

**Sviluppo e validazione di un modello predittivo per valutare il rischio individuale di
abbandono negli Immatricolati ad un Corso di Laurea Triennale o a Ciclo Unico**

Vincenzo Bagnardi e Davide Disalvatore

Dipartimento di Statistica, Università degli Studi di Milano-Bicocca

Lavoro finanziato dalla Commissione Orientamento dell'Ateneo di Milano-Bicocca.

Introduzione

Il monitoraggio delle carriere degli studenti iscritti ai Corsi di Laurea (CdL) è una prassi ormai consolidata nell'Ateneo di Milano-Bicocca, e avviene attraverso l'estrazione e la successiva elaborazione informatica e statistica dei dati contenuti negli Archivi Amministrativi delle Segreterie Studenti.

Dall'analisi delle coorti di studenti iscritti a partire dall'a.a. 2000/01, anno di attivazione dei CdL Triennali, si è evidenziato, a livello di Ateneo, che quasi uno studente immatricolato su quattro abbandona nel primo anno di corso, mentre circa uno studente su tre abbandona entro il terzo anno. Le Facoltà di Economia, Giurisprudenza e Scienze Matematiche, Fisiche e Naturali risultano essere le più colpite da questo fenomeno, con proporzioni che, sebbene variabili nel tempo, si attestano intorno al 30%.

Il rischio individuale di abbandono di un CdL Triennale o a Ciclo Unico è associato sia alle caratteristiche del singolo studente che alla tipologia del corso di studi prescelto. Oltre al voto di maturità, la caratteristica maggiormente associata al rischio di abbandono è risultata essere la scuola superiore di provenienza, con quote inferiori al 20% per i licei e superiori anche al 30% per istituti tecnici e professionali. L'età all'immatricolazione, il genere e l'area di residenza sono risultate anch'esse associate al rischio di abbandono.

L'esigenza di razionalizzare le risorse impiegate dall'Ateneo, sia nell'orientamento in entrata che nelle iniziative di supporto alla didattica, richiede di andare oltre la semplice descrizione di quanto finora osservato. È quindi necessario un ulteriore sforzo finalizzato a individuare fin dal momento dell'immatricolazione i profili degli studenti a maggior rischio di abbandono, in modo da poter attuare tempestivamente degli interventi mirati sui singoli studenti, come attività di tutoraggio o colloqui individuali.

Nell'ambito della ricerca medica, da diversi anni vengono messi a punto modelli statistici utili a identificare soggetti che, sulla base della combinazione di caratteristiche e abitudini individuali,

potrebbero sviluppare con elevata probabilità una particolare patologia o trarre benefici da un particolare trattamento, al fine di pianificare opportuni interventi di prevenzione o cura.

La metodologia utilizzata per sviluppare tali modelli prognostici si basa inizialmente sulla selezione di una appropriata funzione di rischio ottenuta con metodi di regressione multipla e, successivamente, sulla valutazione della sua bontà predittiva, attraverso procedure di validazione e calibrazione.

In analogia a tali studi, ci si propone di ricercare, sviluppare e validare un modello che possa predire il rischio di abbandono nel primo anno di corso negli studenti che si immatricolano per la prima volta ad un CdL Triennale o a Ciclo Unico.

Le previsioni che deriveranno dall'utilizzo di questo strumento consentiranno, una volta individuati i profili di rischio e la loro frequenza, di progettare, sulla base dei costi associati e delle risorse disponibili, i possibili scenari di intervento.

Materiali e Metodi

Dagli Archivi informatizzati delle Segreterie Studenti dell'Ateneo di Milano-Bicocca sono stati identificati 45721 studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico dell'Ateneo negli a.a. 02/03-08/09.

L'evento di interesse per lo sviluppo del modello predittivo è l'abbandono della Facoltà nel primo anno di corso. L'abbandono comprende la rinuncia esplicita agli studi, il trasferimento ad altra Facoltà o Ateneo, o la mancata iscrizione da parte dello studente al secondo anno di corso.

L'insieme dei 39040 studenti immatricolati negli a.a. 02/03-07/08 è stato utilizzato per sviluppare il modello predittivo (“training set”), mentre l'insieme dei 6681 studenti immatricolati nell'a.a. 08/09 è stato utilizzato per valutare, in modo indipendente, l'accuratezza del modello (“validation set”).

La tabella seguente riporta la numerosità degli studenti considerati nei due insiemi (training e validation), suddivisa per le otto Facoltà dell'Ateneo.

Facoltà	Immatricolati a.a. 02/03-07/08 (Training set)	Immatricolati a.a. 08/09 (Validation set)	Totale (%)
Economia	9604	1926	11530 (25.22)
Giurisprudenza	3455	487	3942 (8.62)
Medicina e Chirurgia	3193	557	3750 (8.20)
Psicologia	4612	680	5292 (11.57)
Scienze della Formazione	5839	1046	6885 (15.06)
Scienze Matematiche, Fisiche e Naturali	8167	1294	9461 (20.69)
Scienze Statistiche	766	135	901 (1.97)
Sociologia	3404	556	3960 (8.66)
Totale	39040	6681	45721 (100.00)

Le seguenti variabili sono state considerate predittive nello sviluppo del modello:

- Età all'immatricolazione (in anni compiuti)
- Genere

- Residenza (classificata su 4 livelli: Milano città, Provincia di Milano esclusa Milano città, Lombardia esclusa Milano città e Provincia, Altro)
- Voto di maturità (in centesimi)
- Scuola superiore di provenienza (classificata in 8 livelli: Liceo Classico, Liceo Scientifico, Istituto Tecnico Commerciale, Altri Istituti Tecnici, Magistrale, Istituto Professionale, Liceo Linguistico, Altro)
- CdL di immatricolazione (per utilizzare tutta l'informazione disponibile, sono stati considerati anche gli studenti immatricolati a CdL non più attivi).

L'informazione relativa alla fascia di reddito dichiarata all'immatricolazione era stata inizialmente prevista tra le variabili da predittive. Tuttavia, poiché una quota non trascurabile di studenti che abbandonano nel primo anno di corso non dichiara la fascia di reddito di appartenenza (ricordiamo che la documentazione riguardante il reddito può essere presentata anche nei mesi successivi all'immatricolazione), e poiché le procedure automatizzate delle Segreterie Studenti assegnano la fascia di reddito più elevata a coloro che non presentano la documentazione, l'associazione tra la variabile Reddito e il rischio di abbandono potrebbe risultare distorta. Per questa ragione, il Reddito è stato escluso dal modello finale. I dati relativi alle percentuali di abbandono suddivise per fascia di reddito vengono comunque riportate, a titolo descrittivo, nel presente lavoro.

Modelli di regressione logistica multipla¹ sono stati utilizzati per valutare l'associazione tra le variabili considerate e il rischio di abbandono nel primo anno di corso e per la stima del rischio individuale. L'effetto del voto di maturità è stato valutato separatamente per ogni tipologia di scuola superiore di provenienza, mediante la stima dei parametri di interazione tra la variabile continua “voto” e la variabile nominale “maturità” .

Per ogni livello delle caratteristiche considerate, insieme al rischio di abbandono, è stato calcolato l'intervallo di confidenza (IC) al 95%.

¹ Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. New York, NY: John Wiley & Sons, Inc.; 2000

L’ampiezza dell’intervallo di confidenza in corrispondenza di un determinato livello della variabile considerata (es. Residenza in Milano Città, Voto di Maturità pari a 90è funzione del numero di studenti con quella determinata caratteristica. Minore è l’ampiezza dell’intervallo di confidenza, maggiore è l’affidabilità della stima riportata. Nel confrontare due livelli della stessa variabile (es. Uomini vs Donne), gli IC al 95% permettono di determinare, con un elevato grado di confidenza, se la differenza osservata possa essere attribuita al caso (in caso di sovrapposizione degli intervalli) oppure ad una reale effetto della variabile (nel caso di intervalli che non si sovrappongono).

Il modello finale di rischio è stato rappresentato mediante un nomogramma¹, una semplice rappresentazione grafica che permette sia di valutare il contributo relativo di ogni singola variabile considerata nel modello, sia di calcolare facilmente il rischio di abbandono sulla base delle caratteristiche individuali di uno studente (v. le “istruzioni per l’uso del nomogramma” riportate per ogni modello presentato).

Infine, il modello predittivo è stato validato sulla coorte di studenti immatricolati nell’a.a. 2008/09 (*validation set*). I coefficienti del modello sviluppato sul *training set* sono stati utilizzati per stimare le probabilità di abbandono nel primo anno di corso negli studenti appartenenti al *validation set*.

L’accuratezza predittiva del modello è stata valutata rispetto alla sua calibrazione², definita come il grado di accordo tra le frequenze di abbandoni osservate nel *validation set* e le probabilità predette dal modello. Gli studenti del *validation set* sono stati prima ordinati sulla base del loro rischio predetto, e quindi raggruppati in cinque livelli di uguale numerosità (quintili di rischio). All’interno di ogni livello, è stato calcolato sia il rischio medio previsto di abbandono, sia la frequenza osservata nell’a.a. considerato. Le due frequenze (rischio predetto e rischio osservato) sono state rappresentate congiuntamente in un *calibration plot*. In caso di modello predittivo perfettamente accurato, dove le probabilità osservate corrispondono a quelle predette, si otterebbe, nel calibration

¹ Iasonos A, Schrag D, Raj GV, Panageas KS. How To Build and Interpret a Nomogram for Cancer Prognosis. Journal of Clinical Oncology. 2008; 26:1364-70

² Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine. 1996; 15:361-87

plot, una linea con pendenza pari a 45 gradi (diagonale ideale). La distanza tra le coppie osservato/atteso del modello sviluppato e questa linea ideale misura l'errore assoluto della predizione.

Nel calibration plot sono stati inoltre riportati gli IC al 95% delle percentuali di abbandono osservate. Intervalli che comprendono la linea ideale suggeriscono che le differenze osservate potrebbero essere attribuibili al caso. Tuttavia, bisogna ricordare che l'ampiezza degli IC al 95% dipende dal numero di soggetti inclusi in ogni gruppo: questi saranno quindi maggiori per Facoltà con un numero relativamente basso di immatricolati, molto ridotti per Facoltà con un numero elevato di immatricolati.

Una valutazione contemporanea dell'errore (quanto i punti cadono lontano dalla diagonale ideale) e dell'ampiezza degli IC deve essere effettuata nell'analisi dei calibration plot.

Tutte le analisi sono state condotte a livello di Facoltà.

Tutte le analisi sono state effettuate con il software SAS e il software R (librerie Design e Hmisc).

Risultati

Per ogni Facoltà i dell’Ateneo considerata, vengono presentate, in coda al presente lavoro, le seguenti tabelle e figure:

- **Tabella $i.1$** Legenda dei CdL considerati
- **Tabella $i.2$** Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08)
- **Figura $i.1$** Relazione tra le caratteristiche all’entrata degli studenti immatricolati e rischio di abbandono nel primo anno di corso. Stime puntuali e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08)
- **Figura $i.2$** Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08)
- **Tabella $i.3$** Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09)
- **Figura $i.3$** Calibration plot per il modello predittivo. Validation Set (coorte 08/09)
- **Tabella $i.4$** Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09)
- **Appendice $i.$** Equazione del modello predittivo del rischio di abbandono

Dall’esame dei nomogrammi sviluppati per le Facoltà di Economia e di Giurisprudenza, si osserva che il genere e la residenza risultano essere deboli predittori dell’abbandono, con rischi maggiori associati agli studenti di sesso maschile e ai non residenti in Lombardia (Figure 1.2 e 2.2). L’età all’immatricolazione risulta essere fortemente associata all’abbandono, con rischi crescenti fino ai

30 anni e decrescenti dopo. L'effetto del voto di maturità pare invece essere associato in modo lineare al rischio di abbandono, con andamenti simili e decrescenti per tutti i tipi di maturità considerati (Figure 1.1 e 2.1). Il tipo di maturità risulta essere un forte determinante dell'abbandono. Eccetto che per la categoria di studenti classificati a rischio “molto alto” di abbandono nella Facoltà di Giurisprudenza, dove si è verificato lo stesso numero di eventi osservati nella categoria a rischio “alto” (Tabella 2.4), per tutte le altre categorie di rischio il modello predittivo si dimostra adeguato (Figura 1.3 e 2.3) in entrambe le Facoltà.

Nella Facoltà di Medicina e Chirurgia, caratterizzata da un basso tasso di abbandono nel primo anno, non sembrano esserci caratteristiche che incidono in modo determinante sul rischio, eccetto la residenza. Dal nomogramma si osserva un rischio nettamente maggiore negli studenti provenienti da Milano Città o non residenti in Lombardia, rispetto agli altri studenti lombardi. Il tipo e il voto di maturità hanno un effetto minore rispetto a quanto osservato per le altre Facoltà, mentre la parte maggiore di variabilità nei rischi sembra essere spiegata dal CdL di immatricolazione, con rischi di abbandono più elevati per il CdL in Tecniche di Laboratorio Biomedico o Infermieristica e ridotti per il CdL a Ciclo Unico in Medicina e Chirurgia o per il CdL in Fisioterapia.

L'accuratezza del modello è accettabile, anche se minore rispetto alle altre Facoltà. Gli studenti classificati nella categoria “basso rischio” presentano infatti percentuali di abbandono inferiori solo alla categoria di rischio “molto alto” (Figura 3.3). Buono invece l'accordo tra gli abbandoni osservati nella categoria di rischio “molto alto” e quelli predetti (Tabella 3.4).

Nella Facoltà di Psicologia e di Scienze della Formazione si evidenzia un forte effetto dell'età all'immatricolazione sul rischio di abbandono. Il voto di maturità sembra avere un effetto maggiore nella Facoltà di Scienze della Formazione, soprattutto per chi arriva da una scuola magistrale, mentre nella Facoltà di Psicologia il voto di maturità non pare incidere in modo determinante sul rischio di abbandono (Figura 4.2 e 5.2). In entrambe le Facoltà, un elevato rischio di abbandono si

osserva per gli studenti con maturità professionale, indipendentemente dal voto di maturità conseguito.

In fase di validazione, i rischi predetti dai modelli sviluppati nelle due Facoltà risultano essere molto simili a quelli osservati (Figure 4.3 e 5.3).

Nella Facoltà di Scienze Matematiche, Fisiche e Naturali l'abbandono sembra essere fortemente determinato sia dal tipo di maturità, che dal voto (Figura 6.1). Studenti con maturità classica presentano un elevato rischio di abbandono, e questo non sembra variare al crescere del voto. Al contrario, un voto di maturità alto sembra diminuire di molto il rischio di abbandono per studenti con maturità scientifica o tecnica. Una parte significativa della variabilità nei rischi sembra essere spiegata anche dal CdL di immatricolazione, con rischi di abbandono elevati per il CdL in Scienze Biologiche e bassi per il CdL in Scienze e Tecnologie Chimiche.

La bontà predittiva del modello è ottima: gli scarti tra i rischi previsti e quelli osservati nella coorte di validazione sono minimi, soprattutto agli estremi delle categorie di rischio considerate (Figura 6.3).

Anche nella Facoltà di Scienze Statistiche l'età all'immatricolazione, il voto e il tipo di maturità sembrano essere i fattori che maggiormente condizionano la probabilità di proseguire gli studi oltre il primo anno. Un voto elevato di maturità ha un peso maggiore se conseguito in un liceo classico o scientifico, mentre risulta essere meno discriminante se conseguito in un istituto professionale o tecnico (Figura 7.2). Poco significativi gli effetti del genere, della residenza e del CdL di immatricolazione.

Dato l'esigua numerosità del validation set, il modello predittivo presenta, per alcune categorie di rischio intermedie, differenze non trascurabili tra i rischi predetti e quelli osservati. Tuttavia, nelle classi di rischio estreme (“molto basso” e “molto alto”) l'accordo è quasi perfetto (Figura 7.3).

Dal nomogramma sviluppato per la Facoltà di Sociologia (Figura 8.1) si può notare come non prevalga l'effetto di una caratteristica sulle altre. Età, voto, tipo di maturità, residenza e CdL spiegano tutte una parte significativa della variabilità osservata tra i rischi di abbandono. Il modello predittivo è ben calibrato: la stima dei rischi predetti è pressoché identica alle frequenze osservate per le prime quattro categorie di rischio (Figura 8.3), anche se significativamente inferiore nell'ultima categoria (47% predetto vs 64% osservato).

Conclusioni

A partire dalle informazioni raccolte nel monitoraggio delle carriere degli studenti immatricolati ad un CdL Triennale o a Ciclo Unico dell'Ateneo negli a.a. 2002/03-2007/08, sono stati sviluppati otto modelli statistici, uno per ogni Facoltà dell'Ateneo di Milano-Bicocca, che permettono di predire il rischio di abbandono individuale nel primo anno di corso.

Il rischio medio di abbandono è risultato essere molto eterogeneo tra le Facoltà dell'Ateneo sia quando viene valutato in termini assoluti (si va dal 10-11% nella Facoltà di Medicina e Chirurgia a valori superiori al 30% nella Facoltà di Scienze Matematiche, Fisiche e Naturali), sia quando si considera l'effetto che hanno su di esso le caratteristiche degli studenti all'immatricolazione.

Gli otto modelli sono stati validati nella coorte di studenti immatricolati nell'a.a. 2008/09, al fine di valutare in modo indipendente la loro bontà predittiva.

Per tutte le Facoltà il modello ha rivelato ottime proprietà predittive, in quanto ha permesso di classificare in modo quasi sempre corretto studenti con caratteristiche associate a differenti livelli di rischio.

Le caratteristiche considerate sono state estratte dagli Archivi Amministrativi delle Segreterie Studenti, e comprendono semplici variabili come l'età all'immatricolazione, il genere, l'area di residenza, il diploma di scuola superiore conseguito e il voto di maturità. L'aggiunta di nuove variabili, potenzialmente associate al rischio di abbandono, come ad esempio la professione e il titolo di studio dei genitori o la condizione di studente-lavoratore, porterebbe sicuramente al miglioramento della capacità predittiva dei modelli presentati in questo lavoro. Purtroppo questi dati, non essendo di interesse amministrativo, non vengono rilevati dalle Segreterie Studenti al momento dell'immatricolazione.

Mediante l'utilizzo di un software applicativo sviluppato in ambiente SAS sarà possibile dal prossimo a.a., una volta chiuse le immatricolazioni e dopo avere importato nel software i dati dagli Archivi delle Segreterie, assegnare immediatamente ad ogni nuova matricola il suo profilo di

rischio, basato sulle sue caratteristiche individuali e sui parametri stimati nei modelli predittivi sviluppati in questo lavoro.

Le previsioni che deriveranno dall'utilizzo di questo strumento potranno consentire di progettare, sulla base dei costi associati e delle risorse disponibili, possibili interventi mirati soprattutto verso gli studenti che, con elevata probabilità, avranno maggiori difficoltà nel proseguimento della loro carriera universitaria.

