale	835	497	62,7
	Tecnica geometri	114	30	79,2
	Tecnica industriale	314	118	72,7
	Profess. commercio	115	31	78,8
	Profess. industriale	151	53	74,0
	Altro titolo	49	28	63,6
<i>Voto di maturità</i>	54-60 / 60	290	551	34,5
	45-53 / 60	681	779	46,6
	36-44 / 60	1608	841	65,7
<i>% annualità superate</i>	Oltre il 50%	14	1921	0,7
]30 , 50]	50	129	27,9
]10 , 30]	211	71	74,8
]0 , 10]	395	31	92,7
	Nessuna	1909	19	99,0
<i>In complesso</i>		2579	2171	54,3

x_7 = % annualità superate $\in]10-30]$, x_8 = % annualità superate $\in]30-50]$, x_9 =maturità scientifica, x_{10} =maturità linguistica, x_{11} =maturità magistrale, x_{12} =matur. tecnica commerciale, x_{13} =matur. tecnica geometri, x_{14} =matur. tecnica industriale, x_{15} =matur. Professionale per il commercio, x_{16} =matur. professionale per l'industria, x_{16} =altro tipo di maturità, e nessun regressore di 2° livello, individuando così una *baseline* ove i soggetti sono di sesso femminile, con età all'immatricolazione inferiore a 20 anni, provenienti dal liceo classico, con voto di maturità non minore di 54/60 e che, infine, hanno superato più del 50% delle annualità previste dal proprio corso di studi.

Applicando la procedura di iterazione (con programma *MlwiN* e metodo di Massima Verosimiglianza Limitata) sono state ottenute le stime della Tab.3.

Come è agevole notare, in questo modello il contributo maggiore alla probabilità di insuccesso è fornito dalla percentuale di annualità sostenute, con la preponderante importanza associata alla modalità “nessun esame sostenuto”, e valori via via inferiori al crescere della percentuale suddetta. Detto risultato, per quanto interessante ai fini della descrizione del fenomeno (soprattutto per quanto riguarda gli eccezionali contributi agli *odds* corrispondenti alle varie percentuali), è tuttavia quasi totalmente non esplicativo ai fini dell'analisi dei fattori di rischio: è logico, infatti, che il più rilevante fattore di rischio di abbandono degli studi universitari sia quello di non riuscire

Tab.3 – Stime dei coefficienti del modello multilevel completo, con tutti i regressori fissi di primo livello. Verifica della loro significatività ($t_{22; 0,025}=2,07388$).

Variabili esplicative	Stima coefficiente	Deviazione standard	T	$\exp(\hat{\gamma}_{h0})$
Intercetta	-5,668	0,421	-13,46	0,003
Sesso maschile	0,223	0,196	1,14	1,250
Età ≥ 20 anni	0,148	0,195	0,76	1,160
Voto maturità 45-53	-0,026	0,268	-0,10	0,974
Voto maturità 36-44	-0,234	0,261	-0,90	0,791
% ann.superate [30 , 50]	4,146	0,330	12,56	63,181
]10 , 30]	6,245	0,321	19,45	515,429
]0 , 10]	7,709	0,350	22,03	2228,313
Nessun esame sostenuto	9,644	0,382	25,25	15428,936
Maturità scientifica	0,442	0,276	1,60	1,556
Maturità linguistica	1,537	0,556	2,76	4,651
Maturità magistrale	0,799	0,353	2,26	2,223
Mat. Tecn. commerciale	0,846	0,300	2,82	2,330
Mat. Tecnica geometri	0,631	0,761	0,83	1,879
Mat. Tecnica industriale	0,635	0,389	1,63	1,887
Mat. Profess. commercio	1,061	0,546	1,94	2,889
Mat. Profess. industriale	0,695	0,473	1,47	2,004
Altro titolo	0,665	0,614	1,08	1,944
$\hat{\sigma}^2(\eta)$	0,226	0,150	1,51	

o non essere interessato a superare esami; la precedente Tab. 2, invero, ha posto in luce che quasi tutti coloro che non hanno superato alcun esame nel corso degli anni rientrano fra i rinunciatari o fra gli studenti “ad alto rischio di abbandono”, e la situazione non è molto differente fra coloro che hanno superato non oltre il 10% degli esami del proprio corso. Comunque, potrebbe essere utile indagare ulteriormente intorno a tali aspetti, nell’ottica di quantificare, ad esempio, l’influenza del numero medio di esami superati annualmente sul rischio di abbandono degli studi universitari.

Un modello senz’altro più adeguato agli scopi della presente ricerca è quello che fa riferimento ancora una volta ai soli regressori fissi di primo livello su conside-

Tab.4 – Stime dei coefficienti del modello multilevel comprensivo delle interazioni significative fra i regressori fissi di primo livello. Verifica della loro significatività statistica ($t_{22; 0,025}=2,07388$).

