

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA SPECIALISTICA

IN

STATISTICA E INFORMATICA



Tesi di Laurea Specialistica:

**"FECONDITA', MOBILITA' SOCIALE E
IMMIGRAZIONE. ANALISI DELLE COORTI
1910-1970 NEGLI STATI UNITI
E NEL NORD-OVEST D'ITALIA"**

Relatore: Ch.mo Professore

Giampiero DALLA ZUANNA

Laureanda: Camilla DECIO

Anno Accademico 2007-2008

INDICE	Pag.
1. INTRODUZIONE	5
1.1. SCHEMA DELL'ELABORATO	6
2. TEORIA E IPOTESI ALLA BASE DEL PROGETTO	8
3. DATI E METODO	21
3.1. STATI UNITI: CARATTERISTICHE GENERALI DELLA GSS (GENERAL SOCIAL SURVEY)	23
3.1.1. GSS: I DATI	25
3.1.2. GSS: I NUOVI SVILUPPI	28
3.1.3. GSS: IL DISEGNO DI CAMPIONAMENTO E I PESI	30
3.1.4. GSS: LE VARIABILI	35
3.2. ITALIA: CARATTERISTICHE DELLA MULTISCOPO	37
3.2.1. MULTISCOPO: INDAGINE "FAMIGLIA E SOGGETTI SOCIALI"	39
3.2.2. MULTISCOPO: DISEGNO CAMPIONARIO E TECNICA DI INDAGINE	41
3.2.3. MULTISCOPO: LE VARIABILI	50
4. I RISULTATI SUGLI STATI UNITI	52
4.1. PRIMA FASE	52
4.2. SECONDA FASE	59
4.3. TERZA FASE	78
4.4. QUARTA FASE	85
4.5. QUINTA FASE	90
4.6. CONCLUSIONI STATI UNITI	95
5. I RISULTATI SUL NORD-OVEST D'ITALIA	97
5.1. PRIMA FASE	99
5.2. SECONDA FASE	105
5.3. TERZA FASE	120
5.4. QUARTA FASE	126
5.5. QUINTA FASE	130
5.6. CONCLUSIONI NORD-OVEST D'ITALIA	135
6. CONFRONTI E CONCLUSIONI	137
6.1. PRIMA FASE	138

6.2. SECONDA FASE	140
6.3. TERZA FASE	146
6.4. QUARTA FASE	150
6.5. QUINTA FASE	153
6.6. RIEPILOGO	157
7. ALLEGATI	159
7.1. ALLEGATO A: STATI UNITI	159
7.2. ALLEGATO B: NORD-OVEST D'ITALIA	167

1. INTRODUZIONE

Il rinnovo di una popolazione è frutto della combinazione fra nascite e saldo migratorio. In linea generale, gli studiosi hanno sempre considerato in modo distinto queste due componenti della dinamica demografica, sia nella misura che nell'analisi delle loro determinanti.

In questa tesi cercherò invece di mostrare come, nella moderna società occidentale, nascite e immigrazioni si intrecciano fra loro nel rendere possibili il rinnovo e la crescita delle popolazioni. Definirò quindi un modello omeostatico (dotato di un suo equilibrio intrinseco) in cui sia la bassa fecondità che le immigrazioni vengono collocate all'interno delle spinte alla *mobilità sociale intergenerazionale*, che vedono i figli cercare una collocazione sociale migliore rispetto a quella dei loro genitori.

Partendo dai lavori di Dalla Zuanna (*"Population replacement, social mobility and development in Italy in the twentieth century"*), questa idea verrà prima sviluppata e criticata, e poi empiricamente verificata con riferimento a due popolazioni caratterizzate da importanti flussi immigratori, ossia l'Italia del Nord-Ovest e gli USA, nella seconda metà del Novecento.

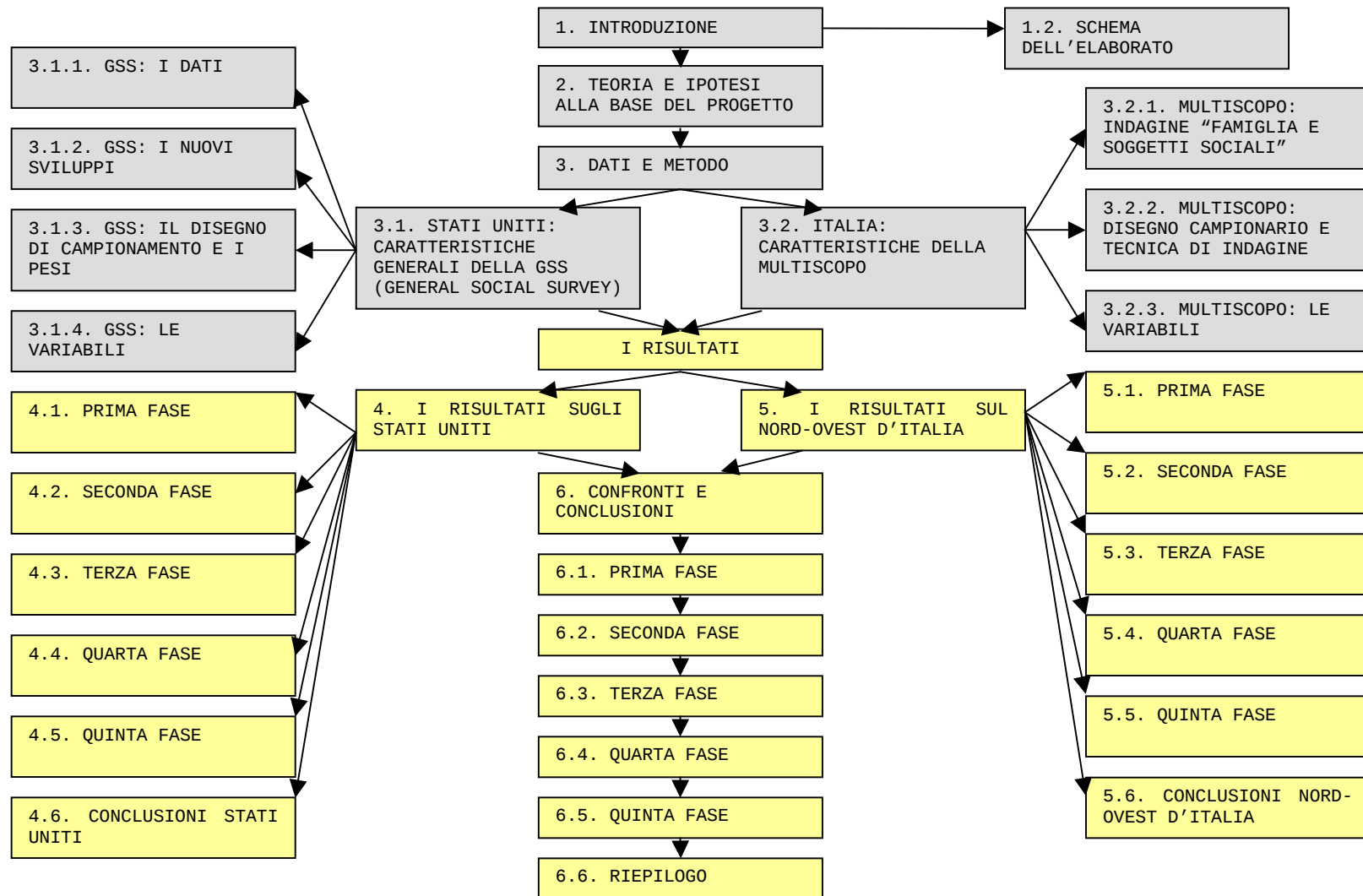
1.1. SCHEMA DELL'ELABORATO

Lo scopo di questo paragrafo è la presentazione dell'intero elaborato attraverso le diverse fasi di costruzione del progetto e le modalità di svolgimento.

Si riportano le fasi necessarie al raggiungimento dell'obiettivo, specificando quali di queste sono state svolte in precedenza, da altri, e quali, invece, riguardano direttamente il progetto di tesi.

Ho deciso di costruire uno schema che segua lo svolgimento della tesi tramite i suoi diversi capitoli, per semplificarne la lettura.

Schema1.2: fasi principali dell'elaborato schematizzate. In giallo sono evidenziate le fasi che sono state elaborate da me, mentre in grigio quelle già svolte da altri, e spiegate nell'elaborato per completezza.



2. TEORIA E IPOTESI ALLA BASE DEL PROGETTO

La teoria su cui si basa l'intero progetto riguarda alcuni aspetti della demografia, importanti per comprendere a fondo le diverse fasi dello sviluppo sociale e produttivo dei paesi sviluppati. Questi concetti vengono presentati ora in generale.

Il primo aspetto è la cosiddetta *mobilità sociale*. Per mobilità sociale si intende il passaggio di un individuo o di un gruppo di individui da uno status sociale ad un altro. La *mobilità sociale* che interessa valutare in questo studio è quella *intergenerazionale*, quella cioè valutata confrontando lo status sociale dell'individuo con quello dei suoi genitori, considerando per semplicità il titolo di studio (o il prestigio) dei soggetti in esame come variabile che indica il loro stato sociale.

Un secondo aspetto considerato è la *fecondità* delle donne, cioè la manifestazione concreta della capacità della donna di procreare, analizzata nel caso specifico attraverso le variabili che indicano o il numero di fratelli (se si analizza la generazione dei figli) o il numero di figli (se si analizza la generazione dei genitori) di un individuo¹.

Terzo aspetto fondamentale è l'*immigrazione*, cioè l'ingresso in un paese, in maniera permanente o semipermanente, di gruppi di persone provenienti da un altro paese.

Una teoria illustrata dal professor Dalla Zuanna, per quanto riguarda il Nord-Ovest d'Italia, prevede cinque fasi principali che collegano questi tre fattori in modo

¹ Massimo Livi Bacci, "Introduzione alla demografia". Terza edizione.

ciclico. Si riporta lo schema riassuntivo in pagina seguente².

La prima generazione osservata (generazione 1) è quella che abbassa la *fecondità*, e ciò porta, nella generazione successiva (generazione 2), a giovani con numero basso di fratelli. Ciò pone le basi per una maggiore accumulazione di capitale umano per questi giovani, che riescono a realizzare processi di *mobilità sociale*. Di conseguenza, nella generazione 2 c'è una carenza di persone adatte a svolgere i lavori più umili.

² Giampiero Dalla Zuanna, "*Population replacement, social mobility and development in Italy in the twentieth century*", *Journal of Modern Italian Studies*."

Tabella1.3: le 5 fasi previste dal modello che si sta verificando.

Fase	Generazione protagonista	Fenomeno osservato
1) In un periodo di sviluppo sostenuto, le coppie che limitano la fecondità possono garantire ai loro figli, che sono già nati, un più elevato capitale umano e maggiori opportunità per la crescita sociale.	1	Residenti: abbassamento fecondità
2) I figli istruiti rifiutano i lavori meno pagati che esigono basso capitale umano. Di conseguenza, c'è una carenza di persone adatte a quei tipi di lavori.	2	Residenti: mobilità sociale
3) Nuovi immigrati prendono il posto dei bambini non nati, svolgendo i lavori meno pagati e meno prestigiosi, riempiendo così gli scalini più bassi della piramide sociale.	2	Immigrazione
4) Gli immigrati si assimilano molto velocemente ai comportamenti sociali dei residenti, abbassando la loro fecondità e investendo nel capitale umano per i loro figli, aumentando così possibilità di crescita sociale di questi ultimi.	3	Immigrati: abbassamento della fecondità
5) Nella nuova generazione, i figli istruiti degli immigrati rifiutano i lavori sotto-pagati dei loro genitori, causando una nuova carenza di lavoratori non esperti. In questo modo il ciclo si ripete (si torna alla fase 3).	3	Immigrati: mobilità sociale

Quindi si assiste alla fase di immigrazione di giovani della stessa età degli autoctoni della seconda generazione (generazione 2). Gli autoctoni della seconda generazione, come i loro genitori, hanno pochi figli (generazione 3). Gli immigrati seguono le abitudini dei loro coetanei autoctoni diminuendo la loro fecondità ed investendo nella *mobilità sociale* dei loro figli (generazione 3). Si assiste quindi nuovamente alla carenza di persone disposte a svolgere i lavori meno specializzati, per cui saranno necessari nuovi ingressi di persone dell'età dei figli della seconda generazione (generazione 3).

In generale lo scopo di questa tesi è verificare se questo tipo di modello può funzionare anche in zone diverse rispetto all'Italia del Nord-Ovest, e per fare ciò si scelgono territori in cui il fenomeno immigratorio sia stato rilevante, e si procede con un'analisi simile per le diverse zone per poter confrontare direttamente i risultati.

Le zone di interesse sono Italia del Nord-Ovest e Stati Uniti, per cui il confronto è tra due zone molto diverse tra loro per vari aspetti che verranno presentati in seguito.

Vi sono alcune condizioni necessarie per la realizzazione di un andamento ciclico come quello illustrato in Tabella1.3:

- 1) il reddito deve essere più alto nella zona di arrivo rispetto a quello della zona dai cui l'immigrato proviene;
- 2) di fatto, ci devono essere poche restrizioni sull'immigrazione nella zona di arrivo;
- 3) sia per gli immigrati che per gli autoctoni l'accesso all'istruzione superiore deve essere agevole e a costi ragionevoli;

4) le persone istruite devono guadagnare di più rispetto a quelle meno istruite.

Inoltre, per quanto riguarda i comportamenti riproduttivi, ci sono altre due condizioni necessarie

5) le coppie devono essere in grado di diminuire la loro fecondità;

6) i genitori devono voler investire sulla *mobilità sociale* dei loro figli.

In questa visione la bassa fecondità e l'immigrazione sostenuta sono viste come possibili motori dello sviluppo. Uno sviluppo economico e di popolazione basato sul continuo afflusso di lavoratori di basso capitale umano, disposti a lavorare per bassi salari, potrebbe sembrare di ostacolo all'incremento di produttività, e quindi allo sviluppo. Ciò può essere vero nel breve periodo. Tuttavia, in una visione più globale, la spinta collettiva alla mobilità sociale ascendente, favorita dall'immigrazione, controbilancia l'abbassamento del capitale umano indotto dall'immigrazione. Inoltre, lo sviluppo di una sempre più numerosa classe di lavoratori intellettuali a reddito medio-alto, rende indispensabile la presenza di una numerosa classe di lavoratori di servizio (per curare i loro figli e i loro anziani, lavare i loro vestiti, tenere pulite le loro case, servire i loro piatti al ristorante, eccetera).

Tutto ciò si scontra con la teoria neoclassica che troppi immigrati non specializzati abbassino la produttività di un paese. Dal punto di vista empirico, d'altro canto, è difficile comprendere se sia l'immigrazione ad aumentare lo sviluppo economico o quest'ultimo ad attirare nuovi immigrati; ciò non toglie che, sempre dal punto di vista empirico, i due fattori siano positivamente correlati.

Gli argomenti trattati sono stati di interesse per molti studiosi. Già nel 1985, Judith Blake pubblicò un articolo

in cui si analizzava la società americana³; in particolare, ella mostrò che, se il livello di istruzione del padre influenza in generale il livello di istruzione dei figli, quest'interazione dipende però dal numero di figli, infatti se questo è basso o medio si osserva una buona correlazione tra le due variabili, se invece è alto la correlazione tra il livello di educazione del padre e quella dei figli è quasi nulla.

Altre ipotesi e teorie più generali sono state pubblicate recentemente (nel 2007) da R. Fernández e A. Fogli⁴, e precedentemente (nel 2006) da R. Fernández⁵. I due hanno studiato l'effetto della cultura sui risultati economici degli Stati Uniti, esaminando variabili riguardanti il lavoro (LFT, "*Labor Force Participation*", partecipazione alla forza lavoro delle donne), e la fecondità (TFR, "*Total Fertility Rate*", Tasso Totale di Fecondità) dei paesi di provenienza delle donne. Queste variabili dovrebbero catturare, oltre ad informazioni sulle condizioni economiche e istituzionali passate, anche le ideologie circa il ruolo delle donne e la numerosità ideale della famiglia. Nell'elaborato si dimostra che le variabili culturali hanno potere esplicativo positivo sui risultati individuali nel lavoro e sulla fecondità (anche considerando eventuali variabili di confondimento). Si prova, inoltre, che l'effetto della cultura è amplificato per quei gruppi etnici che tendono a raggrupparsi nelle stesse zone. E' quindi probabile, secondo l'articolo, che la cultura giochi un ruolo importante nella spiegazione dei cambiamenti, nel paese, per quanto riguarda i risultati

³ Judith Blake, "*Number of Siblings and Educational Mobility*", *American Sociological Review*, vol.50, No.1,1985, pp. 84-94.

⁴ Raquel Fernández, Alessandra Fogli, "*Culture: An Empirical Investigation of Beliefs, Work, and Fertility*". Revised January 2007.

⁵ Raquel Fernández, "*Women, Work, and Culture*", November 2006.

economici fondamentali, come l'attività imprenditoriale, la creatività e l'accumulazione di capitale fisico ed umano. Questa teoria può essere utile nell'analisi dei risultati del nostro progetto, in quanto analizza variabili, come il titolo di studio, che saranno basilari nell'elaborazione, anche se l'articolo si concentra solo sulle variabili riguardanti le donne, mentre in questo contesto si analizzeranno raramente i dati divisi per genere.

Altro argomento basilare, oggetto di studio, è la sopracitata *mobilità sociale*, che verrà esaminata con un confronto diretto tra il livello di istruzione (e di prestigio) dei genitori e dei figli, a confronto.

Anche in questo settore la letteratura è vastissima. Uno studio sugli effetti intergenerazionali dei cambiamenti di livello di istruzione delle donne è stato pubblicato, nel 2006, da R. D. Mare e V. Maralani⁶. In particolare, questi analizzano l'effetto delle caratteristiche socioeconomiche di una generazione sul successo socioeconomico della generazione successiva. Mentre solitamente i ricercatori considerano l'associazione tra le caratteristiche dei genitori e quelle dei loro figli, quindi focalizzandosi solo sulle coppie genitore-figlio, quest'indagine considera non solo i cambiamenti a livello familiare, ma anche quelli a livello generazionale. Applicato ad un campione numeroso di donne Indonesiane e delle loro famiglie, il modello studiato mostra che gli effetti del livello di istruzione delle donne sul livello di istruzione della generazione successiva è positivo. Si osserva inoltre, nell'articolo, che gran parte dei cambiamenti nelle caratteristiche socioeconomiche dei genitori, come ad esempio il variare del livello di istruzione, avvengono relativamente presto nella loro vita ed influiscono sulla fecondità e sul

⁶ Robert D. Mare, Vida Maralani, *"The Intergenerational Effects of Changes in Women's Educational Attainments"*.

matrimonio. Cambiando il livello di istruzione, cambia il comportamento delle donne nei confronti del matrimonio e della fecondità e ciò altera, di conseguenza, il numero relativo di figli nati da donne con diverso livello di istruzione. Una prima conclusione dell'articolo è che l'aumento del livello di istruzione all'interno del matrimonio può influenzare sia il livello di fecondità interno che quello esterno al matrimonio; inoltre si conclude che il miglioramento delle condizioni di vita delle donne nel loro periodo di fecondità aumenta, oltre alla loro esposizione totale alla fecondità, anche l'attenzione che possono dare ai loro figli e la probabilità che questi stessi figli sopravvivano all'età adulta.

Ma quest'indagine ha un limite. Se le donne variano sistematicamente su alcuni fattori non considerati, che congiuntamente condizionano matrimonio, fecondità e allevamento dei figli, allora gli effetti stimati del livello di educazione delle donne sul livello educativo dei loro figli può essere soggetto a distorsione (*selection bias*), data dalla fecondità differenziale. Se, per esempio, tra le donne allo stesso livello di educazione, quelle che possono provvedere ad uno sviluppo positivo per i propri figli sono anche quelle che, a causa delle grandi possibilità personali, hanno meno figli, allora gli effetti stimati dell'educazione della madre sul livello di scolarizzazione dei figli può essere influenzata verso il basso. Questo importante risultato è simile a quello ottenuto dalla Blake in tutt'altro contesto economico, e ribadisce l'idea che la bassa fecondità sia un motore per la mobilità sociale di pochi figli nati.

Per il campione di donne Indonesiane si è dimostrato che le unioni "assortite" (con componenti con diverso livello culturale) rinforzano gli effetti positivi dell'aumento del

livello di scolarizzazione delle donne. Le donne più scolarizzate avvantaggiano i propri figli, sia direttamente che indirettamente, sposando un uomo di livello culturale più alto. Tuttavia, ai più alti livelli di educazione, gli effetti del livello educativo della donna sul matrimonio e sull'educazione delle donne nella generazione successiva diminuiscono.

Gli aspetti di questa teoria di interesse per il progetto sono quelli riguardanti la *mobilità sociale intergenerazionale*, aspetto principale da analizzare.

Un'analisi più legata all'aspetto economico e politico del fattore immigrazione è stata pubblicata, nel 1994, da G. J. Borjas⁷. Lo studio riguarda, principalmente, gli Stati Uniti ma presenta, in generale, alcuni aspetti del fenomeno dell'immigrazione.

Le Nazioni Unite stimarono che più di 60 milioni di persone (1,2% della popolazione mondiale) in quel periodo vivevano in uno Stato in cui non erano nati. Sebbene gli immigrati scelgano solitamente destinazioni "*tradizionali*" (più di metà immigrano negli Stati Uniti, in Canada o in Australia), in molti altri paesi stanno arrivando onde migratorie. E' in atto un'accesa discussione politica sull'immigrazione, concentrata su tre interrogativi principali:

- Come si inseriscono gli immigrati nell'economia del paese?
- Che effetto hanno gli immigrati sulle opportunità di lavoro dei nativi?
- Quale politica sull'immigrazione dà maggiori benefici al paese?

⁷ Borjas George J., "*The Economics of Immigration*", Journal of Economic Literature Vol. XXXII (December 1994), pp. 1667-1717.

Gli immigrati che hanno un alto livello di produttività e si adattano velocemente alla situazione nel mercato del lavoro possono dare un contributo significativo allo sviluppo economico, e questa è una teoria simile a quella che si seguirà nel progetto di tesi.

Il forte dibattito sulle politiche di immigrazione da adottare si è spesso appoggiato all'idea diffusa che le ondate immigratorie abbiano effetti avversi sulle opportunità di lavoro dei nativi. In realtà esistono politiche sull'immigrazione molto diverse nei diversi paesi. Il tema più importante dell'articolo è l'individuazione dei fattori che muovono le persone ad immigrare in alcuni paesi e le conseguenze delle diverse scelte di politiche economiche.

Studi affidabili mostrano che, sebbene gli immigrati, generalmente, arrivino in un paese in una condizione economica svantaggiosa rispetto ai nativi, le loro opportunità economiche aumentano rapidamente nel tempo. In un decennio o due dall'arrivo, i guadagni degli immigrati si avvicinano, fino a raggiungere e talvolta sorpassare, i guadagni dei nativi che si trovano nella stessa condizione sociale. Anche questa teoria supporta l'idea che gli immigrati tendano ad avvicinarsi economicamente, e quindi anche socialmente, alla popolazione di origine, teoria che si avvicina alla fase 4 (si veda Tabella1.3) della teoria qui sostenuta.

Inoltre non ci sono prove che suggeriscano che gli immigranti abbiano un impatto negativo sulle opportunità di lavoro dei nativi. In generale l'evidenza empirica ha sottolineato conseguenze molto ottimistiche del contributo dell'immigrazione nell'economia americana.

La ricerca di Borjas mostra, innanzitutto, alcuni fattori che vanno necessariamente considerati:

- le capacità delle ultime ondate di immigrati sono diminuite, ed è quindi improbabile che gli immigrati più recenti raggiungano i guadagni dei nativi durante la loro vita lavorativa;
- sebbene ci sia solo una debole correlazione negativa tra la presenza di immigrati nel mercato del lavoro e i guadagni dei nativi nello stesso mercato, l'immigrazione recente può essere in parte considerata responsabile della diminuzione dei guadagni della popolazione autoctona non specializzata.

Questa osservazione è importante ma non riguarda direttamente il progetto, in quanto i periodi di immigrazione che verranno considerati andranno dal '50 al '70, quindi si considererà la fase dell'immigrazione che non coinvolge gli ultimi immigrati dagli anni '80 in poi.

Una visione simile sulle conseguenze economiche dell'immigrazione è riportata in un articolo di D. Coleman e R. Rowthorn⁸. Questi trattano la situazione del Regno Unito, per trarre però conclusioni più generali dell'effetto economico dell'immigrazione in un Paese. Dal 1997 una nuova politica sull'immigrazione, nel Regno Unito, cambiò gli obiettivi della precedente politica finalizzata a minimizzare gli ingressi. L'immigrazione su grande scala era considerata ufficialmente essenziale per il benessere economico del Regno Unito e utile per l'intera società; sono state quindi introdotte nuove misure per aumentare gli ingressi.

E' sensato, secondo gli studiosi, temere che l'immigrazione su larga scala possa danneggiare gli interessi degli operai nativi non qualificati. L'idea di fondo dello studio è che le conseguenze dell'immigrazione sono di poco conto,

⁸ David Coleman, Robert Rowthorn, *"The Economic Effects of Immigration into the United Kingdom"*, Population and Development Review 30(4): 579-624 (December 2004).

negative, perlopiù passeggiare e che, né i benefici fiscali né quelli economici, reggono il confronto con l'impatto demografico e ambientale permanente. Affermano infine che gli unici veri beneficiari dell'immigrazione sono proprio gli immigrati.

Il difetto di base delle teorie di Borjas e Coleman è l'approccio di breve periodo alle problematiche immigratorie, mentre per la comprensione globale è necessario uno studio sul lungo periodo.

Si tratta quindi di teorie che si contrappongono in parte a quella di base del progetto, ma ogni teoria va presa in considerazione anche se si discosta da quella che si intende verificare.

Studi più recenti sull'immigrazione, in Italia, hanno trattato argomenti simili. Un articolo pubblicato, nel 2002, da A. Venturini e C. Villosio⁹ propone uno studio dell'effetto dell'immigrazione sulle condizioni dei nativi nel mercato del lavoro. In particolare l'obiettivo di fondo è dimostrare che l'impatto è positivo. Per fare ciò vengono considerati due aspetti:

- il rischio di deposizione (da incarico), misurato con la probabilità di passare dalla condizione di occupato a quella di disoccupato;
- l'efficacia della ricerca di lavoro, misurata con la probabilità di passare dalla condizione di disoccupato a quella di occupato in un anno.

I risultati mostrano che nel Nord Italia, dove si trovano molti immigrati, l'effetto di questi ultimi sulla probabilità di trovare lavoro per coloro che cercano un nuovo lavoro è praticamente nullo, mentre per coloro che cercano il primo lavoro si è riscontrato un effetto

⁹ Alessandra Venturini, Claudia Villosio, *"Are Immigrants Competing with Natives in the Italian Labour Market? The Employment Effect"*, IZA DP No. 467.

negativo solo per un anno di studio (1993), positivo invece per tutti gli altri anni.

Dopo un breve scorcio della letteratura e delle diverse teorie sull'argomento si può entrare nella fase più significativa del progetto, e cioè l'analisi dei dati.

3. DATI E METODO

Si riporta in questo capitolo una descrizione approfondita dei dati utilizzati per l'analisi degli Stati Uniti e per quella del Nord-Ovest d'Italia, con riferimento alla popolazione¹⁰, al campione¹¹ e al tipo di indagine da cui provengono, ai limiti e alle caratteristiche di interesse per il progetto. Si presentano inoltre, brevemente, i metodi statistici che vengono utilizzati per l'analisi.

Per quanto riguarda i dati, una delle informazioni più importanti è il tipo di campionamento.

Come tutte le indagini statistiche, anche quelle da noi utilizzate sono riconducibili a cinque fasi¹².

Tabella3.1: le cinque fasi di un'indagine statistica. Sono evidenziate in giallo le fasi in cui si introduce il progetto di tesi.

¹⁰ Per *popolazione* si intende un insieme di individui, stabilmente costituito, legato da vincoli di riproduzione e identificabile da caratteristiche territoriali, politiche, giuridiche, etniche, religiose o culturali in genere.

¹¹ Per *campione* si intende una parte delle unità della popolazione, selezionate secondo delle regole dette "disegno di campionamento" e questo deve essere rappresentativo della popolazione di riferimento.

¹² L'*indagine statistica* è in generale il complesso di operazioni rivolte ad acquisire informazioni su un insieme di elementi, oggetto dello studio.

1) Astrazione	Si delineano gli obiettivi e le modalità di raccolta scelte.
2) Rilevazione dei dati	Si ottengono i dati utili in modo diretto cioè, ad esempio, tramite un'intervista o in modo indiretto, cioè partendo, ad esempio, da registri ufficiali.
3) Trattamento e analisi dei dati	Si codificano i dati e li si memorizzano in un database, nelle modalità necessarie all'elaborazione.
4) Interpretazione dei risultati	Si traggono le conclusioni dei risultati, quindi segue direttamente l'analisi dei dati. È forse la fase principale, in cui è fondamentale confrontare i risultati con gli obiettivi preposti.
5) Diffusione dei risultati	Si decide con che modalità diffondere i risultati dell'indagine e in particolare con quali mezzi.

Nei paragrafi successivi si analizzano i primi due aspetti, i successivi due saranno alla base dei capitoli sui risultati, il quinto aspetto non verrà trattato, perché non attinente con il progetto di tesi.

3.1. STATI UNITI: CARATTERISTICHE GENERALI DELLA GSS (GENERAL SOCIAL SURVEY)

Per l'analisi sugli U.S.A. si fa riferimento ai dati raccolti dalla *"General Social Survey"*¹³.

Si tratta di una delle indagini più durature disponibile oggi, per quanto riguarda indicatori sociali, culturali e politici. Comprendendo un arco temporale di 34 anni (1972-2006), la GSS è largamente utilizzata da una vasta platea di ricercatori, laboratori dipartimentali e centri di analisi di scienze sociali.

Da un terzo di secolo la GSS monitora i cambiamenti sociali della società americana. Si tratta del maggior progetto finanziato dal *"Sociology Program"* della *"National Science Foundation"*. Con lo *"U.S. Census"*, la GSS è la risorsa di dati forse più utilizzata nelle scienze sociali negli USA.

L'indagine contiene un nucleo standard di domande demografiche e attitudinali, di volta in volta vengono poi aggiunti argomenti di particolare interesse. Molte delle domande del nucleo sono rimaste invariate dal 1972, per facilitare gli studi di trend.

I principali studiosi responsabili della GSS sono James A. Davis, Tom W. Smith e Peter V. Marsden.

Le principali caratteristiche della GSS sono l'ampia copertura, le prospettive nazionali e l'attenzione alla qualità dei dati.

Nel 2006 due grandi cambiamenti sono stati riportati: sono state incluse anche interviste in spagnolo, cosicché la popolazione target è oggi quella degli *adulti, che vivono in abitazioni negli Stati Uniti, che parlano inglese o spagnolo*.

¹³ Per ulteriori informazioni si visiti la pagina <http://gss.norc.org/>

Per quanto riguarda le domande di base nella prima indagine del 1972, circa 105 sociologi e scienziati sociali hanno rivisto le bozze del questionario, suggerito revisioni e aggiunte.

Gli argomenti che compaiono sulle indagini sono di tre tipologie:

- *domande permanenti*, che sono presenti in ogni indagine;
- *domande a rotazione*, che compaiono in due terzi delle indagini;
- *domande occasionali*, che si trovano in una singola indagine.

Uno degli obiettivi principali, che può essere considerato inerente alla fase di *astrazione* sopracitata, cioè della pianificazione degli obiettivi per cui un'indagine viene svolta, è la distribuzione immediata di dati recenti, interessanti e di alta qualità, per una varietà di utilizzatori che non sono soci di grandi centri di ricerche. In conformità con quanto detto, il "*Roper Public Opinion Research Center*" ha permesso la riproduzione e la distribuzione dei dati e del "*Codebook*".

Per raggiungere il mio obiettivo di ricerca, ho costruito un file cumulativo che collega tutte le 26 GSS, in un singolo file contenente i dati dal 1972 al 2006.

3.1.1. GSS: I DATI

Entrando più nei particolari, le *General Social Surveys* sono state condotte durante febbraio, marzo e aprile degli anni indicati in Tabella3.2.

Tabella3.2: tabella di frequenza assoluta, relativa e cumulata dei "GSS YEAR FOR THIS RESPONDENT", cioè degli anni in cui si svolgono le interviste.

