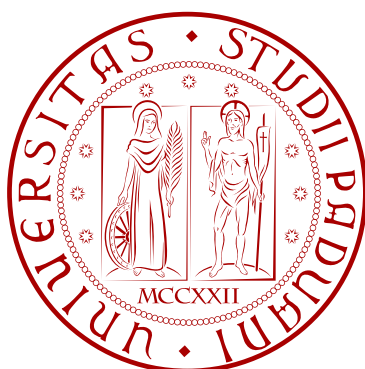


UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA SPECIALISTICA
IN SCIENZE STATISTICHE, ECONOMICHE,
FINANZIARIE E AZIENDALI



TESI DI LAUREA

**LA DATA DI NASCITA INFLUENZA
IL TITOLO DI STUDIO? UN'ANALISI
DEL CONTESTO ITALIANO**

RELATORE: PROF. ADRIANO PAGGIARO

LAUREANDA: BARBARA SARTORI

ANNO ACCADEMICO 2009/2010

Alla mia famiglia

Indice

1	Introduzione	1
2	Gli effetti del mese di nascita	3
2.1	Il concetto di <i>relative age</i>	3
2.2	Rassegna della precedente letteratura sull'argomento	5
2.2.1	La <i>relative age</i> nello sport	5
2.2.2	La <i>relative age</i> nella scuola	6
2.2.3	La <i>relative age</i> e alcune patologie mediche	7
2.3	Breve storia della scuola Italiana	8
3	Il dataset	13
3.1	La data di nascita	14
3.2	Il titolo di studio	21
4	L'analisi dei dati	25
4.1	Il titolo di studio nel tempo	26
4.2	Il modello	30
4.3	L'analisi	32
4.3.1	La scuola dell'obbligo	32
4.3.2	L'università	38

4.4	Interpretazione dei risultati ottenuti	42
5	Controllo della robustezza delle analisi	45
5.1	Modifiche del campione	45
5.2	Modifiche del modello	50
6	Conclusioni	55
	Bibliografia	57

Capitolo 1

Introduzione

Lo scopo di questa tesi è quello di verificare se c'è un effetto del mese di nascita sul titolo di studio conseguito da una persona. In particolare si vuole analizzare la legislazione italiana in materia di iscrizione alla scuola dell'obbligo.

In Italia vengono normalmente iscritti al primo anno di scuola elementare i bambini e le bambine che compiono i sei anni entro il 31 Dicembre dell'anno di riferimento; tra i bambini più grandi e quelli più piccoli di ciascuna classe vi sono quindi ben 11 mesi di differenza. La vasta letteratura sull'argomento ha mostrato che questa differenza di età (chiamata solitamente *relative age*) è tutta a vantaggio dei bambini più grandi che, rispetto ai loro compagni più piccoli, tendono a riuscire meglio sia nella scuola che negli sport. Questa differenza è stata spiegata con la maggiore maturità di questi bambini, ma man mano svanisce col tempo. Il dibattito sulla durata di questo vantaggio però è ancora aperto, infatti la bibliografia sull'argomento è molto vasta e porta a conclusioni parecchio differenti.

Oltre alla presenza di tale effetto, in questo lavoro si è cercato di capire,

qualora fosse presente, fino a quando portava i propri vantaggi. Partendo dal presupposto che siano i bambini più bravi a proseguire gli studi, l'ipotesi iniziale era dunque quella che tra i laureati vi fossero più persone nate nei primi mesi dell'anno e che tra coloro che terminavano gli studi con la scuola dell'obbligo vi fossero più persone nate negli ultimi mesi dell'anno.

La tesi è strutturata nel seguente modo: nel secondo capitolo verranno presentate le principali evidenze della letteratura sull'argomento in questione, nel capitolo terzo verranno illustrati i dati utilizzati per l'analisi, nel quarto capitolo si mostreranno i risultati ottenuti ed infine nel quinto capitolo si presenteranno delle verifiche sulla robustezza delle analisi.

Capitolo 2

Gli effetti del mese di nascita

In questo capitolo verrà prima presentato il concetto di *relative age* e i precedenti studi sull'argomento e in seguito si illustrerà l'evoluzione della scuola in Italia e la relativa legislazione in materia di scuola dell'obbligo.

2.1 Il concetto di *relative age*

Fin da piccoli i bambini vengono suddivisi, sia per la formazione delle classi scolastiche sia per la partecipazione ad attività sportive, in gruppi omogenei per età. Nonostante quasi sempre questa ripartizione venga fatta per anno di nascita (ogni gruppo, cioè, è formato dai bambini nati in uno stesso anno solare) le differenze tra i bambini sono notevoli. Tra i più grandi e i più piccoli, infatti, vi sono praticamente 11 mesi di differenza che, data la loro giovane età, sono ben visibili sia in termini di sviluppo che in termini di prontezza allo svolgimento dell'attività.

Questa differenza di età tra bambini dello stesso gruppo, dovuta alla distribuzione delle date di nascita lungo tutto l'anno, viene definita *relative age*.

La legge italiana prevede che si iscrivano in prima elementare i bambini e le bambine che compiono i 6 anni di età entro il 31 Dicembre dell'anno di riferimento. Possono altresì iscriversi i bambini e le bambine che compiono i 6 anni entro il 30 Aprile dell'anno solare successivo. In generale, però, vengono iscritti i bambini nati entro il 31 Dicembre dell'anno preso in considerazione e raramente quelli nati tra l'inizio del successivo anno e il 30 Aprile, anche perché questa possibilità è vincolata al numero di posti disponibili nella scuola (infatti all'atto di iscrizione hanno precedenza i bambini che compiono i 6 anni entro l'anno). In questo caso, quindi, i bambini più grandi sono quelli nati in Gennaio e i più piccoli quelli nati in Dicembre (la scadenza per iscriversi a scuola cambia da Stato a Stato, per esempio in molte nazioni il termine massimo è dato dal primo Settembre rendendo così i bambini nati in Agosto i più piccoli della classe).

Diversi studi hanno associato questa età relativa ad una varietà di indicatori di performance (come ad esempio i risultati a test scolastici o i successi sportivi) e a svariate patologie, mostrando come i bambini più grandi all'interno del proprio gruppo d'età ottenessero dei risultati scolastici migliori rispetto ai loro coetanei con un'età relativa minore.

Con la presente tesi si è cercato di capire se questo fenomeno, noto appunto come *relative age effect*, influenzi anche il numero di anni di studio di un individuo e quindi il titolo di studio conseguito. Partendo dal presupposto che fossero i bambini più bravi e quindi più dotati a proseguire gli studi, ci si è chiesti se questo vantaggio dovuto all'età relativa fosse così forte e duraturo da influenzare anche le scelte scolastiche oltre l'obbligo scolastico e quindi se fossero le persone nate durante i primi mesi dell'anno a continuare con più

probabilità gli studi.

2.2 Rassegna della precedente letteratura sull'argomento

Come appena accennato, diverse ricerche hanno analizzato gli effetti del raggruppamento dei bambini per età per la partecipazione a diverse attività. In questo paragrafo verranno riportate le principali pubblicazioni sul tema.

2.2.1 *La relative age nello sport*

Il *relative age effect* nello sport fu notato la prima volta tra i giocatori di hockey. Barnsley *et al.* (1985) hanno analizzato sia la National Hockey League (NHL) che le principali leghe minori professionistiche e hanno notato una forte relazione lineare tra il mese di nascita (da Gennaio a Dicembre) e il numero di giocatori di ciascuna lega. Infatti la maggior parte degli atleti era nata durante i primi tre mesi dell'anno, che nell'hockey corrispondono anche ai primi tre mesi del campionato. In seguito a questa scoperta Barnsley e Thompson (1988) hanno studiato le leghe minori di hockey per cercare di scoprire come mai questa relazione fosse così forte. Questo studio ha mostrato che i bambini con un'età relativa maggiore (all'interno della loro lega) continuano a giocare a hockey per più tempo rispetto ai loro compagni più piccoli, che più spesso abbandonano questo sport, e che nelle squadre designate a partecipare ad allenamenti intensivi e a vari tornei vi sono molti più giocatori nati all'inizio dell'anno solare.

Una possibile spiegazione di questo risultato è stata trovata nel maggiore

sviluppo (sia fisico che di coordinazione) dei bambini più grandi durante i primi anni di attività sportiva. Questa maggiore maturità spesso viene scambiata per maggiore abilità dagli allenatori e quindi è più probabile che questi bambini vengano incoraggiati a proseguire l'attività sportiva fino alle leghe maggiori.

Queste scoperte hanno promosso diversi studi su altre discipline sportive: per esempio Thompson *et al.* (1991) hanno studiato il baseball, Barnsley *et al.* (1992) e successivamente Dudink (1994) il calcio ed infine Glamser e Marciani (1992) il football. Tutte queste ricerche hanno portato a conclusioni simili: il relative age effect è evidente in tutti gli sport in cui i bambini sono suddivisi per gruppi di età.

2.2.2 La *relative age* nella scuola

La suddivisione dei bambini in classi mostra molti tratti in comune con la partecipazione dei bambini alle attività sportive: in entrambi i casi, infatti, i bambini vengono raggruppati per età; è quindi facile presumere che lo stesso effetto dovuto all'età relativa dei bambini presente nello sport vi sia anche in materia di risultati scolastici.

Già nel 1980 Dipasquale *et al.* avevano notato che il numero di bambini con un'età relativa inferiore che durante i primi anni di scuola doveva rivolgersi ai Servizi Psicologici era nettamente maggiore del numero dei loro compagni più grandi.

Angrist e Krueger (1991) hanno invece analizzato diversi dataset provenienti da tre censimenti degli Stati Uniti e hanno stabilito che i bambini nati durante i primi mesi dell'anno (e che quindi sono i più grandi della classe) hanno

un livello medio di educazione più basso in quanto possono lasciare la scuola prima dei loro compagni più piccoli (in America, infatti, la legge permette ai ragazzi di abbandonare la scuola non appena hanno compiuto una certa età, tipicamente i 16 o i 17 anni).

Fredriksson e Ockert (2005) per la Svezia e Puhani e Weber (2005) per la Germania hanno studiato gli effetti di un aumento dell'età minima per l'iscrizione a scuola e sono giunti entrambi alla conclusione che posticipare di un anno l'entrata a scuola porta i bambini sia ad ottenere risultati migliori sia ad avere un'educazione maggiore.

Bedard e Dhuey (2006), oltre ad aver mostrato che i bambini con un'età relativa maggiore ottengono risultati migliori, hanno notato che queste differenze a favore dei bambini più grandi sono ancora presenti e significative all'ottavo grado (13-14 anni) di istruzione. La durata di questo effetto è probabilmente amplificata dal sistema scolastico Americano (e più in generale Anglosassone), in quanto durante i primi anni di scuola i bambini vengono suddivisi in classi a seconda del loro livello di abilità: i bambini più bravi nei primi anni di istruzione (cioè quelli più grandi) vengono indirizzati verso curriculum avanzati e quindi hanno maggiori possibilità di ottenere una migliore istruzione.

2.2.3 La *relative age* e alcune patologie mediche

L'età relativa non influenza solo i risultati accademici o quelli sportivi, ma ha effetto anche sulla salute degli individui. Goodman *et al.* (2003) hanno studiato un campione di ragazzi dai 5 ai 15 anni nati in Inghilterra, Galles e Scozia e hanno concluso che i bambini più piccoli di ciascun anno scolastico

sono più a rischio di malattie psichiatriche. In precedenza questo tipo di malattie, come la schizofrenia e il disturbo bipolare, erano state associate alla stagione di nascita (*birthdate effect*) più che all'età relativa degli individui (si veda per esempio Boyd *et al.* (1986)). Questo effetto però non è dovuto solo al mese di nascita ma più probabilmente ai mesi della gravidanza. Watson e McDonald (2007), infatti, hanno dimostrato che una differenza di nutrienti della madre durante la gravidanza dovuta alla stagionalità, influenza lo sviluppo cellulare del feto e ciò li porta a concludere che le diverse malattie dovute alla stagione di nascita siano quindi influenzate da questa variazione stagionale dei nutrienti durante la gestazione.

Nel paragrafo precedente si è detto che i bambini più piccoli all'interno della loro classe hanno dei risultati accademici inferiori rispetto ai loro compagni maggiori, secondo Thompson *et al.* (1999) questa inferiorità li porta ad avere meno stima in sé stessi e in alcuni casi alla depressione: queste sono le spiegazioni trovate dagli autori al fatto che la maggior parte di suicidi dei ragazzi con meno di vent'anni in una città canadese sia stata compiuta dai giovani nati durante gli ultimi sei mesi dell'anno scolastico e quindi più piccoli.

