

Analisi della valenza predittiva dei test di accesso TOLC-S e TOLC-B

Sommario

1. Introduzione.....	2
2. Le fonti dei dati: il dataset MUR e i dataset CISIA-B e CISIA-S.....	2
3.1 TOLC B e TIP B.....	3
3.2 TOLC B e TIP: analisi dei risultati.....	5
A1. CFU e Sezioni TEST (TOLC B/TIP).....	5
A2. CFU e altre variabili (TOLC B/TIP).....	6
A3. Applicazione del modello logit per la predittività.....	9
4.1 TOLC S.....	11
4.2 TOLC S: analisi dei risultati.....	13
B1. CFU e Sezioni TOLC S.....	13
B2. CFU e altre variabili (TOLC S).....	14
B3. Applicazione del modello logit per la predittività.....	18
Appendici.....	20
Appendice A - Matrice origine destinazione.....	20
Appendice B. Odds e odds ratio.....	21
Appendice C. Il modello logit.....	22
Probabilità predette per profilo di studente TOLC B.....	23
Probabilità predette per profilo di studente TOLC S.....	24

1. Introduzione

Questo Rapporto sulla predittività dei test di accesso TOLC-S e TOLC-B è una versione ridotta del *Rapporto sulla Predittività dei Test di Accesso TOLC-S e TOLC-B*. Il Rapporto esteso è la naturale estensione di un lavoro precedentemente condotto sui TOLC-E e sul TOLC-I (Attanasio *et al.*, 2020). L'obiettivo è studiare il "peso predittivo" o la "predittività" (insieme ad altre variabili socio-demografiche e curriculari) del punteggio ottenuto ai test TOLC-B o TOLC-S rispetto al numero di crediti che lo studente ottiene alla fine del primo anno di studio all'università.

La popolazione considerata in questo lavoro è un sottoinsieme degli studenti che nel 2018 si sono immatricolati per la prima volta al primo anno di uno dei corsi di laurea dell'area di ConScienze; i criteri di scelta del sottoinsieme sono illustrati nel capitolo 2 e più dettagliatamente nella versione estesa di questo rapporto. L'anno 2018 è stato scelto per considerare studenti che avessero completato il primo anno prima della pandemia.

Il lavoro è suddiviso in due parti:

1. analisi statistiche esplorativa, per dare una prima descrizione dei dati;
2. analisi statistiche basate su modelli per stimare gli effetti netti di alcune variabili nei confronti della variabile risposta, che è il numero dei crediti conseguiti alla fine del primo anno.

Il Rapporto è organizzato nel seguente modo: il capitolo 2 illustra le fonti dei dati; il capitolo 3 analizza il TOLC B e il TIP B; il capitolo 4 analizza il TOLC S¹.

2. Le fonti dei dati: il dataset MUR e i dataset CISIA-B e CISIA-S

Per condurre le analisi di predittività è stato necessario costruire un algoritmo di "linkage" tra il *dataset* CISIA e il *dataset* dell'Anagrafe Nazionale degli Studenti del Ministero dell'Università e della Ricerca MUR. Il nuovo dataset MUR&CISIA contiene, per ogni studente, informazioni relative alle performance universitarie (misurate attraverso l'acquisizione dei CFU alla fine dell'anno solare 2019); informazioni relative alle performance scolastiche e ai punteggi ai TOLC/TIP. In particolare, suddividiamo i dati a disposizione in tre categorie come segue:

1) Dati anagrafici

- Sesso (F/M); Anno e mese di nascita; Comune di Residenza

2) Dati relativi al percorso scolastico effettuato

- Tipo scuola frequentata, secondo la classificazione del CISIA (LC= Liceo classico; LS=Liceo scientifico; IT= Istituti tecnici; AL=Altre scuole secondarie); Voto di diploma; Regione della scuola frequentata

3) Dati relativi al test effettuato

¹ In generale, l'obiettivo di questo lavoro, insieme al precedente condotto sui TOLC-E TOLC-I, è indagare sul ruolo che i test di accesso hanno nel passaggio dalla scuola di secondo grado all'università, considerato che nella coorte degli immatricolati dell'anno accademico 2018/19 l'abbandono e il cambio di corso di studi al primo anno hanno valori ancora molto elevati, 13,6% per gli abbandoni e 11,2% di cambi di corso di laurea. Per questi ultimi una quota percentuale consistente è data dai cambi verso Medicina, che sono stati il 5,9%. In Appendice A è riportata una matrice origine-destinazione dal 1° al 2° anno con un dettaglio dei passaggi dalle lauree scientifiche di base anche verso i corsi di laurea di area medica. In generale, i tassi di abbandono e di cambio di corso di laurea, infatti, continuano ad essere troppo alti, anche se si è avuto un miglioramento negli ultimi anni.

- Sede e regione in cui è stato svolto in test; Punteggio al TOLC (S o B) e TIP distinto per singole sezioni

4) Dati relativi al percorso accademico intrapreso

- Ateneo di immatricolazione; Classe di laurea; Indicatore relativo alla prima immatricolazione (Sì/No); CFU sostenuti al primo anno o meglio fino alla fine dell'anno solare 2019.

Per entrambi i TOLC, la variabile di interesse, associata ai punteggi riportati nelle Sezioni dei Test di Accesso, è data dai CFU al 1° anno, in quanto quest'ultima da studi precedenti è risultata una variabile predittiva del futuro "successo" universitario, inteso come il conseguimento del titolo triennale entro 4 anni dall'immatricolazione.

3.1 TOLC B e TIP B

Il passaggio ai TOLC non è avvenuto simultaneamente in tutte le sedi, quindi, per l'anno 2018-19 considerato in questo lavoro, abbiamo il TOLC B (ben 8933) e il TIP B (6249), mentre per i test S non verranno considerati i TIP S perché erano stati circoscritti alla sola sede di Firenze.

I record *matched* (appaiati) per Classe di Laurea (CdS) sono riportati nella Tabella 1 e nella Figura 1 e risultano pari a 54,1 per il TOLC-B e al 53,9% per il TIP-B. Una procedura statistica di validazione della procedura di matching è stata applicata. Abbiamo riportato anche i record *matched* associati a immatricolazioni nei corsi di laurea in campo medico e in altri corsi di laurea non vicini alle aree scientifiche specifiche dei TIP e TOLC S.

Tabella 1. Percentuale di matching per ateneo, sede del test e classe di laurea di immatricolazione.

ATENEO SEDE TEST	IMMATRICOLATI L-02	MATCHE D L-02	% L-02	IMMATRICOLATI L-13	MATCHE D L-13	% L-13	Totale immatricolati L-02 e L-13	Totale matched L-02 e L-13	% L-02+L-13	IMMATRICOLATI ALTRI CdS	MATCHED ALTRI CdS	% ALTRI CdS	Totale matched tutti i CdS
Basilicata	76	28	36,8				76	28	36,8	857	16	1,9	44
Cagliari	54	35	64,8	228	131	57,5	282	166	58,9	3646	236	6,5	402
Calabria	43	23	53,5	102	62	60,8	145	85	58,6	3600	56	1,6	141
Catania	54	26	48,1	163	120	73,6	217	146	67,3	5863	711	12,1	857
Genova	51	3	5,9	91	45	49,5	142	48	33,8	5504	47	0,9	95
L'Aquila	76	61	80,3	115	86	74,8	191	147	77,0	1759	39	2,2	186
Messina	32	11	34,4	224	169	75,4	256	180	70,3	3445	166	4,8	346
Milano	296	40	13,5	202	131	64,9	498	171	34,3	10146	261	2,6	432
Modena e Reggio Emilia	63	39	61,9	62	38	61,3	125	77	61,6	5093	137	2,7	214
Napoli Federico II	461	91	19,7	824	274	33,3	1285	365	28,4	12138	322	2,7	687
Pavia	129	42	32,6	209	74	35,4	338	116	34,3	4382	46	1	162
Pisa	54	52	96,3	260	253	97,3	314	305	97,1	6747	337	5	642
Roma La Sapienza	292	147	50,3	328	231	70,4	620	378	61,0	16381	610	3,7	988
Roma Tor Vergata	64	61	95,3	234	186	79,5	298	247	82,9	4405	168	3,8	415
Roma Tre				101	70	69,3	101	70	69,3	5366	26	0,5	96
Salerno				125	120	96	125	120	96,0	5800	347	6	467
Sannio	132	85	64,4	139	81	58,3	271	166	61,3	498	3	0,6	169
Siena	68	4	5,9	127	80	63	195	84	43,1	2312	33	1,4	117
Torino	138	25	18,1	222	73	32,9	360	98	27,2	12774	143	1,1	241
Trento	55	39	70,9				55	39	70,9	2928	66	2,3	105
Udine	53	47	88,7				53	47	88,7	3012	85	2,8	132
Verona	86	70	81,4				86	70	81,4	3971	151	3,8	221
Totale Atenei sede test	2277	929	40,8	3756	2224	59,2	6033	3153	52,3	120627	4006	3,3	7159
Altri Atenei non sede (n)	4413	269	6,1	4419	124	2,8	8832	393	4,4	164921	657	0,4	1050
Totale	6690	1198	-	8175	2348	-	14865	3546	-	285548	4663	-	8209

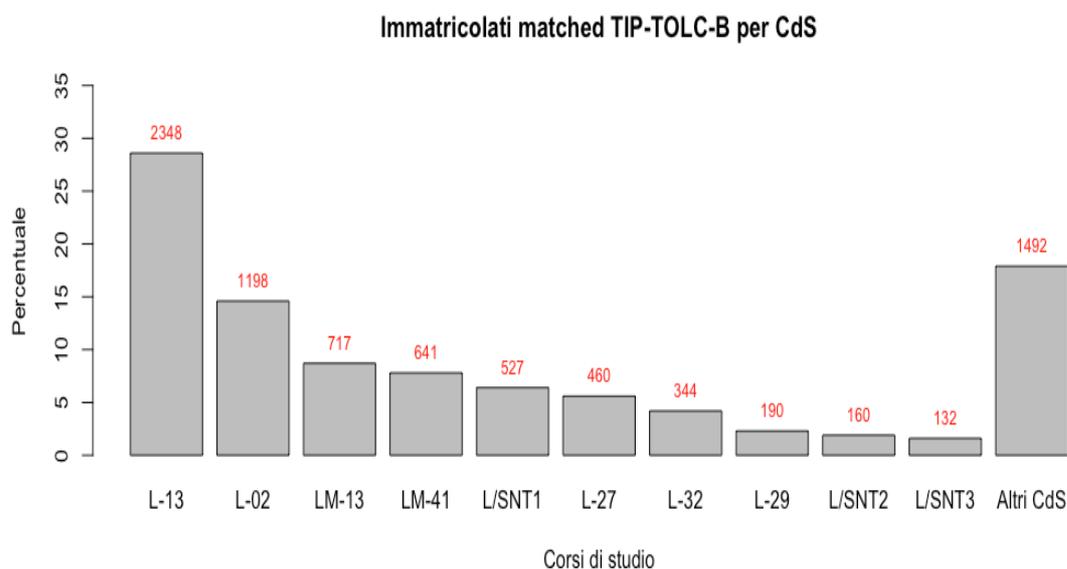


Figura 1 Numero di record matched per corso di studio del MUR&CISIA-B .