Anno	Frequenza	Percentuale	Percentuale cumulata
1972	1613	3.2	3.2
1973	1504	2.9	6.1
1974	1484	2.9	9.0
1975	1490	2.9	11.9
1976	1499	2.9	14.9
1977	1530	3.0	17.9
1978	1532	3.0	20.9
1980	1468	2.9	23.8
1982	1860	3.6	27.4
1983	1599	3.1	30.5
1984	1473	2.9	33.4
1985	1534	3.0	36.4
1986	1470	2.9	39.3
1987	1819	3.6	42.9
1988	1481	2.9	45.8
1989	1537	3.0	48.8
1990	1372	2.7	51.5
1991	1517	3.0	54.5
1993	1606	3.1	57.6
1994	2992	5.9	63.5
1996	2904	5.7	69.2
1998	2832	5.6	74.7
2000	2817	5.5	80.2
2002	2765	5.4	85.6
2004	2812	5.5	91.2
2006	4510	8.8	100.0
Totale	51020	100.0	

Come si vede, il numero di persone intervistate ogni anno rimane pressoché costante fino a metà degli anni '90, per poi quasi raddoppiare negli anni successivi, e subire un nuovo incremento nel 2006.

Ogni indagine, dal 1972 al 2006, ha avuto un diverso campione di persone, che parlano inglese, di 18 anni o più, che vivono in condizioni abitative non istituzionali (né in collegi né in case di cura) negli Stati Uniti.

Per le indagini degli anni 1972 e 1974 e per metà delle indagini del 1975 e 1976 è stato usato il *campionamento per quote*¹⁴.

Un *campione probabilistico* è stato proposto per metà delle indagini del 1975 e 1976 e per gli anni successivi¹⁵. Inoltre, nelle indagini del 2004 e del 2006, sono stati *sotto-campionati i non rispondenti*, riproponendo a parte di questi l'intervista.

La lunghezza mediana dell'intervista di un'ora e mezza.

I dati provenienti dalle interviste sono stati trattati in accordo con le procedure standard del *NORC (National Opinion Research Center)*.

Procedure di pulizia (con l'utilizzo di una combinazione delle specificazioni di codifica con le istruzioni per gli

¹⁴ Il *campionamento per quote* è un particolare campionamento a scelta ragionata, realizzato con modalità organizzative del campionamento di comodo, per il quale si prescinde completamente dalla disponibilità di una lista di appartenenti alla popolazione obiettivo. In altri termini i vincoli imposti al campione sono solo quelli imposti dalle quote (che devono essere aggiornate e statisticamente attendibili). Quindi campione e popolazione devono avere pari quote di persone con un determinato carattere secondo l'idea (presunta e non verificata) che, a parità di composizione strutturale, il campione possa ritenersi rappresentativo della popolazione di riferimento.

¹⁵ In statistica, si parla di *campionamento probabilistico* quando ogni soggetto o oggetto di cui è composta la popolazione ha una probabilità nota di essere incluso nel campione. Questo tipo di campione garantisce la rappresentabilità, mentre nei campioni non probabilistici non si possono generalizzare i risultati di indagine. Infatti il campione probabilistico è quel campione i cui risultati possono essere estesi con un certo livello di fiducia (detto anche livello di confidenza) alla popolazione.

intervistatori) sono state usate per sistemare i codici incongruenti o irregolari.

Il dataset cumulativo che viene utilizzato fonde tutte le 26 indagini in un singolo file, ed ogni anno di indagine è trattato come un sottofile. Ciò semplifica estremamente l'utilizzo dei dati della GSS per indagini di trend, ed è possibile perché le indagini sono fra loro *statisticamente indipendenti*.

In questionario contiene tre tipi di domande:

- le domande che vengono lette al rispondente;
- le domande che vengono presentate al rispondente tramite cartellini;
- le domande che vengono risposte direttamente dall'intervistatore, in base alla migliore corrispondenza con la risposta data dal rispondente.

Le risposte sono comunque sempre mutuamente esclusive, quindi un solo codice di risposta può apparire per ogni domanda.

3.1.2. GSS: I NUOVI SVILUPPI

Nelle GSS degli anni 1993-1997, furono riportati i maggiori cambiamenti nel disegno di indagine. La GSS del 1993 fu l'ultima indagine condotta nel vecchio disegno. Nel 1994 furono introdotte le due più grandi modifiche alla GSS:

- 1) il nucleo tradizionale è stato sostanzialmente ridotto per permettere la creazione di *moduli*. Lo spazio per ogni modulo dà maggiore flessibilità per incorporare innovazioni e per includere importanti voci proposte dalla comunità delle scienze sociali;
- 2) è stato utilizzato un nuovo disegno di *campione spezzato*. Il campione consiste di due *sottocampioni paralleli* approssimativamente di 1500 casi ognuno. I due sottocampioni contengono lo stesso nucleo. Il campione A contiene anche un modulo standard su un particolare argomento di attualità, il mini-modulo, e un modulo *ISSP* ("*International Social Survey Program*") sulle donne, il lavoro e la famiglia. Il campione B ha un secondo modulo su un particolare argomento di attualità, e un modulo *ISSP* sulle condizioni ambientali. In effetti, si può pensare che il campione A rappresenti l'indagine tradizionale per il 1994 e il campione B quella tradizionale per il 1995. Ma invece che essere svolte separatamente nei due anni, sono fatte contemporaneamente.

Mentre solitamente si tengono gli argomenti separati, tramite i moduli sugli argomenti di attualità e i moduli *ISSP*, nei campioni A e B, c'è anche l'opzione di sottoporre alcune domande ad entrambi i campioni, se è necessaria una maggiore numerosità di rispondenti.

Nel 1996 e nei successivi anni pari, è stato riutilizzato lo stesso disegno descritto per il 1994. Inoltre, solo nel

1994, è stato ideato un *disegno di transizione* per calibrare gli impatti delle diverse cancellazioni dal nucleo.

Nel 2002 la GSS subì un cambiamento nel metodo di indagine. Tra il 1972 e il 2002 l'indagine era condotta in formato cartaceo (detto "*carta e penna*"), ma dal 2002, la GSS era condotta da un intervistatore assistito dal computer ("*computer-assisted*").

Nel 2004 ci furono altri due grandi cambiamenti:

- 1) fu adottata una nuova struttura del campione, basata sul Censimento ("*US Census*") del 2000;
- 2) fu utilizzato un *sotto-campione di non-rispondenti*, a cui venne riproposto il questionario, e questo implica che deve essere usata una variabile "*peso*" per le analisi.

Queste due caratteristiche restarono nelle indagini successive.

Nel 2006 altri due cambiamenti sono stati fatti:

- 1) furono introdotte le interviste in Spagnolo (come precedentemente citato);
- 2) un *terzo campione* è stato aggiunto per favorire un numero maggiore di supplementi.

3.1.3. GSS: IL DISEGNO DI CAMPIONAMENTO E I PESI

Tabella3.3: riassunto degli schemi di campionamento.

Anno GSS	Tipo di campionamento	Note
1972 - 1974	A grappolo, o a blocchi	
1975, 1976	$\frac{1}{2}$ a probabilità totali + $\frac{1}{2}$ a grappolo, o a blocchi	
1977, 1978, 1980 e 1982 - 2006	A probabilità totali	
1983	Disegno transitorio	$\frac{1}{2}$ del campione dalla lista del 1970 e $\frac{1}{2}$ dalla lista del 1980
1993	Disegno transitorio	$\frac{1}{2}$ del campione dalla lista del 1980 e $\frac{1}{2}$ dalla lista del 1990
2004	Disegno sul Censimento 2000	

Nell'accordo originale della "*National Science Foundation*", si prevedeva un *campionamento a probabilità variabili*. I campioni dal 1972 al 1974 seguirono questo disegno. Questo campionamento a probabilità variabili, descritto sotto, introduce le *quote* al posto dei livelli fissi. Per le indagini 1975-77 la *NFS* garantì i fondi per un *disegno a probabilità totale*, un disegno che garantisce notoriamente performance superiori.

Così, avendo i mezzi per spostarsi verso un campione a probabilità totale con rispondenti pre-designati, gli studi del 1975 e 1976 furono condotti con un *disegno transitorio*, cioè metà a probabilità totali e metà a quote fisse. Il campione era diviso in due parti per diverse ragioni:

- per fornire i dati per possibili confronti di interesse metodologico;
- nel caso ci fossero alcune differenze nel tempo, perché fosse possibile attribuire queste differenze ai cambiamenti di campionamento, oppure ai cambiamenti di tipologie di rispondenti.

Ci sono considerevoli controversie e ambiguità riguardo ai meriti di questi due campioni.

La GSS allora passò al *campione a probabilità totali* per le indagini degli anni: 1977, 1978, 1980 e 1982-2006¹⁶.

Un disegno transitorio di *campione spezzato* fu utilizzato nel 1983 per misurare l'effetto del cambiamento da dal tipo di campione del 1970 a quello usato nel 1980. Metà del campione fu estratto dalla lista del 1970 e metà da quella del 1980. Ancora nel 1993, un disegno simile fu utilizzato per misurare l'effetto del cambiamento dal campione del 1980 a quello del 1990. Metà del campione fu estratto dalla lista del 1980 e metà da quella del 1990. Nel 2004 fu adottato un nuovo *disegno basato sul Censimento* del 2000. Maggiori dettagli sulle strutture dei campioni del 1970, 1980, 1990 e 2000, così come i campioni a quote fisse, verranno riportati di seguito.

La popolazione adulta che viveva in casa (e non in comunità o case di cura di alcun tipo) negli Stati Uniti, nel 1985, copriva circa il 97.3% della popolazione residente. La *copertura* varia per fasce d'età. Tra i giovani nella classe 18-24, il 9.4% della popolazione nel 1980 viveva fuori dalla famiglia (collegi e quartieri militari). Tra i gruppi di età compresa tra i 25 e i 64 anni solo 0.8-1.4% della popolazione viveva fuori dalla famiglia. Per la classe 75 e più 11.4% viveva in case per anziani.

¹⁶ La variabile *SAMPLE* può essere utilizzata per separare il campione a grappolo (o a blocchi) da quello a probabilità totali per le indagini 1970, 1975, 1976, 1980.

Fino al 2006, come visto, solo la popolazione che parlava inglese veniva campionata, ma l'errore non era molto significativo se si pensa che il 98% delle persone adulte, che viveva in famiglia, parlava inglese. Coloro che parlavano solo spagnolo coprivano, all'incirca, il 60-65% delle esclusioni dovute alla lingua. Circa una dozzina di altre lingue ricoprivano le restanti esclusioni. Dal 2006, come già accennato, la GSS campiona anche coloro che parlano spagnolo, oltre a quelle che parlano inglese. Infine, si riportano alcune spiegazioni aggiuntive sui tipi di campionamenti per maggiore completezza.

- *BLOCK QUOTA* (campione a grappolo, o a blocchi): il campione è a più stadi, consiste in un campione iniziale di aree. Ad ogni livello sono usate quote di campionamento per sesso, età e livello di occupazione. Il costo del campione per quote è sostanzialmente minore di quello per un campione a probabilità totali della stessa numerosità; ma c'è, di certo, la possibilità di *distorsioni di campionamento*, dovute soprattutto alle persone che non sono a casa, che non sono controllate dalle quote. Comunque, per ridurre questa distorsione, agli intervistatori vengono date indicazioni precise, cioè di intervistare solo dopo le 15.00 nei giorni lavorativi, o durante i fine settimana o, infine, durante le vacanze. Questo tipo di disegno di campionamento è il più appropriato quando le esperienze passate suggeriscono che le distorsioni sono piccole in relazione alla precisione degli strumenti di misurazione e alle decisioni che devono essere prese. Le "*Primary Sampling Units*" (*PSUs*, unità principale di campionamento) sono le "*Standard Metropolitan Statistical Areas*" (*SMSAs*) o gli stati non metropolitani selezionati nel "*NORC's Master Sample*". Questi *SMSAs* stati sono stratificati

per regioni, età e razza, prima della selezione. Le unità di selezione di secondo stadio erano i "*block groups*" (BGs) e gli "*enumerations districts*" (ENs). Il terzo stadio è quello dei *blocchi*. I *blocchi* sono selezionati con probabilità proporzionale alla numerosità della popolazione. Nelle zone senza blocchi statistici, le misure dei blocchi sono ottenute con un conto sul campo. La media per *cluster* è di sei rispondenti e ciò produce un buon bilanciamento tra precisione e risparmio. Non si entra in questa sede nei dettagli delle singole indagini¹⁷.

- *SUBCAMPIONE DEI NON RISPONDENTI*: il concetto è quello di ricampionare i non rispondenti, aggiustando i pesi per rendere il campione non distorto. Il sottocampione è pesato per rappresentare tutti i non rispondenti. Il sottocampione permette di focalizzarsi su un gruppo più ristretto dei casi difficili per ulteriori tentativi, riducendo, di conseguenza, sia gli errori delle risposte che la distorsione dovuta alle non risposte. Anche in questo caso non si entra nei dettagli, per maggiori informazioni si rimanda alla lettura del "*Codebook*".

La GSS contiene due variabili peso (*OVERSAMP*, *FORMTW*) che devono essere utilizzate nelle analisi dei dati.

Si descrivono brevemente queste variabili:

- *OVERSAMP*: l'indagine del 1982 include un campione aggiuntivo di neri. Per rendere l'indagine una "*cross-section*" significativa, è possibile o escludere il campione di neri (escludendo i codici 4 e 5 dalla

¹⁷ "*GENERAL SOCIAL SURVEYS, 1972-2006: CUMULATIVE CODEBOOK*", March 2007, Conducted for The National Data Program for the Social Sciences at National Opinion Research Center, University of Chicago, Data Distributed by The Roper Center for Public Opinion Research University of Connecticut.

variabile *SAMPLE*) o pesare il file utilizzando questo peso.

- *FORMWT*: problemi con le procedure di randomizzazione nelle indagini del 1978, 1980, 1982-1985 necessitano l'utilizzo di questo peso quando vengono analizzate le variabili che appaiono solo in un modulo.

3.1.4. GSS: LE VARIABILI

Si prendono ora in considerazione le variabili di interesse, cioè sia quelle già esistenti, sia quelle create nel corso dell'indagine. I dettagli sono riportati in allegato per non appesantire la lettura. Le variabili con l'asterisco ("*") sono quelle modificate, o create da me, in fase di elaborazione.

Tabella3.4: le variabili utilizzate nell'analisi delle cinque fasi per gli Stati Uniti.

Nome esteso	Traduzione	Nome in codice	Ricodifica
Respondent ID number	Numero identificativo del rispondente	ID	-
Prestige of respondent's Occupation*	Prestigio dell'occupazione del rispondente	PRESTIGE_OK_CLASSE	1-basso 2-medio basso 3-medio alto 4-alto
Prestige of father's Occupation*	Prestigio dell'occupazione del padre del rispondente	PA_PRESTIGE_OK_CLASSE	1-basso 2-medio basso 3-medio alto 4-alto
Age of Respondent	Età del rispondente	AGE	18-89+
Age of Respondent by class*	Età del rispondente per classi utili	AGE_CLASSE	18-39 40-59 60+
Number of children*	Numero di figli del rispondente	CHILDS_CLASSE	1-2-3+
RS highest degree	Il più alto grado di istruzione del rispondente	DEGREE	0-"Lt high school" 1-"High School" 2-"Junior College", "Bachelor", "Graduate"
Was born in this country*	Il rispondente è nato in questo paese?	BORN	0-Sì 1-No
Where Respondent's Parents Born in this country?*	I genitori del rispondente erano nati in questo paese?	PARBORN	1-Sì 2-No
Year of Birth	Anno di nascita del rispondente (o coorte)	COHORT	1883-1988

Generazione*	Generazione di appartenenza del rispondente	GENERAZIONE	Generazione1- (1910-1929) Generazione2- (1930-1949) Generazione3- (1950-1969)
Number of brothers and sisters	Numero di fratelli e sorelle del rispondente	SIBS	1-2-3-4+
Weights for black over samples	Pesi per campioni di neri	OVERSAMP	Valori originali tra 0 e 1
Post-stratification weights	Pesi assegnati per neutralizzare la stratificazione	FORMWT	Valori originali tra 0 e 1

In ALLEGATO A si mostra come sono state create le variabili, da quelle esistenti nel dataset originale.

3.2. ITALIA: CARATTERISTICHE DELLA MULTISCOPO

La prima indagine *Multiscopo* risale al periodo tra il 1987 e il 1991, di fronte alla grande domanda di informazioni statistiche in un ambito sociale e demografico. Nel 1993 nasce il nuovo *Sistema integrato di indagini Multiscopo*. Il sistema ha l'obiettivo di produrre informazioni su individui e famiglie. Il sistema si articola su sette indagini principali, di cui si riporta uno schema¹⁸.

Tabella3.5: organizzazione delle indagini Multiscopo con il nuovo Sistema integrato del 1993. E' evidenziata l'indagine utilizzata per le analisi.

Argomento	cadenza	Anni di indagine
1- Aspetti della vita quotidiana	Annuale	1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2005
2- Viaggi e vacanze	Trimestrale	1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005
3- Condizione di salute e ricorso ai servizi sanitari	Quinquennale	1994, 2000, 2005
4- I cittadini e il tempo libero	Quinquennale	1995, 2000, 2006
5- Sicurezza dei cittadini	Quinquennale	1998, 2002
6- Famiglie e soggetti sociali	Quinquennale	1998, 2003
7- Uso del tempo	Quinquennale	1988, 1989, 2003

A rotazione nell'arco di cinque anni

¹⁸ "Il sistema di indagini sociali multiscopo. Contenuti e metodologia delle indagini", ISTAT.

Quindi le indagini sono diverse, ma in tutte c'è una forte attenzione alla qualità dei dati e alle fonti di errore, dalla progettazione alla diffusione. Il nuovo *Sistema di indagini Multiscopo* nasce con l'obiettivo di far fronte a due esigenze: quella di sistematizzare le precedenti esperienze di analisi delle problematiche sociali e quella di soddisfare la crescente domanda di informazione statistica nel campo sociale. Il Sistema è infatti pensato per arricchire le informazioni disponibili, per quanto concerne la situazione sociale del Paese, attraverso la produzione di informazioni dettagliate sulla struttura della famiglia, sulle caratteristiche degli individui, sul concreto svolgersi della vita quotidiana, sui bisogni e problemi della popolazione.

L'indagine "*Famiglie e soggetti sociali*", evidenziata in Tabella3.1 è quella che verrà utilizzata nelle analisi per il Nord-Ovest d'Italia.

3.2.1. MULTISCOPO: INDAGINE "FAMIGLIA E SOGGETTI SOCIALI"

L'indagine *"Famiglia, soggetti sociali e condizione dell'infanzia"* è stata condotta nel 1998 e nel 2003. Essa mira ad approfondire lo studio dell'organizzazione della famiglia, come centro della società contemporanea, attraverso l'analisi della sua struttura interna e del ruolo degli individui che la compongono, visti appunto come *soggetti sociali*.

In questo modo i *soggetti sociali* vengono studiati attraverso l'analisi dei diversi contesti in cui essi vivono: la famiglia, la rete di parenti, la rete delle amicizie e di solidarietà, l'ambito lavorativo e quello scolastico, il tempo libero, le abitudini e le tradizioni familiari.

L'obiettivo principale dell'indagine *Famiglia e soggetti sociali* è cogliere i cambiamenti in atto, in ambito sociale e demografico, che rendono il quadro generale della famiglia italiana piuttosto complesso e singolare, nel contesto dei paesi occidentali: si assiste ad un fenomeno di abbassamento della fecondità, all'invecchiamento della popolazione, alla lunga permanenza dei giovani nella famiglia di origine, a bassi livelli di divorzialità e di nascite al di fuori del matrimonio, ma anche a bassi livelli di occupazione femminile e giovanile, persistenti differenze di genere e forti legami fra generazioni.

La seconda indagine *Famiglia e soggetti sociali* ha avuto luogo alla fine del 2003 e, pur mantenendo inalterata la struttura portante dell'indagine condotta nel 1998, si arricchisce di nuovi contenuti tematici e metodologici, che determinano un vero e proprio salto di qualità della statistica sociale ufficiale.

L'indagine condotta nel 2003 ha, in aggiunta, anche una parte di quesiti su opinioni, intenzioni e atteggiamenti in ambito familiare e lavorativo (uscita dalla famiglia di origine, fecondità, inserimento o consolidamento nel mercato del lavoro, conciliazione lavoro-famiglia).

Gli argomenti principali rilevati con quest'indagine sono schematizzati nella tabella successiva.

Tabella3.6: gli argomenti principali dell'indagine *"Famiglia e soggetti sociali"*.

Sezioni	Argomenti
1- Le strutture familiari	- Dati anagrafici e scheda generale; - i pendolari della famiglia; - giovani da 18 a 34 anni che vivono con almeno un genitore.
2- Le reti di parentela	- Fratelli e sorelle; - genitori e nonni; - figli e nipoti; - altri parenti, amici e vicinato
3- Le reti di aiuto informale	- Aiuti dati; - aiuti ricevuti; - servizi privati alla famiglia; - aiuti e difficoltà economiche; - usanze e tradizioni familiari.
4- La mobilità sociale	- Il lavoro nella sua vita; - il lavoro alle dipendenze; - il lavoro nel passato; - le interruzioni del lavoro; - studio e lavoro dei genitori.
5- La vita di coppia	- Fidanzamento e prime nozze; - ciclo di vita e convivenza prematrimoniale; - la vita di coppia.
6- La vita quotidiana di bambini e ragazzi	- Affidamento dei bambini; - scheda generale; - il gioco e la televisione; - tempo libero e amici; - chiavi di casa e autonomia; - lavoretti in casa e aiuto ai familiari.

3.2.2. MULTISCOPO: DISEGNO CAMPIONARIO E TECNICA DI INDAGINE

L'indagine viene condotta con tecnica *Papi* (*Paper and pencil interview*), mediante intervista diretta e questionario autocompilato. Le sezioni riguardanti le strutture familiari, la rete di parentela e di solidarietà, il ciclo di vita e la vita di coppia, sono rilevate con l'ausilio dell'intervistatore mediante intervista diretta con il questionario base. Le sezioni che riguardano la vita lavorativa, le esperienze di lavoro nel passato, le interruzioni e le intenzioni per il futuro, sono rilevate mediante intervista diretta, per tutte le persone di 18 anni e più. Infine lo studio e il lavoro dei genitori, le opinioni e i valori rispetto alla famiglia e al lavoro, la permanenza dei giovani nella famiglia di origine, le intenzioni di fecondità e di uscire dalla famiglia di origine, sono rilevate mediante questionario compilato direttamente dall'intervistato, di 18 anni e più, senza l'ausilio dell'intervistatore.

Il campione teorico dell'indagine è costituito da 24 mila famiglie ed oltre 50 mila individui.

La popolazione di interesse dell'indagine è costituita dalle famiglie residenti in Italia e dagli individui che le compongono, al netto dei membri permanenti delle convivenze.

I domini di studio, ossia gli ambiti territoriali ai quali sono riferiti i parametri di popolazione oggetto di stima sono i domini tradizionali delle indagini *Multiscopo*, ossia le cinque ripartizioni geografiche (Italia nord-occidentale, Italia nord-orientale, Italia centrale, Italia meridionale, Italia insulare); le regioni (ad eccezione del Trentino Alto Adige, le cui stime sono prodotte distintamente per le province autonome di Bolzano e

Trento); la tipologia comunale, ottenuta suddividendo i comuni italiani in classi, formate in base a caratteristiche socioeconomiche e demografiche.

Il *disegno di campionamento* è a *più stadi*, con *stratificazione* delle unità di primo stadio, rappresentate dai comuni. All'interno di ciascun comune viene effettuato un *campionamento a grappolo*. Le famiglie, cioè i grappoli, sono selezionate in maniera sistematica dalle liste anagrafiche dei comuni-campione estratti. A tutti i componenti appartenenti a tale famiglia, considerando non più la famiglia anagrafica ma quella di fatto, viene somministrata l'intervista.

Produrre un'informazione statistica adeguata significa dedicare forte attenzione alla qualità.

La progettazione dell'indagine assume, pertanto, un'importanza rilevante per definire le strategie e i criteri atti a prevenire gli errori non campionari.

L'attenzione alla qualità e alle possibili fonti di errore rientra nell'attività quotidiana ed è sempre presente, a partire dalla progettazione, fino alla diffusione dei dati. Nella progettazione vengono definiti i contenuti e l'architettura dell'indagine, la sequenza delle attività, le modalità di somministrazione ed i criteri di controllo. Nella progettazione si realizza il disegno dell'indagine attraverso:

- la specificazione del problema,
- il disegno campionario,
- la costruzione del questionario,
- la definizione dei metodi e degli strumenti.

Il disegno campionario è la fase attraverso la quale si definisce:

- la lista di riferimento,
- la numerosità campionaria,

- il sistema di campionamento,
- le procedure di stima.

Questa fase, pur essendo complessa, si differenzia dalle altre nel senso che, basandosi, di fatto, su una teoria ben consolidata che guida alla scelta del campione in modo da avere stime corrette, permette di controllare a priori la variabilità della stima, ed i conseguenti errori dovuti alla dimensione campionaria. Sono le altre fasi che possono generare gli errori non campionari e mettere a rischio la qualità dell'indagine se non si attuano adeguate strategie. È per questo motivo che nelle indagini *Multiscopo*, nella progettazione, vengono definite accuratamente le modalità con cui controllare, mantenere ed integrare le varie fasi d'indagine, in modo che sia possibile intervenire, in corso d'opera, per limitare gli errori. Le fasi sono fortemente interrelate tra loro e ciascuna viene attuata tenendo conto di ciò che è accaduto in quelle precedenti. Per la rilevazione, ad esempio, è determinante la progettazione dei criteri di controllo delle interviste, con riferimento al monitoraggio e all'individuazione degli indicatori per il controllo delle attività degli intervistatori.

Allo stesso modo, è fondamentale la progettazione delle attività che intervengono in corso d'opera, per correggere, oppure prevenire, errori che potrebbero ancora verificarsi durante il trattamento dei dati. Ad esempio, gli errori d'identificazione, riguardanti l'errata trascrizione su un questionario *Papi* dei codici identificativi dell'unità di rilevazione, possono spesso essere eliminati attraverso la fase successiva di revisione.

Determinanti sono, poi, i piani di controllo della registrazione e della codifica, insieme ai piani di controllo e correzione del file dati.

Per garantire efficienza, tutte le attività del *Sistema Multiscopo* sono caratterizzate da una forte interazione tra i ricercatori ed integrazione delle fasi.

Le indagini del *Sistema Multiscopo* considerano come *popolazione di interesse* - intesa come l'insieme delle unità statistiche oggetto di investigazione - le *famiglie* residenti in Italia, e gli individui ad esse appartenenti; ciò vuol dire che tali rilevazioni hanno, in generale, la finalità di produrre sia stime riferite alle famiglie che stime riferite agli individui.

E' importante fare presente che, per quanto riguarda le famiglie, si adotta solitamente la definizione di *famiglia di fatto*¹⁹.

Da ciò si desume, quindi, che chi ha dimora abituale nell'abitazione, ma è assente temporaneamente, viene considerato un membro della famiglia, così come è considerato membro della famiglia anche colui che vi dimora abitualmente, anche se non risulta sullo stato di famiglia. Per quanto riguarda i *parametri di interesse*, fondamentale caratteristica delle indagini considerate è quella di rilevare sia variabili che si riferiscono alle caratteristiche degli individui e delle famiglie (sesso, età, stato civile, livello di istruzione, posizione e condizione professionale), sia variabili relative al numero di eventi di un certo tipo, vissuti dagli individui, e dalle famiglie.

Le indagini Istat sulle famiglie hanno generalmente l'obiettivo di fornire stime dei parametri d'interesse (totali, medie, rapporti, eccetera) riferite a differenti *sottopopolazioni*, che possono essere definite o in base a variabili di tipo territoriale, oppure in base a caratteristiche degli individui, o delle famiglie.

¹⁹ Per *famiglia di fatto* si intende un insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.

Relativamente alla scelta della tecnica di rilevazione, è utile mettere in evidenza i vantaggi e gli svantaggi dell'intervista diretta e dell'intervista telefonica. Gli aspetti positivi della prima consistono essenzialmente nel fatto che essa consente, solitamente, di ottenere tassi di risposta più elevati, e di utilizzare una lista della popolazione (l'anagrafe dei comuni), che almeno in teoria non esclude nessuna sottopopolazione, a fronte di costi più elevati. L'intervista telefonica, d'altra parte, consente di contenere il costo per intervista, e può essere più indicata quando i quesiti dell'indagine riguardano fenomeni di particolare delicatezza o complessità. Con l'intervista telefonica, d'altra parte, non è possibile utilizzare la lista anagrafica, perché non contiene i numeri di telefono delle famiglie, ed è necessario ricorrere a una lista di abbonati telefonici.

Tuttavia, poiché le unità appartenenti a tale lista non coincidono esattamente con la popolazione delle famiglie italiane, ciò può introdurre, nel processo di stima, a distorsioni di difficile misurazione.

Per l'indagine di interesse per l'elaborato, la rilevazione è basata su un'organizzazione del lavoro sul campo a livello comunale. Infatti, per quanto riguarda sia la gestione delle liste di selezione delle famiglie, costituite dagli archivi anagrafici, sia la conduzione delle interviste, i comuni costituiscono gli organi periferici responsabili sul territorio. Si rende pertanto necessario limitare il numero di comuni coinvolti nella rilevazione, sia per tenere sotto controllo i costi relativi all'organizzazione della rilevazione e alle operazioni di istruzione dei rilevatori, sia per garantire a ciascun comune un buon numero di interviste; da ciò scaturisce la necessità di ricorrere a un *disegno campionario a due stadi di selezione: comuni e famiglie*.

Tale scelta, tuttavia, determina, a parità di numerosità campionaria, un aumento della variabilità delle stime prodotte e tale aumento è tanto più elevato quanto più è alto il coefficiente di correlazione intra-classe, e il numero medio di interviste per comune; analogamente, la scelta di intervistare tutti gli individui appartenenti ad ogni famiglia campione è legata alla necessità di ridurre le famiglie, e quindi gli indirizzi, da contattare, per intervistare un dato numero di individui campione. Anche in questo caso, la conseguenza è un aumento della variabilità delle stime prodotte, tanto più elevato quanto più sono elevati il coefficiente di correlazione intra-classe delle variabili di interesse all'interno delle famiglie campione, e la dimensione familiare media.

Il *disegno campionario a due stadi* presenta un doppio livello di stratificazione dei comuni:

- una stratificazione in base a variabili di tipo territoriale;
- una successiva stratificazione in base alla dimensione in termini di popolazione.

In generale, l'obiettivo della stratificazione è quello di formare gruppi (o strati) di unità caratterizzate, relativamente alle variabili oggetto d'indagine, da massima omogeneità interna agli strati e massima eterogeneità fra gli strati. Il raggiungimento di tale obiettivo si traduce in termini statistici in un guadagno nella precisione delle stime, ossia in una riduzione dell'errore campionario a parità di numerosità campionaria.

La lista di selezione delle unità di primo stadio è costituita dall'archivio dei comuni italiani, che riporta per ciascun comune: il numero di famiglie e individui residenti, il codice di provincia, regione e ripartizione e altri riferimenti territoriali, quali la tipologia socio-demografica. Le liste di selezione delle unità di secondo

stadio (le famiglie) sono costituite dagli archivi anagrafici dei comuni selezionati al primo stadio; come è noto, i fogli di famiglia anagrafici, che compongono gli archivi, contengono informazioni relative alle singole famiglie (ampiezza della famiglia, sesso ed età dei singoli componenti), tuttavia, tali informazioni sono difficilmente utilizzabili in fase di stratificazione del campione di secondo stadio.

Da quanto detto risulta chiaro che, nel caso delle indagini sulle famiglie condotte mediante intervista diretta, per la stratificazione del campione è possibile utilizzare le variabili territoriali e l'ampiezza anagrafica dei comuni, mentre non è possibile basarsi sulle caratteristiche strutturali delle famiglie o degli individui.

All'interno di ciascuno strato, il disegno, basato su due differenti schemi di campionamento, prevede che i comuni siano suddivisi in due sottoinsiemi sulla base della popolazione residente:

- l'insieme dei *comuni Auto rappresentativi* (che indicheremo d'ora in avanti come comuni Ar) costituito dai comuni di maggiore dimensione demografica;
- l'insieme dei *comuni Non auto rappresentativi* (o Nar) costituito dai rimanenti comuni.

Nell'ambito dell'insieme dei comuni Ar, ciascun comune viene considerato come uno strato a se stante, e viene adottato un disegno noto con il nome di *campionamento a grappoli*. Le unità primarie di campionamento sono rappresentate dalle famiglie anagrafiche, estratte in modo sistematico dall'anagrafe del comune stesso; per ogni famiglia anagrafica inclusa nel campione vengono rilevate le caratteristiche oggetto di indagine di tutti i componenti di fatto appartenenti alla famiglia medesima.

Nell'ambito dei comuni Nar, invece, viene adottato un *disegno a due stadi con stratificazione delle unità*

primarie, in cui le unità primarie sono i comuni, e le unità finali sono le famiglie anagrafiche.

Sia nella parte Ar che nella parte Nar, per ogni famiglia anagrafica inclusa nel campione vengono rilevate le caratteristiche oggetto di indagine di tutti i componenti di fatto appartenenti alla famiglia medesima.

