

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE STATISTICHE



DOTTORATO DI RICERCA IN STATISTICA APPLICATA ALLE SCIENZE

ECONOMICHE E SOCIALI

CICLO XIX

Il rischio di disagio nelle famiglie venete

Supervisore: Ch.mo Prof. Luigi Fabbris

Coordinatore: Ch.mo Prof. Corrado Provasi

Dottoranda: SARA POFTE

PADOVA, 31/12/2006

Indice

Introduzione	3
1 Metodi e tecniche della ricerca	5
1.1 La famiglia	5
1.2 Le fonti documentarie	9
1.2.1 Osservatori regionali	9
1.2.2 L'indagine multiscopo sulle famiglie condotta dall'Istat . . .	10
1.2.3 L'indagine ILFI sulle famiglie italiane	11
1.2.4 Siti Internet	13
1.3 Il disegno di campionamento	16
1.4 Il metodo di rilevazione	17
1.5 Il questionario	24
2 L'analisi dei dati	29
2.1 Caratteristiche anagrafiche delle famiglie venete	29
2.2 Istruzione e condizione professionale	32
2.3 Composizione delle famiglie	36
2.4 Abitazione	40
2.5 Famiglie unipersonali	44
2.6 Famiglie composte da un genitore con figli a carico	47
2.7 Stili di vita	50
2.7.1 Attitudine al fumo e all'alcool	50
2.7.2 Letture	52
2.7.3 Abitudini alimentari	56

2.8	I figli	59
2.8.1	Figli in età 0-2	60
2.8.2	Figli in età 3-5	61
2.8.3	Figli in età 6-14	62
2.8.4	Figli in età 15-19	63
2.8.5	Problemi di apprendimento scolastico	68
2.8.6	Figli con più di 20 anni	73
2.9	Gli anziani	75
2.10	Stati critici	80
2.10.1	Invalidità	80
2.10.2	Presenza di malati cronico-degenerativi	82
2.10.3	Presenza di disoccupati	84
2.11	Eventi critici	86
2.12	Reti familiari	88
2.13	Servizi	94
3	La ricerca dei gruppi a rischio	99
3.1	L'analisi di segmentazione	99
3.1.1	La variabile dipendente	101
3.1.2	Le variabili predittive	101
3.1.3	La funzione criterio	102
3.2	Misure di separatezza tra popolazioni rispetto a gradazioni di gravità di fenomeni	107
3.2.1	Proposte metodologiche	109
3.2.2	Applicazioni esemplificative	114
3.3	La regressione logistica e i modelli a odds proporzionali	120
4	La percezione del disagio da parte della famiglia	125
4.1	Percezione del disagio	125
4.1.1	Stati critici	127
4.1.2	Eventi critici	139
4.1.3	Influenza di più fattori nella determinazione del disagio	143

4.2	Il rischio di disagio familiare e le sue determinanti	146
4.2.1	Cross-Validation	168
4.3	Percezione del disagio su scala dicotomica	171
5	Modello di analisi neurale	181
5.1	Le reti neurali	181
5.1.1	Pregi e difetti delle reti neurali	186
5.2	Cenni storici	188
5.3	Applicazione	191
5.3.1	Un'ulteriore analisi	217
6	Modelli per dati correlati	221
6.1	Introduzione	221
6.2	I modelli marginali per dati categoriali	222
6.2.1	I modelli Generalized Estimating Equations (GEE)	223
6.2.2	Modello utilizzato nello studio	230
6.3	I modelli di regressione multilivello	245
6.3.1	Cenni storici	247
6.3.2	Formulazione dei modelli multilivello	248
6.3.3	Stima dei modelli multilivello	251
6.3.4	Applicazione	253
7	Conclusioni	261
	Bibliografia	267

Introduzione

Nel DPEF (Documento di Programmazione Economico Finanziaria, L.R. 35/2000) della Regione Veneto, la famiglia è considerata il “*nucleo fondamentale di riferimento del nostro sistema sociale poiché essa rappresenta il luogo di sviluppo e di crescita di tutti i suoi componenti*”. Essa è l’organizzazione sociale che per prima si occupa dei problemi delle persone. È all’interno del nucleo familiare che si intrecciano e si organizzano i legami che costituiscono i nostri contesti di vita e dai quali si dipanano le relazioni più fitte con il mondo esterno (Endrizzi, Fabbris, Martini, 2002).

Il ciclo di vita di gran parte delle famiglie va dalla formazione della coppia alla nascita e alla crescita dei figli, alla loro uscita dalla famiglia e, per concludere, alla morte dei coniugi. Ogni famiglia sperimenta, inoltre, eventi critici¹ che determinano necessariamente un suo cambiamento.

Gli studi sociali, pur nella diversità degli approcci utilizzati, hanno sempre cercato di ricondurre la famiglia a modelli (*pattern*) che oggi non risultano esauritivi per definire un ambito in continuo divenire e sempre più caratterizzato dalla individualità.

Nell’ultimo decennio, la famiglia ha dovuto far fronte ad una crescita delle attese dei soggetti che la compongono, delle sollecitazioni derivanti dal crescente impegno lavorativo, delle esigenze di cura e di assistenza derivanti dall’invecchiamento della popolazione, delle attrazioni provenienti dalla sfera del tempo libero e

¹Per *evento critico* si intende un accadimento di fronte al quale la famiglia trova difficoltà a reagire e comprende che le proprie forze risultano inadeguate per affrontare serenamente la nuova situazione.

del consumo, delle situazioni di crisi soprattutto presenti nell'età giovanile e nelle circostanze di disgregazione dell'unità familiare e delle difficoltà economiche.

Per studiare questi fenomeni, la Regione del Veneto, attraverso la gestione dell'Azienda ULSS 16 di Padova e il Centro Regionale di Documentazione ed Analisi sulla Famiglia², ha voluto promuovere la realizzazione della presente ricerca, commissionata al Dipartimento di Scienze Statistiche dell'Università di Padova, con la finalità di tracciare un quadro sulla situazione delle famiglie venete.

La ricerca si svolge attorno ad un progetto multidisciplinare riguardante approcci concorrenti per lo studio del disagio della famiglia veneta e la possibilità di prevenirlo e di alleviarne le conseguenze quando si manifesta.

Si intendono evidenziare le caratteristiche delle famiglie dal punto di vista demografico, socio-economico, pedagogico, urbanistico-ambientale, mettendo in luce le necessità delle stesse e monitorando gli eventuali cambiamenti della qualità della vita, per consentire agli organi competenti di progettare politiche ed interventi funzionali e formativi.

Quest'analisi costituirà per la Giunta Regionale un valido strumento di studio in grado di fornire indicazioni utili alla programmazione di interventi mirati, atti non solo a sostenere le famiglie in difficoltà ma anche a prevenire le situazioni più frequenti di disagio in un'ottica di promozione di una cultura centrata sulla famiglia.

²Il Centro Regionale di Documentazione e Analisi sulla Famiglia è stato fondato come luogo di analisi della famiglia veneta, di studio delle politiche a favore della stessa, di promozione della ricerca e dell'alta formazione sui temi della prevenzione e della mitigazione del disagio, di proposte progettuali, di sostegno e miglioramento delle competenze e delle conoscenze dei soggetti che, a vario titolo, si occupano di famiglia.

Fra le altre attività, il Centro prevede di costruire anche un servizio di "informa-famiglia", la cui realizzazione, che avviene collegando al Centro una serie di sportelli già presenti o attivandoli laddove assenti, è molto utile in quanto punto di raccolta per lo scambio bidirezionale di informazioni tra realtà familiari venete e Centro di Documentazione. Tale diffusione avviene anche attraverso la realizzazione di convegni e la pubblicazione di riviste volte sia agli operatori dei diversi servizi pubblici, sia alle famiglie della regione. Ciò può consentire di conoscere e capire, collegandoli tra loro, la maggior parte dei fenomeni e dei problemi sociali in maniera più propria di quanto non accada prendendoli in considerazione singolarmente.

Capitolo 1

Metodi e tecniche della ricerca

1.1 La famiglia

Dal punto di vista statistico (ISTAT, 2002), la *famiglia* individua un “*insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, dimoranti abitualmente nella stessa abitazione*”. Viene considerata *famiglia* anche una persona che vive sola. Non sono invece considerate famiglie i gruppi di persone che convivono per motivi religiosi, militari, di salute, di studio o penitenziari. Tale definizione di famiglia, recependo quella data dal nuovo Regolamento anagrafico (1989), differisce da quella adottata nei passati censimenti poiché è stato abolito quello che rappresentava il terzo elemento costitutivo della famiglia anagrafica e cioè l'unicità del bilancio corrispondente alla messa in comune del reddito da parte dei componenti della famiglia.

Un *nucleo familiare* è, invece, l'insieme delle persone che formano una relazione di coppia o di tipo genitore-figlio. Una famiglia può avere al suo interno un nucleo, può essere formata da un nucleo più altri membri, da più nuclei o da nessun nucleo (persone sole, famiglie composte da due fratelli, da un genitore con figlio separato, divorziato o vedovo).

Negli ultimi decenni la società è stata investita da profondi mutamenti quali l'industrializzazione, l'urbanizzazione, la modernizzazione culturale, l'entrata sem-

pre più stabile e costante della donna nel mondo del lavoro.

Questi cambiamenti hanno avuto un forte impatto sulla struttura familiare e sugli equilibri consolidati nel corso dei secoli. Il calo demografico, l'allungamento della vita media e il conseguente invecchiamento della popolazione e il passaggio dalla famiglia di tipo patriarcale a quella di tipo matriarcale hanno messo in discussione il ruolo e il significato stesso di famiglia. Contemporaneamente, l'aumento della cosiddetta *adolescenza lunga*, l'assunzione dell'impegno coniugale e genitoriale in età sempre più avanzata, la maggiore instabilità coniugale, l'intensificarsi dell'esperienza lavorativa sia per gli uomini che per le donne, il progressivo aumento del benessere economico sono solo alcuni fra i molti fattori di mutamento della famiglia.

Tali eventi hanno contribuito ad alterare le strutture familiari, aumentando la quota di famiglie monogenitoriali e ricostruite, di famiglie con anziani e di anziani e complessivamente di famiglie con un basso numero di componenti (Ciucci, 2001).

La famiglia, al pari degli altri sistemi, si struttura e si adatta in corrispondenza delle molteplici influenze ambientali. La configurazione e le funzioni familiari devono considerarsi strettamente connesse al contesto storico, sociale, economico e culturale a cui appartengono. La famiglia non può essere studiata come un'entità omogenea perché ogni famiglia è diversa dalle altre in termini di risorse, fasi del ciclo di vita, modelli culturali ed organizzativi. Come afferma Bersaglio (1998), sempre più spesso si parla di *famiglie* anziché di *famiglia*, proprio per rilevare la nascita di una molteplicità di realtà familiari. Anche se in Italia il matrimonio rimane la forma di convivenza più adottata, si stanno diffondendo modelli di vita familiare alternativi, che rispecchiano il pluralismo di orientamenti culturali dell'odierna società.

In ogni caso, l'essere il più comune luogo di interazione tra persone eterogenee in quanto a sesso, età, professione, rende la famiglia un osservatorio privilegiato per comprendere gli stili di vita della popolazione. Dall'analisi delle interazioni tra i componenti si può dare un significato compiuto ai comportamenti e alle aspettative delle persone, nonché agli eventi che accadono ai singoli e alla società nel complesso. È nella famiglia che si compongono le decisioni dei singoli che,

quando siano osservate in forma aggregata come fenomeni della società, definiscono i tratti fondamentali di una comunità. Alla famiglia, essendo il primo corpo sociale in cui si ammortizzano le situazioni di difficoltà, sono ricondotti anche la maggior parte dei problemi dei singoli. I problemi più rilevanti sono relativi alla salute, alla disponibilità economica, alla qualità dei rapporti sociali e possono essere legati agli stessi componenti o ad altre persone, alla inadeguatezza strutturale, organizzativa o logistica della famiglia oppure ad eventi critici, quali uno sfratto o un lutto (Breda, Fabbris, 2002).

Generalmente le famiglie di piccole dimensioni sono più problematiche delle altre, non per le dimensioni in sé, ma per minori opportunità di sostegno affettivo economico ed organizzativo (Endrizzi, Fabbris e Schievano, 2002). Si pensi, infatti, a quanto più difficoltosa può risultare la cura e la gestione di un anziano o di un invalido all'interno di famiglie numericamente limitate rispetto a quelle che possono contare sull'aiuto reciproco di un maggior numero di componenti.

Oggi più che in passato la famiglia lamenta disagi e manifesta segni di malessere, incontra difficoltà nel conciliare gli impegni di lavoro con le esigenze familiari. Il disagio familiare è una condizione di malessere psicologico o fisico che coinvolge tutti i componenti, anche se l'origine è di tipo individuale; è uno stato dell'animo complesso e multidimensionale che la famiglia vive in modo proprio e mutevole nel tempo.

È importante la distinzione tra disagio individuale e familiare: il primo è avvertito dal singolo individuo e non coinvolge necessariamente altre persone; il secondo, invece, riguarda tutta la famiglia, sebbene possa avere origine da uno solo dei suoi membri.

Breda e Fabbris (2002) definiscono due categorie di possibili cause di disagio:

- gli stati critici, ossia le condizioni negative della famiglia e dei singoli componenti, che possono creare difficoltà;
- gli eventi critici, ossia le gravi situazioni, che possono turbare e rivoluzionare la vita della famiglia, tra cui i lutti improvvisi, gli sfratti, le condizioni di cattiva salute.

Nella maggior parte dei casi la famiglia riesce a far fronte in maniera autonoma ai problemi che le si presentano, trovando una soluzione al proprio interno, sfruttando le basi solide su cui è costruita, o attraverso le risorse che derivano dalla cosiddetta famiglia "allargata", ossia dalle persone che vivono sotto altri tetti e hanno legami di sangue con i membri della famiglia stessa. In questi casi essa riesce a mettere in atto un processo di trasformazione che porta all'acquisizione di nuove e diverse capacità.

Qualora queste risorse risultassero inadeguate al superamento del problema, la famiglia può chiedere aiuto all'esterno (prima il vicinato e il volontariato¹, poi i servizi sociali² e in generale le strutture specializzate, deputate all'assistenza).

Si crea, pertanto, una situazione di crisi, di rottura di un precedente equilibrio e la famiglia, costretta a modificare le relazioni interne ed esterne, i propri tempi e le proprie abitudini, può vivere una situazione traumatica ed insostenibile. Nei casi più drammatici, l'unica cosa che la famiglia può fare è affidarsi al destino e subire passivamente il corso degli eventi. Quest'ultima è la situazione peggiore, una non scelta, a cui la famiglia si rimette quando neppure i servizi sono in grado di essere d'aiuto.

I mutamenti a cui è sottoposta la famiglia hanno spinto a concentrare l'attenzione sulla centralità della stessa e a studiarne le tendenze evolutive per valutare sia dal punto di vista qualitativo sia da quello quantitativo sue capacità e risorse, nonché l'eventuale promozione di servizi e strutture per far fronte alle difficoltà che incontra.

Le istituzioni e i servizi sociali hanno il compito di seguire questi cambiamenti, monitorare i problemi delle famiglie ed intervenire attraverso la sperimentazione di forme di interventi socio-familiari nuovi e sempre più flessibili, in grado di rispondere alle rapide trasformazioni in atto.

È necessario promuovere e far crescere un modello di intervento che riconosca

¹Persone, generalmente riunite in associazioni, che prestano attività volontaria e gratuita a favore della comunità.

²Ente pubblico che presta professionalmente un'attività intesa ad assistere i singoli membri di una categoria o di una comunità in difficili situazioni della loro vita privata.

il valore del mutuo aiuto per offrire alle famiglie concrete situazioni di sviluppo ed un adeguato sostegno nelle situazioni di difficoltà.

Per progettare percorsi di innovazione rispondenti alla famiglia che cambia occorre saper promuovere politiche innovative “dirette” alla famiglia. Per esempio, le famiglie che hanno al loro interno soggetti in difficoltà (anziani, disabili, tossicodipendenti,...) spesso sono oggetto di politiche sociali o assistenziali univocamente rivolte al soggetto in difficoltà, senza che venga presa in considerazione la globalità delle relazioni familiari che sono spesso risorse indispensabili per i servizi stessi. Inoltre, queste famiglie, quando presentano più problemi (le cosiddette *famiglie multiproblematiche*), vengono prese in carico da servizi o associazioni, senza che tra questi vi sia l'indispensabile *raccordo* attorno alla dimensione della famiglia nella sua globalità.

La ragione di questa scelta nuova consiste principalmente in una concezione dei servizi sociali che, volendo tener conto della complessità delle reti relazionali tra i soggetti, guarda alle persone da una prospettiva rinnovata: i servizi alla persona non possono essere progettati senza integrarli con i servizi alla famiglia.

1.2 Le fonti documentarie

1.2.1 Osservatori regionali

Nel Veneto sono stati istituiti diversi Osservatori che si occupano di problematiche connesse a situazioni particolari (handicap e dipendenze ad esempio) o a fasi specifiche della vita di una persona (infanzia, adolescenza, terza età). Ne diamo di seguito un rapido elenco:

- Osservatorio Regionale per l'infanzia e l'adolescenza;
- Osservatorio Regionale permanente sulla condizione giovanile;
- Osservatorio Regionale sull'handicap;
- Osservatorio Regionale sulla condizione della popolazione anziana;

- Osservatorio Regionale sulle dipendenze;
- Osservatorio Regionale per la tutela e la promozione della persona;
- Osservatorio Regionale sulla popolazione detenuta e in esecuzione penale esterna;
- Osservatorio Regionale epidemiologico.

Gli Osservatori funzionano da “sensore sociale”, ossia come uno strumento valutativo ed operativo che approfondisce i diversificati aspetti di particolari aree tematiche. Sono “laboratori” ideati per perseguire un duplice obiettivo:

- interpretare e proporre modelli di analisi dei fenomeni per elaborare strumenti di conoscenza utili alla programmazione sociale;
- svolgere funzioni di servizio, di promozione culturale e di punto di riferimento, in grado di sviluppare un efficiente sistema informativo sociale regionale.

Queste strutture, proprio per la loro stessa natura e per gli obiettivi che devono perseguire, si occupano di problematiche specifiche delle persone, ma mai della famiglia in quanto tale. Sono un centro per il monitoraggio di situazioni problematiche in cui è l'individuo, in quanto portatore delle stesse, a rientrare con il ruolo di protagonista.

1.2.2 L'indagine multiscopo sulle famiglie condotta dall'Istat

A partire da dicembre 1993, l'Istat ha avviato il nuovo corso delle *Indagini Multiscopo sulle famiglie*. Alla fine di ogni anno vengono rilevati gli aspetti fondamentali della vita quotidiana della popolazione ed il livello di soddisfazione dei cittadini rispetto al funzionamento dei servizi di pubblica utilità. I principali contenuti informativi dell'indagine sono: famiglia, abitazione e zona di residenza, istruzione e formazione, lavoro domestico ed extradomestico, spostamenti quotidiani, tempo libero e partecipazione sociale, stili di vita e condizioni di salute, consumo di

farmaci ed utilizzo dei servizi sanitari, criminalità, funzionamento dei servizi di pubblica utilità.

Nell'Indagine Multiscopo sulle famiglie le informazioni vengono rilevate con due questionari, uno autocompilato e l'altro compilato con l'aiuto di un intervistatore; quest'ultimo è diviso in una parte riferita alla famiglia e in una individuale, con una scheda riservata ad ogni componente. La parte familiare raccoglie informazioni relative alle caratteristiche strutturali, all'abitazione e alla zona di residenza, alle condizioni economiche e ai servizi di cui la famiglia fa uso. Le schede individuali riguardano i vari aspetti della vita quotidiana quali il lavoro, il tempo libero, la rete di supporto relazionale, gli spostamenti quotidiani, l'alimentazione, la salute, la fruizione e valutazione dei servizi sanitari e di quelli di pubblica utilità.

Le modalità di diffusione dei dati prevedono la pubblicazione annuale di quattro volumi brevi, ciascuno dedicato ad un particolare aspetto della vita quotidiana degli italiani:

- Stili di vita e condizioni di salute;
- Famiglia, abitazioni e sicurezza dei cittadini;
- Cultura, società e tempo libero;
- I servizi pubblici e di pubblica utilità: utilizzo e soddisfazione.

Il sistema di indagine Multiscopo prevede che, accanto all'indagine *Aspetti della vita quotidiana*, si affianchino, a cadenza quinquennale, altre indagini che approfondiscono tematiche particolari e un'indagine continua a cadenza trimestrale su *Viaggi e vacanze*.

1.2.3 L'indagine ILFI sulle famiglie italiane

Per cercare di descrivere e spiegare i meccanismi e le dinamiche di mutamento che negli ultimi decenni hanno ricevuto un notevole impulso in tutte le società occidentali avanzate, molte istituzioni accademiche hanno dato vita a complessi programmi di ricerca fondati su indagini empiriche molto dettagliate, volte ad analizzare le storie di vita di campioni rappresentativi di cittadini.

L'Indagine Longitudinale sulle Famiglie Italiane si è proposta due finalità principali: la prima ha un carattere essenzialmente descrittivo e consiste nel raccogliere una serie di informazioni sulla composizione, sulle fonti e sui livelli di reddito, sulle caratteristiche sociali e su quelle demografiche di un ampio campione rappresentativo di famiglie italiane; la seconda riguarda lo studio del mutamento sociale e consiste nel raccogliere un ampio insieme di informazioni di tipo retrospettivo e di tipo prospettico su ciascun membro adulto di ogni famiglia inclusa nello studio. Precisamente, la ricerca intende ricostruire la storia di vita di ciascun membro di ogni famiglia - dal momento della nascita al momento dell'ultima ondata di interviste (2005) - in relazione ai seguenti ambiti: mobilità geografica o residenziale, carriera scolastica e formativa, carriera lavorativa, origini sociali e famiglia (costituzione di matrimoni o convivenze e nascita o adozione di figli). Lo studio di queste particolari aree della vita quotidiana consentirà di analizzare i fenomeni migratori e la loro evoluzione nel tempo, i percorsi scolastici degli italiani attraverso le riforme, la diffusione della formazione professionale regionale, il mutamento della struttura occupazionale del nostro paese, l'andamento nel tempo dei flussi di mobilità sociale, l'evoluzione delle forme di convivenza familiare e della nuzialità, le trasformazioni nei comportamenti di fecondità e così via.

L'Indagine è uno studio panel prospettico con prima rilevazione retrospettiva. Nel corso della prima rilevazione, svolta nel 1997, sono state raccolte informazioni su tutti gli eventi rilevanti accaduti ai membri del campione nel periodo compreso fra la data della loro nascita e la data dell'intervista. In ciascuna delle rilevazioni successive tali informazioni sono state aggiornate registrando tutti gli eventi rilevanti accaduti ai membri del campione nel periodo compreso fra la data dell'intervista precedente e la data dell'intervista di volta in volta effettuata.

L'unità di analisi è la famiglia e la popolazione di riferimento è l'insieme delle famiglie residenti sul territorio italiano e registrate presso le anagrafi comunali alla fine del 1996. Per selezionare le famiglie da intervistare nel corso della prima rilevazione è stata utilizzata una procedura di campionamento stratificato a due stadi. La stratificazione è stata effettuata sulla base della zona geografica (Nord-ovest, Nord-est, Centro, Sud e Isole) e del tipo di comune (centro, area metropolitana, co-

muni limitrofi, altri comuni). All'interno di ciascuno strato il comune ha costituito l'unità primaria di campionamento (primo stadio di campionamento) e la famiglia l'unità secondaria di campionamento (secondo stadio di campionamento). All'interno di ciascun comune selezionato nel corso del primo stadio di campionamento le famiglie sono state estratte in modo casuale dalle liste anagrafiche. All'interno di ciascuna famiglia selezionata nel corso del secondo stadio di campionamento sono stati definiti intervistabili tutti i componenti di età uguale o superiore a 18 anni al momento dell'intervista. La popolazione del Trentino-Alto Adige è stata sovracampionata al fine di disporre di un numero sufficiente di casi per effettuare analisi specifiche significative a livello regionale. Complessivamente il campione originario è formato da 10423 individui, appartenenti a 4714 famiglie. Escludendo il sovracampionamento trentino, gli individui diventano 9770 e le famiglie 4407.

Dato il disegno della ricerca prescelto, il campione ha un carattere dinamico: in ciascuna delle rilevazioni successive alla prima sono stati reintervistati tutti gli individui inclusi nel campione iniziale, meno coloro che nel frattempo sono deceduti, più coloro che nel frattempo sono diventati eleggibili all'intervista (ex minorenni divenuti maggiorenni).

La ricerca è stata promossa grazie agli sforzi congiunti dell'Università di Trento, dell'Istituto Trentino di Cultura e dell'Istat, ai quali si sono successivamente aggiunte l'Università degli Studi di Bologna e l'Università degli Studi di Milano Bicocca.

1.2.4 Siti Internet

I siti Internet dedicati all'argomento si articolano generalmente in due grandi aree: alcuni si propongono come sostegno alle famiglie, fornendo informazioni utili su opportunità, servizi o iniziative, altri offrono materiale di documentazione specifico sull'argomento. Molto spesso, però, questa linea di demarcazione si fa piuttosto sottile e le due aree sopra individuate tendono a sovrapporsi. La logica di questa fusione risiede molto probabilmente nell'intento di non fornire informazione fine a se stessa, ma fruibile da parte delle famiglie, finalizzata alla presa di coscienza

za da parte delle stesse delle opportunità offerte dalle istituzioni e dai servizi di volontariato presenti sul territorio.

Di seguito si intende dare una rassegna, non esaustiva, su alcuni dei siti dedicati al tema della famiglia, elencandone caratteristiche e finalità.

- **www.cisf.it**: sito del *Centro Internazionale Studi Famiglia*, promosso dal Ministero del Lavoro e delle Politiche Sociali. Esso intende promuovere una cultura della famiglia, avvalendosi soprattutto dei seguenti mezzi:
 - un Centro di Documentazione informatizzato e specializzato sulle tematiche familiari;
 - un Comitato Scientifico formato da esperti in diverse discipline, che assiste la direzione del CISF nella programmazione annuale dei lavori e nella preparazione di convegni, seminari, incontri e pubblicazioni a carattere scientifico o informativo;
 - la pubblicazione di un rapporto sulla famiglia in Italia a cadenza biennale, affidato a centri di ricerca e a specialisti in varie discipline, ma orientati ai problemi della famiglia in rapporto alla realtà socio-culturale in cui essa vive;
 - la promozione della ricerca scientifica su temi concernenti la vita familiare;
 - la pubblicazione di saggi e studi che siano particolarmente rivolti alla condizione storica della famiglia e alla definizione dei suoi ruoli specifici nella società contemporanea;
 - la collaborazione alla rivista *Famiglia Oggi*, mensile rivolto prevalentemente ad operatori sociali e pastorali, con taglio monografico, interdisciplinare e documentaristico;
- **www.informafamiglie.it**: banca dati relativa al territorio dell'Emilia Romagna, in cui si trovano informazioni raccolte ed organizzate dai Centri per le famiglie in collaborazione con l'Assessorato Regionale alle Politiche Sociali. Suddivise in dodici aree tematiche, si possono trovare informazioni

utili alla famiglia: dall'espletamento delle pratiche burocratiche necessarie per un matrimonio ai consigli per le coppie in attesa di un figlio o che desiderano averne uno, dalle adozioni alle scuole e ai servizi educativi per l'infanzia, al tempo libero. Queste informazioni vengono poi integrate dalle redazioni locali di quattordici comuni che aggiungono per ciascun argomento informazioni specifiche del comune di riferimento;

- **www.osservatorionazionalefamiglie.it**: portale dell'Osservatorio Nazionale della Famiglia, basato su una convenzione tra il Ministero del Lavoro e delle Politiche Sociali e il Comune di Bologna. L'osservatorio nasce con il compito di creare una rete per lo scambio di conoscenze tra le istituzioni nazionali, regionali e locali e le associazioni, ai fini dell'individuazione di nuove strategie di sostegno alle famiglie italiane. Della struttura fanno parte i rappresentanti di 25 comuni italiani. Scopi principali dell'Osservatorio sono l'attivazione di una rete articolata a livello territoriale che consenta un costante e diffuso scambio di conoscenze, esperienze e buone pratiche in materia di sostegno alla famiglia, nonché il rafforzamento della collaborazione interistituzionale in questo ambito ed un consolidamento del confronto con il mondo delle associazioni;
- **www.dirittoefamiglia.it**: sito dell'omonima rivista in cui appare una raccolta ragionata di materiale informativo giuridico, insieme a richiami legislativi e ad articoli che si occupano della famiglia e dei minori, nonché siti e documentazione rintracciabili sulla rete Internet. Lo sforzo sta nel fornire informazioni fruibili da chiunque, addetto ai lavori, cultore della materia o semplice interessato. Nella sezione Documenti vengono riportati articoli e saggi sulla scienza e sulla cultura della famiglia. Si trovano argomenti che spaziano dalle adozioni alla bioetica, alla condizione femminile, ai minori, alla violenza domestica e familiare. Un'altra sezione consente agli utenti di porre domande e quesiti all'attenzione di esperti del settore; le risposte vengono poi pubblicate nel sito.

1.3 Il disegno di campionamento

L'unità di analisi è la famiglia veneta e la popolazione di riferimento è l'insieme delle famiglie reperite attraverso gli elenchi telefonici ed intervistate nel periodo dicembre 2004 - aprile 2005. I dati sono di tipo trasversale, il che consente di rappresentare il fenomeno in un determinato istante temporale, concentrando l'attenzione sulle famiglie che, rispetto a quel riferimento, si trovano in una situazione di disagio.

Ai fini di rilevazione si è considerata la *famiglia di fatto*³, come si presenta al momento della rilevazione, indipendentemente dalla sua composizione anagrafica, dal tipo di legami che intercorrono tra i componenti e dall'eventuale assenza di qualche membro al momento dell'intervista. Inoltre le famiglie non sono state selezionate in base ai loro problemi e non si conoscono possibili situazioni di disagio al loro interno prima dell'intervista.

I criteri che dovrebbero garantire la rappresentatività del campione ai fini di stima delle principali grandezze sono la stratificazione della popolazione e la casualità della selezione dalle liste. La stratificazione è stata effettuata sulla base della provincia di residenza, informazione nota prima dell'intervista; da ciascuno strato (provincia) è stato poi selezionato un campione con metodo sistematico "a serpentina" che garantisce la casualità della scelta e la stratificazione secondo il criterio di ordinamento delle unità (Fabbris, 1989). In base a questo criterio, le unità appartenenti ad uno stesso strato avevano la medesima probabilità di far parte del campione.

La numerosità del campione estratto dallo strato h è data dalla numerosità totale del campione moltiplicata per la frazione di unità che appartengono allo strato h rispetto al totale della popolazione (w_h , peso dello strato):

$$n_h = \frac{nN_h}{N} = nw_h \quad (1.1)$$

³Insieme di persone che vivono sotto lo stesso tetto, unite tra loro da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi. Non sono considerate famiglie i gruppi di persone che convivono per motivi religiosi, militari, di salute, di studio o quant'altro. Una famiglia può inoltre essere composta da una sola persona.

Sono state selezionate complessivamente 3100 famiglie. Al fine di rappresentare i fenomeni su base provinciale con uguale grado di attendibilità, la numerosità attesa del campione di famiglie rispondenti è di circa 350 per provincia. In totale le famiglie che sono risultate collaborative ammontano a 2465.

Nel successivo paragrafo si dà una descrizione esaustiva dell'ammontare di interviste completate, della loro durata, dei motivi di mancato contatto di alcune famiglie e di altre caratteristiche relative all'intervista.

Per la stima basata sull'intero campione sarà necessario introdurre dei pesi, dati dal rapporto tra il numero di famiglie che possiedono un numero di telefono e il numero di famiglie campionate, in ciascuna provincia.

1.4 Il metodo di rilevazione

La rilevazione è stata di tipo CATI (*Computer Assisted Telephone Interviewing*), ossia telefonica con l'assistenza del computer.

Le interviste sono state effettuate presso il centro TECHNÈ (*Telephone and Electronic Computer-Helped New-survey Environment*) del Dipartimento di Scienze Statistiche dell'Università di Padova. Il *software* adottato per la realizzazione del questionario informatico è CAPTOR 2.6.3 (Capiluppi, 2000). Esso consente la gestione delle interviste telefoniche per questionari strutturati nonché la memorizzazione dei dati in un *database* in formato elettronico immediatamente utilizzabile per le analisi. Oltre a comportare costi piuttosto ridotti, un punto di forza di questa tecnica, che la rende particolarmente adeguata per questo tipo di indagine, è quello di controllare e ridurre l'errore dell'intervistatore.

Per le famiglie che non rispondono alla prima chiamata, il *software* prevede una successione di chiamate in ore e giorni diversi per massimizzare la probabilità di contatto. Le famiglie irreperibili a causa dei cosiddetti *errori di lista*, ossia inesattezze ed imprecisioni all'interno della lista di campionamento (nel nostro caso l'elenco telefonico), sono state sostituite con altre famiglie appositamente selezionate in soprannumero nel campione.

Fondamentale è stato inoltre lo studio del contatto telefonico con il rispondente

per ridurre il più possibile i rifiuti. La presentazione dell'indagine doveva essere breve ed incisiva, ma al tempo stesso dare risalto agli aspetti primari dell'indagine ed ai suoi obiettivi per predisporre in modo positivo il rispondente e motivarlo a fornire risposte precise ed accurate.

A ciascuna famiglia facente parte del campione è stata recapitata una lettera nella quale si anticipava la realizzazione dell'intervista spiegandone le finalità e garantendo l'utilizzo dei dati forniti nel rispetto della legge sulla *privacy*. Nella lettera è anche stato fornito un recapito telefonico al quale la famiglia poteva fare riferimento per avere ulteriori delucidazioni ed assicurarsi sulla reale origine dell'indagine.

Per quanto riguarda la persona da interpellare, si richiedeva preferibilmente di intervistare il capofamiglia⁴ o il coniuge. Se questi non erano presenti, si considerava il primo adulto consapevole e responsabile, disposto a dare informazioni sulla famiglia. Poteva quindi presentarsi il problema dei rispondenti *proxy*, ossia i casi in cui chi fornisce le informazioni non è a conoscenza delle abitudini delle persone per cui riferisce. Tale fenomeno è particolarmente sentito per le domande relative alle opinioni o alle abitudini personali o preferenze di ciascun componente, domande comunque non fondamentali per l'indagine. Per gli argomenti di più diretto interesse, quali quelli che riguardano malattie e lavoro, è presumibile che il rispondente fornisca informazioni attendibili anche per gli altri componenti della famiglia.

Durante l'intervista era fondamentale formulare le domande in modo semplice e comprensibile, prestando attenzione alle parole e alle frasi usate, vista l'eterogeneità della popolazione. Il rilevatore doveva essere in grado di condurre l'intervista in modo da non provocare imbarazzo o turbamento nel rispondente, data la delicatezza di alcune tematiche trattate, e di fornirgli le istruzioni opportune per selezionare le informazioni più rilevanti e collocarle nel giusto periodo di riferimento. In quest'ultimo caso era necessario prestare attenzione agli effetti di *telescoping*, che si verificano quando gli eventi accaduti a ridosso dell'inizio di un periodo preso

⁴Persona indicata come riferimento nello stato di famiglia anagrafico ("foglio di famiglia"). È necessaria per determinare la posizione relativa dei componenti della famiglia.

come riferimento vengono comunque riportati dai rispondenti, con evidente effetto di sovrastima.

Per questi motivi, gli intervistatori reclutati per l'indagine sono stati istruiti e supervisionati soprattutto nella fase dell'indagine pilota. Inoltre, per il controllo della loro attività è stata applicata la tecnica dei campioni compenetrati (Fabbris, 1989). I campioni compenetrati sono tali perché selezionati dalla stessa popolazione con il medesimo processo di selezione. La formazione di campioni compenetrati permette di valutare le differenze tra campione e campione attribuibili al caso e, quindi, di stimare le componenti di variabilità delle stime anche quando non sia nota la distribuzione dello stimatore. La tecnica dei campioni compenetrati, mediante assegnazione casuale di un sub-campione ad ogni addetto, si applica, pertanto, ai fini di stima della varianza del rilevatore.

Le interviste completate ammontano, per l'intera regione, a 2465, ossia al 58,8% dei tentativi di contatto telefonico. Le percentuali variano da provincia a provincia, passando dal 53,6% per Venezia, la provincia meno collaborativa, al 63,0% per Treviso. Le famiglie che si sono rifiutate di portare a termine l'intervista sono l'1,8%, mentre quelle che si sono rifiutate di rispondere sono ben il 26,0%. Il restante 13,5% rappresenta, infine, le famiglie che, per diverse ragioni (errori di lista, numeri telefonici corrispondenti a case di villeggiatura), risultano irreperibili (Tab.1.1).

Tab.1.1. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per provincia di residenza ed esito del contatto telefonico*

Provincia	Esito del contatto telefonico				Totale
	Interviste completate	Famiglie irreperibili	Rifiuto di rispondere	Rifiuto di completare l'intervista	
Belluno (n=594)	55,4	19,5	23,6	1,5	100,0
Padova (n=596)	61,4	10,2	25,2	3,2	100,0
Rovigo (n=584)	61,6	11,3	25,2	1,9	100,0
Treviso (n=583)	63,0	11,5	24,0	1,5	100,0
Venezia (n=642)	53,6	14,0	31,0	1,4	100,0
Verona (n=587)	58,9	13,6	25,9	1,5	100,0
Vicenza (n=585)	58,3	13,8	26,7	1,2	100,0
Totale (n=4171)	58,8	13,5	26,0	1,8	100,0

I motivi di mancato contatto sono svariati. I principali risultano essere: non risponde nessuno al telefono o risponde la segreteria telefonica (4,7% delle interviste totali), numero telefonico inesistente, errato, cambiato o non attivo (2,6%) e telefono occupato (1,2%). Nel primo e nel terzo caso sono stati fatti in media, rispettivamente, 10 e 7 tentativi senza ottenere comunque esito positivo.

Per le famiglie che hanno portato a termine l'intervista, sono stati necessari in media 2,5 tentativi di contatto; la maggior parte delle interviste complete (92,1%) sono state effettuate al più dopo 5 tentativi. Per le famiglie che si sono rifiutate di rispondere all'intervista, sono stati effettuati in media 2,5 tentativi prima di determinare completamente la mancata collaborazione da parte della famiglia (Tab.1.2).

Tab.1.2. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per esito dell'intervista e tentativi di contatto telefonico*

Esito del contatto telefonico	Tentativi di contatto					Totale	Media	Sqm	Mediana
	1-2	3-5	6-10	11-20	21 o più				
Interviste completate (n=2493)	66,8	25,3	6,3	1,5	0,1	100,0	2,5	2,2	2,0
Famiglie irreperibili (n=536)	35,1	20,2	21,1	17,9	5,7	100,0	7,1	6,2	5,0
Rifiuto di rispondere (n=1100)	71,1	20,7	6,4	1,4	0,4	100,0	2,5	2,5	2,0
Rifiuto di completare l'intervista (n=74)	63,0	29,0	7,5	0,5	0,0	100,0	2,7	2,1	2,0
Totale (n=4203)	64,0	23,5	8,2	3,5	0,8	100,0	3,1	3,4	2,0

Le interviste sono state condotte per la maggior parte tra le 17 e le 20, fascia oraria in cui è più facile trovare le persone in casa al termine della giornata di lavoro (Tab.1.3).

Tab.1.3. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per orario in cui è avvenuta l'intervista e relativo esito*

Ora	Esito dell'intervista						
	Interviste complete	Rifiuto di rispondere	Altro esito	Totale	Rischio di rifiuto	Interviste complete	Rifiuti
9-11 (n=25)	44,0	36,0	20,0	100,0	45,0	0,4	0,7
11-12 (n=122)	60,2	29,7	10,1	100,0	33,0	3,2	3,5
12-13 (n=163)	56,6	32,1	11,3	100,0	36,2	3,8	4,8
13-14 (n=161)	61,0	22,3	16,7	100,0	26,8	4,0	3,2
14-15 (n=160)	70,7	20,5	8,8	100,0	22,5	4,5	2,9
15-16 (n=395)	60,7	24,7	14,6	100,0	28,9	9,7	8,9
16-17 (n=563)	55,2	30,5	14,3	100,0	35,6	12,8	15,8
17-18 (n=766)	56,3	28,4	15,3	100,0	33,5	17,2	19,4
18-19 (n=788)	57,8	29,0	13,2	100,0	33,4	18,4	20,6
19-20 (n=682)	62,2	23,6	14,2	100,0	27,5	16,6	14,1
20-21 (n=350)	64,8	19,7	15,5	100,0	23,3	8,6	5,8
21-22 (n=28)	75,8	13,7	10,5	100,0	15,3	0,8	0,3
Totale (n=4203)	59,4	26,6	14,0	100,0	30,9	100,0	100,0

Tenendo conto dell'orario in cui sono state effettuate le interviste complete in relazione al numero di componenti per famiglia, si nota che la maggior parte delle

interviste è stata effettuata tra le 17 e le 19, indipendentemente dalla composizione familiare (Tab.1.4).

Tab.1.4. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per orario in cui è avvenuta l'intervista completa e numero di componenti*

Ora	Numero di componenti					Totale (n=2452)
	1 (n=348)	2 (n=690)	3 (n=674)	4 (n=593)	5 o più (n=147)	
9-11	1,1	0,3	0,2	0,7	0,0	0,5
11-12	3,4	5,0	2,0	3,0	0,3	3,2
12-13	3,1	4,3	4,0	3,9	1,0	3,8
13-14	5,1	3,6	4,3	3,5	3,3	4,0
14-15	4,0	5,9	4,4	3,7	3,3	4,5
15-16	13,1	10,1	9,8	8,1	7,3	9,7
16-17	14,0	12,3	12,9	12,0	14,5	12,7
17-18	16,5	17,3	17,8	15,7	21,7	17,2
18-19	17,3	17,8	17,1	21,4	16,9	18,4
19-20	14,4	15,7	16,7	17,8	20,2	16,6
20-21	7,1	7,2	9,7	9,3	10,7	8,6
21-22	0,9	0,5	1,1	0,9	0,8	0,8
Totale	100,0	100,0	100,0	100,0	100,0	100,0

Per quanto riguarda le caratteristiche anagrafiche dei rispondenti, nel 67,8% dei casi si tratta di donne. Le percentuali più elevate si collocano nella classe di età compresa tra i 40 e i 44 anni.

I giovani di età compresa tra i 25 e i 39 anni sono stati contattati perlopiù in serata, tra le 19 e le 20 (19,2% e 21,3% di quelli in età 25-29 e 30-39 rispettivamente). Questo è dovuto al fatto che si tratta di persone in gran parte occupate per motivi di studio e, soprattutto, di lavoro. Gli individui in età tra i 40 e i 49 anni hanno effettuato la maggior parte delle interviste tra le 18 e le 19 (22,4%); per quanto riguarda le altre fasce d'età, la maggior parte delle interviste è stata effettuata nella fascia oraria tra le 17 e le 19 (17,2% di interviste tra le 17 e le 18 e 18,4% tra le 18 e le 19) (Tab.1.5).

Tab.1.5. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per orario in cui è avvenuta l'intervista ed età del rispondente*

Ora	Età del rispondente							Totale (n=2493)
	18-24 (n=127)	25-29 (n=74)	30-39 (n=390)	40-49 (n=567)	50-59 (n=488)	60-69 (n=426)	70-79 (n=421)	
9-11	1,6	0,5	0,3	0,0	1,0	0,4	0,4	0,4
11-12	2,9	3,4	3,3	2,5	1,4	5,2	4,2	3,2
12-13	1,5	6,7	4,2	3,4	3,9	2,7	4,6	3,8
13-14	3,4	7,7	5,3	3,9	3,5	3,7	2,9	4,0
14-15	1,5	5,5	4,2	4,7	5,5	4,7	3,7	4,5
15-16	10,9	9,1	6,7	8,8	9,2	13,0	11,1	9,7
16-17	15,3	10,2	10,9	11,8	10,7	13,5	17,6	12,8
17-18	21,0	14,8	16,9	12,7	20,6	17,1	19,5	17,2
18-19	21,6	15,3	17,3	22,4	16,6	17,6	16,6	18,4
19-20	13,0	19,2	21,3	18,5	16,0	15,4	11,5	16,6
20-21	7,3	6,0	9,3	9,9	10,4	6,6	7,0	8,6
21-22	0,0	1,6	0,3	1,4	1,2	0,1	0,9	0,8
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

La durata delle interviste è variabile: nel 34,5% dei casi è compresa tra i 15 e i 19 minuti, nel 28,9% è inferiore ai 15 minuti e nel 22,7% ha lunghezza variabile tra i 20 ed i 24 minuti. Le interviste che durano da 25 minuti in su sono gradatamente meno frequenti.

Il motivo di queste differenze è dovuto soprattutto al fatto che il questionario è strutturato in sezioni, alcune delle quali individuali e dedicate a soggetti con caratteristiche particolari (invalidi, malati cronico-degenerativi, anziani, donne in gravidanza). Pertanto, tendenzialmente, all'aumentare del numero di componenti e ancor più della complessità delle situazioni familiari aumenta la durata dell'intervista.

Per le famiglie unipersonali la durata media dell'intervista è di 11,8 minuti (il 55,2% di tali famiglie ha svolto l'intervista in un lasso di tempo inferiore ai 14 minuti); essa aumenta progressivamente all'aumentare dei componenti fino a raggiungere i 21,6 minuti nelle famiglie costituite da cinque o più persone (il 32,7% di tali famiglie ha avuto un'intervista di durata tra i 20 e i 24 minuti). Indipendentemente dal numero di componenti, poche sono le interviste che impegnano le famiglie oltre i 35 minuti. La mediana non si discosta in modo apprezzabile dalla

media, segno del fatto che le durate delle interviste sono abbastanza omogenee (Tab.1.6).

Tab.1.6. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per durata dell'intervista (in minuti) e numero di componenti*

Durata dell'intervista	Numero di componenti					Totale (n=2118)
	1 (n=236)	2 (n=562)	3 (n=616)	4 (n=568)	5 o più (n=136)	
0-14	55,2	41,1	23,8	19,1	3,8	28,9
15-19	30,1	37,5	38,3	34,1	25,5	35,2
20-24	10,8	15,6	23,5	30,0	32,7	22,6
25-29	2,4	3,7	9,8	10,9	23,4	8,6
30-34	1,1	1,3	3,2	3,3	10,1	3,0
35-39	0,4	0,7	0,9	1,7	2,2	1,1
40-44	0,0	0,1	0,5	0,7	2,0	0,5
45 o più	0,0	0,0	0,0	0,2	0,3	0,1
Totale	100,0	100,0	100,0	100,0	100,0	100,0
Media	11,8	14,2	16,9	18,6	21,6	15,9
Sqm	11,0	14,0	17,0	18,0	21,5	16,0
Mediana	5,6	6,3	7,0	7,6	8,3	7,6

1.5 Il questionario

Per la raccolta dei dati è stato realizzato un questionario di tipo elettronico dapprima testato mediante un'indagine pilota, per valutarne l'efficienza e poi poterlo utilizzare nella rilevazione di interesse.

Il questionario, frutto di un meticoloso lavoro istruttorio, ha coinvolto esperti in diverse aree tematiche, nelle quali si articola il questionario stesso.

Esso è stato redatto in modo facilmente comprensibile, in quanto rivolto ad un ampio spettro di famiglie, spesso diverse dal punto di vista culturale e della composizione per età dei membri. Doveva inoltre adattarsi a tutte le situazioni familiari che potevano presentarsi nel corso dell'indagine.

Inizialmente si è tracciato uno schema concettuale del questionario e, solo in seconda battuta, si è passati alla stesura delle domande. È stato importante prestare attenzione a tutti i possibili percorsi da creare e a tutti i possibili casi che si sarebbero potuti presentare.

La revisione conclusiva del questionario è avvenuta a seguito dell'indagine pilota che ha coinvolto circa 30 famiglie al fine di determinare la compatibilità dei quesiti con le reali possibilità di rispondere da parte degli interessati e di gestire il questionario nella situazione sopra descritta. La versione definitiva è stata redatta dopo un incontro con gli intervistatori volto a far emergere miglioramenti da apportare al questionario per renderlo più fluido e funzionale all'indagine.

Esso si articola in 32 sezioni allo scopo di raccogliere informazioni sulle variabili più direttamente legate ai fini dell'indagine, ma anche su variabili che permettano di definire il contesto sociale ed economico della famiglia; la prima e l'ultima servono rispettivamente per l'avviamento e la chiusura del questionario, mentre le altre vertono su varie aree tematiche. In particolare si individuano:

- *informazioni generali per ogni componente*: caratteristiche anagrafiche, relazione col capofamiglia, titolo di studio, condizione professionale;
- *informazioni sulla famiglia*: anno in cui si è formata, tipo di famiglia (single, coppia con figli, coppia senza figli...), zona in cui abita;
- *abitazione e beni durevoli*: tipo di fabbricato in cui vive la famiglia e possesso di alcuni beni che indicano il benessere ed il grado di autonomia della famiglia;
- *abitudini alimentari*: si rileva se la famiglia è solita riunirsi a tavola oppure se ciascun componente consuma i pasti separatamente dagli altri. In questo secondo caso se ne indaga il motivo;
- *vacanze*: abitudini familiari per quanto concerne le vacanze (si rileva se tutti i componenti vanno in vacanza assieme, se la famiglia concentra le proprie vacanze in un solo periodo oppure se le diluisce nell'arco dell'anno);
- *letture e televisione*: tipo di libri letti da ciascun componente e tipo di programmi televisivi che la famiglia usualmente segue;
- *invalidità fisica, mentale e sensoriale*: presenza in famiglia di componenti

affetti da invalidità. Nell'eventualità se ne rileva il tipo, il grado di autonomia dell'invalide e della famiglia nel gestire le problematiche legate allo stesso;

- *malattie cronico-degenerative*: presenza in famiglia di componenti affetti da malattie cronico-degenerative ed eventualmente l'apparato colpito ed il grado di autonomia del componente e della famiglia nel gestire le problematiche legate allo stesso;
- *figli da 0 a 2 anni*: si rileva se il componente frequenta l'asilo nido, chi lo accompagna e con quale mezzo, a che ora esce e rientra in casa, chi lo accudisce quando non è all'asilo;
- *figli da 3 a 5 anni*: si rileva se il componente frequenta la scuola materna, chi lo accompagna e con quale mezzo, a che ora esce e rientra in casa, chi lo accudisce quando è a casa, quali sono gli amici con cui gioca prevalentemente;
- *figli da 6 a 14 anni*: si rileva se il componente frequenta la scuola, con che mezzo vi si reca, a che ora esce e rientra a casa, a chi è affidato quando non è a scuola, chi lo segue nei compiti, se presenta problemi di relazione, condotta o apprendimento, con chi gioca prevalentemente, quali attività extrascolastiche segue, se gli viene data una "paghetta" da gestire;
- *figli da 15 a 19 anni*: si rileva se frequenta qualche tipo di scuola superiore, con che mezzo vi si reca, a che ora esce e rientra a casa, se presenta problemi di relazione, condotta o apprendimento, se viene aiutato da qualcuno nello svolgimento dei compiti, se i genitori riescono ad instaurare un dialogo sereno, se esce con gli amici, quali altre attività o sport pratica, se gli viene data una "paghetta" da gestire;
- *figli con 20 anni o più*: si rileva se frequenta scuole superiori o corsi di laurea, il motivo per il quale vive ancora in famiglia, se esce con gli amici, le attività ricreative svolte e gli sport praticati, se gli viene dato un fisso per le sue esigenze economiche;

- *anziani*: si rileva se frequenta circoli, luoghi di ritrovo o l'università per anziani e si cercano informazioni relativamente alle principali attività svolte in una giornata tipo;
- *donne in gravidanza*: si rileva la possibile presenza in famiglia di donne in stato di gravidanza ed eventuali problemi di salute o familiari connessi;
- *lavoro*: si cercano informazioni relativamente alle ore trascorse fuori casa per motivi di lavoro da ciascun componente, se il lavoro prevede turni, il mezzo utilizzato per recarsi al lavoro;
- *perdita di lavoro/fallimento*: sezione riservata a chi, negli ultimi tre anni, ha vissuto problemi legati al lavoro, le ripercussioni degli stessi sul bilancio familiare e sullo stato psicologico proprio e dei famigliari e le misure attuate dalla famiglia per far fronte a tali problemi;
- *disoccupazione/inoccupazione*: si rileva l'eventuale presenza di disoccupati o inoccupati all'interno della famiglia e le possibili ripercussioni psicologiche ed economiche;
- *servizi usufruiti*: tipo di servizi del Comune, dell'ULSS o di qualche associazione di volontariato a cui ha fatto ricorso la famiglia nell'ultimo anno;
- *reti familiari*: rete di parenti sui quali la famiglia può contare in caso di necessità;
- *lutto*: si rileva se nell'ultimo triennio la famiglia ha subito un lutto, di chi si tratta e la causa del decesso;
- *separazioni/divorzi*: si indaga sulla possibile presenza di componenti separati o divorziati e sull'eventuale affidamento dei figli;
- *incidenti*: si rileva se nell'ultimo triennio qualche componente ha subito incidenti e se si è reso necessario un periodo di degenza in ospedale o a casa;

- *furti/intrusioni*: si rileva se negli ultimi tre anni si sono verificati furti o intrusioni nell'abitazione in cui vive la famiglia, se era presente qualcuno al momento dell'intrusione, se sono stati sottratti oggetti di valore, se le autorità competenti sono state informate dell'accaduto;
- *quartiere di residenza*: rispetto alla percezione del rispondente, si chiede qual è il rischio di criminalità della zona in cui abita;
- *percezione del disagio*: si chiede al rispondente di esprimersi relativamente al disagio percepito all'interno della propria famiglia.
- *relazioni*: per la coppia che compone il nucleo familiare si rileva se riesce a ritagliarsi momenti di tempo libero e si chiedono le principali attività svolte;
- *valori familiari*: vengono poste al rispondente alcune domande sulla religione ed altre di opinione a carattere più generale;
- *reddito familiare e situazione economica*: si indaga sulla situazione economica della famiglia anche in un'ottica retrospettiva di confronto, sulle principali fonti di reddito e sul reddito familiare complessivo.

I dati della prima sezione riguardano ogni singolo componente. Analogamente, quelli inerenti agli eventi critici, alle malattie e al possesso di altre caratteristiche che possono provocare disagio sono stati rilevati relativamente a ciascun individuo. La percezione del disagio, le caratteristiche strutturali, sociali, economiche, relazionali e valoriali della famiglia sono invece state rilevate per unità familiare.

Capitolo 2

L'analisi dei dati

2.1 Caratteristiche anagrafiche delle famiglie venete

Le famiglie venete sono generalmente di piccole dimensioni, composte in media da 2,8 componenti, in accordo con il dato del Bilancio Demografico Regionale pari a 2,5 membri per famiglia (ISTAT, 2005). Il 42,1% è costituito da una (13,9%) o due (28,2%) persone. Sono, d'altra parte, poche le famiglie con cinque o più componenti: nel campione selezionato risultano essere il 6,1% del totale.

Il numero di famiglie costituite da una sola persona è una sottostima di quello censuario che ammonta a 400. Ciò è dovuto principalmente a due motivi: innanzitutto al fatto che gli anziani soli e privi di telefono non potevano essere raggiunti dall'indagine ed in secondo luogo alla lista di campionamento (elenco del telefono) che non contiene numeri telefonici della rete mobile, lasciando, pertanto, scoperta quella parte di popolazione (soprattutto giovani) che vive sola e decide di non abbonarsi alla rete fissa.

La percentuale più alta di famiglie unipersonali si trova nella provincia di Belluno (17,5%). È probabile che si tratti prevalentemente di anziani soli visto che nel bellunese si riscontra anche un indice di vecchiaia¹ particolarmente alto (171,15 contro quello italiano di 131,38 (ISTAT, 2001)). Le famiglie composte da due

¹L'indice di vecchiaia è il rapporto tra la popolazione di 65 anni e più e la popolazione di età 0-14 anni moltiplicato per 100.

componenti si concentrano maggiormente a Verona (30,8%), mentre quelle con cinque o più a Padova e a Vicenza.

Il numero medio di componenti non differisce in misura evidente da provincia a provincia, ma meno variabile risulta il numero di componenti a Belluno rispetto alle altre province (sqm = 13,1) (Tab.2.1).

Tab.2.1. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per provincia e numero di componenti*

Provincia	Numero di componenti					Totale	Media	Sqm	Mediana
	1	2	3	4	5 o più				
Belluno (n=343)	17,5	30,6	25,7	21,0	5,2	100,0	2,7	13,1	3,0
Padova (n=391)	12,8	23,0	28,1	28,7	7,4	100,0	2,9	24,1	3,0
Rovigo (n=379)	15,6	26,7	31,1	20,8	5,8	100,0	2,8	23,0	3,0
Treviso (n=377)	13,5	27,1	26,5	27,1	5,8	100,0	2,7	22,1	3,0
Venezia (n=358)	13,7	32,4	28,8	20,6	4,5	100,0	2,7	24,2	3,0
Verona (n=360)	16,1	30,8	21,4	25,8	5,8	100,0	2,7	24,2	3,0
Vicenza (n=351)	12,0	27,6	27,1	25,9	7,4	100,0	2,9	23,0	3,0
Totale (n=2559)	13,9	28,2	26,6	25,2	6,1	100,0	2,8	21,7	3,0

La tabella 2.2 riporta la distribuzione dei componenti delle famiglie intervistate in relazione all'età e al sesso. Il 51,5% del campione è rappresentato da donne, perlopiù in età compresa tra i 40 e i 44 anni (4,9%). Anche per quanto concerne gli uomini, che rappresentano il restante 48,5% del totale della popolazione campionata, la percentuale più elevata si concentra nella classe di età tra i 40 e i 44 anni. Per il 19,0% la popolazione è costituita da giovani sotto i 20 anni e per il 17,9% da anziani sopra i 65 anni.

