

**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, Davide Disalvatore

Progetto finanziato dalla Commissione Orientamento dell'Ateneo.

Premessa e Obiettivi

- ⇒ Un modo per migliorare la capacità di non perdere studenti da parte di un Ateneo/Facoltà/CdL potrebbe essere quello di identificare gli studenti a maggior rischio di abbandono precoce e predisporre per tempo degli interventi, in accordo alla categoria di rischio.
- ⇒ Utilizzo di modelli di regressione logistica multipla

$$\log \left[\frac{p(Y=1 | \mathbf{Z})}{1 - p(Y=1 | \mathbf{Z})} \right] = \boldsymbol{\beta}^T \mathbf{Z}$$

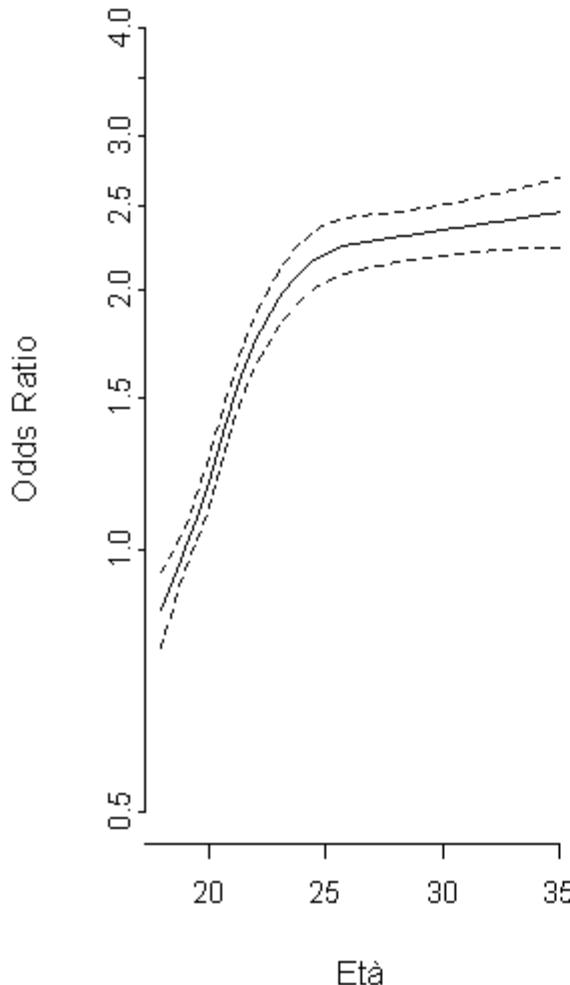
Y=1: abbandono

Vettore caratteristiche dello studente

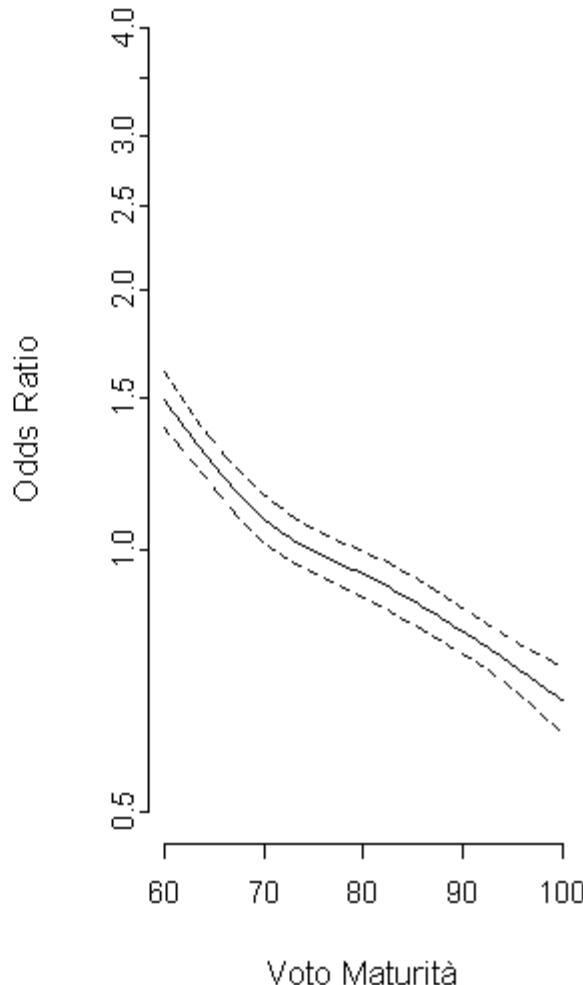
Parametri di regressione

Flessibilità del modello

Età



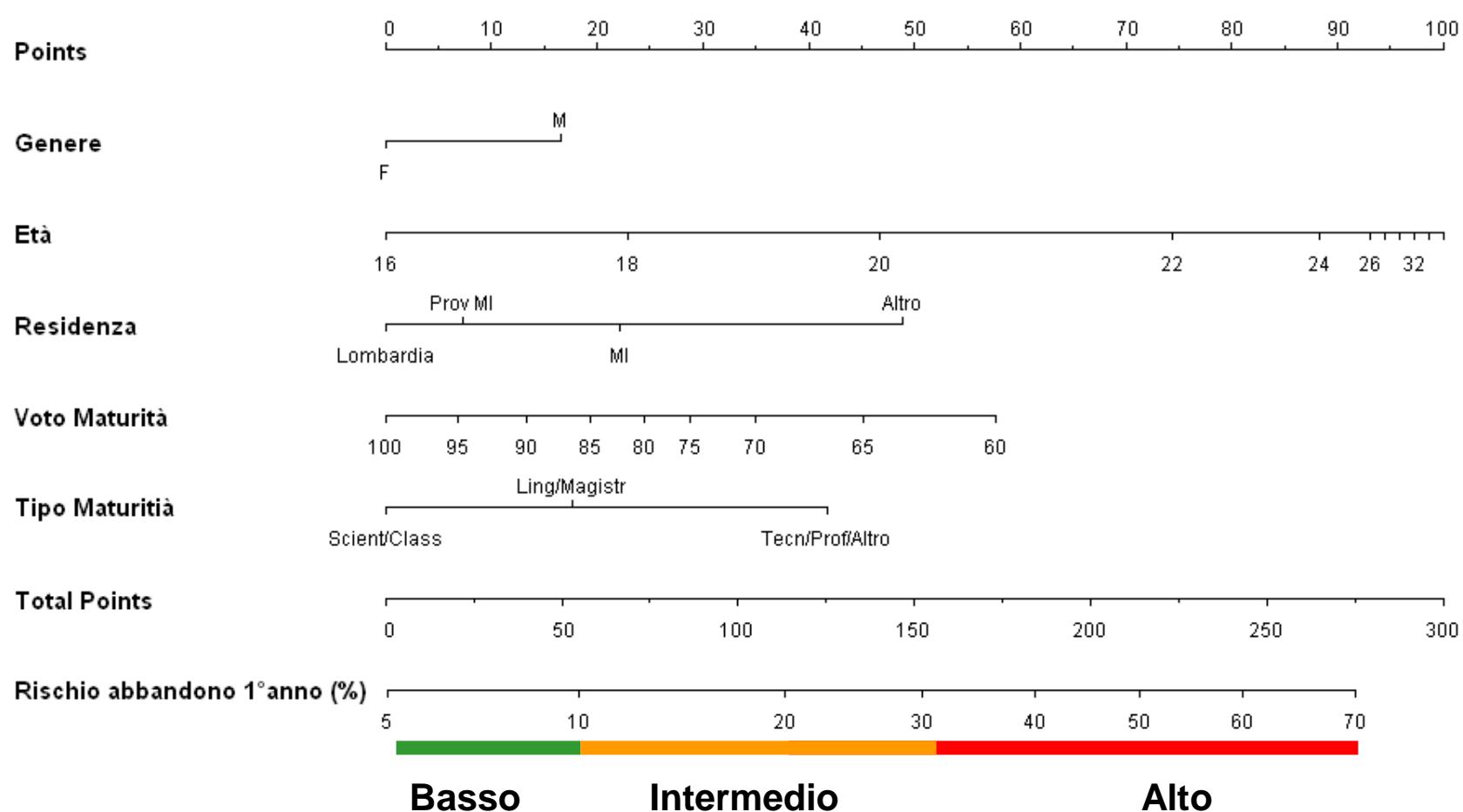
Voto Maturità



Sviluppo di un Risk Score Individuale

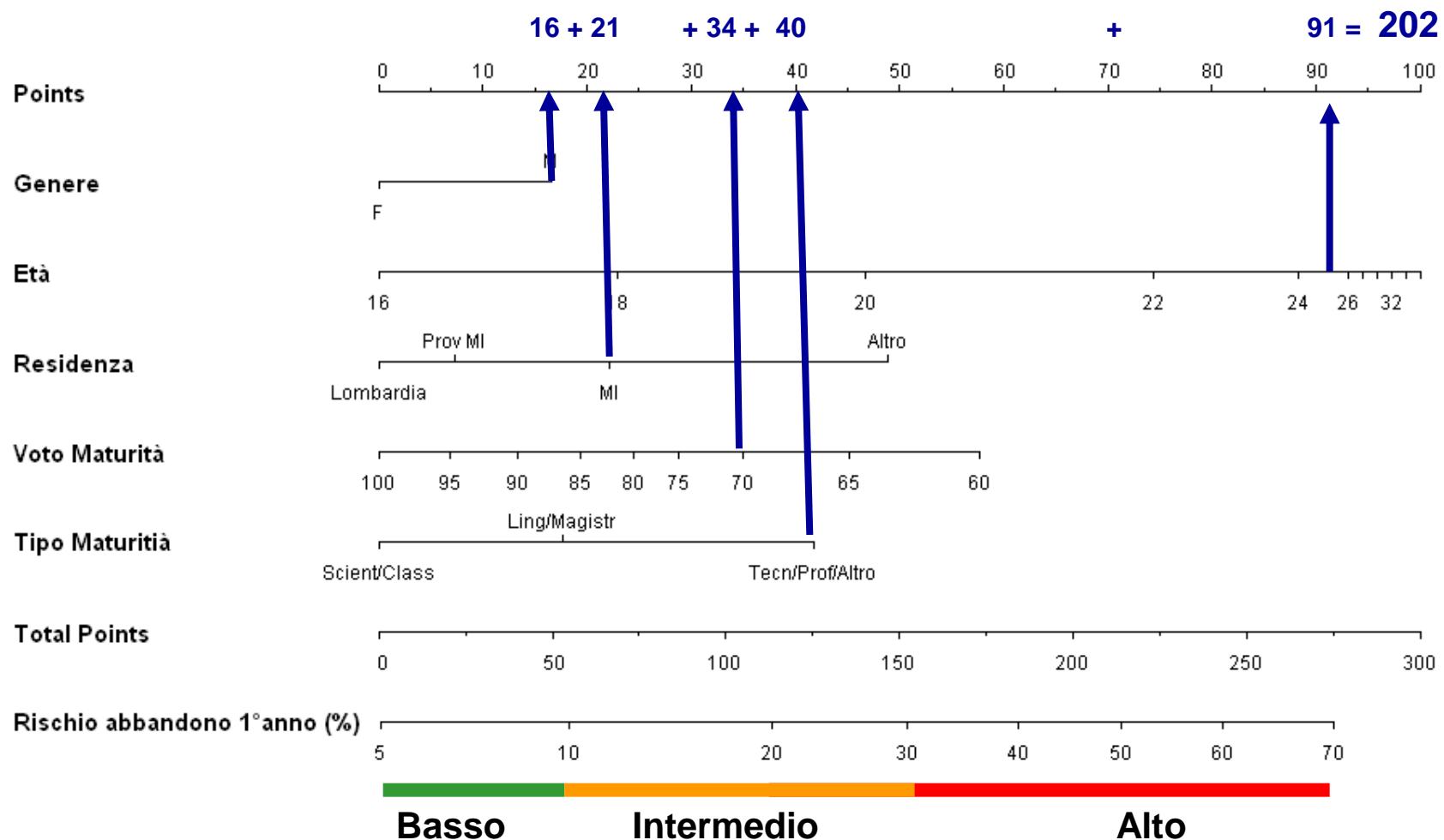


Nomogramma



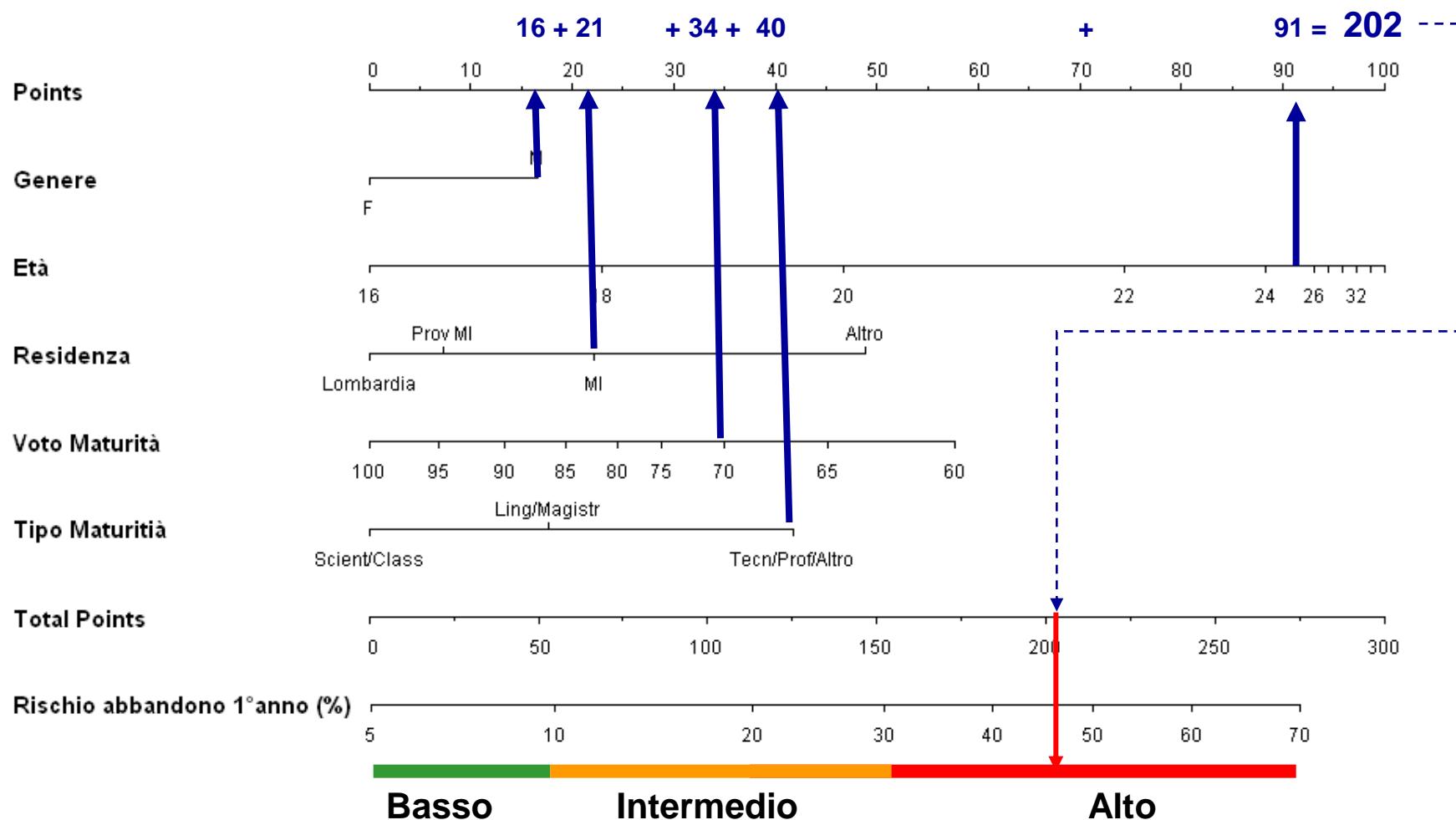
Interpretazione

- ⇒ Qual è il rischio di abbandono per un ragazzo di 25 anni con maturità tecnica (voto 70/100) e residente a Milano?



