

# INDICE

INTRODUZIONE .....	3
CAPITOLO PRIMO .....	5
MATERIALI UTILIZZATI.....	5
1.1.Considerazioni preliminari .....	5
1.2.Dati .....	7
1.3.Variabili .....	8
1.3.1.Formazione pre-universitaria.....	9
1.3.2.Carriera universitaria.....	10
1.4.Scopi e Strumenti .....	12
CAPITOLO SECONDO .....	13
DETERMINANTI DEGLI ESITI .....	13
2.1.Analisi descrittiva .....	13
2.1.1.Corso di Laurea.....	14
2.1.2.Scuola di Provenienza .....	15
2.1.3.Base Voto Diploma .....	17
2.1.4.Voto Diploma .....	18
2.1.5. Provincia Istituto Diploma.....	22
2.1.6.Percorso .....	26
2.1.7. Velocità.....	27
2.1.8.Area Relatore e Area Contro-Relatore .....	29
2.1.9.Voto Finale .....	33
2.2.Analisi attraverso Tabelle di Contingenza .....	37
CAPITOLO TERZO .....	51
DETERMINANTI DEGLI ESITI: STUDIO ATTRAVERSO ANALISI DI SOPRAVVIVENZA .....	51
3.2.1. La variabile risposta.....	55

3.2.2. La censura.....	56
3.3.1. Funzione di sopravvivenza .....	58
3.3.2. Funzione di rischio.....	58
3.4. Stimatori .....	59
3.4.1. La tavola di sopravvivenza.....	59
3.4.2. Metodo di Kaplan Meier .....	62
CAPITOLO QUARTO .....	67
DETERMINANTI DEGLI ESITI:UN APPROFONDIMENTO CON IL MODELLO DI COX.....	67
4.1.Il modello da usare .....	67
4.2. Il modello a rischi proporzionali .....	68
4.3. Stati di destinazione finale e insieme di covariate .....	70
4.4. Selezione delle Variabili.....	71
4.5. Modello di Cox applicato alla popolazione di studio.....	73
CAPITOLO QUINTO .....	81
CONSIDERAZIONI CONCLUSIVE .....	81

# *INTRODUZIONE*

---

Il lavoro presentato in questa relazione finale discute i risultati di analisi statistiche condotte sui dati relativi agli studenti che hanno conseguito la laurea di primo livello del Vecchio Ordinamento (ex DM 509/99), presso la Facoltà di Scienze Statistiche dell'Università degli Studi di Padova tra ottobre dell'anno 2003 e settembre del 2008. I dati di cui si dispone provengono dall'archivio della segreteria studenti dell'Ateneo di Padova.

Questa ricerca ha lo scopo di stabilire se vi siano delle relazioni tra il punteggio assegnato alla relazione finale, il corso di laurea frequentato e l'area scientifico-disciplinare di appartenenza del relatore.

Lo studio dei dati è sviluppato sostanzialmente in due parti: la prima è un'analisi di tipo descrittivo con la quale, attraverso tabelle di contingenza, vengono individuate le variabili caratterizzanti il background dei soggetti in esame e come queste influiscano o intervengano sulla carriera universitaria che si può riassumere nelle variabili che descrivono il voto medio ed il punteggio della relazione finale.

Le variabili esplicative considerate per esaminare il background dei soggetti sono molteplici e coinvolgono variabili legate al tipo di formazione della scuola media superiore di provenienza, alla zona di residenza, al corso di laurea scelto all'interno della facoltà ed alla media dei voti conseguita. Un limite, riscontrato nella parte descrittiva, dei dati a disposizione è la mancanza di variabili legate alla situazione socio-economica della famiglia di provenienza o a eventi intervenuti durante la carriera universitaria.

La seconda parte dello studio è stata sviluppata attraverso l'analisi di sopravvivenza ed il modello di regressione di Cox che ha permesso di prendere in esame anche quei soggetti che presentavano delle informazioni parziali per le ragioni che verranno successivamente esposte e spiegate.



# *CAPITOLO PRIMO*

---

## *MATERIALI UTILIZZATI*

### **1.1.Considerazioni preliminari**

L'arco temporale, nel quale i soggetti in esame si sono laureati, si estende da ottobre 2003 a settembre 2008. La scelta di considerare questo intervallo di tempo è dovuta alla rivoluzione del ciclo universitario avvenuta a partire dall' Anno Accademico 2001-2002.

Tale riforma (ex DM 509/99) stabiliva l'introduzione di un'offerta formativa articolata su una pluralità di livelli che prevedeva:

- un primo livello, di durata di tre anni, alla fine del quale si consegue la laurea triennale;
- un secondo livello, di durata biennale, alla fine del quale si consegue la laurea specialistica;
- un terzo livello, triennale, orientato alla ricerca, alla fine del quale si consegue il dottorato di ricerca.

Tale precisazione circa la riforma è utile in quanto, proprio a partire da settembre 2003 nella Facoltà di Scienze Statistiche di Padova, i primi studenti appartenenti a quello che è stato definito Vecchio Ordinamento (ex DM 509/99) hanno conseguito il diploma di laurea.

In questo modo il presente studio comprende tutti gli studenti che hanno conseguito la laurea di primo livello del Vecchio Ordinamento presso la Facoltà di Scienze Statistiche di Padova dall'entrata in vigore della riforma (ex DM 509/99) fino ad ottobre 2008, momento in cui è iniziata questa ricerca, indipendentemente

dalla data di immatricolazione. Pertanto, è opportuno precisare che il dataset, sul quale sono state condotte le analisi, non comprende la data di immatricolazione degli studenti laureati.

L'offerta didattica della Facoltà di Scienze Statistiche proponeva, all'epoca in cui era in vigore il DM 509/99, quattro corsi per la laurea triennale di primo livello. Tali corsi di laurea erano così denominati: SEF (Statistica Economia e Finanza), SGI (Statistica e Gestione delle Imprese), SPS (Statistica Popolazione e Società) e STI (Statistica e Tecnologie Informatiche).

È opportuno però precisare che il Corso di Laurea SPS è stato attivato nell'anno accademico 2002/03, quindi successivamente rispetto agli altri tre. Precedentemente infatti l'offerta didattica della facoltà di Scienze Statistiche comprendeva i tre corsi di laurea già citati, SEF, SGI e STI, e il Corso di laurea in Statistica e Organizzazioni Sociali (SORS). La Facoltà ha deciso di cambiare il nome di tale corso e di assegnarne uno che fosse più aderente ai contenuti effettivi del corso. Pertanto, il Corso di Laurea SPS ha sostituito il corso SORS, comprendendone tutti gli insegnamenti. Gli studenti che erano iscritti nel 2001/02 al SORS, con il cambio di nome del corso, potevano o chiedere di passare al nuovo corso in Statistica, Popolazione e Società o continuare il corso precedente iscrivendosi agli anni successivi fino al conseguimento della laurea in SORS. È necessario dunque precisare che i dati degli studenti che rientravano nella categoria del Corso di Laurea SORS sono stati accorpati nella categoria che caratterizza il Corso di Laurea SPS.

Il piano degli studi di ciascun corso di laurea comprendeva attività formative per un numero complessivo di 180 crediti, che potevano essere acquisiti come segue:

- 60 crediti relativi ad insegnamenti obbligatori comuni a tutti e quattro i corsi di laurea;
- 30 crediti relativi ad insegnamenti obbligatori specifici ad ogni corso di laurea;

- 54 crediti relativi ad insegnamenti del curriculum scelto all'interno del corso di laurea di appartenenza e ad altri insegnamenti a scelta dello studente coerenti con il progetto formativo;
- 36 crediti relativi ad altre attività formative (lingua straniera, insegnamenti a scelta, altre attività formative, stage e prova finale).

Il personale docente preso in considerazione appartiene a diversi Dipartimenti del cui apporto didattico la Facoltà si avvale.

Di conseguenza tali docenti sono stati suddivisi in sei diverse aree scientifico-disciplinari che sono:

- DSS (Socio Demografica);
- ECO (Economica);
- INF (Informatica);
- MAT (Matematica);
- STAT (Statistica);
- STATEC (Statistico Economica).

## **1.2.Dati**

Gli studenti che hanno conseguito il titolo nell'arco temporale esaminato sono 758, di cui 40 provenienti dall'estero. Per quanto esiguo sia il numero degli studenti stranieri, esso va sottolineato perchè in questi casi vi è la mancanza di alcune informazioni.

I dati di cui non si dispone sono relativi al tipo di formazione pre-universitaria ed il risultato ottenuto in quell'ambito, fondamentali per una parte della ricerca che si propone di valutare se vi siano legami o meno tra l'esito ottenuto nella scuola media superiore di provenienza e quello conseguito al termine degli studi universitari.

Il dataset originariamente presentava ulteriori 47 osservazioni relative ad altrettanti studenti che hanno conseguito il titolo di laurea nel medesimo arco temporale, ma si è ritenuto opportuno eliminare tali osservazioni perchè i relativi record non presentavano le informazioni utili alla ricerca.

L'analisi dell'esito universitario avviene tramite dati di tipo longitudinale. I dati analizzati sono di tipo individuale, ciò significa che ad ogni studente corrisponde un record. L'elemento unico, comune a tutti i soggetti, è il conseguimento del titolo di laurea. Le informazioni sono tratte dagli archivi amministrativi dell'Università di Padova.

La Facoltà di Scienze Statistiche di Padova gode di prestigio in ambito nazionale, conseguentemente gli studenti iscritti provengono da tutta Italia, anche se l'affluenza maggiore è da parte degli studenti della regione Veneto.

Poiché i corsi di laurea offerti dalla facoltà non prevedono frequenza obbligatoria, non è possibile stabilire quanti studenti fossero effettivamente in sede o meno.

Ultima considerazione che è possibile fare in questa fase iniziale di presentazione dei dati, è relativa al fatto che il campione al suo interno contiene varie età anagrafiche.

## **1.3.Variabili**

La scelta delle variabili è frutto di ovvie considerazioni sulle caratteristiche che hanno gli studenti all'inizio del percorso universitario e che in qualche modo possono determinare un iter accademico differenziato.

L'insieme delle variabili coinvolte per condurre lo studio può essere sostanzialmente diviso in due gruppi. Il primo include le variabili legate alla



formazione media superiore, mentre il secondo raggruppa le variabili che descrivono la carriera universitaria.

Al fine di valutare le possibili relazioni tra le condizioni proprie degli studenti o dei docenti e relativi esiti finali, alcune caratteristiche degli studenti costituiranno le variabili stimolo sulla media dei voti e sul voto finale che rappresenteranno invece le variabili risposta.

### **1.3.1. Formazione pre-universitaria**

La formazione pre-universitaria di ciascuno studente può essere analizzata prendendo in considerazione le seguenti variabili:

- Diploma: indica la scuola media superiore di provenienza distinta tra Altri Licei che comprendono Liceo Classico, Liceo linguistico, Liceo delle scienze sociali e altri tipi di licei, il Liceo Scientifico, l'Istituto Tecnico, l'Istituto Professionale, Ragioneria e per finire l' Istituto Straniero;
- Voto diploma: indica il punteggio del diploma di scuola media superiore;
- Base voto diploma: indica la base del voto del diploma e viene distinta in due valori: 60 o 100. Ovvero tra i laureati presi in esame molti avevano conseguito il diploma precedentemente alla riforma che coinvolgeva le scuole medie superiori, conseguendo un voto alla maturità espresso in sessantesimi. A partire dall'Anno Scolastico 1998-99 le valutazioni dell'esame di maturità furono espresse in centesimi;
- Provincia istituto diploma: indica la provincia in cui è situato l'istituto di scuola media superiore nella quale lo studente si è diplomato.

### **1.3.2.Carriera universitaria**

Il processo di analisi della carriera universitaria degli studenti che hanno conseguito la laurea di primo livello del Vecchio Ordinamento presso la Facoltà di Scienze Statistiche si basa sullo studio delle seguenti variabili:

- Corso di laurea: indica il corso di laurea scelto e frequentato dallo studente tra i quattro offerti dalla Facoltà. Come accennato in precedenza, l'offerta formativa (ex DM 509/99) proposta dalla Facoltà di Scienze Statistiche di Padova comprendeva quattro corsi di laurea triennale di primo livello e precisamente:

1. Corso di Laurea in Statistica, Economia e Finanza (SEF);
2. Corso di Laurea in Statistica e Gestione delle Imprese (SGI);
3. Corso di Laurea in Statistica, Popolazione e Società (SPS);
4. Corso di Laurea in Statistica e Tecnologie Informatiche (STI).

A partire dall'a.a. 2008/09 tale offerta formativa è stata progressivamente assorbita dalla nuova offerta didattica (ex DM 270/04), pertanto tutti gli studenti oggetto di studio sono laureati del Vecchio Ordinamento;

- Percorso: indica il percorso scelto dallo studente che gli consente di acquisire 36 crediti per raggiungere il totale di 180 CFU. I percorsi offerti sono quattro. I primi tre (A1-A2-A3) sono caratterizzati da uno stage, mentre l'ultimo (B) da un'attività da concordare e svolgere sotto la supervisione di un docente della Facoltà. Il tipo di stage ed i crediti ad esso associati dipendono dalla sua durata, variabile dai 3 ai 6 mesi. Lo stage costituisce un'attività formativa fortemente raccomandata, ma non obbligatoria. Il percorso B consente al laureando di concordare un lavoro di approfondimento con il proprio relatore, sviluppando un progetto attraverso la raccolta, l'analisi e la rielaborazione del materiale;

- Area relatore: indica l'area scientifico-disciplinare nella quale rientrano le attività didattiche del docente che svolge il ruolo di relatore per la relazione finale dello studente. Le aree sono così nominate: DSS (Socio

Demografica), ECO (Economica), INF (Informatica), MAT (Matematica), STAT (Statistica), STATEC (Statistico Economica);

- Area controrelatore: indica l'area scientifico-disciplinare di appartenenza del docente che funge da controrelatore in sede di discussione della tesi;
- Media voti: esprime il voto medio degli esami conseguito dallo studente;
- Velocità: questa variabile indica il numero di punti che ogni studente può conseguire in base al tempo impiegato per portare a termine il proprio percorso di studi triennale. Può anche essere interpretato come un premio alla "velocità" della carriera dello studente, volto a valorizzare in modo particolare chi tra gli studenti "si laurea in corso". I crediti che questa variabile conferisce sono distribuiti come mostrato in *Tab.1.1*.

*Tabella 1.1: PUNTEGGI AGGIUNTIVI PER LE "LAUREE IN CORSO" dell' EX DM 509/99*

***Punteggi Aggiuntivi***

<i>Lo studente si è immatricolato a settembre dell'anno <math>x</math>; si laurea</i>	<i>Nella sessione</i>	<i>Ottenendo punti aggiuntivi</i>
Entro marzo dell'anno $x+3$	Anticipo rispetto alla durata naturale	7
A giugno/luglio anno $x+3$	I sessione "in corso"	6
A settembre/ottobre/novembre anno $x+3$	II sessione "in corso"	4
A marzo anno $x+4$	Sessione straordinaria "in corso"	2
Da giugno dell'anno $x+4$	I e successive sessioni "fuori corso"	0

- Incremento: rappresenta il punteggio assegnato agli studenti per l'elaborato finale. La valutazione attribuita alla relazione finale e alla attività sottostante conferisce da 0 a 6 punti;
- Voto finale: indica il voto di laurea degli studenti che hanno conseguito il titolo. Il voto finale di laurea è ottenuto dal voto medio degli esami ponderato con il valore dei crediti della relativa attività didattica, espresso in centodecimi e arrotondato all'intero più vicino, più il punteggio in centodecimi conseguito nella prova finale.
- Lode: la lode viene assegnata dalla commissione quando il punteggio complessivo è maggiore o uguale a 112. Oppure su proposta scritta del controrelatore quando il punteggio complessivo è uguale a 109 -110 -111 ed il controrelatore stesso ravvisa nella prova finale del candidato particolari elementi di originalità e documenta capacità di risolvere problematiche concrete in maniera innovativa.

## **1.4.Scopi e Strumenti**

Il presente lavoro si propone di valutare se esistano delle relazioni tra i punteggi assegnati agli studenti laureati per l'elaborato finale e le altre variabili, l'area scientifico disciplinare di appartenenza dei docenti coinvolti in tale ambito in qualità di relatori o controrelatori.

Per giungere al suddetto scopo verranno usati gli strumenti di seguito riportati:

- lo studio dei dati e delle singole variabili contenute nel database;
- la stima della tabella di dati di durata attraverso metodo attuariale;
- la stima e il disegno della funzione di sopravvivenza attraverso l'uso del metodo di Kaplan-Meier;
- la stima e il disegno della funzione di sopravvivenza attraverso l'uso del modello di Cox;

# *CAPITOLO SECONDO*

---

## *DETERMINANTI DEGLI ESITI*

### **2.1. Analisi descrittiva**

Le analisi preliminari consistono nella costruzione di tabelle semplici e successivamente nella costruzione di tabelle di contingenza a due e a tre entrate, e servono essenzialmente per la scelta delle variabili esplicative, per la categorizzazione più appropriata delle stesse e per indagare e descrivere la loro influenza sulla variabile risposta.

Le variabili esplicative inizialmente prese in considerazione sono il Corso di Laurea frequentato dai laureati, il Diploma, e quindi la scuola media superiore di provenienza, la Provincia nella quale è stato conseguito il diploma, la Base del Voto di diploma, l'Area scientifico disciplinare del relatore e del controrelatore, il Percorso scelto dagli studenti per conseguire la laurea e la Velocità.

Vi sono dunque variabili input o esplicative e variabili risposta, date dall'esito universitario descritto dal punteggio assegnato al lavoro di tesi e dal voto finale di laurea.

Di seguito vengono presentati i dati, relativi agli studenti laureati, suddivisi secondo le variabili sopra esposte.

## 2.1.1. Corso di Laurea

Tabella 2.1: DISTRIBUZIONE DEGLI STUDENTI LAUREATI SECONDO IL CORSO DI LAUREA

<i>Corso di Laurea</i>		
	Frequenza	Percentuale
SGI	337	44,5
SEF	166	21,9
STI	164	21,6
SPS	91	12,0
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

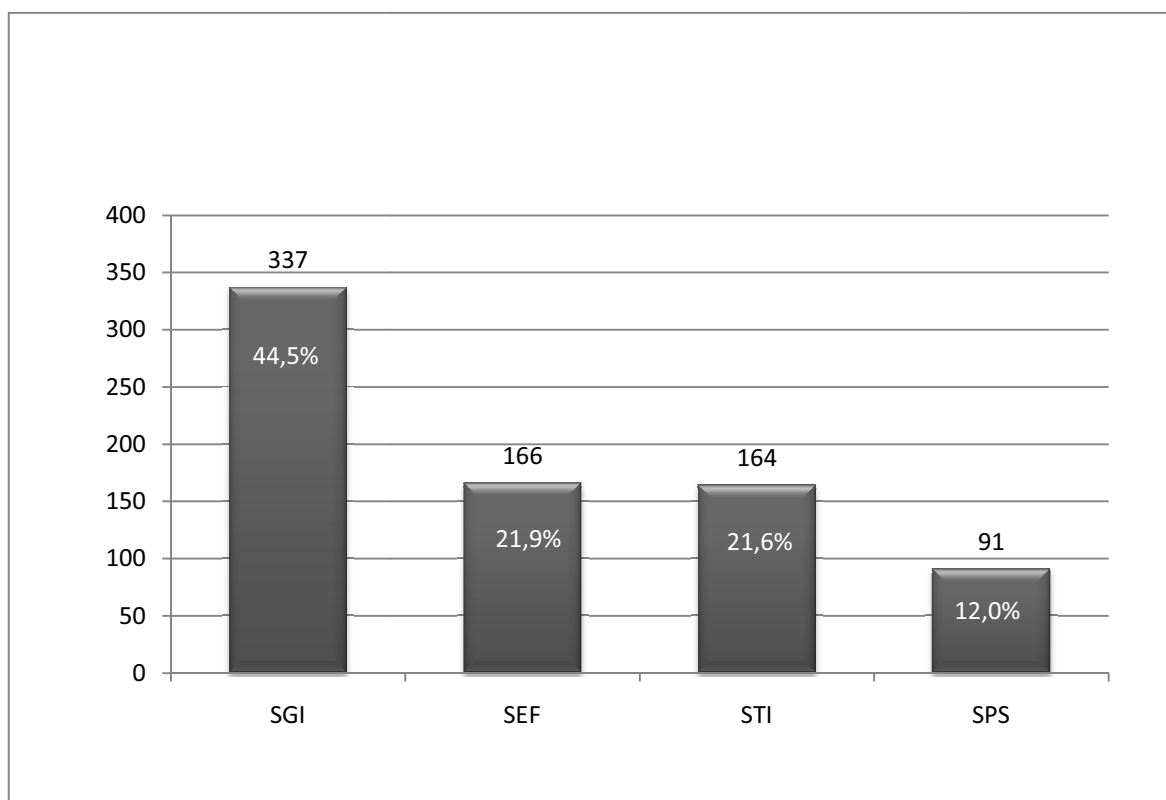


Figura 2.1: Grafico a barre frequenze per Corso di Laurea

## 2.1.2.Scuola di Provenienza

Tabella 2.2: DISTRIBUZIONE DEGLI STUDENTI LAUREATI SECONDO IL DIPLOMA DI PROVENIENZA.