Variabili esplicative	Stima coefficiente	Deviazione standard	T	$\exp(\hat{\gamma}_{h0})$
Intercetta	-2,111	0,142	-14,87	0,121
Sesso maschile	0,104	0,078	1,33	1,110
Età ≥ 20 anni	2,557	0,229	11,17	12,897
Voto maturità 36-44 sessantesimi	1,291	0,114	11,32	3,636
Voto maturità 45-53 sessantesimi	0,565	0,118	4,79	1,759
Maturità scientifica	0,263	0,116	2,27	1,301
Maturità linguistica	0,923	0,211	4,37	2,517
Maturità magistrale	1,411	0,163	8,66	4,100
Maturità Tecnica commerciale	1,382	0,127	10,88	3,983
Maturità Tecnica geometri	1,922	0,338	5,69	6,835
Maturità Tecnica industriale	1,519	0,196	7,75	4,568
Maturità Professionale commercio	2,511	0,306	8,21	12,317
Maturità Professionale per l' industria	2,075	0,260	7,98	7,965
Altro titolo	1,569	0,366	4,29	4,802
<i>Interazioni</i>				
Sesso M \cap Maturità magistrale	-0,901	0,352	-2,56	0,406
Età $\geq 20 \cap$ Voto maturità 36-44	-0,929	0,225	-4,13	0,395
Età $\geq 20 \cap$ Voto maturità 45-53	-0,551	0,243	-2,27	0,576
Età $\geq 20 \cap$ Maturità magistrale	-0,811	0,252	-3,22	0,444
Età $\geq 20 \cap$ Mat. Tecn. commerciale	-0,589	0,184	-3,20	0,555
Età $\geq 20 \cap$ Mat. Tecnica geometri	-1,102	0,444	-2,48	0,332
Età $\geq 20 \cap$ Mat. Tecnica industriale	-0,786	0,259	-3,03	0,456
Età $\geq 20 \cap$ Mat. Profess. commercio	-1,491	0,438	-3,40	0,225
Età $\geq 20 \cap$ Mat. Profess. industriale	-1,475	0,352	-4,19	0,229
Età $\geq 20 \cap$ Altro titolo	-1,467	0,503	-2,92	0,231
$\hat{\sigma}^2(\eta)$	0,036	0,022	1,64	

Tab. 5 – Distribuzione degli studenti del campione per esito degli studi, sesso e tipo di maturità conseguito.

Tipo di maturità	Esito degli studi secondo il sesso						% insuccessi per sesso		
	Insuccesso			Non insuccesso					
	<i>Sesso</i>	<i>F</i>	<i>M</i>	<i>MF</i>	<i>F</i>	<i>M</i>	<i>MF</i>	<i>F</i>	<i>M</i>
Classica	147	97	244	351	168	519	29,5	36,6	32,0
Scientifica	200	207	407	341	287	628	37,0	41,9	39,3
Linguistica	65	9	74	73	4	77	47,1	69,2	49,0
Magistrale	253	23	276	169	21	190	60,0	52,3	59,2
Tecnico Commerciale	374	461	835	277	220	497	57,5	67,7	62,7
Tecnico Geometri	8	106	114	2	28	30	80,0	79,1	79,2
Tecnico Industriale	38	276	314	28	90	118	57,6	75,4	72,7
Profession. Commercio	94	21	115	28	3	31	77,0	87,5	78,8
Professionale Industria	59	92	151	27	26	53	68,6	78,0	74,0
Altro tipo	31	18	49	16	12	28	66,0	60,0	63,6
Campione	1269	1310	2579	1312	859	2171	49,2	60,4	54,3

rati, escludendo però la percentuale di annualità superate, e tenendo presenti anche le interazioni fra i regressori medesimi.

Dopo varie elaborazioni, il modello risulta quello riportato nella Tab. 4. Come è evidente, il coefficiente relativo al sesso non risulta significativo, ma va lasciato nel modello a causa della significatività della sua interazione con la variabile “maturità magistrale”. Detta interazione, evidenziata da un coefficiente negativo pari a $-0,901$, può essere spiegata osservando la Tab. 5, in cui si pone in luce che la percentuale di insuccessi degli studenti provenienti dalle altre scuole è sempre superiore a quelle degli studenti provenienti dal Liceo Classico, per ambo i sessi; per gli studenti in possesso di Maturità Magistrale, tuttavia, detta percentuale è sensibilmente maggiore per le femmine che per i maschi. Lo stesso fenomeno si ripropone anche per gli studenti in possesso di “altro titolo” e, in misura minima, per i diplomati Geometri, ma queste ultime differenze non risultano statisticamente significative.

Sono molto interessanti anche alcune altre relazioni poste in luce dal modello: innanzitutto, l'età di immatricolazione è un fattore di rischio primario (chi si immatricula in età di 20 anni e più fa rilevare un rischio di insuccesso quasi 13 volte superiore agli altri studenti); ciò può essere legato ad una molteplicità di fattori, fra cui, presumibilmente, una contemporanea attività lavorativa degli studenti più anziani (fattore che, ovviamente, riduce le possibilità di frequenza e di studio), oppure, per i maschi, l'aver prima ottemperato agli obblighi di leva; né si deve dimenticare che, se la tardiva immatricolazione è dovuta ad un altrettanto tardivo conseguimento della Maturità, ciò può configurare una effettiva difficoltà di inserimento nel processo culturale da parte dello studente.

Tab. 6 – Distribuzione degli studenti del campione per esito degli studi, classe d'età e tipo di maturità conseguito.