3.2 TOLC B e TIP: analisi dei risultati

Di seguito vengono analizzati:
in via esplorativa:

- (A1) la relazione tra i punteggi ai TOLC B e TIP rispetto al superamento della soglia di 20 o 40 CFU acquisiti al primo anno nei vari corsi di studio;
- (A2) l'effetto delle altre variabili, come il genere, il voto ed il tipo di diploma rispetto al superamento della soglia di 20 o 40 CFU;

con modelli statistici *logit*:

- (A3) la predittività del TOLC B e TIP, misurata come relazione tra punteggio conseguito al test e probabilità di superare le soglie di 20 e 40 CFU, per corso di studio.

A1. CFU e Sezioni TEST (TOLC B/TIP)

In Tabella 2 vengono riportati i punteggi medi ai test condizionati a tre classi di CFU acquisiti al primo anno, separate dalle soglie 20 e 40.

Tabella 2 Punteggi medi nelle sezioni del test e in totale rispetto ai CFU

Sezione test	CFU			Media totale
	≤ 20	20 + 40	> 40	
Punt. Medio	17,4	21,1	26,4	21,4
Punt. Medio Biologia	4,7	5,3	5,9	5,2
Punt. Medio Chimica	3,4	4,0	4,9	4,1
Punt. Medio Fisica	2,1	2,5	3,4	2,6
Punt. Medio Matematica	7,2	9,4	12,2	9,5

Per tutti i test considerati si osservano punteggi medi al test crescenti al crescere dei CFU acquisiti.

A2. CFU e altre variabili (TOLC B/TIP)

La Tabella 3 riporta dei dati riassuntivi sulla distribuzione dei CFU al 1° anno rispetto ai Corsi di Studio L-02 e L-03, al Genere, alla Macroregione, al tipo di Test, al Voto al Diploma e al Tipo di diploma. Per i CFU al 1° anno abbiamo creato 3 classi, in modo da classificare gli studenti come quelli con una performance certamente soddisfacente (>40 CFU), una performance intermedia (20 – 40 CFU) e una ultima poco soddisfacente (<=20 CFU).

Osservando la Tabella 3 si evince chiaramente che:

- la copertura delle 3 classi è nel totale abbastanza omogenea con solo ¼ di studenti nella categoria ≤ 20 ;
- la migliore performance degli studenti di Biotecnologie, infatti quasi il 50% ha più di 40 CFU;
- una performance simile tra i generi;
- la migliore performance al Nord Est seguito dal Nord Ovest;
- migliori risultati per chi aveva sostenuto il TOLC rispetto al TIP.

Tabella 3 Percentuale di riga del numero di studenti per fascia di cfu condizionatamente alle variabili di interesse

Variabile	Modalità	CFU (% di riga)			Totale
		≤ 20	20 – 40	> 40	
CdS	L-13	28,8	39,8	31,4	2024
	L-02	19,4	32,4	48,2	1093
Genere	F	25,3	37,5	37,2	2194
	M	26,0	36,6	37,4	923
Tipo diploma	LC	31,2	36,7	32,1	507
	LS	20,9	36,3	42,8	1925
	IT	30,7	36,9	32,4	244
	AL	36,1	42,2	21,8	441
Voto diploma	<90	31,2	38,2	30,5	2023
	≥ 90	14,9	35,8	49,3	1079
Macro-regione	Nord-ovest	20,2	31,7	48,1	470
	Nord-est	12,1	25,1	62,7	354
	Centro	23,6	37,8	38,6	954
	Sud	32,3	41,0	26,8	896
	Isole	32,5	43,9	23,6	440
Tipo test	TIP	26,3	38,7	35,0	2001
	TOLC	24,1	34,5	41,4	1116
Totale		25,5	37,2	37,3	3117

Consideriamo adesso le distribuzioni dei punteggi in totale e nelle diverse sezioni del test secondo i quartili per L-02 e L-13 (Tabelle 4 e 5).

Tabella 4. Percentuale (%) degli studenti immatricolati che prendono più di 20 CFU, condizionata alle classi di punteggio al test separate dai quartili (Classi); odds e odds ratio (OR) di ciascuna classe nei confronti della prima classe, per classe di CdS.

Soglia 20 CFU	L-02 (n=1198)				L-13 (n=2348)			
Sezione	Classi	%	odds	OR	Classi	%	odds	OR
intero test	min – Q1	51,2	1,05	1	min – Q1	42,2	0,73	1
	Q1 – Q2	70,3	2,37	2,26	Q1 – Q2	59,7	1,48	2,03
	Q2 – Q3	84,4	5,41	5,16	Q2 – Q3	64,9	1,85	2,53
	Q3- max	91,1	10,24	9,76	Q3- max	82,1	4,59	6,28
Biologia	min – Q1	62,8	1,69	1,00	min – Q1	47,9	0,92	1,00
	Q1 – Q2	70,5	2,39	1,42	Q1 – Q2	62,6	1,67	1,82
	Q2 – Q3	79,8	3,95	2,34	Q2 – Q3	64,6	1,82	1,98
	Q3- max	85,8	6,04	3,58	Q3- max	76,4	3,24	3,52
Chimica	min – Q1	57,9	1,38	1,00	min – Q1	51,8	1,07	1,00
	Q1 – Q2	72,6	2,65	1,93	Q1 – Q2	58,2	1,39	1,30
	Q2 – Q3	82	4,56	3,31	Q2 – Q3	66,4	1,98	1,84
	Q3- max	85,5	5,90	4,29	Q3- max	73,9	2,83	2,63
Fisica	min – Q1	61,2	1,58	1,00	min – Q1	55,9	1,27	1,00
	Q1 – Q2	71,9	2,56	1,62	Q1 – Q2	60,1	1,51	1,19
	Q2 – Q3	74,5	2,92	1,85	Q2 – Q3	60,8	1,55	1,22
	Q3- max	90,2	9,20	5,84	Q3- max	72,3	2,61	2,06
Matematica	min – Q1	49,7	0,99	1,00	min – Q1	43,7	0,78	1,00
	Q1 – Q2	71,3	2,48	2,51	Q1 – Q2	58,9	1,43	1,85
	Q2 – Q3	85,8	6,04	6,12	Q2 – Q3	66,8	2,01	2,59
	Q3- max	90,8	9,87	9,99	Q3- max	80,6	4,15	5,35

Tabella 5. Percentuale (%) degli studenti immatricolati che prendono più di 40 CFU, condizionata alle classi di punteggio al test separate dai quartili (Classi), odds e odds ratio (OR) di ciascuna classe nei confronti della prima classe, per classe di CdS.

Soglia 40 CFU	L-02 (n=1198)				L-13 (n=2348)				
	Sezione	Classi	%	odds	OR	Classi	%	odds	OR
Intero test		min – Q1	18,9	0,23	1,00	min – Q1	10,9	0,12	1,00
		Q1 – Q2	40,9	0,69	2,97	Q1 – Q2	22,9	0,30	2,43
		Q2 – Q3	51,8	1,07	4,61	Q2 – Q3	31,5	0,46	3,76
		Q3- max	70,6	2,40	10,30	Q3- max	47,2	0,89	7,31
Biologia		min – Q1	31,1	0,45	1,00	min – Q1	14,7	0,17	1,00
		Q1 – Q2	41,1	0,70	1,55	Q1 – Q2	28,2	0,39	2,28
		Q2 – Q3	52,3	1,10	2,43	Q2 – Q3	31,6	0,46	2,68
		Q3- max	60,3	1,52	3,37	Q3- max	40,2	0,67	3,90
Chimica		min – Q1	29,7	0,42	1,00	min – Q1	19,3	0,24	1,00
		Q1 – Q2	40,8	0,69	1,63	Q1 – Q2	21,9	0,28	1,17
		Q2 – Q3	47,5	0,90	2,14	Q2 – Q3	31,9	0,47	1,96
		Q3- max	65,6	1,91	4,51	Q3- max	40,7	0,69	2,87
Fisica		min – Q1	28	0,39	1,00	min – Q1	21,7	0,28	1,00
		Q1 – Q2	40,7	0,69	1,76	Q1 – Q2	23,9	0,31	1,13
		Q2 – Q3	48,6	0,95	2,43	Q2 – Q3	28	0,39	1,40
		Q3- max	66,1	1,95	5,01	Q3- max	38,9	0,64	2,30
Matematica		min – Q1	21,6	0,28	1,00	min – Q1	12,3	0,14	1,00
		Q1 – Q2	39,5	0,65	2,37	Q1 – Q2	22,8	0,30	2,11
		Q2 – Q3	55,9	1,27	4,60	Q2 – Q3	30	0,43	3,06
		Q3- max	66,2	1,96	7,11	Q3- max	48,4	0,94	6,69

I valori percentuali per riga permettono di confrontare i vari CdS in termini di “difficoltà”, a parità di posizione percentile. Si nota infatti che il gruppo dell’ultimo quartile ha una probabilità di superare 20 CFU tra l’80% e il 90% in tutte le parti del test per la L-02 e che queste percentuali sono circa il 10% in meno per la L-13 tranne che per la sezione di Fisica (Tabella 4). Per quanto riguarda la probabilità di superare 40 CFU (Tabella 5), agli stessi quartili corrispondono ovviamente valori più bassi, compresi tra il 40% e il 60% circa per L-02 e tra 40% e 50% per la L-13.

