

Facoltà di Economia

Corso di Laurea in Economia Aziendale

Tesi di Laurea in Statistica

INDAGINE SUI TEST D'INGRESSO

RELATORE:

Prof. Marco Scarsini

LAUREANDO:

Tommaso Bebi

Matricola 145191

ANNO ACCADEMICO 2009-2010

“Success is the ability to go
from one failure to another
with no loss of enthusiasm.”

(Winston Churchill)

INDICE

RINGRAZIAMENTI.....	5
1 INTRODUZIONE.....	6
1.1 Incipit.....	6
1.2 Altri studi sull'argomento.....	6
1.3 Teoria dei metodi di analisi.....	7
1.4 Software di supporto.....	7
2 I DATI.....	8
2.1 Struttura dei dati.....	8
2.2 Pulizia preliminare dei dati.....	9
2.3 Preparazione dei dati per l'analisi.....	10
3 TEST D'INGRESSO E CARRIERA ACCADEMICA:	
un modello per lo studio del successo accademico.....	12
3.1 Analisi descrittiva.....	12
3.2 Le relazioni tra le variabili.....	13
3.3 Regressione sul voto di Laurea.....	15
3.4 Regressione sul voto medio.....	19

4	LA POSIZIONE GEOGRAFICA DEGLI ISTITUTI DI	
	PROVENIENZA: effetti e dinamiche.....	22
4.1	La variabile latitudine.....	22
4.2	Effetti della posizione geografica dell'istituto superiore sul voto di maturità e sulla media degli esami universitari.....	23
4.3	Bontà di adattamento e stabilità del modello per studenti provenienti da diverse aree geografiche.....	28
4.4	Correlazione tra voto di maturità e latitudine nelle diverse fasce di successo accademico.....	30
5	OSSERVAZIONI E CONCLUSIONI.....	32
5.1	Stabilità del modello nella segmentazione annuale del campione.....	32
5.2	Gli indicatori di successo accademico.....	32
5.3	I predittori di successo accademico.....	33
5.4	Conclusioni.....	34
	APPENDICE.....	36
	BIBLIOGRAFIA.....	39

RINGRAZIAMENTI

Sono arrivato a un altro importante traguardo della mia vita. Prima di proseguire voglio fermarmi almeno un istante per ringraziare tutti coloro che mi hanno aiutato a raggiungere questo obiettivo tanto anelato.

GRAZIE a Paolo De Nardo, per avermi spinto a realizzare la tesi su questo argomento, offrendomi alcuni fondamentali spunti.

GRAZIE a Carlo Di Palma e Alessandro Lubizc, per avermi fornito i dati di cui avevo bisogno e per avermi gentilmente aiutato nella risoluzione di alcune problematiche tecniche.

GRAZIE a Claudio Rossetti, senza il quale i software di analisi sarebbero ancora terreno a me completamente sconosciuto.

GRAZIE a mia cugina Francesca, per essermi stata vicina nei momenti più difficili, quelli in cui le “conclusioni” sembravano essere solamente un miraggio.

CAPITOLO I

Introduzione

1.1 INCIPIT

L'idea di compiere un'indagine sui "test d'ingresso" della facoltà di economia della LUISS emerge dalla necessità di trovare risposta ad alcune domande che molto spesso vengono a trovarsi al centro di accesi dibattiti tra studenti, e non solo.

A cosa servono i test d'ingresso? Selezionano davvero i "migliori" studenti? E' più efficace il test psicoattitudinale o la considerazione della media scolastica? Sono soltanto alcune delle domande cui tenteremo di trovare una risposta.

Per giunta questo studio permette di verificare se e quanto le finalità con cui sono stati adottati e strutturati i test d'ingresso abbiano poi un fattivo riscontro nella realtà di selezione.

1.2 ALTRI STUDI SULL'ARGOMENTO

Preliminarmente abbiamo verificato l'esistenza di studi simili che potessero rappresentare fonte di spunto ed ispirazione del *modus operandi*.

In campo nazionale non si trovano lavori degni di nota.

Estendendo il campo di ricerca all'estero sembra opportuno citare due articoli (entrambi del campo medico) che riportano i risultati di studi simili a quello che stiamo per cominciare.

Sia in "The validity of the Medical College Admission Test for predicting performance in the first two years of medical school" che in "Correlation of Admissions Criteria with Academic Performance in Dental Students", gli autori giungono a verificare modelli soddisfacenti

(anche con $R^2 = 0,77$). Essi utilizzano delle regressioni ponendo come variabile casuale il risultato al test d'ingresso e come variabile spiegata l'andamento accademico degli studenti. Il primo studio risulta particolarmente interessante poiché l'indicatore per l'andamento accademico è simile a quello che, come si vedrà, è stato scelto per il nostro studio e cioè la media aritmetica degli esami sostenuti dei primi due anni del percorso di studi.

Bisogna però rilevare che i test d'ammissione considerati sono strutturati in maniera completamente differente da quelli della LUISS, assomigliando più ai test italiani per l'ammissione alle facoltà di Medicina e Odontoiatria.

1.3 TEORIA DEI METODI DI ANALISI

Le conoscenze di base dei modelli statistici e dei test fanno riferimento al testo adottato per l'esame di Statistica alla facoltà di Economia della LUISS: Monti (2008). Le conoscenze acquisite durante il corso universitario si sono rivelate insufficienti per poter svolgere un'analisi accurata sul tema. Si è pertanto approfondito lo studio soprattutto in materia di analisi multivariata: Fabbris (1997) e Hair, Anderson, Tatham, Black (1998).

1.4 SOFTWARE DI SUPPORTO

Nello svolgere l'analisi ci siamo avvalsi di due software statistici:

- STATA 10.
- SPSS

L'analisi è stata preliminarmente svolta con l'utilizzo del programma STATA 10. Questo programma ha però mostrato alcuni limiti per quanto concerne il layout grafico e si è quindi scelto di operare una seconda analisi con SPSS. Questa analisi ha permesso di verificare i risultati ottenuti con la prima ed ha fornito tabelle e grafici più adatti ad essere riportati. Per svolgere l'analisi su SPSS abbiamo fatto riferimento a due affermati testi: Di Franco (2009) e Field (2000)

CAPITOLO II

I dati

2.1 STRUTTURA DEI DATI

I dati sono stati forniti in formato “.mdb” (Office - Access) dall’Ufficio Sistemi Informativi e Telecomunicazioni della LUISS “Guido Carli”.

Il database è composto da 130.111 record e 29 variabili relative ai dati anagrafici, alla carriera scolastica ed accademica di 5155 studenti della facoltà economia della LUISS dall’anno accademico 1996/97 all’anno accademico 2008/09; laddove gli studenti non abbiano ancora terminato il percorso di studi i dati risultano parziali.

I dati sono strutturati in modo tale che ad ogni record corrisponda un codice studente; più record hanno uno stesso codice studente in relazione al numero di esami sostenuti e registrati e quindi la relazione tra i dati anagrafici di ogni singolo studente e le sue attività formative ha una cardinalità uno a molti.

Di seguito vengono elencate le 29 variabili presenti nell’archivio:

- “codiceelemento” – Codice identificativo del singolo studente.
- “datanascita” – Data di nascita.
- “comunenascita” – Comune di nascita.
- “prnascita” – Provincia di nascita.
- “comuneresidenza” – Comune di residenza.
- “prresidenza” – Provincia di residenza.
- “nazionalita” – Nazionalità.
- “sesso” – Sesso.
- “coddiploma” – Tipologia di diploma conseguito dallo studente.
- “annodiploma” – Anno di conseguimento del diploma.