Modalità dei caratteri	Esito degli studi secondo l'età						% insuccessi per età		
	Insuccesso			Non insuccesso					
	Età <20	≥20	Totale	Età <20	≥20	Totale	Età <20	≥20	Totale
<i>Voto di maturità</i>									
54 - 60 sessant.	164	126	290	508	43	551	24,4	74,6	34,5
45 - 53 sessant.	363	318	681	674	105	779	35,0	75,2	46,6
meno di 45/60	653	955	1608	607	234	841	51,8	80,3	65,7
<i>Tipo di maturità</i>									
Mat. classica	115	129	244	468	51	519	19,7	71,7	32,0
" scientifica	211	196	407	553	75	628	27,6	72,3	39,3
" linguistica	43	31	74	68	9	77	38,7	77,5	49,0
" magistrale	146	130	276	147	43	190	49,8	75,1	59,2
" tec. Comm.	409	426	835	406	91	497	50,2	82,4	62,7
" tec. Geometri	33	81	114	15	15	30	68,8	84,4	79,2
" tec. Industr.	97	217	314	71	47	118	57,7	82,2	72,7
" prof. Comm.	47	68	115	18	13	31	72,3	84,0	78,8
" prof. Industr.	58	93	151	28	25	53	67,4	78,8	74,0
" altro tipo	21	28	49	15	13	28	58,3	68,3	63,6
Campione	1180	1399	2579	1789	382	2171	39,7	78,6	54,3

Un ulteriore fattore di rischio evidenziato nella Tab. 4 è il voto di maturità meno che ottimo: infatti, chi si diploma con voti compresi fra 45 e 53 sessantesimi ha una probabilità di insuccesso quasi doppia rispetto a quella di chi si è diplomato con una votazione maggiore ($\exp(\hat{\gamma}_{h_0})=1,759$); chi si diploma con voti ancor più bassi, poi, ha una probabilità di insuccesso molto più che tripla rispetto alla *baseline* ($\exp(\hat{\gamma}_{h_0})=3,636$).

La Tab. 4 pone in luce, inoltre, che vi è interazione fra età di immatricolazione e voto conseguito agli esami di maturità. Per chiarirne il significato, si osservi (Tab. 6) che la percentuale di insuccessi aumenta al diminuire del voto di maturità, ed anche che è superiore per chi si è immatricolato in età pari o maggiore di 20 anni; tuttavia, nel passare dalla prima alla seconda classe d'età, l'incremento della quota di insuccessi che compete agli studenti con voto di maturità più basso è sensibilmente minore di quello della classe di voto di riferimento (54-60 sessantesimi): ciò produce senz'altro un coefficiente negativo per l'interazione.

È molto ben definita anche la relazione che lega la probabilità di insuccesso al tipo di diploma conseguito: in particolare, la probabilità di insuccesso si incrementa

da quattro a sei volte per chi proviene da istituti tecnici rispetto a chi proviene dal Liceo Classico (*baseline*), ed addirittura di 12 volte per chi proviene dal Professionale per il Commercio. Non si può dimenticare, d'altronde, che il tipo di formazione posto in atto negli istituti Professionali è generalmente orientato alle discipline tecniche nei loro aspetti empirici, con pochi approfondimenti teorici (pur se vi sono eccezioni, soprattutto negli indirizzi informatici del Professionale per l'Industria).

Ancora dalla Tab. 4 si evince, inoltre, che risultano statisticamente significativi tutti i coefficienti relativi al tipo di maturità conseguita, ma è poco rilevante l'incremento di probabilità di insuccesso per chi proviene dal Liceo Scientifico invece che dal Classico: quest'ultima osservazione ben si concilia sia con l'opinione comune, che ritiene quasi equivalenti i due tipi di scuola, sia con i risultati di altri studi (cfr., ad es., F. Delvecchio – N. Tedesco, 1997).

L'interazione fra età di immatricolazione e tipo di maturità presenta *odds* inferiori all'unità per quasi tutti i titoli (escludendo soltanto la Maturità Scientifica e quella Linguistica, per le quali non si rilevano interazioni significative); ancora una volta, però, tali bassi valori non vanno interpretati come una diminuzione del rischio di insuccesso, ma come diminuzione del rapporto di rischio, dovuto all'aumento del rischio anche per la categoria posta a denominatore del rapporto. Ad esempio, per la Maturità Classica l'incremento relativo delle percentuali di insuccesso, nel passare dalla prima alla seconda classe d'età, è pari a $71,7/19,7=3,6396$, mentre per la Maturità Magistrale è $75,1/49,8=1,508$; l'*odds ratio* è, quindi, $1,508/3,6396=0,414$. Tale valore, così come gli altri che si possono calcolare a partire dalla Tab. 6, è molto vicino a quello fornito dalla procedura di analisi multilivello, cosicché le differenze possono essere agevolmente riconducibili alla complessità del modello (che considera tutte le interazioni nel loro complesso, oltre alla variabilità di 2° livello).

L'osservazione più interessante che si evince dal modello con regressori fissi di 1° livello riportato nella Tab. 4, però, è relativa alla caduta della variabilità dell'errore di 2° livello $\hat{\sigma}^2(\eta)$: essa, infatti, è pari a 0,036, valore che, dato l'errore standard di 0,022, è non significativo; detto valore, inoltre, è molto vicino alla stima $\hat{\sigma}^2(\eta)=0,048$ del modello con sola intercetta calcolato per 20 corsi di laurea (senza i tre corsi *outliers*), e del tutto diverso dalla variabilità di 2° livello del modello con sola intercetta calcolato per tutti i 23 corsi rappresentati nel campione ($\hat{\sigma}^2(\eta)=0,224$). Ciò sembra implicare che alcune delle variabili utilizzate come regressori in tale modello siano in grado di spiegare gran parte dell'effetto gruppo prima verificato, effetto che quindi può ritenersi connesso agli *outliers*: in altri termini, l'influenza del corso di laurea sull'insuccesso universitario è presumibilmente legata non tanto a variabili di 2° livello (come potrebbero essere le percentuali di studenti per ogni docente, o altre simili), ma ad alcune variabili a livello studente le quali sembrano essere un più marcato fattore di insuccesso in certi corsi di laurea che non in altri. Si può configura-

re, in tal modo, un “effetto composizione” dovuto alle differenze iniziali fra gli studenti che accedono ad un corso di studi invece che ad un altro, scelta che dipende, appunto, dalla *composizione* delle caratteristiche del corso e dello studente.