	Frequenza	Percentuale
SCIENTIFICO	294	38,8
IST_TECNICI	193	25,5
RAGIONERIA	148	19,5
IST_STRANIERO	40	5,3
ALTRI LICEI	31	4,1
IST_PROFESS	29	3,8
CLASSICO	21	2,8
MANCANTI	2	0,3
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

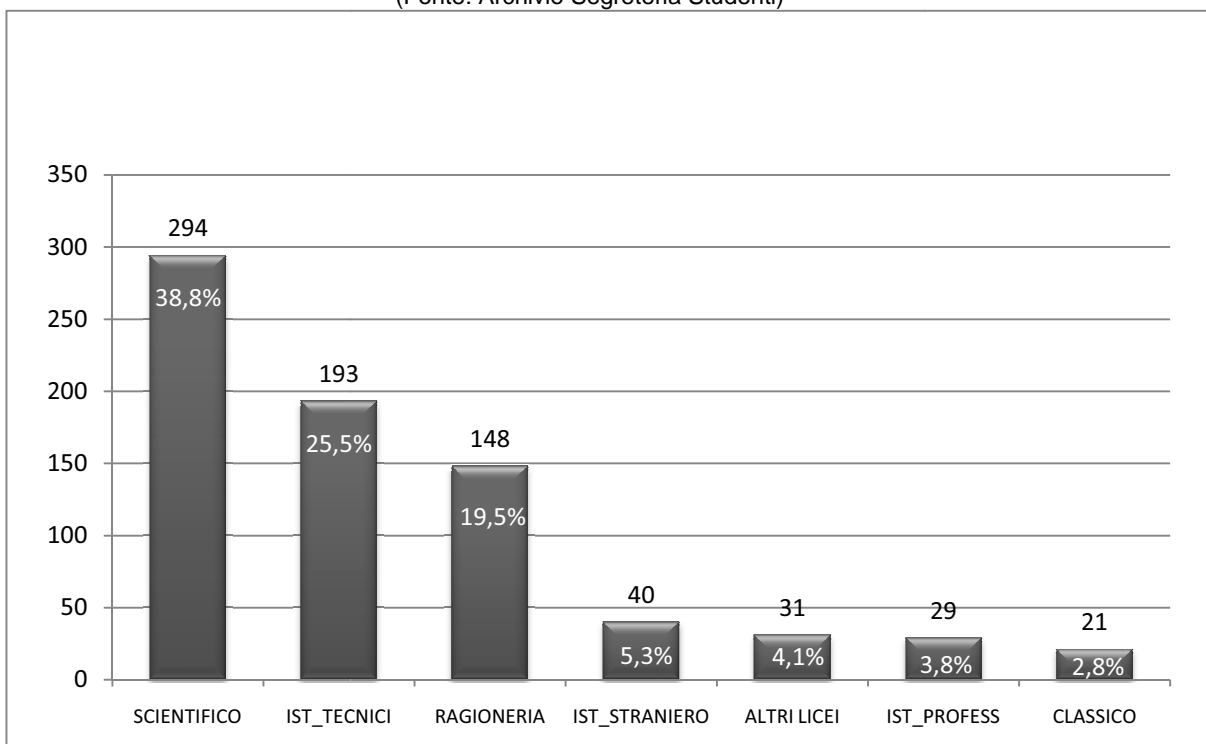


Figura 2.2: Grafico a barre frequenze per Diploma

La percentuale degli studenti provenienti dal liceo scientifico (Tab.2.2) è molto alta, rappresenta il 38.8%. Per quanto riguarda gli istituti tecnici, escluso l'Istituto Tecnico Commerciale, la percentuale degli studenti è pari al 25.5%, pertanto nella categorizzazione della variabile è parso opportuno mantenere separato l'istituto

tecnico commerciale dagli altri istituti tecnici perché in media il 19,5% dei laureati è in possesso di un diploma di ragioneria. Si noti inoltre che il 6,9% degli studenti proviene da altri tipi di liceo o da istituti magistrali e che il 3,8% da istituti professionali. Si attesta sul 5,3% la percentuale di laureati in possesso di un titolo di studio straniero.

Successivamente, le modalità che caratterizzavano la variabile Diploma sono state ridotte, al fine di riunire gli studenti in possesso di un diploma di istituto tecnico, tecnico-professionale e di ragioneria in un'unica modalità che è descritta dalla categoria Istituto Tecnico; stessa operazione è stata applicata agli studenti provenienti dai vari licei, fatta eccezione per il liceo scientifico, e a quelli con diploma di istituto magistrale, inseriti nell'unica categoria denominata Liceo.

Questa riduzione di categorie è avvenuta per rendere più snelle le operazioni necessarie alla costruzione dei modelli adottati nel capitolo 4.

*Tabella 2.3: DISTRIBUZIONE DEGLI STUDENTI LAUREATI SECONDO DIPLOMA/MODALITÀ AMPLIATE*

<i>Diploma</i>		
	Frequenza	Percentuale
IST_TECNICO	370	48,8
SCIENTIFICO	294	38,8
LICEO	52	6,9
IST_STRANIERO	40	5,3
MANCANTI	2	0,3
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

Riducendo le categorie si evidenzia come la percentuale degli studenti in possesso di un diploma di maturità tecnica sia ben superiore a quella relativa ai laureati provenienti dal liceo scientifico.

Quasi la metà dei laureati, più precisamente il 48,8%, proviene dagli istituti tecnici. I dati relativi agli studenti provenienti dal liceo scientifico e agli studenti stranieri rimangono chiaramente invariati, mentre la percentuale che caratterizza i laureati, la cui istruzione media superiore sia umanistica, si attesta al 6,9%.



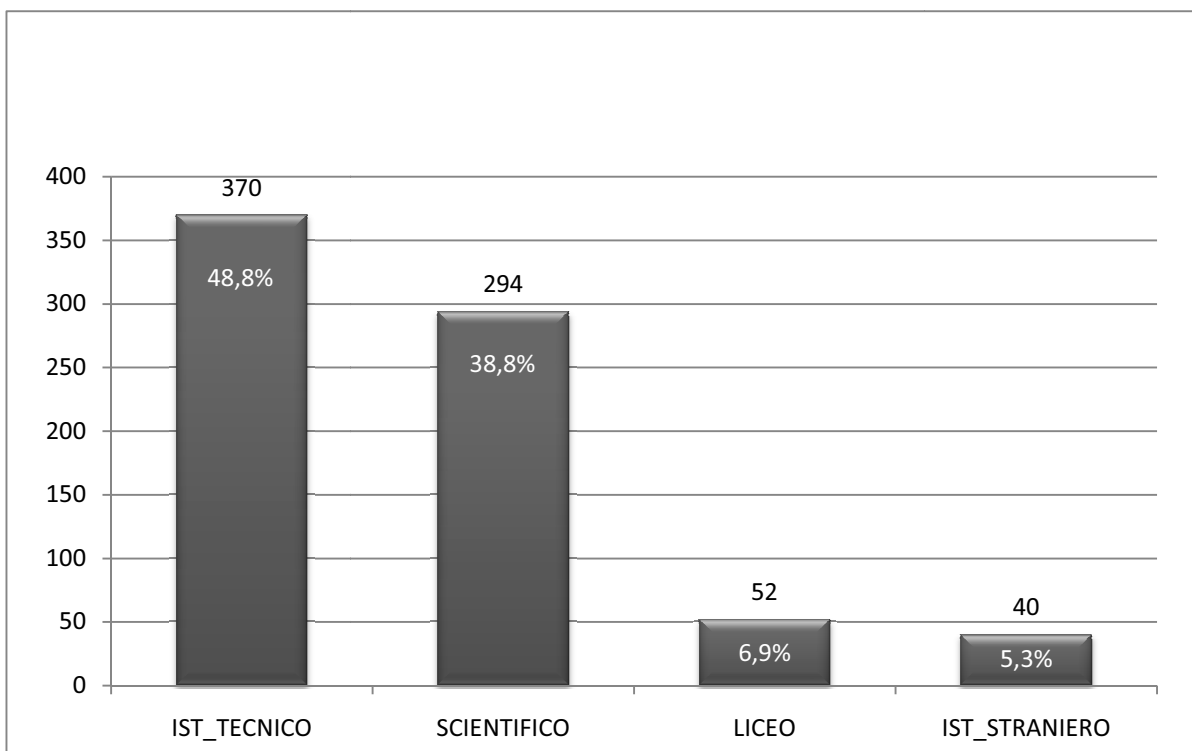


Figura 2.3: Grafico a barre frequenze per Diploma categorie ridotte

### 2.1.3.Base Voto Diploma

Tabella 2.4: *STUDENTI CON VALUTAZIONE DIPLOMA ESPRESSA IN CENTESIMI/SESSANTESIMI*  
*Voto Diploma*

	Frequenza	Percentuale
BASE100	633	83,5
BASE60	123	16,2
MANCANTI	2	0,3
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

La variabile Base Voto Diploma descrive il numero di studenti che hanno conseguito il diploma di maturità con un punteggio espresso in sessantesimi o in centesimi, in base al periodo nel quale è stato superato l'esame di maturità. La distinzione è stata necessaria per stabilire se vi fosse una continuità tra l'andamento della carriera scolastica precedente all'ingresso nel sistema universitario e il rendimento ottenuto nel corso di laurea.

Degli studenti laureati presi in esame, il 16,2% presenta un voto di maturità espresso in sessantesimi, il restante 83,5% ha conseguito il diploma di scuola media superiore con il sistema che è ancor oggi in vigore.

## 2.1.4.Voto Diploma

Di seguito vengono riportate alcune statistiche relative al voto di diploma conseguito dagli studenti, distinguendo le valutazioni espresse in sessantesimi e in centesimi e i relativi istogrammi. Successivamente i voti espressi in sessantesimi, riguardanti i maturi precedenti al 1998, sono stati ottenuti trasformandoli in valutazioni espresse in centesimi, si sono lasciati inalterati quelli successivi al 1998, dopo l'avvento della riforma scolastica.

*Tabella 2.5: VOTO DIPLOMA DISTINTO PER MISURA DI VALUTAZIONE*

<i>Voto Diploma</i>			Statistica	
BASE Voto Diploma	60	Media	44,23	
		Intervallo di confidenza per la media al 95%	Limite inferiore	42,97
			Limite superiore	45,49
		Mediana	43,00	
		Varianza	49,948	
	100	Media	78,86	
		Intervallo di confidenza per la media al 95%	Limite inferiore	77,90
			Limite superiore	79,82
		Mediana	77,00	
		Varianza	151,738	

(Fonte: Archivio Segreteria Studenti)

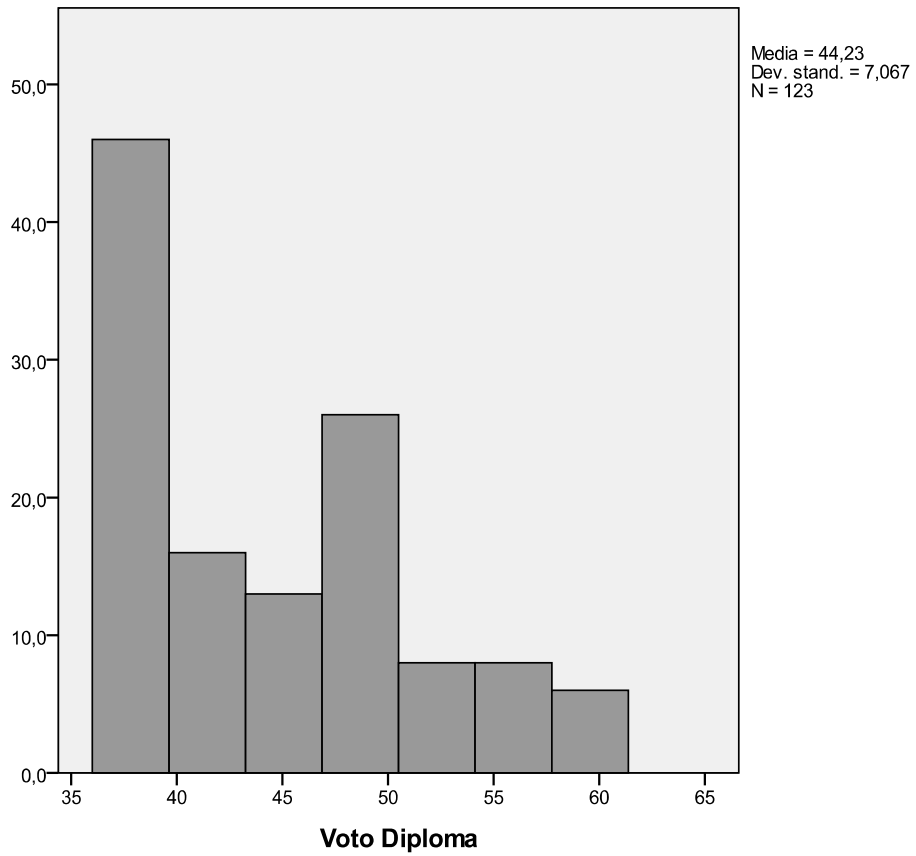


Figura 2.4: Istogramma Voto Diploma con base in sessantesimi

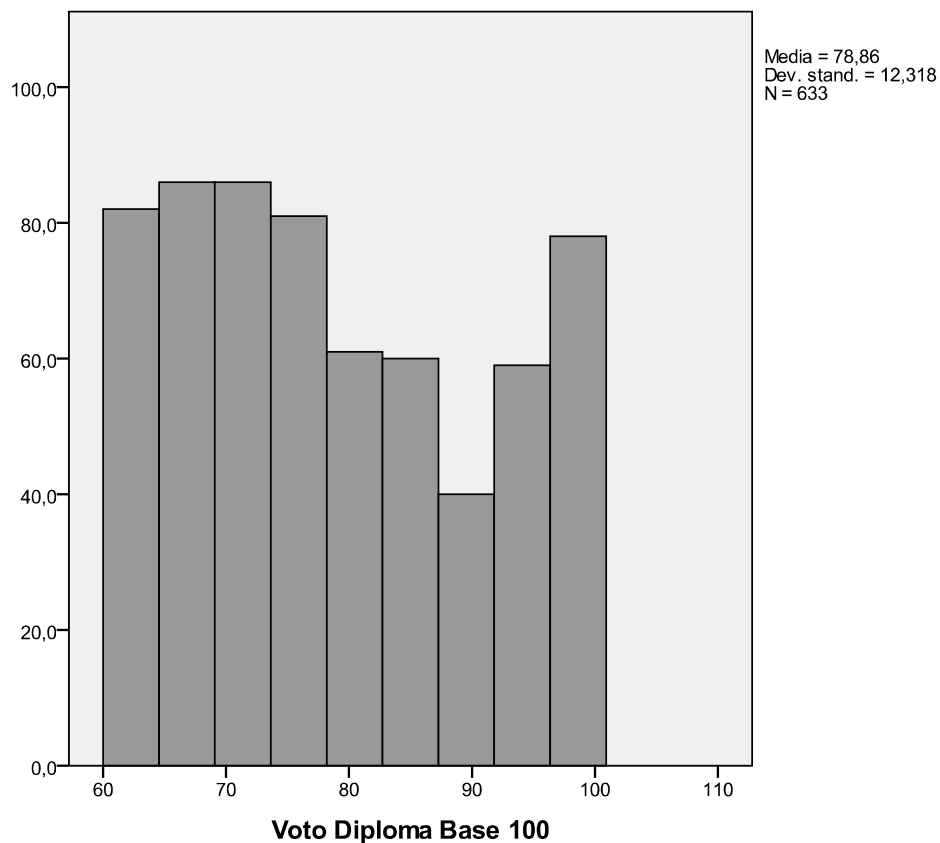


Figura 2.5: Istogramma Voto Diploma con base in centesimi

Successivamente, per uniformare le operazioni di analisi, le valutazioni dei diplomi espresse in sessantesimi sono state trasformate in voti espressi in centesimi.

Tabella 2.6: VOTO DIPLOMA CON UNICA BASE

<i>Voto Diploma</i>		Statistica
UnicaBASE	Media	78,02
	Intervallo di confidenza per la media al 95%	
	Limite inferiore	77,14
	Limite superiore	78,90
	Mediana	77,00
	Varianza	153,011

(Fonte: Archivio Segreteria Studenti)

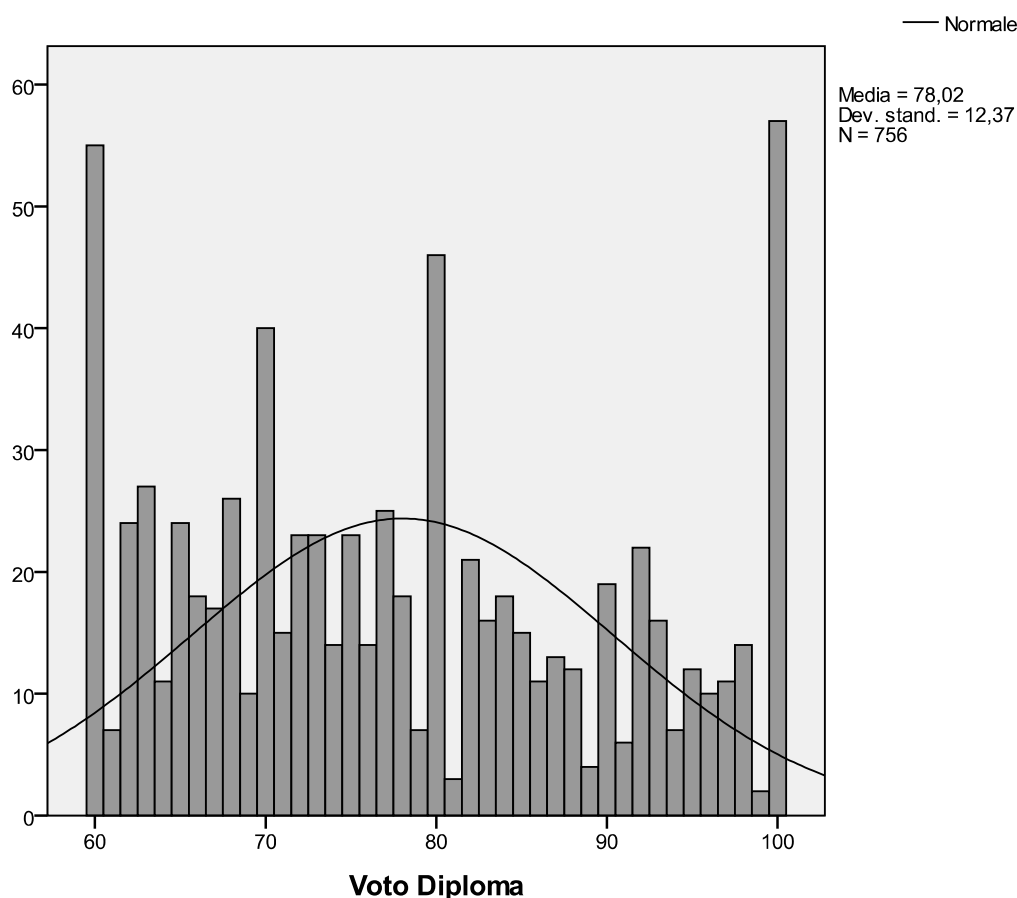


Figura 2.5: Istogramma Voto Diploma totale con base in centesimi

Nella seguente tabella, (Tab.2.7), viene presentata la distribuzione dei laureati secondo il voto di diploma e relativo punteggio di laurea.