All'interno di ciascuno strato territoriale, i comuni vengono stratificati in base alla loro dimensione demografica e nel rispetto delle seguenti condizioni:

- autoponderazione del campione a livello regionale;
- definizione del numero n di comuni da selezionare da ciascuno strato definito sui comuni dell'insieme Nar ($n= 1$ per tutte le indagini, salvo l'indagine sulle condizioni di salute per la quale $n= 4$);
- scelta di un numero minimo di famiglie da intervistare in ciascun comune campione;
- formazione di strati aventi ampiezza approssimativamente costante in termini di popolazione residente.

Effettuata la stratificazione, i comuni Ar sono inclusi con certezza nel campione; per quanto riguarda, invece, i comuni Nar, nell'ambito di ogni strato viene estratto un comune campione con probabilità proporzionale alla dimensione demografica.

La selezione delle famiglie da intervistare in ogni comune campione viene effettuata dalla lista anagrafica di ciascun comune, senza reimmissione, e con probabilità uguali. In particolare, la *tecnica di selezione* è di tipo *sistematico*. I criteri seguiti possono essere sintetizzati nei seguenti punti:

- la dimensione del campione teorico, in termini di famiglie, viene prefissata a livello nazionale, essenzialmente in base a criteri di costo ed operativi, ed è pari a circa 24 mila famiglie per tutte le indagini

dirette, tranne che per l'indagine sulle condizioni di salute;

- il numero di comuni campione interessati non deve essere superiore a una certa soglia (900 circa per le indagini basate su un campione di 24 mila famiglie) in modo da consentire un buon lavoro di controllo e supervisione.

3.2.3. MULTISCOPO: LE VARIABILI

Si prendono ora in considerazione le variabili di interesse, cioè sia quelle già esistenti, sia quelle create nel corso dell'indagine. I dettagli sono riportati in allegato per non appesantire la lettura. Le variabili con l'asterisco ("*") sono quelle modificate, o create da me, in fase di elaborazione.

Tabella3.7: le variabili utilizzate nell'analisi delle cinque fasi per il Nord-Ovest d'Italia.

Nome	Spiegazione ufficiale	Codifica
Codice_famiglia	Progressivo famiglia, univoco a livello indagine	-
Ordine_componente	Numero d'ordine del componente nell'ambito della famiglia	-
Peso*	Coefficiente finale	-
Relazione_parentela	Relazione di parentela con la persona di riferimento del questionario	-
Anno_Nascita*	Calcolato come differenza tra l'anno dell'indagine (2003) e l'età del soggetto	-
Età	Età (scheda generale)	-
Generazione*	Generazione di appartenenza	- Generazione 1 (1910-1929); - Generazione 2 (1930-1949); - Generazione 3 (1950-1969).
Fratelli*	Numero di fratelli avuti	0-1-2-3-4-5-6+

Istruzione*	Titolo di studio più alto conseguito	0- nessun titolo (non sa leggere e/o scrivere), nessun titolo (sa leggere e scrivere), licenza elementare; 1- licenza scuola media inferiore, diploma scuola media superiore (2-3 anni); 2-diploma scuola media superiore (4-5 anni), diploma universitario, laurea, dottorato di ricerca o specializzazione post-laurea.
Provenienza*	Indica se il soggetto è nato o meno nel Nord-Ovest d'Italia	1-autoctono 2-immigrato
Figli_ok_classe*	Numero di figli viventi naturali, adottati o affiliati	0-1-2-3-4-5-6+
Dove_vivono_genitori*	Indica se i genitori vivono a più o meno di 50 chilometri	1- Più di 50 chilometri 2- Meno di 50 chilometri
Prestigio*	Livello di prestigio del soggetto	1- basso 2- medio basso 3- medio alto 4- alto
Prestigio_padre*	Livello di prestigio del padre del soggetto	1- basso 2- medio basso 3- medio alto 4- alto
Prestigio_padre_dicotomico*	Livello di prestigio dicotomico del padre del soggetto	1- basso e medio basso 2- medio alto e alto

In ALLEGATO B si mostra come sono state create le variabili, da quelle esistenti nel dataset originale.

4. I RISULTATI SUGLI STATI UNITI

4.1. PRIMA FASE

“In un periodo di sviluppo sostenuto, le coppie che limitano la fecondità possono garantire ai loro figli, che sono già nati, un più elevato capitale umano e maggiori opportunità per la crescita sociale.”

L'obiettivo è verificare se, effettivamente, la diminuzione di fecondità (nella generazione 1) porta ad un aumento del livello di istruzione della generazione dei figli (generazione 2); quindi studio la correlazione tra il numero di fratelli e il livello di istruzione raggiunto. Si considera, in questo studio, la generazione 2, di coloro che hanno genitori nati negli USA.

Tabella4.1: *numero di fratelli e livello di istruzione della generazione 2 (nati nel 1930-1949) nati negli Stati Uniti, secondo il livello di prestigio del padre. Osservazioni pesate. Classe di età all'intervista '40-59'. Percentuali di riga.*

	Livello di istruzione			Totale
Numero di fratelli	0	1	2	
(A) Padri di livello di prestigio 1 (n = 782.10)				
0	10%	61%	29%	100% (34.34)
1	9%	67%	24%	100% (111.21)
2	17%	57%	26%	100% (108.45)
3	25%	58%	17%	100% (114.92)
4	18%	68%	14%	100% (94.80)
5	36%	55%	9%	100% (69.40)
6+	46%	47%	7%	100% (248.98)
(B) Padri di livello di prestigio 2 (n = 3680.50)				
0	7%	57%	36%	100% (214.41)
1	6%	53%	41%	100% (645.42)
2	10%	56%	34%	100% (659.61)
3	11%	60%	29%	100% (574.47)
4	18%	60%	22%	100% (406.14)
5	18%	65%	17%	100% (288.74)
6+	36%	52%	12%	100% (891.71)
(C) Padri di livello di prestigio 3 (n = 939.16)				
0	6%	36%	58%	100% (96.15)
1	1%	43%	56%	100% (314.13)
2	4%	48%	48%	100% (300.22)
3	3%	51%	46%	100% (203.78)
4+	13%	48%	39%	100% (325.10)
(D) Padri di livello di prestigio 4 (n = 159.91)				
0	0%	8%	92%	100% (12.00)
1	5%	11%	84%	100% (39.95)
2	0%	24%	76%	100% (45.48)
3	3%	24%	73%	100% (33.00)
4+	10%	27%	63%	100% (29.48)

Le percentuali riportate in tabella sono di riga, cioè, per i diversi livelli di prestigio del padre, e per ogni categoria che indica il numero di fratelli, indica la

distribuzione in percentuale del livello di istruzione raggiunto.

L'andamento è tendenzialmente simile per i diversi livelli di prestigio del padre; in particolare, sono state evidenziate le maggiori percentuali per ogni colonna, che indicano, cioè, la percentuale di persone che ha raggiunto il livello di istruzione corrispondente alla colonna, al variare del numero di fratelli. Chiara è la correlazione negativa tra il numero di fratelli e il livello di istruzione: all'aumentare del numero di fratelli, la percentuale di persone che non riesce a superare i livelli inferiori aumenta, e diminuisce invece la percentuale di coloro che raggiungono il maggior livello di istruzione.

Confrontando, invece, le percentuali, al variare del livello di prestigio del padre, si osserva che, per i figli con padre di livello più basso, la *mobilità sociale* è molto più difficile rispetto agli altri, e ciò è rappresentato da percentuali molto più alte di persone al livello più basso di istruzione. Questa percentuale diminuisce notevolmente all'aumentare del livello di prestigio del padre.

Quindi, oltre ad aver provato che il numero di fratelli incide negativamente sulle possibilità di istruzione del soggetto, si è dimostrato anche che c'è una notevole correlazione tra il livello di prestigio del padre, e il livello di istruzione che la famiglia può garantire ai figli.

Per dimostrare, in modo più rigoroso, queste conclusioni si riporta un test Chi-Quadrato²⁰ tra numero di fratelli e titolo di studio, per ogni livello di prestigio del padre.

²⁰ Il test *Chi-Quadrato di Pearson* (o della bontà dell'adattamento) è un test non parametrico, applicato a grandi campioni, quando si è in presenza di variabili nominali, e si vuole verificare se il campione è stato estratto da una popolazione con una predeterminata distribuzione o che due o più campioni derivino dalla stessa popolazione. Fa parte di un'ampia

Tabella4.2: P-value del test Chi-Quadrato tra *numero di fratelli* e *livello di istruzione* della generazione 2 (nati nel 1930-1949) nati negli Stati Uniti, secondo il *livello di prestigio del padre*.

Prestigio del padre	P-value
1	<.0001
2	<.0001
3	<.0001
4	0.1963

La tabella mostra molto chiaramente la relazione forte tra il numero di fratelli e il livello di prestigio. Difatti, per coloro che hanno il padre di livello di prestigio '1', '2' o '3', il P-value del test risulta significativo contro l'ipotesi nulla di indipendenza tra le due variabili. Per coloro che hanno padre benestante sembra, invece, meno

classe di test detti test chi quadrato, in quanto hanno in comune le formule e la variabile casuale Chi quadrato, ma non necessariamente anche le ipotesi di base o le finalità. Questa variante del test utilizza i dati di un solo campione e verifica l'ipotesi nulla che il campione sia stato estratto da una popolazione di cui è nota la distribuzione. Sia

$$X^2 = \sum_{i=1}^g \frac{(n_i - E_i)^2}{E_i} = \sum_{i=1}^g \frac{n_i^2}{E_i} - n$$

n_i è il numero di casi osservati nel campione per la i -esima modalità, E_i è il numero di casi *attesi* nel caso l'ipotesi nulla fosse vera, g è il numero di modalità nella quale si esprime la variabile nominale

$$n = \sum_{i=1}^g n_i = \sum_{i=1}^g E_i$$

è la numerosità del campione. Allora X^2 è distribuita approssimativamente come una variabile casuale Chi Quadrato con $(g-1)$ gradi di libertà $X^2_{(g-1)}$. Si richiede però che tutte le frequenze attese E_i raggiungano un valore minimo (a seconda delle esigenze, almeno pari a 5 oppure almeno pari a 10). Qualora ci siano delle frequenze attese troppo piccole, bisogna procedere ad un raggruppamento di modalità.

rilevante il numero di fratelli nella spiegazione del livello di prestigio, questo perché è intuitivo credere che genitori benestanti riescano a garantire un buon livello di istruzione anche ad un alto numero di figli.

L'obiettivo è ora verificare se l'aumento del numero di fratelli porta ad una diminuzione, oltre che del livello di istruzione, anche del livello di prestigio della seconda generazione.

Tabella4.3: numero di fratelli e livello di prestigio della generazione 2 (nati nel 1930-1949), nati negli Stati Uniti, secondo il *livello di prestigio del padre*. Osservazioni pesate. Classe di età all'intervista '40-59'. Percentuali di riga.

Numero di fratelli	Livello di prestigio				totale
	1- basso	2- medio basso	3- medio alto	4-alto	
(A) Padri di livello di prestigio 1 (n = 756.61)					
0	18%	46%	31%	5%	100% (38.34)
1	22%	43%	32%	3%	100% (110.21)
2	20%	48%	30%	2%	100% (103.21)
3	24%	55%	17%	3%	100% (113.92)
4	26%	49%	23%	2%	100% (89.16)
5	31%	56%	12%	1%	100% (67.96)
6+	27%	56%	15%	1%	100% (233.14)
(B) Padri di livello di prestigio 2 (n = 3580.16)					
0	12%	45%	38%	5%	100% (208.41)
1	8%	49%	38%	5%	100% (629.94)
2	10%	50%	36%	4%	100% (642.71)
3	11%	54%	30%	5%	100% (561.23)
4	17%	55%	25%	3%	100% (393.90)
5	17%	59%	23%	1%	100% (281.5)
6+	24%	56%	19%	1%	100% (862.47)
(C) Padri di livello di prestigio 3 (n = 1227.90)					
0	6%	34%	52%	8%	100% (94.15)
1	3%	43%	47%	7%	100% (313.13)
2	6%	46%	42%	6%	100% (297.22)
3	8%	44%	45%	3%	100% (202.78)
4+	9%	46%	40%	5%	100% (320.62)
(D) Padri di livello di prestigio 4 (n = 156.91)					
0	8%	25%	42%	24%	100% (12.00)
1	3%	30%	52%	15%	100% (38.95)
2	2%	18%	60%	20%	100% (44.48)
3	6%	41%	34%	19%	100% (32.00)
4+	13%	24%	55%	8%	100% (29.48)

Anche in questa tabella sono state riportate le percentuali di riga; queste mostrano, per i diversi livelli di prestigio del padre, e fissato un numero di fratelli, la distribuzione percentuale del livello di prestigio raggiunto.

La tabella mostra che, all'aumentare del numero di fratelli, il livello di prestigio dell'individuo, indipendentemente dal livello di prestigio del padre, tende a diminuire.

Le distribuzioni sono però significativamente diverse al variare del livello di prestigio del padre.

Sono state evidenziate le percentuali maggiori per ogni colonna, queste indicano la percentuale di persone che, per le diverse classi di numero di fratelli, hanno raggiunto il livello di prestigio rappresentato dalla colonna.

Confrontando quindi le tabelle relative ai diversi livelli di prestigio dei padri, si osserva, come già visto per il livello di istruzione, che l'*ascesa sociale* è più difficile per coloro che hanno padre di livello di prestigio basso rispetto agli altri.

Anche in questo caso si preferisce un'evidenza empirica, rispetto alla sola osservazioni delle percentuali. Si costruisce quindi, anche per questi dati, il test Chi-Quadrato.

Tabella4.4: P-value dei test Chi-Quadrato tra *numero di fratelli* e *livello di prestigio* della generazione 2 (nati nel 1930-1949) nati negli Stati Uniti, secondo il *livello di prestigio del padre*.

Prestigio del padre	P-value
1	0.0095
2	<.0001
3	0.1819
4	0.3129

Questa tabella mostra una tendenza abbastanza allineata coi risultati ottenuti con il test tra titolo di studio e numero di fratelli. Anche in questo caso, per coloro che hanno padre di livello di prestigio basso o medio basso è provata, tramite il test Chi-quadrato, la dipendenza tra livello di prestigio e numero di fratelli. Ancora una volta, per coloro che hanno padre benestante, di livello di prestigio medio alto o alto, la relazione non è molto forte, perché genitori benestanti riescono a garantire buone prospettive ai loro figli, anche se molti. La prima fase è quindi verificata.

4.2. SECONDA FASE

“Figli istruiti rifiutano i lavori meno pagati che esigono basso capitale umano. Di conseguenza, c’è una carenza di persone adatte a quei tipi di lavori.”

L’idea è che i figli istruiti (generazione 2) della generazione 1 tendono, e riescono, a superare il livello di prestigio dei loro genitori. Quindi oggetto di studio è ancora la generazione 2 e osservo, ancora una volta, solo i figli di autoctoni.

Tabella4.5: *livello di prestigio* per generazione 1 (1910-1929) e 2 (1930-1949), nati negli USA. Classe di età all’intervista ‘40-59’.

	Livello di prestigio				Totale
	1	2	3	4	
Generazione 1 (1910-1930)	26%	52%	20%	2%	100% (1095.60)
Generazione 2 (1930-1950)	14%	51%	31%	4%	100% (4084.00)

In tabella sono riportate le percentuali per riga, che rappresentano la distribuzione, in percentuale, del livello di prestigio raggiunto dalle due generazioni.

E’ chiara la differenza. Sono state evidenziate, per ogni colonna, le percentuali maggiori; quelle corrispondenti ai due livelli più bassi di prestigio sono più alte per i padri, viceversa per i livelli più alti.

E’ quindi chiaro che i figli desiderano, e riescono, a superare il livello di prestigio dei loro padri.

Si confronta ora, direttamente, il livello di prestigio delle persone appartenenti alla generazione 2 (1930-1949) con il livello di prestigio dei loro padri, per entrare più a fondo nelle *dinamiche familiari*, più che generazionali (trattate alla tabella precedente).

Tabella4.6: *livello di prestigio* dei figli (generazione 2, 1930-1949) sulla base del *livello di prestigio dei loro genitori*. Classe di età all'intervista '40-59'. In riga c'è il prestigio dei padri e in colonna quello dei figli.

	Prestigio dei figli					Totale
		1	2	3	4	
Prestigio dei padri	1	25%	52%	21%	2%	100% (757.70)
	2	14%	53%	29%	4%	100% (3587.60)
	3	6%	44%	44%	6%	100% (1144.90)
	4	6%	27%	50%	17%	100% (148.91)

In Tabella4.6 sono riportate le percentuali di riga, cioè, per ogni livello di prestigio, è riportata la distribuzione percentuale del livello di prestigio dei figli (generazione 2).

Si è evidenziata la percentuale più alta di ogni colonna, e si nota che c'è una tendenza dei dati a posizionarsi sulla diagonale principale: c'è quindi una forte corrispondenza tra il livello di prestigio dei padri e quello dei figli.

Ma si nota che, ad esempio, per livello di prestigio 3 dei figli, la percentuale di riga più alta non corrisponde al livello 3 dei padri, ma al livello 4, quindi si assiste ad una notevole *mobilità sociale* fra le generazioni.

In generale, si è dimostrata la forte tendenza *all'ascesa sociale* per i figli di autoctoni, che porta necessariamente ad una carenza di personale adatto ai lavori più umili.

Anche in questo caso si riporta un test statistico per verificare o meno l'ipotesi di indipendenza.

Tabella4.7: test Chi-quadrato per il *livello di prestigio dei figli* (generazione 2, 1930-1949) sulla base del *livello di prestigio dei loro genitori*. Classe di età all'intervista '40-59'.

Statistica	DF	Valore	Prob
Chi-quadrato	9	332.1809	<.0001

Il P-value del test porta a rifiutare l'ipotesi nulla di indipendenza tra livello di prestigio del padre e livello di prestigio del figlio, anche se non dà informazioni sulla *mobilità sociale* o sull'ordinamento delle variabili.

Si osserva, ora, il livello di istruzione dei figli di autoctoni, per verificare se è vero che i figli di questi, derivanti dalla prima fase di osservazione, raggiungono, all'aumentare del livello di istruzione, livello di prestigio più alto.

Tabella4.8: livello di prestigio sul livello di istruzione dei figli (generazione 2, 1930-1949), con genitori nati negli USA. Classe di età '40-59'. Osservazioni pesate.

	Livello di prestigio					Totale
		1	2	3	4	
Livello di istruzione	0	39%	53%	8%	0%	100% (1096.80)
	1	14%	62%	23%	1%	100% (3508.80)
	2	2%	29%	59%	10%	100% (1957.00)

Anche in questa tabella sono riportate le percentuali di riga, cioè la distribuzione del livello di prestigio, al variare del livello di istruzione.

Si osserva una chiara dipendenza, infatti all'aumentare del livello di istruzione del soggetto, tende ad aumentare il livello di prestigio che questo riesce a raggiungere, l'andamento è *diagonale*, anche se non esiste una vera e propria diagonale, non trattandosi di una tabella quadrata. Si riporta un test per verificare questa dipendenza in modo più preciso.

Tabella4.9: test Chi-quadrato per il *livello di prestigio* sul *livello di istruzione* dei figli (generazione 2, 1930-1949) con genitori nati negli USA. Classe di età all'intervista '40-59'.

Statistica	DF	Valore	Prob
Chi-quadrato	6	1961.6232	<.0001

Il P-value relativo al test porta a rifiutare l'ipotesi nulla di indipendenza tra livello di istruzione e livello di prestigio per la generazione 2.

Quindi, quanto visto finora, dimostra che anche la seconda fase è verificata.

Si decide di costruire dei modelli statistici per verificare le ipotesi con tecniche statistiche più precise. Si sceglie come variabile risposta il *livello di prestigio* del soggetto e come variabili esplicative le seguenti variabili:

- *numero di fratelli,*
- *livello di prestigio del padre,*
- *titolo di studio del soggetto.*

Oggetto di quest'analisi è la seconda generazione (1930-1949).

La fase fondamentale dell'analisi esplorativa è la comprensione e lo studio della *variabile risposta*, cioè quella per cui vengono richieste delle previsioni.

Si considerano tutte le variabili come fattoriali, tranne la risposta, che devo rendere categoriale.

Le unità che contengono un dato mancante in almeno una delle variabili di interesse sono 12734, decido di eliminarle, perché non c'è motivo per supporre che i dati mancanti siano fonte di distorsione.

Ma prima di fare ciò, verifichiamo, in parte, la validità di questa ipotesi, riportando le distribuzioni per alcuni incroci di variabili di interesse anche per i dati

mancanti, in modo da essere sicuri che le persone che non rispondono ad alcuni quesiti siano simili per quanto riguarda le altre variabili.

Le variabili che verranno utilizzate, ora e di seguito per i modelli sono le tre prima elencate, la variabile *prestigio* e la variabile *born* (che indica se il soggetto è nato o meno negli USA). Incrociamo quindi alcune variabili per osservare i dati mancanti. Si riportano le percentuali di riga:

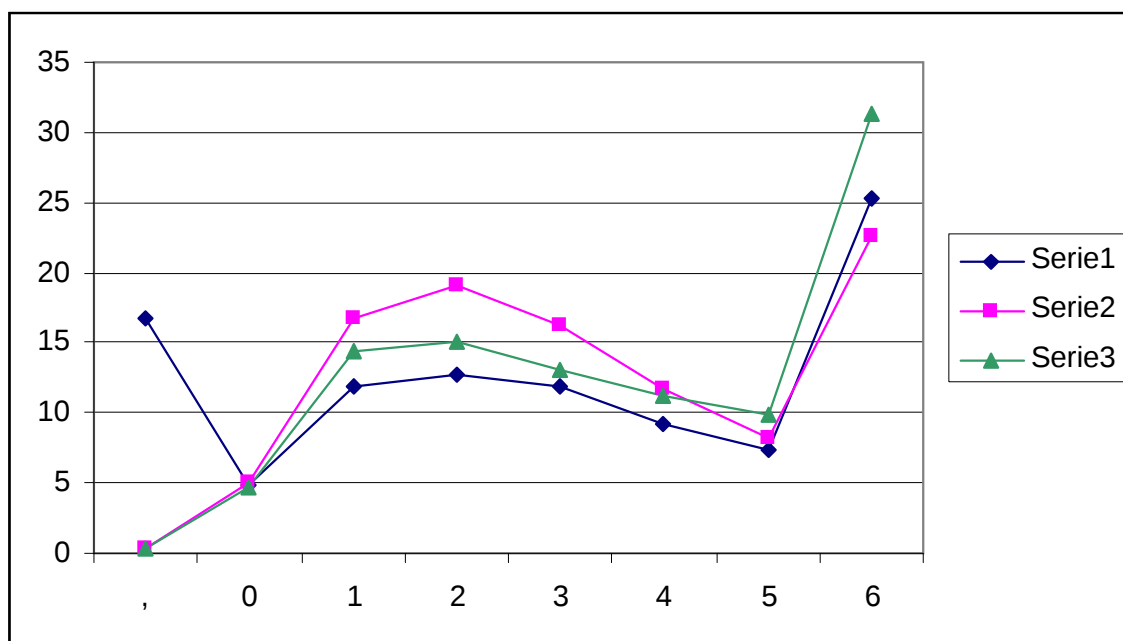
♦ *Born* incrociata col *numero di fratelli*:

Tabella4.10: percentuali di riga per la variabile *numero di fratelli* sulla base della variabile *born* (*provenienza*), considerando le modalità dei valori mancanti.

Born	Numero di fratelli								Totale di riga
	.	0	1	2	3	4	5	6	
.	16.71	4.93	11.97	12.81	11.83	9.13	7.31	25.31	100%
Autoctono	0.27	5.00	16.75	19.16	16.17	11.76	8.21	22.68	100%
Immigrato	0.34	4.69	14.47	15.14	13.04	11.19	9.87	31.26	100%

Si riporta il grafico delle percentuali di riga, che è più leggibile della tabella.

Grafico4.11: percentuali di riga del *numero di fratelli* per le tre modalità della variabile *born* (*provenienza*): dato mancante (serie 1), autoctono (serie 2), immigrato (serie 3).



Osservando il grafico si conclude, come previsto, che la distribuzione di coloro che non hanno risposto alla domanda sulla loro provenienza, per la variabile sul numero di fratelli, non è differente da quella dei rispondenti alla prima. Segue invece l'andamento, anche se ovviamente la curva è un po' più bassa, perché spesso chi non ha risposto alla domanda sulla provenienza, non ha risposto neanche alla variabile sul numero dei fratelli, ma per il resto le percentuali sono simili, e quindi eliminare i dati mancanti non sembra essere fonte di distorsione.

Ma facciamo un altro esempio, con altre variabili:

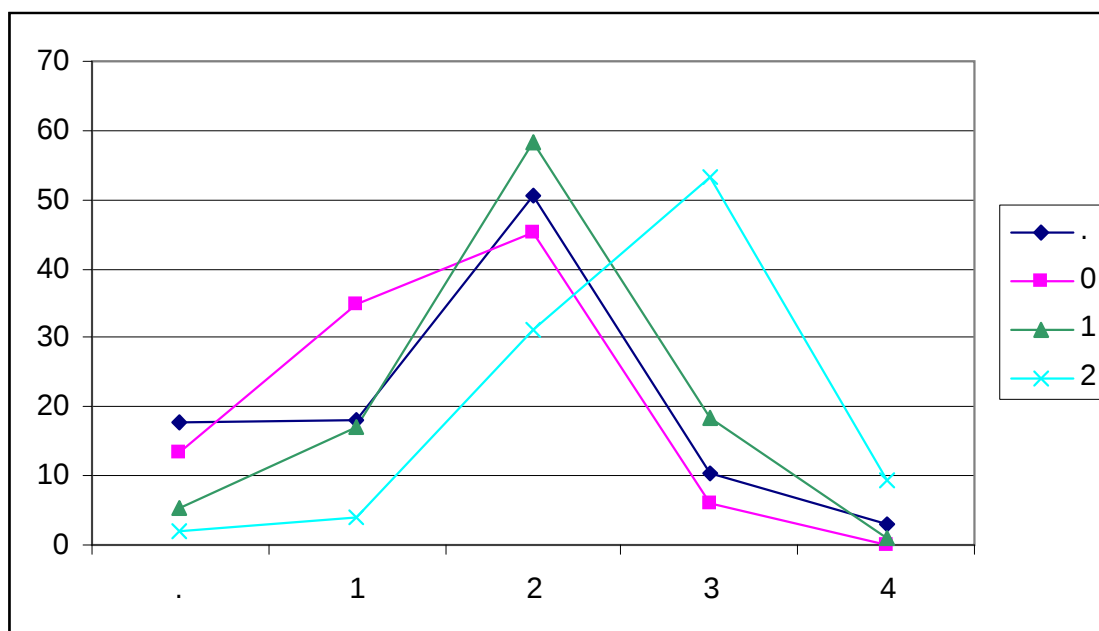
♦ *Prestigio* del soggetto incrociato col *titolo di studio*:

Tabella4.12: percentuali di riga per le variabili: *prestigio* del soggetto e *titolo di studio* considerando le modalità dei valori mancanti.

		Livello di prestigio					
		.	1	2	3	4	Totale
Titolo di studio	.	17.89	18.24	50.51	10.27	3.09	100%
	0	13.56	34.92	45.29	6.07	0.15	100%
	1	5.26	17.18	58.14	18.56	0.86	100%
	2	1.94	4.08	31.10	53.39	9.50	100%

Si riporta il grafico delle percentuali di riga, che è più leggibile della tabella.

Grafico4.13: percentuali di riga del prestigio sul titolo di studio.



Anche da questo grafico sembra che la serie dei dati mancanti per il titolo di studio sia simile alle altre, senza tendenze particolari che facciano supporre una particolare vicinanza ad una delle altre serie. Quindi si è verificato che eliminare i dati mancanti non può portare a grandi distorsioni nelle analisi.

Si procede quindi all'eliminazione.

Quindi restano 38286 persone. Di queste 11960 sono della generazione 2.

Per ogni analisi decido di dividere il dataset in due parti di stima e verifica²¹.

Considero tre casi di interesse:

1) *Variabile categoriale del livello di prestigio: '0' indica i livelli '2', '3', e '4' (medio basso, medio alto e alto) e '1' indica livello '1' (basso).*

Si decide di stimare un modello logistico²² che ha come variabile risposta la dicotomica e come esplicative tutte le variabili rese fattoriali.

²¹ Le tecniche per la selezione del modello sono le seguenti:

-*Insieme di stima e di prova*: si dividono i dati in due gruppi, la prima parte viene usata per la stima dei diversi modelli possibili, la seconda per la verifica della validità.

-*Convalida incrociata*: consiste nell'incrociare i ruoli delle k porzioni di dati, di cui (k-1) gruppi vengono usati per la stima e uno per la verifica, e poi si scambiano i ruoli, quindi si otterranno k stime di cui si stimerà la curva media. K è al massimo uguale ad n. Per n grande è il metodo migliore.

-*Criteri basati sull'informazione*: si tiene conto del numero di parametri nella stima di massima verosimiglianza, grazie ad una penalizzazione del tipo $IC = -2 \cdot \log L(\hat{\theta}) + \text{penalità}(p)$.

-*Selezione automatica delle variabili*: procedimento automatizzato per la scelta dei modelli che scelgono il gruppo di variabili che minimizza l'AIC (selezione passo-a-passo).

²² Per *modello logistico* si intende un modello binomiale con funzione legame logit. In generale, se la variabile risposta è dicotomica, non è conveniente utilizzare un modello lineare, ma può essere utile usare la regressione logistica

$$l(x) = \frac{\exp(x)}{1 + \exp(x)},$$

che è uno dei casi dei GLM (*Modelli Lineari Generalizzati*). La parametrizzazione che si usa è detta d'angolo.

Coefficients:	Estimate	Std.Error	z value	Pr(> z)	
(Intercept)	-0.304840	0.199863	-1.525	0.1272	
sibs_mio1	-0.112099	0.206081	-0.544	0.5865	
sibs_mio2	0.003838	0.200755	0.019	0.9847	
sibs_mio3	0.123928	0.202335	0.612	0.5402	
sibs_mio4	0.408607	0.183916	2.222	0.0263	*
degree1	-1.151535	0.084082	-13.695	< 2e-16	***
degree2	-2.676765	0.154500	-17.325	< 2e-16	***
papres_ok_classe2	-0.453177	0.091123	-4.973	6.58e-07	***
papres_ok_classe3	-0.828987	0.140685	-5.892	3.80e-09	***
papres_ok_classe4	-0.745305	0.367719	-2.027	0.0427	*

Il modello ha lo scopo di verificare quali variabili concorrono ad aumentare la probabilità di abbandonare il livello più basso di prestigio, per raggiungere i livelli '2', '3' e '4' considerati congiuntamente (categoria '0').

Si nota che il *numero di fratelli* non è generalmente significativo per spiegare il *livello di prestigio*, a parte la categoria '4 o più' di fratelli, che è significativa al 95%, e il coefficiente è positivo, quindi avere 4 o più fratelli aumenta la probabilità di passare dalla categoria '0' alla categoria '1', cioè al livello più basso di prestigio.

Inoltre, per quanto riguarda il *titolo di studio*, le categorie '1' e '2' (cioè tutte tranne la '0', corrispondente al minimo livello di istruzione) risultano significative quasi al 100%; il loro coefficiente è negativo e ciò indica che avere titolo di studio medio, o alto, aumenta la probabilità di abbandonare la categoria di prestigio pari a '1', cioè quella corrispondente al livello di prestigio più basso.

Infine si nota che le categorie dei livelli di *prestigio del padre* '2', '3' e '4' (quindi tutte tranne il livello basso, '1') sono tutte significative, ed hanno stima del coefficiente negativa: ciò significa che avere padre con livello di prestigio diverso da '1', diminuisce la

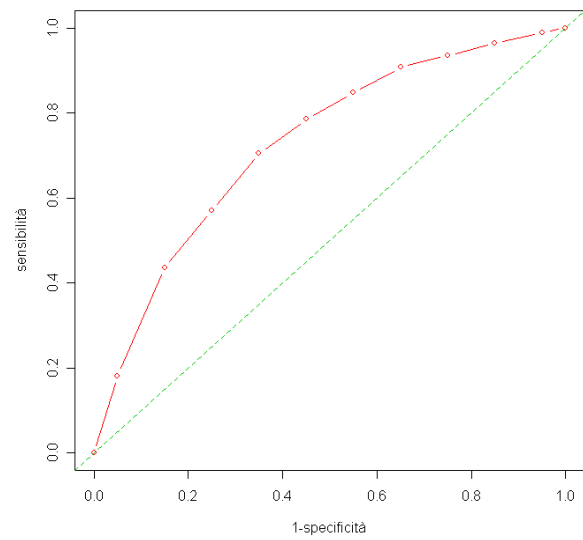
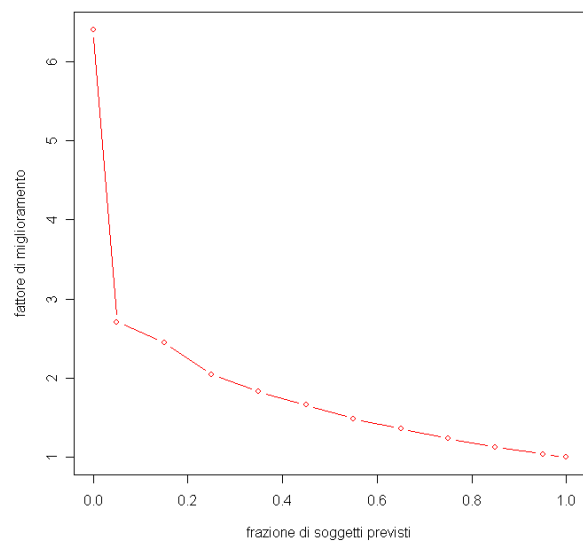
probabilità di raggiungere il livello di categoria '1', cioè di restare al livello inferiore di prestigio.