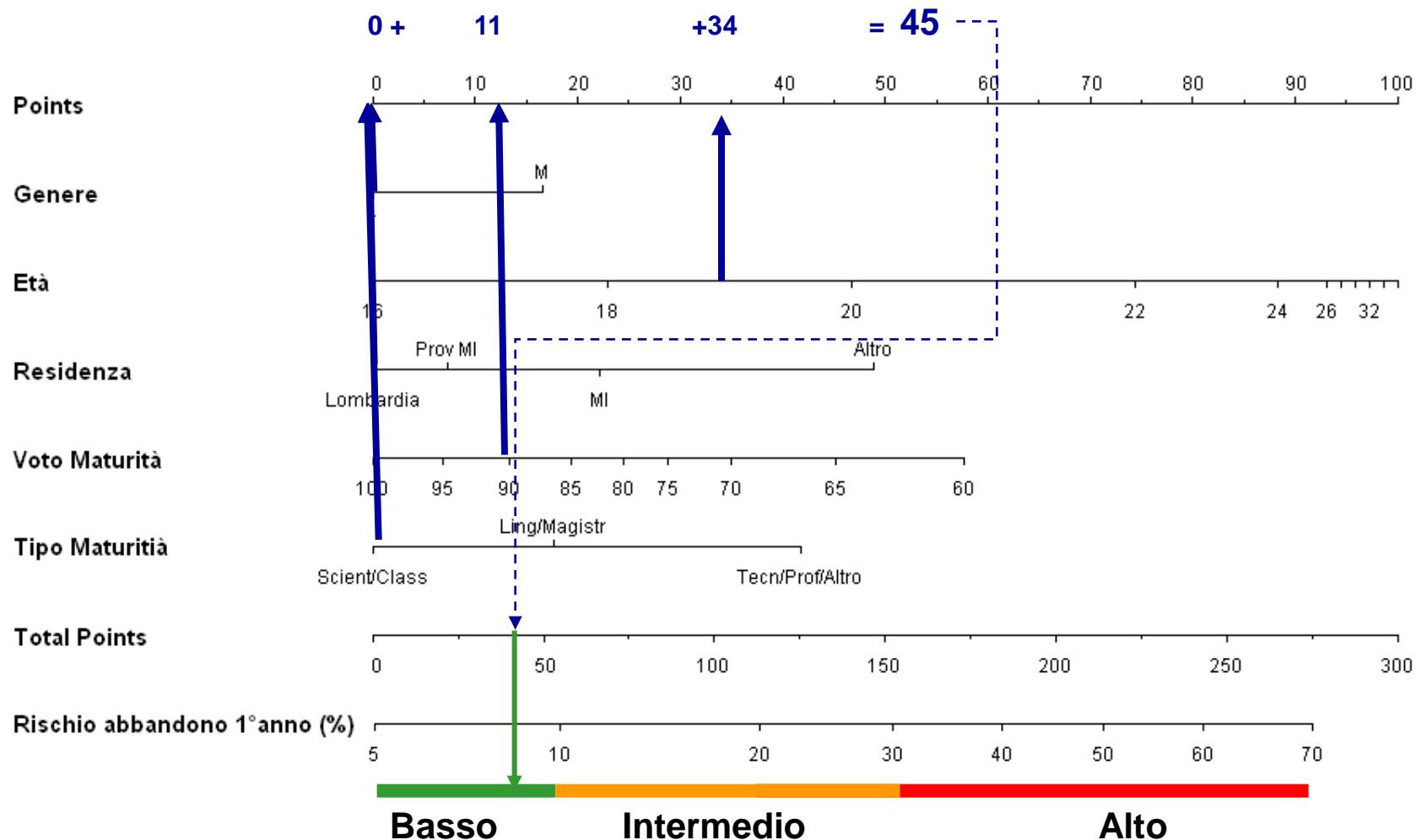
Interpretazione

- ⇒ Qual è il rischio di abbandono per un ragazzo di 25 anni con maturità tecnica (voto 70/100) e residente a Milano?



Previsione

- ⇒ Qual è il rischio di abbandono per una ragazza di 19 anni con maturità scientifica (voto 91/100) e residente a Varese?



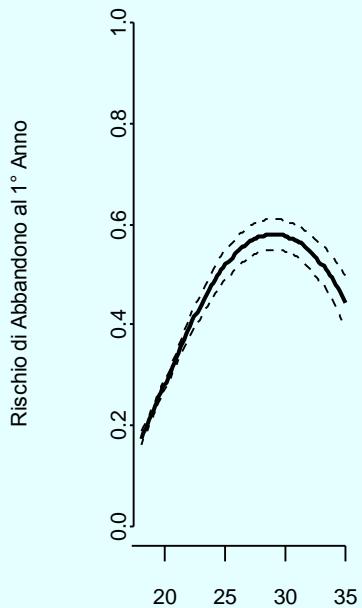
Training set e Validation set

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”).

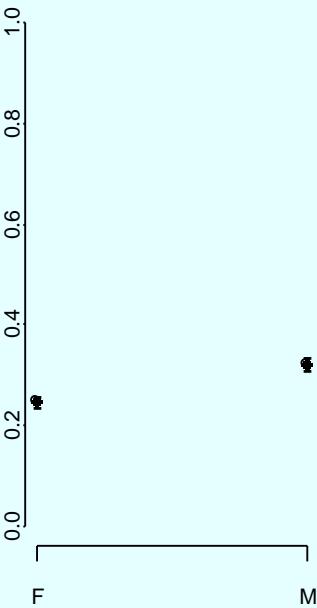
Facoltà	Immatricolati a.a. 02/03-07/08 (Training set)	Immatricolati a.a. 08/09 (Validation set)	Totale (%)
Economia	9604	1926	11530 (25.22)
Giurisprudenza	3455	487	3942 (8.62)
Medicina e Chirurgia	3193	557	3750 (8.20)
Psicologia	4612	680	5292 (11.57)
Scienze della Formazione	5839	1046	6885 (15.06)
Scienze 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)

Economia (Training Set)

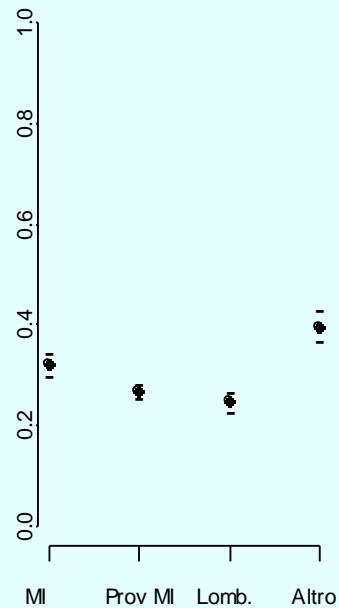
Età all'iscrizione



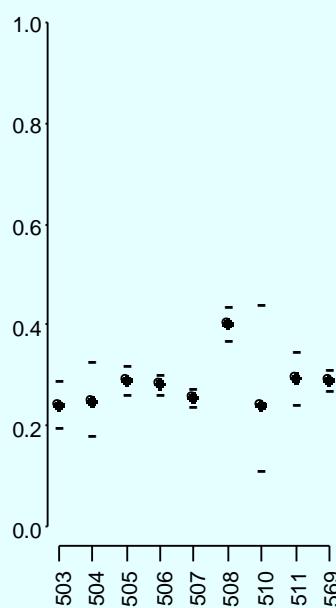
Genere



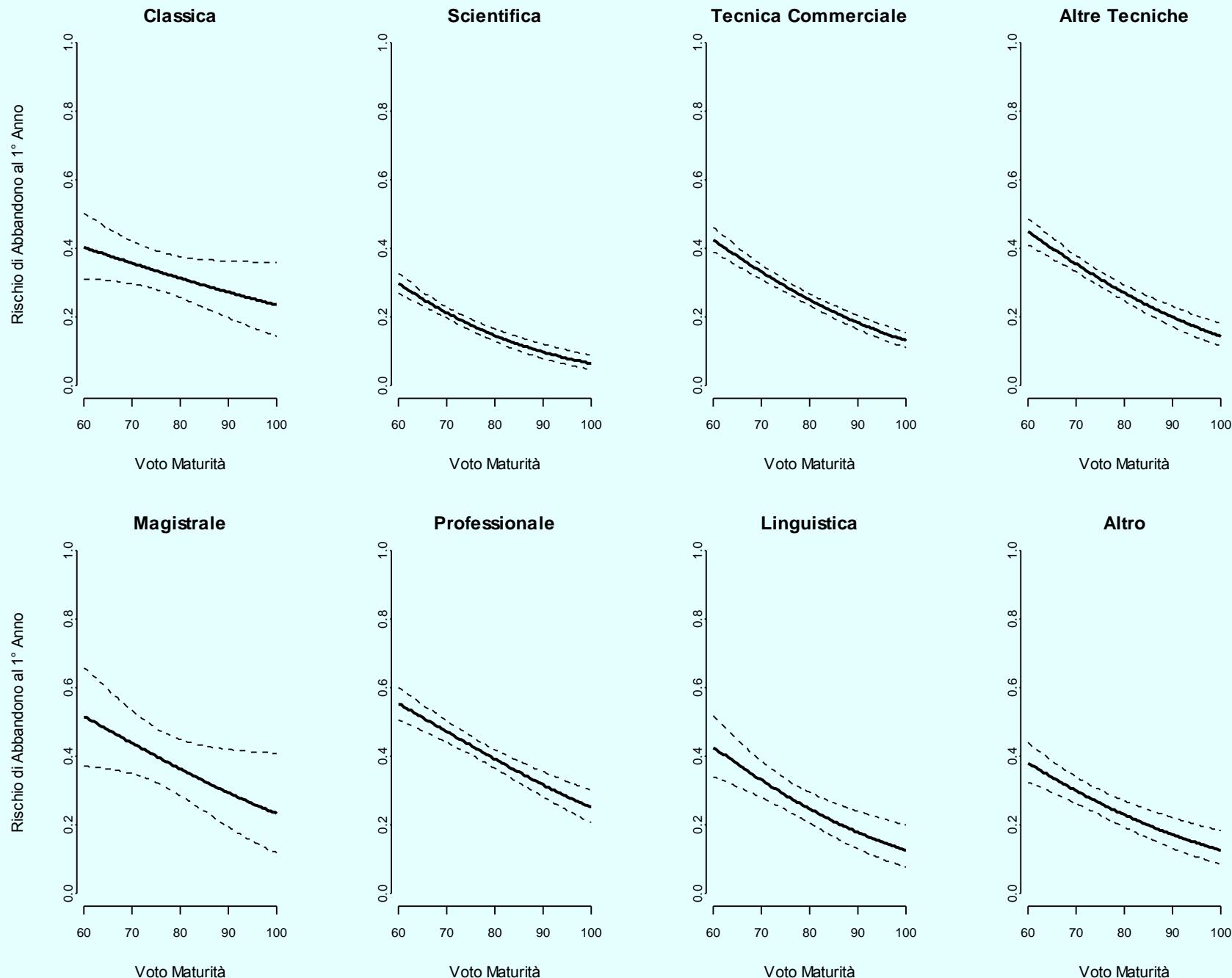
Residenza



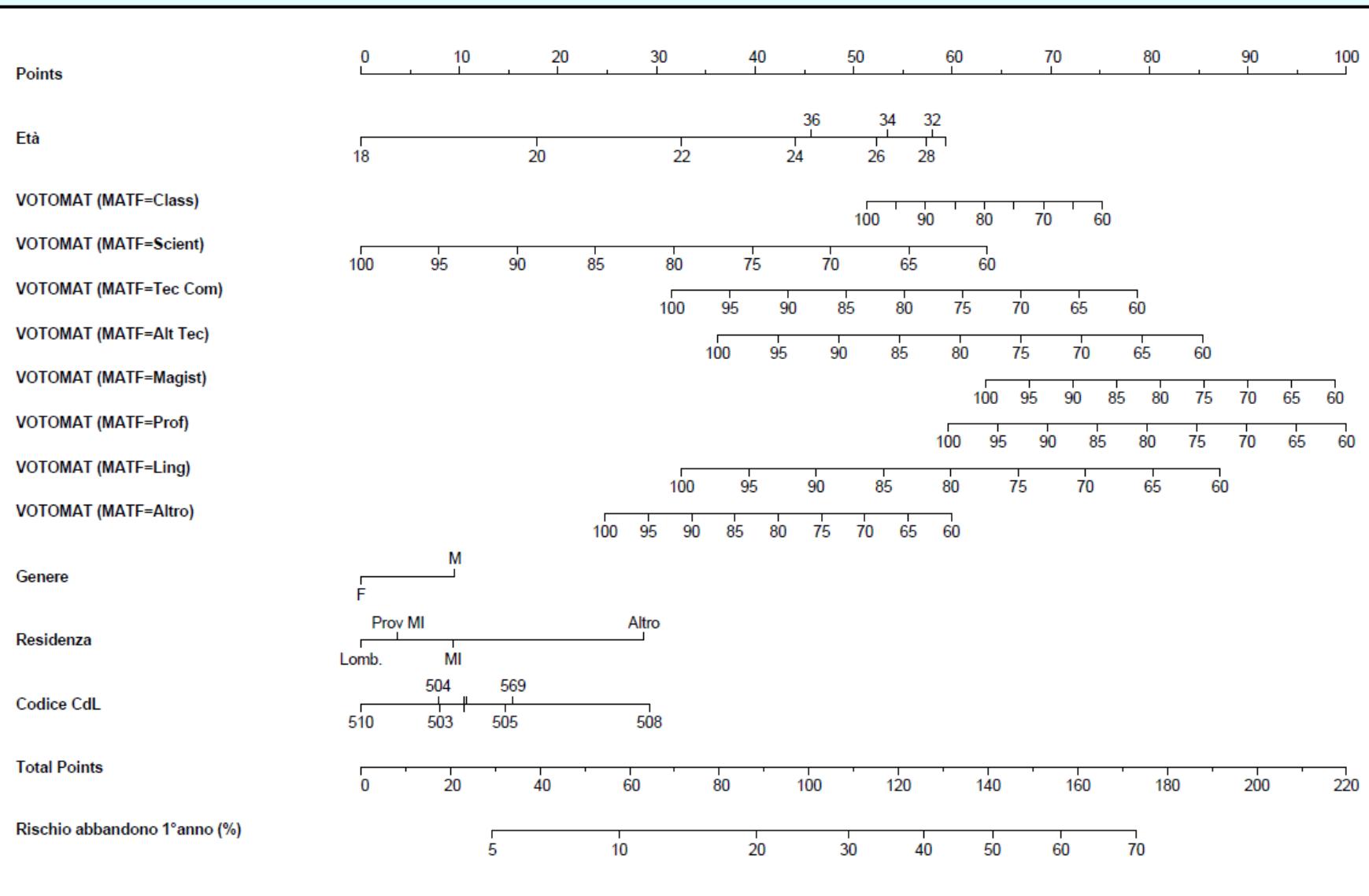
Codice CdL



Economia (Training Set)



