

# Indice

<b>1</b>	<b>Introduzione</b>	<b>3</b>
1.1	Statistiche introduttive . . . . .	4
1.2	Breve sintesi del lavoro . . . . .	7
<b>2</b>	<b>I dati</b>	<b>9</b>
2.1	I dati Banca d'Italia . . . . .	9
2.2	Descrizione dei sottocampioni . . . . .	10
2.3	Creazione dei sottocampioni . . . . .	10
2.4	Pesi . . . . .	11
2.5	Variabili . . . . .	11
<b>3</b>	<b>Popolazione e sottocampioni</b>	<b>15</b>
3.1	Un'analisi descrittiva . . . . .	15
3.2	Composizione e area . . . . .	15
3.3	Caratteristiche del capofamiglia . . . . .	17
3.4	Variabili di reddito . . . . .	19
3.4.1	Reddito familiare . . . . .	21
3.4.2	Differenziazione del reddito . . . . .	24
3.5	Ricchezza familiare e consumi . . . . .	25
<b>4</b>	<b>Confronto in base al titolo di studio del capofamiglia</b>	<b>29</b>
4.1	Capofamiglia laureato o con specializzazione post-laurea . . . . .	30
4.2	Capofamiglia diplomato . . . . .	30
4.3	Capofamiglia con licenza di terza media . . . . .	33

<b>5</b>	<b>Il modello statistico</b>	<b>37</b>
5.1	La scelta del modello . . . . .	37
5.2	Le variabili incluse nel modello . . . . .	37
5.3	Il modello stimato . . . . .	39
5.4	Interpretazione del modello . . . . .	41
<b>6</b>	<b>Conclusioni</b>	<b>45</b>
<b>A</b>	<b>Il programma SAS</b>	<b>49</b>
A.1	La lettura dei dati . . . . .	49
A.2	Variabili . . . . .	54
A.2.1	Creazione delle nuove variabili . . . . .	54
A.2.2	Creazione delle dummies per il modello logit . . . . .	60
A.2.3	Creazione delle classi per il reddito familiare . . . . .	62
A.2.4	Trasformazione a prezzi del 2004 . . . . .	62
A.2.5	Creazione dei redditi familiari equivalenti . . . . .	64
A.2.6	Sottocampioni e popolazione . . . . .	64
<b>B</b>	<b>Le variabili utilizzate</b>	<b>69</b>
<b>C</b>	<b>I modelli stimati</b>	<b>77</b>

# Capitolo 1

## Introduzione

Senza dubbio, uno degli aspetti più discussi e costantemente al centro del dibattito politico è quello riguardante l'istruzione. Negli ultimi anni, infatti, si è molto spesso sentito parlare di riforma universitaria e del sistema scolastico, di nuovi criteri da introdurre nell'insegnamento e di nuovi stimoli e input che possano migliorare l'educazione dei giovani.

Basta guardare alle statistiche per capire come alla base di un paese sviluppato economicamente e all'avanguardia dal punto di vista tecnologico, ci sia una grande partecipazione universitaria, o, più in generale, un livello di istruzione medio piuttosto elevato. Nella letteratura economica infatti, non sono rari studi che mostrino come una maggiore diffusione di elevati titoli di studio tra la popolazione, portino ad una crescita generale del paese, ad un miglioramento delle condizioni di salute dei cittadini, ed ad altre conseguenze, solitamente positive, sulla società.

Data quindi l'importanza che un elevato livello di istruzione ricopre all'interno della società moderna, è opportuno individuare quali siano i fattori che maggiormente incidono sulla decisione di proseguire gli studi dopo il conseguimento del diploma.

Numerose variabili infatti, influenzano la scelta di iscriversi all'università: reddito e ricchezza familiare, livello di istruzione dei genitori, distanza della propria città di residenza da un Ateneo e carriera scolastica precedente, sono solo alcuni esempi degli elementi da considerare nella decisione di proseguire gli studi. Questo lavoro dunque, si pone come obiettivo principale quello di individuare la popola-

zione a rischio di ingresso all'università e di analizzare i fattori che determinano tale decisione.

## 1.1 Statistiche introduttive

Come emerge dalle ricerche sul sistema universitario e scolastico, l'Italia costituisce uno dei paesi con il più basso livello di istruzione tra le nazioni sviluppate. La tabella 1.1 infatti, mostra come le percentuali di frequenza universitaria siano più elevate nelle altre nazioni Europee, in Giappone e negli Stati Uniti.

	scuola elementare	scuola media	scuola professionale	scuola superiore	laurea triennale (3 anni)	laurea specialistica (2 anni)
Francia	18	18	31	10	11	12
Germania	2	16	52	3	15	13
Italia	22	33	8	25	2	10
Giappone	-	17	-	49	15	19
Regno Unito	-	17	42	15	8	18
Stati Uniti	5	8	-	50	9	28

Tabella 1.1: Titolo di studio conseguito dalla popolazione di età compresa tra i 25 e i 64 anni, frequenze percentuali.[Checchi, 2003]

Un'altra differenza non irrilevante che si registra tra l'Italia e gli altri paesi, è quella riguardante le risorse pubbliche investite nell'educazione. Nell'ambito dell'istruzione secondaria, questa differenza non è troppo evidente: l'investimento del governo italiano si pone a metà via tra le risorse messe a disposizione dal Regno Unito e dagli Stati Uniti (più elevate) da un lato e quelle di Francia, Germania e Giappone (più scarse) dall'altro. Una nota negativa per l'Italia invece, si registra quando si passa all'analisi dei fondi pubblici per l'istruzione universitaria: utilizzando come approssimazione di questa variabile il numero di studenti per insegnante, il valore assunto dal rapporto è di 22.8 per le università italiane, ben al di sopra della media degli altri paesi considerati per il confronto (da 12,1 per la Germania a 18,3 per la Francia). Guardando ai dati che si riferiscono direttamen-

te alla spesa effettuata, questa differenza appare ancora più evidente anche se gli investimenti di Italia e Francia non sono troppo distanti tra loro. Vedi tabella 1.2.

	asili	scuola elementare	scuola superiore	università
Francia	3901	4139	7152	7876
Germania	4937	3818	6603	10393
Italia	5133	5354	6518	7552
Giappone	3154	5240	6039	10278
Regno Unito	6233	3627	5608	9554
Stati Uniti	6692	6582	8157	19220

Tabella 1.2: Spesa per studente in base al livello di istruzione (1999)- dollari. [Checchi, 2003]

Tuttavia, se da un lato il confronto con gli altri paesi del mondo occidentale risulta negativo per l'Italia, dall'altro l'evoluzione all'interno della nazione nel decennio 1991-2001 è piuttosto positiva. Dai dati Istat relativi al censimento del 2001, emerge infatti che, sia in termini assoluti che percentuali, si sia verificato un aumento delle persone in possesso dei titoli di studio più elevati ed in particolare modo dei diplomi di laurea, diplomi universitari e dei diplomi di scuola secondaria superiore (tabella 1.3). Parte di questi cambiamenti possono comunque essere attribuibili a numerosi corsi di diploma universitari introdotti tra i due anni.

Un altro importante cambiamento verificatosi nel corso del decennio, è quello riguardante le differenze di genere. Soprattutto nella fascia dei titoli di studio più elevati si nota una diminuzione della diversità nel livello di istruzione raggiunto tra uomini e donne. Il dato più sorprendente, è quello riguardante i diplomi di laurea: se nel 1991 solamente il 42,2% dei laureati erano donne, questa percentuale sale al 49,1 nel 2001. Anche il numero totale dei laureati sale da 2.047.630 a 3.480.535 . Il genere dunque, non può più essere considerato come una variabile discriminante per l'ingresso all'università.

Altri dati interessanti per una generica e introduttiva analisi del problema, sono forniti dalla prima indagine sulle spese sostenute dalle famiglie per l'istruzione e

Grado di istruzione	1991			2001		
	M	F	Totale	M	F	Totale
Diploma di laurea	1.183.480	864.150	2.047.630	1.770.635	1.709.900	3.480.535
%	57,8	42,2	100	50,9	49,1	100
Diploma universitario	82.725	130.942	213.667	197.270	364.454	561.724
%	38,7	61,3	100	35,1	64,9	100
Diploma di maturità	4.219.160	3.890.344	8.109.504	5.828.696	5.621.698	11.450.394
%	52	48	100	50,9	49,1	100

Tabella 1.3: Popolazione residente di età superiore a 11 anni classificata in base al grado di istruzione e al sesso. [Istat, L'istruzione della popolazione al 2001, 2005]

la formazione, nata nell'ambito della convenzione tra l'Istat e il Ministero per l'Università e la Ricerca. Tale indagine, condotta contemporaneamente a quella sui consumi delle famiglie nel 2002, è stata realizzata tramite un questionario aggiuntivo che 9.940 famiglie, con almeno un componente iscritto ad un corso di istruzione o formazione, hanno compilato per un totale di 14.665 questionari individuali. Le spese dichiarate da una famiglia sono poi state classificate in due categorie principali: da un lato le “spese dirette” che comprendono :

- iscrizione;
- mensa e scuolabus;
- corsi a pagamento organizzati dalle scuole;

dall'altro invece, sono state rilevate le “spese esterne” che includono:

- libri, cancelleria, abbigliamento scolastico;
- lezioni private e materiale tecnico;
- gite scolastiche e visite guidate;
- trasporti;
- pasti consumati fuori casa e alloggio.

La spesa media annuale dichiarata da ogni famiglia ammonta a circa 1.009€ annuali, due terzi dei quali per la “spesa esterna”. L’impegno economico delle famiglie aumenta passando dalla scuola dell’obbligo alla superiore e all’università. Guardando in particolar modo alla composizione della spesa per un iscritto all’università, la spesa media dichiarata ammonta a circa 2.311€ l’anno, dei quali 793€ per i costi diretti e i restanti (poco più di 1500€ ) per i costi esterni. Naturalmente per gli studenti fuori sede le spese sono al di sopra di questa media calcolata sull’intero campione.

Attraverso questa breve introduzione, si è quindi visto come, sebbene l’Italia mostri ancora un livello di istruzione medio e di investimento per l’istruzione al di sotto di quello degli altri paesi più sviluppati ( in particolar modo Francia, Regno Unito, Usa, Giappone), la situazione ha registrato un’evoluzione positiva all’interno della nazione con un aumento del numero di iscritti ad un corso di laurea o, più in generale, di formazione. Si sono analizzate infine le spese sostenute dalle famiglie per l’istruzione, concentrandosi in particolar modo sui costi per il mantenimento dei figli iscritti all’università: per alcune fasce di reddito questa spesa potrebbe infatti rivelarsi troppo elevata così da impedire il proseguimento degli studi dopo il diploma.

## 1.2 Breve sintesi del lavoro

I capitoli successivi presenteranno l’analisi di alcune variabili ritenute influenti nel processo di decisione del titolo di studio da conseguire. In particolar modo, saranno ricavati dai dati a disposizione ( dati forniti dalla Banca d’Italia - Indagine sui Bilanci delle Famiglie Italiane) due sottocampioni principali: le famiglie con figli di età compresa tra i 18 e i 22 anni e già occupati o in cerca di prima occupazione e le famiglie con figli nella stessa fascia d’età, ma ancora studenti e quindi iscritti all’università. Questi due sottocampioni saranno poi confrontati in base a delle caratteristiche fondamentali come reddito e ricchezza familiare, titolo di studio del capofamiglia e sua qualifica professionale e altre variabili rilevanti nella scelta di proseguire gli studi dopo il diploma. Le variabili la cui distribuzione è piuttosto

differente tra i due sottocampioni, probabilmente hanno un peso maggiore rispetto a quelle che invece si distribuiscono in modo omogeneo nei due insiemi.

Successivamente le famiglie saranno ulteriormente divise in base al livello di istruzione del capofamiglia (laurea o specializzazione post-laurea, diploma, terza media): per questi sottogruppi, ci si soffermerà sull'analisi delle differenze tra i redditi cercando anche di capire le cause di eventuali divergenze nelle distribuzioni.

Infine, sulla base dei risultati precedenti, verrà stimato un modello logit che predica la probabilità di iscriversi all'università in base alle caratteristiche familiari maggiormente influenti.



# Capitolo 2

## I dati

### 2.1 I dati Banca d'Italia

Lo studio delle variabili coinvolte nella decisione di proseguire gli studi dopo il conseguimento del diploma, è basato sull'analisi di dati tratti dall'Indagine della Banca d'Italia sui Bilanci delle famiglie italiane.

Il questionario viene distribuito a circa 8.000 famiglie selezionate casualmente dalle liste anagrafiche di 344 comuni sparsi nel territorio italiano, per un totale di circa 20.000 individui. Le famiglie fanno parte di un piano di rotazione tale per cui un sottoinsieme di queste viene lasciato nel campione per essere contattato anche durante l'indagine successiva, la parte restante viene invece sostituita. Ogni famiglia può dunque essere intervistata per una o più indagini consecutive: in tal modo si ottengono dei dati panel che permettono di evidenziare eventuali trend o particolari evoluzioni nella distribuzione di alcune variabili.

L'indagine, ripetuta ogni due anni sulla popolazione italiana, ha lo scopo di rilevare le principali caratteristiche delle famiglie e, in particolar modo, gli aspetti riguardanti la struttura delle famiglie e le caratteristiche dei componenti, il reddito, il consumo e il risparmio, la ricchezza.