2.3 Breve storia della scuola Italiana

L'istruzione in Italia, come nel resto dell'Europa, è sempre stata affidata alla Chiesa fin dal Medioevo; i primi esempi di scuola laica si hanno a partire dal XII secolo con la nascita di scuole private e di scuole comunali. Nel corso del XIII secolo cominciano ad apparire anche scuole di secondo grado (per alunni già alfabetizzati) che si dividevano in scuole d'abaco (in cui si apprendevano le tecniche di calcolo) e scuole di grammatica (in cui venivano studiati gli

autori classici e la lingua latina). Durante lo stesso periodo cominciarono a tenersi delle *lectiones magistrales* in cui si discuteva principalmente di filosofia aristotelica. Queste lezioni si diffusero molto rapidamente in tutta Europa, acquistando ben presto il carattere di vere e proprie riunioni assembleari che richiesero ben presto un'organizzazione più razionale. Dall'istituzionalizzazione di questi nuclei nacquero così le prime università e nel 1158 Federico I promulga la *constitutio habita* a seguito della quale l'università diventa un luogo in cui la ricerca si sviluppa in modo indipendente da ogni altro potere. Si è però dovuto aspettare il Settecento e l'Illuminismo affinché il concetto di istruzione assomigliasse al concetto moderno, infatti in questo periodo l'istruzione primaria viene concepita come pubblica, obbligatoria e gratuita, in cui tutti i cittadini (sia maschi che femmine) devono accedervi (Genovesi, 1998). Verso i primi anni del 1800 nacquero i primi licei.

Alla vigilia dell'unità d'Italia, la legge Casati nel 1859 istituiva una scuola elementare formata da due bienni (di cui il primo obbligatorio) e un'istruzione secondaria formata dai licei (a pagamento) e dalle scuole tecniche. Con l'arrivo del Novecento si hanno i primi effetti positivi del sistema e nel 1904 con la legge Orlando si prolungò l'obbligo scolastico fino al dodicesimo anno di età, aumentando gli anni della scuola elementare con le classi quinta e sesta.

Durante il primo governo Mussolini vi fu la riforma Gentile (Regio Decreto n.1054 del 6 Maggio 1923) che in quattro decreti modificava la struttura della scuola primaria ripartendola in un *grado preparatorio*, cioè una scuola materna non obbligatoria, un *grado inferiore* formato dalla I alla III classe, un *grado superiore* formato dalla IV e dalla V classe ed infine da un "doppio

canale”: una scuola media che permetteva il proseguimento degli studi e un *corso integrativo* post-elementare senza sbocchi ulteriori (dalla VI alla VIII classe) che i comuni avevano la facoltà di istituire addossandosene tutti gli oneri. L’obbligo scolastico fu portato così a 14 anni, secondo le direttive di Washington, che avevano fissato di elevare l’età minima degli studi in tutto il mondo (per maggiori informazioni si veda Semeraro, 1996). Tale innalzamento degli anni di scuola restò comunque solo teorico in quanto l’istruzione elementare si interrompeva formalmente con la classe V, dato che solo il primo ciclo poteva dirsi presente su tutto il territorio italiano. La riforma Gentile toccò anche il tema dell’Università, infatti il decreto impose il liceo classico come scuola “principale” che dava accesso a tutte le facoltà.

Per un effettivo prolungamento della scuola dell’obbligo bisogna aspettare il *Piano per lo sviluppo della scuola nel triennio dal 1962 al 1965* presentato alla Camera il 18 Maggio 1962 da Luigi Gui (ministro della pubblica istruzione fino al 1968): la riforma principale di questo piano fu quella dell’unificazione della scuola media con la legge n.1859 del 31 Dicembre 1962, che faceva in modo che questa scuola media unica sostituisse qualsiasi altro tipo di scuola secondaria inferiore e che, in ottemperanza all’art.34 comma 2 della Costituzione (il quale imponeva un obbligo di frequenza scolastica di almeno 8 anni), fosse gratuita e obbligatoria per tutti i ragazzi dagli 11 ai 14 anni.

Nel 1967 compaiono in Italia i primi episodi di rivolta studentesca che portano, con il decreto del presidente della Repubblica n.1236 del 31 Ottobre 1969, alla liberalizzazione degli accessi all’Università eliminando il vincolo imposto da Gentile sul passaggio obbligatorio per il liceo classico per iscriversi all’U-

niversità.

Per l'innalzamento dell'obbligo scolastico a 16 anni si deve aspettare la finanziaria del 2007 (legge n.296 del 27 Dicembre 2006), in quanto, nonostante numerose discussioni in merito, fino a quell'anno era un "diritto all'istruzione fino ai 16 anni".

Capitolo 3

Il dataset

I dati utilizzati per questa analisi sono stati ricavati dalla Rilevazione Continua sulle Forze Lavoro. Tale indagine viene condotta dal 1959 dall'Istat con lo scopo di ottenere informazioni sulla situazione lavorativa, sulla ricerca di lavoro e sugli atteggiamenti verso il mercato del lavoro della popolazione in età lavorativa (De Vitiis *et al.*, 2005). In conformità con le direttive dell'Unione Europea, l'inchiesta ha carattere continuativo: viene cioè eseguita durante tutte le settimane dell'anno (fino al 2003, invece, aveva cadenza trimestrale e veniva eseguita la prima settimana senza festività dei mesi di Gennaio, Aprile, Luglio e Ottobre).

Lo schema di rotazione del campione è di tipo (2,2,2): le famiglie, cioè, partecipano all'indagine per due trimestri consecutivi, poi, dopo una pausa di due trimestri, prendono parte nuovamente all'indagine per altri due trimestri. Gli intervistatori, professionisti selezionati e formati dall'Istat, utilizzano una tecnica mista per somministrare i questionari: la prima intervista viene svolta con tecnica CAPI (*Computer Assisted Personal Interviewing*) in modo tale da instaurare un rapporto con l'intervistato favorendone la collaborazione

e limitando le mancate risposte, le successive con tecnica CATI (*Computer Assisted Telephonic Interviewing*) per ridurre sia la molestia statistica sull'unità campionaria sia i costi.

Per maggiori informazioni sulla rilevazione si veda Istat (2006).

3.1 La data di nascita

I dati disponibili provengono dalle dodici rilevazioni avvenute negli anni 2005, 2006 e 2007. Come prima cosa si sono unite in sequenza temporale tutte le interviste. Il modo in cui è stato formato il dataset implica che alcuni individui siano presenti più di una volta nel campione, infatti ogni individuo ha svolto nell'arco dei tre anni presi in considerazione per l'analisi da una a quattro interviste, a seconda di quando è entrato a far parte del campione. Per poter disporre di una elevata numerosità, si è scelto di tenere tutte le occasioni di colloquio con ciascun soggetto, in quanto si è verificato che questa scelta non influisce sulle evidenze principali (l'analisi tenendo un solo questionario per persona verrà brevemente presentata nel capitolo 5).

Dato che la variabile di maggior interesse per lo studio in questione è la data di nascita, si è deciso di tenere solo gli individui per cui questa informazione fosse completamente disponibile e quindi sono state eliminate tutte le persone per cui mancava anche solo una parte (giorno, mese o anno) di tale data. Sono stati cancellati, inoltre, 22 individui per cui la data di nascita non era ammissibile (29 Febbraio di anni non bisestili). Nella Tabella 3.1 vengono elencate le numerosità e i dati mancanti di ciascun questionario.

Si è poi cercato di capire come siano distribuite nell'arco dell'anno le nascite

Tabella 3.1: Numerosità e dati mancanti di ciascun questionario, RCFL, 2005-2007

	rcfl_051		rcfl_052		rcfl_053		rcfl_054	
	N	%	N	%	N	%	N	%
Numerosità iniziale	182,011		177,723		170,267		174,371	
Solo il giorno mancante	281	0.1544	214	0.1204	310	0.1821	397	0.2277
Solo il mese mancante	33	0.0181	37	0.0208	48	0.0282	48	0.0275
Solo l'anno mancante	18	0.0099	21	0.0118	13	0.0076	17	0.0097
Giorno e mese mancanti	376	0.2066	232	0.1305	384	0.2255	596	0.3418
Giorno e anno mancanti	6	0.0033	3	0.0017	4	0.0023	12	0.0069
Mese e anno mancanti	0	0	0	0	1	0.0006	0	0
Giorno, mese e anno	144	0.0791	132	0.0743	162	0.0951	157	0.0900
<i>Totale dati mancanti</i>	858	0.4714	639	0.3595	922	0.5415	1,227	0.7037
<i>Numero di date inesistenti</i>	3	0.0016	6	0.0034	6	0.0035	4	0.0023
Numerosità finale	181,150	99.5270	177,078	99.6371	169,339	99.4550	173,140	99.2940

	rcfl_061		rcfl_062		rcfl_063		rcfl_064	
	N	%	N	%	N	%	N	%
Numerosità iniziale	171,821		173,363		169,393		169,726	
Solo il giorno mancante	254	0.1478	146	0.0842	167	0.0986	179	0.1055
Solo il mese mancante	22	0.0128	18	0.0104	19	0.0112	22	0.0130
Solo l'anno mancante	13	0.0076	15	0.0087	18	0.0106	619	0.3647
Giorno e mese mancanti	435	0.2532	269	0.1552	426	0.2515	627	0.3694
Giorno e anno mancanti	4	0.0023	6	0.0035	5	0.0030	2	0.0012
Mese e anno mancanti	0	0	0	0	0	0	2	0.0012
Giorno, mese e anno	133	0.0774	112	0.0646	151	0.0891	288	0.1697
<i>Totale dati mancanti</i>	861	0.5011	566	0.3265	786	0.4640	1,739	1.0246
<i>Numero di date inesistenti</i>	1	0.0006	2	0.0012	0	0	0	0
Numerosità finale	170,959	99.4983	172,795	99.6724	168,607	99.5360	167,987	98.9754

	rcfl_071		rcfl_072		rcfl_073		rcfl_074	
	N	%	N	%	N	%	N	%
Numerosità iniziale	174,362		171,496		165,699		166,189	
Solo il giorno mancante	164	0.0941	129	0.0752	126	0.0760	152	0.0915
Solo il mese mancante	15	0.0086	11	0.0064	10	0.0060	15	0.0090
Solo l'anno mancante	7	0.0040	13	0.0076	13	0.0078	11	0.0066
Giorno e mese mancanti	534	0.3063	288	0.1679	402	0.2426	602	0.3622
Giorno e anno mancanti	2	0.0011	1	0.0006	1	0.0006	4	0.0024
Mese e anno mancanti	0	0	0	0	0	0	0	0
Giorno, mese e anno	318	0.1824	226	0.1318	196	0.1183	227	0.1366
<i>Totale dati mancanti</i>	1,040	0.5965	668	0.3895	748	0.4514	1,011	0.6083
<i>Numero di date inesistenti</i>	0	0	0	0	0	0	0	0
Numerosità finale	173,322	99.4035	170,828	99.6105	164,951	99.5486	165,178	99.3917

del campione ottenuto; i risultati sono mostrati in Tabella 3.2.

Tabella 3.2: **Distribuzione delle nascite nell'anno**

	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Luglio	Agosto	Settembre	Ottobre	Novembre	Dicembre	Totale
1	12,223	6,457	6,490	6,229	6,090	6,113	6,088	6,163	6,118	6,410	5,730	5,909	80,020
2	7,811	6,598	6,023	5,763	5,724	5,450	5,597	5,925	5,561	6,084	4,928	5,252	70,716
3	6,447	5,749	6,642	5,680	5,648	5,684	5,645	5,618	5,659	5,606	5,578	5,176	69,132
4	5,945	5,712	5,996	5,965	5,579	5,382	5,910	5,647	5,350	6,043	5,738	5,341	68,608
5	5,936	5,907	5,903	5,636	6,042	5,760	5,472	5,717	5,421	5,862	5,351	5,328	68,335
6	5,793	5,783	5,767	5,709	5,533	5,941	5,791	5,622	5,566	5,586	5,522	5,420	68,033
7	5,550	5,898	5,890	5,661	5,317	5,461	5,825	5,594	5,394	5,821	5,382	5,304	67,097
8	5,770	5,680	6,119	5,641	5,958	5,367	5,613	5,944	5,804	5,664	5,528	5,569	68,657
9	5,633	5,576	5,470	5,285	5,904	5,336	5,486	5,580	5,473	5,440	5,514	5,095	65,792
10	5,974	6,203	6,008	5,685	5,905	5,725	5,886	5,825	5,809	6,320	5,618	5,505	70,463
11	5,457	5,775	5,526	5,309	5,599	5,257	5,551	5,409	5,655	5,478	5,914	5,201	66,131
12	5,967	6,283	5,888	5,328	5,663	5,410	5,608	5,640	5,512	5,449	5,468	5,582	67,798
13	5,530	5,720	5,896	5,484	5,431	5,875	5,409	5,287	5,441	5,520	5,287	5,478	66,358
14	5,625	6,115	5,698	5,861	5,687	5,318	5,533	5,585	5,826	5,414	5,480	5,172	67,314
15	6,022	5,762	5,891	5,617	5,742	5,562	5,477	5,673	6,160	5,720	5,362	5,340	68,328
16	5,825	5,884	5,708	5,658	5,744	5,901	5,505	5,461	6,132	5,642	5,528	5,418	68,406
17	5,667	5,527	5,426	5,189	5,346	5,246	5,583	5,086	5,495	5,221	5,013	4,898	63,697
18	5,865	6,135	5,890	5,707	5,621	5,692	5,803	5,628	5,643	5,697	5,539	5,364	68,584
19	5,505	5,852	6,130	5,573	5,589	5,471	5,529	5,326	5,981	5,569	5,470	5,186	67,181
20	6,109	5,957	5,757	5,466	5,797	5,988	5,697	5,662	5,923	5,852	5,500	5,252	68,960
21	6,127	5,401	5,731	5,570	5,695	5,346	5,616	5,507	5,602	5,228	5,274	5,322	66,419
22	5,468	5,655	5,399	5,392	5,653	5,299	5,750	5,535	6,081	5,512	5,239	5,269	66,252
23	5,946	5,623	5,835	5,253	5,375	5,602	5,582	5,306	5,938	5,469	5,381	4,989	66,299
24	5,764	5,830	5,514	5,767	5,629	5,870	5,653	5,385	5,986	5,441	5,492	4,973	67,304
25	5,879	5,879	5,773	5,200	5,458	5,480	5,688	5,248	5,811	5,513	5,434	4,675	66,038
26	5,853	5,713	5,725	5,404	5,484	5,551	5,614	5,389	5,944	5,506	5,322	4,783	66,288
27	5,879	5,626	5,420	5,576	5,429	5,387	5,740	5,346	5,650	5,447	5,318	4,828	65,646
28	6,148	5,851	5,626	5,573	5,492	5,826	5,696	5,889	5,627	6,422	5,221	4,870	68,241
29	5,538	1,320	5,570	5,161	5,330	5,663	5,480	5,166	5,920	5,331	5,241	4,593	60,313
30	5,352	0	5,218	5,310	5,220	5,294	5,466	5,467	5,705	5,446	5,104	4,264	57,846
31	5,463	0	5,266	0	5,163	0	5,444	4,954	0	5,228	0	3,560	35,078
Totale	188,071	165,471	179,195	166,652	173,847	167,257	174,737	171,584	172,187	174,941	162,476	158,916	2,055,334