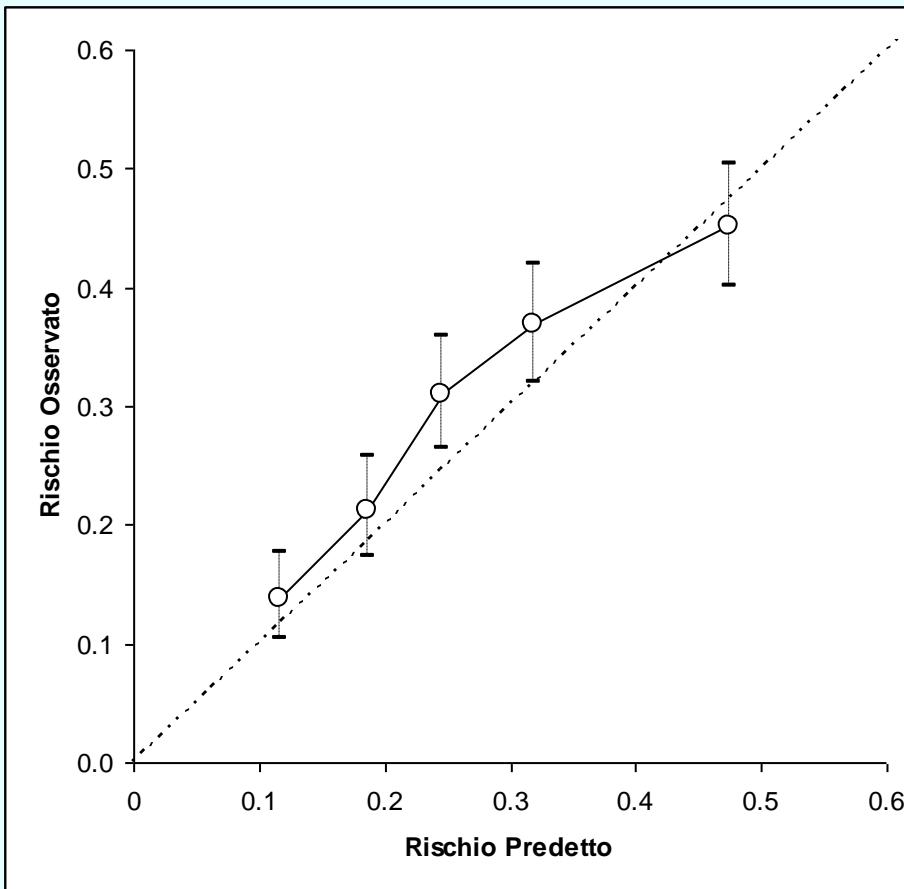
Economia - Nomogramma



Economia – Validazione su coorte 2008/09



Rischio atteso di abbandono sulla base del modello predittivo e rischio osservato, per quintile di rischio predetto. Validation set.

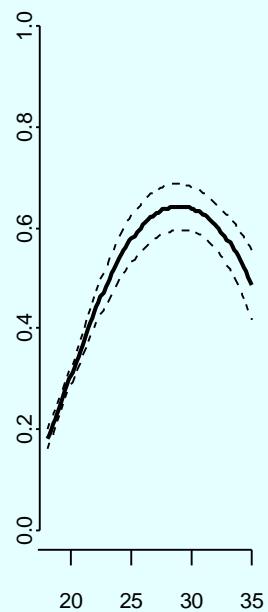


Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	385	53	44.9
basso	385	82	71.5
medio	386	120	94.8
alto	385	142	122.7
molto alto	385	174	182.4

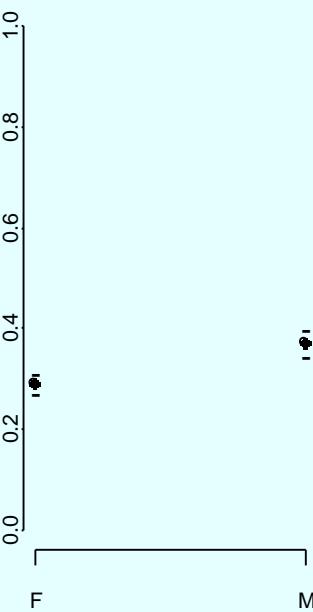
Giurisprudenza (Training Set)

Età all'iscrizione

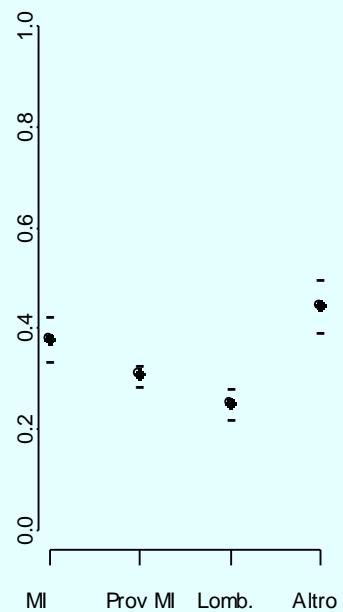
Rischio di Abbandono al 1° Anno



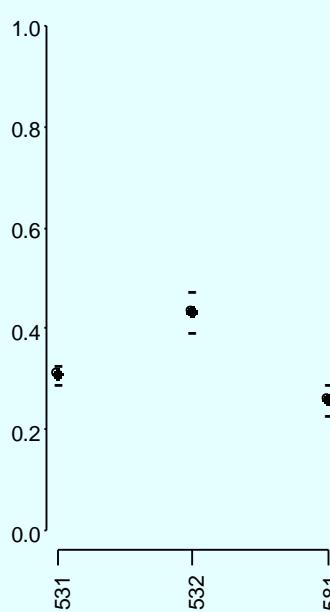
Genere



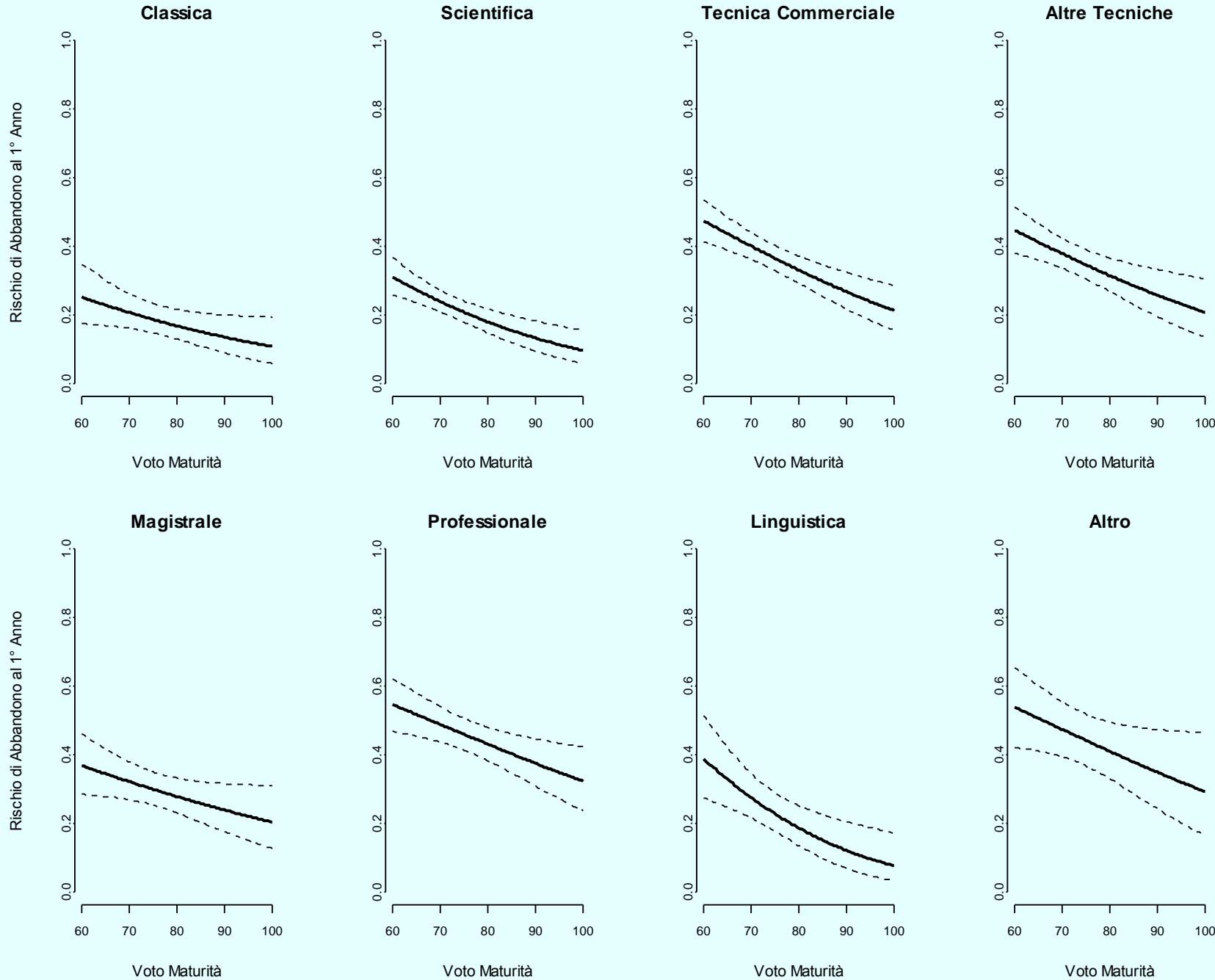
Residenza



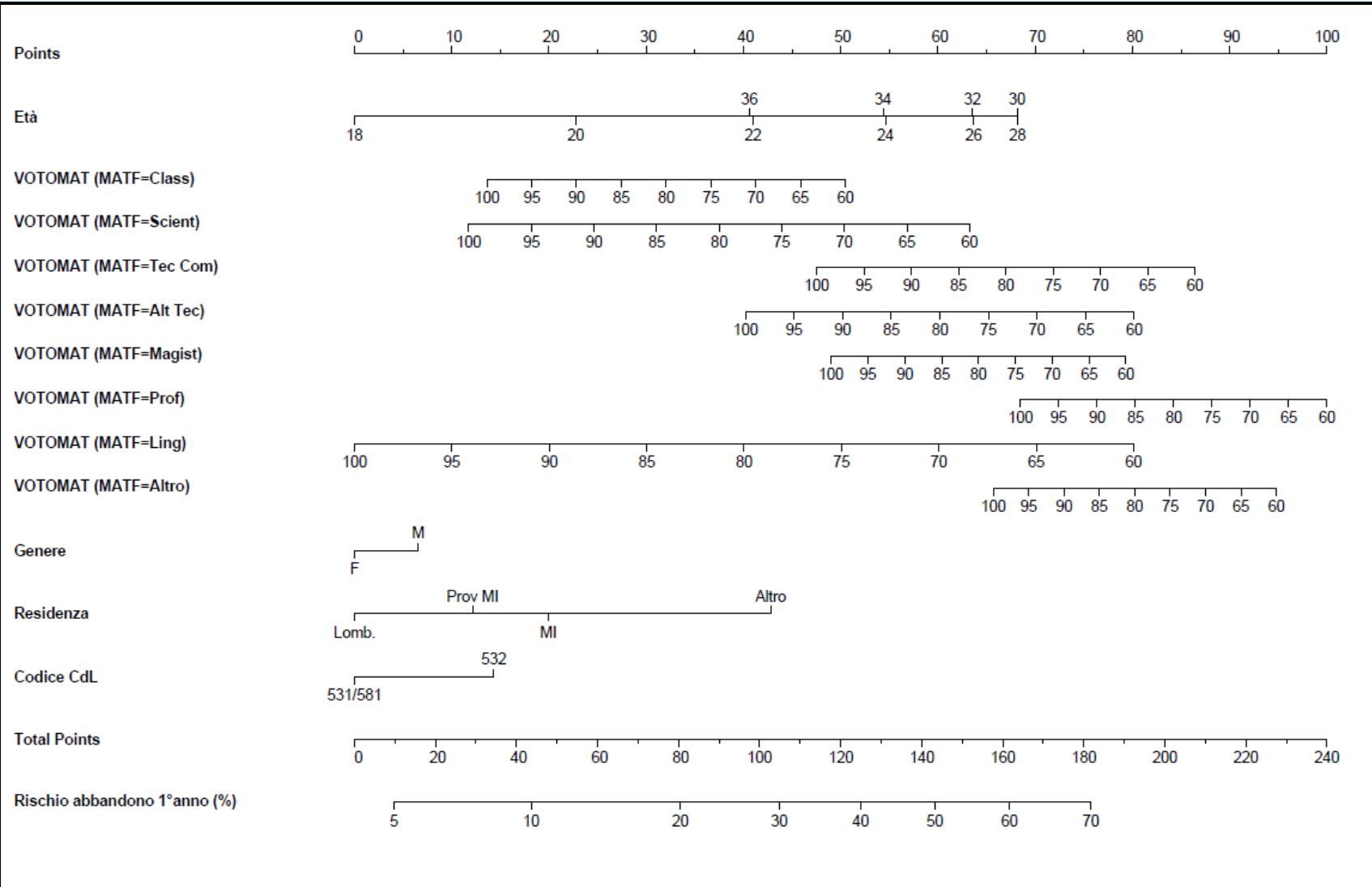
Codice CdL



Giurisprudenza (Training Set)



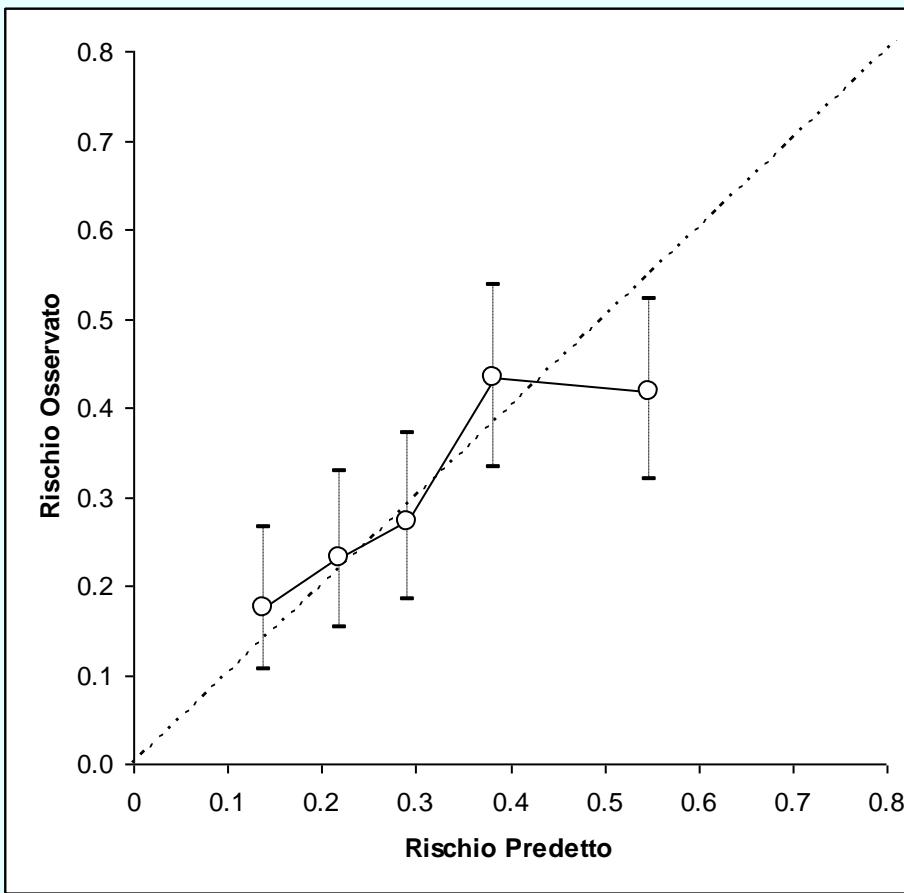
Giurisprudenza - Nomogramma



Giurisprudenza – Validazione su coorte 2008/09



Rischio atteso di abbandono sulla base del modello predittivo e rischio osservato, per quintile di rischio predetto. Validation set.