In accordo con il disegno campionario, a ogni famiglia viene attribuito un peso inversamente proporzionale alla sua probabilità di inclusione nel campione; tali pesi vengono successivamente modificati in modo da tener conto delle mancate

risposte, da aumentare la precisione degli stimatori e, soprattutto, per allineare la struttura del campione a quella della popolazione in base a caratteristiche note.

## 2.2 Descrizione dei sottocampioni

Come affermato nel capitolo introduttivo, lo scopo di tale lavoro è quello di capire quali siano i fattori che incidono sulla decisione di iscriversi all'università. Per condurre le analisi sono quindi stati estratti dall'intero campione, due sottoinsiemi tra loro confrontabili, dai quali trarre delle conclusioni estendibili alla popolazione totale. La scelta più adeguata per delle analisi opportune, sembrava essere costituita dalle famiglie con figli *diplomati*, di età compresa tra i 18 e i 22 anni, divise in modo tale da avere da un lato quelle con figli non più studenti ma già occupati o in cerca di prima occupazione o disoccupati (*Sottocampione1*), dall'altro quelle con componenti iscritti all'università (*Sottocampione2*).

## 2.3 Creazione dei sottocampioni

Con l'obiettivo di ottenere una popolazione sufficientemente elevata dalla quale ricavare risultati attendibili, si è deciso di costruire il campione iniziale attraverso il pooling delle famiglie intervistate negli anni 2000, 2002 e 2004. Come anticipato nei paragrafi precedenti, i dati Banca d'Italia sono dati panel: ciò implica che, se una stessa famiglia viene intervistata più volte nell'arco di tempo considerato (2000-2004), potrebbe essere inclusa nel campione selezionato ripetutamente distorcendo i risultati. Con lo scopo di includere solo una volta le famiglie con le caratteristiche descritte e di delineare una versione dei dati il più possibile aggiornata, è stato adottato il criterio di selezione "*fresh*". Ossia, per le famiglie più volte incluse nel campione, sono state considerate solamente le risposte fornite nell'ultimo anno di inclusione nel panel.

Il campione così costruito è costituito da 16815 unità. All'interno di questo generico insieme, è inoltre possibile distinguere le famiglie a rischio individuate nel paragrafo precedente:

1. quelle con figli che dopo aver conseguito il diploma non proseguono gli studi  
- *Sottocampione1*;
2. quelle con figli che si iscrivono all'università - *Sottocampione2*.

In tal modo, si sono ottenuti dei sottocampioni con una numerosità sufficiente per condurre le analisi: il primo insieme è infatti costituito da 352 famiglie, il secondo da 1043 famiglie.

## 2.4 Pesì

Come già affermato in precedenza, l'Indagine della Banca d'Italia si basa su un campione di circa 8000 famiglie tratto dall'intera popolazione italiana. Dunque, per poter estendere le caratteristiche di questo campione a tutta la popolazione e ottenere delle stime non distorte, è necessario abbinare a ciascuna famiglia un proprio **peso specifico**. Le analisi successive, quindi, saranno basate su questo sistema di pesi, opportunamente riscalato in modo da tener conto della nuova numerosità del campione selezionato.

I pesi esistenti sono stati moltiplicati per 8012, ossia la numerosità della popolazione rilevata nel 2004, e successivamente divisi per 16783, ossia il numero totale degli individui appartenenti al campione finale.

## 2.5 Variabili

Per poter condurre in modo efficace le analisi successive, le variabili iniziali fornite dalla Banca d'Italia, sono state opportunamente modificate: i dati individuali sono stati trasformati in aggregati familiari contenenti le principali caratteristiche sia di ciascun componente, che dell'intera famiglia. In special modo, si è data importanza alle informazioni riguardanti il titolo di studio del capofamiglia e del suo coniuge, al tipo di lavoro e di settore lavorativo di ciascun componente, alle componenti del reddito e della ricchezza e ad altri fattori come il numero di depositi bancari o di carte di credito intestate. Riassumendo si può dire che questo

processo di trasformazione delle variabili, ha come obiettivo principale quello di focalizzare l'attenzione sui fattori che maggiormente caratterizzano una famiglia e che possano essere, direttamente o indirettamente, coinvolti nella decisione di iscriversi o meno all'università.

Poichè il campione analizzato è costituito dal pooling degli anni 2000, 2002 e 2004, i redditi e le altre variabili sul livello di ricchezza, sono state opportunamente moltiplicati per il *deflatore implicito* in modo che rispecchiassero i prezzi correnti del 2004. La struttura del deflatore implicito per trasformare un aggregato dai prezzi dell'anno 0 a quelli dell'anno  $t$ , è data da:

$${}_0D_t = \frac{{}_tX_t}{{}_0X_t} = \frac{\sum_k p_{kt}q_{kt}}{\sum_k p_{k0}q_{kt}}$$

Dove  ${}_tX_t$  rappresenta il Pil dell'anno  $t$  a prezzi correnti e  ${}_0X_t$  rappresenta il Pil dell'anno  $t$  a prezzi dell'anno 0.

Dunque, per l'anno 2000 il deflatore implicito è costituito da:

$${}_{00}D_{04} = \frac{{}_{04}X_{04}}{{}_{00}X_{04}}$$

con

$${}_{00}X_{04} = \frac{{}_{95}X_{04}}{{}_{95}X_{00}}.$$

Si è fatto riferimento all'anno 1995 in quanto l'Istat fornisce gli aggregati a prezzi costanti utilizzando come base tale anno.

Per il 2002 si ha invece:

$${}_{02}D_{04} = \frac{{}_{04}X_{04}}{{}_{02}X_{04}}$$

con

$${}_{02}X_{04} = \frac{{}_{95}X_{04}}{{}_{95}X_{02}}.$$

Si sono quindi ottenuti i seguenti deflatori:

$${}_{00}D_{04} = 1.117$$

$${}_{02}D_{04} = 1.056$$

Le variabili di reddito riferite al 2000 e al 2002, sono state moltiplicate per il corrispettivo deflatore in modo da trasformarle ai prezzi del 2004 e renderle confrontabili tra loro.

Un'ulteriore trasformazione delle variabili di reddito, è quella applicata a “yfam”, il reddito familiare. Un dato livello di reddito infatti, incide diversamente su ciascuna famiglia in base al numero dei suoi componenti. Ad esempio un reddito mensile di 2.000€, ha un diverso peso in una famiglia di due soli componenti piuttosto che in una famiglia formata da quattro individui. Per tener conto di tale disparità, si è ricorso alla Scala di equivalenza di Carbonaro: questa tabella fornisce dei coefficienti, calcolati in base al numero di elementi di ciascuna famiglia, per i quali dividere i redditi familiari ottenendo così i *redditi familiari equivalenti*.

Numero di componenti	Coefficiente
1	0,599
2	1
3	1,335
4	1,632
5	1,905
6	2,150
$\geq 7$	2,401

Tabella 2.1: Scala Carbonaro

In gran parte delle analisi descritte nei prossimi capitoli, la variabile “yfam” sarà spesso sostituita da “yfam\_equiv”: ricorrendo ai redditi familiari equivalenti, il confronto tra famiglie con diverso numero di componenti sarà più oggettivo e affidabile poichè annulla o riduce significativamente tale differenza.



## Capitolo 3

# Popolazione e sottocampioni

### 3.1 Un'analisi descrittiva

Prima di passare all'analisi vera e propria dei fattori che incidono sulla decisione di iscriversi all'università di un neodiplomato, si è ritenuto opportuno delineare le principali caratteristiche dei tre campioni analizzati: la popolazione totale, il primo sottocampione (*Sottocampione1*) e il secondo sottocampione (*Sottocampione2*).

Attraverso delle semplici ma efficaci tecniche di statistica descrittiva, si sono delineati gli aspetti più salienti delle famiglie analizzate. I risultati sono stati quindi confrontati tra loro così da ottenere una visione introduttiva e generale dei dati analizzati.

In particolar modo l'attenzione è stata focalizzata su variabili come il numero di componenti della famiglia, il numero di figli, l'area geografica di residenza, la classe d'età del capofamiglia, il suo titolo di studio, di qualifica lavorativa e settore di lavoro, il reddito e la ricchezza.

### 3.2 Composizione e area

Come è possibile vedere dalle tabelle (devo aggiungere il riferimento all'appendice finale) le famiglie appartenenti alla popolazione totale, sono costituite in

maggioranza da nuclei di una, due, tre o quattro persone al massimo (tabella 3.1).

Osservando i dati sui due sottocampioni (tabella 3.1), si può notare invece che quasi il 47% delle famiglie è composto da quattro elementi. Questa differenza tra la popolazione e i due sottocampioni, è riconducibile al fatto che in tutte le famiglie del *Sottocampione1* e del *Sottocampione2* è presente almeno un componente di età compresa tra i 18 e i 22 anni.

Quanto appena affermato, è confermato dall'analisi del numero dei figli: tra la popolazione totale infatti, più del 50% delle famiglie non ha alcun figlio a differenza delle unità dei due sottocampioni che, invece, nel 50% dei casi, hanno due figli. Nella popolazione totale si devono considerare anche le coppie di anziani che vivono soli e senza figli.

Numero di componenti	<i>Popolazione</i>		<i>Sottocampione1</i>		<i>Sottocampione2</i>	
	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>
1	3.862	22,97	8	2,27	9	0,86
2	4.906	29,18	33	9,38	37	3,55
3	3.600	21,41	127	36,08	201	19,27
4	3.229	19,20	135	38,35	516	49,47
5	942	5,60	40	11,36	210	20,13
6	213	1,27	9	2,56	58	5,56
$\geq 7$	63	0,37	0	0,00	12	1,05
<i>Totale</i>	16.815	100,00	352	100,00	1.043	100,00

Tabella 3.1: Tabella dell'area geografica di residenza di ciascuna famiglia.

Infine, per quanto riguarda la regione di residenza, non si sono registrate evidenti differenze tra i tra gruppi (tabella ??): si ha infatti un' omogenea distribuzione delle famiglie nel territorio. Poichè non emergono divergenze neanche tra il *Sotto-*



*campione1* e il *Sottocampione2*, la variabile “\_area”, probabilmente, non sarà un fattore incidente nella decisione di proseguire gli studi dopo il diploma.

Area	Popolazione		Sottocampione1		Sottocampione2	
	Frequenza	%	Frequenza	%	Frequenza	%
Nordovest	4.279	25,45	87	24,72	216	20,71
Nordest	3.329	19,80	74	21,02	189	18,12
Centro	3.488	20,74	75	21,31	202	19,37
Sud	3.823	22,74	87	24,72	305	29,24
Isole	1.896	11,28	29	8,24	131	12,56
Totale	16.815	100,00	352	100,00	1.043	100,00

Tabella 3.2: Tabella dell'area geografica di residenza di ciascuna famiglia.

### 3.3 Caratteristiche del capofamiglia

In primo luogo, si è ritenuto interessante mostrare le differenze d'età del capofamiglia nei diversi gruppi.

Dalle analisi è emerso che, per la popolazione totale, circa il 30% dei capifamiglia ha un'età compresa tra i 51 e i 65 anni e un'altra identica porzione ha un'età superiore ai 65 anni. Per le famiglie del *Sottocampione1* e del *Sottocampione2* la maggior parte dei capifamiglia ha un'età compresa tra i 41 e i 65 anni (tabella 3.3).

Classe di età (anni)	<i>Popolazione</i>		<i>Sottocampione1</i>		<i>Sottocampione2</i>	
	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>
fino a 30	778	4,63	37	10,51	32	3,07
da 31 a 40	2.562	15,24	8	2,27	24	2,30
da 41 a 50	3.202	19,04	109	30,97	501	48,03
da 51 a 65	4.961	29,50	187	53,13	468	44,87
oltre 65	5.312	31,59	11	3,13	18	1,73
Totale	16.185	100,00	352	100,00	1.043	100,00

Tabella 3.3: Tabella della classe del capofamiglia.

Dato lo scopo finale del lavoro, si è ritenuto importante analizzare anche il **titolo di studio** del capofamiglia: in Italia infatti, il condizionamento dei genitori e il loro livello di istruzione rappresentano variabili fortemente influenti nella carriera scolastica e accademica dei loro figli.

Per la popolazione totale, è emerso che il 90% dei capifamiglia ha un titolo di studio inferiore o pari al diploma. Le loro qualifiche si distribuiscono in modo pressochè omogeneo tra completamento delle elementari(30%), della scuola media (27%) e delle superiori (28%).

Una situazione simile si ha per le famiglie del *Sottocampione1*: solamente il 3% dei capifamiglia è laureato. Un quadro piuttosto differente emerge invece dalle analisi dei capifamiglia del *Sottocampione2*: ben il 14% ha raggiunto la laurea e una parte molto consistente è diplomata (43%). Ciò implica che, in media, ragazzi con genitori istruiti hanno un maggiore sostegno nel proseguire gli studi (tabella 3.4).

Un'altra caratteristica considerata, è la **qualifica professionale**. Osservando i dati riguardanti la popolazione (tabella 3.5), emerge che:

- il 33% dei capifamiglia sono occupati come operai o posizioni simili;
- il 32% invece è impiegato o insegnante in qualsiasi tipo di scuola.