Una lettura dei valori percentuali per colonna dà invece una idea di come i quarti – in termini di studenti – riescono a superare le soglie di 20 CFU e 40 CFU. È interessante confrontare le differenze dei valori percentuali di ogni quarto rispetto al precedente, per avere una idea dell’incremento nelle probabilità rispetto all’essere in un quarto successivo, o anche rispetto alla prima. Per la soglia di 20 CFU, i valori percentuali più distanti si hanno nelle sezioni di Matematica e del Totale, intorno al 40% per entrambi i CdS (Tabella 4), mentre per la soglia di 40 CFU (Tabella 5), il CdS L-02 presenta differenze più pronunciate e variabili (51,7 per la sezione Totale e 29,2 per Biologia) e la L-13 ha differenze meno pronunciate (36,3 per la sezione Totale e 17,2 per Fisica),

Calcoliamo adesso gli *odds* e gli *odds ratio* (rapporti di quote) rispetto al primo quarto di studenti. Per esempio, l’*odds* dell’ultimo quarto L-02 del Totale è pari a $0,91/(1-0,91) = 10,2$: in questo caso, possiamo dire che, per coloro il cui punteggio ricade in questo quarto, la probabilità di superare 20 CFU è 10,2 volte più alta di quella di non superarla. È anche possibile confrontare tale valore con quello ottenuto nella prima classe, facendone il rapporto (*odds ratio*).

Operiamo allo stesso modo per la classe L-13, ovvero considerando la probabilità di superare la soglia di 20 CFU al primo anno per l'ultimo quarto, ovvero per gli studenti che hanno avuto i migliori risultati nelle diverse sezioni del test. In questo CdS, coloro che hanno preso un Punteggio Totale che ricade nell'ultimo quarto, sono per il Punteggio Totale tra poco più dell'80% e poco più del 70%. Gli *odds*, a loro volta, sono tra 4,59 (Punteggio Totale) e 2,61 (Fisica), denotando che la probabilità di superare la soglia di 20 CFU al primo anno è – per gli studenti dell'ultimo quarto - tra 4,59 e 2,61 volte maggiore rispetto a quella di non superarla. Questi valori degli *odds* per L-13, inferiori a quelli riscontrati per L-02, denotano la “maggiore difficoltà” a superare la soglia di 20 CFU a parità di risultato percentile nei test di accesso.

Considerazioni analoghe si possono avanzare quando si considera la soglia di 40 CFU (Tabella 5), infatti i valori degli *odds* per la L-02 variano tra 1,5 e 2, mentre per la L-13 sono inferiori a 1. Infatti, per la L-13, ad esempio, la probabilità di superare 40 CFU è 0,89 volte più bassa rispetto a quella di superarla per gli studenti che hanno riportato un punteggio totale al Test tra 26,5 e 47,8.

A3. Applicazione del modello logit per la predittività

I modelli logit, meglio descritti in appendice, rappresentano un passo successivo e cruciale per valutare quantitativamente gli effetti netti di singoli parametri sul numero dei CFU acquisiti al 1° anno. Essi forniscono la stima numerica del logaritmo dell' odds-ratio (e della loro incertezza) per un insieme di parametri utilizzati per descrivere un insieme di studenti.

Per misurare la capacità del TOLC B a predire il successo universitario sono stati stimati due modelli logit: nel primo la variabile risposta è il superamento o meno della soglia dei 20 CFU, nel secondo la soglia è di 40 CFU. In Tabella 6 vengono riportate le stime dei parametri dei due modelli.

Tabella 6. Stime dei Log-OR dei modelli logistici con soglia 20 e 40 cfu. In grassetto vengono riportati i p-valori significativi.

Predittori	CFU 20			CFU 40		
	Log-OR	Std, Error	p	Log-OR	Std, Error	P
Intercetta	0,424	0,162	0,009	-2.037	0,176	<0,001
Genere: Femmina	0,061	0,104	0,559	-0,001	0,096	0,988
Diploma: IT	0,122	0,19	0,523	0,321	0,202	0,111
Diploma: LC	-0,061	0,151	0,69	0,351	0,165	0,033
Diploma: LS	0,452	0,14	0,001	0,718	0,152	<0,001
Voto di diploma	0,57	0,052	<0,001	0,578	0,05	<0,001
Macro regione: Centro	0,351	0,116	0,003	0,663	0,113	<0,001
Macro regione: Nord-est	0,574	0,206	0,005	0,948	0,163	<0,001
Macro regione: Nord-ovest	0,462	0,15	0,002	1.004	0,136	<0,001
CdS: L-02	0,293	0,102	0,004	0,452	0,091	<0,001
Test: TOLC-B	0,307	0,109	0,005	0,448	0,105	<0,001
Biologia+Chimica	0,197	0,057	0,001	0,136	0,052	0,01
Fisica	-0,068	0,056	0,226	-0,011	0,05	0,83
Matematica	0,394	0,07	<0,001	0,353	0,065	<0,001
Osservazioni	3099					

Studente baseline: maschio, del sud/isole, Altro diploma, che ha fatto il TIP B, che ha punteggi medi al test, un voto medio di diploma, ed è immatricolato in L-13. I Log-OR corrispondono ai valori dei beta stimati nelle equazioni [B2] e [B3]

I valori significativi del modello presentano un $p < 0,05$ e sono evidenziati in grassetto in tabella. La stima dell'intercetta racchiude in sé l'informazione relativa allo studente baseline, ovvero uno studente tipo, definito dal seguente profilo medio: maschio, del sud/isole, diplomato in Altro liceo, che ha punteggi medi al test, un voto medio al diploma (ovvero 83,8/100), e immatricolato in L-13. La stima dell'intercetta, a differenza di tutte le altre variabili, non è misurata in termini di log-OR, ma in termini di odds: l'esponentiale del valore numerico in questo caso fornisce l'odds di uno studente baseline, cioè $e^{0,424} = 1,53$.

Considerando le variabili significative, si nota che:

- la stima dell'effetto del Liceo Scientifico è pari a 0,452, che è un numero positivo. Ciò comporta che, rispetto a chi ha un Altro diploma (categoria baseline), l'Odds Ratio degli studenti del Liceo scientifico è pari a $e^{0,452} = 1,57$: l'aver frequentato il liceo scientifico piuttosto che un Altro liceo aumenta l'odds di uno studente baseline di un fattore 1,57.
- nel modello con CFU 40, un risultato significativo per il liceo scientifico e per il liceo classico, con valori maggiori per il liceo scientifico.
- le stime dei Log-OR sono positive per la L-02 in entrambi i modelli, denotando la "maggiore difficoltà" degli studenti della L-13 a superare entrambe le soglie;
- le altre variabili categoriali (Voto di diploma; macroregione, Tipo di test) hanno parametri statisticamente significative;
- le sezioni del Test di Biologia+Chimica e Matematica sono significative in entrambi i modelli. In particolare, nel modello con soglia 20 CFU, se riferito a L-13, l'effetto di Matematica è particolarmente significativo con un odds pari a $e^{0,394} = 1,48$ (modello con 20 CFU) e a $e^{0,353} = 1,42$ (modello con 40 CFU);
- Fisica non risulta invece significativa, quindi sembra che la Sezione Fisica non abbia capacità discriminante.

A partire dai valori ottenuti con i due modelli logit di tabella 6, sono state calcolate le probabilità di acquisire più di 20 e 40 CFU per alcuni profili di studenti (tabella 7).

Tabella 7 Probabilità di acquisire più di 20 ($P > 20$) o 40 ($P > 40$) CFU, per alcuni differenti profili di studente, predette sulla base dei corrispondenti modelli logistici.

ID	GENERE	MACRO_REG	DIPLOMA	VOTO_DIP	TIPO_TEST	BIO_CHI	FISICA	MATEMATICA	CDS	P(CFU>20)	P(CFU>40)
1	F	Centro	LS	80	TOLC-B	12	6	12	L-13	0,837	0,463
2	F	Centro	LS	80	TOLC-B	20	6	12	L-13	0,885	0,533
3	F	Centro	LS	80	TOLC-B	12	10	12	L-13	0,820	0,458
4	F	Centro	LS	80	TOLC-B	20	10	12	L-13	0,872	0,528
5	F	Centro	LS	80	TOLC-B	12	6	20	L-13	0,901	0,592
6	F	Centro	LS	80	TOLC-B	20	6	20	L-13	0,932	0,657
7	F	Centro	LS	80	TOLC-B	12	10	20	L-13	0,890	0,587
8	F	Centro	LS	80	TOLC-B	20	10	20	L-13	0,924	0,652

4.1 TOLC S

Il test d'accesso per scienze consiste soltanto del TOLC S, considerato che abbiamo escluso il TIP S, perché è stato effettuato soltanto a Firenze. La costruzione del dataset MUR&CISIA-S avviene in maniera analoga a quella del MUR&CISIA-B.

La Tabella 8² riporta le percentuali di matching per ateneo sede del test e corsi di immatricolazione. Tra gli 11888 studenti che hanno fatto il TOLC S, l'algoritmo di matching ne abbina 6739, ovvero il 56,8%. Di questi, 4600 (il 38,7% sul totale dei test) si iscrivono ad un CdS appartenenti alle lauree scientifiche. La percentuale di matching è altamente eterogenea rispetto all'ateneo sede del test e al corso di immatricolazione. La riga Totale della Tabella 8 riporta la distribuzione degli studenti matched per corso di laurea, mentre la Figura 2 riporta il diagramma a barre della stessa distribuzione, includendo anche i CdS più scelti non appartenenti all'area di scienze. Osserviamo che una percentuale che si aggira attorno all'11% si immatricola a biotecnologia (L-02) o biologia (L-13).

² In Tabella 8, e nel resto del rapporto, le classi di laurea sono indicate seguendo le diciture del DM 270: L-27 scienze e tecnologie chimiche; L-28 scienze e tecnologie della navigazione; L-30 scienze e tecnologie fisiche; L-31 scienze e tecnologie informatiche; L-32 scienze e tecnologie per l'ambiente e la natura; L-34 scienze geologiche; L-35 scienze matematiche; L-43 tecnologie per la conservazione e il restauro dei beni culturali.

Tabella 8 Percentuale di matching per ateneo sede del test e corsi di immatricolazione.