Il voto ottenuto al diploma è stato diviso in quattro classi di ampiezza 10, mentre la valutazione che esprime il voto finale conseguito con la laurea è stata classificata in intervalli di ampiezza diversa.

Nella categorizzazione della variabile voto alla maturità è stata usata l'usuale classificazione centesimale, con le classi chiuse a destra.

Emerge come l'esito dell'andamento della carriera scolastica pre-universitaria si discosti poco dall'esito della carriera universitaria.

*Tabella 2.7: DISTRIBUZIONE DEGLI STUDENTI SECONDO IL VOTO ALLA MATURITÀ E CORRISPONDENTE VOTO DI LAUREA*

<b>Distribuzioni delle Valutazioni</b>				
	(76,90]	(90,100]	(100,105]	(105,110]
[60,70]	125	106	17	11
%riga	48,3%	40,9%	6,6%	4,2%
%colonna	52,5%	40,8%	15,9%	7,3%
(70,80]	69	80	33	26
%riga	33,2%	38,5%	15,8%	12,5%
%colonna	29,0%	30,8%	30,9%	17,2%
(80,90]	24	46	25	37
%riga	18,2%	34,8%	18,9%	28,1%
%colonna	10,1%	17,7%	23,3%	24,5%
(90,100]	20	28	32	77
%riga	12,7%	17,8%	20,4%	49,1%
%colonna	8,4%	10,7%	29,9%	51,0%

(Fonte: Archivio Segreteria Studenti)

## **2.1.5. Provincia Istituto Diploma**

Quasi il 90% degli studenti laureati nella Facoltà di Scienze Statistiche di Padova risulta essere residente in Veneto (*Tab.2.8*). Di questi, quasi il 50% è residente nella provincia di Padova; molto probabilmente risiedere in una zona vicina al baricentro della struttura universitaria permette forse di arrivare alla fine degli studi e di essere meglio informati sui corsi di laurea attivati. Inoltre, molte facoltà normalmente svolgono un'opera di sponsorizzazione nelle scuole superiori che meglio rappresentano la loro tipologia di studente, ma molto spesso si limitano a farlo nel Comune e nella Provincia sede del corso di studi. Gli studenti provenienti dalle province di Vicenza e Venezia sono solamente l'11,7% e l'11,6% rispettivamente, quelli provenienti dalla provincia di Treviso sfiorano il 9%, a tal proposito va ricordato che in quegli anni era attivo un Corso di Laurea in Statistica proprio nella città di Treviso; per le altre province della regione l'apporto risulta esiguo.

Ulteriore considerazione, che permette di comprendere maggiormente l'elevato numero degli studenti provenienti dalla provincia di Padova, è riscontrabile nella maggior accessibilità da parte di questi, anche grazie alla rete di trasporto pubblico extraurbano, nel recarsi nella sede universitaria.

Il dato che descrive gli studenti provenienti dalle altre regioni del Nord Italia è del 5%, mentre quelli relativi ai laureati provenienti dal Sud e dal Centro sono molto bassi. Si ricorda che il 5,3% della popolazione in esame proviene dall'estero.

*Tabella 2.8: DISTRIBUZIONE DEGLI STUDENTI LAUREATI SECONDO LA PROVINCIA DI PROVENIENZA*

<i>Provincia Istituto Diploma</i>		
	Frequenza	Percentuale
PADOVA	372	49,1
VICENZA	89	11,7
VENEZIA	88	11,6
TREVISO	66	8,7
VERONA	23	3,0
ROVIGO	17	2,2
BELLUNO	11	1,5
ESTERO	40	5,3
NORD	38	5,0
CENTRO	3	0,4
SUD	9	1,2
MANCANTI	2	0,3
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

*Tabella 2.9: STUDENTI LAUREATI RESIDENTI IN VENETO*

<i>Veneto</i>		
	Frequenza	Percentuale
Veneti	687	90,6
Non Veneti	71	9,4
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

## VENETO

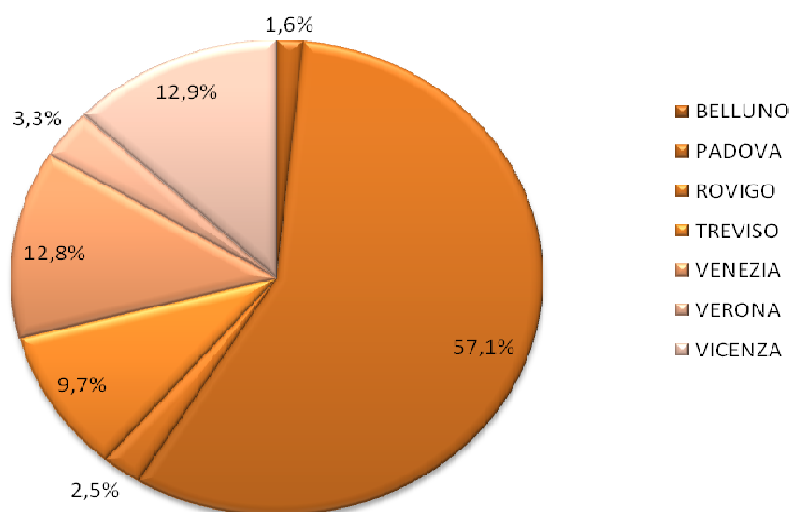


Figura 2.5: Grafico a torta distribuzione studenti nelle province venete

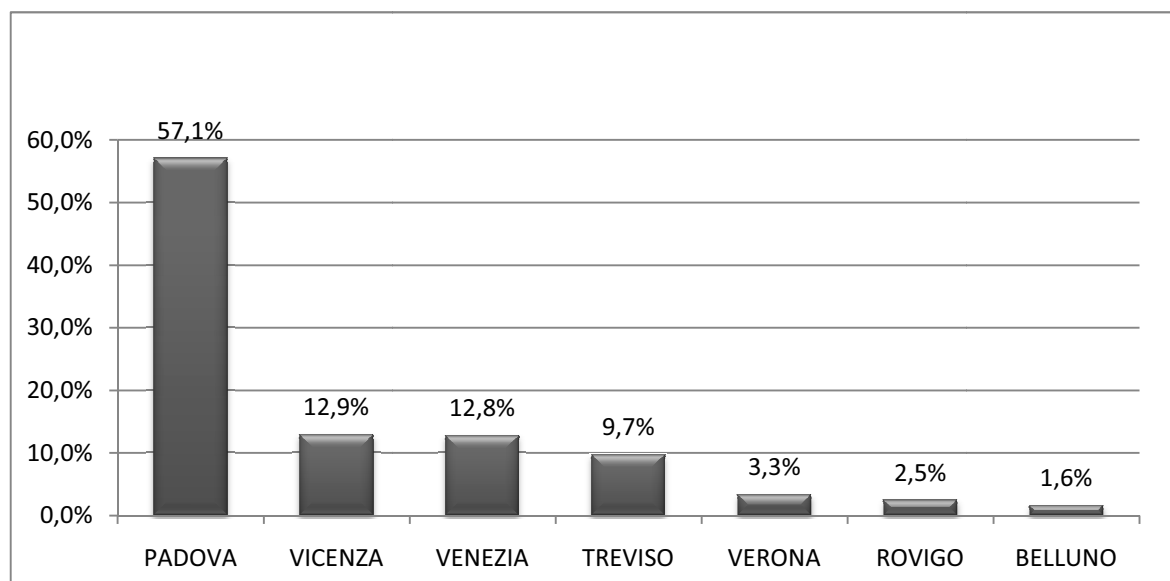


Figura 2.6: Grafico a barre distribuzione studenti nelle province venete



## TOTALE

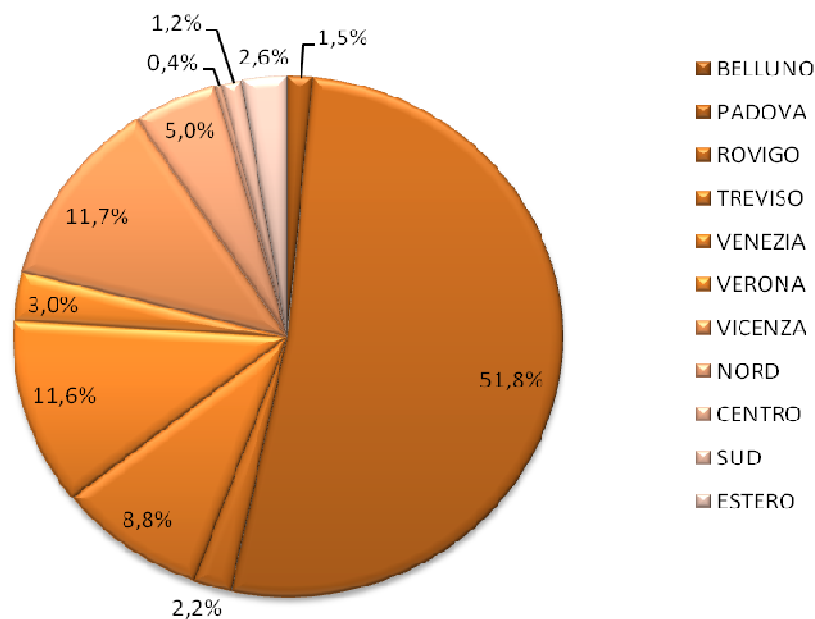


Figura 2.7: Grafico a torta distribuzione studenti per provenienza

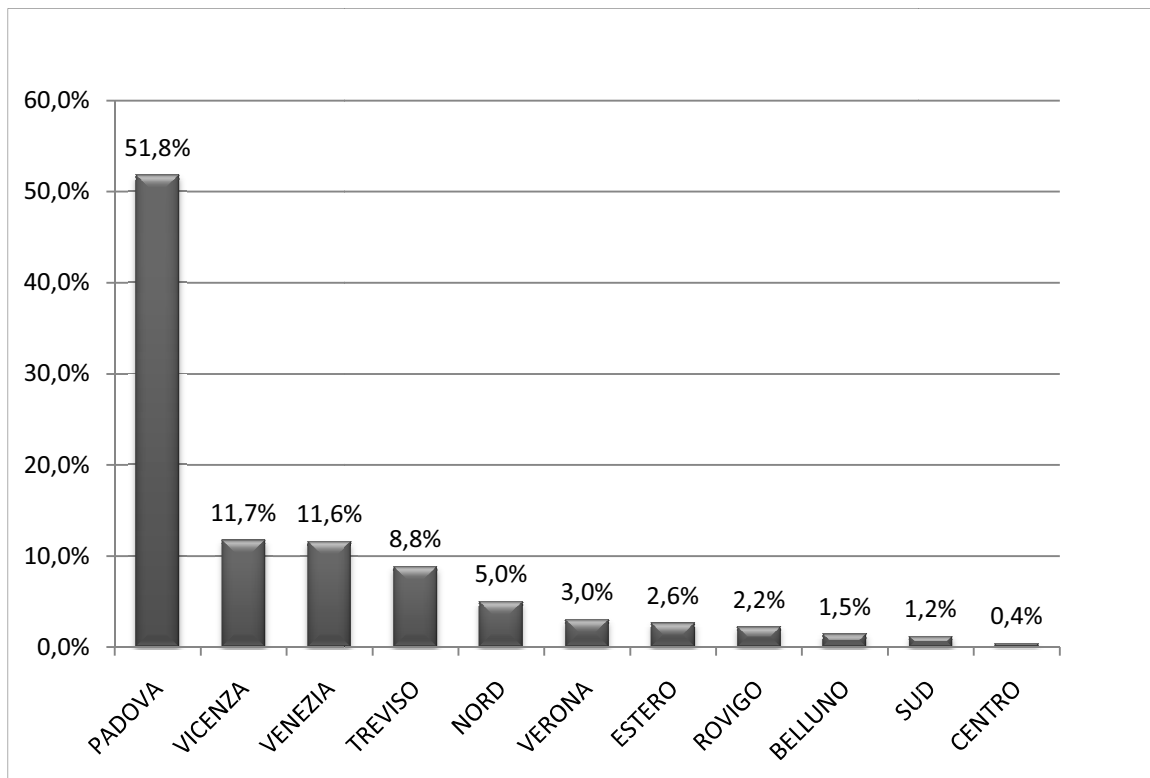


Figura 2.8: Grafico a barre distribuzione studenti per provenienza

## 2.1.6.Percorso

Dalla *Tab.2.10*, emerge come quasi la metà dei laureati, per la precisione il 48,7%, abbia scelto il percorso B; tale propensione da parte degli studenti può essere ricondotta sia ad un aspetto puramente pratico, che coinvolge principalmente gli studenti lavoratori che per ovvie ragioni non possono dedicare il tempo sufficiente ad un'esperienza in azienda/ente, sia ad uno accademico che interviene nella scelta dell'opzione B in quanto uno studente, che abbia deciso di proseguire gli studi iscrivendosi alla Laurea specialistica, reputa più idoneo al suo percorso formativo un progetto di approfondimento.

La scelta dello stage breve o di quello medio caratterizza rispettivamente il 22,8% e il 21,2% dei laureati e solamente il 7,3% dei soggetti in esame ha preferito intraprendere il percorso A3 che prevede uno stage di 6 mesi.

*Tabella 2.10: PERCORSI OFFERTI DALLA FACOLTÀ*

***Percorso***

	Frequenza	Percentuale
B	369	48,7
A1	173	22,8
A2	161	21,2
A3	55	7,3
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

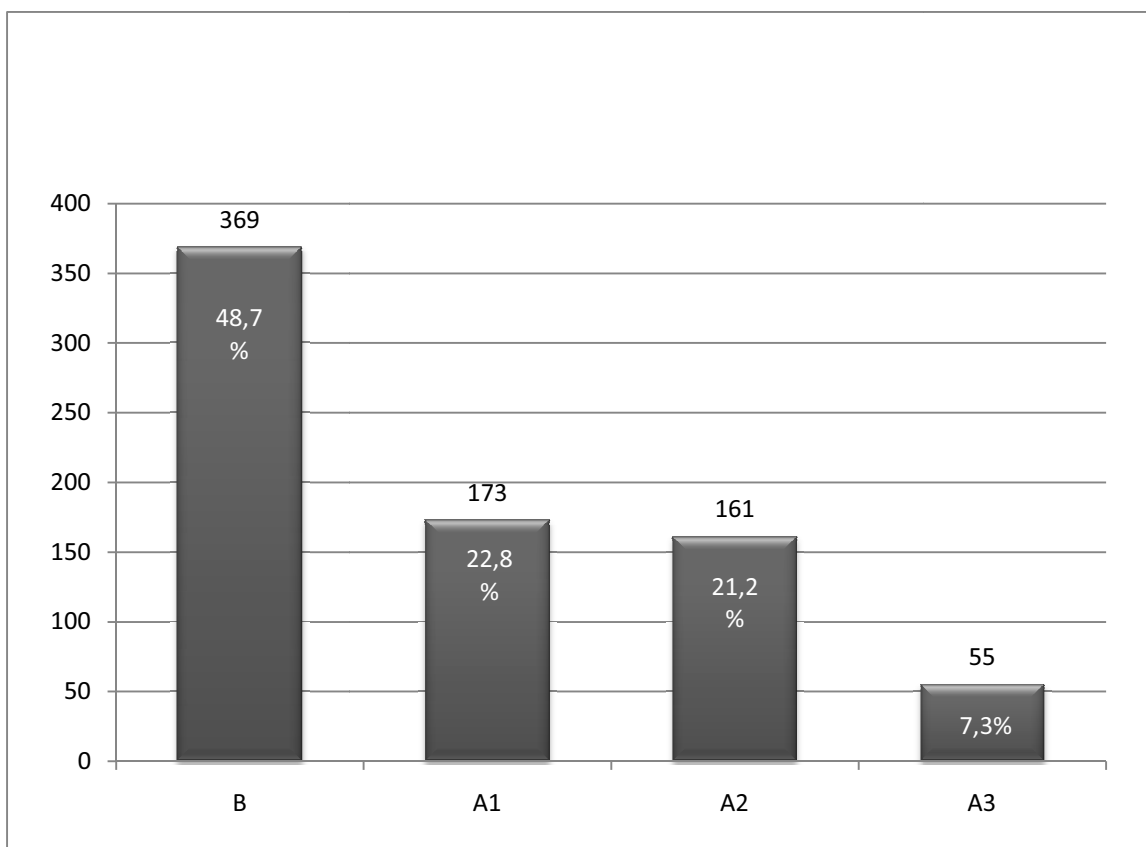


Figura 2.9: Grafico a barre distribuzione per percorso

### 2.1.7. Velocità

Questa variabile indica un punteggio aggiuntivo che viene assegnato agli studenti che si laureano in corso; rappresenta un premio che ha lo scopo di valorizzare gli studenti più meritevoli.

I laureati, che hanno in questo modo ottenuto almeno 2 punti aggiuntivi al voto medio per aver conseguito la laurea in corso, sono più del 60%. Coloro invece che non hanno potuto giovare di tale opportunità, e ai quali dunque non è stato assegnato alcun punto in aggiunta, sono il 38,3%.

Da sottolineare che il 31,9% degli studenti, che risponde al numero di 242 soggetti, ha conseguito la laurea nella prima sessione di laurea “in corso”, e che

ben il 5% delle unità ha ottenuto 7 punti per la velocità perché si sono laureati in anticipo rispetto alla durata naturale del corso.

Tabella 2.11: PUNTEGGIO ASSOCIATO alla VELOCITÀ

<i>Velocità</i>		
	Frequenza	Percentuale
0	290	38,3
2	64	8,4
4	124	16,4
6	242	31,9
7	38	5,0
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

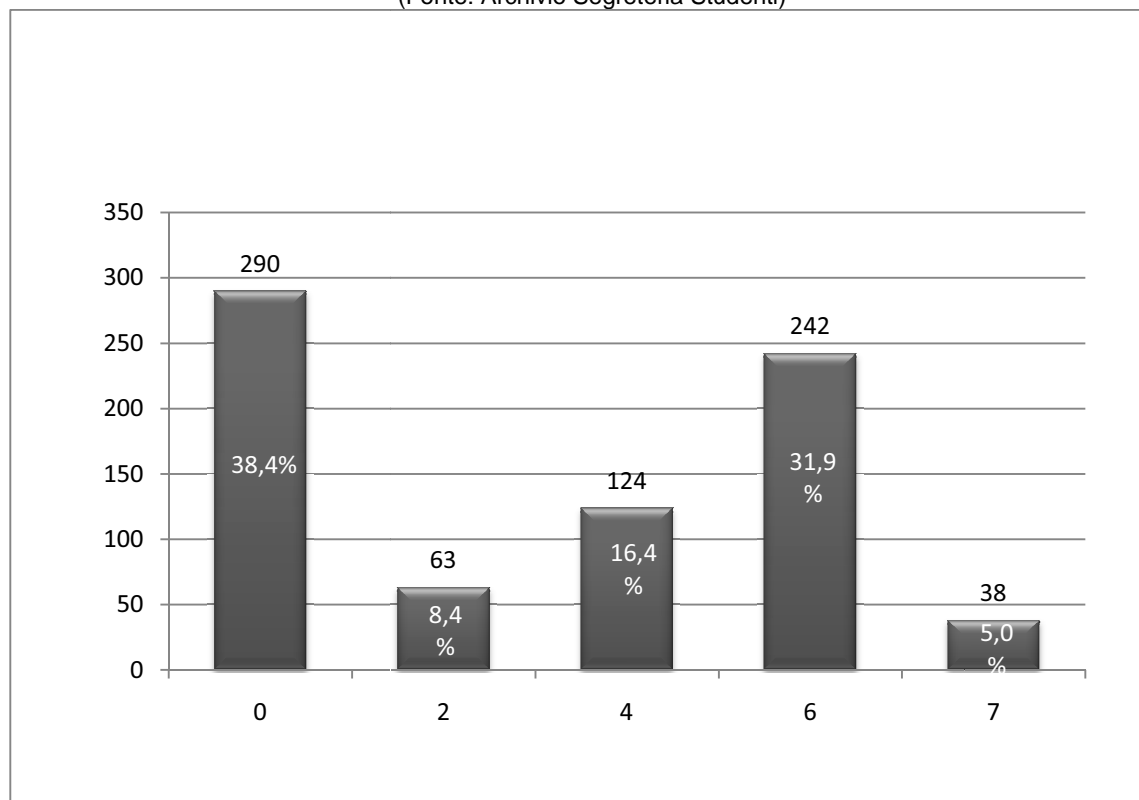


Figura 2.10: Grafico a barre distribuzione per “punti velocità”

## 2.1.8. Area Relatore e Area Contro-Relatore

Le due variabili Area Relatore e Area ControRelatore descrivono l'area scientifico-disciplinare nella quale rientrano le attività didattiche del docente che svolge il ruolo di relatore o di controrelatore per la relazione finale dello studente. Il personale docente viene suddiviso in sei diverse aree scientifico-disciplinari: DSS, ECO, INF, MAT, STAT, STATEC.