Un esempio di influenza di covariate di primo livello sulla variabilità di secondo livello potrebbe essere, ad esempio (come si cercherà di verificare nel seguito della ricerca in corso), l'orientamento fornito dal tipo di scuola media superiore frequentata, essendo ovvio che un Istituto Tecnico Commerciale fornisce basi culturali più adeguate ad un corso di Economia e Commercio che non ad un corso di Scienze Biologiche o di Medicina Veterinaria, mentre il Liceo Linguistico fornisce indubbiamente una preparazione migliore per il corso di Lingue e di Letterature Straniere che non per quello di Matematica, e così via. Presumibilmente, soltanto il possesso di Maturità Scientifica non presenta un rischio di insuccesso superiore a quello dei diplomati del Liceo Classico, mentre la provenienza da un Istituto Professionale appare un pesante fattore di rischio.

Le osservazioni su menzionate lasciano quindi ritenere che l'influenza del tipo di maturità conseguita sull'insuccesso dipenda dal corso di laurea di immatricolazione in modo più complesso di quanto possa essere espresso da un singolo modello, anche se costruito con metodologie multilivello. La composizione fra le caratteristiche degli studenti (per esempio, attitudini ed aspettative) e le caratteristiche dei corsi di studio, infatti, può variare sensibilmente, comprendendo, in alcuni corsi, variabili e modalità che non hanno effetto in corsi di diverso genere. A questo punto, forse il miglior tipo di analisi potrebbe essere la scelta di elaborare una serie di modelli logit separatamente per ognuno dei corsi di laurea, con opportuni confronti dei risultati.

Tuttavia, in questo lavoro, diretto a fornire alcune osservazioni preliminari sull'analisi dei fattori di rischio di abbandono, ci si limiterà (anche per brevità) a definire due modelli logit separati per gli opposti *outliers* (uno per il c.d.l. di minore rischio d'insuccesso, ossia in Medicina e Chirurgia, ed uno per i due corsi con maggiori quote di insuccesso, Scienze Ambientali e Scienze Politiche) ed un terzo modello logit per l'insieme degli altri corsi del campione, per i quali, invero, l'analisi multilevel a componenti di varianza, con modello a sola intercetta, non ha rilevato un effetto gruppo statisticamente significativo.

5. Modelli logit per gruppi di corsi di laurea

Tenendo conto della citata decisione di condurre tre studi distinti e non dovendo più sottostare, quindi, ai precedenti limiti dimensionali dovuti a necessità computazionali, si è ritenuto conveniente, per aumentare l'affidabilità dei risultati, incrementare il numero di unità estratte dal c.d.l. in Medicina e da quello in Scienze Ambientali; per mantenere all'incirca costante il peso di quest'ultimo corso di laurea nel relativo cam-

pione, è stata incrementata anche l'ampiezza del subcampione estratto dal c.d.l. in Scienze Politiche, mantenendo la proporzionalità fra i corsi e quella in base all'esito.

A questo punto, date le numerose modalità di cui sono composte le quattro variabili esplicative considerate in questo studio (le quali sono, come nella precedente analisi: sesso, classe d'età all'immatricolazione, tipo di maturità, classe di voto di maturità), per determinare il modello logit più appropriato è stata condotta in primo luogo un'analisi log-lineare volta ad individuare i fattori significativi della relazione, (e quindi i modelli logit più adeguati).

Detta analisi parte da un modello log-lineare saturato per tabelle di contingenza pentadimensionali, che, come è noto (cfr., ad es. F. Delvecchio, 2002), è

$$\begin{aligned} \ln n_{ijkq}^* = & \mu + \lambda_i^X + \lambda_j^Y + \lambda_h^Z + \lambda_k^W + \lambda_{ij}^{XY} + \lambda_{ih}^{XZ} + \lambda_{ik}^{XW} + \lambda_{iq}^{XV} + \lambda_{jh}^{YZ} + \lambda_{jk}^{YW} + \lambda_{jq}^{YV} + \\ & + \lambda_{hk}^{ZW} + \lambda_{hq}^{ZV} + \lambda_{kq}^{WV} + \lambda_{ijh}^{XYZ} + \lambda_{ijk}^{XYW} + \lambda_{ijq}^{XYV} + \lambda_{ihk}^{XZW} + \lambda_{ihq}^{XZV} + \lambda_{ikq}^{XWV} + \\ & + \lambda_{jhk}^{YZW} + \lambda_{jhq}^{YZV} + \lambda_{jkq}^{YVW} + \lambda_{hkq}^{ZVW} + \lambda_{ijk}^{XYZW} + \lambda_{ijkq}^{XYVW} + \lambda_{jkhq}^{YZVW} + \lambda_{ijhkq}^{XYZVW} \end{aligned}$$

con i vincoli

$$\sum_i \lambda_i^X = \sum_j \lambda_j^Y = \dots = \sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = \dots = 0,$$

per poi eliminare, uno per volta, gli effetti non significativi mediante il test $G^2 = -2 \ln \Lambda$, ove $\lambda = L(\hat{\omega})/L(\hat{\Omega})$ è il rapporto fra il massimo $L(\hat{\omega})$ della verosimiglianza quando è vera l'ipotesi H_0 che l'effetto in esame non sia significativo e il massimo $L(\hat{\Omega})$ della verosimiglianza in tutto lo spazio dei parametri¹⁴.