ATENEOSede TEST	Immatricolati			Totale Imm. Sci.	Totale Sci.	% Totale	Imm. Altro CdS	Matched Altro CdS	% Altri CdS																					
	L-27	Matched L-27	% L-27	L-28	Matched L-28	% L-28	L-30	Matched L-30	% L-30	L-31	Matched L-31	% L-31	L-32	Matched L-32	% L-32	L-34	Matched L-34	% L-34	L-35	Matched L-35	% L-35							L-43	Matched L-43	% L-43
Basilicata	18	3	16,7							84	39	46,4				6	2	33,3	14	7	50				122	51	41,8	811	3	0,4
Cagliari	66	41	62,1				65	50	76,9	115	83	72,2	70	2	2,9	16	8	50	47	28	59,6				379	212	55,9	3549	126	3,6
Calabria	46						50			89			59	11	18,6	46	9	19,6	63						353	20	5,7	3392	5	0,1
Camerino	40	19	47,5				23	14	60,9	163	92	56,4	9	2	22,2	12	5	41,7	41	22	53,7	19	5	26,3	307	159	51,8	945	3	0,3
Ferrara	102	2	2				44	24	54,5	90	1	1,1				26	12	46,2	37	2	5,4				299	41	13,7	5506	51	0,9
Firenze	152	9	5,9				126	1	0,8	131	3	2,3	120	9	7,5	23			76	2	2,6	24			652	24	3,7	7787	298	3,8
Insubria	104	18	17,3				25	3	12	175	9	5,1	83	1	1,2				40						427	31	7,3	2408	60	2,5
L'Aquila	50	29	58				27	16	59,3	78	36	46,2	19	1	5,3				36	24	66,7				210	106	50,5	1740	15	0,9
Messina	85	47	55,3	46	27	58,7	12	7	58,3	63	41	65,1	32	11	34,4	3	1	33,3	20	15	75				261	149	57,1	3440	27	0,8
Milano	160	123	76,9				236	172	72,9	566	382	67,5	328	152	46,3	65	32	49,2	157	93	59,2	53	32	60,4	1565	986	63,0	9079	695	7,7
Milano Bicocca	118	48	40,7				269	132	49,1	318	86	27	68	21	30,9	90	47	52,2	144	90	62,5				1007	424	42,1	4733	183	3,9
Modena e Reggio Emilia	65	1	1,5				55	2	3,6	158	2	1,3	87	24	27,6	21	11	52,4	56	2	3,6				442	42	9,5	4776	11	0,2
Pavia	84	39	46,4				73	32	43,8				145	40	27,6	29	6	20,7	62	35	56,5				393	152	38,7	4327	102	2,4
Perugia	109	69	63,3				47	1	2,1	115						16			49						336	70	20,8	4132	10	0,2
Pisa	114						170	110	64,7	334	210	62,9	124	60	48,4	28	19	67,9	132	76	57,6				902	475	52,7	6159	81	1,3
Roma La Sapienza	548	321	58,6				348	258	74,1	312	4	1,3	284	135	47,5	60	42	70	209	160	76,6	18	8	44,4	1779	928	52,2	15222	293	1,9
Roma Tor Vergata	72	10	13,9				59	18	30,5	108	25	23,1							61	7	11,5				300	60	20,0	4403	46	1
Salerno	56	4	7,1				32	24	75	357			62	38	61,3				122	7	5,7				629	73	11,6	5296	39	0,7
Siena	67	2	3				7	1	14,3				39	1	2,6	9	6	66,7	30	22	73,3				152	32	21,1	2355	7	0,3
Trento							75	51	68	151	8	5,3							93	65	69,9				319	124	38,9	2664	35	1,3
Trieste	38	1	2,6				87	6	6,9				48	1	2,1	43	19	44,2	70	48	68,6				286	75	26,2	2483	6	0,2
Udine										316	148	46,8	71						32	18	56,2				419	166	39,6	2646	6	0,2
Urbino Carlo Bo										87	32	36,8				13									100	32	32,0	2584	2	0,1
Verona										188	130	69,1							61	38	62,3				249	168	67,5	3808	35	0,9
Totale Atenei sede test	2094	786	37,5	46	27	58,7	1830	922	50,4	3998	1331	33,3	1648	509	30,9	506	219	43,3	1652	761	46,1	114	45	39,5	11888	4600	38,7	104245	2139	2,1
Altri Atenei non sede (n)	1883	18	1	271			1439	25	1,7	2970	26	0,9	1171	10	0,9	424	6	1,4	1449	27	1,9	18			9625	112	1,2	174655	453	0,3
Totale	3977	804		317	27		3269	947		6968	1357		2819	519		930	225		3101	788		132	45		21513	4712		278900	2592	

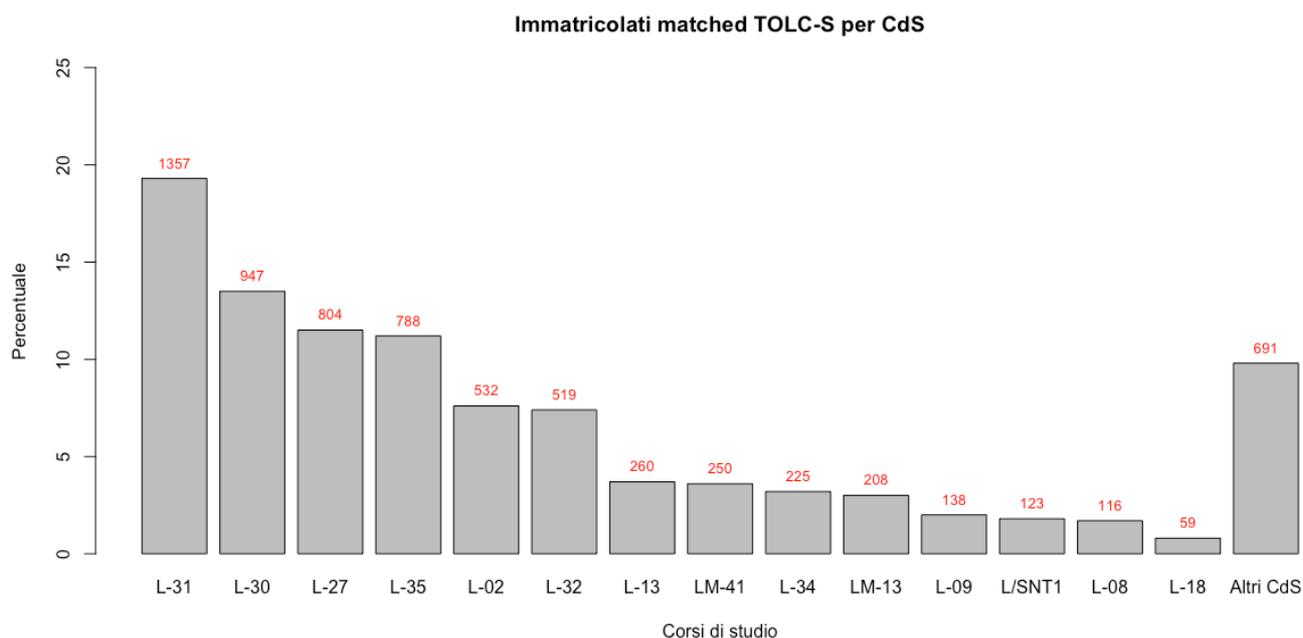


Figura 2. Distribuzione del numero di studenti matched per corso di studio nel dataset MUR&CISIA-S.

4.2 TOLC S: analisi dei risultati

Di seguito vengono analizzati:
in via esplorativa:

- (B1) la relazione tra i punteggi ai TOLC S rispetto al superamento della soglia di 20 o 40 CFU acquisiti al primo anno nei vari corsi di studio;
- (B2) l'effetto delle altre variabili, come il genere, il voto ed il tipo di diploma rispetto al superamento della soglia di 20 o 40 CFU;

con modelli logit:

- (B3) la predittività del TOLC S, misurata come relazione tra classi di punteggio e probabilità di superare le soglie di 20 e 40 CFU, per corso di studio.

B1. CFU e Sezioni TOLC S

In Tabella 9 vengono riportati i punteggi medi ai test condizionati a tre classi di CFU acquisiti al primo anno, separate dalle soglie 20 e 40.

Tabella 9 Punteggi medi nelle sezioni del test e in totale rispetto ai CFU

SEZIONI test	CFU			Media totale
	≤ 20	20 + 40	> 40	
Punt. Medio generale	17,5	22,1	28,0	23,3
Punt. Medio Testo: totale	4,1	5,0	5,9	5,1
Punt. Medio Testo: testo	2,7	3,1	3,4	3,1
Punt. Medio Testo: domande	1,4	1,9	2,5	2,0
Punt. Medio Ragionamento e Probabilità	2,9	3,6	4,5	3,8
Punt. Medio Scienze Di Base	2,6	3,5	4,4	3,6
Punt. Medio Matematica	7,9	10,2	13,1	10,8

Per tutti i test considerati si evincono gradienti crescenti nell'acquisizione dei CFU al crescere dei punteggi medi al test, specificando che i punteggi medi in matematica sono superiori in quanto il test di matematica è composto da più domande.

B2. CFU e altre variabili (TOLC S)

Nella Tabella 10 vengono mostrate le distribuzioni marginali percentuali dei CFU acquisiti dagli studenti al primo anno, rispetto alle classi di $CFU \leq 20$, $20 < CFU \leq 40$ e $CFU > 40$, e quelle condizionate al Corso di Studi, al genere, al voto al diploma (dicotomizzata per valori superiori o uguali al voto 90/100), e alla macroregione.

Tabella 10 Percentuali di riga del numero di studenti e in totale per fascia di cfu condizionatamente alle variabili di interesse.