Indice delle Tabelle e Figure

1. Facoltà di Economia	pag. 15-23
2. Facoltà di Giurisprudenza	pag. 24-32
3. Facoltà di Medicina	pag. 33-41
4. Facoltà di Psicologia	pag. 42-50
5. Facoltà di Scienze della Formazione	pag. 51-59
6. Facoltà di Scienze Matematiche, Fisiche e Naturali	pag. 60-68
7. Facoltà di Scienze Statistiche	pag. 69-77
8. Facoltà di Sociologia	pag. 78-86

1. Facoltà di Economia

Tabella 1.1 Legenda dei CdL considerati. Facoltà di Economia.

Codice CdL	Descrizione
503	Commercio Estero *
504	Discipline Economiche e Sociali *
505	Economia delle Banche, delle Assicuraz. e degli Interm. Fin.
506	Economia e Amministrazione delle Imprese
507	Economia e Commercio
508	Economia e Gestione dei Servizi Turistici
510	Economia e Gestione delle Amministrazioni Pubbliche *
511	Economia, Statistica e Informatica per l'Azienda
569	Marketing, Comunicazione Aziendale e Mercati Internazionali

* CdL non più attivi

Tabella 1.2 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08). Facoltà di Economia.

	N. immatricolati 9604	N. Abbandoni 2748	% Abbandono 28.6
Classe di età			
<20	5696	1243	21.8
=>20	3908	1505	38.5
Genere			
F	4584	1134	24.7
M	5020	1614	32.2
Residenza			
Milano	1795	575	32.0
Prov. MI(no Milano)	4956	1327	26.8
Lombardia(no MI/Prov)	1888	464	24.6
Altro	965	382	39.6
Voto Maturità *			
Alto	1860	294	15.8
Basso	7744	2454	31.7
Maturità			
Classica	272	91	33.5
Scientifica	2462	501	20.3
Tecnico commerciale	2633	711	27.0
Altre Tecniche	1778	565	31.8
Magistrale	151	59	39.1
Professionale	1309	545	41.6
Linguistico	406	113	27.8
Altro	593	163	27.5
Coorte			
2002	1302	382	29.3
2003	1430	404	28.3
2004	1754	546	31.1
2005	1682	460	27.3
2006	1706	494	29.0
2007	1730	462	26.7
CdL			
503	316	76	24.1
504	129	32	24.8
505	1069	311	29.1
506	2050	580	28.3
507	2983	766	25.7
508	864	347	40.2
510	25	6	24.0
511	293	86	29.4
569	1875	544	29.0
Fascia reddito #			
Bassa	2380	657	27.6
Media	5609	1539	27.4
Alta	1615	552	34.2

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta)

Figura 1.1 Relazione tra le caratteristiche all'entrata degli studenti immatricolati e il rischio di abbandono nel primo anno di corso. Stime puntuale e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08). Facoltà di Economia.

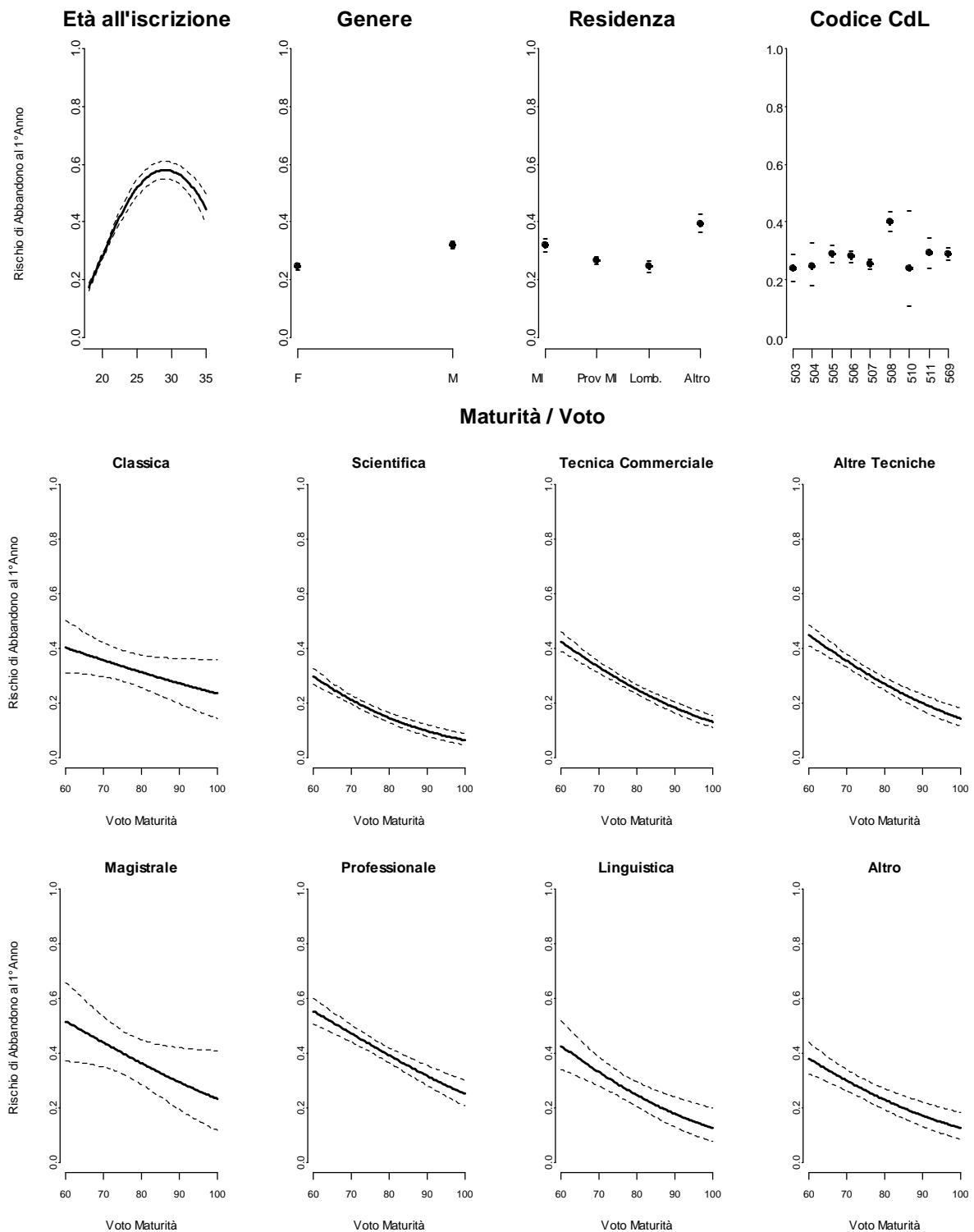
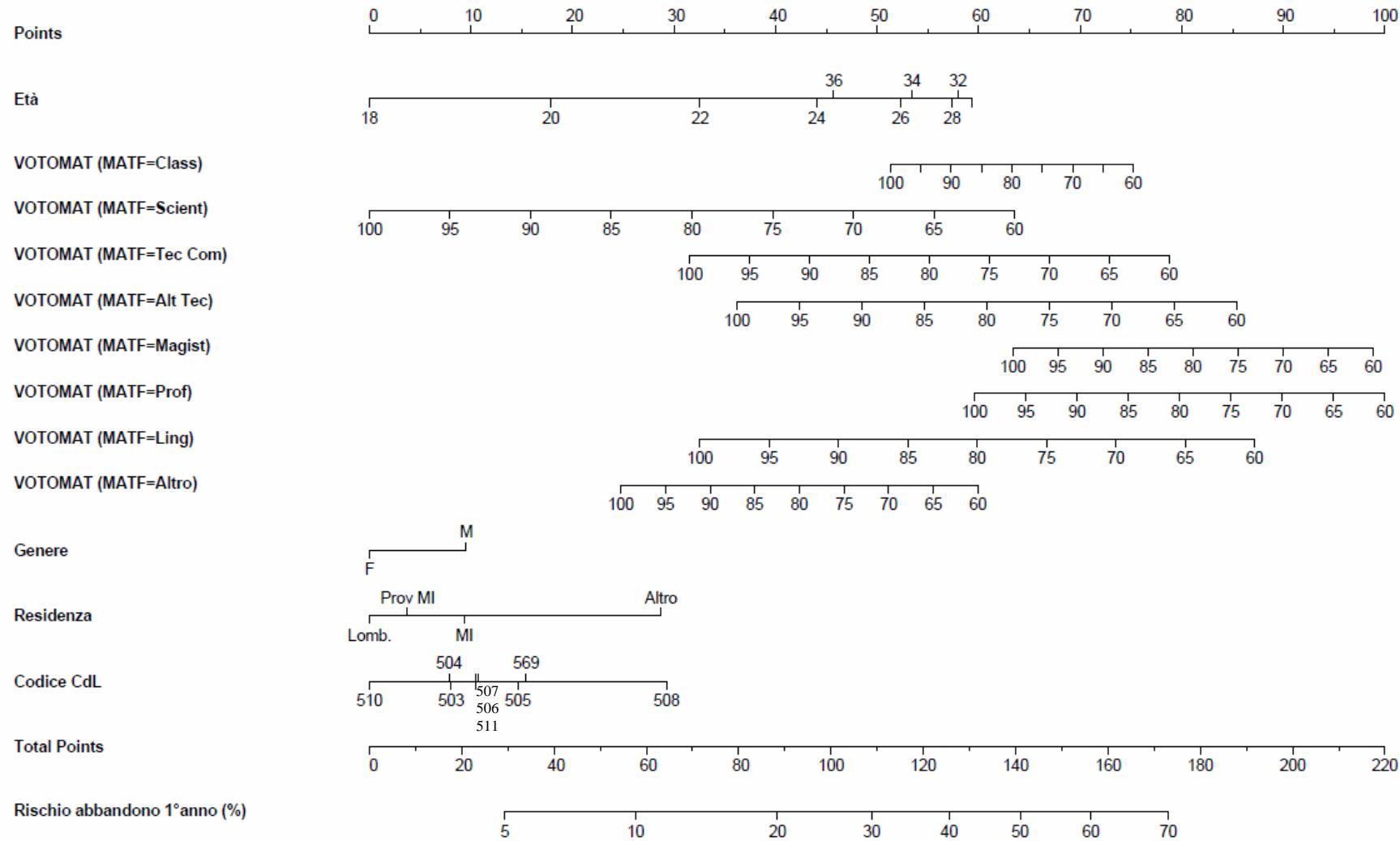


Figura 1.2 Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08). Facoltà di Economia.



Istruzioni per l'uso del nomogramma.

Sul secondo asse delle ascisse a partire dall'alto si identifichi l'età all'immatricolazione dello studente per il quale si vuole calcolare il rischio di abbandono nel primo anno. Si tracci una linea verticale fino all'ascissa **Points** (la prima in alto), indicante il numero di punti prognostici che lo studente riceverà per la sua età. Si faccia lo stesso per tutte le altre variabili prognostiche (attenzione: per assegnare il punteggio prognostico relativo al voto di maturità è necessario identificare l'ascissa corrispondente al tipo di maturità conseguita).

Una volta che i punteggi prognostici sono stati determinati per tutte le caratteristiche, si provveda a calcolare il punteggio totale. Il rischio individuale di abbandono nel primo anno di corso si determina tracciando una linea verticale dal punto corrispondente al punteggio totale sull'ascissa **Total Points** (la seconda dal basso) all'ultima ascissa (**Rischio di abbandono nel 1° anno**).

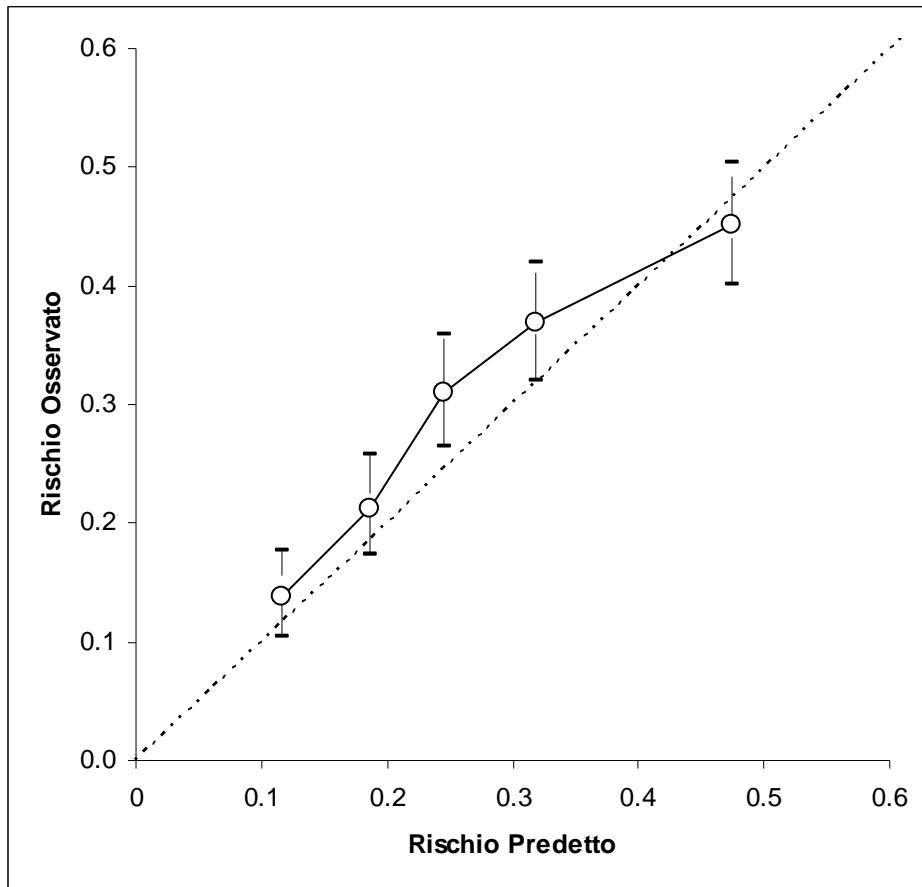
Tabella 1.3 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09). Facoltà di Economia.

	N. immatricolati 1926	N. Abbandoni 571	% Abbandono 29.6
Classe di età			
<20	1152	292	25.3
=>20	774	279	36.0
Genere			
F	868	222	25.6
M	1058	349	33.0
Residenza			
Milano	391	132	33.8
Prov. MI(no Milano)	897	247	27.5
Lombardia(no MI/Prov)	419	103	24.6
Altro	219	89	40.6
Voto Maturità *			
Alto	262	41	15.6
Basso	1664	530	31.9
Maturità			
Classica	90	26	28.9
Scientifica	691	173	25.0
Tecnico commerciale	397	111	28.0
Altre Tecniche	312	100	32.1
Magistrale	49	17	34.7
Professionale	138	63	45.7
Linguistico	84	19	22.6
Altro	165	62	37.6
CdL			
505	263	83	31.6
506	519	149	28.7
507	484	133	27.5
508	109	39	35.8
511	54	11	20.4
569	497	156	31.4
Fascia reddito #			
Bassa	456	115	25.2
Media	625	142	22.7
Alta	845	314	37.2

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta).

Figura 1.3 Calibration plot per il modello predittivo. Validation Set (coorte 08/09). Facoltà di Economia.



La linea tratteggiata indica la linea ideale, dove il rischio predetto corrisponde esattamente alle percentuali di abbandono osservate. I punti rappresentano le coppie di rischi osservati/rischi predetti per ognuno dei cinque quintili di rischio considerati. Le linee rappresentano gli IC al 95% calcolati per i rischi osservati.

Tabella 1.4 Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09). Facoltà di Economia.

Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	385	53	44.9
basso	385	82	71.5
medio	386	120	94.8
alto	385	142	122.7
molto alto	385	174	182.4

Appendice 1. Equazione del modello predittivo del rischio di abbandono. Facoltà di Economia.

$$\text{Prob}\{\text{ABB} = \text{Abbandono}\} = \frac{1}{1 + \exp(-X\hat{\beta})}, \text{ where}$$

$$\begin{aligned}
 X\hat{\beta} = & \\
 & -8.371203 + 0.63673511 \text{ETA}_I \text{SC} - 0.01055017 \text{ETA}_I \text{SC}^2 \\
 & + 0.2518818 \{\text{M}\} \\
 & - 0.1479068 \{\text{Prov MI}\} - 0.2469601 \{\text{Lomb.}\} + 0.5107376 \{\text{Altro}\} \\
 & - 0.005785645 \{504\} + 0.172535101 \{505\} + 0.063332304 \{506\} + 0.069606557 \{507\} \\
 & + 0.559373396 \{508\} - 0.212859450 \{510\} + 0.062504030 \{511\} + 0.194869700 \{569\} \\
 & + 1.25957414 \{\text{Scient}\} + 1.01446590 \{\text{Tec Com}\} + 1.26872651 \{\text{Alt Tec}\} \\
 & + 1.08336376 \{\text{Magist}\} + 1.30733790 \{\text{Prof}\} + 1.53067653 \{\text{Ling}\} + 0.04275082 \{\text{Altro}\} \\
 & - 0.01577172 \text{VOTOMAT} \\
 & + \text{VOTOMAT}[-0.026133628 \{\text{Scient}\} - 0.015361772 \{\text{Tec Com}\} - 0.016673686 \{\text{Alt Tec}\} \\
 & - 0.007661206 \{\text{Magist}\} - 0.010892178 \{\text{Prof}\} - 0.020261418 \{\text{Ling}\} \\
 & - 0.007434417 \{\text{Altro}\}]
 \end{aligned}$$

and $\{c\} = 1$ if subject is in group c , 0 otherwise.

2. Facoltà di Giurisprudenza

Tabella 2.1 Legenda dei CdL considerati. Facoltà di Giurisprudenza.

Codice CdL	Descrizione
531	Scienze Giuridiche *
532	Scienze per Operatori dei Servizi Giuridici
581	Giurisprudenza #

* CdL non più attivo

CdL a ciclo unico

Tabella 2.2 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08). Facoltà di Giurisprudenza.