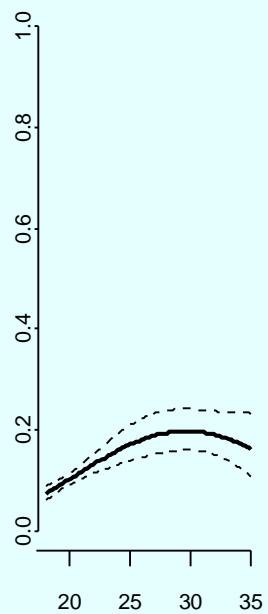


Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	97	17	13.4
basso	99	23	21.7
medio	96	26	27.9
alto	97	42	37.1
molto alto	98	41	53.7

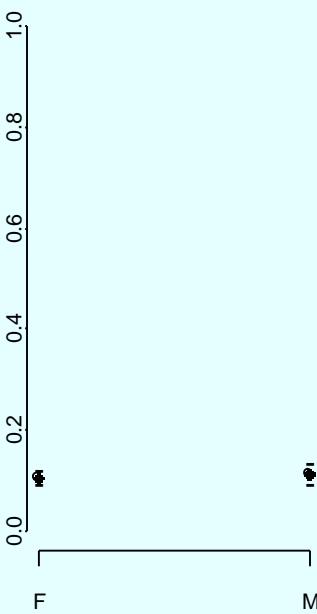
Medicina e Chirurgia (Training Set)

Età all'iscrizione

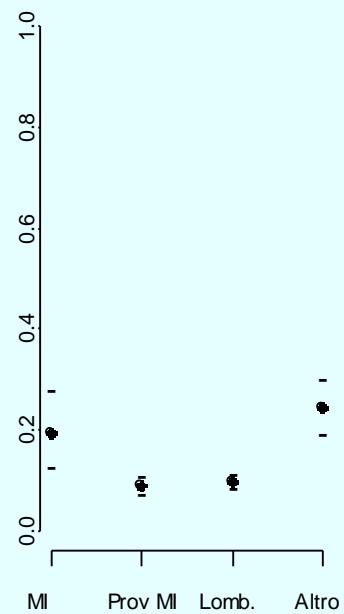
Rischio di Abbandono al 1° Anno



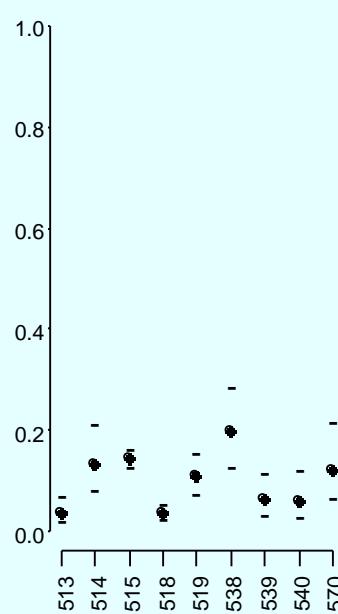
Genere



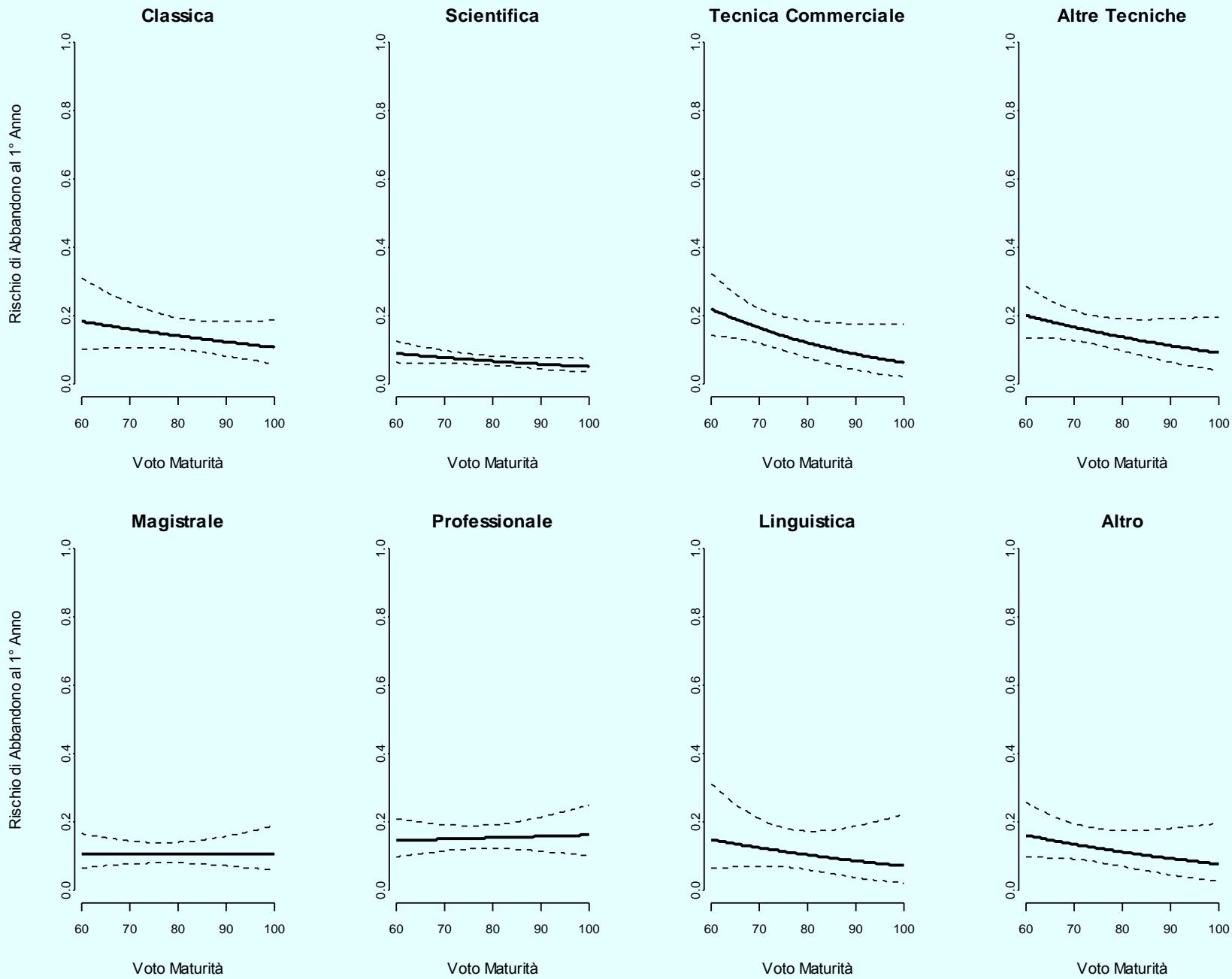
Residenza



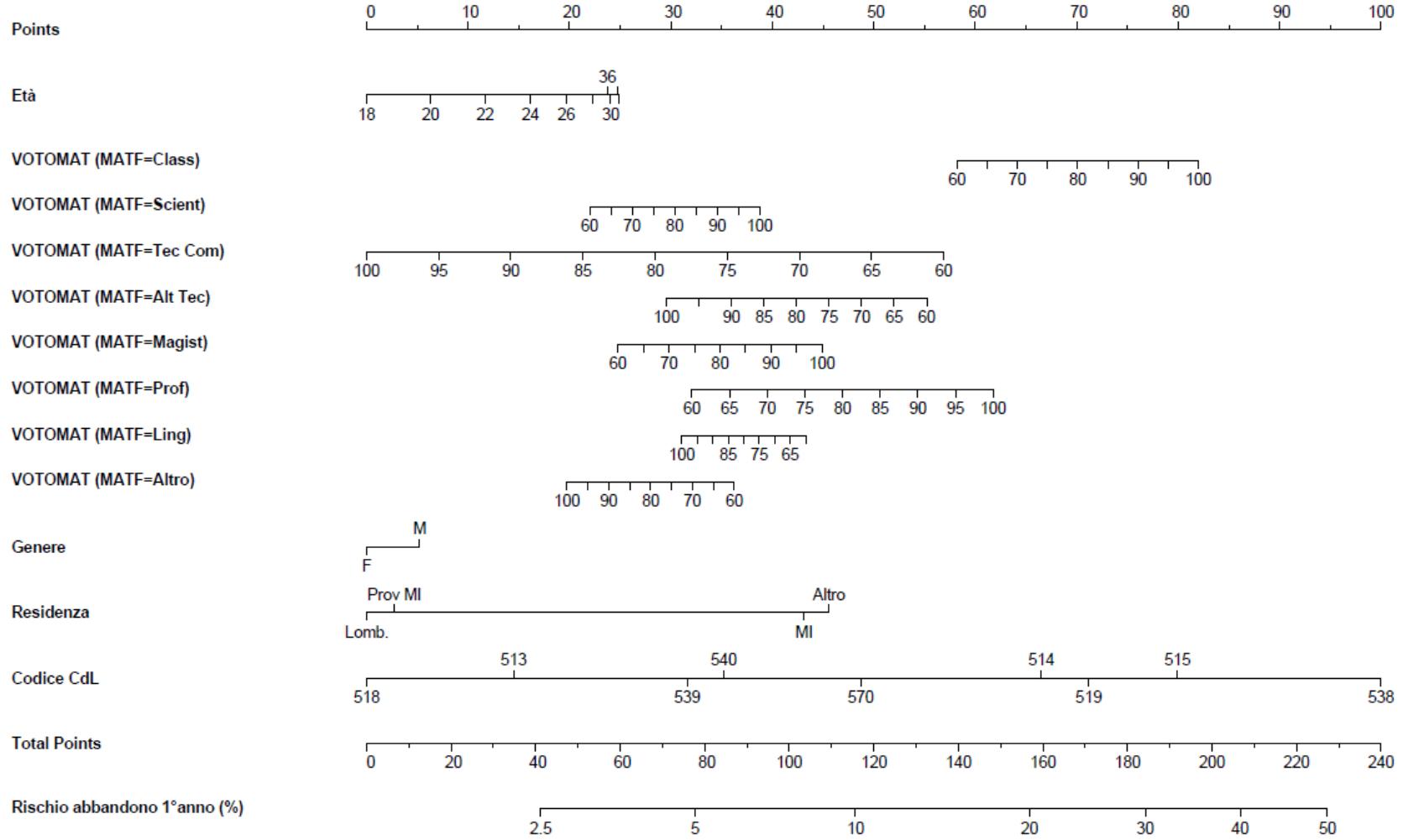
Codice CdL



Medicina e Chirurgia (Training Set)



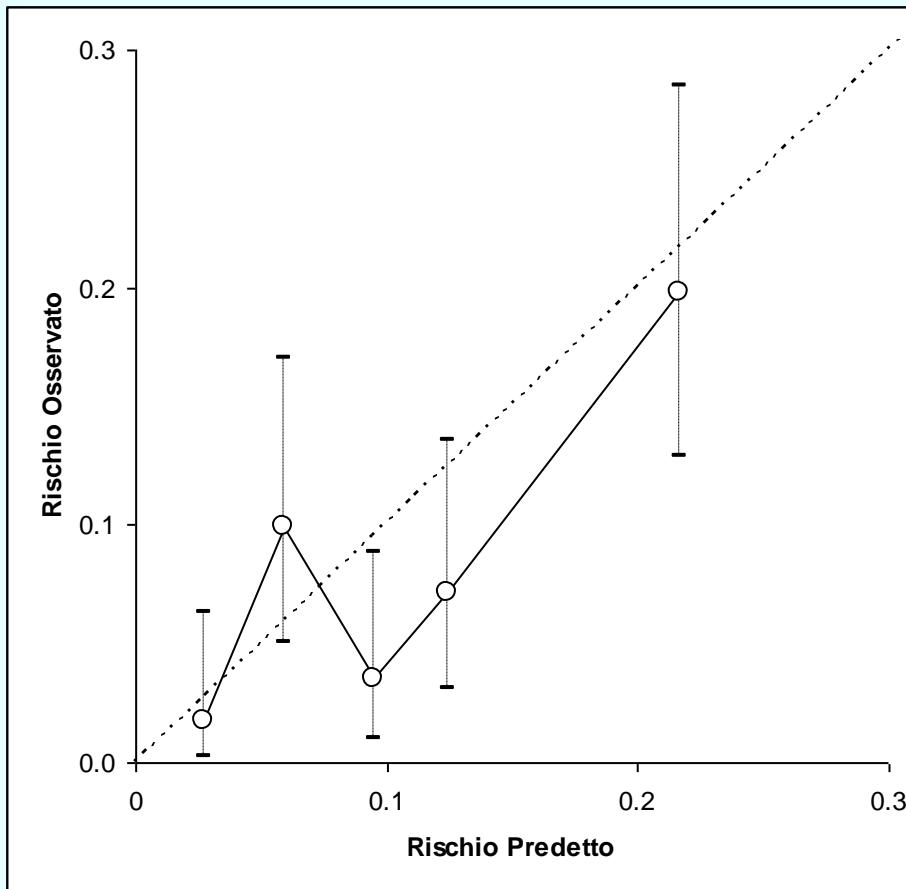
Medicina e Chirurgia - Nomogramma



Medicina e Chirurgia – Validazione su coorte 2008/09



Rischio atteso di abbandono sulla base del modello predittivo e rischio osservato, per quintile di rischio predetto. Validation set.