La suddetta procedura è stata reiterata (tramite software SPSS) fino all'eliminazione di tutti i fattori non significativi, ottenendo i risultati che la Tab. 7 descrive in riferimento ai soli fattori di interazione: è noto, invero, che la presenza di un'interazione implica la necessità di mantenere nel modello (a causa della natura gerarchica del medesimo) tutte le variabili che hanno originato l'interazione stessa.

Prendendo in considerazione soltanto le interazioni con la variabile *Esito* (del percorso didattico), a questo punto è immediato affermare che, per il c.d.l. in Medicina e Chirurgia, il modello logit più adeguato comprenderà fra le covariate innanzitutto l'età di immatricolazione e il tipo di diploma conseguito, senza effetti di interazione; il modello logit per i corsi di laurea in Scienze Ambientali e Scienze Politiche comprenderà le medesime variabili. Invece, faranno parte del modello logit costruito per il campione relativo al complesso degli altri venti corsi di laurea dell'Ateneo di Bari l'età all'immatricolazione, il tipo di maturità ed il voto di maturità, tenendo conto dell'interazione fra l'età ed ognuna delle altre due variabili.

¹⁴ Poiché G^2 , al variare del campione, si distribuisce come la v.c. χ_g^2 con $g = g_1 - g_2$ pari alla differenza fra i gradi di libertà g_1 del modello comprendente l'effetto di cui si sta verificando la significatività e i gradi di libertà g_2 del modello privo di tale effetto, fissato α , se $G^2 < \chi_{g;\alpha}^2$ si accetta l'ipotesi di indipendenza dell'effetto omissso da quelli inclusi nel modello (Cfr., ad es., F. Delvecchio, 2002).

Tab. 7 – Effetti significativi dei modelli log-lineari saturati relativi ai tre campioni di corsi di laurea: “Medicina e chirurgia”, “Scienze Ambientali e Scienze Politiche”, “Restanti corsi di laurea”.

Corsi di laurea	Effetti significativi	g	G ²	$\chi^2_{g; 0,01}$
Medicina e chirurgia (n = 557)	<i>Esito</i> ∩ <i>Età di immatricolazione</i> <i>Esito</i> ∩ <i>Tipo di maturità</i> Età immatricol. ∩ Tipo di maturità Età immatricol. ∩ Voto di maturità Sesso ∩ Tipo di maturità Tipo di maturità ∩ Voto di maturità	177	134,81	223,69
Scienze Ambientali Scienze Politiche (n = 1084)	<i>Esito</i> ∩ <i>Età di immatricolazione</i> <i>Esito</i> ∩ <i>Tipo di maturità</i> Età di immatricolazione ∩ Sesso Età di immatr. ∩ Voto di maturità Sesso ∩ Tipo di maturità.	183	206,89	230,42
Restanti corsi di laurea (n = 4040)	<i>Esito</i> ∩ <i>Età imm.</i> ∩ <i>Tipo di maturità</i> <i>Esito</i> ∩ <i>Età imm.</i> ∩ <i>Voto di maturità</i> Età di immatricolazione ∩ Sesso Sesso ∩ Tipo di maturità Sesso ∩ Voto di maturità Tipo di maturità ∩ Voto di maturità	161	194,53	205,66

Partendo da tutte le modalità delle variabili che sono risultate significativamente associate all'esito degli studi nei modelli log-lineari di ciascun sub-campione, per la costruzione dei modelli logit ottimali ci si è avvalsi della procedura *backward stepwise* (eliminazione delle modalità non significative ai fini della verosimiglianza del modello), ancora una volta utilizzando il test G² per determinare i coefficienti delle singole modalità la cui omissione non inficia la verosimiglianza del modello¹⁵.

Il risultato dell'analisi è sintetizzato nella Tab. 8, che consente alcune considerazioni di notevole interesse: innanzitutto, per quanto riguarda il c.d.l. in Medicina e Chirurgia si rileva che la modalità che maggiormente influenza l'esito degli studi (in senso di insuccesso) è il possesso di maturità Professionale per il Commercio, i cui titolari iscritti al c.d.l. hanno una probabilità di insuccesso superiore di oltre 1000 volte rispetto alla *baseline* (i possessori di Maturità Classica), in quanto hanno fatto rilevare insuccesso *tutti* i pochi studenti provenienti da detta scuola che si sono immatricolati nel periodo osservato al corso in oggetto.

¹⁵ Essendo il test G² fornito, in tal caso, dalla differenza fra log-verosimiglianza del modello con *u-1* coefficienti e log-verosimiglianza del modello con *u* coefficienti, ossia $G^2 = G^2(M_{u-1}) - G^2(M_u)$, è dunque necessario confrontare il valore empirico del test così ottenuto con il valore critico della distribuzione χ^2 con $g = g(M_{u-1}) - g(M_u) = 1$, ed accettando la non significatività dei coefficienti (e quindi la loro esclusione dal modello) se $G^2 < \chi^2_{1,\alpha}$ (Cfr., ad es., F. Delvecchio, 2002).

Tab. 8 – Effetti significativi dei modelli logit a risposta dicotomica (con le variabili esplicative categoriche “sesso”, “età all’immatricolazione”, “tipo di maturità conseguito”, “voto di maturità”) relativi ai tre gruppi di corsi di laurea: “Medicina e chirurgia”, “Scienze Ambientali e Scienze Politiche”, “Restanti corsi di laurea”. $\chi^2_{1, 0,01} = 6,6349$.