- “votodiploma” – Voto conseguito al diploma.
- “basevoto” – Base del voto conseguito al diploma.
- “istitutodiploma” – Nome dell’istituto di provenienza.
- “comuneistituto” – Comune in cui è sito l’istituto di provenienza.
- “pristituto” – Provincia in cui è sito l’istituto di provenienza.
- “indistituto” – Indirizzo dell’istituto di provenienza.
- “pcarriera” – Punteggio assegnato in base alla carriera scolastica.
- “ptest” – Punteggio assegnato in base al test psico-attitudinale.
- “annoaccingresso” – Anno dell’ingresso dello studente in LUISS.
- “codmateria” – Codice esame registrato.
- “denominazione” – Nome dell’esame registrato.
- “dataesame” – Data di registrazione dell’esame.
- “crediti” – Numero di crediti assegnati all’esame.
- “voto” – Voto conseguito all’esame.
- “esamisostenuti_lode” – Eventuale conseguimento della lode.
- “data laurea” – Data di laurea.
- “incorso” – Numero di anni di fuori corso.
- “votolaurea” – Voto conseguito alla laurea.
- “laureati_lode” – Eventuale conseguimento della lode.

2.2 PULIZIA PRELIMINARE DEI DATI

Partendo dal database fornito dall’Ufficio Sistemi Informativi e Telecomunicazioni ne è stato ricavato un altro in cui ad ogni record corrisponde uno studente con dati relativi alla sua persona e alla sua carriera scolastica-accademica e una variabile voto ricavata come media aritmetica dei voti degli esami verbalizzati durante il percorso accademico.

L’archivio è stato, quindi, nuovamente modificato procedendo alla cancellazione permanente di alcuni record ed alcune variabili:

- 1) I record relativi agli studenti che hanno conseguito un diploma in un istituto superiore secondario all’estero.

- 2) I record relativi agli studenti che non hanno terminato il percorso di studi universitario conseguendo la laurea.

Al termine delle operazioni di “depurazione” preliminare il database utilizzato per l’analisi risulta costituito da 3.762 studenti.

2.3 PREPARAZIONE DEI DATI PER L’ANALISI

Per poter svolgere l’analisi si è resa necessaria la creazione di alcune variabili:

- 1) “votodiploma2”

La variabile è stata determinata convertendo le votazioni in sessantesimi degli esami di maturità in centesimi. Sono rimaste invariate le valutazioni che erano già espresse in centesimi. Si riporta la formula utilizzata:

$$\text{“votodiploma2”} = \text{“votodiploma”} / \text{“basevoto”} * 100$$

Questa nuova variabile rende confrontabili le valutazioni agli esami di stato del vecchio e del nuovo ordinamento, permettendo un’analisi omogenea.

- 2) “pcarriera100”

Il punteggio attribuito alla carriera scolastica degli studenti presenta due distinte problematiche. In primis, le regole di attribuzione del punteggio cambiano di anno in anno e, talvolta anche all’interno dello stesso anno d’ingresso tra il test di aprile e quello di settembre. In secundis, di anno in anno cambia il peso del punteggio assegnato alla carriera nella formazione del punteggio totale del test (vedi tavola “criteri generali di formazione delle graduatorie e prove di ammissione ai corsi di laurea triennale” in appendice).

La prima problematica non è risolvibile con i dati a disposizione e verrà considerata come esogena al modello spiegato.

La seconda problematica è stata risolta attraverso la creazione di due nuove variabili. Per prima cosa è stata creata la variabile “basecarriera”, inserendo per ogni anno il peso della carriera. Successivamente è stata creata la variabile “pcarriera100”, secondo la formula:

$$\text{“pcarriera100”} = \text{“pcarriera”} / \text{“basecarriera”} * 100$$

“pcarrieria100” non è altro che il punteggio attribuito alla carriera dello studente riportato in centesimi. Questa nuova variabile rende confrontabili i punteggi attribuiti in anni diversi ed è stata utilizzata per l’analisi.

3) “ptest100”

Anche per il punteggio attribuito al test psicoattitudinale si è dovuto risolvere il problema dei diversi pesi. Poiché la somma dei pesi del punteggio carriera e del punteggio test è sempre uguale a cento, la soluzione è stata trovata nella creazione di una nuova variabile,

“basetest”, talechè:

$$\text{“basetest”} = 100 - \text{“basecarriera”}$$

Successivamente si è provveduto alla creazione della variabile “ptest100”, secondo la formula:

$$\text{“ptest100”} = \text{“ptest”} / \text{“basetest”} * 100$$

“ptest100” non è altro che il punteggio dello studente al test psicoattitudinale riportato in centesimi. Questa nuova variabile rende confrontabili i punteggi conseguiti in anni diversi ed è stata utilizzata per l’analisi.

4) “punttotest”

Questa variabile è stata determinata sommando i valori della variabile “pcarrieria” e “ptest” ed indica il risultato totale ottenuto dal singolo studente al test d’ammissione (con questo punteggio vengono stilate le graduatorie per l’ammissione).

5) “votomedio”

Questa variabile indica la media aritmetica dei voti degli esami verbalizzati da ogni singolo studente.

CAPITOLO III

Test d'ingresso e carriera accademica: un modello per lo studio del successo accademico

3.1 ANALISI DESCRITTIVA

La tavola 3.1.1 riporta le principali caratteristiche delle variabili che sono state utilizzate per l'analisi.

TAVOLA 3.1.1:

variabile	Obs	Mean	Std. Dev.	Min	Max
voto laurea	3762	104.2701	6.211841	80	110
voto medio	3762	26.88802	1.619384	21.30769	30
voto diploma2	3762	90.45348	10.01329	60	100
p.carriera100	3762	83.51123	10.54222	60	100
p.test100	3762	72.76622	11.37722	35	98.4
punttottest	3762	77.46884	7.927403	60	98.56

Il campione è composto da 3762 studenti, laureati tra il 2000 ed il 2009, che abbiano compiuto un percorso di studi pre-universitario in Italia.

E' interessante notare come la media dei voti di laurea sia sensibilmente diminuita con l'avvento della riforma del sistema universitario (tavola 3.1.2).

TAVOLA 3.1.2:

AnnoLaurea	VotoLaurea
2000	107,75
2001	108,52
2002	107,17
2003	107,29
2004	106,02
2005	105,58
2006	103,89
2007	101,53
2008	101,66
2009	100,26
Totale	104,27

3.2 LE RELAZIONI TRA LE VARIABILI

In via preliminare sono stati esaminati gli scatter plot ed è stata effettuata l'analisi di correlazione tra le variabili a disposizione con lo scopo di individuare quelle maggiormente correlate con le variabili "media esami" e "voto di laurea".

Da una dettagliata lettura della tavola 3.2.1 si osserva che:

- Sia il voto medio sia il voto di laurea risultano moderatamente e positivamente correlate con il voto di maturità;
- La correlazione tra il voto medio e il voto di laurea con il Punteggio al test di ingresso è positiva ma notevolmente più bassa rispetto a quella evidenziata con il voto di maturità;
- Il voto di maturità e il voto al test risultano, invece, scarsamente correlate;

- Sia il voto medio sia il voto di laurea risultano positivamente correlate con il punteggio carriera e con punteggio totale test ma meno intensamente di quanto non lo siano con il voto di maturità;

TAVOLA 3.2.1: Matrice di correlazione

		Correlazioni					
		Voto Maturita'	Punteggio Carriera (x/100)	Punteggio Test (x/100)	Punteggio Totale Test	votomedio	Voto Laurea
Voto Maturita'	Correlazione di Pearson	1	,710**	-,066**	,359**	,503**	,455**
	Sig. (2-code)		,000	,000	,000	,000	,000
	N	3762	3762	3762	3762	3762	3762
Punteggio Carriera (x/100)	Correlazione di Pearson	,710**	1	,089**	,611**	,384**	,361**
	Sig. (2-code)	,000		,000	,000	,000	,000
	N	3762	3762	3762	3762	3762	3762
Punteggio Test (x/100)	Correlazione di Pearson	-,066**	,089**	1	,824**	,162**	,180**
	Sig. (2-code)	,000	,000		,000	,000	,000
	N	3762	3762	3762	3762	3762	3762
Punteggio Totale Test	Correlazione di Pearson	,359**	,611**	,824**	1	,350**	,349**
	Sig. (2-code)	,000	,000	,000		,000	,000
	N	3762	3762	3762	3762	3762	3762
votomedio	Correlazione di Pearson	,503**	,384**	,162**	,350**	1	,951**
	Sig. (2-code)	,000	,000	,000	,000		,000
	N	3762	3762	3762	3762	3762	3762
Voto Laurea	Correlazione di Pearson	,455**	,361**	,180**	,349**	,951**	1
	Sig. (2-code)	,000	,000	,000	,000	,000	
	N	3762	3762	3762	3762	3762	3762

** . La correlazione è significativa al livello 0,01 (2-code).

I test effettuati per valutare se la correlazione è “significativamente diversa da zero” confermano comunque, l’esistenza di una relazione tra le nostre variabili di studio; le correlazioni evidenziate nella tavola risultano, infatti, tutte significativamente diverse da zero ($p < 0,001$).