	N. immatricolati	N. Abbandoni	% Abbandono
	3455	1100	31,8
Classe di età			
<20	1989	465	23,4
=20	1466	635	43,3
Genere			
F	2193	633	28,9
M	1262	467	37,0
Residenza			
Milano	458	173	37,8
Prov. MI(no Milano)	1919	591	30,8
Lombardia(no MI/Prov)	734	183	24,9
Altro	344	153	44,5
Voto Maturità *			
Alto	525	119	22,7
Basso	2930	981	33,5
Maturità			
Classica	324	59	18,2
Scientifica	746	169	22,7
Tecnico commerciale	677	250	36,9
Altre Tecniche	512	185	36,1
Magistrale	343	102	29,7
Professionale	453	206	45,5
Linguistico	232	54	23,3
Altro	168	75	44,6
Coorte			
2002	549	178	32,4
2003	576	197	34,2
2004	642	212	33,0
2005	624	184	29,5
2006	538	156	29,0
2007	526	173	32,9
CdL			
531	2033	626	30,8
532	608	264	43,4
581	814	210	25,8
Fascia reddito #			
Bassa	874	296	33,9
Media	2023	618	30,5
Alta	558	186	33,3

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta)

Figura 2.1 Relazione tra le caratteristiche all'entrata degli studenti immatricolati e il rischio di abbandono nel primo anno di corso. Stime puntuali e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08). Facoltà di Giurisprudenza

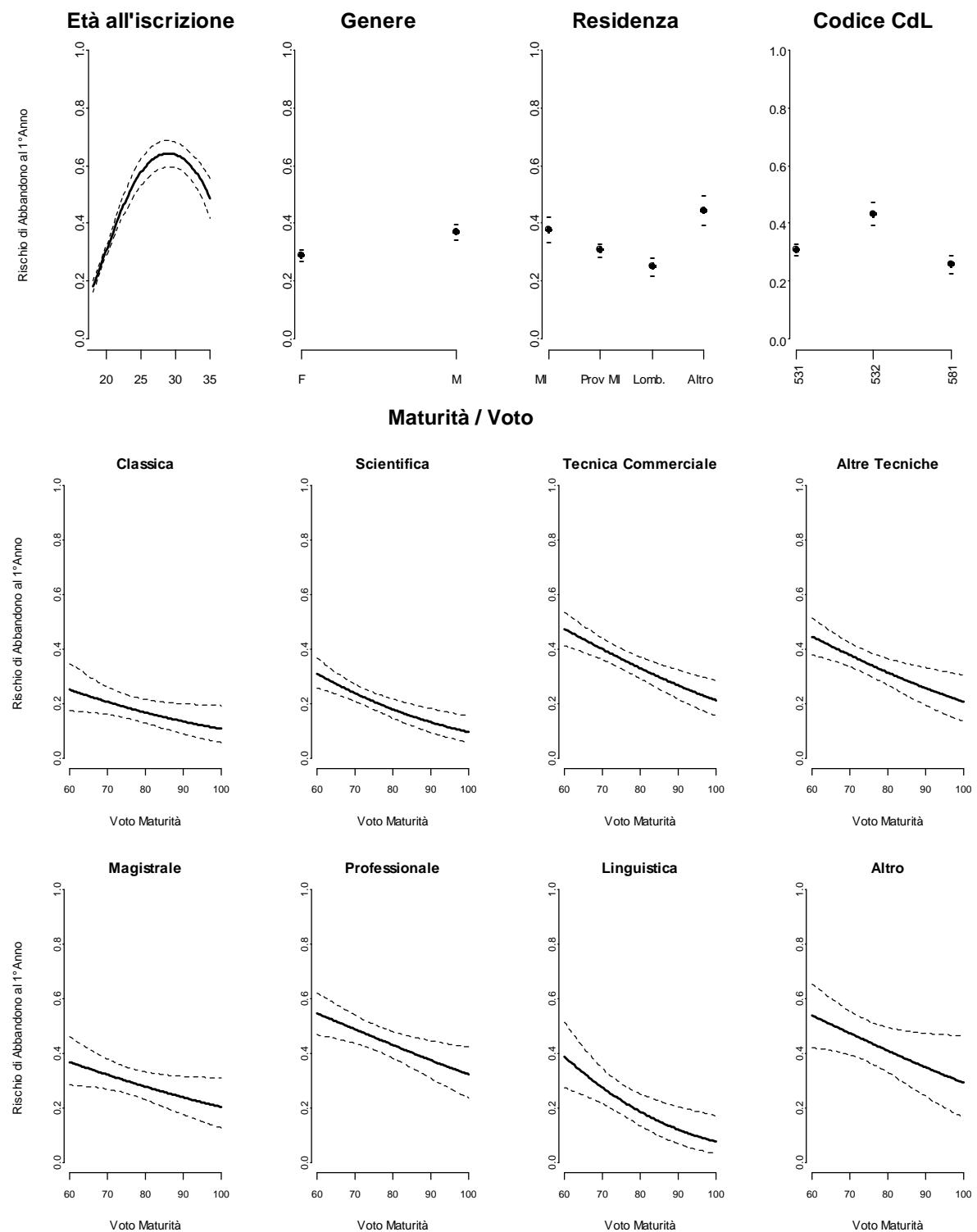
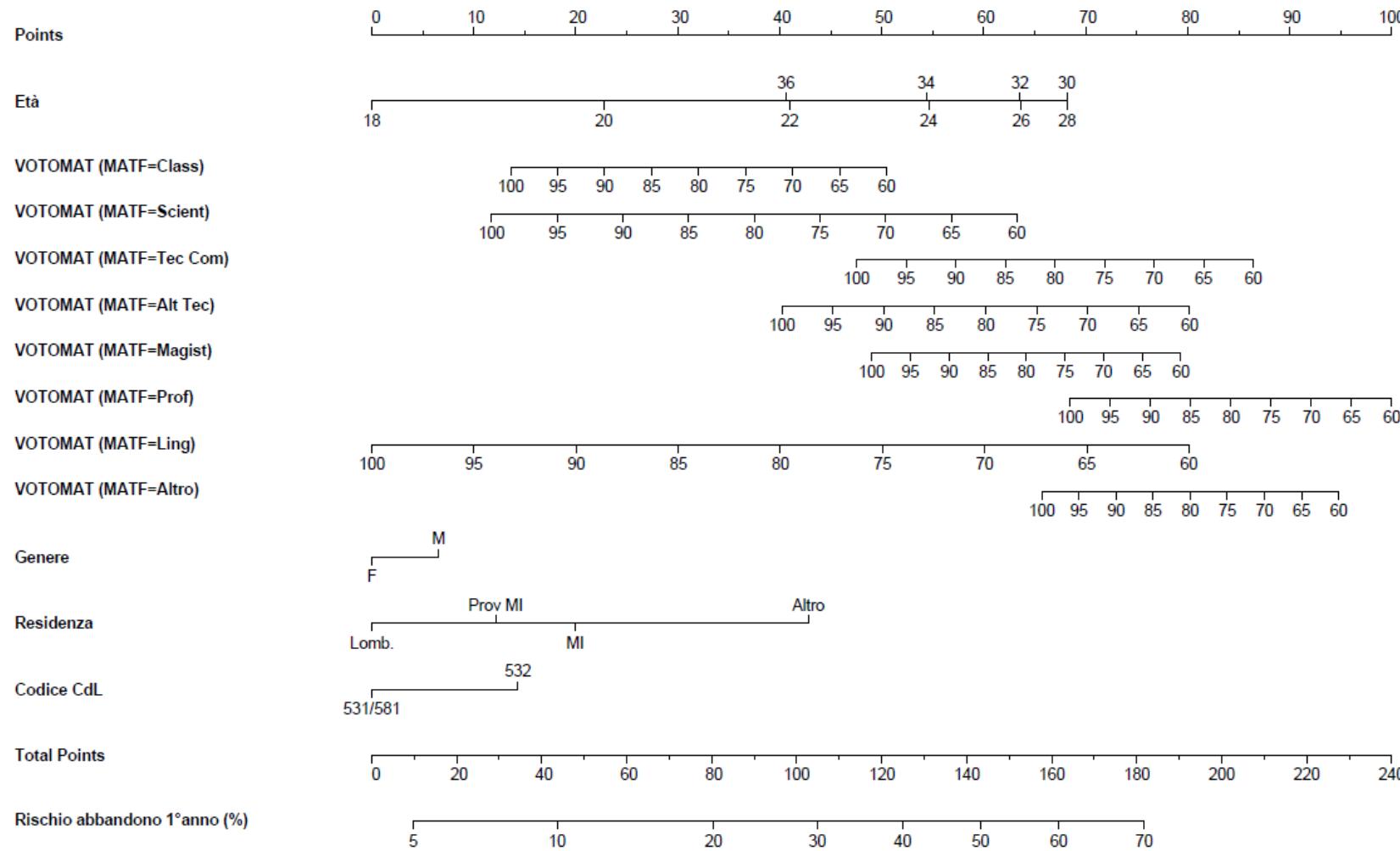


Figura 2.2 Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08). Facoltà di Giurisprudenza.



Istruzioni per l'uso del nomogramma.

Sul secondo asse delle ascisse a partire dall'alto si identifichi l'età all'immatricolazione dello studente per il quale si vuole calcolare il rischio di abbandono nel primo anno. Si tracci una linea verticale fino all'ascissa **Points** (la prima in alto), indicante il numero di punti prognostici che lo studente riceverà per la sua età. Si faccia lo stesso per tutte le altre variabili prognostiche (attenzione: per assegnare il punteggio prognostico relativo al voto di maturità è necessario identificare l'ascissa corrispondente al tipo di maturità conseguita).

Una volta che i punteggi prognostici sono stati determinati per tutte le caratteristiche, si provveda a calcolare il punteggio totale. Il rischio individuale di abbandono nel primo anno di corso si determina tracciando una linea verticale dal punto corrispondente al punteggio totale sull'ascissa **Total Points** (la seconda dal basso) all'ultima ascissa (**Rischio di abbandono nel 1° anno**).

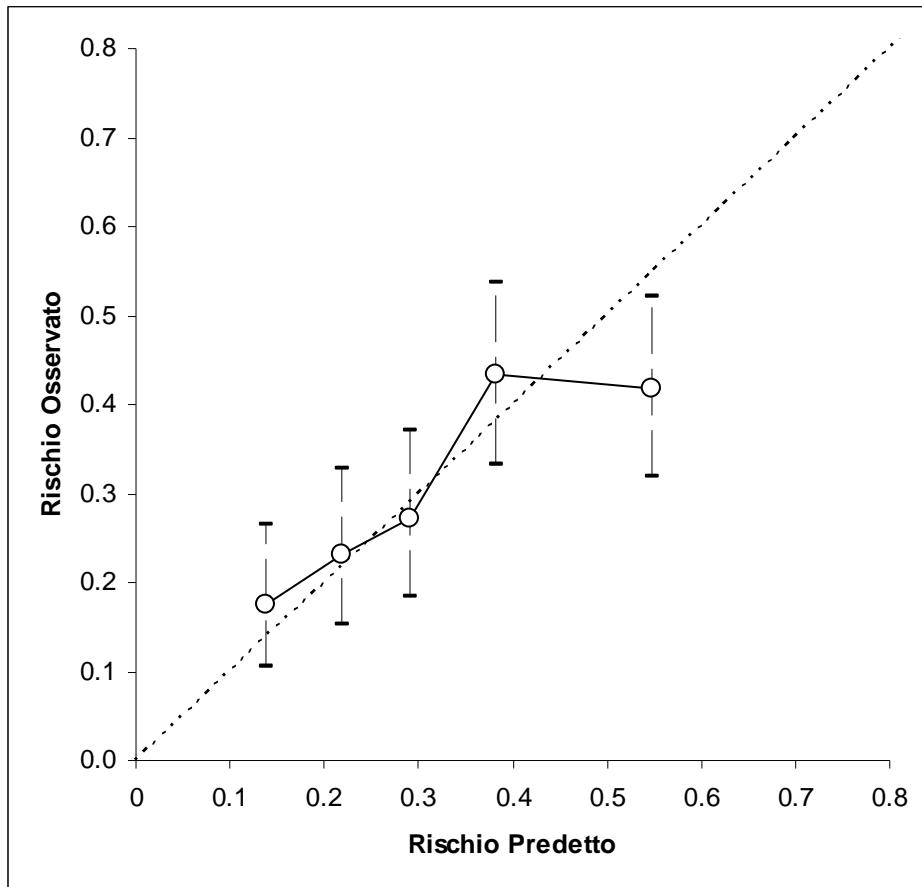
Tabella 2.3 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09). Facoltà di Giurisprudenza.

	N. immatricolati	N. Abbandoni	% Abbandono
	487	149	30,6
Classe di età			
<20	288	65	22,6
=20	199	84	42,2
Genere			
F	302	89	29,5
M	185	60	32,4
Residenza			
Milano	56	22	39,3
Prov. MI(no Milano)	270	81	30,0
Lombardia(no MI/Prov)	99	30	30,3
Altro	62	16	25,8
Voto Maturità *			
Alto	64	16	25,0
Basso	423	133	31,4
Maturità			
Classica	50	10	20,0
Scientifica	119	29	24,4
Tecnico commerciale	74	22	29,7
Altre Tecniche	87	37	42,5
Magistrale	57	15	26,3
Professionale	42	22	52,4
Linguistico	30	5	16,7
Altro	28	9	32,1
CdL			
532	132	61	46,2
581	355	88	24,8
Fascia reddito #			
Bassa	112	28	25,0
Media	177	44	24,9
Alta	198	77	38,9

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta).

Figura 2.3 Calibration plot per il modello predittivo. Validation Set (coorte 08/09). Facoltà di Giurisprudenza.



La linea tratteggiata indica la linea ideale, dove il rischio predetto corrisponde esattamente alle percentuali di abbandono osservate. I punti rappresentano le coppie di rischi osservati/rischi predetti per ognuno dei cinque quintili di rischio considerati. Le linee rappresentano gli IC al 95% calcolati per i rischi osservati.

Tabella 2.4 Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09). Facoltà di Giurisprudenza.

Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	97	17	13.4
basso	99	23	21.7
medio	96	26	27.9
alto	97	42	37.1
molto alto	98	41	53.7

Appendice 2. Equazione del modello predittivo del rischio di abbandono. Facoltà di Giurisprudenza.

$$\text{Prob}\{\text{ABB} = \text{Abbandono}\} = \frac{1}{1 + \exp(-X\hat{\beta})}, \text{ where}$$

$$\begin{aligned}
X\hat{\beta} = & \\
& -9.116268 + 0.72885944 \text{ETA}_I \text{SC} - 0.01257579 \text{ETA}_I \text{SC}^2 \\
& + 0.1440715\{\text{M}\} \\
& - 0.1719434\{\text{Prov MI}\} - 0.4395442\{\text{Lomb.}\} + 0.5041447\{\text{Altro}\} \\
& - 0.3162656\{531/581\} \\
& + 0.7658322\{\text{Scient}\} + 0.8605645\{\text{Tec Com}\} + 0.7583446\{\text{Alt Tec}\} + 0.4199387\{\text{Magist}\} \\
& + 0.9158650\{\text{Prof}\} + 2.0850183\{\text{Ling}\} + 0.7217539\{\text{Altro}\} - 0.02031729 \text{VOTOMAT} \\
& + \text{VOTOMAT}[-0.008069437 \{\text{Scient}\} - 0.001130713 \{\text{Tec Com}\} - 0.001721000 \{\text{Alt Tec}\} \\
& + 0.003603559 \{\text{Magist}\} + 0.002937129 \{\text{Prof}\} - 0.023854996 \{\text{Ling}\} \\
& + 0.004260197 \{\text{Altro}\}]
\end{aligned}$$

and $\{c\} = 1$ if subject is in group c , 0 otherwise.

3. Facoltà di Medicina e Chirurgia

Tabella 3.1 Legenda dei CdL considerati. Facoltà di Medicina e Chirurgia.

Codice CdL	Descrizione
513	Fisioterapia
514	Igiene Dentale
515	Infermieristica
518	Medicina e Chirurgia #
519	Ostetricia
538	Tecniche di Laboratorio Biomedico
539	Tecniche di Radiologia Medica, per Immagini e Radioterapia
540	Terapia della Neuro e Psicomotricità dell'Età Evolutiva
570	Odontoiatria e Protesi dentaria #

CdL a ciclo unico

Tabella 3.2 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08). Facoltà di Medicina e Chirurgia.

	N. immatricolati 3193	N. Abbandoni 346	% Abbandono 10,8
Classe di età			
<20	1928	161	8,4
≥20	1265	185	14,6
Genere			
F	2282	243	10,6
M	911	103	11,3
Residenza			
Milano	104	20	19,2
Prov. MI(no Milano)	1211	108	8,9
Lombardia(no MI/Prov)	1634	159	9,7
Altro	244	59	24,2
Voto Maturità *			
Alto	789	63	8,0
Basso	2404	283	11,8
Maturità			
Classica	226	31	13,7
Scientifica	1279	86	6,7
Tecnico commerciale	221	34	15,4
Altre Tecniche	304	48	15,8
Magistrale	426	45	10,6
Professionale	428	65	15,2
Linguistico	128	14	10,9
Altro	181	23	12,7
Coorte			
2002	529	65	12,3
2003	536	51	9,5
2004	538	54	10,0
2005	555	66	11,9
2006	498	52	10,4
2007	537	58	10,8
CdL			
513	243	9	3,7
514	106	14	13,2
515	1632	236	14,5
518	559	19	3,4
519	233	25	10,7
538	97	19	19,6
539	132	8	6,1
540	117	7	6,0
570	74	9	12,2
Fascia reddito #			
Bassa	743	98	13,2
Media	1944	179	9,2
Alta	506	69	13,6

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta)

Figura 3.1 Relazione tra le caratteristiche all'entrata degli studenti immatricolati e il rischio di abbandono nel primo anno di corso. Stime puntuali e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08). Facoltà di Medicina e Chirurgia