Effetti significativi relativi ai tre gruppi di corsi di laurea	Stime dei parametri	Errori standard	G ²	Odds ratio
<i>Medicina e chirurgia</i>				
Intercetta	-1,6125	0,1294		
Età di immatricolazione ≥ 20 anni	1,1610	0,2432	22,01	3,2
Maturità Profession. per il Commercio	7,1446	12,5011	7,27	1267,2
Maturità Professionale per l’Industria	1,6468	0,2903	32,44	5,2
<i>Scienze Ambientali e Scienze Politiche</i>				
Intercetta	-0,0044	0,1282		
Età di immatricolazione ≥ 20 anni	1,3150	0,1525	75,04	3,7
Maturità Tecnica Commerciale	0,5704	0,1656	12,13	1,8
Maturità Tecnica per Geometri	1,5656	0,6139	9,65	4,8
Maturità Tecnica Industriale	0,6398	0,2518	6,96	1,9
<i>Restanti corsi di laurea</i>				
Intercetta	-1,8618	0,1141		
Età di immatricolazione ≥ 20 anni	1,9727	0,2249	88,46	7,2
Voto di maturità 45-53 sessantesimi	0,5595	0,1217	21,76	1,7
Voto di maturità 36-44 sessantesimi	1,2894	0,1177	131,02	3,6
Maturità Scientifica	0,1736	0,1283	1,83	1,2
Maturità Linguistica	0,6474	0,1995	10,29	1,9
Maturità Magistrale	1,0002	0,1229	67,18	2,7
Maturità Tecnica Commerciale	1,1506	0,0922	161,14	3,2
Maturità Tecnica per Geometri	1,6429	0,2516	50,65	5,2
Maturità Tecnica Industriale	1,2403	0,1427	79,30	3,5
Maturità Professionale Commercio	1,8957	0,2418	73,29	6,7
Maturità Professionale per l’Industria	1,4959	0,2058	58,49	4,5
Altro titolo	1,0469	0,2926	13,70	2,8
Età $\geq 20 \cap$ Voto 36-44 sessantesimi	-0,7316	0,2671	7,72	0,5
Età $\geq 20 \cap$ Voto 45-53 sessantesimi	-0,8856	0,2485	13,35	0,4
Età $\geq 20 \cap$ Maturità Scientifica	0,5990	0,1851	10,83	1,8

Una probabilità di insuccesso significativamente superiore alla categoria di riferimento (*odds ratio*=5,2) viene fatta rilevare anche dai titolari di maturità Professionale per l’Industria: si consideri, tuttavia, che molti di tali diplomati si sono iscritti al c.d.l. in Odontoiatria (i quali per motivi di ridotta ampiezza numerica sono inseriti nel campione di Medicina e Chirurgia), presumibilmente provenendo da un Istituto Professionale Odontotecnico, e quindi con qualche conoscenza almeno empirica del

ramo di studi intrapreso. Inoltre, com'è noto, gli iscritti a tale corso di laurea sono fortemente motivati: si consideri che poco più del 6% di costoro fa rilevare un insuccesso (quota che, comunque, sale quasi al 10% per coloro che hanno conseguito una Maturità Professionale per l'Industria).

L'essersi immatricolati dal ventesimo anno di età in poi è anch'esso un notevole fattore di rischio per il c.d.l. in oggetto (*odds ratio*=3,2 rispetto a coloro che si sono immatricolati prima del 20.mo compleanno). Ciò può essere dovuto alla minore preparazione di base di questi studenti, che è una fra le cause del tardivo superamento del concorso d'ammissione, insieme all'adempimento degli obblighi di leva prima dell'immatricolazione.

Per quanto riguarda i c.d.l. in Scienze Ambientali e Scienze Politiche, il fattore di maggior rilievo per l'insuccesso risulta la provenienza dall'Istituto Tecnico per Geometri (*odds ratio*=4,8) e, come nel caso precedente, l'immatricolazione avvenuta in età pari o superiore a 20 anni (*odds ratio*=3,7). Si rileva un rischio di insuccesso quasi doppio rispetto alla *baseline* (maturità classica) anche per coloro che hanno conseguito la Maturità Tecnica Industriale o Commerciale.

L'ultimo dei modelli logit, relativo all'esito degli studi degli immatricolati presso i restanti venti corsi di laurea dell'Università di Bari, fa rilevare alcuni fattori simili ai due precedenti, ma anche molti altri del tutto peculiari (com'era prevedibile, data la maggiore ampiezza numerica e la maggiore articolazione del campione): anche qui si nota un cospicuo peso della modalità "età ≥ 20 anni" sul rischio di insuccesso (quasi 8 volte rispetto agli immatricolati più giovani), seguita dal possesso di Maturità Professionale per il Commercio (*odds ratio*=7,2), di Maturità Tecnica per Geometri (*odds ratio*=5,5) e di Maturità Professionale per l'Industria, con un rischio di insuccesso quasi quintuplo rispetto ai titolari di Maturità Classica (*baseline*). In realtà, qualsiasi diploma di scuola media superiore diverso dalla Maturità Classica risulta fattore di rischio, sia pure con livelli differenti; fa eccezione soltanto la Maturità Scientifica, la cui stima del parametro di regressione logit non è significativa. Tuttavia, i diplomati al Liceo Scientifico che si immatricolano da 20 anni in poi fanno rilevare un incremento significativo del rapporto (*odds ratio*=1,8), la cui interpretazione è analoga a quella esposta nel corso dell'analisi multilivello.