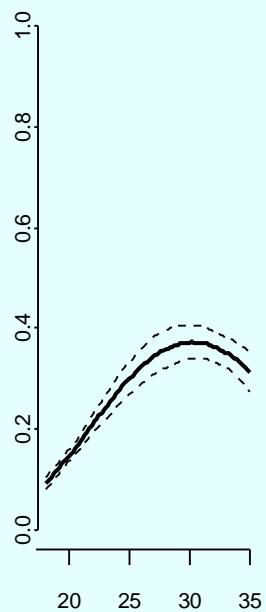


Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	111	2	3.0
basso	111	11	6.5
medio	112	4	10.6
alto	112	8	13.9
molto alto	111	22	24.0

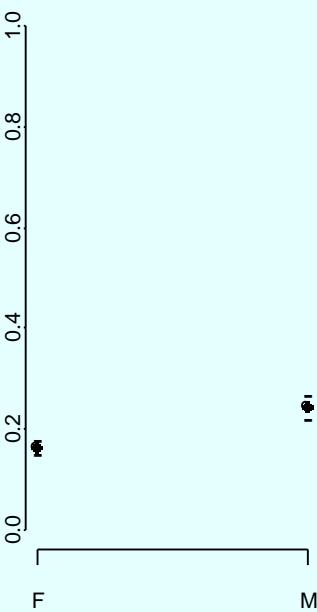
Psicologia (Training Set)

Età all'iscrizione

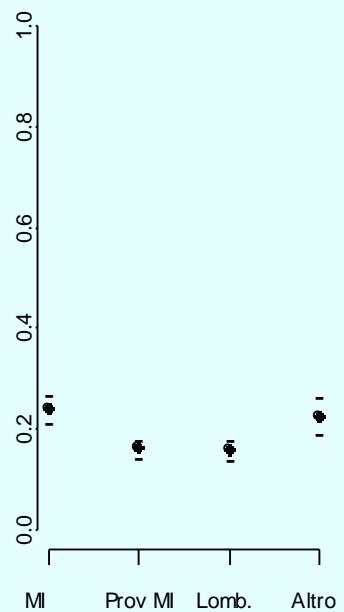
Rischio di Abbandono al 1° Anno



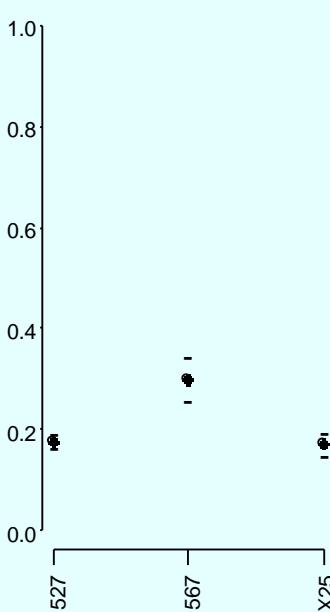
Genere



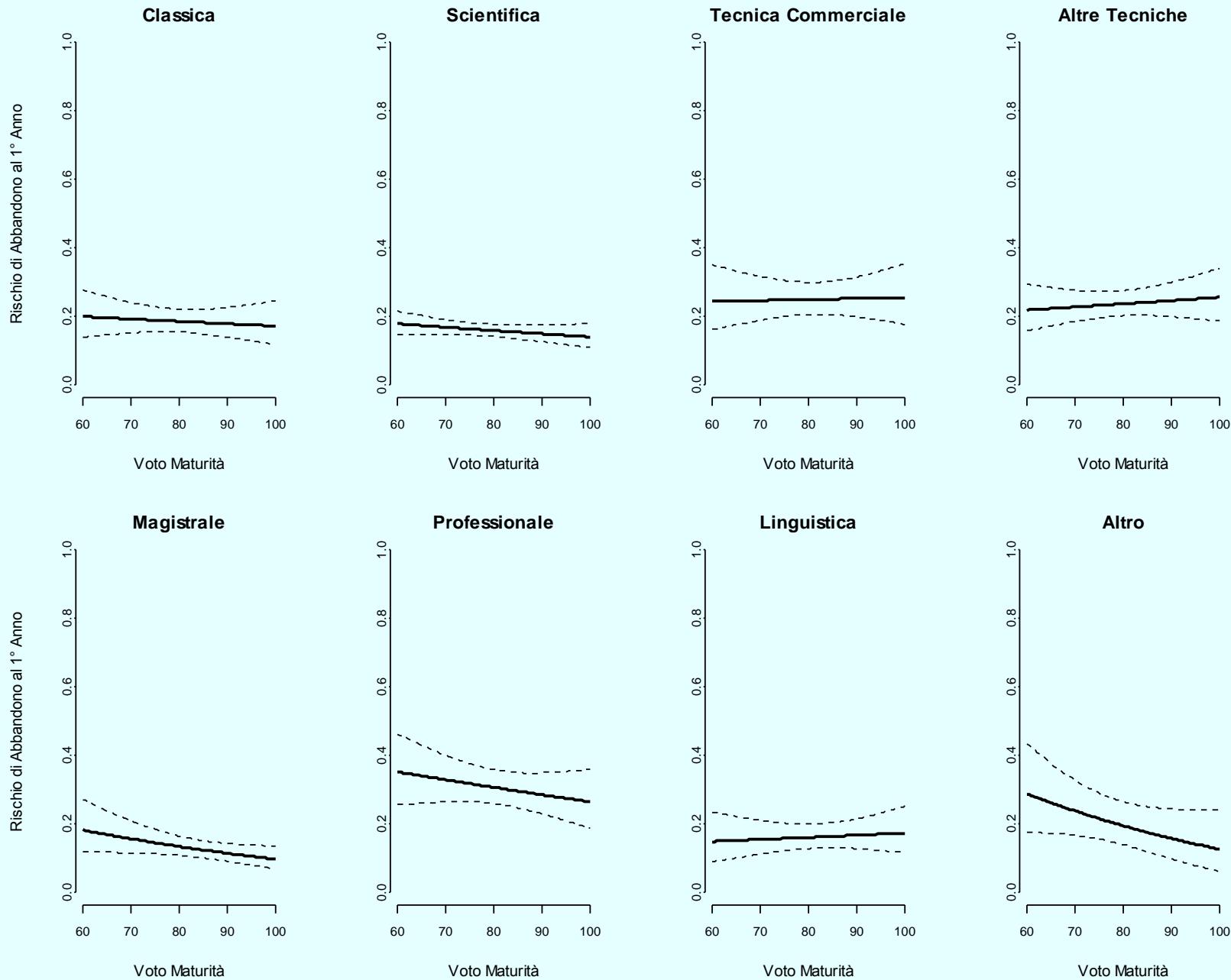
Residenza



Codice CdL

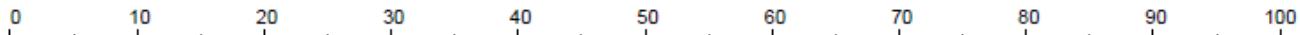


Psicologia (Training Set)

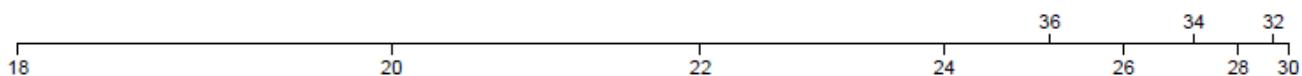


Psicologia - Nomogramma

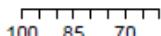
Points



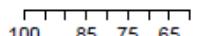
Età



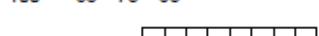
VOTOMAT (MATF=Class)



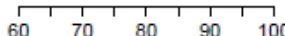
VOTOMAT (MATF=Scient)



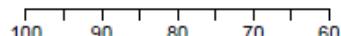
VOTOMAT (MATF=Tec Com)



VOTOMAT (MATF=Alt Tec)



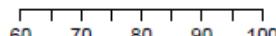
VOTOMAT (MATF=Magist)



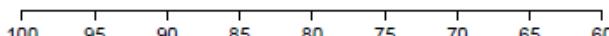
VOTOMAT (MATF=Prof)



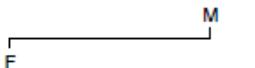
VOTOMAT (MATF=Ling)



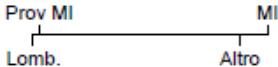
VOTOMAT (MATF=Altro)



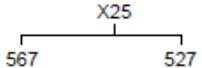
Genere



Residenza



Codice CdL



Total Points



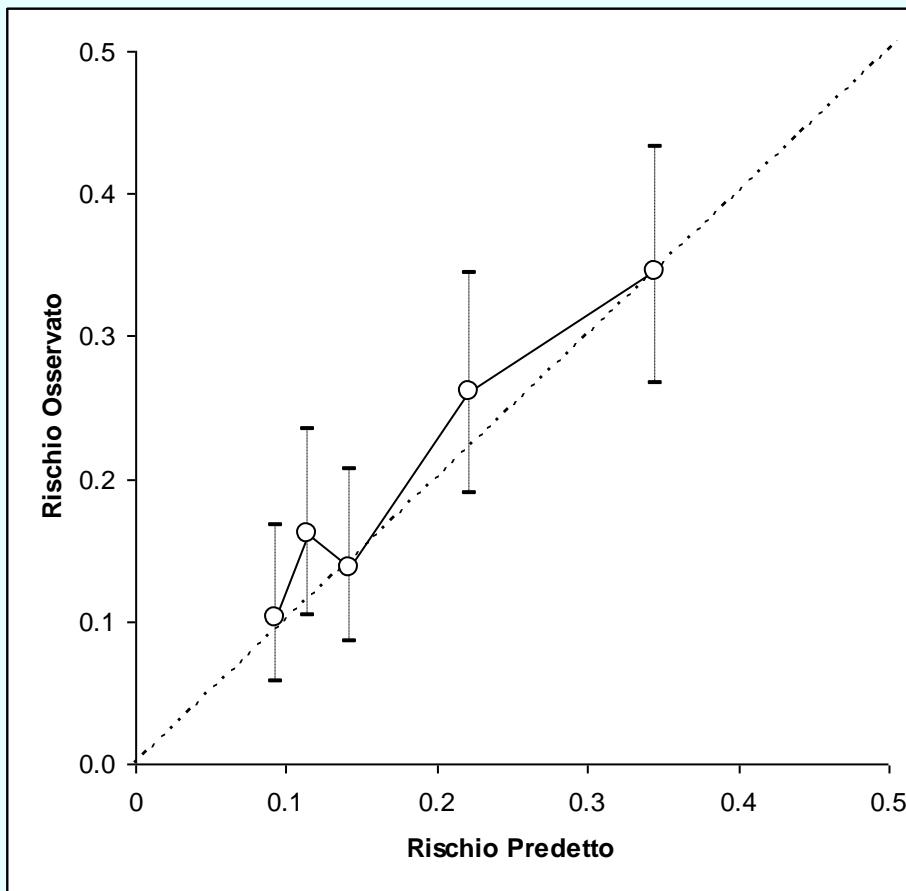
Rischio abbandono 1°anno (%)



Psicologia – Validazione su coorte 2008/09



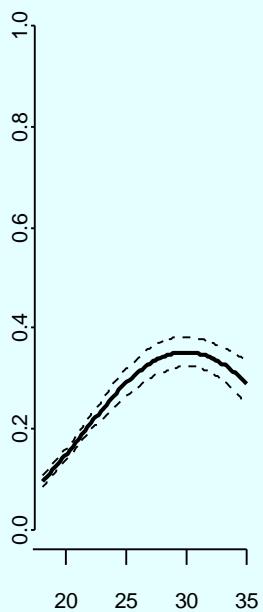
Rischio atteso di abbandono sulla base del modello predittivo e rischio osservato, per quintile di rischio predetto. Validation set.



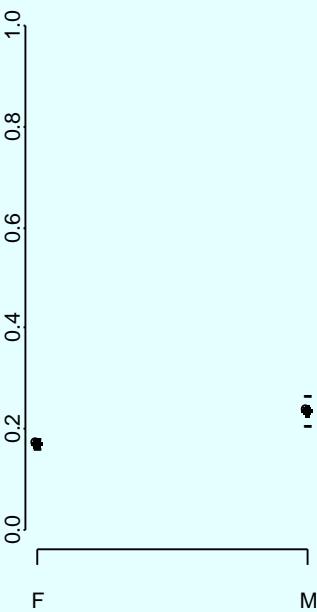
Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	136	14	12.7
basso	136	22	15.6
medio	138	19	19.6
alto	134	35	29.7
molto alto	136	47	46.9

Scienze della Formazione (Training Set)