Variabile		CFU (% di riga)			Totale
		≤ 20	$20 < 40$	> 40	
CdS	L-27	35,6	25,5	38,9	795
	L-28	25,9	33,3	40,7	27
	L-30	27,3	20,4	52,3	947
	L-31	37,1	25,1	37,8	1357
	L-32	41,8	24,3	33,9	510
	L-34	27,1	24,9	48,0	225
	L-35	34,6	24,5	40,9	788
	L-43	20,0	24,4	55,6	45
	Altri CdS	23,2	26,9	49,9	2305
Genere	F	27,7	26,6	45,7	3056
	M	32,9	23,7	43,4	3943
Voto diploma	<90	38,4	26,9	34,7	4734
	≥ 90	13,7	21,4	64,9	2187
Tipo diploma	IT	38,7	24,6	36,7	1263
	LC	31,8	28,5	39,8	840
	LS	22,2	24,7	53,1	3795
	Altro	49,3	23,8	26,9	1101
Macro-regione	Nord-ovest	28,6	22,9	48,6	2856
	Nord-est	23,8	21,2	55,1	799
	Centro	32,5	27,3	40,2	2033
	Sud	34,2	28,9	36,9	661
	Isole	37,7	28,5	33,8	634
Totale		30,6	25,0	44,4	6999

Gli elementi di rilievo sono:

- i CdS in L-28 ed L-43 non vengono inclusi a causa dell'esiguo numero di osservazioni disponibili;
- il CdS in cui sembrano esserci le maggiori difficoltà nell'acquisizione dei CFU è L-32, per cui quasi il 42% degli studenti non supera la soglia dei 20 CFU alla fine del primo anno;
- il L-30 è il CdS con la percentuale più alta nella fascia CFU>40, con più della metà degli studenti (52%);
- performance migliori delle femmine nella classe CFU≤ 20;
- il voto al diploma è associato ai CFU acquisiti al primo anno; infatti la percentuale di coloro con voto superiore a 90 che supera la soglia CFU>40 è circa il 65%;
- migliori risultati al Nord.

Nelle due tabelle che seguono vengono riportate, per Corso di Studi, le percentuali degli studenti immatricolati che acquisiscono rispettivamente, più di 20 (Tabella 11) o di 40 (Tabella 12) CFU, condizionatamente alle classi di punteggio al test separate dai quartili (Classi).

Osserviamo, per il CdS in L-27, che tra coloro che hanno preso un punteggio Totale che ricade nell'ultimo quarto, ovvero tra 27,9 (terzo quartile) e 45,2 (punteggio massimo osservato), la probabilità di superare la soglia di 20 CFU al primo anno è di poco maggiore dell'85%. In termini di *odd* (quota), calcolato come $0,85/(1-0,85) = 5,7$, possiamo dire che, per coloro il cui punteggio ricade in questa classe, la probabilità di superare 20 CFU è 5,7 volte più alta di quella di non superarla. È anche possibile confrontare tale valore con quello ottenuto nella prima classe, facendone il rapporto. Otteniamo così l'*odds ratio* (OR) riportato nella colonna accanto, che è pari a 9,3. Esso esprime di quante volte l'*odds* di superare 20 di CFU, per uno studente il cui punteggio totale ricade nell'ultimo quarto, è superiore a quello di uno studente il cui punteggio totale ricade nel primo quarto. Se si osserva un gradiente positivo in termini di percentuali o di OR, passando da una classe di punteggio a quella successiva, allora è verosimile che la sezione corrispondente sarà predittiva per il superamento della soglia, nel CdS considerato. Poiché le sezioni hanno un numero di domande differente, e quindi punteggi differenti, non è agevole confrontare quale sezione è più predittiva condizionandosi ai valori assoluti, ovvero ad un CdS. Conviene invece leggere la Tabella 11 per riga, ovvero ci si condiziona alla sezione considerata e la si confronta tra i vari CdS. Possiamo così notare che nei CdS L-30 ed L-35 si hanno OR molto più alti rispetto agli altri CdS, e questo è vero per ogni sezione considerata, mentre L-32 ha gli OR più bassi. Da notare che si riesce ad osservare gradienti crescenti degli OR in tutte le sezioni e tutti i CdS, eccetto che per la sezione di *Ragionamento e Problemi* nel CdS L-32, ed in Scienze di base per il CdS L-31. Fenomeni di questo tipo si ritrovano anche nella Tabella 12, dove si considera il superamento della soglia di 40 CFU, guardando, ad esempio, alla sezione *Comprensione testo – sezione testo* nei CdS L-31 ed L-34. Comunque, la Tabella 12 conferma una certa predittività delle sezioni per tutti i CdS considerati, con OR molto alti per L-31 ed L-35, mentre si hanno OR più bassi per L-32 ed L-34.

Tabella 11. Percentuale (%) degli studenti immatricolati che prendono più di 20 CFU, condizionata alle classi di punteggio al test in Totale e nelle varie Sezioni suddivise in quartili (Classi), odds ratio (OR) di ciascuna classe rispetto alla prima classe, per CdS. Le tabelle per L-28 ed L-43 non sono state riportate per numero di osservazioni insufficiente.

≥20 CFU	L-27 (n=786)			L-30 (n=922)			L-31 (n=1331)			L-32 (n=509)			L-34 (n=219)			L-35 (n=761)		
Sezioni	Classi	%	OR	Classi	%	OR	Classi	%	OR	Classi	%	OR	Classi	%	OR	Classi	%	OR
Intero TEST	min - Q1	38,7	1	min - Q1	41,3	1	min - Q1	40,5	1	min - Q1	46,5	1	min - Q1	62,1	1	min - Q1	30,8	1
	Q1 - Q2	56,4	2,2	Q1 - Q2	73,8	4	Q1 - Q2	58,4	2,1	Q1 - Q2	52,8	1,3	Q1 - Q2	60	0,9	Q1 - Q2	60,1	3,4
	Q2 - Q3	77,7	5,5	Q2 - Q3	84,3	7,6	Q2 - Q3	69,3	3,3	Q2 - Q3	65,4	2,2	Q2 - Q3	84,2	3,3	Q2 - Q3	81	9,6
	Q3- max	85,4	9,3	Q3- max	94,1	22,7	Q3- max	84,5	8	Q3- max	72,4	3	Q3- max	87,3	4,2	Q3- max	91,1	23
Comp. Testo totale	min - Q1	49,3	1	min - Q1	50,8	1	min - Q1	50	1	min - Q1	47,9	1	min - Q1	61,3	1	min - Q1	42,2	1
	Q1 - Q2	59,6	1,5	Q1 - Q2	70,2	2,3	Q1 - Q2	61,7	1,6	Q1 - Q2	59,2	1,6	Q1 - Q2	70,6	1,5	Q1 - Q2	59,3	2
	Q2 - Q3	70,5	2,5	Q2 - Q3	86	5,9	Q2 - Q3	63,8	1,8	Q2 - Q3	62,1	1,8	Q2 - Q3	77,4	2,2	Q2 - Q3	76,8	4,5
	Q3- max	79,9	4,1	Q3- max	92,7	12,3	Q3- max	79,8	4	Q3- max	69,9	2,5	Q3- max	86	3,9	Q3- max	89,3	11,4
Comp. Testo: testo	min - Q1	53,1	1	min - Q1	56	1	min - Q1	54	1	min - Q1	45,9	1	min - Q1	66,1	1	min - Q1	49,5	1
	Q1 - Q2	64,1	1,6	Q1 - Q2	74	2,2	Q1 - Q2	65,4	1,6	Q1 - Q2	59,8	1,8	Q1 - Q2	65,2	1	Q1 - Q2	66,9	2,1
	Q2 - Q3	68,6	1,9	Q2 - Q3	82,5	3,7	Q2 - Q3	64,1	1,5	Q2 - Q3	63,2	2	Q2 - Q3	83,6	2,6	Q2 - Q3	76	3,2
	Q3- max	73,4	2,4	Q3- max	NA	NA	Q3- max	70	2	Q3- max	71,2	2,9	Q3- max	82,2	2,4	Q3- max	NA	NA
Comp. Testo: domande	min - Q1	49,8	1	min - Q1	49,8	1	min - Q1	50,6	1	min - Q1	47,6	1	min - Q1	75	1	min - Q1	43,6	1
	Q1 - Q2	61,3	1,6	Q1 - Q2	72,4	2,6	Q1 - Q2	58,7	1,4	Q1 - Q2	61,2	1,8	Q1 - Q2	61,8	0,5	Q1 - Q2	56,6	1,7
	Q2 - Q3	68,6	2,2	Q2 - Q3	83,8	5,2	Q2 - Q3	67,9	2,1	Q2 - Q3	62,2	2	Q2 - Q3	73,6	0,9	Q2 - Q3	75,5	4
	Q3- max	81,3	4,4	Q3- max	92,5	12,4	Q3- max	80,2	4	Q3- max	68,9	2,9	Q3- max	90	3	Q3- max	89,1	10,6
Ragionamento e Problemi	min - Q1	44,8	1	min - Q1	52,3	1	min - Q1	46,9	1	min - Q1	46,4	1	min - Q1	63,3	1	min - Q1	44,2	1
	Q1 - Q2	64	2,2	Q1 - Q2	74,4	2,7	Q1 - Q2	63,5	2	Q1 - Q2	65,7	2,2	Q1 - Q2	77,1	2	Q1 - Q2	60,9	2
	Q2 - Q3	69,6	2,8	Q2 - Q3	82,2	4,2	Q2 - Q3	67,9	2,4	Q2 - Q3	64,6	2,1	Q2 - Q3	75	1,7	Q2 - Q3	75	3,8
	Q3- max	80,3	5	Q3- max	87,1	6,2	Q3- max	76,4	3,7	Q3- max	61,7	1,9	Q3- max	78,4	2,1	Q3- max	86,4	8
Scienze di base	min - Q1	51,7	1	min - Q1	58,5	1	min - Q1	50,6	1	min - Q1	46,8	1	min - Q1	68,3	1	min - Q1	40,8	1
	Q1 - Q2	57,3	1,3	Q1 - Q2	63,9	1,3	Q1 - Q2	66,2	1,9	Q1 - Q2	59,5	1,7	Q1 - Q2	71,7	1,2	Q1 - Q2	61	2,3
	Q2 - Q3	73,7	2,6	Q2 - Q3	82,9	3,4	Q2 - Q3	61,3	1,5	Q2 - Q3	60,7	1,8	Q2 - Q3	73,2	1,3	Q2 - Q3	74	4,1
	Q3- max	80,1	3,8	Q3- max	91,2	7,4	Q3- max	77,4	3,3	Q3- max	71,4	2,8	Q3- max	81,1	2	Q3- max	89	11,7
Matematica	min - Q1	37,1	1	min - Q1	41,2	1	min - Q1	40,2	1	min - Q1	49,3	1	min - Q1	59,7	1	min - Q1	29,8	1
	Q1 - Q2	59,3	2,5	Q1 - Q2	73,2	3,9	Q1 - Q2	58,6	2,1	Q1 - Q2	57	1,4	Q1 - Q2	68,6	1,5	Q1 - Q2	64	4,2
	Q2 - Q3	75,7	5,3	Q2 - Q3	86,1	8,8	Q2 - Q3	73,2	4,1	Q2 - Q3	58,7	1,5	Q2 - Q3	75	2	Q2 - Q3	83,7	12,1
	Q3- max	87,4	11,8	Q3- max	93,2	19,6	Q3- max	82,1	6,8	Q3- max	73,4	2,8	Q3- max	91,1	6,9	Q3- max	85,9	14,4

Tabella 12. Percentuale (%) degli studenti immatricolati che prendono più di 40 CFU, condizionata alle classi di punteggio al test in Totale e nelle varie Sezioni suddivise in quartili (Classi), odds ratio (OR) di ciascuna classe rispetto alla prima classe, per CdS. Le tabelle per L-28 ed L-43 non sono state riportate per numero di osservazioni insufficiente.