Chi ha conseguito la maturità con meno di 45/60 ha una probabilità di abbandono pari a 3,6 volte quella degli elementi meglio valutati all'esame di maturità, mentre è giustamente intermedio (1,7) l'*odds ratio* della categoria intermedia "voto di maturità di 45-53 sessantesimi". Inoltre, l'essersi immatricolati più tardivamente implica un'interazione che, come nella precedente analisi, è legata all'aumento di rischio per gli studenti che hanno conseguito le votazioni migliori. Si può supporre, al proposito, l'esistenza di altri fattori connessi (non evidenti dai dati a disposizione), come, ad esempio, la condizione lavorativa: è indiscusso, invero, che chi si diploma con i migliori risultati ha una maggiore probabilità di inserirsi nel mercato del lavoro, e che ottenga

anche occupazioni più remunerative (ma impegnative) di altri diplomati meno bravi; tuttavia, nel caso di una successiva scelta universitaria, detto vantaggio può risolversi in uno svantaggio, in quanto aumentano le difficoltà di conciliare studio e lavoro.

Considerazioni conclusive

Dall'analisi qui condotta, volta ad esplorare i possibili fattori del rischio di insuccesso negli studi universitari - insuccesso che, come specificato in un precedente lavoro (F. Delvecchio – F. d'Ovidio, 2002), può esprimersi sia come abbandono formalmente definito (rinuncia o decadimento dagli studi), sia come elevata propensione all'abbandono stesso (individuata dall'appartenenza alla categoria di studenti da noi definiti "ad alto rischio di abbandono") -, sono scaturiti alcuni punti di notevole interesse.

In primo luogo, in tema di insuccesso universitario è molto verosimile che esista realmente un'influenza del corso di laurea di immatricolazione, influenza più marcata in alcuni corsi come Medicina e Chirurgia (ove si riscontra una quota di insuccessi molto inferiore alla media, verosimilmente a causa della selezione che ammette all'immatricolazione soltanto i più preparati) o Scienze Ambientali e Scienze Politiche (che invece fanno rilevare il fenomeno opposto).

In secondo luogo, detta influenza non sembra dovuta in misura sensibile alle caratteristiche strutturali dei vari corsi, escludendo quelle peculiarità che fungono o possono fungere da filtro al momento dell'immatricolazione¹⁶, ma a caratteristiche pregresse di quegli studenti che a detti c.d.l. fanno poi afferenza, ossia a variabili definite a livello studente: in particolare alla provenienza da scuole secondarie il cui intento prevalente non è fornire basi teoriche per futuri studi, ma basi tecniche (alcune di notevole valore e complessità) che consentano un più agevole inserimento nel mercato del lavoro: istituti professionali e (seppure con una certa variabilità) istituti tecnici.

Certamente, ha un notevole peso anche l'età all'immatricolazione, soprattutto tenendo conto che un ritardo nell'immatricolazione, per una scarsa attitudine pregressa o per aver adempiuto agli obblighi di leva o per aver scelto di proseguire gli studi in ultima istanza (dopo aver cercato inutilmente un'occupazione o averla persa), compromette generalmente *l'abitudine allo studio*, che è un importante fattore di successo in molti campi, ma specialmente in ambito universitario.

¹⁶ Si pensi, ad esempio, alle prove istituite per l'accesso al c.d.l. in Medicina e Chirurgia, oppure alle specifiche conoscenze umanistiche che, nel comune pensiero, sono presupposto naturale per seguire la maggior parte degli indirizzi di studio attivati nel c.d.l. in Lettere o, infine, alle attitudini richieste agli studenti per seguire i corsi di Matematica e Fisica: condizioni ragionevoli, queste, ma non abbastanza da assumere un ruolo ufficiale di filtro al momento dell'iscrizione.

Infine, com'è logico, gli studenti a cui la scuola secondaria ha fornito una preparazione inadeguata agli studi universitari risultano avere una maggior probabilità di non concludere positivamente il proprio corso di laurea; d'altra parte, a causa della crescente specializzazione dei vari corsi di laurea e la sempre maggiore complessità delle conoscenze, la preparazione scolastica difficilmente può coprire le competenze richieste a priori da molti corsi di laurea.

La successiva analisi con modelli logit ha confermato le osservazioni precedenti, sottolineando che le scuole più orientate alla tecnica applicativa ed alla formazione professionale riescono generalmente a fornire basi teoriche poco adeguate al proseguimento degli studi in ambiente universitario, pur se con qualche differenza di gradazione; nel caso del c.d.l. in Medicina e Chirurgia, tuttavia, l'effetto "scuola di provenienza" è molto meno notevole (interessando solo gli Istituti Professionali), presumibilmente perché a tale corso si iscrivono prevalentemente diplomati presso i Licei classici e scientifici.

Nel caso dei corsi di laurea in Scienze Ambientali e in Scienze Politiche si riscontra, invece, una significativa influenza del possesso di una Maturità Tecnica di qualunque tipo sul rischio di insuccesso, che comunque è generalmente molto elevato in tali corsi universitari: si ricordi che oltre tre quarti di coloro che vi si immatricolano abbandonano gli studi o comunque se ne estraneano.

Nel caso dell'insieme degli altri corsi di laurea dell'Università di Bari, il dettaglio informativo che si ricava dall'analisi è notevole, anche a causa della cospicua ampiezza del campione. In generale, le conclusioni tratte costituiscono una semplice precisazione di quanto osservato nell'analisi precedente, ma vi sono alcune differenze nel peso dei singoli fattori, ed anche qualche osservazione originale. In definitiva, lo studio condotto consente di definire una sorta di graduatoria di rischio per gli studenti in base al tipo di scuola frequentato prima di iscriversi all'Università, con gli Istituti Professionali e Tecnici ai primi posti ed i Licei agli ultimi, fors'anche per le minori possibilità lavorative degli studenti che sono in possesso di una maturità liceale e quindi per le maggiori motivazioni che li spingono.