Quindi, per verificare la validità del modello, si provano delle previsioni, per vedere, usando il dataset di verifica, quanto si indovinerebbe la categoria di livello di prestigio, avendo a disposizione le altre tre variabili. Si costruiscono le curve *Lift*²³ e *ROC*²⁴, per avere un'indicazione grafica della bontà di previsione del modello.

Le due curve, riportate, nell'ordine, in seguito, dimostrano entrambe che il modello è buono. Ciò significa che, volendo prevedere la categoria del titolo di studio utilizzando le variabili risposta note, si ha un guadagno in efficienza maggiore rispetto alla scelta casuale di queste, sulla base della proporzione nota di soggetti appartenenti alle due categorie.

²³ *Curva Lift*: è la rappresentazione grafica della funzione lift, che indica il miglioramento ottenuto rispetto ad una classificazione casuale al variare della frazione di soggetti "positivi" previsti nel campione.

²⁴ *Curva ROC*: in ascissa si ha (1-specificità), dove per specificità si intende $n_{11}/(n_{11}+n_{21})$, cioè la proporzione tra i veri negativi e il totale dei negativi reali. In ordinata si ha la sensibilità, $n_{22}/(n_{12}+n_{22})$, cioè la proporzione tra i veri positivi e il totale dei positivi reali. Il modello va bene se la curva si discosta molto dalla bisettrice, che indica la classificazione casuale.



2) *Variabile categoriale del livello di prestigio: '0' indica i livelli '3' e '4' (medio alto e alto) e '1' indica livelli '1' e '2' (basso e medio basso)*

Si riportano i primi risultati del modello logistico.

Coefficients:	Estimate	Std.Error	z_value	Pr(> z)	
(Intercept)	2.44750	0.19243	12.719	< 2e-16	***
sibs_mio1	0.07656	0.14072	0.544	0.5864	
sibs_mio2	0.05031	0.14151	0.356	0.7222	
sibs_mio3	0.16222	0.14679	1.105	0.2691	
sibs_mio4	0.29675	0.13440	2.208	0.0272	*
degree1	-1.18148	0.13462	-8.776	< 2e-16	***
degree2	-3.15716	0.13993	-22.563	< 2e-16	***
papres_ok_classe2	-0.10078	0.09763	-1.032	0.3019	
papres_ok_classe3	-0.53050	0.11252	-4.715	2.42e-06	***
papres_ok_classe4	-0.21327	0.21447	-0.994	0.3200	

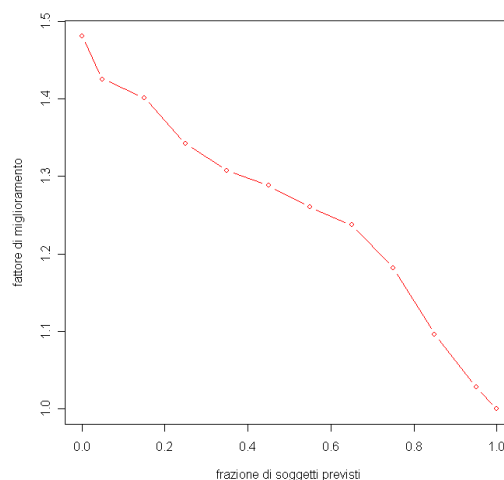
A differenza del modello precedente l'intercetta risulta significativa. Osservando il *numero di fratelli* del soggetto, si nota che le classi più basse non sono significative, mentre è di interesse la classe '4 o più', e il coefficiente corrispondente è positivo: ciò indica che passare da non avere fratelli ad averne 4 o più aumenta la probabilità di passare al livello di prestigio di categoria '1' (corrispondente alle classi di prestigio più basse, '1' e '2').

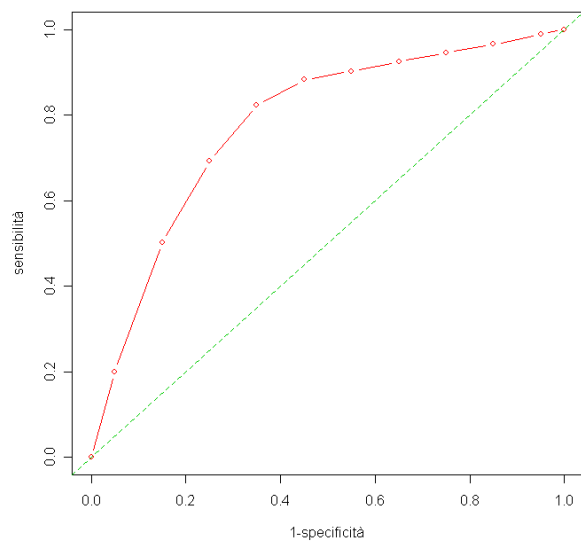
Le categorie relative al *titolo di studio* '1' e '2' (medio e alto) del soggetto, sono entrambe significative, e le stime dei loro coefficienti sono negative, ciò significa che raggiungere livelli di istruzione più alti aumenta la probabilità di raggiungere la categoria di prestigio '0', cioè quella più alta (relativa ai livelli '3' e '4' di prestigio).

Infine, per quanto riguarda *livello di prestigio del padre*, questo è significativo solo per la classe '3' e il

corrispondente coefficiente è negativo, quindi avere il padre di livello di prestigio '3', quindi medio alto, aumenta la probabilità, per il soggetto, di raggiungere un livello di prestigio alto (classe '0' della categoria, corrispondente ai livelli '3' e '4' di prestigio). Risulta un po' anomalo il fatto che risulti significativa la categoria '3' e non la '4', ma possono accadere certe anomalie quando si lavora con dati in ambito sociale, in quanto possono interferire una serie di altri fattori (*background sociale*) non osservabili, considerati, con ipotesi un po' ottimistiche, costanti.

Le curve *Lift* e *ROC* sono simili al caso precedente, ma riportate per analogia di analisi. Anche questo modello è efficace nella previsione della categoria di livello di prestigio del soggetto, note le tre variabili di interesse.





3) *Variabile categoriale del livello di prestigio: '0' indica il livello '4' di prestigio (alto) e '1' indica i livelli '1', '2' e '3' di prestigio (basso, medio basso e medio alto)*

Si riportano, anche per quest'ultimo caso, i risultati del modello logistico.

Coefficients:	Estimate	Std.Error	z_value	Pr(> z)	
(Intercept)	6.81333	1.05116	6.482	9.07e-11	***
sibs_mio1	-0.08758	0.27170	-0.322	0.7472	
sibs_mio2	0.08064	0.27775	0.290	0.7716	
sibs_mio3	-0.01646	0.29478	-0.056	0.9555	
sibs_mio4	0.33714	0.28142	1.198	0.2309	
degree1	-2.27523	1.01586	-2.240	0.0251	*
degree2	-4.49004	1.00750	-4.457	8.33e-06	***
papres_ok_classe2	-0.11861	0.27044	-0.439	0.6610	
papres_ok_classe3	-0.43391	0.28116	-1.543	0.1228	
papres_ok_classe4	-0.60803	0.37190	-1.635	0.1021	

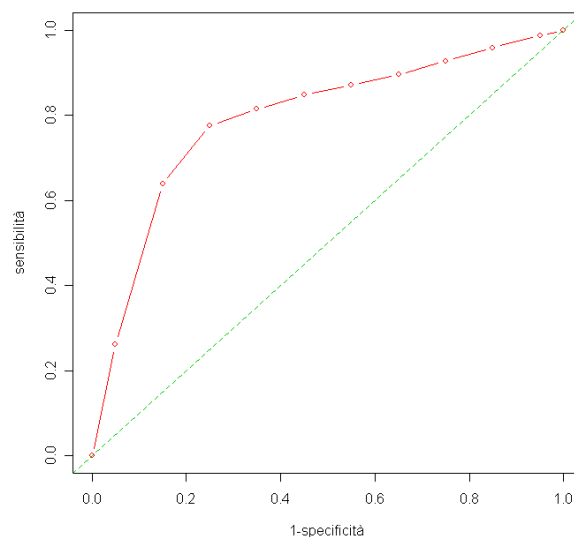
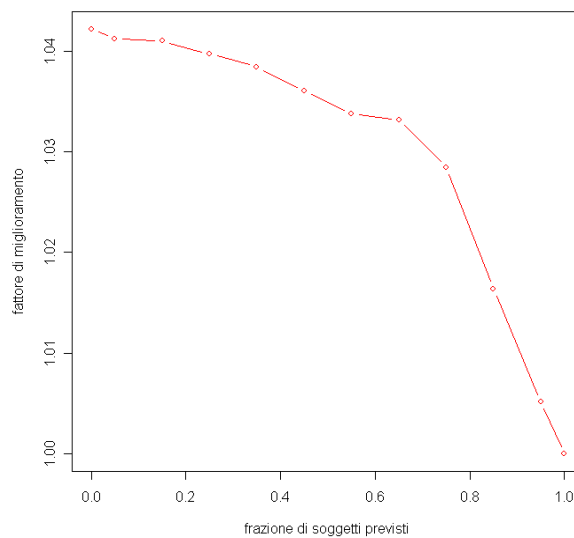
I risultati di questo nuovo modello sono un po' meno interessanti, infatti solo poche variabili risultano significative, vediamole singolarmente.

L'intercetta è significativa, ma ciò non fornisce informazioni interessanti per lo studio che sto conducendo. Sono inoltre significative le due categorie del *titolo di studio* più alte ('1' e '2'); le stime dei coefficienti di entrambe sono negative, cioè significa che raggiungere un titolo di studio medio o alto aumenta la probabilità di raggiungere la categoria di livello di prestigio '0', cioè quella corrispondente al più alto livello di prestigio ('4').

Non risultano, però, significativi, né il numero di fratelli del soggetto, né il livello di prestigio del padre, questo porta a dedurre che, per spiegare il raggiungimento del livello di prestigio più alto, è

sufficiente conoscere il titolo di studio del soggetto, ma ciò non toglie che, se non conoscessimo il titolo di studio, le altre variabili potrebbero diventare significative.

Ancora una volta le curve *Lift* e *ROC* non si differenziano dalle precedenti, quindi anche quest'ultimo modello è utile per predire la categoria di livello di prestigio del soggetto, conoscendo le tre variabili di interesse.



Infine si riporta una tabella riassuntiva dei tre modelli: sono riportate, per ogni categoria delle variabili risposta, le stime dei coefficienti e il suo esponenziale. La lettura degli asterischi (livelli di significatività) è la seguente:

- ◆ '***' indica la significatività al 100%;
- ◆ '**' indica la significatività al 99,9%;
- ◆ '*' indica la significatività al 99%;
- ◆ '.' indica la significatività al 95%;
- ◆ ' ' indica che la variabile non è significativa.

Tabella4.13: stime dei coefficienti, loro livello di significatività rappresentato dal numero degli asterischi ed esponenziale del coefficiente.

	Modello 1 0-livelli 2,3,4 1-livello 1		Modello 2 0-livelli 3,4 1-livelli 1,2		Modello 3 0-livello 4 1-livelli 1,2,3	
esplicativa	Stima coeff	Funzione coeff	Stima coeff	Funzione coeff	Stima coeff	Funzione coeff
Numero di fratelli						
<i>Sibs_mio0 (di riferimento)</i>	-	1	-	1	-	1
Sibs_mio1	-0.112099	0.8939558	0.07656	1.079567	-0.08758	0.9161456
Sibs_mio2	0.003838	1.003845	0.05031	1.051597	0.08064	1.083981
Sibs_mio3	0.123928	1.131934	0.16222	1.176119	-0.01646	0.9836747
Sibs_mio4	0.408607 *	1.504720	0.29675 *	1.345479	0.33714	1.400935
Titolo di studio						
<i>Degree0 (di riferimento)</i>	-	1	-	1	-	1
Degree1	-1.151535 ***	0.3161511	-1.18148 ***	0.3068243	-2.27523 *	0.1027733
Degree2	-2.676765 ***	0.06878532	-3.15716 ***	0.0425464	-4.49004 ***	0.01122019
Livello di prestigio del padre						
<i>Papres_ok_classe0 (di riferimento)</i>	-	1	-	1	-	1
Papres_ok_classe2	-0.453177 ***	0.6356056	-0.10078	0.904132	-0.11861	0.8881541
Papres_ok_classe3	-0.828987 ***	0.4364912	-0.53050 ***	0.5883107	-0.43391	0.6479706
Papres_ok_classe4	-0.745305 *	0.4745895	-0.21327	0.807938	-0.60803	0.5444223

La tabella riassuntiva fornisce alcune informazioni non deducibili dai risultati dei singoli modelli; quindi si considerano, una per volta, tutte le variabili esplicative. La variabile *numero di fratelli* è significativa, per la categoria '4', solo nei primi due modelli, ma non lo è nel terzo; questo perché, in quest'ultimo, la categoria di livello '0' è relativa al solo livello più alto di prestigio, ed ha senso presupporre che, se una famiglia è benestante, il numero dei figli è meno influente sul loro livello di prestigio, perché i genitori riusciranno comunque a garantire a tutti i loro figli un tenore di vita e un livello di istruzione alto. La forte relazione tra il numero dei fratelli e il livello di prestigio è invece molto forte in tutti gli altri casi.

Anche il *titolo di studio*, pur essendo significativo in tutte le sue categorie e in tutti i modelli proposti, risulta meno significativo nel terzo modello, il che significa che il fatto di raggiungere il livello più alto di prestigio è meno connesso col titolo di studio rispetto al fatto di raggiungere i livelli inferiori: ciò è sempre legato al fatto che i figli di benestanti riescono, spesso, a raggiungere i livelli di prestigio più alti anche senza aver raggiunto i titoli di studio superiori.

Vediamo, infine il *livello di prestigio del padre*. Questo influenza molto nel passaggio dai livelli di prestigio più bassi a quelli medi e medio alti, mentre non è sufficiente per il passaggio al livello superiore di prestigio (terzo modello).

Si sono provate strade differenti per verificare le ipotesi della seconda fase, e tutte hanno avuto esiti soddisfacenti, si procede quindi con l'analisi delle fasi successive.

4.3. TERZA FASE

“Nuovi immigrati prendono il posto dei bambini non nati, svolgendo i lavori meno pagati e meno prestigiosi, riempiendo così gli scalini più bassi della piramide sociale”.

Si procede quindi con un confronto diretto tra livello di istruzione e di prestigio di autoctoni e immigrati per vedere se ci sono differenze sostanziali tra queste due sottopopolazioni. I soggetti di interesse appartengono alla seconda generazione.

Innanzitutto osserviamo la variabile titolo di studio per autoctoni e immigrati e ci aspettiamo che gli immigrati siano meno istruiti degli autoctoni perché, generalmente, il paese di origine degli immigrati è meno sviluppato del paese in cui questi si trasferiscono.

Tabella4.14: *titolo di studio* in base alla *provenienza*. Osservazioni pesate. Classe di età all'intervista '40-59'. Generazione 2 (1930-1949).

	Titolo di studio			
	0	1	2	Totale
Autoctoni	18%	53%	29%	100% (6899.50)
Immigrati	25%	36%	39%	100% (537.69)

In Tabella4.14 sono riportate le percentuali di riga, cioè, data la provenienza, è riportata la distribuzione in percentuale del titolo di studio.

La differenza è molto forte, e ciò si nota in particolare per quanto riguarda il livello più basso di istruzione, raggiunto dal 18% degli autoctoni e il 25% degli immigrati.

Un'altra grande differenza è la moda²⁵: la percentuale più alta per gli autoctoni è quella relativa al livello '1' di istruzione (medio), mentre, inaspettatamente, la moda per gli immigrati il livello '2' (alto) di istruzione.

Si nota, inoltre, che gli immigrati sono molto più *polarizzati* rispetto agli autoctoni; infatti, alle tre modalità di titolo di studio corrispondono percentuali molto più alte, nelle classi estreme, rispetto agli autoctoni: infatti gli individui si distribuiscono nelle tre classi, senza concentrarsi, come accade per gli autoctoni, in una singola classe centrale. Quindi vi sono immigrati che, in realtà, sono molto diversi tra loro, e si possono suddividere nelle due seguenti modalità:

- ◆ immigrati *skilled*, cioè immigrati con un alto livello di istruzione che si spostano, ad esempio, per specializzarsi in qualche settore lavorativo, o per coprire ruoli di alto livello;
- ◆ immigrati *unskilled*, cioè immigrati che solitamente hanno livello di istruzione abbastanza basso, e provengono da paesi poveri e poco sviluppati, e si spostano con l'obiettivo di migliorare le condizioni di vita, e che riempiono il livello inferiore della piramide sociale.

Sono soprattutto questi ultimi i soggetti di interesse in questa fase di analisi. Quindi, mentre si introduce lo studio considerando tutti gli immigrati, si procederà in seguito, ad una focalizzazione sugli immigrati *unskilled*.

Si riporta, per confermare quanto detto sulla distribuzione differente del titolo di studio tra autoctoni e immigrati, un test Chi-Quadrato tra la variabile born (provenienza) e il titolo di studio.

²⁵ Per *moda* si intende la modalità di risposta con frequenza assoluta, o percentuale, maggiore.

Tabella4.15: test Chi-quadrato per il *titolo di studio* sulla variabile *born* (provenienza). Classe di età '40-59'. Generazione 2 (1930-1949).

Statistica	DF	Valore	Prob
Chi-quadrato	2	55.8863	<.0001

Il test porta a rifiutare l'ipotesi nulla di indipendenza tra titolo di studio e provenienza, ciò conferma l'ipotesi che i due sottogruppi di interesse sono molto diversi per quanto riguarda il titolo di studio.

Riportiamo ora un'analisi simile per il livello di prestigio, sempre per la stessa popolazione.

Tabella4.16: *livello di prestigio* in base al fatto di essere autoctoni o immigrati (*born*). Osservazioni pesate. Classe '40-59'. Generazione 2 (1930-1949).

	Livello di prestigio				
	1	2	3	4	Totale
Autoctoni	15%	50%	31%	4%	100% (6713.80)
Immigrati	19%	45%	30%	6%	100% (515.45)

La tabella è un po' diversa dalle aspettative, infatti era più sensato prevedere che anche il livello di prestigio fosse molto diverso tra immigrati e autoctoni, invece la differenza non è poi così forte.

Se si osservano, però, solo le percentuali del livello inferiore di prestigio (livello '1'), questo è raggiunto da una percentuale maggiore di immigrati rispetto agli autoctoni e ciò dimostra il fatto che gli immigrati coprono i lavori meno pagati più degli autoctoni. Ma, d'altro canto, la percentuale di immigrati che raggiunge il livello più alto di istruzione è maggiore della corrispondente percentuale degli autoctoni.

Ciò è in accordo con quanto visto per il titolo di studio: gli immigrati hanno, nelle classi estreme, percentuali più elevate, perché molto differenziati tra loro, mentre gli autoctoni tendono sempre a stabilirsi nelle classi intermedie. Si propone un test Chi-Quadrato tra provenienza e livello di prestigio dei soggetti.

Tabella4.17: test Chi-quadrato tra il *livello di prestigio* e la variabile *born* (autoctoni o immigrati). Classe di età '40-59'. Generazione 2 (1930-1949).

Statistica	DF	Valore	Prob
Chi-quadrato	3	18.5943	0.0003

Il test porta a rifiutare l'ipotesi nulla di indipendenza tra le due variabili (al 99%). Quindi, anche se non è completamente dimostrato che gli immigrati riempiano i gradini inferiori della piramide sociale, è comunque dimostrato che le due popolazioni sono molto diversificate tra loro, sia per titolo di studio che per livello di prestigio.

Si studiano, ora, in modo autonomo, gli immigrati, in particolare si vuole vedere se questi tendono a migliorare le loro condizioni di vita attraverso la *mobilità sociale*. Si costruisce, quindi, una tabella che mette in relazione il livello di prestigio con il livello di istruzione, della generazione 2, per gli immigrati (per gli autoctoni si veda la Tabella4.8).

L'obiettivo è quello di osservare se, per gli immigrati, la correlazione tra titolo di studio e prestigio è inferiore rispetto agli autoctoni.

Tabella4.18: *livello di prestigio per grado di istruzione*, per i soggetti che hanno valore della variabile *born* pari a '2', cioè nati fuori dagli USA. Osservazioni pesate. Classe '40-59'. Generazione 2 (1930-1949).

	Livello di prestigio					Totale
		1	2	3	4	
Livello di istruzione	0	41%	51%	8%	0%	100% (121.15)
	1	22%	55%	23%	0%	100% (188.20)
	2	4%	32%	49%	15%	100% (205.10)

In tabella sono riportate le percentuali di riga, quindi, per ogni livello di istruzione, è riportata la distribuzione, in percentuale, del livello di prestigio. Come per gli autoctoni, c'è una forte correlazione tra livello di istruzione e livello di prestigio.

Un'anomalia, (che però conferma i risultati precedenti), è che il 51% degli immigrati col titolo di studio inferiore (classe '0') raggiunge il livello di prestigio '2' (medio basso), mentre ci saremmo aspettati che questi soggetti non superassero il livello '1' (basso).

Si può ipotizzare quindi che l'attenzione e gli investimenti per l'ascesa sociale siano più forti negli immigrati che negli autoctoni e, cosa alquanto interessante, non solo nei confronti dei loro figli, ma anche per loro stessi.

Si riporta il test Chi-Quadrato tra il livello di prestigio e il titolo di studio.

Tabella4.19: test Chi-quadrato tra il *livello di prestigio* e il *titolo di studio*. Classe di età '40-59'. Generazione 2 (1930-1949). Immigrati.

Statistica	DF	Valore	Prob
Chi-quadrato	6	158.2420	<.0001

Anche in questo caso il test Chi-quadrato conferma la forte correlazione tra livello di istruzione e prestigio sociale. Proviamo a vedere un modello di particolare interesse. Si utilizzano sempre le unità appartenenti alla generazione 2,

ma si introduce nell'analisi la variabile *born*, che vale '1' se il soggetto è nato negli USA e '0' se è nato fuori dagli USA.

- ♦ *esplicativa categoriale che vale '0' se il prestigio vale '2', '3' o '4' (medio basso, medio alto e alto) e '1' se il prestigio vale '1' (basso)*

Eliminati i dati mancanti (si è già dimostrata la validità di questa scelta nella fase precedente), e selezionati i soggetti della seconda generazione, restano 9652 unità e su queste si procede con l'analisi.

Si riportano i primi risultati del modello logistico.

Coefficients:	Estimate	Std.Error	z_value	Pr(> z)	
(Intercept)	-0.21159	0.11160	-1.896	0.05797	.
born2	0.16850	0.17015	0.990	0.32204	
degree1	-1.19839	0.09369	-12.791	< 2e-16	***
degree2	-2.69906	0.16336	-16.522	< 2e-16	***
papres_ok_classe2	-0.30586	0.10773	-2.839	0.00452	**
papres_ok_classe3	-0.86088	0.16526	-5.209	1.89e-07	***
papres_ok_classe4	-0.42079	0.37481	-1.123	0.26157	

Quasi tutte le modalità delle variabili di questo modello sono significative.

I *livelli di istruzione* pari a '1' (medio) e '2' (alto) sono significativi, e le stime dei loro coefficienti sono negative, ciò significa che, avere titoli di studio non basso diminuisce la probabilità di raggiungere il livello di prestigio inferiore (rappresentato dalla categoria '1'). Infine il *livello di prestigio del padre* è significativo per le categorie medie '2' (medio basso) e '3' (medio alto), ed entrambi i coefficienti sono negativi, quindi avere il padre con livello di prestigio medio aumenta la probabilità che il soggetto raggiunga livello di prestigio rappresentato dalla categoria '0', cioè il livello più alto.

In realtà la variabile di maggior interesse del modello è *born*, ma questa non risulta significativa. Può essere comunque di interesse osservare la stima del coefficiente relativo alla categoria '2' della variabile *born* (soggetto immigrato); questa risulta positiva, quindi essere immigrato aumenta (anche se di poco) la probabilità di raggiungere la categoria di prestigio pari a '1' (basso). Quindi, anche se questo risultato non è molto preciso, la fase 3 è parzialmente verificata.

Si procede quindi con la verifica delle fasi successive.

4.4. QUARTA FASE

“Gli immigrati si assimilano molto velocemente ai comportamenti sociali dei residenti, abbassando la loro fecondità e investendo nel capitale umano per i loro figli, aumentando così la possibilità di crescita sociale di questi ultimi.”

L'obiettivo è, quindi, verificare se effettivamente gli immigrati diminuiscono la loro fecondità in modo simile agli autoctoni. Per fare ciò si decide di confrontare il numero di fratelli con il numero di figli per la generazione 2 (1930-1949). Innanzitutto osserviamo se si assiste, in generale, alla diminuzione di fecondità, non suddividendo, per questa prima analisi, gli immigrati dagli autoctoni.

Tabella4.20: *numero di figli sul numero di fratelli* per la generazione 2 (1930-1949). Classe di età all'intervista '40-59'. Osservazioni pesate. La seconda tabella è una sintesi della prima, si riportano solo le percentuali marginali.

Numero di fratelli	Numero di figli								
		0	1	2	3	4	5	6	Totale
	0	1.05%	1.05%	2.03%	1.44%	0.52%	0.50%	0.32%	6.75%
	1	2.69%	2.22%	6.15%	3.72%	1.83%	0.69%	0.58%	18.18%
	2	2.44%	2.28%	6.05%	3.57%	1.97%	0.74%	0.89%	17.85%
	3	1.93%	1.80%	4.30%	3.51%	1.92%	0.80%	0.76%	15.02%
	4	1.31%	1.04%	3.04%	2.29%	1.69%	0.74%	0.73%	10.86%
	5	0.90%	1.03%	2.39%	1.81%	0.95%	0.42%	0.72%	8.03%
	6	2.19%	2.61%	5.67%	4.82%	3.41%	2.03%	2.55%	23.28%
	Totale	12.49%	12.22%	29.64%	21.16%	12.40%	5.73%	6.34%	100%

Numero di figli (per la variabile <i>figli</i> si riporta il valore, per la variabile <i>fratelli</i> il valore sommato di 1)								
	0	1	2	3	4	5	6	7
Fratelli	-	6.75%	18.18%	17.85%	15.02%	10.86%	8.03%	23.28%
Figli	12.49%	12.22%	29.64%	21.16%	12.40%	5.73%	6.24%	-

Si nota che le percentuali marginali sono parecchio diverse. In particolare c'è una tendenza a diminuire il numero di figli, infatti raddoppia praticamente il numero di persone con un figlio, mentre non c'è modo di confrontare le persone che non hanno figli. Anche la percentuale di coloro che hanno due e tre figli aumenta moltissimo, quindi, di conseguenza, diminuisce quella di coloro che hanno più di quattro figli.

Si procede, ora, con un'osservazione di confronto tra immigrati ed autoctoni, confrontando la distribuzione del numero di fratelli e del numero di figli, aspettandosi che la prima sia molto diversa tra le due popolazioni di interesse (immigrati e autoctoni), perché dipende molto dalle abitudini del luogo di origine, mentre per verificare l'ipotesi occorre dimostrare che le distribuzioni del numero di figli tendono ad avvicinarsi.

Tabella4.21: percentuali marginali sul *numero di fratelli* e di *figli* sulla base della variabile *born* (immigrati o autoctoni). Generazione 2 (1930-1949). Classe di età all'intervista '40-59'.

Numero di fratelli (n = 7423.81)							
	0	1	2	3	4	5	6
Autoctoni	7%	19%	18%	15%	11%	8%	22%
Immigrati	6%	16%	15%	11%	11%	10%	30%
Differenza percentuale in modulo	1%	3%	3%	4%	0%	2%	8%
Numero di figli (n = 7434.44)							
	0	1	2	3	4	5	6
Autoctoni	13%	12%	30%	21%	12%	6%	6%
Immigrati	15%	13%	34%	18%	10%	4%	6%
Differenza percentuale in modulo	2%	1%	4%	3%	2%	2%	0%

Nella tabella sono riportate le percentuali di riga cioè, per ogni modalità della variabile *born* (autoctoni e

immigrati), si ha la distribuzione del numero di fratelli e del numero di figli.

Per ogni classe di fratelli e figli è riportata la differenza tra le percentuali degli immigrati e degli autoctoni, in valore assoluto.

La somma dei valori in percentuale delle differenze dei fratelli è 21, mentre per i figli è 14; anche se non si tratta di valori con particolare significato statistico, sono comunque utili per comprendere la differenza sostanziale. Ciò conferma la tendenza degli immigrati ad omologarsi alle abitudini sociali ed al livello di fecondità delle donne degli USA, anche se l'assimilazione non è completa.

Si riportano i due test Chi-Quadrato tra la variabile *born* e il numero di figli e di fratelli.

Tabella4.22: test Chi-quadrato tra la variabile *born* e il *numero di fratelli* (prima tabella) e tra la variabile *born* e il *numero di figli* (seconda tabella). Classe di età '40-59'. Generazione 2 (1930-1949).

Statistica	DF	Valore	Prob
Chi-quadrato	6	25.6154	0.0003

Statistica	DF	Valore	Prob
Chi-quadrato	8	11.5233	0.1738

Il test sui fratelli porta a rifiutare l'ipotesi di indipendenza tra provenienza e numero di fratelli, mentre il secondo test porta ad accettare l'ipotesi di indipendenza tra provenienza e numero di figli.

Questo risultato è molto importante perché rappresenta, statisticamente, l'omologarsi degli immigrati al livello di fecondità degli autoctoni, partendo invece da abitudini molto diverse.

A questo punto si concentra l'analisi, in modo più diretto, sugli immigrati *unskilled*, cioè quelli con titolo di studio pari a '0' (è una scelta di comodità, avremmo potuto scegliere come condizioni il livello di prestigio basso), per vedere se questi soggetti si omologano alla popolazione residente con lo stesso titolo di studio, diminuendo la loro fecondità.

Confronto, ad esempio, il numero di fratelli e di figli tra immigrati ed autoctoni con titolo di studio '0' (basso).

Tabella4.23: percentuali marginali sul *numero di fratelli* e di *figli* sulla base della variabile *born* (immigrati o autoctoni). Generazione 2 (1930-1949). *Titolo di studio* pari a '0'. Classe di età all'intervista '40-59'.

Numero di fratelli							
	0	1	2	3	4	5	6
Autoctoni	4%	7%	10%	11%	11%	9%	48%
Immigrati	4%	8%	11%	10%	12%	12%	43%
Differenza percentuale in modulo	-	1%	1%	1%	1%	3%	5%
Numero di figli							
	0	1	2	3	4	5	6
Autoctoni	9%	11%	19%	20%	15%	10%	16%
Immigrati	14%	9%	24%	20%	14%	6%	13%
Differenza percentuale in modulo	5%	2%	5%	-	1%	4%	3%

La tabella mostra una tendenza molto interessante, che non si poteva osservare senza considerare singolarmente le classi più basse delle società. Gli immigrati meno istruiti diminuiscono la loro fecondità, così come gli autoctoni con lo stesso livello di istruzione. Ma ciò che stupisce è che il loro tentativo di *mobilità sociale* per i propri figli sembra più forte che per gli autoctoni: una percentuale superiore di immigrati tende ad avere da nessuno a tre figli, e una percentuale molto più bassa di autoctoni che di immigrati raggiunge i cinque o sei figli.

Quindi anche la quarta fase è verificata.

Si procede con l'analisi della quinta e ultima fase, quella incentrata sui figli degli immigrati, cioè dei soggetti appartenenti alla generazione 3 (1950-1969).

4.5. QUINTA FASE

“Nella nuova generazione, i figli istruiti degli immigrati rifiutano i lavori sotto-pagati dei loro genitori, causando una nuova carenza di lavoratori non esperti. In questo modo il ciclo si ripete (si torna alla fase 3).”

L'obiettivo è quindi quello di verificare come si comportano i figli degli immigrati della generazione 2 (1930-1949), in particolare i figli dei cosiddetti immigrati *unskilled*.