Tab.2.2. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Totale (n=7106)
	Femmine (n=3662)	Maschi (n=3444)	
0-4	1,7	1,9	3,6
5-9	2,7	2,2	4,9
10-14	2,5	2,5	5,0
15-19	2,7	2,8	5,5
20-24	2,6	2,7	5,3
25-29	2,5	3,1	5,6
30-34	3,2	2,9	6,1
35-39	4,0	3,6	7,6
40-44	4,9	4,5	9,4
45-49	4,1	3,8	7,9
50-54	3,8	3,9	7,7
55-59	3,9	3,8	7,7
60-64	3,1	2,7	5,8
65-69	3,2	3,1	6,3
70-74	2,6	2,3	4,9
75-79	2,0	1,4	3,4
80-84	1,2	0,8	2,0
85-89	0,5	0,3	0,8
90 e più	0,3	0,2	0,5
Totale	51,5	48,5	100,0

Per quanto riguarda lo stato civile della popolazione intervistata, emerge che il 55,7% è coniugato, il 36,7% è celibe/nubile, il 5,8% vedovo e una piccola minoranza separato/divorziato (1,8%).

Le separazioni e di divorzi sono in continuo aumento negli ultimi anni, a causa dei cambiamenti di costume e delle concezioni di coppia e famiglia che hanno caratterizzato il Veneto, l'Italia nel suo complesso e in genere tutto il mondo occidentale (Maggioni, 1997).

Rispetto alle donne, gli uomini risultano in percentuale maggiore coniugati/conviventi e in percentuale decisamente inferiore vedovi (1,5% contro il 9,8% delle donne) e separati/divorziati (1,2% contro il 2,4% delle donne) (Tab.2.3). La differenza tra i due sessi è spiegata, per quanto riguarda la vedovanza soprattutto dalla maggior longevità delle donne rispetto agli uomini, per quanto riguarda la

separazione dalla maggior frequenza con cui gli uomini tendono a risposarsi o ad iniziare una convivenza con un nuovo partner.

Tab.2.3. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per stato civile e sesso*

Stato civile	Sesso		
	Femmine (n=3634)	Maschi (n=3430)	Totale (n=7064)
Coniugato / convivente	53,8	57,7	55,7
Celibe / nubile	34,0	39,6	36,7
Separato/a, divorziato/a	2,4	1,2	1,8
Vedovo/a	9,8	1,5	5,8
Totale	100,0	100,0	100,0

2.2 Istruzione e condizione professionale

Per completare la descrizione anagrafica dei singoli componenti della popolazione intervistata, è necessario considerare il grado di istruzione che, per ovvie ragioni, riguarderà esclusivamente le persone con età superiore ai 6 anni. Il titolo di studio più diffuso è la licenza di scuola media inferiore (29,8% del totale degli individui con più di 6 anni), in cui rientrano soprattutto le persone prossime all'anzianità; la differenza non è molto rilevante per i due sessi, sebbene la percentuale sia leggermente superiore per gli uomini. Segue un 25,1% di individui in possesso di diploma di scuola media superiore (4-5 anni); come nel caso precedente non c'è grande differenza tra uomini e donne. La percentuale di persone in possesso della sola licenza elementare (21,2%) è leggermente più alta per le donne (23,9% contro 18,3% per gli uomini). Tra gli uomini è inoltre più elevata la percentuale di diplomi di scuola media superiore (2-3 anni) rispetto alle donne (8,4% contro 6,5%) (Tab.2.4).

Per quanto concerne la condizione professionale, emerge che il 49,5% della popolazione con età pari o superiore ai 15 anni risulta occupato, il 26,6% è costituito da pensionati, il 12,1% da casalinghe e il 9,1% da studenti. Una bassa percentuale di individui, è invece disoccupata (1,6%), inabile al lavoro (0,3%) e in servizio di leva o servizio civile (0,1%) (Tab.2.6).

Tab.2.6. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per condizione professionale*

Condizione professionale (n=6129)	% (a)	% (b)
Occupato	48,8	49,5
Disoccupato alla ricerca di nuova occupazione	1,2	1,3
Disoccupato in cerca di prima occupazione	0,3	0,3
In servizio di leva o servizio civile sostitutivo	0,1	0,1
Casalinga	12,0	12,1
Studente	10,3	9,1
Inabile al lavoro	0,3	0,3
Pensionato/a	26,3	26,6
In altra condizione	0,7	0,7
Totale	100,0	100,0

% (a) = percentuale sul totale della popolazione

% (b) = percentuale sul totale della popolazione con più di 15 anni

Tra gli occupati il 71,3% svolge un lavoro dipendente, mentre il restante 28,7% esercita un lavoro autonomo. La percentuale più elevata di lavoratori dipendenti è costituita da impiegati (44,5%) ed operai (28,3%) (Tab.2.7).

Tab.2.7. *Distribuzione percentuale della popolazione occupata residente nel Veneto nel 2004-2005 per posizione nel lavoro dipendente*

Posizione lavorativa (n=2123)	
Dirigente	3,4
Direttivo	5,7
Impiegato	44,5
Intermedio	0,9
Operaio	28,3
Uscere, guardiano, bidello	15,9
Apprendista	0,6
Lavorante a domicilio per conto di imprese	0,2
Militare di carriera	0,5
Totale	100,0

I lavoratori autonomi sono principalmente imprenditori (60,0%) e liberi professionisti (28,1%) (Tab.2.8).

Tab.2.8. *Distribuzione percentuale della popolazione occupata residente nel Veneto nel 2004-2005 per posizione nel lavoro autonomo*

Posizione lavorativa (n=747)	
Imprenditore	60,0
Libero professionista	28,1
Socio di cooperativa	1,2
Coadiuvante in azienda familiare	2,0
Tirocinante	0,4
Lavoratore con contratto atipico	3,6
Altro	4,7
Totale	100,0

La tabella 2.9 mostra il numero di persone occupate in relazione al numero di componenti in famiglia. Nelle famiglie unipersonali lavora solo il 22,6% degli individui; ciò accade perché si tratta principalmente di anziani pensionati che vivono soli. Anche le famiglie di due componenti sono composte per oltre la metà dei casi da persone non occupate, perlopiù da coppie di anziani pensionati. Nelle famiglie con tre o più componenti, la situazione più diffusa è data invece dalla presenza di due componenti occupati (44,1% per le famiglie con tre componenti, 55,8% per quelle con quattro e 39,1% per quelle con cinque o più).

Tab.2.9. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per numero di componenti e numero di membri occupati*

Numero di componenti	Numero di componenti occupati				Totale	Media	Sqm	Mediana
	0	1	2	3 o più				
1 (n=361)	77,4	22,6	-	-	100,0	0,2	0,4	0,0
2 (n=715)	53,9	24,5	21,6	-	100,0	0,7	0,8	0,0
3 (n=683)	9,2	41,2	44,1	5,5	100,0	1,5	0,7	1,5
4 (n=611)	1,9	29,9	55,8	12,4	100,0	1,8	0,7	2,0
5 o più (n=154)	2,9	31,9	39,1	26,0	100,0	1,9	0,8	2,0
Totale (n=2524)	29,0	30,5	34,3	6,2	100,0	1,1	0,9	1,0

Solo l'1,6% della popolazione con età superiore ai 15 anni risulta disoccupato; il fenomeno, pertanto, non assume proporzioni tali da destare preoccupazione. Il 26,1% dei disoccupati si concentra tra i 25 e i 29 anni, il 18,9% ha invece un'età

compresa tra i 30 e i 34 anni (Tab.2.10). Tra i due sessi si nota una marcata differenza in alcune classi di età: in percentuale decisamente maggiore rispetto agli uomini, si trovano le donne disoccupate nelle classi di età 15-19 e 40-44; inferiore è, invece, la percentuale di donne disoccupate tra i 25 e i 29 anni e oltre i 50.

Tab.2.10. *Distribuzione percentuale della popolazione disoccupata / inoccupata residente nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Totale (n=94)
	Femmine (n=52)	Maschi (n=42)	
15-19	12,1	0,8	7,0
20-24	11,7	11,2	11,5
25-29	22,3	30,8	26,1
30-34	18,4	19,5	18,9
35-39	8,4	6,6	7,6
40-44	14,5	0,8	8,3
45-49	7,1	3,7	5,6
50-54	3,1	12,0	7,1
55-59	2,4	14,6	7,9
Totale	100,0	100,0	100,0

2.3 Composizione delle famiglie

La maggior parte delle famiglie intervistate si sono formate più di 25 anni fa (46,2% del totale), nel 20,1% dei casi si sono formate tra i 15 e i 25 anni fa; le percentuali vanno via via diminuendo fino ad arrivare ad un 1,2% di famiglie giovanissime, formatesi meno di un anno fa (Tab.2.11).

Tab.2.11. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per età della famiglia*

Età della famiglia (n=2495)	
Meno di un anno	1,2
Da 1 a 3 anni	4,3
Da 3 a 7 anni	10,4
Da 7 a 15 anni	17,8
Da 15 a 25 anni	20,1
Oltre 25 anni	46,2
Totale	100,0

Per quanto riguarda la composizione, si sono prese in considerazione diverse tipologie di famiglie: unipersonali, coppie con o senza figli, un solo genitore con figli, famiglie polinucleari, dette anche “estese”² (ISTAT, 2003).

Con i censimenti del 1951 e del 1961 si nota una riduzione del numero di nuclei familiari estesi e/o multipli e, pertanto, la tendenza verso il processo di nuclearizzazione della famiglia. Nel 1991 il 61,7% delle famiglie costituite da un solo nucleo era di tipo coppia con figli (Saraceno, 2003).

Oggi, nonostante le coppie con figli rappresentino la tipologia familiare quantitativamente più rilevante (51,7% del campione), altre percentuali importanti sono rappresentate dalle coppie senza figli (23,3%) e dalle famiglie unipersonali (13,8%). Non trascurabile risulta anche il numero di famiglie composte da un solo genitore con figli, che rappresentano il 6,7% del totale, anche se, restringendo l’osservazione ai soli figli a carico³, si passa ad un 2,7% (Tab.2.12).

Tab.2.12. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per tipo di famiglia*

Tipo di famiglia (n=2503)	
Unipersonale	13,8
Coppia senza figli	23,3
Coppia con figli	51,7
Coppia senza figli con altri componenti	0,8
Coppia con figli ed altri componenti	2,2
Un genitore con figli a carico	4,7
Un genitore con almeno un figlio occupato	4,0
Un genitore con figli ed altri componenti	0,8
Fratelli / sorelle	0,5
Altro	0,2
Totale	100,0

Considerando solo le famiglie con almeno due componenti, il 30,5% non ha figli, il 34,4% ne ha soltanto uno e, infine, le famiglie con più di due figli rappresentano il 5,9%. Il numero medio di figli per le famiglie con almeno due componenti è

²Le famiglie estese corrispondono a quelle tipologie familiari all’interno delle quali si individuano almeno due nuclei (coppie o nuclei monogenitoriali) oppure un solo nucleo con membri aggregati alla famiglia.

³Come famiglie monogenitoriali con figli a carico si sono considerate quelle in cui il figlio non riceve un reddito da lavoro.

di 0,9, valore prossimo a quello mediano, indice del fatto che il numero di figli è piuttosto omogeneo all'interno delle famiglie (Tab.2.13).

Tab.2.13. *Distribuzione percentuale delle famiglie (con almeno due componenti) residenti nel Veneto nel 2004-2005 per numero di figli*

Numero di figli (n=2190)	
0	30,5
1	34,4
2	29,2
3	5,0
4	0,7
5	0,2
Totale	100,0
Media	0,9
Sqm	0,9
Mediana	1,0

È evidente che la prolungata permanenza nella famiglia di origine e la posticipazione del comportamento riproduttivo, fanno sì che le famiglie con prole numerosa siano sempre più rare. Il calo della fecondità, oltre a produrre una diminuzione della coppie con figli, favorisce un costante aumento delle coppie con un solo figlio e la diminuzione della coppie con tre o più figli.

Questo aspetto è ancora più evidente se si esamina la presenza di uno o più componenti in determinate fasce d'età (Tab.2.14). Le famiglie con figli di età 0-2 anni rappresentano il 5,6% del totale delle famiglie, quelle con figli tra i 3 e i 5 anni l'8,5%, tra i 6 e i 14 il 20,2%, quelle con adolescenti tra i 15 e i 19 anni sono il 13,7%. Gli anziani, tra i 65 e i 79 anni sono invece presenti nel 29,1% delle famiglie intervistate, quelli con più di 80 anni nell'8,0%. Questo dato dà anche la misura di come sia vecchia la struttura per età della popolazione nella nostra regione.

Tab.2.14. *Distribuzione percentuale delle famiglie residenti nel Veneto nel per presenza di componenti da 0 a 2 anni, da 3 a 5, da 6 a 14, da 15 a 19, con più di 65 anni e numero totale di membri in famiglia*

Classi	Numero di componenti			Totale
	0	1	2 o più	
0-2 (n=2556)	94,4	5,3	0,3	100,0
3-5 (n=2555)	91,5	8,2	0,3	100,0
6-14 (n=2555)	79,8	15,1	5,1	100,0
15-19 (n=2555)	86,3	12,6	1,1	100,0
65-79 (n=2551)	70,9	18,5	10,6	100,0
80 e più (n=2526)	92,0	6,8	1,2	100,0

Il tasso di fecondità del Veneto è inferiore a quello italiano, pari a 1,6 figli per le donne che nascono nel 1963. Nella coorte femminile del 1920 il numero medio di figli ammonta a 2,5. Pertanto, nell'arco di 44 coorti la fecondità italiana ha subito un drastico calo, non interrompendo mai il suo declino, a differenza di quanto è avvenuto in altri paesi dell'Europa occidentale.

All'interno del nostro paese è tuttavia sempre stata presente una forte differenzialità regionale: nel settentrione nascono meno bambini in contrasto con le popolazioni meridionali. Nel corso del tempo i livelli di fecondità caratteristici delle aree settentrionali e di quelle meridionali si sono avvicinati, anche se la situazione non ha subito sostanziali modifiche.

Con il declino del numero medio di figli, si sono riscontrati in Italia mutamenti nel comportamento riproduttivo. Nelle famiglie del centro-nord, e pertanto anche in quelle venete, la tendenza va verso la scelta del figlio unico (34,4%); sono poco diffuse le nascite di ordine superiore al secondo, mentre elevata è la percentuale di donne senza figli (30,5%). Insieme, madri senza figli o con un figlio solo rappresentano il 64,9% delle donne in famiglie con almeno due componenti. Di contro, nelle famiglie del sud prevale la famiglia con due figli, mentre la fecondità si riduce solo negli ordini superiori al secondo. La particolarità di questa area è la maggiore presenza di donne senza figli rispetto a quella di donne con un figlio solo: è più probabile rinunciare alla maternità, che avere un solo figlio (Santini, 1997).

Infine, si sono considerate le famiglie in cui è presente almeno un componente anziano: è importante capire se, in famiglia, siano presenti anche componenti di età inferiore ai 65 anni.

Il 24,6% delle famiglie in cui sono presenti anziani è unipersonali, ossia si tratta di anziani che vivono soli (8,5% delle famiglie totali), il 28,6% è composto da una coppia di anziani (9,9% delle famiglie totali) e il restante 46,8% è rappresentato da famiglie in cui è presente almeno un anziano oltre a componenti di età inferiore a 65 anni (Tab.2.15).

Tab.2.15. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per composizione delle famiglie con membri anziani (sopra i 65 anni)*

Composizione della famiglia (n=910)	% (a)	% (b)
Famiglie unipersonali di anziani	8,5	24,6
Famiglie di una coppia di anziani	9,9	28,6
Famiglie con un anziano	12,2	35,2
Famiglie con più anziani	4,0	11,6
Totale	34,6	100,0

% (a) = percentuale sul totale delle famiglie

% (b) = percentuale sul totale delle famiglie con membri anziani

2.4 Abitazione

La casa è un bene indispensabile per la convivenza familiare. La mancanza di un ambiente idoneo alle esigenze della famiglia può essere fonte di tensioni e malessere per e tra i componenti del nucleo stesso.

La maggior parte delle famiglie venete (86,4%) vive in un'abitazione di proprietà, mentre solo l'11,6% occupa l'abitazione in affitto o subaffitto e il restante 2,0% secondo un altro titolo (gratuito o prestazione di servizio). Il possesso dell'abitazione di residenza è più diffuso nelle zone isolate e periferiche, soprattutto perché gli ampi spazi lo permettono, mentre nei centri urbani la percentuale è leggermente più bassa (Tab.2.16).

Tab.2.16. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per zona di residenza e titolo a cui la famiglia occupa l'abitazione*

Zona di residenza	Titolo a cui la famiglia occupa l'abitazione			Totale
	Proprietà	Affitto	Altro	
Centro (n=1029)	83,4	13,8	2,8	100,0
Periferia (n=1287)	88,0	10,6	1,4	100,0
Zona isolata (n=121)	95,0	2,9	2,1	100,0
Totale (n=2437)	86,4	11,6	2,0	100,0

La condizione abitativa della famiglia veneta risulta profondamente cambiata rispetto agli anni del dopoguerra e ai successivi: oggi la proprietà è così diffusa che non permette di fare distinzione tra differenti ceti e posizioni sociali (Cremaschi, 1996).

A ciò va aggiunto il miglioramento delle condizioni abitative, sebbene, accanto ad una maggioranza di famiglie che si ritiene soddisfatta della propria abitazione, ne esistano di quelle che la ritengono inadeguata. Il costo delle abitazioni è in crescita e, per le famiglie in difficoltà, ciò si traduce in una situazione di disagio, che può, col tempo, coinvolgere un numero maggiore di famiglie, visto il crescente divario tra costi e redditi.

Il 94,9% ritiene la propria abitazione adeguata o proporzionata alle esigenze quotidiane della famiglia e di tutti i componenti; i giudizi sono diversi secondo le diverse tipologie abitative (Tab.2.17). È evidente che le case singole o plurifamiliari possono essere le più confortevoli, probabilmente per il fatto che spesso sono dotate di un piccolo giardino o di zone verdi e sono più adeguate anche alle esigenze dei bambini.

Tab.2.17. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per tipo di abitazione e relativa adeguatezza*

Tipo di abitazione	Adeguatezza dell'abitazione			Totale
	Più che sufficiente	Proporzionata	Insufficiente	
Casa singola (n=918)	62,1	36,6	1,3	100,0
Plurifamiliare (n=605)	54,1	42,4	3,5	100,0
Appartamento (n=953)	38,4	52,3	9,3	100,0
Altro (n=4)	16,2	83,8	0	100,0
Totale (n=2480)	50,5	44,4	5,1	100,0

All'aumentare del numero di redditi in famiglia aumentano le percentuali di famiglie che abitano in case singole: il 51,3% delle famiglie in cui i redditi sono

almeno tre occupa una casa singola, mentre ben il 45,2% di quelle che percepiscono un solo reddito abita in un appartamento (Tab.2.18).

Tab.2.18. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per tipo di abitazione e numero di componenti che percepiscono reddito (stipendio o pensione)*

Tipo	Numero redditi			Totale (n=2488)
	1 (n=890)	2 (n=1192)	3 o più (n=406)	
Casa singola	31,5	32,4	51,3	35,2
Casa plurifamiliare	23,3	25,0	24,9	24,4
Appartamento	45,2	42,6	23,8	40,4
Totale	100,0,0	100,0,0	100,0,0	100,0,0

Per ovvie ragioni di spazio, le famiglie di piccole dimensioni ritengono più che sufficiente, in quanto a dimensione, l'abitazione che occupano. All'aumentare del numero di componenti, la percentuale di famiglie soddisfatte della loro abitazione diminuisce: il 56,1% delle famiglie unipersonali ritiene adeguata la propria abitazione contro il 39,8% di quelle con 5 o più componenti (Tab.2.19).

Tab.2.19. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per adeguatezza dell'abitazione e numero di componenti*

Adeguatezza	Numero componenti					Totale (n=2487)
	1 (n=357)	2 (n=703)	3 (n=676)	4 (n=603)	5 o più (n=148)	
Più che sufficiente	65,2	56,9	48,1	40,6	39,8	50,6
Proporzionata	31,7	41,4	47,6	49,8	49,9	44,3
Insufficiente	3,1	1,7	4,3	9,6	10,3	5,1
Totale	100,0	100,0	100,0	100,0	100,0	100,0

Anche la professione svolta dal capofamiglia influisce, sebbene in misura leggera, sul giudizio che la famiglia ha riguardo l'adeguatezza della propria abitazione. Ben il 56,1% e il 51,6% delle famiglie in cui il capofamiglia è rispettivamente dirigente/direttivo e lavoratore in proprio ritiene più che sufficiente la propria abitazione (Tab.2.20).

2.5 Famiglie unipersonali

In questi ultimi anni, nella nostra regione, come in tutte le società occidentali, sono in continuo aumento e acquistano sempre più interesse dal punto di vista statistico le cosiddette *famiglie unipersonali*, ossia composte da una persona che vive da sola.

Esse costituivano il 20,3% di tutte le famiglie nel 1989/1990 e il 21,3% nel 1997/1998.

Le famiglie unipersonali rappresentano, quindi, una tipologia familiare estremamente instabile, difficile da “fissare” in cifre e percentuali, sia in età giovanile che in età adulta. È una realtà abbastanza variegata che deriva da una serie di motivazioni sociali, culturali ed economiche diverse secondo l'età e il sesso dell'individuo.

Mentre per la persona anziana, la solitudine è, nella maggior parte dei casi, una condizione definitiva, sinonimo della fine del processo esistenziale, per i giovani la solitudine rappresenta una fase transitoria, spesso frutto di una scelta di maggior indipendenza prima dell'approdo a diverse forme di convivenza: un intervallo più o meno lungo, più o meno frequente, secondo le circostanze.

Quella di single è una condizione che può essere interrotta dalla costituzione di una convivenza e soprattutto può assumere connotazioni assai differenti a seconda della consistenza delle reti relazionali che, con il progredire dell'età, tendono generalmente ad affievolirsi.

Le principali caratteristiche della famiglia unipersonale italiana emersero già con i censimenti degli anni 1981 e 1991. In quest'ultimo anno, il 54,3% di tutte le famiglie unipersonali era composto da un individuo di oltre 65 anni, nella maggior parte dei casi (69%) di sesso femminile. Nell'Indagine Multiscopo su “famiglia, soggetti sociali e condizioni dell'infanzia” del 1998, si è rilevato che l'81,2% delle persone sole oltre i 65 anni erano donne, mentre nelle età più giovani vivevano soli con maggior frequenza gli uomini (63%) rispetto alle donne (37%). Pertanto, l'esperienza di vivere soli era tipicamente femminile in età avanzata e maschile in età giovane, a causa, rispettivamente, della maggior probabilità di una vedovanza delle donne e della maggior tendenza degli uomini a vivere soli, piuttosto che con

i figli, dopo una separazione o divorzio (Saraceno, 2003).

Oggi, la condizione di persone sole riguarda in misura differenziata uomini e donne, nelle varie fasce di età. Il 70,5% delle famiglie unipersonali è costituito da una donna, perlopiù di età superiore ai 75 anni (29,7%). La struttura per età dell'uomo single è completamente diversa. Gli uomini, infatti, che rappresentano il restante 29,5% della popolazione campionata, si concentrano soprattutto nella fascia d'età 45-64 (12,4%) (Tab.2.22). Questa differente diffusione nelle varie classi di età della condizione di persona sola tra uomini e donne è dovuta principalmente all'effetto della più elevata speranza di vita femminile e all'età al matrimonio più bassa rispetto a quella maschile. Sul totale delle famiglie unipersonali, il 35,3% è costituito da anziani sopra i 75 anni; solo l'11,6% comprende individui con meno di 45 anni.

Tab.2.22. *Distribuzione percentuale delle famiglie unipersonali residenti nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Totale (n=363)
	Femmine (n=262)	Maschi (n=101)	
18-44	5,5	6,1	11,6
45-64	14,2	12,4	26,6
65-74	21,1	5,4	26,5
75 e più	29,7	5,6	35,3
Totale	70,5	29,5	100,0

Per quanto concerne lo stato civile, oltre la metà delle persone sole sono vedovi/e (56,1%), il 32,5% è celibe/nubile, l'11,4% separato/divorziato. Anche in questo caso le differenze tra i sessi sono notevoli: le donne che vivono sole sono vedove nel 67,2% dei casi, separate o divorziate nel 7,6%, nubili nel 25,2%. Gli uomini soli sono nel 50,7% celibi, nel 20,8% separati o divorziati e nel 28,5% vedovi. La proporzione di separati/divorziati negli uomini soli si mantiene superiore a quella delle donne anche perché, in seguito a separazione o divorzio, le donne si ritrovano a vivere più frequentemente con i figli in nuclei monogenitoriali (Tab.2.23).

Tab.2.23. *Distribuzione percentuale delle famiglie unipersonali residenti nel Veneto nel 2004-2005 per stato civile e sesso*

Stato civile	Sesso		
	Femmine (n=257)	Maschi (n=93)	Totale (n=350)
Celibe / nubile	25,2	50,7	32,5
Separato/a / divorziato/a	7,6	20,8	11,4
Vedovo/a	67,2	28,5	56,1
Totale	100,0	100,0	100,0

Le persone che vivono sole per oltre la metà sono pensionate (69,9%), per il 23,0% occupate. La situazione è, ancora una volta, differente tra i due sessi e riflette le differenze esplicitate sopra: il numero di donne pensionate è maggiore di quello degli uomini (77,2% contro 51,8%), mentre il numero di donne occupate è minore di quello degli uomini (14,8% contro 43,2%) (Tab.2.24).

Tab.2.24. *Distribuzione percentuale delle famiglie unipersonali residenti nel Veneto nel 2004-2005 per condizione professionale e sesso*

Condizione professionale	Sesso		
	Femmine (n=261)	Maschi (n=98)	Totale (n=359)
Occupato	14,8	43,2	23,0
Disoccupato	0,0	2,9	0,8
Casalinga	6,0	0,0	4,3
Studente	0,5	1,3	0,7
Inabile al lavoro	0,5	0,0	0,4
Pensionato/a	77,2	51,8	69,9
Altro	1,0	0,8	0,9
Totale	100,0	100,0	100,0

Il quadro della situazione lavorativa indica con chiarezza che generalmente chi vive solo rientra in una delle seguenti categorie (Donati e Di Nicola, 1996):

- anziani pensionati;
- giovani non sempre economicamente indipendenti, che vivono fuori casa perlopiù per motivi di studio;
- adulti celibi/nubili in una situazione transitoria prima di dar vita ad un'eventuale convivenza;

- adulti separati/divorziati.

Le condizioni socio-economiche delle persone che vivono sole sono, pertanto, alquanto differenziate, come dimostrano anche ricerche effettuate dall'ISTAT relative ai consumi e al titolo di studio dei single. Le persone sole si concentrano ai due poli estremi della agiatezza e della povertà.

Gli anziani soli sono, infatti, generalmente in condizioni economiche di semi-povertà o anche di emarginazione sociale. Le classi più giovani sono in condizioni sociali più elevate, in quanto possiedono risorse culturali, ma generalmente non economiche. Gli adulti, celibi/nubili o separati, possiedono spesso risorse culturali ed economiche che consentono loro di superare le difficoltà abitative o lavorative.

2.6 Famiglie composte da un genitore con figli a carico

Un'altra tipologia di famiglia che, gradatamente, sta prendendo piede nella nostra regione è quella costituita da un genitore con figli a carico.

Sul totale di queste famiglie, nel 70,0% dei casi il genitore è una donna e le fasce d'età più frequenti sono quelle comprese tra i 18 e i 44 anni (33,3%) e tra i 45 e i 64 anni (23,7%). I genitori maschi che vivono con i soli figli, realtà molto meno diffusa (30,0%), hanno perlopiù un'età compresa tra i 45 e i 64 anni (19,1%) (Tab.2.25)⁴.

⁴Le famiglie con un solo genitore e almeno un figlio occupato, d'altro canto, che rappresentano il 50,9% delle famiglie composte da un genitore con figli, sono perlopiù costituite da madri o padri anziani e figlio adulto, a volte anche con più di 40 anni.

Tab.2.25. *Distribuzione percentuale delle famiglie composte da un genitore con figli a carico residenti nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Totale (n=69)
	Femmine (n=48)	Maschi (n=21)	
18-44	33,3	6,6	40,0
45-64	23,7	19,1	42,8
65-74	7,5	1,8	9,3
75 e più	5,5	2,4	7,9
Totale	70,0	30,0	100,0

Proprio la forte prevalenza femminile in questi nuclei monogenitoriali porta gli studiosi di altri paesi europei a verificare se è possibile parlare di “famiglie a rischio”: non è in sé la presenza di un capofamiglia donna a determinare una situazione di disagio, ma molte ricerche sottolineano come queste famiglie siano più colpite da problemi di carattere socio-economico dovuti in larga misura alla sfavorevole posizione della donna nel mercato del lavoro e alla sua esclusiva responsabilità nei confronti dei figli. In Italia il 9,7% delle famiglie monogenitoriali con capofamiglia donna vive in condizioni di povertà, contro l'8,5% di quelle costituite da una coppia con figli. Si tratta di una percentuale contenuta rispetto a quelle registrate in altri paesi occidentali dal momento che in Italia la maggior parte delle madri sole ha un elevato livello di istruzione, è inserita nel mercato del lavoro e appartiene a classi sociali medio-alte; a ciò si aggiunge il fatto che nel nostro paese non esistono misure di sostegno a favore delle madri sole, fatta eccezione per le modeste detrazioni fiscali e le relative maggiorazioni degli assegni familiari e ciò è un incentivo per la donna a farsi strada nel mercato del lavoro (Zanatta, 1997).

Il 44,8% dei genitori sono separati/divorziati, il 30,6% vedovi. Le differenze tra i sessi sono evidenti: le madri sono perlopiù separate/divorziate (35,0%) o vedove (28,2%), mentre i padri celibi (59,4%) (Tab.2.26).

Tab.2.26. *Distribuzione percentuale delle famiglie composte da un genitore con figli a carico residenti nel Veneto nel 2004-2005 per stato civile e sesso*

Stato civile	Sesso		Totale (n=69)
	Femmine (n=48)	Maschi (n=21)	
Celibe / nubile	6,8	59,4	24,6
Separato/a / divorziato/a	35,0	32,6	44,8
Vedovo/a	28,2	18,0	30,6
Totale	100,0	100,0	100,0

Il fatto che neanche la metà (44,8%) di questi genitori, che già fanno parte di una piccola porzione di famiglie (quelle monogenitoriali con figli a carico), è separata/divorziata, è indice del fatto che la fragilità coniugale, anche se sta iniziando a diffondersi, è ancora poco frequente nella nostra regione.

I genitori sono perlopiù occupati (67,7%), e, in percentuale minore, pensionati (24,9%). La situazione non differisce molto fra i due sessi: più numerose sono le occupate rispetto agli occupati (70,3% vs. 61,9%), mentre tra i pensionati sono in numero maggiore gli uomini (28,3%). Inoltre, il 3,5% delle donne è casalinga, contro lo 0,0% degli uomini e il 7,9% degli uomini è inabile al lavoro contro lo 0,0% delle donne (Tab.2.27).

Tab.2.27. *Distribuzione percentuale delle famiglie composte da un genitore con figli a carico residenti nel Veneto nel 2004-2005 per condizione professionale e sesso*

Condizione professionale	Sesso		Totale (n=69)
	Femmine (n=48)	Maschi (n=21)	
Occupato	70,3	61,9	67,7
Disoccupato	0,0	1,9	0,6
Casalinga	3,5	0,0	2,5
Inabile al lavoro	0,0	7,9	2,4
Pensionato/a	23,5	28,3	24,9
Altro	2,7	0,0	1,9
Totale	100,0	100,0	100,0

Le famiglie composte da un solo genitore sono una realtà che esisteva anche nel passato, sebbene con caratteristiche diverse rispetto ad oggi. Infatti, un tempo, queste famiglie avevano origine soprattutto dalla morte di uno dei due coniugi,

molto spesso della madre a causa delle gravidanze e dei parti, dall'emigrazione dei mariti o dall'abbandono di giovani donne rese madri. Oggi, d'altra parte, esse derivano per buona parte da scelte volontarie dei coniugi e dal diffondersi delle separazioni e dei divorzi (Zanatta, 1997).

2.7 Stili di vita

Si analizzano ora le abitudini familiari e gli stili di vita dei singoli componenti: informazioni relative all'abitazione e ad altri beni durevoli, alle abitudini alimentari, alle vacanze, all'attitudine al fumo o all'alcool, alle letture.

2.7.1 Attitudine al fumo e all'alcool

Per le persone con più di 14 anni si hanno informazioni relative al consumo di sigarette e di bevande alcoliche. Per questi argomenti può presentarsi il problema delle risposte compiacenti, nonché dei rispondenti *proxy*. Per esempio, se il genitore ignora che il figlio fuma, ciò comporterà una sottostima del numero di fumatori.

In ogni caso, l'82,2% della popolazione con più di 14 anni non fuma. Le donne fumatrici rappresentano il 10,8% della popolazione femminile, gli uomini il 17,2% della popolazione maschile (Tab.2.28).

Tab.2.28. *Distribuzione percentuale della popolazione (sopra i 14 anni) residente nel Veneto nel 2004-2005 per attitudine al fumo e sesso*

Attitudine al fumo	Sesso		
	Femmine (n=3141)	Maschi (n=2972)	Totale (n=6113)
Non fumatori	86,1	78,0	82,2
Fumatori occasionali	3,1	4,8	3,9
Fumatori abituali	10,8	17,2	13,9
Totale	100,0	100,0	100,0

In tabella 2.29 si sono calcolati i tassi della popolazione che fuma abitualmente come rapporto tra gli individui, maschi e femmine rispettivamente, che fumano e gli individui totali in ciascuna classe d'età. Separatamente per i due sessi, si sono calcolati inoltre i tassi standardizzati per età come somma dei prodotti tra i

tassi in ciascuna classe di età e le frequenze relative di una popolazione tipo nella medesima classe d'età. Come popolazione tipo si è considerata quella rispondente al questionario.

Il tasso più elevato si ha in corrispondenza dei maschi (16,6%, contro l'11,0% delle femmine). Per entrambi i sessi i tassi più alti riguardano la classe d'età 45-54. Rispetto alle donne i maschi fumano di più ad ogni età, in particolare in quella più giovane (15-24), in cui si riscontra un approccio al fumo più anticipato tra i maschi.

Tab.2.29. *Tassi percentuali di fumatori abituali residenti nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Popolazione tipo (n=6177)
	Femmine (n=3192)	Maschi (n=2985)	
15-24	7,3	15,0	0,125
25-34	15,1	20,2	0,137
35-44	13,2	20,6	0,196
45-54	16,6	21,3	0,179
55-64	9,5	16,4	0,155
65 e più	4,6	7,6	0,208
Totale	11,0	16,6	1,0

Per quanto riguarda il consumo di alcolici, circa un terzo della popolazione (35,7%) è astemia, il 32,1% beve solo occasionalmente o quando è in compagnia ed il rimanente 31,2% fa consumo abituale di alcolici, in particolar modo durante i pasti. In questo caso le differenze tra maschi e femmine sono notevoli: mentre le donne astemie sono il 45,3%, gli uomini sono il 25,6%; di contro, i consumatori abituali sono il 23,9% tra le donne ed il 40,8% tra gli uomini; molto simile per i due sessi risulta, invece, la percentuale di coloro che hanno un consumo occasionale (30,8% delle donne e 33,6% degli uomini) (Tab.2.30).

Tab.2.30. *Distribuzione percentuale della popolazione (sopra i 14 anni) residente nel Veneto nel 2004-2005 per attitudine all'alcool e sesso*

Attitudine all'alcool	Sesso		Totale (n=6113)
	Femmine (n=3141)	Maschi (n=2972)	
Astemi	45,3	25,6	35,7
Consumatori occasionali	30,8	33,6	32,1
Consumatori abituali	23,9	40,8	32,2
Totale	100,0	100,0	100,0

Le donne astemie sono decisamente più frequenti rispetto agli uomini (45,1% contro 24,4%). Per entrambi i sessi, il tasso più elevato corrisponde alla classe d'età 15-24 (68,3% quello femminile e 56,0% quello maschile) e va diminuendo passando dalle classi più giovani a quelle più anziane (eccezion fatta per gli over 65). All'aumentare dell'età aumentano anche le differenze tra i due sessi: ad esempio, per gli individui tra i 55 e i 64 anni, quelli che contano meno astemi, i tassi per le femmine e i maschi ammontano rispettivamente a 33,5% e 12,3% (Tab.2.31).

Tab.2.31. *Tassi percentuali di astemi residenti nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Popolazione tipo (n=6177)
	Femmine (n=3192)	Maschi (n=2985)	
15-24	68,3	56,0	0,125
25-34	51,5	32,9	0,137
35-44	44,1	22,4	0,196
45-54	40,2	18,0	0,179
55-64	33,5	12,3	0,155
65 e più	40,9	16,4	0,208
Totale	45,1	24,4	1,0

2.7.2 Letture

L'indagine aveva anche lo scopo di analizzare i passatempi e gli interessi dei componenti, in particolar modo le preferenze relative alla lettura. Si noti che, anche per questo tipo di domande, si presenta il problema dei rispondenti *proxy*.

Dai dati rilevati emerge che il 70,6% delle persone coinvolte nell'indagine legge quotidiani, riviste o entrambi, il restante 29,4% non legge né l'uno né l'altro (Tab.2.32).

Tab.2.32. *Distribuzione percentuale della popolazione (sopra i 5 anni) residente nel Veneto nel 2004-2005 per lettura di quotidiani e/o riviste*

Letture (n=6685)	
Quotidiani	14,1
Riviste	15,9
Entrambi	40,6
Nessuno dei due	29,4
Totale	100,0

La maggior parte della popolazione legge sia quotidiani che riviste (il tasso standardizzato per età è pari a 37,5%). Seguono coloro che non si dedicano a nessuna di queste letture (27,5%). Minori risultano i tassi di chi legge solo quotidiani o solo riviste, a indicare che la maggior parte degli individui o legge o non legge affatto. Coloro che si dedicano a queste letture sono perlopiù adulti tra i 25 e i 54 anni; i giovani tra i 15 e i 24 anni comunque non rimangono del tutto estranei a queste letture ma prediligono in genere le riviste specializzate (il tasso in questo caso ammonta a 19,8%). (Tab.2.33).

Tab.2.33. *Tassi percentuali della popolazione (sopra i 5 anni) residente nel Veneto nel 2004-2005 per lettura di quotidiani e/o riviste e classi di età*

Classi di età	Letture				Popolazione tipo (n=7106)
	Quotidiani (n=983)	Riviste (n=1080)	Entrambi (n=2674)	Nessuno dei due (n=1948)	
6-14	0,5	6,4	5,8	57,5	0,124
15-24	8,6	17,4	34,6	29,8	0,119
25-34	10,7	19,8	45,3	22,6	0,119
35-44	14,9	16,7	44,8	22,3	0,170
45-54	18,6	16,0	47,0	17,2	0,155
55-64	20,2	16,3	45,3	17,0	0,134
65 o più	18,2	13,7	35,4	30,2	0,179
Totale	13,7	15,9	37,5	27,5	1,0

Il 30,5% degli intervistati legge regolarmente libri di narrativa, romanzi o altro, il 26,4% lo fa solo occasionalmente, mentre il 43,1% non lo fa mai (Tab.2.34).

Tab.2.34. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per lettura di libri*

Letture (n=6645)	
Regolarmente	30,5
Occasionalmente	26,4
Mai	43,1
Totale	100,0

Il tasso standardizzato per età di coloro che non si dedicano affatto alla lettura di libri ammonta a 40,8%. Osservando i tassi di tali individui nelle varie classi di età, si nota che essi vanno aumentando nel passaggio dalle classi più giovani a quelle più anziane; in particolare per gli anziani con più di 65 anni il tasso è pari a 58,6% (Tab.2.35).

Tab.2.35. *Tassi percentuali delle famiglie residenti nel Veneto nel 2004-2005 per lettura di libri e classi di età*

Classi di età	Lettura			
	Regolarmente (n=2004)	Occasionalmente (n=1771)	Mai (n=2910)	Popolazione tipo (n=7106)
5-14	22,1	18,5	24,1	0,134
15-24	38,0	30,6	30,5	0,109
25-34	33,3	27,5	37,5	0,119
35-44	31,1	28,0	39,7	0,170
45-54	32,1	26,0	40,6	0,155
55-64	26,8	25,6	46,3	0,134
65 o più	18,7	20,2	58,6	0,179
Totale	28,3	30,3	40,8	1,0

Per quanto riguarda la presenza di libri in casa, il 40,1% delle famiglie intervistate afferma di avere tra 100 e 500 libri, il 28,6% meno di 50 e il 22,2% tra 50 e 100.

Nelle famiglie in cui il massimo titolo di studio è il dottorato di ricerca o la specializzazione post-lauream il 50,9% possiede più di 500 libri. In quelle in cui il massimo titolo di studio è la laurea o il diploma di scuola media superiore (4-5 anni) oltre il 50% possiede tra 100 e 500 libri. Tra le famiglie meno istruite, la percentuale di quelle che possiedono oltre 500 libri è molto bassa (Tab.2.36).

Si tenga presente che ci può essere una sovrastima del numero di libri posseduti dalle famiglie a causa del comportamento di alcuni rispondenti che, associando questo dato al livello culturale della famiglia, potrebbero ritenere socialmente più “corretto” affermare di avere una buona quantità di libri a disposizione.

Tab.2.36. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per titolo di studio del componente più istruito e numero di libri in famiglia*

Titolo di studio	Numero di libri				Totale	Media	Sqm	Mediana
	Meno di 50	Tra 50 e 100	Tra 100 e 500	Più di 500				
Dottorato o specializzazione post-lauream (n=4)	0,0	0,0	49,1	50,9	100,0	3,5	12,0	4,0
Laurea (vecchio ordinamento o specialistica) (n=57)	4,7	4,5	52,5	38,3	100,0	3,2	14,4	3,0
Diploma universitario o laurea triennale (n=8)	0,0	13,7	59,1	27,2	100,0	3,1	13,0	3,0
Diploma di scuola media superiore (4-5 anni) (n=260)	6,9	13,1	57,6	22,4	100,0	2,9	14,7	3,0
Diploma di scuola media superiore (2-3 anni) (n=75)	20,0	23,8	44,0	12,2	100,0	2,5	16,4	3,0
Licenza di scuola media inferiore o avviamento (n=556)	21,1	23,0	46,3	9,6	100,0	2,4	17,0	3,0
Licenza elementare (n=880)	39,0	24,5	31,6	4,9	100,0	2,0	17,2	2,0
Nessun titolo ma sa leggere e scrivere (n=265)	37,5	22,8	36,3	3,4	100,0	2,0	17,5	2,0
Nessun titolo e non sa leggere e/o scrivere (n=335)	31,0	25,3	37,2	6,5	100,0	2,2	17,6	2,0
Totale (n=2440)	28,6	22,2	40,1	9,1	100,0	2,3	18,0	2,0

2.7.3 Abitudini alimentari

Per quanto concerne gli stili di vita della famiglia, è interessante capire se i componenti siano soliti riunirsi a tavola durante i pasti, evento che può essere occasione, spesso unica, di discussione e confronto e momento di condivisione della giornata. Questa informazione, che evidentemente riguarda solo le famiglie con almeno due componenti, può dare una misura di quanto gli impegni lavorativi assorbano i componenti a discapito dei rapporti familiari.

L'88,7% delle famiglie intervistate con più di un componente si trova a mangiare assieme anche nei giorni feriali, il 5,5% solo nei fine settimana e il 3,5% sporadicamente (Tab.2.37).

Tab.2.37. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per momento in cui la famiglia mangia riunita*

Momento (n=2130)	
Anche nei giorni feriali	88,7
Solo nei fine settimana	5,5
Sporadicamente	3,5
Mai	2,3
Totale	100,0

Il tasso, standardizzato per numero di componenti, di coloro che condividono il momento del pasto anche nei giorni feriali ammonta a 85,5%, un valore molto elevato che, tenendo conto del numero di membri, raggiunge il 90,9% nelle famiglie con due persone (spesso anziani pensionati) e diminuisce all'aumentare delle dimensioni della famiglia. Una tendenza inversa si ha nelle famiglie che condividono il pasto solo nei fine settimana: il tasso percentuale aumenta all'aumentare del numero di componenti. In questo conteggiano pesano le famiglie in cui sono presenti persone che lavorano e studiano fuori casa. Nelle famiglie che non mangiano mai assieme il tasso più elevato si ha, per ovvie ragioni, in corrispondenza di quelle costituite da almeno cinque componenti (5,2%) (Tab.2.38).

Tab.2.38. *Tassi percentuali delle famiglie residenti nel Veneto nel 2004-2005 per momento in cui la famiglia mangia riunita e numero di componenti*

Numero di componenti	Momento					Popolazione tipo (n=2193)
	Anche nei giorni feriali (n=1875)	Solo nei fine settimana (n=124)	Sporadicamente (n=83)	Mai (n=51)		
2	90,9	2,3	1,9	2,2		0,326
3	84,2	6,6	4,8	2,5		0,310
4	82,7	7,8	4,3	1,6		0,293
5 o più	77,3	7,8	5,8	5,2		0,071
Totale	85,5	5,7	3,8	2,3		1,0

In tabella 2.39 sono riportati i tassi percentuali delle abitudini alimentari della famiglia in relazione all'età del capofamiglia. Il tasso, standardizzato per età ammonta a 71,2%

Tab.2.39. *Tassi percentuali delle famiglie residenti nel Veneto nel 2004-2005 per momento in cui la famiglia mangia riunita e età del capofamiglia*

Eta del CF	Momento					Popolazione tipo (n=2525)
	Anche nei giorni feriali (n=1870)	Solo nei fine settimana (n=124)	Sporadicamente (n=83)	Mai (n=51)		
20-29	81,3	0,0	0,0	0,0		0,007
30-39	81,5	4,3	2,2	3,3		0,115
40-49	83,4	6,4	3,6	1,0		0,232
50-59	77,5	5,9	6,0	2,5		0,236
60-69	70,6	5,3	2,8	2,6		0,200
70-79	63,2	3,0	1,2	1,0		0,151
80 e più	51,3	0,6	0,6	2,6		0,059
Totale	71,2	4,1	2,2	1,2		1,0

Il motivo principale per il quale molte famiglie non riescono a consumare tutti i pasti assieme è di tipo lavorativo o di studio (92,6% del totale di famiglie che mangiano raramente assieme) (Tab.2.40).

Tab.2.40. *Distribuzione percentuale delle famiglie che non mangiano sempre assieme) residenti nel Veneto nel 2004-2005 per motivazione*

Motivazione (n=1773)	
Lavoro e/o studio	92,6
Malattia	1,6
Non vanno d'accordo	3,5
Nessun vero motivo	0,7
Altro	1,6
Totale	100,0

Nella maggioranza dei casi (48,5%), il pasto che le famiglie consumano assieme è la cena, alla fine della giornata di lavoro; il 28,7% si riunisce a tavola sia a pranzo che a cena, mentre una discreta percentuale di famiglie (17,9%) condivide tutti i pasti della giornata.

La situazione differisce chiaramente tra le famiglie in cui il capofamiglia lavora e quelle in cui è in pensione o in altra condizione (disoccupato o inabile al lavoro). In quest'ultimo caso, infatti, aumenta considerevolmente la percentuale di famiglie che consumano tutti i pasti assieme (passando dal 6,2% nelle famiglie in cui il capofamiglia lavora al 35,1% in quelle in cui è in pensione), mentre diminuisce quella che condivide esclusivamente la cena (passando dal 63,7% nelle famiglie in cui il capofamiglia lavora al 26,2% in quelle in cui è in pensione) (Tab.2.41).

Tab.2.41. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per pasti consumati assieme e condizione professionale del capofamiglia*

Pasti assieme	Condizione professionale del CF		Totale (n=2096)
	CF lavoratore (n=1237)	CF in altra condizione (n=859)	
Solo la prima colazione	0,2	0,0	0,1
Solo il pranzo	3,5	4,0	3,7
Solo la cena	63,7	26,2	48,5
Colazione e pranzo	0,1	0,1	0,1
Colazione e cena	1,2	0,6	1,0
Pranzo e cena	25,1	34,0	28,7
Tutti i pasti	6,2	35,1	17,9
Totale	100,0	100,0	100,0

2.8 I figli

In questa sezione si concentra l'attenzione sulla presenza in famiglia di figli da 0 a 19 anni.

L'1,8% della popolazione veneta ha un'età compresa tra i 0 e i 2 anni, il 2,9% da 3 a 5, l'8,8% da 6 a 14, il 5,5% da 15 a 19, il 14,5% da 65 a 79 e il 3,4% oltre i 79. La proporzione di maschi e femmine è all'incirca la stessa nelle diverse classi d'età, eccezion fatta per gli over 80, per cui il numero di donne è superiore rispetto a quello degli uomini (4,0% vs. 2,8%) (Tab.2.42).

Tab.2.42. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Totale (n=7106)
	Femmine (n=1442)	Maschi (n=782)	
0-2	1,7	1,8	1,8
3-5	2,7	3,2	2,9
6-14	8,9	8,6	8,8
15-19	5,3	5,7	5,5
20-24	5,1	5,7	5,4
25-29	5,1	6,4	5,7
30-39	14,0	13,6	13,8
40-54	24,8	25,0	24,8
55-64	13,4	13,3	13,4
65-79	15,0	13,9	14,5
80 e oltre	4,0	2,8	3,4
Totale	100,0	100,0	100,0

Il 53,8% delle famiglie in cui sono presenti componenti in età 0-19 (19,3% delle famiglie totali) ha al proprio interno un solo membro in questa fascia d'età. Il restante 46,2% (16,6% delle famiglie totali) è costituito da due o più figli in età 0-19 (Tab.2.43).

Tab.2.43. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per composizione delle famiglie con componenti in età 0-19*

Composizione della famiglia (n=894)	% (a)	% (b)
Famiglie con un minore	19,3	53,8
Famiglie con più minori	16,6	46,2
Totale	35,9	100,0

% (a) = percentuale sul totale delle famiglie

% (b) = percentuale sul totale delle famiglie con membri in età 0-19

Al fine di orientare gli interventi e le politiche a favore della famiglia con bambini in età prescolare, la presente ricerca ha analizzato le problematiche legate alla cura ed accudimento dei figli, dall'inserimento negli asili nidi o nelle scuole materne alla rete parentale ed amicale alla quale la famiglia è solita appoggiarsi. Oggigiorno, infatti, il compito genitoriale si assume con crescente difficoltà vista anche la complessità che si riscontra nella conciliazione tra gli impegni lavorativi e familiari.

2.8.1 Figli in età 0-2

Il 67,9% dei bambini in età compresa tra i 0 e i 2 anni passa la maggior parte della giornata con la madre o il padre, l'11,8% con altri familiari e il restante 6,7% frequenta l'asilo nido pubblico (Tab.2.44).

Ciò che determina lo scarso ricorso ai servizi è la presenza di familiari che si occupano in prima persona della cura del bambino. I servizi possono rappresentare una soluzione qualora venissero a mancare tali supporti interni alla famiglia.

Tab.2.44. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per persona che si occupa del bambino (0-2 anni) prevalentemente*

Chi se ne occupa (n=121)	
Mamma / papà	67,9
Altri famigliari	11,8
Baby sitter	3,2
Nido pubblico	6,7
Nido privato	10,4
Totale	100,0

Tra tutti i bambini che non frequentano il nido, il 10,5% lo frequenterà l'anno successivo, il 3,5% ha presentato domanda ma non è stata accettata, soprattutto per mancanza di posti, e il restante 86,0% non ha presentato domanda o, dopo averla presentata, ha cambiato idea e l'ha ritirata (Tab.2.45).

Tab.2.45. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per motivo per cui il bambino (0-2 anni) non frequenta il nido*

Motivo per cui non frequenta il nido (n=112)	
Lo frequenterà prossimamente	10,5
Non è stato accettato	3,5
Non ha presentato domanda	86,0
Totale	100,0

I bambini che frequentano il nido, quando sono a casa, vengono accuditi prevalentemente dalla madre (75,5%) (Tab.2.46).

Tab.2.46. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per persona che si occupa del bambino (0-2 anni) quando non è al nido*

Chi se ne occupa (n=21)	
Mamma	75,5
Altri familiari	24,5
Totale	100,0

2.8.2 Figli in età 3-5

I primi servizi per l'infanzia nacquero alla fine del XIX secolo come conseguenza di due esigenze: la custodia dei bambini delle lavoratrici dei ceti più poveri e la formazione di abitudini e atteggiamenti adeguati alle esigenze di una società che stava urbanizzandosi ed industrializzandosi. Proprie per le finalità educative che li caratterizzavano, in poco tempo attirarono l'attenzione anche delle famiglie dei ceti medi istruiti e oggi sono un'esperienza normale e condivisa che riguarda tutti i bambini in età prescolare.

I figli in età 3-5 nel 71,2% dei casi frequentano la scuola materna (Tab.2.47).

Tab.2.47. *Distribuzione percentuale delle famiglie (con figli in età 3-5) residenti nel Veneto nel 2004-2005 per frequentazione della scuola materna*

Frequenta (n=201)	
Si	71,2
No	28,8
Totale	100,0

Del 28,8% che non va all'asilo, l'87,5% andrà prossimamente e il 5,6% non frequenta perché accudito da un genitore a casa (Tab.2.48).

Tab.2.48. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per motivo per cui il bambino (3-5 anni) non frequenta la scuola materna*

Frequenta (n=57)	
Non è stato accettato	2,7
Frequenterà prossimamente	87,5
Un genitore è a casa	5,6
Lo accudisce un familiare	2,2
Lo accudisce una persona retribuita	2,0
Totale	100,0

L'83,5% dei bambini che vanno all'asilo, a casa sono accuditi prevalentemente da un genitore, il 15,3% da un altro familiare (Tab.2.49).

Tab.2.49. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per persona che si occupa del bambino (3-5 anni) prevalentemente*

Chi se ne occupa (n=201)	
Mamma / papà	83,5
Altri familiari	15,3
Persona retribuita	1,2
Totale	100,0

2.8.3 Figli in età 6-14

Per i bambini/ragazzi in età compresa tra i 6 e i 14 anni il problema dell'accudimento si fa sentire in misura minore, da un lato perché gli impegni scolastici occupano perlomeno la metà della loro giornata e dall'altro perché, col passare del tempo, diventano progressivamente più autonomi. D'altro canto, però, questo periodo di crescita e sviluppo, anche caratteriale, richiede cure ed attenzioni che spesso vanno al di là del semplice accudimento: è necessario un impegno costante e a volte si rende indispensabile il ricorso ai servizi e a strutture deputate alla cura di specifiche problematiche.

Per questi motivi, l'interesse è anche rivolto agli eventuali problemi di apprendimento, di relazione con i compagni e di condotta in classe. Fortunatamente si sono riscontrati pochi casi problematici: il 3,1% presenta problemi di apprendimento, il 2,4% di relazione e il 2,6% di condotta (Tab.2.50).

Tab.2.50. *Distribuzione percentuale della popolazione (6-14 anni) residente nel Veneto nel 2004-2005 per problemi di apprendimento, di relazione e di condotta*

	Problemi		
	Apprendimento (n=576)	Relazione (n=575)	Condotta (n=575)
Si, gravi	0,7	0,4	0,4
Si, lievi	2,4	2,0	2,2
No	96,9	97,6	97,4
Totale	100,0	100,0	100,0

Alle famiglie con figli in età 6-14 è stato chiesto anche se il figlio viene aiutato nello svolgimento dei compiti scolastici: nel 54,3% il bambino è autonomo, mentre nel 40,4% è la madre che aiuta il figlio. Le mamme aiutano i più piccoli, segno della cura del genitore nei confronti del figlio, mentre il 54,3% rappresenta perlopiù i ragazzi quattordicenni o quasi (Tab.2.51).

Tab.2.51. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per persona che aiuta il bambino/ragazzo (6-14 anni) nello svolgimento dei compiti*

Chi lo aiuta (n=575)	
Nessuno	54,3
Mamma	40,4
Papà	1,0
Altro familiare convivente	2,5
Altro familiare non convivente	0,3
Persona retribuita	1,1
Persona non retribuita	0,4
Totale	100,0

2.8.4 Figli in età 15-19

La quasi totalità dei figli in età 15-19, e precisamente il 93,4%, va ancora a scuola. Tra coloro che hanno smesso di studiare, oltre la metà (52,3%) ha già trovato un lavoro (Tab.2.52).

Tab.2.52. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 per motivo per cui non va a scuola*

Motivo (n=23)	
Si è diplomato e andrà a lavorare	1,4
Già lavora	52,3
Ha smesso di studiare e sta cercando lavoro	36,6
Motivi di salute (fisici o mentali)	4,8
Altro	4,9
Totale	100,0

L'interesse verso gli adolescenti tra i 15 e i 19 anni era particolarmente concentrato sul rapporto e il dialogo con i genitori relativamente ad uscite, amicizie, profitto scolastico e spese sostenute dal ragazzo.

I giovani che escono abitualmente con gli amici rappresentano il 90,7% della popolazione totale. Di questi, il 26,7% frequenta perlopiù case di amici, il 24,6% piazze o giardini pubblici, il 24,1% birrerie, bar o pub e il 21,5% patronati; solo lo 0,5% va con gli amici in discoteca. Tra i due sessi non vi sono notevoli differenze: rispetto ai maschi (18,2%), le femmine frequentano molto di più piazze o giardini pubblici (26,2%) (Tab.2.53).