Titolo di studio	Popolazione		Sottocampione1		Sottocampione2	
	Frequenza	%	Frequenza	%	Frequenza	%
nessun	1.173	6,98	6	1,70	9	0,86
elementari	5.007	29,78	95	26,99	119	11,41
medie	4.501	26,77	121	34,38	318	30,49
diploma	4.727	28,11	119	33,81	447	42,86
laurea	1.367	8,13	11	3,13	145	13,90
specializ.	40	0,24	0	0,00	5	0,48
Totale	16.185	100,00	352	100,00	1.043	100,00

Tabella 3.4: Tabella della classe del capofamiglia.

Passando ai risultati del *Sottocampione1* e del *Sottocampione2*, si possono notare delle differenze anche tra i due gruppi (tabella 3.5).

**Sottocampione1:** i genitori dei ragazzi che hanno deciso di non proseguire gli studi, sono generalmente operai (37%), impiegati o insegnanti (29%) anche se si deve sottolineare la presenza di lavoratori in proprio (13%), di imprenditori (6%) e di impiegati-dirigenti(5.6%).

**Sottocampione2:** in questo insieme, invece, si ha una netta diminuzione dei capifamiglia occupati come operai (la percentuale scende al 26%) e un aumento degli impiegati o insegnanti (40%); la distribuzione delle altre qualifiche lavorative è praticamente invariata.

Le differenze appena descritte, contribuiscono alla formazione di un ambiente familiare più o meno stimolante per i figli che, senza dubbio, ne vengono influenzati. Questi fattori, anche se in modo indiretto, incidono sui giovani nel momento della scelta dell'iscrizione all'università.

### 3.4 Variabili di reddito

Alcune delle variabili più importanti rilevate dall'Indagine sui Bilanci delle Famiglie italiane, sono logicamente, quelle che descrivono i **redditi familiari**. In

Qualifica profession.	<i>Popolazione</i>		<i>Sottocampione1</i>		<i>Sottocampione2</i>	
	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>
operaio	2.498	33,16	80	36,70	165	22,21
impiegato, insegnante	2.471	32,80	63	28,90	291	39,17
impiegato direttivo	439	5,83	12	5,50	56	7,54
dirigente	216	2,87	4	1,83	30	4,04
libero profession.	446	5,92	5	2,29	53	7,13
imprendit.	371	4,92	13	5,96	31	4,17
in proprio	879	11,67	28	12,84	92	12,38
coadiuvante	214	2,84	13	5,96	25	3,36
Totale	16.185	100,00	218	100,00	743	100,00

Tabella 3.5: Tabella della qualifica professionale del capofamiglia.

particolar modo, i dati forniti dalla Banca d'Italia, permettono la classificazione di tali redditi in base alla loro origine.

Poichè si tratta di variabili continue, questa prima analisi si basa sul confronto delle fondamentali misure di posizione come media e mediana, e sugli indici di variabilità.

Per le famiglie appartenenti al *Sottocampione1* e al *Sottocampione2*, è stata condotta anche un'*analisi della varianza* (ANOVA) per capire se le differenze tra questi due insiemi fossero significative o meno. Il reddito potrebbe infatti rive-

larsi una delle variabili che determinano la decisione di iscriversi all'università: il mantenimento dei figli all'università, come sottolineato nel capitolo introduttivo, implica infatti un impegno economico che, forse, non tutte le famiglie possono sostenere.

### 3.4.1 Reddito familiare

La prima variabile di reddito analizzata è la più generale, costituita dalla somma di tutte le variabili di reddito analizzate di seguito. Inoltre si è tenuto conto della numerosità familiare: si sono quindi utilizzati i redditi familiari *equivalenti*.

Inaspettatamente, il reddito che risulta più elevato è quello corrispondente al *Sottocampione1*. Tuttavia, l'analisi della varianza su questa variabile, indica che questa differenza non è statisticamente significativa.

	<b>Popolazione</b>	<b><i>Sottocampione1</i></b>	<b><i>Sottocampione2</i></b>
<b>Media</b>	32.451,58	33.009,54	30.230,10
<b>Mediana</b>	25.528,36	26.970,36	24.286,36
<b>Deviazione std</b>	32.219	24.485	24.335

Tabella 3.6: Statistiche di base sui redditi familiari *equivalenti*; valori espressi in Euro.

Di seguito sono riportati i boxplots dei redditi familiari equivalenti e gli istogrammi delle classi di reddito per la popolazione totale e per i due *Sottocampioni*.

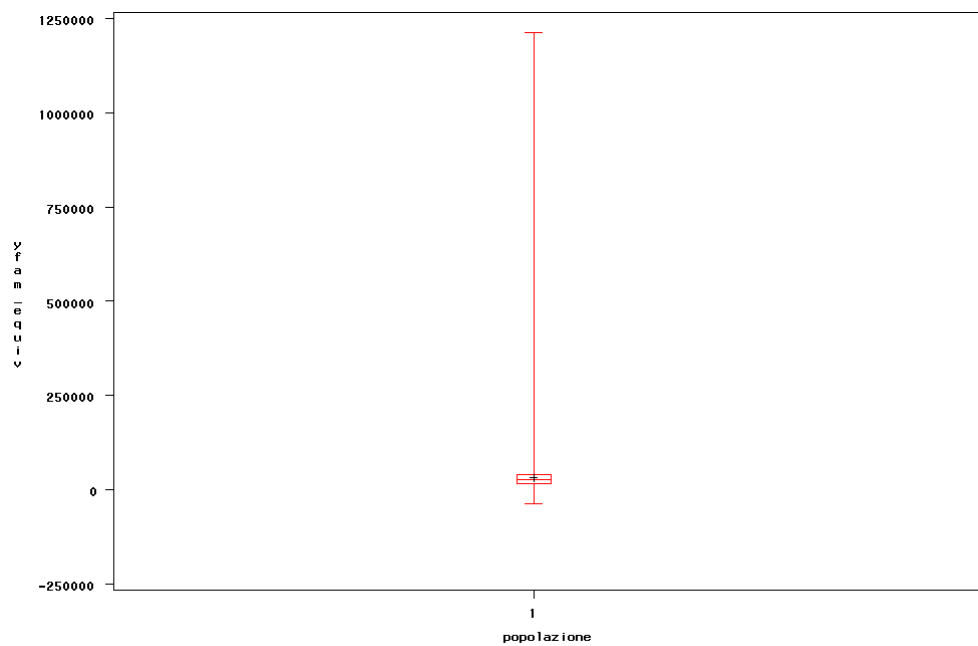


Figura 3.1: Boxplot del reddito familiare equivalente per la popolazione totale.

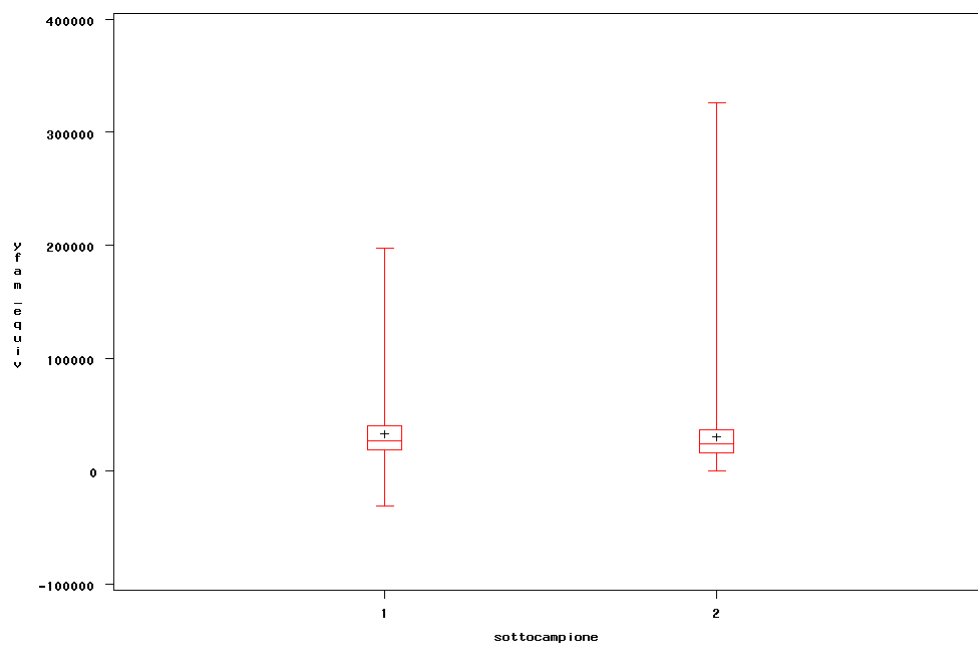


Figura 3.2: Boxplot del reddito familiare equivalente per i due Sottocampioni.

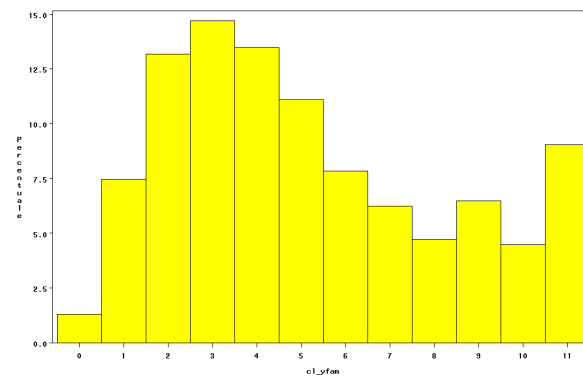


Figura 3.3: Istogramma delle classi di reddito della popolazione totale.

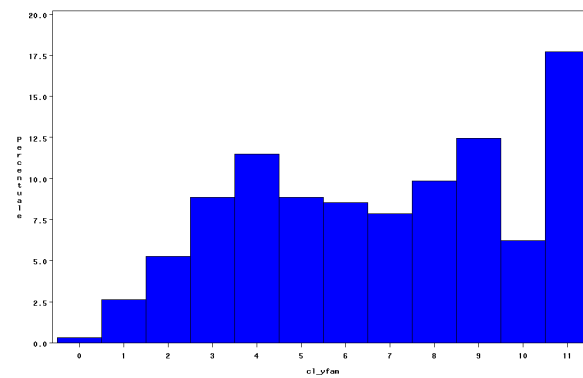


Figura 3.4: Istogramma delle classi di reddito delle famiglie del Sottocampione1.

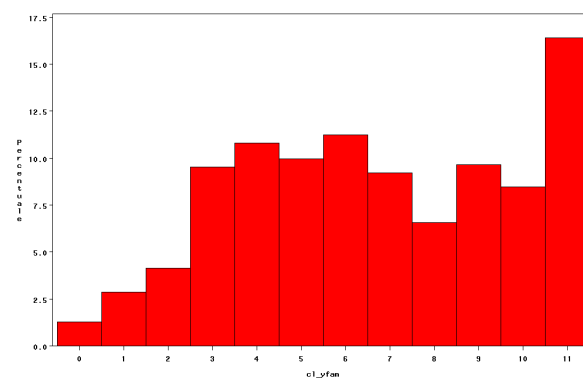


Figura 3.5: Istogramma delle classi di reddito delle famiglie del Sottocampione2.

### 3.4.2 Differenziazione del reddito

Nei prossimi paragrafi sono analizzati: il reddito da lavoro dipendente, il reddito da lavoro autonomo, il reddito da trasferimenti e il reddito da capitale. Le medie sono state calcolate sulla parte di campione per il quale la variabile di interesse registra un valore positivo.

#### Reddito da lavoro dipendente

Il primo tipo di reddito analizzato è quello che ha origine da lavoro dipendente. Il valore più elevato per la media si registra per il *Sottocampione2*; la differenza tra i due sottogruppi non è però significativa.

	Popolazione	<i>Sottocampione1</i>	<i>Sottocampione2</i>
<b>Media</b>	26.944,67	29.478,3	31.372,20
<b>Mediana</b>	22.000	26.083,20	26.249,75
<b>Deviazione std</b>	20.089	21.584	22.633

Tabella 3.7: Statistiche di base sui redditi familiari da lavoro dipendente; valori espressi in Euro.

#### Reddito da lavoro autonomo

Anche in questo caso, è il sottocampione delle famiglie con figli iscritti all'università a registrare il valore più elevato, ma il test ANOVA indica che tale differenza non è significativa (tabella 3.8).

#### Reddito da trasferimenti

Poichè i redditi da trasferimenti includono anche le pensioni, il valore più elevato si registra per la popolazione totale. Tra i due sottocampioni è il *Sottocampione1* ad avere media e mediana più elevate (tabella 3.9).



	Popolazione	<i>Sottocampione1</i>	<i>Sottocampione2</i>
<b>Media</b>	30.589	31.679	33.291,40
<b>Mediana</b>	20.106	20.000	23.000
<b>Deviazione std</b>	44.657	37.898	38.225

Tabella 3.8: statistiche di base sui redditi familiari da lavoro autonomo; valori espressi in Euro.

	Popolazione	<i>Sottocampione1</i>	<i>Sottocampione2</i>
<b>Media</b>	16.471,23	15.701,30	14.494,47
<b>Mediana</b>	23.650,00	13.650	12.255,73
<b>Deviazione std</b>	11.707	10.304	13.554

Tabella 3.9: Statistiche di base sui redditi familiari da trasferimenti; valori espressi in Euro.

### Reddito da capitale

Per i redditi da capitale si ha invece una differenza (significativa) tra i due sottocampioni: in questo caso, i valori sono più elevati per il *Sottocampione2*.