Poiché non possediamo l'informazione sul livello di istruzione dei padri, scegliamo come caratteristica distintiva dei padri (generazione 2) dei soggetti della generazione 3, quelli con livello di prestigio basso e medio basso (pari a '1' e '2').

Quindi i soggetti di interesse sono, per la prima volta dall'inizio dell'analisi, la generazione 3, cioè i nati tra il 1950-1969, figli di genitori *unskilled*.

Innanzitutto osservo se questi individui tendono a “disprezzare” i lavori sotto-pagati dei loro genitori, raggiungendo un livello di istruzione e di prestigio più alto del loro, e confronto i risultati con quelli dei figli di autoctoni con padri *unskilled*.

Questa prima analisi serve per verificare se c'è *mobilità* a livello *generazionale*.

Quindi, innanzitutto, confronto, tra tutti coloro che hanno padre *unskilled*, i figli di autoctoni con i figli di immigrati, sia per il loro titolo di studio che per il loro prestigio.

Tabella4.24: *titolo di studio* dei figli di autoctoni e immigrati della generazione 3 (1950-1969). Osservazioni

pesate. Classe di età all'intervista '30-59'. Tengo i soggetti con *prestigio del padre* pari a '1' o '2'.

	Titolo di studio			
	0	1	2	Totale
Figli di autoctoni	10%	58%	32%	100% (5717.49)
Figli di immigrati	22%	43%	35%	100% (580.81)

In Tabella4.24 sono riportate le percentuali di riga, cioè, per ogni sottogruppo della generazione 3 di interesse (figli di autoctoni e figli di immigrati), è riportata la distribuzione del titolo di studio.

Si osserva che una percentuale del 12% maggiore di figli di immigrati, rispetto ai figli di autoctoni, non riesce a raggiungere titolo di studio superiore alla livello minimo, sorprendente un 35% dei figli di immigrati *unskilled* che riesce a raggiungere il livello più alto di istruzione, magari spinto dal desiderio di *riscatto sociale* dei loro genitori.

Quindi, prima di confrontare le stesse due sottopopolazioni per quanto riguarda il titolo di studio, si decide di confrontare il livello di prestigio degli immigrati *unskilled* (livello '1' e '2' di prestigio) della generazione 2 (1930-1949), con il livello di prestigio dei loro figli (generazione 3, 1950-1969), per vedere se c'è, per questi, una *mobilità sociale* (come è previsto dall'ipotesi).

Tabella4.25: *livello di prestigio* degli immigrati della generazione 2 (1930-1949) e dei loro figli della generazione 3 (1950-1969). Osservazioni pesate. Classe di età all'intervista '30-59'. Della generazione 2 tengo i soggetti con prestigio '1' e '2', e della generazione 3 quelli con padre con prestigio '1' e '2'.

	Livello di prestigio				
	1	2	3	4	Totale
Immigrati (generazione 2)	30%	70%	-	-	100% (420.58)
Figli di Immigrati (generazione 3)	16%	51%	28%	5%	100% (558.81)

La Tabella4.25 riporta le percentuali di riga quindi, per ogni generazione di interesse, si può osservare la distribuzione percentuale del livello di prestigio (si sono tenuti solo i padri di livello di prestigio '1' e '2'). Ciò permette di notare che i figli di immigrati *unskilled* tendono a crescere socialmente rispetto ai loro genitori: si nota infatti che un buon 33% dei figli di immigrati *unskilled* raggiunge un livello di prestigio medio alto o alto. Quest'informazione dimostra, in parte, la quinta fase, ma per verificare la *mobilità sociale* è più interessante confrontare, direttamente, il livello di prestigio dei genitori con quello dei loro figli. Si studia, quindi, la *mobilità sociale* di tipo *familiare*, costruendo una tabella che confronta il livello di prestigio dei genitori immigrati *unskilled* (generazione 2) con quello dei loro figli (generazione 3).

Tabella4.26: *livello di prestigio* dei figli di immigrati *unskilled* (livello di prestigio '1' e '2') della generazione 3 (1950-1969), in base al *livello di prestigio* del padre (generazione 2, 1930-1949). Osservazioni pesate. Classe di età all'intervista '30-59'.

		Prestigio dei figli				
		1	2	3	4	Totale
Prestigio dei padri	1	24%	54%	19%	3%	100% (91.96)
	2	15%	50%	30%	5%	100% (456.84)

Nella tabella sono riportate le percentuali di riga, al variare del livello di prestigio del padre (considerando solo i livelli di prestigio pari a '1' e '2').

La tabella mostra la forte relazione tra il livello di prestigio dei padri e quello dei figli. Infatti coloro che hanno padre con livello di prestigio '2' raggiungono, in genere, livello di prestigio più alto di coloro che hanno padre con livello di prestigio '1', cioè quello più basso.

E' interessante notare che il 76% dei figli di padre con livello di prestigio '1' raggiunge un livello di prestigio superiore, e un buon 22% raggiunge, addirittura, livello medio alto o alto. Mentre, tra i figli di genitore con livello di prestigio '2', il 35% riesce a raggiungere livelli di prestigio alti.

Si assiste quindi al fenomeno detto *mobilità sociale* per i figli di immigrati *unskilled*.

Si riporta, per completezza, il test Chi-Quadrato tra il livello di prestigio dei padri e quello dei figli.

Tabella4.27: test Chi-Quadrato tra il *livello di prestigio* dei figli di immigrati (generazione 3, 1950-1969) e il *livello di prestigio dei padri* immigrati (generazione 2, 1930-1949). Osservazioni pesate. Classe di età all'intervista '30-59'.

Statistica	DF	Valore	Prob
Chi-quadrato	9	49.6767	<.0001

Come prevedibile, il test risulta significativo contro l'ipotesi nulla di indipendenza tra le due variabili.

Infine si confrontano i livelli di prestigio delle due sottopopolazioni della generazione 3 (figli di autoctoni e figli di immigrati) per vedere se c'è una tendenza all'omologazione, causando una carenza di lavoratori *unskilled*.

Tabella4.28: *livello di prestigio* dei figli di autoctoni e immigrati della generazione 3 (1950-1969). Osservazioni pesate. Classe di età all'intervista '30-59'. Tengo i soggetti con padre con prestigio '1' e '2'.

	Livello di prestigio				
	1	2	3	4	Totale
Figli di autoctoni	15%	53%	29%	3%	100% (5581.31)
Figli di immigrati	16%	51%	28%	5%	100% (558.81)

Come ipotizzato, c'è una forte omologazione dei figli di immigrati ai figli di autoctoni, infatti le due distribuzioni sono praticamente uguali, e sembra non esserci correlazione tra titolo di studio e provenienza del genitore. Quindi ci sono pochi lavoratori *unskilled* di generazione 3, questo porterà a nuovi flussi di immigrati, sempre di generazione 3, che svolgeranno i lavori meno pagati.

Si riportano quindi delle conclusioni riassuntive e globali sulle cinque fasi per gli Stati Uniti.

4.6. CONCLUSIONI STATI UNITI

Scorrendo velocemente le cinque fasi, si nota che queste sono verificate quasi completamente, ma c'è un'anomalia di particolare interesse.

Infatti gli immigrati, negli USA, durante il periodo 1950-1969, risultano molto più istruiti e benestanti del previsto.

Per spiegare ciò abbiamo introdotto la distinzione fra immigrati *skilled* e *unskilled*; con il primo termine intendiamo quelli benestanti, istruiti, che si spostano fondamentalmente per cambiare lavoro, migliorando le proprie condizioni di vita, mentre con il secondo termine intendiamo gli immigrati che partono da zero, cioè coloro che seguono il cosiddetto *sogno americano*, e si spostano con la speranza di trovare un lavoro per sé e provvedere, di conseguenza, prima alla sopravvivenza, e poi al futuro dei loro figli.

Le diverse fasi, pur essendo verificate, non sono così evidenti come ci si aspettava.

Questo può essere spiegato dal fatto che non si tratta, in realtà, di un periodo di forte immigrazione, difatti la percentuale di immigrati sul totale si aggira intorno all'8%.

Si sono quindi cercati alcuni dati per osservare il fattore immigrazione da una prospettiva più generale.

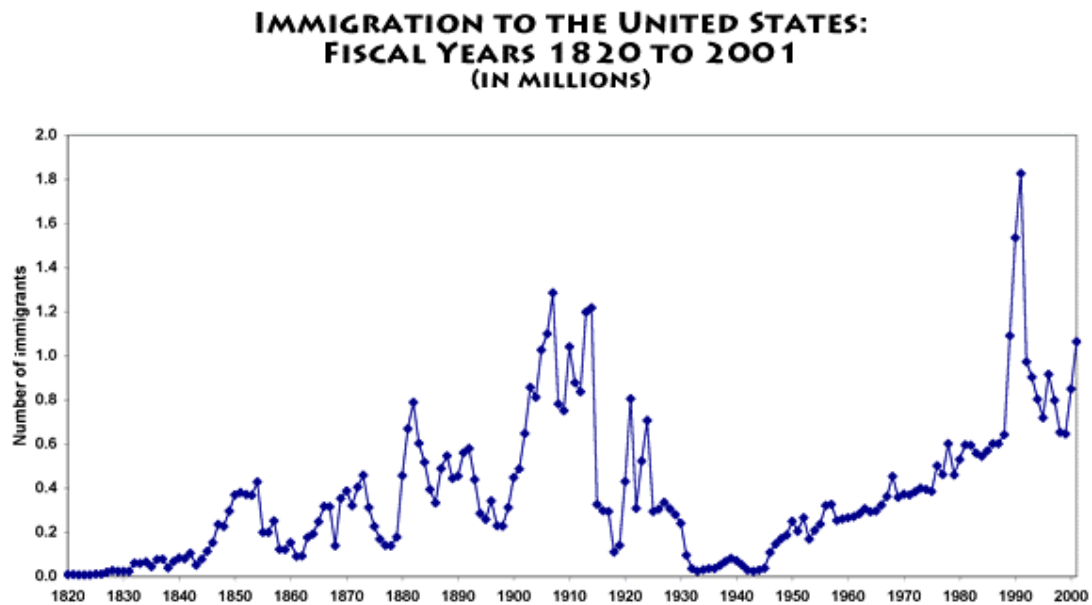
Un primo risultato ci viene fornito dal Grafico4.29.

La generazione di immigrati di interesse per questo studio è quella nata tra il 1930 e il 1949 (generazione 2). Supponendo un'età media, al momento dell'immigrazione, intorno ai 20 anni, gli anni di interesse sono tra il 1950 e in 1969.

Osservando questi anni nel grafico alla pagina successiva, si osserva che si tratta di un periodo in cui, in media,

sono entrati negli USA circa 300 mila immigrati l'anno, un numero assai superiore rispetto al ventennio precedente, ma molto inferiore rispetto a quello successivo.

Grafico4.29: Numero di immigrati l'anno negli Stati Uniti dal 1820 al 2000.



"Source: US Department of Justice Immigration and Naturalization Service, Statistical Yearbook of the Immigration and Naturalization Service 2001, US Government Printing Office, Washington, DC, forthcoming"

5. I RISULTATI SUL NORD OVEST D'ITALIA

Si riportano, di seguito, i risultati per il Nord-Ovest d'Italia, cercando di mantenere, per quanto possibile, un'analogia con l'analisi fatta per gli Stati Uniti.

Necessariamente ci saranno delle piccole divergenze dovute ad una disponibilità di variabili differente.

Ad esempio, per quando riguarda la variabile sul *prestigio*, questa era già calcolata nel dataset utilizzato per gli Stati Uniti, in base alla posizione professionale del soggetto (e del padre del soggetto) e al suo reddito; è stata invece calcolata, solo sulla base della posizione professionale, (il reddito non è richiesto nella Multiscopo) per il Nord-Ovest d'Italia.

Forse la differenza principale riguarda, però, la generazione 3 (nati tra il 1950 e il 1969), in particolare nella loro distinzione tra i figli di immigrati e i figli di autoctoni:

- ◆ per gli Stati Uniti è presente la variabile che indica la provenienza dei genitori, perciò è stato semplice creare la distinzione, per la generazione 3, tra figli di autoctoni e di immigrati, utilizzando solo questa variabile. In questo modo si considerano figli di immigrati coloro che hanno genitore immigrato, indipendentemente dal fatto che i soggetti siano nati all'estero (e quindi loro stessi immigrati) o negli Stati Uniti;
- ◆ per il Nord-Ovest d'Italia non è presente la variabile sulla provenienza dei genitori, quindi si ricorre ad una soluzione più complessa, ma che permette di risalire ad una buona percentuale dei soggetti di interesse. Si prendono, come figli (di generazione 3, 1950-1969) di immigrati (della generazione 2, 1930-1949), coloro che sono nati fuori dal Nord-Ovest e

residenti nel Nord-ovest, con genitori che vivono a meno di 50 chilometri da loro. Prendo, invece, come figli di autoctoni del Nord-Ovest i nati e residenti nel Nord-Ovest con genitori che vivono a meno di 50 chilometri. In questo modo il limite è quello di inserire tra i figli di autoctoni anche i figli di immigrati che hanno avuto il figlio direttamente nel Nord-Ovest.

Andiamo quindi a studiare non tutti i figli degli immigrati, ma solo quelli nati nelle aree di emigrazione, giunti poi nel Nord-Ovest assieme ai loro genitori. Per converso, questo porta a considerare come autoctoni gli immigrati di seconda generazione (ossia i figli di immigrati nati nell'area di arrivo). Nello studio della *mobilità sociale*, questa scelta dovrebbe condurre a risultati sbilanciati verso esiti più negativi, poiché è stato dimostrato come, fra i figli di immigrati nel Nord-Ovest, i peggiori esiti scolastici sono stati raggiunti dai ragazzi nati nelle regioni di provenienza, mentre chi è nato in Piemonte, Liguria o Lombardia ha conseguito titoli di studio simili a quelli degli autoctoni (Roberto Impicciatore, Gianpiero Dalla Zuanna, 2006, *"Una difficile mobilità sociale. L'istruzione dei figli dei meridionali emigrati verso il Centro e Nord Italia"*).

La conoscenza della direzione della distorsione ci permette di controllare la distorsione stessa nel corso dell'analisi.

5.1. PRIMA FASE

“In un periodo di sviluppo sostenuto, le coppie che limitano la fecondità possono garantire ai loro figli, che sono già nati, un più elevato capitale umano e maggiori opportunità per la crescita sociale.”

In questa fase devo verificare che, al diminuire del numero di figli della generazione 1 (1910-1929), si assiste ad un aumento del livello di istruzione dei soggetti della generazione 2 (1930-1949).

In questa prima fase mi interessano solo i soggetti autoctoni: studio quindi la correlazione tra il numero di fratelli e il livello di istruzione per i soggetti nati e residenti nel Nord-Ovest d'Italia²⁶.

Tabella5.1: *numero di fratelli e livello di istruzione della generazione 2 (1930-1949) nati nel Nord-Ovest d'Italia, secondo il livello di prestigio del padre. Osservazioni pesate. Classe di età all'intervista '40-59'. Percentuali di riga.*

²⁶ E' necessaria, per alcuni livelli di prestigio del padre, una modifica rispetto allo studio sugli Stati Uniti, infatti i soggetti della generazione 2 per in Nord-Ovest d'Italia sono in numero nettamente inferiore rispetto a quelli disponibili nella GSS statunitense, perciò è necessario creare delle classi meno specifiche per il numero di fratelli. Sempre per lo stesso problema della numerosità metto insieme, per alcune analisi, i livelli di prestigio del padre '1' e '2' e '3' e '4'. Queste modifiche verranno utilizzate in entrambe le tabelle di questa fase, e anche in alcune fasi successive.

	Livello di istruzione			Totale
Numero di fratelli	0	1	2	
(A-B) Padri di livello di prestigio '1' e '2' (n = 335.30)				
0-1	21%	44%	35%	100% (171.79)
2+	44%	35%	21%	100% (163.51)
(C-D) Padri di livello di prestigio '3' e '4' (n = 171.18)				
0-1	20%	33%	47%	100% (84.36)
2+	29%	43%	28%	100% (86.82)

La tabella mostra che, effettivamente, all'aumentare del numero di fratelli tende a diminuire il grado di istruzione.

Dalle percentuali di riga sembra che la soglia scelta per quanto riguarda il numero di fratelli (due fratelli) sia particolarmente interessante: si nota, infatti, un notevole cambiamento, proprio in quella soglia, nelle percentuali di riga, che indicano la distribuzione, dato il numero di fratelli e il livello di prestigio del padre, del titolo di studio dei soggetti.

Quindi è dimostrato che le coppie della generazione 1 (1910-1929), diminuendo il loro numero di figli, riescono a garantire, a questi pochi figli, un buon livello di istruzione.

Inoltre è di interesse notare che, all'aumentare del livello di prestigio del padre, c'è una tendenza per i soggetti a raggiungere un titolo di studio superiore, per entrambe le classi del numero di fratelli; questo significa che, se il padre è benestante, riesce a garantire una buona educazione anche ad un alto numero di figli.

Per dimostrare, in modo più rigoroso, queste conclusioni, si riporta un test Chi-Quadrato tra livello di istruzione e numero di fratelli.

Tabella5.2: P-value dei test Chi-Quadrato²⁷ tra *numero di fratelli* e *livello di istruzione* della generazione 2 (1930-1949) nati nel Nord-Ovest d'Italia, secondo il livello di prestigio del padre.

Prestigio del padre	p-value
1-2	<.0001
3-4	0.0308

E' dimostrato, quindi, che c'è una forte dipendenza tra il numero di fratelli e il livello di istruzione. L'ipotesi di indipendenza è rifiutata senza dubbio per coloro che hanno padre di livello di prestigio basso o medio basso, mentre è

²⁷ Il *test chi quadrato di Pearson* (o della bontà dell'adattamento) è un test non parametrico, applicato a grandi campioni, quando si è in presenza di variabili nominali, e si vuole verificare se il campione è stato estratto da una popolazione con una predeterminata distribuzione o che due o più campioni derivino dalla stessa popolazione. Fa parte di un'ampia classe di test detti test chi quadrato, in quanto hanno in comune le formule e la variabile casuale Chi quadrato, ma non necessariamente anche le ipotesi di base o le finalità. Questa variante del test utilizza i dati di un solo campione e verifica l'ipotesi nulla che il campione sia stato estratto da una popolazione di cui è nota la distribuzione. Sia

$$X^2 = \sum_{i=1}^g \frac{(n_i - E_i)^2}{E_i} = \sum_{i=1}^g \frac{n_i^2}{E_i} - n$$

n_i è il numero di casi osservati nel campione per la i -esima modalità, E_i è il numero di casi *attesi* nel caso l'ipotesi nulla fosse vera, g è il numero di modalità nella quale si esprime la variabile nominale

$$n = \sum_{i=1}^g n_i = \sum_{i=1}^g E_i$$

è la numerosità del campione. Allora X^2 è distribuita approssimativamente come una variabile casuale Chi Quadrato con $(g-1)$ gradi di libertà $X^2_{(g-1)}$. Si richiede però che tutte le frequenze attese E_i raggiungano un valore minimo (a seconda delle esigenze, almeno pari a 5 oppure almeno pari a 10). Qualora ci siano delle frequenze attese troppo piccole, bisogna procedere ad un raggruppamento di modalità.

un po' in dubbio la relazione per coloro che hanno padre benestante (di livello di prestigio alto o medio alto), per cui si rifiuta l'ipotesi di indipendenza al 95%, ma la si accetterebbe al 99%.

Il secondo obiettivo di questa prima fase è verificare se l'aumento del numero di fratelli porta ad una diminuzione del livello di prestigio della generazione 2 (1930-1949), dopo aver dimostrato che porta chiaramente ad una diminuzione del suo livello di istruzione.

Tabella5.3: numero di fratelli e livello di prestigio della generazione 2 (1930-1949), nati nel Nord-Ovest d'Italia, secondo il livello di prestigio del padre. Osservazioni pesate. Classe di età all'intervista '40-59'. Percentuali di riga.

	Livello di prestigio				totale
Numero di fratelli	1- basso	2- medio basso	3- medio alto	4-alto	
(A-B) Padri di livello di prestigio 1 e 2 (n = 124.50)					
0-1	18%	33%	30%	19%	100% (68.53)
2+	37%	11%	37%	15%	100% (55.97)
(C-D) Padri di livello di prestigio 3 e 4 (n = 83.63)					
0-1	8%	28%	40%	24%	100% (40.45)
2+	23%	20%	47%	10%	100% (43.17)

Anche in Tabella5.3 sono state riportate le percentuali di riga: queste mostrano, per i diversi livelli di prestigio del padre, e fissata una classe indicante il numero di fratelli, la distribuzione percentuale del livello di prestigio raggiunto.

E' chiaro il salto di qualità, per quanto riguarda il livello di prestigio, nel passare da meno di due fratelli a due fratelli o più.

In particolare, si nota che, indipendentemente dal livello di prestigio del padre, una percentuale maggiore dei soggetti con due o più fratelli si ferma al livello inferiore di prestigio, rispetto a coloro che ne hanno meno

di due; si osserva, inoltre, che una percentuale maggiore di soggetti con pochi fratelli riesce a raggiungere il livello più alto di prestigio.

Per i figli di padre di livello di prestigio medio alto e alto questa distinzione sembra, però, meno evidente: ciò avviene perché i genitori benestanti riescono, più facilmente, a garantire buon livello di prestigio ai loro figli, anche se molti.

Quindi l'ipotesi che, all'aumentare del livello di prestigio del padre, tenda ad aumentare il livello di prestigio del figlio, è verificata.

Si riporta un test Chi-Quadrato per verificare la dipendenza o indipendenza tra le due variabili in questione.

Tabella5.4: P-value dei test Chi-Quadrato tra *numero di fratelli* e *livello di prestigio* della generazione 2 (1930-1949) nati nel Nord-Ovest d'Italia, secondo il livello di prestigio del padre.

Prestigio del padre	P-value
1-2	0.0021
3-4	0.1140

La tabella mostra che la relazione tra numero di fratelli e livello di prestigio è più debole della relazione tra numero di fratelli e livello di istruzione. In particolare, per i figli di padri *unskilled*, si rifiuta l'ipotesi di indipendenza, mentre si accetta per i figli di padri benestanti, questo perché genitori di alto livello sociale riescono a garantire buon livello di prestigio ai loro figli, anche se molti.

Quindi la prima fase è dimostrata per il Nord-Ovest d'Italia, anche se risulta esser più forte la relazione tra

il numero di fratelli e il titolo di studio che tra il numero di fratelli e il livello di prestigio.

5.2. SECONDA FASE

“Figli istruiti rifiutano i lavori meno pagati che esigono basso capitale umano. Di conseguenza, c’è una carenza di persone adatte a quei tipi di lavori.”

Si cerca, quindi, in questa fase, di dimostrare che i pochi figli della generazione 1 (soggetti della generazione 2, 1930-1949) tendono a superare il livello di prestigio dei loro genitori.

Quindi si riporta un confronto generazionale per la variabile del livello di prestigio.

Tabella5.5: *livello di prestigio* dei figli (generazione 2, 1930-1949) sulla base del livello di prestigio dei loro genitori. Classe di età all’intervista ‘40+’²⁸. In riga c’è il prestigio dei padri e in colonna quello dei figli. Osservazioni pesate.

	Prestigio dei figli					Totale
		1	2	3	4	
Prestigio dei padri	1	29%	17%	41%	13%	100% (126.56)
	2	12%	25%	18%	45%	100% (35.47)
	3	18%	12%	52%	18%	100% (100.62)
	4	2%	22%	34%	42%	100% (32.86)

In Tabella5.5 sono riportate le percentuali di riga, cioè, per ogni livello di prestigio, è riportata la distribuzione percentuale del livello di prestigio dei figli (generazione 2).

Si sono evidenziate le percentuali di riga dei soggetti che hanno superato il livello di prestigio dei loro genitori, e queste percentuali sono molto elevate, quindi interessanti:

²⁸ Mentre per gli Stati Uniti era sufficiente utilizzare la classe ‘40-59’, per l’Italia risulta necessario allargare la classe, a causa della bassa numerosità dei soggetti nelle sottopopolazioni di interesse. Questa soluzione verrà utilizzata anche in alcune analisi successive.

un notevole 71% dei soggetti il cui padre ha livello di prestigio basso riesce a raggiungere livelli di prestigio superiori al primo, e addirittura il 54% raggiunge livelli medio alti e alti; quindi si assiste ad una notevole *mobilità sociale* fra le generazioni.

Degno di attenzione è anche il 45% dei soggetti con padre di livello medio basso ('2'), che riesce a raggiungere il livello più alto di prestigio ('4'), scavalcando notevolmente le condizioni sociali del padre.

Quindi si è dimostrata la forte tendenza all'*ascesa sociale* per i nati e residenti nel Nord-Ovest d'Italia, nel passaggio dalla generazione 1 (1910-1929) alla generazione 2 (1930-1949): ciò porta, necessariamente, ad una carenza di personale adatto ai lavori più umili.

Si riporta un test statistico per verificare o meno la dipendenza tra le due variabili.

Tabella5.6: test Chi-Quadrato per il *livello di prestigio* dei figli (generazione 2, 1930-1949) sulla base del livello di prestigio dei loro genitori. Classe di età all'intervista di età all'intervista '40+'.

Statistica	DF	Valore	Prob
Chi-quadrato	9	43.7069	<.0001

Il P-value del test risulta significativo contro l'ipotesi nulla di indipendenza tra livello di prestigio del padre e livello di prestigio del figlio.

Si osserva, ora, il livello di istruzione dei nati e residenti nel Nord-Ovest d'Italia, della generazione 2, per verificare se è vero che i figli degli autoctoni della generazione 1, raggiungono, all'aumentare del livello di istruzione, livello di prestigio più alto.

Tabella5.7: *livello di prestigio* sul *livello di istruzione* dei figli (generazione 2, 1930-1949), con genitori nati nel

Nord-Ovest d'Italia. Classe di età '40+'. Osservazioni pesate.

	Livello di prestigio					Totale
		1	2	3	4	
Livello di istruzione	0	44%	5%	51%	-	100% (80.88)
	1	25%	14%	59%	3%	100% (106.93)
	2	8%	28%	20%	44%	100% (145.41)

Anche in questa tabella sono riportate le percentuali di riga, cioè la distribuzione del livello di prestigio dei soggetti, al variare del loro livello di istruzione.

Si nota che, in genere, all'aumentare del livello di istruzione, tende ad aumentare il livello di prestigio. A conferma di ciò, nessun soggetto con titolo di studio basso ('0') riesce a raggiungere il più alto livello di prestigio, mentre riesce a raggiungerlo il 44% dei soggetti con titolo di studio alto.

Per verificare questa dipendenza si riporta un test statistico.

Tabella5.8: test Chi-quadrato per il *livello di prestigio* sul *livello di istruzione* dei figli (generazione 2) con genitori autoctoni. Classe di età all'intervista '40+'.

Statistica	DF	Valore	Prob
Chi-quadrato	6	147.2670	<.0001

Il P-value relativo al test porta a rifiutare l'ipotesi nulla di indipendenza tra livello di istruzione e livello di prestigio per la generazione 2 (1930-1949).

Quindi anche la seconda fase sembra verificata, ma si decide di costruire alcuni modelli, uguali a quelli costruiti per gli Stati Uniti, per poter fare un confronto diretto sui risultati, al capitolo successivo.

Si sceglie, come variabile risposta, il *livello di prestigio* del soggetto, e come variabili esplicative le seguenti variabili:

- *numero di fratelli*,
- *livello di prestigio del padre*,
- *titolo di studio del soggetto*.

Oggetto di quest'analisi è la generazione 2 (1930-1949).

La fase fondamentale dell'analisi esplorativa è la comprensione, e lo studio, della variabile risposta, cioè quella per cui vengono richieste delle previsioni.

Si considerano tutte le variabili come fattoriali, tranne la risposta, che rendo categoriale.

Le unità che contengono un dato mancante, in almeno una delle variabili di interesse, sono 6024, decido di eliminarle, perché non c'è motivo per supporre che i dati mancanti siano fonte di distorsione.

Per poter fare ciò, prima verifichiamo la validità di questa ipotesi, riportando le distribuzioni per alcuni incroci di variabili, osservando anche i dati mancanti.

Le variabili che verranno utilizzate, ora e di seguito, per i modelli sono le tre prima elencate, e, in aggiunta, la variabile *prestigio* e la variabile *provenienza* (che indica se il soggetto è nato o meno nel Nord-Ovest d'Italia).

Si mostrano alcuni incroci per vedere se coloro che non hanno risposto ad una domanda hanno risposto in modo polarizzato ad altre domande, o sono invece nella media.

Si riportano le percentuali di riga:

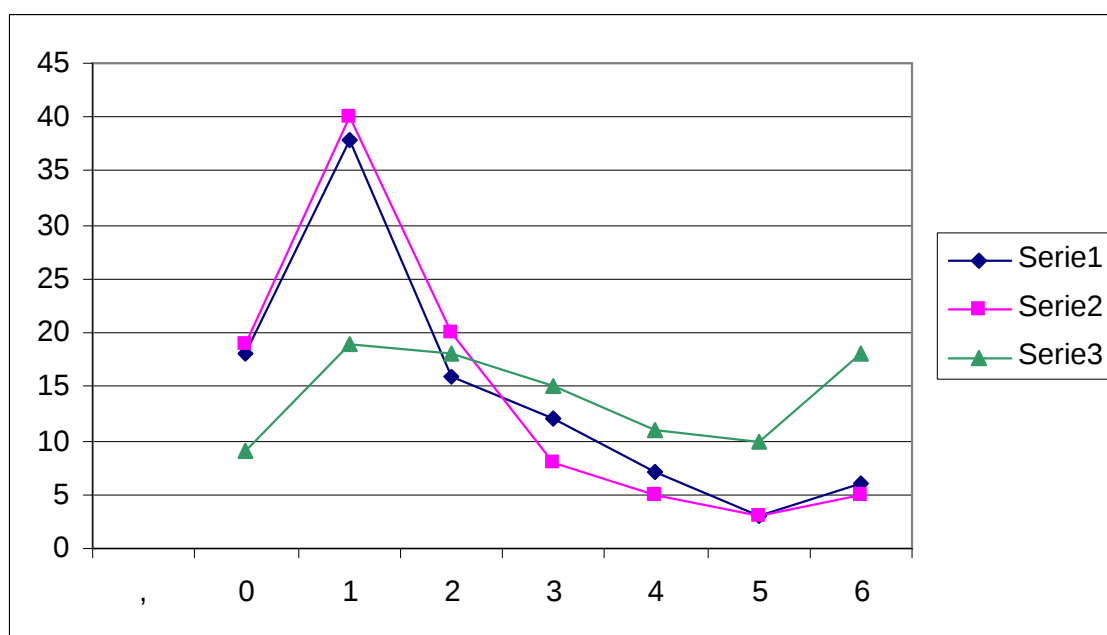
♦ *Provenienza incrociata col numero di fratelli:*

Tabella5.9: percentuali di riga per la variabile *numero di fratelli* sulla base della variabile *provenienza*, considerando le modalità dei valori mancanti.

	Numero di fratelli								
Born	.	0	1	2	3	4	5	6	Totale di riga
.	-	18%	38%	16%	12%	7%	3%	6%	100% (298.61)
Autoctono	-	19%	40%	20%	8%	5%	3%	5%	100% (7567.40)
Immigrato	-	9%	19%	18%	15%	11%	10%	18%	100% (2191.00)

Si riporta il grafico delle percentuali di riga, che è più leggibile della tabella.

Grafico5.10: percentuali di riga del *numero di fratelli* per le tre modalità della variabile *provenienza*: dato mancante (serie 1), autoctono (serie 2), immigrato (serie 3).