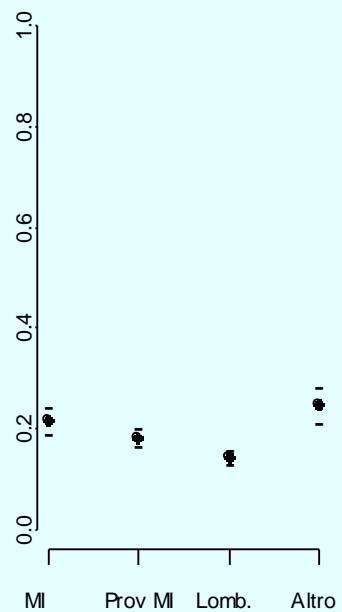
Età all'iscrizione



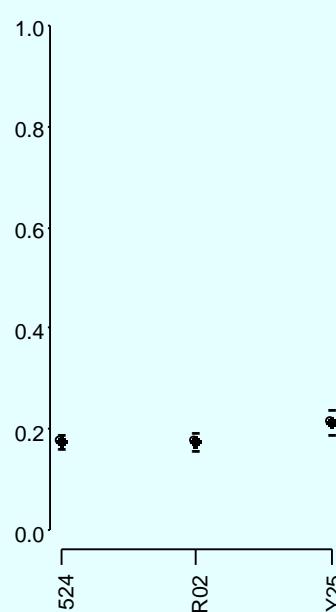
Genere



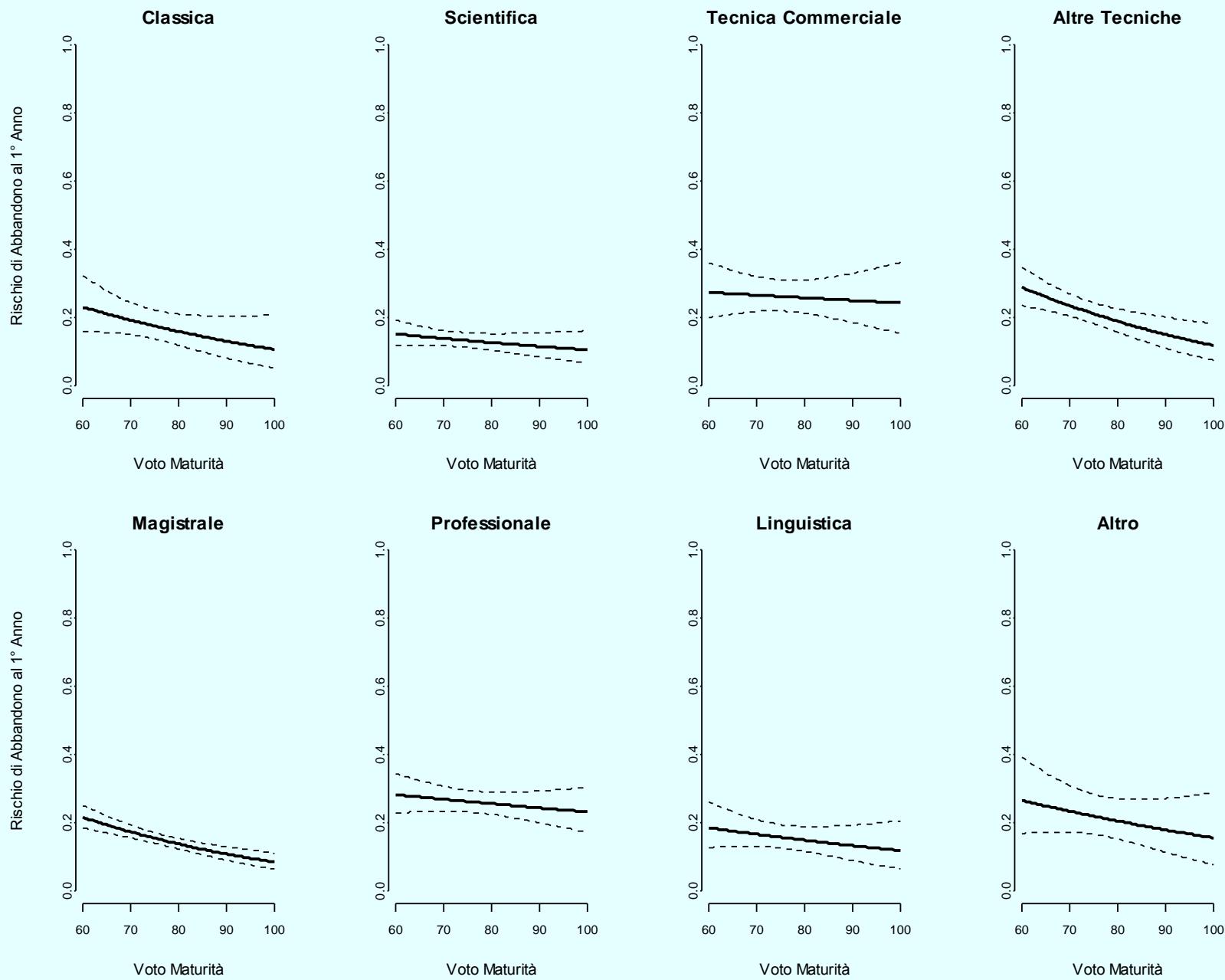
Residenza



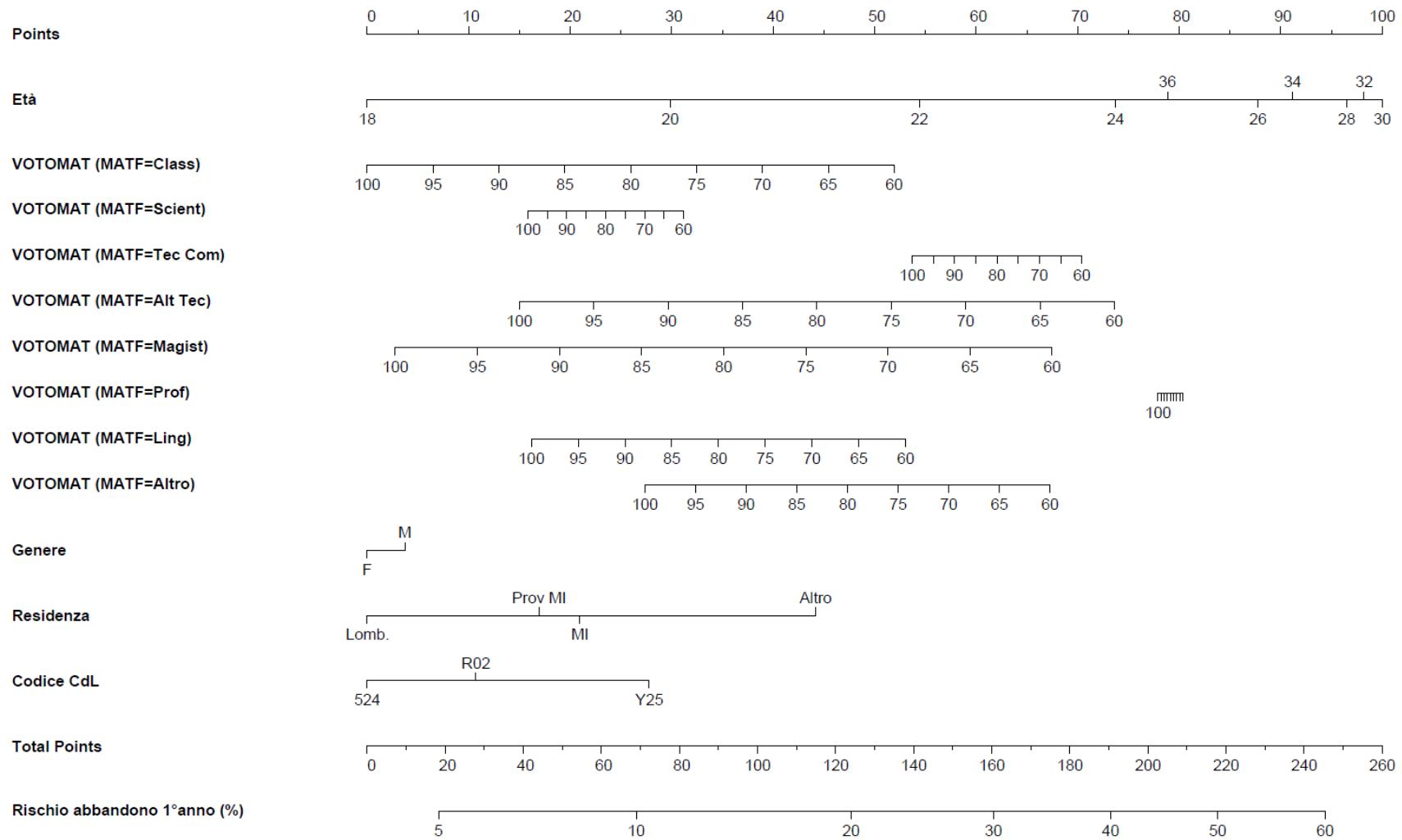
Codice CdL



Scienze della Formazione (Training Set)



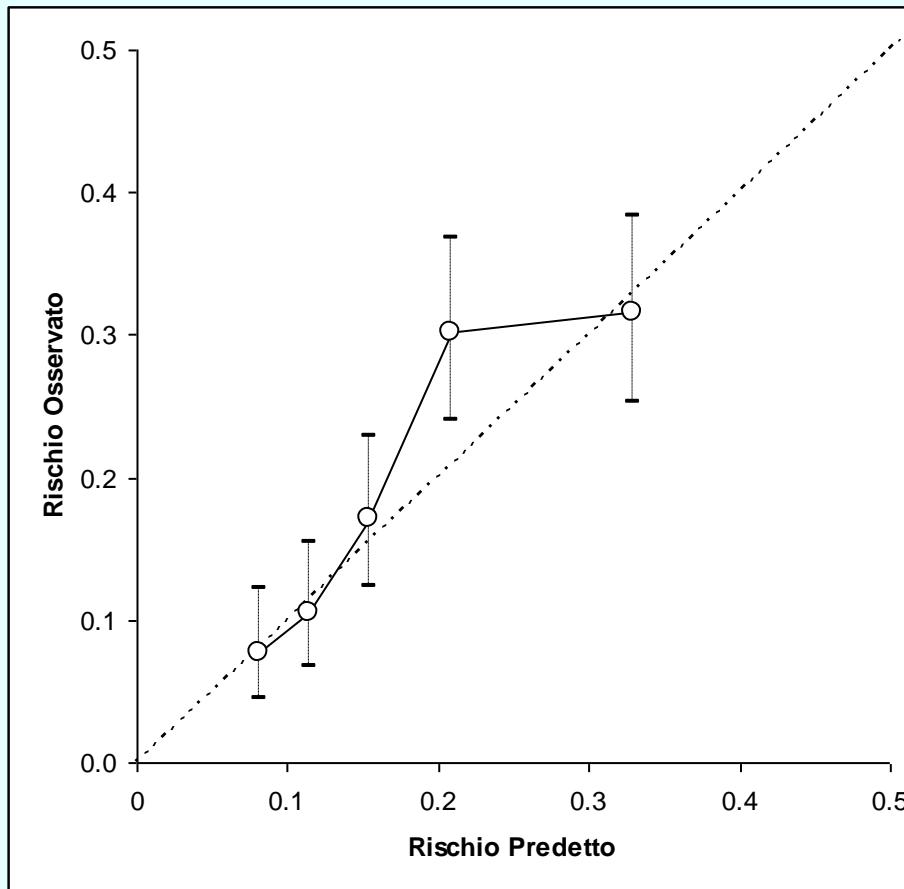
Scienze della Formazione - Nomogramma



Scienze della Formazione – Validazione su coorte 2008/09



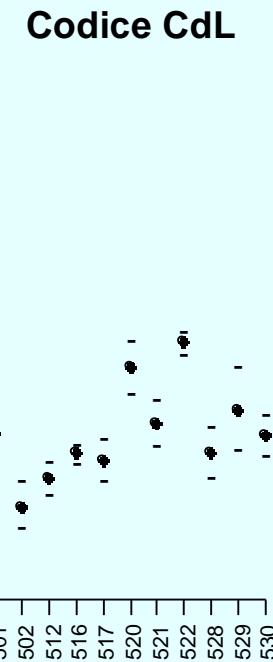
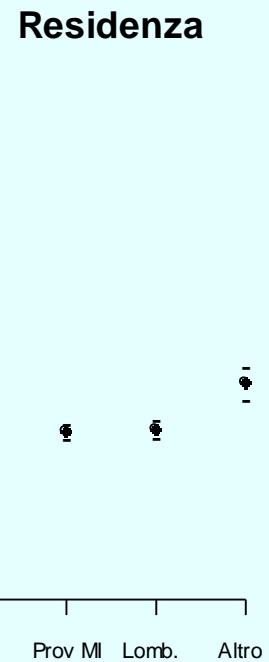
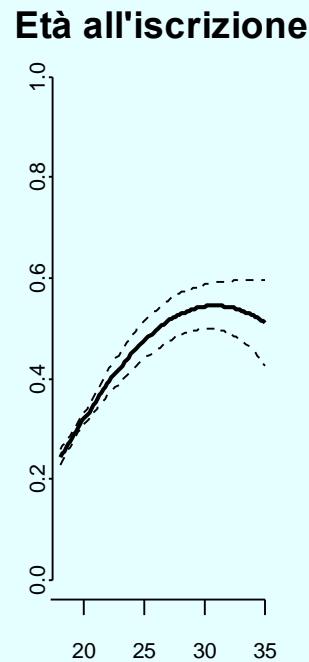
Rischio atteso di abbandono sulla base del modello predittivo e rischio osservato, per quintile di rischio predetto. Validation set.



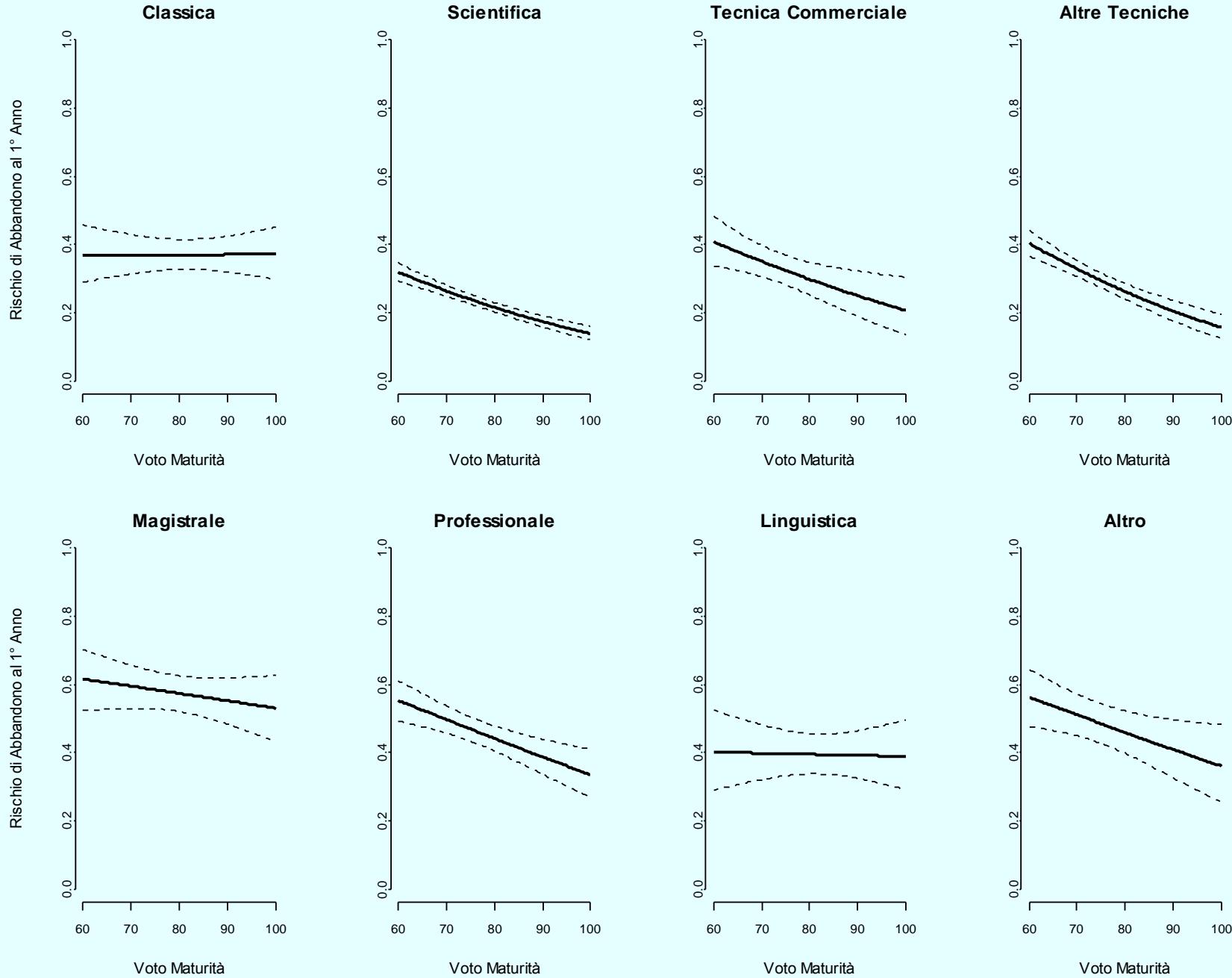
Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	208	16	16.8
basso	210	22	24.0
medio	210	36	32.4
alto	209	63	43.5
molto alto	209	66	68.8

Scienze MMFFNN (Training Set)

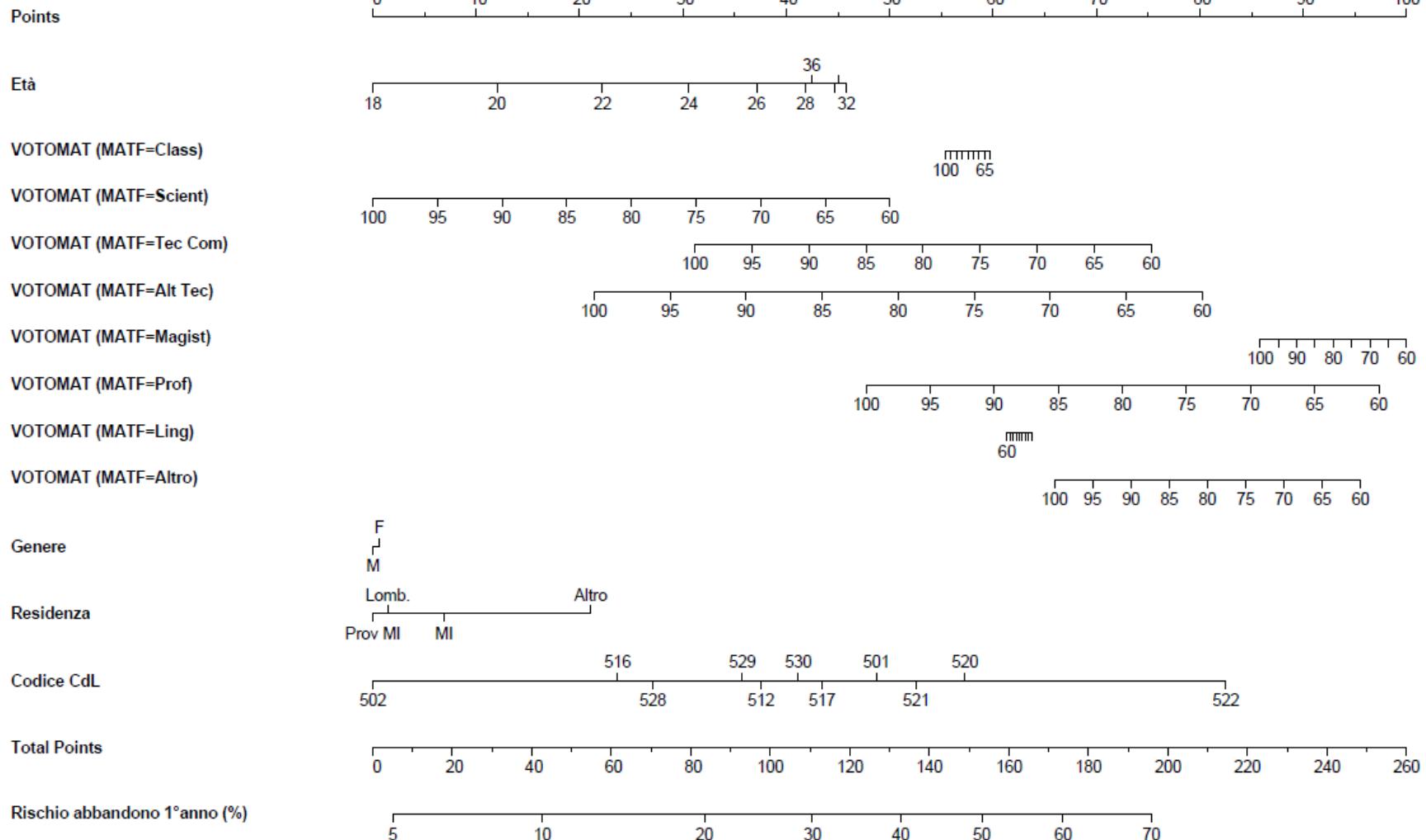
Rischio di Abbandono al 1° Anno



Scienze MMFFNN (Training Set)