I professori che maggiormente rivestono il ruolo di relatore, per il 39,3% delle volte, sono i docenti appartenenti all'area economica, questo dato sottolinea come gli studenti prediligano sviluppare principalmente dei lavori di tesi che si occupino soprattutto degli aspetti propri della sfera economica. Seguono gli insegnanti appartenenti all'area statistico-economica, svolgendo la funzione di relatore nel 20% dei casi, mentre ai docenti dell'area disciplinare puramente statistica viene chiesto, da parte dei laureandi, di seguirli nella relazione finale, solamente nel 14% dei casi. Il dato relativo ai professori dell'ambito informatico è dell'8.3%, mentre quello che caratterizza il corpo docente prettamente matematico è leggermente superiore ai 2.5 punti percentuali.

I valori, che descrivono il numero di casi in cui i docenti provenienti dalle varie aree scientifico-disciplinari hanno ricoperto il ruolo di contro-relatore, differiscono di poco, per le diverse aree di appartenenza, rispetto ai dati sopra citati.

Tabella 2.12: *DOCENTI-RELATORI*

<i>Area_relatore</i>		
	Frequenza	Percentuale
ECO	298	39,3
STATEC	152	20,1
DSS	119	15,7
STAT	106	14,0
INF	63	8,3
MAT	20	2,6
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

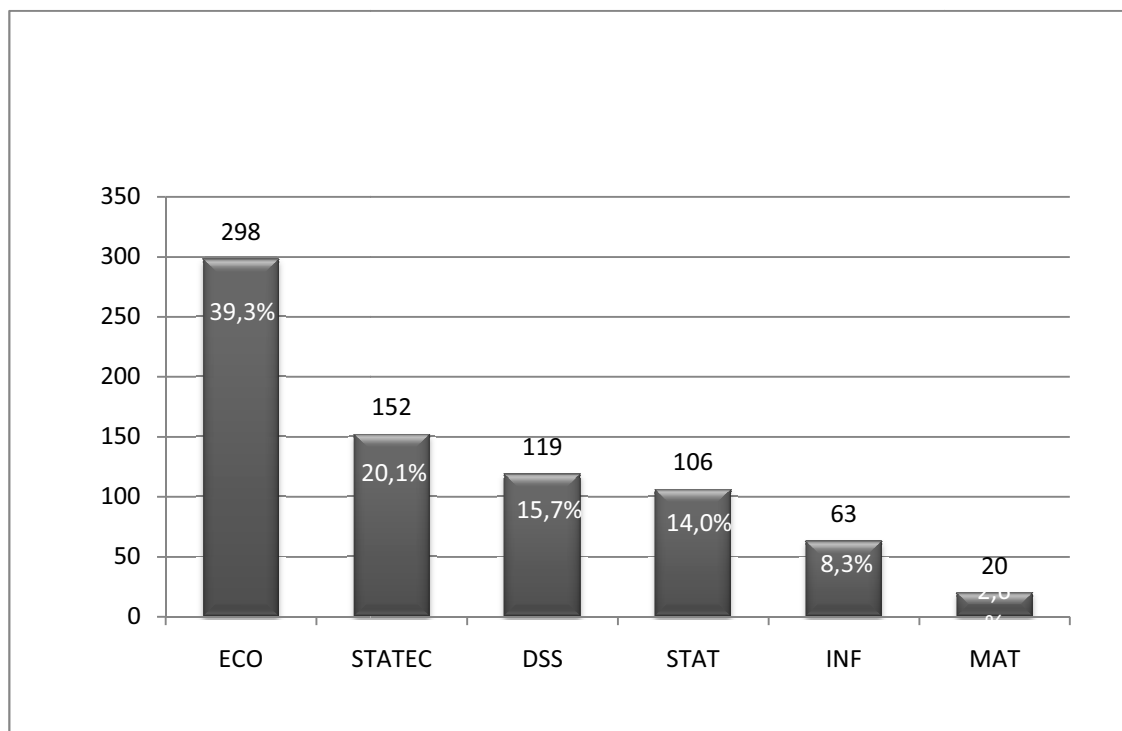


Figura 2.11: Grafico a barre distribuzione per Area Relatore

Nella seguente tabella, (Tab.2.13), vengono presentate le percentuali relative alla scelta, da parte degli studenti, sul percorso che permette di completare il piano di studi e la scelta del relatore o più in generale l'area disciplinare corrispondente al relatore.

In questo modo si vuole capire quali siano gli interessi degli studenti e come siano orientate le scelte del percorso utile al conseguimento del titolo. Ci si potrebbe aspettare che gli studenti, interessati alla frequentazione di uno stage, chiedano ai docenti dell'area economica di coordinare il proprio percorso e ciò trova riscontro nei dati presentati. Ma allo stesso tempo, si sottolinea che la percentuale maggiore di studenti che si rivolgono a relatori del settore economico, 42,5%, è caratterizzata da coloro che decidono di presentare alla prova finale un progetto di approfondimento, seguiti nel 19,2% dei casi, dagli studenti che scelgono i propri relatori in ambito Statistico. Si rileva inoltre la tendenza da parte dei relatori, appartenenti alle aree Statistica e Statistico Economica, a richiedere di

intervenire in qualità di controrelatori ai docenti delle altre aree disciplinari come riporta la *Tab.2.15*. Questa propensione a un confronto con controrelatori di aree diverse da parte dei docenti dei suddetti settori, viene seguita in misura minore dai colleghi delle aree DSS,ECO e INF che vedono, per circa il 90% delle prove finale, relatore e controrelatore appartenenti allo stesso settore disciplinare.

*Tabella 2.13: AREA RELATORE E PERCORSO*

		Percorso				Totale
		A1	A2	A3	B	
Area Relatore	DSS	16,2%	14,9%	23,6%	14,6%	15,7%
	ECO	36,4%	38,5%	30,9%	42,5%	39,4%
	INF	9,8%	10,6%	18,2%	5,1%	8,3%
	MAT	3,5%	1,2%	0,0%	3,3%	2,6%
	STAT	11,0%	6,8%	7,3%	19,2%	13,9%
	STATEC	23,1%	28,0%	20,0%	15,2%	20,1%
Totale		100,0%	100,0%	100,0%	100,0%	100,0%

(Fonte: Archivio Segreteria Studenti)

*Tabella 2.14: DOCENTI-CONTRORELATORI*

***Area\_ControRelatori***

	Frequenza	Percentuale
ECO	293	38,7
STAT	171	22,6
EC	121	16,0
DSS	92	12,1
INF	67	8,8
MAT	14	1,8
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

Tabella 2.15: AREA

		Area ControRelatore
		UGUALE
Area Relatore	DSS	89,1%
	ECO	92,0%
	INF	93,7%
	MAT	50,0%
	STAT	71,4%
	STATEC	82,9%

(Fonte: Archivio Segreteria Studenti)

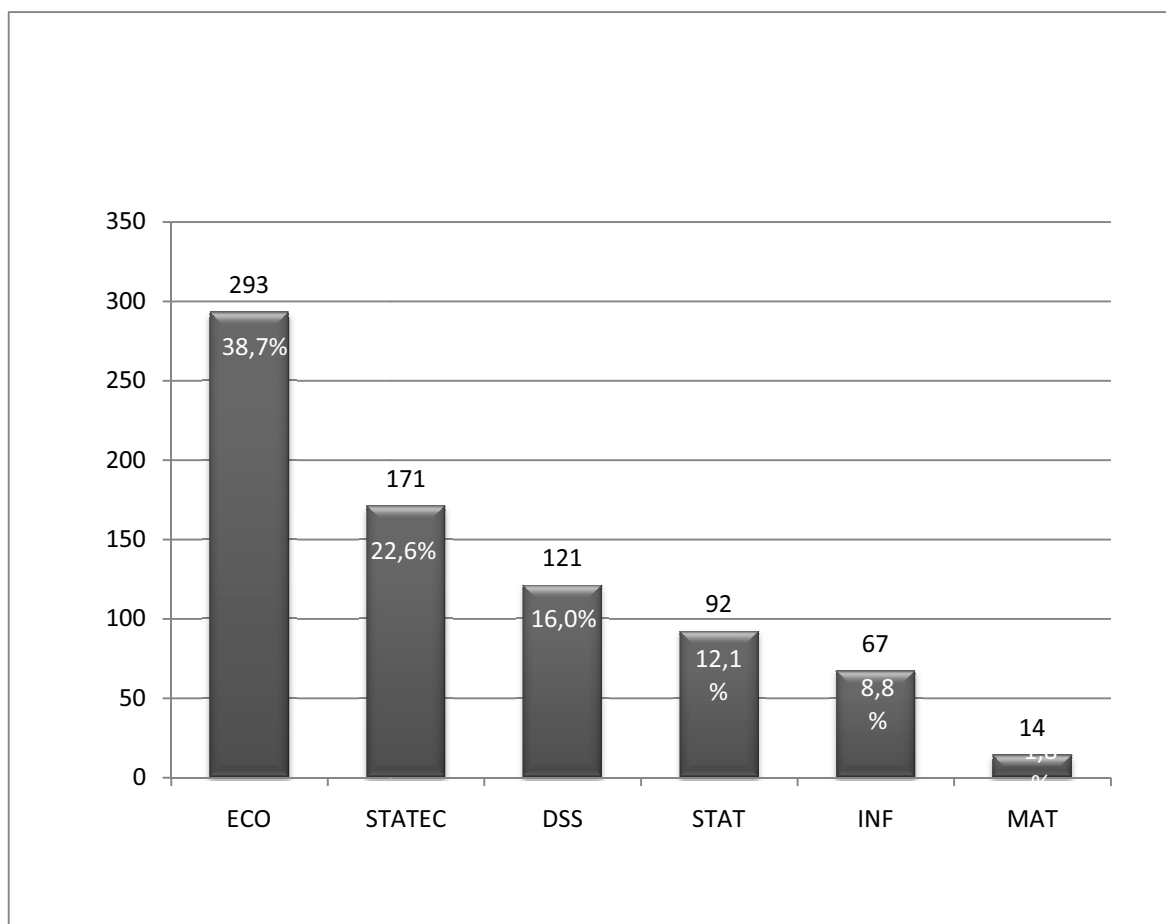


Figura 2.12: Grafico a barre distribuzione per Area Contro Relatore



## 2.1.9.Voto Finale

Questa variabile indica il voto finale di laurea che è costituito dal voto medio degli esami espresso in centodecimi e arrotondato all'intero più vicino, più il punteggio in centodecimi conseguito per la relazione finale, più eventuale premio per la velocità.

Nella *Tab.2.16*, il voto finale è presentato in 4 intervalli di ampiezza diversa.

Gli studenti laureati sono stati suddivisi, in base al punteggio finale ottenuto, come segue:

- da 76/110 a 90/110 con il 31,4%;
- da 91/110 a 100/110 con il 34,6%;
- da 101/110 a 110/110 con il 34%.

La ragione che ha portato alla suddivisione in ampiezze diverse è riscontrabile nella poca differenza tra le percentuali presentate in tali blocchi.

Si nota infatti che ogni intervallo è caratterizzato da una percentuale di poco superiore al 30%.

In un secondo momento è stata effettuata un'altra suddivisione dell'intervallo 100/110-110/110 in ulteriori due intervalli:

- da 101/110 a 105/110 con il 14,1%;
- da 106/110 a 110/110 con il 19,9%.

In questo caso si è optato per tale specificazione per sottolineare la percentuale considerevole, quasi il 20%, di studenti meritevoli che hanno concluso il loro percorso di studi nella Facoltà di Scienze Statistiche di Padova con un voto finale compreso tra 106 e 110.

Poco meno del 20% dei laureati, più precisamente 151 soggetti, ha conseguito un punteggio finale compreso tra 106-110; di questi, 104, cioè il 13,7% dell'intero gruppo in esame si è laureato con voto finale di 110/110.

Inoltre, il dato relativo agli studenti che hanno suggellato il loro percorso di studi con il riconoscimento della lode è poco inferiore al 10%.

Tabella 2.16: *VOTO FINALE*

<i>Voto finale</i>		
	Frequenza	Percentuale
76-90	238	31,4
91-100	262	34,6
101-105	107	14,1
106-110	151	19,9
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

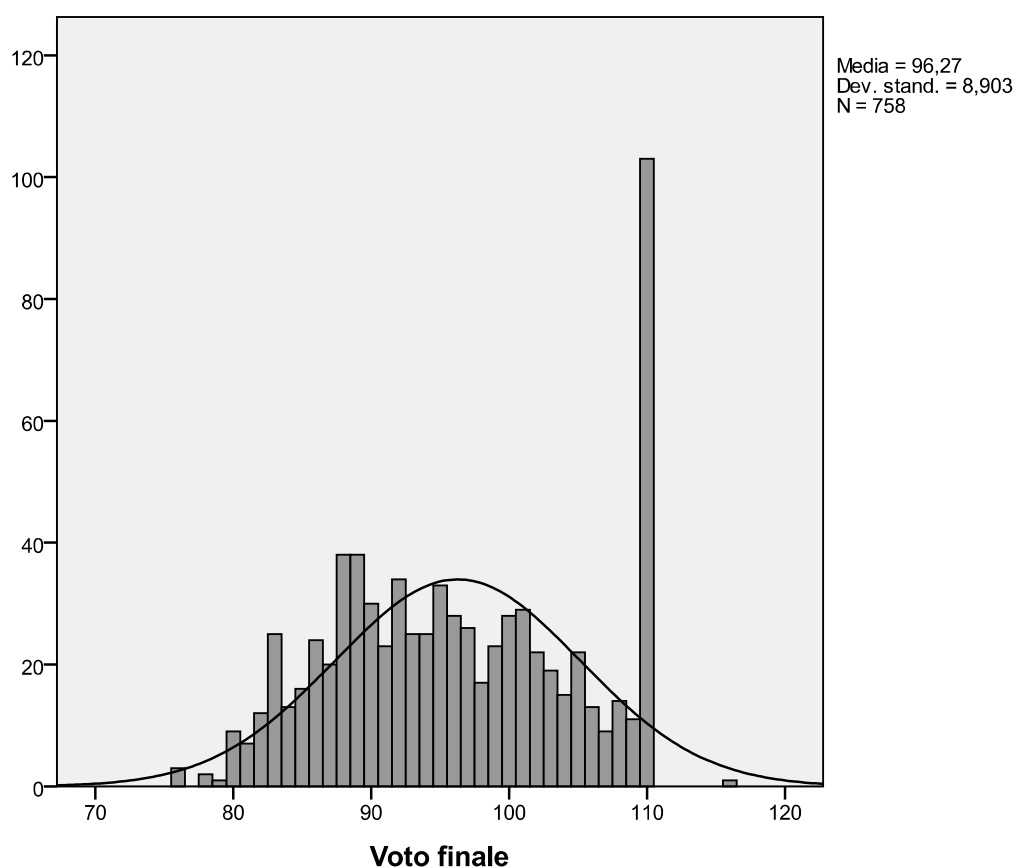


Figura 2.13: Istogramma Voto Finale

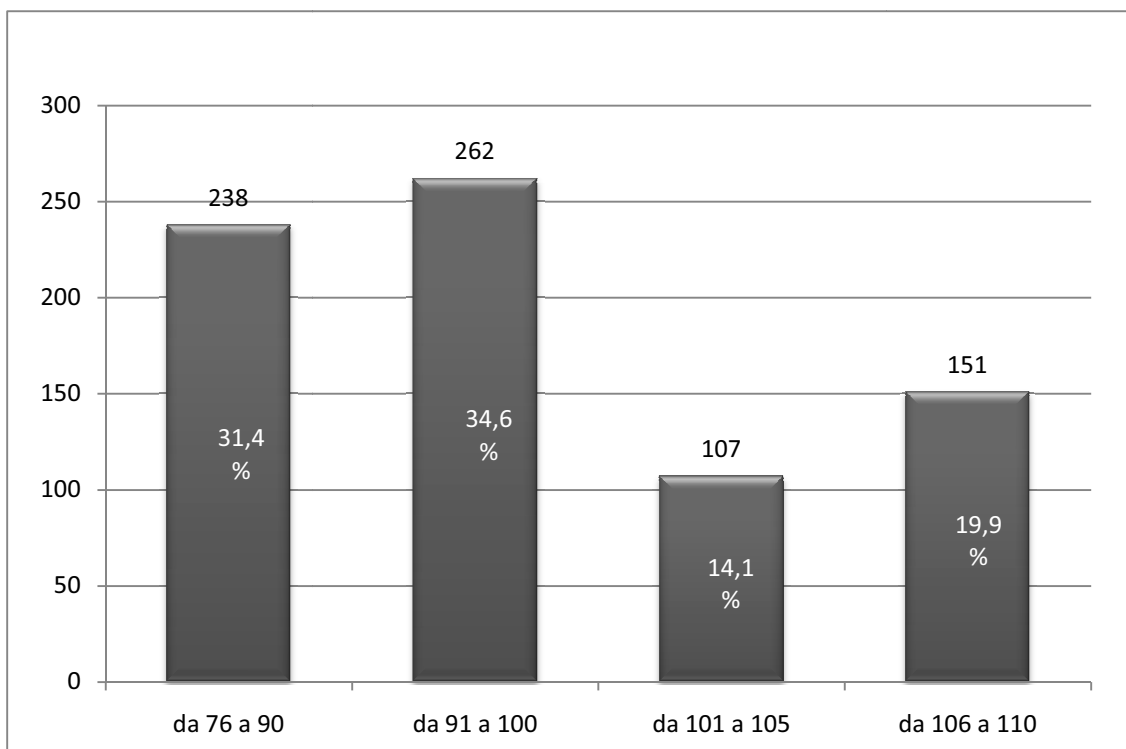


Figura 2.14: Grafico a barre Voto Finale

Tabella 2.17: *STUDENTI CHE HANNO CONSEGUITO VOTO MASSIMO*

**Studenti 110**

	Frequenza	Percentuale
<110	654	86,3
110	30	3,9
110 <i>lode</i>	74	9,8
Totale	758	100,0

(Fonte: Archivio Segreteria Studenti)

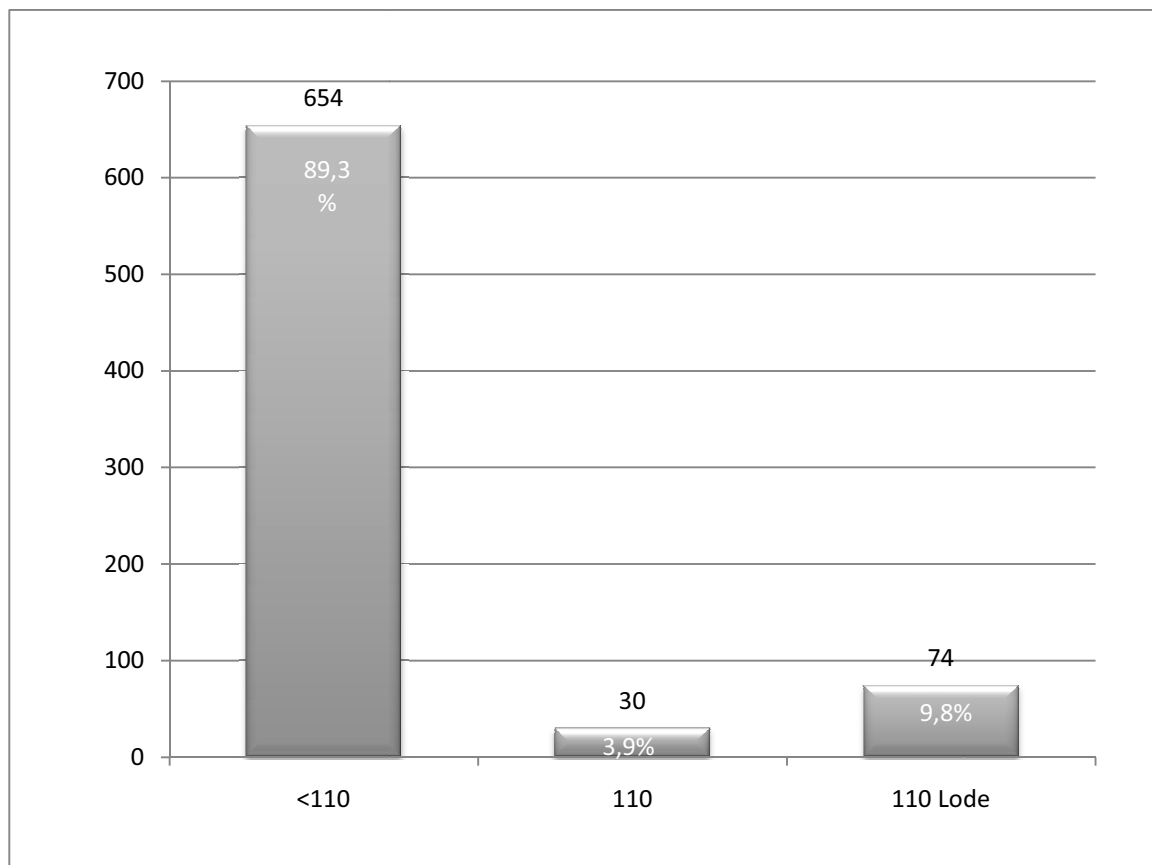


Figura 2.15: Grafico a barre frequenze studenti per Voto Finale

## **2.2. Analisi attraverso Tabelle di Contingenza**

Sembra opportuno operare una scelta delle variabili esplicative e una loro categorizzazione attraverso l'analisi di tabelle di contingenza che mostrino la distribuzione secondo la scuola di provenienza, il voto alla maturità, il luogo di provenienza, il corso di laurea prescelto e frequentato.