Dai risultati di correlazione si possono fare utili supposizioni per la definizione di tecniche di analisi statistica multivariata (regressione lineare) volte all’esplorazione più approfondita dei dati . L’esistenza di una associazione positiva moderatamente intensa tra l’esito della carriera universitaria degli studenti (voto medio e voto di laurea) e il voto conseguito alla maturità suggeriscono una relazione lineare tra le capacità acquisite dallo studente durante il percorso di studio pre-universitario e il successo accademico espresso indifferentemente da una buona media agli esami o da una votazione elevata conseguita alla Laurea. Il voto di

maturità, può essere utilizzato, quindi, non solo come un indicatore delle capacità acquisite nel percorso di studi pre-universitari ma anche come un predittore di quello che sarà il percorso accademico e di conseguenza appropriatamente utilizzato nella definizione di punteggi e graduatorie per l'accesso al percorso di studi. Il punteggio test, a sua volta, risulta positivamente correlato al successo accademico degli studenti ma, con un'intensità minore rispetto al voto di maturità facendo presupporre una minore capacità predittiva rispetto a quello che sarà la carriera universitaria futura dello studente. Infine, è utile sottolineare come una correlazione meno intensa tra quello che può essere definito l'esito della carriera universitaria degli studenti (voto medio e voto di laurea) e le variabili "punteggio carriera" e "punteggio totale test" possa essere l'indicazione ad una parziale ridefinizione di questi indicatori in quanto meno legati al risultato accademico finale rispetto a quanto non lo siano il voto di maturità e il punteggio al test.

In considerazione, dello studio della matrice delle correlazioni, si è proceduto quindi ad ipotizzare una relazione lineare tra il successo accademico e le due variabili "voto di maturità" e "punteggio al test" d'ingresso utilizzando modelli di regressione lineare volti a "quantificare" la forza delle correlazioni finora evidenziate.

3.3 REGRESSIONE SUL VOTO DI LAUREA

In questo paragrafo viene proposto un modello di regressione lineare per lo studio delle relazioni che legano il successo accademico contestualmente valutato attraverso il voto conseguito alla laurea e le due variabili "voto di maturità" e "punteggio conseguito al test" che descrivono rispettivamente le capacità acquisite durante gli studi pre-universitari e una misura dell'attitudine (potenziale) dello studente ad affrontare il percorso accademico. Le tavole 3.3.1 e 3.3.2 illustrano, quindi, i risultati della regressione ottenuta considerando come variabile indipendente il voto di laurea e come variabili esplicative il "voto di maturità" e il "punteggio al test".

TAVOLA 3.3.1: Riepilogo della regressione sul voto di laurea

Modello	R	R-quadrato	R-quadrato corretto	Deviazione standard Errore della stima
1	,501 ^a	,251	,251	5,376

a. Predittori: (Costante), Punteggio Test (x/100), Voto Maturita'

TAVOLA 3.3.2: ANOVA della regressione sul voto di laurea

Modello		Somma dei quadrati	df	Media dei quadrati	F	Sig.
1	Regressione	36477,134	2	18238,567	631,015	,000 ^a
	Residuo	108648,476	3759	28,904		
	Totale	145125,610	3761			

a. Predittori: (Costante), Punteggio Test (x/100), Voto Maturita'

Dall'analisi dell'output riportato nelle tavole sovrastanti, è evidente che il modello risulta sostanzialmente accettabile per quanto il valore del coefficiente di determinazione non sia estremamente elevato ($R^2 = 0,251$); del resto è plausibile che sia il voto di maturità sia il punteggio conseguito al test d'ingresso possano solo in parte spiegare (determinare/predire) il successo accademico di uno studente che sarà determinato da una serie di altre variabili riferite direttamente alla carriera universitaria, alle capacità acquisite durante gli studi universitari, nonché alle caratteristiche socio-culturali dello studente. Entrambi i regressori ("voto di maturità" "punteggio al test") risultano significativi ($p < 0,001$) e presentano coefficienti positivi suggerendo che ad un voto di maturità più elevato e ad un punteggio migliore conseguito al test consegua un voto di laurea più elevato (tavola 3.3.3). Confrontando i coefficienti standardizzati, però, è evidente come sia maggiore il peso della variabile "voto di maturità" rispetto al "punteggio al test" nello spiegare il voto di laurea.

TAVOLA 3.3.3: Coefficienti della regressione sul voto medio

Modello	Coefficienti non standardizzati		Coefficienti standardizzati	t	Sig.	Statistiche di collinearità	
	B	Deviazione standard Errore	Beta			Tolleranza	VIF
1 (Costante)	69,589	1,006		69,185	,000		
Voto Maturita'	,291	,009	,469	33,164	,000	,996	1,004
Punteggio Test (x/100)	,115	,008	,210	14,882	,000	,996	1,004

Dalla matrice di correlazione (paragrafo 3.2) non emerge in maniera evidente un problema di multicollinearità; le nostre variabili esplicative, infatti, non risultano fortemente correlate (-0,063). Tuttavia, per approfondire questo aspetto, è stata impostata l'analisi includendo le statistiche per verificare la multicollinearità (Tollerance e VIF), che sono riportate nella tavola dei coefficienti. I test calcolati per il modello confermano l'assenza di multicollinearità; i valori della statistica Tollerance risultano, infatti, sufficientemente elevati mentre quelli della statistica VIF prossimi ad 1 (valore soglia $P=10$).

I due grafici riportati in figura 3.3.4 permettono di verificare l'assunzione di normalità dei residui. Come possiamo osservare i residui non seguono una distribuzione normale, evidenziando una notevole asimmetria dei dati. Nel Normal Probability plot, i punti tendono a disporsi con approssimazione insufficiente lungo la retta. Vi è quindi l'evidenza di una significativa violazione dell'ipotesi di normalità dei residui. Inoltre, il plot dei residui standardizzati rispetto ai valori stimati standardizzati riportato in 3.3.5, che permette di verificare l'assunzione di omoschedasticità, suggerisce la presenza di eteroschedasticità con varianza decrescente. Non potendo essere verificate, pertanto, le assunzioni di eteroschedasticità e normalità dei residui, il modello analizzato non risulta caratterizzato da una forte stabilità.