Si nota subito come il numero dei nati l'1 e il 2 Gennaio sia nettamente superiore al numero dei nati in tutti gli altri giorni dell'anno; si è cercato quindi di capire se questa anomalia fosse condensata su determinate età degli intervistati e si è osservato che questo effetto è concentrato soprattutto negli individui dai 35 anni in su (come si può notare in Tabella 3.3). Questo aumento del numero dei nati nei primi due giorni dell'anno non è spiegabile

con la stagionalità delle nascite, ma è probabilmente dovuto ad errori di misura.

Tabella 3.3: **Distribuzione per classi d'età dei nati l'1 e il 2 Gennaio**

Classe d'età	1 Gennaio	2 Gennaio
0-14	616	674
15-24	698	560
25-34	1,153	638
35-44	1,990	1,105
45-54	2,131	1,249
55-64	2,481	1,413
65-74	1,555	994
75 e +	1,599	1,178

Nel dataset sono presenti anche delle variabili che permettono di identificare univocamente ogni individuo presente nelle varie occasioni di intervista:

- l'anno di estrazione della famiglia campione;
- il codice della provincia di residenza;
- il codice del comune di residenza;
- il codice di quartina della famiglia;
- il codice della famiglia all'interno della quartina (famiglia base, prima sostituta, seconda sostituta o terza sostituta);
- il codice identificativo dell'individuo all'interno della famiglia.

Tramite queste variabili è stato quindi possibile controllare se per ogni individuo la data di nascita fosse identica in ciascuna intervista effettuata; il risultato è mostrato nella Tabella 3.4.

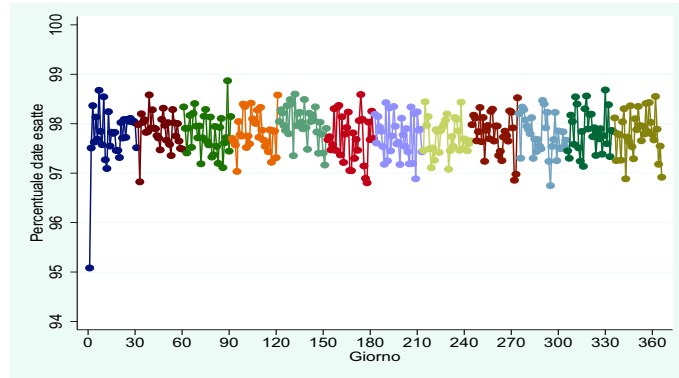
Tabella 3.4: **Frequenze delle risposte coerenti alla domanda sulla data di nascita**

Num di interviste effettuate	Num di volte che l'individuo ha risposto con la stessa data al questionario				Totale
	1	2	3	4	
1	173,220	0	0	0	173,220
2	9,994	581,990	0	0	591,984
3	522	22,464	384,189	0	407,175
4	72	10,524	21,612	835,052	867,260
Totale	183,808	614,978	405,801	835,052	2,039,639

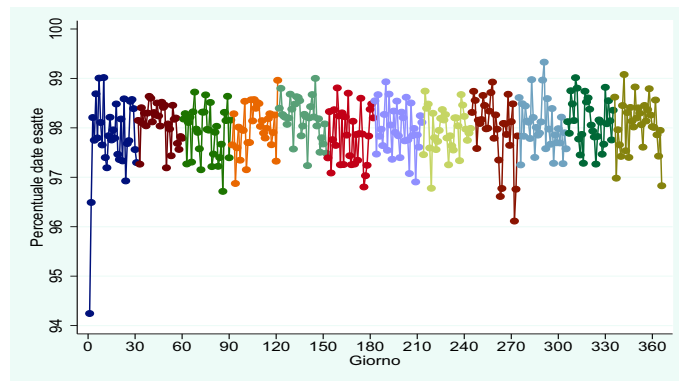
Per essere più sicuri della correttezza delle analisi svolte si è deciso di tenere solo gli individui per cui la data di nascita fosse stata ripetuta almeno una volta: si sono quindi eliminati gli individui che hanno svolto un solo questionario (in quanto non era possibile verificare la veridicità dell'informazione) e quelli che hanno risposto con date differenti nelle diverse occasioni di intervista. La numerosità del campione si è perciò ridotta da 2,039,639 a 1,801,231. In seguito si sono eliminati 3,216 individui che nei vari episodi hanno risposto in modi differenti alla domanda sul sesso del componente della famiglia, portando la numerosità totale a 1,798,015.

Si ritiene che la causa principale degli errori e delle mancate risposte in queste variabili sia il fatto che queste informazioni contenute nella scheda generale possono essere date da un solo individuo per tutti i membri della famiglia. Vi è però la possibilità di un errore nei codici di abbinamento, per cui le persone confrontate nelle diverse occasioni non sono le medesime. La scelta di eliminare qualsiasi incoerenza è quindi conservativa rispetto ai falsi positivi. Si è poi deciso di togliere dal campione i cittadini stranieri (UE e non UE) e le persone non nate in Italia, portando la numerosità finale a 1,721,620; questa scelta è stata fatta sia per migliorare la percentuale di risposte coerenti alla domanda sulla data di nascita sia per ridurre i problemi relativi al percorso scolastico per le analisi successive. Infatti analizzando il titolo di studio conseguito si è deciso di tenere solo coloro che quasi certamente lo hanno conseguito in Italia, coerentemente con l'obiettivo della tesi di valutare la legislazione italiana in materia di iscrizioni alla scuola dell'obbligo.

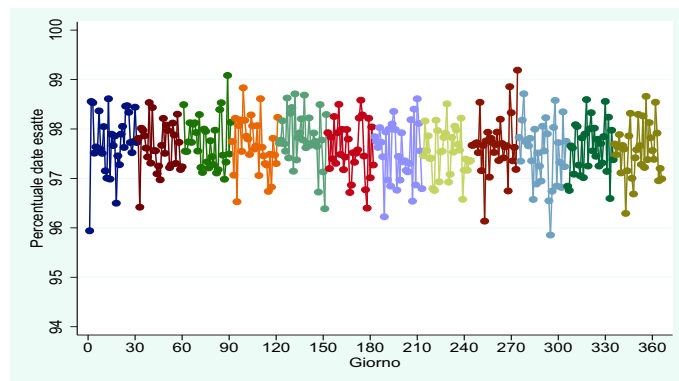
Come ulteriore controllo sulla distribuzione delle date di nascita errate, è stato fatto il grafico della percentuale di date coerenti all'interno dell'anno (ovviamente togliendo gli individui che hanno partecipato ad una sola intervista): sull'asse delle ascisse si possono vedere i 366 giorni dell'anno mentre sull'asse delle ordinate ci sono le percentuali di date coerenti sul totale di nati in ogni giorno. Si può subito notare dalla Figura 3.1 come le percentuali di risposte coerenti alla domanda sulla data di nascita siano mediamente più elevate nei maschi rispetto alle femmine, ma anche come l'andamento delle percentuali all'interno di ciascun mese sia diverso fra i vari mesi dell'anno, escludendo la possibilità di un errore sistematico nella rilevazione di questa variabile. Vi è inoltre la conferma della presenza di errori di misura sui primi



(a) Totale



(b) Maschi



(c) Femmine

Figura 3.1: Percentuale di date coerenti per giorno di nascita

due giorni dell'anno in quanto, nonostante siano stati eliminati dal campione la maggior parte dei nati in quei giorni, la percentuale di date coerenti si abbassa in modo considerevole.

3.2 Il titolo di studio

Un'altra variabile di grande interesse per l'analisi è il titolo di studio. Nella RCFL le possibili risposte alla domanda sul più alto titolo di studio conseguito al momento dell'intervista sono dodici:

1. nessun titolo;
2. licenza elementare;
3. licenza media (o avviamento professionale);
4. diploma di scuola media superiore di 2 o 3 anni che non permette l'iscrizione all'università;
5. diploma di scuola media superiore di 4 o 5 anni che permette l'iscrizione all'università;
6. Accademia belle arti, Istituto superiore industrie artistiche, Accademia di arte drammatica, perfezionamento Accademia di danza, perfezionamento Conservatorio e Istituto musica pareggiato, Scuola interpreti e traduttori, Scuola archivista, paleografia e diplomatica;
7. diploma universitario, laurea breve (vecchio ordinamento), scuola diretta a fini speciali, scuola parauniversitaria;
8. laurea di 3 anni di primo livello (nuovo ordinamento);

9. laurea specialistica di 2 anni di secondo livello (nuovo ordinamento);
10. laurea di 4 anni o più (vecchio ordinamento o nuova laurea specialistica a ciclo unico);
11. specializzazione post-laurea (compresi master di 1° e 2° livello);
12. dottorato di ricerca.

Per semplificare l'analisi si sono raggruppate queste dodici modalità in quattro gruppi:

- nessun titolo;
- licenza elementare o licenza media (unendo la seconda e la terza modalità);
- diploma (unendo la quarta e la quinta modalità);
- laurea o post laurea (unendo le modalità dalla sesta alla dodicesima).

Poiché per le analisi si è interessati al più elevato titolo di studio conseguito da ciascun individuo, si è poi cercato di tenere nel campione solo le persone che presumibilmente avessero terminato gli studi; si sono quindi eliminati gli individui fino ai 29 anni e quelli che dichiaravano di essere ancora iscritti a scuola (cioè coloro che hanno risposto affermativamente alla domanda “Nelle quattro settimane ‘dal...al...’ lei era iscritto a scuola o all’università?”). Sono stati inoltre esclusi i soggetti con più di 65 anni in quanto per questi individui la possibilità di continuare gli studi è molto probabilmente vincolata al contesto sociale dell’Italia di quegli anni. Il numero di individui rimasti è quindi 838,452.

Tra le persone rimanenti nel campione ve ne sono 19,091 che dichiarano di non essere in possesso di alcun titolo di studio al momento dell'intervista. La distribuzione dell'età di queste persone è mostrata in Tabella 3.5. Come

Tabella 3.5: **Distibuzione per età delle persone senza titolo di studio**

Età	Numero	Percent	Età	Numero	Percent
30	60	0.31	48	306	1.60
31	78	0.41	49	356	1.86
32	90	0.47	50	425	2.23
33	88	0.46	51	426	2.23
34	94	0.49	52	443	2.32
35	96	0.50	53	441	2.31
36	86	0.45	54	509	2.67
37	123	0.64	55	567	2.97
38	150	0.79	56	711	3.72
39	161	0.84	57	762	3.99
40	153	0.80	58	888	4.65
41	195	1.02	59	1,096	5.74
42	211	1.11	60	1,176	6.16
43	220	1.15	61	1,279	6.70
44	235	1.23	62	1,346	7.05
45	282	1.48	63	1,623	8.50
46	283	1.48	64	1,837	9.62
47	314	1.64	65	1,981	10.38

si può notare vi sono diverse persone che, pur essendo di giovane età, non possiedono alcun titolo di studio. Poiché risulta difficile pensare che tali informazioni siano reali, si è preferito eliminare dal campione tutti questi individui, portando così la numerosità totale a 819,361. Nel capitolo 5 si è

provato a non eliminare queste persone ma ad unirle al gruppo di coloro che hanno ottenuto al massimo la licenza elementare o la licenza media e si è verificato che i risultati sostanziali non cambiano.

Capitolo 4

L'analisi dei dati

Come è stato detto in precedenza, in Italia si possono iscrivere a scuola i bambini nati tra il primo Gennaio e il 31 Dicembre dell'anno di riferimento. Un bambino nato il 31 Dicembre, quindi, comincia la scuola un anno prima rispetto ad un bambino nato un solo giorno dopo, il primo Gennaio. Questo implica anche che questo bambino nato il primo Gennaio abbia praticamente un anno in più rispetto ad un suo compagno di classe nato l'ultimo giorno dell'anno. È facile presumere quindi che questa differenza di età relativa e di maturità porti i bambini nati in diversi periodi dell'anno ad ottenere dei differenti risultati scolastici.