Tab.2.53. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 per luogo di ritrovo abituale con gli amici e sesso*

Luogo di ritrovo	Sesso		
	Femmine (n=168)	Maschi (n=165)	Totale (n=333)
Casa di amici	25,0	27,3	26,1
Piazza/giardino pubblico	26,2	18,2	22,2
Patronato/ricreatorio	17,2	24,8	21,0
Discoteca	0,6	0,6	0,6
Birreria/bar/pub	27,4	26,7	27,1
Altro	3,6	2,4	3,0
Totale	100,0	100,0	100,0

Il 9,3% dei ragazzi non esce di casa la sera; la maggior parte di questi lo fa perché non ha amici (40,1%), per mancanza di interesse (31,1%) o per dedicare il tempo allo studio (20,6%). Tra i due sessi sono evidenti le differenze: gli impegni legati allo studio riguardano soprattutto le femmine (34,3% contro il 13,7% dei

maschi), mentre la mancanza di interesse ad uscire tocca perlopiù i maschi (41,5% contro il 10,8% delle femmine) (Tab.2.54).

Tab.2.54. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 per motivo per cui non esce la sera e sesso*

Motivo	Sesso		
	Femmine (n=11)	Maschi (n=24)	Totale (n=35)
Studio	34,3	13,7	20,6
I genitori non glielo permettono	10,8	6,8	8,2
Non ha amici	44,1	38,0	40,1
Non ci tiene ad uscire	10,8	41,5	31,1
Totale	100,0	100,0	100,0

Gran parte dei genitori riesce ad avere un dialogo sereno con i figli adolescenti; solo l'1,0% della popolazione non riesce affatto ad avere un rapporto di dialogo. Le discussioni riguardano soprattutto le uscite serali (27,4% delle famiglie con almeno un figlio in età compresa tra i 15 e i 19 anni) e il profitto scolastico (23,0%). Solo per quanto riguarda il profitto scolastico vi sono differenze fra i due sessi: le femmine, più diligenti, hanno solo nel 14,6% discussioni con i genitori, mentre i maschi nel 31,1% (Tab.2.55).

Tab.2.55. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 per argomenti di discussione con i genitori e sesso*

Argomento di discussione	Sesso		
	Femmine (n=177)	Maschi (n=188)	Totale (n=365)
Uscite pomeridiane	12,5	11,1	11,8
Uscite serali	28,4	26,4	27,4
Profitto a scuola	14,6	31,1	23,0
Spese per lui/lei	16,0	12,2	14,1
Amicizie	100,0	100,0	100,0

Un'ulteriore analisi è andata ad individuare se la presenza di discussioni può essere influenzata dal sesso dell'adolescente o dal fatto che pratici o meno attività extrascolastiche (sport, musica, recitazione). A livello familiare, invece, si è voluto scoprire se le discussioni sono collegate al fatto che la famiglia percepisca qualche tipo di disagio (sociale, economico, sanitario) o ad altre variabili che rilevano, ad

esempio, se la famiglia va in vacanza, se guarda la TV o consuma i pasti insieme. Le unità con mancate risposte, presenti quasi esclusivamente per la variabile che rileva se i componenti guardano la TV assieme, sono state escluse dall'analisi.

Osservato che nel campione la frazione di adolescenti che hanno contrasti con i genitori è del 46,9%, si sono andate ad effettuare le analisi sopra enunciate, utilizzando come misura della forza della relazione il rapporto crociato, una stima del rischio relativo (Agresti, 1984).

L'età del ragazzo sembra non influire in maniera pesante sul rapporto con i genitori, sebbene i figli in età 18-19 abbiano più motivi di scontro rispetto ai più giovani. Il rischio di avere discussioni è del 14,7% maggiore per i ragazzi in età 18-19 rispetto a quelli in età 15-17. Differenziando per sesso, si nota che, sia nella fascia d'età 15-17 che in quella 18-19, hanno più litigi coi genitori i maschi rispetto alle femmine; tuttavia mentre per le femmine in età 15-17 il rischio di litigi è del 7,6% maggiore rispetto alle maggiorenni, per i maschi è più alto e precisamente pari a 1,282 (Tab.2.56).

Tab.2.56. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 che ha discussioni coi genitori per fascia d'età e sesso*

Età	Sesso		
	Femmine	Maschi	Totale
15-17 (n=224)	42,5	45,3	44,1
18-19 (n=137)	44,3	51,5	47,5

Il numero di attività (sportive o artistiche) praticate dagli adolescenti influisce sul rapporto tra genitori e figli. Il 58,2% dei ragazzi impegnati in almeno due attività extra scolastiche hanno discussioni con i loro genitori.

Di qui, il rischio di avere discussioni è del 74,4% maggiore se il figlio pratica almeno due attività rispetto al non praticarne alcuna (rapporto crociato = 1,744): i genitori si oppongono alle numerose uscite pomeridiane o ad attività svolte dai figli. Distinguendo tra i due sessi, si nota che tale rischio è maggiore per i maschi (2,101) che per le femmine (1,589) (Tab.2.57).

Tab.2.57. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 che ha discussioni coi genitori per numero di attività extra scolastiche e sesso*

Numero di attività	Sesso		
	Femmine	Maschi	Totale
0 (n=119)	37,1	53,1	44,4
1 (n=196)	46,2	39,8	42,6
2 o più (n=48)	48,2	70,4	58,2

In modo evidente per i maschi, le discussioni sono più frequenti quando entrambi i genitori lavorano, rispetto a quando uno solo dei due lavora (rapporto crociato = 2,053). Per le femmine, al contrario si hanno più discussioni laddove solo un genitore lavora. (Tab.2.58).

Tab.2.58. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 che ha discussioni coi genitori per genitori occupati e sesso*

Genitori occupati	Sesso		
	Femmine	Maschi	Totale
Solo uno dei due (n=232)	45,8	36,0	41,0
Entrambi (n=132)	41,3	53,7	47,7

La relazione tra discussioni e abitudine a condividere i pasti assieme appare più complessa: da una parte condividere il momento dei pasti assieme rende la famiglia più unita e comporta un rapporto più sereno con i figli (nel passaggio da “talvolta” a “spesso” il numero di litigi diminuisce in maniera evidente); dall'altra, nei casi in cui i genitori e i figli non pranzano mai o quasi mai assieme, il numero di contrasti risulta particolarmente basso, probabilmente a causa delle minori occasioni di discussioni e controversie (Tab.2.59).

Tab.2.59. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 che ha discussioni coi genitori per pasti consumati assieme in famiglia e sesso*

Pasti assieme	Sesso		
	Femmine	Maschi	Totale
Sempre (n=310)	42,8	47,7	45,2
Talvolta (n=43)	47,9	54,3	51,7
Mai (n=11)	31,0	18,3	23,0

A fronte di un rischio relativo di 1,484, si può dedurre che la relazione tra i contrasti tra genitori e figli e l'abitudine a guardare la TV assieme sia piuttosto

significativa: nelle famiglie i cui componenti guardano assieme la televisione le discussioni tra genitori e figli sono minori (Tab.2.60).

Tab.2.60. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 che ha discussioni coi genitori per abitudine a guardare la tv assieme in famiglia e sesso*

Tv assieme	Sesso		
	Femmine	Maschi	Totale
Spesso (n=204)	40,4	43,8	42,3
Raramente (n=146)	48,7	55,8	52,1

Infine, la presenza di problemi (sociali, economici o sanitari) in famiglia fa aumentare la probabilità che il figlio abbia contrasti con i genitori del 26,1% (rapporto crociato=1,261). In questo risultato pesano quasi esclusivamente le femmine: il 61,6% delle ragazze nelle cui famiglie vi sono problemi importanti hanno discussioni con i loro genitori, contro il 42,2% di quelle che vivono in famiglie senza o con problemi di poco conto (Tab.2.61).

Tab.2.61. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 che ha discussioni coi genitori per disagio familiare e sesso*

Disagio	Sesso		
	Femmine	Maschi	Totale
Nulla o lieve (n=344)	42,2	47,4	46,8
Grave o estremo (n=19)	61,6	48,3	52,6

Il risultato necessiterebbe, tuttavia, di ulteriori approfondimenti dal momento che causa ed effetto si intersecano.

In conclusione si può affermare che, come le attività extrascolastiche degli adolescenti, anche quelle relative alla famiglia nel suo complesso sembrano influire sul rapporto genitori-figli: la condivisione di alcuni momenti per certi versi ha un effetto positivo sul dialogo tra genitori e figli, portando ad un ambiente familiare più sereno, per altri può avere un effetto contrario nel senso che i figli si sentono eccessivamente controllati da parte dei genitori.

2.8.5 Problemi di apprendimento scolastico

Tra i ragazzi in età 6-19 anni si è andata a studiare la relazione esistente tra la presenza o meno di problemi di apprendimento scolastico e le variabili che rilevano

rispettivamente se il ragazzo è affidato a qualcuno quando non è a scuola, il titolo di studio e lo stato civile dei genitori.

Solo il 4,0% dei ragazzi in questo intervallo di età risulta avere qualche problema scolastico di apprendimento, una percentuale piuttosto ridotta che, tuttavia, risulta sufficiente al fine di ottenere valide conclusioni (Tab.2.62).

Tab.2.62. *Distribuzione percentuale della popolazione (6-19 anni) residente nel Veneto nel 2004-2005 per problemi scolastici di apprendimento*

Problemi (n=918)	
Si	4,0
No	96,0
Totale	100,0

Una preliminare analisi mostra che tra i ragazzi con problemi di apprendimento, la maggior parte (48,7%) ha un fratello/sorella, il 27,0% è figlio unico e il restante 24,3% ha due o più fratelli/sorelle (Tab.2.63).

Tab.2.63. *Distribuzione percentuale della popolazione con problemi di apprendimento (6-19 anni) residente nel Veneto nel 2004-2005 per numero di fratelli*

Fratelli (n=37)	
Zero	27,0
Uno	48,7
Due o più	24,3
Totale	100,0

Inoltre, nell'11,0% delle famiglie in cui almeno un figlio ha problemi scolastici di apprendimento, i genitori sono separati/divorziati.

Analizzando il comportamento della variabile dipendente (problemi scolastici di apprendimento) al variare delle modalità di quella indipendente (il bambino/ragazzo è affidato a qualcuno quando non è a scuola), risulta che, tra i ragazzi con problemi, è maggiore la percentuale di coloro che sono seguiti da qualcuno (4,6%) rispetto a quella di coloro che non lo sono (3,2%) (rapporto crociato = 1,449) (Tab.2.64).

Tab.2.64. *Distribuzione percentuale della popolazione (6-19 anni) residente nel Veneto nel 2004-2005 per problemi scolastici di apprendimento e affidamento*

Problemi	Affidamento		
	Si (n=544)	No (n=373)	Totale (n=917)
Si	4,6	3,2	4,0
No	95,4	96,8	96,0
Totale	100,0	100,0	100,0

Questo risultato può essere dovuto al fatto che chi ha problemi di apprendimento ha maggior bisogno di essere affidato a qualcuno che lo aiuti nello svolgimento dei compiti.

In questa interpretazione il ruolo delle due variabili è invertito: la causa (il non essere seguiti da qualcuno) diventa l'effetto dell'aver problemi di apprendimento.

Risultati analoghi si ottengono se si analizza il sottoinsieme di ragazzi in età 6-14 (Tab.2.65).

Tab.2.65. *Distribuzione percentuale della popolazione (6-14 anni) residente nel Veneto nel 2004-2005 per problemi scolastici di apprendimento e affidamento*

Problemi	Affidamento		
	Si (n=515)	No (n=60)	Totale (n=575)
Si	2,9	1,7	2,8
No	97,1	98,3	97,2
Totale	100,0	100,0	100,0

Il rapporto crociato, pari a 1,77, indica che i ragazzi affidati a qualcuno hanno un rischio di avere problemi del 77,0% maggiore rispetto ai ragazzi che non sono affidati a nessuno.

Molto più alto, e precisamente pari a 14,45, è il rischio relativo se si considerano i ragazzi tra i 15 e i 19 anni. Gli adolescenti risultano più a rischio per il fatto che sono in una fase in cui si scontrano con vari tipi di problemi, tra cui quelli scolastici (Tab.2.66).

Tab.2.66. *Distribuzione percentuale della popolazione (15-19 anni) residente nel Veneto nel 2004-2005 per problemi scolastici di apprendimento e affidamento*

Problemi	Affidamento		Totale (n=342)
	Si (n=29)	No (n=313)	
Si	34,5	3,5	6,1
No	65,5	96,5	93,9
Totale	100,0	100,0	100,0

Successivamente, si è voluto indagare se il titolo di studio dei genitori abbia influenza sui problemi scolastici di apprendimento dei figli in età 6-19. Il titolo di studio è stato ripartito in tre classi (alto, medio e basso) corrispondenti rispettivamente a:

- diploma universitario o laurea di 3 anni o titolo superiore;
- diploma di scuola media superiore (2-3 anni o 4-5 anni);
- licenza di scuola media inferiore o titolo inferiore.

La percentuale più elevata di bambini/ragazzi senza problemi scolastici hanno genitori con titolo di studio basso, mentre la percentuale più elevata di quelli con problemi hanno genitori con titolo alto (Tab.2.67 e Tab.2.68).

Tab.2.67. *Distribuzione percentuale della popolazione (6-19 anni) residente nel Veneto nel 2004-2005 per problemi scolastici di apprendimento e titolo di studio del padre*

Problemi	Titolo di studio del padre			Totale (n=616)
	Alto (n=69)	Medio (n=268)	Basso (n=279)	
Si	8,7	4,8	4,7	5,2
No	91,3	95,2	95,3	94,8
Totale	100,0	100,0	100,0	100,0

Tab.2.68. *Distribuzione percentuale della popolazione (6-19 anni) residente nel Veneto nel 2004-2005 per problemi scolastici di apprendimento e titolo di studio della madre*

Problemi	Titolo di studio della madre			
	Alto (n=71)	Medio (n=285)	Basso (n=291)	Totale (n=647)
Si	9,9	4,2	5,5	5,4
No	90,1	95,8	94,5	94,6
Totale	100,0	100,0	100,0	100,0

Questo risultato mostra che nelle famiglie dove il livello di istruzione dei genitori è elevato i figli hanno meno propensione ad ottenere buoni risultati scolastici.

Il rischio di avere problemi di apprendimento è del 150,6% maggiore se le madri hanno titolo di studio alto rispetto all'averlo medio e dell'88,8% maggiore se hanno titolo di studio alto rispetto all'averlo basso.

Se si considera il titolo di studio della madre, i risultati sono leggermente diversi: i figli delle madri con titolo di studio medio hanno un rischio del 25% inferiore di avere problemi, i figli delle madri con titolo di studio alto hanno un rischio dell'87% maggiore rispetto ai figli delle madri con titolo di studio basso.

Una spiegazione dei risultati ottenuti può cercarsi nel fatto che i genitori con titolo di studio elevato hanno in genere un lavoro che li impegna maggiormente e, pertanto, meno tempo per prendersi cura dei figli.

Da ultimo, si è voluta studiare l'eventuale relazione che intercorre tra i problemi scolastici e lo stato civile dei genitori.

I bambini/ragazzi dei genitori coniugati/conviventi sembrano avere minori problemi di apprendimento scolastico (Tab.2.69 e Tab.2.70).

Tab.2.69. *Distribuzione percentuale della popolazione (6-19 anni) residente nel Veneto nel 2004-2005 per problemi scolastici di apprendimento e stato civile del padre*

Problemi	Stato civile del padre		
	Coniugato/convivente (n=608)	Altro (n=11)	Totale (n=619)
Si	5,1	9,1	5,2
No	94,9	90,9	94,8
Totale	100,0	100,0	100,0

Tab.2.70. *Distribuzione percentuale della popolazione (6-19 anni) residente nel Veneto nel 2004-2005 per problemi scolastici di apprendimento e stato civile della madre*

Problemi	Stato civile della madre		Totale (n=651)
	Coniugata/convivente (n=615)	Altro (n=36)	
Si	5,2	8,3	5,4
No	94,8	91,7	94,6
Totale	100,0	100,0	100,0

Il 5,1% dei padri coniugati contro il 9,1% di quelli in un'altra condizione civile, ha il figlio con problemi. Per le madri tali percentuali risultano rispettivamente pari a 5,2 e 8,3.

Il calcolo dei rapporti crociati mostra che i figli di padri non coniugati/conviventi hanno un rischio di avere problemi dell'86,0% superiore rispetto a quello dei figli di padri coniugati/conviventi. Tale percentuale ammonta a 66,0% per le madri.

È, pertanto, probabile che nelle famiglie in cui i genitori sono coniugati/conviventi ci sia una maggiore stabilità e una maggiore disponibilità dei genitori nei confronti dei figli.

2.8.6 Figli con più di 20 anni

Per quanto riguarda i giovani con più di 20 anni, solo il 23,8% frequenta un qualche corso di studi. Di questi, la maggior parte è iscritto ad un corso di laurea (83,8%) (Tab.2.71).

Tab.2.71. *Distribuzione percentuale della popolazione (sopra i 20 anni) residente nel Veneto nel 2004-2005 per tipo di corso di studi frequentato*

Corso di studi (n=274)	
Dottorato o specializzazione post-lauream	3,3
Laurea specialistica o vecchio ordinamento	41,1
Laurea triennale	42,7
Scuola superiore	7,0
Altro	5,8
Totale	100,0

Si riscontra una marcata tendenza a vivere sempre più a lungo in casa con i genitori, ritardando l'uscita dalla famiglia di origine e quindi l'assunzione di responsabilità.

Osservando la situazione italiana, si nota come spesso tale scelta persista anche quando i giovani hanno già un lavoro e un reddito propri: il 61,8% dei maschi e il 54,63% delle femmine italiane che lavorano vivono con almeno un genitore (Stella, 1997). Rimangono più a lungo in famiglia i maschi rispetto alle femmine, ma ciò è dovuto soprattutto alla differenza di età al matrimonio e non ad una diversa modalità di uscita dalla famiglia di origine (Saraceno, 2003).

Le motivazioni della permanenza nella famiglia d'origine sono molteplici, legate prima di tutto a questioni di comodità o comunque ad una scelta personale (44,0%), ragioni economiche (20,9%) e di studio (20,6%) (Tab.2.72).

Tab.2.72. *Distribuzione percentuale della popolazione (sopra i 20 anni) residente nel Veneto nel 2004-2005 per motivo di permanenza in famiglia*

Motivazione (n=1020)	
Studio	20,6
Sta per trasferirsi	6,0
Scelta	44,0
Precarietà economica	20,9
Motivi familiari	2,7
Problemi di salute	2,3
Altro	3,5
Totale	100,0

Cavalli e Galland (1995) ritengono riduttivo attribuire ad un unico fattore il rallentamento dei giovani nell'ingresso dell'età adulta. Sarebbe fuorviante pensare, scrive Galland, "che una ripresa economica e una riduzione della disoccupazione possa segnare automaticamente il ritorno a modelli tradizionali di transizione alla maturità".

Si riscontra una marcata differenza tra le regioni del Centro-Nord e quelle del Sud: in quest'ultime solo una piccola porzione di giovani lavoratori sceglie di vivere con i genitori. Il motivo di questa differenza non è ben chiaro, ma si ritiene credibile il fatto che i giovani settentrionali, rispetto ai coetanei meridionali, preferiscano la comodità e la scarsa dispendiosità assicurate dalla famiglia di origine alla spesa

richiesta da un'abitazione propria. Inoltre, i giovani occupati che risultano coniugati sono più numerosi al Sud che al Nord: il tempo che intercorre tra l'inizio dell'attività lavorativa e il matrimonio è minore al Sud che al Nord. Non verosimile sembra, invece, imputare la differenza al fattore disoccupazione, che anzi riguarda maggiormente le zone del Sud (Stella, 1997).

Spesso gli studiosi si interrogano sul perché le generazioni più giovani prolunghino la propria permanenza, come figli, nella casa dei genitori, ma non si chiedono che cosa succeda in queste famiglie in cui sono presenti due generazioni di adulti: come sono allocate le responsabilità, organizzata la vita quotidiana, gestite le risorse finanziarie, trovato l'equilibrio tra le esigenze di autonomia di ciascuno. Che i giovani stiano a lungo in famiglia è un dato di fatto, che però non sembra far nascere la curiosità riguardo ai rapporti e alle strategie relazionali che sollecita. Ciò che viene messa in luce è la solidità del concetto di appartenenza familiare, ma non vi sono ricerche riguardo al lavoro di reciproco adattamento che una situazione di questo tipo richiede (Barbagli e Saraceno, 1997).

2.9 Gli anziani

Il rapporto annuale ISTAT sull'evoluzione demografica dell'Italia presenta una situazione alquanto allarmante. L'Italia detiene l'indice di vecchiaia più elevato del mondo.

La speranza di vita alla nascita nel 2005 (http://demo.istat.it/altridati/indicatori/2005/Tab_6.pdf) è di 77,6 anni per gli uomini e 83,2 per le donne, valori, in entrambi i casi, superiori alla media europea. Tuttavia, se da un lato l'aumento della speranza di vita è sinonimo di benessere e progresso, dall'altro non è sempre garanzia di una buona qualità di vita. Questo fenomeno, tipicamente italiano, interessa anche la regione del Veneto e richiede un'attenzione particolare nella programmazione di politiche ed interventi a favore delle problematiche legate alla terza età. I problemi che riguardano gli anziani sono spesso intrinsecamente legati alla famiglia, non solo laddove il nucleo è composto da soli anziani ma anche nel

caso in cui la famiglia si prenda cura di un anziano facendolo risiedere nella propria abitazione.

Nel 29,1% delle famiglie intervistate è presente almeno un anziano tra i 65 ed i 79 anni, mentre nell'8,0% delle famiglie c'è almeno un ultraottantenne.

Il 24,6% delle famiglie in cui sono presenti anziani sono unipersonali, ossia si tratta di anziani che vivono da soli (8,5% del totale delle famiglie), il 28,6% è composto da una coppia di anziani (9,9% del totale della famiglia) ed il restante 46,8% è rappresentato da famiglie in cui è presente almeno un anziano oltre a componenti di età inferiore a 65 anni (Tab.2.16).

Nonostante queste percentuali sono pochi gli anziani che effettivamente risultano isolati: la maggior parte fa parte della rete di sostegno primaria, ha figli che vivono vicini, da cui riceve frequenti visite e telefonate (Facchini, 1997).

Per analizzare la qualità della vita degli anziani è importante approfondire, oltre alle condizioni di salute, anche come normalmente impiegano il proprio tempo, cercando di capire se viene speso in modo costruttivo, dedicandosi ad attività ricreative, hobby, piccoli lavori manuali, coltivando interessi culturali o magari prendendosi cura dei nipoti. Questo, infatti, fa sì che l'anziano non sviluppi un senso di inadeguatezza e di disagio rispetto all'ambiente circostante, ma che al contrario si senta inserito in una realtà stimolante in grado di mantenere vivi i suoi interessi e valorizzare le sue potenzialità. È emerso che il 75,5% della popolazione al di sopra dei 65 anni si prende cura della casa, dell'orto o degli animali da cortile, il 62,3% si dedica a lavori manuali (quali cucito, ricamo, bricolage), il 36,0% accudisce i figli di familiari o conoscenti, il 21% frequenta circoli o luoghi di ritrovo, il 20,1% partecipa con una certa frequenza a gite o attività ricreative, il 9,3% fa volontariato ed il 3,3% frequenta l'università per anziani (Tab.2.73).

Tab.2.73. *Distribuzione percentuale della popolazione (sopra i 65 anni) residente nel Veneto nel 2004-2005 per attività praticate nel tempo libero*

Attività	
Cura della casa, orto, animali da cortile (n=952)	75,5
Lavori manuali (n=768)	62,3
Lettura (n=733)	59,2
Ascolto di musica (n=572)	47,3
Accudimento di bambini (n=431)	36,0
Circoli e luoghi di ritrovo (n=256)	21,0
Gite ed attività ricreative (n=257)	21,0
Volontariato (n=115)	9,3
Università degli anziani (n=40)	3,3

Per quanto riguarda le condizioni di salute, il 31,9% della popolazione in età compresa fra i 65 e i 79 anni ed il 39,9% di quella con più di 79 anni, è affetta da qualche malattia cronico-degenerativa. La tabella 2.74 dà una panoramica completa del tipo di affezione o degli apparati principalmente coinvolti per le due classi di età. Eccezion fatta per le malattie delle ossa e degli arti e per le neoplasie, è evidente come, all'aumentare dell'età, aumentino le percentuali degli anziani che presentano qualche tipo di patologia.

Tab.2.74. *Distribuzione percentuale della popolazione (sopra i 65 anni) residenti nel Veneto nel 2004-2005 per principali malattie cronico-degenerative*

Tipo di malattia	Età	
	65-79	80 o più
Metabolismo (n=132)	21,8	24,8
Sistema cardiocircolatorio (n=300)	47,6	49,6
Vie respiratorie (n=91)	12,9	21,8
Apparato digerente (n=46)	7,5	11,1
Ossa ed articolazioni (n=79)	14,1	10,3
Artrite artrosi (n=155)	30,3	27,6
Sistema immunitario (n=5)	0,4	0,0
Neoplasie (n=48)	7,7	6,9
Sistema Nervoso (n=23)	4,7	11,9
Disturbi psichici (n=44)	6,2	6,9

Secondo i dati dell'ultima Indagine Istat sulla salute, le persone in Italia che hanno bisogno di assistenza in seguito a problemi di salute sono 3,9 milioni. Gli anziani con più di 65 anni ne rappresentano il 62% e la maggior parte ha bisogno di un aiuto continuativo.

Il grado di indipendenza ed autonomia diminuisce con l'aumentare dell'età: l'anziano ultraottantenne che presenta qualche patologia necessita in misura maggiore di aiuto per svolgere le attività quotidiane come mangiare, vestirsi, lavarsi, e muoversi (Tab.2.75).

Tab.2.75. *Distribuzione percentuale della popolazione (sopra i 65 anni) residente nel Veneto nel 2004-2005 attività quotidiane per le quali necessita di aiuto*

Attività	Età	
	65-79	80 o più
Mangiare (n=58)	12,1	25,4
Lavarsi (n=61)	15,8	36,3
Vestirsi (n=58)	14,7	33,8
Muoversi in casa (n=85)	20,4	43,6
Salire / scendere le scale (n=101)	22,5	47,6

Ciò comporta naturalmente un impegno da parte della famiglia che si prende cura di questa persona sia nel caso viva sotto lo stesso tetto, sia nel caso abiti altrove. Nella maggior parte delle situazioni l'anziano è autosufficiente (81,0%). Chi si occupa prevalentemente in famiglia di coloro che non sono autonomi è il capofamiglia o coniuge (15,1%), in pochi casi il figlio (2,2%) (Tab.2.76).

Tab.2.76. *Distribuzione percentuale della popolazione (sopra i 65 anni) residente nel Veneto nel 2004-2005 per persona che si prende cura dell'anziano malato e sesso*

Chi se ne occupa	Sesso		
	Femmine (n=339)	Maschi (n=289)	Totale (n=628)
E' autosufficiente	43,9	37,1	81,0
Capofamiglia/coniuge	7,5	7,6	15,1
Figlio/a	1,6	0,6	2,2
Fratello/a	0,0	0,2	0,2
Altro parente	0,0	0,2	0,2
Altro	1,0	0,3	1,3
Totale	54,0	46,0	100,0

Analisi analoghe a quelle sopra riportate, sono state condotte per i casi di anziani che presentano qualche forma di disabilità.

La tabella 2.77 mostra le disabilità più comuni. In questo caso si può notare come sia soprattutto la disabilità motoria ad avere una maggiore incidenza sugli ultraottantenni.

Tab.2.77. *Distribuzione percentuale della popolazione (sopra i 65 anni) residente nel Veneto nel 2004-2005 per tipo di disabilità più frequenti*

Tipo di disabilità	Età	
	65-79	80 o più
Cecità (n=25)	15,6	8,9
Sordomutismo (n=4)	1,1	0,0
Sordità (n=14)	8,7	11,7
Ritardo mentale (n=73)	12,5	11,5
Invalità motoria totale (n=34)	21,3	28,4
Invalità motoria parziale (n=143)	32,3	47,8

Anche per quanto concerne il grado di autonomia nella conduzione delle attività quotidiane è molto evidente come all'avanzare dell'età aumenti la richiesta di aiuto da parte dell'anziano con disabilità (Tab.2.78).

Tab.2.78. *Distribuzione percentuale della popolazione (sopra i 65 anni) residente nel Veneto nel 2004-2005 per attività quotidiane per le quali necessita di aiuto*

Attività	Età	
	65-79	80 o più
Mangiare (n=46)	27,3	40,1
Lavarsi (n=75)	43,2	53,3
Vestirsi (n=74)	43,2	53,3
Muoversi in casa (n=84)	52,5	67,7
Salire / scendere le scale (n=110)	64,7	73,1

La tabella 2.79 rispecchia l'analogia (Tab.2.76) presentata nel caso di malattie cronico-degenerative: a prendersi prevalentemente cura dell'anziano disabile è il capofamiglia o il coniuge (70,9%) (famiglie di anziani in cui il marito/la moglie si prende cura del coniuge invalido), a cui seguono i figli (13,3%). Manca in questo caso la categoria degli autosufficienti: la persona anziana e disabile ha, in ogni caso, bisogno di assistenza.

Tab.2.79. *Distribuzione percentuale della popolazione (sopra i 65 anni) residente nel Veneto nel 2004-2005 per persona che si prende cura dell'anziano disabile e sesso*

Chi se ne occupa	Sesso		Totale (n=226)
	Femmine (n=118)	Maschi (n=108)	
Capofamiglia/coniuge	28,6	42,3	70,9
Figlio/a	10,8	2,5	13,3
Fratello/a	1,6	1,2	2,7
Altro parente	2,1	0,8	3,0
Altro	7,8	2,3	10,1
Totale	50,9	49,1	100,0

All'aumento del rischio di disabilità e malattia dovuto all'avanzare dell'età si accompagna un modesto aumento del fenomeno della istituzionalizzazione. Adirittura, la forte prevalenza di donne vedove e di celibi e nubili mostra come il ricorso agli istituti sia più frequente non tanto, o non solo, tra chi ha un grado di non autosufficienza troppo grave, ma tra coloro che non possono contare sull'aiuto di familiari in grado di accudirli (Saraceno, 2003).

2.10 Stati critici

Gli "stati critici" rappresentano condizioni negative della famiglia e dei singoli componenti, che possono creare difficoltà. Ne sono esempi l'invalidità, la malattia e la disoccupazione.

2.10.1 Invalidità

Per quanto riguarda l'invalidità fisica, mentale e sensoriale, nel 7,7% delle famiglie intervistate è presente almeno un invalido. Il 93,8% delle famiglie in cui sono presenti invalidi (7,2% delle famiglie totali) è composto da un solo membro affetto da invalidità. Il restante 6,2% (0,5% delle famiglie totali) è costituito da due o più invalidi (Tab.2.80).

Tab.2.80. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per composizione delle famiglie con membri invalidi*

Composizione della famiglia (n=211)	% (a)	% (b)
Famiglie con un invalido	7,2	93,8
Famiglie con più invalidi	0,5	6,2
Totale	7,7	100,0

% (a) = percentuale sul totale delle famiglie

% (b) = percentuale sul totale delle famiglie con membri invalidi

L'invalidità più diffusa è di tipo motorio parziale (47,9% del totale della popolazione affetta da qualche tipo di invalidità) (Tab.2.81).

Tab.2.81. *Distribuzione percentuale della popolazione (invalida) residente nel Veneto nel 2004-2005 per tipo di invalidità*

Tipo di invalidità (n=211)	% (a)
Cecità	12,1
Sordomutismo	2,4
Sordità	6,2
Ritardo mentale lieve	11,9
Ritardo mentale grave	9,6
Invalidità motoria parziale	47,9
Invalidità motoria totale	15,3

% (a) = percentuale sul totale dei membri invalidi in ciascuna tipologia di invalidità

Il 12,1% delle famiglie con invalidi è composto da una persona che vive sola, principalmente con più di 70 anni, il 26,9% è, invece, composto da due membri.

L'invalidità interessa tanto le donne quanto gli uomini (il tasso ammonta a 3,1% per le prime e a 3,4% per i secondi) e le età più colpite sono quelle oltre gli ottant'anni. È naturale che all'aumentare dell'età aumenti il numero di invalidi in quanto l'invalidità più frequente è quella di tipo motorio, che colpisce soprattutto le persone più anziane.

Per le donne con più di 90 anni il tasso di invalidità è pari a 55,6%, mentre per i coetanei uomini a 30,4%, risultato che deve essere letto tenendo conto del fatto che le donne sono più longeve (Tab.2.82).

Tab.2.82. *Tassi percentuali di invalidità della popolazione residenti nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Popolazione tipo (n=7106)
	Femmine (n=3662)	Maschi (n=3444)	
0-9	0,3	1,0	0,085
10-19	0,6	1,6	0,104
20-29	0,8	1,5	0,111
30-39	1,4	1,1	0,138
40-49	1,9	1,7	0,173
50-59	2,4	2,6	0,152
60-69	4,0	4,8	0,121
70-79	7,7	10,4	0,082
80-89	23,8	20,7	0,029
90 e più	55,6	30,8	0,005
Totale	3,1	3,4	1,0

2.10.2 Presenza di malati cronico-degenerativi

La presenza di individui affetti da malattie cronico-degenerative può rappresentare per la famiglia uno stato critico analogo a quello di famiglie con componenti invalidi.

Nel 20,9% delle famiglie intervistate è presente almeno un malato; in particolare, il 17,8% è costituito da un malato e il 3,1% da due o più. Il 20,9% delle famiglie in cui sono presenti malati è unipersonale, ossia si tratta di malati che vivono soli (4,4% delle famiglie totali), il 9,3% è composto da soli malati (1,9% delle famiglie totali) e il restante 69,8% è rappresentato da famiglie in cui è presente almeno un malato oltre a componenti in buona salute (Tab.2.83).

Tab.2.83. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per composizione delle famiglie con membri affetti da malattie cronico-degenerative*

Composizione della famiglia (n=563)	% (a)	% (b)
Famiglie unipersonali di malati	4,4	20,9
Famiglie pluripersonali con solo malati	1,9	9,3
Famiglie con un malato	13,4	64,1
Famiglie con più malati	1,2	5,7
Totale	20,9	100,0

% (a) = percentuale sul totale delle famiglie

% (b) = percentuale sul totale delle famiglie con membri malati

Le malattie più diffuse sono quelle del cuore e del sistema circolatorio (infarto, ipertensione arteriosa, angina pectoris, etc.), che si presenta nel 44,3% dei casi (Tab.2.84).

Tab.2.84. *Distribuzione percentuale della popolazione (malata) residente nel Veneto nel 2004-2005 per tipo di malattia*

Tipo di malattia (n=562)	% (a)
Metabolismo	21,1
Sistema cardi-circolatorio	44,3
Vie respiratorie	14,9
Apparato digerente	6,9
Ossa ed articolazioni	11,6
Artrite, artrosi	23,6
Sistema immunitario	0,7
Neoplasie	7,5
Sistema nervoso	3,9

% (a) = percentuale sul totale dei membri malati in ciascuna tipologia di malattia

Il 20,7% delle famiglie con presenza di componenti affetti da malattie cronicodegenerative è composto da un solo membro, il 33,2% da due.

Le malattie cronicodegenerative colpiscono in misura leggermente superiore le donne (i tassi di morbosità sono pari a 9,5% per le donne e a 9,1% per gli uomini) e l'andamento non presenta notevoli differenze tra i due sessi. Le persone più colpite hanno più di 70 anni. Le persone giovani, al di sotto dei 50 anni, presentano più di rado malattie cronicodegenerative. Dopo questa età la percentuale di malati aumenta progressivamente; questa tendenza è legata al fatto che le malattie più frequenti, ossia quelle del cuore e del sistema circolatorio, si manifestano prevalentemente in età avanzata (Tab.2.85).

Tab.2.85. *Tassi percentuali di morbosità della popolazione residente nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Popolazione tipo (n=7106)
	Femmine (n=3662)	Maschi (n=3444)	
0-9	0,7	0,3	0,085
10-19	0,6	1,1	0,104
20-29	1,9	1,7	0,111
30-39	3,6	3,1	0,138
40-49	4,0	4,8	0,173
50-59	9,4	7,5	0,152
60-69	19,6	19,1	0,121
70-79	35,0	33,2	0,082
80-89	41,5	37,9	0,029
90 e più	29,6	38,5	0,005
Totale	9,5	9,1	1,0

2.10.3 Presenza di disoccupati

L'1,6% della popolazione con età superiore ai 15 anni è disoccupato; tale percentuale comprende sia gli individui in cerca di nuova occupazione (1,3%), sia quelli in cerca di prima occupazione (0,3%) (Tab.2.6).

In tabella 2.86 si riportano i tassi di disoccupazione per età e sesso, calcolati come rapporto tra le persone disoccupate e le forze lavoro. Il tasso di disoccupazione standardizzato per età è del 4,4% per le donne e del 3,0% per gli uomini. Per le donne il tasso più elevato riguarda la fascia d'età tra i 15 e i 24 anni (20,2%), mentre per gli uomini quella tra i 25 e i 34 anni (7,6%), sebbene poco inferiore risulti il tasso per i maschi in età 15-24 (6,1%). Si tratta, verosimilmente, di persone diplomate o laureate in cerca di prima occupazione o di un lavoro fisso.

Tab.2.86. *Tassi percentuali di disoccupazione della popolazione residente nel Veneto nel 2004-2005 per classi di età e sesso*

Classi di età	Sesso		Popolazione tipo (n=3716)
	Femmine (n=1921)	Maschi (n=1795)	
15-24	20,2	6,1	0,061
25-34	8,8	7,6	0,203
35-44	2,9	1,1	0,320
45-54	1,3	1,3	0,285
55-64	0,3	3,1	0,131
Totale	4,4	3,0	1,0

Studiando il fenomeno della disoccupazione in relazione al grado di istruzione, si nota che nella fascia d'età 15-34 il tasso di disoccupazione ammonta a 8,8%, mentre nella fascia 35-64 si riduce all'1,6%. Per i più giovani il tasso è più elevato, e precisamente pari a 10,0%, tra coloro che sono in possesso del diploma di scuola media superiore (4-5 anni) o di un titolo maggiore (verosimilmente persone perlopiù in cerca di prima occupazione) (Tab.2.87).

Tab.2.87. *Tassi percentuali di disoccupazione della popolazione residente nel Veneto nel 2004-2005 per titolo di studio e classi di età*

Titolo di studio	Classi di età		
	15-34 (n=975)	35-64 (n=2741)	Popolazione tipo (n=3755)
Diploma di scuola media superiore (4-5 anni) o titolo maggiore	10,0	0,9	0,470
Diploma di scuola media inferiore (2-3 anni) o titolo minore	7,9	2,2	0,530
Totale	8,8	1,6	1,0

I tassi di disoccupazione calcolati sul campione sono tuttavia inferiori a quelli nazionali, ma ciò è spiegato soprattutto dal fatto che nella nostra regione il tasso di disoccupazione non conosce livelli elevati come in certe zone del mezzogiorno e del meridione.

Il 30,9% dei disoccupati si trova in questa situazione da un periodo di tempo che va dai due ai sei mesi, solo il 14,5% da oltre due anni (Tab.2.88).

Tab.2.88. *Distribuzione percentuale della popolazione (disoccupata) residente nel Veneto nel 2004-2005 per periodo di disoccupazione*

Periodo di disoccupazione (n=92)	
Meno di un mese	12,4
Da due a sei mesi	30,9
Da 7 a 12 mesi	18,2
Da 1 a 2 anni	18,6
Oltre i 2 anni	14,5
Non cerca lavoro	5,4
Totale	100,0

L'82,9% dei disoccupati cerca attivamente lavoro, andando di persona a candidarsi come possibile lavoratore (39,1%), attraverso le agenzie interinali (29,0%) o facendo uso di altri possibili mezzi (Tab.2.89).

Tab.2.89. *Distribuzione percentuale della popolazione (disoccupata) residente nel Veneto nel 2004-2005 per azioni di ricerca di lavoro svolte*

Azioni di ricerca di lavoro (n=92)	
Agenzie interinali	29,0
Inserzioni sui giornali	12,9
Web	4,3
Diffusione di curricula	14,7
Andando di persona	39,1
Totale	100,0

2.11 Eventi critici

Per “evento critico” si intende una grave situazione che può turbare e rivoluzionare la vita della famiglia. Si tratta di episodi che rompono l’equilibrio familiare e l’abituale organizzazione dei componenti e che costringono la famiglia a dar vita a nuovi meccanismi che contribuiscano a riportare una certo equilibrio al suo interno.

Il primo evento critico preso in considerazione è il lutto di una persona vicina, parente o amica, non necessariamente abitante nella stessa dimora della famiglia. Si sono presi in considerazione i lutti avvenuti negli ultimi tre anni e, nel caso di un numero di decessi superiore ad uno, si è fatto riferimento a quello più recente.

Il 30,5% delle famiglie venete ha subito almeno un lutto negli ultimi tre anni. In particolare, il 25,1% ne ha subito uno e il 5,4% due o più. Con riferimento al capofamiglia, il decesso riguarda per il 44,8% dei casi il genitore/suocero e per il 15,2% il fratello/sorella (Tab.2.90).

Tab.2.90. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per persone care decedute negli ultimi tre anni lavoro svolte*

Deceduto (n=767)	
Coniuge/convivente	8,7
Figlio/a	1,2
Genitore/suocero	44,8
Fratello/sorella	15,2
Cognato/a	6,5
Altro	23,6
Totale	100,0

Nell’83% delle famiglie colpite da un lutto, il decesso riguarda una persona che

non viveva all'interno della famiglia e, pertanto, generalmente non va ad influenzare in maniera pesante l'equilibrio al suo interno.

Un altro evento critico è lo sfratto che riguarda le sole famiglie che vivono in affitto: lo 0,4% delle famiglie totali (il 3,1% di quelle che vivono in affitto) è sotto sfratto. Di quest'ultime l'84,4% è attualmente in cerca di una nuova sistemazione mentre solo il 12,2% l'ha già trovata e sta per trasferirsi.

Alle famiglie è stato poi chiesto se, negli ultimi tre anni, qualcuno fosse entrato abusivamente o con l'inganno nella loro abitazione: il 2,5% ha subito un furto nell'ultimo anno, il 4,0% da più di un anno, ma da meno di tre (Tab.2.91).

Tab.2.91. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per furti subiti negli ultimi tre anni*

Furto (n=2465)	
Nell'ultimo anno	2,5
Da uno a tre anni fa	4,0
No o più di tre anni fa	93,5
Totale	100,0

Il 66,3% delle famiglie che hanno subito un furto nell'ultimo anno ha presentato denuncia alla Polizia, ai Carabinieri o alle autorità competenti in generale. La percentuale di denunce è abbastanza elevata ed è abbastanza in linea con l'Indagine Multiscopo sulle famiglie 1989-1990 secondo cui il 58% delle famiglie denunciava i furti avvenuti nella propria abitazione.

La famiglia può trovarsi in difficoltà anche quando si verificano incidenti di una certa importanza: il 7,1% delle famiglie ha subito un incidente nell'ultimo anno, il 5,9% da più di un anno, ma da meno di tre (Tab.2.92).

Tab.2.92. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per incidenti subiti negli ultimi tre anni*

Incidenti (n=2458)	
Nell'ultimo anno	7,1
Da uno a tre anni fa	5,9
No o più di tre anni fa	87,0
Totale	100,0

Nel primo caso, per il 39,7% delle persone (l'1,3% della popolazione totale) si è reso necessario un ricovero in ospedale, mentre il 73,5% (il 2,3% della popolazione totale) è stato costretto a letto per almeno due giorni

2.12 Reti familiari

La famiglia è estesa, nel senso che è inserita in una rete di rapporti con la parentela, il vicinato, il gruppo amicale, e lo scambio dei beni e servizi non presuppone né la coabitazione né la stretta vicinanza geografica (Litwak, 1960).

Di estrema importanza per la programmazione di interventi e politiche a favore della famiglia è l'analisi e la comprensione delle reti di supporto di cui la famiglia dispone in caso di bisogno. Per *reti di supporto* si intendono quelle persone, soprattutto parenti, alle quali la famiglia si rivolge in caso di necessità per ottenere aiuto e sostegno. Le reti di supporto sono una delle risorse più considerevoli di cui la famiglia dispone e sono spesso garanti di equilibrio e stabilità.

Di particolare rilievo è il modo in cui concretamente entro una rete parentale vengono percepite e allocate le responsabilità nei confronti di chi, per diversi motivi, ha bisogno di aiuto: la famiglia si rivolge alle reti parentali non solo per problemi di grossa entità (malattie, incidenti, problemi economici) ma anche per le piccole necessità quotidiane come l'accudimento momentaneo di un minore o di un anziano, l'espletamento di pratiche burocratiche, gli spostamenti e molte altre necessità che altrimenti graverebbero per intero sulla famiglia. Da non sottovalutare, inoltre, è l'appoggio morale che una famiglia trova all'interno della propria rete di supporto in una dimensione affettiva completamente disinteressata e gratuita.

Generalmente, quando sorgono problemi, la famiglia in prima battuta ricerca al proprio interno, tra le risorse che ben conosce, l'aiuto di cui necessita; solo in un secondo momento, si rivolge ai servizi. Le reti di supporto sono dunque una sorta di *filtro*: esse "fanno passare" i problemi che sono in grado di risolvere e "trattengono" quelli più gravi che vengono successivamente affidati alla cura dei servizi.

Esping-Andersen (1990; 1995), in un confronto su diversi modelli di welfare, affianca lo stato sociale italiano ad altri quali quelli tedesco, francese e belga, essendo la famiglia la principale fornitrice di servizi e di risorse. Tuttavia, esiste tra questi welfare una profonda ed esplicita differenza: è vero che in alcuni paesi ci si attendono dalla famiglia servizi e sostegni per i suoi componenti, ma essa viene anche sostenuta e protetta dallo stato attraverso generose disponibilità economiche.

L'assenza di reti di sostegno può generare due tipi di reazione: da una parte si possono ritrovare famiglie che cercano autonomamente di far fronte alle proprie necessità imponendo a se stesse uno sforzo inadeguato per le risorse di cui dispongono, dall'altra famiglie che fanno ricorso ai servizi in modo improprio ed eccessivo sovraccaricando la domanda di assistenza.

I dati disponibili oggi in Italia, a partire da quelli dell'Indagine Multiscopo dell'Istat, hanno iniziato a delineare un quadro dei rapporti parentali presenti nel nostro paese, sebbene risulti difficile andare a conoscere i protagonisti e l'oggetto di questi rapporti.

Nel nostro paese gli obblighi di parentela sono in larga misura prescritti per legge, soprattutto per quanto concerne il bisogno economico, ma molti studiosi si chiedono se l'esistenza di una rete parentale funzionante è conseguenza di un modello culturale condiviso o della scarsità nell'offerta dei servizi. Inoltre, tali servizi a sostegno delle responsabilità familiari - servizi per l'infanzia, per i disabili ad esempio - sono sempre stati visti in Italia con sospetto per quanto concerne la loro opportunità: si temeva che gli asili nidi de-responsabilizzassero i genitori e che i servizi per i disabili minassero la solidarietà familiare. Tutto ciò ha fatto sì che in Italia sia più facile ricorrere al supporto parentale per far fronte ai problemi interni alla famiglia.

Il 25,7% delle famiglie venete ha dichiarato di non sapere a chi rivolgersi in caso di bisogno. È una percentuale piuttosto alta, nonostante che, in caso di reale necessità, qualcuna di queste famiglie sarebbe comunque in grado di chiedere aiuto a qualche conoscente o parente al quale nel corso dell'intervista non ha pensato. Nelle situazioni più gravi si è sostanzialmente in grado di attivare una serie di risorse che, in condizioni normali, la famiglia non è consapevole di avere.

In particolare, facendo una distinzione per tipologia di famiglia, si ottengono i risultati riportati in tabella 2.93.

Tab.2.93. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per possibilità di rivolgersi a qualcuno in caso di bisogno e tipo di famiglia*

Tipo di famiglia	
Almeno un invalido (n=208)	70,0
Almeno due malati (n=25)	71,7
Un genitore con figli (n=167)	71,2
Famiglie di soli anziani (n=456)	84,0

Il 30,0% delle famiglie con almeno un invalido al loro interno ha dichiarato di non sapere a chi rivolgersi in caso di bisogno, contro il 16,0% di quelle composte da soli anziani.

Alle famiglie è stato chiesto a quale parente si rivolgerebbero in caso di bisogno e quanto tempo questo impiegherebbe a raggiungere l'abitazione.

Le famiglie si appoggiano più frequentemente alla rete primaria di aiuti, rappresentata dalla famiglia d'origine oppure dai figli o da altri componenti che hanno lasciato l'abitazione per costituire nuclei autonomi, persone che possono, dunque, fornire anche un supporto morale. Il rapporto genitori-figli costituisce l'asse portante attorno a cui si costruiscono gli altri rapporti con i parenti.

Il 37,9% delle famiglie venete sa di poter contare sui figli o sui coniugi dei figli, il 38,0% sui genitori o suoceri, il 54,1% sui fratelli o sorelle, il 9,7% sui nipoti e il 3,5% su altri parenti.

Per quanto riguarda la tempestività dell'aiuto, all'aumentare della distanza tra le abitazioni delle famiglie che si scambiano l'aiuto, diminuisce tale scambio: l'87,4% ha risposto che entro 15 minuti la persona che fornisce l'aiuto riesce a

raggiungere l'abitazione della famiglia, il 9,9% entro un'ora, il 2,5% dopo più di un'ora.

In particolare il 32,1% delle famiglie può far affidamento sui figli che vivono entro 15 minuti dalla propria abitazione, il 44,3% sui fratelli e il 33,8% sui genitori o suoceri (Tab.2.94).

Tab.2.94. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per aiuto e tempo impiegato per raggiungere l'abitazione*

Aiuto per la famiglia	Tempo per raggiungere l'abitazione					Totale
	No	Entro 15 minuti	Entro 1 ora	Oltre 1 ora	Oltre 1 giorno	
Figli / figlie o nuore / generi (n=1883)	62,1	32,1	4,5	1,3	0,0	100,0
Fratelli / sorelle o cognati / e (n=1864)	45,9	44,3	7,0	2,6	0,2	100,0
Nipoti con almeno 15 anni (n=1859)	90,3	8,6	0,6	0,5	0,0	100,0
Genitori / suoceri (n=1859)	62,0	33,8	3,2	0,8	0,2	100,0
Altri parenti (n=1856)	96,5	3,3	0,1	0,1	0,0	100,0
Totale	0,0	87,4	9,9	2,5	0,2	100,0

L'intensità dei rapporti tra parenti, e soprattutto tra genitori e figli, è una caratteristica tipica del Veneto e di tutta l'Italia, che contrasta con il comportamento della maggior parte degli altri paesi europei (Barbagli, 1990).

Sono soprattutto le giovani coppie ad avere rapporti con i genitori, mentre le persone non sposate intensificano tali rapporti dopo i 50 anni: il fatto di non essere sposati e soprattutto di non avere figli porta le persone a prestare cura e sostegno ai genitori anziani, mentre chi ha una propria discendenza fa generalmente affidamento ai genitori per l'accudimento dei figli. Ciò che si sottolinea è, così, un'asimmetria nei rapporti tra genitori e figli, basata sul rendere ai propri figli il sostegno ricevuto dai genitori (Cioni, 1997).

Nonostante la varietà delle motivazioni in cui si presta e si riceve aiuto entro la rete parentale, si possono individuare tre tipi prevalenti di interventi (Sabbadini, 1994):

- sostegni economici tra diversi nuclei familiari, in particolare tra genitori e figli;

- assistenza agli anziani e ai disabili;
- accudimento nei confronti dei nuclei con figli piccoli, soprattutto qualora la madre abbia un lavoro extra-domestico.

Oggi, sul piano economico, il 6,0% delle famiglie intervistate ha dichiarato di aver avuto necessità di aiuto nell'ultimo anno. In particolare, hanno avuto bisogno di aiuto economico l'11,6% delle famiglie in affitto, il 10,4% di quelle con reddito inferiore ai 1500 euro mensili e il 7,9% di quelle in cui vi è al più un componente che percepisce uno stipendio (o pensione) (Tab.2.95).

Tab.2.95. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per necessità di aiuto economico e caratteristiche familiari*

Caratteristiche familiari	
Reddito \leq 1500 € (n=63)	10,4
Al più uno stipendio (n=65)	7,9
Casa in affitto (n=30)	11,6

Il 35,0% delle famiglie con necessità economiche si è rivolta ad istituti bancari, il 25,9% ai genitori di uno dei coniugi, il 12,2% ai figli di uno dei due coniugi e il 9,0% ai servizi sociali (Tab.2.96).

Tab.2.96. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per persone a cui si rivolge per ottenere un aiuto economico*

Persona (n=146)	
Genitori/suoceri	25,9
Figli/coniugi dei figli	12,2
Fratelli/sorelle	3,5
Altri parenti	1,6
Altri non parenti	4,9
Servizi sociali dell'ULSS	1,1
Servizi sociali del Comune	9,0
Regione del Veneto	3,3
Banche	35,0
Altro	3,5
Totale	100,0

Gli scambi di carattere economico all'interno della rete parentale avvengono prevalentemente nella direzione genitori-figli e riguardano soprattutto gli uomini:

il 22% dei padri ha fornito aiuto economico ai figli contro il 10% delle madri, mentre solo il 6% dei figli e il 3% delle figlie ha dato aiuto ai genitori. Ad esempio, un'importante aiuto economico è quello che i genitori offrono alle giovani coppie di figli che decidono di sposarsi o di convivere per l'acquisto della casa e dei beni durevoli necessari (Cioni, 1997).

I problemi di tipo economico sono per una famiglia argomento estremamente delicato. Spesso ammettere di essere in condizioni poco agiate genera imbarazzo e la ricerca di aiuto avviene al di fuori della rete parentale ed amicale, in modo anonimo e riservato. I dati rilevati possono rappresentare, pertanto, una sottostima.

Dall'analisi emerge che le famiglie che lamentano problemi socio-relazionali sono perlopiù quelle formatesi da uno a tre anni fa (18,3%), mentre tra le altre famiglie la distribuzione è abbastanza omogenea.

Altri tipi di problemi sono stati presi in considerazione nell'indagine: quelli economici colpiscono prevalentemente le famiglie più giovani (16,1% di quelle formatesi meno di un anno fa e 16,0% di quelle formatesi da uno a tre anni fa). I problemi di salute riguardano perlopiù le famiglie formatesi oltre i 25 anni fa (19,2%), costituite soprattutto da persone anziane. Infine, poco rilevanti risultano i problemi socio-relazionali. (Tab.2.97).

Tab.2.97. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per tipo di problema ed età della famiglia*

Tipo di problema	Età della famiglia						Totale (n=2457)
	Meno di un anno (n=31)	Da 1 a 3 anni (n=106)	Da 3 a 7 anni (n=229)	Da 7 a 15 anni (n=446)	Da 15 a 25 anni (n=499)	Oltre 25 anni (n=1146)	
Economici	16,1	16,0	12,2	12,6	13,0	11,5	12,3
Di salute	16,1	17,0	7,4	8,5	8,0	19,2	13,8
Socio-relazionali	6,4	8,4	3,9	5,6	5,6	3,2	4,5
Altro	3,2	2,8	0,4	2,0	1,0	0,4	1,0

Insieme a quelli ricevuti dall'esterno, gli aiuti che la famiglia presta all'esterno concorrono a delineare il quadro delle reti di relazioni interfamiliari. Il 30,6% delle famiglie ha dichiarato di aver prestato nell'ultimo anno assistenza/aiuto a

titolo gratuito a persone non coabitanti. Il 61,0% delle famiglie offre compagnia, assistenza in casa e il 32,1% aiuto in attività domestiche (Tab.2.98). Il tipo di aiuto che viene offerto più frequentemente è, pertanto, la compagnia a persone sole, che, a livello nazionale, riguarda il 27,7% delle famiglie (ISTAT, 1998).

Tab.2.98. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per aiuti prestati*

Tipo di aiuto	%(a)
Prestazioni sanitarie	16,2
Compagnia, assistenza	61,0
Aiuto in attività domestiche	32,1
Accompagnamento fuori casa	16,1
Espletamento di pratiche burocratiche	7,4
Aiuto economico	21,4
Altro	3,0

%(a) = percentuale sul totale delle famiglie che prestano aiuto, in ciascun tipo di aiuto

Principali destinatari di questo tipo di aiuto sono i genitori o suoceri del capofamiglia o comunque persone anziane (35,4%) (Tab.2.99).

Tab.2.99. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per destinatari degli aiuti prestati*

Destinatari	
Genitori/suoceri (n=189)	35,4
Figli (n=31)	5,9
Fratelli/cognati (n=31)	5,9
Nipoti (n=22)	4,3
Nonni (n=10)	2,2
Altri parenti (n=64)	10,7
Altri non parenti (n=189)	35,6

2.13 Servizi

In Italia manca una vera politica sociale per la famiglia. La famiglia è sempre stata la risorsa principale a cui un individuo può ricorrere in caso di bisogno e questa realtà ha determinato una scarsa diffusione dei servizi alle persone - servizi

socio-sanitari, sociali - ritenuti necessari laddove manchi una famiglia e una rete parentale adeguate. Ad esempio, in molti comuni l'assistenza a persone anziane non autonome è fornita solamente nei casi in cui l'anziano è povero o non ha familiari che si possono occupare di lui (Saraceno, 1997).

La legge 328/2000 ha rappresentato un momento importante per la modalità di programmazione e di finanziamento degli interventi e dei servizi sociali. Il tentativo è stato quello di dotare gli ambiti territoriali di "Piani di zona dei Servizi Sociali", costituiti da progetti innovativi per il soddisfacimento dei bisogni socio-assistenziali e socio-sanitari. La particolarità di tali progetti è la compartecipazione di tutti i soggetti attivi (Comuni, Aziende ULSS, Associazioni e gruppi di volontariato) nell'erogazione diretta ed indiretta di interventi e servizi sociali⁵. La legge ha individuato i seguenti obiettivi:

- rafforzare e sostenere le responsabilità familiari;
- rafforzare i diritti dei minori;
- potenziare gli interventi di contrasto alla povertà;
- sostenere con servizi domiciliari il diritto delle persone non autosufficienti a rimanere nel proprio ambiente.