FIGURA 3.3.4: istogramma e Normality Probability plot

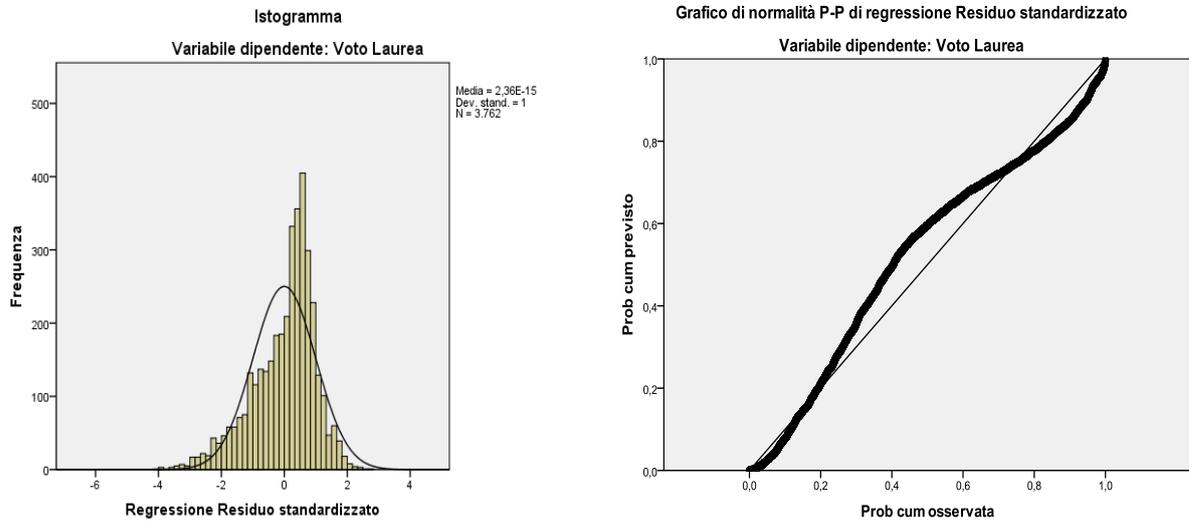
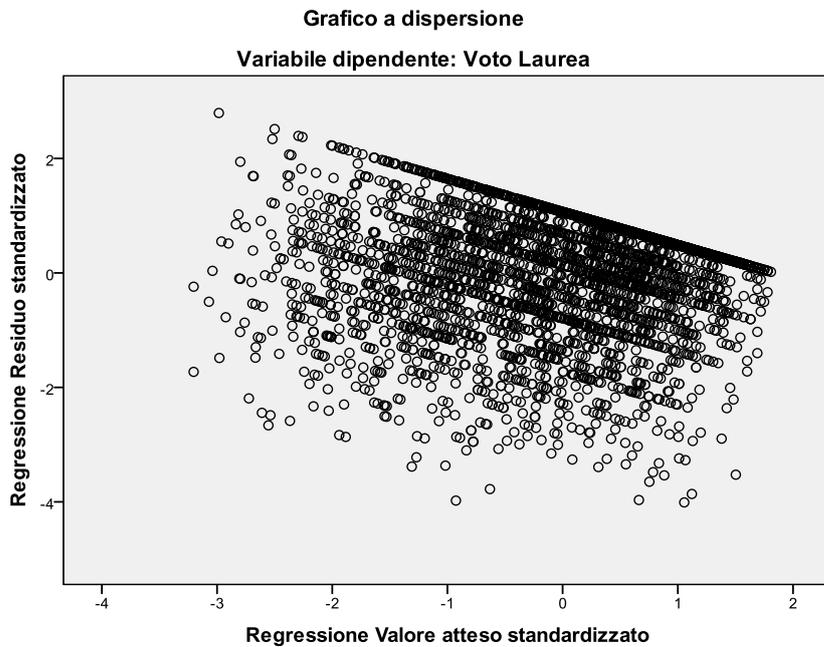


FIGURA 3.3.5: Scatter plot dei residui standardizzati vs valori stimati standardizzati



3.4 REGRESSIONE SUL VOTO MEDIO

In questo paragrafo viene proposto un modello di regressione lineare per lo studio delle relazioni che legano il successo accademico contestualmente valutato attraverso la “media” conseguita negli esami durante il percorso accademico e le due variabili “voto di maturità” e “punteggio conseguito al test”. Rispetto all’analisi svolta nel paragrafo precedente, viene sostituita la variabile dipendente “voto di laurea” con la variabile dipendente “media aritmetica dei voti”.

Le tavole 3.4.1 e 3.4.2 illustrano i risultati della regressione ottenuta considerando come variabile indipendente il “voto medio” agli esami e come variabili esplicative il “voto di maturità” e il “punteggio al test”.

TAVOLA 3.4.1: Riepilogo della regressione sul voto medio

Modello	R	R-quadrato	R-quadrato corretto	Deviazione standard Errore della stima
1	,539 ^a	,291	,291	1,36395

a. Predittori: (Costante), Punteggio Test (x/100), Voto Maturita'

TAVOLA 3.4.2: ANOVA della regressione sul voto medio

Modello	Somma dei quadrati	df	Media dei quadrati	F	Sig.	
1	Regressione	2869,814	2	1434,907	771,311	,000 ^a
	Residuo	6993,046	3759	1,860		
	Totale	9862,861	3761			

a. Predittori: (Costante), Punteggio Test (x/100), Voto Maturita'

Dall’analisi dell’output riportato nelle tavole sovrastanti è evidente che il modello risulta sostanzialmente accettabile per quanto il valore del coefficiente di determinazione non sia estremamente elevato ($R^2 = 0,29$). Il valore del coefficiente di determinazione di questo

modello risulta comunque più elevato rispetto a quello del modello analizzato nel paragrafo precedente. Entrambi i regressori (“voto di maturità” “punteggio al test”) risultano significativi ($p < 0,001$) e presentano coefficiente positivi suggerendo che ad voto di maturità più elevato e ad un punteggio migliore conseguito al test consegua una media più elevata agli esami. Anche in questo modello confrontando i coefficienti standardizzati (tavola 3.4.3) risulta evidente come sia maggiore il peso della variabile “voto di maturità” rispetto al “punteggio al test” nello spiegare la media degli esami sostenuti durante il percorso accademico degli studenti.

TAVOLA 3.4.3: Coefficienti della regressione sul voto medio

Modello	Coefficienti non standardizzati		Coefficienti standardizzati	t	Sig.	Statistiche di collinearità	
	B	Deviazione standard Errore	Beta			Tolleranza	VIF
1 (Costante)	17,316	,255		67,857	,000		
Voto Maturita'	,083	,002	,516	37,457	,000	,996	1,004
Punteggio Test (x/100)	,028	,002	,196	14,244	,000	,996	1,004

Anche per questo modello i test Tollerance e VIF confermano l’assenza di multicollinearità delle variabili esplicative.

I due grafici riportati in figura 3.4.4 permettono di verificare l’assunzione di normalità dei residui. Come possiamo osservare i residui seguono approssimativamente una distribuzione normale, sebbene sia riscontrabile una lieve asimmetria dei dati e l’istogramma risulti più appuntito. Nel Normal Probability plot, i punti tendono a disporsi, anche se con approssimazione non del tutto soddisfacente, lungo una retta. Si può concludere, quindi, che non vi è l’evidenza di una forte violazione dell’ipotesi di normalità. Il plot dei residui standardizzati rispetto a valori stimati standardizzati riportato in 3.4.5 permette, infine di verificare l’assunzione di omoschedasticità; la nuvola di punti si dispone in modo casuale e non suggerisce la presenza di eteroschedasticità.