In questa tesi si è quindi cercato di capire se per un bambino nato il primo Gennaio il fatto di aver cominciato la scuola un anno dopo rispetto ad un bambino nato il giorno precedente abbia influenzato la scelta sulla lunghezza della carriera scolastica. Per fare ciò si è dunque stimata la differenza nel grado scolastico ultimato (e quindi di titolo di studio conseguito) tra gli individui nati ai due estremi dell'anno (Gennaio e Dicembre).

Di seguito, dopo alcune analisi preliminari sulla distribuzione del titolo di

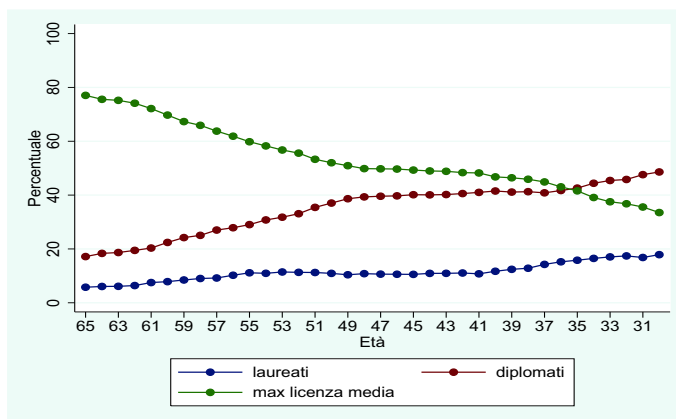
studio per età, verrà presentato il modello utilizzato per l'analisi e i risultati ottenuti.

4.1 Il titolo di studio nel tempo

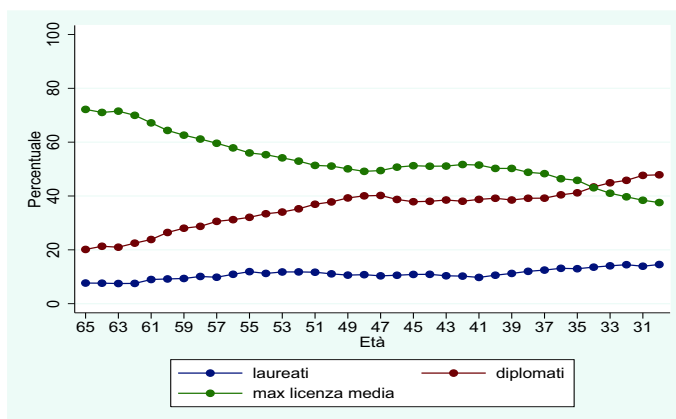
Come descritto nel paragrafo 2.3 la scuola italiana è passata attraverso varie trasformazioni nel corso del tempo, anche in anni relativamente recenti, e questo ha portato, assieme al cambiamento dei tempi e del tenore di vita, ad una diversa distribuzione del titolo di studio. La liberalizzazione degli accessi all'università, la fine dei conflitti mondiali, il boom economico degli anni Cinquanta e Sessanta, infatti, hanno portato ad una maggiore possibilità di proseguire gli studi oltre l'obbligo scolastico previsto dalla legge.

La Figura 4.1 mostra le distribuzioni del titolo di studio per età e per sesso. Come era facile aspettarsi, la percentuale di donne che hanno terminato gli studi con la licenza elementare o la licenza media è maggiore di quella dei maschi per la fascia d'età 47-65 anni. Tra i sessantacinquenni, infatti, i maschi che hanno concluso la loro carriera scolastica con al massimo la licenza media sono il 72.22% mentre le donne sono addirittura l'81.95%; le donne laureate invece sono solo il 3.91% al contrario degli uomini che raggiungono il 7.67%.

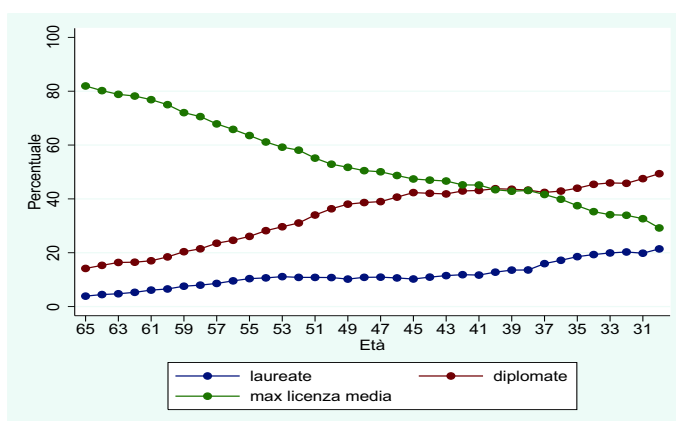
Tale relazione si inverte però nelle persone più giovani; fra le donne con meno di 50 anni, infatti, vi è un netto aumento delle diplomate e uno più graduale ma comunque significativo delle laureate. Tra i trentenni la percentuale dei laureati si ferma al 14.53% mentre quella delle donne sale al 21.45%, invece i maschi che terminano gli studi con la scuola dell'obbligo sono il 37.68% mentre le donne appena il 29.26%.



(a) Totale



(b) Maschi



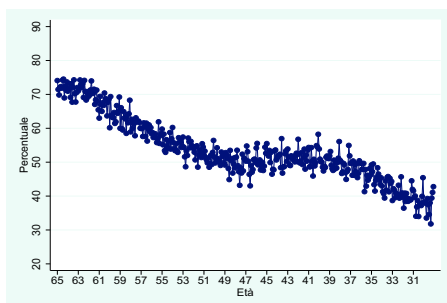
(c) Femmine

Figura 4.1: Distribuzione del titolo di studio degli individui dai 30 ai 65 anni intervistati fra il 2005 ed il 2007

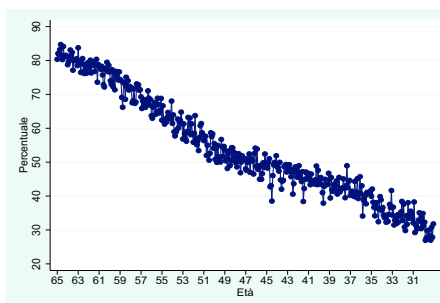
Guardando più in dettaglio queste distribuzioni (vedi Figura 4.2) si nota che l'aumento della percentuale dei laureati non è lineare ma vi è un arresto attorno ai 53 anni e una ripresa dopo i 42 anni. Lo stesso si ha con i diplomati anche se in questo caso la crescita si arresta attorno ai 47 anni per poi riprendere pressappoco sui 37 anni. Questo blocco nell'aumento del numero dei laureati e diplomati corrisponde all'incirca alla fine degli anni Settanta e agli anni Ottanta: dopo il boom di iscrizioni degli anni Settanta dovuto alla liberalizzazione degli accessi all'università avvenuta nel 1969, il numero di persone che hanno proseguito gli studi accademici è rimasto stabile in tale decennio per avere un nuovo incremento a partire dagli anni Novanta.

Da questi grafici si può notare anche una forte variabilità all'interno di ogni anno, soprattutto per quanto riguarda la distribuzione dei laureati: la percentuale di laureati nati in due mesi differenti può variare anche di cinque punti percentuali; le analisi successive serviranno a verificare se vi sia una relazione sistematica tra il mese di nascita e il titolo di studio conseguito.

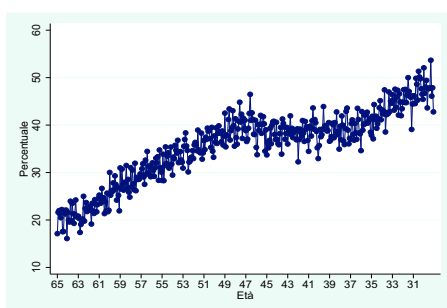
Figura 4.2: Distribuzione del titolo di studio per mese di nascita di ciascuno individuo dai 30 ai 65 anni



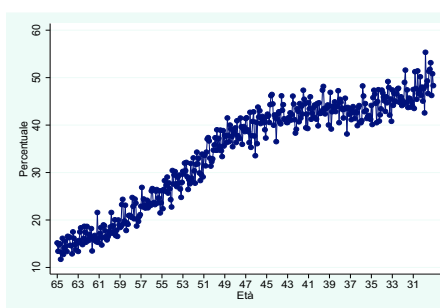
(a) Distribuzione max-medie maschi



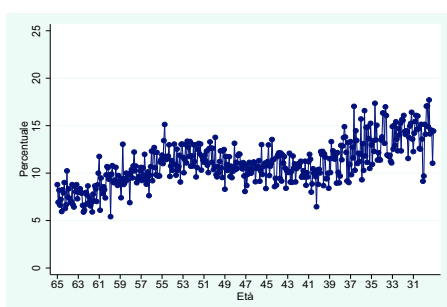
(b) Distribuzione max-medie femmine



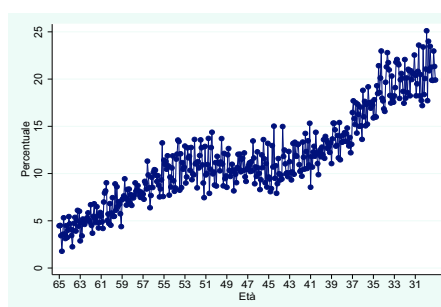
(c) Distribuzione diplomati maschi



(d) Distribuzione diplomate femmine



(e) Distribuzione laureati maschi



(f) Distribuzione laureate femmine

4.2 Il modello

Per esaminare l'effetto del mese di nascita sul titolo di studio si è stimato il seguente modello:

$$y = \alpha + \beta M + \gamma X + \delta Et\grave{a} + \epsilon \quad (4.1)$$

dove M rappresenta un insieme di undici variabili dummy per i mesi di nascita (la dummy per Gennaio è omessa) ed $Et\grave{a}$ un trend polinomiale per l'età. Per essere certi di identificare correttamente l'effetto del mese di nascita sul titolo di studio conseguito eliminando effetti spuri, sono state aggiunte al modello delle variabili di controllo (X) per la zona d'Italia di residenza (Nord, Centro o Sud con la dummy per il Nord omessa) e per il sesso.

La variabile di outcome y rappresenta una dummy che vale:

$$y_i = \begin{cases} 1 & \text{se l'individuo } i \text{ ha conseguito un certo titolo di studio;} \\ 0 & \text{altrimenti.} \end{cases} \quad (4.2)$$

Il parametro di interesse è dunque il coefficiente della dummy per il mese di Dicembre ($\beta_{Dicembre}$), che rappresenta la differenza percentuale di persone che hanno concluso la carriera scolastica con un certo titolo di studio tra Dicembre e Gennaio.

Il modello 4.1 è un modello di regressione lineare multipla, la variabile casuale y , cioè, viene spiegata utilizzando una combinazione lineare di k variabili esplicative e parametri. In formule:

$$E[y] = \mu = \sum_{k=1}^K \theta_k x_k. \quad (4.3)$$

Questa relazione lineare tra la variabile risposta e le variabili indipendenti è stata stimata con i Minimi Quadrati Ordinari (OLS dall'inglese *Ordinary*

Least Square). Tale tecnica permette di stimare i valori dei parametri in modo tale da minimizzare la somma dei quadrati dei residui, definiti come la differenza tra valori effettivi della variabile dipendente e corrispettivi valori teorici. In simboli:

$$\begin{aligned} \min_{\theta_1, \dots, \theta_k} \sum_{k=1}^K \epsilon_k^2 &= \min_{\theta_1, \dots, \theta_k} \sum_{k=1}^K (y_k - \hat{y}_k)^2 = \\ &= \min_{\theta_1, \dots, \theta_k} \sum_{k=1}^K (y_k - \theta_0 - \theta_1 x_1 - \dots - \theta_k x_k)^2. \end{aligned} \quad (4.4)$$

In realtà la tecnica appena descritta non è quella più adatta a stimare il modello, in quanto la variabile risposta è una dummy. Ad esempio il modello 4.1 potrebbe essere scritto nella forma:

$$\log \left[\frac{Pr(y = 1)}{1 - Pr(y = 1)} \right] = \alpha + \beta M + \gamma X + \delta Et\grave{a} + \epsilon \quad (4.5)$$

e stimato con una regressione logit. In questa tecnica la relazione tra la variabile risposta e le variabili esplicative non è più lineare ma logistica grazie alla seguente funzione link:

$$\eta = \log \left(\frac{\mu}{1 - \mu} \right). \quad (4.6)$$

La forma rappresentata dall'equazione 4.5 può essere espressa anche in termini di probabilità per meglio evidenziare come questa tecnica stimi la probabilità che un certo evento si realizzi (per maggiori informazioni si veda Liao, 1994):

$$P(y = 1) = \frac{e^{\sum_{k=1}^K \theta_k x_k}}{1 + e^{\sum_{k=1}^K \theta_k x_k}}. \quad (4.7)$$

Si è provato a stimare tutti i modelli con entrambe le tecniche appena descritte e si è osservato che i risultati sostanziali ottenuti erano pressoché identici, quindi si è preferito riportare solo le stime ottenute con i Minimi Quadrati

Ordinari per una più intuitiva interpretazione dei risultati. Nel paragrafo 4.3, in cui verranno illustrati i risultati ottenuti, verrà mostrato con un esempio l'equivalenza dei due metodi di stima.