≥ 40 CFU	L-27 (n=786)			L-30 (n=922)			L-31 (n=1331)			L-32 (n=509)			L-34 (n=219)			L-35 (n=761)		
Sezioni	Classi	%	OR	Classi	%	OR	Classi	%	OR	Classi	%	OR	Classi	%	OR	Classi	%	OR
Intero Test	min – Q1	14,6	1	min – Q1	20,2	1	min – Q1	14,1	1	min – Q1	16,3	1	min – Q1	29,3	1	min – Q1	9,6	1
	Q1 – Q2	23,5	1,8	Q1 – Q2	42,1	2,9	Q1 – Q2	30,2	2,6	Q1 – Q2	26	1,8	Q1 – Q2	32,7	1,2	Q1 – Q2	29,3	3,9
	Q2 – Q3	50,3	5,9	Q2 – Q3	67,4	8,2	Q2 – Q3	45,5	5,1	Q2 – Q3	40,2	3,5	Q2 – Q3	63,2	4,1	Q2 – Q3	50,5	9,6
	Q3- max	69,3	13,2	Q3- max	83,9	20,6	Q3- max	63,4	10,6	Q3- max	54,3	6,1	Q3- max	70,9	5,9	Q3- max	76,6	30,8
Comp. Testo totale	min – Q1	27,3	1	min – Q1	28,4	1	min – Q1	25,8	1	min – Q1	25,7	1	min – Q1	33,9	1	min – Q1	18,1	1
	Q1 – Q2	26,3	1	Q1 – Q2	45	2,1	Q1 – Q2	33,6	1,5	Q1 – Q2	28	1,1	Q1 – Q2	43,1	1,5	Q1 – Q2	29,9	1,9
	Q2 – Q3	43,7	2,1	Q2 – Q3	68,2	5,4	Q2 – Q3	37,3	1,7	Q2 – Q3	38,6	1,8	Q2 – Q3	58,1	2,7	Q2 – Q3	53,5	5,2
	Q3- max	61,9	4,3	Q3- max	81,8	11,3	Q3- max	59,6	4,2	Q3- max	46	2,5	Q3- max	62	3,2	Q3- max	70,2	10,7
Comp. Testo: testo	min – Q1	30,4	1	min – Q1	35,3	1	min – Q1	29,8	1	min – Q1	25,3	1	min – Q1	42,4	1	min – Q1	25,5	1
	Q1 – Q2	33,3	1,1	Q1 – Q2	52,2	2	Q1 – Q2	34,9	1,3	Q1 – Q2	31,6	1,4	Q1 – Q2	37,9	0,8	Q1 – Q2	40,1	2
	Q2 – Q3	44,6	1,8	Q2 – Q3	64,3	3,3	Q2 – Q3	40,2	1,6	Q2 – Q3	39,7	1,9	Q2 – Q3	56,4	1,8	Q2 – Q3	53,8	3,4
	Q3- max	50	2,3	Q3- max	NA	NA	Q3- max	48,8	2,2	Q3- max	41,4	2,1	Q3- max	64,4	2,5	Q3- max	NA	NA
Comp. Testo: domande	min – Q1	24,9	1	min – Q1	28,5	1	min – Q1	27	1	min – Q1	24,8	1	min – Q1	46,9	1	min – Q1	21,6	1
	Q1 – Q2	32	1,4	Q1 – Q2	50	2,5	Q1 – Q2	31,7	1,3	Q1 – Q2	32,8	1,5	Q1 – Q2	38,2	0,7	Q1 – Q2	24,7	1,2
	Q2 – Q3	41,8	2,2	Q2 – Q3	58,7	3,6	Q2 – Q3	40,2	1,8	Q2 – Q3	34,3	1,6	Q2 – Q3	47,2	1	Q2 – Q3	50,5	3,7
	Q3- max	61,5	4,8	Q3- max	83,9	13,1	Q3- max	60,5	4,1	Q3- max	48,1	2,8	Q3- max	72,5	3	Q3- max	69,8	8,4
Ragionamento e Probabilità	min – Q1	22,7	1	min – Q1	30,8	1	min – Q1	23,9	1	min – Q1	18,6	1	min – Q1	26,7	1	min – Q1	21,2	1
	Q1 – Q2	30,5	1,5	Q1 – Q2	47,9	2,1	Q1 – Q2	33,7	1,6	Q1 – Q2	35	2,4	Q1 – Q2	55,7	3,5	Q1 – Q2	31,6	1,7
	Q2 – Q3	43,6	2,6	Q2 – Q3	61,7	3,6	Q2 – Q3	42,5	2,4	Q2 – Q3	40,7	3	Q2 – Q3	63,6	4,8	Q2 – Q3	48,9	3,6
	Q3- max	61,7	5,5	Q3- max	75,4	6,9	Q3- max	55,3	3,9	Q3- max	45	3,6	Q3- max	52,9	3,1	Q3- max	68,4	8
Scienze di base	min – Q1	21,1	1	min – Q1	35,7	1	min – Q1	26,4	1	min – Q1	20,6	1	min – Q1	41,3	1	min – Q1	13,4	1
	Q1 – Q2	32,6	1,8	Q1 – Q2	43,4	1,4	Q1 – Q2	37,2	1,7	Q1 – Q2	26,7	1,4	Q1 – Q2	47,2	1,3	Q1 – Q2	34,9	3,5
	Q2 – Q3	50,3	3,8	Q2 – Q3	60,8	2,8	Q2 – Q3	37,5	1,7	Q2 – Q3	43,8	3	Q2 – Q3	50	1,4	Q2 – Q3	49,3	6,3
	Q3- max	59,1	5,4	Q3- max	79,5	7	Q3- max	53,7	3,2	Q3- max	48,4	3,6	Q3- max	58,5	2	Q3- max	70,5	15,4
Matematica	min – Q1	14,8	1	min – Q1	18,9	1	min – Q1	13,1	1	min – Q1	17,8	1	min – Q1	27,4	1	min – Q1	11,6	1
	Q1 – Q2	23,3	1,7	Q1 – Q2	43,3	3,3	Q1 – Q2	29,5	2,8	Q1 – Q2	27,2	1,7	Q1 – Q2	39,2	1,7	Q1 – Q2	30,5	3,3
	Q2 – Q3	48,1	5,3	Q2 – Q3	66,4	8,5	Q2 – Q3	51	6,9	Q2 – Q3	40,5	3,1	Q2 – Q3	53,6	3,1	Q2 – Q3	57,8	10,4
	Q3- max	72,6	15,3	Q3- max	85,1	24,5	Q3- max	61,6	10,6	Q3- max	53,2	5,2	Q3- max	76,8	8,8	Q3- max	68,3	16,4

B3. Applicazione del modello logit per la predittività

Come per il TOLC B, per misurare la capacità del TOLC S a predire il successo universitario sono stati stimati due modelli logit: nel primo la variabile risposta è il superamento o meno della soglia dei 20 CFU, nel secondo la soglia è di 40 CFU. In Tabella 12 vengono riportate le stime dei parametri dei due modelli.

Tabella 12. Stime dei Log-OR dei modelli logistici con soglia 20 e 40 cfu. In grassetto vengono riportati i p-valori significativi.

Predittori	CFU 20			CFU 40		
	Log-OR	std, Error	p	Log-OR	std, Error	p
Intercetta	0,747	0,21	< 0,001	-0,308	0,214	0,151
Genere: Femmina	0,115	0,089	0,194	-0,106	0,087	0,227
Diploma: IT	0,243	0,132	0,067	0,231	0,154	0,135
Diploma: LC	0,636	0,168	< 0,001	0,478	0,185	0,01
Diploma: LS	0,746	0,132	< 0,001	0,724	0,15	< 0,001
Voto di diploma	0,798	0,047	< 0,001	0,919	0,048	< 0,001
Macro regione: Centro	0,037	0,104	0,718	0,057	0,107	0,59
Macro regione: Nord-est	0,09	0,143	0,531	0,438	0,139	0,002
Macro regione: Nord-ovest	-0,274	0,114	0,016	0,159	0,114	0,163
CdS: L-27	-0,197	0,178	0,268	-0,596	0,165	< 0,001
CdS: L-30	-1,017	0,177	< 0,001	-1,425	0,181	< 0,001
CdS: L-31	0,088	0,169	0,603	-0,322	0,162	0,046
CdS: L-34	0,802	0,284	0,005	0,479	0,239	0,045
CdS: L-35	-1,634	0,193	< 0,001	-2,278	0,222	< 0,001
Comp. del testo – sez. testo	0,047	0,04	0,236	0,041	0,041	0,32
Comp. del testo – sez. domande	0,144	0,051	0,005	0,091	0,049	0,066
Ragionamento e Problemi	0,174	0,052	0,001	0,166	0,05	0,001
Scienze di base	0,191	0,049	< 0,001	0,202	0,047	< 0,001
Matematica	0,058	0,126	0,643	0,281	0,132	0,033
Matematica*CdS: L-27	0,593	0,16	< 0,001	0,413	0,168	0,014
Matematica*CdS: L-30	0,6	0,165	< 0,001	0,407	0,173	0,019
Matematica*CdS: L-31	0,376	0,146	0,01	0,298	0,155	0,055
Matematica*CdS: L-34	0,261	0,245	0,286	0,067	0,23	0,772
Matematica*CdS: L-35	0,706	0,178	< 0,001	0,554	0,201	0,006
Osservazioni	4567					

Studente baseline: maschio, del sud/isole, Altro diploma, che ha punteggi medi al test, un voto medio al diploma, ed è immatricolato in L-32.