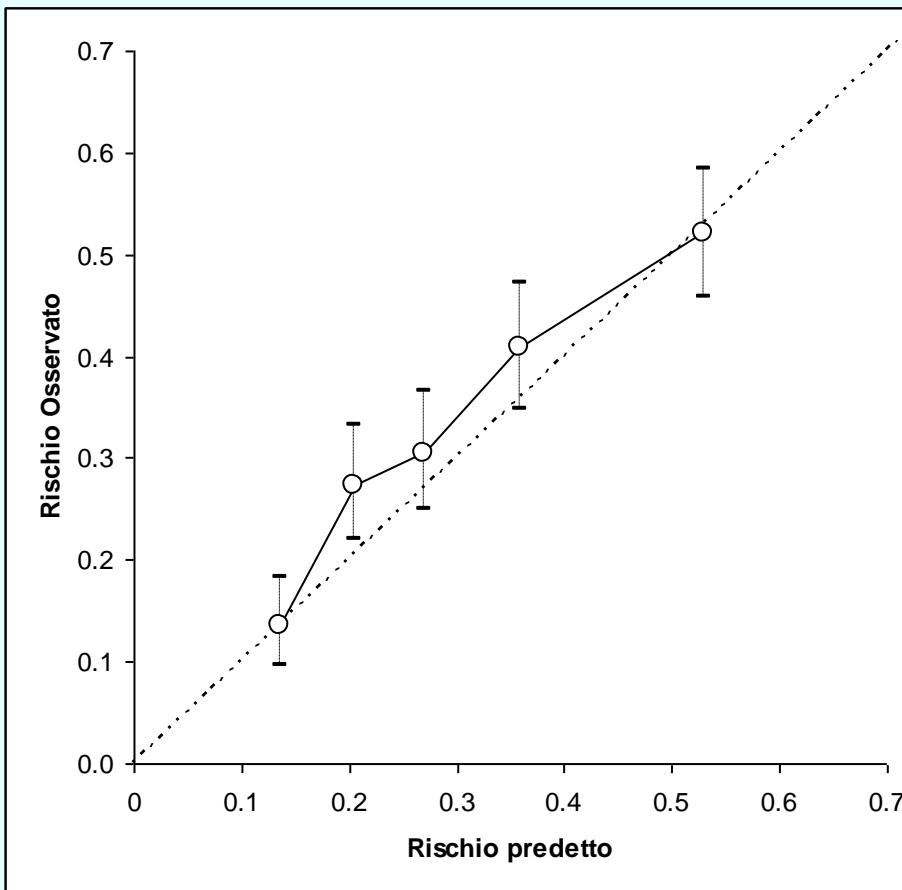
Scienze MMFFNN - Nomogramma



Scienze MMFFNN – Validazione su coorte 2008/09



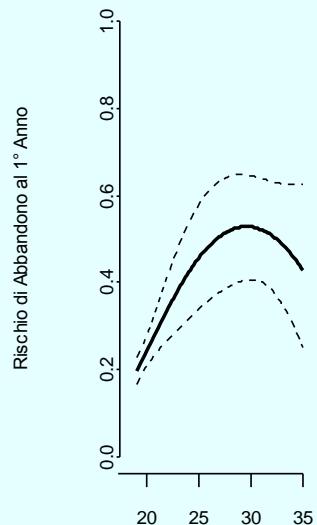
Rischio atteso di abbandono sulla base del modello predittivo e rischio osservato, per quintile di rischio predetto. Validation set.



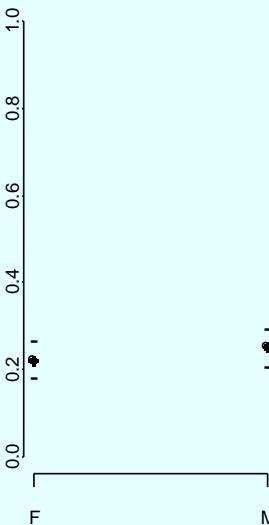
Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	258	35	35.0
basso	259	71	53.1
medio	259	79	69.7
alto	259	106	92.6
molto alto	259	135	137.1

Scienze Statistiche (Training Set)

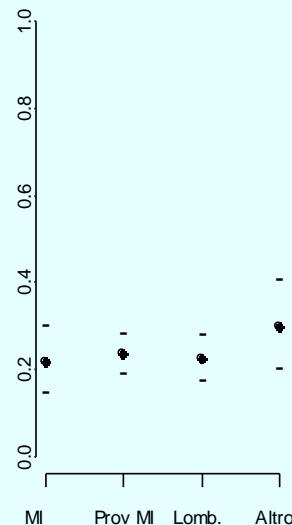
Età all'iscrizione



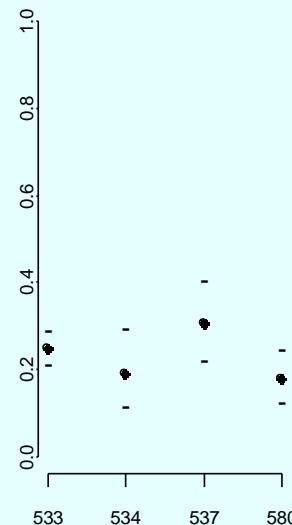
Genere



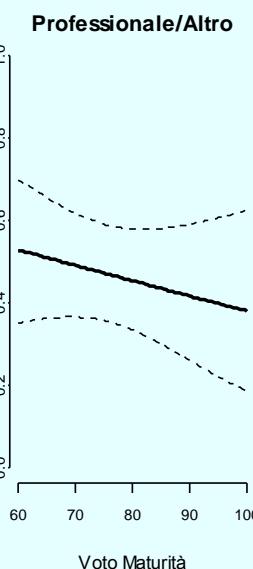
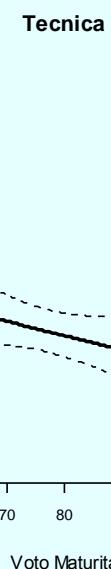
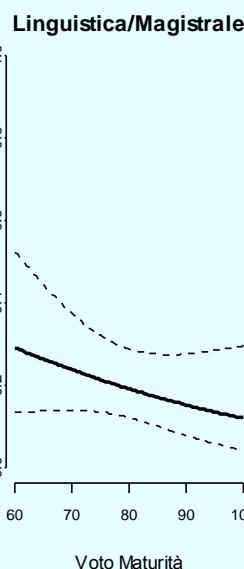
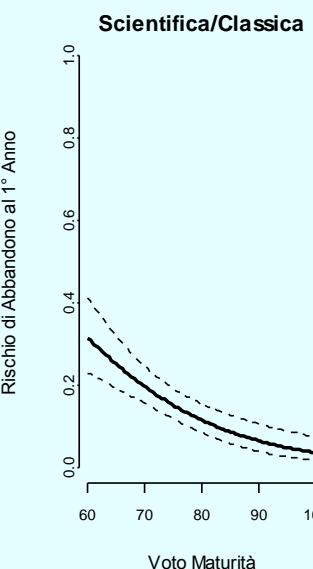
Residenza



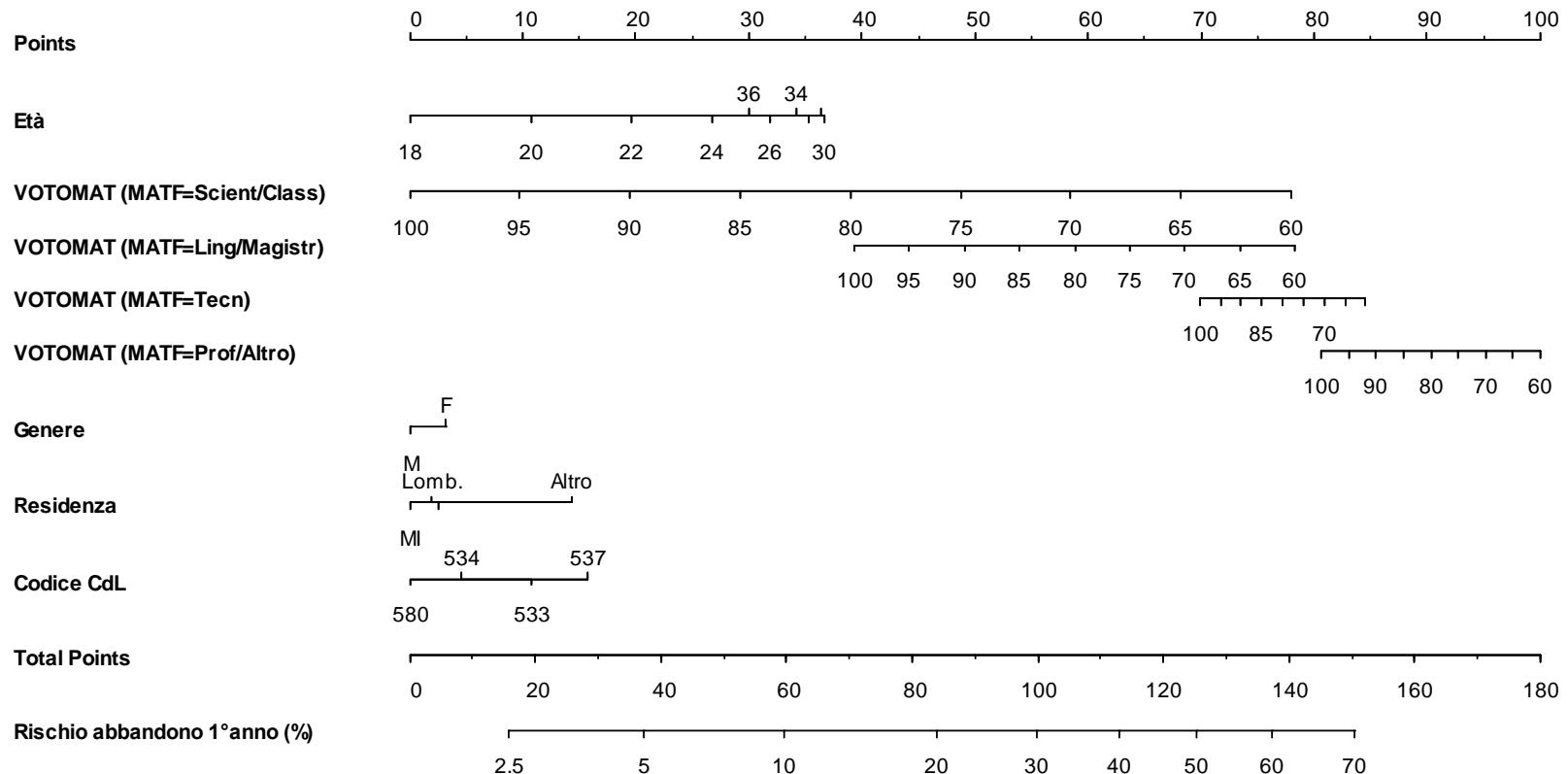
Codice CdL



Maturità / Voto



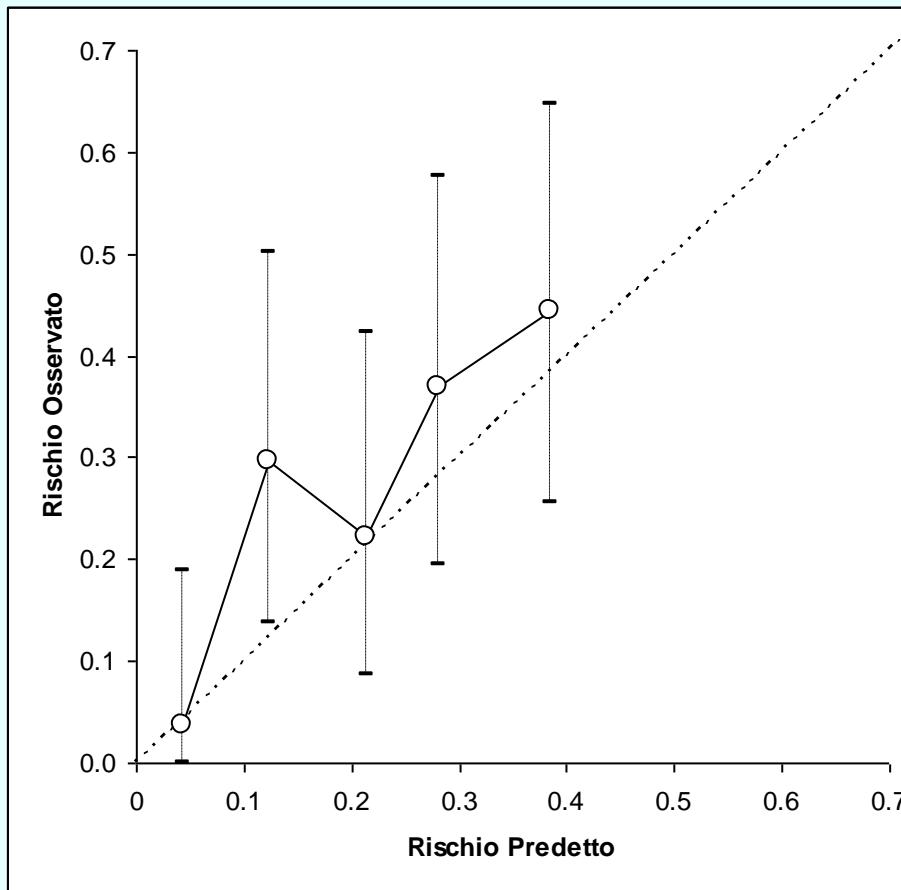
Scienze Statistiche - Nomogramma



Scienze Statistiche – Validazione su coorte 2008/09



Rischio atteso di abbandono sulla base del modello predittivo e rischio osservato, per quintile di rischio predetto. Validation set.