L'analisi che segue si propone di valutare se nel quartiere di residenza della famiglia esistono attività o servizi a favore della stessa. Anche in questo caso

⁵Un confronto interessante è quello con il Friuli Venezia Giulia, il cui progetto di legge è l'integrazione delle politiche socio-assistenziali (Sole 24 ore, 13 luglio 2005). Il testo, intitolato "Sistema integrato di interventi e servizi per la promozione e la tutela dei diritti di cittadinanza sociale", introduce una politica che rafforza il ruolo dei distretti sanitari e dei Comuni e rende partecipe il mondo del terzo settore. Tale proposta prevede l'introduzione di istituti in grado di rispondere alle emergenze sociali e la creazione di una rete di servizi diretta all'intero sistema regione: alle famiglie, agli anziani, ai minori, ai disabili e ai cittadini immigrati. Il progetto mira all'integrazione delle politiche socio-assistenziali, sanitarie, abitative, dei trasporti, dell'educazione, formative, del lavoro, culturali, ambientali ed urbanistiche, dello sport e del tempo libero e di altri interventi finalizzati al benessere della persona e alla prevenzione delle condizioni di disagio sociale.

si presenta il problema delle risposte *proxy*: se si immagina che per una famiglia risponda il figlio maggiorenne, è probabile che questi non sia a conoscenza di servizi per la famiglia anche se magari la stessa madre ne ha fatto uso. Complessivamente, il 43,8% delle famiglie ritiene che nel proprio quartiere esistano servizi a favore della famiglia, il 9,5% che non ne esistano e il 46,7% non sa se esistano. La conoscenza della presenza di servizi non differisce in maniera evidente tra i rispondenti, sebbene i più informati risultino i capofamiglia o i coniugi (44,2%) (Tab.2.100).

Tab.2.100. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per presenza di servizi di sostegno nel quartiere e rispondente al questionario*

Servizi	Rispondente al questionario			Totale (n=2462)
	Capofamiglia o coniuge (n=2288)	Figlio (n=140)	Altro parente (n=34)	
Sì	44,2	40,3	32,1	43,8
No	9,3	10,7	15,1	9,5
Non so	46,5	49,0	52,8	46,7
Totale	100,0	100,0	100,0	100,0

Le differenze tra provincia e provincia sono notevoli: sembra che questo tipo di servizi scarseggi nelle zone del Bellunese e del Rodigino mentre la provincia di Vicenza sembra esserne la più fornita. Rimane in ogni caso molto alta la percentuale delle famiglie che hanno risposto di non sapere se questi servizi esistano o meno. Questo aspetto, che a prima vista può sembrare importante, in realtà generalmente non lo è, dal momento che, in caso di bisogno, una famiglia si adopera per avere informazioni sui servizi che possono venire loro offerti (Tab.2.101).

Tab.2.101. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per presenza di servizi di sostegno nel quartiere e provincia*

Servizi	Provincia							Totale (n=2464)
	Belluno (n=323)	Padova (n=368)	Rovigo (n=363)	Treviso (n=370)	Venezia (n=350)	Verona (n=349)	Vicenza (n=341)	
Sì	35,3	39,1	36,9	45,4	50,0	51,0	39,3	42,4
No	14,2	9,0	15,2	9,2	7,7	6,8	11,4	10,5
Non so	50,5	51,9	51,9	45,4	42,3	42,2	49,3	47,1
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

La tabella 2.102 aiuta a quantificare il ricorso ai servizi da parte delle famiglie. Quelli più utilizzati sono senza dubbio i servizi socio-sanitari dell'ULSS: il 57,7%

delle famiglie intervistate vi accede in modo regolare o saltuario; seguono con netto scarto i consultori familiari (5,5%) e i servizi sociali del Comune (4,5%).

Tab.2.102. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per ricorso a diversi tipi di servizi*

Ricorso al servizio	Tipo di servizio			
	Consultorio familiare (n=2442)	Servizi socio-sanitari dell'ULSS (n=2463)	Servizi sociali del Comune (n=2460)	Associazioni di volontariato (n=2455)
Regolarmente	1,4	21,9	2,4	0,7
Occasionalmente	4,4	35,8	2,1	0,3
Mai	94,2	42,3	95,5	99,0
Totale	100,0	100,0	100,0	100,0

Analizzando la frequenza dell'utilizzo di alcuni servizi da parte delle famiglie in relazione al livello di disagio dichiarato dalle stesse, si nota che fanno regolarmente ricorso al consultorio familiare il 6,2% di quelle che hanno gravi problemi e l'1,0% di quelle che non disagate (Tab.2.103).

Tab.2.103. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per ricorso al consultorio familiare e livello di disagio*

Ricorso al servizio	Livello di disagio				Totale (n=2439)
	Problemi gravi (n=1773)	Problemi risolvibili con aiuto esterno (n=478)	Problemi risolvibili con le forze della famiglia (n=140)	Nessun vero problema (n=48)	
Regolarmente	1,2	1,3	2,9	7,3	1,4
Occasionalmente	4,8	3,4	2,1	5,6	4,4
Mai	94,0	95,3	95,0	87,1	94,2
Totale	100,0	100,0	100,0	100,0	100

Il ricorso ai servizi socio-sanitari dell'ULSS è abbastanza frequente. Rispetto al consultorio familiare, ancora più marcata è la differenza tra le famiglie senza e con gravi problemi rispettivamente. Usufruiscono di questo servizio il 15,9% delle prime, contro il 44,9% delle seconde. La percentuale di famiglie che fa ricorso a questo servizio occasionalmente è, d'altro canto, più elevata tra quelle che non hanno problemi (37,2%) rispetto a quelle che avvertono un forte disagio (24,9%) (Tab.2.104).

Tab.2.104. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per ricorso ai servizi socio-sanitari dell'ULSS e livello di disagio*

Ricorso al servizio	Livello di disagio				Totale (n=2439)
	Problemi gravi (n=1792)	Problemi risolvibili con aiuto esterno (n=480)	Problemi risolvibili con le forze della famiglia (n=141)	Nessun vero problema (n=47)	
Regolarmente	15,9	39,2	38,8	44,9	21,9
Occasionalmente	37,2	32,1	33,1	24,9	35,8
Mai	46,9	28,7	28,1	30,2	42,3
Totale	100,0	100,0	100,0	100,0	100,0

I risultati relativi al ricorso ai servizi sociali del Comune sono riportati nella tabella 2.105, da cui emerge in modo ancor più netto la differenza rispetto all'utilizzo del servizio tra le famiglie che non percepiscono disagio e quelle che avvertono gravi problemi

Tab.2.105. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per ricorso ai servizi sociali del comune e livello di disagio*

Ricorso al servizio	Livello di disagio				Totale (n=2439)
	Problemi gravi (n=1789)	Problemi risolvibili con aiuto esterno (n=480)	Problemi risolvibili con le forze della famiglia (n=141)	Nessun vero problema (n=47)	
Regolarmente	0,7	3,6	14,3	20,0	2,4
Occasionalmente	1,6	4,3	2,4	3,2	2,1
Mai	97,7	92,1	83,3	76,8	95,5
Totale	100,0	100,0	100,0	100,0	100,0

La distribuzione delle famiglie rispetto all'aiuto chiesto alle associazioni di volontariato e al disagio non è stata presa in considerazione a causa delle basse numerosità.

Indipendentemente dal tipo di servizio, sarebbero auspicabili campagne informative che pubblicizzassero maggiormente i servizi e una sensibilizzazione rivolta soprattutto alle famiglie risultate più problematiche, affinché possano percepire il sostegno delle istituzioni e alleggerire il loro carico di problemi.

Capitolo 3

La ricerca dei gruppi a rischio

3.1 L'analisi di segmentazione

L'analisi di segmentazione è una tecnica di analisi multivariata che permette di evidenziare le relazioni esistenti tra una variabile criterio Y e più variabili esplicative X . Per esplorare tali relazioni, il processo di segmentazione suddivide il campione di dati iniziale in gruppi di unità, sulla base delle variabili indipendenti che danno il contributo maggiore all'interpretazione dei valori di una variabile dipendente.

Terminata l'analisi di segmentazione del campione di dati, il ricercatore potrà individuare i fattori che influenzano maggiormente il fenomeno d'interesse, isolandoli da una moltitudine di variabili osservate e riuscendo ad individuare se, fra di essi, vi siano effetti di interazione importanti.

La tecnica di segmentazione mira, infatti, a risolvere un problema con il quale si misurano gli analisti dei dati: l'*interazione* tra variabili, ossia l'effetto che le modalità combinate di variabili esplicative hanno sulla variabile dipendente.

La suddivisione dei dati, detta partizione, viene fatta in modo da ottimizzare una funzione criterio della variabile dipendente, al fine di individuare sottogruppi massimamente omogenei al loro interno e disomogenei tra loro.

Al primo passo il campione viene suddiviso in due o più sottoinsiemi, caratterizzati dai valori assunti da una o più variabili esplicative, ciascuno dei quali viene poi ulteriormente suddiviso finché il processo non è interrotto da una regola di

arresto.

Perché l'analisi abbia significato e il frazionamento delle unità nei sottogruppi porti a stime di associazione con la variabile dipendente non casuali, è bene disporre di campioni di numerosità piuttosto elevata.

La segmentazione viene denominata a seconda del numero di sottoinsiemi che si vengono a formare ad ogni suddivisione: se si considerano partizioni a due vie la segmentazione è detta "binaria" (le caratteristiche dei predittori sono riunite in due sottoinsiemi creando, su un generico nodo genitore, una partizione dicotomica), se, invece, la suddivisione è a tre vie si denomina "ternaria", a più vie "multipla".

Una rappresentazione della struttura di dipendenza tra le variabili è l'*albero di segmentazione* o *dendogramma*, dove i *nodi* sono i gruppi che si vengono a formare dal processo di segmentazione e i *rami* le condizioni che hanno determinato la suddivisione. I nodi terminali, per i quali non è ritenuta necessaria un'ulteriore suddivisione, vengono denominati *foglie* (Breiman, 1984).

Sulla struttura ad albero si possono individuare i *livelli*, ciascuno dei quali identifica i nodi creati dopo un certo numero di segmentazioni a partire dal gruppo di partenza.

In ambito statistico, la segmentazione è utile per diversi impieghi:

- ricerca di predittori, ossia determinazione di possibili effetti che una o più combinazioni di modalità dei predittori possono avere sulla variabile dipendente;
- identificazione di gruppi devianti, ossia valori che si collocano sulle code delle distribuzioni;
- generazione di modelli di interpretazione causale del fenomeno rappresentato dalla variabile dipendente;
- sintesi delle informazioni contenute nei dati, cioè individuazione delle variabili indipendenti più esplicative del fenomeno all'interno di ogni sottogruppo;
- ricerca di relazioni non lineari tra variabili quantitative, o non monotone tra variabili ordinali, nel caso di segmentazione multipla;

- sostituzione di un dato mancante con uno appartenente alla medesima classe;
- creazione di regole di previsione o classificazione allo scopo di determinare il valore che assumerà la variabile dipendente di una nuova unità statistica, note le modalità assunte dalle variabili esplicative.

3.1.1 La variabile dipendente

Le procedure di segmentazione si diversificano a seconda delle variabili dipendenti coinvolte nel processo.

Per variabili criterio su scala qualitativa, la segmentazione tratta il caso univariato analizzando una variabile alla volta. Per ridurre i problemi di instabilità, è bene che la variabile qualitativa in esame abbia un numero relativamente basso di modalità così da evitare probabilità prossime a zero per alcune modalità.

Per variabili dipendenti quantitative, invece, l'analisi è stata sviluppata anche per il caso multivariato, considerando un vettore di variabili criterio contemporaneamente. Anche in questo caso, è bene studiare la distribuzione delle variabili per escludere quelle che presentano una forte asimmetria, dato che quest'ultima potrebbe portare ad un isolamento dei gruppi di piccola dimensione sulla coda della distribuzione.

3.1.2 Le variabili predittive

Le conoscenze a priori sul fenomeno in esame, derivanti da ricerche analoghe o da preliminari analisi statistiche, sono di aiuto per selezionare le variabili predittive che, plausibilmente, meglio spiegano la distribuzione della variabile criterio.

I predittori vengono generalmente suddivisi, a seconda del tipo di categorie di cui sono formati, in:

- *nominale* (o *liberi*), se le categorie non sono ordinate né ordinabili e, quindi, possono combinarsi in tutti i modi possibili per formare una partizione;
- *ordinabili* (o *monotoni*), se le categorie sono ordinate o ordinabili cosicché da potersi combinare solo in modo adiacente per dar luogo ad una partizione;

- *floating*, se le categorie sono ordinate o ordinabili ad eccezione di una, che solitamente rappresenta i dati mancanti o risposte del tipo “non so”, “altro”. Le partizioni che si vengono a formare avranno categorie ordinate secondo il loro ordine, con la possibilità, per la modalità fluttuante, di combinarsi con qualsiasi altra categoria.

Con una partizione dicotomica su un nodo, le M modalità di un predittore possono essere combinate nei due gruppi in $2^{M-1} - 1$ modi. Quando le modalità sono ordinabili, dovrà essere rispettato l'ordine e le possibili combinazioni risultano $M - 1$.

Poiché risulterebbe computazionalmente dispendioso cercare la miglior suddivisione se per ogni variabile con M modalità si dovesse determinare l'effetto di tutte le $2^{M-1} - 1$ possibili partizioni, Fisher (1958) ha dimostrato che è sufficiente esaminare solo $M - 1$ partizioni sia per variabili ordinabili che nominali. Per quest'ultime, si ordinano le M categorie sulla base del valor medio della variabile dipendente condizionata ad ognuna di esse, verificando successivamente quale, tra le $M - 1$ aggregazioni contigue è la più conveniente, rispetto al criterio scelto per valutare l'efficacia della segmentazione.

Al fine di combinare le categorie in una partizione, è necessario che tutte le variabili esplicative siano trasformate in variabili categoriali. Generalmente non si devono superare le tre-quattro categorie per predittore. Questa scelta la si fa per ridurre i tempi di calcolo e per evitare di favorire l'ingresso di predittori con molte categorie, che darebbero luogo ad un maggior numero di possibili partizioni. Inoltre, nei predittori con distribuzione asimmetrica è necessario aggregare le classi con frequenze basse in modo da ottenere una variabile con frequenze quasi uniformi.

3.1.3 La funzione criterio

Dato un insieme S di possibili partizioni e un generico nodo t , per ogni partizione $s \in S$, si indichi con $\Phi(s, t)$ la *funzione criterio*, che permetta di misurare la diversità dei gruppi e che, massimizzandola, porti ad individuare la migliore partizione.

Inoltre, poiché lo scopo di ogni partizione è quello di ridurre l'impurità in un insieme suddividendolo in sottogruppi, si definisca una funzione $i(t)$ di disomogeneità di un gruppo rispetto alla variabile dipendente. Per determinare la bontà della partizione, in termini di riduzione di variabilità della variabile criterio, si calcola la differenza tra la disomogeneità del nodo t e quella dei suoi nodi figli t_r :

$$\Delta i(s, t) = i(t) - \sum_r p_r i(t_r)$$

dove $p_r = p(t_r)/p(t)$ è la frequenza delle unità appartenenti al nodo figlio, data dal rapporto tra la frequenza del sottogruppo t_r e quella del nodo genitore t . La funzione criterio è data da:

$$\Phi(s, t) = p(t)i(t) - p(t)\left[\sum_r p_r i(t_r)\right] = I(t) - \sum_r I(t_r) = \Delta I(s, t)$$

Distinguendo le variabili criterio quantitative da quelle qualitative e a seconda del tipo di funzione di disomogeneità adottata, si ottengono diversi tipi di funzione criterio.

Variabile dipendente quantitativa

Una misura di distanza tra nodi è determinata sulla base di indici generati dai valori assunti dalla variabile criterio, quali media e varianza.

- **Criteri basati sulle misure di distanza tra le medie.**

Si indichi con:

- \bar{Y}_t la media del nodo genitore t di numerosità n ;
- \tilde{Y}_t la mediana del nodo genitore t di numerosità n ;
- \bar{Y}_{t_r} la media del nodo figlio t_r di numerosità n_r , ossia la media della variabile criterio condizionata all'insieme di modalità che hanno determinato il t_r -esimo nodo;
- \tilde{Y}_{t_r} la mediana del nodo figlio t_r di numerosità n_r , ossia la mediana della variabile criterio condizionata all'insieme di modalità che hanno determinato il t_r -esimo nodo;

- Y_{jr} , $j = 1, \dots, n_r$ i valori di Y che cadono nel nodo r .

Alcuni di questi criteri sono:

- la distanza euclidea:

utilizzando come funzione di impurità $i(t) = \sum_i (Y_i - \bar{Y}_t)^2$, si ottiene la funzione criterio

$$\Phi(s, t) = \sum_r (\bar{Y}_{t_r} - \bar{Y}_t)^2 n_r$$

che rappresenta la somma delle distanze euclidee tra i valori attesi dei nodi figli ed il valore atteso del nodo genitore. Rapportando questo valore con il suo massimo si ottiene come funzione criterio l'indice η^2 di Fisher:

$$\eta^2 = \frac{\sum_r (\bar{Y}_{t_r} - \bar{Y}_t)^2 n_r}{\sum_{j=1}^{n_r} (Y_{jr} - \bar{Y}_t)^2}$$

- la distanza assoluta:

utilizzando come funzione di impurità $i(t) = \sum_i |Y_i - \tilde{Y}_t|$, si ottiene la funzione criterio

$$\Phi(s, t) = \left[\sum_r |\tilde{Y}_{t_r} - \tilde{Y}_t| n_r \right]$$

che rappresenta la deviazione assoluta della mediana;

- la metrica di Minkowski:

se $i(t) = \left[\sum_i |Y_i - \tilde{Y}_t|^\lambda \right]^{\frac{1}{\lambda}}$, si ottiene come funzione di disomogeneità

$$\Phi(s, t) = \left[\sum_r |\tilde{Y}_{t_r} - \tilde{Y}_t|^\lambda w_{t_r} \right]^{\frac{1}{\lambda}}$$

con w_{t_r} coefficiente di ponderazione. Questa funzione criterio è utilizzata quando la variabile è su scala intervallare e permette di valutare efficientemente le distanze tra i nodi.

Ognuna delle misure di distanza tra le medie può essere utilizzata per valutare la dipendenza in media tra la variabile criterio e ciascun predittore. Se tutte le medie condizionate sono uguali tra loro e, quindi, uguali alla media generale, la capacità predittiva della singola variabile esplicativa rispetto alle medie di Y è nulla.

Variabile dipendente qualitativa

Per determinare la segmentazione di un nodo sono necessarie misure che esprimano la capacità del predittore di spiegare la mutabilità della variabile dipendente. Sia per le variabili ordinali che per quelle nominali si possono calcolare varie statistiche ed utilizzarle come base per la costruzione di funzioni criterio che indicheranno la riduzione nella frequenza di errori che si commetterebbero nel prevedere la variabile dipendente conoscendo, per ogni unità, la modalità di ogni singolo predittore. Si consideri una variabile dipendente divisa in K classi e un generico predittore X_i con m modalità raggruppate in gruppi t_r .

- **Criterio basato sulle mode condizionate.**

Se la funzione di impurità è $i(t) = n'(t) - n'_Y$, una possibile funzione criterio è il coefficiente λ di Goodman-Kruskal (1954):

$$\lambda = \frac{\sum_r n'(t_r) - n'_Y}{n - n'_Y} = \frac{\sum_r p'(t_r) - p'_Y}{1 - p'_Y}$$

dove n'_Y e $p'_Y = \frac{n'_Y}{n}$ sono rispettivamente le frequenze assoluta e relativa della moda della distribuzione marginale di Y e $n'(t_r)$ e $p'(t_r) = \frac{n'(t_r)}{n}$ le frequenze assoluta e relativa della moda di Y condizionata al nodo t_r . Questo coefficiente valuta la predittività delle modalità di Y conoscendo quelle di X_i sulla base delle sole frequenze delle mode. Esso varia tra 0 e 1: raggiunge il valore massimo quando le frequenze delle mode condizionate sono gli unici valori non nulli della matrice delle frequenze. E' inefficiente nel caso in cui una modalità della variabile dipendente prevalga sulle altre.

- **Criterio basato sull'entropia.**

La funzione di disomogeneità $i(t)$ uguale all'entropia o incertezza (Shannon

e Weaver, 1949) interna al gruppo:

$$i(t) = - \sum_k p(k|t) \log(p(k|t))$$

può essere utilizzata per misurare la riduzione nell'incertezza (entropia) di Y che risulta dalla conoscenza di X_i rispetto all'incertezza che si avrebbe conoscendo solo la distribuzione di Y .

Si ottiene così la funzione criterio:

$$\Phi(s, t) = - \sum_k p(k|t) \log(p(k|t)) + \sum_r \sum_k p(k|t_r) \log(p(k|t_r))$$

dove $p(k|t)$ e $p(k|t_r) = n_{kr}$ sono le probabilità che un'unità del gruppo t e del gruppo t_r rispettivamente appartengano alla classe k della variabile Y .

- **Criterio basato sul coefficiente χ^2 .**

Per valutare l'associazione tra variabili confrontando le frequenze osservate con quelle che si otterrebbero nell'ipotesi di indipendenza, si può utilizzare come funzione criterio l'indice χ^2 di Pearson (1900):

$$\chi^2 = \sum_r \sum_k \frac{(n_{k,tr} - \hat{n}_{k,tr})^2}{\hat{n}_{k,tr}}$$

la cui funzione di impurità è data da:

$$i(t) = \sum_k (n_{k,t} - \hat{n}_{k,t})^2$$

$\hat{n}_{k,t} = \frac{n_{.r} \times n_{k.}}{n}$ è la frequenza attesa nell'ipotesi di indipendenza. Il coefficiente χ^2 viene utilizzato per valutare la significatività di una segmentazione, pur essendo un indice simmetrico, in quanto non considera la relazione causa-effetto tra le variabili.

- **Criterio basato sull'indice di mutabilità di Gini.**

L'indice di Gini (1912):

$$i(t) = 1 - \sum_k p^2(k|t)$$

conduce ad una funzione criterio pari all'indice τ_B di Goodman-Kruskal (1954):

$$\tau_B = \frac{1 - \sum_r \frac{p(t_r)}{p(t)} (1 - \sum_k p^2(k|t_r))}{1 - \sum_k p^2(k|t)} = \frac{1 - \sum_r \frac{p(t_r)}{p(t)} i(t_r)}{i(t)}$$

dove t è il gruppo genitore, diviso nei figli $t_1, t_2, \dots, t_r, \dots$ e $p(t_r) = n_{kr}$ è la probabilità che un caso appartenga al nodo t_r . Il coefficiente varia tra 0 e 1: è nullo nell'ipotesi di indipendenza tra X_i e Y , mentre è massimo quando le frequenze di ciascuna distribuzione condizionata si concentrano su una sola modalità.

- **Criterio basato sulla mediana.**

Il calcolo della mediana sulla variabile criterio ha un significato statistico solo se la variabile è ordinale. Per misurare il grado di predittività della variabile ordinale in base alla conoscenza delle modalità di un predittore si può utilizzare il coefficiente di differenziazione Θ di Freeman (cfr. Par. 4.2).

Non esistono regole specifiche per la scelta della funzione criterio ed è, pertanto, consigliabile utilizzare più criteri in modo da verificarne il comportamento nei specifici casi di analisi.

3.2 Misure di separatezza tra popolazioni rispetto a gradazioni di gravità di fenomeni

Nell'ambito degli studi sociali, e di quelli sanitari in particolare, l'interesse è spesso rivolto a fenomeni caratterizzati da gradi progressivi di gravità che vanno da un livello di assenza del fenomeno a punte di gravità estrema, con almeno una gradazione intermedia. La distribuzione delle frequenze è caratterizzata tipicamente da una forte concentrazione sulla gradazione che denota la nullità o l'irrelevanza del fenomeno rappresentato e da un decadimento molto rapido in cor-

rispondenza della coda destra, su cui si collocano le massime intensità di gravità¹. I fenomeni così rappresentati possono essere la manifestazione di una malattia di una persona, il grado di difficoltà sofferto da un'impresa economica oppure, come nel nostro caso, il disagio avvertito da una famiglia.

Sia f_A la distribuzione di un campione da una ipotetica popolazione A e f_B la distribuzione di un campione dalla popolazione B . La distribuzione dei due campioni sia ripartita in K classi, come nella tabella 3.1.

Tab.3.1. *Distribuzione delle frequenze relative (p), assolute (n) e cumulate (N) di due campioni da due popolazioni rispetto a gradazioni di gravità di fenomeni*

Livello di gravità	A			B			Totale		
1	p_{1A}	n_{1A}	N_{1A}	p_{1B}	n_{1B}	N_{1B}	$p_{1\cdot}$	$n_{1\cdot}$	$N_{1\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	p_{kA}	n_{kA}	N_{kA}	p_{kB}	n_{kB}	N_{kB}	$p_{k\cdot}$	$n_{k\cdot}$	$N_{k\cdot}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
K	p_{KA}	n_{KA}	N_{KA}	p_{KB}	n_{KB}	N_{KB}	$p_{K\cdot}$	$n_{K\cdot}$	$N_{K\cdot}$
Totale	1	$n_{\cdot A}$	-	1	$n_{\cdot B}$	-	1	$n_{\cdot\cdot}$	-

L'obiettivo è quello di cogliere l'eventuale differenza tra le due distribuzioni A e B , misurandone le gradazioni su una scala ordinale ad almeno tre livelli.

Il problema sarebbe di poco conto se si volesse dare a ciascuna classe di gravità la stessa importanza. L'intenzione è quella di dare maggior peso alle variazioni nelle categorie estreme (superiori), che rappresentano i livelli di maggior gravità, che a quelle nelle categorie inferiori, che caratterizzano la nullità o l'irrilevanza del fenomeno.

S'intendono proporre metodi che forniscano come risultato non solo un livello di significatività ma anche, per facilità di comprensione ed interpretazione, misure della separatezza, ossia della differenza tra le due distribuzioni.

Al fine di individuare se il fattore di separazione, le cui modalità sono A e B , sia connesso con la gravità del fenomeno, si prendono in considerazione alcune classi di metodi per identificare la più efficiente, secondo standard prefissati, rispetto al

¹Se i dati fossero quantitativi, le distribuzioni sottostanti sarebbero concentrate attorno all'origine con una lunga coda a destra, tale da prefigurare una distribuzione di tipo esponenziale-negativo o log-normale.

problema in questione. Le classi considerate fanno riferimento a quattro approcci metodologici caratterizzati dal mantenere l'ordinalità della gravità e dal valorizzare le classi con frequenze molto basse (per maggiori dettagli cfr. Par. 3.2.1):

- massimizzazione della distanza distributiva;
- massimizzazione della differenza tra le mediane condizionate;
- massimizzazione dei rischi relativi;
- massimizzazione della significatività mediante soluzione non parametrica.

Potrebbero esserci delle compensazioni nelle due distribuzioni tra danni di diversa importanza sebbene sia molto più probabile il caso in cui le differenze riguardino tutte le categorie di gravità estreme.

3.2.1 Proposte metodologiche

I metodi che si analizzano sono appropriati per l'analisi asimmetrica, ossia per l'analisi della dipendenza tra la variabile che identifica i gruppi e quella che descrive la gravità del fenomeno in esame.

Il primo metodo fa riferimento al concetto di distanza distributiva. Si considerino le distribuzioni di frequenze su K categorie ordinate di due campioni tratti dagli strati A e B di una popolazione e si indichino con $p_{kj} = n_{kj}/n_{k\cdot}$ ($k = 1, \dots, K; j = A, B$), le frequenze relative dei due campioni e con \bar{p}_k il valor medio della k -esima categoria:

$$\bar{p}_k = \frac{n_{\cdot A}p_{kA} + n_{\cdot B}p_{kB}}{n_{\cdot\cdot}}$$

dove $n_{\cdot\cdot} = n_{\cdot A} + n_{\cdot B}$ è la somma delle numerosità dei gruppi tratti dalle popolazioni A e B .

La dipendenza in distribuzione può essere misurata attraverso il seguente indice (Fleiss, 1981):

$$\chi^2 = n_{\cdot A}n_{\cdot B} \sum_{k=1}^K \frac{\left(\frac{n_{kA}}{n_{\cdot A}} - \frac{n_{kB}}{n_{\cdot B}}\right)^2}{n_{k\cdot}} = \frac{n_{\cdot A}n_{\cdot B}}{n_{\cdot\cdot}} \sum_{k=1}^K \frac{(p_{kA} - p_{kB})^2}{\bar{p}_k} \quad (3.1)$$

Si può notare che, per ciascuna delle K categorie, le differenze tra le frequenze sono ponderate con peso inversamente proporzionale alla frequenza della categoria. Tale indice si distribuisce approssimativamente come un χ^2 con $K - 1$ gradi di libertà. Quanto più il valore empirico risulterà superiore al valore di un χ_{K-1}^2 , tanto più la differenza tra i due campioni sarà significativa.

La seconda misura di dipendenza è data dalla media delle misure di dipendenza calcolabili per ciascun gruppo. Tale misura può essere la differenza tra due proporzioni, il logaritmo dell'odds ratio e così via. Siano y_j il valore che la misura assume per il gruppo j -esimo, supposto indipendente dagli altri, $s(y_j)$ il suo standard error e $w_j = [s(y_j)]^{-2}$ il rispettivo peso. Un valore elevato dello standard error di y_j sta ad indicare che y_j non è un valore preciso e pertanto w_j risulta piccolo. Questo è ragionevole dal momento che stime imprecise non possono essere accompagnate da pesi elevati. D'altra parte, se lo standard error di y_j assume un valore piccolo, y_j è un valore piuttosto preciso e il relativo peso w_j è elevato. Anche questo è ragionevole dal momento che stime precise sono associate a pesi elevati. Supponendo un valore di y_j pari a zero in caso di assenza di associazione per il gruppo j -esimo, la quantità $\chi_j^2 = w_j y_j^2$ ha distribuzione χ_1^2 . Poiché l'interesse non è rivolto al j -esimo, o ad un altro specifico, gruppo, si valuta il grado medio di associazione attraverso la quantità χ_{assoc}^2 (Gart, 1962; 1970):

$$\chi_{assoc}^2 = \bar{y}^2 \sum_{j=1}^G \frac{1}{[s(y_j)]^2} = \frac{\left(\sum_{j=1}^G w_j y_j \right)^2}{\sum_{j=1}^G w_j} \quad (3.2)$$

dove:

- G è il numero di modalità della variabile "gruppi" che nel caso più semplice assume valore 2 ($j = A, B$);

- $\bar{y} = \frac{\sum_{j=1}^G w_j y_j}{\sum_{j=1}^G w_j}$ è una media pesata dei G gruppi.

Tale indice si distribuisce approssimativamente come un χ^2 con un grado di libertà. Il suo campo di variazione è rappresentato dai reali positivi e quanto più il suo valore risulterà superiore a quello di un χ^2_1 , tanto più sarà evidente la differenza in distribuzione tra i due campioni.

Un terzo metodo statistico, basato sulla distanza distributiva, va sotto il nome di analisi Ridit (*Relative to an Identified Distribution*) e il suo obiettivo è quello di mettere a confronto, nell'ambito di diverse popolazioni, una delle quali di riferimento ("standard"), la distribuzione di uno stesso carattere di tipo ordinale. L'analisi si basa sul Ridit medio, dato da:

$$\bar{r} = \sum_{k=1}^K r_k \frac{n_{kA}}{n_{\cdot A}} \quad (3.3)$$

dove $r_k = \frac{n_{1B} + \dots + n_{kB}}{2n_{\cdot B}}$ è il Ridit della categoria k -esima (Selvin, 1991).

Tale indicatore varia tra 0 e 1, assumendo così il significato di probabilità che un'unità appartenente al campione da verificare A possa avere livelli di gravità maggiori di una appartenente al campione di riferimento B . Un valore di 0,5 rappresenta l'equivalenza in gravità tra le due distribuzioni a confronto; valori superiori a 0,5 indicano che il campione A manifesta gravità superiore a quello di controllo B ; nel caso di un valore inferiore a 0,5 il campione A manifesta gravità inferiori al campione B . Per registrare lo scostamento di un certo livello nelle distribuzioni di frequenze è opportuno che l'indice si allontani da 0,5 in maniera apprezzabile. La statistica test per verificare la significatività dello scostamento è:

$$t = \frac{\bar{r} - 0,5}{\sqrt{\text{var}(\bar{r})}} \quad \text{con} \quad \text{var}(\bar{r}) = \frac{1}{12n_{\cdot A}} + \frac{1}{12n_{\cdot B}}$$

Questi tre indici sono non particolarmente adatti per la misura della dipendenza perché danno a ciascuna categoria una importanza proporzionale alla frequenza e, quindi, non esaltano l'importanza delle categorie di gravità estrema. Infatti:

- il χ^2 presenta a numeratore una differenza prossima allo zero per le gradazioni più elevate di gravità e il valore diminuisce ulteriormente quando viene elevato al quadrato;

- il χ^2_{assoc} non è efficiente nel caso di categorie con frequenze molto diverse tra loro, essendo una media delle misure di dipendenza calcolate in ognuno dei gruppi considerati;
- anche il Ridit, essendo una media ponderata, non valorizza le classi caratterizzate da basse frequenze.

Un altro metodo plausibile è basato sulla variabilità tra le mediane condizionate. L'indice che misura tale separatezza è il coefficiente di differenziazione Θ di Freeman (1965):

$$\Theta = \frac{\sum_{i=1}^G \sum_{j>i}^G |C_{ij} - D_{ij}|}{\sum_{i=1}^G \sum_{j>i}^G n_{.i} n_{.j}} \quad (3.4)$$

dove:

- G è il numero di strati da confrontare ($G \geq 2$);
- C_{ij} rappresenta il numero di confronti in cui le unità appartenenti alla categoria di gravità i del gruppo A hanno rango inferiore a quello delle unità della categoria j del gruppo B . Se $G = 2$: $C = n_{1A}(n_{2B} + \dots + n_{kB} + \dots + n_{KB}) + n_{2A}(n_{3B} + \dots + n_{kB} + \dots + n_{KB}) + \dots + n_{K-1A} n_{KB}$
- D_{ij} rappresenta, viceversa, il numero di confronti in cui le unità della categoria di gravità i del gruppo A hanno rango superiore a quello delle unità della categoria j del gruppo B . Se $G = 2$: $D = n_{KA}(n_{1B} + \dots + n_{kB} + \dots + n_{K-1B}) + n_{K-1A}(n_{1B} + \dots + n_{kB} + \dots + n_{K-2B}) + \dots + n_{2A} n_{1B}$

Il coefficiente Θ varia tra 0 e 1, estremi compresi. Il valore 0 si ottiene quando in corrispondenza di ogni gruppo le mediane condizionate del livello di gravità sono tra loro uguali e, quindi, si ha indipendenza in mediana; il valore 1 è raggiunto quando c'è una sola frequenza non nulla per ogni gruppo (perfetta dipendenza in mediana). Questo indice risulta del tutto inappropriato per le nostre analisi,

proprio per l'insensibilità della mediana alle frequenze estreme, quelle di maggior interesse, e perché attribuisce ad ogni classe importanza proporzionale alla rispettiva frequenza.

Si può considerare un indicatore di dipendenza basato sui rischi relativi. Sia $R_{k|1}$ ($k = 2, \dots, K$) il rischio del livello di gravità k -esimo in rapporto alla situazione di base che, senza perdere in generalità, si identifica con la modalità 1 (in genere, si tratta della nullità del fenomeno):

$$R_{k|1} = \frac{n_{kB} \times n_{1A}}{n_{kA} \times n_{1B}} \quad k = 2, \dots, K$$

$R_{k|1} = 1$ in caso di uguaglianza tra i rischi; $R_{k|1} > 1$ nel caso in cui si sia in presenza di un fattore di rischio; $R_{k|1} < 1$ nel caso di fattore protettivo.

La media dei rischi delle $K - 1$ categorie rapportate alla nullità del fenomeno:

$$\bar{R} = \frac{\sum_{k=1}^K R_{k|1}}{K} \quad (3.5)$$

evidenzia a sufficienza l'esistenza di differenze nei rischi di maggiore gravità. Nonostante questo aspetto positivo, esiste in questo approccio un grosso limite dato dalla mancanza di un test che valuti la significatività statistica del risultato ottenuto.

Per questo si fa ricorso ai modelli a odds proporzionali (McCullagh e Nelder, 1989), che assumono la forma:

$$\log\{\gamma_k(\mathbf{x})/(1 - \gamma_k(\mathbf{x}))\} = \theta_k - \boldsymbol{\beta}^T \mathbf{x} \quad k = 1, \dots, K - 1 \quad (3.6)$$

dove $\gamma_k(\mathbf{x}) = \Pr(Y \leq k|\mathbf{x})$ è la probabilità cumulata fino alla k -esima categoria, \mathbf{x} è il vettore delle variabili esplicative e $\boldsymbol{\beta}$ è un vettore di parametri ignoti (Agresti, 2002). Si veda al riguardo il paragrafo 4.2.

Questo criterio sembra appropriato perché riesce a cogliere differenze tra due distribuzioni anche in presenza di frequenze molto basse per alcune categorie della variabile risposta.

Il modello di cui questa funzione verifica l'adattamento è denominato da Agresti (2002) *cumulative link model*.

In alternativa alle soluzioni parametriche, si è tentato un approccio non parametrico, basato sulla metodologia dei test di permutazione (Pesarin, 2001). Il metodo è particolarmente indicato in presenza di basse numerosità campionarie, come accade nei fenomeni che si stanno considerando. In questo caso, una soluzione al problema è data dalla statistica test di Anderson-Darling per dati categoriali ordinati.

Indicando col simbolo * i dati permutati, ovvero ottenuti generando tutte le possibili permutazioni dall'unione dei sottocampioni per ciascuna delle K modalità, la statistica test di Anderson-Darling assume nel nostro caso la seguente forma:

$$T_{AD}^* = \sum_{k=2}^K (N_{kB}^* - N_{kA}^*) \left[4 \frac{N_{k\cdot}}{n_{\cdot\cdot}} \left(\frac{n_{\cdot\cdot} - N_{k\cdot}}{n_{\cdot\cdot}} \right) \frac{n_{\cdot A} n_{\cdot B}}{n_{\cdot\cdot} - 1} \right]^{-\frac{1}{2}} \quad (3.7)$$

dove N_{kA}^* e N_{kB}^* ($k = 2, \dots, K$) sono le frequenze cumulate ottenute dopo aver permutato i dati e $N_{k\cdot} = N_{kA} + N_{kB} = N_{kA}^* + N_{kB}^*$. Si noti che $n_{k\cdot}^* = n_{k\cdot}$ e $N_{k\cdot}^* = N_{k\cdot}$ per tutte le possibili permutazioni.

Questa statistica, oltre ad evidenziare eventuali differenze distributive tra i due campioni, ha anche il vantaggio di identificarne le direzioni, rilevando quale dei due presenti livelli di gravità superiori (Pesarin, 2001)². Questo criterio fornisce una misura della significatività, ma non fornisce una misura di dipendenza.

Per valutare la capacità dei metodi sopra esposti di cogliere l'importanza delle classi estreme si utilizzano i dati raccolti negli anni 2005 e 2006 dal *Centro Regionale di Documentazione ed Analisi sulla Famiglia* relativi al numero di componenti invalidi e di componenti affetti da malattie cronico-degenerative per famiglia.

3.2.2 Applicazioni esemplificative

Nella tabella 3.2 si riportano la distribuzione dell'intero campione di famiglie rispetto a quattro livelli di disagio e alcune misure dell'interazione tra il numero

²Altre statistiche test per confrontare, in ambito non parametrico, le distribuzioni di due campioni non sono esaminate in questa nota perché si tratta di approcci riconducibili a quelli qui trattati.

di invalidi e il numero di malati rilevati presso le famiglie campione³. I livelli di disagio sono così definiti:

- disagio nullo o fisiologico. Al momento dell'intervista, il rispondente ha dichiarato che la sua famiglia non ha problemi;
- disagio lieve, affrontabile con impiego di risorse esclusivamente familiari. La famiglia si ritiene colpita da problemi che, però, può risolvere con le proprie forze o ricorrendo alla famiglia allargata. Le famiglie che dichiarano una situazione di questo tipo possono essere assimilate a quelle che dichiarano di non avere problemi, essendo la vita, in sè, una sequenza di "sfide" più o meno risolvibili;
- disagio grave, tale da richiedere l'intervento esterno di istituzioni (servizi) o associazioni di volontariato;
- disagio molto grave, che permane anche dopo essere ricorsi all'aiuto esterno e crea una situazione a cui non è possibile porre rimedio sostanziale. I problemi dichiarati tanto gravi sono perlopiù relativi all'invalidità o alle malattie cronico-degenerative di uno o più componenti, problemi che richiedono un impegno spesso insostenibile.

Tab.3.2. *Distribuzione delle famiglie venete rispetto al disagio percepito e alla presenza di componenti invalidi e/o malati*

Livello di disagio	Numero invalidi		Malati cronico-degenerativi		Effetto di interazione			Totale
	0	≥ 1	≤ 1	≥ 2	0 invalidi ≤ 1 malati	0 invalidi ≥ 2 malati	≥ 1 invalido	
Nulla, fisiologica	1741	54	1756	39	1707	34	54	1795
Lieve	392	89	454	27	372	20	89	481
Acuto	97	44	130	11	88	9	44	141
Estremo	27	21	42	6	26	1	21	48
Totale	2257	208	2382	83	2193	64	208	2465

³Per problemi di frequenze estremamente basse, le famiglie in cui è presente almeno un invalido sono state raggruppate in un'unica modalità, indipendentemente dal numero di componenti affetti da malattie cronico-degenerative.

Si consideri dapprima il metodo basato sulla distanza tra distribuzioni. Tutti i risultati sono riportati nella tabella 3.3. Relativamente al numero di invalidi il χ^2 assume un valore pari a 307,7 che, se confrontato con un χ^2 con tre gradi di libertà, evidenzia una significativa differenza tra le distribuzioni del disagio delle famiglie con invalidi rispetto a quelle senza invalidi. Anche per l'effetto di interazione tra "zero invalidi, al più un malato" e "almeno un invalido" si ottiene un risultato di considerevole significatività ($\chi^2 = 317,0$). Nel caso dell'interazione, è possibile ricavare tre valori del χ^2_{assoc} , confrontando a coppie le tre modalità da cui è costituita. Per le altre variabili i risultati sono significativi, sebbene i valori assunti dal χ^2 siano notevolmente inferiori.

Nei campioni in cui sono presenti più malati e/o almeno un invalido, le frequenze dei livelli di gravità più importanti (acuto ed estremo) sono decisamente più elevate che nei casi di assenza di componenti con deficit. Ipotizzando un raddoppio di tali frequenze, l'indice non riesce a rilevare le differenze tra i due campioni. L'evidenza della differente distribuzione emerge solo nel caso in cui le frequenze che fanno riferimento ai livelli più elevati di gravità aumentino di almeno tre volte da un gruppo all'altro. Per i nostri fini, invece, anche piccole differenze possono risultare importanti.

Con l'indicatore χ^2_{assoc} si può valutare il grado di associazione tra i due campioni relativamente alla prima variabile combinando i rischi calcolati nei tre studi che confrontano il livello di disagio nullo o fisiologico con ciascuna delle altre tre gradazioni di gravità. Relativamente al numero di invalidi e all'interazione tra "zero invalidi, al più un malato" e "almeno un invalido", tali quantità risultano 348,3 e 356,5, rispettivamente. Anche i risultati relativi al numero di malati e agli altri due effetti di iterazione mostrano, sebbene in modo meno esplicito, evidenti differenze in distribuzione tra i gruppi.

Come con l'indice precedente, questa apparente adeguatezza è dovuta esclusivamente alle elevate frequenze nei livelli di maggiore gravità del campione in cui sono presenti più malati e/o almeno un invalido rispetto a quello in cui non sono presenti componenti con qualche tipo di deficit.

L'indice Ridit calcolato sui dati in esame coglie perfettamente le differenze tra i due campioni per tutte e tre le variabili: il Ridit medio si colloca, infatti, molto lontano da 0,5, che rappresenta l'ipotesi di indipendenza, come confermato dalla statistica test t . Tuttavia, se nel campione di "casi" le frequenze di disagio acuto o estremo al più raddoppiassero, il Ridit resterebbe prossimo a 0,5 e non renderebbe conto della forte diversità tra i gruppi. Ciò dipende dall'essere costruito come una media (ponderata con le frequenze cumulate) che non valorizza adeguatamente le classi di bassa numerosità.

Per le tre variabili a disposizione, il coefficiente Θ non evidenzia una cospicua diversità tra le mediane condizionate della gravità, nonostante i due campioni messi a confronto presentino frequenze decisamente diverse nei livelli più gravi di disagio per ognuna delle tre variabili. Rispetto al numero di invalidi, l'indice è 0,54, rispetto al numero di malati 0,28 e all'effetto di interazione $-0,48$.

Per ciascuna variabile si è calcolata la media aritmetica semplice dei rischi. Questo indicatore riesce ad evidenziare le differenze tra i rischi dei due livelli di maggior gravità (acuto ed estremo) nel caso di aumento di frequenze da un campione all'altro. Esso assume valori elevati per tutte e tre le variabili a disposizione. Indicando con \bar{R} la media aritmetica dei rischi, il risultato relativo alla prima variabile è pari a:

$$\bar{R}\{\text{invalidi}\} = \frac{R(\text{estremo} \mid \text{nullo}) + R(\text{acuto} \mid \text{nullo}) + R(\text{lieve} \mid \text{nullo})}{3} = 15,7$$

Per la seconda variabile il valore assunto dall'indicatore risulta 4,3, per la terza 3,3 confrontando la prima modalità (nessun invalido, al più un malato) con la seconda (nessun invalido, due o più malati), 16,3 confrontando la prima con la terza (uno o più invalidi) e 6,4 confrontando la seconda con la terza.

Relativamente all'applicazione del modello a odds proporzionali, la stima puntuale pari a 8,847 mostra come il rischio di disagio per le famiglie con almeno un invalido in casa sia di quasi nove volte maggiore rispetto a quelle con zero invalidi. Una differenza di gran lunga inferiore, in termini di percezione di disagio, esiste tra le famiglie con nessun invalido e almeno due malati e quelle con almeno un invalido (odds ratio = 2,825).

Questa metodologia fornisce prime buone risposte al nostro problema dal momento che mette in risalto le caratteristiche delle famiglie a maggior rischio di disagio. Come esemplificative si sono considerate le variabili numero di invalidi e numero di malati cronico-degenerativi, ma anche per altri aspetti si ricavano risultati analoghi.

I metodi sopra riportati sono stati applicati in un secondo momento invertendo l'ordinamento della gravità, ossia considerando una scala di misura che va da una gradazione di gravità molto elevata alla nullità del fenomeno. Le conclusioni alle quali si perviene sono essenzialmente le stesse.

Da ultimo si è preso in considerazione l'approccio non parametrico mediante la statistica test di Anderson-Darling⁴. La significatività che ne deriva sembra indicare l'adeguatezza del metodo rispetto al problema in questione. Per la prima variabile si evidenzia che le gradazioni di gravità sono maggiori nelle famiglie con uno o più invalidi; per la seconda nelle famiglie con due o più malati e nella terza nelle famiglie con almeno un invalido e almeno due malati.

⁴Per l'analisi è stato utilizzato il software NPC Test (Pesarin, 2001).

Tab.3.3. Risultati ottenuti applicando i diversi metodi statistici considerati al fenomeno del disagio familiare

	Variabile esplicativa e relative modalità	χ^2	χ^2_{assoc}	<i>Ridit</i>	Θ di <i>Freeman</i>	<i>Media dei rischi</i>	Odds Ratio	<i>Intervallo di confidenza</i>	<i>Anderson Darling</i>
Gravità misurata su quattro livelli	Numero di invalidi (nessuno/uno o più)	307,7	348,3	$\bar{r} = 0,02$ $t = -22,9$	0,54	15,7	8,847	8,713 - 8,983	0,0001
	Numero di malati cronico-degenerativi (al più uno/due o più)	39,4	45,4	$\bar{r} = 0,01$ $t = -15,1$	0,28	4,3	3,373	3,296 - 3,452	0,0001
	Effetto di interazione (nessun invalido, al più un malato/nessun invalido, due o più malati)	33,5	30,9	$\bar{r} = 0,01$ $t = -13,4$	-0,48	3,3	3,264	3,177 - 3,353	0,0001
	Effetto di interazione (nessun invalido, al più un malato/uno o più invalidi)	317,0	356,5	$\bar{r} = 0,02$ $t = -22,8$		16,3	9,113	8,972 - 9,256	
	Effetto di interazione (nessun invalido, due o più malati/uno o più invalidi)	17,8	21,1	$\bar{r} = 1,12$ $t = 14,9$		6,4	2,825	2,741 - 2,912	
Gravità misurata su tre livelli	Numero di invalidi (nessuno/uno o più)	194,9	155,7	$\bar{r} = 0,03$ $t = -22,3$	0,26	9,2	8,038	7,886 - 8,193	0,0001
	Numero di malati cronico-degenerativi (al più uno/due o più)	63,1	23,2	$\bar{r} = 0,01$ $t = -15,0$	0,13	3,8	3,922	3,807 - 4,041	0,0097
	Effetto di interazione (nessun invalido, al più un malato/nessun invalido, due o più malati)	58,2	15,0	$\bar{r} = 0,01$ $t = -13,3$	0,22	2,7	3,290	1,626 - 6,657	0,0001
	Effetto di interazione (nessun invalido, al più un malato/uno o più invalidi)	202,4	160,1	$\bar{r} = 0,03$ $t = -22,2$		9,5	8,366	5,921 - 11,821	
	Effetto di interazione (nessun invalido, due o più malati/uno o più invalidi)	7,3	4,6	$\bar{r} = 1,3$ $t = 20,7$		4,9	2,554	1,220 - 5,346	

In conclusione si può affermare che i metodi basati sulla massimizzazione della distanza distributiva e della differenza tra le mediane condizionate si sono rivelati inadeguati alla risoluzione del problema. D'altro canto, quelli basati sulla massimizzazione dei rischi relativi e della significatività mediante approccio non parametrico sono adatti allo studio di fenomeni che presentano un decadimento molto rapido della frequenze.

I metodi sono stati applicati anche considerando tre livelli di gravità (nullo o lieve, acuto, estremo). I risultati sono riportati nella seconda parte della tabella 3.3. Pur valorizzando le modalità che denotano la gravità del fenomeno a dispetto di quella che rappresenta la nullità o irrilevanza dello stesso, essi sono una conferma di ciò a cui si è giunti utilizzando quattro livelli di gravità. Solo gli indici χ^2 e χ^2_{assoc} non evidenziano la differenza in distribuzione tra le famiglie con nessun invalido e almeno due malati e quelle con almeno un invalido. D'altra parte, anche nell'analisi a quattro modalità, nonostante anche questi due indici risultassero significativi, la differenza tra le due tipologie di famiglie, per quanto riguarda la percezione di disagio da parte delle stesse, non era così evidente quanto negli altri confronti ("0 invalidi vs. 1 o più invalidi", "al più un malato vs. 2 o più malati", "0 invalidi e al più un malato vs. 1 o più invalidi", "0 invalidi e al più un malato vs. 1 o più invalidi").

3.3 La regressione logistica e i modelli a odds proporzionali

In questo paragrafo si approfondiscono alcuni aspetti della regressione logistica e dei modelli a odds proporzionali al fine di determinare gruppi di famiglie venete a rischio di disagio, con riferimento al caso di variabile dipendente che assume più di due modalità, e precisamente quattro, tra loro ordinate.

Il modello di regressione logistica (binaria) si distingue da quello lineare in quanto la variabile indipendente Y è una variabile casuale Bernoulliana; essa assume valore 0 o 1 e la probabilità che assuma valore 1 è funzione di un regressore

$\mathbf{x} = (\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_m)$:

$$P(Y = 1|X = \mathbf{x}) = \pi(\mathbf{x}) \quad (3.8)$$

In particolare:

$$\pi(\mathbf{x}) = \frac{\exp(\beta_0 + \beta_1 \mathbf{x}_1 + \dots + \beta_m \mathbf{x}_m)}{1 + \exp(\beta_0 + \beta_1 \mathbf{x}_1 + \dots + \beta_m \mathbf{x}_m)} \quad (3.9)$$

Per l'interpretazione dei parametri si definisce l'odds, ossia il rapporto fra la probabilità che Y assuma valore 1 e quella che assuma valore 0, condizionatamente al valore assunto dai regressori:

$$odds(\mathbf{x}) = \frac{P(Y = 1|X = \mathbf{x})}{P(Y = 0|X = \mathbf{x})} = \frac{P(Y = 1|X = \mathbf{x})}{1 - P(Y = 1|X = \mathbf{x})} = \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \quad (3.10)$$

Il logit, ossia il logaritmo dell'odds, è una funzione lineare dei parametri:

$$logit(\mathbf{x}) = \ln[odds(\mathbf{x})] = \beta_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_m \mathbf{x}_m \quad (3.11)$$

Se il parametro β_j è positivo, un incremento della variabile esplicativa \mathbf{x}_j induce un incremento nella probabilità che la variabile dipendente assuma valore 1.

Indicando con y_1, \dots, y_n i valori osservati della variabile dipendente e con \mathbf{x}_i i vettori contenenti i corrispondenti valori delle variabili esplicative, si ottiene la funzione di verosimiglianza:

$$L(\beta) = \prod_{i=1}^n \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{1-y_i} \quad (3.12)$$

e la log-verosimiglianza:

$$l(\beta) = \sum_{i=1}^n \{y_i \ln[\pi(\mathbf{x}_i)] + (1 - y_i) \ln[1 - \pi(\mathbf{x}_i)]\} \quad (3.13)$$

Per l'analisi dei dati nel caso di variabile risposta ordinale, una possibilità consiste nel dicotomizzare la variabile risposta ed implementare un modello di regressione logistica binaria con la conseguente perdita di informazione. Inoltre i rapporti tra odds risultanti possono dipendere dal modo in cui si effettua la dicotomizzazione, scelta spesso arbitraria.

Esistono, d'altra parte, tre diverse impostazioni per descrivere la probabilità che la variabile dipendente assuma le diverse modalità, le quali si differenziano per il logit che considerano e i conseguenti vincoli che sono imposti sui parametri. In particolare si distingue tra il *proportional-odds*, il *continuation-ratio* e l'*adjacent-category model*.

Per il nostro problema, il modello più idoneo risulta essere quello a odds proporzionali, secondo cui l'odds ratio per ciascuna delle variabili esplicative è costante tra tutte le possibili aggregazioni di modalità adiacenti della variabile risposta. I modelli a odds proporzionali, con il logit come funzione legame, furono introdotti da McCullagh (1980) in ambito biologico, per valutare le misure effettuate su farmaci, dove "la variabile latente corrisponde generalmente ad una misura di 'tolleranza' (al farmaco), che si assume avere una distribuzione continua nella popolazione. Le tolleranze non sono direttamente osservabili, ma l'aumento di tolleranza si rispecchia in un aumento della probabilità di sopravvivenza. Le categorie sono viste come intervalli contigui su una scala continua".

La scelta dell'utilizzo dei modelli a odds proporzionali è dovuta al fatto che sono stati utilizzati nell'ambito della regressione logistica con variabile risposta ordinale e rappresentano un'estensione multivariata dei modelli lineari generalizzati. Sono stati sviluppati nel contesto della regressione, dove l'obiettivo è stimare l'associazione tra fattori di rischio e una risposta ordinale di interesse.

Si consideri una variabile risposta ordinale Y a K modalità con probabilità $\pi_k(\mathbf{x}) = P(Y = k|\mathbf{x})$, $k = 1, \dots, K$, $\sum_{k=1}^K \pi_k(\mathbf{x}) = 1$ e un vettore di variabili esplicative \mathbf{x} .

Nel caso di variabile dipendente ordinale, è importante, al fine di arrivare a validi risultati, che le conclusioni non dipendano dal numero di modalità della variabile risposta, sebbene, riducendo il numero di modalità, diminuisca l'informazione disponibile e cambino le stime (McCullagh & Neleder, 1989).

L'attenzione è posta su modelli basati non sulle singole probabilità $\pi_k(\mathbf{x})$, bensì

sulle probabilità cumulate:

$$\begin{aligned}\gamma_1(\mathbf{x}) &= \pi_1(\mathbf{x}) \\ \gamma_2(\mathbf{x}) &= \pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) \\ &\vdots \\ \gamma_K(\mathbf{x}) &\equiv 1\end{aligned}$$

E in generale:

$$\gamma_k(\mathbf{x}) = \Pr(Y \leq k | \mathbf{x}) \quad \text{con} \quad k = 1, \dots, K \quad (3.14)$$

I due insiemi di probabilità sono equivalenti, ma modelli che sfruttano le probabilità cumulate risultano verosimilmente avere, nel caso di variabile risposta ordinale, proprietà migliori. Ovviamente ha senso lavorare con le probabilità cumulate solo nel caso in cui le categorie della variabile risposta siano ordinate (McCullagh & Nelder, 1989).

Nei modelli a odds proporzionali, per ogni modalità k della variabile dipendente Y , con $k \geq 0$, si considera il logit fra le categorie precedenti o pari a k rispetto alle successive:

$$\log \frac{\gamma_k(\mathbf{x})}{1 - \gamma_k(\mathbf{x})} = \theta_k - \boldsymbol{\beta}^T \mathbf{x} \quad \text{con} \quad k = 1, \dots, K - 1 \quad (3.15)$$

dove $\gamma_k(\mathbf{x}) = \Pr(Y \leq k | \mathbf{x})$ è la probabilità cumulata fino alla categoria k , quando il vettore delle covariate è \mathbf{x} , e $\boldsymbol{\beta}$ è un vettore di parametri ignoti.