Le variabili appena citate rientrano nell'insieme di variabili accettate da numerosi studiosi (Trivellato e Zuliani, 1977, Gori e Rampichini, 1989) come fattori determinanti dell'esito universitario.

Inoltre verrà valutata anche la distribuzione secondo l'area scientifico-disciplinare del relatore.

Il corso di laurea, tra quelli offerti dalla Facoltà di Scienze Statistiche di Padova, che presenta il maggior numero di iscritti conseguentemente laureati, è il corso SGI con un totale di 337 studenti; di questi, quasi la metà, che corrisponde a poco più del 21% dell'intera popolazione presa in considerazione, detiene un diploma tecnico; gli studenti provenienti dagli istituti tecnici, inoltre, scelgono nel 15% dei casi il corso di Statistica e Tecnologie Informatiche che è composto per il 70% proprio dai diplomati in ambito tecnico. Tale propensione è riconducibile al fatto che la maturità tecnica fornisce delle conoscenze e basi tecnologiche che potranno poi essere sviluppate e approfondite proprio tramite il corso di laurea STI.

Il 39% degli iscritti laureati presso la suddetta Facoltà provengono dal liceo scientifico; quasi la metà, il 18% dell'intera popolazione, ha optato per l'iscrizione al corso di Statistica e Gestione delle Imprese, mentre solo il 4% di loro intraprende il piano di studi offerto dal corso STI. La seconda preferenza che caratterizza i diplomati al liceo scientifico è data dal corso in Statistica Economia e Finanza che conta tra gli iscritti proprio un 45% con maturità scientifica.

I dati relativi alle percentuali dei laureati con diplomi diversi da quelli appena citati risultano piuttosto esigui.

**Tabella 2.18: DISTRIBUZIONE DEGLI STUDENTI LAUREATI SECONDO IL CORSO DI LAUREA E LA SCUOLA SUPERIORE DI PROVENIENZA**

			lauree triennali_Corso di Laurea			
			SEF	SGI	SPS	STI
Diploma	IST_STRANIERO	Corso di Laurea	15,1%	2,4%	0,0%	4,3%
		totale	3,3%	1,1%	0,0%	,9%
	IST_TECNICO	Corso di Laurea	34,3%	48,4%	37,4%	70,7%
		totale	7,5%	21,5%	4,5%	15,3%
	LICEO	Corso di Laurea	5,4%	7,1%	14,3%	3,7%
		totale	1,2%	3,2%	1,7%	,8%
	SCIENTIFICO	Corso di Laurea	45,2%	41,5%	48,4%	21,3%
		totale	9,9%	18,5%	5,8%	4,6%
Totale		Corso di Laurea	100,0%	100,0%	100,0%	100,0%
		totale	21,9%	44,5%	12,0%	21,6%

(Fonte: Archivio Segreteria Studenti)

In media gli studenti che hanno conseguito il voto più alto alla maturità sono quelli provenienti dai licei, fatta eccezione per il liceo scientifico, con voto di diploma medio pari a 80,96/100; poco inferiore, 79,70/100, è il voto medio alla maturità dei laureati che precedentemente avevano intrapreso un percorso formativo tecnico; 78,10/100 è il dato che contraddistingue gli studenti che provengono da istituti stranieri, mentre gli studenti del liceo scientifico presentano in media il voto di diploma più basso che si attesta sul 75,38/100, (Tab.2.19). Tutte le valutazioni conseguite dagli studenti all'esame di maturità sono espresse in centesimi, inclusi i dati che esprimevano le valutazioni in sessantesimi che sono stati opportunamente trasformati in centesimi.

Tabella 2.19: STATISTICHE SUL VOTO MEDIO DI DIPLOMA DEGLI STUDENTI DIVISI PER PERCORSO SCOLASTICO PRE-UNIVERSITARIO

Voto Diploma	Diploma				Totale
	IST_STRANIERO	IST_TECNICO	LICEO	SCIENTIFICO	
Media	78,10	79,70	80,96	75,38	77,82
N	40	370	52	294	758
Deviazione std.	15,906	11,422	13,331	12,380	12,986
Mediana	77,00	79,00	81,50	72,00	77,00
Varianza	253,015	130,460	177,724	153,260	168,647
E. S. della media	2,515	0,594	1,849	0,722	0,472

(Fonte: Archivio Segreteria Studenti)

#### TEST ANOVA

	VotoDiploma * Diploma		
	Fra gruppi	Entro gruppi	Totale
Somma dei quadrati	3619.523	111904.138	115523.661
df	3	752	755
Media dei quadrati	1206.508	148.809	
F	8.108		
Sig.	.000		

(Fonte: Archivio Segreteria Studenti)

Risulta dunque interessante presentare un'ulteriore tabella (Tab.2.20) che evidenzia come gli stessi studenti provenienti dallo scientifico, che hanno in media il voto di diploma più basso, siano al contempo, insieme ai laureati che precedentemente avevano intrapreso degli studi umanistici, coloro che maggiormente riescono a terminare il loro percorso di studi in corso. Emerge inoltre un dato, che contraddistingue gli studenti provenienti dal liceo scientifico da altri licei, relativo alle percentuali di quanti, tra gli studenti appena menzionati, abbiano ricevuto come “premio” per la velocità, il massimo o il minimo dei punti.

Più precisamente, si evidenzia come all'interno della stessa categoria, studenti con formazione intrapresa nel liceo scientifico, i dati sulle percentuali di coloro che hanno ricevuto “0 punti velocità” e “6 punti velocità” siano molto simili, infatti dei laureati che provengono dal liceo scientifico, il 35,4% non ha conseguito alcun credito aggiuntivo, mentre il 35% ha ricevuto 6 crediti ulteriori, come “premio” per aver portato a termine il proprio piano di studi “in corso”. Una considerazione

analoga può essere condotta anche per i laureati che abbiano, precedentemente all'ingresso nel sistema universitario, frequentato altri tipi di licei.

Di questi infatti il 34,6% ha ottenuto “0 punti velocità”, mentre il 38,5% si è laureato nella prima sessione “in corso” accreditandosi 6 punti aggiuntivi. Si è reso opportuno sottolineare questo aspetto perché i dati relativi ai punti “velocità”, senza considerare la distribuzione per formazione pre-universitaria, mostrano come sia maggiore la percentuale dei laureati che non ha ricevuto alcun punto ulteriore, 38,3%, rispetto al dato che caratterizza quanti abbiano invece avuto un incremento di 6 punti, il 31,9%.

Il 5% degli studenti presi in esame ha concluso il ciclo di studi legato alla laurea di primo livello anticipatamente rispetto alla lunghezza naturale del corso di laurea.

Emerge anche, (*Tab.2.21a*), che tra i laureati che conseguono il titolo con il massimo dei voti, proprio coloro che arrivano dal liceo scientifico rappresentano circa il 53% degli studenti che si sono laureati con un voto di 110/110, a seguire, con il 36,5%, i laureati che hanno intrapreso un percorso di tipo tecnico nella scuola secondaria.

Il 31,7% degli studenti che si sono laureati con il massimo dei voti ha frequentato il percorso SGI, (*Tab.2.21b*), il 26% arriva dal corso SEF, mentre, approssimativamente, il 22% e il 20% dei laureati aveva svolto i propri studi rispettivamente nei corsi STI e SPS.



Tabella 2.20: PUNTI ASSEGNATI PER LA VELOCITÀ AI LAUREATI DIVISI PER DIPLOMA

			Diploma				Totale
			IST_STRANIERO	IST_TECNICO	LICEO	SCIENTIFICO	
velocità	0	% Colonna	47,5%	40,0%	34,6%	35,4%	38,3%
		% totale	2,5%	19,5%	2,4%	13,7%	38,3%
	2	% Colonna	5,0%	9,5%	7,7%	7,1%	8,3%
		% totale	,3%	4,6%	,5%	2,8%	8,3%
	3	% Colonna		0,3%			0,1%
		% totale		0,1%			0,1%
	4	% Colonna	12,5%	17,3%	13,5%	16,3%	16,4%
		% totale	,7%	8,4%	,9%	6,3%	16,4%
	6	% Colonna	30,0%	28,9%	38,5%	35,0%	31,9%
		% totale	1,6%	14,1%	2,6%	13,6%	31,9%
	7	% Colonna	5,0%	4,1%	5,8%	6,1%	5,0%
		% totale	,3%	2,0%	0,4%	2,4%	5,0%
	Totale	% Diploma	100,0%	100,0%	100,0%	100,0%	100,0%
		% totale	5,3%	48,8%	6,9%	38,8%	100,0%

(Fonte: Archivio Segreteria Studenti)

	Valore	df	Sig. asint. (2 vie)
Chi-quadrato di Pearson	9.175 <sup>a</sup>	15	0,868
Rapporto di verosimiglianza	9.562	15	.846
N. di casi validi	758		

Attraverso le seguenti tabelle, si evidenziano quali siano le principali caratteristiche degli studenti che si sono laureati con il massimo dei voti.

Tabella 2.21a: STUDENTI CON VOTO FINALE 110/110 DIVISI per DIPLOMA

		Voto Finale		Totale
		<110	110	
Diploma	IST_STRANIERO	39	1	40
	% entro Voto Finale	6,0%	1,0%	5,3%
	IST_TECNICO	332	38	370
	% entro Voto Finale	50,8%	36,5%	48,8%
	LICEO	42	10	52
	% entro Voto Finale	6,4%	9,6%	6,9%
	SCIENTIFICO	239	55	294
	% entro Voto Finale	36,5%	52,9%	38,8%
Totale		654	104	758
% entro Voto Finale		100,0%	100,0%	100,0%

(Fonte: Archivio Segreteria Studenti)

Tabella 2.21b: STUDENTI CON VOTO FINALE 110/110 DIVISI per CORSO di LAUREA

		Voto Finale		Totale
		<110	110	
Corso di Laurea	SEF	139	27	166
	% entro Riga	83,7%	16,3%	100,0%
	% entro Colonna	21,3%	26,0%	21,9%
	SGI	304	33	337
	% entro Riga	90,2%	9,8%	100,0%
	% entro Colonna	46,5%	31,7%	44,5%
	SPS	70	21	91
	% entro Riga	76,9%	23,1%	100,0%
	% entro Colonna	10,7%	20,2%	12,0%
	STI	141	23	164
	% entro Riga	86,0%	14,0%	100,0%
	% entro Colonna	21,6%	22,1%	21,6%
Totale		654	104	758
% entro Riga		86,3%	13,7%	100,0%
% entro Colonna		100,0%	100,0%	100,0%

(Fonte: Archivio Segreteria Studenti)

Il 13,7% della popolazione ha conseguito il titolo della laurea con il massimo punteggio di 110/110. Circa il 23% dei frequentanti il corso SPS ha ottenuto il massimo punteggio. Si evidenzia con la seguente tabella, (Tab.2.21c), come l'87,5% degli studenti che si sono laureati con voto pari a 110/110, abbia raggiunto tale risultato anche grazie ai punti che sono stati concessi loro come "premio" per aver portato a termine il proprio percorso di studi entro i termini previsti o addirittura in anticipo rispetto alla durata naturale del corso.

Tabella 2.21c: *STUDENTI SUDDIVISI per VOTO FINALE 110/110 e per PUNTI AGGIUDICATI per VELOCITÀ*

			Voto Finale		Totale
			<110	110	
velocità 0	Frequenze		288	2	290
	% entro Riga		99,3%	0,7%	100,0%
	% entro Colonna		44,0%	1,9%	38,3%
2	Frequenze		61	2	63
	% entro Riga		96,8%	3,2%	100,0%
	% entro Colonna		9,3%	1,9%	8,3%
3	Frequenze		1	0	1
	% entro Riga		100,0%	0,0%	100,0%
	% entro Colonna		0,2%	0,0%	0,1%
4	Frequenze		115	9	124
	% entro Riga		92,7%	7,3%	100,0%
	% entro Colonna		17,6%	8,7%	16,4%
6	Frequenze		167	75	242
	% entro Riga		69,0%	31,0%	100,0%
	% entro Colonna		25,5%	72,1%	31,9%
7	Frequenze		22	16	38
	% entro Riga		57,9%	42,1%	100,0%
	% entro Colonna		3,4%	15,4%	5,0%
Totale	Frequenze		654	104	758
	% entro Riga		86,3%	13,7%	100,0%
	% entro Colonna		100,0%	100,0%	100,0%

(Fonte: Archivio Segreteria Studenti)

Nelle tabelle a seguire, (*Tab. 2.21d, Tab.2.21e*), viene presentata la distribuzione dei laureati in base al voto di laurea pari a 110/110 o inferiore a 110/110, al Corso di Laurea frequentato e ai punti per la velocità ottenuti. L'attenzione viene posta soprattutto su coloro che hanno ottenuto un incremento ulteriore pari a 0, 6 o 7 punti. È stata adottata questa scelta perchè i dati relativi a queste categorie sono più elevati. Infatti, il numero degli studenti con 0 punti velocità è pari a 290, quello che caratterizza i laureati con 2 punti aggiuntivi si attesta sui 63 laureati e la frequenza che individua coloro che hanno conseguito un incremento di 4 punti risulta uguale a 124. Infine, gli studenti che hanno concluso il loro percorso di studi triennale “in corso” o addirittura anticipatamente, sfruttato l'opportunità di ottenere un ulteriore punteggio al fine di conseguire una maggiore valutazione finale, sono 280.

Il Corso di Laurea che presenta la percentuale maggiore di laureati con il massimo punteggio di 110/110 è il Corso di Laurea in Statistica, Popolazione e Società con il 23% dei suoi iscritti, a seguire il Corso di Laurea in Statistica, Economia e Finanza con un dato pari al 16,3%, il Corso in Statistica e Tecnologie Informatiche con il 14%, infine il Corso in Statistica e Gestione delle Imprese con il valore che individua gli studenti con voto finale massimo pari a 9.8%. Tale dato va sottolineato sia per la differenza rispetto ai dati che caratterizzano gli altri corsi, sia per il fatto che il Corso di laurea SGI è il corso, tra quelli offerti dalla Facoltà di Scienze Statistiche, con il maggior numero di iscrizioni uguale a 337. Infatti gli altri corsi di laurea presentano dati legati alle iscrizioni particolarmente distanti, il Corso SPS, caratterizzato come precedentemente accennato dalla percentuale più alta di laureati con il punteggio massimo, ha avuto 91 iscrizioni nel corso degli anni presi in esame; il Corso SEF conta 166 laureati sempre nello stesso arco temporale, mentre il Corso STI 164.

Tabella 2.21d: **STUDENTI con 6/7 PUNTI VELOCITÀ DIVISI per CORSO di LAUREA**

velocità				Voto Finale		Totale
				<110	110	
6	Corso di Laurea	SEF	Frequenze	41	19	60
			% entro Colonna	24,6%	25,3%	24,8%
	SGI	Frequenze	79	21	100	
		% entro Colonna	47,3%	28,0%	41,3%	
	SPS	Frequenze	21	17	38	
		% entro Colonna	12,6%	22,7%	15,7%	
	STI	Frequenze	26	18	44	
		% entro Colonna	15,6%	24,0%	18,2%	
	Totale		Frequenze	167	75	242
			% entro Colonna	100,0%	100,0%	100,0%
7	Corso di Laurea	SEF	Frequenze	5	6	11
			% entro Colonna	22,7%	37,5%	28,9%
	SGI	Frequenze	14	6	20	
		% entro Colonna	63,6%	37,5%	52,6%	
	SPS	Frequenze	1	2	3	
		% entro Colonna	4,5%	12,5%	7,9%	
	STI	Frequenze	2	2	4	
		% entro Colonna	9,1%	12,5%	10,5%	
	Totale		Frequenze	22	16	38
			% entro Colonna	100,0%	100,0%	100,0%

(Fonte: Archivio Segreteria Studenti)

Tabella 2.21e: **STUDENTI SUDDIVISI PER VOTO FINALE 110/110, PER CORSO DI LAUREA E PER DIPLOMA DI MATURITÀ**

Voto Finale 110		Corso di Laurea				Totale
		SEF	SGI	SPS	STI	
Diploma	IST_STRANIERO	1	0	0	0	1
	IST_TECNICO	2	4	3	1	10
	LICEO	17	19	15	4	55
	SCIENTIFICO	7	10	3	18	38
Totale		27	33	21	23	104

(Fonte: Archivio Segreteria Studenti)

**Tabella 2.21f: STUDENTI SUDDIVISI PER VOTO FINALE 110/110, PER CORSO DI LAUREA E PER PUNTI ASSEGNATI PER LA VELOCITÀ**

Voto Finale 110		Corso di Laurea				Totale
		SEF	SGI	SPS	STI	
velocità	0	0	1	0	1	2
	2	1	1	0	0	2
	4	1	4	2	2	9
	6	19	21	17	18	75
	7	6	6	2	2	16
Totale		27	33	21	23	104

(Fonte: Archivio Segreteria Studenti)

**Tabella 2.21g: STUDENTI SUDDIVISI PER VOTO FINALE e PUNTI VELOCITÀ**

			Voto finale diviso per classi				Totale
			(76-90]	(90-100]	(100-105]	(105-110]	
Velocità	0	Frequenze	186	87	15	2	290
		% riga	64,1%	30,0%	5,2%	0,7%	100,0%
		% colonna	78,2%	33,2%	14,0%	1,3%	38,3%
2	2	Frequenze	31	27	2	3	63
		% riga	49,2%	42,9%	3,2%	4,8%	100,0%
		% colonna	13,0%	10,3%	1,9%	2,0%	8,3%
3	3	Frequenze	0	1	0	0	1
		% riga	0,0%	100,0%	0,0%	0,0%	100,0%
		% colonna	0,0%	0,4%	0,0%	0,0%	0,1%
4	4	Frequenze	15	69	22	18	124
		% riga	12,1%	55,6%	17,7%	14,5%	100,0%
		% colonna	6,3%	26,3%	20,6%	11,9%	16,4%
6	6	Frequenze	6	70	58	108	242
		% riga	2,5%	28,9%	24,0%	44,6%	100,0%
		% colonna	2,5%	26,7%	54,2%	71,5%	31,9%
7	7	Frequenze	0	8	10	20	38
		% riga	0,0%	21,1%	26,3%	52,6%	100,0%
		% colonna	0,0%	3,1%	9,3%	13,2%	5,0%
Totale	Totale	Frequenze	238	262	107	151	758
		% riga	31,4%	34,6%	14,1%	19,9%	100,0%
		% colonna	100,0%	100,0%	100,0%	100,0%	100,0%

(Fonte: Archivio Segreteria Studenti)

La seguente tabella, (*Tab.2.21h*), mette in luce che tra i laureati con voto finale pari a 110, nessuno ha optato per il percorso A3 che è caratterizzato dallo stage di più lunga durata; essi hanno invece preferito nella maggior parte dei casi, più del 62% delle volte, sviluppare il progetto di relazione finale concordato e coordinato dal proprio relatore.