La stima dell'intercetta racchiude in sé l'informazione relativa allo studente baseline, ovvero uno studente tipo, definito dal seguente profilo medio: maschio, del sud/isole, diplomato in Altro liceo, che ha punteggi medi al test, un voto medio al diploma (ovvero 82/100), e immatricolato in L-32. La stima dell'intercetta, come già detto, è tuttavia un'eccezione perché, a differenza di tutte le altre variabili, non è misurata in termini di log-OR, ma in termini di log-Odds.

I risultati più rilevanti sono:

- le femmine hanno una maggiore probabilità di superare la soglia dei 20 CFU rispetto ai maschi del 12%. Tale valore non è però statisticamente significativo anche nel modello con soglia 40 CFU.
- le sezioni Comp. del testo - sottosezione Ragionamento e Problemi, e Scienze di base risultano significativi con 20 e con 40 CFU;
- l'effetto base della sezione di Matematica è significativo nel modello con soglia 40 CFU;
- le interazioni del punteggio di Matematica con i CdS sono tutte significative, ad eccezione di L-34 in entrambi i modelli, e di L-31 nel modello a soglia 40 CFU. Per la spiegazione e

- l'interpretazione dell'interazione si rimanda al rapporto esteso o, per vedere il loro effetto, occorre vedere le probabilità predette di seguito riportate;
- il voto di diploma risulta essere significativamente predittivo in entrambi i modelli. Tenendo sempre a mente che la categoria baseline per il tipo di diploma è AL (altro diploma), osserviamo che il liceo classico (LC) e soprattutto il liceo scientifico (LS) hanno coefficienti positivi e significativi, specialmente LC;
 - marginali sono le differenze tra le macroregioni.

Concludendo, in entrambi i modelli, matematica è il più predittivo tra i test del TOLC S, ed ha un peso diverso a seconda del corso di studi scelto. Si è visto che il peso maggiore, neanche a dirlo, lo si ha ad L-35 (Scienze Matematiche), seguito da L-27 (Scienze e Tecnologie Chimiche) ed L-30 (Scienze e Tecnologie Fisiche), mentre il peso minore lo si ha in L-32 (Scienze e Tecnologie per l'Ambiente e la Natura) ed in L-34 (Scienze Geologiche). Anche le sezioni Scienze di Base e Ragionamento e Probabilità del TOLC S hanno coefficienti significativi, mentre la comprensione del testo risulta significativa soltanto nella sezione domande del modello a soglia 20 CFU. Aumentano significativamente la probabilità di superare le entrambe le soglie alti voti di diploma e l'aver frequentato il liceo classico o scientifico (in particolare quest'ultimo). Infine, il Nord-Est presenta performance migliori rispetto al Sud/Isole nel modello a 40 CFU.

Nella Tabella 13 vengono riportate le probabilità di acquisire rispettivamente più di 20 o 40 CFU per uno studente, basate sui due modelli di Tabella 12, per alcuni profili con punteggi al test e voto al diploma su scala originaria.

Tabella 13. Probabilità di acquisire più di 20 ($P>20$) o 40 ($P>40$) CFU, per alcuni differenti profili di studente, predette sulla base dei corrispettivi modelli logistici.

ID	GENERE	C. T. TESTO	C. T. DOM	RAG. PROB.	SCL. BASE	MAT	VOTO DIP.	DIPLOMA	MACRO REG	CDS	P>20 CFU	P>40 CFU
1	F	3	3	6	6	12	80	LS	Centro	L-35	0,6	0,18
2	F	3	3	6	6	12	80	LS	Centro	L-31	0,89	0,59
3	F	3	3	6	6	12	80	LS	Centro	L-34	0,94	0,75
4	F	3	3	6	6	12	80	LS	Centro	L-32	0,87	0,65
5	F	3	3	6	6	12	80	LS	Centro	L-30	0,73	0,33
6	F	3	3	6	6	12	80	LS	Centro	L-27	0,86	0,53

È allegata al presente documento il file Excel "prob_predette_tolc_s.xlsx" contenente una versione più estesa delle Tabelle 7 e 13, con una più ampia gamma di profili di studente

Appendici

Appendice A - Matrice origine destinazione

Appendice A – Matrice origine-destinazione 1°-2° anno per corso di laurea in Italia, coorte 2018/19.

CdS 1°anno	CdS 2°anno															STAYER	MOVER	ABB	TOT	
	ALTRO	BIO	BIOT	CHI	FIS	GEO	INFO	MAT	SC.AMB	STAT	SAN	MED	FAR	VET	ABB	Med	Altro			
ALTRO	185806	190	130	73	67	36	309	80	110	46	872	344	252	108	31544	84,5	0,4	0,8	14,3	219967
BIO	565	4865	64	11	7	5	6	2	21	1	580	810	64	46	1031	60,2	7,2	19,8	12,8	8078
BIOT	462	67	3792	18	5	0	5	6	13	2	491	1078	53	32	675	56,6	7,3	26,0	10,1	6699
CHI	300	46	38	2629	7	6	7	5	17	5	129	91	59	6	634	66,1	3,2	14,8	15,9	3979
FIS	212	19	9	12	2516	9	20	17	13	2	19	13	4	0	406	76,9	0,6	10,1	12,4	3271
GEO	48	15	8	4	4	623	1	0	10	0	17	2	6	0	193	66,9	1,8	10,5	20,7	931
INFO	272	5	0	0	5	1	5166	11	6	5	13	4	1	0	1486	74,1	0,2	4,4	21,3	6975
MAT	314	13	6	1	9	1	34	2218	6	23	24	11	0	0	442	71,5	0,8	13,5	14,2	3102
SC.AMB	250	114	33	14	8	4	1	2	1655	0	114	30	30	12	554	58,7	4,0	17,7	19,6	2821
STAT	73	1	1	0	1	0	1	0	1	927	2	1	0	0	131	81,4	0,2	6,9	11,5	1139
SAN	417	43	25	7	2	3	4	4	6	0	14506	357	39	10	1330	86,6	2,1	3,3	7,9	16753
MED	17	2	0	2	2	0	0	3	0	1	6	6052	0	0	162	96,9	-	0,5	2,6	6247
FAR	289	29	38	7	2	4	1	3	3	0	280	732	4536	23	472	70,7	4,4	17,6	7,4	6419
VET	5	0	0	0	0	0	0	0	0	0	1	23	0	393	9	91,2	0,2	6,5	2,1	431
TOT	189030	5409	4144	2778	2635	692	5555	2351	1861	1012	17054	9548	5044	630	39069	75,1	5,9	5,3	13,6	286812

Altro= tutti corsi di laurea non-stem e non di ambito medico-veterinario

Stayer = studenti iscritti al 2° anno nel medesimo CdS

Mover = studenti iscritti al 2° anno in un altro CdS, distinti Medicina (Med) e altro CdS

Abb = studenti che non hanno formalizzato l'iscrizione al 2° anno o che non hanno sostenuto esami positivamente al 2° anno

Appendice B. Odds e odds ratio

Una misura alternativa alla probabilità

Il rapporto tra la probabilità p di un evento A e la probabilità $(1-p)$ dell'evento complementare $\text{Non}(A)$ prende il nome di *Odds* (o quota)

$$\text{Odds} = p(A) / (1-p(A)) = p(A) / p(\text{Non}(A))$$

Se $p=0,1$ è la probabilità allora l'*odds* è pari a $0,1/0,9=0,111$. Quindi in questo caso la probabilità $(0,1)$ e l'*odds* $(0,111)$ sono abbastanza simili. Infatti, ogni volta che p è piccolo, la probabilità e l'*odds* saranno simili. Ma quando p non è piccolo, la probabilità e l'*odds* saranno generalmente molto diverse. Per esempio, se $p=0,5$, abbiamo l'*odds* pari a $0,5/0,5=1$. All'aumentare di p , gli *odds* diventano sempre più grandi. Ad esempio, con $p=0,99$, l'*odds* è $0,99/0,01=99$.

Nel linguaggio delle scommesse “5 a 1” (indicato come 5-1 *odds*) significa che su 6 possibili risultati, 5 sono di un tipo (A) e 1 è di un secondo tipo ($\text{Non}A$). L' *odds* in questo caso è 5 e significa che su 6 scommesse la tua scelta vincerà 5 volte e perde 1 volta, se scegli A ; vincerà 1 volta e perde 5 volte se scegli $\text{Non}A$.

NEL NOSTRO CASO

L'*Odds* è una misura di chance (o di rischio, a seconda dell'accezione che vi si vuole dare) che esprime quante volte è più probabile avere un successo che un insuccesso, nel nostro caso l'evento di interesse (successo) A è l'acquisizione di almeno 20 CFU (o 40).

$$\text{Odds} = \frac{\text{Pr}(A = \text{almeno } 20 \text{ CFU} / \text{ Liceo Scientifico})}{1 - \text{Pr}(A = \text{almeno } 20 \text{ CFU} / \text{ Liceo Scientifico})} = \frac{0,7}{1 - 0,7} = 2,333$$

L'*Odds ratio* mette in relazione gli *odds* di due gruppi di interesse. Se A è l'evento “successo” (nel nostro caso è avere > 20 CFU o > 40 CFU), l'OR esprime quante volte l'*odds* del successo per uno studente dello Scientifico è maggiore dell'*odds* di successo per uno studente del Classico.

$$\text{Odds ratio (OR)} = \frac{\text{odds}(\text{Pr}(A = \text{almeno } 20 \text{ CFU} / \text{ Liceo Scientifico}))}{\text{odds}(\text{Pr}(A = \text{almeno } 20 \text{ CFU} / \text{ Liceo Classico}))} = \frac{\frac{0,7}{1 - 0,7}}{\frac{0,6}{1 - 0,6}} = 1,55$$

Oppure, parimenti, in relazione ai punteggi ottenuti al TOLC:

$$\text{OR} = \frac{\text{odds}(\text{Pr}(A = \text{almeno } 20 \text{ CFU} / \text{ punteggio totale TOLC nel } 4^{\circ} \text{ Q}))}{\text{odds}(\text{Pr}(A = \text{almeno } 20 \text{ CFU} / \text{ punteggio totale TOLC nel } 1^{\circ} \text{ Q}))} = \frac{\frac{0,9}{1 - 0,9}}{\frac{0,6}{1 - 0,6}} = 3,85$$

Appendice C. Il modello logit

Al fine di quantificare la predittività del TOLC B/ TIP B e del TOLC S rispetto al successo universitario, misurato dall'acquisizione dei CFU al primo anno, sono stati stimati due modelli logit: nel primo la variabile risposta è il superamento o meno della soglia dei 20 CFU, nel secondo la soglia è di 40 CFU. Le variabili esplicative sono le caratteristiche individuali degli studenti e il punteggio al test nelle varie sezioni del test. I modelli logit applicati sono due. Il primo definisce la variabile risposta binaria attraverso la seguente relazione:

$$D_{CFU}^{20} = \{1, \text{se il numero di CFU A.S. 19 è } \geq 20; 0, \text{altrimenti}\} \quad [B1]$$

e il secondo attraverso la relazione:

$$D_{CFU}^{40} = \{1, \text{se il numero di CFU A.S. 19 è } \geq 40; 0, \text{altrimenti}\} \quad [B2]$$

Abbiamo scelto come valori soglie 20 e 40, in quanto questi valori sono stati adottati da ANVUR come soglie di riferimento di alcuni indicatori sulla didattica nella Scheda di Monitoraggio Annuale (SMA) per l'attività di autovalutazione dei Corsi di Studio (<https://www.anvur.it/news/ava-indicatori-per-il-monitoraggio-annuale-dei-cds/>).