	Popolazione	<i>Sottocampione1</i>	<i>Sottocampione2</i>
<b>Media</b>	8.382,342	8.444,503	10.325,56
<b>Mediana</b>	5.715,456	6.089,705	6.764,18
<b>Deviazione std</b>	14.569	8.814	14.789

Tabella 3.10: Statistiche di base sui redditi familiari da capitale; valori espressi in Euro.

## 3.5 Ricchezza familiare e consumi

Altre variabili ritenute importanti nel delineare le caratteristiche principali delle famiglie analizzate, sono quelle riguardanti la ricchezza familiare e i consumi.

### Ricchezza familiare netta

Il valore più alto registrato per la media della ricchezza familiare netta, è quello che si riferisce al *Sottocampione2*. Tuttavia, la differenza con il *Sottocampione1* non risulta statisticamente significativa.

	Popolazione	<i>Sottocampione1</i>	<i>Sottocampione2</i>
<b>Media</b>	242.999	295.402,7	348.439,9
<b>Mediana</b>	138.960	171.343	220.505,4
<b>Deviazione std</b>	471.405	534.007	605.054

Tabella 3.11: Statistiche di base sulla ricchezza netta familiare; valori espressi in Euro.

### Consumi

Come nel caso precedente anche per i consumi si registra il valore più elevato per il *Sottocampione2*, anche se, ancora una volta, tale distanza tra i due gruppi non è significativa.

	Popolazione	<i>Sottocampione1</i>	<i>Sottocampione2</i>
<b>Media</b>	26.943,96	34.155,38	36.650,9
<b>Mediana</b>	21.542,4	28.250	30.200
<b>Deviazione std</b>	19.986	20.479	23.970

Tabella 3.12: Statistiche di base sui consumi familiari; valori espressi in Euro.

### Altre variabili

Sono state poi considerate altre variabili come il titolo di proprietà dell'abitazione, il possesso di carta di credito o bancomat e il numero di depositi bancari o postali, che potrebbero rivelarsi utili nella descrizione e nel confronto dei tre campioni. Tra i risultati ottenuti, il più interessante è quello che riguarda il possesso di carta di

credito: se nella popolazione totale la porzione di famiglie che la posseggono è più o meno uguale alla porzione di famiglie che non la possiede, un dato significativo emerge dal confronto dei due sottocampioni. Il 77% di tutte le famiglie che hanno figli di età compresa tra i 18 e i 22 anni e che hanno almeno una carta di credito, appartiene infatti al *Sottocampione2* e solamente il restante 23% fa parte del *Sottocampione1*.



## Capitolo 4

### Confronto in base al titolo di studio del capofamiglia

Dalle analisi svolte nel capitolo precedente si può vedere come vi siano delle differenze significative tra i due sottocampioni e la popolazione totale. In questo capitolo si vuole concentrare l'attenzione solamente sul *Sottocampione1* - famiglie con figli di età compresa tra i 18 e i 22 anni occupati o in cerca di prima occupazione- e sul *Sottocampione2* - famiglie con figli di età compresa tra i 18 e i 22 anni ancora studenti - suddividendo le famiglie in base al titolo di studio del capofamiglia. In altre parole, le unità saranno ulteriormente divise e i nuovi gruppi saranno individuati dal livello di istruzione del principale percettore di reddito del nucleo familiare. Si otterrà quindi l'insieme delle famiglie con capofamiglia laureato o con specializzazione post laurea, l'insieme delle famiglie con capofamiglia diplomato, ed infine, l'insieme delle famiglie con capofamiglia con licenza di terza media. Una volta ottenuti tali sottocampioni, saranno condotte ulteriori analisi con lo scopo di capire se, controllando per il titolo di studio del capofamiglia, il reddito sia una variabile distribuita in modo diverso tra i due insiemi. Dalle analisi precedenti, infatti, era emerso che non vi erano differenze significative tra il livello del reddito familiare *equivalente* delle famiglie del *Sottocampione1* e del *Sottocampione2*.

## 4.1 Laurea o specializzazione post-laurea

Il primo sottocampione è costituito da tutte le famiglie con capofamiglia che abbia conseguito almeno la laurea. Le unità selezionate sono 161 di cui solamente 5 con capofamiglia che abbia anche una specializzazione post-laurea. Il 93% dei giovani di età compresa tra i 18 e i 22 anni appartenenti alle famiglie di questo sottocampione, sono iscritti all'università. Si può quindi concludere che, il fatto di avere un genitore con un grado di istruzione molto elevato, influenzi anche l'educazione dei figli, motivandoli a proseguire gli studi.

Il confronto tra i redditi delle famiglie di *Sottocampione1* e di *Sottocampione2* è dunque inutile visto che sono quasi tutte appartenenti al *Sottocampione2*.

	<i>Sottocampione1</i>	<i>Sottocampione2</i>	<i>Totale</i>
<i>Frequenze</i>	11	150	161
%	6,83	93,17	100,00

Tabella 4.1: Tabella di classificazione delle famiglie con capofamiglia laureato o con specializzazione post laurea, in base al sottocampione di appartenenza.

## 4.2 Capofamiglia diplomato

Il secondo sottocampione è costituito da tutte le famiglie il cui capofamiglia abbia conseguito il diploma. Sono state individuate 566 famiglie con tale caratteristica delle quali 119 provenienti dal *Sottocampione1* e 447 dal *Sottocampione2*.

In primo luogo si è cercato di capire se vi fosse una differenza nella distribuzione del reddito tra le famiglie con figli ancora studenti e quelle con figli già occupati o in cerca di un primo lavoro.

E' stata quindi creata una tabella a doppia entrata che mettesse a confronto la classe di reddito familiare con il sottocampione (1 o 2) alla quale ciascuna famiglia appartiene. (Tabella 4.2).

Classe di reddito <i>Euro</i>	<i>Sottocampione1</i>		<i>Sottocampione2</i>	
	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>
$reddito < 10.000$	4	3,36	13	2,91
$10.000 \leq reddito < 20.000$	13	10,92	42	9,40
$20.000 \leq reddito < 30.000$	20	16,81	72	16,11
$30.000 \leq reddito < 40.000$	22	18,49	80	17,90
$40.000 \leq reddito < 50.000$	16	13,45	84	18,79
$50.000 \leq reddito < 60.000$	13	10,92	34	7,61
$60.000 \leq reddito < 70.000$	10	8,40	37	8,28
$reddito \geq 70.000$	21	17,65	85	19,02
Totale	119	100,00	447	100,00

Tabella 4.2: Tabella di classificazione della classe di reddito delle famiglie con capofamiglia diplomato in base al sottocampione di appartenenza.

Il test del  $\chi^2$  condotto su tale tabella rivela che non vi sono significative divergenze tra le famiglie del *Sottocampione1* e quelle appartenenti al *Sottocampione2*. In base a questo test si potrebbe quindi concludere che non vi è dipendenza tra le due variabili. Tuttavia, si è ritenuto opportuno condurre anche un'analisi della varianza sulla distribuzione del reddito equivalente (ossia una variabile continua, non più categoriale).

Diversamente da quanto ci si sarebbe attesi in base ai risultati del test del  $\chi^2$ , per l'ANOVA si ottiene un livello di significatività dello 0.005%. Non si può dunque affermare che per i campioni considerati ci sia indipendenza tra reddito familiare e il fatto che i figli di età compresa tra i 18 e i 22 anni siano ancora studenti o meno.

Tuttavia, i redditi familiari equivalenti sono mediamente più elevati per le famiglie del *Sottocampione1*. Per queste infatti, la media è di 35854.0€, mentre per le famiglie del *Sottocampione2* è di 31961.3€.

Per avere un'ulteriore conferma del risultato, si è cercato di capire se tale differenza sia riconducibile al fatto che nelle famiglie del *Sottocampione2* si ha un

possibile aumento del numero dei percettori di reddito riconducibile alla presenza di giovani che smettono di studiare per entrare nel mondo del lavoro.

Sono state quindi costruite delle tabelle a doppia entrata che confrontassero la classe di reddito familiare con il fatto che ci fosse almeno un componente nella famiglia di età compresa tra i 18 e i 22 anni già occupato (tabella 4.3) o in cerca di prima occupazione (tabella 4.4).

Classe di reddito	<i>Giovani (18-22 anni) occupati</i>			
	<i>Nessuno</i>		<i>Almeno uno</i>	
	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>
<i>Euro</i>				
$reddito < 10.000$	1	2,13	3	4,17
$10.000 \leq reddito < 20.000$	7	14,89	6	8,33
$20.000 \leq reddito < 30.000$	8	17,02	12	16,67
$30.000 \leq reddito < 40.000$	13	27,66	9	12,50
$40.000 \leq reddito < 50.000$	3	6,38	13	18,06
$50.000 \leq reddito < 60.000$	7	14,89	6	8,33
$60.000 \leq reddito < 70.000$	3	6,38	7	9,72
$reddito \geq 70.000$	5	10,64	16	22,22
Totale	47	100,00	72	100,00

Tabella 4.3: Tabella di classificazione della classe di reddito delle famiglie del *Sottocampione1* con capofamiglia diplomato in base alla presenza di giovani occupati

I test del  $\chi^2$  sulle due tabelle, indicano che non ci sono particolari relazioni tra le due variabili. Ciò implica che non c'è alcuna dipendenza tra il fatto che un giovane tra i 18 e i 22 anni smetta di studiare per iniziare a lavorare o a cercare la sua prima occupazione e la classe di reddito familiare. La scelta di non proseguire gli studi è quindi indipendente dalla necessità di trovare presto un'occupazione. Dunque, almeno per le famiglie con capofamiglia diplomato, la decisione di non iscriversi all'università non è legata al reddito familiare.



Classe di reddito <i>Euro</i>	<i>Giovani (18-22 anni) in cerca di prima occupazione</i>			
	<i>Nessuno</i>		<i>Almeno uno</i>	
	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>
$reddito < 10.000$	4	5,13	0	0,00
$10.000 \leq reddito < 20.000$	9	11,54	4	9,76
$20.000 \leq reddito < 30.000$	12	15,38	8	19,51
$30.000 \leq reddito < 40.000$	10	12,82	12	29,27
$40.000 \leq reddito < 50.000$	11	14,10	5	12,20
$50.000 \leq reddito < 60.000$	8	10,26	5	12,20
$60.000 \leq reddito < 70.000$	8	10,26	2	4,88
$reddito \geq 70.000$	16	20,51	5	12,10
Totale	78	100,00	41	100,00

Tabella 4.4: Tabella di classificazione della classe di reddito delle famiglie del *Sottocampione1* con capofamiglia diplomato in base alla presenza di giovani in cerca di prima occupazione.

### 4.3 Capofamiglia con licenza di terza media

Infine sono state isolate tutte le famiglie il cui capofamiglia abbia conseguito la licenza di scuola media inferiore. Il campione è costituito da 439 osservazioni delle quali 121 appartenenti al *Sottocampione1* e 318 al *Sottocampione2*.

Come per i gruppi precedenti anche per le famiglie di questo insieme sono stati analizzati i redditi equivalenti, soffermandosi in particolar modo sulla distribuzione di questa variabile all'interno del *Sottocampione1* e del *Sottocampione2*. Come per le famiglie il cui capofamiglia era diplomato, anche in questo caso, si ha un risultato significativo nel test ANOVA: esiste quindi una differenza significativa tra il valore della media del reddito familiare equivalente del *Sottocampione1* e del *Sottocampione2*. Ancora una volta, però, i redditi più elevati si registrano per le famiglie con figli di età compresa tra i 18 e i 22 anni che decidono di non proseguire gli studi. Si ha infatti che, per il *Sottocampione1* la media è di 31004,00€, mentre per le famiglie del *Sottocampione2* la media è di 23692,1€.

Tuttavia, a differenza di quanto visto per le famiglie il cui capofamiglia era diplomato, in questo caso è possibile riscontrare una dipendenza tra il fatto che un ragazzo di età compresa tra i 18 e i 22 anni lavori (tabella 4.5) o sia in cerca di prima occupazione (tabella 4.6) con la classe di reddito familiare.

Classe di reddito	<i>Giovani (18-22 anni) occupati</i>			
	<i>Nessuno</i>		<i>Almeno uno</i>	
	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>
<i>Euro</i>				
$reddito < 10.000$	2	3,57	0	0,00
$10.000 \leq reddito < 20.000$	17	30,36	0	0,00
$20.000 \leq reddito < 30.000$	9	16,07	7	10,77
$30.000 \leq reddito < 40.000$	8	14,29	14	21,54
$40.000 \leq reddito < 50.000$	8	14,29	13	20,00
$50.000 \leq reddito < 60.000$	6	10,71	12	18,46
$60.000 \leq reddito < 70.000$	1	1,79	4	6,15
$reddito \geq 70.000$	5	8,93	15	23,08
Totale	56	100,00	65	100,00

Tabella 4.5: Tabella di classificazione della classe di reddito delle famiglie del *Sottocampione1* con capofamiglia con licenza di terza media in base alla presenza di giovani occupati.