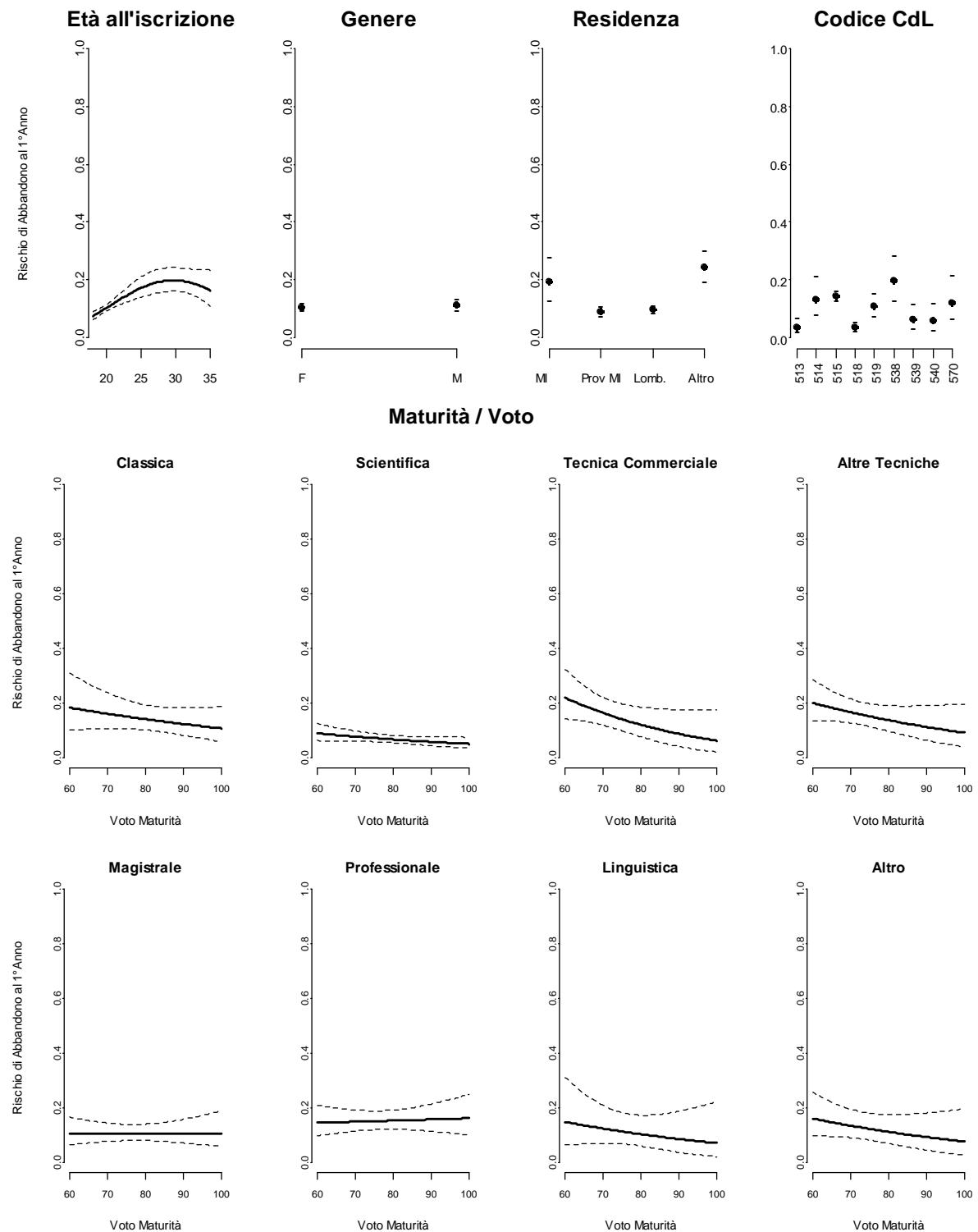
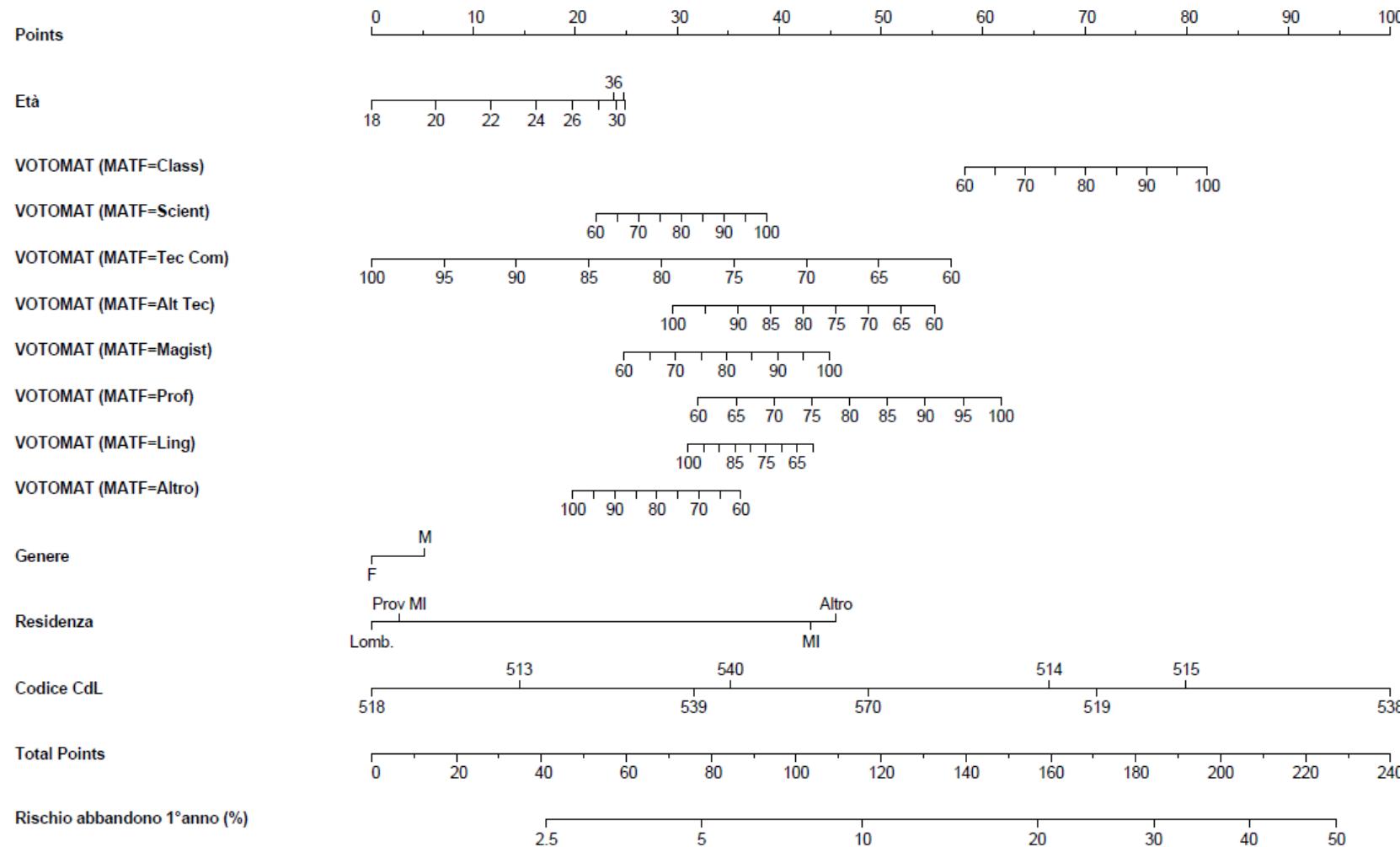


Figura 3.2 Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08). Facoltà di Medicina e Chirurgia.



Istruzioni per l'uso del nomogramma.

Sul secondo asse delle ascisse a partire dall'alto si identifichi l'età all'immatricolazione dello studente per il quale si vuole calcolare il rischio di abbandono nel primo anno. Si tracci una linea verticale fino all'ascissa **Points** (la prima in alto), indicante il numero di punti prognostici che lo studente riceverà per la sua età. Si faccia lo stesso per tutte le altre variabili prognostiche (attenzione: per assegnare il punteggio prognostico relativo al voto di maturità è necessario identificare l'ascissa corrispondente al tipo di maturità conseguita).

Una volta che i punteggi prognostici sono stati determinati per tutte le caratteristiche, si provveda a calcolare il punteggio totale. Il rischio individuale di abbandono nel primo anno di corso si determina tracciando una linea verticale dal punto corrispondente al punteggio totale sull'ascissa **Total Points** (la seconda dal basso) all'ultima ascissa (**Rischio di abbandono nel 1° anno**).

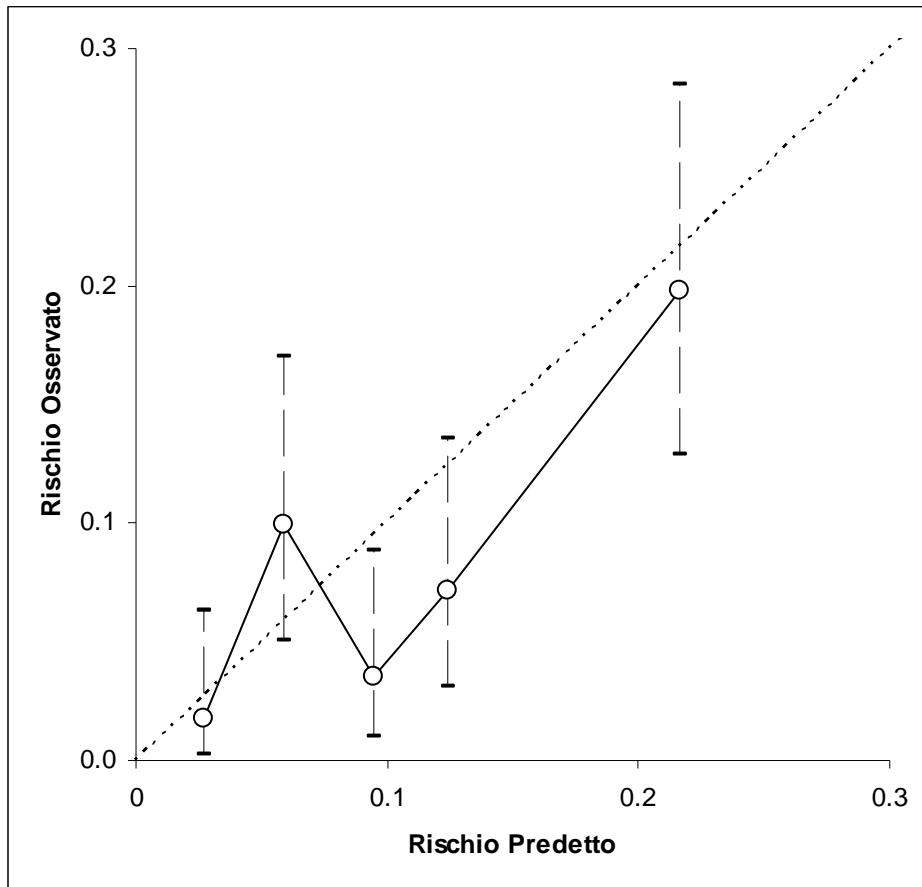
Tabella 3.3 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09). Facoltà di Medicina e Chirurgia.

	N. immatricolati	N. Abbandoni	% Abbandono
	557	47	8,4
Classe di età			
<20	370	23	6,2
=20	187	24	12,8
Genere			
F	393	32	8,1
M	164	15	9,1
Residenza			
Milano	14	4	28,6
Prov. MI(no Milano)	221	16	7,2
Lombardia(no MI/Prov)	281	20	7,1
Altro	41	7	17,1
Voto Maturità *			
Alto	140	8	5,7
Basso	417	39	9,4
Maturità			
Classica	63	9	14,3
Scientifica	266	10	3,8
Tecnico commerciale	19	3	15,8
Altre Tecniche	42	7	16,7
Magistrale	75	3	4,0
Professionale	43	8	18,6
Linguistico	14	1	7,1
Altro	35	6	17,1
CdL			
513	45	2	4,4
514	14	0	0,0
515	277	31	11,2
518	101	2	2,0
519	35	1	2,9
538	21	2	9,5
539	26	2	7,7
540	21	7	33,3
570	17	0	0,0
Fascia reddito #			
Bassa	102	16	15,7
Media	233	13	5,6
Alta	222	18	8,1

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta).

Figura 3.3 Calibration plot per il modello predittivo. Validation Set (coorte 08/09). Facoltà di Medicina e Chirurgia.



La linea tratteggiata indica la linea ideale, dove il rischio predetto corrisponde esattamente alle percentuali di abbandono osservate. I punti rappresentano le coppie di rischi osservati/rischi predetti per ognuno dei cinque quintili di rischio considerati. Le linee rappresentano gli IC al 95% calcolati per i rischi osservati.

Tabella 3.4 Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09). Facoltà di Medicina e Chirurgia.

Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	111	2	3.0
basso	111	11	6.5
medio	112	4	10.6
alto	112	8	13.9
molto alto	111	22	24.0

Appendice 3. Equazione del modello predittivo del rischio di abbandono. Facoltà di Medicina e Chirurgia.

$$\text{Prob}\{\text{ABB} = \text{Abbandono}\} = \frac{1}{1 + \exp(-X\beta)}, \text{ where}$$

$$\begin{aligned}
X\hat{\beta} = & \\
& -4.821993 + 0.147537875 \text{ETA}_I \text{SC} - 0.002251627 \text{ETA}_I \text{SC}^2 \\
& + 0.1030543\{\text{M}\} \\
& - 0.79533482\{\text{Prov MI}\} - 0.84886905\{\text{Lomb.}\} + 0.04820316\{\text{Altro}\} \\
& + 1.0237757\{514\} + 1.2875970\{515\} - 0.2855638\{518\} + 1.1153093\{519\} + 1.6828787\{538\} \\
& + 0.3362375\{539\} + 0.4072780\{540\} + 0.6730997\{570\} \\
& - 0.5060242\{\text{Scient}\} + 2.3564804\{\text{Tec Com}\} + 1.4007419\{\text{Alt Tec}\} \\
& - 0.5497853\{\text{Magist}\} - 0.6895444\{\text{Prof}\} + 0.7699112\{\text{Ling}\} + 0.7563310\{\text{Altro}\} \\
& + 0.01170098 \text{VOTOMAT} \\
& + \text{VOTOMAT}[-0.003447849 \{\text{Scient}\} - 0.039701065 \{\text{Tec Com}\} - 0.024325629 \{\text{Alt Tec}\} \\
& - 0.001800498 \{\text{Magist}\} + 0.002915519 \{\text{Prof}\} - 0.017722705 \{\text{Ling}\} \\
& - 0.019809610 \{\text{Altro}\}]
\end{aligned}$$

and $\{c\} = 1$ if subject is in group c , 0 otherwise.

4. Facoltà di Psicologia

Tabella 4.1 Legenda dei CdL considerati. Facoltà di Psicologia.

Codice CdL	Descrizione
527	Scienze e Tecniche Psicologiche
567	Discipline della Ricerca Psicologico-Sociale
X25	Scienze Comunicazione (Psicologia delle Comunicazione)

Tabella 4.2 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08). Facoltà di Psicologia.

	N. immatricolati	N. Abbandoni	% Abbandono
	4612	850	18,4
Classe di età			
<20	2414	292	12,1
≥20	2198	558	25,4
Genere			
F	3386	553	16,3
M	1226	297	24,2
Residenza			
Milano	978	234	23,9
Prov. MI(no Milano)	1728	279	16,1
Lombardia(no MI/Prov)	1391	222	16,0
Altro	515	115	22,3
Voto Maturità*			
Alto	1293	220	17,0
Basso	3319	630	19,0
Maturità			
Classica	525	97	18,5
Scientifica	1659	266	16,0
Tecnico commerciale	325	81	24,9
Altre Tecniche	518	122	23,6
Magistrale	668	82	12,3
Professionale	335	102	30,4
Linguistico	422	68	16,1
Altro	160	32	20,0
Coorte			
2002	765	131	17,1
2003	814	147	18,1
2004	816	168	20,6
2005	800	166	20,8
2006	692	107	15,5
2007	725	131	18,1
CdL			
527	3170	553	17,4
567	412	122	29,6
X25	1030	175	17,0
Fascia reddito#			
Bassa	750	126	16,8
Media	2716	417	15,4
Alta	1146	307	26,8

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta)

Figura 4.1 Relazione tra le caratteristiche all'entrata degli studenti immatricolati e il rischio di abbandono nel primo anno di corso. Stime puntuali e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08). Facoltà di Psicologia.

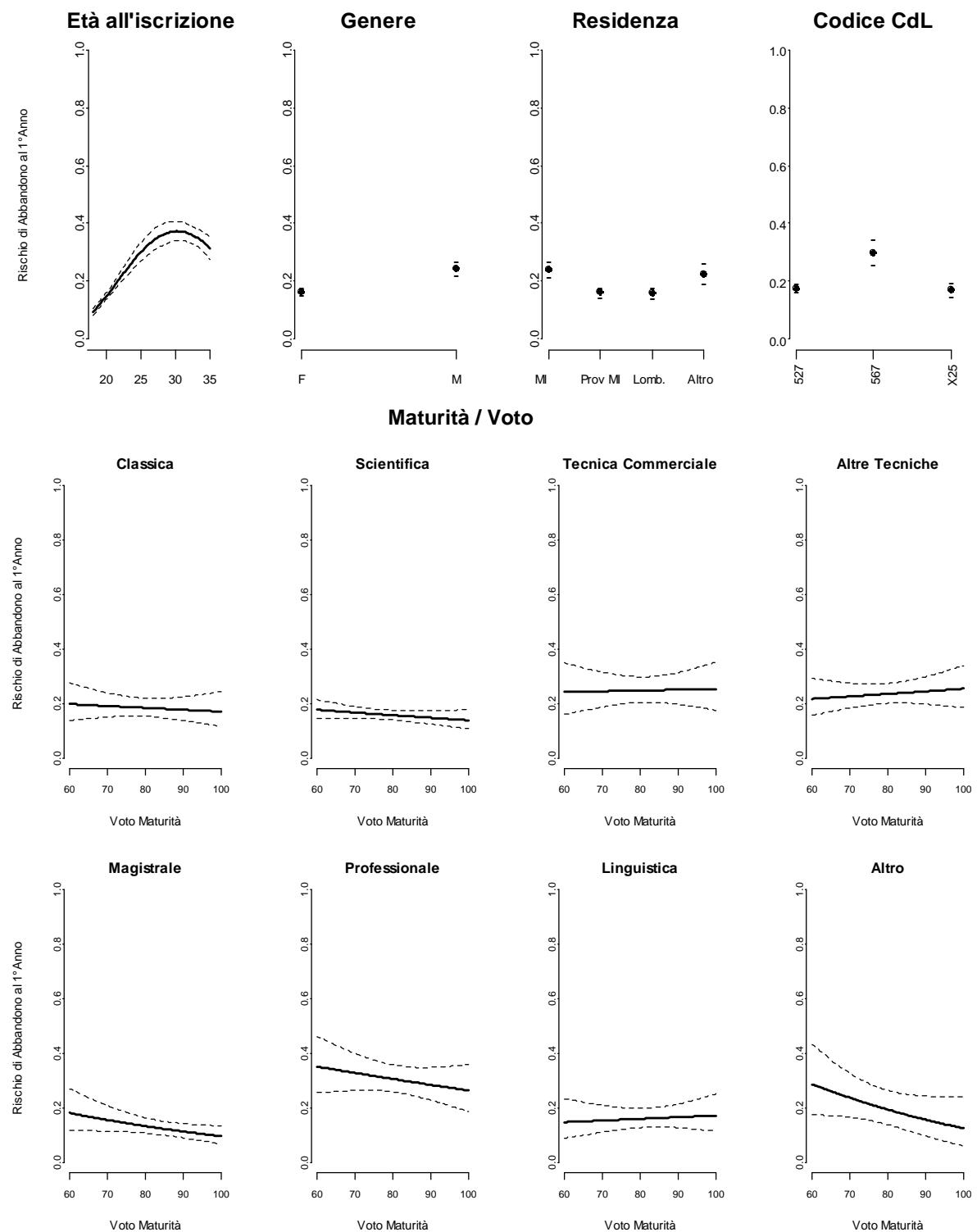
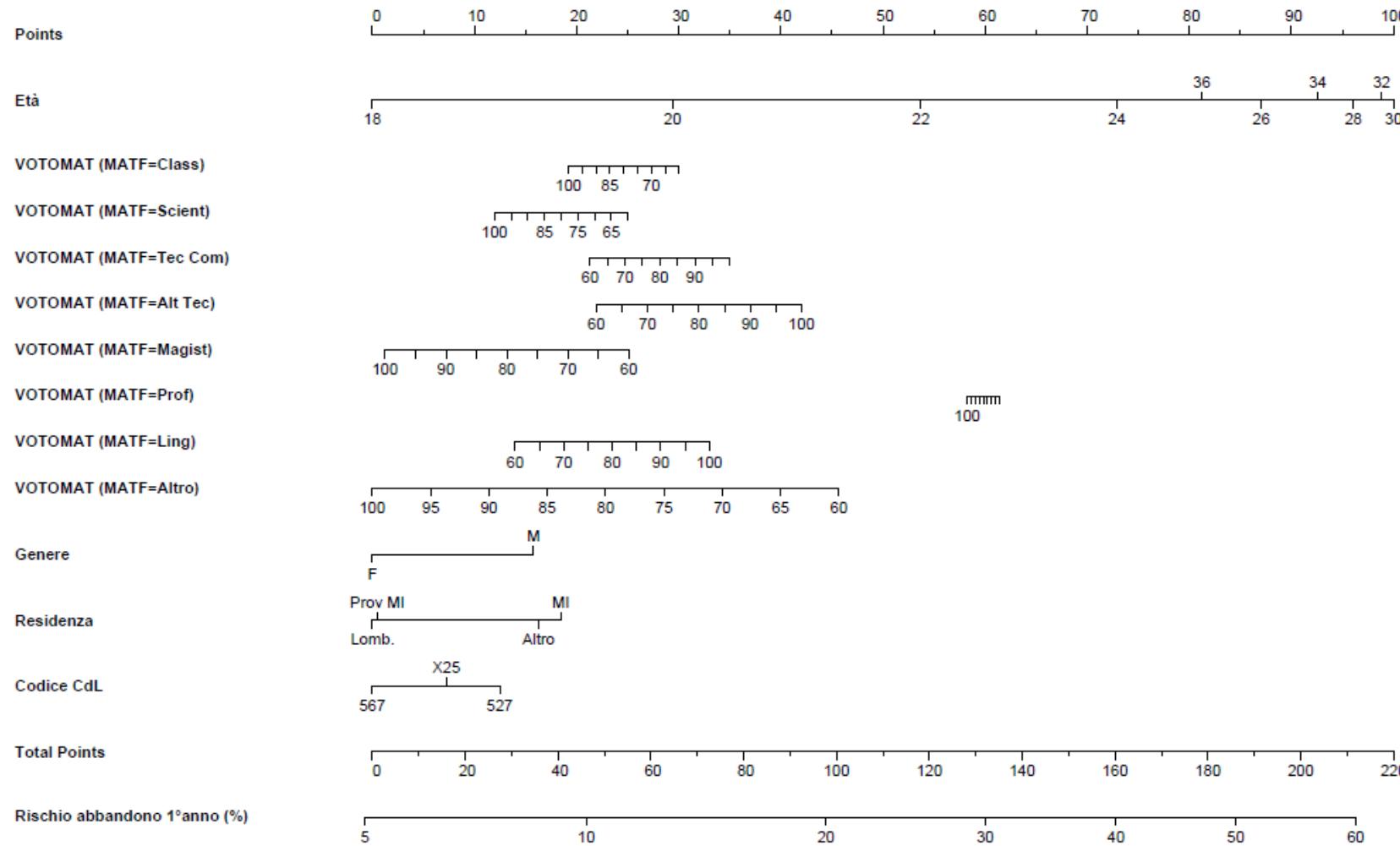


Figura 4.2 Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08). Facoltà di Psicologia.