Tale modello prevede $K - 1$ intercette e un unico coefficiente angolare, β . Le intercette $\theta_1, \dots, \theta_{K-1}$, rappresentano i punti di discontinuità di una variabile latente continua e non osservabile, ma il nostro interesse è rivolto al coefficiente β , ossia al logaritmo del rapporto crociato dell'evento $Y > k$ versus $Y \leq k$ per ogni cambiamento unitario di \mathbf{x} .

Il nome proportional-odds deriva dal fatto che il rapporto tra odds dell'evento $Y \leq k$ in corrispondenza dei vettori $\mathbf{x} = \mathbf{x}_1$ e $\mathbf{x} = \mathbf{x}_2$ è:

$$\frac{\gamma_k(\mathbf{x}_1)/(1 - \gamma_k(\mathbf{x}_1))}{\gamma_k(\mathbf{x}_2)/(1 - \gamma_k(\mathbf{x}_2))} = \exp \{ -\boldsymbol{\beta}^T (\mathbf{x}_1 - \mathbf{x}_2) \} \quad (3.16)$$

che, come si può notare, risulta indipendente dalla scelta della categoria k : l'effetto delle variabili esplicative è costante al variare di k .

Si noti che la (3.15) è identica all'equazione del modello lineare logistico. Nel caso in cui $K = 2$, la (3.15) è equivalente al modello lineare logistico per dati binari (Cox, 1970) e al modello log-lineare. Tuttavia, quando il numero di categorie è superiore a due, la (3.15) non rappresenta un modello con struttura log-lineare.

Ogni logit ha la propria intercetta. I termini θ_k , al crescere di k , assumono via via valori più elevati poiché $\Pr(Y \leq k|\mathbf{x})$ aumenta all'aumentare di k , per \mathbf{x} fissato, e il logit è una funzione crescente di questa probabilità (Agresti, 2002).

Il punto di forza dei modelli a odds proporzionali è che il rapporto tra odds relativo ad ogni variabile esplicativa può essere interpretato come sintesi dei rapporti tra odds delle regressioni logistiche binarie ottenute cumulando le modalità adiacenti della variabile risposta in tutti i modi possibili (Scott et al., 1997). Mentre una regressione logistica binaria modella un singolo logit, i modelli a odds proporzionali ne modellano più di uno alla volta, in sequenza.

Nel nostro caso, la variabile di interesse è ordinale a quattro modalità, che determinano conseguentemente tre punti di discontinuità: tra $k \leq 1$ e $k > 1$, tra $k \leq 2$ e $k > 2$ e tra $k \leq 3$ e $k > 3$. In questo modo k assume valori 1, 2 e 3 nell'equazione (3.15).

La significatività statistica dell'associazione tra la risposta e le variabili esplicative può essere verificata testando l'ipotesi $H_0 : \boldsymbol{\beta} = 0$ o, in termini di odds ratio, $H_0 : \theta_k = 1$. In questo modo, se $\theta_k = 1$ le variabili risultano non associate. Nel caso in cui $1 \leq \theta_k < \infty$, le unità per cui $\mathbf{x} = \mathbf{x}_1$ hanno una maggior probabilità di avere gradazioni inferiori nella risposta rispetto a quelle per le quali $\mathbf{x} = \mathbf{x}_2$, mentre se $0 \leq \theta_k < 1$, le unità per cui $\mathbf{x} = \mathbf{x}_1$ hanno minor probabilità avere gradazioni inferiori nella risposta rispetto a quelle per cui $\mathbf{x} = \mathbf{x}_2$. Le ipotesi sui parametri $\boldsymbol{\beta}$ possono essere verificate mediante la statistica test di Wald $W = \hat{\boldsymbol{\beta}}' V^{-1} \hat{\boldsymbol{\beta}}$, che ha distribuzione χ_1^2 e V^{-1} è la matrice di informazione stimata.

Nel pacchetto statistico SAS, utilizzato per le analisi che seguono, la procedura utilizzata per implementare un modello a odds proporzionali è la PROC LOGISTIC.

Capitolo 4

La percezione del disagio da parte della famiglia

4.1 Percezione del disagio

Un'analisi preliminare mostra che il 74,6% delle famiglie venete non si ritiene colpito da problemi generatori di disagio. Questo è un dato confortante che rappresenta le famiglie a cui non è necessario un aiuto dall'esterno, ma che fa rientrare in una situazione di normalità le usuali difficoltà della vita. Il 17,9% si ritiene colpito da problemi risolvibili con le proprie forze, il 5,4% da problemi risolvibili con aiuto esterno alla famiglia e il restante 2,1% da gravi problemi non risolvibili neppure con l'intervento dei servizi. La causa principale di disagio è la salute di uno o più membri della famiglia (64,7% delle famiglie con disagio estremo o risolvibile con aiuto esterno, 7,6% delle famiglie totali); seguono i problemi economici (39,6% delle famiglie con disagio, 7,1% del campione totale) e quelli di tipo socio-relazionali (11,9% delle famiglie disagiate, 2,4% del totale).

Sfruttando la teoria relativa ai modelli a odds proporzionali, è interessante analizzare la percezione del disagio rispetto ad alcune variabili considerate nell'analisi. Dapprima le variabili sono considerate singolarmente, poi se ne studiano le interazioni per avere una visione globale delle famiglie a maggior rischio di

disagio.

I fattori che possono dare origine ad una situazione di non agio sono molteplici; nella nostra indagine sono stati presi in considerazione alcune variabili strutturali, alcuni stati critici tra cui l'invalidità, la malattia cronico-degenerativa, la disoccupazione e altre situazioni che possono turbare in qualche modo la stabilità familiare.

Ciò che si intende valutare è la relazione asimmetrica che intercorre tra ogni fattore possibilmente fonte di disagio e la percezione stessa del disagio, che rispecchia l'effetto del fattore sulla famiglia.

Per quanto riguarda l'età media della famiglia si può notare che al di sopra dei 70 anni è maggiore la probabilità di disagio. Questa caratteristica si riscontra soprattutto nelle famiglie che dichiarano di essere afflitte da gravi problemi non risolvibili (se si escludono le famiglie con età media inferiore ai 30 anni, per le quali non si hanno casi di grave disagio). Tali famiglie disagiate sono perlopiù costituite da anziani che vivono soli o in coppia e che non possono fare affidamento su una rete parentale o amicale in caso di necessità (Tab.4.1).

Tab.4.1. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per età media dei componenti e gravità del disagio*

Età media	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologica	
Fino a 29 (n=474)	0,0	4,0	16,4	79,6	100,0
30-49 (n=961)	0,8	4,5	15,7	79,0	100,0
50-69 (n=642)	2,6	4,4	22,0	71,0	100,0
70 o più (n=388)	7,1	11,2	19,1	62,6	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

La tabella seguente, che riporta le stime degli odds ratio ottenute facendo i confronti a coppie tra le diverse classi di età, mostra che la probabilità di disagio è significativamente maggiore nelle famiglie con età media dei componenti maggiore di 69 anni rispetto a quelle più giovani.

In particolare le famiglie i cui componenti hanno un'età media pari a 70 anni o più hanno probabilità di passare da un certo livello di disagio ad uno immedia-

tamente superiore del 169,2% maggiore rispetto a quelle con componenti di età media inferiore ai 30 anni.

Confronti	Odds Ratio	Intervallo di confidenza
30-49 vs. fino a 29	1,065	0,818 - 1,386
50-69 vs. fino a 29	1,491	1,133 - 1,963
70 o più vs. fino a 29	2,692	2,010 - 3,605
50-69 vs. 30-49	1,401	1,118 - 1,755
70 o più vs. 30-49	2,529	1,977 - 3,234
70 o più vs. 50-69	1,805	1,396 - 2,335

Questo risultato può essere spiegato dal fatto che in età avanzate, tendenzialmente, iniziano a manifestarsi problemi di salute o altri legati alla vita familiare che, mentre nelle famiglie con componenti giovani possono essere facilmente superati, nelle famiglie di anziani fanno sentire maggiormente il loro peso e appaiono più gravi.

4.1.1 Stati critici

Invalidità

Nel 7,7% delle famiglie intervistate è presente almeno un invalido. L'invalidità più diffusa è quella motoria parziale (47,9% del totale della popolazione affetta da qualche tipo di invalidità) (Tab.2.80 e 2.81).

La presenza di uno o più componenti invalidi in famiglia comporta un notevole aumento della percezione del disagio: tra le famiglie che non hanno invalidi in casa solo l'1,2% dichiara di avere gravi problemi non risolvibili, contro il 12,1% di quelle che hanno al loro interno almeno un membro affetto da qualche tipo di invalidità. Anche osservando gli altri livelli di disagio si può giungere alla conclusione che la presenza di invalidi è un fattore che spesso minaccia la situazione di agio familiare. Mentre tra le famiglie che non hanno invalidi in casa il 78,5% dichiara di non avere problemi, tra quelle in cui è presente almeno un membro invalido, tale percentuale scende a 28,6% (Tab.4.2).

Tab.4.2. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di invalidi e gravità del disagio*

Invalidi	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
0 (n=2256)	1,2	4,2	16,1	78,5	100,0
1 o più (n=208)	12,1	18,8	40,5	28,6	100,0
Totale (n=2464)	2,1	5,4	17,9	74,6	100,0

Analizzando la dipendenza tra la variabile che conta il numero di invalidi e quella che misura la percezione del disagio, si nota che la probabilità di una famiglia con invalidi di passare da un certo livello di disagio ad uno superiore è di quasi nove volte maggiore rispetto alla stessa probabilità calcolata per una famiglia che non ha membri invalidi al suo interno.

Confronti	Odds Ratio	Intervallo di confidenza
1 o più invalidi vs. 0 invalidi	8,847	8,713 - 8,983

L'handicap, come si può notare anche nelle analisi presentate nei successivi capitoli, si configura come lo stato critico per eccellenza, che "mette alla prova" la famiglia, evidenziandone gli stili di funzionamento, le risorse e il rapporto con l'ambiente che la circonda.

Le fasi di vita di una famiglia con componenti affetti da invalidità sono caratterizzate da un accumulo di eventi stressanti che minacciano la stabilità dei singoli individui e dell'intero nucleo familiare. La famiglia ha la necessità di attivare una serie di risorse interne ed esterne per fronteggiare le esigenze specifiche determinate dall'invalidità e superare i momenti di crisi e di tensione (Valtolina, 2002). Beresford (1994) individua, tra le risorse, fattori sia personali sia socio-ecologici. Tra i primi, i più importanti sono la salute fisica, i valori e le credenze personali, il carattere; per quanto riguarda la sfera sociale, invece, emergono gli aiuti interni ed esterni alla famiglia, il contesto socio-economico in cui la famiglia si trova inserita e lo stato occupazionale dei componenti. In particolare, il supporto esterno alla famiglia ristretta può provenire dalla famiglia d'origine o allargata, dagli amici più vicini, dalle istituzioni e dalle associazioni di volontariato ed è indispensabile al fine di permettere ai componenti di riprendere energia fisica e psicologica.

Oltre alle risorse, la famiglia può mettere in atto una serie di strategie e comportamenti per affrontare l'evento invalidità. Quelle individuate da Hepple (1989) ed utilizzate dai genitori di figli portatori di handicap, ad esempio, sono il parlare apertamente con il proprio coniuge, condividere con lui/lei i propri sentimenti e valorizzare gli aspetti positivi della malattia del figlio.

Malattie cronico-degenerative

Nel 21,6% delle famiglie intervistate è presente almeno un componente affetto da una qualche malattia cronico-degenerativa. Le malattie più diffuse sono quelle del cuore e del sistema circolatorio (infarto, ipertensione arteriosa, angina pectoris, etc.), che si presentano nel 44,3% dei casi.

Come nel caso degli invalidi, anche la presenza di malati comporta la percezione di un livello di disagio maggiore nelle famiglie: il 50,7% delle famiglie in cui sono presenti almeno due componenti affetti da malattie cronico-degenerative afferma di avere problemi. La stessa percentuale calcolata nelle famiglie in cui il numero di componenti malati è al più uno, è notevolmente più bassa (24,5%) (Tab.4.3).

Tab.4.3. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di malati cronico-degenerativi e gravità del disagio*

Malati cronico- degenerativi	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologica	
Al più 1 malato (n=2382)	1,9	5,0	17,6	75,5	100,0
Almeno 2 malati (n=83)	7,3	15,3	28,1	49,3	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

La dipendenza tra la percezione del disagio e la presenza di malati cronico-degenerativi è significativa: la probabilità di disagio aumenta di oltre tre volte quando in famiglia sono presenti almeno due componenti malati.

Confronti	Odds Ratio	Intervallo di confidenza
Almeno 2 malati vs. al più 1 malato	3,373	3,296 - 3,452

Livello di autonomia degli invalidi e dei malati

La mancanza di autonomia, che porta una persona a dipendere da altri membri, può comportare problemi di carattere organizzativo e relazionale all'interno della famiglia.

Nel 58,7% delle famiglie al cui interno è presente almeno un componente malato, questo risulta non autonomo. In questi casi la famiglia si ritiene colpita da gravi problemi talvolta risolvibili con le proprie forze, talvolta con l'aiuto esterno, talvolta, infine, non risolvibili. Solo il 16,1% delle famiglie in cui è presente almeno un invalido non autonomo non avverte alcun tipo di disagio, contro il 46,9% delle famiglie in cui è presente almeno un invalido, ma autonomo (Tab.4.4).

Tab.4.4. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per autonomia dell'invalido e gravità del disagio*

Autonomia dell'invalido	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologica	
No (n=123)	14,9	23,1	45,9	16,1	100,0
Si (n=85)	8,0	12,3	32,8	46,9	100,0
Totale (n=208)	12,1	18,7	40,6	28,6	100,0

La percezione del disagio dipende in maniera significativa dalla mancanza di autonomia dei membri invalidi: in presenza di membri invalidi non autonomi, la probabilità di passare da un certo livello di disagio ad uno subito superiore aumenta di circa tre volte.

Confronti	Odds Ratio	Intervallo di confidenza
Invalidi non autonomi vs. invalidi autonomi	2,956	1,738 - 5,029

Nelle famiglie in cui sono presenti malati cronico-degenerativi il 25,1% dei malati risulta non autonomo.

Anche in questo caso, in presenza di malati non autonomi, la percezione del disagio aumenta, sebbene la maggior parte delle famiglie dichiara di non avere problemi o, al più, problemi risolvibili al proprio interno (Tab.4.5).

Tab.4.5. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per autonomia del malato e gravità del disagio*

Autonomia del malato	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
No (n=138)	16,9	16,7	40,7	25,8	100,0
Si (n=423)	3,4	11,6	26,7	58,3	100,0
Totale (n=561)	6,8	12,9	30,3	50,0	100,0

Rispetto alle famiglie in cui i malati sono autonomi, in quelle in cui è presente almeno un malato non autonomo il rischio di disagio aumenta, in maniera ancora più evidente rispetto al caso di presenza di invalidi non autonomi.

Confronti	Odds Ratio	Intervallo di confidenza
Malati non autonomi vs. malati autonomi	3,726	2,583 - 5,374

Disoccupazione

Le persone disoccupate sono l'1,5% della popolazione intervistata e nel 63,5% dei casi hanno un'età inferiore ai 35 anni. La presenza in famiglia di persone senza lavoro è fonte di disagio. Si tratta nella maggior parte dei casi di un problema risolvibile con le forze della famiglia, che si supera una volta che il componente trova lavoro: si ritiene colpito da problemi che si risolvono entro le mura domestiche il 33,5% delle famiglie in cui almeno un membro è senza lavoro, contro il 17,5% di quelle in cui non vi sono disoccupati (Tab.4.6).

Tab.4.6. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di disoccupati e gravità del disagio*

Disoccupati	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
0 (n=2392)	2,1	5,3	17,5	75,2	100,0
1 o più (n=73)	2,2	7,6	33,5	56,7	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

Nonostante il problema non sia allarmante nel territorio del Veneto (solo l'1,5% della popolazione è disoccupata (Tab.2.6)) la dipendenza tra disagio e presenza di

membri disoccupati è statisticamente significativa. L'aumento della probabilità, più che doppio, di disagio nel caso di disoccupati in famiglia ($OR = 2,300$) è, tuttavia, attribuibile prevalentemente alla diminuzione della percentuale di famiglie, con almeno un membro disoccupato, che si dichiarano senza problemi a discapito dell'aumento di quelle che hanno dichiarato di aver problemi, ma risolvibili al loro interno.

Confronti	Odds Ratio	Intervallo di confidenza
1 o più disoccupati vs. 0 invalidi	2,300	1,465 - 3,610

Componenti della famiglia

Nel corso degli ultimi anni si è assistito alla progressiva diminuzione della dimensione familiare. Le famiglie composte da uno o due componenti rappresentano il 42,1% del totale; in particolare il 13,9% è costituito da un membro e il restante 28,2% da due.

Tendenzialmente, le famiglie di piccole dimensioni sono sede di maggior problemi e disagi. Effettivamente, all'aumentare del numero di componenti diminuisce in maniera significativa la probabilità di disagio: questo perché nelle famiglie più numerose c'è maggior possibilità di reciproco aiuto in caso di necessità. E' evidente la diminuzione percentuale di famiglie con problemi più gravi nel passare dalle famiglie con uno a quelle con cinque componenti (Tab.4.7).

Tab.4.7. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di componenti e gravità del disagio*

Numero di componenti	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
1 (n=352)	5,1	8,9	15,1	70,9	100,0
2 (n=694)	2,8	5,0	20,4	71,8	100,0
3 (n=677)	1,9	4,3	19,6	74,1	100,0
4 (n=595)	0,0	5,2	15,3	79,5	100,0
5 o più (n=147)	1,1	4,3	16,8	76,8	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

La dipendenza tra numero di componenti e percezione del disagio risulta significativa solo nei confronti tra le famiglie con quattro componenti e con uno, due e tre, rispettivamente. Tendenzialmente, però, le famiglie di diversa dimensione non differiscono tra loro rispetto al disagio dichiarato.

Confronti	Odds Ratio	Intervallo di confidenza
2 componenti vs. 1 componente	0,844	0,643 - 1,108
3 componenti vs. 1 componente	0,766	0,581 - 1,008
4 componenti vs. 1 componente	0,557	0,415 - 0,746
5 componenti vs. 1 componente	0,728	0,476 - 1,112
3 componenti vs. 2 componenti	0,907	0,719 - 1,144
4 componenti vs. 2 componenti	0,660	0,512 - 0,849
5 componenti vs. 2 componenti	0,862	0,579 - 1,283
4 componenti vs. 3 componenti	0,727	0,563 - 0,940
5 componenti vs. 3 componenti	0,951	0,637 - 1,418
5 componenti vs. 4 componenti	1,307	0,866 - 1,974

Figli in età 0-5

Un'ipotesi che inizialmente si era ritenuta plausibile è che la presenza di figli possa essere fonte di disagio per la famiglia, la quale deve reimpostare le sue abitudini quotidiane sulla base delle esigenze dei più piccoli. In realtà, i figli si rivelano una risorsa per la famiglia e non compromettono in maniera negativa il suo equilibrio: la gioia che deriva dalla presenza di un bambino non è paragonabile alle piccole difficoltà di organizzazione che la famiglia incontra. Il ruolo di genitore può talvolta essere vissuto con paura, ma solo in casi eccezionali comporta problemi di tale entità da minare l'equilibrio e la stabilità familiari. Anche i dati a disposizione confermano quanto detto: l'80,6% delle famiglie con almeno un bambino in età 0-5 in casa ha affermato di non avere problemi, contro il 73,7% delle famiglie senza bambini. Inoltre, i problemi più gravi sono avvertiti nel 2,4% dei casi tra le famiglie senza figli e nello 0,1% tra le famiglie con uno o più figli (Tab.4.8).

Tab.4.8. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di bambini in età 0-5 e gravità del disagio*

Componenti 0-5	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
0 (n=2167)	2,4	5,6	18,3	73,7	100,0
1 o più (n=298)	0,1	3,9	15,4	80,6	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

Come si può notare anche dalla tabella sotto riportata, vi è dipendenza tra percezione del disagio e presenza di bambini in età 0-5 in famiglia. Per le famiglie senza figli la probabilità di disagio è del 51,7% maggiore rispetto alle altre.

Confronti	Odds Ratio	Intervallo di confidenza
1 vs. 0	0,659	0,649 - 0,670

Figli in età 6-14

All'età di 6 anni il bambino inizia ad avere nuove esigenze: se da una parte si alleggerisce per la famiglia il peso richiesto per il suo accudimento, dall'altra nasce il bisogno di gestire l'inizio della sua vita sociale. Il bambino inizia ad andare a scuola, a frequentare gli amici e ad interessarsi a diverse attività extra scolastiche. Nonostante gli impegni richiesti al genitore, anche in questo caso il figlio risulta una fonte di beneficio per la famiglia. Le famiglie con almeno un componente in questa fascia d'età dichiarano meno problemi rispetto a quelle senza figli in età 6-14 (Tab.4.9).

Tab.4.9. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di bambini/ragazzi in età 6-14 e gravità del disagio*

Componenti 6-14	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
0 (n=1984)	2,6	5,8	19,2	72,4	100,0
1 o più (n=481)	0,1	3,7	13,0	83,2	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

Si nota, come nel precedente caso, il vantaggio della presenza di un figlio in età 6-14, che fa all'incirca dimezzare la probabilità di disagio.

Confronti	Odds Ratio	Intervallo di confidenza
1 vs. 0	0,521	0,514 - 0,528

Figli in età 15-19

Anche la fase adolescenziale, sebbene possa comportare per i genitori difficoltà di relazione e dialogo con i figli, non determina un aumento della percezione del disagio; al contrario, la presenza di figli in età 15-19 è, come nel caso di figli più piccoli, un fattore protettivo per la famiglia.

La tabella 4.10 riporta le distribuzioni percentuali delle famiglie venete per percezione di disagio in relazione alla presenza di adolescenti; a seguire vengono presentate le stime ottenute attraverso i modelli a odds proporzionali. Quest'ultime rilevano una situazione di indipendenza tra la presenza di componenti in età 15-19 e la percezione del disagio.

Tab.4.10. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di bambini/ragazzi in età 15-19 e gravità del disagio*

Componenti 15-19	Livello di disagio				Totale
	Grave / non risolvibile	Risolubile con aiuto esterno	Risolubile dalla famiglia	Nulla o fisiologico	
0 (n=2135)	2,2	5,4	18,0	74,4	100,0
1 o più (n=330)	1,1	5,0	17,7	76,2	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

Confronti	Odds Ratio	Intervallo di confidenza
1 vs. 0	0,891	0,878 - 0,904

Le conclusioni a cui è possibile giungere sono pressoché analoghe a quelle ottenute analizzando le famiglie con figli di età inferiore ai 15 anni: la percezione del disagio è del 12,2% maggiore nelle famiglie in cui non sono presenti giovani dai 15 ai 19 anni, rispetto a quelle con almeno un componente in questa fascia d'età.

Anziani con più di 65 anni

Se i giovani, da una parte, rappresentano una risorsa per la famiglia, gli anziani, dall'altra, possono costituire nel lungo termine una fonte di stress, soprattutto quando manifestano qualche tipo di invalidità, malattia o comunque non sono totalmente autonomi.

Il 34,6% delle famiglie intervistate è formato da componenti anziani (sopra i 65 anni d'età).

La tabella 4.11 mostra che i problemi gravi sono avvertiti in percentuale maggiore nelle famiglie in cui sono presenti almeno due anziani, mentre solo lo 0,8% delle famiglie in cui nessun componente ha più di 65 anni dichiara un forte livello di disagio. Una conferma di quanto detto si ha confrontando le percentuali delle famiglie che si dichiarano senza problemi in base alla presenza di zero, uno o più anziani: il 79,2% delle famiglie con tutti i componenti di età inferiore ai 65 anni non ha problemi, contro il 62,1% di quelle con almeno due anziani.

Tab.4.11. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di anziani con più di 65 anni e gravità del disagio*

Componenti con più di 65 anni	Livello di disagio				
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	Totale
0 (n=1586)	0,8	3,9	16,1	79,2	100,0
1 (n=514)	3,4	8,2	20,6	67,8	100,0
2 o più (n=340)	5,9	8,6	23,4	62,1	100,0
Totale (n=2440)	2,0	5,4	18,0	74,6	100,0

La dipendenza tra presenza di membri anziani e rischio di disagio è significativa. In particolare la presenza di due o più componenti di età superiore ai 65 anni comporta, rispetto alle famiglie con nessun anziano, un aumento più che doppio della probabilità di passare da un certo livello di disagio ad uno subito superiore.

Confronti	Odds Ratio	Intervallo di confidenza
1 vs. 0	1,876	1,514 - 2,323
2 vs. 0	2,419	1,900 - 3,080
2 vs. 1	1,290	0,979 - 1,699

Le famiglie composte da soli anziani, che rappresentano il 18,4% delle famiglie totali (Tab. 2.15), hanno problemi maggiori rispetto agli altri tipi di famiglie e

lamentano più frequentemente situazioni di disagio anche piuttosto grave: tra le famiglie di soli anziani, il 5,6% e il 10,6% hanno, rispettivamente, problemi gravi e risolvibili con aiuto esterno, contro l'1,3% e il 4,2% delle altre famiglie (Tab.4.12).

Tab.4.12. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per presenza di soli anziani e gravità del disagio*

Tipo di famiglia	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologica	
Famiglia di soli anziani (n=454)	5,6	10,6	18,5	65,3	100,0
Altro tipo (n=2011)	1,3	4,2	17,9	76,6	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

Come è ovvio, la probabilità di disagio è maggiore (quasi il doppio) nelle famiglie in cui tutti i componenti hanno un'età superiore ai 65 anni.

Confronti	Odds Ratio	Intervallo di confidenza
Famiglia di soli anziani vs. altro tipo di famiglia	1,871	1,516 - 2,309

Considerando le famiglie di soli anziani in relazione all'età degli stessi, si nota, come ci si può attendere, che le persone con più di 80 anni hanno più problemi di quelle con età compresa tra i 65 e i 79 anni. Ad esempio il 14,1% e il 18,4% degli over 80 hanno rispettivamente problemi gravi e risolvibili con aiuto esterno; tali percentuali si riducono a 3,4 e 9,2% nei meno anziani.

Famiglie con un solo genitore e figli a carico

Oltre alle famiglie composte da soli anziani, un altro tipo di famiglia che può essere a rischio di disagio è quella in cui è presente un solo genitore con figli a carico.

In realtà, le famiglie con un solo genitore non sono tanto più problematiche delle altre. Esse dichiarano perlopiù problemi risolvibili con le forze della famiglia: il 33,4% delle famiglie composte da un solo genitore con figli ha problemi risolvibili con le forze della famiglia, contro il 17,6% delle altre (Tab.4.13).

Tab.4.13. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per presenza o meno di un solo genitore e gravità del disagio*

Tipo di famiglia	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
Famiglia con un solo genitore (n=68)	1,9	9,3	31,0	57,8	100,0
Altro tipo (n=2397)	2,1	5,3	17,6	75,0	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

Sebbene la dipendenza tra presenza di un solo genitore e percezione del disagio risulti significativa ($OR = 2,065$), si può parlare per queste famiglie di affanno esistenziale piuttosto che di vero disagio.

Confronti	Odds Ratio	Intervallo di confidenza
Famiglia con un solo genitore vs. altro tipo di famiglia	2,065	2,010 - 2,120

Reddito familiare

Anche il reddito familiare, se non è sufficiente a garantire un certo livello di benessere alla famiglia, può essere all'origine di disagio per la stessa. La tabella 4.14 mostra come la percentuale di famiglie con problemi di una certa importanza sia più alta laddove il reddito mensile è inferiore a 1500 euro: il 14,0% delle famiglie con reddito inferiore a 1500 euro dichiara problemi, contro il 5,1% di quelle con reddito compreso tra 1500 e 3000 euro e il 4,2% di quelle con reddito superiore a 3000 euro.

Tab.4.14. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per fascia di reddito e gravità del disagio*

Reddito mensile (in euro)	Livello di disagio				Totale
	Grave non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
> 3000 (n=492)	0,5	3,7	12,4	83,4	100,0
> 1500 e ≤ 3000 (n=1017)	1,6	3,5	20,2	74,7	100,0
≤ 1500 (698)	4,6	9,7	20,6	65,1	100,0
Totale (n=2207)	2,3	5,4	18,6	73,7	100,0

La dipendenza tra livello di reddito e percezione del disagio è evidente. Come è ovvio aspettarsi, la differenza più evidente si riscontra tra le famiglie con reddito inferiore a 1500 euro e quelle con reddito superiore a 3000 euro: il rischio di disagio per una famiglia con reddito inferiore a 1500 euro è del 96,5% maggiore rispetto a quello di una famiglia con reddito di almeno 3000 euro.

Confronti	Odds Ratio	Intervallo di confidenza
> 1500 e < 3000 vs. > 3000	1,425	1,085 - 1,871
< 1500 vs. > 3000	1,965	1,472 - 2,624
< 1500 vs. > 1500 e < 3000	1,379	1,113 - 1,709

Oltre alle famiglie di anziani, quelle maggiormente esposte ad un basso livello di reddito sono le coppie con figli, in particolare *single income*, mentre le più estranee al fenomeno sono le coppie senza figli, più frequentemente *double income* (Malerba, 2001).

4.1.2 Eventi critici

Come eventi critici si sono presi in considerazione alcuni episodi avvenuti nell'ultimo anno o negli ultimi tre anni, che possono aver minato l'equilibrio interno alla famiglia e costretto la stessa a riorganizzare la propria vita quotidiana.

Lutto

Un lutto può determinare uno stravolgimento nella vita di una famiglia, tanto maggiore quanto più la persona deceduta è vicina ai componenti, perché coabitante o per motivi affettivi. Un tale evento critico può comportare seri problemi, che talvolta si risolvono nel breve periodo, talvolta, invece, si protraggono nel tempo, ad esempio quando il lutto determina difficoltà economiche per la famiglia. D'altra parte, un lutto può essere visto come momento di liberazione della famiglia da un periodo frustrante di assistenza ad una persona malata o anziana. In quest'ultimo caso, il dolore per la perdita della persona cara viene meno nel momento in cui la famiglia si risollewa e ritrova quell'equilibrio venutole a mancare durante il periodo di accudimento alla persona malata.

Il 30,5% delle famiglie venete ha subito almeno un lutto negli ultimi tre anni. In particolare, il 25,1% ne ha subito uno e il 5,4% due o più. Con riferimento al capofamiglia, i lutti presi in considerazione riguardano principalmente i coniugi/conviventi, i figli, i genitori/suoceri, i fratelli/sorelle e i cognati.

Le famiglie che hanno subito almeno un lutto nell'ultimo anno hanno affermato di avere problemi soprattutto risolvibili con le forze della famiglia stessa (17,4%), ma ben il 7,4% necessita di un aiuto dall'esterno. Tale percentuale ammonta a 4,8% e 5,1% nelle famiglie che rispettivamente hanno subito un lutto da uno a tre anni fa e non hanno subito alcun lutto (Tab.4.15).

Tab.4.15. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per lutti in famiglia e gravità del disagio*

Lutti	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologica	
Si, negli ultimi dodici mesi (n=332)	2,1	7,4	17,4	73,1	100,0
Si, da uno a tre anni fa (n=423)	2,0	4,8	18,5	74,7	100,0
No (n=1703)	2,1	5,1	18,0	74,8	100,0
Totale (n=2458)	2,1	5,4	18,0	74,5	100,0

La percezione del disagio non è avvertita in misura diversa dalle famiglie in cui è avvenuto almeno un lutto nell'ultimo anno rispetto a quelle in cui non ne è avvenuto alcuno o ne è avvenuto almeno uno da uno a tre anni fa.

Confronti	Odds Ratio	Intervallo di confidenza
Lutto da uno a tre anni fa vs. nessun lutto	1,045	0,826 - 1,323
Lutto negli ultimi dodici mesi vs. nessun lutto	1,041	0,802 - 1,350
Lutto negli ultimi dodici mesi vs. lutto da uno a tre anni fa	0,996	0,725 - 1,368

Solo qualora il lutto riguardi un componente della famiglia quale il coniuge o il figlio, il disagio che genera è avvertito in maniera pesante: il 2,0% delle famiglie

che non hanno subito un tale lutto dichiara gravi problemi, contro il 5,3% delle famiglie che hanno avuto questa perdita. Anche la percentuale di famiglie che dichiara problemi risolvibili con l'aiuto esterno o con le proprie forze è maggiore laddove viene dichiarato il lutto (Tab.4.16).

Tab.4.16. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per perdita di un familiare stretto quale il coniuge/convivente o il figlio del capofamiglia e gravità del disagio*

Perdita del coniuge/convivente o del figlio	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologica	
Si (n=76)	5,3	11,6	28,6	54,5	100,0
No (n=2382)	2,0	5,2	17,7	75,2	100,0
Totale (n=2458)	2,1	5,4	18,0	74,5	100,0

In questo caso particolare, la stima dell'odds ratio conferma quanto detto sopra: la perdita del coniuge o di un figlio determina un aumento significativo della probabilità di passare da un certo livello di disagio ad uno immediatamente superiore (aumento di oltre due volte e mezzo).

Confronti	Odds Ratio	Intervallo di confidenza
Perdita del coniuge o del figlio vs. altro lutto o nessun lutto	2,522	1,627 - 3,910

Furto

Un altro possibile evento critico, che riguarda il 6,5% delle famiglie, è il furto. In particolare, il 2,5% l'ha subito nell'ultimo anno, mentre il 4,0% da più di un anno, ma da meno di tre.

Nonostante le basse numerosità, si nota che le famiglie che hanno subito un furto necessitano di maggior aiuto esterno e avvertono maggiori problemi risolvibili con le loro forze rispetto a quelle che non l'hanno subito (Tab.4.17).

Tab.4.17. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per furti subiti nell'ultimo anno e gravità del disagio*

Furto	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
Si (n=57)	2,0	7,2	25,9	64,9	100,0
No (n=2407)	2,1	5,3	17,8	74,8	100,0
Totale (n=2464)	2,1	5,4	17,9	74,6	100,0

Nonostante questo, la stima ottenuta attraverso l'applicazione dei modelli a odds proporzionali mette in evidenza che il rischio di disagio non differisce tra le famiglie che hanno subito un furto e quelle che non l'hanno subito.

Confronti	Odds Ratio	Intervallo di confidenza
Furto vs. non furto	1,550	0,907 - 2,646

Incidenti

La famiglia può trovarsi in difficoltà anche quando si verificano incidenti di una certa importanza: il 7,1% delle famiglie ha subito un incidente nell'ultimo anno. Le famiglie in cui un componente ha subito un incidente sono più problematiche delle altre, ma si tratta principalmente di problemi risolvibili con le proprie forze (Tab.4.18).

Tab.4.18. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per incidenti subiti nell'ultimo anno e gravità del disagio*

Incidenti	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
Si (n=324)	1,7	6,6	28,5	63,2	100,0
No (n=2138)	2,1	5,2	16,4	76,3	100,0
Totale (n=2462)	2,1	5,4	18,0	74,5	100,0

La probabilità di passare da un certo livello di disagio ad uno immediatamente superiore è del 61,4% maggiore quando un componente subisce un incidente.

Confronti	Odds Ratio	Intervallo di confidenza
Incidente vs. non incidente	1,614	1,267 - 2,056

Aiuto economico per la famiglia

Come ultimo evento critico si è considerata la necessità di aiuto economico da parte della famiglia nell'ultimo anno.

La tabella 4.19 mostra che il disagio, soprattutto quello più grave e quello risolvibile con aiuto esterno, è decisamente maggiore nelle famiglie che hanno avuto bisogno di aiuti economici nel corso dell'anno: l'8,5% e il 20,7% delle famiglie che hanno chiesto aiuti dichiarano rispettivamente disagio grave e risolvibile con aiuto esterno. Tali percentuali si riducono a 1,6% e 4,4% nelle altre famiglie.

Tab.4.19. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per necessità di aiuto economico nell'ultimo anno e gravità del disagio*

Aiuto economico	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologico	
Si (n=146)	8,5	20,7	27,7	43,1	100,0
No (n=2319)	1,6	4,4	17,3	76,7	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

La dipendenza tra percezione del disagio e necessità di aiuto economico da parte della famiglia è significativa. Il rischio di passare da un certo livello di disagio ad uno superiore nelle prime famiglie aumenta di oltre cinque volte.

Confronti	Odds Ratio	Intervallo di confidenza
Necessità di aiuto economico vs. mancanza di bisogno economico	5,336	3,893 - 7,314

4.1.3 Influenza di più fattori nella determinazione del disagio

Le ricerche svolte in ambito psicosociale (McCubbin, Dahl e coll., 1975; McCubbin e coll., 1976; Muton, 1990) hanno mostrato che non è tanto la presenza di ogni singolo fattore, quanto la combinazione di più situazioni difficili a determinare il venir meno dell'equilibrio familiare e a provocare le crisi più difficili da gestire per la famiglia. Pertanto, si è deciso di sommare in ogni famiglia il numero delle situazioni particolari prese in considerazione precedentemente.

Il 71,5% delle famiglie venete non presenta alcun evento o stato critico. All'aumentare del numero di tali eventi le percentuali vanno via via diminuendo, essendo più difficile trovare famiglie con più di un problema importante: il 15,8% ne presenta uno, l'8,4% due e il 4,3% tre o più.

Come si può notare dalla tabella 4.20, all'aumentare del numero di eventi o stati critici aumenta in maniera evidente la percezione del disagio da parte della famiglia.

I problemi più gravi sono avvertiti perlopiù dalle famiglie che contano tre o più eventi critici (24,9% contro il 9,7% delle famiglie con due eventi critici, l'8,4% delle famiglie con un evento critico e il 4,8% di quelle con nessun evento critico). Dalla distribuzione delle famiglie che dichiarano disagio nullo o fisiologico si giunge alla medesima conclusione: tra le famiglie colpite da almeno tre eventi critici solo il 41,7% non si ritiene problematica, mentre tra quelle che non dichiarano alcun evento critico tale percentuale ammonta a 80,8%.

Tab.4.20. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di eventi e stati critici e gravità del disagio*

Numero di eventi critici	Livello di disagio				Totale
	Grave / non risolvibile	Risolvibile con aiuto esterno	Risolvibile dalla famiglia	Nulla o fisiologica	
0 (n=1200)	1,4	3,4	14,4	80,8	100,0
1 (n=675)	1,4	7,0	19,3	72,3	100,0
2 (n=370)	3,0	6,7	25,2	65,1	100,0
3 o più (n=220)	9,0	15,9	33,4	41,7	100,0
Totale (n=2465)	2,1	5,4	17,9	74,6	100,0

Considerando la distribuzione delle sole famiglie che hanno dichiarato tre o più eventi critici, il 9,0% avverte un livello di disagio grave, il 15,9% risolvibile con l'aiuto proveniente dall'esterno, il 33,4% risolvibile con le forze della famiglia ed infine il 41,7% nulla o fisiologica. In particolare, si sono individuati gli eventi critici e le caratteristiche strutturali che caratterizzano queste famiglie al fine di capire quali influiscono maggiormente sulla percezione del disagio.

Il 23,1% delle famiglie con tre o più eventi critici che hanno dichiarato di non avere problemi ha un reddito inferiore a 1500 euro mensili, mentre il 21,6% è formata da componenti la cui età media è superiore o uguale a 70 anni. Entrambe

le categorie di famiglia non necessitano di aiuti da parte dell'esterno, ma ritengono il problema gestibile al loro interno.

La maggior parte delle famiglie con tre o più eventi critici è formata da individui la cui età media è almeno pari a 70 anni e ha invalidi non autonomi al suo interno. Di queste famiglie, il 10,3% si ritiene colpito da gravi problemi, spesso non risolvibili, il 29,4% ha bisogno di un aiuto esterno alla famiglia, il 21,6% ha problemi a cui riesce a far fronte con le proprie forze e solo il 38,7% si ritiene privo di problemi.

Altrettanto importante è la quota di famiglie con componenti di età media superiore a 70 anni, con malati cronico-degenerativi non autonomi e reddito mensile inferiore a 1500 euro. Di queste, ben il 20,5% dichiara problemi gravi, mentre il 44,0% non si ritiene colpito da problemi.

Infine, anche le famiglie con componenti invalidi non autonomi e basso reddito costituiscono una quota importante all'interno della categoria di famiglie colpite da almeno tre eventi critici. Solo il 38,0% di queste non ha problemi mentre ben il 39,8% dichiara problemi non risolvibili (7,9%) o risolvibili ricorrendo all'aiuto esterno alla famiglia (31,9%).

Nella tabella 4.21 si presenta la percentuale di disagio nelle famiglie con zero, uno, due, tre o più eventi o stati critici. In questo caso il disagio riflette sia i problemi gravi non risolvibili, sia quelli risolvibili con aiuto esterno alla famiglia.

Tab.4.21. *Percentuale della popolazione residente nel Veneto nel 2004-2005 che ha dichiarato disagio per numero di eventi e stati critici*

Eventi critici	%
0 (n=22)	1,8
1 (n=45)	6,5
2 (n=49)	13,5
3 o più (n=73)	32,9

La percezione del disagio dipende in maniera significativa dal numero di eventi critici. La differenza più evidente la si ottiene, come ci si può attendere, mettendo a confronto le famiglie con 3 o più e quelle con nessun evento critico; in questo

caso il rischio di passare da un livello di disagio ad uno immediatamente superiore aumenta di oltre sette volte.

Confronti	Odds Ratio	Intervallo di confidenza
1 vs. 0	1,792	1,433 - 2,241
2 vs. 0	2,418	1,857 - 3,149
3 o più vs. 0	7,360	5,250 - 10,318
2 vs. 1	1,350	1,008 - 1,807
3 o più vs. 1	4,108	2,866 - 5,888
3 o più vs. 2	3,044	2,068 - 4,479

4.2 Il rischio di disagio familiare e le sue determinanti

In questo paragrafo ci si propone di individuare le determinanti del disagio, misurato su una scala ordinale a quattro livelli: assenza di gravi problemi, problemi risolvibili con le forze della famiglia, problemi risolvibili con aiuto esterno, gravi problemi non risolvibili. In questo modo sarà possibile intervenire in maniera quasi personalizzata a sostegno delle famiglie che presentano le tipologie più frequenti ed evidenti di disagio.

Nel panorama della letteratura nazionale ed internazionale è facile trovare ricerche focalizzate su particolari aspetti, anche individuali, che possono tradursi in forme di disagio per la famiglia. Primi fra tutti si posizionano quelli relativi allo studio delle famiglie di fronte alla cura di disabili o comunque componenti non autonomi (Valtolina, 2002). Altri argomenti che riscontrano particolare interesse in letteratura riguardano i problemi di lavoro (disoccupazione, inoccupazione) (Accornero e Carmignani, 1986), di carattere organizzativo e soprattutto strutturale (famiglie composte da un solo genitore, che spesso trova difficoltà nella gestione dei figli) (Zanatta, 1997).

In molti casi tali studi sono condotti con metodi di analisi dei rischi volti a ricercare le determinanti del disagio. Endrizzi, Fabbris e Schievano (2002) studiano il disagio familiare mediante analisi di segmentazione binaria con riferimento a due livelli di disagio, uno “grave”, per il quale è necessario ricorrere ai servizi, e uno

comprensivo anche del disagio risolvibile all'interno della famiglia. Mastrovita e Siciliani (2002) esaminano il rischio di permanenza in una situazione di disagio economico applicando una procedura di segmentazione ternaria, ricavando che le determinanti del disagio sono molteplici e agiscono in interazione.

La realizzazione del nostro obiettivo passa attraverso la rilevazione di gruppi di famiglie a maggior rischio di disagio, concentrando l'interesse sui rischi che riguardano i caratteri strutturali e prevedibili delle stesse.

Per la determinazione di tali gruppi omogenei si è adottato il metodo di analisi della segmentazione binaria. Si tratta di una procedura *stepwise* che consente, passo dopo passo, la segmentazione binaria di un campione rispetto alla combinazione delle modalità di una variabile esplicativa. In particolare, nel nostro caso la variabile criterio è una trasformata del disagio in un odds proporzionale (McCullagh, 1980). Le modalità della variabile esplicativa sono combinate in modo tale che il rapporto tra gli odds calcolati per i due gruppi risultanti sia massimo. Indicati con \mathbf{x}_1 e \mathbf{x}_2 i vettori di covariate dei gruppi 1 e 2, il rapporto tra gli odds è:

$$OR = \frac{\gamma_k(\mathbf{x}_1)/(1 - \gamma_k(\mathbf{x}_1))}{\gamma_k(\mathbf{x}_2)/(1 - \gamma_k(\mathbf{x}_2))} = \exp \{ -\boldsymbol{\beta}^T (\mathbf{x}_1 - \mathbf{x}_2) \}$$

dove $\gamma_k(\mathbf{x}) = \Pr(Y \leq k | \mathbf{x})$, $k = 1, \dots, K - 1$, è la probabilità cumulata fino alla categoria k e Y è la variabile dipendente (ordinale) a K modalità (McCullagh e Nelder, 1989).

Essendo l'unità di analisi la famiglia, la segmentazione è stata condotta sintetizzando le variabili predittive, relative ai singoli componenti, in variabili riassuntive della famiglia stessa. Per esempio, le eventuali malattie cronico-degenerative relative ai componenti di ciascuna famiglia sono state riassunte nella variabile numero di componenti affetti da malattie cronico-degenerative.

Nella tabella 4.22 si riportano le variabili utilizzate nell'analisi di segmentazione, l'etichetta e le relative modalità.

Tab.4.22. Variabili utilizzate nell'analisi di segmentazione

Variabile	Modalità
Numero componenti	1 = 1 componente 2 = 2 componenti 3 = 3 o più componenti
Età media dei componenti	1 = < 30 anni 2 = ≥ 30 e < 50 anni 3 = ≥ 50 e < 70 anni 4 = ≥ 70 anni
Stato civile del capofamiglia	1 = coniugato 2 = celibe/nubile 3 = separato/divorziato 4 = vedovo
Massimo titolo di studio in famiglia	0 = basso (diploma di scuola media superiore - 2/3 anni- o inferiore) 1 = alto (diploma di scuola media superiore - 4/5 anni- o superiore)
Componenti disoccupati	0 = 0 disoccupati 1 = 1 o più disoccupati
Tipo di famiglia	1 = unipersonale 2 = coppia senza figli 3 = coppia con figli 4 = un genitore con figli a carico 5 = un genitore con almeno un figlio occupato 6 = altro
Titolo a cui la famiglia occupa l'abitazione	1 = proprietà od usufrutto 2 = affitto o subaffitto 3 = altro (gratuito o prestazione di servizio)
Numero di automobili	0 = 0 automobili 1 = 1 automobile 2 = 2 automobili 3 = 3 o più automobili
Pasti condivisi dai componenti	1 = i componenti pranzano insieme anche nei giorni feriali 2 = pranzano insieme solo nei fini settimana 3 = pranzano insieme sporadicamente 4 = non pranzano mai insieme
Numero di invalidi	0 = 0 invalidi 1 = 1 o più invalidi
Autonomia dell'invalido	0 = no 1 = si
Numero di malati cronico-degenerativi	1 = al più 1 2 = 2 o più
Autonomia del malato cronico-degenerativo	0 = no 1 = si
Possibilità per la famiglia di aiuto da parte di parenti	0 = no 1 = si
Necessità di aiuti economici per la famiglia nell'ultimo anno	0 = no 1 = si

Lutti in famiglia	0 = no 1 = si
Separazione/divorzio negli ultimi 3 anni	0 = no 1 = si
Incidenti negli ultimi 3 anni	0 = no 1 = si
Furti negli ultimi 3 anni	0 = no 1 = si
Eventuali problemi in famiglia	1 = sanitari 2 = economici 3 = socio-relazionali 4 = altro
Livello di reddito mensile familiare	1 => 6000 € 2 =<= 6000 e > 3000 € 3 =<= 3000 e > 1500 € 4 =<= 1500 €

Per la scelta delle variabili da utilizzare nel modello ci si è, innanzitutto basati sui suggerimenti forniti dalla teoria (Endrizzi, Fabbris e Martini, 2002; Endrizzi, Fabbris e Schievano, 2002; Breda e Fabbris, 2002; Blangiardo, 2002) e sulle considerazioni risultanti dall'analisi svolta precedentemente in questo capitolo.

Tre risultano le tipologie di variabili prese come riferimento:

1. *strutturali*: non durature, osservabili nel tempo (numero di componenti, tipo di famiglia);
2. *conoscibili*: anagrafiche o ricostruibili con l'analisi (età, stato civile del capofamiglia, massimo titolo di studio in famiglia, numero di invalidi e loro autonomia, numero di malati cronico-degenerativi e loro autonomia, lutti, separazioni, incidenti, furti, tipologie di problemi presenti in famiglia);
3. *transitorie*: mutevoli nel tempo (disoccupazione, titolo a cui la famiglia occupa l'abitazione, numero di automobili possedute, pasti, possibilità di aiuto da parenti, necessità di aiuti economici, reddito dichiarato).

Variabili psicologiche che rilevano sensazioni e giudizi non vengono messe in evidenza da nessuna analisi dal momento che una variabile rilevata in un preciso momento non può essere esplicativa del disagio.

Un problema che si è presentato è stato quello dell'elevato numero di dati mancanti. In molti casi le variabili, risposta o esplicative, con valori mancanti vengono escluse dall'analisi con la conseguente perdita di informazione e possibilità di distorsione. Alternativamente, i valori mancanti possono essere sostituiti da statistiche riassuntive, il valore medio dei dati per cui si dispone della risposta ad esempio, ma anche questa operazione può tradursi in stime distorte. Nel nostro caso, per trattare le mancate risposte si è sfruttato il criterio della modalità fluttuante, la quale viene esclusa dal calcolo delle statistiche per la definizione della miglior partizione e, una volta determinati i sottogruppi ottimali, assegnata prima ad uno e poi all'altro e il valore della statistica ricalcolato. La modalità si associa al gruppo di modalità per il quale la differenza tra il valore della statistica ricalcolata e quello della statistica iniziale è minima.

Per ottimizzare la formazione dell'albero si sono poste restrizioni al campo d'azione del processo, utilizzando particolari metodi che aumentassero le potenzialità predittive della struttura, ne riducessero la complessità e permettessero di rilevare le relazioni tra le variabili.

La procedura di segmentazione si arresta per i nodi la cui numerosità risulta inferiore alle 20 unità. Se almeno uno dei due sottogruppi in cui viene diviso un nodo ha numerosità inferiore alla dimensione minima imposta, il processo verrà arrestato al passo precedente, quando tutti i nodi hanno numerosità superiore o uguale a quella fissata e sono in grado di fornire stime attendibili delle statistiche.

Il valore di significatività che deve raggiungere il test che valuta la bontà della partizione prima che questa possa essere applicata è stato posto pari a 0,15. Questo fa sì che l'analisi non possa proseguire per i gruppi la cui divisione su un qualsiasi predittore non crei dei nodi figli sufficientemente omogenei. Sotto tale soglia, infatti, un gruppo genitore è considerato compatto e omogeneo e, quindi, non divisibile.

Infine, per limitare la dimensione dell'albero, si è posto a 10 il numero massimo di nodi terminali. Raggiunto tale limite, l'analisi viene terminata su tutta la struttura fino a quel punto generata. Il numero massimo di foglie che può raggiungere l'albero di classificazione corrisponde al numero massimo di suddivisioni che si è

interessati ad ottenere, meno uno.

Nella figura 4.1 si riportano i risultati dell'analisi di segmentazione. Il livello di disagio associato a ciascun nodo è determinato dalla distanza tra la distribuzione di frequenza di ogni gruppo e quella del gruppo genitore.

Le variabili che hanno concorso alla segmentazione del campione sono:

- *il bisogno di aiuti economici.* Questo bisogno è avvertito dal 5,9% delle famiglie venete. Le difficoltà economiche dipendono dalla mancanza di lavoro e dalle basse pensioni. Queste famiglie sono colpite da un livello estremo di disagio nell'8,5% dei casi e da un disagio che richiede l'intervento dei servizi pubblici nel 20,7% ($OR = 11,402$).
- *la presenza di invalidi, anche nelle famiglie economicamente autonome.* Il rischio di disagio aumenta di oltre nove volte se almeno un componente della famiglia è affetto da invalidità: $OR = 9,292$ in queste famiglie rispetto a quelle in cui tutti i componenti sono sani. Solo il 30,5% delle famiglie con invalidi non avverte un disagio persistente (nel gruppo in cui non sono presenti invalidi il disagio riguarda il 4,3% delle famiglie, ossia un po' più della metà del disagio del campione complessivo, pari a 7,5%);
- *la presenza di componenti non autosufficienti affetti da malattie cronico-degenerative nelle famiglie con reddito inferiore a 3000 euro* ($OR = 4,671$). Questo tipo di difficoltà determina un alto rischio di grave disagio (19,4% contro lo 0,4% delle famiglie con componenti autonomi). Anche il bisogno di servizi pubblici e il ricorso al volontariato aumentano in presenza di membri non autosufficienti (10,0% rispetto al 3,3% nel caso di componenti autonomi);
- *un'anomala struttura familiare.* In particolare, ad alto rischio di disagio si trovano le famiglie formate da un solo genitore con figli a carico, sebbene un adulto professionalmente occupato riesca generalmente a far fronte alle

Legenda

- p1: % nessun problema
 p2: % problemi risolvibili con le forze della famiglia
 p3: % problemi risolvibili con aiuto esterno
 p4: % problemi non risolvibili

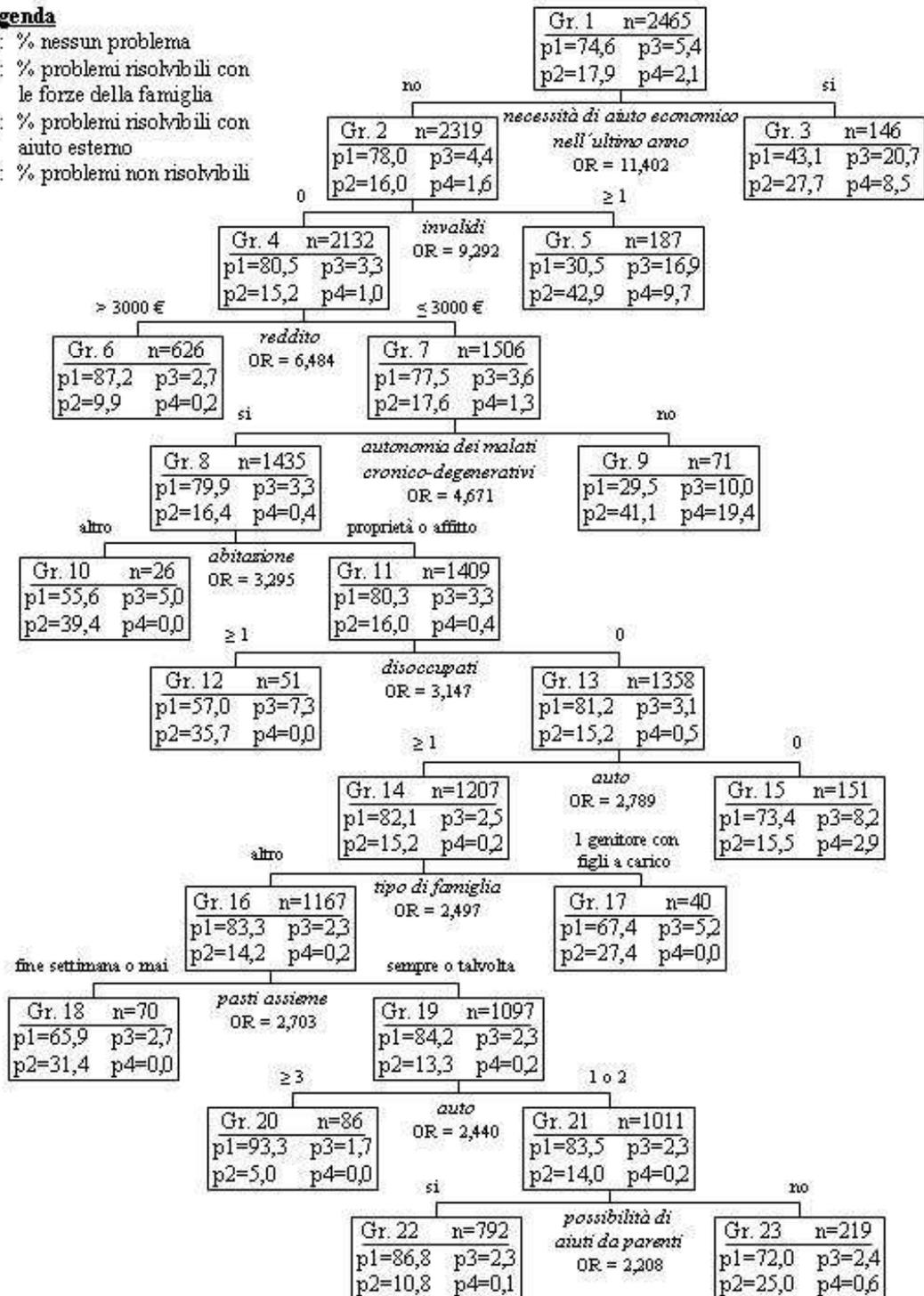


Figura 4.1: Analisi di segmentazione binaria relativa al “rischio di disagio” (variabile politomica) nelle famiglie venete.

difficoltà con le proprie risorse o ricorrendo all'aiuto della famiglia "allargata" e dei servizi pubblici (32,6% di disagio vs. 16,7% nelle altre famiglie);

- *la mancanza di automobili.* Nelle famiglie economicamente autonome, senza invalidi e componenti non autosufficienti, reddito inferiore a 3000 euro mensili netti e senza componenti disoccupati (i.e. nelle comuni famiglie con componenti anziani), il rischio di disagio è elevato qualora queste non possiedano nessuna automobile: 11,1% vs. 2,7% ($OR = 2,789$) di quelle che ne possiedono almeno una. Si vedrà in seguito che non è in sé la mancanza di automobili a determinare il disagio: nel gruppo in questione entrano a far parte famiglie, soprattutto unipersonali con redditi da pensione e problemi, oltre che economici, di tipo sanitario. La variabile che rileva la presenza di automobili è, pertanto, *proxy* dell'autonomia dei componenti e della loro età. Le famiglie che possiedono tale bene sono quelle in cui i componenti ne fanno uso e da cui è possibile anche valutare la loro autonomia; di contro, le altre sono quelle in cui i componenti non possono economicamente permettersi tale bene o, più spesso, non sono in grado di farne uso per motivi di età e di salute.