4.3 L'analisi

In questo paragrafo verranno presentati i principali risultati ottenuti stimando il modello descritto nel paragrafo precedente. Per le analisi si è deciso di non utilizzare tutto il campione ma di concentrarsi solo sulla fascia centrale. Questa scelta è stata effettuata per diversi motivi: primo, eliminare gli individui più anziani, in modo tale da avere un sistema scolastico identico per tutto il campione (in quanto la scuola media unificata non era ancora presente per queste persone); secondo, eliminare gli individui troppo giovani in quanto negli anni più recenti è aumentata la frequenza di iscrizioni anticipate per i nati fino ad Aprile; terzo, analizzare un periodo in cui il trend sia approssimativamente costante. A questo punto la scelta finale è caduta nella fascia 42-51 anni, per rimanere entro il periodo di stabilizzazione della distribuzione del titolo di studio descritto a pagina 28 e non includere un così netto cambiamento di pendenza. La numerosità del campione è quindi di 244,098 persone (di cui 119,341 maschi e 124,757 femmine).

4.3.1 La scuola dell'obbligo

Come prima cosa si è voluto vedere se vi fosse una relazione tra il mese di nascita e il proseguimento degli studi oltre l'obbligo scolastico. La variabile

risposta è quindi data da una dummy che vale:

$$y_i = \begin{cases} 1 & \text{se l'individuo } i \text{ ha conseguito al massimo la licenza media;} \\ 0 & \text{se ha conseguito un titolo di studio più elevato.} \end{cases}$$

I risultati ottenuti stimando il modello 4.1 tramite la regressione logistica e la regressione lineare classica sono mostrati nella Tabella 4.1.

La Tabella 4.1(a) riporta le stime dei parametri e degli *odds ratio* della re-

(a) Logistica				(b) Lineare		
Variabile	Coefficiente	Odds ratio	(Std. Err.)	Variabile	Coefficiente	(Std. Err.)
Intercetta	-0.12947**		(0.01664)	Intercetta	0.46773**	(0.00412)
Febbraio	0.00743	1.00746	(0.02015)	Febbraio	0.00184	(0.00496)
Marzo	-0.02180	0.97844	(0.01915)	Marzo	-0.00540	(0.00485)
Aprile	-0.03783†	0.96288	(0.01934)	Aprile	-0.00938†	(0.00498)
Maggio	-0.07361**	0.92903	(0.01844)	Maggio	-0.01825**	(0.00492)
Giugno	-0.08899**	0.91486	(0.01841)	Giugno	-0.02206**	(0.00499)
Luglio	-0.14663**	0.86361	(0.01716)	Luglio	-0.03633**	(0.00492)
Agosto	-0.10058**	0.90431	(0.01807)	Agosto	-0.02494**	(0.00495)
Settembre	-0.03123	0.96925	(0.01928)	Settembre	-0.00774	(0.00493)
Ottobre	-0.03032	0.97013	(0.01934)	Ottobre	-0.00751	(0.00494)
Novembre	0.01136	1.01142	(0.02042)	Novembre	0.00282	(0.00500)
Dicembre	0.00359	1.00360	(0.02038)	Dicembre	0.00090	(0.00503)
Centro	-0.06327**	0.93869	(0.01143)	Centro	-0.01570**	(0.00302)
Sud	0.33087**	1.39217	(0.01231)	Sud	0.08236**	(0.00219)
Femmina	-0.05672**	0.94485	(0.00769)	Femmina	-0.01406**	(0.00202)
Eta	0.01894**	1.01912	(0.00144)	Eta	0.00469**	(0.00035)

Livelli di significatività : † : 10% * : 5% ** : 1%

Tabella 4.1: **Regressione logistica e regressione lineare sulla probabilità di fermarsi alla scuola dell'obbligo**

gressione logistica, ad esempio per i mesi il rapporto tra l'odds della baseline (in questo caso Gennaio) e quello di ciascuna variabile. Un odds ratio maggiore di 1 indica che la probabilità che un certo evento accada è maggiore della probabilità che l'evento non accada, viceversa un odds ratio minore di

1 indica che è più probabile che tale evento non si verifichi; in questo caso, per esempio, l'odds ratio della della variabile Febbraio che è maggiore di uno indica che un individuo nato in Febbraio è più probabile che termini gli studi con la scuola dell'obbligo e che quindi non prosegua gli studi rispetto ad un individuo nato in Gennaio.

Come si può notare le stime dei coefficienti e degli standard error ottenute con la regressione lineare classica sono coerenti con le stime ottenute con la regressione logistica e i livelli di significatività delle stime sono uguali in entrambi i casi. Data la corrispondenza delle due tecniche di stima, verranno nel seguito riportati solo i risultati ottenuti tramite la regressione lineare per rendere più intuitiva l'interpretazione delle stime.

La baseline di questo modello è data da un uomo del Nord di 42 anni nato in Gennaio e le stime di ciascun coefficiente rappresentano l'effetto marginale della corrispondente variabile su y , *coeteris paribus*; ad esempio la stima del coefficiente β_{Marzo} rappresenta la differenza tra la percentuale di uomini di 42 anni del Nord nati in Marzo che terminano gli studi con al massimo la licenza media e la percentuale di uomini di 42 anni del Nord nati in Gennaio che concludono la carriera scolastica con la scuola dell'obbligo.

Il coefficiente della variabile Dicembre non è significativo e quindi la percentuale di persone nate in Dicembre che non prosegue gli studi oltre l'obbligo scolastico è statisticamente uguale a quella dei nati in Gennaio. Tra i coefficienti delle dummy per i mesi di nascita, gli unici significativi sono quelli relativi ai mesi di Maggio, Giugno, Luglio e Agosto (anche Marzo se si sceglie il 10% come massima probabilità di commettere un errore del I tipo) e con segno negativo. Tutto ciò induce a pensare che le persone nate durante i mesi

estivi siano più portate a proseguire gli studi rispetto a quelle nate durante i mesi invernali.

Per vedere meglio l'andamento della relazione tra il mese di nascita ed il numero di persone che terminano gli studi è stato fatto un grafico delle stime dei parametri della regressione lineare (Figura 4.3). Come si può facilmente

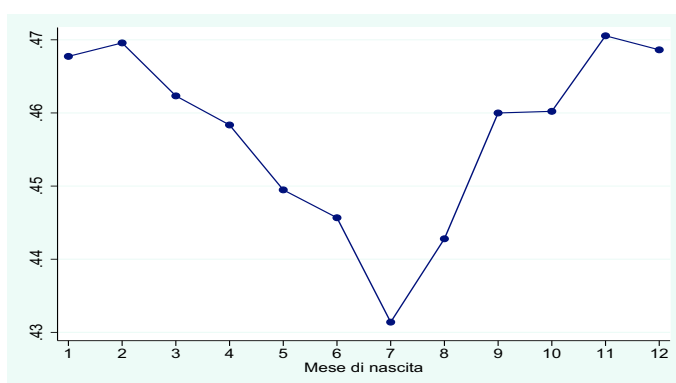


Figura 4.3: **Rappresentazione grafica delle stime dei coefficienti della regressione lineare sulla probabilità di fermarsi alla scuola dell'obbligo**

notare, il minimo è in Luglio e quindi la probabilità di terminare gli studi con la scuola dell'obbligo decresce fino a tale mese per poi aumentare e riportarsi ai livelli di Gennaio. Il divario tra la percentuale di individui nati in Luglio che terminano gli studi e la percentuale di persone che conseguono al massimo la licenza media nati in Gennaio o Dicembre è di ben 4 punti percentuali: tale differenza, oltre ad essere statisticamente significativa, è rilevante anche in termini assoluti.

In relazione alle variabili di controllo, è da notare che mentre nel Centro Italia è più probabile che un individuo prosegua gli studi rispetto ad una persona del Nord, la relazione si inverte per il Sud Italia; per quanto riguarda il sesso

degli individui, il coefficiente per la dummy “femmina” porta a dire che è più probabile che una donna prosegua gli studi rispetto ad un uomo, a parità di tutte le altre condizioni. Per il trend per l'età si è utilizzato un polinomio di primo grado in quanto si è ritenuto che la relazione tra l'età e il titolo di studio fosse lineare per questa fascia d'età (come si è visto nelle Figure 4.1(a) e 4.2). Si è comunque provato a stimare il modello inserendo altri gradi per il polinomio ma si è notato che o non erano significativi o non miglioravano significativamente il modello (queste analisi verranno illustrate nel capitolo 5 sulla robustezza delle analisi), quindi si è preferito continuare con un trend lineare. Il coefficiente del trend per l'età degli intervistati è positivo quindi, come è già stato illustrato nel paragrafo 4.1, con il passare degli anni (ovvero con il diminuire dell'età) la percentuale di persone che non terminano gli studi con al massimo la licenza media cresce linearmente.

Si è poi stimato lo stesso modello per il sottocampione dei maschi e per quello delle femmine. I risultati delle stime sono mostrati nella Tabella 4.2. Anche nei due sottocampioni l'andamento è lo stesso, infatti, vi è una forte stagionalità con picco negativo in estate e non c'è una differenza significativa tra i mesi di Gennaio e Dicembre. A differenza delle femmine e del campione totale però, nei maschi questa tendenza è più netta (infatti sono significativi i coefficienti dei mesi da Maggio a Ottobre mentre per le donne solo i mesi da Giugno ad Agosto) e il punto di minimo viene toccato nel mese di Giugno e non di Luglio (sebbene la differenza tra Giugno e Luglio non sia statisticamente significativa).

Anche in questo caso sono stati fatti i grafici delle stime dei parametri delle due regressioni (Figura 4.4).

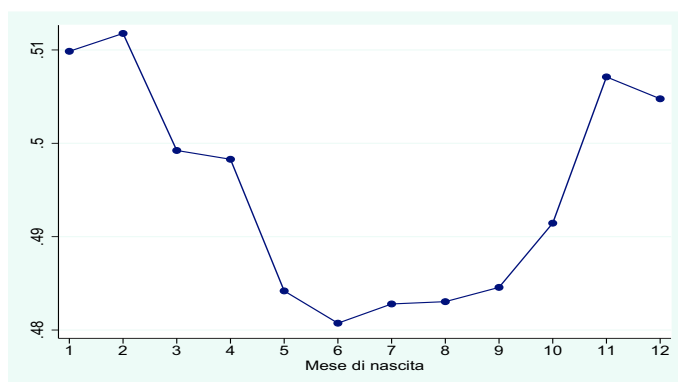
(a) Maschi			(b) Femmine		
Variabile	Coefficiente	(Std. Err.)	Variabile	Coefficiente	(Std. Err.)
Intercetta	0.50985**	(0.00572)	Intercetta	0.41270**	(0.00558)
Febbraio	0.00191	(0.00710)	Febbraio	0.00192	(0.00691)
Marzo	-0.01062	(0.00694)	Marzo	-0.00109	(0.00677)
Aprile	-0.01156	(0.00716)	Aprile	-0.00773	(0.00691)
Maggio	-0.02566**	(0.00705)	Maggio	-0.01125	(0.00686)
Giugno	-0.02912**	(0.00715)	Giugno	-0.01566*	(0.00695)
Luglio	-0.02706**	(0.00705)	Luglio	-0.04550**	(0.00686)
Agosto	-0.02682**	(0.00710)	Agosto	-0.02314**	(0.00690)
Settembre	-0.02529**	(0.00703)	Settembre	0.00999	(0.00691)
Ottobre	-0.01841**	(0.00708)	Ottobre	0.00307	(0.00689)
Novembre	-0.00274	(0.00716)	Novembre	0.00849	(0.00698)
Dicembre	-0.00509	(0.00722)	Dicembre	0.00549	(0.00701)
Centro	-0.03115**	(0.00433)	Centro	-0.00054	(0.00420)
Sud	0.05599**	(0.00314)	Sud	0.10799**	(0.00305)
Eta	-0.00103*	(0.00050)	Eta	0.01024**	(0.00049)

Livelli di significatività : † : 10% * : 5% ** : 1% Livelli di significatività : † : 10% * : 5% ** : 1%

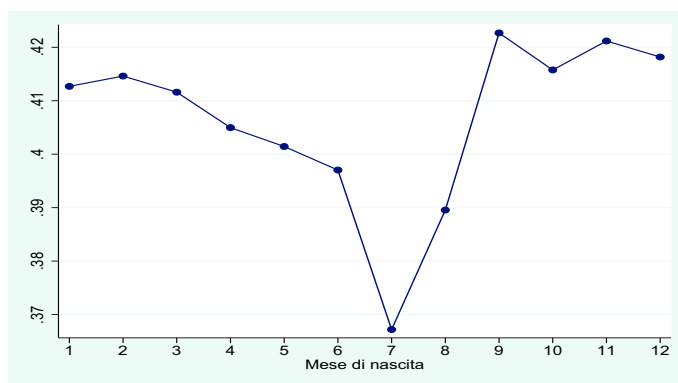
Tabella 4.2: **Confronto maschi e femmine nella regressione sulla sulla probabilità di fermarsi alla scuola dell'obbligo**

Per quanto riguarda le variabili di controllo, mentre per i maschi la stima del parametro della variabile Centro rimane significativa, per le donne non è più così, segno che non vi è differenza tra il Nord e il Centro Italia.