*Tabella 2.17h: STUDENTI SUDDIVISI PER VOTO FINALE 110/110, PER CORSO DI LAUREA E PER PERCORSO DI STAGE*

Voto Finale			Corso di Laurea				Totale
			SGI	SEF	SPS	STI	
110	Percorso	A1	8	5	6	3	22
		A2	6	4	2	5	17
		B	19	18	13	15	65
Totale			33	27	21	23	104

(Fonte: Archivio Segreteria Studenti)

Di seguito si presenta pure una tabella, (Tab.2.22), che mostra la distribuzione dei laureati secondo il corso di laurea frequentato e la provincia/posizione geografica di provenienza.

*Tabella 2.22: DISTRIBUZIONE DEGLI STUDENTI SECONDO LA PROVINCIA DI PROVENIENZA E IL CORSO DI LAUREA FREQUENTATO*

		Corso di Laurea				Totale
		SEF	SGI	SPS	STI	
PROVENIENZA	BELLUNO	1	4	2	4	11
	PADOVA	68	184	41	79	372
	ROVIGO	4	5	6	2	17
	TREVISO	16	24	6	20	66
	VENEZIA	16	35	16	21	88
	VERONA	2	14	3	4	23
	VICENZA	14	45	11	19	89
	NORD	16	11	5	6	38
	CENTRO	1	1	1	0	3
	SUD	3	4	0	2	9
	ESTERO	25	8	0	7	40
	MANCANTI	0	2	0	0	2
Totale		166	337	91	164	758

(Fonte: Archivio Segreteria Studenti)

Infine, risulta interessante proporre una tabella, (Tab.2.23), che mostri la tendenza da parte degli studenti laureati a scegliere maggiormente, come relatori, docenti appartenenti ad una determinata area scientifico disciplinare rispetto ad altri.

I settori didattici proposti sono 7; emerge subito che i docenti che assumono il ruolo di relatore si dividono principalmente tra le aree disciplinari in ambito economico, infatti tra tutti spicca il dato relativo ai casi in cui il relatore è stato un docente del settore economico-aziendale, 25%. Nel 14% dei casi gli studenti si sono rivolti a relatori che appartengono pure essi alla sfera economica ma con indirizzo politico. Altro dato rilevante è quello che caratterizza i docenti del campo statistico-



economico, che sono stati interpellati per coordinare il progetto della relazione finale nel 20% delle occasioni. Tali osservazioni sottolineano che i laureati scelgono nel 59% delle volte sviluppare e approfondire gli argomenti propri del contesto economico.

Il valore che caratterizza la scelta da parte degli studenti di appoggiarsi agli insegnanti propri del settore demografico (DSS) si attesta sul 16%; quello che invece individua la preferenza verso un lavoro di tesi finale prettamente statistico è del 14%.

**Tabella 2.23: DISTRIBUZIONE DEGLI STUDENTI SECONDO L'AREA SCIENTIFICO DISCIPLINARE DI APPARTENENZA DEL PROPRIO RELATORE**

	Corso di Laurea				Totale
	SEF	SGI	SPS	STI	
Area_Relatore DSS	2	15	83	19	119
ECO AZ	19	153	0	18	190
ECO POL	79	24	1	4	108
INF	0	18	1	44	63
MAT	10	2	0	8	20
STAT	12	27	4	63	106
STATEC	44	98	2	8	152
Totale	166	337	91	164	758

(Fonte: Archivio Segreteria Studenti)



## *CAPITOLO TERZO*

---

# *DETERMINANTI DEGLI ESITI: STUDIO ATTRAVERSO ANALISI DI SOPRAVVIVENZA*

### **3.1. Considerazioni preliminari**

L'indagine di partenza consiste nello stabilire se professori appartenenti a determinate aree scientifico disciplinari, rispetto ad altre, proponano tendenzialmente di sviluppare ai propri laureandi lavori di approfondimento più semplici o con garanzia di riuscita maggiore.

La variabile “Incremento” che indica il punteggio assegnato alla prova finale che può attestarsi tra gli 0 e 6 punti è una variabile positiva e di tipo quantitativo.

Inoltre durante la fase di analisi preliminare è emerso che 104 studenti, per la precisione il 13,7% della popolazione di riferimento, ha concluso il ciclo universitario triennale con il punteggio massimo di 110/110, inoltre di queste 104 unità, il 71%, 74 soggetti, ha ricevuto il riconoscimento aggiuntivo della lode.

Considerate dunque le variabili che sono state menzionate, valutando il background scolastico precedente e l'andamento degli studi universitari, è emerso che 82 laureati, sui 104 che hanno conseguito il punteggio massimo, hanno ottenuto un punteggio di tesi che forse non corrisponde all'effettivo valore della relazione finale.

L'idea che gli elaborati di alcuni laureati con 110 abbiano ricevuto un punteggio che non corrisponde all'effettiva valutazione nasce anche dall'osservazione di alcune caratteristiche dei laureati che hanno raggiunto il massimo dei punti. Infatti, gli studenti che conseguono il titolo di laurea con il voto più alto si presentano alla prova finale con un punteggio di partenza che si attesta mediamente sui 102,45/110, mentre i colleghi che si laureano con una valutazione minore a 110 sono caratterizzati da un voto di partenza decisamente inferiore, pari a circa 87,5 punti su 110. Inoltre, mentre i laureati con 110 ricevono in media circa 6 punti aggiuntivi per la velocità, gli altri studenti se ne aggiudicano 2,7, ma al contempo sono proprio gli studenti con le valutazioni finali inferiori a 110 a ricevere mediamente più punti per l'elaborato finale. Essi infatti ricevono per la prova finale in media circa 5,5 punti, mentre gli elaborati dei laureati con 110 ottengono poco più di 2 punti dei sei conseguibili.

Pertanto, sulla base di tali considerazioni, risulta anomalo che proprio gli studenti con il voto medio più alto e con il voto massimo di laurea siano coloro che si aggiudicano il minor punteggio per la prova finale.

*Tabella.3.1: LAUREATI con VOTO=110 e LAUREATI con VOTO<110*

<b>Laureati Voto110</b>				
	Minimo	Massimo	Media	Deviazione std.
Media Voti	98	109	102,45	2,961
Velocità	2	7	5,96	0,867
Incremento	0	5	2,29	1,940

(Fonte: Archivio Segreteria Studenti)

### *Laureati Voto<110*

	Minimo	Massimo	Media	Deviazione std.
Media Voti	74	104	87,49	5,961
Velocità	0	7	2,74	2,655
Incremento	0	8	4,49	1,327

(Fonte: Archivio Segreteria Studenti)

Sembra logico pensare che i dati relativi alla valutazione degli elaborati dei laureati con 110 non presentano l'effettivo valore della prova finale.

Tali dati forniscono dunque delle informazioni incomplete che non permettono di sapere quanto realmente valgano quei determinati progetti di tesi. La scarsità di informazioni dei dati succitati lascia spazio a considerazioni inerenti la possibilità da parte degli studenti laureatisi con 110/110 di conseguire nella prova finale un numero maggiore di punti rispetto a quello che hanno ottenuto.

Sulla base delle considerazioni fino a qui riportate, si è ritenuto interessante utilizzare strumenti di analisi di sopravvivenza, adattandoli al problema e ai dati a disposizione.

## **3.2. Nozioni e terminologia statistica**

L'idea consiste infatti nell'adattare il concetto di tempo di sopravvivenza alla variabile Incremento, come viene di seguito chiarito, pur non essendo quest'ultima caratterizzata da intervalli temporali.

Prima di entrare in merito allo studio di questi dati si presentano le principali nozioni relative all'analisi dei dati di sopravvivenza utilizzate per arrivare alle conclusioni presentate a fine capitolo (vedi Clark, et al.,2003).

Nel seguente paragrafo si presentano i principali requisiti chiave dell'analisi di sopravvivenza che sono:

- la determinazione della variabile risposta;
- il tipo di censura utilizzato per i dati in esame;

### **3.2.1. La variabile risposta**

La variabile risposta risulta essere la variabile Incremento, cioè quella variabile che descrive i punti che uno studente può conseguire con il progetto di elaborato finale.

È una variabile positiva e di tipo quantitativo. Normalmente il metodo della sopravvivenza prevede la rilevazione del numero e della proporzione di casi validi, del numero di eventi terminali, e del numero di casi che sono “usciti” dall’indagine che si dicono censurati.

In questo contesto insolito per l’analisi di durata, viene presentata la suddivisione della variabile Incremento in intervalli di valore unitario che descrivono il numero di punti che ciascun studente laureato ha conseguito con il progetto di tesi.

Nella prova finale attualmente si possono conseguire da 0 a 6 punti. Alcuni studenti però hanno ricevuto un punteggio maggiore rispetto a quello appena indicato per il loro elaborato. Infatti 9 laureati hanno ricevuto una valutazione pari a 7 punti, mentre l’elaborato di uno studente è stato addirittura valutato con 8 punti. È consentita l’assegnazione di punti ulteriori su proposta scritta e motivata da parte del controrelatore quando il punteggio complessivo del candidato è pari a 109,110,111 punti per il conseguimento della lode, ma solo due dei dieci studenti che hanno ricevuto un punteggio così alto hanno concluso con il punteggio di 110 e lode, altri due candidati hanno conseguito il titolo con il punteggio di 110, mentre i rimanenti sei, nonostante la valutazione particolarmente alta dell’elaborato finale, hanno terminato con un punteggio finale inferiore a 110. Pertanto in questi casi, si può presumere che tali valutazioni generose derivino essenzialmente dalla predisposizione da parte di alcuni docenti nel concedere più punti. Per questa ragione nella tabella di sopravvivenza che di seguito viene presentata sono presenti pure i valori, tra i punti di tesi, 7 e 8.

Se di norma nell’analisi di durata la variabile oggetto di studio è il tempo intercorrente tra evento iniziale e evento finale, in questo specifico caso, la variabile

positiva e quantitativa è Incremento che descrive il numero di punti assegnati per il lavoro di tesi. Con evento iniziale si definisce il conseguire almeno 0 punti per la tesi da parte di uno studente, mentre con evento finale viene indicato un qualsiasi punteggio ottenuto dalla valutazione della prova finale.

### **3.2.2. La censura**

L'analisi di sopravvivenza è caratterizzata normalmente dalla presenza di dati censurati, cioè osservazioni di cui si hanno informazioni parziali, ciò accade per la possibilità che una certa unità statistica venga, per così dire, “persa” prima dell'evento finale.

Può accadere infatti che il tempo di sopravvivenza di alcuni soggetti sia superiore alla durata dell'osservazione, in altre parole che in tali soggetti non si verifichi l'evento terminale nell'arco di tempo intercorrente tra il loro ingresso nello studio e la fine dello studio stesso. In altri casi invece può accadere di perdere traccia durante il periodo di studio senza che sia stato osservato l'evento finale, ciò comporta la presenza di mancate osservazioni e conseguentemente di informazioni incomplete che portano a considerazioni sulla probabilità per un'unità, oggetto di studio, di essere “viva” all'osservazione nel tempo  $t$ , considerato che era viva al tempo  $t-1$ . Quindi la censura avviene proprio per il fatto che non si conosce il tempo di “sopravvivenza” esatto.

Gli schemi di censura più ricorrenti, in pratica, possono essere raggruppati sostanzialmente in tre classi (vedi Klein e Moeschberger, 2003):

- Censura 1° tipo: i soggetti sono osservati per un periodo di tempo fissato. Alla fine dello studio i soggetti che non presentano fallimento risultano censurati;



- Censura 2° tipo: il numero totale di fallimenti è stabilito a priori. La lunghezza dello studio quindi non risulta fissata;
- Censura casuale: il totale del periodo di osservazione è fissato, ma i soggetti entrano in studio in tempi differenti. Alcuni individui falliscono, altri individui risultano persi, altri ancora non presentano fallimento alla fine dello studio .

Nello studio condotto si è in presenza di una censura di primo tipo.

### **3.3. Analisi di sopravvivenza**

Le tecniche di analisi della sopravvivenza sono molto flessibili e perciò applicabili ad una vasta tipologia di problematiche e ambiti. Queste considerazioni hanno permesso di adattare questo modello di analisi ai dati oggetto di studio nel modo che di seguito viene riportato.

Nel presente lavoro e nei dati in esame si riscontrano delle censure che caratterizzano il punteggio per l'elaborato finale di quegli studenti che si sono laureati con il massimo dei voti, ma che non hanno conseguito il massimo punteggio nella prova finale.

Il concetto alla base di tale procedimento può essere così spiegato: considerati tutti gli studenti che hanno conseguito un voto finale di 110/110, sono stati presi coloro che hanno ricevuto per la loro relazione finale un punteggio pari al massimo a 5 punti. Infatti, per gli studenti con un voto di partenza maggiore a 104 punti, si sa che la loro tesi è stata valutata almeno "*i*" *punti Incremento*, (con *i*= numero di punti che descrivono la variabile Incremento), ma non si sa quanto l'elaborato sia stato valutato effettivamente, assegnando quindi un limite inferiore anziché un valore assoluto. Si è in presenza quindi di informazioni parziali relative alle valutazioni delle prove finali di tali studenti. Pertanto sono state considerate censure i voti che caratterizzavano i lavori di tesi di tutti i laureati 110 che si siano presentati con un voto di partenza maggiore a 104.

### 3.3.1. Funzione di sopravvivenza

La *funzione di sopravvivenza*  $S(i)$ , dove  $i$  indica il valore assunto dalla variabile di interesse Incremento, fornisce la probabilità che uno studente consegua più punti rispetto a quanti specificati in  $i$ , ovvero  $S(i) = Pr(I > i)$ .

Questa funzione risulta essere di fondamentale importanza poichè ottenendo probabilità di sopravvivenza per differenti valori di  $i$ , si possono avere informazioni riassuntive sui dati di sopravvivenza.

Dalla funzione di sopravvivenza deriva la *funzione di distribuzione* della variabile Incremento che è definita come segue  $F(i) = Pr(I \leq i) = 1 - S(i)$ , mentre la derivata della funzione  $F(i)$ ,  $f(i) = \frac{d}{di} F(i)$ , caratterizza la funzione di densità della variabile Incremento.

### 3.3.2. Funzione di rischio

La funzione di rischio è  $\lambda(i) = f(i)/S(i)$ , dove  $i$  indica il valore assunto dalla variabile di interesse Incremento.

Tale funzione rappresenta la probabilità di transizione condizionata alla “sopravvivenza”, o più nello specifico, descrive il “rischio istantaneo” che l’evento di interesse si verifichi in corrispondenza di un determinato esito finale, dato che non si è verificato in corrispondenza del precedente.

La funzione di rischio  $\lambda(i)$  è così espressa:

$$\lambda(i) = \lim_{\Delta i \rightarrow 0} \frac{P(i \leq I < i + \Delta i \mid I \geq i)}{\Delta i},$$

con  $\Delta i$  che denota un intervallo di Incremento piccolo.

In termini pratici la funzione di rischio  $\lambda(i)$  dà un potenziale istantaneo (vedi Kleinbaum e Klein, 2005) per unità di incremento di un determinato evento che

avviene, dato che non è avvenuto in corrispondenza del precedente. A differenza della funzione di sopravvivenza, la quale si focalizza sul non fallimento, la funzione di rischio si focalizza sul fallimento, che è l'evento accaduto.

### **3.4. Stimatori**

L'analisi della sopravvivenza può essere condotta con differenti strumenti. La funzione di sopravvivenza presentata viene studiata con due metodi: la tavola di sopravvivenza del metodo attuariale e lo stimatore di Kaplan Meier. Una differenza tra i due metodi consiste nel fatto che lo stimatore di Kaplan Meier viene utilizzato quando si conosce il dato esatto in corrispondenza del quale avviene l'evento terminale, mentre nel metodo attuariale si conosce l'intervallo relativo a tale dato. Nel presente lavoro questa differenza non si presenta in quanto gli intervalli sono unitari.

#### **3.4.1. La tavola di sopravvivenza**

La tavola di sopravvivenza, che di seguito viene presentata, permette di stimare in maniera non parametrica, considerando anche le osservazioni censurate, la funzione di sopravvivenza della variabile incremento.

Questa tabella può essere pensata come una tabella di distribuzione della frequenza.

Il metodo di calcolo delle tavole di sopravvivenza comporta, come già accennato, la suddivisione dell'asse dell'unità di misura della variabile risposta, cioè della variabile Incremento, in  $J+1$  intervalli  $H_j = [t_{j-1}, t_j)$  con  $j=1,2, \dots, 7$ ,  $t_0=0$ ,  $t_{j+1}=8$ , dove 8 è il limite superiore dei punti di tesi assegnati. Gli intervalli hanno ampiezza unitaria e si rilevano, per ciascuna unità statistica che partecipa allo studio, i corrispondenti punti assegnati per il lavoro della tesina finale e si contano quanti soggetti presentano l'evento terminale, che in precedenza abbiamo definito come il conseguimento di qualsiasi punteggio, o l'osservazione censurata in ciascuno degli intervalli.

Più specificatamente gli elementi necessari per la costruzione della tavola di sopravvivenza sono:

- $t_{mj}$  = punto medio dell'intervallo  $H_j = [t_{j-1}, t_j)$ , con  $j = 1, 2, \dots, 8$ ;
- $h_j = t_j - t_{j-1}$ , ampiezza dell'intervallo  $H_j$ , con  $j = 1, 2, \dots, 8$ ;
- $n_j$  = numero di entrati nell'intervallo: rappresenta il numero di studenti laureati il cui elaborato finale ha sicuramente ottenuto almeno il punteggio indicato dall'intervallo;
- $d_j$  = numero di rimossi nell'intervallo: è il numero di eventi terminali; indica il numero di studenti che hanno conseguito con la prova finale il punteggio indicato dal numero dell'intervallo;
- $w_j$  = numero di censure: cioè il numero di studenti che ha ottenuto per il proprio progetto di tesi un punteggio pari al numero di punti indicato dall'intervallo, ma che probabilmente avrebbero potuto conseguire con il proprio lavoro un punteggio superiore.

A partire da questi numeri e proporzioni, si possono stimare anche altre statistiche importanti che in questo contesto possono essere così definite:

- *Numero di casi a "rischio"*: rappresenta la differenza tra il numero di studenti che sono entrati nell'intervallo in esame e la metà del numero dei casi rimossi o censurati nell'intervallo stesso perchè si fa riferimento al punto medio dell'intervallo;
- *Proporzione di terminazione*: stimata come il rapporto fra il numero di eventi terminali nell'intervallo e il numero di casi a rischio;
- *Proporzione di sopravvivenza*: stima complementare della proporzione di terminazione;
- $S(i)$  = *Funzione di sopravvivenza* (o *Proporzione di sopravvivenza cumulata*): poiché le probabilità di sopravvivenza si assumono indipendenti nei vari intervalli, questa probabilità viene calcolata moltiplicando le probabilità di sopravvivenza di tutti gli intervalli precedenti.

Viene di seguito presentata la Tavola di sopravvivenza costruita sui dati relativi gli studenti oggetto di studio.

*Tabella 3.3: TAVOLA DI SOPRAVVIVENZA PER I LAUREATI SECONDO PUNTI PER TESI*

Punti di tesi	N. di entrati nell'intervallo	N. di rimossi nell'intervallo	N. di casi a rischio	N. di eventi terminali	Proporzione di terminazione	Proporzioni e di sopravvivenza	Prop. di sopravvivenza cumulata	Densità di probabilità	Err. stand. di densità di probabilità	Tso di rischio
0	758	27	744,50	2	0,00	1,00	1,00	0,003	0,002	0,00
1	729	4	727,00	13	0,02	0,98	0,98	0,018	0,005	0,02
2	712	13	705,50	49	0,07	0,93	0,91	0,068	0,009	0,07
3	650	9	645,50	100	0,15	0,85	0,77	0,141	0,013	0,17
4	541	14	534,00	168	0,31	0,69	0,53	0,242	0,016	0,37
5	359	15	351,50	197	0,56	0,44	0,23	0,296	0,018	0,78
6	147	0	147,00	137	0,93	0,07	0,02	0,216	0,016	1,75
7	10	0	10,000	9	0,90	0,10	0,00	0,014	0,005	1,64
8	1	0	1,000	1	1,00	0,00	0,00	0,000	0,000	0,00

(Fonte: Archivio Segreteria Studenti)

### 3.4.2. Metodo di Kaplan Meier

Attraverso le tavole attuariali e il metodo non parametrico di Kaplan-Meier è possibile descrivere le caratteristiche del processo in esame senza fare ipotesi circa la distribuzione degli eventi e tenendo conto dei dati di sopravvivenza censurati che descrivono il non verificarsi dell'evento.