Istruzioni per l'uso del nomogramma.

Sul secondo asse delle ascisse a partire dall'alto si identifichi l'età all'immatricolazione dello studente per il quale si vuole calcolare il rischio di abbandono nel primo anno. Si tracci una linea verticale fino all'ascissa **Points** (la prima in alto), indicante il numero di punti prognostici che lo studente riceverà per la sua età. Si faccia lo stesso per tutte le altre variabili prognostiche (attenzione: per assegnare il punteggio prognostico relativo al voto di maturità è necessario identificare l'ascissa corrispondente al tipo di maturità conseguita).

Una volta che i punteggi prognostici sono stati determinati per tutte le caratteristiche, si provveda a calcolare il punteggio totale. Il rischio individuale di abbandono nel primo anno di corso si determina tracciando una linea verticale dal punto corrispondente al punteggio totale sull'ascissa **Total Points** (la seconda dal basso) all'ultima ascissa (**Rischio di abbandono nel 1° anno**).

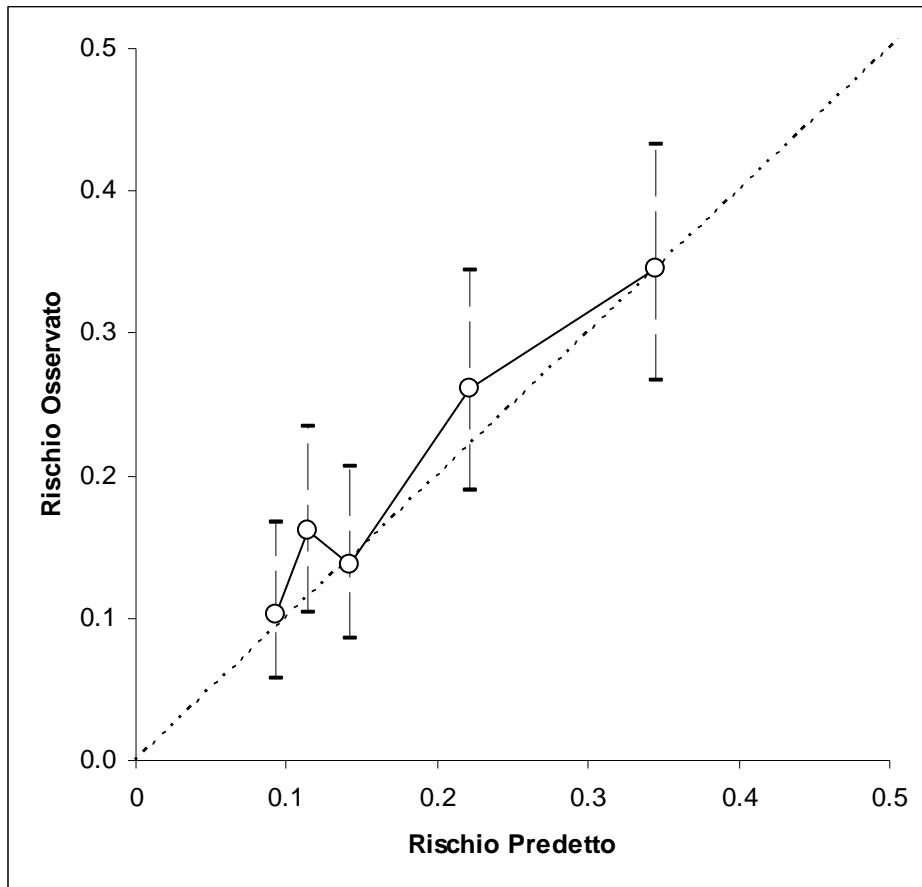
Tabella 4.3 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09). Facoltà di Psicologia.

	N. immatricolati	N. Abbandoni	% Abbandono
	680	137	20,1
Classe di età			
<20	329	44	13,4
≥20	351	93	26,5
Genere			
F	489	91	18,6
M	191	46	24,1
Residenza			
Milano	138	37	26,8
Prov. MI(no Milano)	265	46	17,4
Lombardia(no MI/Prov)	194	36	18,6
Altro	83	18	21,7
Voto Maturità			
Alto	158	34	21,5
Basso	522	103	19,7
Maturità			
Classica	98	20	20,4
Scientifica	246	41	16,7
Tecnico commerciale	33	12	36,4
Altre Tecniche	65	19	29,2
Magistrale	119	20	16,8
Professionale	38	6	15,8
Linguistico	59	12	20,3
Altro	22	7	31,8
CdL			
527	455	81	17,8
567	84	33	39,3
X25	141	23	16,3
Fascia reddito			
Bassa	85	15	17,6
Media	221	26	11,8
Alta	374	96	25,7

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta).

Figura 4.3 Calibration plot per il modello predittivo. Validation Set (coorte 08/09). Facoltà di Psicologia.



La linea tratteggiata indica la linea ideale, dove il rischio predetto corrisponde esattamente alle percentuali di abbandono osservate. I punti rappresentano le coppie di rischi osservati/rischi predetti per ognuno dei cinque quintili di rischio considerati. Le linee rappresentano gli IC al 95% calcolati per i rischi osservati.

Tabella 4.4 Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09). Facoltà di Psicologia.

Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	136	14	12.7
basso	136	22	15.6
medio	138	19	19.6
alto	134	35	29.7
molto alto	136	47	46.9

Appendice 4. Equazione del modello predittivo del rischio di abbandono. Facoltà di Psicologia.

$$\text{Prob}\{\text{ABB} = \text{Abbandono}\} = \frac{1}{1 + \exp(-X\beta)}, \text{ where}$$

$$\begin{aligned}
 X\hat{\beta} = & \\
 & -9.46234 + 0.610742578 \text{ETA}_I \text{SC} - 0.009998092 \text{ETA}_I \text{SC}^2 \\
 & + 0.2477773\{\text{M}\} \\
 & - 0.28246918\{\text{Prov MI}\} - 0.28975601\{\text{Lomb.}\} - 0.03438012\{\text{Altro}\} \\
 & - 0.1958711\{567\} - 0.0822231\{\text{X25}\} \\
 & - 0.02575472\{\text{Scient}\} - 0.71448541\{\text{Tec Com}\} - 0.85587595\{\text{Alt Tec}\} \\
 & + 0.23009295\{\text{Magist}\} + 0.31119804\{\text{Prof}\} - 0.95524739\{\text{Ling}\} + 1.06665918\{\text{Altro}\} \\
 & - 0.004250932 \text{VOTOMAT} \\
 & + \text{VOTOMAT}[-0.0008795157\{\text{Scient}\} + 0.0096258177\{\text{Tec Com}\} \\
 & + 0.0121413177\{\text{Alt Tec}\} - 0.0051205968\{\text{Magist}\} + 0.0030186651\{\text{Prof}\} \\
 & + 0.0117187753\{\text{Ling}\} - 0.0136818004\{\text{Altro}\}]
 \end{aligned}$$

and $\{c\} = 1$ if subject is in group c , 0 otherwise.

5. Facoltà di Scienze della Formazione

Tabella 5.1. Legenda dei CdL considerati. Facoltà di Scienze della Formazione.

Codice CdL	Descrizione
524	Scienze dell'Educazione
Y25	Scienze Comunicazione
R02	Scienze della Formazione Primaria #

CdL a ciclo unico

Tabella 5.2 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze della Formazione.

	N. immatricolati 5839	N. Abbandoni 1055	% Abbandono 18,1
Classe di età			
<20	2798	329	11,8
≥20	3041	726	23,9
Genere			
F	4980	851	17,1
M	859	204	23,7
Residenza			
Milano	864	187	21,6
Prov. MI(no Milano)	2327	426	18,3
Lombardia(no MI/Prov)	2062	298	14,5
Altro	586	144	24,6
Voto Maturità*			
Alto	987	123	12,5
Basso	4852	932	19,2
Maturità			
Classica	316	56	17,7
Scientifica	999	134	13,4
Tecnico commerciale	356	93	26,1
Altre Tecniche	708	155	21,9
Magistrale	2086	314	15,1
Professionale	742	193	26,0
Linguistico	436	68	15,6
Altro	196	42	21,4
Coorte			
2002	1070	211	19,7
2003	987	175	17,7
2004	1042	194	18,6
2005	924	181	19,6
2006	809	159	19,7
2007	1007	135	13,4
CdL			
524	3134	547	17,5
R02	1805	316	17,5
Y25	900	192	21,3
Fascia reddito#			
Bassa	1280	238	18,6
Media	3873	648	16,7
Alta	686	169	24,6

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta)

Figura 5.1 Relazione tra le caratteristiche all'entrata degli studenti immatricolati e il rischio di abbandono nel primo anno di corso. Stime puntuali e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze della Formazione.

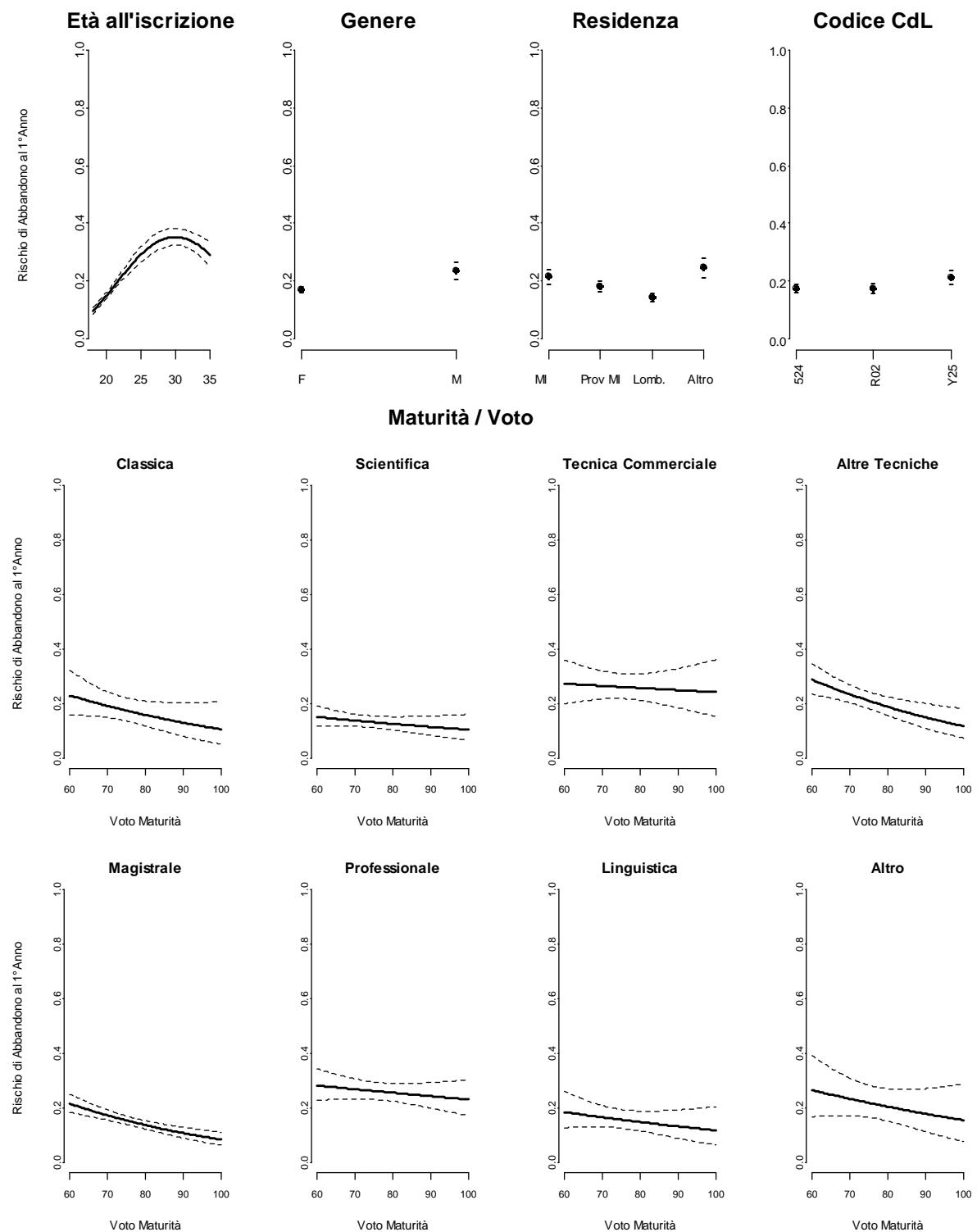
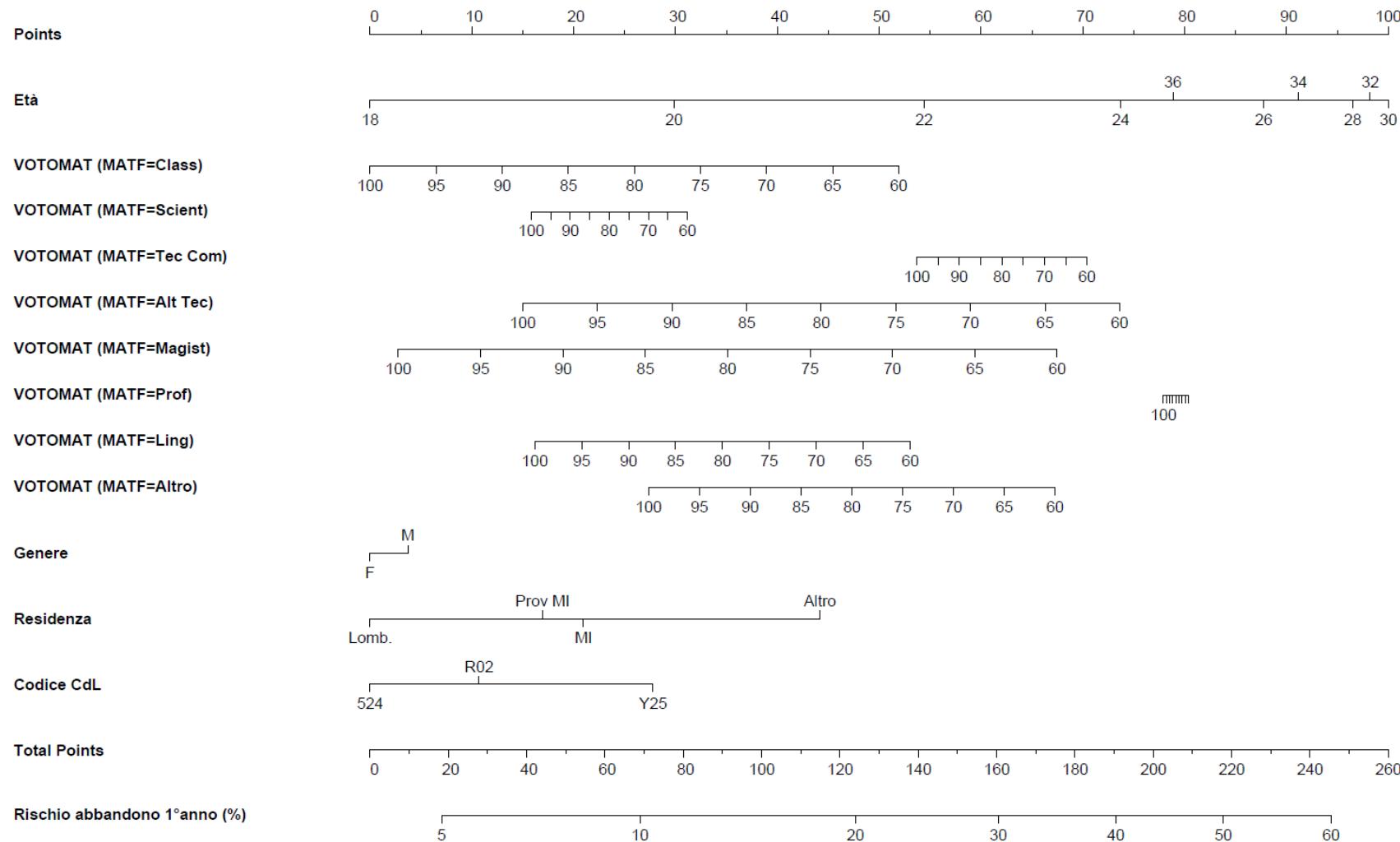


Figura 5.2 Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze della Formazione.



Istruzioni per l'uso del nomogramma.