In conclusione di questa prima parte di analisi, si può quindi dire che il cosiddetto *relative age effect* che ci si aspettava di trovare in realtà non è presente, in quanto la percentuale di persone nate nei mesi di Gennaio e Dicembre che termina gli studi con la licenza media non differisce; si è però osservato un effetto stagionale: le persone nate durante i mesi estivi proseguono gli studi con una probabilità più elevata rispetto alle persone nate durante i mesi invernali.



(a) Maschi



(b) Femmine

Figura 4.4: **Grafico dei parametri delle regressioni per i maschi e per le femmine sulla probabilità di fermarsi alla scuola dell'obbligo**

4.3.2 L'università

In questo paragrafo si analizzerà lo stesso problema precedente visto però da un altro punto di vista: si cercherà cioè di capire se vi è una relazione tra il mese di nascita e il conseguimento della laurea o di un titolo superiore. La variabile risposta diventa quindi una dummy che vale:

$$y_i = \begin{cases} 1 & \text{se l'individuo } i \text{ ha conseguito la laurea o un titolo post-laurea;} \\ 0 & \text{altrimenti.} \end{cases}$$

I risultati ottenuti stimando il modello 4.1 con questa nuova variabile indipendente sono mostrati nella Tabella 4.3.

Tabella 4.3: **Regressione lineare sulla frequenza di laureati**

Variabile	Coefficiente	(Std. Err.)
Intercetta	0.09882**	(0.00257)
Febbraio	0.00571 [†]	(0.00309)
Marzo	0.00349	(0.00303)
Aprile	0.00364	(0.00310)
Maggio	0.00828**	(0.00307)
Giugno	0.00237	(0.00311)
Luglio	0.02038**	(0.00307)
Agosto	0.00918**	(0.00309)
Settembre	0.00264	(0.00308)
Ottobre	0.00188	(0.00308)
Novembre	-0.00101	(0.00312)
Dicembre	0.00569 [†]	(0.00314)
Centro	0.01556**	(0.00188)
Sud	0.00300*	(0.00137)
Femmina	0.00150	(0.00126)
Eta	-0.00003	(0.00022)

Significance levels : † : 10% * : 5% ** : 1%

Come si può notare, la probabilità di continuare gli studi fino al raggiungimento di almeno un diploma di laurea cresce fino al mese di Luglio per poi decrescere. Questo è consistente con quanto detto nel paragrafo precedente: gli individui nati durante i mesi estivi oltre a proseguire con più probabilità gli studi, riescono anche più probabilmente a laurearsi.

La Figura 4.5 mostra graficamente le stime ottenute.

Il coefficiente per il sesso dell'individuo mostra che in questo caso non vi è differenza nella percentuale di laureati tra i due sessi: pur essendo positivo il segno, la sua stima non è statisticamente significativa. Riguardo alla zona di residenza in questa occasione entrambe le stime hanno segno positivo e ciò porta a dire che nel Centro e Sud Italia è più probabile che un individuo raggiunga almeno la laurea rispetto a un uomo del Nord.

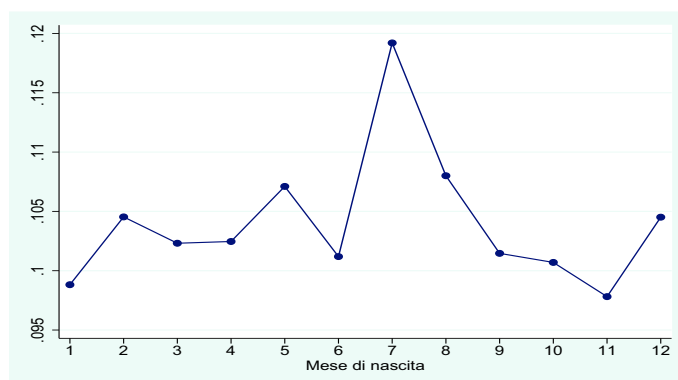


Figura 4.5: **Rappresentazione grafica delle stime dei coefficienti della regressione sulla frequenza di laureati**

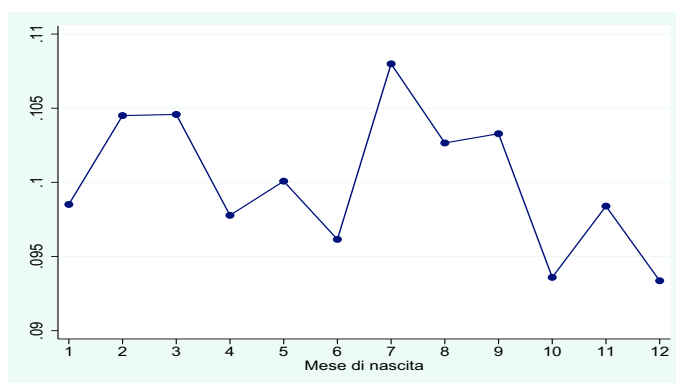
Anche in questo contesto si è ristimato il modello dividendo il campione per sesso. Le stime ottenute sono presentate in Tabella 4.4.

(a) Maschi			(b) Femmine		
Variabile	Coefficiente	(Std. Err.)	Variabile	Coefficiente	(Std. Err.)
Intercetta	0.09851**	(0.00355)	Intercetta	0.10072**	(0.00350)
Febbraio	0.00599	(0.00441)	Febbraio	0.00546	(0.00434)
Marzo	0.00607	(0.00431)	Marzo	0.00107	(0.00425)
Aprile	-0.00074	(0.00445)	Aprile	0.00789†	(0.00434)
Maggio	0.00157	(0.00438)	Maggio	0.01459**	(0.00430)
Giugno	-0.00236	(0.00444)	Giugno	0.00690	(0.00436)
Luglio	0.00948*	(0.00437)	Luglio	0.03081**	(0.00431)
Agosto	0.00414	(0.00440)	Agosto	0.01397**	(0.00433)
Settembre	0.00477	(0.00436)	Settembre	0.00033	(0.00433)
Ottobre	-0.00492	(0.00440)	Ottobre	0.00824†	(0.00432)
Novembre	-0.00012	(0.00445)	Novembre	-0.00196	(0.00438)
Dicembre	-0.00515	(0.00448)	Dicembre	0.01612**	(0.00440)
Centro	0.01182**	(0.00269)	Centro	0.01899**	(0.00264)
Sud	0.00316	(0.00195)	Sud	0.00268	(0.00192)
Eta	0.00097**	(0.00031)	Eta	-0.00098**	(0.00031)

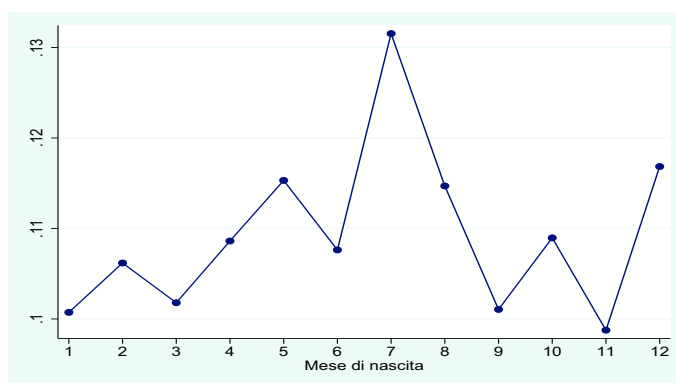
Significance levels : † : 10% * : 5% ** : 1%

Tabella 4.4: **Confronto maschi e femmine nella regressione sulla frequenza di laureati**

In questo caso le differenze tra maschi e femmine sono più evidenti rispetto a prima, ora, infatti, l'effetto stagionale tra il mese di nascita e il possesso della laurea è molto più marcato per le donne. Per le donne, però, diventa significativo il coefficiente della variabile per il mese di Dicembre, anche se l'andamento generale all'interno dell'anno non fa supporre che sia dovuto ad un reale effetto dell'età relativa tra i nati in Gennaio e i nati in Dicembre. I grafici delle stime dei parametri delle due regressioni sono mostrati in Figura 4.6.



(a) Maschi



(b) Femmine

Figura 4.6: **Grafico dei parametri delle regressioni per i maschi e per le femmine sulla frequenza di laureati**

Il parametro del trend lineare ha segno opposto nei due casi, questo perché per le donne con l'aumento dell'età vi è meno probabilità di arrivare a conseguire la laurea, mentre per i maschi la relazione tra età e laurea si inverte: aumentando l'età cresce la probabilità di laurearsi (come mostrato nel paragrafo 4.1); questo spiega perché nella regressione su tutto il campione la stima di questo coefficiente non sia significativa (si veda la Tabella 4.3). Per quanto riguarda le variabili relative alla zona di residenza, è significativa solo la dummy per il Centro Italia a riprova che vi sono più residenti in tale zona che hanno raggiunto la laurea.

4.4 Interpretazione dei risultati ottenuti

In questa analisi si è partiti dal presupposto che siano i bambini più bravi a proseguire gli studi e, data la precedente letteratura sull'argomento (si veda per esempio Shepard e Smith (1986)) che indicava che i bambini che ottengono risultati migliori a scuola sono quelli più grandi all'interno del loro anno scolastico (quindi in Italia quelli nati in Gennaio), ci si aspettava di trovare un maggior numero di laureati nati durante i primi mesi dell'anno. I risultati ottenuti mostrano invece che non vi è alcuna differenza tra i mesi di Gennaio e Dicembre, facendo emergere però una forte stagionalità.

In questo paragrafo verranno analizzate le possibili motivazioni delle evidenze presentate nel paragrafo precedente. In primo luogo, le ricerche precedenti si concentravano su studenti più piccoli e, dato che è ampiamente dimostrato che questo effetto età svanisce nel tempo, è possibile che con il termine della scuola dell'obbligo questo fenomeno si sia esaurito eliminando la differenza

tra i bambini nati ai due estremi dell'anno.

In secondo luogo, Buckles e Hungerman (2008) hanno dimostrato che i bambini nati in inverno sono sproporzionalmente nati da madri che sono teenager, non sposate e con un basso titolo di studio. Questo porta a pensare che questi bambini abbiano meno possibilità di continuare gli studi e quindi la maggior parte di coloro che proseguono gli studi siano nati durante i mesi estivi, in accordo con i risultati ottenuti in questa tesi.

Infine, diverse ricerche mostrano una correlazione tra la temperatura durante la gestazione e l'intelligenza. Orme (1963), per esempio, ha dimostrato che il quoziente intellettivo dei pazienti con ritardo mentale nati in estate era significativamente più alto di quello dei pazienti nati in inverno a causa del progressivo aumento della temperatura durante tutta la gravidanza. In tale articolo ha inoltre postulato che tale relazione possa essere simile anche negli individui che si trovano nel normale range di intelligenza. Dato che si è supposto che siano gli individui più bravi (e quindi più intelligenti) a proseguire gli studi, questa scoperta può spiegare come mai la maggioranza dei laureati sia nata durante i mesi estivi.

Capitolo 5

Controllo della robustezza delle analisi

In questo capitolo verranno riportate alcune verifiche (utilizzando un nuovo campione o una nuova specificazione del modello) fatte a riprova della robustezza della analisi svolte nel capitolo 4. Per brevità qui di seguito si mostreranno solo i risultati ottenuti per il campione totale (senza quindi riportare le analisi divise per maschi e femmine), in quanto le principali evidenze erano pressoché identiche.

5.1 Modifiche del campione

Nella fase di formazione del campione sono state effettuate delle scelte dettate dalla volontà di arrivare ad ottenere il campione migliore per l'obiettivo della ricerca. In questa sezione si verificherà se queste scelte hanno influenzato i risultati ottenuti.

In primo luogo si è controllato se il fatto di aver tenuto ogni occasione di col-

loquio con ciascun individuo (e quindi di avere alcuni individui presenti più di una volta all'interno del campione come spiegato nella sezione 3.1) abbia creato dei problemi di selezione del campione. I risultati ottenuti dalle due regressioni (quella con variabile di risposta la dummy per il raggiungimento di al massimo la licenza media e quella con variabile risposta la dummy per la laurea) tenendo solo la prima occasione di intervista sono mostrati nella Tabella 5.1.