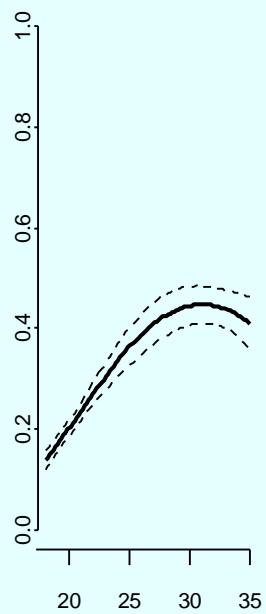


Quintile di rischio	N. immatricolati	N. Abbandoni Osservati	N. Abbandoni Predetti
molto basso	27	1	1.1
basso	27	8	3.3
medio	27	6	5.8
alto	27	10	7.6
molto alto	27	12	10.4

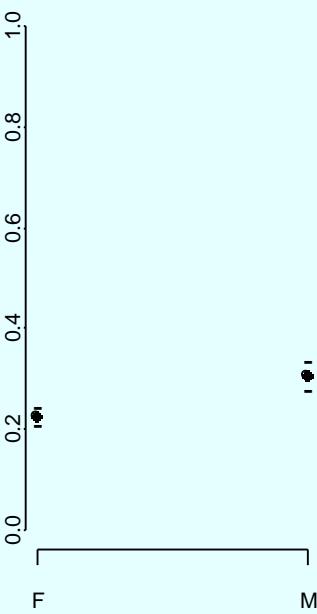
Sociologia (Training Set)

Età all'iscrizione

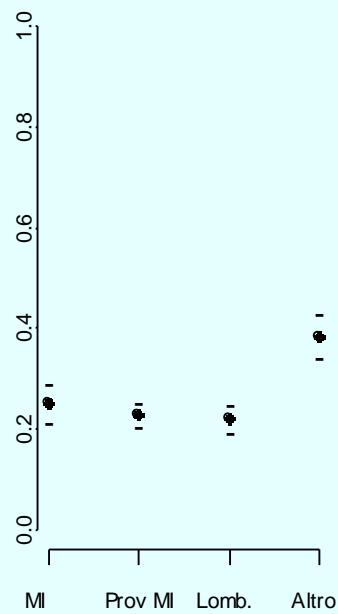
Rischio di Abbandono al 1° Anno



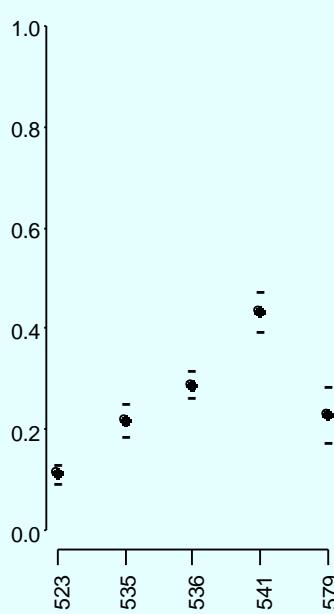
Genere



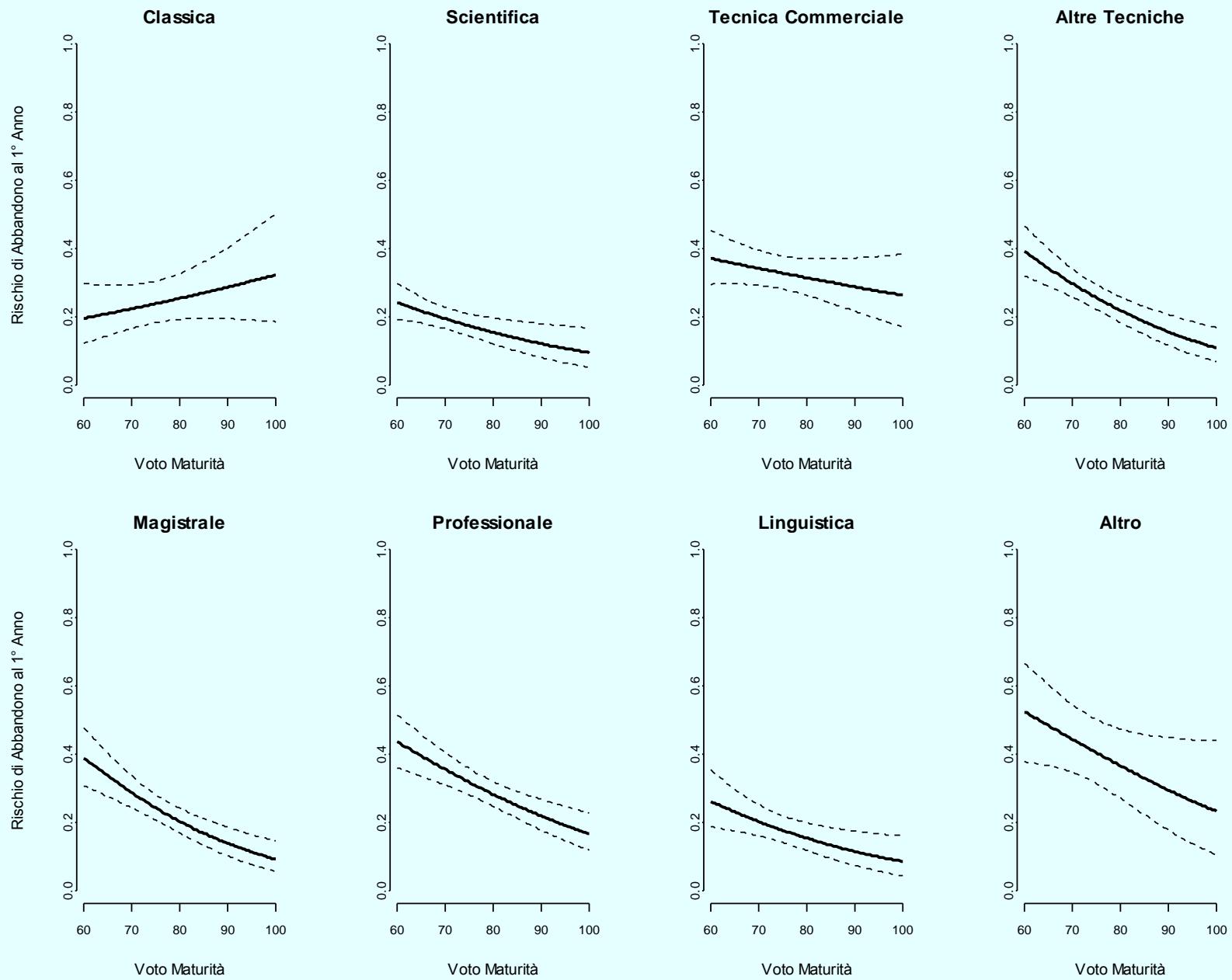
Residenza



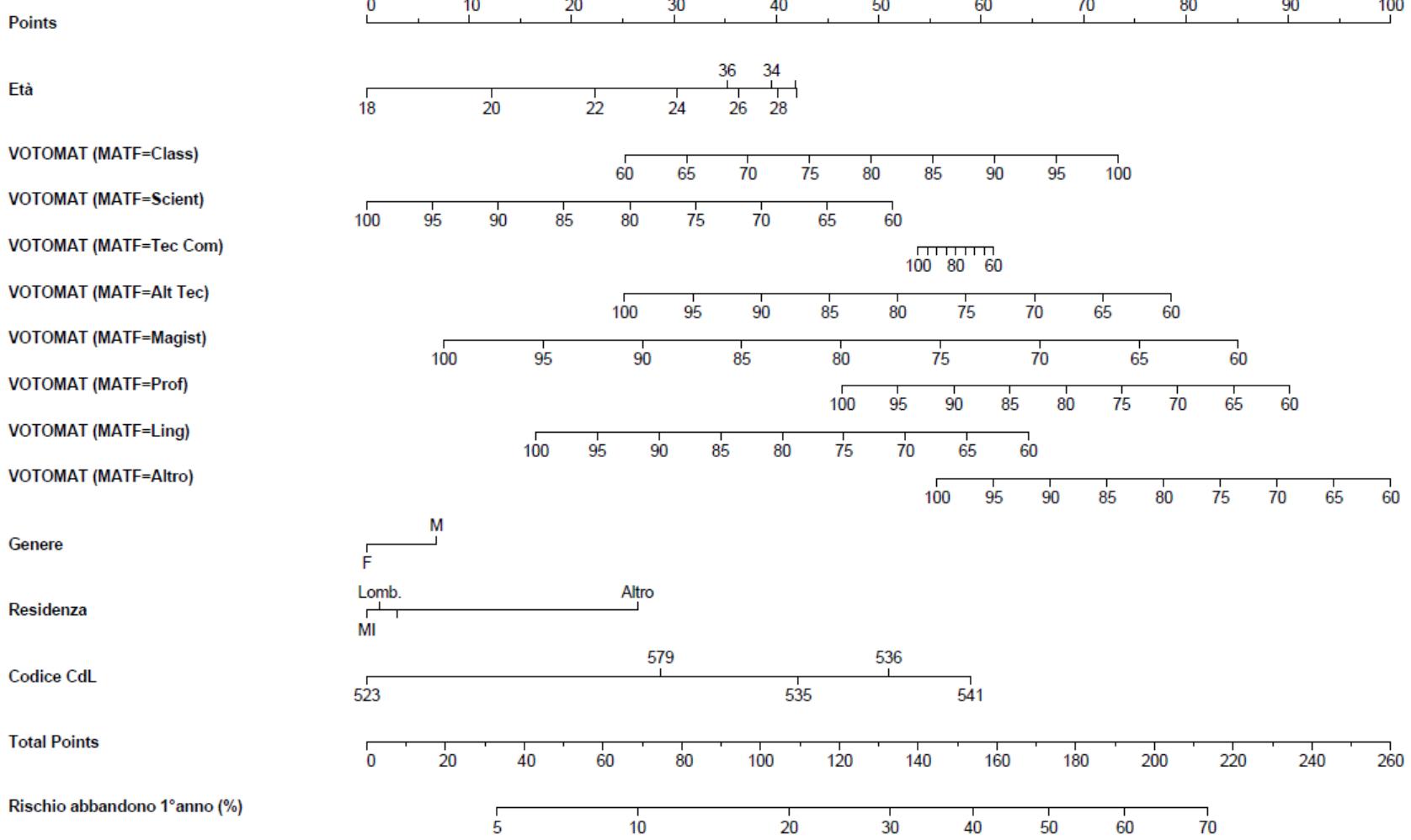
Codice CdL



Sociologia (Training Set)



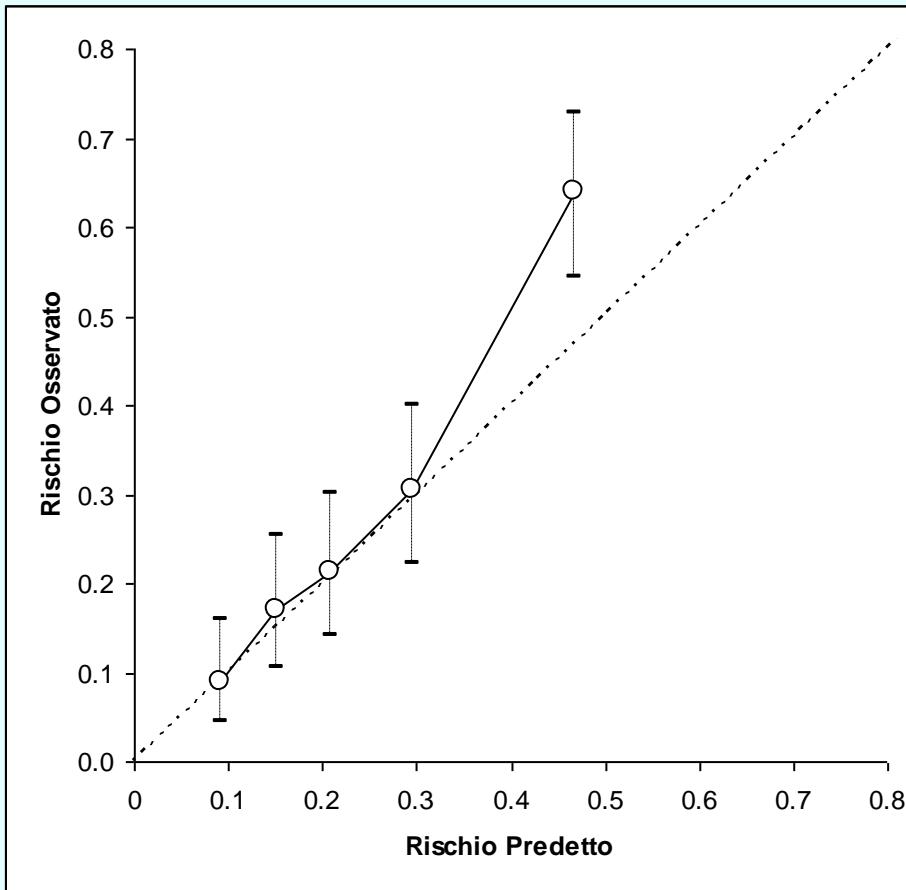
Sociologia - Nomogramma



Sociologia – Validazione su coorte 2008/09



Rischio atteso di abbandono sulla base del modello predittivo e rischio osservato, per quintile di rischio predetto. Validation set.



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

Conclusioni

- ⇒ Il rischio medio di abbandono è risultato essere molto **eterogeneo** tra le Facoltà dell'Ateneo sia quando viene valutato in termini assoluti sia quando si considera l'**effetto** che hanno su di esso le **caratteristiche degli studenti all'immatricolazione**

- ⇒ 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

Sviluppi futuri

- ⇒ Applicativo software per assegnare immediatamente ad ogni nuova matricola il suo **profilo di rischio**

- ⇒ 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**

Esempio: Scienze Statistiche

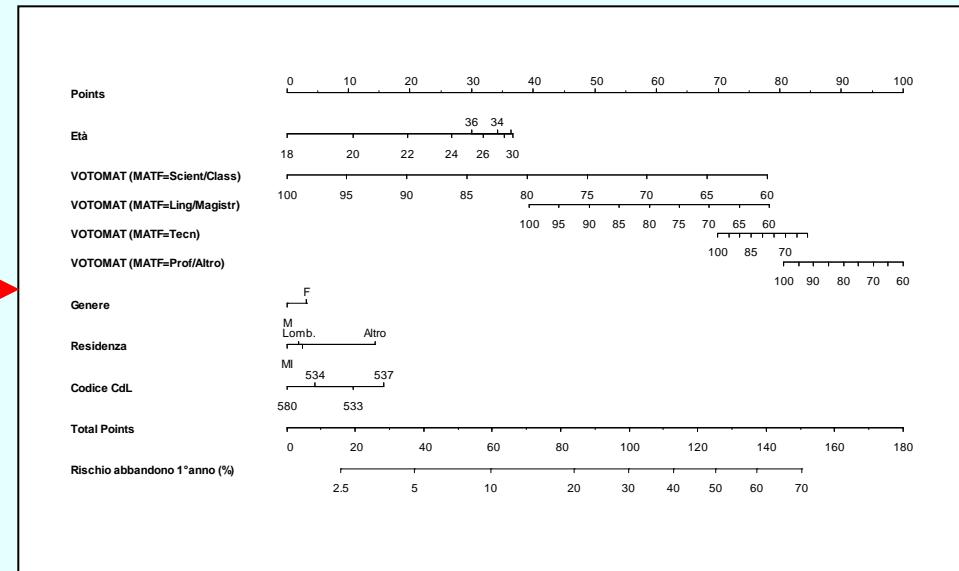
La Facoltà di Scienze Statistiche associa ad ogni studente iscritto al primo anno un docente con funzione di mentore, cioè di guida e punto di riferimento per il percorso accademico.

Si potrebbe porre maggiore attenzione agli studenti ad “alto rischio” di abbandono.

Esempio: Scienze Statistiche

⇒ Caratteristiche immatricolati 2009/10

No. immatricolati 104	
Classe di età	
<20	72
=>20	32
Genere	
F	55
M	49
Residenza	
Milano	11
Prov. MI(no Milano)	43
Lombardia(no MI/Prov)	43
Altro	7
Voto Maturità	
Alto	12
Basso	92
Maturità	
Classica	2
Scientifica	49
Tecnico commerciale	18
Altre Tecniche	20
Magistrale	5
Professionale	3
Linguistico	5
Altro	2
CdL	
533	51
580	53



Livello di rischio	No. Immatricolati (%)
Basso (0-10%)	30 (29%)
Intermedio (10-30%)	51 (49%)
Alto (>30%)	23 (22%)