FIGURA 3.4.4: istogramma e Normality Probability plot

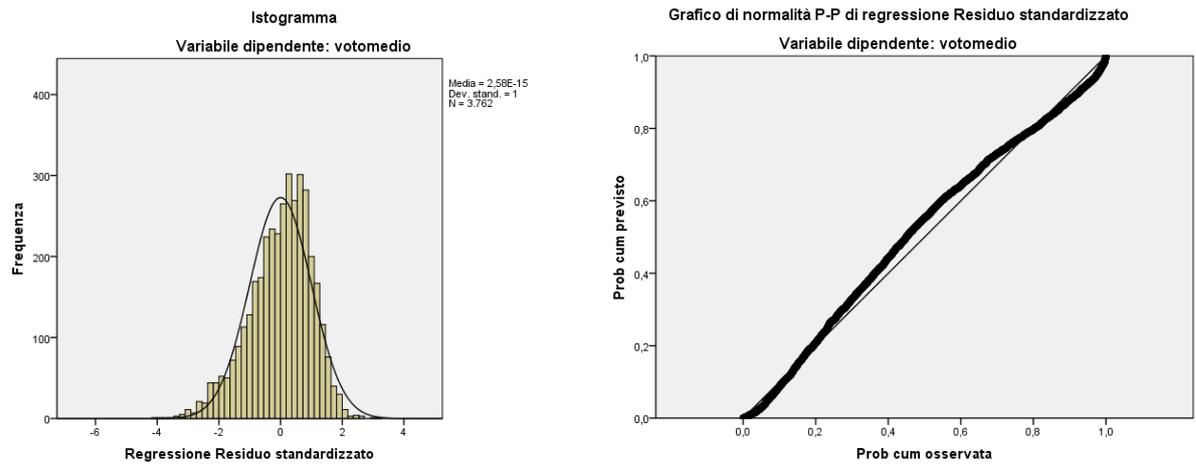
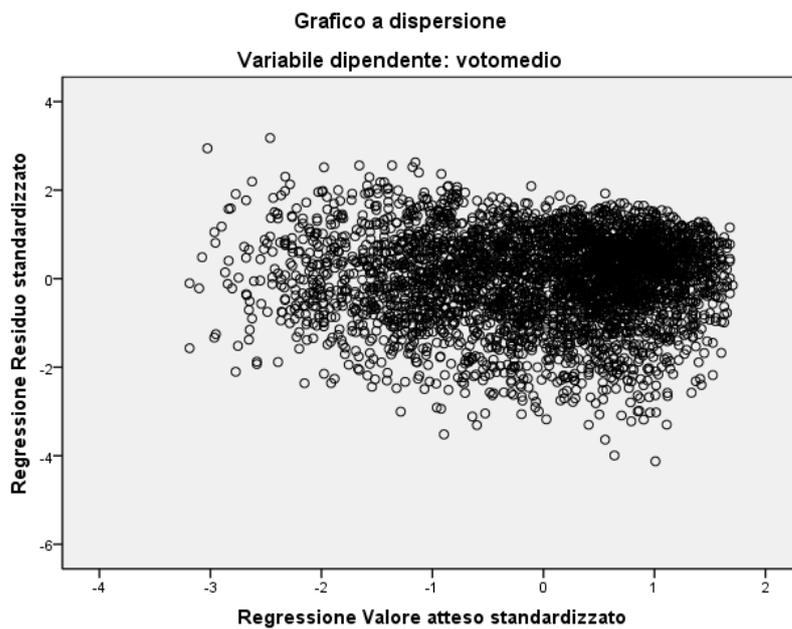


FIGURA 3.4.5: Scatter plot dei residui standardizzati vs valori stimati standardizzati



CAPITOLO IV

La posizione geografica degli istituti di provenienza: effetti e dinamiche

4.1 LA VARIABILE LATITUDINE

La creazione della variabile “latitudine” si pone come obiettivo la realizzazione di un indicatore quantitativo che possa indicare il posizionamento geografico degli istituti secondari di provenienza dei laureati.

Avendo deciso di dividere il campione su base provinciale si è provveduto a reperire in rete un elenco attendibile delle latitudini delle 103 province italiane.

In un archivio Excel sono state riportate le sigle delle singole province affiancate dal valore di latitudine riportato in gradi.

Dopo aver individuato la provincia più a sud, che presenta un valore di latitudine pari a 36.9333, nella terza colonna del foglio Excel si è sottratto questo valore alle latitudini di tutte le province, secondo la formula:

$$\text{“colonna3”} = \text{“colonna2”} - 36.9333$$

In questo modo si è assegnato valore “0” alla provincia più a sud.

Si è individuata, quindi, la provincia più a nord e utilizzando il valore della “terza colonna” (9.5667) come base per la creazione di una quarta colonna, secondo la formula:

$$\text{“colonna4”} = \text{“colonna3”} / 9.5667 * 100$$

I valori contenuti nella quarta colonna del foglio Excel indicano, in una scala di valori compresi tra 0 e 100, quanto una provincia si trova a nord. Per Ragusa, provincia più a sud, si riscontra il valore 0. Per Bolzano, provincia più a nord, si riscontra valore 100. Per tutte le altre province si riscontrano valori intermedi proporzionali alla loro latitudine.

I valori così determinati sono stati utilizzati per l’inserimento all’interno del nostro archivio di una variabile, denominata “latitudine” ottenuta assegnando ad ogni provincia dell’istituto di provenienza (variabile “pristituto”) il corrispondente valore della sopracitata “colonna4”.

4.2 EFFETTI DELLA POSIZIONE GEOGRAFICA DELL’ISTITUTO SUPERIORE SUL VOTO DI MATURITA’ E SULLA MEDIA AGLI ESAMI UNIVERSITARI.

Per poter evidenziare gli effetti della posizione geografica degli istituti superiori di provenienza degli studenti è stato necessario suddividere il campione in tre macroclassi (nord, centro e sud), basandosi sulla variabile latitudine.

Si è proceduto, quindi, ad una stratificazione degli studenti in tre gruppi omogenei rispetto alla localizzazione geografica (nord, centro e sud); in considerazione dell’esiguo numero di studenti “del nord” che abbiano conseguito la laurea in economia alla LUISS, la suddetta suddivisione è stata operata cercando anche di garantire una consistente numerosità di studenti in ogni singolo gruppo.

Le tavole 4.2.1 e 4.2.2 descrivono rispettivamente la distribuzione di frequenza della variabile “gruppi latitudine” ottenuta dalle procedure di stratificazione e la composizione dei tre gruppi denominati “nord”, “centro” e “sud” rispetto alle province di provenienza dell’istituto d’istruzione secondaria superiore frequentato dagli studenti. La figura 4.2.3 è dedicata, infine, ad una rappresentazione grafica della suddivisione operata.

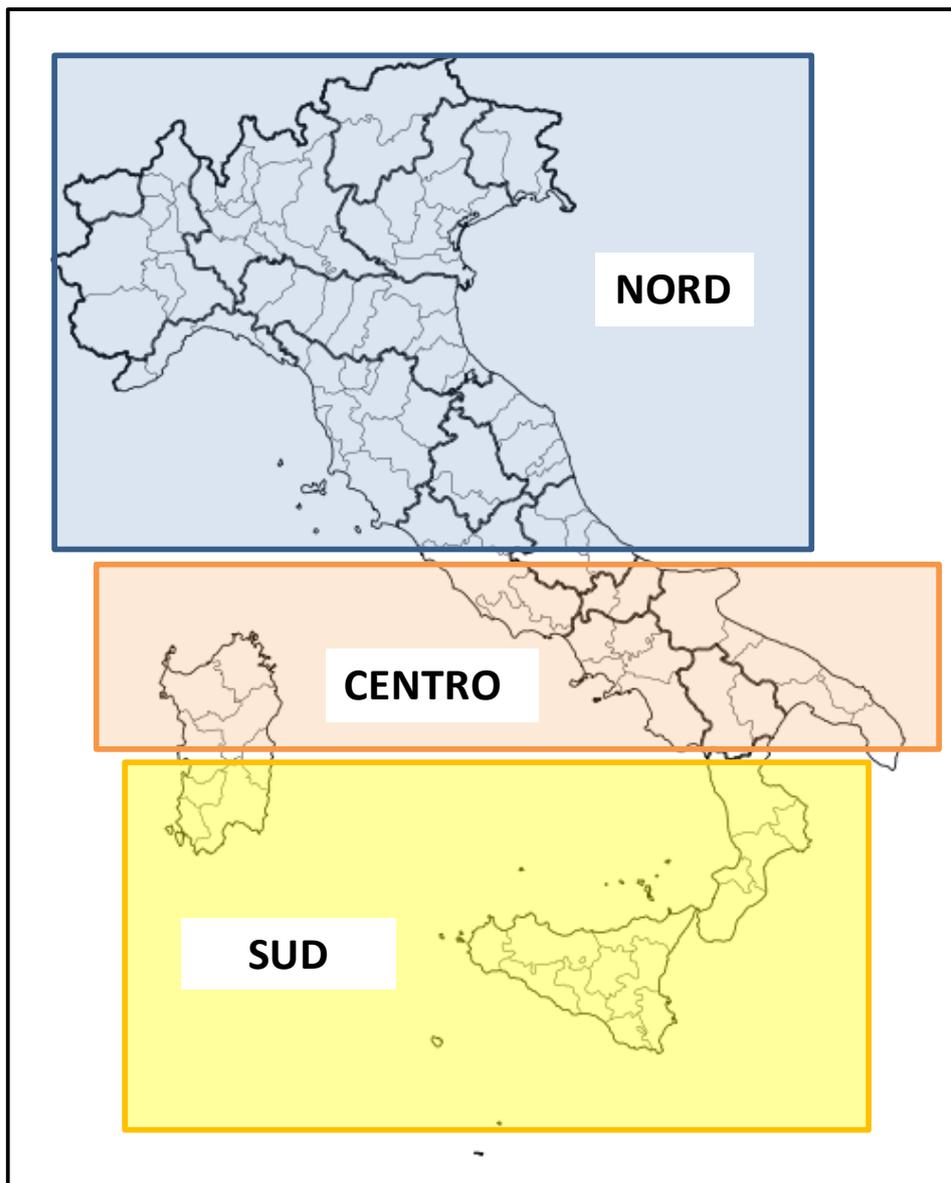
FIGURA 4.2.1:

		Gruppi latitudine			
		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	NORD	466	12,4	12,4	12,4
	CENTRO	2968	78,9	78,9	91,3
	SUD	328	8,7	8,7	100,0
	Totale	3762	100,0	100,0	

FIGURA 4.2.2:

NORD		CENTRO		SUD	
AQ	67	RM	1657	CS	75
AP	45	NA	176	CZ	42
CH	40	SA	169	RC	34
PE	35	LT	121	SR	32
VT	34	FR	110	KR	27
TE	33	FG	105	ME	24
PG	32	CE	93	CL	15
TR	32	BA	78	PA	14
RI	31	PZ	78	RG	14
AN	24	AV	76	TP	12
MC	14	LE	74	CA	11
GR	12	TA	50	VV	11
AR	7	BR	40	CT	8
BZ	7	BN	38	AG	7
FI	7	MT	30	EN	1
PS	4	IS	29	OR	1
PD	3	CB	27		
PU	3	SS	14		
TN	3	NU	3		
TO	3				
TV	3				
VE	3				
BO	2				
LI	2				
PI	2				
RA	2				
SI	2				
SV	2				
CO	1				
FE	1				
FO	1				
LU	1				
MI	1				
PO	1				
PV	1				
RE	1				
SP	1				
UD	1				
VA	1				
VI	1				

FIGURA 4.2.3:



La tavola 4.2.4 contiene la media del voto di maturità e del voto medio agli esami universitari conseguito dagli studenti nei tre gruppi costituiti rispetto la latitudine e denominati: nord, centro e sud.

TAVOLA 4.2.4:

Descrittivi									
		N	Media	Deviazione std.	Errore std.	Intervallo di confidenza 95% per la media		Minimo	Massimo
						Limite inferiore	Limite superiore		
Voto Maturita'	NORD	466	90,3970	10,16892	,47107	89,4713	91,3227	61,67	100,00
	CENTRO	2968	90,0186	10,13242	,18599	89,6540	90,3833	60,00	100,00
	SUD	328	94,4685	7,53308	,41594	93,6502	95,2868	68,00	100,00
	Totale	3762	90,4535	10,01329	,16326	90,1334	90,7736	60,00	100,00
votomedio	NORD	466	27,0093	1,63893	,07592	26,8601	27,1585	21,60	30,00
	CENTRO	2968	26,8949	1,61003	,02955	26,8370	26,9529	21,67	29,93
	SUD	328	26,6533	1,65691	,09149	26,4733	26,8332	21,31	29,63
	Totale	3762	26,8880	1,61938	,02640	26,8363	26,9398	21,31	30,00

Per quanto riguarda il voto di maturità , l'analisi della varianza con l'applicazione del test ANOVA (tavola 4.2.5) per i confronti multipli conferma l'esistenza di una significativa differenza tra le medie dei voti di maturità nei tre gruppi ($P=0,000$). Il test di Bonferroni (tavola 4.2.6), inoltre, indica che tale differenza è significativa soltanto nella classe di studenti appartenenti al sud e che votazioni più elevate sono conseguite dagli studenti del sud rispetto agli studenti del nord e del centro.

Anche relativamente al voto medio agli esami si registra una differenza significativa delle valutazioni nei tre gruppi ($P=0,008$) ed anche in questo caso il confronto multiplo indica che tale differenza è significativa soltanto nella classe di studenti appartenenti al sud.

TAVOLA 4.2.5: ANOVA

ANOVA univariata

		Somma dei quadrati	df	Media dei quadrati	F	Sig.
Voto Maturita'	Fra gruppi	5850,160	2	2925,080	29,617	,000
	Entro gruppi	371250,419	3759	98,763		
	Totale	377100,579	3761			
votomedio	Fra gruppi	25,068	2	12,534	4,789	,008
	Entro gruppi	9837,792	3759	2,617		
	Totale	9862,861	3761			

TAVOLA 4.2.6: Test di Bonferroni

Confronti multipli

Bonferroni

Variabile dipendente	(I) Gruppi latitudine	(J) Gruppi latitudine	Differenza fra medie (I-J)	Errore std.	Sig.	Intervallo di confidenza 95%	
						Limite inferiore	Limite superiore
Voto Maturita'	NORD	CENTRO	,37835	,49519	1,000	-,8077	1,5644
		SUD	-4,07150*	,71627	,000	-5,7870	-2,3560
	CENTRO	NORD	-,37835	,49519	1,000	-1,5644	,8077
		SUD	-4,44985*	,57826	,000	-5,8348	-3,0649
	SUD	NORD	4,07150*	,71627	,000	2,3560	5,7870
		CENTRO	4,44985*	,57826	,000	3,0649	5,8348
votomedio	NORD	CENTRO	,11434	,08061	,468	-,0787	,3074
		SUD	,35600*	,11660	,007	,0767	,6353
	CENTRO	NORD	-,11434	,08061	,468	-,3074	,0787
		SUD	,24166*	,09413	,031	,0162	,4671
	SUD	NORD	-,35600*	,11660	,007	-,6353	-,0767
		CENTRO	-,24166*	,09413	,031	-,4671	-,0162

*. La differenza media è significativa al livello 0.05

Possiamo, quindi, affermare che la latitudine incide:

- Sulla media aritmetica degli esami sostenuti: gli studenti del nord e del centro ottengono risultati lievemente superiori a quelli del sud.
- Sulla votazione conseguita all'esame di maturità: gli studenti del sud ottengono risultati decisamente migliori di quelli del centro e del nord.

4.3 BONTÀ DI ADATTAMENTO E STABILITÀ DEL MODELLO PER STUDENTI PROVENIENTI DA DIVERSE AREE GEOGRAFICHE

In via preliminare è stata effettuata l'analisi di correlazione del "voto medio" con le due variabili "voto di maturità" e "punteggio al test psico-attitudinale". Si è subito notato come la correlazione di entrambe più evidente al nord decresca gradualmente passando al centro ed ancora più evidentemente al sud.

Applicando il modello sul voto medio (proposto nel paragrafo 3.4) ai tre sottocampioni "nord", "centro" e "sud", si hanno inoltre differenti riscontri.