Sulla base di questi risultati, dunque si potrebbe concludere che, per le famiglie il cui capofamiglia ha conseguito la licenza di terza media, il reddito familiare incide significativamente sulla decisione di proseguire o meno gli studi. Si potrebbe infatti spiegare questo risultato ipotizzando che i giovani che rinunciano alla laurea sono indotti a prendere questa decisione da motivi economici: l'ulteriore reddito che potrebbero percepire potrebbe aumentare significativamente il reddito familiare. Per avere conferma di questa ipotesi, si è pensato di sottrarre dai redditi familiari di queste famiglie il reddito percepito dai giovani di età compresa tra i 18 e i 22 anni. Una volta eseguita questa operazione, si sono confrontati i redditi ottenuti con quelli delle famiglie del *Sottocampione2*. Diversamente da quanto ci si sarebbe attesi, i redditi delle famiglie del *Sottocampione1* ai quali sono stati

Classe di reddito <i>Euro</i>	<i>Giovani (18-22 anni) in cerca di prima occupazione</i>			
	<i>Nessuno</i>		<i>Almeno uno</i>	
	<i>Frequenza</i>	<i>%</i>	<i>Frequenza</i>	<i>%</i>
$\text{reddito} < 10.000$	0	0,00	2	4,35
$10.000 \leq \text{reddito} < 20.000$	2	2,67	15	32,61
$20.000 \leq \text{reddito} < 30.000$	6	8,00	10	21,74
$30.000 \leq \text{reddito} < 40.000$	14	18,67	8	17,39
$40.000 \leq \text{reddito} < 50.000$	16	21,33	5	10,87
$50.000 \leq \text{reddito} < 60.000$	14	18,67	4	8,70
$60.000 \leq \text{reddito} < 70.000$	3	4,00	2	4,35
$\text{reddito} \geq 70.000$	20	26,67	0	0,00
Totale	75	100,00	46	100,00

Tabella 4.6: Tabella di classificazione della classe di reddito delle famiglie del *Sottocampione1* con capofamiglia diplomato in base alla presenza di giovani in cerca di prima occupazione.

sottratti i redditi dei giovani tra i 18 e i 22 anni occupati, restano, in media, più elevati rispetto ai redditi delle famiglie con figli iscritti all'università. Si ottiene infatti che il reddito equivalente medio equivale a 30205.68€ per le famiglie del *Sottocampione1* e a 23692.11€ per le famiglie del *Sottocampione2*.



## Capitolo 5

# Stima del modello logit per la previsione della probabilità

### 5.1 La scelta del modello

Dopo le analisi descrittive esposte nei capitoli precedenti, è stato stimato un modello logit per la previsione della probabilità di iscriversi o meno all'università. Poiché la variabile di output, ossia la decisione di proseguire gli studi, è binaria, si è deciso di utilizzare un modello logistico. Partendo dalla popolazione costituita dall'unione del *Sottocampione1* e del *Sottocampione2*, si è costruito il modello inserendo come variabile dipendente il fatto di appartenere al *Sottocampione1*, che rappresenta le famiglie con figli occupati o in cerca di prima occupazione, o al *Sottocampione2*, che include invece le famiglie con figli ancora studenti. Il modello risultante, stimerà per ciascun regressore un coefficiente che ne misuri l'effetto sulla probabilità predetta.

### 5.2 Le variabili incluse

Nel delineare il modello statistico, si è rivelata di fondamentale importanza la scelta dei regressori da includere.

Nei primi tentativi di stima, si è infatti cercato di includere tutte le variabili

che potessero incidere nella decisione. Sono state quindi considerate nel primo modello stimato:

- *tutte le variabili di reddito;*
- *i consumi familiari;*
- *la ricchezza familiare;*
- *il numero dei figli;*
- *la presenza di altri componenti nel nucleo familiare;*
- *il fatto che i genitori fossero sposati o separati;*
- *il tipo di lavoro del capofamiglia (dipendente o autonomo);*
- *condizione lavorativa del coniuge del capofamiglia (occupato o non occupato);*
- *il titolo di studio del capofamiglia e del coniuge.*

Inizialmente le variabili di reddito (da lavoro dipendente, da lavoro autonomo, da trasferimenti, da capitale) sono state inserite come variabili continue; nei modelli successivi, sono state invece sostituite con il logaritmo del reddito familiare e poi da tre dummies che rappresentassero il secondo, terzo e quarto quartile della distribuzione della variabile reddito.

I consumi e la ricchezza familiare sono state considerate solamente come variabili continue e quindi non sono state trasformate in variabili categoriali ordinate.

Il numero dei figli è stato inserito nel modello come variabile discreta: una dummy per indicare che la famiglia ha due figli, una seconda dummy per indicare che ci sono almeno tre figli nel nucleo familiare. In ciascun caso il punto di riferimento erano quindi le famiglie con solamente un figlio: attraverso l'utilizzo delle variabili binarie si è potuto osservare la differenza nell'impatto sulla probabilità di iscriversi all'università, tra avere uno o due, uno o tre o più figli.

Dummies sono state utilizzate anche per indicare la presenza di altri componenti nella famiglia, per differenziare le coppie separate, per il tipo di lavoro del capofamiglia, per il tipo di attività del coniuge (occupato in altra condizione) e per il

titolo di studio. In questo ultimo caso, si sono raggruppati i livelli di istruzione in tre categorie principali: titolo di studio inferiore alla terza media, diploma di scuola superiore, laurea o specializzazione post laurea. Nel modello sono poi state inserite le variabili binarie corrispondenti ai due livelli di istruzione più elevato per rilevare il diverso impatto che diploma e laurea hanno rispetto al fatto che il capofamiglia abbia un basso titolo di studio.

Da questo primo modello di partenza, attraverso il metodo del rapporto di verosimiglianza nel caso di modelli annidati, o ricorrendo al criterio AIC nel caso di modelli non annidati, si sono selezionate le variabili significative che hanno un impatto decisivo nella scelta di iscriversi all'università.

### 5.3 Il modello stimato

Le variabili mantenute nel modello definitivo sono riportate nella seguente tabella.

<i>Variabile</i>	<i>Descrizione</i>
<i>figli2</i> =1	se nella famiglia ci sono due figli
<i>figli2</i> =0	altrimenti
<i>figli3</i> =1	se nella famiglia ci sono almeno tre figli
<i>figli3</i> =0	altrimenti
<i>co_lav</i> =1	se il coniuge è occupato
<i>co_lav</i> =0	altrimenti
<i>yfam2</i> =1	se il reddito familiare appartiene al secondo quartile
<i>yfam2</i> =0	altrimenti
<i>yfam3</i> =1	se il reddito familiare appartiene al terzo quartile
<i>yfam3</i> =0	altrimenti

---

$yfam_4=1$	se il reddito familiare appartiene al quarto quartile
$yfam_4=0$	altrimenti
$cf\_diploma=1$	se il capofamiglia è diplomato
$cf\_diploma=0$	altrimenti
$cf\_lausp=1$	se il capofamiglia è laureato
$cf\_lausp=0$	altrimenti
$co\_diploma=1$	se il coniuge del capofamiglia è diplomato
$co\_diploma=0$	altrimenti
$co\_lausp=1$	se il coniuge del capofamiglia è laureato
$co\_lausp=0$	altrimenti
$con2=1$	se i due coniugi vivono insieme
$con2=0$	altrimenti

---

Il modello che meglio spiega la decisione di iscriversi o meno all'università, risulta essere (tabella 5.1):



Parametro	DF	Stima	Errore standard	Wald Chi-quadr	P-value
Intercept	1	3,0843	0,2999	105,7483	<,0001
figli2	1	0,5941	0,0762	60,7429	<,0001***
figli3	1	0,8877	0,1084	67,0043	<,0001***
co_lav	1	0,1923	0,0772	6,1987	0,0128**
yfam2	1	-0,1095	0,0940	1,3576	0,2440
yfam3	1	-0,2661	0,0965	7,6125	0,0058***
yfam4	1	-0,3733	0,1010	13,6538	0,0002***
cf_diploma	1	0,2963	0,0766	14,9713	0,0001***
cf_lausp	1	0,7522	0,1801	17,4419	<,0001***
co_diploma	1	0,3444	0,0912	14,2542	0,0002***
co_lausp	1	0,7617	0,2319	10,7850	0,0010***
con2	1	-0,2362	0,1063	4,9356	0,0263**

Tabella 5.1: Stime dei parametri del modello logistico.

La stima dei rapporti di quote è invece data da (tabella 5.2):

## 5.4 Interpretazione del modello

Tutte le variabili inserite nel modello hanno quindi un effetto significativo sulla probabilità di iscriversi all'università (tutte tranne “yfam2” ma non avrebbe senso toglierla visto che come, “yfam3” e “yfam4” rappresenta un quartile del reddito familiare).

Come ci si sarebbe atteso dalle analisi precedenti, il **reddito** ha un influenza negativa sulla variabile risposta. Nel terzo e nel quarto capitolo è infatti emerso che, in media, i redditi delle famiglie del *Sottocampione1* erano più elevati di quelle del *Sottocampione2*. Questo impatto, risulta inoltre essere lineare: il valore assoluto dei coefficienti riferiti a “yfam2”, “yfam3”, “yfam4” aumenta passando dal primo al secondo, dal primo al terzo e dal primo al quarto quartile della variabile reddito. Di conseguenza, considerando i rapporti di quote (*log odds-ratios*), la

Effetto		Stima puntuale	95% Limiti di confidenza di Wald	
figli2	1 vs 0	3.281	2.433	4.423
figli3	1 vs 0	5.903	3.859	9.030
co_lav	1 vs 0	1.469	1.085	1.988
yfam2	1 vs 0	0.803	0.556	1.161
yfam3	1 vs 0	0.587	0.402	0.857
yfam4	1 vs 0	0.474	0.319	0.704
cf_diploma	1 vs 0	1.809	1.340	2.442
cf_lausp	1 vs 0	4.501	2.222	9.120
co_diploma	1 vs 0	1.991	1.393	2.847
co_lausp	1 vs 0	4.587	1.848	11.387
con2	1 vs 0	0.623	0.411	0.946

Tabella 5.2: Stime dei rapporti di quote del modello logistico.

probabilità di iscriversi all'università diminuisce a mano a mano che aumenta il livello del reddito familiare: per le famiglie il cui reddito appartiene al secondo quartile, tale probabilità è pari a 0,803 volte la stessa probabilità per una famiglia il cui reddito è nel primo quartile. Il rapporto scende a 0,587 per le famiglie del terzo quartile rispetto a quelle del primo e per le famiglie con reddito più elevato la probabilità è addirittura dimezzata (0,474) rispetto alle famiglie più povere.

Interessanti sono poi i risultati che si riferiscono alle variabili “**co\_lav**” e “**con2**”. Per la prima infatti, si ha un impatto positivo sulla variabile dipendente: il fatto che il coniuge lavori aumenta la probabilità di iscriversi all'università. Tale probabilità è infatti pari a 1,469 volte la probabilità di continuare gli studi per le famiglie in cui il coniuge del capofamiglia non lavora. Al contrario, il fatto che i due coniugi non vivano insieme perchè separati o divorziati o per altri motivi, ha un impatto negativo sulla variabile risposta.

Inaspettatamente, il **numero dei figli** ha un effetto positivo e crescente sulla probabilità di iscriversi all'università. Provando ad inserire le interazioni tra numero di figli e reddito, il risultato resta lo stesso: impatto negativo del reddito

considerando lo stesso numero di figli, impatto positivo delle dummies del numero dei figli controllando per il reddito. Ciò implica che non vi è interazione (tali variabili risultano anche non significative) tra numero di figli e reddito familiare nella decisione di iscriversi all'università e che questa interazione non è funzionale nella spiegazione del risultato. Probabilmente, invece, va ricondotto al fatto che, indipendentemente dal numero di figli, se i maggiori si iscrivono all'università anche i fratelli minori intraprenderanno la stessa strada a prescindere dalla possibilità economica. Il reddito, come visto precedentemente, non preclude infatti la scelta di proseguire gli studi.

Infine è importante sottolineare i coefficienti delle variabili riferite al **titolo di studio** del coniuge e del capofamiglia. Il fatto di avere uno dei due genitori diplomati, infatti, raddoppia o quasi la probabilità di iscriversi all'università. Adirittura, nel caso in cui il capofamiglia o il coniuge siano laureati, la probabilità di conseguire una laurea per i loro figli è pari a circa 4,5 volte la stessa probabilità per i figli di genitori con un basso titolo di studio. Il livello di istruzione dei genitori è quindi una variabile molto significativa nel processo decisionale. Diversamente dagli altri paesi europei, in Italia si ha una scarsa mobilità intergenerazionale per quanto riguarda il titolo di studio: è piuttosto più probabile che il figlio di due genitori laureati, prosegua gli studi dopo il diploma piuttosto che un ragazzo con genitori con un basso livello di istruzione.



## Capitolo 6

### Conclusioni

Nei capitoli precedenti, sono state analizzate le differenze fondamentali tra le famiglie con figli di età compresa tra 18 e 22 anni già occupati o in cerca di prima occupazione, il *Sottocampione1*, e le famiglie con figli nella stessa fascia d'età ma ancora studenti, il *Sottocampione2*. Questo confronto è stato condotto osservando le divergenze tra i due gruppi nella distribuzione di alcune variabili ritenute importanti nella decisione di proseguire gli studi.

Nel terzo capitolo, si è visto che, tra le variabili categoriali, quella che maggiormente differisce tra i due gruppi è il titolo di studio del capofamiglia: in media infatti, il livello di istruzione dei capifamiglia delle famiglie del *Sottocampione2* sono più elevati di quello dei capifamiglia del *Sottocampione1*. Questo risultato è anche confermato dal modello logit stimato nel quinto capitolo: i coefficienti delle dummies del titolo di studio del capofamiglia e del suo coniuge sono infatti significative allo 0,001%. In particolar modo, si è visto che la presenza di un genitore laureato può aumentare la probabilità di iscrizione all'università per i figli di quasi 5 volte.