Nelle tabelle 4.23 - 4.34 si vogliono mettere in risalto le tipologie di problemi maggiormente frequenti nella totalità delle famiglie venete intervistate (Tab.4.23) e nei gruppi a rischio sopra individuati (Tab.4.24 - 4.34).

Il 12,8% delle famiglie venete ha problemi di salute, il 12,0% economici. Solo una piccola percentuale (4,0%) avverte problemi socio-relazionali (Tab.4.23).

Tab.4.23. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=345)	12,8
Economici (n=312)	12,0
Socio-relazionali (n=103)	4,0
Altro (n=35)	1,4

Come si nota dalla figura 4.1, la prima variabile che entra nel processo di segmentazione, con l'obiettivo di evidenziare i gruppi di famiglie a maggior rischio

di disagio, è la necessità di aiuti economici nell'ultimo anno. Il gruppo di famiglie che hanno avuto tale necessità (Gr.3), come ci si può attendere, si caratterizza per il fatto di avere un'elevatissima percentuale (43,3%) di problemi di tipo economico. I problemi di salute, nonostante siano meno frequenti, sono presenti nel 21,1% dei casi. Si noti che considerando la totalità delle famiglie, questo tipo di problema è presente con una percentuale minore, e precisamente pari a 12,8%. Questo aspetto sta ad indicare che diversi problemi spesso si presentano contemporaneamente nella stessa famiglia: una famiglia povera ha più difficoltà ad affrontare spese sanitarie rispetto ad una benestante, nella quale i problemi di salute, se non di estrema gravità, si possono affrontare con una relativamente maggiore facilità (Tab.4.24).

Tab.4.24. *Distribuzione percentuale delle famiglie del gruppo 3 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=30)	21,1
Economici (n=70)	43,3
Socio-relazionali (n=9)	6,1
Altro (n=2)	1,6

Tra le famiglie che non hanno avuto bisogno di aiuti economici nell'ultimo anno e che hanno almeno un componente invalido al loro interno (Gr.5), ben oltre la metà (67,2%) sono affette da problemi di tipo sanitario: l'invalidità implica una condizione di disagio importante per la famiglia, la quale solo nel 43,1% dei casi non si ritiene colpita da problemi. Per questo gruppo, molto meno frequenti, rispetto a quelli sanitari, sono i problemi economici, che coinvolgono il 13,5% delle famiglie (Tab.4.25).

Tab.4.25. *Distribuzione percentuale delle famiglie del gruppo 5 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=129)	67,2
Economici (n=27)	13,5
Socio-relazionali (n=4)	2,3
Altro (n=1)	0,7

L'11,5% delle famiglie che non ha avuto recente bisogno di aiuti economici, che non conta individui affetti da invalidità al proprio interno e che ha un reddito

inferiore a 3000 euro mensili netti (Gr.7) avverte problemi di tipo economico. Non si tratta di una percentuale rilevante, se confrontata con gli altri gruppi a rischio; queste famiglie, infatti, hanno un reddito mensile piuttosto basso, ma abbastanza sufficiente da non richiedere aiuti esterni (Tab.4.26).

Tab.4.26. *Distribuzione percentuale delle famiglie del gruppo 7 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=161)	9,5
Economici (n=180)	11,5
Socio-relazionali (n=64)	4,0
Altro (n=26)	1,6

In presenza di malati cronico-degenerativi non autonomi (Gr.9), la percentuale di famiglie colpite da problemi di salute è molto alta (65,6%). Molto meno importanza assumono i problemi di tipo economico, che toccano l'11,2% delle famiglie con reddito inferiore a 3000 euro mensili netti e malati non autonomi al loro interno (Tab.4.27).

Tab.4.27. *Distribuzione percentuale delle famiglie del gruppo 9 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=46)	65,6
Economici (n=9)	11,2
Socio-relazionali (n=3)	4,1
Altro (n=0)	0,0

Ceteris paribus, altre caratteristiche che rendono problematica una famiglia sono la presenza di disoccupati (Gr.12) e l'essere la famiglia composta da un genitore con figli a carico (Gr.17). Nel primo caso i problemi maggiormente avvertiti sono di tipo economico e coinvolgono il 40,0% delle famiglie (Tab.4.28), nel secondo assumono molto più importanza quelli di tipo sociale-relazionale (23,6%) (Tab.4.29)

Tab.4.28. *Distribuzione percentuale delle famiglie del gruppo 12 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=4)	6,4
Economici (n=20)	40,0
Socio-relazionali (n=2)	1,4
Altro (n=1)	0,7

Tab.4.29. *Distribuzione percentuale delle famiglie del gruppo 17 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=3)	5,0
Economici (n=9)	21,4
Socio-relazionali (n=9)	23,6
Altro (n=2)	1,9

Nelle tabelle 4.30-4.34 si riportano le percentuali di problemi avvertiti nelle famiglie facenti parte dei restanti gruppi a rischio (Gr.10-15-18-21-23).

Tab.4.30. *Distribuzione percentuale delle famiglie del gruppo 10 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=5)	21,8
Economici (n=6)	16,0
Socio-relazionali (n=3)	11,6
Altro (n=2)	6,7

Tab.4.31. *Distribuzione percentuale delle famiglie del gruppo 15 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=29)	16,7
Economici (n=16)	9,6
Socio-relazionali (n=4)	2,3
Altro (n=2)	1,7

Tab.4.32. *Distribuzione percentuale delle famiglie del gruppo 18 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=6)	6,1
Economici (n=13)	22,0
Socio-relazionali (n=6)	9,9
Altro (n=1)	1,8

Tab.4.33. *Distribuzione percentuale delle famiglie del gruppo 21 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=67)	5,6
Economici (n=99)	9,3
Socio-relazionali (n=37)	3,4
Altro (n=17)	1,6

Tab.4.34. *Distribuzione percentuale delle famiglie del gruppo 23 residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=22)	7,5
Economici (n=38)	17,9
Socio-relazionali (n=15)	7,1
Altro (n=6)	2,9

Oltre ad essere caratterizzati da problemi, i gruppi a rischio hanno al loro interno anche risorse a cui possono fare riferimento in caso di necessità: la definizione dell'evento "stressante" e della crisi porta con sé, necessariamente, la definizione delle strategie di cui la famiglia fa uso per far fronte alla situazione creatasi. Le risorse di cui la famiglia può immediatamente usufruire possono essere di tipo materiale (reddito, livello di istruzione) o relazionale (caratteristiche del funzionamento familiare e rete di supporto parentale e sociale in generale).

Si è studiata la distribuzione delle famiglie appartenenti a tali gruppi rispetto alle seguenti quattro variabili: *età media dei componenti, tipo di famiglia, possibilità di aiuto da parte di parenti e livello di reddito mensile* (Tab.4. 36 - 4.38 - 4.40 - 4.42).

L'essere la famiglia composta da individui in giovane età può comportare alcune difficoltà, soprattutto di carattere economico, ma molto spesso rappresenta un punto di forza: le famiglie giovani affrontano generalmente i problemi con maggior serenità e ottimismo, facendo appello alle forze di cui dispongono e alla speranza di vita ancora lunga. Il 4,0% delle famiglie con componenti di età media inferiore a 30 anni e il 5,3% di quelle con componenti tra i 30 e i 50 anni si ritiene colpito da problemi risolvibili con l'intervento dei servizi o da gravi problemi non risolvibili. Nelle famiglie meno giovani queste percentuali aumentano in maniera considerevole: 7,0% per le famiglie con età media tra i 50 e i 70 e 18,3% per quelle con età media superiore ai 70 anni (Tab.4.35)

Tab.4.35. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per percezione di disagio e età media dei componenti*

Livello di disagio	Età				Totale (n=2465)
	< 30 (n=474)	≥ 30 e < 50 (n=961)	≥ 50 e < 70 (n=642)	≥ 70 (n=388)	
Nulla	79,6	79,0	71,0	62,6	74,6
Lieve	16,4	15,7	22,0	19,1	17,9
Grave	4,0	4,5	4,4	11,2	5,4
Estremo	0,0	0,8	2,6	7,1	2,1
Totale	100,0	100,0	100,0	100,0	100,0

Il 59,4% delle famiglie venete (Gr.1) è composto da individui con un'età media inferiore a 50 anni. I gruppi a rischio individuati dall'analisi di segmentazione (Fig.4.1) sono caratterizzati in alcuni casi da famiglie meno giovani (Gr.5-9-15), in altri da famiglie più giovani (Gr.3-12-17-18) rispetto alla totalità delle famiglie venete campionate. In particolare, l'82,5% delle famiglie che hanno reddito inferiore a 3000 euro mensili netti, ma non hanno avuto bisogno di aiuti economici nell'ultimo anno, che non contano invalidi al proprio interno e i malati cronico-degenerativi sono autonomi, la cui abitazione è di proprietà o in affitto e in cui è presente almeno un componente disoccupato (Gr.12) ha un'età media inferiore a 50 anni. Tali percentuali ammontano a 82,9% per le famiglie con le stesse caratteristiche di quelle del gruppo 12, e che inoltre si distinguono per il fatto di possedere almeno un'automobile e per l'essere composte da un genitore con figli a carico (Gr.17) e a 73,9% per quelle che non pranzano mai assieme o lo fanno solo

nei fine settimana (Gr.18). Infine, tra le famiglie che hanno avuto recente bisogno di sostegno economico (Gr.3), la percentuale di quelle con età media inferiore a 50 anni è pari a 71,0%. Si noti che la giovane età delle famiglie del gruppo 17 è dovuta alla composizione della famiglia stessa (un genitore con figli a carico): l'età del figlio determina una bassa età media e, in questo particolare caso, ciò non rappresenta una risorsa, bensì una difficoltà in più che la famiglia si trova ad affrontare. Anche per le famiglie appartenenti al gruppo 3 la giovane età non è una risorsa; questo gruppo è infatti caratterizzato dalle famiglie appena formatesi, che non hanno ancora una loro autonomia economica (Tab.4.36).

Tab.4.36. *Distribuzione percentuale delle famiglie nei diversi gruppi a rischio per età media dei componenti*

Età	Gr.1		Gr.3		Gr.5		Gr.7	
	n	%	n	%	n	%	n	%
< 30	474	19,9	42	28,7	10	6,6	301	20,6
≥ 30 e < 50	961	39,5	62	42,3	45	25,1	506	34,3
≥ 50 e < 70	642	25,6	27	17,9	69	38,0	423	27,6
≥ 70	388	15,0	15	11,1	63	30,3	276	17,5
Totale	2465	100,0	146	100,0	187	100,0	1506	100,0

Gr.9		Gr.10		Gr.12		Gr.15	
n	%	n	%	n	%	n	%
2	2,3	12	43,2	11	21,9	4	3,4
7	9,8	4	13,0	30	60,6	8	6,3
22	29,2	4	20,2	10	17,5	41	29,4
40	58,7	6	23,6	0	0,0	98	60,9
71	100,0	26	100,0	51	100,0	151	100,0

Gr.17		Gr.18		Gr.21		Gr.23	
n	%	n	%	n	%	n	%
17	44,0	17	27,6	234	23,6	30	14,8
14	38,9	35	46,3	340	34,6	97	45,2
6	12,1	17	25,6	309	29,6	71	31,8
3	5,0	1	0,5	128	12,2	21	8,2
40	100,0	70	100,0	1011	100,0	219	100,0

Le famiglie di tipo coppia con figli sono risultate le più “forti” dal punto di vista del disagio percepito: il 94,9% di queste famiglie non avverte alcun tipo di problema o ha problemi risolvibili al proprio interno. Più problematiche sono,

invece, le famiglie unipersonali e quelle composte da un genitore con figli a carico: ben il 14,0% delle prime e l'11,2% delle seconde si ritiene colpito da problemi risolvibili con l'aiuto esterno o da gravi problemi non risolvibili (Tab.4.37).

Tab.4.37. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per percezione di disagio e tipo di famiglia*

livello di disagio	Tipo di famiglia						Totale (n=2465)
	Unipersonale (n=352)	Coppia senza figli (n=549)	Coppia con figli (n=1313)	Un genitore con figli a carico (n=68)	Un genitore con almeno un figlio occupato (n=99)	Altro (n=39)	
Nulla	70,9	73,5	78,5	57,8	60,9	51,2	74,6
Lieve	15,1	18,6	16,4	31,0	31,9	30,5	17,9
Grave	8,9	4,6	4,2	9,3	7,2	15,0	5,4
Estremo	5,1	3,3	0,9	1,9	0,0	3,3	2,1
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Tra le famiglie venete (Gr.1), oltre la metà (54,1%) è di tipo coppia con figli. Solo in due degli undici gruppi a rischio individuati dall'analisi di segmentazione, tale percentuale è più elevata: nelle famiglie che hanno reddito inferiore a 3000 euro mensili netti, ma non hanno avuto bisogno di aiuti economici nell'ultimo anno, che non contano invalidi al proprio interno e i malati cronico-degenerativi sono autonomi, la cui abitazione è di proprietà o in affitto e in cui è presente almeno un componente disoccupato (Gr.12) la percentuale di coppie con figli ammonta a 77,9%; se, inoltre, queste famiglie non pranzano mai assieme o lo fanno solo nei fine settimana tale percentuale aumenta fino a raggiungere l'76,7% (Gr.18). In quest'ultimo caso, tuttavia, non sono considerate le famiglie composte da un genitore con figli a carico, che sono andate a costituire un gruppo a se stante (Tab.4.38).

Tab.4.38. *Distribuzione percentuale delle famiglie nei diversi gruppi a rischio per tipo di famiglia*

Tipo di famiglia	Gr.1		Gr.3		Gr.5		Gr.7	
	n	%	n	%	n	%	n	%
Unipersonale	352	13,8	30	22,7	24	11,6	271	17,3
Coppia senza figli	594	24,0	20	13,9	50	26,7	393	26,2
Coppia con figli	1313	54,1	76	50,5	78	43,6	724	48,7
Un genitore con figli a carico	68	2,6	7	4,5	9	3,9	50	3,3
Un genitore con almeno un figlio occupato	99	4,0	6	4,4	15	8,8	58	3,7
Altro	39	1,5	7	4,0	11	5,4	10	0,8
Totale	2465	100,0	146	100,0	187	100,0	1506	100,0

Gr.9		Gr.10		Gr.12		Gr.15	
n	%	n	%	n	%	n	%
17	21,6	6	23,6	1	0,8	109	66,8
32	46,9	6	26,8	5	10,3	27	20,9
17	22,6	13	48,1	39	77,9	7	6,0
2	3,6	1	1,5	5	8,6	2	1,7
2	3,5	0	0,0	1	2,4	5	3,7
1	1,8	0	0,0	0	0,0	1	0,9
71	100,0	26	100,0	51	100,0	151	100,0

Gr.17		Gr.18		Gr.21		Gr.23	
n	%	n	%	n	%	n	%
-	-	0	0,0	138	13,9	25	12,7
-	-	9	10,5	312	30,3	63	27,9
-	-	52	76,7	517	51,8	118	53,9
40	100,0	-	-	-	-	-	-
-	-	9	12,8	37	3,2	10	3,9
-	-	0	0,0	7	0,8	3	1,6
40	100,0	70	100,0	1011	100,0	219	100,0

Rispetto alle famiglie che non hanno la possibilità di ricevere aiuti da parte di parenti, quelle che dispongono di tale risorsa avvertono un livello di disagio inferiore, anche se la differenza non è così evidente. Il 6,9% delle famiglie che non hanno parenti a cui fare affidamento in caso di bisogno si ritiene colpito da problemi al più risolvibili con le forze esterne alla famiglia, contro il 9,0% di quelle che possono contare sull'aiuto di qualche membro della famiglia allargata (Tab.4.39).

Tab.4.39. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per percezione di disagio e possibilità di aiuto da parte di parenti*

Livello di disagio	Aiuto		Totale (n=2465)
	Si (n=1856)	No (n=609)	
Nulla	77,6	66,0	74,6
Lieve	15,5	25,0	17,9
Grave	5,0	6,5	5,4
Estremo	1,9	2,5	2,1
Totale	100,0	100,0	100,0

La percentuale di famiglie venete che ritiene di poter fare affidamento su parenti in caso di necessità ammonta a 74,4%. Nei gruppi a rischio 7, 9, 10, 15 e 21 tale percentuale aumenta. In particolare, l'84,0% delle famiglie, che hanno reddito inferiore a 3000 euro mensili netti, ma non hanno avuto bisogno di aiuti economici nell'ultimo anno, che non contano invalidi al proprio interno e i malati cronico-degenerativi sono autonomi, la cui abitazione è di proprietà o in affitto, in cui non sono presenti componenti disoccupati e che non possiedono nessuna automobile (Gr.15), ritiene di avere un supporto dalla famiglia allargata qualora si presentasse il bisogno. Probabilmente è proprio la mancanza di un proprio mezzo di trasporto che attiva queste famiglie nella ricerca efficace di aiuti. Negli altri gruppi, dove le percentuali sono più basse, è probabile che più spesso le famiglie affermino di non avere la possibilità di chiedere aiuto ai parenti, anche quando, in caso di reale bisogno, possono contare sul loro aiuto. In caso di estrema necessità si riesce, quindi, ad attivare una rete di aiuti anche laddove, in condizioni di assenza di disagio, non sembra possibile (Tab.4.40).

Tab.4.40. *Distribuzione percentuale delle famiglie nei diversi gruppi a rischio per possibilità di aiuto da parte di parenti*

Aiuto	Gr.1		Gr.3		Gr.5		Gr.7	
	n	%	n	%	n	%	n	%
Si	1856	74,4	102	69,5	135	71,3	1167	76,6
No	609	25,6	44	30,5	52	28,7	339	23,4
Totale	2465	100,0	146	100,0	187	100,0	1506	100,0

Gr.9		Gr.10		Gr.12		Gr.15	
n	%	n	%	n	%	n	%
58	81,9	21	81,7	34	69,0	129	84,0
13	18,1	5	18,3	17	31,0	22	16,0
71	100,0	26	100,0	51	100,0	151	100,0

Gr.17		Gr.18		Gr.21		Gr.23	
n	%	n	%	n	%	n	%
13	37,8	51	72,3	792	77,7	-	-
27	62,2	19	27,7	219	22,3	219	100,0
40	100,0	70	100,0	1011	100,0	219	100,0

All'aumentare del livello di reddito, diminuisce il disagio percepito dalla famiglia: il 14,3% delle famiglie che hanno un reddito inferiore a 1500 euro mensili netti si ritiene colpito da problemi di una certa importanza, non risolvibili o risolvibili al più ricorrendo ad aiuti esterni alla famiglia; di contro, solo lo 0,7% di quelle che ricevono oltre i 6000 euro avverte problemi di tale importanza (Tab.4.41).

Tab.4.41. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per percezione di disagio e livello di reddito mensile*

Livello di disagio	Reddito				Totale (n=2207)
	> 6000€	≤ 6000 e > 3000 €	≤ 3000 e > 1500 €	≤ 1500 €	
	(n=53)	(n=439)	(n=1017)	(n=698)	
Nulla	91,6	82,4	74,7	65,1	73,7
Lieve	7,7	13,0	20,2	20,6	18,6
Grave	0,7	4,1	3,5	9,7	5,4
Estremo	0,0	0,5	1,6	4,6	2,3
Totale	100,0	100,0	100,0	100,0	100,0

Il 22,9% delle famiglie venete ha un reddito superiore a 3000 euro mensili netti; il 46,5% compreso tra i 1500 e i 3000 euro e il 30,6% inferiore a 1500 euro (Gr.1).

Come ci si può attendere, in quasi tutti i gruppi a rischio risultanti dall'analisi di segmentazione la percentuale di famiglie con reddito basso (inferiore a 1500 euro) è superiore. Solo nelle famiglie che hanno reddito inferiore a 3000 euro mensili netti, ma non hanno avuto bisogno di aiuti economici nell'ultimo anno, che non contano invalidi al proprio interno e i malati cronico-degenerativi sono autonomi, la cui abitazione è di proprietà o in affitto, in cui non sono presenti componenti disoccupati, che possiedono almeno un'automobile e che non pranzano mai assieme o lo fanno solo nei fine settimana (Gr.18) avviene il contrario: "solo" l'8,7% ha un reddito inferiore a 1500 euro, mentre ben l'91,3% compreso fra 1500 e 3000 euro. C'è da osservare, tuttavia, che tra le famiglie di questo gruppo non ve ne sono con reddito superiore a 3000 euro (Tab.4.42).

Tab.4.42. *Distribuzione percentuale delle famiglie nei diversi gruppi a rischio per livello di reddito mensile*

Reddito	Gr.1		Gr.3		Gr.5		Gr.7	
	n	%	n	%	n	%	n	%
> 6000€	53	2,4	1	0,9	2	1,6	-	-
≤ 6000 e > 3000 €	439	20,5	17	11,5	25	17,1	-	-
≤ 3000 e > 1500 €	1017	46,5	52	36,8	75	43,5	890	62,5
≤ 1500 €	698	30,6	63	50,8	71	37,8	564	37,5
Totale	2207	100,0	133	100,0	173	100,0	1454	100,0

Gr.9		Gr.10		Gr.12		Gr.15	
n	%	n	%	n	%	n	%
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
24	35,8	16	63,1	33	62,9	27	21,9
44	64,2	10	32,9	18	37,1	118	78,1
68	100,0	26	100,0	51	100,0	145	100,0

Gr.17		Gr.18		Gr.21		Gr.23	
n	%	n	%	n	%	n	%
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
19	48,0	60	91,3	644	61,1	140	65,2
21	52,0	8	8,7	333	38,9	73	34,8
40	100,0	68	100,0	977	100,0	213	100,0

Si è ritenuto opportuno sintetizzare le caratteristiche dei gruppi a rischio risultanti dall'analisi di segmentazione, per avere una più precisa visione delle famiglie a maggior rischio di disagio.

La prima variabile che entra nel processo di segmentazione è quella che rileva la necessità di recenti aiuti economici da parte della famiglia. Essa dà origine al gruppo 3, per il quale la percentuale di famiglie con disagio grave, non risolvibile o risolvibile al più con l'intervento dei servizi, è pari a 29,2%. Il bisogno di aiuti economici è molto elevato (43,3%) per questo gruppo di famiglie (Tab.4.24), perlopiù coppie con figli (50,5%) (Tab.4.38) e con età media inferiore ai 50 anni (71,0%) (Tab.4.36). Esse sono inoltre caratterizzate per il fatto di avere un basso livello di reddito: il 50,8% non supera i 1500 euro mensili netti. Entrano a far parte di questo gruppo anche famiglie unipersonali (22,7%) con componenti in giovane età, ragazzi che lasciano la famiglia d'origine per formarne una propria, ma che non sono ancora completamente autonomi dal punto di vista economico.

Le famiglie che non hanno avuto bisogno di un sostegno economico nell'ultimo anno e che contano al loro interno almeno un componente affetto da invalidità (Gr.5) si ritengono colpite da disagio di una certa importanza nel 26,6% dei casi e proprio la presenza di invalidi fa sì che il 67,2% di queste abbia problemi di salute (Tab.4.25). Si tratta di coppie con figli, con componenti di età media superiore rispetto a quella della famiglie con necessità economiche (il 38,0% ha età media compresa tra i 50 e i 70 anni).

Tra le famiglie il cui reddito è inferiore a 3000 euro, ma che non hanno avuto recente bisogno di un sostegno economico, che non hanno invalidi al proprio interno ma malati cronico-degenerativi non autonomi (Gr.9), il 29,4% si ritiene colpito da gravi problemi, nel 65,6% dei casi di salute (Tab.4.27). In questo gruppo entrano a far parte soprattutto coppie di anziani (il 46,9% di queste famiglie è costituito da una coppia senza figli (Tab.4.38) e il 58,7% ha un'età media superiore a 70 anni (Tab.4.36)), con redditi da pensione, inferiori a 1500 euro mensili nel 64,2% dei casi (Tab.4.42).

Ceteris paribus, se i malati cronico-degenerativi sono autonomi e la famiglia ha un'abitazione precaria (Gr.10), il rischio di disagio è del 5,0%. Queste famiglie

hanno soprattutto problemi di salute (21,8%) ed economici (16,0%) (Tab.4.30) e sono costituite nel 43,2% dei casi da giovani di età inferiore ai 30 anni e nel 23,6% dei casi da anziani oltre i 70 (Tab.4.36).

Se l'abitazione non è precaria, bensì di proprietà o in affitto, e tra i componenti almeno uno è disoccupato (Gr.12), la percezione del disagio ammonta a 7,3%. Come nel gruppo 10, tale percentuale è imputabile ai soli problemi risolvibili con aiuto esterno, mentre non vi sono famiglie che ritengono irrisolvibile la propria situazione di disagio. I problemi dichiarati sono perlopiù economici (40,0%) (Tab.4.28): le famiglie hanno un reddito basso, inferiore a 3000 euro mensili netti nella totalità dei casi. L'età media dei componenti è inferiore ai 50 anni nell'82,5% di queste famiglie (Tab.4.36); si tratta di famiglie giovani, perlopiù costituite da una coppia con figli (77,9%) in cui uno dei due coniugi lavora e può sopperire alle difficoltà provocate dalla disoccupazione dell'altro (Tab.4.38).

Nel gruppo 15 entrano a far parte le famiglie che hanno reddito inferiore a 3000 euro mensili netti, ma non hanno avuto bisogno di aiuti economici nell'ultimo anno, che non contano invalidi al proprio interno e i malati cronico-degenerativi sono autonomi, la cui abitazione è di proprietà o in affitto, in cui non sono presenti componenti disoccupati e che non possiedono automobili. Questo gruppo è costituito prevalentemente da anziani con oltre 70 anni di età, che vanno a comporre famiglie unipersonali, con problemi di salute e un livello di reddito che non supera i 1500 euro. Si vedano al riguardo le tabelle 4.31-4.36-4.38-4.42.

Le famiglie di soli anziani sono più problematiche rispetto alle altre e ciò è vero soprattutto nel caso in cui l'anziano non sia completamente autonomo o colpito da qualche forma di invalidità o malattia.

Un'ulteriore analisi di questo gruppo mira a verificare il sesso degli individui (Tab.4.43).

Tab.4.43. *Distribuzione percentuale degli individui del gruppo 15 residenti nel Veneto nel 2004-2005 per sesso*

Sesso	
Femmine (n=154)	73,1
Maschi (n=51)	26,9
Totale (n=205)	100,0

Gli anziani che vivono soli sono perlopiù di sesso femminile: si tratta in larga misura di donne vedove che percepiscono la sola pensione di anzianità e con problemi di salute dovuti alla vecchiaia.

Ceteris paribus, le famiglie che possiedono almeno un'automobile e sono composte da un genitore con figli a carico (Gr.17) hanno, perlopiù, problemi di tipo socio-relazionale. In ogni caso, queste famiglie, soprattutto quando il genitore è professionalmente occupato riescono a far fronte alle difficoltà con le proprie risorse o ricorrendo all'aiuto della famiglia allargata (32,6% di disagio vs. 16,7% nelle altre famiglie). Esse sono famiglie giovani (l'82,9% ha età media inferiore a 50 anni (Tab.4.36)), con problemi socio-relazionali (23,6%) ed economici (21,4%) (Tab.4.29).

L'ultimo gruppo a rischio di disagio preso in considerazione per le sue peculiarità è quello costituito da famiglie con le stesse caratteristiche di quelle del gruppo precedente, tranne per il fatto di non essere di tipo monogenitoriale e di non avere l'abitudine di pranzare mai assieme o di farlo solo nei fine settimana (Gr.18). Queste sono famiglie soprattutto giovani (Tab.4.36), di tipo coppia con figli (Tab.4.38) e con reddito non basso (solo per l'8,7% di queste è inferiore a 1500 euro mensili contro il 30,6% calcolato per l'intero campione di famiglie) (Tab.4.42). Il 31,4% di queste famiglie avverte un disagio risolvibile con le proprie forze o con l'aiuto della famiglia allargata. Nel complesso, le famiglie di questo tipo ammontano a 5,5% e contano al loro interno, generalmente, uno o più figli che scelgono di studiare in un'altra città e sono, perciò, costretti a vivere lontano da casa o un coniuge, solitamente il marito, che vive fuori casa per motivi di lavoro. Oltre alla metà di queste persone torna a casa settimanalmente. Apparentemente risulta inspiegabile la suddivisione avvenuta a questo livello dell'albero: la nostra logica portava ad includere le famiglie che solo talvolta condividono il momento del pasto insieme a quelle che non pranzano mai insieme o solo nei fine settimana e non a quelle i cui componenti pranzano sempre insieme. Ulteriori analisi hanno mostrato come possa accadere che, riunirsi sporadicamente, limiti quelle possibili fonti di tensione e di contrasto tra i componenti e preservi, pertanto, dal rischio di disagio.

Un'attenzione particolare va posta nel confronto tra i gruppi 22 e 23 che di-

scriminano le famiglie in base alla loro possibilità di chiedere aiuti o meno a parenti: a prima vista, osservando le percentuali di disagio nullo o risolvibile all'interno della famiglia, sembrerebbe che chi ha questa possibilità ne abbia meno bisogno; in realtà, le famiglie del gruppo 22 sono famiglie ben strutturate, che non presentano problemi importanti (non entrano, infatti, a far parte dei gruppi a rischio individuati nei livelli precedenti dell'albero di segmentazione) e che, inoltre, hanno una rete parentale su cui fare affidamento in caso di bisogno.

In conclusione, si può affermare che è la combinazione di problemi sanitari ed economici la causa di disagio grave e persistente, che rende una famiglia debole e vulnerabile anche nell'ambito sociale.

4.2.1 Cross-Validation

La figura 4.1 rappresenta la soluzione ottima dell'individuazione dei gruppi di famiglie a maggior rischio di disagio. Per la ricerca di soluzioni sub-ottimali si è utilizzato il criterio della Cross-Validation. Ad ogni split dell'albero si considerano tutte le variabili e si sceglie la migliore, ossia quella per cui l'errore della Cross-Validation è il minore.

Sfruttando al meglio l'informazione disponibile, si sono create V suddivisioni, ognuna con il $(n/V)\%$ dei dati, dove n è la numerosità campionaria, e poi, a rotazione, si sono usate $(V - 1)$ porzioni per il ruolo di stima e una porzione per quello di verifica. Con questo procedimento, si scambiano i ruoli delle porzioni di dati, e ciò richiede di ripetere le operazioni di stima e di verifica del modello V volte. Il massimo valore che può assumere V è n , quando si usano $n - 1$ dati per la stima e un dato per la verifica, il che comporta un notevole incremento della complessità computazionale. Le V stime ottenute, diverse fra di loro anche se presumibilmente non di molto, possono essere combinate, ad esempio, tramite la loro media.

Nel nostro caso i risultati si riferiscono a quest'ultimo tipo di implementazione, nota anche come *leave-one-out*.

L'albero ottenuto è riportato nella figura 4.2.

Legenda

- p1: % nessun problema
- p2: % problemi insolubili con le forze della famiglia
- p3: % problemi insolubili con aiuto esterno
- p4: % problemi non risolubili

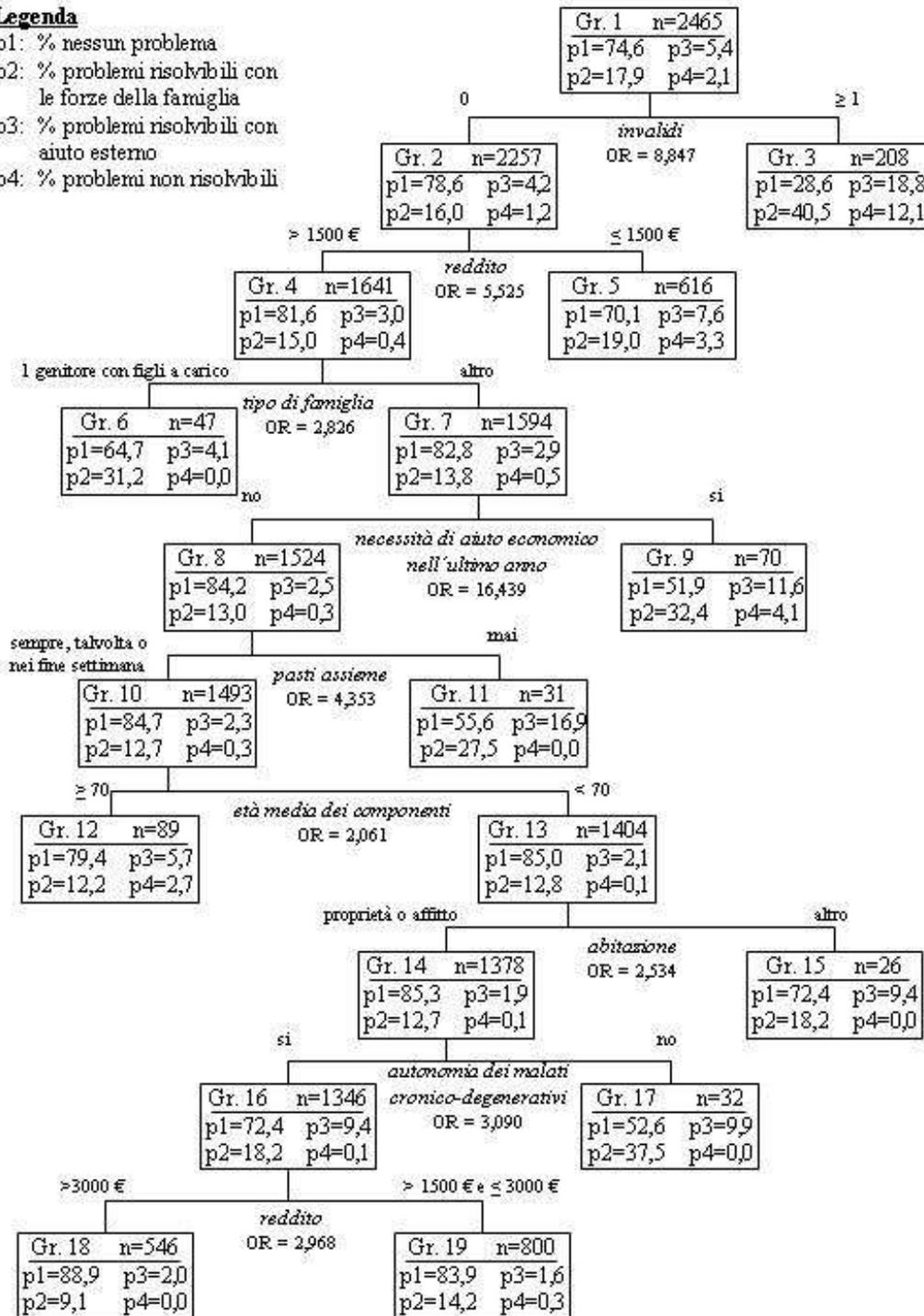


Figura 4.2: Analisi di segmentazione binaria relativa al “rischio di disagio” (variabile politomica) nelle famiglie venete. La soluzione è stata ottenuta attraverso il criterio della Cross-Validation.

Nella tabella 4.44 sono riportati, come misura di accuratezza della regola di classificazione, i tassi di errore relativi alle variabili non correttamente classificate nel processo di validazione incrociata, ossia la proporzione di unità assegnate ad una classe diversa da quella a cui appartengono.

La descrizione delle variabili si trova nella tabella 4.22.

Tab.4.44. *Errori di classificazione (%) ad ogni passo di segmentazione dell'albero*

Variabile	tasso di errore(%)
Numero di invalidi	24,415
Livello di reddito mensile	15,889
Tipo di famiglia	19,153
Necessità di aiuti economici nell'ultimo anno	20,847
Pasti assieme	18,754
Età media dei componenti	24,175
Titolo a cui la famiglia occupa l'abitazione	30,775
Autonomia del malato cronico-degenerativo	26,290
Livello di reddito mensile	20,326

Rispetto all'analisi di segmentazione ottenuta sfruttando tutti i dati a disposizione per la stima (cfr. Par. 4.2), la nuova variabile che entra nella soluzione ottenuta via Cross-Validation è l'*età media dei componenti*. Il rischio di disagio aumenta del 106,1% nelle famiglie in cui l'età media dei componenti supera i 70 anni: l'8,4% di queste famiglie prova disagio di una certa importanza, contro il 2,2% di quelle con componenti più giovani.

Altre variabili che concorrono alla formazione dei gruppi sono le seguenti:

- *il bisogno di aiuti economici*. Nelle famiglie senza invalidi e con reddito superiore a 1500 euro, la necessità da parte della famiglia di chiedere aiuti economici esterni è associata ad un notevole stato di disagio ($OR = 16,439$). Il 15,7% di queste famiglie avverte disagio, contro il 2,8% di quelle economicamente autonome;
- *la presenza di invalidi*. Il rischio di disagio aumenta di quasi sette volte se la famiglia ha al proprio interno almeno un componente invalido ($OR = 8,847$). Ben il 30,9% delle famiglie con invalidi avverte un disagio persistente;

- *la presenza di componenti non autosufficienti affetti da malattie cronico-degenerative* ($OR = 3,090$). Tra le famiglie senza invalidi, con età media dei componenti inferiore ai 70 anni e con malati no autonomi, solo il 52,6% non avverte alcuna forma di disagio, contro il 72,4% di quelle con malati autonomi.

Le variabili che non entrano in questo secondo processo di selezione sono il *numero di disoccupati*, il *numero di automobili* e la *possibilità di aiuto da parte di parenti*.

Nelle due soluzioni le variabili non hanno lo stesso ordine di ingresso; ad esempio, nella segmentazione originaria entra per prima la variabile che rileva l'eventuale necessità per la famiglia di aiuti economici ($OR = 11,402$), mentre nella soluzione ottenuta via Cross-Validation quella che rileva il numero di invalidi ($OR = 8,857$).

In ogni caso, le variabili rilevanti sono all'incirca le stesse nelle due segmentazioni, indice del fatto che la soluzione sub-ottimale si avvicina a quella originaria. Anche i valori degli odds ratio sono simili nelle due soluzioni.

4.3 Percezione del disagio su scala dicotomica

Nel precedente paragrafo sono stati evidenziati, mediante segmentazione binaria, alcuni gruppi di famiglie venete soggetti a diversi livelli di disagio. Le caratteristiche di tali famiglie riguardano diversi ambiti e sono relative sia ai singoli componenti sia alla famiglia nel suo insieme. Per la misura del disagio, si è utilizzata una scala ordinale a quattro modalità, la quale consente di evidenziare, con buon dettaglio, i tratti delle famiglie venete soggette a differenti livelli di disagio.

A questo punto, per individuare le determinanti del disagio grave, si è utilizzata una scala dicotomica dello stesso:

- disagio nullo o lieve, affrontabile con le forze della famiglia;
- disagio grave o estremo, risolvibile al più attraverso l'intervento esterno di istituzioni (servizi) o associazioni di volontariato.

Il 92,5% delle famiglie venete si ritiene non colpito da problemi generatori di disagio. Questa porzione di famiglie considera i propri problemi risolvibili con le proprie forze o comunque con l'aiuto di parenti stretti. Il restante 7,5% delle famiglie considera i propri problemi piuttosto gravi, risolvibili al più attraverso il ricorso all'aiuto esterno alla famiglia.

Come nel caso precedente, in cui la variabile che misura la percezione del disagio è misurata su una scala a quattro modalità, per la formazione dei gruppi omogenei in funzione del disagio percepito, si è adottato un metodo multivariato di segmentazione binaria. In particolare, tale metodo permette di segmentare il campione di famiglie con una procedura di suddivisione binaria *stepwise* in base alla massima differenza tra i rischi, ossia le probabilità di rischio, nei due sottogruppi che di volta in volta si vengono a formare. In questo modo, ad ogni passo dell'analisi, si formano due sottogruppi caratterizzati da odds ratio massimamente differenti. Tale differenza è valutata in funzione del logit della speranza matematica del verificarsi del disagio Y (variabile dicotomica) condizionatamente al preventivo succedere di X (generico predittore):

$$\text{logit}(\pi(Y|X)) = \ln \left[\frac{\pi(Y|X)}{1 - \pi(Y|X)} \right]$$

Le variabili entranti nel processo di segmentazione sono 9: la *presenza di invalidi*, lo *stato di autonomia dei malati cronico-degenerativi*, il *bisogno di recenti aiuti economici*, il *numero di automobili*, il *tipo di famiglia*, la *presenza di componenti disoccupati*, il *livello di reddito*, il *titolo a cui la famiglia occupa l'abitazione* e l'*età media dei componenti*.

Si riportano, in figura 4.3 i risultati ottenuti.

L'albero ottenuto è formato da 19 nodi, di cui 8 foglie. Tutte le variabili che risultano significative nell'analisi dicotomica, ad eccezione di quella che rileva l'età media dei componenti, sono significative anche nell'analisi politomica. Questo sta ad indicare che le cause di grave disagio, evidenziate sfruttando una scala a due modalità, sono messe in rilievo anche attraverso un'analisi più dettagliata, la quale, inoltre, evidenzia altre caratteristiche che determinano una situazione di disagio meno importante perché risolvibile attraverso il ricorso ai servizi.

Legenda

p0: % problemi nulli o lievi
 p1: % problemi gravi o estremi

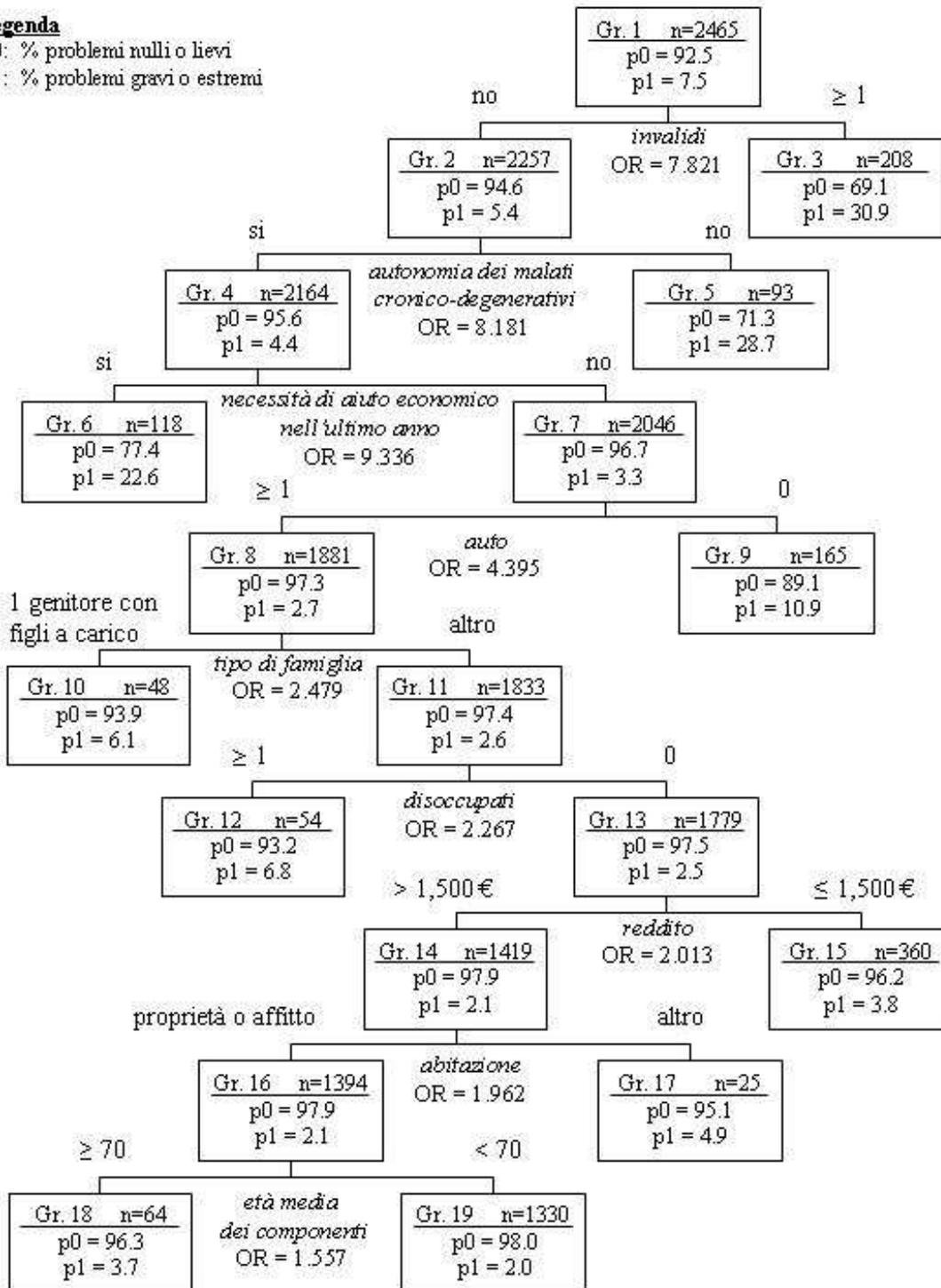


Figura 4.3: Analisi di segmentazione binaria relativa al “rischio di disagio” (variabile dicotomica) nelle famiglie venete.

In particolare lo stato di disagio è determinato da:

- il *bisogno di aiuti economici*. A differenza dell'analisi politomica, in cui la discriminazione tra le famiglie che hanno avuto bisogno di sostegno economico e quelle autosufficienti sul piano finanziario è avvenuta al primo livello di segmentazione dell'albero, nell'analisi dicotomica tale discriminazione ha comportato, al terzo livello di segmentazione, la formazione di un gruppo di 118 famiglie di cui il 22,6% ad alto rischio di disagio;
- la *presenza di invalidi*. La variabile che rileva la presenza di componenti affetti da invalidità entra al primo livello di segmentazione, rilevando, per le famiglie con almeno un componente invalido, un rischio di disagio di quasi otto volte superiore rispetto alle altre ($OR = 7,821$), rischio che riguarda il 30,9% di queste famiglie;
- la *presenza di componenti non autosufficienti affetti da malattie cronico-degenerative*. Il rischio di disagio è di oltre otto volte superiore ($OR = 8,181$) in presenza di malati cronico-degenerativi non autonomi (il disagio grave coinvolge il 28,7% di queste famiglie) rispetto al caso di malati autonomi (disagio grave nel 4,4% dei casi);
- la *manca di automobili*. Tra le famiglie con malati cronico-degenerativi autonomi, 165 non possiedono alcuna automobile e, tra queste, il 10,9% è soggetta ad un alto rischio di disagio.

È, dunque, la combinazione di problemi sanitari ed economici la causa di disagio grave e persistente, che rende una famiglia debole e vulnerabile anche sul piano sociale.

Altre situazioni problematiche, che tuttavia si possono collocare in una zona grigia, e non nera, di disagio, ossia in una zona di affanno più che di disagio vero e proprio, sono quelle a cui la famiglia riesce a far fronte con le proprie forze. Caratteristiche quali un'*anomala struttura familiare*, la *presenza di disoccupati* e

il *possesso dell'abitazione a titolo gratuito*, sono messe in evidenza a partire dal quinto livello di segmentazione, a seguito di quelle che rilevano problemi economici e sanitari.

Esse comportano la formazione di coppie di gruppi diversi tra loro, rispetto alla percezione del disagio, anche se tale diversità non è così evidente come quella che riguarda le coppie di gruppi segmentati sulla base delle variabili che rilevano la presenza di invalidi, l'autonomia dei malati cronico-degenerativi, il bisogno di aiuti economici e il numero di automobili.

Eventi critici quali sfratti, furti o incidenti si sono rivelati poco o per nulla significativi nella determinazione del disagio e, come si può notare, non sono entrati nel processo di segmentazione.

Non si è ritenuto opportuno riportare le caratteristiche strutturali dei gruppi a rischio individuati, dal momento che i risultati concordano con quelli ottenuti misurando il disagio su scala politomica. Si ritiene, invece, opportuno sottolineare che ciò che l'analisi a quattro modalità ci ha permesso di evidenziare sono gli affanni, le situazioni di disagio non estremo, che non si sono rivelate essere l'anticamera del disagio grave, bensì condizioni risolvibili all'interno della famiglia stessa.

L'analisi dei dati è stata condotta, in un secondo momento, con gli algoritmi di segmentazione implementati nel software LAID-OUT 2.1 (Schievano, 2002)¹.

I parametri utilizzati per ottimizzare la formazione dell'albero sono analoghi a quelli sfruttati nell'analisi politomica (cfr. Par. 4.2).

Delle 21 variabili prese in esame (4.22), sette entrano con un ruolo fondamentale nel processo di segmentazione: il *numero di invalidi*, l'*età media dei componenti*,

¹Per quanto riguarda i criteri di scelta per l'arresto della procedura, si sono imposti i seguenti parametri:

- | | |
|-------------------------------|------|
| • livello di significatività | 0,15 |
| • massimo numero di foglie | 10 |
| • dimensione minima di gruppo | 20 |
| • minimo numero di modalità | 2 |
| • massimo numero di modalità | 30 |

il livello di reddito, l'autonomia dei malati cronico-degenerativi, il titolo a cui la famiglia occupa l'abitazione, il tipo di famiglia e l'autonomia degli invalidi.

Nella figura 4.4 si riportano i risultati dell'analisi, con indicate le modalità delle variabili che generano i gruppi, le numerosità di gruppo e le frequenze relative $P(Y|X)$ delle famiglie che presentano disagio di una certa importanza.

L'albero ottenuto è formato da 21 nodi, di cui 11 terminali (o foglie), a ciascuno dei quali è associato un particolare stato di disagio determinato dalla combinazione delle modalità delle variabili che hanno concorso a creare quel nodo.

La soluzione di individuare gruppi omogenei rispetto al rischio di disagio, ha messo in luce le interazioni tra variabili che possono essere le cause principali di disagio grave per le famiglie. Tali cause sono di varia natura e riguardano i caratteri sia esogeni che individuali e comportamentali. In particolare lo stato di disagio è determinato da:

- *presenza di invalidi in famiglia.* Il rischio di disagio aumenta di quasi otto volte nel caso di una o più persone invalide in casa:

$$\hat{\Psi} = \frac{P(\text{disagio} \mid \text{invalidi})}{P(\text{disagio} \mid \text{nessun invalido})} = 7,8$$

Il 30,9% delle famiglie in cui è presente almeno un componente invalido avverte disagio, contro il 5,4% di quelle in cui tutti sono sani (e il 7,4% del gruppo genitore, che rappresenta l'intero campione di famiglie). Il problema è talmente grave che la famiglia necessita di aiuti dall'esterno per farvi fronte;

- *assenza di autonomia dei componenti invalidi.* Nelle famiglie in cui il reddito è inferiore a 3000 euro mensili netti, l'abitazione è in proprietà o usufrutto e la famiglia non è di tipo unipersonale, la presenza di uno o più invalidi con grado di autonomia nullo o parziale è fonte di maggior disagio rispetto a quello avvertito, *ceteris paribus*, nelle famiglie in cui gli invalidi sono autonomi ($\hat{\Psi} = 2,2$);

Legend

p0: % problemi nulli o lievi
 p1: % problemi gravi o estremi

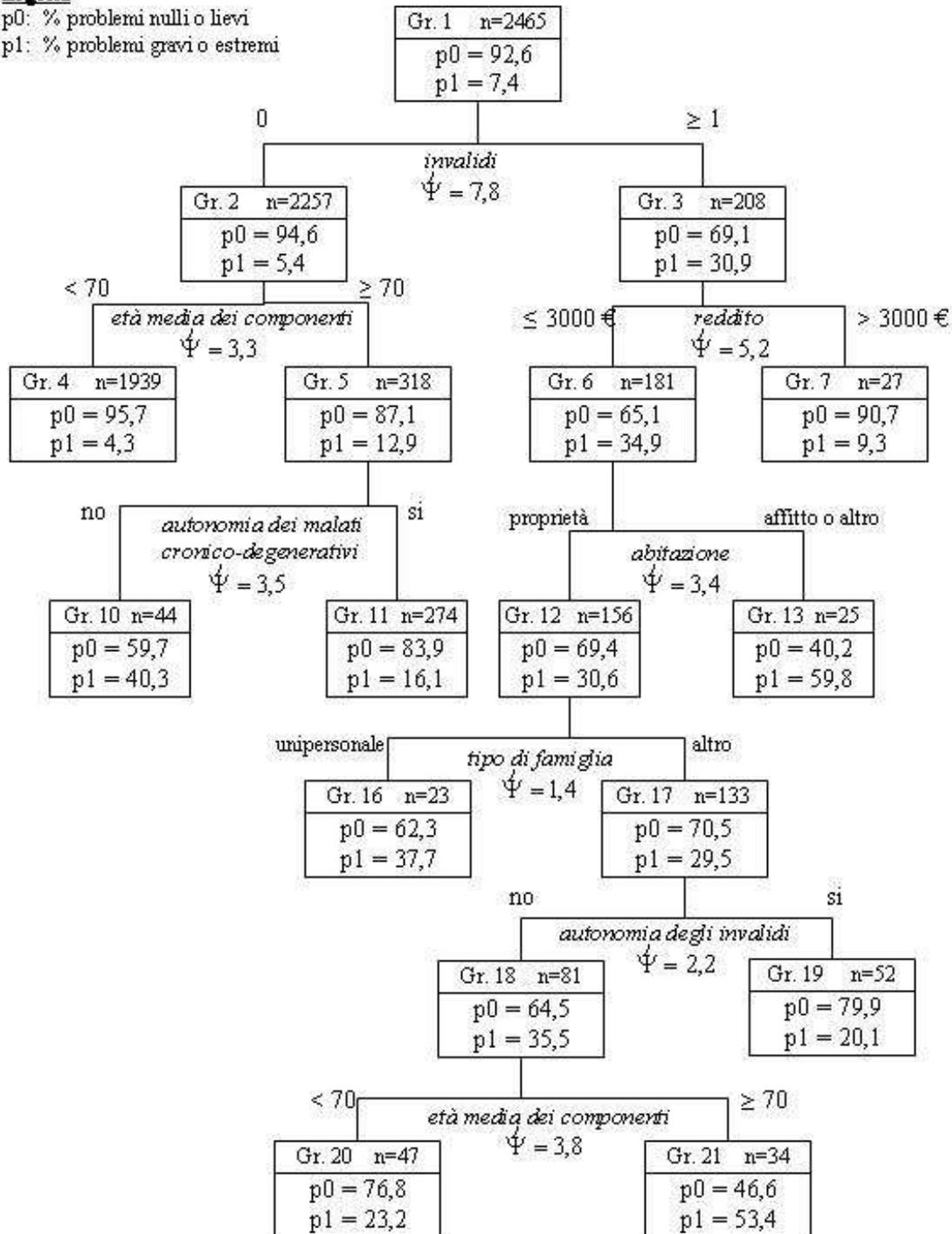


Figura 4.4: Analisi di segmentazione binaria relativa al “rischio di disagio” (variabile dicotomica) nelle famiglie venete.

- *assenza di autonomia dei malati cronico-degenerativi.* Risulta evidente che le famiglie in cui sono presenti almeno due malati cronico-degenerativi totalmente non autonomi avvertono un disagio superiore rispetto alle altre. Il rischio in questo caso è pari a $\hat{\Psi} = 3,5$. Avverte disagio il 40,3% delle famiglie in cui l'età media dei componenti è superiore a 70 anni e vi sono almeno due malati cronico-degenerativi non autonomi. *Ceteris paribus* tale percentuale scende a 16,1% per le famiglie in cui i malati cronico-degenerativi risultano essere totalmente autonomi;
- *famiglie unipersonali.* Per quanto riguarda le famiglie in cui è presente almeno un componente invalido, il reddito è inferiore a 3000 euro mensili netti e l'abitazione è in proprietà o usufrutto, il rischio di disagio è maggiore in quelle unipersonali ($\hat{\Psi} = 1,4$). Questo risultato è dovuto soprattutto all'evidente mancanza di aiuto reciproco tra i componenti nelle famiglie unipersonali, aiuto tanto più necessario per il fatto che nel gruppo in questione entrano a far parte le famiglie unipersonali di invalidi;
- *età media dei componenti superiore a 70 anni.* Il rischio di disagio è maggiore nelle famiglie in cui l'età media dei componenti è superiore a 70 anni ($\hat{\Psi} = 3,3$). Il 12,9% delle famiglie in cui non sono presenti invalidi e l'età media dei componenti è superiore a 70 anni prova forte disagio, contro il 4,3% di quelle in cui l'età media è inferiore a 70 anni;
- *reddito familiare inferiore a 3000 euro.* Il rischio di disagio dovuto a bassi livelli di reddito familiare (inferiore a 3000 euro mensili netti) è elevato. Nelle famiglie con almeno un invalido il rischio di disagio aumenta di oltre cinque volte ($\hat{\Psi} = 5,2$) se il reddito è inferiore a 3000 euro rispetto al caso di redditi superiori. Il 34,9% delle famiglie con reddito inferiore a 3000 euro prova forte disagio, contro il 9,3% di quelle con reddito superiore;
- *titolo a cui la famiglia occupa l'abitazione.* Le famiglie che hanno l'abitazione in affitto/subaffitto avvertono un disagio più che triplo rispetto a quelle per cui l'abitazione è di loro proprietà: $\hat{\Psi} = 3,4$.