(a) Massimo licenza media			(b) Laurea o post-laurea		
Variabile	Coefficiente	(Std. Err.)	Variabile	Coefficiente	(Std. Err.)
Intercetta	0.46872**	(0.00666)	Intercetta	0.10061**	(0.00417)
Febbraio	-0.00038	(0.00811)	Febbraio	0.00292	(0.00508)
Marzo	-0.00493	(0.00795)	Marzo	0.00241	(0.00498)
Aprile	-0.01248	(0.00815)	Aprile	0.00208	(0.00510)
Maggio	-0.01834*	(0.00807)	Maggio	0.00643	(0.00505)
Giugno	-0.02498**	(0.00817)	Giugno	0.00234	(0.00512)
Luglio	-0.03454**	(0.00807)	Luglio	0.01731**	(0.00505)
Agosto	-0.02454**	(0.00811)	Agosto	0.00943 [†]	(0.00508)
Settembre	-0.00866	(0.00808)	Settembre	0.00179	(0.00506)
Ottobre	-0.00796	(0.00808)	Ottobre	-0.00040	(0.00506)
Novembre	0.00038	(0.00820)	Novembre	-0.00053	(0.00513)
Dicembre	0.00364	(0.00825)	Dicembre	0.00343	(0.00517)
Centro	-0.01644**	(0.00493)	Centro	0.01611**	(0.00308)
Sud	0.08531**	(0.00359)	Sud	0.00215	(0.00225)
Femmina	-0.01391**	(0.00330)	Femmina	0.00207	(0.00207)
Eta	0.00523**	(0.00055)	Eta	0.00003	(0.00035)

Tabella 5.1: **Regressioni inserendo solo la prima intervista effettuata da ciascun individuo.**

Come si può vedere le stime sono molto simili a quelle ottenute inserendo tutte le interviste effettuate da ciascun individuo; ovviamente cambiano gli standard error essendo diminuita considerevolmente la numerosità del campione (da 244,098 a 90,770 persone). I relativi grafici delle stime dei coefficienti delle dummy per i mesi di nascita sono mostrati nella Figura 5.1.

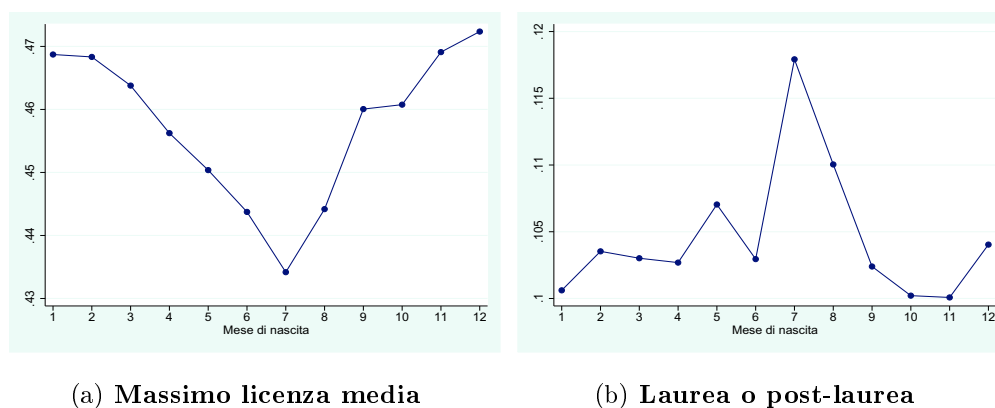


Figura 5.1: **Grafici dei parametri delle regressioni inserendo solo la prima intervista effettuata**

Un'altra decisione presa è stata quella di escludere coloro che dichiaravano di non essere in possesso di alcun titolo di studio in quanto si è ritenuto che tali informazioni non fossero totalmente veritiere. Per controllare che questa scelta non abbia distorto le stime si è provato a ristimare entrambi i modelli inserendo le 3,058 persone che non hanno ottenuto neanche la licenza elementare. Le stime ottenute sono illustrate nella Tabella 5.2 e nella Figura 5.2.

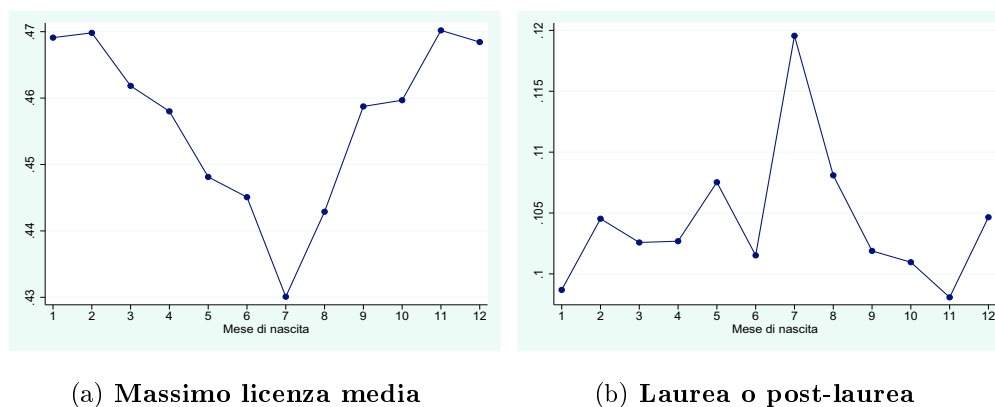


Figura 5.2: **Grafici dei parametri delle regressioni inserendo gli individui senza titolo di studio**

(a) Massimo licenza media			(b) Laurea o post-laurea		
Variabile	Coefficiente	(Std. Err.)	Variabile	Coefficiente	(Std. Err.)
Intercetta	0.46910**	(0.00409)	Intercetta	0.09869**	(0.00254)
Febbraio	0.00072	(0.00491)	Febbraio	0.00585 [†]	(0.00305)
Marzo	-0.00726	(0.00481)	Marzo	0.00389	(0.00299)
Aprile	-0.01110*	(0.00494)	Aprile	0.00400	(0.00306)
Maggio	-0.02097**	(0.00488)	Maggio	0.00884**	(0.00303)
Giugno	-0.02402**	(0.00495)	Giugno	0.00284	(0.00307)
Luglio	-0.03901**	(0.00488)	Luglio	0.02087**	(0.00303)
Agosto	-0.02621**	(0.00491)	Agosto	0.00941**	(0.00305)
Settembre	-0.01036*	(0.00489)	Settembre	0.00320	(0.00304)
Ottobre	-0.00942 [†]	(0.00490)	Ottobre	0.00227	(0.00304)
Novembre	0.00109	(0.00496)	Novembre	-0.00061	(0.00308)
Dicembre	-0.00067	(0.00499)	Dicembre	0.00597 [†]	(0.00310)
Centro	-0.01550**	(0.00301)	Centro	0.01547**	(0.00187)
Sud	0.09112**	(0.00217)	Sud	0.00078	(0.00135)
Femmina	-0.01300**	(0.00200)	Femmina	0.00127	(0.00124)
Eta	0.00516**	(0.00035)	Eta	-0.00015	(0.00022)

Tabella 5.2: **Regressioni inserendo gli individui senza titolo di studio**

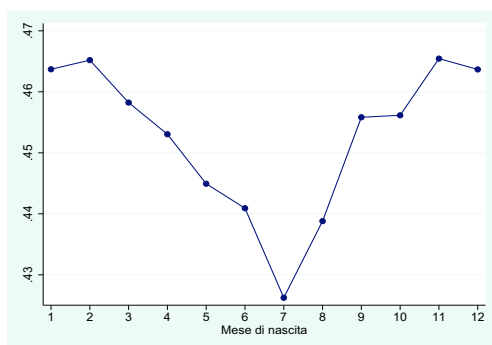
Anche in questa occasione i risultati non differiscono da quelli ottenuti nel capitolo precedente e quindi è possibile dire che la decisione di eliminare tali individui non ha influito sui risultati ottenuti.

Infine si è controllato che anche la decisione di escludere dal campione finale gli individui ancora iscritti a scuola non abbia alterato i risultati. In Tabella 5.3 sono riportate le stime dei parametri delle regressioni inserendo anche i 2,082 individui ancora iscritti a scuola.

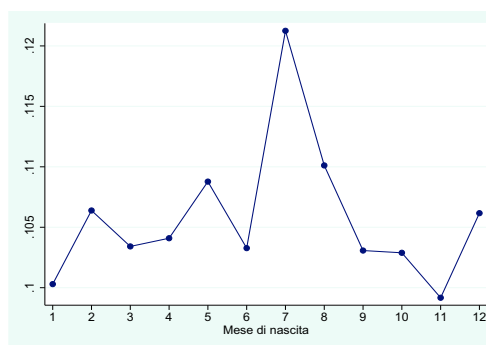
Come nei casi precedenti, i risultati ottenuti sono coerenti con quelli ottenuti escludendo queste persone. In conclusione, quindi, tutte le scelte effettuate in materia di formazione del dataset non hanno influito significativamente sui risultati ottenuti.

(a) Massimo licenza media			(b) Laurea o post-laurea		
Variabile	Coefficiente	(Std. Err.)	Variabile	Coefficiente	(Std. Err.)
Intercetta	0.46369**	(0.00411)	Intercetta	0.10030**	(0.00258)
Febbraio	0.00150	(0.00494)	Febbraio	0.00609*	(0.00310)
Marzo	-0.00546	(0.00483)	Marzo	0.00312	(0.00303)
Aprile	-0.01065*	(0.00496)	Aprile	0.00380	(0.00311)
Maggio	-0.01876**	(0.00490)	Maggio	0.00848**	(0.00308)
Giugno	-0.02278**	(0.00497)	Giugno	0.00298	(0.00312)
Luglio	-0.03744**	(0.00490)	Luglio	0.02095**	(0.00307)
Agosto	-0.02489**	(0.00493)	Agosto	0.00982**	(0.00310)
Settembre	-0.00786	(0.00491)	Settembre	0.00278	(0.00308)
Ottobre	-0.00754	(0.00492)	Ottobre	0.00259	(0.00309)
Novembre	0.00175	(0.00498)	Novembre	-0.00113	(0.00313)
Dicembre	-0.00003	(0.00501)	Dicembre	0.00587 [†]	(0.00315)
Centro	-0.01679**	(0.00300)	Centro	0.01603**	(0.00188)
Sud	0.08126**	(0.00218)	Sud	0.00342*	(0.00137)
Femmina	-0.01411**	(0.00201)	Femmina	0.00200	(0.00126)
Eta	0.00506**	(0.00035)	Eta	-0.00019	(0.00022)

Tabella 5.3: Regressioni inserendo gli individui ancora iscritti a scuola



(a) Massimo licenza media



(b) Laurea o post-laurea

Figura 5.3: Grafici dei parametri delle regressioni inserendo gli individui ancora iscritti a scuola

5.2 Modifiche del modello

Un altro momento cruciale è stato quello di scelta del modello da utilizzare (e quindi di decisione di quali variabili includere e quali escludere dal modello) in modo tale da identificare chiaramente l'effetto che ci si era proposti di verificare e stimare. In questa sezione si mostrerà come la specificazione finale del modello non abbia condizionato i risultati raggiunti.

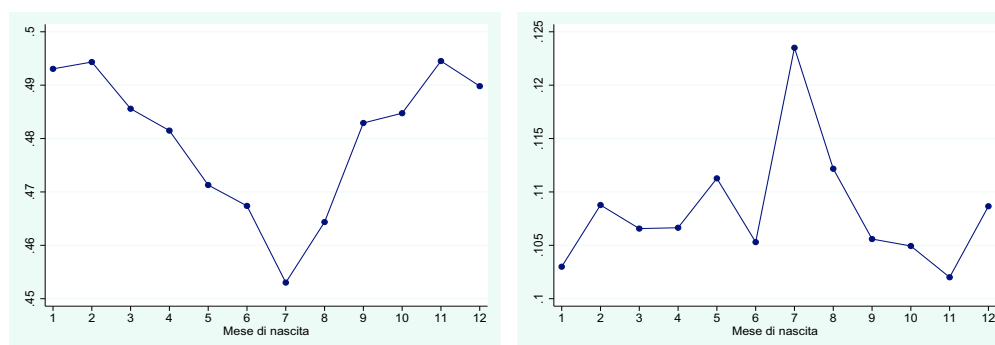
Come prima cosa si è provato a eliminare dal modello il set di variabili di controllo in modo da vedere se tali variabili influissero significativamente sui risultati ottenuti. La Tabella 5.4 mostra le stime dei due modelli senza variabili di controllo.

(a) Massimo licenza media			(b) Laurea o post-laurea		
Variabile	Coefficiente	(Std. Err.)	Variabile	Coefficiente	(Std. Err.)
Intercetta	0.49306**	(0.00385)	Intercetta	0.10300**	(0.00239)
Febbraio	0.00128	(0.00497)	Febbraio	0.00578 [†]	(0.00309)
Marzo	-0.00747	(0.00487)	Marzo	0.00357	(0.00303)
Aprile	-0.01157*	(0.00500)	Aprile	0.00365	(0.00311)
Maggio	-0.02176**	(0.00494)	Maggio	0.00828**	(0.00307)
Giugno	-0.02567**	(0.00500)	Giugno	0.00230	(0.00311)
Luglio	-0.04003**	(0.00494)	Luglio	0.02051**	(0.00307)
Agosto	-0.02870**	(0.00497)	Agosto	0.00918**	(0.00309)
Settembre	-0.01014*	(0.00495)	Settembre	0.00259	(0.00308)
Ottobre	-0.00831 [†]	(0.00496)	Ottobre	0.00194	(0.00308)
Novembre	0.00147	(0.00502)	Novembre	-0.00099	(0.00312)
Dicembre	-0.00324	(0.00505)	Dicembre	0.00566 [†]	(0.00314)
Eta	0.00466**	(0.00035)	Eta	-0.00002	(0.00022)

Tabella 5.4: **Regressioni eliminando le variabili di controllo**

È facile notare che i risultati anche in questa nuova specificazione non differiscono molto da quelli ottenuti con la forma del modello scelta. La Figura 5.4

mostra graficamente questi risultati.