Il "modello nord" (tavola 4.3.1) risulta addirittura migliore di quello generale. Il test ANOVA conferma la significatività delle variabili dipendenti ed il coefficiente ($R^2 = 0,414$) è sensibilmente più elevato. I grafici dei residui permettono di verificare le assunzioni di omoschedasticità e normalità dei residui.

TAVOLA 4.3.1: Riepilogo del modello nord

Riepilogo del modello ^b				
Modello	R	R-quadrato	R-quadrato corretto	Deviazione standard Errore della stima
1	,646 ^a	,417	,414	1,25428

a. Predittori: (Costante), Punteggio Test ($\times 100$), Voto Maturita'

b. Variabile dipendente: votomedio

Il "modello centro" (tavola 4.3.2) risulta simile a quello generale. Il test ANOVA conferma la significatività delle variabili dipendenti ed il coefficiente ($R^2 = 0,293$) rimane sostanzialmente invariato. Anche in questo caso i grafici dei residui permettono di verificare le assunzioni di omoschedasticità e normalità dei residui.

TAVOLA 4.3.2: Riepilogo del modello centro

Riepilogo del modello^b

Modello	R	R-quadrato	R-quadrato corretto	Deviazione standard Errore della stima
1	,541 ^a	,293	,293	1,35419

a. Predittori: (Costante), Punteggio Test (x/100), Voto Maturita'

b. Variabile dipendente: votomedio

Il “modello sud” (tavola 4.3.3) risulta fortemente peggiorato rispetto a quello generale. Nonostante il test ANOVA confermi la significatività delle variabili dipendenti, il coefficiente ($R^2 = 0,205$) subisce un evidente crollo.

Inoltre, diversamente da quanto accaduto per i modelli del “nord” e del “centro”, l’analisi dei residui (figura 4.3.4) mette in evidenza una sostanziale instabilità in questo caso i grafici dei residui risultano compromessi.

TAVOLA 4.3.3: Riepilogo del modello sud

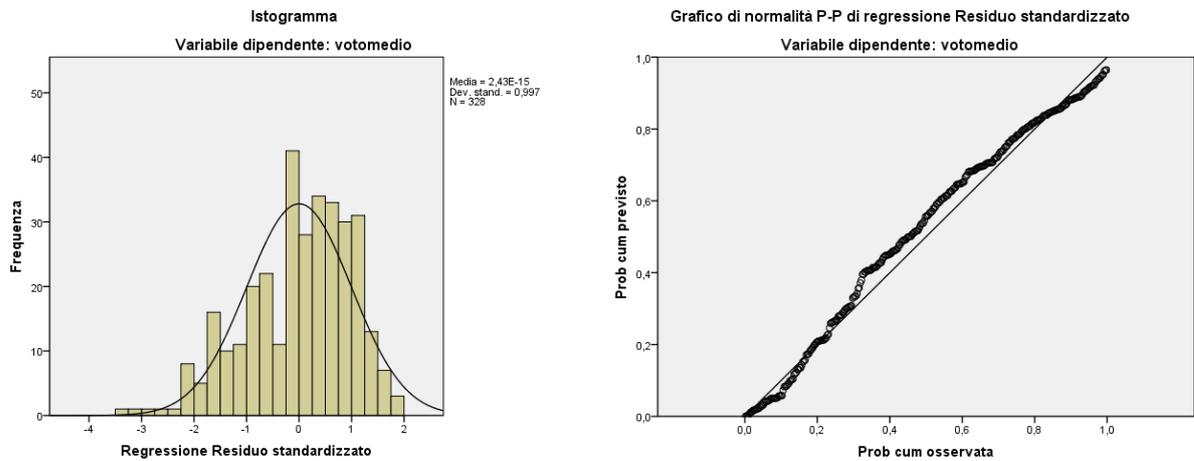
Riepilogo del modello^b

Modello	R	R-quadrato	R-quadrato corretto	Deviazione standard Errore della stima
1	,458 ^a	,210	,205	1,47729

a. Predittori: (Costante), Punteggio Test (x/100), Voto Maturita'

b. Variabile dipendente: votomedio

FIGURA 4.3.4: istogramma e Normality Probability plot



4.4 CORRELAZIONE TRA VOTO DI MATURITÀ E LATITUDINE NELLE DIVERSE FASCE DI SUCCESSO ACCADEMICO

Utilizzando i percentili si è suddiviso il campione in cinque fasce rispetto alla variabile “votomedio”, che è risultato essere il miglior indicatore di successo accademico a nostra disposizione.

La tavola 4.4.1 che riporta la distribuzione di frequenza della variabile “gruppi voto medio”, riassume efficacemente la composizione dei cinque microgruppi.

FIGURA 4.4.1:

Gruppi voto medio

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Voto medio inferiore a 25,5	771	20,5	20,5	20,5
Voto medio compreso tra 25,5 e 26,6	697	18,5	18,5	39,0
Voto medio compreso tra 26,6 e 27,5	772	20,5	20,5	59,5
Voto medio compreso tra 27,5 e 28,4	814	21,6	21,6	81,2
Voto medio superiore a 28,4	708	18,8	18,8	100,0
Totale	3762	100,0	100,0	

All'interno di ogni singola fascia è stata effettuata un'analisi di correlazione tra le variabili "latitudine" e "voto di maturità". La correlazione risulta sempre negativa e significativa. Possiamo quindi affermare che, studenti con risultati accademici simili, hanno ottenuto votazioni superiori agli esami di maturità quanto più a sud era situato l'istituto superiore di provenienza.

Come evidenziato dalla tabella che segue (tavola 4.4.2) la correlazione tende a diminuire con l'aumentare del rendimento accademico. Questo significa che nella valutazione dei "bravi" studenti le differenze tra nord e sud sono meno evidenti, mentre risultano maggiori per i "cattivi" studenti.

TAVOLA 4.4.2:

INTERVALLO	Coefficiente di Correlazione
18 - 25,5	-0,267
25,5 - 26,6	-0,241
26,6 - 27,5	-0,205
27,5 - 28,4	-0,122
28,4 - 30	-0,100

CAPITOLO V

Osservazioni e Conclusioni

5.1 STABILITÀ DEL MODELLO NELLA SEGMENTAZIONE ANNUALE DEL CAMPIONE

Durante l'analisi è stata verificata la stabilità del modello segmentando il campione in base all'anno d'immatricolazione degli studenti. La variazione degli R^2 riscontrata, sebbene spesso si rivelasse abbastanza ampia, non ha fornito evidenze che potessero indicare la non casualità del fenomeno.

L'unica osservazione degna di nota, non tanto per la sua rilevanza quanto per la sua sintomaticità deriva dall'analisi della regressione lineare tra il voto di laurea ed il punteggio al test psicoattitudinale, immatricolati nell'a.a. 96/97. In questa circostanza il coefficiente della retta di regressione risulta negativo. Sebbene l' R^2 sia insignificante (0,003) ed il modello non sia da considerarsi attendibile, è paradossale osservare che, per l'anno in questione, un alto punteggio al test è associabile ad un basso risultato accademico.

5.2 GLI INDICATORI DI SUCCESSO ACCADEMICO

Nel corso dell'analisi ci siamo trovati a dover individuare il miglior indicatore di successo accademico.

Il voto di laurea, che è stato lungamente analizzato nel terzo capitolo, oltre ad aver mostrato problematiche di stabilità nel modello, risente in maniera eccessiva degli "arrotondamenti" della commissione.

Quando è stata considerata la media ponderata dei voti degli esami verbalizzati, si è dovuto scegliere tra la media aritmetica e la media ponderata. Sono state tentate entrambe le

strade, ma il modello creato con la media ponderata si è dimostrato essere meno stabile e significativo. Inoltre, se anche il modello fosse risultato migliore, si sarebbe dovuto affrontare il problema della non completa confrontabilità degli studenti pre- e post- riforma. Per gli studenti pre-riforma si sarebbe infatti comunque dovuto utilizzare la media aritmetica, non esistendo ancora il sistema dei crediti.