Sempre nel terzo capitolo, si è analizzata la differenza tra la qualifica professionale dei capifamiglia nei due sottocampioni: per il *Sottocampione1* è molto più consistente la presenza di operai (37%) che scende notevolmente nel *Sottocampione2* (26%) lasciando spazio alla percentuale di insegnanti e impiegati.

Per quanto riguarda invece le variabili continue, è emerso che l'unica differenza

nella loro distribuzione è presente per il reddito da capitale che però non incide, almeno non direttamente, nella scelta di proseguire gli studi.

Le variabili di reddito sono state poi analizzate in modo più approfondito nel quarto capitolo: dopo aver diviso le famiglie del *Sottocampione1* e del *Sottocampione2* in base al titolo di studio del capofamiglia si sono confrontati i redditi familiari controllando per il livello di istruzione del capofamiglia. Da questa analisi è emerso che:

- per le famiglie con capofamiglia laureato o con specializzazione post-laurea, il fattore che maggiormente incide sulla decisione di proseguire gli studi è proprio l'ereditarietà del titolo di studio. E' quindi molto probabile che i ragazzi che abbiano genitori con un elevato livello di istruzione abbiano una maggiore propensione a proseguire gli studi dopo la scuola superiore;
- per le famiglie con capofamiglia diplomato invece, si è osservata una diversa distribuzione del reddito tra le famiglie del *Sottocampione1* e del *Sottocampione2*. Inaspettatamente però, sono le famiglie con figli già occupati o in cerca di prima occupazione ad avere un reddito mediamente più elevato. Inoltre il fatto di interrompere gli studi per entrare nel mondo del lavoro non può essere ricondotto a motivi economici: dal test del  $\chi^2$ , infatti, non emerge nessuna relazione tra la presenza di giovani occupati o in cerca di prima occupazione e la classe di reddito familiare;
- un dato interessante emerge anche per le famiglie con capofamiglia che abbia conseguito la licenza di terza media. Anche in questo caso i redditi sono mediamente più elevati per le famiglie appartenenti al *Sottocampione1*: in questo caso però esiste una relazione tra il fatto che un ragazzo sia occupato o in cerca di prima occupazione e il reddito familiare. Tuttavia, sottraendo dal reddito familiare, il reddito percepito dai più giovani, si ottiene che il reddito familiare di questo sottoinsieme di famiglie sia più elevato di quello delle famiglie con capofamiglia con licenza di terza media appartenenti al *Sottocampione2* (ossia con figli iscritti all'università).

I risultati precedenti sono stati confermati anche nell'ultimo capitolo: il modello logit stimato sottolinea l'influenza che il titolo di studio dei genitori ha sul livello di istruzione dei figli. In Italia, infatti, si osserva una scarsa mobilità intergenerazionale rispetto al titolo di studio: giovani con padre o madre laureato, conseguiranno un titolo di studio elevato molto più probabilmente rispetto a ragazzi con genitori che abbiano solamente la licenza di terza media. Diversamente da quanto ci si potrebbe aspettare, il reddito ha una minore influenza rispetto all'istruzione dei genitori: dai dati emerge come questa variabile abbia un effetto negativo sulla probabilità che un giovane di età compresa tra i 18 e i 22 anni si iscriva all'università.

Concludendo, si può affermare che la scelta di proseguire gli studi non dipende dalla reale possibilità economica di sostenere la spesa che l'iscrizione all'università comporta; tradizione familiare, attitudini e interessi personali rappresentano infatti gli elementi fondamentali che incidono sulla decisione del titolo di studio che un giovane vuole conseguire.





# Appendice A

## Il programma SAS

Le analisi viste nei capitoli precedenti sono state svolte utilizzando il programma statistico SAS. Di seguito, sarà quindi riportata la sintassi utilizzata per la lettura dei dati, per la creazione e la trasformazione delle variabili.

### A.1 La lettura dei dati

Di seguito è riportato il programma per la lettura in **SAS** dei dati forniti dalla Banca. I dataset utilizzati sono:

**RFAM0x:** redditi familiari;

**RISFAM0x:** risparmi e consumi familiari;

**RICFAM0x:** ricchezza familiare;

**RPER0x:** redditi individuali;

**CARCOM0x:** caratteristiche dei componenti;

**Q0xC1:** strumenti di pagamento e attività finanziarie;

**Q0xD:** beni immobili

Poichè i dati utilizzati sono stati ricavati da tre diverse indagini (2000, 2002, 2004), è stato trascritto solamente il programma generale, sostituendo l'anno di riferimento (00, 02, 04) con 0x.

### Lettura di Carcom

```
libname dati200x 'c:\tesi\datiBI\sas_200x';  
libname datiaggr 'c:\tesi\datiBI\aggr';  
run;
```

```
data carcom0x (drop=studio parent);  
set dati200x.carcom0x;  
anno=200x;
```

```
if studio='1' then studiox='1';  
if studio='2' then studiox='2';  
if studio='3' then studiox='3';  
if studio='4' then studiox='4';  
if studio='5' then studiox='4';  
if studio='6' then studiox='5';  
if studio='7' then studiox='5';  
if studio='8' then studiox='6';
```

```
if apqual='1' then apqual1='1';  
if apqual='2' then apqual1='2';  
if apqual='3' then apqual1='2';  
if apqual='4' then apqual1='3';  
if apqual='5' then apqual1='4';  
if apqual='6' then apqual1='5';  
if apqual='7' then apqual1='6';  
if apqual='8' then apqual1='7';  
if apqual='9' then apqual1='8';  
if apqual='10' then apqual1='6';  
if apqual='20' then apqual1='7';
```

```
if apqual='11' then apnonoc1='1';  
if apqual='12' then apnonoc1='2';  
if apqual='13' then apnonoc1='3';
```

```
if apqual='14' then apnonoc1='4';
if apqual='15' then apnonoc1='5';
if apqual='16' then apnonoc1='5';
if apqual='17' then apnonoc1='6';
if apqual='18' then apnonoc1='8';
if apqual='19' then apnonoc1='7';
```

```
if apsett='1' then apsett1='1';
if apsett='2' then apsett1='3';
if apsett='3' then apsett1='2';
if apsett='4' then apsett1='4';
if apsett='5' then apsett1='5';
if apsett='6' then apsett1='6';
if apsett='7' then apsett1='7';
if apsett='8' then apsett1='8';
if apsett='9' then apsett1='8';
if apsett='10' then apsett1='8';
```

```
parenx=parent;
if parent=4 or parent=5 then parenx=4;
if parent=6 then parenx=5;
```

```
if apqual1='' then do;
apqual1=.;
apsett1=.;
end;\
```

```
if apnonoc1='' then apnonoc1=.;
run;
```

```
data carcom0x;
set carcom0x;
```

```
rename parenx=parent;
rename studiox=studio;
run;
```

### Lettura di RPER

```
data rper0x;
set dati200x.rper0x;
AF=0;
AF= sum(AF1, AF2, AF3, AF4);
keep nquest nord Y YL YT YTP YTA YTA1 YTA2 YM YC YCA YCF AF;
run;
```

### Lettura di RFAM

```
data rfam0x;
set dati200x.rfam0x;
rename y=yfam
      yl=ylfam
      yt=ytfam
      ym=ymfam
      yc=ycfam;
run;

data rfam0x;
set rfam0x;
keep nquest yfam ylfam ytfam ytp yta yta1 yta2 ymfam ycfam yca ycf;
run;
```

### Lettura di RISFAM

```
data confam0x (keep=nquest consumi);
set dati200x.risfam0x;
rename c=consumi;
run;
```

### Lettura di Q0xD

```
data g0x (keep=nquest godabit);  
set dati200x.q0xd;  
run;
```

### Lettura di RICFAM

```
data ricfam00 (keep= nquest AR AF PF BD W WR);  
set dati2000.ricfam00;  
run;
```

### Lettura di Q0xC1

```
data attfin00 (keep= nquest depbanc depbanr depposc depposr bancomat  
carta);  
set dati2000.q00c1;  
run;
```

### Unione dei dataset

```
proc sort data=carcom00;  
by nquest nord; run;
```

```
proc sort data=rper00;  
by nquest nord;  
run;
```

```
proc sort data=rfam00;  
by nquest;  
run;
```

```
proc sort data=confam00;  
by nquest;  
run;
```

```
proc sort data=g00;
by nquest;
run;
```

```
proc sort data=ricfam00;
by nquest;
run;
```

```
proc sort data=attfin00;
by nquest;
run;
```

```
data datiaggr.dati1_00;
merge carcom00 rper00;
by nquest nord;
run;
```

```
data datiaggr.dati2_00;
merge confam00 ricfam00 g00 rfam00 attfin00;
by nquest;
run;
```

## A.2 Variabili

Di seguito è invece riportato il programma per la creazione e la trasformazione delle nuove variabili che saranno utilizzate per le analisi descrittive e per la stima del modello logit.

### A.2.1 Creazione delle nuove variabili

```
data individuali0x;
set datiaggr.dati1_0x(keep=nquest nord pesoFL
parent acom4c area5 ncomp nperc sex cleta eta staciv studio
```

```
apqual1 apnonoc1 apsett1 Y) ;  
by nquest;
```

```
if first.nquest then D0;  
cf_sesso=0;  
cf_cleta=0;  
cf_staci=0;  
cf_stud=0;  
cf_apqua=0;  
cf_apnon=0;  
cf_apset=0;  
co_stud=0;  
co_apqua=0;  
co_apnon=0;  
co_apset=0;  
so_apnon=0;
```

```
coniugi2=0;  
con2=0;  
cf_donna=0;  
cf_att=0;  
c2_att=0;  
con2at=0;  
n_figli=0;  
n_altri=0;  
n_anz=0;  
n_bamb=0;  
n_bscuo=0;  
n_campione_18_22=0;  
n_diploma_18_22=0;  
n_studente_16_20=0;  
n_studente_18_22=0;  
n_occupato_18_22=0;  
n_disoccupato_18_22=0;
```

```
n_primaocc_18_22=0;
n_disoccupato_18_22=0;
campione_16_20=0;
campione_18_22=0;
licmedia_16_20=0;
diploma_18_22=0;
studente_18_22=0;
occupato_18_22=0;
disoccupato_18_22=0;
primaocc_18_22=0;
disoccupato_18_22=0;
p_attiva=0;
n_flav=0;
n_occ=0;
n_dis=0;
n_1a_oc=0;
n_dip=0;
n_aut=0;
n_nflav=0;
n_pens=0;
n_cas=0;
n_stud=0;
n_comp=0;
n_perc=0;
_area=0;
_acom=0;
_peso=0;
Yfigli=0;
END;
```

```
if parent=1 then cf_sesso=sex;
if parent=1 then cf_cleta=cleta;
if parent=1 then cf_staci=staciv;
if parent=1 then cf_stud=studio;
```



```
if parent=1 then cf_apqua=apqual1;
if parent=1 then cf_apnon=apnonoc1;
if parent=1 then cf_apset=apsett1;
if parent=2 then co_stud=studio;
if parent=2 then co_apqua=apqual1;
if parent=2 then co_apnon=apnonoc1;
if parent=2 then co_apset=apsett1;
if parent=3 then so_apnon=apnonoc1;

if parent=3 and (18<= eta <=22) then Yfigli=Y;
if parent in (1,2) then coniugi2+1;
if coniugi2=2 then con2=1;
if coniugi2=1 or coniugi2=0 then con2=2;
if parent=1 and sex=2 then cf_donna+1;
if parent=1 and apnonoc1=. and (14<= eta <=60) then cf_att+1;
if (parent IN (1,2) and apnonoc1=. ) then c2_att+1;
if c2_att=2 then con2at=1;
if c2_att=1 or c2_att=0 then con2at=0;
if parent=3 then n_figli+1;
if parent=4 or parent=5 then n_altri+1;
if (eta>=65) then n_anz+1;
if (0<= eta <=4) then n_bamb+1;
if (5<= eta <=14) then n_bscuo+1;
if (14<= eta <=60) then p_attiva+1;
if (18<= eta <=22) then n_campione_18_22+1;
if n_campione_18_22>=1 then campione_18_22=1;

if campione_18_22=1 and studio=4 then n_diploma_18_22+1;
if n_diploma_18_22>=1 then diploma_18_22=1;
if campione_18_22=1 and apnonoc1=6 then n_studente_18_22+1;
if n_studente_18_22>=1 then studente_18_22=1;
if campione_16_20=1 and apnonoc1=6 then n_studente_16_20+1;
if n_studente_16_20>=1 then studente_16_20=1;
if campione_18_22=1 and apqual1 in (1,2,3,4,5,6,7,8)
```

```

    then n_occupato_18_22+1;
if n_occupato_18_22=>1 then occupato_18_22=1;
if campione_18_22=1 and apnonoc1=1 then n_primaocc_18_22+1;
if n_primaocc_18_22=>1 then primaocc_18_22=1;
if campione_18_22=1 and apnonoc1=2 then n_disoccupato_18_22+1;
if n_disoccupato_18_22=>1 then disoccupato_18_22=1;

if (apnonoc1=1 or apnonoc1=2) or (apnonoc1=. and 14<= eta <=60)
    then n_flav+1;
if apnonoc1=. and (14<= eta <=60) then n_occ+1;
if apnonoc1=2 then n_dis+1;
if apnonoc1=1 then n_1a_oc+1;
if (apqual1=1 or apqual1=2 or apqual1=3 or apqual1=4)
    and (apnonoc1=. and 14<= eta <= 60) then n_dip+1;
if (apqual1=5 or apqual1=6 or apqual1=7 or apqual1=8)
    and (apnonoc1=. and 14<= eta <=60) then n_aut+1;
if (apnonoc1=3 or apnonoc1=4 or apnonoc1=5 or apnonoc1=6 or
    apnonoc1=7) then n_nflav+1;
if apnonoc1=5 then n_pens+1;
if apnonoc1=3 then n_cas+1;
if apnonoc1=6 then n_stud+1;
if (parent IN (1,2,3,4,5)) then n_comp+1;
if parent=1 then n_perc=nperc;
if parent=1 then _area=area5;
if parent=1 then _acom=acom4c;
if parent=1 then _peso=pesoFL;

RETAIN cf_sesso cf_cleta cf_staci cf_stud cf_apqua cf_apnon
       cf_apset co_stud co_apqua co_apnon co_apset so_apnon
       coniugi2 con2 cf_donna cf_att c2_att con2at n_figli
       n_altri n_anz n_bamb n_bscuo p_attiva n_flav n_occ
       n_dis n_1a_oc n_dip n_aut n_nflav n_pens n_cas
       n_stud n_comp n_perc n_campione_16_20 n_campione_18_22