Osservando il grafico, si nota che la distribuzione di coloro che non hanno risposto alla domanda sulla loro provenienza (serie 1), non si differenzia per quanto riguarda il numero di fratelli, infatti la curva dei dati mancanti è praticamente sempre intermedia, e non mostra nessuna anomalia degna di nota. Ovviamente la distribuzione

è più vicina a quella degli autoctoni che a quella degli immigrati, perché sono differenti le percentuali assolute. Vediamo un'altra distribuzione:

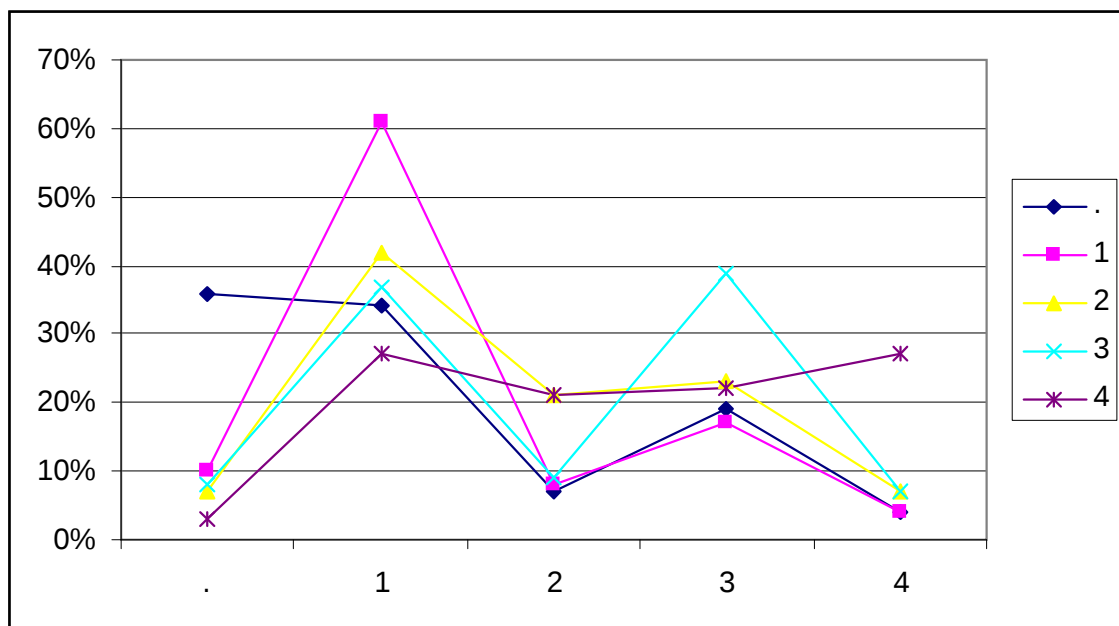
♦ *Prestigio_padre incrociato con prestigio*

Tabella5.11: percentuali di riga per le variabili: *prestigio del padre* del soggetto e *prestigio del soggetto* considerando le modalità dei valori mancanti.

		Prestigio padre					
		.	1	2	3	4	Totale
Prestigio	.	36%	34%	7%	19%	4%	100%
	1	10%	61%	8%	17%	4%	100%
	2	7%	42%	21%	23%	7%	100%
	3	8%	37%	9%	39%	7%	100%
	4	3%	27%	21%	22%	27%	100%

Si riporta il grafico delle percentuali di riga, che è più leggibile della tabella.

Grafico5.12: percentuali di riga del prestigio sul prestigio del padre.



Per la serie dei valori mancanti la distribuzione è un po' alterata, poiché è semplice credere che, chi non ha risposto alla domanda sul proprio prestigio, non abbia risposto neanche alla domanda sul prestigio dl padre.

Per tutte le altre categorie del prestigio del soggetto non sembra esserci alcuna anomalia per la serie dei dati mancanti rispetto alle altre, anzi questa si trova, praticamente sempre, in una posizione intermedia.

Provando anche altri incroci sembra poco rischioso eliminare i dati mancanti. Si procede quindi all'eliminazione.

Quindi restano 4033 persone. Di queste 489 sono della generazione 2 (1930-1949).

Faccio l'analisi su queste, tenendo conto che la numerosità è abbastanza bassa, ma dovrebbe riuscire ugualmente a garantire risultati attendibili.

Per ogni analisi decido di dividere il dataset in due parti, di stima e verifica²⁹.

²⁹ Le tecniche per la selezione del modello sono le seguenti:

-*Insieme di stima e di prova*: si dividono i dati in due gruppi, la prima parte viene usata per la stima dei diversi modelli possibili, la seconda per la verifica della validità.

-*Convalida incrociata*: consiste nell'incrociare i ruoli delle K porzioni di dati, di cui (k-1) gruppi vengono usati per la stima e uno per la verifica, e poi si scambiano i ruoli, quindi si otterranno K stime di cui si stimerà la curva media. K è al massimo uguale ad n. Per n grande è il metodo migliore.

-*Criteri basati sull'informazione*: si tiene conto del numero di parametri nella stima di massima verosimiglianza, grazie ad una penalizzazione del tipo $IC = -2\log L(\hat{\theta}) + \text{penalità}(p)$. L'approccio è detto AIC, ce ne sono altri con assunzioni diverse.

-*Selezione automatica delle variabili*: procedimento automatizzato per la scelta dei modelli che scelgono il gruppo di variabili che minimizza l'AIC (selezione passo-a-passo).

Considero due casi di interesse³⁰:

1) *variabile categoriale del livello di prestigio: '0' indica i livelli '2', '3' o '4' (medio basso, medio alto, alto) e '1' indica livello '1' (basso).*

Si decide di stimare un modello logistico³¹ che ha come variabile risposta la dicotomica e come esplicative tutte le variabili, rese fattoriali.

Coefficients:	Estimate	Std.Error	z.value	Pr(> z)
(Intercept)	0.1520	0.4731	0.321	0.7480
fratelli1	-0.5616	0.5114	-1.098	0.2721
fratelli2	0.4236	0.5481	0.773	0.4396
fratelli3	0.3249	0.4657	0.698	0.4853
istruzione1	-0.6120	0.3524	-1.737	0.0825 .
istruzione2	-1.8474	0.4483	-4.121	3.78e-05 ***
prestigio_padre2	-0.2660	0.5968	-0.446	0.6557
prestigio_padre3	-0.7379	0.3559	-2.073	0.0381 *
prestigio_padre4	-0.9682	0.6943	-1.395	0.1631

Dai risultati di questo primo modello si possono riportare alcune conclusioni interessanti.

Considerando, innanzitutto, le variabili che risultano significative, si osserva che il *livello di istruzione* del

³⁰ Mentre per gli Stati Uniti sono stati costruiti tre modelli, per l'Italia si decide di costruirne solo due, in quanto non è possibile costruire quello che confronta i primi tre livelli di prestigio con livello '4', perché troppo pochi soggetti raggiungono quel livello di prestigio, e un modello darebbe risultati distorti.

³¹ Per *modello logistico* si intende un modello binomiale con funzione legame logit. In generale se la variabile risposta è dicotomica non è conveniente utilizzare un modello lineare, ma può essere utile usare la regressione logistica

$$l(x) = \frac{\exp(x)}{1 + \exp(x)},$$

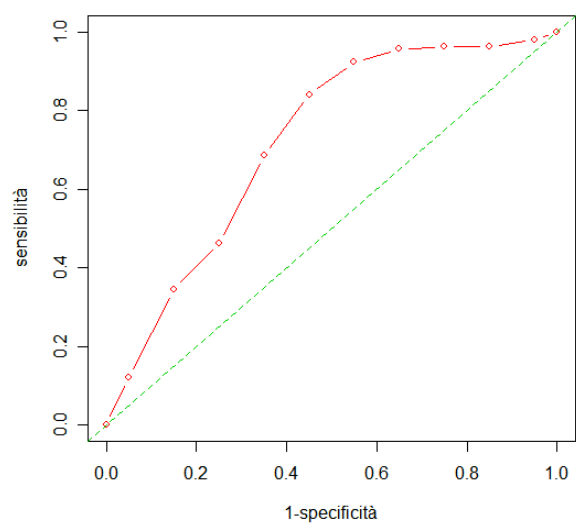
che è uno dei casi dei GLM (Modelli Lineari Generalizzati). La parametrizzazione che si usa è detta d'angolo.

soggetto è significativo sia nella categoria '1' (livello medio) che nella categoria '2' (livello alto). Le corrispondenti stime dei coefficienti sono negative, ciò significa che avere titolo di studio medio o alto aumenta la probabilità di passare dal livello di prestigio più basso (corrispondente alla categoria '1') a tutti i livelli superiori (corrispondenti alla categoria '0'). Quindi è verificato, anche attraverso questo modello, che, all'aumentare del livello di istruzione, diminuisce la probabilità di raggiungere i livelli inferiori di prestigio.

Risulta, inoltre, significativa, la categoria del *livello di prestigio del padre* pari a '3' (livello medio alto), e tutte le categorie presenti nel risultato hanno coefficiente negativo, quindi avere padre con livello di prestigio diverso da quello più basso (pari a '1') fa diminuire la probabilità, per il soggetto, di raggiungere il livello inferiore di prestigio. E' dimostrato, quindi, che all'aumentare del livello di prestigio del padre aumenta la probabilità, per il soggetto, di superare il livello '1' (basso) di prestigio.

Non sembrano, invece, significative le categorie indicanti il *numero di fratelli*; osservando ugualmente i coefficienti corrispondenti, si prova che il passaggio da meno di due fratelli a due fratelli o più, è di interesse, in quanto cambia il segno della stima del coefficiente, cioè all'aumentare del numero di fratelli aumenta la probabilità di raggiungere il livello inferiore di prestigio.

Per valutare la validità del modello si riporta la curva ROC. Questa è abbastanza discostata dalla bisettrice, quindi conoscere le tre variabili esplicative, permette di migliorare notevolmente la previsione della variabile risposta.



2) *variabile categoriale del livello di prestigio: '0' indica i livelli '3' e '4' (medio alto e alto) e '1' indica livelli '1' e '2' (basso e medio basso)*

Si riportano i primi risultati del modello logistico.

Coefficients:	Estimate	Std.Error	z.value	Pr(> z)
(Intercept)	0.3209	0.4200	0.764	0.4449
fratelli1	-0.5985	0.4075	-1.469	0.1419
fratelli2	-0.4026	0.4611	-0.873	0.3826
fratelli3	0.0490	0.3961	0.124	0.9016
istruzione1	0.1217	0.3496	0.348	0.7277
istruzione2	-0.4553	0.3496	-1.303	0.1927
prestigio_padre2	-0.1949	0.4767	-0.409	0.6827
prestigio_padre3	-0.6504	0.3168	-2.053	0.0401 *
prestigio_padre4	-0.3759	0.5406	-0.695	0.4869

A differenza del modello precedente, in questo caso solo una variabile risulta significativa, e cioè la categoria '3' del *livello di prestigio del padre*.

Ad ogni modo osserviamo le stime dei coefficienti, per farci un'idea più accurata sui risultati. Le categorie '2', '3' e '4' del livello di prestigio del padre hanno stime dei coefficienti negative, quindi avere padre con livello diverso da quello più basso (pari a '1') aumenta la probabilità del soggetto di raggiungere livello di prestigio medio alto o alto.

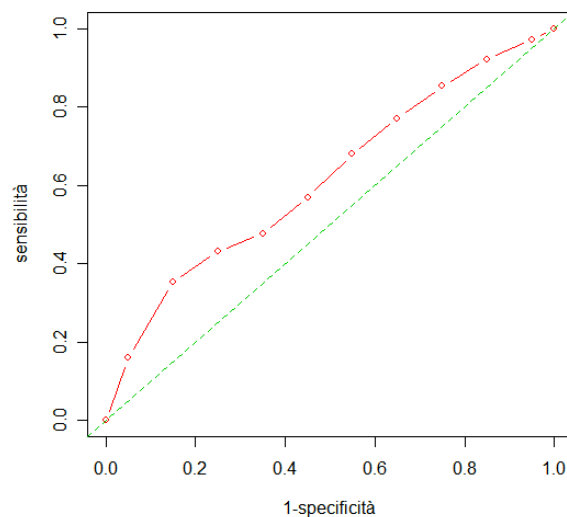
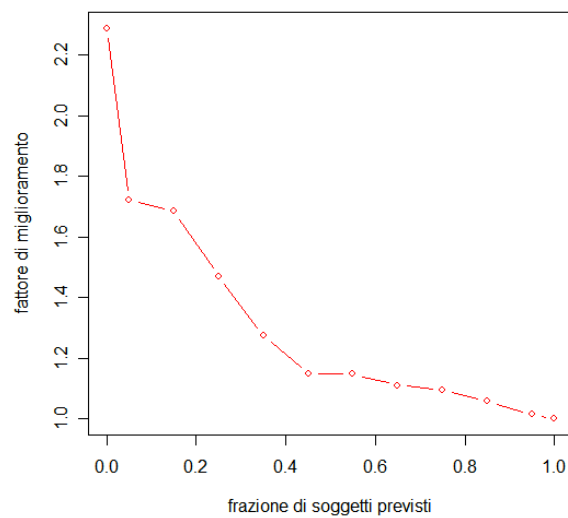
Per quanto riguarda, invece, il *numero di fratelli*, si nota che passare da meno di tre fratelli a tre fratelli o più aumenta la probabilità di raggiungere livelli di prestigio bassi o medio bassi.

Infine, passare da *livello di istruzione* medio al livello di istruzione alto, aumenta la probabilità di raggiungere livelli di prestigio alti o medio alti per il soggetto.

Quindi, anche se i risultati di questo modello non sono evidenti come i risultati del modello precedente, è

comunque dimostrato che: all' aumentare del numero di fratelli tende a diminuire il livello di prestigio; una buona istruzione garantisce livelli di prestigio alti e all'aumentare del livello di prestigio del padre tende ad aumentare il livello di prestigio del figlio.

Si riportano, nell'ordine, la curva *Lift* e la curva *ROC* per confermare la validità del modello.



Infine si riporta una tabella riassuntiva dei tre modelli: sono riportate, per ogni categoria delle variabili risposta, le stime dei coefficienti e il suo esponenziale. La lettura dei simboli (livelli di significatività) è la seguente:

- ◆ '***' indica la significatività al 100%;
- ◆ '**' indica la significatività al 99.9%;
- ◆ '*' indica la significatività al 99%;
- ◆ '.' indica la significatività al 95%;
- ◆ ' ' indica che la variabile non è significativa.

Tabella5.13: stime dei coefficienti, loro livello di significatività rappresentato dal numero degli asterischi ed esponenziale del coefficiente.

	Modello 1 0-livelli 2,3,4 1-livello 1		Modello 2 0-livelli 3,4 1-livelli 1,2	
esplicativa	Stima coeff	Funzione coeff	Stima coeff	Funzione coeff
<i>Numero di fratelli</i>				
<i>Fratelli0 (di riferimento)</i>	-	1	-	1
Fratelli1	-0.5616	0.5702959	-0.5985	0.5496355
Fratelli2	0.4236	1.527450	-0.4026	0.6685795
Fratelli3	0.3249	1.383892	0.0490	1.050220
<i>Titolo di studio</i>				
<i>Istruzione0 (di riferimento)</i>	-	1	-	1
Istruzione1	-0.6120 .	0.5422653	0.1217	1.129415
Istruzione2	-1.8474 ***	0.1576465	-0.4553	0.6342577
<i>Livello di prestigio del padre</i>				
<i>Prestigio_padre1 (di riferimento)</i>	-	1	-	1
Prestigio_padre2	-0.2660	0.7664391	-0.1949	0.822917
Prestigio_padre3	-0.7379	0.4781169	-0.6504 *	0.521837
Prestigio_padre4	-0.9682	0.379766	-0.3759	0.686671

Questa tabella riassuntiva permette di fare dei confronti interessanti tra i modelli; quindi si considerano le diverse esplicative e si traggono le conclusioni.

Il *numero di fratelli* non risulta di particolare interesse in questi due modelli.

Dal primo modello si deduce che, passare da avere un fratello ad averne due o più, diminuisce la probabilità di riuscire ad abbandonare il livello inferiore di prestigio; mentre dal secondo si deduce che, passare da avere meno di due fratelli ad averne tre o più, aumenta la probabilità di raggiungere livelli di prestigio bassi, o medio bassi.

Confrontando, invece, i risultati sul *livello di istruzione*, si osserva che una sola categoria di questa esplicativa risulta significativa, e solo nel primo modello; quindi, un buon livello di istruzione, è molto utile per riuscire ad abbandonare il livello inferiore di prestigio.

Osservando tutte le stime dei coefficienti, si nota che avere titolo di studio medio, o alto, aumenta la probabilità di riuscire ad abbandonare il livello inferiore di prestigio, e allo stesso tempo aumenta la probabilità di raggiungere il livello massimo di prestigio.

Infine, una modalità del *livello di prestigio del padre* risulta significativa nel primo modello. Tutte le categorie di prestigio del padre presenti (quindi tutte tranne quella inferiore, '0') hanno coefficienti negativi, in tutti e due i modelli. Da ciò si deduce che, se il padre non ha livello di prestigio basso, cioè '0', è probabile che anche il figlio riesca a raggiungere livelli di prestigio superiori al primo.

Quindi tutte le ipotesi sono state, almeno in parte, verificate. Si procede con le analisi relative alle fasi successive.

5.3. TERZA FASE

“Nuovi immigrati prendono il posto dei bambini non nati, svolgendo i lavori meno pagati e meno prestigiosi, riempiendo così gli scalini più bassi della piramide sociale”.

In questa fase si decide di confrontare, all'interno della generazione 2 (1930-1949), gli immigrati e gli autoctoni, per vedere se è vero che i primi raggiungono livelli di istruzione inferiori e svolgono lavori meno prestigiosi dei secondi.

Quindi osserviamo, prima di tutto, la distribuzione del titolo di studio per le due sotto-popolazioni della generazione 2.

Tabella5.14: *titolo di studio* in base alla *provenienza*. Osservazioni pesate. Classe di età all'intervista '40-59'. Generazione 2 (1930-1949).

	Titolo di studio			
	0	1	2	Totale
Autoctoni	30%	40%	30%	100% (560.83)
Immigrati	44%	36%	20%	100% (273.86)

In Tabella5.14 sono riportate le percentuali di riga, cioè, data la provenienza, è riportata la distribuzione, in percentuale, del titolo di studio.

C'è una grandissima differenza tra la distribuzione del titolo di studio degli autoctoni e quella degli immigrati. Senza dubbio gli autoctoni sono molto più istruiti degli immigrati, infatti una percentuale maggiore di immigrati raggiunge il livello inferiore di istruzione, e una percentuale maggiore di autoctoni riesce a raggiungere il livello superiore.

La moda³², per gli autoctoni, è il livello medio di istruzione, mentre per gli immigrati è il livello inferiore, perciò la tesi sembra verificata.

E' da notare, però, che una certa percentuale di immigrati entra nel Nord-Ovest d'Italia con un livello di istruzione alto, quindi è necessaria, come per gli Stati Uniti, una distinzione tra gli immigrati *skilled* (cioè quelli che immigrano per assumere magari ruoli di alto livello nelle aziende o per specializzarsi in alcuni settori) e quelli *unskilled* (di livello di istruzione basso e basso capitale umano), che sarà particolarmente importante, soprattutto nelle prossime fasi.

Di particolare interesse, nelle ipotesi delle ultime fasi, sono gli immigrati *unskilled* ed i loro figli della generazione 3 (1950-1969).

Si riporta, per dimostrare la correlazione tra la provenienza e il titolo di studio, un test Chi-Quadrato.

Tabella5.15: test Chi-Quadrato per il *titolo di studio* sulla variabile *provenienza* (autoctoni o immigrati). Classe di età '40-59'. Generazione 2 (1930-1949).

Statistica	DF	Valore	Prob
Chi-quadrato	2	17.4530	0.0002

Il P-value del test porta a rifiutare l'ipotesi nulla di indipendenza tra titolo di studio e provenienza, ciò conferma l'ipotesi che i due sottogruppi sono statisticamente differenti per la variabile titolo di studio.

Procediamo con un'analisi parallela per il livello di prestigio dei soggetti della generazione 2.

³² Per *moda* si intende la modalità di risposta con frequenza assoluta, o percentuale, maggiore.

Tabella5.16: *livello di prestigio* in base alla variabile *provenienza*. Osservazioni pesate. Classe '40-59'. Generazione 2 (1930-1949).

	Livello di prestigio				
	1	2	3	4	Totale
Autoctoni	24%	24%	37%	15%	100% (233.24)
Immigrati	32%	24%	30%	14%	100% (114.42)

La tabella mostra che gli immigrati hanno prestigio inferiore rispetto agli autoctoni, ma ci si sarebbe aspettati una differenza più lampante. E' alta la differenza tra le due percentuali di riga per quanto riguarda il livello inferiore di prestigio: questo significa che, per gli immigrati, è più alta la probabilità di trovarsi a svolgere i lavori più umili, rispetto agli autoctoni.

E' evidente, anche da questa tabella, la necessità di distinguere le due classi di immigrati della generazione 2 (1930-1949), cioè gli immigrati *skilled* e gli *unskilled*.

Si propone un test Chi-Quadrato tra la variabile provenienza e livello di prestigio dei soggetti della generazione 2.

Tabella5.17: test Chi-Quadrato tra il *livello di prestigio* e la variabile *provenienza* (autoctoni o immigrati). Classe di età '40-59'. Generazione 2 (1930-1949).

Statistica	DF	Valore	Prob
Chi-quadrato	3	3.1782	0.3650

Le differenze, riportate in Tabella5.16, non sono sufficienti a rendere il test Chi-Quadrato, tra provenienza e livello di prestigio, significativo.

Concentrandoci, ora, solo sugli immigrati della generazione 2, osserviamo se questi sono o meno soggetti al fenomeno della *mobilità sociale*.

Innanzitutto si costruisce la tabella che incrocia il livello di istruzione con il livello di prestigio, per studiare questa relazione per gli immigrati (per gli autoctoni si veda la Tabella5.8).

Tabella5.18: *livello di prestigio per grado di istruzione*, per i soggetti con *provenienza* pari a '2', cioè nati fuori dal Nord-Ovest d'Italia. Osservazioni pesate. Classe '40-59'. Generazione 2 (1930-1949).

	Livello di prestigio					Totale
		1	2	3	4	
Livello di istruzione	0	50%	15%	35%	-	100% (34.18)
	1	32%	30%	36%	2%	100% (41.27)
	2	13%	28%	15%	44%	100% (34.04)

In tabella sono riportate le percentuali di riga, quindi, per ogni livello di istruzione, è riportata la distribuzione percentuale del livello di prestigio.

C'è, come ci si poteva aspettare, una forte correlazione tra il livello di istruzione e il livello di prestigio per gli immigrati, come si era mostrato per gli autoctoni.

Si riporta il test Chi-Quadrato tra il livello di prestigio e il titolo di studio, per confermare statisticamente quest'ipotesi.

Tabella5.19: test Chi-Quadrato tra il *livello di prestigio* e il *titolo di studio*. Classe di età '40-59'. Generazione 2 (1930-1949). Immigrati.

Statistica	DF	Valore	Prob
Chi-quadrato	6	45.7191	<.0001

Il test Chi-Quadrato conferma la forte correlazione tra livello di istruzione e prestigio sociale.

Si riportano, quindi, i risultati di un modello di interesse.

Si utilizzano sempre le unità appartenenti alla generazione 2, ma si introduce nell'analisi la variabile *provenienza*, che vale '1' se il soggetto è nato nel Nord-Ovest d'Italia e '2' se è nato altrove:

- ♦ *esplicativa categoriale che vale '0' se il prestigio vale '2', '3' o '4' (medio basso, medio alto e alto) e '1' se il prestigio vale '1' (basso).*

Eliminati i dati mancanti (si è già dimostrata la validità di questa scelta nella fase precedente), e selezionati i soggetti della generazione 2 (1930-1949), restano 482 unità e su queste si procede con l'analisi.

Si riportano i primi risultati del modello logistico.

Coefficients:	Estimate	Std.Error	z.value	Pr(> z)
(Intercept)	0.0568	0.3121	0.182	0.8556
provenienza2	0.2158	0.3254	0.663	0.5073
istruzione1	-0.7093	0.3550	-1.998	0.0457 *
istruzione2	-2.4404	0.4804	-5.080	3.78e-07 ***
prestigio_padre2	0.4687	0.5988	0.783	0.4338
prestigio_padre3	-0.4726	0.3538	-1.336	0.1816
prestigio_padre4	0.2008	0.6953	0.289	0.7727

In questo modello risulta significativa solo la variabile *livello di istruzione*, in particolare i due coefficienti risultano entrambi negativi, questo significa che avere un titolo di studio diverso da quello inferiore (classe '0') diminuisce la probabilità di restare al livello inferiore di prestigio.

Vediamo, comunque, come si comportano le altre variabili per avere un quadro complessivo.

I risultati sulla *provenienza*, pur non risultando questa significativa, sono di particolare interesse; la stima del coefficiente della categoria della provenienza relativa agli immigrati risulta positivo, ciò significa che essere immigrato aumenta la probabilità di raggiungere il livello '1' di prestigio, cioè quello più basso. Questo risultato dimostra proprio le ipotesi alla base della terza fase.

La variabile sul *prestigio del padre* non risulta significativa; quindi, in conclusione, a determinare il livello di prestigio del soggetto interviene soprattutto il *titolo di studio*.

Si procede, quindi, con la verifica delle fasi successive.

5.4. QUARTA FASE

“Gli immigrati si assimilano molto velocemente ai comportamenti sociali dei residenti, abbassando la loro fecondità e investendo nel capitale umano per i loro figli, aumentando così la possibilità di crescita sociale di questi ultimi.”

In questa fase, ci si propone di verificare che gli immigrati della generazione 2 (1930-1949) diminuiscono la loro fecondità, in modo simile agli autoctoni della stessa generazione.

Si confrontano, innanzitutto, il numero di fratelli e il numero di figli della generazione 2, senza distinguere gli autoctoni dagli immigrati, per vedere se si assiste, in generale, ad una diminuzione della fecondità.

Tabella5.20: *numero di figli sul numero di fratelli* per la generazione 2 (1930-1949). Classe di età all'intervista '40-59'. Osservazioni pesate. Percentuali. La seconda tabella è una sintesi della prima, si riportano solo le percentuali marginali.

		Numero figli				
		0	1	2	3+	Totale
Numero fratelli	0 1 figlio	2.93%	7.31%	4.78%	1.65%	16.67%
	1 2 figli	3.33%	9.00%	10.27%	1.34%	23.94%
	2 3 figli	2.85%	5.49%	7.97%	3.85%	20.16%
	3+ 4+ figli	4.72%	9.91%	16.76%	7.84%	39.23%
	Totale	13.82%	31.71%	39.78%	14.69%	100%

Numero di figli (se la variabile è 'figli' si riporta il valore, se è 'fratelli' il valore sommato di 1)					
	0	1	2	3	4
Fratelli	-	16.67%	23.94%	20.16%	39.23%
Figli	13.82%	31.71%	39.78%	14.69%	-

Le percentuali marginali sono molto diverse. E' chiara la tendenza a diminuire il numero di figli; infatti, mentre la maggior parte dei soggetti della generazione 2 ha quattro fratelli o più, la maggior parte di loro ha solo due figli. Quindi si procede con un confronto tra immigrati e autoctoni per quanto riguarda il numero di fratelli e il numero di figli, aspettandoci, per i primi, una grossa differenza, e, invece, una maggiore assimilazione per i secondi.

Questo perché il numero di fratelli è molto legato al luogo di origine, solitamente zona con fecondità maggiore rispetto a quella in cui ci si trasferisce, mentre il numero di figli dipende molto, secondo le ipotesi, dalle abitudini degli autoctoni.

Tabella5.21: percentuali marginali sul *numero di fratelli* e di *figli* sulla base della variabile *provenienza* (immigrati o autoctoni). Generazione 2 (1930-1949). Classe di età all'intervista '40-59'. Osservazioni pesate.

	Numero di fratelli (n = 834.39)			
	0	1	2	3
Autoctoni	20%	30%	21%	29%
Immigrati	9%	13%	19%	59%
Differenza percentuale in modulo	11%	17%	2%	30%
	Numero di figli (n = 834.69)			
	0	1	2	3
Autoctoni	15%	35%	38%	12%
Immigrati	10%	25%	44%	21%
Differenza percentuale in modulo	5%	10%	6%	9%

Nella tabella sono riportate le percentuali di riga, cioè per ogni modalità della variabile provenienza (autoctoni e immigrati) si ha la distribuzione del numero di fratelli, e del numero di figli.

Per le diverse classi di fratelli e di figli è riportata la differenza assoluta tra le percentuali degli autoctoni e

quelle degli immigrati, non si tratta di valori con valenza statistica, ma sono, comunque, utili, per comprendere l'andamento.

Si nota, chiaramente, che le distribuzioni dei figli, per autoctoni e immigrati, sono molto più simili delle distribuzioni dei fratelli: è verificato che c'è una tendenza, per gli immigrati, ad assimilarsi alle abitudini del luogo in cui si trasferiscono.

Si riportano i due test Chi-Quadrato tra la variabile provenienza e il numero di figli, e tra la prima e il numero di fratelli.

Tabella5.22: test Chi-Quadrato tra la variabile *provenienza* e il *numero di fratelli* (prima tabella) e tra la variabile *provenienza* e il *numero di figli* (seconda tabella). Classe di età '40-59'. Generazione 2 (1930-1949).

Statistica	DF	Valore	Prob
Chi-quadrato	3	80.5412	<.0001

Statistica	DF	Valore	Prob
Chi-quadrato	3	19.5596	0.0002

Entrambi i test portano a rifiutare l'ipotesi nulla di indipendenza, ma il primo è molto più forte del secondo. Quindi c'è una dipendenza fortissima tra provenienza e numero di fratelli, mentre la dipendenza è un po' meno forte tra la provenienza e il numero dei figli.

Anche se l'assimilazione non è completa, e i legami con le origini sono forti per quanto riguarda il numero di figli, tutti i risultati portano a dimostrare una progressiva assimilazione degli immigrati, della generazione 2, agli autoctoni della stessa generazione.

Quindi si decide di concentrarsi su quegli immigrati che abbiamo definito *unskilled*, al paragrafo precedente, scegliendo come variabile discriminante il loro titolo di

studio: quindi definiamo *unskilled* gli immigrati con titolo di studio basso (pari a '0').

Ovviamente possiamo confrontare questi soggetti solo con gli autoctoni con lo stesso titolo di studio, per mantenere l'analogia.

Decido, quindi, di confrontare il numero di fratelli e di figli, restringendomi ai soggetti in questione.

Tabella5.23: percentuali marginali sul *numero di fratelli* e di *figli* sulla base della variabile *provenienza* (immigrati o autoctoni). Generazione 2 (1930-1949). Titolo di studio '0'. Classe di età all'intervista '40-59'.

	Numero di fratelli (n=288.26)			
	0	1	2	3
Autoctoni	13%	20%	17%	50%
Immigrati	10%	6%	14%	70%
Differenza percentuale in modulo	3%	14%	3%	20%
	Numero di figli (n=288.26)			
	0	1	2	3
Autoctoni	13%	34%	31%	22%
Immigrati	10%	23%	35%	32%
Differenza percentuale in modulo	3%	11%	4%	10%

Sembra evidente l'assimilarsi degli immigrati *unskilled* agli autoctoni con lo stesso titolo di studio: le differenze tra le distribuzioni del numero di figli sono decisamente inferiori alle differenze tra quelle del numero di fratelli.

Quindi anche la quarta fase è verificata. Si procede con l'analisi della quinta fase, concentrando l'attenzione sui figli degli immigrati della generazione 2 (1930-1949), cioè dei soggetti appartenenti alla generazione 3 (1950-1969).

5.5. QUINTA FASE

“Nella nuova generazione, i figli istruiti degli immigrati rifiutano i lavori sotto-pagati dei loro genitori, causando una nuova carenza di lavoratori non esperti. In questo modo il ciclo si ripete (si torna alla fase 3).”

L'attenzione è quindi, in questa fase, rivolta ai figli (generazione 3) degli immigrati *unskilled* della generazione 2, per verificare se questi individui sono soggetti alla *mobilità sociale*. E' inoltre nostro obiettivo confrontare i figli di immigrati con i figli di autoctoni.

Come discriminante per definire i padri *unskilled* scegliamo coloro che hanno livello di prestigio basso o medio basso (pari a '1' e '2'), in analogia con quando operato per gli Stati Uniti.

Innanzitutto confronto la distribuzione del titolo di studio tra le due sottopopolazioni dei figli di padre *unskilled*.

Tabella5.24: *titolo di studio* dei figli di autoctoni e immigrati della generazione 3 (1950-1969). Osservazioni pesate. Classe di età all'intervista '30-59'. Tengo i soggetti con padre con prestigio '1' e '2'.

	Titolo di studio			
	0	1	2	Totale
Figli di autoctoni	6%	53%	41%	100% (1316.64)
Figli di immigrati	14%	57%	29%	100% (246.18)

In Tabella5.24 sono riportate le percentuali di riga: per ogni sottopopolazione della generazione 3, è riportata la distribuzione del titolo di studio.

Si osserva che il livello di istruzione dei figli di immigrati è decisamente inferiore rispetto a quello dei

figli di autoctoni. Questa grande differenza è conseguenza anche delle scelte effettuate che, come è stato sottolineato nell'introduzione all'analisi sul Nord-Ovest d'Italia, portano ad una sottostima del livello di istruzione dei figli di immigrati della generazione 3 (nati nel 1950-1969).

Prima di effettuare un controllo analogo per il livello di prestigio, si confronta il livello di prestigio dei padri, immigrati *unskilled*, di generazione 2, con il prestigio dei loro figli (generazione 3).

Tabella5.25: *livello di prestigio* degli immigrati della generazione 2 (1930-1949) e dei loro figli della generazione 3 (1950-1969). Osservazioni pesate. Classe di età all'intervista '30-59'. Della generazione 2 tengo i soggetti con prestigio '1' e '2', e della generazione 3 quelli con padre con prestigio '1' e '2'.