L'indagine ha posto in luce, inoltre, che è generalmente marcato il rischio di abbandono per chi si immatricola in età di 20 anni e più: detto rischio è presumibilmente legato sia alle oggettive difficoltà di chi consegue in ritardo la maturità (per motivi di difficoltà ad integrarsi nel processo culturale, oppure per motivi economici o familiari), sia alle difficoltà di altro genere che preoccupano anche chi si è diplomato nelle migliori condizioni, ma che si è iscritto con ritardo per motivi occupazionali (in tal caso, il rischio di insuccesso è legato alla difficoltosa conciliazione fra gli impegni lavorativi e quelli di studio).

In definitiva, si può quindi affermare che è molto consistente l'impatto della "selezione differita" dovuta principalmente ad una preparazione scolastica poco adeguata alla prosecuzione degli studi a livello universitario (ma anche ad una valutazione poco ponderata delle proprie attitudini e delle proprie conoscenze di base), se-

lezione che causa molti problemi agli studenti, alle loro famiglie ed alle medesime istituzioni universitarie.

Per ridurre questo fenomeno, è forse opportuno prendere in attenta considerazione l'attivazione di corsi pre-universitari di formazione (facoltativi, ma fortemente consigliati per quelle tipologie di studenti che fanno verificare la presenza di uno o più dei fattori di rischio qui posti in luce), i quali, pur senza comportare un ritorno agli sbarramenti del passato, possano fornire a coloro che ne sono carenti le basi necessarie per avviare un serio progetto didattico in ambiente universitario. Tali corsi avrebbero uno scopo di "alfabetizzazione", nel senso di *consolidamento* delle nozioni di base per chi le possiede e di *aggiornamento* per quegli studenti che non hanno precedentemente svolto o approfondito taluni argomenti, allo scopo di rendere più agevole lo studio delle discipline universitarie e di offrire un valido aiuto per le selezioni legate al "numero programmato", finora istituito solo in alcuni corsi di laurea ma che potrebbe maggiormente diffondersi in futuro.

I suddetti corsi formativi fornirebbero senz'altro qualche strumento conoscitivo in più a chi vuole intraprendere un percorso universitario serio, pur senza apportare, ovviamente, benefici sensibili a coloro che, nell'isciversi all'Università, obbediscono prevalentemente a motivi spuri, come, ad esempio, gli eventuali vantaggi economici derivanti dalla condizione di studente a carico dei genitori e motivi consimili.

Riferimenti bibliografici

- BRYK A.S., RAUDENBUSH S.W. (1988) Methodology for cross-level organization research, in *Research of Organizational Behavior*.
- BRYK A.S., RAUDENBUSH S.W. (1992) *Hierarchical Linear Models*, Sage, Newbury Park, CA.
- COMMENGES D., JACQMIN H. (1994) The intraclass correlation coefficient: distribution-free definition and test, *Biometrics*, 50, 517-526.
- DELVECCHIO F., TEDESCO N. (1997) Un'analisi esplorativa sull'influenza di alcuni fattori sul voto di maturità a Matera, in "Socializzare la Statistica", *Atti della Giornata di studio in memoria di Bruno Grazia-Resi, Roma, 24 gennaio 1997*, SIAE.
- DELVECCHIO F., *Cenni sui modelli multilevel - Appunti ad uso degli studenti*, Cattedra di Statistica Sociale, Dipartimento di Scienze Statistiche, Università di Bari, a.a. 2000-01.
- DELVECCHIO F. (2002) *Statistica per la ricerca sociale*, Cacucci, Bari.

- DELVECCHIO F., D'OVIDIO F. (2002) Alcuni risultati relativi ai percorsi universitari nell'Ateneo barese, *Workshop: Indicatori e metodi per l'analisi dei percorsi universitari e post-universitari*, Bari, 22 febbraio 2002.
- GOLDSTEIN H. (1987) *Multilevel Models in Educational and Social Research*, Griffin, London.
- GOLDSTEIN H. (1995) *Multilevel Statistical Models*, Arnold, London / Halsted, New York..
- HOX J. J. (1995) *Applied Multilevel Analysis*, TT-Publikaties - Amsterdam.
- MC CULLAGH P., NELDER J.A. (1989) *Generalized linear model*, Chapman & Hall, London.
- MOSTELLER F., TUKEY J. W. (1977) *Data Analysis and Regression*, Addison-Wesley, Reading (Mass.).
- SNIJDERS T., BOSKER R. (1999) *Multilevel Analysis*, SAGE Publications, London.
- VERBEKE G., LESAFFRE E. (1997) The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data, in *Computational Statistics and Data Analysis*, 23, 541-556.

Some risk factors of failure in university studies

Summary. *The observation of data on the university population of Bari shows, among the different graduation courses, some differentiation in the risk of failure in university studies, due to reasons related to the characteristics of students or to the structure of the courses themselves. In order to verify whether such differences exist and what they are determined by, logit models were used, even with multilevel structure, that enabled us to determine matriculation age, type and grade of high-school diploma as primary factors of risk for the students of Bari University, with a differentiated influence according to the graduation course chosen, thus providing us with interesting hints of interpretation.*

Keywords: *Failure in university studies, risk factors, multilevel models, logit models.*