```

```

n_licmedia_16_20      n_diploma_18_22      n_studente_16_20
n_studente_18_22  n_occupato_18_22      n_disoccupato_18_22
n_primaocc_18_22      n_disoccupato_18_22      campione_16_20
campione_18_22      licmedia_16_20      diploma_18_22
studente_16_20      studente_18_22      occupato_18_22
disoccupato_18_22      primaocc_18_22      disoccupato_18_22
_area _acom _peso Yfigli;

IF last.nquest;

keep  cf_sesso  cf_cleta  cf_staci  cf_stud  cf_apqua  cf_apnon
      cf_apset  co_stud  co_apqua  co_apnon  co_apset  so_apnon
      coniugi2  con2  cf_donna  cf_att  c2_att  con2at  n_figli
      n_altri  n_anz  n_bamb  n_bscuo  p_attiva  n_flav  n_occ
      n_dis  n_1a_oc  n_dip  n_aut  n_nflav  n_pens  n_cas
      n_stud  n_comp  n_perc  n_campione_16_20  n_campione_18_22
      n_licmedia_16_20      n_diploma_18_22      n_studente_16_20
      n_studente_18_22  n_occupato_18_22      n_disoccupato_18_22
      n_primaocc_18_22      n_disoccupato_18_22      campione_16_20
      campione_18_22      licmedia_16_20      diploma_18_22
      studente_16_20      studente_18_22      occupato_18_22
      disoccupato_18_22      primaocc_18_22      disoccupato_18_22
      _area _acom _peso Yfigli;

run;

data individuali0x;
set individuali0x;
if con2=0 then do;
co_apqua=.;
co_stud=.;
co_apset=.;
co_apnon=.;
end;
if n_figli=0
then so_apnon=.;

```

```
run;
```

### A.2.2 Creazione delle dummies per il modello logit

```
data popol00;  
set popol00;  
cf_notit=2;  
cf_elem=2;  
cf_media=2;  
cf_diploma=2;  
cf_lausp=2;  
cf_no_el=2;  
co_notit=2;  
co_elem=2;  
co_media=2;  
co_diploma=2;  
co_lausp=2;  
co_no_el=2;  
propabit=2;  
d_banc_post=2;  
cart_banc=2;  
figli2=2;  
figli3=2;  
altri=2;  
l_aut=2;  
co_lav=2;  
yfam1=2;  
yfam2=2;  
yfam3=2;  
yfam4=2;  
figli2_y2=2;  
figli2_y3=2;  
figli2_y4=2;  
figli3_y2=2;
```

```
figli3_y3=2;
figli3_y4=2;
nord_ovest=2;
nord_est=2;
centro=2;
sud=2;
isole=2;
if cf_stud=1 then cf_notit=1;
if cf_stud=1 or cf_stud=2 then cf_no_el=1;
if cf_stud=2 then cf_elem=1;
if cf_stud=3 then cf_media=1;
if cf_stud=4 then cf_diploma=1;
if cf_stud in (5,6) then cf_lausp=1;
if co_stud=1 then co_notit=1;
if co_stud=1 or co_stud=2 then co_no_el=1;
if co_stud=2 then co_elem=1;
if co_stud=3 then co_media=1;
if co_stud=4 then co_diploma=1;
if co_stud in (5,6) then co_lausp=1;
if godabit=1 then propabit=1;
if depbanc=1 or depposc=1 then d_banc_post=1;
if bancomat=1 or carta=1 then cart_banc=1;
if n_figli=2 then figli2=1;
if n_figli>=3 then figli3=1;
if n_altri>=1 then altri=1;
if cf_apqua in (5,6,7) then l_aut=1;
if co_apqua in (1,2,3,4,5,6,7,8) then
co_lav=1;
if yfam<25241.5 then yfam1=1;
if 25241.5<=yfam<39371.1
then yfam2=1;
if 39371.2<=yfam<59426.3 then yfam3=1;
if yfam>=59426.3 then yfam4=1;
if figli2=1 and yfam2=1 then figli2_y2=1;
```

```
if figli2=1 and yfam3=1 then figli2_y3=1;
if figli2=1 and yfam4=1 then figli2_y4=1;
if figli3=1 and yfam2=1 then figli3_y2=1;
if figli3=1 and yfam3=1 then figli3_y3=1;
if figli3=1 and yfam4=1 then figli3_y4=1;
if _area=1 then nord_ovest=1;
if _area=2 then nord_est=1;
if _area=3 then centro=1;
if _area=4 then sud=1;
if _area=5 then isole=1;
lyfam= Log(yfam);
run;
```

### A.2.3 Creazione delle classi per il reddito familiare

```
data familiari00;
set datiaggr.dati2_00;

if yfam=0 then cl_yfam=.;
if (0< yfam <=10000) then cl_yfam=1;
if(10000< yfam <=20000) then cl_yfam=2;
if (20000< yfam <=30000) then cl_yfam=3;
if (30000< yfam <=40000) then cl_yfam=4;
if (40000< yfam <=50000) then cl_yfam=5;
if (50000< yfam <=60000) then cl_yfam=6;
if (60000< yfam <=70000) then cl_yfam=7;
if (yfam >70000 ) then cl_yfam=8;
```

### A.2.4 Trasformazione a prezzi del 2004

Tutte le variabili riferite a reddito, consumi, ricchezza sono state trasformate attraverso l'utilizzo del deflatore implicito a prezzi del 2004.

**Prezzi del 2000**

```
data datiaggr.nuove00;  
set datiaggr.nuove00;  
consumi=consumi*1.117;  
BD=BD*1.117;  
PF=PF*1.117;  
AR=AR*1.117;  
AF=AF*1.117;  
WR=WR*1.117;  
W=W*1.117;  
yfam=yfam*1.117;  
ylfam=ylfam*1.117;  
ytfam=ytfam*1.117;  
YTP=YTP*1.117;  
YTA=YTA*1.117;  
YTA1=YTA1*1.117;  
YTA2=YTA2*1.117;  
ymfam=ymfam*1.117;  
ycfam=ycfam*1.117;  
YCA=YCA*1.117;  
YCF=YCF*1.117;  
run;
```

**Prezzi del 2002**

```
data datiaggr.nuove02;  
set datiaggr.nuove02;  
consumi=consumi*1.056;  
BD=BD*1.056;  
PF=PF*1.056;  
AR=AR*1.056;  
AF=AF*1.056;  
WR=WR*1.056;  
W=W*1.056;
```

```
yfam=yfam*1.056;  
ylfam=ylfam*1.056;  
ytfam=ytfam*1.056;  
YTP=YTP*1.056;  
YTA=YTA*1.056;  
YTA1=YTA1*1.056;  
YTA2=YTA2*1.056;  
ymfam=ymfam*1.056;  
ycfam=ycfam*1.056;  
YCA=YCA*1.056;  
YCF=YCF*1.056;  
run;
```

### A.2.5 Creazione dei redditi familiari equivalenti

```
if n_comp=1 then yfam_equiv=yfam/0.599;  
if n_comp=2 then yfam_equiv=yfam/1;  
if n_comp=3 then yfam_equiv=yfam/1.335;  
if n_comp=4 then yfam_equiv=yfam/1.632;  
if n_comp=5 then yfam_equiv=yfam/1.905;  
if n_comp=6 then yfam_equiv=yfam/2.150;  
if n_comp=>7 then yfam_equiv=yfam/2.401;  
run;
```

### A.2.6 Sottocampioni e popolazione

#### Famiglie

Il data set A200x contiene per ogni anno il prospetto di come le famiglie vengono incluse o eliminate dal campione anno dopo anno. Attraverso questo programma si sono selezionate le famiglie in modo tale che comparissero nel campione solamente una volta. (Criterio *fresh*)

```
proc sort data=individuali0x;  
by nquest;  
run;
```



```
proc sort
data=familiari0x;
by nquest;
run;
data datiaggr.nuove0x;
merge  individuali0x familiari0x;
bynquest;
run;
proc sort data=selez.A200x;
by nquest;
run;
data popol0x;
merge datiaggr.nuove0x selez.A200x;
run;
data datiaggr.popol0x;
if CTR='W6-1' or CTR='W5-2' or CTR='W4-3'
    or CTR='W3-4' or CTR='W2-5' or CTR='S-00';
run;
```

### Popolazione

```
data datiaggr.popol;
merge datiaggr.popol00 datiaggr.popol02 datiaggr.popol04;
by nquest;
run;
```

### Sottocampioni

Con il seguente programma si sono selezionate tutte le famiglie con figli diplomati di età compresa tra i 18 e i 22 anni; sono state quindi divise in base al tipo di occupazione dei figli: non studenti nel *Sottocampione1* e studenti nel *Sottocampione2*. Questa prima operazione è stata eseguita per ciascuno degli anni considerati, quindi, i campioni ottenuti per ciascun anno sono stati uniti attraverso il pooling.

```
data sottocampione_0x;
set datiaggr.popol0x;
```

```
if campione_18_22=1  
and diploma_18_22=1;  
run;
```

```
data rischio1_0x;  
set sottocampione_0x;  
if studente_18_22=0;  
run;
```

```
data rischio2_0x;  
set sottocampione_0x;  
if studente_18_22=1;  
run;
```

```
data rischio.sottocampione1;  
set rischio1_00 rischio2_02 rischio2_04;  
run;
```

```
data rischio.sottocampione2;  
set rischio2_00 rischio2_02 rischio2_04;  
run;
```

### Creazione dei pesi

```
data rischio.popolp;  
set datiaggr.popol;  
pesop=_peso*8012/16783.1;  
run;
```

```
data p_weight;  
set rischio.popolp;  
keep nquest pesop;  
run;
```

```
proc sort data=p_weight;  
by nquest;  
run;
```

```
proc sort data=rischio.sottocampione1;  
by nquest;  
run;
```

```
proc sort data=rischio.sottocampione2;  
by nquest;  
run;
```

```
data rischio.sottocampione1;  
merge p_weight rischio.sottocampione1;  
if n_comp>0;  
run;
```

```
data rischio.sottocampione2;  
merge p_weight rischio.rischio3;  
if n_comp>0;  
run;
```



# Appendice B

## Le variabili utilizzate

Nell'Appendice A, è stato riportato il programma in SAS per la creazione delle variabili utilizzate per l'analisi dei dati e alle quali si fa spesso riferimento nei capitoli centrali. Di seguito è quindi spiegato il significato di ciascuna variabile

### Variabili categoriali utilizzate nel terzo capitolo

**area=** regione di residenza della famiglia

1 = *Nord – Ovest*; (B.1)

2 = *Nord Est*; (B.2)

3 = *Centro*; (B.3)

4 = *Sud*; (B.4)

5 = *Isole* (B.5)

**cleta=** classe d'età del capofamiglia

1 = *fino a 30 anni*; (B.6)

2 = *da 31 a 40 anni*; (B.7)

3 = *da 41 a 50 anni*; (B.8)

4 = *da 51 a 65 anni*; (B.9)

5 = *oltre 65 anni* (B.10)

**studio**= titolo di studio

1 = *nessuno*; (B.11)

2 = *licenza elementare*; (B.12)

3 = *licenza media*; (B.13)

4 = *diploma*; (B.14)

5 = *laurea*; (B.15)

6 = *specializzazione post – laurea*; (B.16)

**apqual**= qualifica lavorativa

1 = *operaio e posizioni assimilate*; (B.17)

2 = *impiegato, insegnante*; (B.18)

3 = *impiegato direttivo/quadro*; (B.19)

4 = *dirigente, preside, direttore didattico*, (B.20)

*docente universitario, magistrato*; (B.21)

5 = *libero professionista*; (B.22)

6 = *imprenditore*; (B.23)

7 = *lavoratore in proprio*; (B.24)

8 = *coadiuvante*; (B.25)