Sul secondo asse delle ascisse a partire dall'alto si identifichi l'età all'immatricolazione dello studente per il quale si vuole calcolare il rischio di abbandono nel primo anno. Si tracci una linea verticale fino all'ascissa **Points** (la prima in alto), indicante il numero di punti prognostici che lo studente riceverà per la sua età. Si faccia lo stesso per tutte le altre variabili prognostiche (attenzione: per assegnare il punteggio prognostico relativo al voto di maturità è necessario identificare l'ascissa corrispondente al tipo di maturità conseguita).

Una volta che i punteggi prognostici sono stati determinati per tutte le caratteristiche, si provveda a calcolare il punteggio totale. Il rischio individuale di abbandono nel primo anno di corso si determina tracciando una linea verticale dal punto corrispondente al punteggio totale sull'ascissa **Total Points** (la seconda dal basso) all'ultima ascissa (**Rischio di abbandono nel 1° anno**).

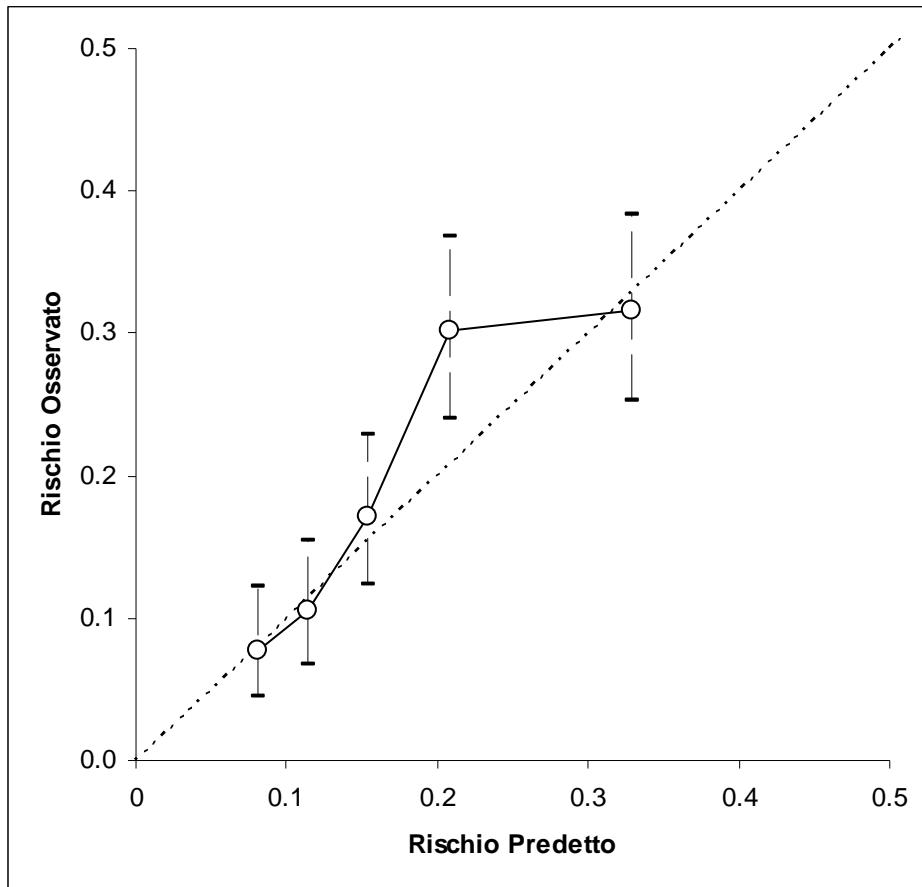
Tabella 5.3 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09). Facoltà di Scienze della Formazione.

	N. immatricolati	N. Abbandoni	% Abbandono
	1046	203	19,4
Classe di età			
<20	474	61	12,9
=20	572	142	24,8
Genere			
F	940	177	18,8
M	106	26	24,5
Residenza			
Milano	148	34	23,0
Prov. MI(no Milano)	383	81	21,1
Lombardia(no MI/Prov)	395	60	15,2
Altro	120	28	23,3
Voto Maturità*			
Alto	161	18	11,2
Basso	885	185	20,9
Maturità			
Classica	59	8	13,6
Scientifica	175	31	17,7
Tecnico commerciale	56	13	23,2
Altre Tecniche	94	26	27,7
Magistrale	415	62	14,9
Professionale	139	48	34,5
Linguistico	73	8	11,0
Altro	35	7	20,0
CdL			
524	598	128	21,4
R02	369	62	16,8
Y25	79	13	16,5
Fascia reddito#			
Bassa	214	37	17,3
Media	474	74	15,6
Alta	358	92	25,7

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta).

Figura 5.3 Calibration plot per il modello predittivo. Validation Set (coorte 08/09). Facoltà di Scienze della Formazione.



La linea tratteggiata indica la linea ideale, dove il rischio predetto corrisponde esattamente alle percentuali di abbandono osservate. I punti rappresentano le coppie di rischi osservati/rischi predetti per ognuno dei cinque quintili di rischio considerati. Le linee rappresentano gli IC al 95% calcolati per i rischi osservati.

Tabella 5.4 Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09). Facoltà di Scienze della Formazione.

Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	208	16	16.8
basso	210	22	24.0
medio	210	36	32.4
alto	209	63	43.5
molto alto	209	66	68.8

Appendice 5. Equazione del modello predittivo del rischio di abbandono. Facoltà di Scienze della Formazione.

$$\text{Prob}\{\text{ABB} = \text{Abbandono}\} = \frac{1}{1 + \exp(-X\beta)}, \text{ where}$$

$$\begin{aligned}
X\hat{\beta} = & \\
& -8.45038 + 0.58939453 \text{ETA}_I \text{SC} - 0.00971814 \text{ETA}_I \text{SC}^2 \\
& + 0.05484024 \{\text{M}\} \\
& - 0.05952527 \{\text{Prov MI}\} - 0.30917373 \{\text{Lomb.}\} + 0.34248654 \{\text{Altro}\} \\
& + 0.1578077 \{\text{R02}\} + 0.4099479 \{\text{Y25}\} \\
& - 1.11533474 \{\text{Scient}\} - 0.50653586 \{\text{Tec Com}\} + 0.46776546 \{\text{Alt Tec}\} \\
& + 0.51212983 \{\text{Magist}\} - 0.67514232 \{\text{Prof}\} - 0.31669637 \{\text{Ling}\} - 0.04005676 \{\text{Altro}\} \\
& - 0.01914945 \text{VOTOMAT} \\
& + \text{VOTOMAT}[0.013487086 \{\text{Scient}\} + 0.012984439 \{\text{Tec Com}\} - 0.002465336 \{\text{Alt Tec}\} \\
& - 0.004711886 \{\text{Magist}\} + 0.018242165 \{\text{Prof}\} + 0.005553318 \{\text{Ling}\} \\
& + 0.004437756 \{\text{Altro}\}]
\end{aligned}$$

and $\{c\} = 1$ if subject is in group c , 0 otherwise.

6. Facoltà di Scienze Matematiche, Fisiche e Naturali

Tabella 6.1 Legenda dei CdL considerati. Facoltà di Scienze Matematiche, Fisiche e Naturali.

Codice CdL	Descrizione
501	Bioteconomie
502	Scienze e Tecnologie Chimiche
512	Fisica
516	Informatica
517	Matematica
520	Ottica e Optometria
521	Scienze dei Materiali
522	Scienze Biologiche
528	Scienze e Tecnologie Geologiche
529	Scienze e Tecnologie Orafe
530	Scienze e Tecnologie per l'Ambiente

Tabella 6.2 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze Matematiche, Fisiche e Naturali.

	N. immatricolati 8167	N. Abbandoni 2537	% Abbandono 31,1
Classe di età			
<20	5598	1542	27,5
=>20	2569	995	38,7
Genere			
F	3247	1125	34,6
M	4920	1412	28,7
Residenza			
Milano	1202	396	32,9
Prov. MI(no Milano)	3771	1110	29,4
Lombardia(no MI/Prov)	2376	712	30,0
Altro	818	319	39,0
Voto Maturità*			
Alto	1897	434	22,9
Basso	6270	2103	33,5
Maturità			
Classica	464	172	37,1
Scientifica	3751	864	23,0
Tecnico commerciale	462	153	33,1
Altre Tecniche	1818	547	30,1
Magistrale	342	197	57,6
Professionale	778	359	46,1
Linguistico	263	104	39,5
Altro	289	141	48,8
Coorte			
2002	1212	363	30,0
2003	1222	330	27,0
2004	1447	480	33,2
2005	1444	448	31,0
2006	1363	402	29,5
2007	1479	514	34,8
CdL			
501	1638	475	29,0
502	221	32	14,5
512	589	119	20,2
516	2053	515	25,1
517	357	85	23,8
520	351	149	42,5
521	398	124	31,2
522	1646	779	47,3
528	281	71	25,3
529	121	41	33,9
530	512	147	28,7
Fascia reddito#			
Bassa	1457	501	34,4
Media	5158	1503	29,1
Alta	1552	533	34,3

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta)

Figura 6.1 Relazione tra le caratteristiche all'entrata degli studenti immatricolati e il rischio di abbandono nel primo anno di corso. Stime puntuali e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze Matematiche, Fisiche e Naturali.

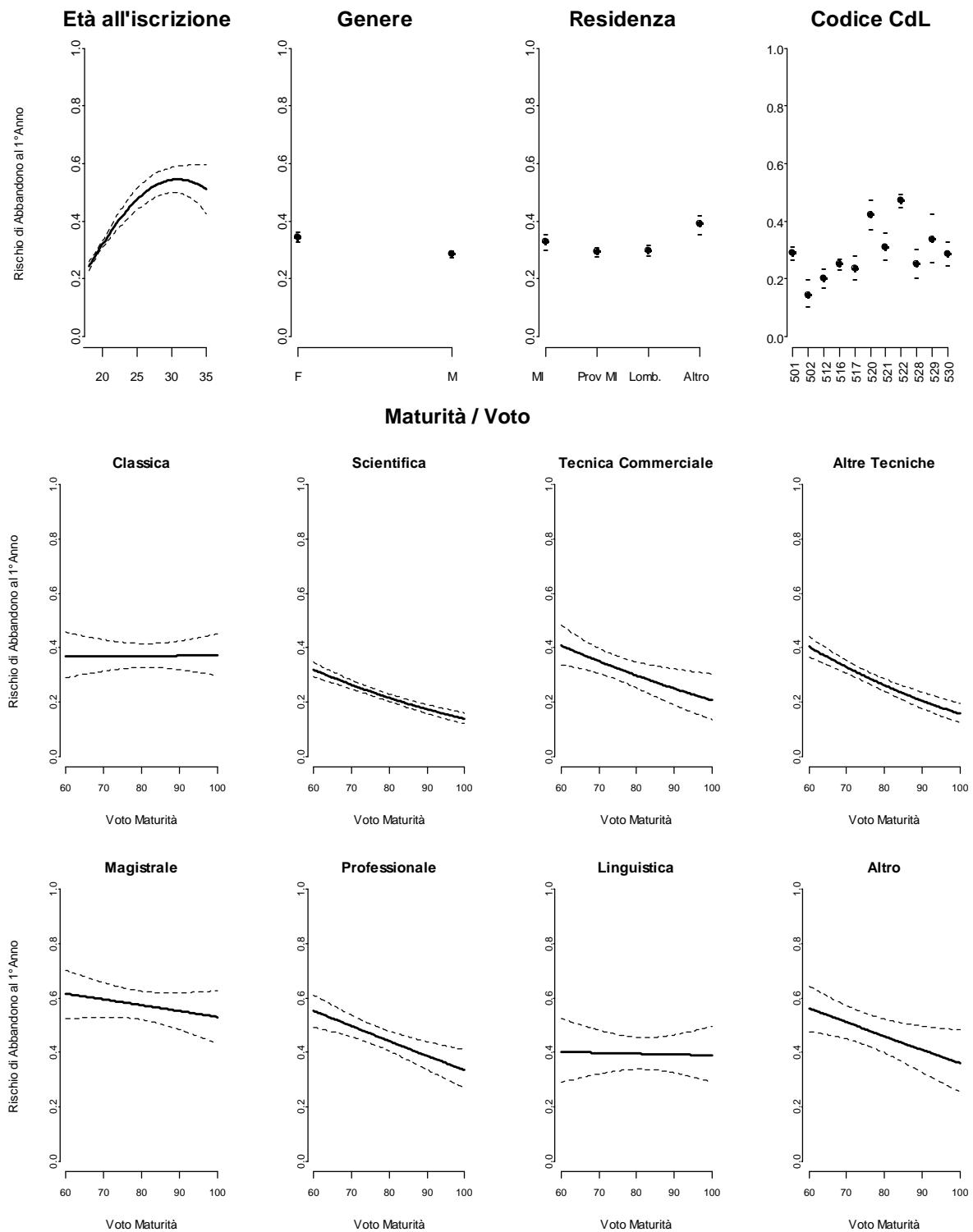
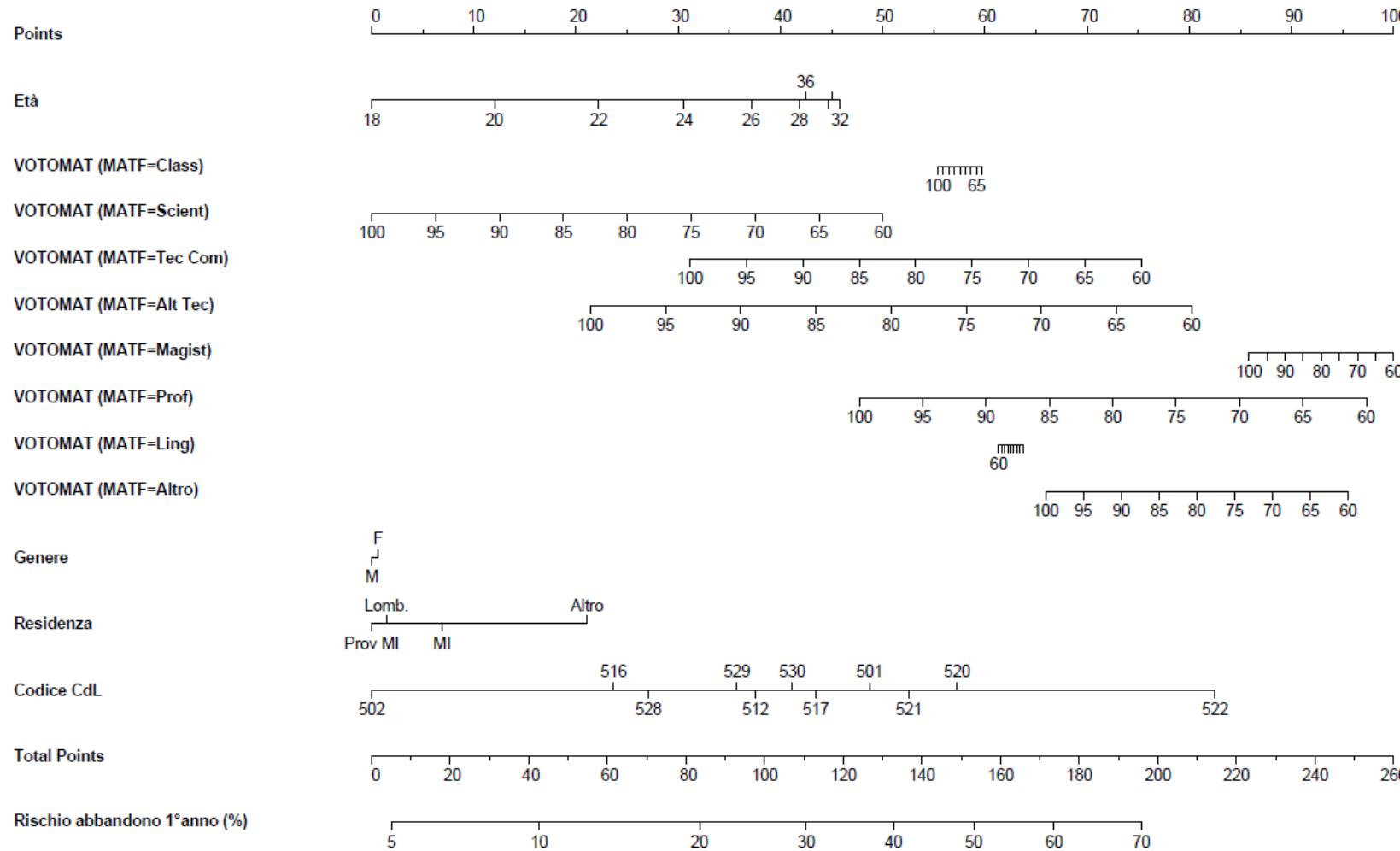


Figura 6.2 Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze Matematiche, Fisiche e Naturali.



Istruzioni per l'uso del nomogramma.

Sul secondo asse delle ascisse a partire dall'alto si identifichi l'età all'immatricolazione dello studente per il quale si vuole calcolare il rischio di abbandono nel primo anno. Si tracci una linea verticale fino all'ascissa **Points** (la prima in alto), indicante il numero di punti prognostici che lo studente riceverà per la sua età. Si faccia lo stesso per tutte le altre variabili prognostiche (attenzione: per assegnare il punteggio prognostico relativo al voto di maturità è necessario identificare l'ascissa corrispondente al tipo di maturità conseguita).

Una volta che i punteggi prognostici sono stati determinati per tutte le caratteristiche, si provveda a calcolare il punteggio totale. Il rischio individuale di abbandono nel primo anno di corso si determina tracciando una linea verticale dal punto corrispondente al punteggio totale sull'ascissa **Total Points** (la seconda dal basso) all'ultima ascissa (**Rischio di abbandono nel 1° anno**).

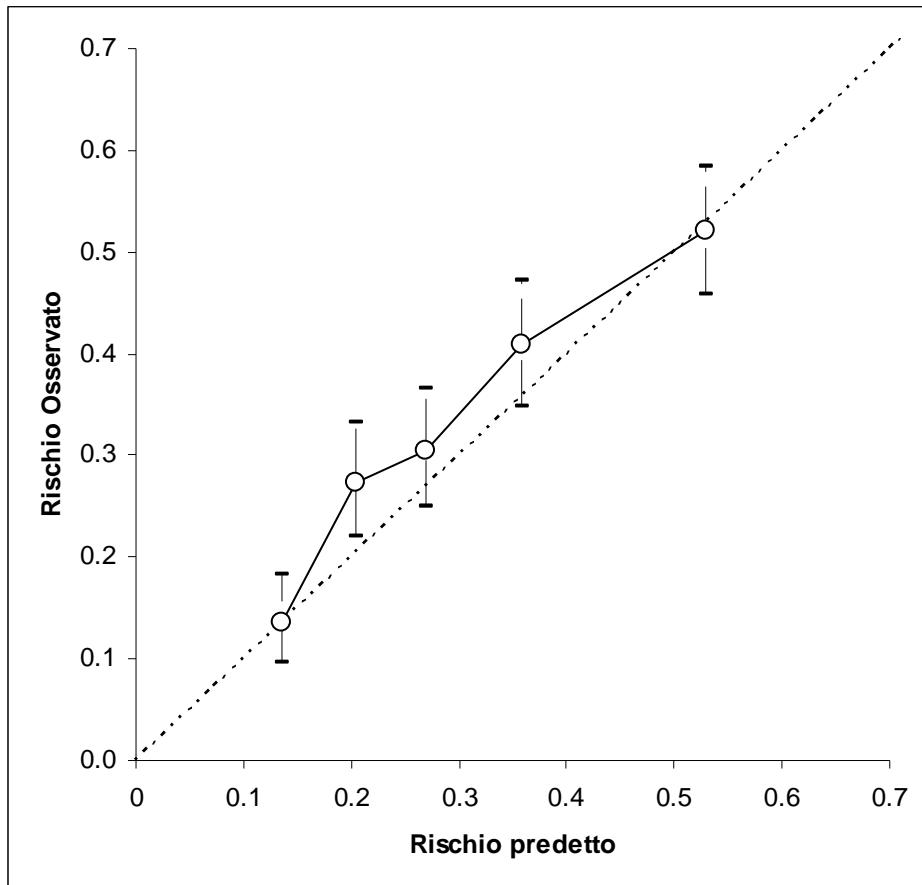
Tabella 6.3 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09). Facoltà di Scienze Matematiche, Fisiche e Naturali.