Partendo dalle [B1] e [B2], si ottiene il modello *logit* nel seguente modo:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + x_{1i} \beta_1 + x_{2i} \beta_2 + \dots + x_{ki} \beta_k$$

dove:

- p_i , la probabilità che lo studente i -esimo ha di acquisire almeno 20 (o almeno 40) CFU nell'A.A 2019;
- β_0 è l'intercetta, ed è il parametro corrispondente al profilo di studente detto baseline;
- $x'_i = [x_{1i}, x_{2i}, \dots, x_{ki}]$, il vettore delle k variabili esplicate osservate relative allo studente i -esimo;
- $\beta' = [\beta_1, \beta_2, \dots, \beta_k]$, il vettore dei coefficienti che verranno stimati dal modello.

Il rapporto tra la probabilità p_i e il suo complementare $(1-p_i)$ prende il nome di *Odd*, una misura di chance (o di rischio, a seconda dell'accezione che vi si vuole dare) che esprime quante volte è più probabile avere un successo che un insuccesso, nel nostro caso l'acquisizione di almeno 20 CFU (o 40). Si definisce

$$\text{Odd ratio (OR)} = \frac{\text{odd}(\text{Pr}(D = 1/x_i = 1))}{\text{odd}(\text{Pr}(D = 1/x_i = 0))}$$

il rapporto tra due odd che differiscono per i valori assunti dalle x_i . Ad esempio, se D è l'evento "successo" (nel nostro caso è avere ≥ 20 CFU o ≥ 40 CFU), l'OR esprime quante volte l'odd del successo per uno studente con $x_i = 1$ è maggiore dell'odd di successo per uno studente con $x_i = 0$. In particolare, la relazione che lega il coefficiente β_k della k -esima variabile (che assumiamo dicotomica per semplicità) e l'OR di avere, ad esempio, 20 CFU, è esplicitata come segue:

$$\frac{P(D_{CFU}^{20} = 1 | x_k = 1)}{1 - P(D_{CFU}^{20} = 1 | x_k = 1)} / \frac{P(D_{CFU}^{20} = 1 | x_k = 0)}{1 - P(D_{CFU}^{20} = 1 | x_k = 0)} = \exp(\beta_k).$$

Per trovare la probabilità di interesse, si applicherà la seguente formula:

$$p_i = \frac{\exp(\beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \dots + x_{ki}\beta_k)}{1 + \exp(\beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + \dots + x_{ki}\beta_k)} \quad [B3]$$

Per trovare la probabilità predetta di superare 20 (o 40) CFU occorre specificare il profilo di studente di interesse, ovvero sostituire nella [B3] i beta stimati e le x_i del profilo dello studente. Ad esempio, se si è interessati allo studente baseline, la probabilità predetta sarà funzione soltanto dell'intercetta, poiché per definizione tutte le altre variabili saranno poste a zero:

$$p_i = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$$

Probabilità predette per profilo di studente TOLC B

Nei modelli applicati alle classi L-02 e L-03, sia i punteggi al test che il voto di diploma sono stati standardizzati, ovvero per una generica variabile numerica x si è usata la trasformazione lineare: $z = (x - \text{Media}) / \text{DS}$, dove DS indica la deviazione standard campionaria. La finalità di tale trasformazione è quella di poter confrontare gli effetti dei punteggi al test, e del voto di diploma, indipendentemente dall'unità di misura delle variabili. Per riottenere il punteggio x sulla scala originaria, partendo da quello standardizzato z , si può ricorrere alla formula $x = z * \text{DS} + \text{Media}$. I valori stimati per le classi L-02 e L-13 di tali medie e deviazioni standard sono riportati nella Tabella seguente.

Media e Deviazione standard del voto di diploma e dei punteggi al test, su scala originale.

Variabile	Non condizionata	
	DS	Media
Voto di diploma	11,17	83,85
Biologia+Chimica	3,83	9,41
Fisica	2,18	2,56
Matematica	5,32	9,34

Pertanto, se la variabile risposta è la soglia di 40 CFU, confrontiamo i due coefficienti beta di due sezioni standardizzate, per esempio quelli di Biologia+Chimica e Matematica, osserviamo che quest'ultimo ha un coefficiente beta maggiore, indicando un maggiore peso della matematica sulla probabilità di acquisizione dei CFU.

Nel nostro caso, sempre riguardo al modello con soglia 40 CFU, poiché i punteggi al test ed il voto al diploma sono stati standardizzati, il profilo baseline corrisponderà ad un *Maschio*, del *Sud/Isole*, diplomato in *Altro*, immatricolato in L-13, che ha fatto il TIP B, con punteggi medi ai test e voto medio al diploma. Per questo studente, la probabilità predette di acquisire più di 40 CFU sarà uguale a

$$p_i = \frac{\exp(-2,037)}{1 + \exp(-2,037)} = 0,115,$$

mentre per trovare la corrispondente probabilità predetta per uno studente con un profilo simile, che differisce dal fatto di aver preso un voto pari a 100 al diploma, occorrerà dapprima standardizzare il voto: $(100 - 83,85) / 11,17 = 1,45$, e calcolare:

$$p_i = \frac{\exp(-2,037 + 1,45 * 0,578)}{1 + \exp(-2,037 + 1,45 * 0,578)} = 0,231,$$

dove 0,578 è il beta stimato per il voto di diploma nel modello in Tabella 3.8. Se per lo stesso studente, invece siamo interessati al fatto che proviene dal liceo scientifico:

$$p_i = \frac{\exp(-2,037 + 1,45 * 0,578 + 0,351)}{1 + \exp(-2,037 + 1,45 * 0,578 + 0,351)} = 0,30.$$

Probabilità predette per profilo di studente TOLC S

Nei modelli utilizzati, sia i punteggi al test che il voto di diploma sono stati standardizzati, ovvero per una generica variabile numerica x si è usata la trasformazione lineare: $z=(x-Media)/DS$, dove DS indica la deviazione standard campionaria. La finalità di tale trasformazione è quella di poter confrontare gli effetti dei punteggi al test, e del voto di diploma, indipendentemente dall'unità di misura delle variabili, e dal numero di domande contenute nei test. Per riottenere il punteggio x sulla scala originaria, partendo da quello standardizzato z , si può ricorrere alla formula $x=z*DS+Media$. I valori stimati di tali medie e deviazioni standard sono riportati nella seguente tabella.

Media e Deviazione standard del voto di diploma standard e dei punteggi al test, su scala originale.

Variabile	Media	DS
Voto di diploma	82,34	12,57
Comprensione testo testo	3,05	1,64
Comprensione testo domande	2,05	1,65
Ragionamento e prob.	3,78	2,63
Scienze di base	3,46	2,73
Matematica	10,82	5,78

Pertanto, se nel modello in Tabella 4.8 (soglia 40 CFU), confrontiamo i due coefficienti beta di due test standardizzati, per esempio quelli di Comp. del testo - sez. testo e Scienze di base, osserviamo che quest'ultimo ha un coefficiente beta maggiore, indicando un maggiore peso di Scienze di base sulla probabilità di acquisizione dei CFU. Diverso è il caso del punteggio al test di Matematica, in quanto è presente nel modello un termine di interazione con il CdS. Questo significa che non si avrà un coefficiente beta unico per tutti i CdS come nel caso di Comp. del testo - sez. testo e Scienze di base, ma un coefficiente specifico per CdS. Per ottenere il coefficiente beta del punteggio al test di Matematica in L-35, occorre sommare il coefficiente del termine base e cioè 0,281 con il corrispettivo termine di interazione Matematica*L-35, e cioè 0,554. Tale somma 0,835, e rappresenta il coefficiente più alto fra i CdS, più alto del coefficiente di Scienze di base (0,202), ma un po' più basso di quello del voto al diploma (0,919).

A partire dai coefficienti stimati nel modello con soglia 40 CFU è possibile calcolare la probabilità predetta di acquisire più di 40 CFU per un profilo definito di studente, ovvero uno studente a cui assegniamo i valori delle variabili. Ad esempio, il profilo baseline corrisponderà ad uno studente maschio, con punteggi medio al test, e voto medio al diploma, del sud/isole, diplomato in Altro liceo e immatricolato in L-32. L'informazione relativa a tale studente è contenuta nell'intercetta, pertanto la probabilità predette di acquisire più di 40 CFU sarà uguale a

$$p_i = \frac{\exp(-0,31)}{1 + \exp(-0,31)} = 0,42$$

mentre, per trovare la corrispondente probabilità predetta per uno studente con un profilo simile, che differisce soltanto per l'aver preso un voto pari a 100 al diploma, occorrerà dapprima standardizzare il voto: $(100-82,34)/12,57= 1,40$, e calcolare:

$$p_i = \frac{\exp(-0,31 + 1,40 * 0,92)}{1 + \exp(-0,31 + 1,40 * 0,92)} = 0,73,$$

dove 0,92 è il beta stimato per il voto di diploma nel modello in Tabella 5.1. Se per lo stesso studente, siamo invece interessati alla probabilità di superare 40 CFU, sapendo che proviene dal liceo scientifico:

$$p_i = \frac{\exp(-0,31 + 1,40 * 0,92 + 0,72)}{1 + \exp(-0,31 + 1,40 * 0,92 + 0,72)} = 0,84.$$