Inizialmente si era pensato che il percorso accademico comune potesse appianare le differenze tra gli studenti. Sia il voto di maturità, sia il test psicoattitudinale avrebbero dovuto predire con efficacia migliore i risultati dei primi esami affrontati dagli studenti. Nel tentativo di verificare questa ipotesi è stato creato un modello che avesse come variabile dipendente la media aritmetica dei primi cinque esami verbalizzati da ogni singolo studente corrispondenti sostanzialmente agli esami del primo semestre del primo anno accademico. I risultati di questa parte di analisi non sono stati riportati perché scarsamente significativi. Basti sapere che l'ipotesi formulata può essere decisamente respinta: sia il test, sia il voto di maturità spiegano molto meglio il successo dello studente piuttosto che le sue abilità acquisite.

Per rendere completa l'analisi si è verificato anche quanto il voto di maturità ed il punteggio del test potessero essere utilizzati per prevedere la tendenza degli studenti a laurearsi in corso. Un voto alto alla maturità sembra favorire la celerità dello studente nel conseguire la laurea, il punteggio al test è risultato essere praticamente non correlato.

5.3 I PREDITTORI DI SUCCESSO ACCADEMICO

La scelta di focalizzare la nostra attenzione sul voto di maturità, piuttosto che sul punteggio attribuito alla carriera non è certo casuale. Confrontando i modelli creati con le due differenti variabili ci siamo subito resi conto che il voto di maturità risulta essere un predittore di successo accademico migliore del punteggio carriera. Considerando che il punteggio carriera è frequentemente determinato proprio dal voto di maturità, se ne deduce che gli altri indicatori utilizzati (ad esempio le medie dei voti conseguiti in terza e quarta) sono predittori peggiori. Questa tesi è stata anche avvalorata dalle verifiche di stabilità del modello svolte stratificando il campione per anno d'immatricolazione. Negli anni

in cui il “punteggio carriera” non è determinato dal voto di maturità, il modello perde di efficacia.

5.4 CONCLUSIONI

La nostra analisi ci porta a dover fare considerazioni differenti in riferimento alle due componenti del punteggio totale del test d'ingresso (test psicoattitudinale e punteggio attribuito alla carriera pre-universitaria), con il quale vengono stilate le classifiche per l'ammissione alla facoltà di economia della LUISS.

Per quanto riguarda il test psicoattitudinale, i risultati sono al quanto deludenti. Così come strutturato, il test risulta avere scarsa rilevanza nella determinazione del successo accademico dello studente. Dall'analisi è evidente la totale assenza di correlazione tra il test psicoattitudinale e il voto di maturità. Questo potrebbe indicare l'utilità del test nel mettere in luce qualità diverse da quelle espresse dal voto di maturità. Questa deduzione, però, appare forzata: non bisogna dimenticare infatti che un elevato valore della somma delle due variabili è *conditio sine qua non* per l'ammissione in LUISS e che la nostra analisi non poteva tenere in considerazione gli studenti non ammessi.

Uno dei principali limiti dell'analisi svolta è riscontrabile proprio nell'impossibilità di reperire i dati per la creazione di un gruppo di controllo.

Non sembra essere pienamente condivisibile, pertanto, la scelta di utilizzare il test psicoattitudinale (almeno così com'è strutturato) per la selezione degli studenti da ammettere ai corsi di laurea.

Per quanto riguarda il punteggio attribuito alla carriera pre-universitaria, è necessario fare delle considerazioni diverse.

Innanzitutto, possiamo affermare con certezza che il voto di maturità rappresenta il migliore predittore di successo accademico e dovrebbe essere preferito agli altri criteri che nel corso degli anni sono stati utilizzati per la determinazione del “punteggio carriera”.

Inoltre si riscontra che il voto di maturità pesa quasi il 30% nella determinazione della media dei voti conseguiti nel percorso accademico e risulta quindi essere un indicatore abbastanza rilevante.

Sembra essere opportuna la scelta di utilizzare questo parametro per la selezione degli studenti da ammettere, sebbene anche in questo parametro si possano trovare delle controindicazioni.

E' stato dimostrato come in Italia non vi sia uniformità nei criteri di valutazione degli studenti agli esami di stato. Si è visto come, a parità di preparazione al termine della carriera accademica, al sud vi sia la tendenza a assegnare valutazioni più elevate rispetto al centro/nord.

Utilizzando il parametro senza correttivi si rischia quindi di facilitare l'ingresso di studenti meno validi provenienti dal sud a scapito di studenti migliori provenienti dal centro/nord.

APPENDICE

CRITERI GENERALI DI FORMAZIONE DELLE GRADUATORIE**PROVE DI AMMISSIONE AI CORSI DI LAUREA TRIENNALE**

ANNO ACCADEMICO	APRILE		SETTEMBRE	
	PESO TEST	PESO CARRIERA	PESO TEST	PESO CARRIERA
1996/1997	(prova solo a settembre)		risultati test (50%)	media III, IV e voto di maturità (50%)
1997/1998	(prova solo a settembre)		risultati test (50%)	media III, IV e voto di maturità (50%)
1998/1999	(prova solo a settembre)		risultati test (50%)	media III, IV e voto di maturità (50%)
1999/2000	(prova solo a settembre)		risultati test (60%)	voto di maturità (40%)
2000/2001	(prova solo a settembre)		risultati test (60%)	voto di maturità (40%)
2001/2002	(prova solo a settembre)		risultati test (60%)	voto di maturità (40%)
2002/2003	(prova solo a settembre)		risultati test (70%)	voto di maturità (30%)
2003/2004	risultati test (50%)	media voti penultimo e terzultimo anno di scuola (50%)	risultati test (50%)	voto di maturità (50%)
2004/2005	risultati test (50%)	media voti penultimo e terzultimo anno di scuola (50%)	risultati test (50%)	voto di maturità (50%)
2005/2006	risultati test (50%)	media voti penultimo e terzultimo anno di scuola (50%)	risultati test (50%)	media voti penultimo e terzultimo anno di scuola (50%)

2006/2007	risultati test (50%)	media voti penultimo e terzultimo anno di scuola (50%)	risultati test (50%)	media voti penultimo e terzultimo anno di scuola (50%)
2007/2008	risultati test (50%)	media voti penultimo e terzultimo anno di scuola (50%)	risultati test (50%)	media voti penultimo e terzultimo anno di scuola (50%)
2008/2009	risultati test (40%)	media voti penultimo e terzultimo anno di scuola (60%)	risultati test (40%)	voto di maturità (60%)

N.B. Per gli anni evidenziati non è stato possibile trovare alcun documento che potesse comprovare la veridicità delle informazioni.

BIBLIOGRAFIA

- "The validity of the Medical College Admission Test for predicting performance in the first two years of medical school" (Wiley A. & Koenig J.A. - Association of American Medical Colleges – 1996)
- "Correlation of Admissions Criteria with Academic Performance in Dental Students"(Donald A. Curtis; Samuel L. Lind; Octavia Plesh; Frederick C. Finzen)
- Anna Clara Monti (2008): Introduzione alla statistica. Napoli: edizioni Scientifiche Italiane.
- Luigi Fabbris (1997): Statistica multivariata: analisi esplorativa dei dati Ed. McGraw-Hill, Milano
- Hair J.F., Anderson R.E., Tatham R.L., Black W.C. (1998): Multivariate data analysis, Prentice Hall International Edition, New Jersey, USA.
- Giovanni Di Franco (2009) L'analisi dei dati con spss, Franco Angeli, Roma.
- Andy Field (2000): Discovering Statistics using SPSS for windows, Cromwell Press Ltd, Trowbritge, Wiltshire.
- <http://www.matematicamente.it/staticfiles/approfondimenti/astronomia/CoordGeogProvince.pdf>