	Livello di prestigio				
	1	2	3	4	Totale
Immigrati (generazione 2)	57%	43%	-	-	100% (64.71)
Figli di immigrati (generazione 3)	43%	28%	23%	6%	100% (193.57)

La Tabella5.25 riporta le percentuali di riga, cioè per ogni generazione di interesse, si può osservare la distribuzione percentuale del livello di prestigio (si sono tenuti solo i soggetti con padri di livello '1' e '2' per la generazione 3 e i soggetti di prestigio '1' e '2' per la generazione 2).

Si osserva che il 29% dei figli di immigrati *unskilled* riesce a raggiungere livelli medio alti, o alti, di prestigio: si assiste quindi al fenomeno della *mobilità sociale generazionale*.

Se, invece, ci vogliamo concentrare sulla *mobilità a livello familiare*, costruiamo una tabella che confronti i livelli di prestigio dei figli di immigrati *unskilled* con il prestigio dei loro genitori (pari a '1' o '2').

Tabella5.26: *livello di prestigio* dei figli di immigrati *unskilled* (livello di prestigio '1' e '2') della generazione 3 (1950-1969), in base al livello di prestigio dei genitori (generazione 2, 1930-1949). Osservazioni pesate. Classe di età all'intervista '30-59'.

		Prestigio dei figli				
		1	2	3	4	Totale
Prestigio dei padri	1	48%	26%	21%	5%	100% (158.39)
	2	19%	40%	30%	11%	100% (35.18)

In Tabella5.26 sono riportate le percentuali di riga, al variare del livello di prestigio del padre (di cui si sono considerati solo i livelli '1' e '2').

La tabella mostra una relazione evidente tra livello di prestigio del padre e quello del figlio. Inoltre, ancora più interessante, si può osservare che il 52% dei figli con padre di prestigio '1', riesce a superare il livello del padre, e addirittura il 26% riesce a raggiungere prestigio medio alto o alto.

Tra i soggetti con padre di livello di prestigio medio basso, invece, il 41% raggiunge livello di prestigio superiore al padre.

Si assiste, quindi, ad una forte *mobilità sociale* dei figli degli immigrati *unskilled*, come ipotizzato.

Si riporta, inoltre, il test Chi-Quadrato tra il livello di prestigio dei padri e quello dei figli, per i soggetti in questione.

Tabella5.27: test Chi-Quadrato tra il *livello di prestigio* dei figli di immigrati (generazione 3, 1950-1969) e il

livello di prestigio dei genitori immigrati (generazione 2, 1930-1949). Osservazioni pesate. Classe di età all'intervista '30-59'.

Statistica	DF	Valore	Prob
Chi-quadrato	3	10.9187	0.0122

Il P-value è significativo al 95%, mentre non lo sarebbe al 90%. Decido di tenere come soglia il 95% perché si tratta di variabili demografiche. Concludo che c'è una forte dipendenza tra livello di prestigio del padre e quello del figlio.

Infine si confrontano i livelli di prestigio delle due sottopopolazioni della generazione 3 (figli di autoctoni e figli di immigrati) per vedere se c'è una tendenza all'omologazione, causando una carenza di lavoratori *unskilled*.

Tabella5.28: *livello di prestigio* dei figli di autoctoni e immigrati della generazione 3 (1950-1969). Osservazioni pesate. Classe di età all'intervista '30-59'. Tengo i soggetti con padre con prestigio '1' e '2'.

	Livello di prestigio				
	1	2	3	4	Totale
Figli di autoctoni	32%	41%	15%	12%	100% (1074.24)
Figli di immigrati	43%	28%	23%	6%	100% (193.57)

Come ipotizzato, c'è una certa omologazione dei figli di immigrati ai figli di autoctoni per quanto riguarda il livello di prestigio.

Le due distribuzioni, pur essendo diverse, mostrano che i figli di immigrati *unskilled* riescono, per il 29%, a superare il livello di prestigio del padre ('1' o '2'), e un 27% dei figli di autoctoni fa lo stesso, quindi c'è una *mobilità sociale* per entrambi gruppi di soggetti rispetto

ai loro genitori. Ciò porta probabilmente ad una carenza di personale lavorativo *unskilled*, come ipotizzato.

Tutte le cinque fasi sono state, quindi, verificate. Si traggono le conclusioni per il Nord-Ovest d'Italia, per procedere, nel capitolo successivo, con un confronto tra le due zone di interesse.

5.6. CONCLUSIONI NORD-OVEST D'ITALIA

Anche per quanto riguarda il Nord-Ovest d'Italia, le cinque fasi sembrano funzionare.

L'unica carenza dell'analisi è la numerosità dei dati un po' bassa, ma ciò non ha impedito di ottenere risultati importanti.

In particolare, avendo utilizzato la *Multiscopo* del 2003, è stato un po' difficile trovare i soggetti della generazione 1, cioè i nati tra il 1910 e il 1929; utilizzando, però, le variabili sui genitori, si è riusciti a ricostruire il quadro generale di tutte le generazioni di interesse.

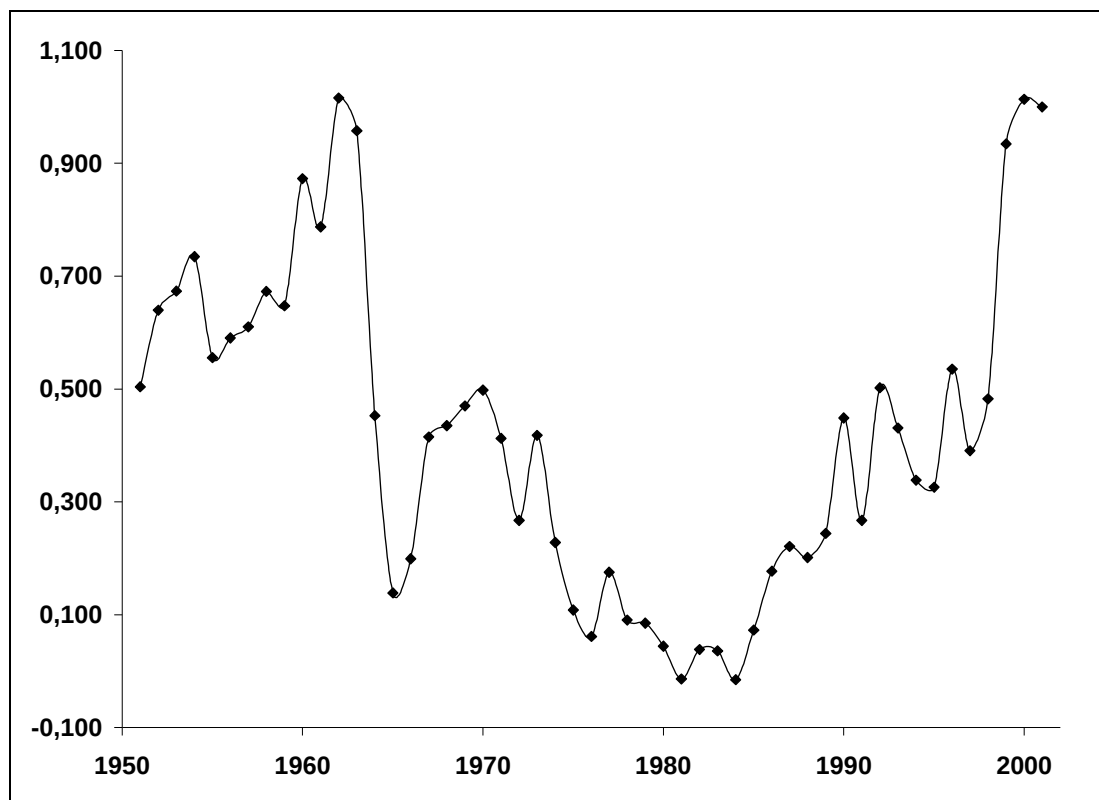
Gli immigrati di interesse in quest'elaborato sono quelli della generazione 2, cioè nati nel periodo 1930-1949. Supponendo che siano immigrati intorno ai 20 anni, gli anni di interesse sono 1950-1969.

Si tratta di un periodo di forte immigrazione nel Nord-Ovest d'Italia: immigrati provengono dal Centro e dal Nord-Est d'Italia.

Anche per l'Italia, come per gli Stati Uniti, è stato necessario suddividere gli immigrati nelle due categorie *skilled* e *unskilled*.

Vediamo un grafico che riporta l'andamento dell'immigrazione nella zona, e che mostra come, negli anni 1950-1969, la misura del rapporto tra saldo migratorio (ingressi a cui vengono sottratte le uscite) e il numero di nati, tocca i suoi massimi picchi, in particolare intorno al 1960.

Grafico5.29: rapporto tra saldo migratorio e nascite nel Nord-Ovest d'Italia dal 1951 al 2000.



Sources: Demographic Yearbooks of Italy.

Si procede, nel capitolo successivo, ad un confronto dei risultati per le due zone di interesse.

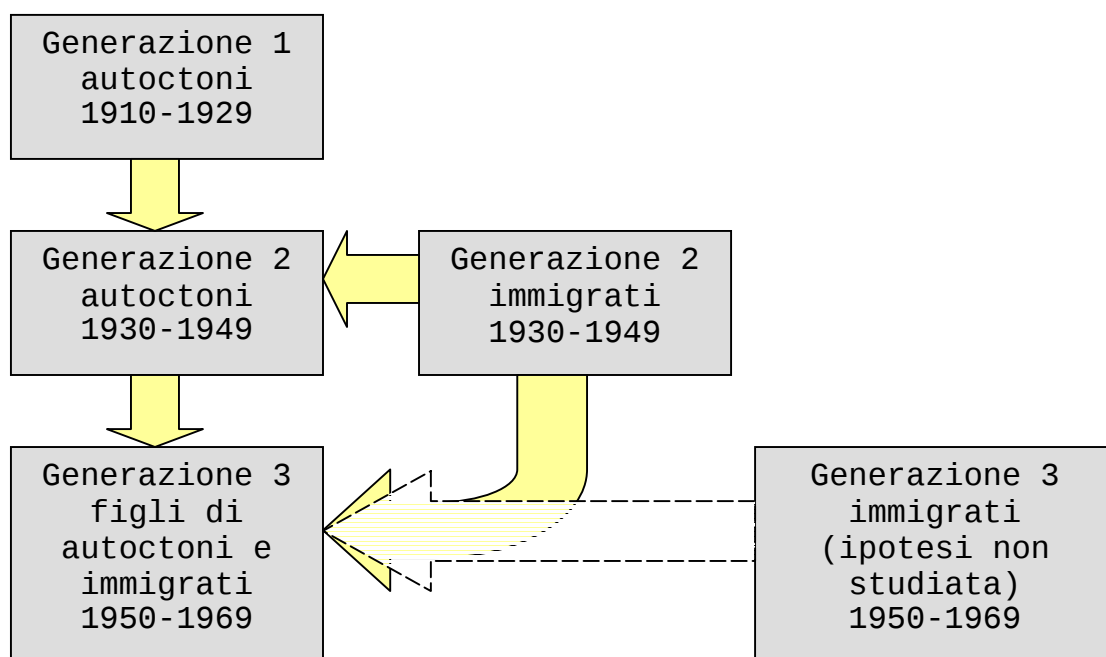
6. CONFRONTI E CONCLUSIONI

Il capitolo conclusivo è organizzato mantenendo lo stesso schema utilizzato durante lo svolgimento dell'elaborato, cioè distinguendo le cinque fasi da dimostrare, e confrontando i risultati ottenuti per gli Stati Uniti e quelli per il Nord-Ovest d'Italia.

Verranno riportate le tabelle più significative per il confronto. Si presterà particolare attenzione ai risultati statistici, come i test Chi-Quadrato, riportati per alcune coppie di variabili, e i modelli binomiali. Infine si concluderà con alcune considerazioni.

Ricordiamo quali generazioni sono state oggetto di studio tramite uno schema riassuntivo.

Schema6.1: Riassunto delle generazioni oggetto di studio.



6.1. PRIMA FASE

“In un periodo di sviluppo sostenuto, le coppie che limitano la fecondità possono garantire ai loro figli, che sono già nati, un più elevato capitale umano e maggiori opportunità per la crescita sociale.”

Quindi si studia la generazione 2 (1930-1949) di autoctoni, nelle due zone di interesse, e si vede se il numero di fratelli è correlato al titolo di studio e al livello di prestigio, per verificare se è vero che, al diminuire del numero di figli, la generazione 1 (1910-1929) riesce a offrire ai propri figli migliori possibilità di crescita sociale.

Tabella6.2: P-value dei test Chi-Quadrato tra *numero di fratelli* e *livello di istruzione* della generazione 2 (nata nel 1930-1949), autoctoni, secondo il *livello di prestigio del padre*.

Prestigio padre	Stati Uniti	Nord-Ovest d'Italia
	P-value	P-value
1	<.0001	<.0001
2	<.0001	
3	<.0001	0.0308
4	0.1963	

Tabella6.3: P-value dei test Chi-Quadrato tra *numero di fratelli* e *livello di prestigio* della generazione 2 (nata nel 1930-1949), autoctoni, secondo il *livello di prestigio del padre*.

Prestigio padre	Stati Uniti	Nord-Ovest d'Italia
	P-value	P-value
1	0.0095	0.0021
2	<.0001	
3	0.1819	0.114
4	0.3129	

Osservando le due tabelle precedenti si nota che le relazioni tra numero di fratelli e, alternativamente, il livello di istruzione e il livello di prestigio, sono molto simili. Infatti, per i soggetti con padre *unskilled*, si rifiuta l'ipotesi di indipendenza tra numero di fratelli e sia livello di istruzione che livello di prestigio, e ciò sia per gli Stati Uniti che per il Nord-Ovest d'Italia. Sempre per entrambe le zone, a parte per gli statunitensi con padre di livello di prestigio medio alto ('3'), il prestigio di figli di padre benestante sembra non essere influenzato così tanto dal numero di fratelli, infatti i genitori benestanti riescono a garantire un buon titolo di studio e un buon livello di prestigio ai loro figli, anche se questi sono in numero considerevole.

Quindi c'è una dipendenza forte tra numero di fratelli e crescita sociale per i figli di *unskilled*, mentre il legame è più debole per i figli di genitori benestanti.

6.2. SECONDA FASE

“Figli istruiti rifiutano i lavori meno pagati che esigono basso capitale umano. Di conseguenza, c’è una carenza di persone adatte a quei tipi di lavori.”

Quindi si studiano i figli della generazione 1 (1910-1929), cioè la generazione 2 (1930-1949), sempre dei soli autoctoni, per vedere se c’è una *mobilità sociale* dei soggetti rispetto ai loro genitori, con conseguente carenza di personale lavorativo *unskilled* di generazione 2.

Si confrontano i risultati tramite una tabella che mette in relazione il livello di prestigio del padre e quello del figlio.

Tabella 6.4: *livello di prestigio dei figli (generazione 2, 1930-1949) sulla base del livello di prestigio dei loro genitori*. Classe di età all’intervista ‘40-59’ per gli Stati Uniti e ‘40+’ per il Nord-Ovest d’Italia. In riga c’è il prestigio dei padri e in colonna quello dei figli. La prima tabella riguarda gli Stati Uniti e la seconda il Nord-Ovest d’Italia. Osservazioni pesate.

Stati Uniti						
	Prestigio dei figli					Totale
		1	2	3	4	
Prestigio dei padri	1	25%	52%	21%	2%	100% (757.70)
	2	14%	53%	29%	4%	100% (3587.60)
	3	6%	44%	44%	6%	100% (1144.90)
	4	6%	27%	50%	17%	100% (148.91)
Nord-Ovest d’Italia						
	Prestigio dei figli					Totale
		1	2	3	4	
Prestigio dei padri	1	29%	17%	41%	13%	100% (126.56)
	2	12%	25%	18%	45%	100% (35.47)
	3	18%	12%	52%	18%	100% (100.62)
	4	2%	22%	34%	42%	100% (32.86)

Si nota subito che c'è una forte differenza per quanto riguarda le numerosità assolute tra Stati Uniti e Nord-Ovest d'Italia, inoltre bisogna ricordare che le due variabili sul prestigio sono costruite in modo diverso, quindi ha senso aspettarsi distribuzioni molto diverse.

Ad ogni modo, la tendenza diagonale è rappresentata abbastanza bene (con lievi anomalie) da entrambe le distribuzioni, quindi i soggetti, in genere, tendono a raggiungere lo stesso livello di prestigio del padre.

Vediamo, quindi, se c'è una *mobilità sociale* generazionale dei figli rispetto ai loro genitori nelle due aree. Per gli Stati Uniti, il 75% dei soggetti con padre di livello di prestigio basso raggiunge un livello di prestigio superiore rispetto al padre, la corrispondente percentuale per il Nord-Ovest d'Italia è il 71%. Tra coloro che, invece, hanno padre di prestigio medio basso, il 33% degli statunitensi e il 63% dei nativi del Nord-Ovest d'Italia superano il livello del padre.

Un altro interessante confronto può essere fatto sulla relazione tra titolo di studio e prestigio, sempre tra gli stessi soggetti.

Tabella6.5: P-value dei test Chi-Quadrato per il *livello di prestigio* sul *livello di istruzione* dei figli (generazione 2) con genitori autoctoni. Classe di età all'intervista '40-59' per gli Stati Uniti e '40+' per il Nord-Ovest d'Italia.

	p-value
Stati Uniti	<.0001
Italia	<.0001

La tabella mostra che, sia per gli Stati Uniti che per il Nord-Ovest d'Italia, è forte la dipendenza tra il livello di istruzione e il livello di prestigio. In effetti questa

relazione è abbastanza ovvia, e potrebbe valere in qualsiasi zona di studio.

Si sono costruiti tre modelli per gli Stati Uniti e due per il Nord-Ovest d'Italia, in cui la variabile risposta è il prestigio (in particolare una categoria del prestigio, reso dicotomico) e come esplicative:

- il *numero di fratelli*;
- il *livello di istruzione*;
- il *livello di prestigio del padre*.

Si stimano modelli logistici che hanno come variabile risposta la dicotomica indicante il livello di prestigio e come esplicative tutte le altre variabili, rese fattoriali. Per confrontare i risultati si riporta una tabella riassuntiva dei modelli: sono riportate, per ogni categoria delle variabili risposta, le stime dei coefficienti e il loro esponenziale.

La lettura degli asterischi (livelli di significatività) è la seguente:

- ◆ '***' indica la significatività al 100%;
- ◆ '**' indica la significatività al 99,9%;
- ◆ '*' indica la significatività al 99%;
- ◆ '.' indica la significatività al 95%;
- ◆ ' ' indica che la variabile non è significativa.

Tabella6.6: stime dei coefficienti, loro livello di significatività rappresentato dal numero degli asterischi ed esponenziale della stima del coefficiente. La prima tabella rappresenta il risultato per gli Stati Uniti e la seconda quello per il Nord-Ovest d'Italia.

Si sono evidenziate in giallo le modalità delle variabili che risultano significative.

Stati Uniti				
	Modello 1 0-livelli 2,3,4 1-livello 1		Modello 2 0-livelli 3,4 1-livelli 1,2	
esplicativa	Stima coeff	Funzione coeff	Stima coeff	Funzione coeff
Numero di fratelli				
<i>Fratelli0 (di riferimento)</i>	-	1	-	1
Fratelli1	-0.112099	0.8939558	0.07656	1.0795670
Fratelli2	0.003838	1.0038450	0.05031	1.0515970
Fratelli3	0.123928	1.1319340	0.16222	1.1761190
Fratelli4	0.408607 *	1.5047200	0.29675 *	1.3454790
Livello di istruzione				
<i>Istruzione0 (di riferimento)</i>	-	1	-	1
Istruzione1	-1.151535 ***	0.3161511	-1.18148 ***	0.3068243
Istruzione2	-2.676765 ***	0.0687853	-3.15716 ***	0.0425464
Livello di prestigio del padre				
<i>Prestigio_padre1 (di riferimento)</i>	-	1	-	1
Prestigio_padre2	-0.453177 ***	0.6356056	-0.10078	0.9041320
Prestigio_padre3	-0.828987 ***	0.4364912	-0.53050 ***	0.5883107
Prestigio_padre4	-0.745305 *	0.4745895	-0.21327	0.8079380
Nord Ovest d'Italia				
Numero di fratelli				
<i>Fratelli0 (di riferimento)</i>	-	1	-	1
Fratelli1	-0.5616	0.5702959	-0.5985	0.5496355
Fratelli2	0.4236	1.5274500	-0.4026	0.6685795
Fratelli3	0.3249	1.3838920	0.0490	1.0502200
Livello di istruzione				
<i>Istruzione0 (di riferimento)</i>	-	1	-	1
Istruzione1	-0.6120 .	0.5422653	0.1217	1.1294150
Istruzione2	-1.8464 ***	0.1576465	-0.4553	0.6342577
Livello di prestigio del padre				
<i>Prestigio_padre1 (di riferimento)</i>	-	1	-	1
Prestigio_padre2	-0.2660	0.7664391	-0.1949	0.8229170
Prestigio_padre3	-0.7379	0.4781169	-0.6504 *	0.5218370
Prestigio_padre4	-0.9682	0.3797660	-0.3759	0.6866710

Innanzitutto è possibile confrontare solo i primi due modelli, perché il terzo non è stato stimato per il Nord-Ovest d'Italia (quindi non viene riportato nel confronto). Confrontiamo variabile per variabile, in ogni aspetto di interesse:

- Il *numero di fratelli*³³ è significativo per i dati degli Stati Uniti nella categoria dei quattro fratelli o più, in entrambi i modelli proposti; non risulta invece mai significativo per il Nord-Ovest d'Italia. Traiamo comunque alcune conclusioni importanti per il confronto, a partire dalle stime dei coefficienti. Nel primo modello, per entrambe le aree di interesse, nel passaggio dalla categoria di un fratello a quella di due fratelli i coefficienti, da negativi, diventano positivi, quindi avere due fratelli o più aumenta la probabilità, per il soggetto, di raggiungere, e non superare, il livello inferiore di prestigio.

Nel secondo modello, invece, mentre per il Nord-Ovest d'Italia il salto di qualità sembra essere il passaggio da due fratelli a tre fratelli o più, per gli Stati Uniti sembra, addirittura, essere il passaggio da non avere fratelli ad averne anche solo uno.

In ogni caso il numero di fratelli sembra più incisivo nella spiegazione del livello di prestigio dei soggetti statunitensi che di quelli italiani.

- Il *livello di istruzione* è la variabile che risulta più significativa, in generale, nella spiegazione del livello di prestigio dei soggetti. In particolare, sia la categoria '1' che la '2' di istruzione sono significative (per entrambe le zone) nel primo

³³ Si ricorda che per gli Stati Uniti l'analisi è stata fatta distinguendo fino alla classe di quattro fratelli o più, mentre per il Nord-Ovest d'Italia ci si è fermati alla classe di tre fratelli o più.

modello, mentre nel secondo modello sono entrambe significative per gli Stati Uniti, ma non lo sono per il Nord-Ovest d'Italia.

Nel primo modello tutti i coefficienti sono negativi quindi, sia in Italia che negli Stati Uniti, superare il livello inferiore di istruzione aiuta ad abbandonare il livello inferiore di prestigio. Nel secondo modello si traggono le medesime conclusioni per gli Stati Uniti, mentre per l'Italia sembra più importante il passaggio dal livello medio al livello alto di istruzione, affinché il soggetto riesca a raggiungere livello medio alto, o alto, di prestigio.

- Il livello di prestigio del padre, infine, risulta spesso significativo per gli Stati Uniti, mentre è significativo in un solo modello, e in una sola categoria, per l'Italia del Nord-Ovest.

Ma entriamo nei dettagli. Nel primo modello tutte le stime dei coefficienti risultano negative quindi, per entrambe le zone di interesse, avere padre con livello di prestigio diverso da '1' (basso) aumenta la probabilità, per il soggetto, di raggiungere livello di prestigio superiore al più basso. Anche tutte le stime del secondo modello sono negative e la spiegazione è la medesima.

Quindi il livello di prestigio del padre è molto utile per prevedere il livello di prestigio del soggetto, in particolare negli Stati Uniti, ma anche in Italia.

Osservate le differenze e le analogie tra i risultati nelle prime due fasi, si procede con il confronto per la terza fase.

6.3. TERZA FASE

“Nuovi immigrati prendono il posto dei bambini non nati, svolgendo i lavori meno pagati e meno prestigiosi, riempiendo così gli scalini più bassi della piramide sociale”.

In questa fase si confrontano gli autoctoni e gli immigrati della generazione 2 (1930-1949), su diversi aspetti, per vedere se è vero che gli immigrati riempiono gli scalini più bassi della piramide sociale.

Per confrontare i risultati degli Stati Uniti con quelli dell'Italia del Nord-Ovest si riportano e poi i test Chi-Quadrato tra titolo di studio e provenienza e tra livello di prestigio e provenienza.

Tabella6.7: *titolo di studio* in base alla *provenienza*. Osservazioni pesate. Classe di età all'intervista '40-59'. Generazione 2 (1930-1949).

Stati Uniti				
Titolo di studio				
	0	1	2	Totale
Autoctoni	18%	53%	29%	100% (6899.50)
Immigrati	25%	36%	39%	100% (537.69)
Nord-Ovest d'Italia				
Autoctoni	30%	40%	30%	100% (560.83)
Immigrati	44%	36%	20%	100% (273.86)

Tabella6.8: *livello di prestigio* in base alla *provenienza*. Osservazioni pesate. Classe di età all'intervista '40-59'. Generazione 2 (1930-1949).

Stati Uniti					
	Livello di prestigio				
	1	2	3	4	Totale
Autoctoni	15%	50%	31%	4%	100% (6713.80)
Immigrati	19%	45%	30%	6%	100% (515.45)
Nord-Ovest d'Italia					
	Livello di prestigio				
	1	2	3	4	Totale
Autoctoni	24%	24%	37%	15%	100% (233.24)
Immigrati	32%	24%	30%	14%	100% (114.42)

Prima di trarre le conclusioni si confrontano i test relativi alle due zone di interesse.

Tabella6.9: P-value dei test Chi-Quadrato per il *titolo di studio* sulla variabile *provenienza* (autoctoni o immigrati). Classe di età '40-59'. Generazione 2 (1930-1949).

	P-value
Stati Uniti	<.0001
Italia	0.0002

Tabella6.10: P-value dei test Chi-Quadrato per il *livello di prestigio* sulla variabile *provenienza* (autoctoni o immigrati). Classe di età '40-59'. Generazione 2 (1930-1949).

	P-value
Stati Uniti	0.0003
Italia	0.3650

Per fare un confronto si osservano i P-value, riportati, dei quattro test. In generale sembra più forte la relazione tra provenienza e titolo di studio che quella tra provenienza e livello di prestigio, questo dipende

probabilmente dal fatto che il titolo di studio è legato ad una fase giovanile della vita, che quindi, per molti immigrati, è strettamente connessa con la zona di provenienza, solitamente meno sviluppata di quella di arrivo, il prestigio invece è più legato ad una fase adulta della vita dei soggetti, quindi entrano in gioco, nel determinarlo, altre componenti legate molto alla zona di immigrazione.

Entrando più nel dettaglio, si nota che si rifiuta l'ipotesi di indipendenza tra titolo di studio e provenienza sia per gli Stati Uniti che per il Nord-Ovest d'Italia. L'ipotesi di indipendenza tra livello di prestigio e provenienza è rifiutata per gli Stati Uniti, mentre è accettata per l'Italia.

Quindi la terza fase sembra verificata meglio per gli Stati Uniti che per l'Italia, ma in ogni caso le conclusioni sono che gli immigrati hanno titolo di studio nettamente inferiore rispetto agli autoctoni, quindi, almeno nei primi anni dopo l'immigrazione, riempiono gli scalini più bassi della piramide sociale, ma ciò non toglie che poi possano scalare la piramide sociale, e nel farlo possano essere, in un certo senso, più motivati gli stranieri rispetto agli autoctoni, e ciò spiegherebbe perché questa differenza, in Italia, tende a diminuire notevolmente con il livello di prestigio.

Si confrontano, quindi, sempre i risultati del test Chi-Quadrato tra livello di istruzione e livello di prestigio, ma per i soggetti immigrati della generazione 2 (1930-1949).

Tabella6.11: P-value dei test Chi-Quadrato tra il *livello di prestigio* e il *titolo di studio*. Classe di età '40-59'. Generazione 2 (1930-1949). Immigrati.

	P-value
Stati Uniti	<.0001
Italia	<.0001

La tabella mostra, come prevedibile, che sia per gli Stati Uniti che per il Nord-Ovest d'Italia, si rifiuta l'ipotesi di indipendenza tra livello di prestigio e titolo di studio, anche per i soggetti immigrati.

Le ipotesi alla base della terza fase sono quindi state verificate e confrontate.

6.4. QUARTA FASE

“Gli immigrati si assimilano molto velocemente ai comportamenti sociali dei residenti, abbassando la loro fecondità e investendo nel capitale umano per i loro figli, aumentando così la possibilità di crescita sociale di questi ultimi.”

Quindi si confrontano i risultati tra Stati Uniti e Nord-Ovest d'Italia.

La prima analisi riguarda l'abbassamento di fecondità da parte degli immigrati (in confronto con la fecondità degli autoctoni) di generazione 2 (1930-1949). Si riporta una tabella riassuntiva che contiene sia la distribuzione del numero di fratelli che quella del numero di figli, e una misura della differenza tra le due distribuzioni.

Tabella6.12: percentuali marginali sul *numero di fratelli* e di *figli* sulla base della variabile *provenienza* (immigrati o autoctoni)³⁴. Generazione 2. Classe di età all'intervista '40-59'. La prima tabella riguarda gli Stati Uniti e la seconda il Nord-Ovest d'Italia. Osservazioni pesate.

³⁴ Ricordiamo che per gli Stati Uniti si sono mantenute le classi dei fratelli e dei figli fino alla '6 o più', mentre per il Nord-Ovest d'Italia, a causa dell'esigua numerosità, si è tenuta come classe superiore per entrambe le variabili la '3 o più'.

Stati Uniti							
Numero di fratelli							
	0	1	2	3	4	5	6
Autoctoni	7%	19%	18%	15%	11%	8%	22%
Immigrati	6%	16%	15%	11%	11%	10%	30%
Differenza percentuale in modulo	1%	3%	3%	4%	0%	2%	8%
Numero di figli							
	0	1	2	3	4	5	6
Autoctoni	13%	12%	30%	21%	12%	6%	6%
Immigrati	15%	13%	34%	18%	10%	4%	6%
Differenza percentuale in modulo	2%	1%	4%	3%	2%	2%	0%
Nord-Ovest d'Italia							
Numero di fratelli							
	0	1	2	3	4	5	6
Autoctoni	20%	30%	21%	29%	-	-	-
Immigrati	9%	13%	19%	59%	-	-	-
Differenza percentuale in modulo	11%	17%	2%	30%	-	-	-
Numero di figli							
	0	1	2	3	4	5	6
Autoctoni	15%	35%	38%	12%	-	-	-
Immigrati	10%	25%	44%	21%	-	-	-
Differenza percentuale in modulo	5%	10%	6%	9%	-	-	-

A primo impatto si nota che la distribuzione dei fratelli degli autoctoni, tra Stati Uniti e Italia del Nord-Ovest, è molto diversa, in particolare la fecondità della generazione 1 statunitense (rappresentata dai fratelli della generazione 2) è molto più alta della corrispondente fecondità degli autoctoni italiani, quindi bisogna leggere i risultati anche sulla base di questa deduzione. Si utilizza, per il confronto, la somma delle differenze percentuali in valore assoluto.

La somma delle differenze tra numero di fratelli di autoctoni e immigrati vale 21 per gli Stati Uniti e 60 per l'Italia, mentre la somma delle differenze tra numero di figli di autoctoni e immigrati vale 14 per gli Stati Uniti e 30 per l'Italia: sembra chiaro che in entrambe le zone l'assimilazione è molto forte. Si nota che le abitudini riguardanti la fecondità degli immigrati, negli Stati

Uniti, sono molto più simili alle abitudini degli autoctoni statunitensi, mentre gli immigrati nel Nord-Ovest d'Italia sono molto diversi, per le stesse abitudini, agli autoctoni. Di conseguenza si parte da una divario molto differente, ed è più difficile per gli immigrati nel Nord-Ovest d'Italia assimilarsi agli autoctoni di quanto non lo sia per gli immigrati negli Stati Uniti.

Ad ogni modo è evidente la tendenza ad avvicinarsi alle caratteristiche degli autoctoni, rappresentata da distribuzioni più simili per il numero di figli che per il numero di fratelli (caratteristica legata al luogo di provenienza).

6.5. QUINTA FASE

“Nella nuova generazione, i figli istruiti degli immigrati rifiutano i lavori sotto-pagati dei loro genitori, causando una nuova carenza di lavoratori non esperti. In questo modo il ciclo si ripete (si torna alla fase 3).”

Oggetto di studio è quindi la generazione 3 (1950-1969), in particolare le due sottopopolazioni:

- figli di autoctoni *unskilled*;
- figli di immigrati *unskilled*.