A differenza della costruzione della tabella di sopravvivenza presentata precedentemente, nella quale il numero e la dimensione degli intervalli vengono stabiliti a priori dal ricercatore, è possibile stimare la funzione di sopravvivenza direttamente dall'ordinamento stesso dei dati di accadimento.

Tale metodo di calcolo delle tavole di sopravvivenza con cui viene eseguita la stima delle probabilità di sopravvivenza e delle relative curve di sopravvivenza avviene in ogni punto in cui si verifica un evento e non in base alla suddivisione in intervalli di incremento uguali e predefiniti.

Supponiamo che i dati di sopravvivenza, considerate anche le osservazioni censurate, di un certo gruppo di  $n$  soggetti (in ordine crescente) siano  $t_1, t_2, \dots, t_n$ .

Moltiplicando le probabilità di sopravvivenza per ogni singola osservazione è possibile stimare la funzione di sopravvivenza come segue:

$$S(i) = \frac{(n_1 - d_1)}{n_1} \times \frac{(n_2 - d_2)}{n_2} \times \dots \times \frac{(n_p - d_p)}{n_p},$$

dove  $i_j$  è il dato di sopravvivenza più grande che sia minore o uguale a  $i$ ,  $n_j$  è il numero di studenti che si presentano fino al dato  $i_j$ , quindi fino al  $j$ -esimo dato di sopravvivenza e  $d_j$ , come è stato spiegato in precedenza, denota il numero di eventi terminali, o meglio indica il numero degli studenti che hanno conseguito con la

prova finale il punteggio indicato dal dato  $i_j$  ( con  $j= 1,2 ,..., p$  ). Per le osservazioni censurate  $d_j= 0$  .

Tale metodo consente di stimare  $S(i)$  come segue:

$$S(i) = \prod_{i_j \leq i} \left( 1 - \frac{d_j}{n_j} \right).$$

Tale stimatore è anche conosciuto come stimatore di Kaplan-Meier o del prodotto limite.

Come è stato anticipato, il vantaggio dello stimatore di Kaplan Meier per l'analisi dei dati di sopravvivenza sta nel fatto che le stime risultanti non dipendono dal raggruppamento di dati in un certo numero di intervalli. In pratica i due metodi risultano identici se gli intervalli della tavola di sopravvivenza, che è stata presentata nel paragrafo precedente, contengono al più un'osservazione.

Di seguito vengono presentati i dati elaborati tramite l'analisi di sopravvivenza eseguita, suddivisi per i vari corsi di laurea e per settori disciplinari dei docenti relatori.

*Tabella 3.4.*

***Riepilogo dell'elaborazione dei casi***

Corso di Laurea	N totale	N. di eventi	Censure	
			N	Percentuale
SGI	337	312	25	7,4%
SEF	166	144	22	13,3%
SPS	91	72	19	20,9%
STI	164	148	16	9,8%
Globale	758	676	82	10,8%

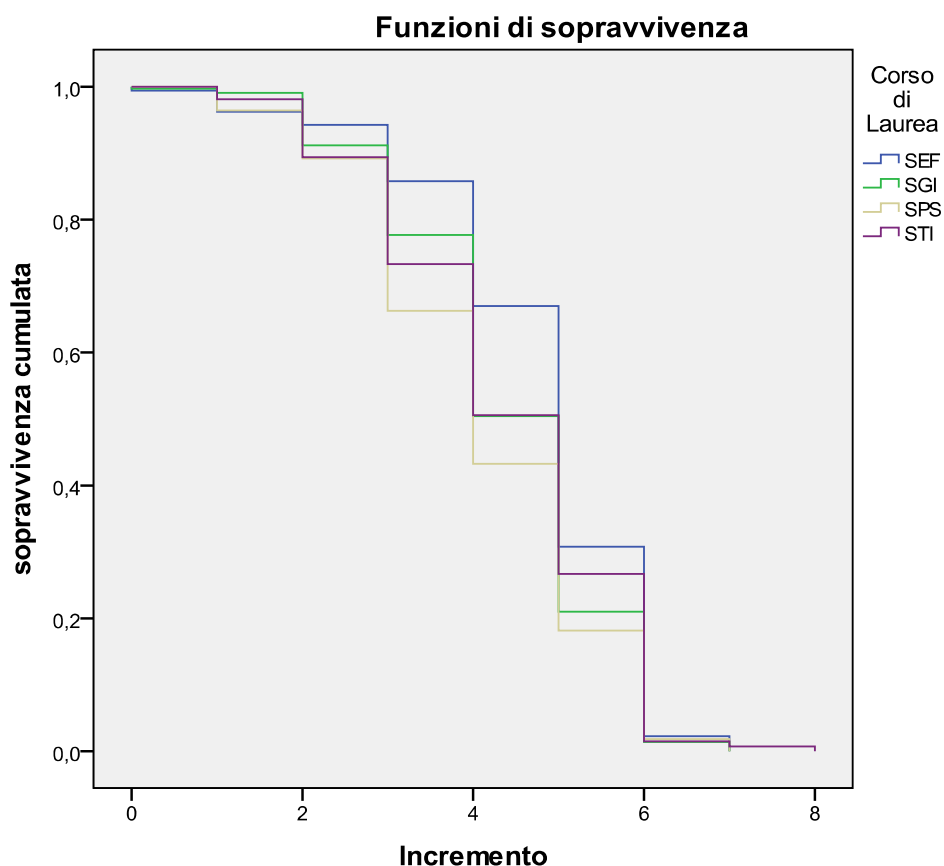


Figura 3.1: Stima curva sopravvivenza metodo Kaplan Meier Corsi di Laurea

Per verificare l'uguaglianza o meno delle curve di sopravvivenza relative alle diverse categorie del fattore preso in esame, nello specifico il fattore è il Corso di Laurea frequentato dagli studenti laureati, è stato utilizzato il *Log-Rank Test*.

L'ipotesi nulla  $H_0$  che il Log-Rank Test consente di verificare è l'uguaglianza tra le varie funzioni di sopravvivenza dei vari gruppi messi a confronto.

***Livello di significatività del Log-Rank Test***

	Chi-quadrato	df	Sig
Log Rank (Mantel-Cox)	12,954	3	0,005



L'ipotesi nulla sull'uguaglianza delle curve di sopravvivenza per i diversi Corsi di Laurea viene rifiutata, il *p-value* è altamente significativo, pertanto le curve di sopravvivenza dei vari corsi sono significativamente diverse e quindi le valutazioni della prova finale differiscono anche sulla base del Corso di laurea che è stato frequentato.

Tabella 3.5.

***Riepilogo dell'elaborazione dei casi***

Area_rel	N totale	N. di eventi	Censure	
			N	Percentuale
DSS	119	103	16	13,4%
ECO	299	268	31	10,4%
INF	61	60	1	1,6%
ING	2	1	1	50,0%
MAT	20	17	3	15,0%
STAT	105	89	16	15,2%
STATEC	152	138	14	9,2%
Globale	758	676	82	10,8%

(Fonte: Archivio Segreteria Studenti)

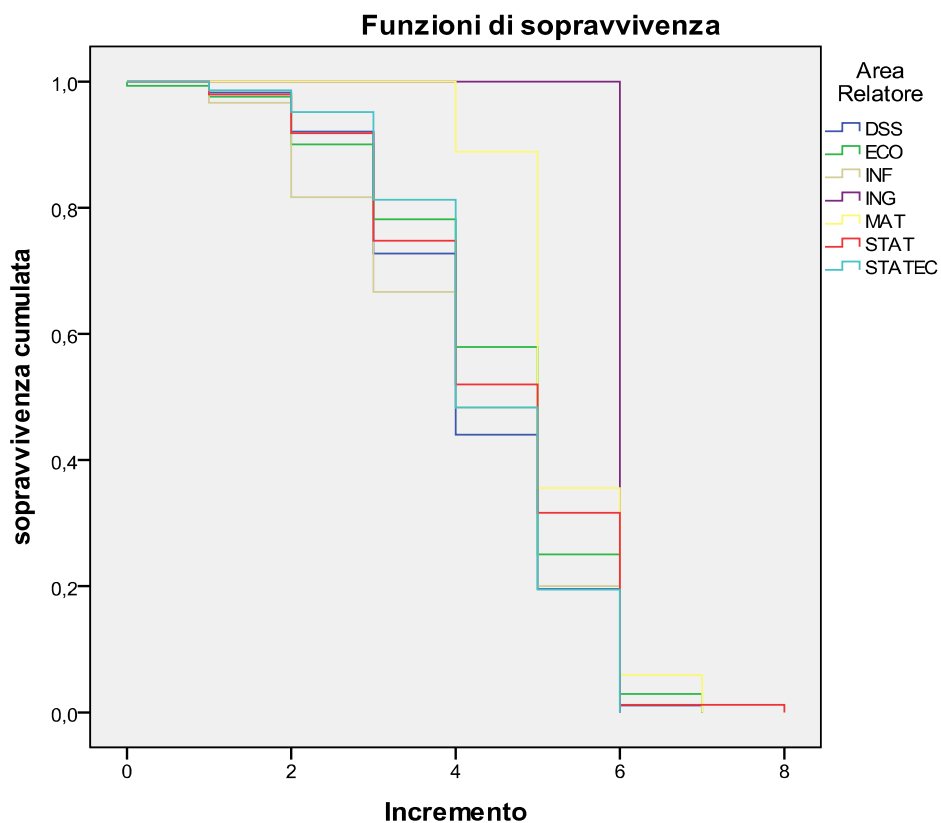


Figura 3.2: Stima curva sopravvivenza metodo Kaplan Meier Area Relatore

*Livello di significatività del Log-Rank Test*

	Chi-quadrato	df	Sig
Log Rank (Mantel-Cox)	15,960	6	0,014

Anche nel caso del confronto tra i settori disciplinari dei docenti, l'ipotesi nulla sull'uguaglianza tra le differenti curve di sopravvivenza delle relative aree viene rifiutata. La valutazione della prova finale cambia in modo significativo sulla base dell'area scientifico disciplinare di appartenenza del proprio docente relatore.

# *CAPITOLO QUARTO*

---

## *DETERMINANTI DEGLI ESITI: UN APPROFONDIMENTO CON IL MODELLO DI COX*

### **4.1. Il modello da usare**

Sulla base delle analisi fino a qui presentate, si è condotto uno studio su come e con quale peso le diverse variabili intervengono sull'esito finale.

Nei grafici che presentano le diverse funzioni di sopravvivenza, la popolazione è stata suddivisa sulla base dei corsi di laurea frequentati dagli studenti e in base all'area disciplinare di appartenenza dei relatori, in modo da poter confrontare le rispettive curve di sopravvivenza.

Il predittore scelto è pertanto la variabile Corso di Laurea, ma esso non fornisce alcuna informazione sul peso che assume la regola con la quale sono stati divisi i gruppi, più precisamente non quantifica l'influenza della variabile Corso di Laurea sulla "sopravvivenza", ossia sulla probabilità di conseguire il maggior punteggio per il proprio elaborato finale. Analoga considerazione viene fatta per il predittore Area Relatore. Dall'osservazione delle curve di sopravvivenza si evince che Corso di Laurea e Area Relatore intervengono sul punteggio che uno studente può conseguire per il proprio elaborato finale ma non si sa in quale misura.

Risulta opportuno costruire un modello che permetta di cogliere l'effetto di un determinato fattore al netto di tutte le altre variabili considerate.

Per ogni singolo studente verrà quindi calcolata la probabilità di conseguire un determinato punteggio e la probabilità di ottenere un punteggio superiore, o più

precisamente di passare da una valutazione di “n” punti a “n + 1”, distinguendo su tale *rischio di transizione* gli effetti contemporanei di differenti fattori, quali il valore assunto dalla variabile Incremento e alcune altre variabili esplicative. Un metodo tradizionale di stimare tali effetti prevede l’utilizzo di modelli di durata (Cox e Oakes, 1984, Lancaster,1990), specificati attraverso una funzione di rischio. Modelli che soddisfano le esigenze sopra elencate sono i modelli semiparametrici. In tali modelli non viene fatta alcuna ipotesi sulla funzione di sopravvivenza e quindi sul rischio. Si specifica solo la forma funzionale per valutare l’influenza delle covariate sui tempi di sopravvivenza e la forma dell’andamento del rischio viene lasciata non specificata.

Il modello più classico, e maggiormente usato, risulta essere il modello semiparametrico a rischi proporzionali di Cox.

## 4.2. Il modello a rischi proporzionali

La *funzione di rischio* del modello semiparametrico a rischi proporzionali è

$$\lambda(i) = f(i)/S(i),$$

dove  $i$  indica il valore della variabile di interesse Incremento,  $S(i) = Pr(I > i)$  la *funzione di sopravvivenza* e  $f(i) = \frac{d}{di} F(i)$ , caratterizza la funzione di densità. La funzione di rischio rappresenta la probabilità istantanea di transizione condizionata alla “sopravvivenza”, o, più nello specifico, descrive il “rischio istantaneo” che l’evento di interesse si verifichi in corrispondenza di un determinato esito finale.

Nel presente studio, l’evento di interesse è rappresentato dal conseguimento di un determinato punteggio per l’elaborato finale, rappresentato dalla variabile Incremento, e per ciascuno dei suoi possibili valori verrà calcolata la probabilità di

sopravvivenza, ossia la probabilità di assumere un dato valore condizionata al fatto che non ha assunto nessuno dei valori inferiori.

Il modello di Cox è un modello di regressione semiparametrico con funzione di rischio parzialmente specificata, sviluppato proprio per poter analizzare il peso di determinate variabili sulla sopravvivenza.

In presenza di un determinato numero di queste variabili esplicative la funzione di rischio si arricchisce di altrettanti fattori di rischio, o covariate, che vanno a costituire il vettore  $z$  delle covariate, dove  $z$  rappresenta un insieme di covariate  $(z_1, z_2, \dots, z_p)$ . Infatti una tipica assunzione prevede che tali variabili agiscano moltiplicativamente su una *funzione di rischio base*  $\lambda_B(i)$  :

$$\lambda(i; z) = \lambda_B(i) \varphi(z).$$

La forma di  $\lambda_B(i)$  assume particolare importanza se si è interessati alla *dipendenza dal punteggio conseguibile*, ossia alla variazione del rischio di transizione per effetto dell' incremento da un punteggio all'altro.

La stima della funzione  $\varphi(z)$ , specificata solitamente in forma pienamente parametrica, fornisce invece gli effetti delle singole variabili esplicative come fattori di scala sulla funzione di rischio, la cui forma è definita esclusivamente da  $\lambda_B(i)$ .

Infatti  $\varphi(z)$  è l'espressione esponenziale della somma lineare dei  $\beta_i z_i$ , dove la somma è sulle  $p$  variabili esplicative, ossia

$$\varphi(z) = \exp \left( \sum_{i=1}^p \beta_i z_i \right).$$

La trasformata esponenziale viene invocata per assicurare che il modello stimato possa sempre dare delle stime dei rischi non negative. Questo modello impone l'assunzione di rischio proporzionale, ovvero che il rapporto di rischio sia costante o equivalentemente che il rischio per un individuo sia proporzionale al rischio di ogni altro individuo, dove la proporzionalità è indipendente dall'incremento.

Infatti se si rapportano due rischi, si ha quanto segue

$$\widehat{HR} = \frac{\lambda(i; z^*)}{\lambda(i; z)} = \frac{\lambda_B(i) \exp(\sum_i \beta_i z_i^*)}{\lambda_B(i) \exp(\sum_i \beta_i z_i)} = \exp[\sum_i \beta_i (z_i^* - z_i)],$$

non dipende dall'effetto dell'incremento dei punti.

La quantità  $\widehat{HR}$  viene chiamata *hazard ratio* (Kleinbaum e Klein, 2005).

### 4.3. Stati di destinazione finale e insieme di covariate

Come spiegato nel capitolo precedente, l'analisi della sopravvivenza permette di studiare i dati relativi ad una variabile casuale, positiva e di tipo quantitativo, solitamente il tempo intercorrente fra un evento iniziale ed una certa tipologia di eventi finali.

Nel caso in esame, risulta più appropriato riferirsi a stati di destinazione finale piuttosto che a uno stato di sopravvivenza/eliminazione.

Gli stati di destinazione finali, come emerso precedentemente dalla curva di sopravvivenza, sono 7, corrispondenti a ciascun punteggio conseguibile con il proprio elaborato finale, e corrispondono quindi agli eventi “laureato con 0 punti per l'elaborato finale”; “laureato con 1 punto per l'elaborato finale” e così via fino a 6 punti, punteggio massimo a cui può ambire un laureando per il lavoro di approfondimento finale.

Particolare attenzione va posta ai dati censurati, cioè alle osservazioni di cui si hanno informazioni parziali, che riguardano, come è stato detto, le prove finali che hanno raggiunto “*i*” *punti Incremento*, ma per le quali non si sa quale sia la loro

valutazione effettiva e se il punteggio assegnato corrisponda a tale valutazione. Pertanto in questi casi le informazioni sono da considerarsi parziali, o meglio censurate.

Infine l'insieme delle covariate è costituito dalle variabili presentate all'inizio dello studio, le stesse utilizzate anche durante la fase di analisi descrittiva, mantenendo la suddivisione in categorie già adottata in precedenza.

## **4.4. Selezione delle Variabili**

La selezione delle variabili è un'operazione importante per capire quali variabili risultino significative per il problema e quali no.

Per sviluppare un modello di Cox che rispondesse alle esigenze e che descrivesse il fenomeno oggetto di studio sono stati fatti diversi tentativi aggiungendo progressivamente le covariate che si è ritenuto potessero influire maggiormente.

Come più volte è stato sottolineato, lo scopo principale dello studio consiste nel rilevare eventuali relazioni tra il punteggio assegnato agli studenti per l'elaborato finale e l'area scientifico disciplinare di appartenenza dei docenti, pertanto risultava fondamentale inserire nel presunto modello la variabile `Area_Relatore`.

Il percorso per individuare il modello più appropriato ha visto alcuni tentativi che includevano le variabili che sembravano più significative o che potevano incidere maggiormente e successive combinazioni tra queste covariate.

Inizialmente si sono condotte analisi su modelli che coinvolgevano pure le variabili quali Diploma e Voto di Diploma che caratterizzano il percorso scolastico precedente all'ingresso nel sistema universitario degli studenti, per appurare se caratteristiche come la formazione secondaria e il voto conseguito all'esame di

maturità potessero intervenire in modo significativo. I coefficienti ottenuti, propri delle due covariate appena menzionate, non sono risultati significativi pertanto si è passati all'analisi di modelli che non coinvolgessero questo tipo di covariate. Sono stati analizzati pure modelli che studiavano l'incidenza della variabile Percorso, per verificare se la scelta di svolgere o meno uno stage, indipendentemente dalla durata, potesse intervenire in misura significativa, ma anche in questo caso non sono emerse relazioni notevoli.

Di altra natura è la considerazione relativa alla variabile che indica il voto che descrive la carriera universitaria degli studenti, cioè il punteggio espresso in centodecimi della media ponderata degli esiti degli esami universitari e quella legata alla covariata che esprime il numero di punti assegnati come premio alla velocità. Queste due variabili, infatti, raccolgono gran parte delle informazioni che riassumono l'esito della carriera universitaria degli studenti.

Il Corso di Laurea rappresenta una caratteristica fondamentale del profilo dei laureati, pertanto si è posta l'attenzione su modelli che includessero questa peculiarità. Infine, il modello che è stato preso maggiormente in considerazione propone delle analisi che si concentrano in particolar modo sul Corso di Laurea di Statistica, Economia e Finanza in quanto i risultati emersi dalle analisi dei dati hanno evidenziato alcune differenze tra il Corso di Laurea SEF e gli altri corsi di laurea offerti dalla facoltà, SGI, SPS e STI.