Oltre ai fattori che possono generare forme di malessere all'interno della famiglia, esistono caratteri protettivi contro il rischio di disagio, ad esempio:

- *presenza di due o più componenti.* *Ceteris paribus*, la presenza di più di due componenti (per le famiglie con reddito basso) e più di un componente (per le famiglie con reddito medio/basso) riduce il rischio relativo di disagio grave di quasi quattro volte rispetto a quello che si ha nelle famiglie in cui sono presenti uno o due componenti (reddito basso) o un solo componente (reddito medio/basso);
- *presenza di entrambi i genitori in famiglia.* Per esempio, in una famiglia in cui il reddito sia elevato, non siano presenti invalidi e al più un componente sia affetto da malattie cronico-degenerative, la presenza di una coppia con figli, o perlomeno con altri componenti, riduce il rischio di disagio grave. Le famiglie unipersonali che sembrano soffrire di minor disagio sono quelle per cui il livello di reddito è alto;
- *livello di reddito alto.* Per le famiglie in cui non sono presenti componenti affetti da invalidità il rischio di disagio diminuisce di oltre dieci volte quando il reddito familiare è alto o medio/alto.

I fattori più influenti nella determinazione del disagio sono la presenza di invalidi in famiglia, l'età media dei componenti superiore a 70 anni, il basso livello di reddito, la mancanza di autonomia degli eventuali componenti invalidi e malati cronico-degenerativi, l'essere la famiglia in affitto/subaffitto e l'essere composta da un solo componente.

I risultati, visti nel dettaglio, si discostano da quelli ottenuti nelle due precedenti analisi, ma, nella sostanza, evidenziano le stesse tipologie di problemi e le famiglie a rischio individuate hanno le stesse caratteristiche strutturali. La diversità dei risultati è imputabile esclusivamente all'utilizzo di due diversi pacchetti statistici, il SAS per le prime due analisi e il LAID-OUT per la terza.

Capitolo 5

Modello di analisi neurale

5.1 Le reti neurali

Le reti neurali artificiali (ANN, Artificial Neural Network) sono modelli matematici che “emulano” il sistema nervoso centrale umano, utilizzando un certo numero di unità indipendenti tra loro interconnesse e raggruppate in strati differenti. Nel cervello umano le unità sono i neuroni e i collegamenti gli assoni e le sinapsi.

Esse si basano sul concetto che, in determinate situazioni, è possibile far “apprendere” a un apparato matematico determinate leggi che non si conoscono a priori, semplicemente facendogli analizzare un elevato numero di casi reali.

Elaborando una mole notevole di dati nella fase di apprendimento, le reti propongono una linearizzazione dei legami tra variabili di analisi per mezzo di una formula matematica. Tale funzione lineare stima la probabilità che ha un nuovo caso mai esaminato di collocarsi nell’area delle “soluzioni” già verificatesi nella fase di addestramento.

In una rete neurale alcuni neuroni sono utilizzati come ingressi, altri come uscite, altri ancora, non connessi con l’esterno, svolgono una funzione di “calcolo” e vengono chiamati neuroni “nascosti” (hidden).

Ogni singolo neurone rappresenta una mappa, che solitamente riceve molteplici input e fornisce un unico output.

Nel cervello umano un impulso (elettrico) viaggia da un neurone all’altro seguen-

do i collegamenti; se il segnale è abbastanza forte, il neurone si “attiva” e invia il segnale a tutti i neuroni ad esso collegati.

Allo stesso modo, in una rete neurale, quando un’unità riceve uno stimolo (input) di sufficiente intensità dalle unità che la precedono, essa si attiva e invia il segnale (output) alle successive unità, elaborando una funzione detta “funzione di attivazione” o “funzione di trasferimento” o “funzione soglia”.

Alla connessione tra due unità viene associato un valore numerico (peso), che rappresenta la forza con cui due unità sono collegate. L’input di un’unità equivale alla somma pesata dei segnali provenienti da tutte le unità ad essa collegate e sarà alta se quell’input viene giudicato importante, bassa nel caso contrario. Da un collegamento all’altro, il segnale, trasformazione della somma opportunamente pesata degli input, viene attenuato fino a smorzarsi quando non risulta più sufficiente ad attivare neuroni.

I singoli neuroni potranno ricevere come input valori direttamente provenienti dall’esterno e/o output di neuroni situati a livelli precedenti nell’architettura neurale.

Le capacità della rete sono determinate sia dal numero di neuroni, sia, soprattutto dalla tipologia (cioè da come sono collegati tra loro i neuroni).

Non esiste, infatti, un’unica architettura possibile di rete neurale artificiale; in letteratura si trovano le descrizioni di molteplici strutture possibili. Le più utilizzate si possono schematizzare in due classi: le reti alimentate in avanti (*feed forward*) e le reti ricorrenti (o ricorsive) (Bishop, 1995).

Nel primo tipo di rete, che di norma è considerato uno schema base da cui partire, i neuroni sono interconnessi in strati differenti in modo tale che il flusso di informazioni segua una sola direzione: ogni strato di neuroni riceve input solo dallo strato precedente e li trasmette solo allo strato successivo. Questo aspetto rende le reti abbastanza semplici da studiare e addestrare, ma non permette loro di mantenere una memoria di quello che è successo nel passato.

Le reti ricorrenti hanno delle capacità più estese, permettendo anche un movimento *backward* dell’informazione: alcuni input della rete corrispondono ad output da essa calcolati precedentemente. Questo tipo di architettura permette alla rete

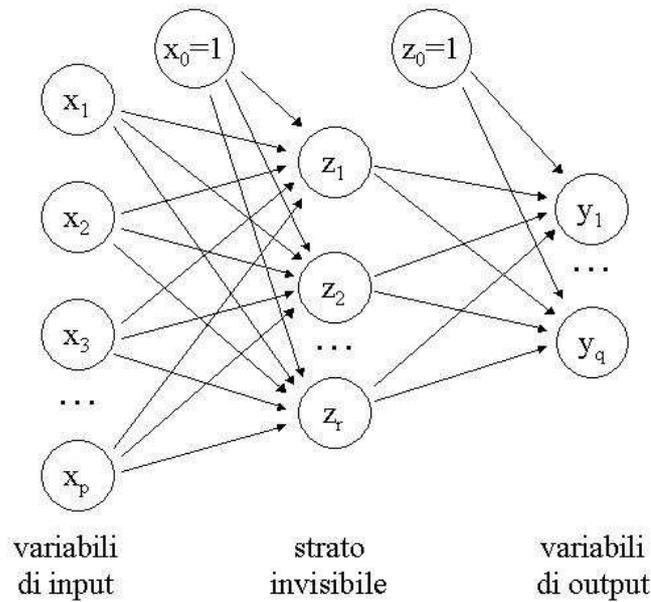


Figura 5.1: Diagramma di una semplice rete neurale.

di avere una memoria associativa. Infatti, ogni volta che un nuovo input è passato alla rete, la risposta non dipenderà solo da esso, ma anche dall'analisi svolta precedentemente. Ad oggi, tuttavia, le reti ricorrenti non sono state ancora ben comprese e non si conosce sempre un metodo adatto per addestrarle.

I principali campi di applicazione delle reti neurali sono quelli in cui l'analisi statistica delle variabili di un problema risulta difficoltosa o dispendiosa in termini di calcolo, ma soprattutto dove non è chiaro a priori quali relazioni esistano tra le diverse variabili che caratterizzano il problema.

Per illustrare l'idea alla base delle reti neurali, si osservi la figura 5.1, nella quale p variabili esplicative, o di input, sono messe in relazione con q variabili risposta, o di output, tramite uno strato di r variabili latenti, quindi non osservate. Nella figura sono, inoltre, rappresentate delle costanti aggiuntive pari a 1, introdotte per eliminare possibili fonti di distorsione. Pertanto, una rete neurale è uno schema di regressione a due stadi, generalmente di tipo non lineare.

In formule matematiche:

$$z_j = f_0 \sum_{h \rightarrow j} \alpha_{hj} x_h \quad , \quad h_k = f_1 \sum_{j \rightarrow k} \beta_{jk} z_j \quad (5.1)$$

dove x_h , z_j e y_k sono, rispettivamente, le generiche variabili di input, latenti e di output e α_{hj} e β_{jk} i parametri da stimare, che rappresentano i pesi associati alla connessione tra coppie di unità.

Le sommatorie si riferiscono a variabili tra le quali è prevista una relazione di dipendenza, specificata in figura 5.1 tramite frecce orientate.

Infine f_0 e f_1 sono le *funzioni di attivazione*, poste, nei problemi di regressione, solitamente pari a:

$$f_0(u) = \frac{e^u}{1 + e^u} \quad , \quad f_1(u) = u \quad (5.2)$$

L'argomento delle funzioni f_0 e f_1 è il valore della somma pesata degli input, che andrà confrontato con un valore soglia: se esso sarà superiore, si genererà un segnale, se inferiore se ne produrrà un altro.

Per far sì che la rete non diventi un insieme di relazioni lineari, formato dalle sole variabili di input e di output, almeno una delle due funzioni f_0 e f_1 deve necessariamente essere non lineare. Se così non fosse, l'output risulterebbe semplicemente proporzionale all'input e risulterebbe difficile trovare una soluzione (Azzalini e Scarpa, 2004).

Una rete risulta spesso caratterizzata dalla funzione di trasferimento che utilizza, dal momento che l'output prodotto dipende strettamente da essa.

Altri elementi da determinare sono i coefficienti α e β . Supponendo di aver fissato r , scelto empiricamente provando diversi valori, la stima dei coefficienti α e β si ottiene minimizzando la funzione di perdita:

$$D = \sum_i \|y^{(i)} - f(x^{(i)})\|^2 \quad (5.3)$$

dove $y^{(i)}$ e $x^{(i)}$ sono i vettori q-dimensionale e p-dimensionale, di variabili risposta ed esplicative, rispettivamente, relative alla i-esima osservazione e $f(x)$ è il vettore la cui k-esima componente è:

$$f(x)^{(k)} = f_1 \sum_{j \rightarrow k} \beta_{jk} f_0 \left(\sum_{h \rightarrow j} \alpha_{hj} x_h \right) \quad (k = 1, \dots, q) \quad (5.4)$$

Per evitare problemi di sovra-adattamento e favorire il processo di ottimizzazione, alla funzione obiettivo precedente si aggiunge un termine di penalizzazione:

$$D_0 = D + \lambda J(\alpha, \beta) \quad (5.5)$$

dove λ è un parametro positivo di regolazione e $J(\alpha, \beta)$ è una funzione di penalizzazione, pari alla somma del quadrato dei pesi (tra le varie unità) (Azzalini e Scarpa, 2004).

Per la scelta di λ , Venables & Ripley (1997) consigliano valori compresi tra 10^{-4} e 10^{-2} .

Per la minimizzazione di D_0 sono stati sviluppati numerosi algoritmi, il più utilizzato dei quali prende il nome di *back-propagation*, che provvede alla minimizzazione dell'errore partendo dall'ultimo strato di neuroni fino ad arrivare al primo.

Va tenuto presente che la funzione obiettivo D_0 ha spesso molti punti di minimo locale e risulta necessario, pertanto, implementare l'algoritmo a partire da diversi punti iniziali.

Rispetto ad altri sistemi di calcolo, la rete non è programmata per svolgere un determinato compito, ma viene "addestrata", mostrandole degli esempi (sotto forma di coppie ingresso-uscita). Tali esempi costituiscono l'intervallo di allenamento (*Training set*).

All'inizio le connessioni sono tarate in modo casuale e l'output fornito sarà, probabilmente, privo di senso.

Gli input che ogni neurone riceve vengono confrontati con un valore soglia. Se superano questo valore si avrà un determinato segnale, differente da quello ottenibile nel caso opposto. Questo segnale potrà poi essere inviato, sempre opportunamente pesato, agli eventuali neuroni dello strato successivo, o esternamente alla rete, fornendo in tal modo la risposta a quel determinato problema.

La rete, controllata da un algoritmo di apprendimento, confronta l'output effettivo con quello atteso e, a poco a poco, modifica le connessioni per minimizzare tale differenza.

La rete impara così ad associare ad ogni ingresso un'uscita e, inoltre, a generalizzare quello che ha imparato per ingressi che non le erano stati presentati in fase di addestramento (Ripley, 1999).

Esistono due tipi di apprendimento: con supervisione o senza supervisione (Bishop, 1995).

Nel primo caso la procedura di apprendimento è pilotata esternamente. Per ciascuno dei dati di input, che costituiscono l'insieme di addestramento, è associato un output (output desiderato), tenendo conto del quale vengono aggiornati i pesi dei collegamenti e presentato alla rete un nuovo input. L'apprendimento termina quando l'errore ha raggiunto il suo livello minimo. La "supervisione" sta nella conoscenza dell'output, dato l'input. Questo tipo di procedura è spesso usato in contesto previsionale, dove come valore desiderato normalmente si fornisce il valore vero che il fenomeno studiato assume nell'istante di previsione.

Le reti non supervisionate sono, invece, in grado di imparare autonomamente. I pesi e le distorsioni della rete sono modificati solo in base agli input, poiché ad essa non sono forniti valori target di output. La rete cerca regolarità fra i segnali di input, dividendoli in categorie omogenee, e opera opportuni adattamenti a seconda del tipo di funzione sottostante alla rete stessa. La rete possiede, infatti, nella sua topologia e nelle sue regole di apprendimento, le informazioni su come organizzarsi. Tra queste reti, le più famose sono quelle di Kohonen (1989) (SOM, Self-Organizing Maps), che creano una "mappa" degli input mettendo vicini tra di loro gli input simili.

Caratteristica comune dei due tipi di apprendimento è la presenza di un tasso di apprendimento (*learning rate*) che misura la velocità di convergenza dei pesi iniziali verso quelli finali; non esistendo ad oggi criteri universalmente validi per stabilire la dimensione di questo parametro.

5.1.1 Pregi e difetti delle reti neurali

Le reti neurali sono uno strumento molto utilizzato per i numerosi vantaggi che presentano. Fra questi, i principali sono:

- la capacità di esplorazione dei dati, ossia di trovare relazioni nascoste in

masse di dati, pur non essendo le reti programmate per risolvere uno specifico problema;

- la capacità di generalizzazione, cioè, di produrre output a partire da input non codificati in precedenza;
- la flessibilità, nel senso di ottenere buone approssimazioni di qualsiasi funzione di regressione $f(x)^{(k)}$;
- la compattezza della rappresentazione, essendo la funzione di regressione stimata identificata da un numero limitato di componenti;
- l'aggiornabilità sequenziale, che permette, attraverso varianti dell'algoritmo di *back-propagation*, di aggiornare sequenzialmente i coefficienti α e β , via via che si inseriscono nuovi dati.

Come tutti i metodi di analisi, le reti neurali presentano una serie di limiti, che la ricerca sta studiando e cercando di superare. I più evidenti risultano essere:

- la mancanza di precise regole a cui attenersi nella progettazione di una rete neurale. E' arbitraria la scelta del numero di variabili latenti r ; poche indicazioni si hanno anche per la scelta del parametro di *weight decay* λ . Probabilmente non esiste un modello di rete neurale ottimo, valido per ogni tipo di problema: la dipendenza della rete dal tipo di applicazione è di estrema importanza per la scelta della sua architettura;
- la difficoltà della stima della funzione obiettivo D , per la quale è evidente la presenza di molti minimi locali. Inoltre, avviando l'algoritmo da punti diversi si ottengono risultati diversi;
- l'impossibilità di ottenere errori standard per i coefficienti;
- la necessità di molti esempi nel caso di allenamento supervisionato: per addestrare bene una rete è necessario disporre di una grande quantità di dati e questo rende l'apprendimento un processo piuttosto lungo;

- l'assenza di modularità e la scarsa trasparenza: la conoscenza è disposta su tutta la rete e tutta la rete partecipa alla soluzione del problema. Questo è da una parte un vantaggio, ma dall'altra uno svantaggio in quanto, se le prestazioni della rete non risultano soddisfacenti, sarà difficile individuarne il motivo.
- la difficoltà dell'interpretazione del risultato, soprattutto al crescere di r .

5.2 Cenni storici

L'idea dei modelli di reti neurali artificiali risale al 1943, quando il neurofisiologo McCulloch e il matematico Pitts (McCulloch e Pitts, 1943) modellarono una rete neurale semplice con circuiti elettrici.

Alla fine degli anni '40, Hebb (1949) propose una rete neurale in grado di imparare grazie ad una alterazione dei pesi sulle connessioni fra i neuroni (*Hebbian learning rule*). In seguito, Minsky (1954) costruì e testò dei neurocomputer e Rosenblatt (1958) della Cornell University creò un modello di neurone, ispirandosi al lavoro di McCulloch e Pitts, conosciuto oggi come il modello del *perceptron*, struttura in grado di imparare a classificare modelli modificando le connessioni fra gli elementi. Sebbene il *perceptron* fosse un modello molto semplice, contribuì ad aumentare l'interesse verso lo sviluppo della ricerca nel campo delle reti neurali artificiali. Tuttavia, tale entusiasmo fu frenato da Minsky e Papert (1969), che mostrarono come il *perceptron* fosse un modello inadeguato della mente umana, non essendo in grado di eseguire calcoli elementari. Anche la necessità di trovare una soluzione nel caso di problemi non linearmente separabili portò a complicare la struttura del *perceptron*. Fu, infatti, concepita una struttura di rete, il *multilayer perceptron*, basata su uno o più strati di neuroni, connessi in modo tale che ogni neurone ricevesse come input l'output dello strato precedente, fino all'ultimo strato che costituisce l'output definitivo della rete. La legge di apprendimento che governa questo tipo di architetture è del tipo *back-propagation*.

Nei primi anni '60 Widrow e Hoff (1960) svilupparono l'importante regola di ap-

prendimento, nota come *Windrow-Hoff learning rule*, basata sulla minimizzazione dell'errore quadratico durante la procedura di addestramento.

Le reti sviluppate da Widrow e Hoff, note come reti adattive, aggiornano le risposte e i pesi in ogni intervallo temporale. In tal modo, l'apprendimento è di tipo dinamico: la rete riceve inizialmente un singolo vettore di input e, sulla base dell'errore di previsione, provvederà alla modifica dei pesi, che saranno usati nell'intervallo temporale successivo come punto di partenza per allenare la rete, a cui sarà presentato un nuovo vettore di input. L'obiettivo sarà allora quello di allenare una rete, su una base temporale continua, in modo tale che il segnale di output si avvicini il più possibile al valore obiettivo.

L'identificazione del valore del *learning rate* assume molta importanza. Dal momento che la rete non conosce a priori il tipo di dati in input, esso dovrà assumere un valore relativamente piccolo, per indagare in modo accurato lo spazio di ricerca, ma allo stesso tempo deve esplorare in modo esaustivo tutte le possibili soluzioni.

Questa struttura di rete è risultata di grande utilità nelle applicazioni riguardanti il controllo dei segnali e la rimozione di possibili componenti di rumore (Windrow e Sterns, 1985). Purtroppo, però, non si era in grado all'epoca di verificare la validità e applicabilità dei teoremi.

Negli anni '80 Hopfield (1982) dimostrò come una rete neurale fosse adatta alla risoluzione del problema del commesso viaggiatore (*Travelling Salesman Problem*). Egli, prendendo le distanze dal modello del *perceptron* e avvicinandosi al concetto odierno di rete neurale, mostrò, attraverso una precisa analisi matematica, la potenza di tale strumento.

La funzione principale delle reti di Hopfield è quella di riconoscere modelli con molto rumore come modelli che sono stati "immagazzinati nella sua memoria", cioè già classificati: ad un modello con rumore viene associato in output un modello privo di rumore e, quindi, perfettamente identificato.

La rete di Hopfield è caratterizzata da un singolo strato di neuroni. Ogni neurone è connesso simmetricamente ad ogni altro neurone della rete (*fully connected*). Gli input sono passati contemporaneamente a tutti i neuroni della rete; durante

l'allenamento ogni output di un neurone è un nuovo input per lo stesso neurone che produrrà così un nuovo output fino alla convergenza della rete (*recurrent network*). Questo tipo di rete ha, però, una limitata capacità mnemonica: talvolta possono essere presenti punti stabili spuri, che portano a risposte sbagliate e non è sempre detto che lo stato finale della rete corrisponda necessariamente al caso più simile all'input. Il suo punto di forza sta, invece, nella semplicità della sua architettura e nella necessità di un tempo computazionale limitato per la risoluzione dei problemi.

Nello stesso periodo, Rumelhart e McClelland (1986), insieme ad altri studiosi, presentarono l'algoritmo di retropropagazione dell'errore per l'addestramento di reti multistrato di *perceptron*. Tale meccanismo consente di usare l'informazione contenuta nella componente di errore per modificare i pesi degli strati neurali precedenti.

Negli stessi anni, Kohonen (1989) propose un tipo di rete in grado di apprendere in maniera non supervisionata, senza basarsi su informazioni conosciute a priori, avendo la capacità di ricevere diverse configurazioni in ingresso e di classificarle per somiglianza.

Esse contengono due strati di neuroni: il primo ha la sola funzione di calcolare la somma pesata degli input; nel secondo, bidimensionale, i neuroni sono disposti a formare un reticolo. Quest'ultimo strato fornirà l'output finale della rete.

Ogni neurone del primo strato è collegato a tutti i neuroni del secondo e l'apprendimento coinvolge i neuroni fra loro vicini, essendo le aree di vicinanza scelte dal programmatore.

La strategia di apprendimento, di tipo *competitivo*, prevede il computo delle distanze tra i pesi di ogni neurone del reticolo con i neuroni del primo strato. La distanza minima viene selezionata e si procede alla modifica dei pesi, anche dei neuroni vicini, dal momento che la rete procede all'apprendimento creando regioni costituite da un insieme di valori attorno all'input.

Proprio per questo motivo, la rete di Kohonen meglio si adatta al problema del commesso viaggiatore rispetto alla rete di Hopfield (Angeniol, 1988). Quando un neurone viene attivato, si considerano anche tutti i neuroni vicini, minimizzando

contemporaneamente le distanze di tutta l'area di neuroni di interesse.

I campi di applicazione in cui trovano sviluppo queste reti sono soprattutto quelle della compressione di immagini.

Oggi la ricerca nell'ambito delle reti neurali è molto sviluppata, sebbene addestrare una rete possa richiedere molte ore di lavoro e necessitare di macchine potenti.

5.3 Applicazione

Si è proceduto alla costruzione di una rete di tipo *feed-forward*, formata da un unico strato di variabili latenti e senza collegamenti tra le unità di ingresso e quelle di uscita¹.

La rete conta ventuno variabili di ingresso (Tab.4.22) e una variabile di uscita (la percezione del disagio), misurata su scala ordinale a quattro modalità:

- disagio nullo o fisiologico;
- disagio lieve, affrontabile con impiego di risorse esclusivamente familiari;
- disagio grave, tale da richiedere l'intervento esterno di istituzioni (servizi) o associazioni di volontariato;
- disagio molto grave, che permane anche dopo essere ricorsi all'aiuto esterno.

Un problema che si è incontrato è dato dalla determinazione dei parametri strutturali che compongono la rete, quali il numero di neuroni nascosti, le funzioni di attivazione da utilizzare, le modalità di connessione dei neuroni.

Le funzioni di attivazione utilizzate sono date dalle formule (5.2).

Per la scelta del parametro di *weight decay* λ e del numero di variabili che vanno a costituire lo strato latente, necessari per controllare l'apprendimento, si sono adottati due metodi.

¹In generale, è possibile definire più di uno strato di variabili latenti e, inoltre, stabilire connessioni dirette tra le variabili di input e quelle di output.

Il primo prevede la divisione dell'insieme di dati in due parti, una va a formare il cosiddetto *campione di stima*, costituito dalle prime 2000 famiglie e utilizzato per stimare la rete neurale, l'altra il *campione di verifica*, costituito dalle restanti 465 famiglie e utilizzato per il calcolo del tasso di errata classificazione² della rete stimata col campione di stima. Facendo variare il parametro λ tra 10^{-4} e 10^{-2} , in accordo con Venables e Ripley (1997), e il numero di variabili latenti tra 1 e 10, si sono considerate tutte le possibili combinazioni di questi due parametri e si è calcolato, in corrispondenza di ognuna, il tasso di errata classificazione della rete.

La seconda metodologia usata è la *10-fold cross-validation*. Il dataset iniziale d viene suddiviso casualmente in 10 insiemi d_1, \dots, d_{10} . Per valori di λ compresi tra 10^{-4} e 10^{-2} e del numero di variabili latenti tra 1 e 10, su ogni insieme $d - d_i$, ($i = 1, \dots, 10$) viene stimata la rete neurale e su ogni insieme d_i calcolato il corrispondente tasso di errata classificazione.

Per ciascuna delle due metodologie si è scelta la combinazione del parametro λ e del numero di variabili latenti che comportava il minor tasso di errata classificazione. Si sono confrontati i due tassi così ottenuti e si è adottato il modello di rete neurale corrispondente al tasso di errata classificazione minore.

Infine, per la scelta ottima dei pesi α e β si è applicato l'algoritmo di *backpropagation* (Ripley, 1996).

Scelti tutti i parametri, si costruisce la rete neurale sul dataset originale e si calcola l'errore di previsione³, come somma degli scarti al quadrato tra il valore osservato e il valore previsto della variabile risposta, relativi all' i -esima osservazione:

$$R(\theta) = \sum \sum (y_{ik} - f_k(x_i))^2 \quad (5.6)$$

La metodologia delle reti neurali è stata applicata al fine di confrontare la distribuzione osservata del disagio con la distribuzione del disagio prevista per una

²Il tasso di errata classificazione è stato calcolato come rapporto tra il numero di famiglie non correttamente classificate nella relativa categoria di disagio e il numero totale di famiglie presenti nel campione.

³L'errore di previsione calcolato sul campione di stima sarà sicuramente minore di quello calcolato sul campione di verifica, quello di diretto interesse, perché si usano gli stessi dati per la stima e la verifica.

serie di dataset simulati, ottenuti da quello osservato facendo variare alcune variabili esplicative. Le variabili esplicative prese in considerazione al riguardo sono il *numero di invalidi*, il *numero di malati cronico-degenerativi*, gli *eventuali problemi in famiglia*, il *tipo di famiglia*, il *numero di componenti* e l'*età media dei componenti*.

I dataset simulati sono andati a costituire l'input della rete neurale stimata per il dataset osservato e, per ciascuno, si sono calcolati l'errore di previsione e il tasso di errata classificazione, nonché la distribuzione delle famiglie rispetto alla percezione del disagio (si veda a tal proposito la tabella 5.1).

Indicati tali dataset simulati con $S1, \dots, S14$, essi si differenziano da quello osservato per le seguenti caratteristiche:

- S1: nessuna famiglia ha invalidi in casa;
- S2: nessuna famiglia ha più di un malato cronico-degenerativo in casa;
- S3: nessuna famiglia ha invalidi e più di un malato cronico-degenerativo in casa;
- S4: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale;
- S5: i problemi di tipo sanitario sono assimilati a problemi di tipo economico;
- S6: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale;
- S7: le famiglie unipersonali si trasformano in famiglie di tipo coppia senza figli;
- S8: le famiglie costituite da un genitore con figli a carico si trasformano in famiglie di tipo coppia con figli;
- S9: le famiglie con uno si trasformano in famiglie con due componenti;
- S10: le famiglie con due si trasformano in famiglie con tre componenti;

- S11: le famiglie con tre si trasformano in famiglie con quattro o più componenti;
- S12: le famiglie con componenti di età media compresa fra i 30 e i 50 anni si trasformano in famiglie con componenti di età media inferiore ai 30 anni;
- S13: le famiglie con componenti di età media compresa fra i 50 e i 70 anni si trasformano in famiglie con componenti di età media compresa tra i 30 e i 50 anni;
- S14: le famiglie con componenti di età media superiore ai 70 anni si trasformano in famiglie con componenti di età media compresa tra i 50 e i 70 anni.

Il confronto tra la distribuzione osservata del disagio con la distribuzione del disagio prevista per ciascun dataset simulato è stato fatto sia attraverso l'indice RIDIT (Par.3.2.1), sia attraverso i modelli a odds proporzionali (McCullagh, 1980).

Per questo secondo caso, si consideri la tabella 3.1 dove n_{kj} , ($k = 1, \dots, K$; $j = A, B$), rappresenta le frequenze assolute di due campioni da due popolazioni rispetto a gradazioni di gravità di fenomeni. Siano, inoltre, R_{kj} le somme cumulate di colonna, cosicché $n_j = R_{kj}$ è il totale della colonna j -esima. Sotto l'assunzione che ciascuna colonna abbia distribuzione multinomiale, la distribuzione marginale di R_{kj} condizionatamente al totale di colonna n_j è una binomiale di parametri n_j e $\gamma_k(\mathbf{x}_A)$, il secondo dei quali soddisfa la (3.14). Di qui:

$$\hat{\lambda}_{kj} = \log\left\{\left(R_{kj} + \frac{1}{2}\right) / \left(n_j - R_{kj} + \frac{1}{2}\right)\right\} \quad (5.7)$$

ha valore atteso $\lambda_{kj} + O(n_j^{-2})$ (Cox, 1970; Plackett, 1974).

Utilizzando come pesi (Clayton, 1974):

$$w_k \propto \gamma_k(1 - \gamma_k)(\pi_k + \pi_{k+1}) \quad (5.8)$$

dove γ_k è il valore comune di $\gamma_k(\mathbf{x}_A)$ e $\gamma_k(\mathbf{x}_B)$ sotto l'ipotesi in cui:

$$\Delta = \frac{\gamma_k(\mathbf{x}_B)/(1 - \gamma_k(\mathbf{x}_B))}{\gamma_k(\mathbf{x}_A)/(1 - \gamma_k(\mathbf{x}_A))} = 0 \quad (5.9)$$

e π_k e π_{k+1} le probabilità relative alle categorie k -esima e $(k+1)$ -esima, la varianza asintotica della stima di Δ è data da:

$$Var(\hat{\Delta}) = \left\{ \frac{n \cdot A \cdot n \cdot B}{n} \sum_{k=1}^{K-1} \gamma_k (1 - \gamma_k) (\pi_k + \pi_{k+1}) \right\}^{-1} + O(\Delta^2) \quad (5.10)$$

Con queste informazioni, è possibile calcolare l'odds ratio come $exp(\hat{\Delta})$ e il relativo intervallo di confidenza al 90% elevando all'esponente l'intervallo di confidenza per $\hat{\Delta}$, dato da $(\hat{\Delta} \pm 1,64s.e.(\hat{\Delta}))^4$.

Inoltre, dividendo $\hat{\Delta}^2$ per la sua varianza si ottiene un test chi-quadro che verifica l'ipotesi nulla di uguaglianza delle distribuzioni dei due gruppi (Clayton, 1974).

Il χ^2 così ottenuto ha un grado di libertà e lo si confronta con il test complessivo che verifica il grado di associazione tra le due distribuzioni, che ha due gradi di libertà.

Il pacchetto statistico utilizzato è *R* e la libreria a cui si è fatto riferimento è *nnet*, che richiede a sua volta la libreria *MASS*. Entrambe queste librerie sono state scritte da Venables & Ripley (1997). La funzione principale della prima è *nnet*, che consente di stimare reti neurali con un singolo strato nascosto e collegamenti tra le unità di ingresso e di uscita.

Il comportamento della rete è stato studiato sfruttando tutta la conoscenza relativa alla percezione del disagio, ossia considerando tutte e quattro le modalità di cui tale variabile è costituita.

Il tasso di errata classificazione più basso, e precisamente pari a 0,109, si ha in corrispondenza della scelta di un numero di variabili latenti pari a nove e di un valore di λ pari a 10^{-3} . A questi valori si è giunti ponendo il numero massimo di iterazioni a 1000 e implementando l'algoritmo a partire da diversi punti iniziali per evitare di incorrere in minimi locali della funzione obiettivo D_0 .

Nella tabella 5.1 si riportano i risultati ottenuti dall'adattamento del modello di analisi neurale al dataset osservato e a tutti i dataset simulati. In particolare,

⁴Nel caso in cui $K = 2$ l'espressione $\sum \gamma_k (1 - \gamma_k) (\pi_k + \pi_{k+1})$ si riduce alla formula della varianza di una binomiale $p(1 - p)$.

si presentano l'errore di previsione (e), il tasso di errata classificazione (err), la distribuzione delle frequenze assolute (n), l'indice ridit (r) e la relativa statistica test (t), l'odds ratio (\hat{OR}) e il relativo intervallo di confidenza (\hat{IC}), al 10%, 5% e 1%, e la statistica chi-quadro ($\hat{\chi}^2$), ottenuta dividendo $\hat{\Delta}^2$ per la sua varianza, con il relativo valore p .

Tab.5.1. *Errore di previsione (e), tasso di errata classificazione (err), frequenze assolute (n), indice ridit (r) e relativa significatività (t), Odds Ratio ($\hat{O}R$) e relativo Intervallo di Confidenza ($\hat{I}C$) e statistica chi-quadro ($\hat{\chi}^2$) e p-value, ottenuti come output della rete neurale con nove variabili nello strato latente e $\lambda = 10^{-3}$. La percezione del disagio è misurata su una scala a quattro modalità*

dataset	e	% err	n_1	n_2	n_3	n_4	r	t	$\hat{O}R$	90%- $\hat{I}C$	95%- $\hat{I}C$	99%- $\hat{I}C$	$\hat{\chi}^2$	p-value
S1	332,9	11,7	1874	508	70	13	0,479 ***	-2,4	0,459 ***	0,411 - 0,513	0,402 - 0,524	0,386 - 0,546	133,49 ***	< 0,0001
S2	297,2	11,3	1881	521	53	10	0,478 ***	-2,7	0,391 ***	0,350 - 0,437	0,343 - 0,446	0,328 - 0,465	193,53 ***	< 0,0001
S3	338,6	11,5	1885	497	68	15	0,477 ***	-2,6	0,481 ***	0,431 - 0,537	0,421 - 0,549	0,404 - 0,573	117,40 ***	< 0,0001
S4	567,0	17,5	1908	430	122	3	0,475 ***	-3,1	0,329 ***	0,296 - 0,366	0,289 - 0,374	0,278 - 0,389	288,82 ***	< 0,0001
S5	615,8	20,3	1981	410	71	1	0,458 ***	-5,1	0,193 ***	0,173 - 0,216	0,169 - 0,220	0,163 - 0,230	604,01 ***	< 0,0001
S6	387,6	11,6	1880	495	62	28	0,479 ***	-2,6	0,606 ***	0,545 - 0,674	0,533 - 0,688	0,512 - 0,716	59,62 ***	< 0,0001
S7	443,7	16,6	1999	452	4	10	0,452 ***	-5,8	0,211 ***	0,189 - 0,236	0,185 - 0,241	0,178 - 0,252	533,84 ***	< 0,0001
S8	436,1	15,0	1777	539	104	18	0,496	-0,5	0,329 ***	0,297 - 0,366	0,291 - 0,373	0,280 - 0,388	306,10 ***	< 0,0001
S9	425,7	12,9	1838	529	82	16	0,487 *	-1,6	0,535 ***	0,482 - 0,595	0,472 - 0,607	0,454 - 0,632	94,85 ***	< 0,0001
S10	432,7	12,3	1881	510	53	19	0,477 ***	-2,8	0,493 ***	0,443 - 0,549	0,434 - 0,560	0,417 - 0,583	118,26 ***	< 0,0001
S11	410,9	13,5	1810	528	115	11	0,493	-0,8	0,528 ***	0,475 - 0,586	0,466 - 0,598	0,448 - 0,622	100,69 ***	< 0,0001
S12	418,2	11,9	1900	480	56	29	0,475 ***	-3,1	0,592 ***	0,532 - 0,659	0,521 - 0,673	0,500 - 0,701	64,52 ***	< 0,0001
S13	393,6	11,7	1853	509	81	18	0,483 **	-2,0	0,552 ***	0,496 - 0,613	0,486 - 0,626	0,467 - 0,651	85,07 ***	< 0,0001
S14	446,6	12,4	1889	428	131	17	0,480 ***	-2,4	0,607 ***	0,546 - 0,676	0,535 - 0,690	0,514 - 0,718	58,86 ***	< 0,0001

Significatività: *** = 1%, ** = 5%, * = 10%

Osservando l'indice ridit con la relativa significatività, si nota come tutti sottolineino un'evidente differenza tra la distribuzione del disagio nel dataset osservato e quella nei dataset simulati, tranne quelli relativi all'ottavo e all'undicesimo dataset simulati, che non risultano significativi. Quest'ultimo caso indica che non vi è differenza, in termini di percezione di disagio, tra le famiglie di tipo coppia con figli e quelle costituite da un genitore con figli e tra le famiglie con tre e quelle con quattro o più componenti.

Dal momento che, come osservato nel capitolo 3, l'indice ridit non è tra i più adatti per trattare i dati a nostra disposizione, ad esso si è affiancato il calcolo dell'odds ratio.

Tutti gli odds e i rispettivi intervalli di confidenza mostrano significativa differenza tra la distribuzione del disagio nel dataset osservato e quella nei dataset simulati.

Ad esempio, se tutti i problemi sanitari rilevati fossero in realtà problemi economici il rischio di disagio diminuirebbe di oltre cinque volte ($\hat{OR} = 0,193$; $90\% - \hat{IC} = (0,173 - 0,216)$). Si noti, tuttavia, che per questo dataset si registra il tasso di errata classificazione più alto ($err = 20,3$).

Ancora, se tutte le famiglie unipersonali fossero famiglie costituite da una coppia senza figli, il rischio di disagio diminuirebbe di quasi cinque volte ($\hat{OR} = 0,211$; $90\% - \hat{IC} = (0,189 - 0,236)$). Anche in questo caso il tasso di errata classificazione è piuttosto elevato ($err = 16,6$).

La statistica chi-quadro conferma i risultati ottenuti con i modelli a odds proporzionali, mettendo in luce, in corrispondenza di ciascuno dei dataset simulati, la diversità tra le distribuzioni dei due gruppi.

Se si considerano livelli di significatività del 5% e dell'1% per il calcolo degli intervalli di confidenza e della significatività della statistica chi-quadro, i risultati che si ottengono portano alle medesime conclusioni del caso $\alpha = 0,10$, nonostante l'intervallo sia, ovviamente, più ampio. Anche la statistica chi-quadro rimane significativa per tutti i dataset simulati. La statistica t , ad un livello del 5%, non risulta più significativa per il dataset S9, ossia non sembra esserci differenza, in termini di disagio, tra le famiglie con uno e con due componenti, e ad un livello

dell'1% nemmeno per il dataset S13, che confronta le famiglie con componenti di età media compresa fra i 50 e i 70 anni e quelle con componenti di età media tra i 30 e i 50 anni.

La stessa analisi è stata condotta studiando la percezione del disagio su scala tricotomica:

- disagio nullo o lieve;
- disagio grave;
- disagio molto grave.

e dicotomica:

- disagio nullo o lieve;
- disagio grave o molto grave.

Nel primo caso la rete neurale ottima è costituita da otto variabili nello strato latente e un valore di λ pari a 10^{-3} . Il tasso di errata classificazione è pari a 0,101. Nel secondo, il numero di variabili latenti è pari a 10 e λ a 10^{-4} . Il tasso di errata classificazione ammonta a 0,030.

Nella tabella 5.2 si riportano i risultati ottenuti misurando il disagio su scala tricotomica e considerando un livello di significatività del 10%, del 5% e dell'1% rispettivamente.

Tab.5.2. *Errore di previsione (e), tasso di errata classificazione (err), distribuzione di frequenze assolute (n), indice ridit (r) e relativa significatività (t), Odds Ratio (OR) e relativo Intervallo di Confidenza (IC) e statistica chi-quadro (χ^2), ottenuti come output della rete neurale con otto variabili nello strato latente e $\lambda = 10^{-3}$. La percezione del disagio è misurata su una scala a tre modalità*

dataset	e	% err	n_1	n_2	n_3	r	t	\hat{OR}	90%- \hat{IC}	95%- \hat{IC}	99%- \hat{IC}	$\hat{\chi}^2$	p-value
S1	103,6	4,4	2372	73	20	0,479 ***	-2,4	0,292 ***	0,238 - 0,360	0,228 - 0,375	0,211 - 0,405	94,40 ***	< 0,0001
S2	116,5	5,0	2359	93	13	0,482 **	-2,1	0,253 ***	0,207 - 0,310	0,198 - 0,323	0,184 - 0,349	122,41 ***	< 0,0001
S3	103,7	4,3	2346	104	15	0,484 **	-1,8	0,291 ***	0,238 - 0,356	0,229 - 0,370	0,213 - 0,399	102,35 ***	< 0,0001
S4	160,5	5,1	2375	53	36	0,479 ***	-2,6	0,383 ***	0,312 - 0,469	0,300 - 0,488	0,278 - 0,527	60,00 ***	< 0,0001
S5	173,3	6,0	2364	85	15	0,481 ***	-2,4	0,264 ***	0,216 - 0,322	0,208 - 0,335	0,193 - 0,361	119,99 ***	< 0,0001
S6	154,7	5,2	2357	80	28	0,482 **	-2,2	0,373 ***	0,306 - 0,454	0,295 - 0,472	0,274 - 0,509	67,35 ***	< 0,0001
S7	57,6	1,5	2405	26	34	0,473 ***	-3,2	0,304 ***	0,245 - 0,376	0,235 - 0,393	0,217 - 0,426	83,12 ***	< 0,0001
S8	151,4	5,0	2372	67	26	0,479 ***	-2,5	0,333 ***	0,272 - 0,407	0,261 - 0,424	0,242 - 0,457	79,87 ***	< 0,0001
S9	169,8	5,5	2379	42	44	0,478 ***	-2,6	0,416 ***	0,339 - 0,510	0,326 - 0,531	0,301 - 0,573	49,59 ***	< 0,0001
S10	165,4	6,4	2312	147	6	0,490	-1,2	0,213 ***	0,177 - 0,256	0,171 - 0,266	0,159 - 0,285	189,34 ***	< 0,0001
S11	158,7	5,6	2348	65	42	0,482 **	-2,1	0,457 ***	0,375 - 0,556	0,361 - 0,578	0,335 - 0,623	42,40 ***	< 0,0001
S12	191,6	6,5	2324	128	13	0,488 *	-1,4	0,294 ***	0,244 - 0,355	0,235 - 0,368	0,219 - 0,395	114,46 ***	< 0,0001
S13	160,4	5,6	2352	93	20	0,483 **	-2,1	0,323 ***	0,266 - 0,393	0,256 - 0,409	0,238 - 0,440	89,70 ***	< 0,0001
S14	165,6	5,7	2316	133	16	0,490	-1,2	0,355 ***	0,279 - 0,404	0,269 - 0,419	0,251 - 0,449	93,45 ***	< 0,0001

Significatività: *** = 1%, ** = 5%, * = 10%

Rispetto all'analisi effettuata misurando il disagio su scala a quattro modalità, nell'analisi tricotomica si nota come gli intervalli di confidenza degli odds siano più lontani dal valore 1, sottolineando una più marcata diversità, in termini di percezione disagio, tra le famiglie di volta in volta confrontate. Ad esempio, se tutte le famiglie con due fossero famiglie costituite da tre componenti, il rischio di disagio diminuirebbe di quasi cinque volte ($\hat{OR} = 0,213$; $90\% - \hat{IC} = (0,177 - 0,256)$) Misurando il disagio su scala a quattro modalità, la diminuzione di disagio passando dalla prima tipologia di famiglia alla seconda è di poco più di due volte ($\hat{OR} = 0,493$; $90\% - \hat{IC} = (0,443 - 0,549)$).

Inoltre, a differenza della scala a quattro modalità, misurando il disagio su scala tricotomica, l'indice ridit risulta sempre significativo, tranne nel confronto tra le famiglie con due e quelle con tre componenti (S10) e tra le famiglie con componenti di età media superiore ai 70 anni e quelle con componenti di età media compresa tra i 50 e i 70 anni (S14).

La statistica t assume tendenzialmente valori minori, in termini assoluti, rispetto al caso di variabile risposta a quattro modalità, mostrando in modo meno evidente la diversità tra ciascuna coppia di famiglie.

Ad un livello di significatività del 5% e dell'1% i risultati ottenuti dal calcolo degli intervalli di confidenza degli odds e della statistica chi-quadro portano alle stesse conclusioni a cui si è giunti ponendo $\alpha = 0,10$. Ciò che varia è l'interpretazione della statistica t : mentre ad un livello del 10% la mancanza di significatività per tale statistica è evidente solo nel confronto tra il dataset originale e i dataset S10 e S14, ad un livello del 5% essa perde la significatività anche nel confronto tra il dataset originale e il dataset S12 e all'1% anche nel confronto tra il dataset originale e i dataset S2, S3, S6, S11 e S13. Così, ad un livello del 5%, non risulta esserci differenza in termini di percezione di disagio tra le famiglie con componenti di età media compresa fra i 30 e i 50 anni e quelle con componenti di età media inferiore ai 30 anni. La variabile *età* si rileva, pertanto, non decisiva nel discriminare le famiglie rispetto alla variabile di interesse. Variabili che risultano ancor meno rilevanti (perché non significative all'1%) nel distinguere le famiglie disagiate da quelle che non si ritengono tali, sono la *presenza di malati cronico-degenerativi in*

famiglia, la presenza contemporanea di malati e invalidi, la tipologia di problemi avvertiti dalla famiglia e il numero di componenti in famiglia.

Nella tabella 5.3 si riportano i risultati ottenuti misurando il disagio su scala dicotomica e considerando un livello di significatività del 10%, del 5% e dell'1% rispettivamente.

Tab.5.3. *Errore di previsione (e), tasso di errata classificazione (err), distribuzione di frequenze assolute (n), indice ridit (r) e relativa significatività (t), Odds Ratio (\hat{OR}) e relativo Intervallo di Confidenza (\hat{IC}) e statistica chi-quadro ($\hat{\chi}^2$), ottenuti come output della rete neurale con dieci variabili nello strato latente e $\lambda = 10^{-4}$. La percezione del disagio è misurata su scala dicotomica*

dataset	e	% err	n_1	n_2	r	t	\hat{OR}	90%- \hat{IC}	95%- \hat{IC}	99%- \hat{IC}	$\hat{\chi}^2$	p - value
S1	63,3	3,3	2366	99	0,482 **	-2,1	0,507 ***	0,413 - 0,622	0,396 - 0,647	0,367 - 0,699	29,60 ***	< 0,0001
S2	55,3	3,3	2364	101	0,482 **	-2,1	0,519 ***	0,423 - 0,636	0,406 - 0,662	0,376 - 0,715	27,77 ***	< 0,0001
S3	75,9	4,7	2410	55	0,473 ***	-3,2	0,276 ***	0,222 - 0,344	0,212 - 0,359	0,195 - 0,390	92,12 ***	< 0,0001
S4	83,4	3,8	2329	136	0,489 *	-1,3	0,704 ***	0,583 - 0,850	0,562 - 0,881	0,523 - 0,947	9,35 ***	0,0093
S5	110,8	5,6	2405	60	0,474 ***	-3,2	0,302 ***	0,244 - 0,374	0,234 - 0,390	0,216 - 0,423	84,67 ***	< 0,0001
S6	100,4	5,0	2366	99	0,482 **	-2,2	0,505 ***	0,414 - 0,616	0,398 - 0,641	0,369 - 0,691	31,62 ***	< 0,0001
S7	17,9	0,9	2427	38	0,469 ***	-3,7	0,190 ***	0,152 - 0,238	0,146 - 0,249	0,134 - 0,270	148,81 ***	< 0,0001
S8	75,5	3,5	2330	134	0,489 *	-1,4	0,693 ***	0,574 - 0,837	0,553 - 0,868	0,515 - 0,933	10,15 ***	0,0062
S9	100,3	4,9	2362	102	0,482 **	-2,1	0,521 ***	0,427 - 0,635	0,411 - 0,660	0,382 - 0,712	29,06 ***	< 0,0001
S10	103,1	5,1	2334	131	0,488 *	-1,4	0,677 ***	0,560 - 0,818	0,539 - 0,849	0,502 - 0,912	11,41 ***	0,0033
S11	85,1	4,5	2307	158	0,494	-0,8	0,825 *	0,687 - 0,990	0,663 - 1,026	0,619 - 1,100	2,98	0,2256
S12	83,7	4,1	2308	154	0,493	-0,9	0,804 *	0,669 - 0,966	0,646 - 1,001	0,602 - 1,073	3,80	0,1498
S13	100,4	5,2	2333	132	0,488 *	-1,4	0,682 ***	0,564 - 0,824	0,544 - 0,855	0,506 - 0,919	10,98 ***	0,0041
S14	91,8	4,6	2359	106	0,483 **	-2,0	0,542 ***	0,445 - 0,660	0,428 - 0,686	0,398 - 0,739	25,97 ***	< 0,0001

Significatività: *** = 1%, ** = 5%, * = 10%

Rispetto all'analisi effettuata misurando il disagio su scala a quattro modalità, la statistica chi-quadro relativa ai dataset S11 e S12 risulta non significativa ad un livello del 10%, mostrando l'uguaglianza, in termini di percezione di disagio, tra le famiglie con tre e quelle con quattro o più componenti ($\hat{\chi}^2 = 2,98$, $p - value = 0,2256$) e tra le famiglie con componenti di età media compresa fra i 30 e i 50 anni e quelle in cui l'età media è inferiore ai 30 anni ($\hat{\chi}^2 = 3,80$, $p - value = 0,1498$).

L'odds ratio relativo all'undicesimo dataset simulato è al limite della significatività ($\hat{OR} = 0,825$; $90\% - IC = (0,687 - 0,990)$): esso mostra una differenza tra le famiglie con tre e quelle con quattro o più componenti, ma tale differenza è molto meno marcata rispetto al caso di disagio misurato su scala a quattro modalità.

Ciò che emerge dai risultati è, inoltre, il fatto che, se tutte le famiglie unipersonali fossero famiglie costituite da una coppia senza figli, il rischio di disagio diminuirebbe di oltre cinque volte ($\hat{OR} = 0,190$; $90\% - IC = (0,152 - 0,238)$).

Quasi tutti gli odds ratio assumono un valore superiore rispetto a quelli ottenuti dall'analisi a quattro modalità, sottolineando in maniera meno marcata la differenza tra le due tipologie di famiglie di volta in volta confrontate. Ad esempio, l'odds ratio che verifica la differenza tra le famiglie con problemi di tipo socio-relazionale e quelle con problemi di tipo sanitario risulta pari a $\hat{OR} = 0,704$, contro $\hat{OR} = 0,329$, calcolato in corrispondenza di una scala di disagio su quattro modalità, evidenziando in modo meno forte come il disagio diminuisca se i problemi sanitari sono socio-relazionali.

Solo in corrispondenza dei dataset S3, S6 e S7, l'odds ratio assume un valore maggiore nel caso di disagio misurato su scala a quattro modalità rispetto a scala dicotomica. Così, mentre nel caso di scala a quattro modalità il rischio di disagio diminuisce poco più di due volte nelle famiglie che non hanno invalidi e al più un malato cronico-degenerativo in casa rispetto a quelle che hanno almeno un invalido e due o più malati ($\hat{OR} = 0,481$; $90\% - IC = (0,431 - 0,537)$), nel caso di scala dicotomica tale rischio diminuisce di quasi quattro volte ($\hat{OR} = 0,276$; $90\% - IC = (0,222 - 0,344)$).

Infine, la statistica chi-quadro assume valori di gran lunga minori nel caso dicotomico rispetto al caso di scala su quattro modalità.

Questo risultato è dovuto al fatto che, aggregando le modalità della variabile che misura la percezione del disagio, le famiglie tendono ad essere distribuite in maniera più omogenea nelle diverse categorie di gravità e si perdono, pertanto, informazioni.

Ad un livello di significatività del 5%, osservando sia l'intervallo di confidenza dell'odds ratio che il valore p della statistica $\hat{\chi}^2$, risulta non significativa la differenza, in termini di percezione di disagio, tra le famiglie con tre e quelle con quattro componenti (95% - $\hat{IC} = (0,663 - 1,026)$); $p\text{-value} = 0,2256$) e tra le famiglie con componenti di età media compresa fra i 30 e i 50 anni e quelle in cui l'età media è inferiore ai 30 anni (95% - $\hat{IC} = (0,646 - 1,001)$); $p\text{-value} = 0,1498$). Questo risultato, ad un livello di significatività del 10% veniva evidenziato solamente dalla statistica chi-quadro, mentre gli intervalli di confidenza, sebbene prossimi a uno, non includevano tale valore. La statistica t , inoltre, non mette più in evidenza la differenza nella percezione di disagio tra le tipologie di famiglie messe a confronto nei dataset S4, S8, S10 e S13.

Ad un livello di significatività dell'1% le conclusioni a cui è possibile giungere sono le medesime di quelle ottenute ponendo $\alpha = 0,05$, sebbene gli intervalli di confidenza per gli odds siano più ampi. Per quanto riguarda la statistica t , invece, emerge l'uguaglianza tra le coppie di famiglie prese in considerazione nei dataset S1, S2, S6, S9 e S14, oltre a quella evidenziata ad un livello di significatività del 10% e del 5%.

In un secondo momento, si sono costruiti altri dataset simulati, ottenuti da quello osservato modificando, per ciascun dataset, le stesse variabili considerate nelle precedenti simulazioni, ma a livello combinato.

Indicati tali dataset con $S15, \dots, S56$, essi si caratterizzano nel modo seguente:

- S15: nessuna famiglia ha invalidi in casa e tutte le famiglie costituite da un genitore con figli a carico si trasformano in famiglie di tipo coppia con figli;
- S16: nessuna famiglia ha invalidi in casa e tutte le famiglie con due si trasformano in famiglie con tre componenti;

- S17: nessuna famiglia ha invalidi in casa e tutte le famiglie con tre si trasformano in famiglie con quattro o più componenti;
- S18: nessuna famiglia ha invalidi in casa e tutte le famiglie con componenti di età media compresa fra i 30 e i 50 anni si trasformano in famiglie con componenti di età media inferiore ai 30 anni;
- S19: nessuna famiglia ha invalidi in casa e tutte le famiglie con componenti di età media compresa fra i 50 e i 70 anni si trasformano in famiglie con componenti di età media compresa tra i 30 e i 50 anni;
- S20: nessuna famiglia ha invalidi in casa e tutte le famiglie con componenti di età media superiore ai 70 anni si trasformano in famiglie con componenti di età media compresa tra i 50 e i 70 anni;
- S21: nessuna famiglia ha più di un malato cronico-degenerativo in casa e tutte le famiglie costituite da un genitore con figli a carico si trasformano in famiglie di tipo coppia con figli;
- S22: nessuna famiglia ha più di un malato cronico-degenerativo in casa e tutte le famiglie con due si trasformano in famiglie con tre componenti;
- S23: nessuna famiglia ha più di un malato cronico-degenerativo in casa e tutte le famiglie con tre si trasformano in famiglie con quattro o più componenti;
- S24: nessuna famiglia ha più di un malato cronico-degenerativo in casa e tutte le famiglie con componenti di età media compresa fra i 30 e i 50 anni si trasformano in famiglie con componenti di età media inferiore ai 30 anni;
- S25: nessuna famiglia ha più di un malato cronico-degenerativo in casa e tutte le famiglie con componenti di età media compresa fra i 50 e i 70 anni si trasformano in famiglie con componenti di età media compresa tra i 30 e i 50 anni;

- S26: nessuna famiglia ha più di un malato cronico-degenerativo in casa e tutte le famiglie con componenti di età media superiore ai 70 anni si trasformano in famiglie con componenti di età media compresa tra i 50 e i 70 anni;
- S27: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie unipersonali si trasformano in famiglie di tipo coppia senza figli;
- S28: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie costituite da un genitore con figli a carico si trasformano in famiglie di tipo coppia con figli;
- S29: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con uno si trasformano in famiglie con due componenti;
- S30: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con due si trasformano in famiglie con tre componenti;
- S31: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con tre si trasformano in famiglie con quattro o più componenti;
- S32: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con componenti di età media compresa fra i 30 e i 50 anni si trasformano in famiglie con componenti di età media inferiore ai 30 anni;
- S33: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con componenti di età media compresa fra i 50 e i 70 anni si trasformano in famiglie con componenti di età media compresa tra i 30 e i 50 anni;

- S34: i problemi di tipo sanitario sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con componenti di età media superiore ai 70 anni si trasformano in famiglie con componenti di età media compresa tra i 50 e i 70 anni;
- S35: i problemi di tipo sanitario sono assimilati a problemi di tipo economico e tutte le famiglie unipersonali si trasformano in famiglie di tipo coppia senza figli;
- S36: i problemi di tipo sanitario sono assimilati a problemi di tipo economico e tutte le famiglie costituite da un genitore con figli a carico si trasformano in famiglie di tipo coppia con figli;
- S37: i problemi di tipo sanitario sono assimilati a problemi di tipo economico e tutte le famiglie con uno si trasformano in famiglie con due componenti;
- S38: i problemi di tipo sanitario sono assimilati a problemi di tipo economico e tutte le famiglie con due si trasformano in famiglie con tre componenti;
- S39: i problemi di tipo sanitario sono assimilati a problemi di tipo economico e tutte le famiglie con tre si trasformano in famiglie con quattro o più componenti;
- S40: i problemi di tipo sanitario sono assimilati a problemi di tipo economico e tutte le famiglie con componenti di età media compresa fra i 30 e i 50 anni sono assimilate a famiglie con componenti di età media inferiore ai 30 anni;
- S41: i problemi di tipo sanitario sono assimilati a problemi di tipo economico e tutte le famiglie con componenti di età media compresa fra i 50 e i 70 anni sono assimilate a famiglie con componenti di età media compresa tra i 30 e i 50 anni;
- S42: i problemi di tipo sanitario sono assimilati a problemi di tipo economico e tutte le famiglie con componenti di età media superiore ai 70 anni sono assimilate a famiglie con componenti di età media compresa tra i 50 e i 70 anni;

- S43: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie unipersonali sono assimilate a famiglie di tipo coppia senza figli;
- S44: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie costituite da un genitore con figli a carico sono assimilate a famiglie di tipo coppia con figli;
- S45: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con uno sono assimilate a famiglie con due componenti;
- S46: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con due sono assimilate a famiglie con tre componenti;
- S47: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con tre sono assimilate a famiglie con quattro o più componenti;
- S48: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con componenti di età media compresa fra i 30 e i 50 anni si trasformano in famiglie con componenti di età media inferiore ai 30 anni;
- S49: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con componenti di età media compresa fra i 50 e i 70 anni si trasformano in famiglie con componenti di età media compresa tra i 30 e i 50 anni;
- S50: i problemi di tipo economico sono assimilati a problemi di tipo socio-relazionale e tutte le famiglie con componenti di età media superiore ai 70 anni si trasformano in famiglie con componenti di età media compresa tra i 50 e i 70 anni;

- S51: le famiglie unipersonali si trasformano in famiglie di tipo coppia senza figli e quelle con componenti di età media compresa fra i 30 e i 50 anni in famiglie con componenti di età media inferiore ai 30 anni;
- S52: le famiglie unipersonali si trasformano in famiglie di tipo coppia senza figli e quelle con componenti di età media compresa fra i 50 e i 70 anni in famiglie con componenti di età media compresa tra i 30 e i 50 anni;
- S53: le famiglie unipersonali si trasformano in famiglie di tipo coppia senza figli e quelle con componenti di età media superiore ai 70 anni si considerano famiglie con componenti di età media compresa tra i 50 e i 70 anni;
- S54: tutte le famiglie costituite da un genitore con figli a carico si trasformano in famiglie di tipo coppia con figli e quelle con componenti di età media compresa fra i 30 e i 50 anni in famiglie con componenti di età media inferiore ai 30 anni;
- S55: tutte le famiglie costituite da un genitore con figli a carico si trasformano in famiglie di tipo coppia con figli e quelle con componenti di età media compresa fra i 50 e i 70 anni in famiglie con componenti di età media compresa tra i 30 e i 50 anni;
- S56: tutte le famiglie costituite da un genitore con figli a carico si trasformano in famiglie di tipo coppia con figli e quelle con componenti di età media superiore ai 70 anni in famiglie con componenti di età media compresa tra i 50 e i 70 anni.