Per identificare gli immigrati *unskilled* della generazione 2 (1930-1949) si scelgono coloro che hanno prestigio basso o medio basso (cioè '1' o '2').

Si confrontano, quindi, alcune tabelle per vedere le differenze tra la situazione negli Stati Uniti e quella nel Nord-Ovest d'Italia.

Innanzitutto, confronto il titolo di studio (e poi il prestigio) dei figli di autoctoni e di immigrati *unskilled*.

Tabella6.13: *titolo di studio* dei figli di autoctoni e immigrati della generazione 3 (1950-1969). Osservazioni pesate. Classe di età all'intervista '30-59'. Tengo i soggetti con padre con prestigio '1' e '2'. La prima tabella riguarda gli Stati Uniti e la seconda il Nord-Ovest d'Italia.

Stati Uniti				
Titolo di studio				
	0	1	2	Totale
Figli di autoctoni	10%	58%	32%	100% (5717.49)
Figli di immigrati	22%	43%	35%	100% (580.81)
Nord-Ovest d'Italia				
Figli di autoctoni	6%	53%	41%	100% (1316.64)
Figli di immigrati	14%	57%	29%	100% (246.18)

Tabella6.14: *livello di prestigio* dei figli di autoctoni e immigrati della generazione 3 (1950-1969). Osservazioni pesate. Classe di età all'intervista '30-59'. Tengo i soggetti con padre con prestigio '1' e '2'. La prima tabella riguarda gli Stati Uniti e la seconda il Nord-Ovest d'Italia.

Stati Uniti					
Livello di prestigio					
	1	2	3	4	Totale
Figli di autoctoni	15%	53%	29%	3%	100% (5581.31)
Figli di immigrati	16%	51%	28%	5%	100% (558.81)
Nord-Ovest d'Italia					
Figli di autoctoni	32%	41%	15%	12%	100% (1074.24)
Figli di immigrati	43%	28%	23%	6%	100% (193.57)

Da queste due tabelle si traggono le seguenti conclusioni. Le distribuzioni degli Stati Uniti e del Nord-Ovest d'Italia sono abbastanza simili, per il titolo di studio, per quanto riguarda i figli di autoctoni, diversificate invece per quanto riguarda i figli di immigrati. In particolare, i figli di autoctoni statunitensi sono un po' meno istruiti dei figli di autoctoni italiani, mentre i figli di immigrati statunitensi hanno titoli di studio più diversificati rispetto ai figli di immigrati in Italia, che

si posizionano perlopiù nella classe intermedia del livello di istruzione.

Entrando nel dettaglio delle ipotesi, si vede che, sia per gli Stati Uniti che per il Nord-Ovest d'Italia, i figli di immigrati hanno titolo di studio inferiore rispetto ai figli di autoctoni, ma la differenza non è forte in nessuno dei due casi: ciò conferma l'ipotesi di una tendenza all'omologazione dei figli di immigrati ai loro coetanei, figli di autoctoni.

Dalla tabella sul livello di prestigio si vede, innanzitutto, che le distribuzioni dell'Italia e degli Stati Uniti sono molto diverse; ciò è dovuto anche a conseguenze delle differenze nella costruzione delle due variabili.

Ad ogni modo, sembra che i figli di immigrati negli Stati Uniti siano ben omologati ai figli di autoctoni per il livello di prestigio: le due distribuzioni sono infatti praticamente uguali.

Per il Nord-Ovest d'Italia, invece, le due distribuzioni sono più differenziate, ma non abbastanza da confutare l'ipotesi di una progressiva omologazione dei figli di immigrati ai figli di autoctoni.

Infine si confronta la tendenza alla *mobilità sociale* dei figli di immigrati e di autoctoni.

Tabella6.15: *livello di prestigio* dei figli di immigrati *unskilled* (livello di prestigio '1' e '2') della generazione 3 (1950-1969), in base al *livello di prestigio dei genitori* (generazione 2, 1930-1949). Osservazioni pesate. Classe di età all'intervista '30-59'. La prima tabella riguarda gli Stati Uniti, la seconda il Nord-Ovest d'Italia.

Stati Uniti						
	Prestigio dei figli					
Prestigio dei padri		1	2	3	4	Totale
	1	24%	54%	19%	3%	100% (91.96)
	2	15%	50%	30%	5%	100% (456.84)
Nord-Ovest d'Italia						
	Prestigio dei figli					
Prestigio dei padri		1	2	3	4	Totale
	1	48%	26%	21%	5%	100% (158.39)
	2	19%	40%	30%	11%	100% (35.18)

Da quest'ultima tabella si nota che c'è una forte *mobilità sociale* per i figli di immigrati *unskilled*, in particolare il 76% dei soggetti figli di immigrati negli Stati Uniti, e il 52% dei figli di immigrati in Italia, con padre di livello di prestigio basso, superano il livello di prestigio del padre. Inoltre, tra i figli con padre di livello di prestigio medio basso, il 35% dei soggetti figli di immigrati negli Stati Uniti e il corrispondente 41% per l'Italia, riesce a raggiungere livelli medio alti o alti di prestigio, superando il livello del padre.

Quindi anche l'ultima fase è valida sia per gli Stati Uniti che per il Nord-Ovest d'Italia, anche se sembra esserci una mobilità maggiore per gli immigrati negli Stati Uniti che in Italia.

6.6. RIEPILOGO

Dopo aver rivisitato e confrontato i risultati di tutte le cinque fasi, si conclude che lo schema teorico presentato all'inizio della tesi non viene confutato dai dati. Il modello viene sostanzialmente seguito in entrambe le aree di interesse, a prescindere dalle grandi differenze di *background sociale e culturale*, al quadro storico delle due aree nei periodi studiati, etc.

La generazione 2 (1930-1949) è interessata (sia negli Stati Uniti, che nel Nord-Ovest d'Italia) da una forte tendenza alla *mobilità sociale intergenerazionale*.

L'ipotesi che il numero di fratelli incida sul titolo di studio raggiunto dai soggetti è verificata, mentre il numero di fratelli sembra incidere meno sul livello di prestigio, in particolare in Italia. Una considerazione, risultata da diverse analisi, è che i soggetti benestanti riescono ad offrire un buon titolo di studio e un buon livello di prestigio anche a molti figli. I soggetti *unskilled*, invece, sono spinti a diminuire il numero di figli, e questa scelta permette di offrire loro un futuro più promettente.

Sia negli Stati Uniti che in Italia si assiste ad una carenza di personale *unskilled* di generazione 2 (1930-1949), e questo fenomeno sembra più intenso in Italia che negli Stati Uniti.

Si assiste, quindi, all'ingresso di flussi migratori di generazione 2 (1930-1949) nelle due aree di interesse, flussi che coprono gli scalini più bassi della piramide sociale, ma che portano con sé, specialmente negli USA, anche immigrati istruiti e che svolgono lavori prestigiosi. Quindi conviene distinguere fra gli immigrati *skilled* e quelli *unskilled*, e concentrare le nostre attenzioni principalmente sui secondi.

Si è poi notato che gli immigrati, pur avendo un titolo di studio mediamente inferiore rispetto a quello degli autoctoni, svolgono lavori migliori rispetto a quelli che ci aspetteremmo data la loro istruzione.

Ciò dimostra un gran desiderio di *ascesa sociale* fra gli immigrati, probabilmente per una personale ricerca di *riscatto sociale*.

Questa tendenza si rispecchia anche nella progressiva assimilazione dei comportamenti degli immigrati a quelli degli autoctoni, anche per quanto riguarda il livello di fecondità.

Ma la tendenza al *riscatto sociale* si compie pienamente solo nella generazione dei figli (generazione 3, nata nel 1950-1969). Le due sottopopolazioni di figli di *unskilled*, autoctoni e immigrati, tendono ad assomigliarsi gradualmente, sia per titolo di studio che per livello di prestigio, grazie alla forte *mobilità sociale* di cui possono giovare. Questo fatto crea una carenza di lavoratori *unskilled* di generazione 3.

Nuovi immigrati, di generazione 3, stanno ora entrando negli Stati Uniti e nel Nord-Ovest d'Italia. E' facile immaginare che, dopo aver coperto, inizialmente, gli scalini più bassi della piramide sociale, tenderanno poi la *scalata sociale*, anche attraverso i loro figli, creando così carenza di personale *unskilled* nella generazione (nata dopo il 1970) che si affaccia, ora, al mondo del lavoro.

7. ALLEGATI

7.1. ALLEGATO A: STATI UNITI

In questo paragrafo si riportano le procedure utilizzate per la costruzione delle variabili che vengono utilizzate nell'analisi dei dati sugli Stati Uniti. Si procede con un elenco di tutte le variabili con la corrispondente spiegazione.

PRESTIGE OF RESPONDENT'S OCCUPATION e PRESTIGE OF FATHER'S OCCUPATION

Le variabili che assumevano valore '0' nel dataset iniziale assumono ora il simbolo '.' dei dati mancanti.

I livelli di prestigio assegnati alle occupazioni erano presi dai sistemi di classificazione sviluppato alla *NORC*, nel periodo 1963-1965, in un progetto sul prestigio occupazionale³⁵. Questo concetto di prestigio è definito come una stima del livello sociale dell'occupazione del rispondente. Il livello di prestigio veniva generato chiedendo al rispondente di stimare il livello sociale dell'occupazione in una scala da 1 a 9, dal più basso al più alto, disegnato su un cartellino e presentato al rispondente. I *box* nella scala erano etichettati inoltre con: 'basso', 'medio', 'alto'.

- 1970. Il livello di prestigio era inizialmente destinato ad essere usato con i codici occupazionali del Censimento del 1960 ed è stato adattato nel 1970 come segue:
 - se l'occupazione era nuova ed assegnata ad una precedente suddivisione esistente, riceveva il livello di prestigio della categoria generale di quella suddivisione;

³⁵ Robert W. Hodge, Paul S. Siegel, and Peter H. Rossi.

- se l'occupazione era raggruppata con una occupazione simile nel 1960 ma divisa dal 1970, entrambe le occupazioni ricevevano lo stesso livello di prestigio;
- se l'occupazione era completamente nuova per la lista e non era assegnata a nessuna suddivisione esistente, riceveva il livello di prestigio della categoria più numerosa;
- 1980. La GSS replicò ed estese la precedente stima del prestigio occupazionale. Il modulo del prestigio fu ridisegnato³⁶. Il modulo fu discusso in dettaglio nel *"GSS Methodological Reports"*.

Si assiste quindi ad una duplice classificazione della variabile sul prestigio, sia per i soggetti che i padri dei soggetti.

Poiché la scala è però praticamente la stessa si decide di unificarle in un'unica variabile riassuntiva di tutte, e nei casi di passaggio inserire un valore, funzione dei due valori assegnati (si sceglie di utilizzare il massimo). Tutto ciò è stato fatto solo dopo aver considerato che le variazioni erano trascurabili.

Si riportano quindi le due analisi svolte prima della creazione delle variabili del prestigio del soggetto e del padre del soggetto, per dimostrare la validità delle scelte effettuate, in particolare si incrociano le vecchie variabili con le nuove costruite, per vedere se vi sono disallineamenti consistenti o meno. Si ricorda che la variabile costruita è la variabile risposta se solo una delle due di partenza ha un valore in risposta, assume valore mancante se entrambe le variabili di partenza erano valori mancanti, e il valore massimo tra le due, se erano presenti entrambe.

³⁶Robert W. Hodge, Judith Treas and Keiko Nakao in consultation with Smith and Davis.

Tabella7.1: *prestige_classe* per *prestige_ok_classe*.

prestige_classe		prestige_ok_classe				
Frequenza Percentuale		1	2	3	4	Totale
1	5628	367	24	0	6019	
	23.19	1.51	0.10	0.00	24.80	
2	0	12113	293	12	12418	
	0.00	49.92	1.21	0.05	51.17	
3	0	0	5223	42	5265	
	0.00	0.00	21.52	0.17	21.70	
4	0	0	0	565	565	
	0.00	0.00	0.00	2.33	2.33	
Totale		5628	12480	5540	619	24267

Noto che il 96.96% dei dati sta sulla diagonale quindi l'errore è trascurabile.

Tabella7.2: test Chi-Quadrato tra le variabili *prestige_classe* e *prestige_ok_classe*.

Statistica	DF	Valore	Prob
Chi-quadrato	9	66332.2162	<.0001

Anche i test sulla correlazione sembrano non scontrarsi con la mia approssimazione.

Tabella7.3: *prestige80_classe* per *prestige_ok_classe*.

prestige80_classe		prestige_ok_classe				
Frequenza Percentuale		1	2	3	4	Totale
1	4115	184	22	0	4321	
	14.90	0.67	0.08	0.00	15.64	
2	0	14135	286	30	14451	
	0.00	51.17	1.04	0.11	52.32	
3	0	0	7843	33	7876	
	0.00	0.00	28.39	0.12	28.51	
4	0	0	0	975	975	
	0.00	0.00	0.00	3.53	3.53	
Totale		4115	14319	8151	1038	27623

Vedo che il 97.99% dei dati sta sulla diagonale quindi l'errore è trascurabile.

Tabella7.4: test Chi-Quadrato tra le variabili prestige_classe per prestige_ok_classe.

Statistiche per la tabella di prestige80_classe per prestige_ok_classe

Statistica	DF	Valore	Prob
Chi-quadrato	9	77809.1931	<.0001

Si conclude che la scelta effettuata è sensata e viene utilizzata nell'analisi. Faccio, quindi, la stessa analisi per il prestigio del padre.

Tabella7.5: *pa_prestige_classe* per *papres_ok_classe*.

pa_prestige_classe		papres_ok_classe				
	Frequenza					Totale
	Percentuale	1	2	3	4	
1	3980 17.85	3980	337	24	1	4342
2	0 0.00	0	13086	206	2	13294
3	0 0.00	0	0	4100	40	4140
4	0 0.00	0	0	0	524	524
Totale		3980	13423	4330	567	22300

Vedo che il 97.27% dei dati sta sulla diagonale, quindi l'errore è trascurabile.

Tabella7.6: test Chi-Quadrato tra *pa_prestige_classe* e *papres_ok_classe*.

Statistica	DF	Valore	Prob
Chi-quadrato	9	61136.8687	<.0001

Tabella7.7: *pa_prestige80_classe* per *papres_ok_classe*.

pa_prestige80_classe		papres_ok_classe			
		Frequenza			
		Percentuale			
		1	2	3	4 Totale
1	2109	88	14	0	2211
	9.47	0.39	0.06	0.00	9.92
2	0	13740	271	6	14017
	0.00	61.67	1.22	0.03	62.92
3	0	0	5171	25	5196
	0.00	0.00	23.21	0.11	23.32
4	0	0	0	855	855
	0.00	0.00	0.00	3.84	3.84
Totale	2109	13828	5456	886	22279

Le osservazioni che stanno sulla diagonale sono il 98.19%.

Tabella7.8: test Chi-Quadrato tra *pa_prestige80_classe* e *papres_ok_classe*.

Statistica	DF	Valore	Prob
Chi-quadrato	9	63215.5670	<.0001

AGE OF RESPONDENT:

Dove, nel dataset in SAS compaiono i valori '98' o '99', questi sono stati sostituiti con il simbolo '.' dei valori mancati.

Quindi ho raggruppato la variabile dell'età in classi di età utilizzabili con maggiore semplicità, quelle di interesse per le analisi.

NUMBER OF CHILDREN:

Originariamente la classe superiore era di otto figli, si è deciso di costruire delle classi intermedie in cui la classe superiore fosse di tre figli; queste classi sono utilizzate in momenti diversi delle analisi sugli Stati Uniti.

In alcune analisi è però necessario tenere classi fino a 6 o 7 figli, in quei casi viene utilizzata la variabile iniziale, non suddivisa in classi (con alcune operazioni per riportare i risultati alle classi di interesse).

RS HIGHEST DEGREE:

La variabile indica il massimo titolo di studio raggiunto dal soggetto, la riclassificazione è fatta come segue:

- '0' = "LT HIGH SCHOOL" (corrisponde a titoli inferiori alla scuola media italiana);
- '1' = "HIGH SCHOOL" (corrisponde a scuola media e scuola superiore italiana);
- '2' = "JUNIOR COLLEGE", "BACHELOR", "GRADUATE" (corrisponde ad ogni titolo di studio superiore alla scuola superiore italiana).

WAS R BORN IN THIS COUNTRY:

La variabile indica se il soggetto è nato o meno negli Stati Uniti, e serve per valutare se è un soggetto autoctono o immigrato.

Sistemo i valori mancanti inizialmente codificati con codici diversi, con l'unico simbolo '.'.

WHERE RS PARENTS BORN IN THIS COUNTRY:

La variabile indica le diverse modalità di provenienza dei due genitori, secondo lo schema seguente, interpretato grazie al '*Codebook*':

1. Entrambi i genitori sono nati negli US;
2. La madre è nata negli US il padre no;
3. La madre non è nata negli US, il padre sì;
4. La madre è nata negli US, il padre è ignoto;
5. La madre non è nata negli US, il padre è ignoto;
6. La madre è ignota, il padre è nato negli US;

7. La madre è ignota, il padre non è nato negli US;
8. Sia il padre che la madre sono ignoti;
9. Nessuno dei due è nato negli US;

Innanzitutto sistemo i valori mancanti codificandoli con lo stesso simbolo '.'.

Quindi ricodifico, riportandomi ad una risposta dicotomica, scegliendo come *persone con genitori stranieri* quelle che hanno o entrambi i genitori stranieri, o uno straniero e l'altro ignoto, con il seguente schema:

- 0-1-2-3-4-7 → '0'-Sì
- 5-6-8 → '1'-No

YEAR OF BIRTH:

La variabile indica l'anno di nascita del soggetto. Sistemo i valori mancanti. Quindi costruisco la variabile GENERAZIONE che serve per le mie analisi, con il seguente schema:

- 1910-1929 → generazione1
- 1930-1949 → generazione2
- 1950-1969 → generazione3

A me interessano solo le tre generazioni quindi le altre coorti non sono state codificate, e non verranno utilizzate.

NUMBER OF BROTHERS AND SISTERS:

La variabile iniziale distingue fino a 34 fratelli e sorelle (è indicato il totale). Sistemo i valori mancanti, codificandoli tutti con lo stesso simbolo '.', quindi costruisco delle classi di età più maneggevoli, cioè fino a tre fratelli o più. In alcune analisi è necessaria una suddivisione più accurata fino a sei o sette fratelli, in

quel caso si trovano degli espedienti, man mano che si procede con l'analisi.

II PESI:

- OVERSAMP: il peso per i sottocampioni di soggetti di colore, assumono valore tra 0.3778 e 1.2410 e vengono utilizzati in ogni analisi;
- FORMWT: il peso per la post-stratificazione, varia da 0.402 a 7.180, e viene utilizzato, anche questo secondo valore, in ogni analisi.

7.2. ALLEGATO B: NORD-OVEST D'ITALIA

Oggetto di studio della seconda parte della tesi è il Nord-Ovest d'Italia. Quindi creo un dataset a partire da quello iniziale di 49541 individui, in cui tengo solo i soggetti intervistati nel Nord-Ovest di Italia.

Tabella7.9: distribuzione della variabile RIPARTIZIONE.

ripartizione	Frequenza	Percentuale	Frequenza cumulata	Percentuale cumulata
1	10057	20.30	10057	20.30
2	10175	20.54	20232	40.84
3	9015	18.20	29247	59.04
4	14828	29.93	44075	88.97
5	5466	11.03	49541	100.00

Le ripartizioni seguono la seguente legenda:

- ◆ 1→ Nord-Ovest: Piemonte, Valle D'Aosta, Liguria, Lombardia;
- ◆ 2→ Nord-Est: Trentino Alto Adige, Veneto, Friuli Venezia Giulia, Emilia Romagna;
- ◆ 3→ Centro: Marche, Toscana, Umbria, Lazio;
- ◆ 4→ Sud: Campania, Abruzzo, Molise, Puglia, Basilicata, Calabria;
- ◆ 5→ Isole: Sicilia, Sardegna.

Quindi il dataset che riguarda il solo Nord-Ovest contiene 10057 soggetti.

Si procede con la descrizione e creazione delle variabili.

COEFFICIENTE

Ad ogni soggetto è associato un coefficiente, da questi coefficienti mi calcolo i pesi, facendo in modo che il totale dei pesi sia pari al totale dei soggetti, 10057. Per fare ciò divido ogni coefficiente per il totale dei coefficienti (150.150.069.873) e poi moltiplico il

risultato per 10057. Costruisco in questo modo la variabile PESO. Queste due variabili non contengono dati mancanti.

ORDINE COMPONENTE e CODICE FAMIGLIA

Sono le due variabili che servono per identificare il soggetto.

L'ordine del componente assume valori da 1 a 10, mentre il codice della famiglia è di sei cifre ed è un numero progressivo, univoco per la famiglia, all'interno dell'indagine.

ETÀ

La variabile età nel dataset assume valori da 1 a 97. Semplificando leggermente il calcolo dell'anno di nascita, e decidendo di non lavorare, per comodità, con i mesi di nascita, si decide di calcolare l' ANNO DI NASCITA del soggetto a partire dalla sua età, supponendo un errore minimo: $\text{anno di nascita} = 2003 - \text{età}$.

GENERAZIONE

A partire dall'anno di nascita mi costruisco le tre generazioni di interesse:

- ♦ Generazione 1: nati 1910-1929;
- ♦ Generazione 2: nati 1930-1949;
- ♦ Generazione 3: nati 1950-1969.

FRATELLI_O_SORELLE, FRATELLI_MASCHI, SORELLE_FEMMINE

Queste tre variabili servono, congiuntamente, a creare la variabile FRATELLI, che indica il numero totale di fratelli e sorelle avuti; è necessaria la variabile aggiuntiva

FRATELLI O SORELLE in cui si chiede al soggetto se ha o meno fratelli o sorelle perché, nel caso in cui non ne abbia, il soggetto non risponde alla variabile successiva, in questo modo si recupera l'informazione, se il soggetto non ha fratelli o sorelle.

Si costruisce quindi la variabile FRATELLI CLASSE: dove i soggetti con sei fratelli o più sono riuniti in un'unica classe, e la maggior parte delle analisi vengono fatte su questa variabile, o raggruppando ancora di più le classi, per raggiungere numerosità accettabili.

TITOLO DI STUDIO

La legenda è la seguente:

1	dottorato di ricerca o specializzazione post-laurea
2	Laurea
3	diploma universitario
4	diploma scuola media superiore (4-5 anni)
5	diploma scuola media superiore (2-3 anni)
6	licenza scuola media inferiore
7	licenza elementare
8	nessun titolo (sa leggere e scrivere)
9	nessun titolo (non sa leggere e/o scrivere)

Si decidono di raggruppare le diverse categorie come segue (per renderle paragonabili agli Stati Uniti):

- ◆ '0': Titolo basso (7, 8, 9)
- ◆ '1': Titolo medio (5, 6)
- ◆ '2': Titolo alto (1, 2, 3, 4)

Si crea quindi la variabile ISTRUZIONE con queste tre modalità.

REGIONE NASCITA

La variabile indica la regione di nascita del soggetto e si distribuisce come segue:

Tabella7.9: distribuzione della regione di nascita.

regione_nascita	Frequenza	Percentuale	Frequenza cumulata	Percentuale cumulata
.	306	3.04	306	3.04
0	301	2.99	607	6.04
1	1540	15.31	2147	21.35
2	430	4.28	2577	25.62
3	2058	20.46	4635	46.09
4	17	0.17	4652	46.26
5	232	2.31	4884	48.56
6	25	0.25	4909	48.81
7	518	5.15	5427	53.96
8	146	1.45	5573	55.41
9	107	1.06	5680	56.48
10	5	0.05	5685	56.53
11	20	0.20	5705	56.73
12	47	0.47	5752	57.19
13	42	0.42	5794	57.61
14	10	0.10	5804	57.71
15	188	1.87	5992	59.58
16	230	2.29	6222	61.87
17	61	0.61	6283	62.47
18	307	3.05	6590	65.53
19	280	2.78	6870	68.31
20	90	0.89	6960	69.21
99	3097	30.79	10057	100.00

La cui legenda è:

00	Estero
01	Piemonte
02	Valle d'Aosta
03	Lombardia
04	Trentino Alto Adige
05	Veneto
06	Friuli-V.G.
07	Liguria
08	Emilia-Romagna
09	Toscana
10	Umbria
11	Marche
12	Lazio
13	Abruzzo
14	Molise
15	Campania
16	Puglia
17	Basilicata
18	Calabria
19	Sicilia
20	Sardegna
99	stessa regione in cui risiede

A noi interessa avere una variabile, che chiamiamo PROVENIENZA, che si costruisce come segue:

- ◆ '1' = il soggetto è nato nel Nord-Ovest;
- ◆ '2' = il soggetto non è nato nel Nord-Ovest

Poiché ho, nel dataset attuale, solo soggetti residenti nel Nord-Ovest, associo le due categorie come segue:

- ◆ 1 → 99, 01, 02, 07, 03;
- ◆ 2 → rimanenti.

Quindi ottengo la distribuzione seguente.

Tabella7.10: distribuzione della provenienza.

provenienza	Frequenza	Percentuale	Frequenza cumulata	Percentuale cumulata
.	306	3.04	306	3.04
1	7643	76.00	7949	79.04
2	2108	20.96	10057	100.00

HA_FIGLI e FIGLI

Come per i fratelli ci serviamo della domanda filtro, cioè quella in cui si chiede al soggetto se ha figli (viventi, suoi o adottati), le risposte possibili sono:

- ◆ '1': no;
- ◆ '2': sì;

Creo una nuova variabile, detta FIGLI_OK, che vale esattamente il numero di figli, ed inoltre vale 0 se il soggetto ha risposto '1' alla prima domanda.

Quindi raggruppo le ultime tre classi nella categoria '6 o più', creando la variabile FIGLI_OK_CLASSE. Data la bassa numerosità di soggetti, è talvolta necessario utilizzare come classe superiore la '3 o più', e in quel caso si fanno alcuni adattamenti sulle percentuali riportate.

POSIZIONE_PROFESSIONE e POSIZIONE_PROFESSIONE_PADRE (quando soggetto aveva 14 anni)

La legenda della posizione professionale dei soggetti è:

01	dirigente
02	direttivo, quadro
03	insegnante di scuola media inferiore o superiore
04	insegnante di scuola materna o elementare
05	impiegato, intermedio
06	capo operaio, operaio subalterno e assimilati
07	apprendista
08	lavorante a domicilio per conto d'impresa
09	imprenditore
10	libero professionista
11	lavoratore in proprio
12	socio cooperativa produzione beni e/o prestazioni di servizio
13	coadiuvante
14	collaborazione coordinata e continuativa (co.co.co)
15	prestazione d'opera occasionale

Mentre la legenda per la posizione professionale dei padri dei soggetti è leggermente diversa:

01	dirigente
02	direttivo, quadro
03	insegnante di scuola media inferiore o superiore
04	insegnante di scuola materna o elementare
05	impiegato, intermedio
06	capo operaio, operaio subalterno e assimilati
07	apprendista
08	lavorante a domicilio per conto d'impresa
09	imprenditore
10	libero professionista
11	lavoratore in proprio
12	socio cooperativa produzione beni e/o prestazioni di servizio
13	coadiuvante
14	non so

Sia per il soggetto che per il padre, l'obiettivo è quello di creare le variabili PRESTIGIO e PRESTIGIO_PADRE, sulla base del livello sociale del lavoro svolto, per poter confrontare i risultati italiani con quelli statunitensi. Ciò viene fatto come segue.

PRESTIGIO (da POSIZIONE_PROFESSIONE):

- ♦ '1' basso → 6, 7, 8, 14, 15;
- ♦ '2' medio basso → 3, 4, 5;

- ♦ '3' medio alto→ 9, 11, 12, 13;
- ♦ '4' alto→ 1, 2, 10.

PRESTIGIO_PADRE (da POSIZIONE_PROFESSIONE_PADRE):

- ♦ '1' basso→ 6, 7, 8;
- ♦ '2' medio basso→ 3, 4, 5;
- ♦ '3' medio alto→ 9, 11, 12, 13;
- ♦ '4' alto→ 1, 2, 10;
- ♦ '14'→ '.'

DOVE VIVE MADRE e DOVE VIVE PADRE

Le due variabili servono per capire se il soggetto è figlio di immigrati o meno. Infatti se il soggetto è nato fuori dal Nord-Ovest d'Italia, e almeno uno dei suoi genitori vive a meno di 50 chilometri da lui, si decide che il soggetto è figlio di immigrati ed è immigrato insieme a loro da giovanissimo.

Le legende sono:

	Dove abita la madre
1	insieme a lei
2	in un altro appartamento dello stesso caseggiato
3	entro 1 km
4	nel resto del comune
5	meno di 16 km
6	da 16 a 50 km
7	più di 50 km
8	all'estero
9	è deceduta
	Dove abita il padre
1	insieme a lei
2	in un altro appartamento dello stesso caseggiato
3	entro 1 km
4	nel resto del comune
5	meno di 16 km
6	da 16 a 50 km
7	più di 50 km
8	all'estero
9	è deceduto

Mi creo una variabile DOVE VIVONO GENITORI, che si presenta come segue:

- ◆ ‘.’: dato non interessante (se soggetto ha risposto ‘9’ sia per il padre che per la madre, cioè entrambi i genitori sono deceduti) ;
- ◆ ‘2’: vivono nella stessa ripartizione (madre o padre rispondono 1, 2, 3, 4, 5, 6);
- ◆ ‘1’: non immigrati (restanti casi).

BIBLIOGRAFIA

PUBBLICAZIONI

Livi Bacci Massimo (2004), *"Introduzione alla demografia. Terza edizione"*;

Dalla Zuanna Giampiero, Rosina Alessandro, Rossi Fiorenzo (2004), *"Il Veneto - Storia della popolazione dalla caduta di Venezia ad oggi"* ;

Fabbris Luigi, *"L'indagine campionaria- Metodi, disegni e tecniche di campionamento"*;

Bonarini Franco, *"Guida alle Fonti Statistiche socio-demografiche"*;

Dalla Zuanna Giampiero, *"Population replacement, social mobility and development in Italy in the twentieth century"*, Journal of Modern Italian Studies;

Davis James, Tom W. Smith, Peter Marsde, *"General Social Survey, 1972-2002: Cumulative Codebook. Chicago: NORC, 2003"*;

Blake Judith, *"Number of Siblings and Educational Mobility"*, American Sociological Review vol.50, No.1, 1985, pp. 84-94;

Borjas George J., *"The Economics of Immigration"*, Journal of Economic Literature Vol. XXXII (December 1994), pp. 1667-1717;

Coleman David, Rowthorn Robert, *"The Economic Effects of Immigration into the United Kingdom"*, Population and Development Review 30(4): 579-624 (December 2004);

Venturini Alessandra, Villosio Claudia, *"Are Immigrants Competing with Natives in the Italian Labour Market? The Employment Effect"*, IZA DP No. 467;

Fernández Raquel, Fogli Alessandra, *"Culture: An Empirical Investigation of Beliefs, Work, and Fertility"*. Revised January 2007;

Fernández Raquel, *"Women, Work, and Culture"*, November 2006;

Mare Robert D., Maralani Vida, *"The Intergenerational Effects of Changes in Women's Educational Attainments"*;

University of Chicago, The Roper Center for Public Opinion Research University of Connecticut, *"General Social Surveys, 1972-2006: Cumulative Codebook"*, March 2007, Conducted for The National Data Program for the Social Sciences at National Opinion Research Center;

US Department of Justice Immigration and Naturalization Service, *"Statistical Yearbook of the Immigration and Naturalization Service 2001"*, US Government Printing Office, Washington, DC, forthcoming;

Pace Luigi, Salvan Alessandra, *"Introduzione alla Statistica. II Inferenza, verosimiglianza, modelli"*;

ISTAT, *"Il sistema di indagini sociali multiscopo. Contenuti e metodologia delle indagini"*.

SITI INTERNET

[http://gss.norc.org/;](http://gss.norc.org/)

[http://www.ropercenter.uconn.edu/data_access/data/datasets/general_social_survey.html;](http://www.ropercenter.uconn.edu/data_access/data/datasets/general_social_survey.html)

[http://www.icpsr.umich.edu/cocoon/ICPSR/STUDY/04697.xml;](http://www.icpsr.umich.edu/cocoon/ICPSR/STUDY/04697.xml)

[http://www.wordreference.com/;](http://www.wordreference.com/)

<http://www.migrationinformation.org/DataHub/charts/final.immigbyyear.shtml>

<http://www.migrationinformation.org/USfocus/>

http://www.migrationpolicy.org/pubs/fact_sheets.php

<http://www.migrationinformation.org/datahub/charts/final.fb.shtml>

<http://www.istat.it/>

<http://www.istat.it/strumenti/metodi/notemet/NM154.htm>

<http://www.ais-sociologia.it/multiscopo.php>