	N. immatricolati	N. Abbandoni	% Abbandono
	1294	426	32,9
Classe di età			
<20	915	281	30,7
=>20	379	145	38,3
Genere			
F	567	208	36,7
M	727	218	30,0
Residenza			
Milano	175	55	31,4
Prov. MI(no Milano)	521	174	33,4
Lombardia(no MI/Prov)	443	139	31,4
Altro	155	58	37,4
Voto Maturità*			
Alto	263	66	25,1
Basso	1031	360	34,9
Maturità			
Classica	86	41	47,7
Scientifica	714	187	26,2
Tecnico commerciale	59	24	40,7
Altre Tecniche	225	65	28,9
Magistrale	50	26	52,0
Professionale	76	38	50,0
Linguistico	38	18	47,4
Altro	46	27	58,7
CdL			
501	259	100	38,6
502	48	10	20,8
512	106	14	13,2
516	247	68	27,5
517	99	34	34,3
520	69	23	33,3
521	41	11	26,8
522	262	120	45,8
528	59	17	28,8
529	21	6	28,6
530	83	23	27,7
Fascia reddito#			
Bassa	226	78	34,5
Media	478	141	29,5
Alta	590	207	35,1

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta).

Figura 6.3 Calibration plot per il modello predittivo. Validation Set (coorte 08/09). Facoltà di Scienze Matematiche, Fisiche e Naturali.



La linea tratteggiata indica la linea ideale, dove il rischio predetto corrisponde esattamente alle percentuali di abbandono osservate. I punti rappresentano le coppie di rischi osservati/rischi predetti per ognuno dei cinque quintili di rischio considerati. Le linee rappresentano gli IC al 95% calcolati per i rischi osservati.

Tabella 6.4 Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09). Facoltà di Scienze Matematiche, Fisiche e Naturali.

Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	258	35	35.0
basso	259	71	53.1
medio	259	79	69.7
alto	259	106	92.6
molto alto	259	135	137.1

Appendice 6. Equazione del modello predittivo del rischio di abbandono. Facoltà di Scienze Matematiche, Fisiche e Naturali.

$$\text{Prob}\{\text{ABB} = \text{Abbandono}\} = \frac{1}{1 + \exp(-X\beta)}, \text{ where}$$

$$\begin{aligned}
X\hat{\beta} = & \\
& -4.392633 + 0.291796686 \text{ETA}_I \text{SC} - 0.004535604 \text{ETA}_I \text{SC}^2 \\
& -0.01281912\{\text{M}\} \\
& -0.1370683\{\text{Prov MI}\} - 0.1087316\{\text{Lomb.}\} + 0.2819036\{\text{Altro}\} \\
& -0.96840235\{502\} - 0.22218707\{512\} - 0.49779482\{516\} - 0.10489404\{517\} \\
& +0.17007150\{520\} + 0.07648006\{521\} + 0.67169604\{522\} - 0.43005866\{528\} \\
& -0.25790525\{529\} - 0.15014035\{530\} \\
& +1.1691556\{\text{Scient}\} + 1.4971752\{\text{Tec Com}\} + 2.0317270\{\text{Alt Tec}\} \\
& +1.0941111\{\text{Magist}\} + 2.0958341\{\text{Prof}\} - 0.1679508\{\text{Ling}\} + 1.4651198\{\text{Altro}\} \\
& -0.002156339 \text{VOTOMAT} \\
& +\text{VOTOMAT}[-0.022699121 \{\text{Scient}\} - 0.019788826 \{\text{Tec Com}\} - 0.027071994 \{\text{Alt Tec}\} \\
& -0.004892389 \{\text{Magist}\} - 0.022483331 \{\text{Prof}\} + 0.003335527 \{\text{Ling}\} \\
& -0.012556417 \{\text{Altro}\}]
\end{aligned}$$

and $\{c\} = 1$ if subject is in group c , 0 otherwise.

7. Facoltà di Scienze Statistiche

Tabella 7.1 Legenda dei CdL considerati. Facoltà di Scienze Statistiche.

Codice CdL	Descrizione
533	Scienze Statistiche ed Economiche
534	Scienze Statistiche Demografiche e Sociali*
537	Statistica*
580	Statistica e Gestione delle Informazioni

* CdL non più attivo

Tabella 7.2 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze Statistiche.

	N. immatricolati	N. Abbandoni	% Abbandono
	766	181	23,6
Classe di età			
<20	564	105	18,6
≥20	202	76	37,6
Genere			
F	361	80	22,2
M	405	101	24,9
Residenza			
Milano	110	24	21,8
Prov. MI(no Milano)	342	81	23,7
Lombardia(no MI/Prov)	240	54	22,5
Altro	74	22	29,7
Voto Maturità*			
Alto	181	30	16,6
Basso	585	151	25,8
Maturità			
Classica	30	9	30,0
Scientifica	326	43	13,2
Tecnico commerciale	95	35	36,8
Altre Tecniche	170	48	28,2
Magistrale	44	9	20,5
Professionale	44	20	45,5
Linguistico	36	6	16,7
Altro	21	11	52,4
Coorte			
2002	97	24	24,7
2003	109	16	14,7
2004	144	43	29,9
2005	129	24	18,6
2006	145	32	22,1
2007	142	42	29,6
CdL			
533	459	114	24,8
534	74	14	18,9
537	92	28	30,4
580	141	25	17,7
Fascia reddito#			
Bassa	123	30	24,4
Media	505	111	22,0
Alta	138	40	29,0

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta)

Figura 7.1 Relazione tra le caratteristiche all'entrata degli studenti immatricolati e il rischio di abbandono nel primo anno di corso. Stime puntuali e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze Statistiche.

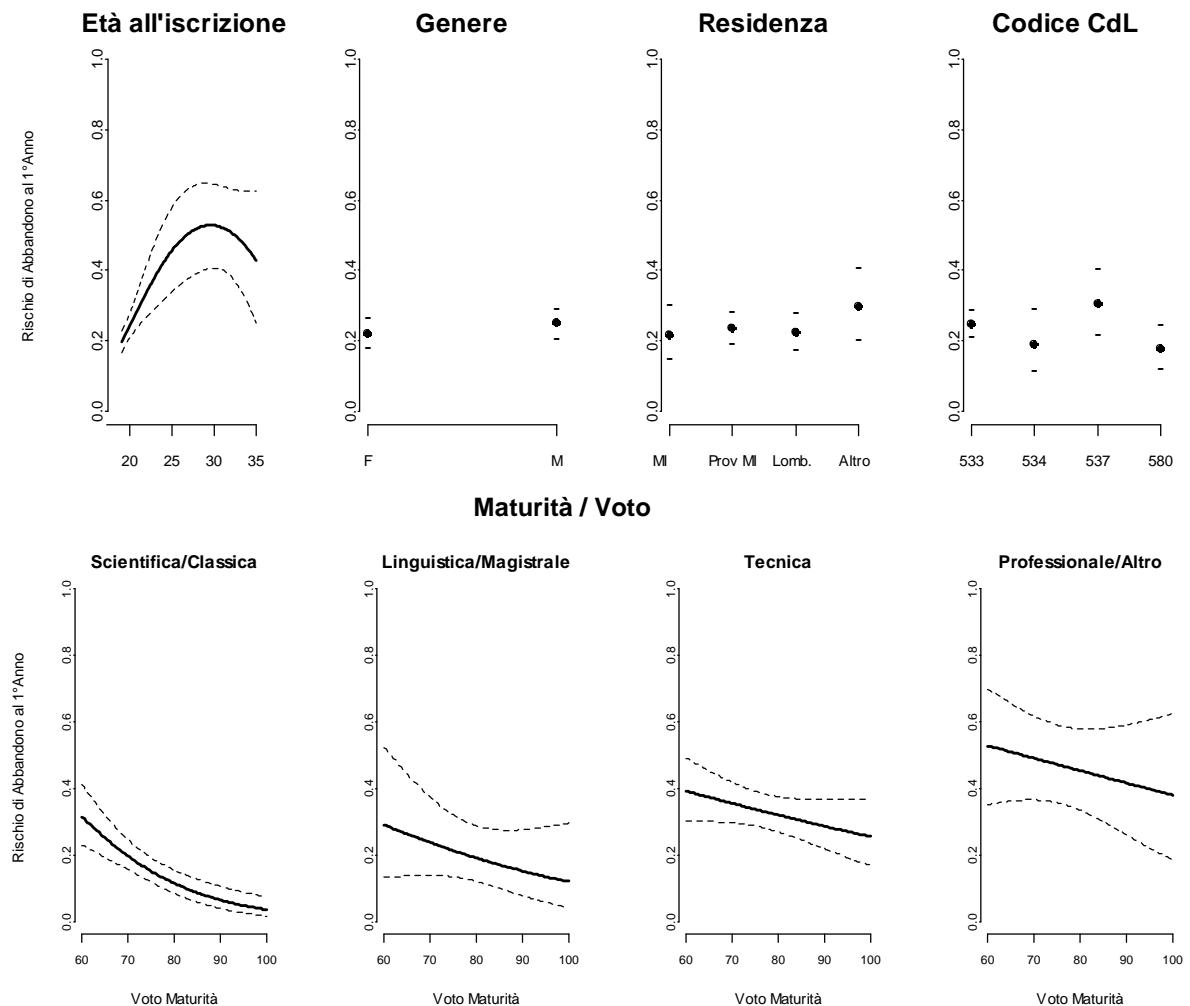
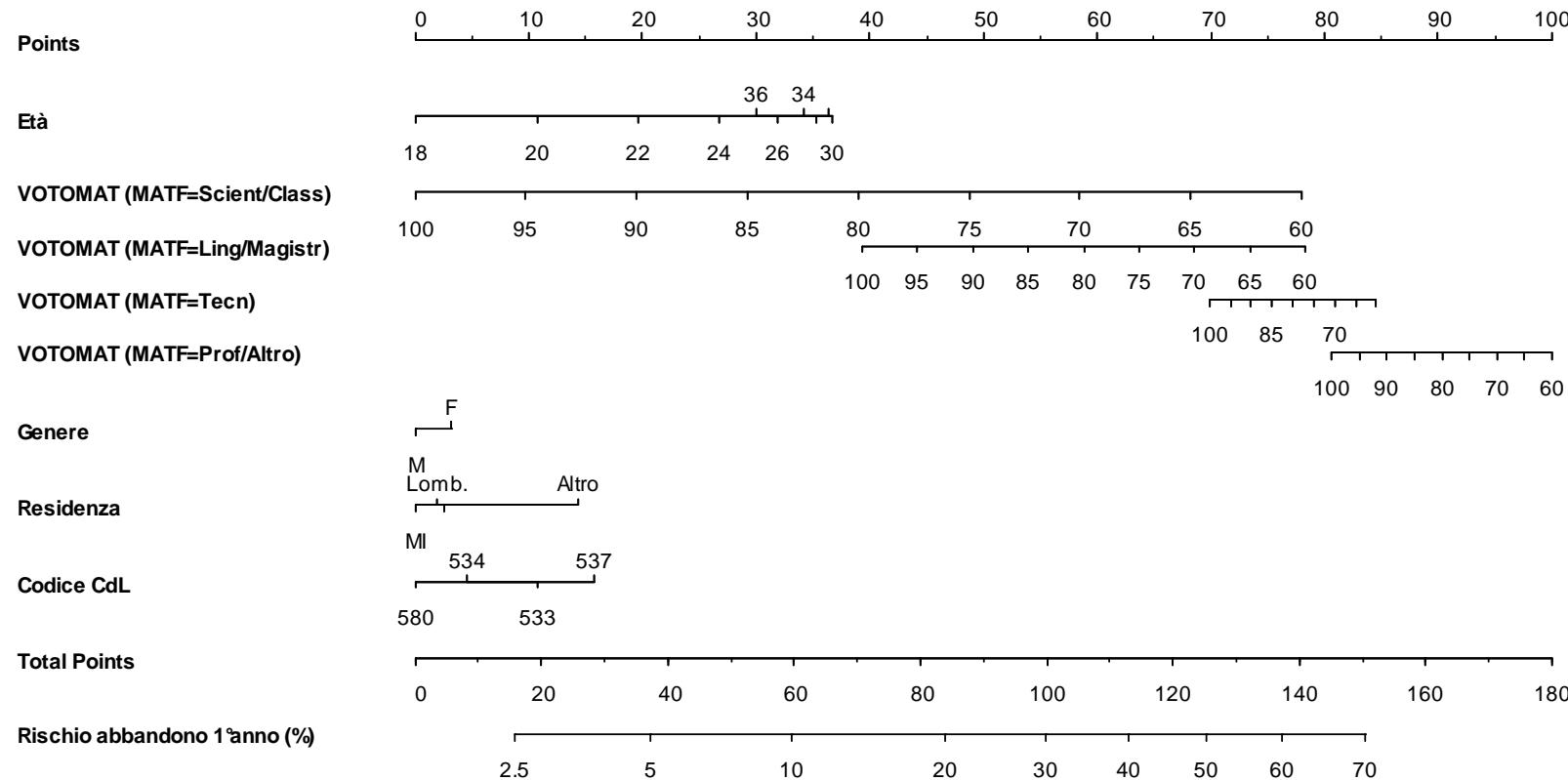


Figura 7.2 Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08). Facoltà di Scienze Statistiche.



Istruzioni per l'uso del nomogramma.

Sul secondo asse delle ascisse a partire dall'alto si identifichi l'età all'immatricolazione dello studente per il quale si vuole calcolare il rischio di abbandono nel primo anno. Si tracci una linea verticale fino all'ascissa **Points** (la prima in alto), indicante il numero di punti prognostici che lo studente riceverà per la sua età. Si faccia lo stesso per tutte le altre variabili prognostiche (attenzione: per assegnare il punteggio prognostico relativo al voto di maturità è necessario identificare l'ascissa corrispondente al tipo di maturità conseguita).

Una volta che i punteggi prognostici sono stati determinati per tutte le caratteristiche, si provveda a calcolare il punteggio totale. Il rischio individuale di abbandono nel primo anno di corso si determina tracciando una linea verticale dal punto corrispondente al punteggio totale sull'ascissa **Total Points** (la seconda dal basso) all'ultima ascissa (**Rischio di abbandono nel 1° anno**).

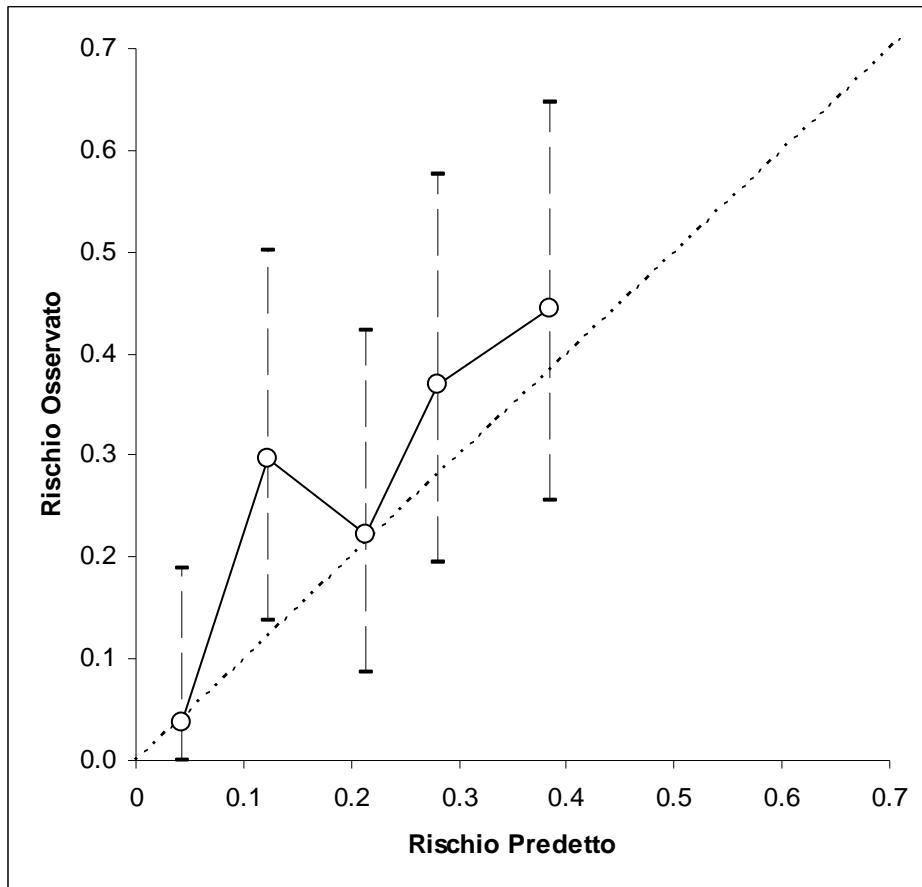
Tabella 7.3 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09). Facoltà di Scienze Statistiche.

	N. immatricolati	N. Abbandoni	% Abbandono
	135	37	27,4
Classe di età			
<20	105	25	23,8
≥20	30	12	40,0
Genere			
F	61	14	23,0
M	74	23	31,1
Residenza			
Milano	19	7	36,8
Prov. MI(no Milano)	53	14	26,4
Lombardia(no MI/Prov)	51	12	23,5
Altro	12	4	33,3
Voto Maturità*			
Alto	34	6	17,6
Basso	101	31	30,7
Maturità			
Classica	5	0	0,0
Scientifica	64	11	17,2
Tecnico commerciale	16	9	56,3
Altre Tecniche	39	14	35,9
Magistrale	4	1	25,0
Professionale	3	1	33,3
CdL			
533	93	26	28,0
580	42	11	26,2
Fascia reddito#			
Bassa	21	6	28,6
Media	54	13	24,1
Alta	60	18	30,0

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta).

Figura 7.3 Calibration plot per il modello predittivo. Validation Set (coorte 08/09). Facoltà di Scienze Statistiche.