## Variabili di reddito

**yfam**= reddito totale familiare

**ylfam**= reddito da lavoro dipendente familiare

**ymfam**= reddito da libera professione e impresa familiare

**ytfam**= reddito da pensioni e trasferimenti netti familiari

**ycfam**= reddito da capitale familiare

**WR**= reddito da ricchezza familiare

**consumi**= consumi familiari

**BD**= beni durevoli

**PF**= passività finanziarie

**AF**= attività finanziarie

**W**= ricchezza netta familiare

**YTP**= pensioni e arretrati

**YTA**= altri trasferimenti

**YTA1**= assistenza economica

**YTA2**= borse di studio

**YCA**= reddito da fabbricati

**YCF**= reddito da capitale finanziario

**YC**= reddito da capitale

**cl\_yfam**= classi di reddito familiare

$$1 = \text{se } 0 < yfam \leq 10000; \quad (\text{B.26})$$

$$2 = \text{se } 10000 < yfam \leq 20000; \quad (\text{B.27})$$

$$3 = \text{se } 20000 < yfam \leq 30000; \quad (\text{B.28})$$

$$4 = \text{se } 30000 < yfam \leq 40000; \quad (\text{B.29})$$

$$5 = \text{se } 40000 < yfam \leq 50000; \quad (\text{B.30})$$

$$6 = \text{se } 50000 < yfam \leq 60000; \quad (\text{B.31})$$

$$7 = \text{se } 60000 < yfam \leq 70000; \quad (\text{B.32})$$

$$8 = \text{se } yfam > 70000; \quad (\text{B.33})$$

### **Altre variabili**

**nquest**= numero del questionario

**nord**= numero di ordine del componente

**parent=** relazione con il capofamiglia

$$1 = \text{capofamiglia}; \quad (\text{B.34})$$

$$2 = \text{coniuge/convivente del capofamiglia}; \quad (\text{B.35})$$

$$3 = \text{figlio del capofamiglia}; \quad (\text{B.36})$$

$$4 = \text{altro parente}; \quad (\text{B.37})$$

$$5 = \text{altrocomponente non parente}; \quad (\text{B.38})$$

**apnon=** non occupati

$$1 = \text{in cerca di prima occupazione}; \quad (\text{B.39})$$

$$2 = \text{disoccupato}; \quad (\text{B.40})$$

$$3 = \text{casalinga}; \quad (\text{B.41})$$

$$4 = \text{benestante}; \quad (\text{B.42})$$

$$5 = \text{pensionato}; \quad (\text{B.43})$$

$$6 = \text{studente}; \quad (\text{B.44})$$

$$7 = \text{altra condizione}; \quad (\text{B.45})$$

$$8 = \text{bambini fino a 6 anni}; \quad (\text{B.46})$$

**godabit=** titolo di godimento dell'abitazione

$$1 = \text{propria}; \quad (\text{B.47})$$

$$2 = \text{in affitto}; \quad (\text{B.48})$$

$$3 = \text{a riscatto}; \quad (\text{B.49})$$

$$4 = \text{in usufrutto}; \quad (\text{B.50})$$

$$5 = \text{in usograttuito}; \quad (\text{B.51})$$

**cf\_sesso=** sesso del capofamiglia

**cf\_cleta=** classe d'età del capofamiglia

**cf\_staci=** stato civile del capofamiglia

$$1 = \text{coniugato/convivente}; \quad (\text{B.52})$$



$$2 = \textit{celibe/nubile}; \quad (\text{B.53})$$

$$3 = \textit{separato/divirziato}; \quad (\text{B.54})$$

$$4 = \textit{vedovo/vedova}; \quad (\text{B.55})$$

**cf\_stud**= titolo di studio del capofamiglia

**cf\_apqua**= qualifica professionale del capofamiglia

**cf\_apset**= settore lavorativo del capofamiglia

**cf\_apnon**= motivo per cui il capofamiglia non è occupato

**co\_stud**= titolo di studio del coniuge del capofamiglia

**co\_apqua**= qualifica professionale del coniuge del capofamiglia

**co\_apset**= settore lavorativo del coniuge del capofamiglia

**co\_apnon**= motivo per cui il coniuge del capofamiglia non è occupato

**n\_figli**= numero di figli

**n\_altri**= numero di altri componenti

**n\_anz**= numero di individui di età superiore ai 65 anni

**n\_bamb**= numero di individui di età compresa tra 0 e 3 anni

**n\_bscuo**= numero di bambini di età compresa tra 4 e 14 anni

**n\_campione\_18\_22**= numero di giovani di età compresa tra 18 e 22 anni

**p\_attiva**= numero di individui di età compresa tra i 14 e i 65 anni

**n\_flav**= numero di individui forza lavoro

**n\_nflav**= numero di individui non forza lavoro

**n\_occ**= numero di occupati

**n\_dis**= numero di disoccupati

**n\_perc**= numero di percettori di reddito

**n\_1a\_occ**= numero di individui in cerca di prima occupazione

**n\_dip**= numero di lavoratori dipendenti

**n\_aut**= numero di lavoratori autonomi

**n\_pens**= numero di pensionati

**n\_cas**= numero di casalinghe

**n\_stud**= numero di studenti

**n\_studente\_18\_22**= numero di studenti di età compresa tra i 18 e i 22 anni

**n\_diploma\_18\_22**= numero di diplomati tra i 18 e i 22 anni

**n\_occupato\_18\_22**= numero di occupati tra i 18 e i 22 anni

**n\_primaocc\_18\_22**= numero di giovani in cerca di prima occupazione tra i 18 e i 22 anni

### Variabili per il modello logit

Sono riportate di seguito le altre variabili costruite per la stima del modello che non sono già state spiegate nel quinto capitolo perchè non incluse nel modello finale.

$$propabit = \begin{cases} 1 & \text{se la casa è di proprietà} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.56})$$

$$d_{banc\_post} = \begin{cases} 1 & \text{se la famiglia possiede un deposito} \\ & \text{bancario o postale} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.57})$$

$$cart_{banc} = \begin{cases} 1 & \text{se la famiglia possiede almeno una carta} \\ & \text{di credito o un bancomat} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.58})$$

$$altri = \begin{cases} 1 & \text{se ci sono altri componenti nella famiglia} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.59})$$

$$l_{aut} = \begin{cases} 1 & \text{se il capofamiglia svolge un lavoro autonomo} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.60})$$

$$figli2\_y2 = \begin{cases} 1 & \text{se la famiglia ha due figli e appartiene al} \\ & \text{secondo quartile del reddito} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.61})$$

$$figli2\_y3 = \begin{cases} 1 & \text{se la famiglia ha due figli e appartiene al} \\ & \text{terzo quartile del reddito} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.62})$$

$$figli2\_y4 = \begin{cases} 1 & \text{se la famiglia ha due figli e appartiene al} \\ & \text{quarto quartile del reddito} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.63})$$

$$figli3\_y2 = \begin{cases} 1 & \text{se la famiglia ha tre figli e appartiene al} \\ & \text{secondo quartile del reddito} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.64})$$

$$figli3\_y3 = \begin{cases} 1 & \text{se la famiglia ha tre figli e appartiene al} \\ & \text{terzo quartile del reddito} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.65})$$

$$figli3\_y4 = \begin{cases} 1 & \text{se la famiglia ha tre figli e appartiene al} \\ & \text{quarto quartile del reddito} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.66})$$

$$nord\_ovest = \begin{cases} 1 & \text{se la famiglia risiede nel nord-ovest} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.67})$$

$$nord\_est = \begin{cases} 1 & \text{se la famiglia risiede nel nord\_est} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.68})$$

$$centro = \begin{cases} 1 & \text{se la famiglia risiede nel centro} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.69})$$

$$sud = \begin{cases} 1 & \text{se la famiglia risiede nel sud} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.70})$$

$$isole = \begin{cases} 1 & \text{se la famiglia risiede nelle isole} \\ 0 & \text{altrimenti} \end{cases} \quad (\text{B.71})$$

$$(\text{B.72})$$



# Appendice C

## I modelli stimati

Prima di raggiungere il modello logit definitivo, sono stati stimati alcuni altri modelli che però, non spiegavano in modo adeguato la decisione di iscriversi o meno all'università.

Il primo modello stimato includeva le seguenti variabili:

- consumi;
- W;
- ylfam;
- ytfam;
- ymfam;
- ycfam;
- figli2;
- figli3;
- altri;
- co\_diploma;
- co\_lausp;
- cf\_diploma;

- cf\_lausp;
- l\_aut;
- co\_lav.

Tra queste tuttavia risultavano significative solamente le variabili di reddito, il titolo di studio del capofamiglia e del suo coniuge, il numero dei figli, il fatto che i genitori vivessero insieme o separati e il fatto che il coniuge lavorasse.

Si è quindi stimato un ulteriore modello che escludesse tali variabili: il confronto della log-verosimiglianza indica senza dubbio di accettare il secondo modello ridotto.

In questo secondo modello, tuttavia, il coefficiente della dummy riferita al tipo di lavoro del capofamiglia (dipendente o autonomo), non risulta più significativo: probabilmente l'effetto di questa variabile è catturato dall'utilizzo dei diversi tipi di reddito.

Per il modello successivo, si è quindi deciso di utilizzare al posto delle quattro variabili di reddito solamente la variabile yfam (reddito familiare): in tal modo infatti, la dummy l\_aut diviene significativa, come ci si attendeva, e anche il criterio AIC per il confronto di modelli non annidati suggerisce di accettare questo terzo modello.

Infine, per ottenere un coefficiente più facilmente interpretabile, si è utilizzata la trasformazione logaritmica del reddito, ottenendo così il seguente modello:

#### Analisi delle stime di massima verosimiglianza

Parametro	DF	Stima	Errore standard	Wald Chi-quadr	P-value
Intercept	1	9.9326	1.3130	57.2230	<.0001
lyfam	1	-0.5940	0.1142	27.0558	<.0001
con2	1	-0.2221	0.1074	4.2801	0.0386
cf_diploma	1	0.2995	0.0771	15.0815	0.0001
cf_lausp	1	0.7841	0.1817	18.6288	<.0001
co_diploma	1	0.3542	0.0920	14.8305	0.0001
co_lausp	1	0.8014	0.2328	11.8540	0.0006
figli2	1	0.6012	0.0768	61.3467	<.0001
figli3	1	0.8846	0.1089	66.015	<.0001
co_lav	1	0.2281	0.0783	8.4919	0.0036

---

l_aut	1	0.1820	0.0986	3.4082	0.0649
-------	---	--------	--------	--------	--------

Il modello appena stimato potrebbe essere accettato: tutte le variabili sono significative e il criterio della log-verosimiglianza e di AIC, in seguito ai confronti con gli altri tentativi, suggeriscono di accettare questa stima. In linea con le analisi viste nei capitoli 3, 4 e 5, infatti, le variabili riferite al titolo di studio del capofamiglia e del coniuge sono significative e hanno un impatto positivo e crescente sulla probabilità di iscriversi all'università. Il reddito invece ha un effetto negativo: all'aumentare della disponibilità economica diminuisce la probabilità di iscriversi all'università.

Tuttavia si è deciso di stimare un ulteriore modello, quello definitivo (capitolo 6), che considerasse il reddito non più come variabile continua ma come variabile discreta. Al posto del logaritmo del reddito sono quindi state introdotte nel modello tre variabili dummies che indicassero il secondo, il terzo ed il quarto quartile del reddito. In tal modo si è potuto verificare il diverso e graduale impatto che l'aumento del reddito ha sulla variabile risposta.

In questo modello però la dummy riferita al tipo di lavoro (tabella successiva) del capofamiglia non è più significativa: è stata quindi eliminata dal modello ottenendo in tal modo le stime definitive.

Parametro	DF	Stima	Errore standard	Wald Chi-quadr	P-value
Intercept	1	3.1839	0.3076	107.1569	<.0001
figli2	1	0.5946	0.0763	60.7406	<.0001
figli3	1	0.8906	0.1085	67.3264	<.0001
co_lav	1	0.2077	0.0780	7.0936	0.0077
yfam2	1	-0.1156	0.0942	1.5062	0.2197
yfam3	1	-0.2770	0.0968	8.1776	0.0042
yfam4	1	-0.3944	0.1020	14.9343	0.0001
cf_diploma	1	0.2916	0.0767	14.4392	0.0001
cf_lausp	1	0.7391	0.1804	16.7845	<.0001
co_diploma	1	0.3407	0.0914	13.9002	0.0002
co_lausp	1	0.7713	0.2322	11.0312	0.0009
con2	1	-0.2499	0.1067	5.4870	0.0192
l_aut	1	0.1568	0.0980	2.5631	0.1094

Si è cercato poi di verificare se le dummies riferite alla regione di residenza fossero significative o meno ma, introducendole in questo ultimo modello stimato, si ottiene che i loro coefficienti non sono significativi.



# Bibliografia

- [1] Banca d'Italia (2002, 2004, 2006), I Bilanci delle famiglie italiane nell'anno [2000, 2002, 2004]. Supplemento al Bollettino statistico, Roma, Banca d'Italia.
- [2] Checchi D. (2003), The Italian educational system, family background and social stratification, WP Gennaio 2003, Dipartimento di Economia Politica, Università degli Studi di Milano
- [3] Istat (2005), L'istruzione della popolazione al 2001.
- [4] Istat (2003), La prima indagine sulla spesa per l'istruzione delle famiglie italiane.
- [5] Trivellato U. (2004), Appunti di statistica economica a.a. 2004/2005.