Tabella 4.1.: *VARIABILI INSERITE nella COSTRUZIONE del MODELLO*

<b>MODELLI USATI</b>
(Area Relatore+Corso di Laurea)
(Area Relatore+Corso di Laurea+Diploma)
(Area Relatore+Corso di Laurea+Diploma+VotoDiploma)
(Area Relatore+Corso di Laurea+Diploma+VotoDiploma +Voto di Partenza)
(Area Relatore+ CorsodiLaurea +DiplomaVotodiPartenza)
(Area Relatore+Corso di Laurea+Diploma+Voto di Partenza+Velocità)
(Area Relatore+Corso di Laurea+Diploma+Velocità)
(Area Relatore+Corso di Laurea+Voto di Partenza)
(Area Relatore+Corso di Laurea+Voto di Partenza+Percorso)
(Area Relatore+Corso di Laurea+Voto di Partenza+Percorso+Velocità)
(Area Relatore+Corso di Laurea+Votodi Partenza +Velocità)
[Area Relatore+(Corso di Laurea="SEF")+Voto di Partenza+Velocità]

## 4.5. Modello di Cox applicato alla popolazione di studio

Il modello a cui si perviene contiene dunque i seguenti regressori:

- **Area Relatore;** area scientifico disciplinare di appartenenza dei docenti;
- **Corso di Laurea;** corso di laurea frequentato dai laureati oggetto di studio;
- **Voto di Partenza;** punteggio che descrive il voto medio degli esami ponderato con il valore dei crediti della relativa attività didattica, espresso in centodecimi e arrotondato all'intero più vicino;

- **Punti Velocità;** numero di punti aggiuntivi che ogni studente può conseguire per incrementare il proprio voto finale sulla base del tempo impiegato per portare a termine gli studi.

Il modello sviluppato propone una specificazione della variabile Corso di Laurea, infatti si è posta l'attenzione su una particolare categoria, quella relativa al Corso di Laurea SEF posta a confronto con gli altri tre corsi di laurea.

Dopo essere pervenuti al modello con i regressori mostrati, si presentano anche i valori:

- dei coefficienti delle variabili;
- degli *hazard ratio*;
- dell'errore standard;
- della statistica del test di Wald;
- del *p-value* della statistica test di Wald;
- del limite inferiore dell'intervallo di confidenza con  $\alpha=0.05$ ;
- del limite superiore dell'intervallo di confidenza con  $\alpha=0.05$ ;
- relativi alla frequenza;
- del livello di significatività.

La verifica della significatività di ogni variabile avviene attraverso il test di Wald (vedi Klein e Moeschberger, 2003).

L'ipotesi che si vuole verificare con questo test risulta

$$\begin{cases} H_0: \beta_1 = \beta_0 \\ H_1: \beta_1 \neq \beta_0 \end{cases}$$

dove generalmente  $\beta_0 = 0$ .

Nel test di Wald la stima di massima verosimiglianza del parametro di interesse  $\beta_1$  è confrontata con un valore preposto,  $\beta_0$ .

Il test per l'ipotesi sopra riportata è

$$Q_W = (\hat{\beta}_1 - \beta_0)^t I(\hat{\beta}_1) (\hat{\beta}_1 - \beta_0),$$

dove  $I(\hat{\beta}_1)$  è la matrice di informazione attesa (Pace e Salvan, 2001) e  $\hat{\beta}_1$  lo stimatore di massima verosimiglianza non vincolato per  $\beta_1$ .

Il test viene quindi confrontato con una  $\chi^2$  con  $p$  gradi di libertà, dove  $p$  è il numero di componenti scalari.

La colonna indicata con “z”, della tabella di seguito presentata, rappresenta il *Test Wald* per ogni regressore. La colonna contraddistinta da “*exp(coef)*” indica gli *hazard ratio* che rappresentano il tasso di eventi per studente-punti incremento.

Dalla tabella emerge che un docente dell'area disciplinare Economico-Aziendale assegna circa il 30% in meno di punti ad un elaborato finale rispetto ad un docente del settore Socio Demografico e che i relatori dell'area Economico Politica assegnano un punteggio inferiore del 44%.

Tabella 4.2: MODELLO di COX APPLICATO alla POPOLAZIONE

	coef	exp(coef)	se(coef)	z	p	lower .95	upper .95	N	Sign.
<b>Rif. SOCIO- DEMOGRAFICA</b>								119	*
<b>EcoAZIENDALE</b>	-0.32	0.73	0.13	-2.47	0.013	0.57	0.94	190	
<b>EcoPOLITICA</b>	-0.57	0.56	0.17	-3.45	0.000	0.41	0.78	108	***
<b>INFORMATICA</b>	-0.22	0.80	0.16	-1.32	0.186	0.58	1.11	63	
<b>MATEMATICA</b>	-0.61	0.54	0.27	-2.28	0.023	0.32	0.92	20	*
<b>STATISTICA</b>	-0.19	0.82	0.15	-1.34	0.180	0.62	1.09	106	
<b>STATISTICO- ECO</b>	-0.18	0.83	0.13	-1.34	0.180	0.64	1.09	152	
<b>Corso LaureaSEF</b>	-0.19	0.82	0.11	-1.75	0.080	0.66	1.02	166	.
<b>Rif. Altri Corsi Laurea</b>								592	
<b>Voto PARTENZA</b>	-0.06	0.94	0.00	-8.92	<2e-16	0.92	0.95	758	***
<b>Punti VELOCITÀ</b>	0.07	1.08	0.02	3.53	0.000	1.03	1.12	758	***

(Fonte: Archivio Segreteria Studenti)

## Curva di sopravvivenza del modello

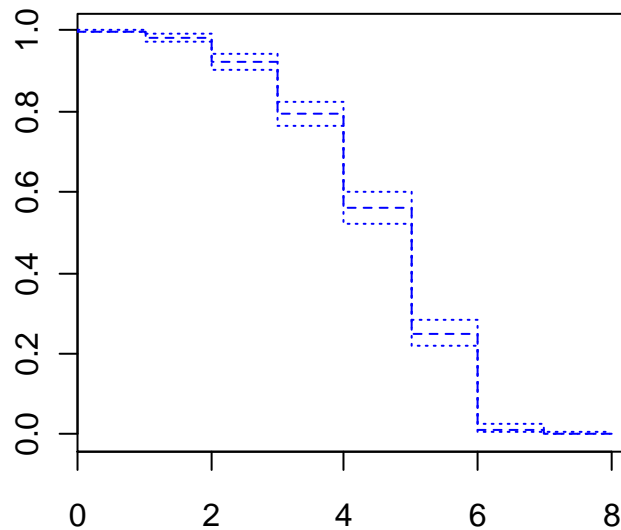


Figura 4.1: Stima curva sopravvivenza Modello di Cox

Tutte le variabili presentate che concorrono a costruire il modello, tranne tre livelli della variabile Area\_Relatore, e più nello specifico la categoria propria dell'area scientifico disciplinare Informatica, Statistica e Statistico-Economica, presentano dei coefficienti significativi contro l'ipotesi nulla. Tale precisazione, circa le categorie che non sono caratterizzate da coefficienti significativi, risulta opportuna in quanto i risultati emersi dal modello sottolineano come non vi siano particolari differenze o lontananze tra il sistema di valutazione adottato dai docenti provenienti dall'area Socio Demografica e quello dei docenti delle aree Informatica, Statistica e Statistico-Economica.

Infatti, nel modello sopra riportato, l'area disciplinare presa a riferimento è quella Socio-Demografica. Nella tabella si evidenzia che in 119 occasioni sono stati i docenti socio-demografici a svolgere il ruolo di relatore. I risultati significativi sono quelli legati principalmente all'area Economica, che è stata ulteriormente suddivisa

in altre due categorie, Economico-Aziendale ed Economico-Politica, e quelli che contraddistinguono l'area Matematica. Si evince come i docenti socio-demografici siano più di tutti propensi ad assegnare punteggi più alti per l'elaborato finale. Sarebbe inoltre interessante capire se il maggior numero di voti attribuito dall'area demografica sia da attribuire ad una predisposizione da parte dei docenti che appartengono a tale area o dipenda indubbiamente da migliori elaborati finali, che però riflettono una certa propensione da parte dei relatori nell'assegnare progetti con maggiore probabilità di esito positivo. Considerazione analoga può essere fatta per i docenti "informatici", "statistici" e "statistico-economico", infatti non vi sono notevoli differenze tra i punti assegnati da questi ultimi e i docenti presi a riferimento.

In misura significativa si discostano invece i risultati prodotti per i relatori che provengono dalle aree economica e matematica. In questi casi si nota una propensione ad assegnare punteggi inferiori per gli elaborati finali. Attraverso l'interpretazione dell'*hazard ratio* infatti, si sottolinea come un docente Matematico assegni per un progetto conclusivo un punteggio pari quasi al 46% in meno dei punti che conferirebbe un collega Socio Demografico, un relatore dell'area Economico-Aziendale attribuisce una valutazione inferiore circa del 30% e un relatore Economico-Politico assegna il 44% in meno di punti rispetto a quelli garantiti dal docente "demografo" per l'elaborato finale.

Tali considerazioni emergono dalla lettura della tabella sopra riportata che propone il modello al quale si è pervenuti. È necessario, tuttavia, leggere la tabella tenendo ben presente anche le corrispondenti ampiezze degli intervalli di confidenza e le relative frequenze assolute. Partendo da questa considerazione, non è possibile sapere con certezza se il risultato relativo all'area Informatica non sia significativo perchè realmente non sussiste una notevole differenza con l'area Demografica o se la mancanza di significatività sia da attribuire al basso numero di osservazioni che la caratterizza.

Infine si sottolinea che le variabili che descrivono il percorso universitario, attraverso il punteggio conseguito con la media degli esami e i punti volti a premiare il minor tempo impiegato, risultano chiaramente significative.





# *CAPITOLO QUINTO*

---

## *CONSIDERAZIONI CONCLUSIVE*

La ricerca proposta nel presente lavoro è stata condotta seguendo due ottiche.

Il primo approccio mira ad un'analisi di background, mentre il secondo si concentra su un'analisi che comunemente viene definita temporale, ma, che come è stato spiegato, ha visto degli adattamenti e delle rivisitazioni per le quali risulta più corretto definire il tipo di analisi adottata come analisi di sopravvivenza, anche se il carattere puramente temporale non interviene nei dati in esame.

La prima parte dell'elaborato, sulla base delle informazioni contenute nell'Archivio della Segreteria, ha avuto come scopo la valutazione dell'influenza di alcune caratteristiche degli studenti all'ingresso nel sistema sull'esito universitario. La seconda invece, introdotta da un'analisi con tavole di eliminazione, ha mirato a sperimentare l'uso di modelli propri dell'analisi dei dati di sopravvivenza, per analizzare la valutazione della prova finale in funzione dell'area scientifico disciplinare di provenienza dei diversi relatori e in funzione dei Corsi di Laurea offerti dalla facoltà.

Per quanto concerne l'analisi di background si è pervenuti alle seguenti considerazioni. È possibile innanzitutto evidenziare che la distanza dal luogo di studio per gli studenti della Facoltà di Scienze Statistiche non incide sull'esito della carriera, semmai risulta interessante sottolineare che poco meno del 90% degli studenti che hanno conseguito il titolo di laurea tra ottobre 2003 e settembre 2008 è residente nella regione Veneto e di questi, quasi il 50% risiede nella provincia di Padova.

Il Corso di Laurea maggiormente frequentato è il Corso SGI che conta il 44.5% dei laureati, a seguire il Corso SEF con un'adesione del 21.9%, simile a quella che caratterizza il Corso STI, descritto da una percentuale pari al 21.6% e infine il Corso di Laurea SPS che vede il 12% delle preferenze. Il 48.9% dei laureati è in possesso di un diploma di scuola media superiore tecnico, seguono poi i laureati che hanno conseguito la maturità al liceo scientifico, con il 38.8%. Gli studenti che arrivano da un formazione media superiore di tipo tecnico scelgono principalmente l'offerta formativa proposta dal corso SGI, a seguire il corso STI che è composto per il 70% proprio da studenti in possesso di un diploma tecnico. I dati relativi invece agli studenti provenienti dal liceo scientifico vedono come principale scelta da parte di questi, ancora il corso SGI, seguito però dal corso SEF caratterizzato dal 45% degli iscritti con maturità scientifica. Questi dati portano a ipotizzare come possano incidere le conoscenze e basi tecnologiche apprese con la maturità tecnica sulla scelta orientata proprio verso il Corso di Laurea in Statistica e Tecnologie Informatiche al fine di approfondire e sviluppare tale bagaglio. L'analisi ha messo in luce che la composizione delle caratteristiche con cui uno studente si presenta al primo anno di corso incidono nella previsione del suo esito universitario. Il voto alla maturità ha mostrato maggiore efficacia predittiva rispetto alla scuola di provenienza, anche se il liceo scientifico ha dato conferma del suo supposto ruolo predittivo. Inoltre, gli studenti provenienti dal liceo scientifico, che presenta mediamente il voto di maturità leggermente più basso, insieme ai colleghi con maturità classica rappresentano gli studenti che maggiormente portano a termine il loro piano di studi in corso.

Quasi la metà dei laureati, il 48,7%, una volta conclusi gli esami previsti ha scelto di completare il proprio piano di studi con il percorso B che prevede un lavoro di approfondimento da concordare con il proprio relatore, mentre le altre tre opzioni offerte dalla Facoltà di Scienze Statistiche che prevedono uno stage della durata variabile di tre, quattro o sei mesi sono state accolte rispettivamente dal

22.8% per il percorso A1, dal 21.2% per il percorso A2 e dal 7.3% degli studenti per il percorso più lungo A3.

Questa scelta può trovare fondamento sia per ragioni accademiche nel caso in cui uno studente sia intenzionato a proseguire i propri studi e quindi reputerà più idoneo alla propria situazione un lavoro di approfondimento, sia per ragioni pratiche che coinvolgono ad esempio studenti lavoratori che chiaramente non hanno il tempo per sviluppare un progetto proposto dallo stage in un'azienda o ente.

Il 36,9% degli studenti in esame si è laureato nella prima sessione di laurea “in corso” o addirittura in anticipo rispetto alla durata naturale del corso, il 38,3% invece ha terminato gli studi oltre il tempo che consentiva loro di aggiudicarsi dei punti aggiuntivi volti a premiare la velocità, pertanto più del 60% degli studenti ha conseguito la laurea “in corso”. Il 5% dei laureati ha conseguito il titolo anticipatamente rispetto alla lunghezza naturale del Corso di Laurea.

Il voto finale si distribuisce principalmente nell'intervallo di punti 91/110-100/110 con il 34,6% dei laureati, ma si nota come una buona percentuale degli studenti, quasi il 20%, abbia conseguito il titolo con un voto finale superiore a 105/110 punti. Inoltre si sottolinea che il 13,7% ha conseguito il titolo con il punteggio massimo, quasi il 10% della popolazione ha avuto il riconoscimento della lode. Tra coloro che hanno avuto il massimo dei voti, spicca il dato relativo agli studenti di SGI, infatti più del 30% degli studenti laureatisi con 110/110 aveva precedentemente frequentato il corso in Statistica e Gestione delle Imprese. È opportuno però ricordare la diversa numerosità che caratterizza i quattro Corsi di Laurea offerti dalla Facoltà. Sulla base di tale considerazione si riscontra che meno del 10% dei frequentanti SGI consegue il voto più alto, mentre di altra natura sono le percentuali che riguardano gli altri Corsi di Laurea, con SPS che spicca tra tutti, presentando più del 23% degli studenti che hanno raggiunto la massima valutazione. Si evidenzia che il conseguimento del titolo con il voto di 110/110 risulta agevolato anche dal contributo dei punti assegnati per la velocità volti a

premiare la conclusione della carriera “in corso”, infatti più dell’87% degli studenti con voto finale massimo ha raggiunto tale risultato anche grazie questi punti ulteriori.

I docenti coinvolti il maggior numero di volte in qualità di relatori sono quelli che appartengono all’area economica. Tali docenti vengono scelti, principalmente, non solo dagli studenti orientati verso il lavoro di uno stage, come ci si aspetterebbe, ma sono i preferiti anche tra gli studenti che per la prova finale hanno deciso di sviluppare un lavoro di approfondimento da concordare con il proprio relatore.

Nello sviluppare le analisi che hanno portato ai dati appena esposti, si è sottolineato che gli studenti che si laureano con il massimo dei voti, mediamente conseguono il titolo nella prima sessione “in corso” e hanno un voto di partenza maggiore ai 102/110 punti, ma ricevono in media solo 2 punti per la prova finale. Tale punteggio è inferiore sia alla valutazione che si suppone possa ricevere un laureato con il massimo dei voti sia ai punti che vengono assegnati per l’elaborato ai laureati che conseguono il titolo con valutazioni inferiori, quindi tali dati forniscono delle informazioni parziali. Si tratta chiaramente di un problema strutturale del sistema di valutazione in quanto uno studente che si presenti con un voto di partenza maggiore a 104 può ricevere una valutazione per l’elaborato finale che non esprima effettivamente il valore di tale progetto in quanto i punti che si possono conseguire nella prova finale sono 6 e il punteggio massimo 110/110.

Pertanto attraverso lo studio di sopravvivenza, sono state condotte delle analisi che permettessero di cogliere eventuali relazioni tra le variabili che determinano l’esito della prova finale e si è costruito un modello per vedere in che modo l’area scientifico disciplinare del relatore intervenisse al netto delle altre variabili.

Attraverso le curve di sopravvivenza è stato evidenziato come il punteggio della prova finale cambi a seconda del Corso di Laurea frequentato e in base anche al settore disciplinare del relatore, ma per capire in quale misura si discostino le

valutazioni sulla base delle aree di appartenenza dei docenti si è usato il modello di Cox. Il modello elaborato ha permesso di vedere come contribuisca la variabile *Area\_Relatore* a parità dell'influenza esercitata dalle altre variabili inserite nell'analisi. È emerso che la valutazione delle prove finali da parte dei docenti delle aree Statistica, Statistico-Economico e Informatica non è significativamente differente dal punteggio concesso dai relatori dell'area Socio Demografica, che costituisce l'area di riferimento del modello. Invece, dai risultati ottenuti con il modello, i docenti del settore Economico risultano più severi nell'assegnazione dei punti. Infatti un relatore le cui attività didattiche rientrano nell'area Economico Aziendale assegna circa 30% in meno di punti rispetto a quelli conferiti da un collega socio demografico mentre un docente economico-politico assegna una valutazione che risulta inferiore del 44% sempre rispetto all'area di riferimento.

Risulta inoltre che i relatori più severi sono quelli dell'area matematica, caratterizzati da una propensione a dare punteggi del 46% inferiori, ma in questo caso è giusto sottolineare che il numero di osservazioni relative alle occasioni in cui a fungere il ruolo di relatore era un docente matematico è di gran lunga inferiore rispetto agli altri casi.

Infine va sottolineato che queste considerazioni sono la semplice lettura dei dati emersi dal modello e potrebbe essere che i dati relativi a punteggi inferiori assegnati dai docenti dell'area Economica non riflettano una reale predisposizione da parte di questi nel conferire voti più bassi, ma che i voti più bassi risultino invece dalla qualità degli studenti.



# BIBLIOGRAFIA

---

- [1] Cox, D.R. & Oakes, D. (1984) *Analysis of Survival Data*, New York, Chapman & Hall.
- [2] Ewens, W.J. & Grant G.R. (2005) *Statistical methods in Bioinformatics: An Introduction, Second Edition*, New York, Springer.
- [3] Kleinbaum, D.G. & Klein M. (2005) *Survival Analysis: a self learning text, Second edition*, New York, Springer.
- [4] Klein J.P. & Moeschberger M.L. (2003) *Survival Analysis : Techniques for Censored and Truncated Data*, New York, Springer.
- [5] Marubini, E. & Valsecchi, M.G. (1987) *Analisi della Sopravvivenza in Sperimentazioni Cliniche Controllate e nelle Osservazioni Pianificate*, Milano, Edizioni a cura del “Centro Zambon”.
- [6] Pace, L. & Salvan A. (2001) *Introduzione alla Statistica II*, Padova, Cedam.
- [7] Therneau T.M. & Grambsch P. (2000), *Modelling Survival Data: Extending the Cox Model*, New York, Springer
- [8] Tripepi, G. & Catalano F. (2004) *L’Analisi di Sopravvivenza con il Metodo di Kaplan Meier*, *Giornale Italiano di Nefrologia*, 6, pp. 540-546.
- [9] Venables, W. N. & Ripley B. D. (2002) *Modern Applied Statistics with S*, Springer.