La linea tratteggiata indica la linea ideale, dove il rischio predetto corrisponde esattamente alle percentuali di abbandono osservate. I punti rappresentano le coppie di rischi osservati/rischi predetti per ognuno dei cinque quintili di rischio considerati. Le linee rappresentano gli IC al 95% calcolati per i rischi osservati.

Tabella 7.4 Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09). Facoltà di Scienze Statistiche.

Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	27	1	1.1
basso	27	8	3.3
medio	27	6	5.8
alto	27	10	7.6
molto alto	27	12	10.4

Appendice 7. Equazione del modello predittivo del rischio di abbandono. Facoltà di Scienze Statistiche.

$$\text{Prob}\{\text{ABB} = \text{Abbandono}\} = \frac{1}{1 + \exp(-X\beta)}, \text{ where}$$

$$\begin{aligned}
 X\hat{\beta} = & \\
 & -3.255528 + 0.476573131 \text{ETA}_I \text{SC} - 0.007789216 \text{ETA}_I \text{SC}^2 \\
 & -0.1037987\{\text{M}\} \\
 & +0.08673778\{\text{Prov MI}\} + 0.06460582\{\text{Lomb.}\} + 0.48216237\{\text{Altro}\} \\
 & -0.2081176\{534\} + 0.1674086\{537\} - 0.3585492\{580\} \\
 & -1.940704\{\text{Ling/Magistr}\} - 2.959657\{\text{Tecn}\} - 2.207166\{\text{Prof/Altro}\} \\
 & -0.06532683 \text{VOTOMAT} \\
 & +\text{VOTOMAT}[0.03256878 \{\text{Ling/Magistr}\} + 0.05301851 \{\text{Tecn}\} + 0.04910436 \{\text{Prof/Altro}\}]
 \end{aligned}$$

and $\{c\} = 1$ if subject is in group c , 0 otherwise.

8. Facoltà di Sociologia

Tabella 8.1. Legenda dei CdL considerati. Facoltà di Sociologia.

Codice CdL	Descrizione
523	Scienze del Turismo e Comunità Locale
535	Servizio Sociale
536	Sociologia
541	Scienze del Turismo e Comunità Locale (Corso a Distanza)
579	Scienze dell'Organizzazione

Tabella 8.2 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Training Set (coorti a.a. 02/03-07/08). Facoltà di Sociologia.

	N. immatricolati 3404	N. Abbandoni 849	% Abbandono 24,9
Classe di età			
<20	1492	247	16,6
≥20	1912	602	31,5
Genere			
F	2406	544	22,6
M	998	305	30,6
Residenza			
Milano	512	128	25,0
Prov. MI(no Milano)	1410	321	22,8
Lombardia(no MI/Prov)	1026	225	21,9
Altro	456	175	38,4
Voto Maturità*			
Alto	593	102	17,2
Basso	2811	747	26,6
Maturità			
Classica	188	45	23,9
Scientifica	626	119	19,0
Tecnico commerciale	365	121	33,2
Altre Tecniche	589	150	25,5
Magistrale	537	122	22,7
Professionale	618	182	29,4
Linguistico	374	66	17,6
Altro	107	44	41,1
Coorte			
2002	564	128	22,7
2003	504	119	23,6
2004	572	136	23,8
2005	656	183	27,9
2006	528	140	26,5
2007	580	143	24,7
CdL			
523	951	106	11,1
535	545	118	21,7
536	1086	312	28,7
541	610	265	43,4
579	212	48	22,6
Fascia reddito#			
Bassa	576	125	21,7
Media	1868	356	19,1
Alta	960	368	38,3

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta)

Figura 8.1 Relazione tra le caratteristiche all'entrata degli studenti immatricolati e il rischio di abbandono nel primo anno di corso. Stime puntuali e IC al 95% basati su modelli di regressione logistica. Training Set (coorti a.a. 02/03-07/08). Facoltà di Sociologia.

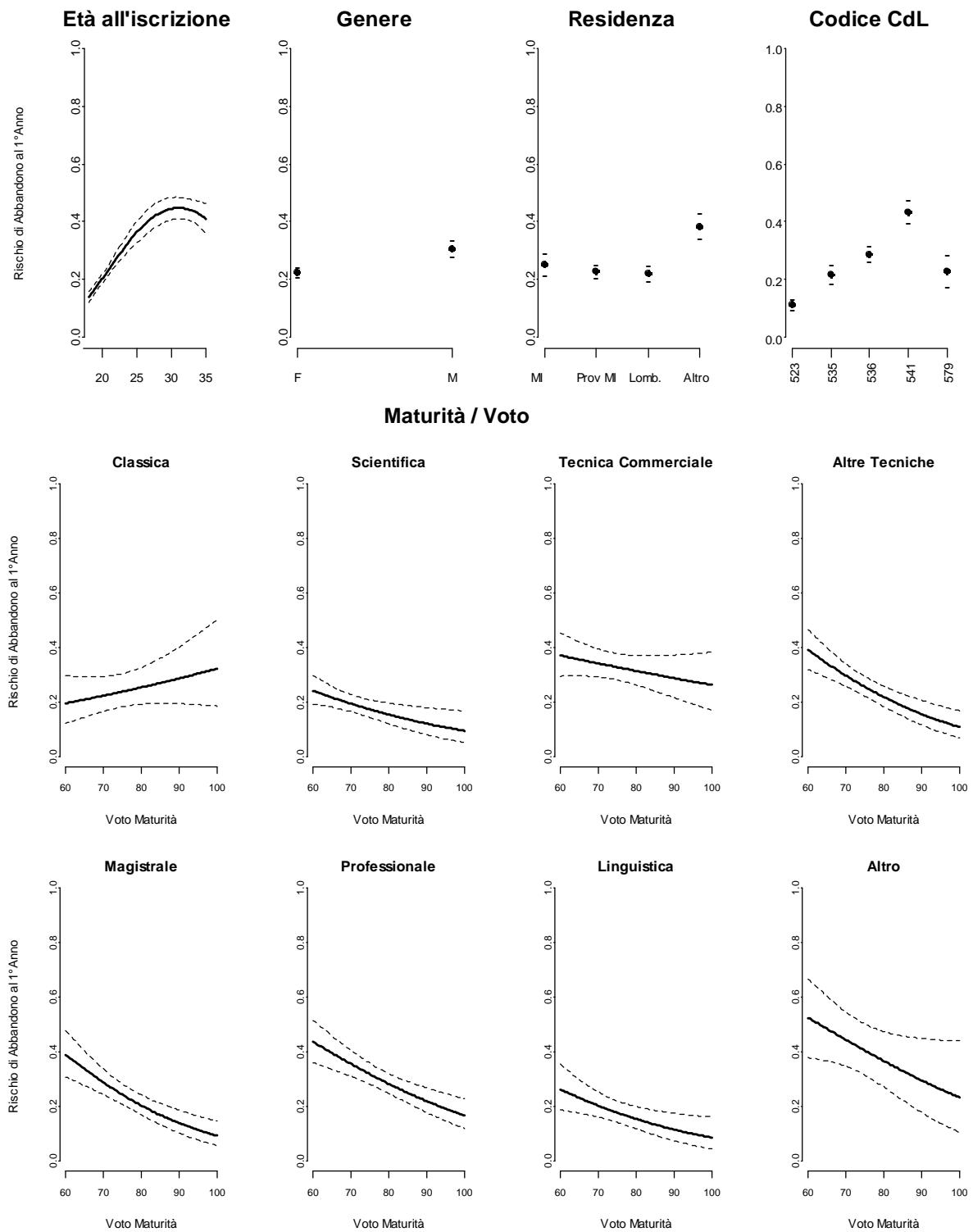
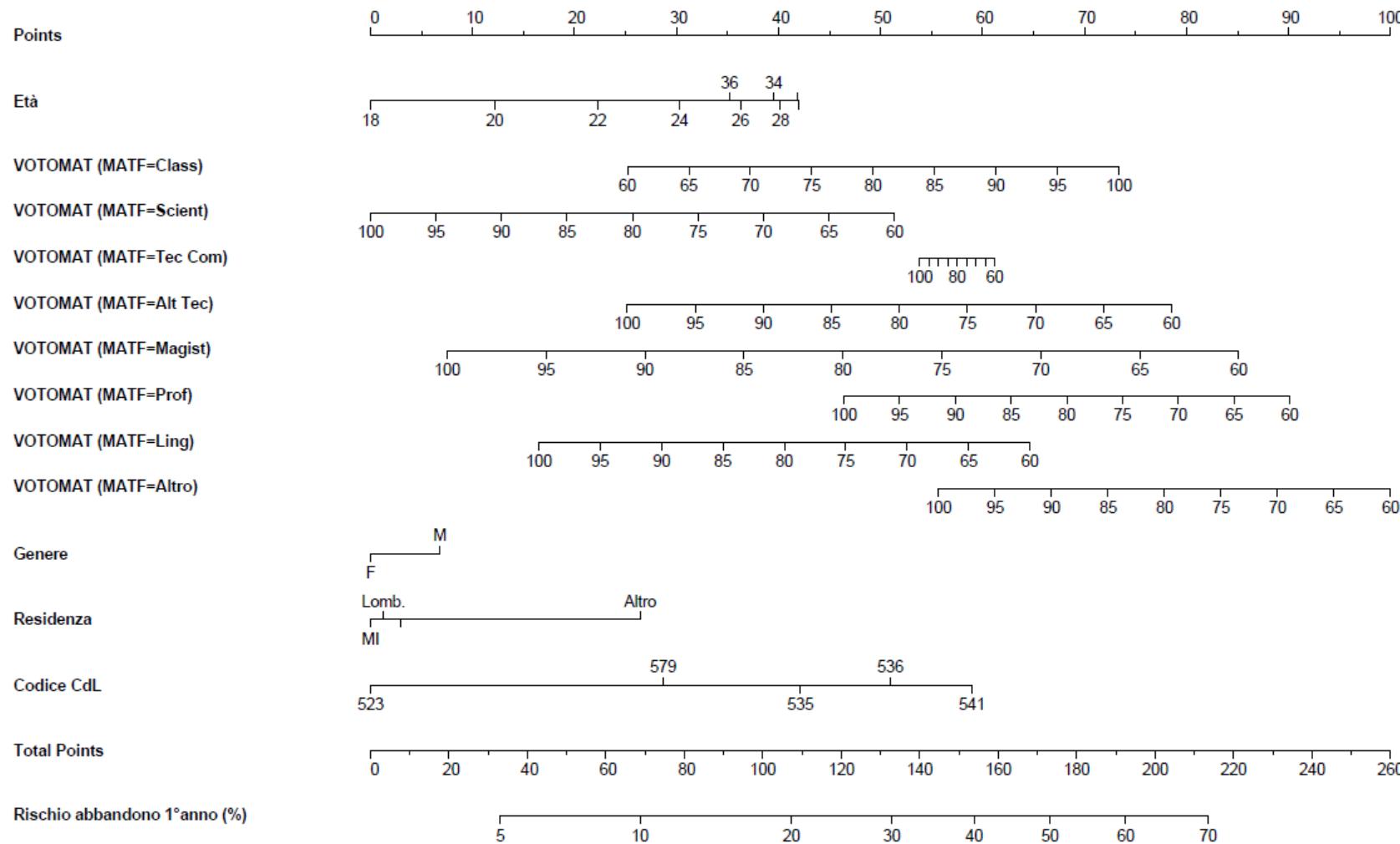


Figura 8.2 Nomogramma del rischio di abbandono nel primo anno di corso. Stime basate su un modello di regressione logistica multiplo. Training Set (coorti a.a. 02/03-07/08). Facoltà di Sociologia.



Istruzioni per l'uso del nomogramma.

Sul secondo asse delle ascisse a partire dall'alto si identifichi l'età all'immatricolazione dello studente per il quale si vuole calcolare il rischio di abbandono nel primo anno. Si tracci una linea verticale fino all'ascissa **Points** (la prima in alto), indicante il numero di punti prognostici che lo studente riceverà per la sua età. Si faccia lo stesso per tutte le altre variabili prognostiche (attenzione: per assegnare il punteggio prognostico relativo al voto di maturità è necessario identificare l'ascissa corrispondente al tipo di maturità conseguita).

Una volta che i punteggi prognostici sono stati determinati per tutte le caratteristiche, si provveda a calcolare il punteggio totale. Il rischio individuale di abbandono nel primo anno di corso si determina tracciando una linea verticale dal punto corrispondente al punteggio totale sull'ascissa **Total Points** (la seconda dal basso) all'ultima ascissa (**Rischio di abbandono nel 1° anno**).

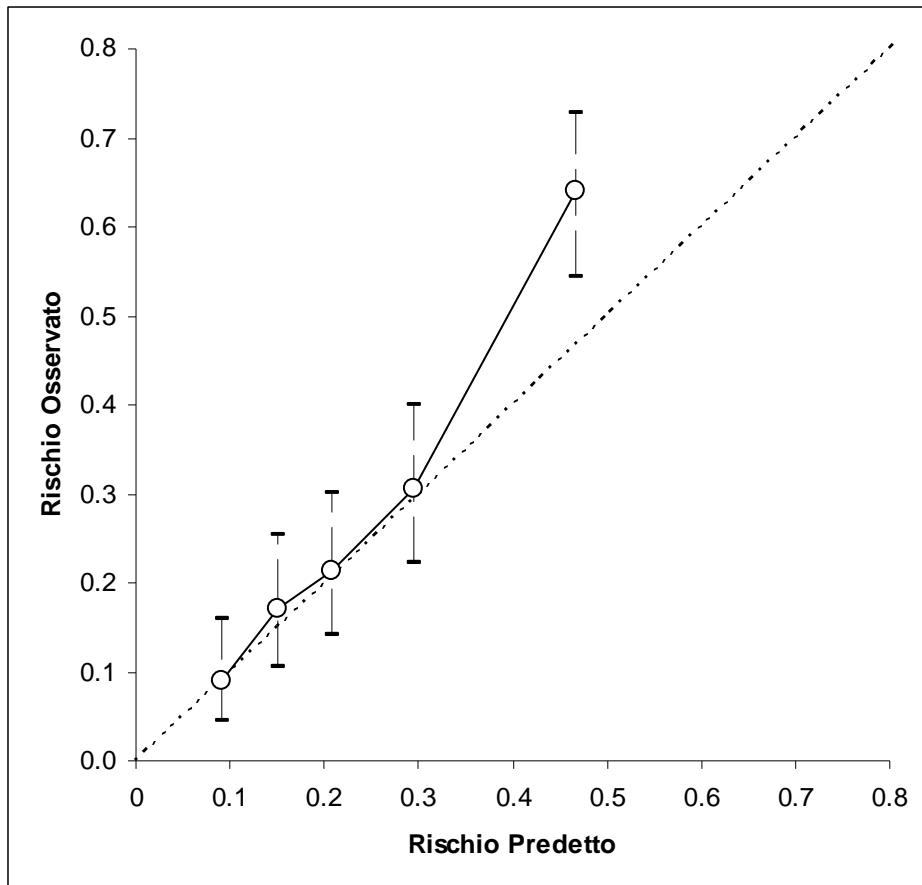
Tabella 8.3 Caratteristiche degli studenti immatricolati per la prima volta ad un CdL Triennale o a Ciclo Unico della Facoltà e percentuali di abbandono nel primo anno di corso. Validation Set (coorte 08/09). Facoltà di Sociologia.

	N. immatricolati	N. Abbandoni	% Abbandono
	556	158	28,4
Classe di età			
<20	251	41	16,3
=20	305	117	38,4
Genere			
F	407	103	25,3
M	149	55	36,9
Residenza			
Milano	78	24	30,8
Prov. MI(no Milano)	234	49	20,9
Lombardia(no MI/Prov)	161	44	27,3
Altro	83	41	49,4
Voto Maturità*			
Alto	66	16	24,2
Basso	490	142	29,0
Maturità			
Classica	30	8	26,7
Scientifica	92	26	28,3
Tecnico commerciale	48	19	39,6
Altre Tecniche	119	38	31,9
Magistrale	94	21	22,3
Professionale	86	28	32,6
Linguistico	63	9	14,3
Altro	24	9	37,5
CdL			
523	153	22	14,4
535	84	16	19,0
536	145	54	37,2
541	87	44	50,6
579	87	22	25,3
Fascia reddito#			
Bassa	90	21	23,3
Media	208	41	19,7
Alta	258	96	37,2

* Alto: ≥90/100; Basso: <90/10

Il numero e la definizione delle classi di reddito riportati negli archivi delle segreterie studenti varia al variare dell'anno accademico considerato. Nella presente analisi si è cercato di ridurle e uniformarle in tre soli livelli (Bassa, Media e Alta).

Figura 8.3 Calibration plot per il modello predittivo. Validation Set (coorte 08/09). Facoltà di Sociologia.



La linea tratteggiata indica la linea ideale, dove il rischio predetto corrisponde esattamente alle percentuali di abbandono osservate. I punti rappresentano le coppie di rischi osservati/rischi predetti per ognuno dei cinque quintili di rischio considerati. Le linee rappresentano gli IC al 95% calcolati per i rischi osservati.

Tabella 8.4 Numero di abbandoni previsti dal modello e numero di abbandoni osservati, per quintile di rischio predetto. Validation Set (coorte 08/09). Facoltà di Sociologia.

Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	111	10	10.2
basso	111	19	16.8
medio	112	24	23.2
alto	111	34	32.7
molto alto	111	71	51.8

Appendice 8. Equazione del modello predittivo del rischio di abbandono. Facoltà di Sociologia.

$$\text{Prob}\{\text{ABB} = \text{Abbandono}\} = \frac{1}{1 + \exp(-X\hat{\beta})}, \text{ where}$$

$$\begin{aligned}
 X\hat{\beta} = & \\
 & -8.882\,793 + 0.33387\,2890 \text{ETA}_1\text{SC} - 0.005421906 \text{ETA}_1\text{SC}^2 \\
 & + 0.1423326\{\text{M}\} \\
 & + 0.06302293\{\text{Prov MI}\} + 0.02609729\{\text{Lomb.}\} + 0.55554661\{\text{Altro}\} \\
 & + 0.8859165\{535\} + 1.0712566\{536\} + 1.2394676\{541\} + 0.6028958\{579\} \\
 & + 3.688927\{\text{Scient}\} + 2.508055\{\text{Tec Com}\} + 4.322601\{\text{Alt Tec}\} + 5.225112\{\text{Magist}\} \\
 & + 4.261654\{\text{Prof}\} + 3.867173\{\text{Ling}\} + 4.488219\{\text{Altro}\} + 0.02531129 \text{VOTOMAT} \\
 & + \text{VOTOMAT}[-0.05231078 \{\text{Scient}\} - 0.02918743 \{\text{Tec Com}\} - 0.05335597 \{\text{Alt Tec}\} \\
 & - 0.06609303 \{\text{Magist}\} - 0.04828268 \{\text{Prof}\} - 0.05062778 \{\text{Ling}\} - 0.04859498 \{\text{Altro}\}]
 \end{aligned}$$

and $\{c\} = 1$ if subject is in group c , 0 otherwise.