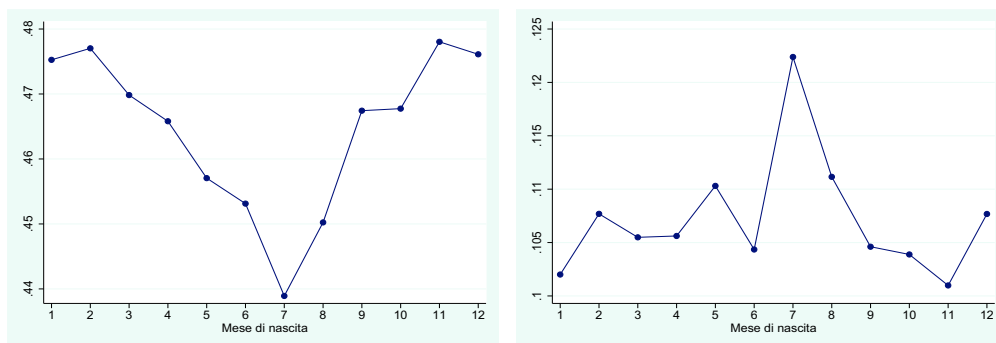
(a) Massimo licenza media

(b) Laurea o post-laurea

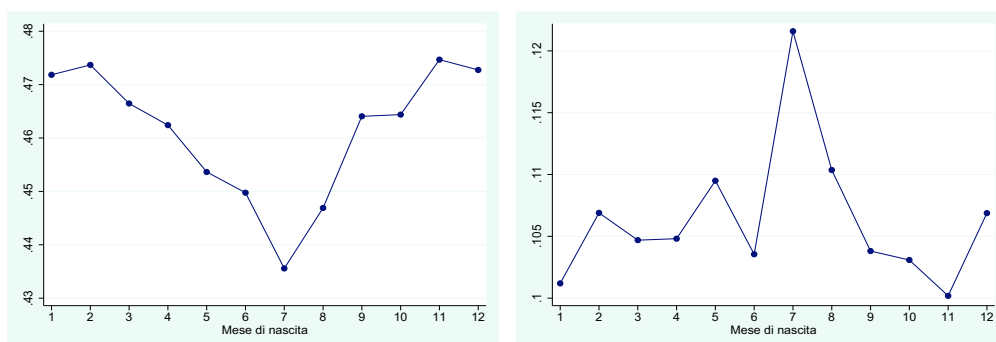
Figura 5.4: **Grafici dei parametri delle regressioni eliminando le variabili di controllo**

Un'ultima verifica è stata fatta sul grado del polinomio del trend per l'età. Per il modello finale si è deciso di utilizzare in tutti i casi un polinomio di primo grado in quanto il trend per l'età nella fascia 42-51 anni sembra essere lineare e per una maggiore semplicità del modello. Qui di seguito si mostrano le stime dei modelli inserendo, invece che un trend lineare, un trend polinomiale di secondo e di terzo grado (Tabella 5.5 e Figura 5.5).

Come si può notare gradi del polinomio superiori al primo risultano talvolta significativi, ma anche nel caso in cui lo siano (si veda la Tabella 5.5(b)) i risultati di interesse sono molto simili a quelli del modello con un polinomio di primo grado. Per una maggiore semplicità del modello si è pertanto preferito tenere il grado più basso.



(a) Massimo licenza media, polinomio di 2° grado (b) Laurea o post-laurea, polinomio di 2° grado



(c) Massimo licenza media, polinomio di 3° grado (d) Laurea o post-laurea, polinomio di 3° grado

Figura 5.5: Grafici dei parametri delle regressioni con diversi gradi del polinomio per l'età

(a) Massimo licenza media

Variabile	Coefficiente	(Std. Err.)
Intercetta	0.47526**	(0.00442)
Febbraio	0.00177	(0.00496)
Marzo	-0.00542	(0.00485)
Aprile	-0.00945 [†]	(0.00498)
Maggio	-0.01821**	(0.00492)
Giugno	-0.02213**	(0.00499)
Luglio	-0.03634**	(0.00492)
Agosto	-0.02502**	(0.00495)
Settembre	-0.00782	(0.00493)
Ottobre	-0.00751	(0.00494)
Novembre	0.00278	(0.00500)
Dicembre	0.00085	(0.00503)
Centro	-0.01563**	(0.00302)
Sud	0.08244**	(0.00219)
Femmina	-0.01406**	(0.00202)
Eta	-0.00109	(0.00128)
Eta2	0.00065**	(0.00014)

(b) Laurea o post-laurea

Variabile	Coefficiente	(Std. Err.)
Intercetta	0.10201**	(0.00276)
Febbraio	0.00568 [†]	(0.00309)
Marzo	0.00349	(0.00303)
Aprile	0.00361	(0.00310)
Maggio	0.00830**	(0.00307)
Giugno	0.00234	(0.00311)
Luglio	0.02038**	(0.00307)
Agosto	0.00915**	(0.00309)
Settembre	0.00261	(0.00308)
Ottobre	0.00188	(0.00308)
Novembre	-0.00102	(0.00312)
Dicembre	0.00567 [†]	(0.00314)
Centro	0.01559**	(0.00188)
Sud	0.00303*	(0.00137)
Femmina	0.00149	(0.00126)
Eta	-0.00248**	(0.00080)
Eta2	0.00027**	(0.00009)

(c) Massimo licenza media

Variabile	Coefficiente	(Std. Err.)
Intercetta	0.47184**	(0.00464)
Febbraio	0.00185	(0.00496)
Marzo	-0.00538	(0.00485)
Aprile	-0.00943 [†]	(0.00498)
Maggio	-0.01820**	(0.00492)
Giugno	-0.02208**	(0.00499)
Luglio	-0.03627**	(0.00492)
Agosto	-0.02495**	(0.00495)
Settembre	-0.00778	(0.00493)
Ottobre	-0.00746	(0.00494)
Novembre	0.00284	(0.00500)
Dicembre	0.00090	(0.00503)
Centro	-0.01565**	(0.00302)
Sud	0.08241**	(0.00219)
Femmina	-0.01406**	(0.00202)
Eta	0.00526 [†]	(0.00289)
Eta2	-0.00123	(0.00078)
Eta3	0.00014*	(0.00006)

(d) Laurea o post-laurea

Variabile	Coefficiente	(Std. Err.)
Intercetta	0.10120**	(0.00289)
Febbraio	0.00570 [†]	(0.00309)
Marzo	0.00350	(0.00303)
Aprile	0.00362	(0.00310)
Maggio	0.00830**	(0.00307)
Giugno	0.00235	(0.00311)
Luglio	0.02039**	(0.00307)
Agosto	0.00917**	(0.00309)
Settembre	0.00262	(0.00308)
Ottobre	0.00189	(0.00308)
Novembre	-0.00101	(0.00312)
Dicembre	0.00568 [†]	(0.00314)
Centro	0.01559**	(0.00188)
Sud	0.00303*	(0.00137)
Femmina	0.00150	(0.00126)
Eta	-0.00097	(0.00180)
Eta2	-0.00017	(0.00049)
Eta3	0.00003	(0.00004)

Tabella 5.5: Regressioni cambiando il grado del polinomio del trend per l'età

Capitolo 6

Conclusioni

Lo scopo di questa tesi era quello di valutare se vi fosse una relazione tra il mese di nascita di un individuo e la durata degli studi. In particolare si voleva verificare se fosse presente un effetto dell'età relativa di un individuo all'interno delle classe scolastica di appartenenza sul titolo di studio conseguito.

Per far ciò si sono regrediti i titoli di studio di 244,098 individui dai 42 ai 51 anni, su alcune variabili tra le quali i mesi di nascita.

I risultati ottenuti sembrano indicare che il metodo con cui vengono formate le classi non comporti alcun effetto sulla carriera scolastica. La probabilità di continuare gli studi oltre l'obbligo scolastico è, infatti, uguale per gli individui nati in Gennaio e in Dicembre (rispettivamente i più grandi e i più piccoli di ciascun anno scolastico). Una possibile spiegazione di tale risultato sta nel fatto che è possibile che per la fine della scuola dell'obbligo questo *relative age effect* sia già svanito e che quindi non sia osservabile più alcuna differenza tra i bambini più grandi e più piccoli di ciascuna classe.

Si è però osservato un netto effetto stagionale: gli individui nati in estate,

infatti, hanno una maggiore probabilità di proseguire gli studi rispetto ai bambini nati ai due estremi dell'anno (vi sono ben il 4% in più di persone nate in Luglio che proseguono gli studi rispetto alle persone nate in Gennaio o Dicembre). Un'altra evidenza emersa è che queste persone nate in estate, oltre a proseguire con più probabilità gli studi, più verosimilmente arrivano anche a conseguire la laurea. Questo risultato è interpretabile grazie a degli studi che dimostrano una maggiore intelligenza dei bambini nati durante i mesi estivi e ad una precaria situazione familiare ed economica dei bambini nati nel corso dei mesi invernali.

Bibliografia

ANGRIST, J. D. e KRUEGER, A. B. (1991), «Does compulsory school attendance affect schooling and earnings?», *Quarterly Journal of Economics*, vol. 106 (4), p. 979–1014.

BARNSLEY, R. H. e THOMPSON, A. H. (1988), «Birthdate and success in minor hockey: The key to the NHL», *Canadian Journal of Behavioural Science*, vol. 20, p. 167–176.

BARNSLEY, R. H., THOMPSON, A. H. e BARNSLEY, P. E. (1985), «Hockey success and birthdate: The relative age effect», *Canadian Association for Health, Physical Education and Recreation*, vol. 51, p. 23–28.

BARNSLEY, R. H., THOMPSON, A. H. e LEGAULT, P. (1992), «Family planning: Football style. The relative age effect in football.», *International Review for the Sociology of Sport*, vol. 227 (1), p. 77–88.

BEDARD, K. e DHUEY, E. (2006), «The persistence of early childhood maturity: International evidence of long-run age effects», *The Quarterly Journal of Economics*, vol. 121 (4), p. 1437–1472.

- BOYD, J. H., PULVER, A. E. e STEWART, W. (1986), «Season of birth: Schizophrenia and bipolar disorder», *Schizophrenia Bulletin*, vol. 12 (2), p. 173–188.
- BUCKLES, K. e HUNGERMAN, D. M. (2008), «Season of birth and later outcomes: Old questions, new answers», *NBER working paper series*.
- DE VITIIS, C., DI CONSIGLIO, L. e FALORSI, S. (2005), *Studio del disegno campionario per la nuova rilevazione continua sulla Forze di Lavoro*, ISTAT - Servizio Progettazione e Supporto Metodologico.
- DIPASQUALE, G. W., MOULE, A. D. e FLEWELLING, R. W. (1980), «The birthdate effect», *Journal of Learning Disabilities*, vol. 13, p. 4–8.
- DUDINK, A. (1994), «Birth date and sporting success», *Nature*, vol. 368, p. 592.
- FREDRIKSSON, P. e OCKERT, B. (2005), «Is early learning really more productive? The effect of school starting age on school and labor market performance», *IZA Discussion Paper No.1659*.
- GENOVESI, G. (1998), *Storia della scuola in Italia dal Settecento a oggi*, Laterza, Roma-Bari.
- GLAMSER, F. D. e MARCIANI, L. M. (1992), «The birthdate effect and college athletic participation: Some comparisons», *Journal of Sport Behavior*, vol. 15, p. 227–238.
- GOODMAN, R., GLEDHILL, J. e FORD, T. (2003), «Child psychiatric disorder and relative age within school year: Cross sectional survey of large population sample», *British Medical Journal*, vol. 327, p. 472–475.

- ISTAT (2006), *La rilevazione sulle forze lavoro: contenuti, metodologie, organizzazione*, Roma.
- LIAO, T. F. (1994), *Interpreting probability models: logit, probit and other generalized linear models*, Sage University Paper.
- ORME, J. E. (1963), «Intelligence, season of birth and climatic temperature», *British Journal of Psychology*, vol. 54 (3), p. 273–276.
- PUHANI, P. e WEBER, A. (2005), «Does the early bird catch the worm? Instrumental variable estimates of educational effects of age of school entry in Germany», *IZA Discussion Paper No. 1827*.
- SEMERARO, A. (1996), *Il sistema scolastico italiano*, La Nuova Italia Scientifica, Roma.
- SHEPARD, L. A. e SMITH, M. L. (1986), «Synthesis of research on school readiness and kindergarten retention», *Educational Leadership*, vol. 44 (3), p. 78–86.
- THOMPSON, A. H., BARNSLEY, R. H. e DYCK, R. J. (1999), «A new factor in youth suicide: The relative age effect», *Canadian Journal of Psychiatry*, vol. 44 (1), p. 82–85.
- THOMPSON, A. H., BARNSLEY, R. H. e STEBELSKY, G. (1991), «Born to play ball: The relative age effect and major league baseball», *Sociology of Sport Journal*, vol. 8, p. 146–151.
- WATSON, P. E. e McDONALD, B. W. (2007), «Season variation of nutrient intake in pregnancy: Effects on infant measure and possible influence on

diseases related to season of birth», *European Journal of Clinical Nutrition*,
vol. 61, p. 1271–1280.