Tali dataset sono stati usati come input della rete neurale con nove variabili nello stato latente e $\lambda = 10^{-3}$, ovvero della rete neurale stimata come ottima, in termini di minimizzazione del tasso di errata classificazione, per il dataset osservato.

Nella tabella 5.4 si riportano i risultati ottenuti.

Tab.5.4. *Errore di previsione (e), tasso di errata classificazione (err), distribuzione di frequenze assolute (n), indice ridit (r) e relativa significatività (t), Odds Ratio (\hat{OR}) e relativo Intervallo di Confidenza (\hat{IC}) e statistica chi-quadro ($\hat{\chi}^2$), ottenuti come output della rete neurale con nove variabili nello strato latente e $\lambda = 10^{-3}$. La percezione del disagio è misurata su una scala a quattro modalità*

dataset	e	% err	n_1	n_2	n_3	n_4	r	t	\hat{OR}	90%- \hat{IC}	95%- \hat{IC}	99%- \hat{IC}	$\hat{\chi}^2$	p-value
S15	325,0	10,9	1917	465	71	12	0,471 ***	-3,4	0,432 ***	0,386 - 0,483	0,378 - 0,493	0,362 - 0,515	152,12 ***	< 0,0001
S16	344,3	11,5	1894	490	68	13	0,475 ***	-2,9	0,448 ***	0,401 - 0,500	0,392 - 0,511	0,376 - 0,533	140,88 ***	< 0,0001
S17	343,1	12,3	1962	413	78	12	0,463 ***	-4,4	0,411 ***	0,367 - 0,460	0,359 - 0,471	0,344 - 0,491	166,30 ***	< 0,0001
S18	335,5	11,9	1939	477	33	16	0,478 ***	-4,0	0,395 ***	0,353 - 0,442	0,345 - 0,452	0,331 - 0,471	183,73 ***	< 0,0001
S19	323,3	12,4	1883	490	82	10	0,464 ***	-2,6	0,444 ***	0,398 - 0,496	0,389 - 0,507	0,373 - 0,529	144,65 ***	< 0,0001
S20	375,5	12,3	1948	459	50	12	0,475 ***	-4,2	0,381 ***	0,340 - 0,426	0,333 - 0,436	0,319 - 0,455	197,42 ***	< 0,0001
S21	293,4	10,7	1886	506	59	13	0,479 ***	-2,9	0,403 ***	0,360 - 0,450	0,353 - 0,460	0,338 - 0,479	181,29 ***	< 0,0001
S22	268,7	10,4	1884	466	105	10	0,485 ***	-2,4	0,480 ***	0,430 - 0,536	0,420 - 0,548	0,403 - 0,571	118,47 ***	< 0,0001
S23	277,5	11,8	1871	417	162	15	0,488 **	-1,8	0,635 ***	0,569 - 0,709	0,557 - 0,725	0,534 - 0,756	45,46 ***	< 0,0001
S24	222,1	9,0	1835	520	92	17	0,481 *	-1,4	0,576 ***	0,516 - 0,642	0,505 - 0,656	0,485 - 0,684	68,66 ***	< 0,0001
S25	329,5	12,4	1858	549	49	8	0,483 ***	-2,2	0,355 ***	0,318 - 0,397	0,312 - 0,405	0,299 - 0,423	237,48 ***	< 0,0001
S26	272,7	11,0	1864	485	100	15	0,477 **	-2,0	0,547 ***	0,490 - 0,611	0,480 - 0,624	0,460 - 0,651	80,52 ***	< 0,0001
S27	536,7	17,0	1900	420	136	6	0,468 ***	-2,8	0,426 ***	0,382 - 0,474	0,374 - 0,484	0,360 - 0,504	171,31 ***	< 0,0001
S28	510,4	15,9	1930	448	66	19	0,489 ***	-3,9	0,492 ***	0,441 - 0,548	0,432 - 0,559	0,415 - 0,583	116,13 ***	< 0,0001
S29	491,8	16,5	1832	472	140	10	0,466 *	-1,4	0,516 ***	0,464 - 0,573	0,455 - 0,585	0,437 - 0,609	106,39 ***	< 0,0001
S30	531,4	16,7	1959	349	145	9	0,467 ***	-4,1	0,474 ***	0,425 - 0,529	0,416 - 0,540	0,400 - 0,563	126,18 ***	< 0,0001
S31	509,0	15,1	1948	404	92	21	0,470 ***	-4,0	0,565 ***	0,507 - 0,630	0,497 - 0,644	0,477 - 0,671	74,35 ***	< 0,0001
S32	513,2	15,4	1927	429	85	21	0,471 ***	-3,6	0,562 ***	0,504 - 0,626	0,494 - 0,639	0,474 - 0,666	76,80 ***	< 0,0001
S33	544,7	16,4	1919	451	78	17	0,487 ***	-3,5	0,508 ***	0,457 - 0,566	0,447 - 0,578	0,429 - 0,602	106,30 ***	< 0,0001
S34	489,4	16,5	1847	475	129	13	0,475 *	-1,6	0,566 ***	0,509 - 0,629	0,499 - 0,642	0,479 - 0,668	78,51 ***	< 0,0001

dataset	e	% err	n_1	n_2	n_3	n_4	r	t	\hat{OR}	90%- \hat{IC}	95%- \hat{IC}	99%- \hat{IC}	$\hat{\chi}^2$	p - value
S35	580,1	20,2	1898	452	103	7	0,477 ***	-3,0	0,409 ***	0,367 - 0,455	0,360 - 0,464	0,345 - 0,484	187,76 ***	< 0,0001
S36	614,1	21,0	1895	443	123	1	0,463 ***	-2,8	0,250 ***	0,224 - 0,278	0,220 - 0,284	0,211 - 0,295	452,96 ***	< 0,0001
S37	590,3	19,3	1962	410	93	0	0,454 ***	-4,5	0,149 ***	0,133 - 0,166	0,130 - 0,169	0,125 - 0,176	822,80 ***	< 0,0001
S38	603,2	19,5	2001	398	54	11	0,465 ***	-5,5	0,362 ***	0,324 - 0,404	0,317 - 0,413	0,304 - 0,431	227,71 ***	< 0,0001
S39	613,3	20,5	1950	414	99	0	0,460 ***	-4,2	0,153 ***	0,137 - 0,171	0,135 - 0,175	0,129 - 0,182	801,40 ***	< 0,0001
S40	594,2	18,9	1980	381	93	10	0,479 ***	-4,8	0,419 ***	0,375 - 0,467	0,367 - 0,478	0,353 - 0,498	169,51 ***	< 0,0001
S41	597,0	21,6	1877	469	109	4	0,463 ***	-2,5	0,353 ***	0,318 - 0,393	0,311 - 0,401	0,299 - 0,418	256,73 ***	< 0,0001
S42	587,3	19,4	1964	402	86	13	0,482 ***	-4,5	0,456 ***	0,409 - 0,508	0,400 - 0,519	0,384 - 0,541	139,66 ***	< 0,0001
S43	420,8	12,6	1868	479	87	30	0,483 **	-2,2	0,685 ***	0,616 - 0,761	0,603 - 0,777	0,579 - 0,809	34,30 ***	< 0,0001
S44	447,4	15,2	1850	522	89	0	0,483 **	-2,0	0,159 ***	0,143 - 0,177	0,141 - 0,181	0,135 - 0,188	810,86 ***	< 0,0001
S45	415,7	13,3	1862	494	89	19	0,484 **	-2,1	0,575 ***	0,517 - 0,640	0,507 - 0,653	0,487 - 0,680	73,27 ***	< 0,0001
S46	435,5	12,9	1859	474	120	11	0,479 **	-1,9	0,517 ***	0,465 - 0,574	0,455 - 0,586	0,437 - 0,610	104,82 ***	< 0,0001
S47	382,3	11,8	1879	480	83	20	0,483 ***	-2,5	0,568 ***	0,511 - 0,632	0,500 - 0,645	0,480 - 0,672	75,89 ***	< 0,0001
S48	413,2	12,6	1853	521	74	13	0,476 **	-2,1	0,475 ***	0,427 - 0,528	0,418 - 0,539	0,402 - 0,561	133,40 ***	< 0,0001
S49	422,2	12,7	1902	433	115	14	0,475 ***	-2,9	0,523 ***	0,470 - 0,583	0,461 - 0,595	0,442 - 0,619	98,35 ***	< 0,0001
S50	422,9	12,1	1900	463	85	16	0,475 ***	-3,0	0,517 ***	0,464 - 0,575	0,454 - 0,587	0,436 - 0,611	102,45 ***	< 0,0001
S51	432,0	12,7	1874	511	69	11	0,479 ***	-2,5	0,431 ***	0,387 - 0,479	0,379 - 0,489	0,364 - 0,509	168,77 ***	< 0,0001
S52	380,7	11,4	1886	487	62	29	0,477 ***	-2,7	0,612 ***	0,550 - 0,681	0,539 - 0,696	0,518 - 0,724	56,90 ***	< 0,0001
S53	403,7	13,7	1820	477	138	29	0,493	-0,8	0,783 ***	0,705 - 0,870	0,691 - 0,888	0,664 - 0,923	14,65 ***	0,0007
S54	393,9	11,8	1881	499	58	25	0,478 ***	-2,7	0,567 ***	0,509 - 0,630	0,499 - 0,644	0,479 - 0,670	76,38 ***	< 0,0001
S55	446,0	13,5	1859	489	101	14	0,483 **	-2,0	0,533 ***	0,480 - 0,593	0,470 - 0,605	0,452 - 0,630	94,83 ***	< 0,0001
S56	433,3	12,7	1877	514	66	6	0,478 ***	-2,7	0,342 ***	0,528 - 0,381	0,301 - 0,389	0,290 - 0,405	272,56 ***	< 0,0001

Significatività: *** = 1%, ** = 5%, * = 10%

Rispetto all'analisi dei dataset S_1, \dots, S_{14} , modificando le variabili a coppie la statistica t è maggiore in valore assoluto, evidenziando in modo più marcato la differenza tra le famiglie oggetto del confronto.

L'indice ridit risulta non significativo solo per i dataset S53 ($r = 0,493$).

Anche gli odds ratio mettono in evidenza, per ciascun confronto, la differenza tra la distribuzione del disagio nel dataset osservato e quella di ciascuno dei dataset simulati.

Sia a livello 10% che a livelli 5% e 1%, i loro valori sono inferiori rispetto a quelli calcolati in corrispondenza dei primi quattordici dataset simulati, indice del fatto che la modifica contemporanea di due variabili sottolinea in modo più marcato la differenza tra la distribuzione osservata del disagio e quella stimata per ciascun dataset simulato. Gli intervalli di confidenza sono più lontani da uno, valore che rappresenta l'indipendenza tra le due distribuzioni, e, anzi, nessuno contiene il valore uno, a dimostrazione della evidente differenza, in termini di disagio, tra le famiglie confrontate.

Ad esempio, ad un livello del 10%, se i problemi di tipo sanitario fossero assimilati a problemi di tipo economico e tutte le famiglie con uno si trasformassero in famiglie con due componenti, il rischio di disagio diminuirebbe di quasi sette volte ($\hat{OR} = 0,149; 90\% - \hat{IC} = (0,133 - 0,166)$). Come si può osservare dai risultati nella tabella 5.1, tale differenza in distribuzione è dovuta principalmente alla "trasformazione" dei problemi di tipo sanitario in problemi di tipo economico, piuttosto che da quella delle famiglie con uno in famiglie con due componenti ($\hat{OR} = 0,193$ per il dataset S5 e $\hat{OR} = 0,535$ per S9).

Anche la statistica chi-quadro sottolinea la diversità in distribuzione tra ciascuna coppia di gruppi.

Clayton (1974) propone un'alternativa per il calcolo degli odds ratio, adatta qualora la probabilità cumulata γ_k assuma un valore troppo vicino a 0 o a 1, ossia qualora le categorie estreme della variabile di interesse siano caratterizzate da frequenze molto basse. Tale approccio è analogo alla procedura di Mantel-Haenszel per la combinazione degli odds ratio di più tabelle di contingenza 2×2

(Mantel-Haenszel, 1959).

Indicati con \mathbf{x}_A e \mathbf{x}_B i vettori di covariate di due gruppi A e B (cfr. Tab.3.1), la stima dell'odds ratio è data da:

$$\exp(\tilde{\Delta}) = \frac{\sum_{k=1}^{K-1} \gamma_k(\mathbf{x}_A)(1 - \gamma_k(\mathbf{x}_B))(\pi_k + \pi_{k+1})}{\sum_{k=1}^{K-1} \gamma_k(\mathbf{x}_B)(1 - \gamma_k(\mathbf{x}_A))(\pi_k + \pi_{k+1})} \quad (5.11)$$

Sotto l'ipotesi nulla di uguaglianza tra le distribuzioni dei due gruppi, la varianza asintotica di $\tilde{\Delta}$ è uguale a quella di $\hat{\Delta}$ (Eq. 5.10).

Anche in questo caso, un test chi-quadro con un grado di libertà permette di testare l'ipotesi nulla di uguaglianza distributiva.

Le stime calcolate sono presentate in tabella 5.5 ad un livello di significatività del 10%.

Tab.5.5. Odds Ratio (\tilde{OR}), relativo Intervallo di Confidenza (\tilde{IC}) e statistica chi-quadro ($\tilde{\chi}^2$), ottenuti come output della rete neurale con nove variabili nello strato latente e $\lambda = 10^{-3}$. La percezione del disagio è misurata su una scala a quattro modalità

dataset	\tilde{OR}	\tilde{IC}	$\tilde{\chi}^2$	dataset	\tilde{OR}	\tilde{IC}	$\tilde{\chi}^2$
S1	0, 805	0, 721 – 0, 899	10, 35	S29	0, 734	0, 659 – 0, 817	22, 20
S2	0, 786	0, 704 – 0, 878	12, 72	S30	0, 794	0, 714 – 0, 883	12, 56
S3	0, 788	0, 706 – 0, 811	12, 38	S31	0, 712	0, 639 – 0, 793	26, 46
S4	0, 764	0, 686 – 0, 851	16, 89	S32	0, 850	0, 765 – 0, 944	6, 39
S5	0, 627	0, 562 – 0, 700	48, 52	S33	0, 797	0, 716 – 0, 886	12, 25
S6	0, 801	0, 720 – 0, 891	11, 66	S34	0, 666	0, 598 – 0, 743	37, 36
S7	0, 583	0, 522 – 0, 651	64, 45	S35	0, 718	0, 645 – 0, 800	25, 19
S8	0, 960	0, 865 – 1, 065	0, 42	S36	0, 662	0, 593 – 0, 738	38, 25
S9	0, 874	0, 787 – 0, 971	4, 37	S37	0, 794	0, 714 – 0, 884	12, 52
S10	0, 788	0, 708 – 0, 877	13, 43	S38	0, 723	0, 649 – 0, 806	24, 21
S11	0, 860	0, 774 – 0, 956	5, 50	S39	0, 686	0, 615 – 0, 764	32, 50
S12	0, 766	0, 688 – 0, 853	16, 65	S40	0, 618	0, 554 – 0, 690	51, 18
S13	0, 844	0, 760 – 0, 939	6, 88	S41	0, 753	0, 676 – 0, 839	18, 71
S14	0, 810	0, 728 – 0, 901	10, 49	S42	0, 775	0, 697 – 0, 863	15, 19
S15	0, 669	0, 598 – 0, 749	34, 10	S43	0, 913	0, 822 – 1, 014	2, 04
S16	0, 739	0, 661 – 0, 826	19, 74	S44	0, 798	0, 717 – 0, 888	12, 02
S17	0, 805	0, 721 – 0, 899	10, 35	S45	0, 945	0, 852 – 1, 049	0, 77
S18	0, 718	0, 642 – 0, 803	23, 66	S46	0, 808	0, 727 – 0, 900	10, 77
S19	0, 699	0, 625 – 0, 782	27, 35	S47	0, 872	0, 785 – 0, 969	4, 50
S20	0, 750	0, 671 – 0, 839	17, 95	S48	0, 798	0, 718 – 0, 888	12, 06
S21	0, 857	0, 768 – 0, 956	5, 35	S49	0, 896	0, 807 – 0, 996	2, 92
S22	0, 777	0, 696 – 0, 868	13, 95	S50	0, 805	0, 723 – 0, 895	11, 22
S23	0, 862	0, 772 – 0, 961	4, 96	S51	0, 824	0, 741 – 0, 916	9, 02
S24	0, 898	0, 805 – 1, 002	2, 61	S52	0, 830	0, 747 – 0, 923	8, 30
S25	0, 839	0, 752 – 0, 937	6, 76	S53	0, 771	0, 693 – 0, 858	15, 92
S26	0, 808	0, 724 – 0, 902	10, 06	S54	0, 781	0, 702 – 0, 869	14, 43
S27	0, 684	0, 614 – 0, 762	32, 80	S55	0, 784	0, 704 – 0, 872	13, 97
S28	0, 847	0, 762 – 0, 941	6, 65	S56	0, 798	0, 717 – 0, 888	12, 03

Rispetto all'analisi precedente (Tab. 5.1 e 5.4), la differenza tra la distribuzione del disagio osservata e quella di ciascun dataset simulato è meno significativa. Tendenzialmente gli odds ratio sono più vicini a uno, valore che rappresenta l'indipendenza tra le due distribuzioni.

Per alcuni confronti, addirittura, l'odds ratio non evidenzia la diversità distributiva. Questo risultato è, probabilmente, dovuto al fatto che, essendo il metodo

proposto da Clayton più “attento” alle categorie caratterizzate da basse frequenze (disagio estremo), molte più famiglie vengono ora classificate in questa categoria e la distribuzione risultante si avvicina a quella osservata.

Non vi è differenza, rispetto alla percezione di disagio, tra le famiglie costituite da un genitore con figli a carico e quelle di tipo coppia con figli ($\tilde{OR} = 0,960$; $\tilde{IC} = (0,865 - 1,065)$), tra le famiglie con almeno due malati cronico-degenerativo in casa e componenti di età media compresa fra i 30 e i 50 anni e quelle con al più un malato e componenti di età media inferiore ai 30 anni ($\tilde{OR} = 0,898$; $\tilde{IC} = (0,805 - 1,002)$), tra le famiglie unipersonali con problemi di tipo economico e quelle costituite da una coppia senza figli e problemi di tipo socio-relazionale ($\tilde{OR} = 0,913$; $\tilde{IC} = (0,822 - 1,014)$) e infine tra le famiglie unipersonali con problemi di tipo economico e quelle costituite da due componenti e problemi di tipo socio-relazionale ($\tilde{OR} = 0,945$; $\tilde{IC} = (0,852 - 1,049)$).

Per quanto riguarda gli ultimi due confronti, si tenga presente che nelle famiglie con due componenti entrano a far parte, oltre alle coppie senza figli, anche le famiglie composte da un genitore con figli a carico e altre tipologie meno diffuse (ad esempio coppia di fratelli).

La statistica chi-quadro rispecchia i risultati ottenuti con il calcolo degli odds ratio.

Utilizzando la metodologia proposta da Clayton, pertanto, sembrano non aver influenza nella distinzione tra le famiglie riguardo alla percezione di disagio le seguenti combinazioni di variabili: *tipo di famiglia*, *presenza di malati cronico-degenerativi* e *età media dei componenti*, *tipo di famiglia* e *tipo di problema percepito dalla stessa*.

Questo risultato desta un po' di sospetto se confrontato con le analisi effettuate in questo e nei precedenti capitoli. Il capitolo successivo, che sfrutta delle metodologie adatte a dati correlati, funge anche da supporto per verificare quali siano le caratteristiche familiari, nonché individuali, che determinano un maggior rischio di disagio per la famiglia.

5.3.1 Un'ulteriore analisi

Armitage (1971) ha proposto il calcolo di un test chi-quadro che riuscisse a cogliere una ben determinata deviazione dall'ipotesi nulla.

Tra i numerosi test chi-quadro esistenti in letteratura, molti sono fondati su statistiche distribuite come chi-quadro a più gradi di libertà: in questi casi il test è sensibile ad ogni deviazione dall'ipotesi nulla. Ad esempio, in una tabella di contingenza $K * 2$ l'ipotesi nulla fissa l'uguaglianza delle proporzioni attese di individui che, in ogni riga, cadono nella prima colonna; se vi sono K proporzioni, l'ipotesi nulla può essere disattesa se una sola di esse differisce dalle altre.

Questi test permettono la risoluzione di diversi tipi di problemi, ma l'interesse può essere anche rivolto a test sensibili a più specifiche condizioni, anche se meno efficienti del chi-quadro globale nel determinare altri tipi di deviazione.

Talvolta è possibile ottenere test orientati a ipotesi particolari suddividendo la statistica globale chi-quadro in componenti che seguono la distribuzione del chi-quadro con un numero inferiore di gradi di libertà.

Si consideri la tabella di contingenza $K * 2$ (3.1).

Il normale test χ_{K-1}^2 è adatto per mettere in evidenza differenze tra le K proporzioni di osservazioni che cadono nella prima colonna della tabella. Più specificatamente, si potrebbe verificare l'esistenza di un andamento significativo delle proporzioni passando dalla modalità 1 alla modalità K .

Per fare questo Armitage associa una variabile quantitativa (Z) alle K modalità. Si indichi, inoltre, con p_{iA} la proporzione di unità che appartengono al gruppo A. Misurando il disagio su scala a quattro modalità, la tabella 3.1 può essere riscritta nel modo seguente (Tab.5.6):

Tab.5.6. *Distribuzione delle frequenze assolute (n) di due campioni da due popolazioni rispetto alla percezione di disagio*

Livello di disagio	A	B	Totale	Proporzione del gruppo A	Z
Nulla	n_{1A}	n_{1B}	n_1	p_{1A}	z_1
Lieve	n_{2A}	n_{2B}	n_2	p_{2A}	z_2
Grave	n_{3A}	n_{3B}	n_3	p_{3A}	z_3
Estremo	n_{4A}	n_{4B}	n_4	p_{4A}	z_4
Totale	$n_{.A}$	$n_{.B}$	$n_{..}$	$p_{.A}$	-

La statistica test proposta da Armitage (1971) è data da:

$$\chi_2^2 = \chi^2 - \chi_1^2 \quad (5.12)$$

dove χ^2 è la statistica chi-quadro globale:

$$\frac{\sum n_i \cdot (p_{iA} - p_{\cdot A})^2}{p_{\cdot A}(1 - p_{\cdot A})} \quad (5.13)$$

e:

$$\chi_1^2 = \frac{n_{\cdot\cdot} \cdot (n_{\cdot\cdot} \sum n_{iA} z_{i\cdot} - n_{\cdot A} \sum n_{i\cdot} z_{i\cdot})^2}{n_{\cdot A} (n_{\cdot\cdot} - n_{\cdot A}) (n_{\cdot\cdot} \sum n_{i\cdot} z_{i\cdot}^2 - (\sum n_{i\cdot} z_{i\cdot})^2)} \quad (5.14)$$

Tale statistica ha distribuzione $\chi_{(K-2)}^2$ e verifica l'ipotesi dello scarto dalla linearità della regressione di p_{iA} rispetto a $z_{i\cdot}$.

Il valore di questa statistica è stato calcolato considerando come gruppi A e B la distribuzione del disagio osservata (gruppo A) e quella di ciascun dataset simulato (gruppo B). I risultati sono riportati in tabella 5.7.

Tab.5.7. Valori della statistica χ_2^2 per testare l'ipotesi dello scarto dalla linearità della regressione della proporzione di famiglie disagiate rispetto alla variabile quantitativa Z. Le due distribuzioni rispetto alle quali è stata calcolata la statistica sono quella del disagio nel dataset originale e quella in ciascuno dei 56 dataset simulati.

S1	18,00	S15	6,29	S29	11,17	S43	18,09
S2	29,39	S16	24,31	S30	9,68	S44	29,19
S3	15,20	S17	10,81	S31	3,98	S45	15,48
S4	21,25	S18	30,31	S32	25,50	S46	19,29
S5	12,30	S19	22,81	S33	11,07	S47	27,70
S6	16,79	S20	47,74	S34	13,80	S48	18,45
S7	60,12	S21	18,46	S35	8,57	S49	39,76
S8	17,80	S22	39,68	S36	28,28	S50	10,89
S9	20,10	S23	16,14	S37	14,01	S51	37,19
S10	26,11	S24	20,24	S38	15,97	S52	15,69
S11	16,46	S25	12,30	S39	4,72	S53	55,04
S12	17,44	S26	41,89	S40	8,83	S54	11,81
S13	14,66	S27	1,88	S41	5,46	S55	7,79
S14	7,07	S28	11,79	S42	15,98	S56	7,89

Quasi tutti i test per la deviazione dalla tendenza lineare risultano significativi, mostrando, pertanto, una tendenza che comporta incrementi non costanti della

proporzione di famiglie disagiate (gruppo A) al passaggio da un livello di disagio al successivo.

Questo risultato è coerente con le aspettative: la distanza tra le modalità che vanno a costituire la percezione del disagio non è costante. Infatti, mentre è sottile la differenza tra le famiglie che non avvertono disagio e quelle che si ritengono colpite da problemi risolvibili all'interno della famiglia, sia pure quella allargata, molto più rilevante è la differenza tra quest'ultime e quelle che necessitano dell'intervento dei servizi e, ancor più, tra queste e quelle colpite da gravi problemi non risolvibili.

Solo se si confrontano le famiglie unipersonali con problemi di tipo sanitario (gruppo A) e quelle di tipo coppia senza figli con problemi di tipo socio-relazionale, gli incrementi della proporzione di famiglie disagiate (gruppo A) al passaggio da un livello di disagio al successivo risultano costanti.

Capitolo 6

Modelli per dati correlati

6.1 Introduzione

In vari ambiti disciplinari (sociologia, economia, epidemiologia, demografia) si analizzano spesso fenomeni con struttura gerarchica, in cui i dati si presentano a più livelli: individuale, familiare, territoriale, sociale. In particolare l'analisi delle relazioni esistenti tra l'individuo e la società cui appartiene, relazioni che influenzano i suoi atteggiamenti e comportamenti, può essere ricondotta all'analisi di fenomeni a struttura gerarchica.

L'organizzazione dei dati in strutture gerarchiche comporta la necessità di tener conto della dipendenza tra le osservazioni, dipendenza tipica della struttura a gruppi.

In una struttura gerarchica, il livello micro fa riferimento all'individuo, alle sue caratteristiche e alle sue azioni, mentre il livello macro caratterizza il contesto di cui l'individuo fa parte. Nel nostro caso, la dimensione macro individua la famiglia con le sue tipicità strutturali, sociali, ambientali, relazionali, economiche e culturali, condivise dai membri di cui è costituita.

6.2 I modelli marginali per dati categoriali

Una variabile categoriale può, per definizione, assumere un determinato numero di modalità. La scala di misura per tale variabile può essere nominale, se i valori non sono ordinabili, ordinale, se le modalità sono in ordine, intervallo, se i livelli sono valori successivi con una determinata ampiezza.

L'analisi dei dati categoriali viene spesso affrontata tramite l'utilizzo dei modelli log-lineari, adatti allo studio della distribuzione congiunta di due o più variabili. In alcuni casi, come nel nostro, è opportuno tener conto della correlazione tra le osservazioni, dovuta al raggruppamento delle stesse in cluster o al fatto che derivano da misure ripetute in più occasioni sullo stesso singolo individuo.

Le tecniche statistiche che assumono l'indipendenza delle osservazioni, come l'analisi di regressione lineare e quella logistica, non possono essere utilizzate in questi studi, essendo necessario "correggere" le correlazioni tra le unità dello stesso cluster.

In questi casi ciò che può interessare è la distribuzione marginale delle variabili stesse. Al riguardo, sono stati introdotti i *modelli marginali*.

Le situazioni che si prestano all'applicazione dei modelli marginali come strumento di analisi sono molteplici; prima fra tutte si trova l'analisi di dati provenienti da indagini panel. Si pensi, ad esempio, all'analisi delle variazioni nella dimensione nello stock di occupati o disoccupati. Un'informazione di questo tipo può essere ottenuta esaminando le distribuzioni marginali di una tabella di contingenza ricavata classificando le osservazioni, ottenute in istanti temporali successivi, per condizione lavorativa.

Anche in ambito medico l'uso dei modelli marginali è particolarmente indicato. Un esempio è dato dall'osservazione dei pazienti sottoposti ad una determinata terapia in relazione al loro quadro clinico prima e dopo il trattamento. Le marginali ci permettono, in questo caso, di valutare il grado di efficacia della terapia.

Il problema della modellazione marginale è stato affrontato in passato da diversi autori. Un modo di procedere, proposto tra gli altri da Liang (1986), è quello di modellare direttamente i valori marginali, mettendoli in relazione con una serie di covariate opportune.

6.2.1 I modelli Generalized Estimating Equations (GEE)

Sia y_{ij} la variabile di interesse e \mathbf{x}_{ij} un vettore di p variabili esplicative, rilevate per l'individuo $i = 1, \dots, n_j$ appartenente al cluster $j = 1, \dots, J$.

La media e la varianza di y_{ij} sono rappresentate da $E(y_{ij}) = \mu_{ij}$ e $Var(y_{ij}) = v_{ij}$.

L'insieme delle risposte relative agli individui facenti parte dello stesso cluster j è rappresentato da un vettore di lunghezza n_j , $\mathbf{y}_j = (y_{1j}, \dots, y_{n_jj})$, con media $E(\mathbf{y}_j) = \boldsymbol{\mu}_j$ e matrice di varianze e covarianze ($n_j \times n_j$) $Var(\mathbf{y}_j) = V_j$. Sia R_j la matrice ($n_j \times n_j$) di correlazione di \mathbf{y}_j .

Inoltre, $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_J)$ è il vettore N -dimensionale delle risposte relative a ciascun cluster, con $N = \sum_{j=1}^J n_j$.

Le stesse considerazioni possono essere fatte se si considerano studi longitudinali dove l'indice i caratterizza ciascun individuo e l'indice j ciascuna osservazione rilevata su ogni soggetto.

Si consideri il modello di regressione:

$$y_{ij} = \beta_1 x_{ij1} + \beta_2 x_{ij2} + \dots + \beta_p x_{ijp} + \epsilon_{ij} = \mathbf{x}_{ij} \boldsymbol{\beta} + \epsilon_{ij}$$

dove $\boldsymbol{\beta}$ è un vettore p -dimensionale di coefficienti di regressione ignoti e ϵ_{ij} è il termine di errore che rappresenta la parte di variabilità di y_{ij} non riducibile alla dipendenza da x_{ijp} .

Diggle et al. (1994) fanno notare che l'unità dell'esperimento non è la misura individuale y_{ij} , ma la sequenza \mathbf{y}_j delle misure effettuate su ciascun cluster.

L'analisi di dati correlati risultanti da misure ripetute che si suppongono avere distribuzione normale multivariata è stata ampiamente studiata. Tuttavia, l'assunzione di normalità non è sempre verificata. Una delle procedure di analisi più utilizzata per tener conto di questa correlazione, nel caso di dati non distribuiti normalmente, è quella dei *Generalized Estimating Equations* (GEE), sviluppata da Liang e Zeger (1986). Si tratta di un'estensione dei modelli lineari generalizzati (GLM) al caso di dati correlati, basata sulle equazioni di quasi-verosimiglianza introdotte da Wedderburn (1974).

I modelli stimati tramite GEE sono modelli *marginali*, cioè esprimono la media della distribuzione marginale della variabile di interesse come funzione delle variabili esplicative attraverso un GLM e separatamente specificano la correlazione tra osservazioni mediante una matrice quadrata $R(\alpha)$.

L'interesse è rivolto non tanto alla matrice di varianze-covarianze delle misure ripetute, trattata come elemento di disturbo, quanto ai parametri di regressione, stimati utilizzando tutti i dati disponibili, anche quelli provenienti da unità con osservazioni mancanti.

I GEE forniscono stime dei parametri consistenti e indipendenti dalla specificazione di R : qualora questa sia appropriata, tali stime risulteranno essere anche efficienti.

Come i GLM, i GEE richiedono la specificazione di un previsore lineare:

$$\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} \quad (6.1)$$

e di una funzione legame:

$$g(\mu_{ij}) = \eta_{ij} \quad (6.2)$$

Le scelte comuni per la funzione legame possono essere, tra le altre, le funzioni identità, logit e logaritmo per dati, rispettivamente, continui, binari e categoriali (come nel nostro caso).

La varianza è definita come funzione della media:

$$Var(y_{ij}) = \phi V(\mu_{ij}) \quad (6.3)$$

dove $V(\mu_{ij})$ è una funzione nota della varianza e ϕ è un parametro di scala che può essere noto o stimato da:

$$\hat{\phi} = \frac{1}{N-p} \sum_{j=1}^J \sum_{i=1}^{n_j} e_{ij}^2 \quad (6.4)$$

dove $N = \sum_{j=1}^J n_j$ è il numero totale di osservazioni e e_{ij} sono i residui di Pearson:

$$e_{ij} = \frac{y_{ij} - \mu_{ij}}{\sqrt{\frac{V(\mu_{ij})}{w_{ij}}}} \quad (6.5)$$

con pesi w_{ij} .

È possibile calcolare intervalli di confidenza al $100(1 - \alpha)\%$ approssimati per il previsore lineare η_{ij} nel modo seguente:

$$g^{-1}(\eta_{ij} \pm z_{1-\alpha/2} \sqrt{Var(y_{ij})}) \quad (6.6)$$

dove z_p è il percentile 100p-esimo della distribuzione normale standard.

Nella tabella 6.1 sono presentate le distribuzioni di probabilità della variabile risposta y , il parametro di scala, la varianza e la devianza di y in relazione a varie distribuzioni ipotizzate:

Tab.6.1. *Distribuzioni di probabilità $f(y)$, parametro di dispersione ϕ , parametro di scala, varianza e devianza di y*

Nome	$f(y)$	Parametro di dispersione	Parametro di scala	$Var(y)$	Devianza
Normale	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], -\infty < y < +\infty$	σ^2	σ	σ^2	$\sum_j w_j (y_j - \mu_j)^2$
Gaussiana inversa	$\frac{1}{\sigma\sqrt{2\pi y^3}} \exp\left[-\frac{1}{2y}\left(\frac{y-\mu}{\mu\sigma}\right)^2\right], 0 < y < \infty$	σ^2	σ	$\sigma^2 \mu^3$	$\sum_j \frac{w_j (y_j - \mu_j)^2}{\mu_j^2 y_j}$
Gamma	$\frac{1}{\Gamma(\nu)y} \left(\frac{y\mu}{\mu}\right)^2 \exp\left(-\frac{y\nu}{\mu}\right), 0 < y < \infty$	ν^{-1}	ν	$\frac{\mu^2}{\nu}$	$2 \sum_j w_j \left[-\log\left(\frac{y_j}{\mu_j}\right) + \frac{y_j - \mu_j}{\mu_j}\right]$
Binomiale negativa	$\frac{\Gamma\left(\frac{y+1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \times \frac{(k\mu)^y}{(1+k\mu)^{\frac{y+1}{k}}}, y = 0, 1, 2, \dots$	k		$\mu + k\mu^2$	$2 \sum_j \left[y \log\left(\frac{y}{\mu}\right) - \left(y + \frac{w_j}{k}\right) \log\left(\frac{y + \frac{w_j}{k}}{\mu + \frac{w_j}{k}}\right)\right]$
Poisson	$\frac{\mu^y e^{-\mu}}{y!}, y = 0, 1, 2, \dots$	1		μ	$2 \sum_j w_j \left[y_j \log\left(\frac{y_j}{\mu_j}\right) - (y_j - \mu_j)\right]$
Binomiale	$\binom{n}{r} \mu^r (1-\mu)^{n-r}, y = \frac{r}{n} \text{ e } r = 0, 1, 2, \dots, n$	1		$\frac{\mu(1-\mu)}{n}$	$2 \sum_j w_j m_j \left[y_j \log\left(\frac{y_j}{\mu_j}\right) + (1-y_j) \log\left(\frac{1-y_j}{1-\mu_j}\right)\right]$
Multinomiale	$\frac{m!}{y_1! y_2! \dots y_k!} p_1^{y_1} p_2^{y_2} \dots p_k^{y_k}$				$\sum_j \sum_i w_j y_{ij} \log\left(\frac{y_{ij}}{p_{ij} m_j}\right)$

La distribuzione binomiale negativa contiene un parametro k , che non coincide con ϕ , ma è un addizionale parametro di distribuzione che deve essere stimato o a cui si deve attribuire un valore fisso.

Nel caso in cui sia presente una variabile w contenente dei pesi, il parametro di dispersione ϕ deve essere sostituito con ϕ/w .

Rispetto ai GLM, la specificazione addizionale nei GEE è per la struttura della matrice di correlazione $R(\boldsymbol{\alpha})$, che può essere definita arbitrariamente. Assumendo costante e pari a n la numerosità di cluster, tale matrice ha dimensione $(n \times n)$. Tuttavia, non disponendo necessariamente di gruppi di uguale dimensione, si suppone di disporre di matrici R_j di dimensione $(n_j \times n_j)$, con $n_j \leq n$.

Si assume che R , e perciò R_j , dipenda da un vettore di parametri $\boldsymbol{\alpha}$, che è lo stesso per tutte le unità. $\boldsymbol{\alpha}$ rappresenta la dipendenza media tra le osservazioni di ogni unità appartenente al medesimo cluster.

Esistono diverse strutture della matrice di correlazione. Nella tabella 6.2 si illustrano i casi di matrici indipendente, cambiabile, non strutturata, autoregressiva, M-dipendente, fissata.

Tab.6.2. *Tipi di struttura della matrice di correlazione*

Struttura	Definizione $R_{u,v}$	Esempio	#Parametri
Indipendente	$\begin{cases} 1 & \text{se } u = v \\ 0 & \text{altrimenti} \end{cases}$	$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$	0
Cambiabile	$\begin{cases} 1 & \text{se } u = v \\ \rho & \text{altrimenti} \end{cases}$	$\begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots & 1 \end{pmatrix}$	1
Non strutturata	$\begin{cases} 1 & \text{se } u = v \\ \rho_{u,v} & \text{altrimenti} \end{cases}$	$\begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1t} \\ \rho_{21} & 1 & \dots & \rho_{2t} \\ \dots & \dots & \dots & \dots \\ \rho_{t1} & \rho_{t2} & \dots & 1 \end{pmatrix}$	$\frac{t(t-1)}{2}$
Autoregressiva	$\begin{cases} 1 & \text{se } u = v \\ \rho^{ u-v } & \text{altrimenti} \end{cases}$	$\begin{pmatrix} 1 & \rho & \dots & \rho^{ 1-t } \\ \rho & 1 & \dots & \rho^{ 2-t } \\ \dots & \dots & \dots & \dots \\ \rho^{ t-1 } & \rho^{ t-2 } & \dots & 1 \end{pmatrix}$	1
M-dipendente	$\begin{cases} 1 & \text{se } u = v \\ \rho_{ u-v } & \text{altrimenti} \end{cases}$	$\begin{pmatrix} 1 & \rho_1 & \dots & \rho_{ 1-t } \\ \rho_1 & 1 & \dots & \rho_{ 2-t } \\ \dots & \dots & \dots & \dots \\ \rho_{ t-1 } & \rho_{ t-2 } & \dots & 1 \end{pmatrix}$	$0 < M \leq t - 1$
Fissata	$\begin{cases} 1 & \text{se } u = v \\ r_{u,v} & \text{altrimenti} \end{cases}$	$\begin{pmatrix} 1 & r_{12} & \dots & r_{1t} \\ r_{21} & 1 & \dots & r_{2t} \\ \dots & \dots & \dots & \dots \\ r_{t1} & r_{t2} & \dots & 1 \end{pmatrix}$	0

Una guida alla scelta della matrice R si trova in Diggle et al. (1994); in ogni caso, un'analisi delle stime dei parametri e dei loro errori standard è d'aiuto alla scelta.

Indicate con A_j e con W_j le matrici diagonale di dimensione $(n_j \times n_j)$, con elementi diagonali i -esimi rispettivamente la funzione della varianza $V(\mu_{ij})$ e il peso w_{ij} , la matrice di varianze-covarianze di lavoro per \mathbf{y}_j è pari a:

$$V_j(\boldsymbol{\alpha}) = \phi A_j^{\frac{1}{2}} W_j^{-\frac{1}{2}} R_j(\boldsymbol{\alpha}) W_j^{-\frac{1}{2}} A_j^{\frac{1}{2}} \quad (6.7)$$

Se $R_j(\boldsymbol{\alpha})$ fosse la vera matrice di correlazione di \mathbf{y}_j , $V_j(\boldsymbol{\alpha})$ sarebbe la vera matrice di covarianza di \mathbf{y}_j .

Lo stimatore GEE di $\boldsymbol{\beta}$ è soluzione di:

$$\sum_j^J D_j' V(\hat{\boldsymbol{\alpha}}_j)^{-1} (\mathbf{y}_j - \boldsymbol{\mu}_j) = \mathbf{0} \quad (6.8)$$

dove $\hat{\boldsymbol{\alpha}}$ è una stima consistente di $\boldsymbol{\alpha}$ e $D_j = \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\beta}}$.

La soluzione GEE può, in questo modo, essere vista come una generalizzazione della soluzione GLM per dati correlati.

Ad esempio, nel caso normale:

$$\boldsymbol{\mu}_j = X_j \boldsymbol{\beta} \quad (6.9)$$

$$D_j = X_j \quad (6.10)$$

$$V(\hat{\boldsymbol{\alpha}}) = R_j(\hat{\boldsymbol{\alpha}}) \quad (6.11)$$

Da cui:

$$\sum_j X_j' [R_j(\hat{\boldsymbol{\alpha}})]^{-1} (\mathbf{y}_j - X_j \boldsymbol{\beta}) = \mathbf{0} \quad (6.12)$$

e, risolvendo rispetto a $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \left[\sum_j X_j' [R_j(\hat{\boldsymbol{\alpha}})] X_j \right]^{-1} \left[\sum_j X_j' [R_j(\hat{\boldsymbol{\alpha}})]^{-1} \mathbf{y}_j \right] \quad (6.13)$$

Tale stimatore è lo stesso che si ottiene con i *minimi quadrati ponderati* con matrice dei dati $[R_j(\hat{\boldsymbol{\alpha}})]^{-1}$, ma, mentre nei minimi quadrati ponderati si conosce la matrice dei pesi, nei GEE essa dipende dal vettore di parametri $\boldsymbol{\alpha}$ da stimare. La soluzione si ottiene tramite la procedura dei *minimi quadrati ponderati iterati* (McCullagh e Nelder, 1989), dove le stime iterative di $\boldsymbol{\alpha}$ sono utilizzate per ottenere nuove stime di $\boldsymbol{\beta}$ con una procedura che continua fino alla convergenza.

6.2.2 Modello utilizzato nello studio

Il modello utilizzato vuole individuare le caratteristiche individuali e familiari che determinano un rischio di disagio familiare ed è stato realizzato come modello lineare generalizzato stimato tramite procedura GEE.

Il software utilizzato è il SAS System e la procedura la GENMOD, che fa uso delle stime di massima verosimiglianza per adattare un modello lineare generalizzato e implementa un'analisi della varianza attraverso il rapporto di verosimiglianza e il test di Wald sugli effetti fissi (Hosmer and Lemeshow, 2000).

In un'analisi logit ordinaria le stime di massima verosimiglianza possono essere ottenute attraverso l'algoritmo dei minimi quadrati iterativamente ripesati: ad ogni step dell'algoritmo si ottengono le stime ai minimi quadrati pesati, modificando, ad ogni iterazione, sia i pesi che il valore della variabile dipendente in base ai risultati ottenuti alla precedente iterazione. La matrice dei pesi W ha gli elementi sulla diagonale pari a $p_i(1 - p_i)$, con p_i probabilità calcolata nella precedente iterazione ed elementi non diagonali pari a zero.

L'algoritmo GEE estende questo approccio ai minimi quadrati iterati *generalizzati*. La matrice W ha elementi diagonali non uguali a zero, ma funzione della correlazione tra le osservazioni, stimata ad ogni iterazione. Si assume, infatti, che le osservazioni siano indipendenti tra cluster e correlate entro cluster.

Nella procedura GENMOD è possibile specificare la struttura della matrice di correlazione tra le osservazioni R entro ciascun cluster e richiedere in output la sua stima.

Le variabili utilizzate per la stima dei parametri sono riportate nella tabella 6.3. Le prime sei sono variabili misurate a livello individuale: le risposte sono rilevate per tutti i 6919 componenti appartenenti alle famiglie campionate. Le altre dieci variabili, invece, sono misurate a livello familiare e interessano le 2465 famiglie intervistate.

Accanto ai parametri è riportata la relativa modalità. Sono riportate tutte le modalità tranne una, quella relativa al parametro preso come riferimento per il calcolo delle stime.

Tab.6.3. Variabili utilizzate nel modello

Variabile	Modalità
Età	1 = < 30 2 = ≥ 30 e < 50 3 = ≥ 50 e < 70 4 = ≥ 70
Stato civile	1 = coniugato 2 = celibe/nubile 3 = separato/divorziato 4 = vedovo/a
Titolo di studio	0 = basso (diploma di scuola media superiore - 2/3 anni- o inferiore) 1 = alto (diploma di scuola media superiore - 4/5 anni- o superiore)
Disoccupato	0 = no 1 = si
Autonomia dell'invalido	0 = no 1 = si 2 = non è invalido
Autonomia del malato cronico-degenerativo	0 = no 1 = si 2 = non è malato
Tipo di famiglia	1 = unipersonale 2 = coppia senza figli 3 = coppia con figli 4 = un genitore con figli a carico 5 = un genitore con almeno un figlio occupato 6 = altro
Titolo a cui la famiglia occupa l'abitazione	1 = proprietà od usufrutto 2 = affitto o subaffitto 3 = altro (gratuito o prestazione di servizio)
Numero di automobili in famiglia	0 = 0 1 = 1 2 = 2 o più
Pasti assieme	1 = i componenti pranzano insieme anche nei giorni feriali 2 = pranzano insieme solo nei fini settimana 3 = pranzano insieme sporadicamente 4 = non pranzano mai insieme
Possibilità per la famiglia di aiuto da parte di parenti	0 = no 1 = si
Necessità di aiuti economici per la famiglia	0 = no 1 = si
La famiglia ha subito un lutto negli ultimi 3 anni	0 = no 1 = si
Qualcuno in famiglia si è separato/ha divorziato negli ultimi 3 anni	0 = no 1 = si
La famiglia ha subito un furto negli ultimi 3 anni	0 = no 1 = si
Livello di reddito mensile familiare	1 = > 6000 € 2 = ≤ 6000 e > 3000 € 3 = ≤ 3000 e > 1500 € 4 = ≤ 1500 €

La matrice di correlazione scelta ha struttura indipendente, ossia forza la correlazione intra-cluster a zero. In questo modo, la stima GEE si riduce ad una stima logit ordinaria. In ogni caso, le stime empiriche dell'errore standard tengono conto della correlazione tra gli individui appartenenti alla stessa famiglia.

La distribuzione di probabilità scelta per la variabile di interesse, la percezione del disagio, è la multinomiale.

Nella tabella 6.4 si riportano le stime dei parametri e le stime empiriche dell'errore standard, gli intervalli di confidenza, le statistiche Z e i valori p . Le stime dei parametri rappresentano i logaritmi degli odds ratio calcolati come rapporto tra l'odds di ciascuna modalità di una determinata variabile e l'odds della modalità presa come riferimento per quella variabile. Le stime empiriche dell'errore standard sono stime robuste ottenute con il metodo di White (1980), che non dipendono dalla correttezza della struttura imposta alla matrice di correlazione R . Tali stime sono preferibili quando il numero di cluster è elevato; al contrario, quando i gruppi sono meno di 20, si calcolano generalmente le stime dell'errore standard basate sul modello, che sono consistenti quando la matrice di varianze e covarianze è correttamente specificata, ma, anche in assenza di questa condizione, hanno buone proprietà (Prentice, 1988). Ciò dipende dal fatto che lo stimatore della varianza è asintoticamente corretto, ma può essere distorto in presenza di uno scarso numero di cluster. Gli intervalli di confidenza sono ottenuti con il metodo di Wald: l'intervallo di confidenza al $100(1 - \alpha)\%$ per il parametro β è definito come:

$$\hat{\beta} \pm z_{1-\alpha/2} \hat{\sigma} \quad (6.14)$$

dove $\hat{\beta}$ è la stima del parametro e $\hat{\sigma}$ quella del suo errore standard.

La statistica Z è pari alla radice quadrata della statistica χ^2 di Pearson:

$$\chi^2 = \sum_j \frac{w_j (y_j - \mu_j)^2}{V(\mu_j)} \quad (6.15)$$

I risultati si sono ottenuti adattando come funzione legame la logit cumulativa. Gli intervalli di confidenza sono calcolati per valori di α pari a 0.01, 0.05 e 0.10.

Tab.6.4. *Stime (log odds ratio) di un modello GEE con intervalli di confidenza al 99%, 95% e 90%, struttura di corr. indipendente e legame logit cumulativo*

Parametro	Modalità	Stima	Err. Std.	99%–IC		95%–IC		90%–IC		Z	Pr > Z	
Età	< 30	-0,111	0,154	-0,509	0,287	-0,413	0,192	-0,365	0,143	-0,7	0,4725	
Età	≥ 50 e < 70	0,047	0,111	-0,239	0,332	-0,171	0,264	-0,136	0,229	0,4	0,6743	
Età	≥ 70	0,209	0,158	-0,199	0,617	-0,102	0,520	-0,052	0,470	1,3	0,1873	
Stato civile	celibe/nubile	0,101	0,156	-0,300	0,621	-0,204	0,407	0,155	-0,357	0,6	0,5169	
Stato civile	separato/divorziato	0,037	0,227	-0,547	0,503	-0,407	0,482	-0,336	0,410	0,2	0,8693	
Stato civile	vedovo/a	0,148	0,178	-0,309	0,309	-0,200	0,496	-0,144	0,440	0,8	0,4034	
Titolo di studio	alto	-0,137	0,080	-0,344	0,070	-0,295	0,020	-0,270	-0,005	-1,7	0,0874	*
Disoccupato	no	-0,793	0,261	-1,465	-0,122	-1,304	-0,282	-1,222	-0,364	-3,0	0,0023	***
Autonomia dell'invalido	no	2,401	0,235	1,795	3,007	1,940	2,862	2,014	2,788	10,2	<,0001	***
Autonomia dell'invalido	sì	1,213	0,269	0,520	1,906	0,686	1,740	0,770	1,656	4,5	<,0001	***
Autonomia del malato cronico-degenerativo	no	2,060	0,194	1,561	2,559	1,681	2,440	1,742	2,379	10,6	<,0001	***
Autonomia del malato cronico-degenerativo	sì	0,999	0,126	0,673	1,325	0,751	1,247	0,791	1,207	7,9	<,0001	***
Tipo di famiglia	unipersonale	0,468	0,248	-0,172	1,107	-0,019	0,954	0,059	0,876	-1,9	0,0595	*
Tipo di famiglia	coppia senza figli	0,014	0,142	-0,353	0,381	-0,265	0,293	-0,220	0,249	0,1	0,9197	
Tipo di famiglia	un gen.con f.a carico	0,524	0,228	-0,062	1,110	0,078	0,970	0,149	0,898	2,3	0,0214	**
Tipo di famiglia	un gen.con f.occupato/i	0,227	0,343	-0,657	1,110	-0,446	0,899	-0,338	0,791	0,7	0,5089	
Tipo di famiglia	altro	0,466	0,456	-0,710	1,642	-0,429	1,361	-0,285	1,217	1,0	0,3073	
Titolo a cui la famiglia occupa l'abitazione	affitto o subaffitto	0,269	0,170	-0,169	0,707	-0,064	0,603	-0,011	0,549	1,6	0,1139	
Titolo a cui la famiglia occupa l'abitazione	altro	1,070	0,321	0,244	1,896	0,442	1,698	0,543	1,597	3,3	0,0008	***
Numero di automobili in famiglia	0	0,534	0,236	-0,074	1,141	0,071	0,996	0,146	0,922	2,3	0,0236	**
Numero di automobili in famiglia	1	-0,083	0,141	-0,445	0,279	-0,358	0,193	-0,314	0,149	-0,6	0,5566	
Pasti assieme	solo fine settimana	0,512	0,233	-0,088	1,112	0,055	0,968	0,128	0,895	2,2	0,0281	**
Pasti assieme	sporadicamente	-0,090	0,259	-0,759	0,578	-0,599	0,418	-0,517	0,336	-0,3	0,7275	
Pasti assieme	mai	1,358	0,310	0,558	2,158	0,750	1,967	0,847	1,869	4,4	<,0001	***
Possibilità per la famiglia di aiuto da parenti	sì	-0,539	0,123	-0,857	-0,221	-0,781	-0,297	-0,742	-0,336	-4,4	<,0001	***
Necessità di aiuti economici per la famiglia	no	-1,410	0,208	-1,945	-0,874	-1,817	-1,002	-1,752	-1,068	-6,8	<,0001	***
Lutti in famiglia	sì	0,003	0,121	-0,307	0,314	-0,233	0,240	-0,195	0,202	0,1	0,9779	
Separazione/divorzio	sì	0,024	0,430	-1,084	1,132	-0,819	0,867	-0,683	0,732	0,1	0,9551	
Furti negli ultimi 3 anni	sì	-0,389	0,232	-0,988	0,210	-0,845	0,067	-0,771	-0,006	-1,7	0,0943	*
Livello di reddito mensile familiare	≤ 6000 e > 3000 €	0,897	0,572	-0,577	2,372	-0,225	2,019	-0,044	1,839	1,6	0,1171	
Livello di reddito mensile familiare	≤ 3000 e > 1500 €	1,093	0,560	-0,350	2,535	-0,005	2,190	0,171	2,014	1,9	0,0511	*
Livello di reddito mensile familiare	≤ 1500 €	1,542	0,572	0,068	3,016	0,420	2,664	2,483	0,601	2,7	0,0071	***

Significatività: *** = 1%, ** = 5%, * = 10%

Ad un livello di significatività del 99%, sono significative le variabili che rilevano la presenza di componenti disoccupati, l'autonomia di eventuali componenti invalidi o affetti da malattie cronico-degenerative, l'occupare l'abitazione a titolo gratuito, le abitudini alimentari con riferimento all'attitudine dei componenti a condividere o meno il momento del pasto, la possibilità di aiuto da parte di parenti, la necessità di recenti aiuti economici per la famiglia e il livello di reddito.

Si noti, tuttavia, che, per queste variabili, non tutti i confronti sono significativi: il titolo a cui la famiglia occupa l'abitazione risulta determinante, nel discriminare le famiglie con e senza disagio, se si confrontano i due gruppi composti, rispettivamente, da famiglie che hanno l'abitazione in proprietà o usufrutto e da quelle che la occupano secondo un altro titolo (prevalentemente gratuito) ($p\text{-value} = 0,0008$).

Il disagio avvertito in questa seconda tipologia di famiglia è spiegato dal fatto che le famiglie che occupano a titolo gratuito l'abitazione sono tendenzialmente più povere di quelle proprietarie della loro abitazione. Solo il 17,4% delle prime ha un reddito superiore a 3000 euro netti mensili, contro il 25,2% delle seconde (6.5).

Tab.6.5. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per titolo secondo cui occupa l'abitazione e livello di reddito*

Titolo	Reddito				Totale
	$\leq 1500 \text{ €}$	$> 1500 \text{ e}$ $\leq 3000 \text{ €}$	$> 3000 \text{ e}$ $\leq 6000 \text{ €}$	$> 6000 \text{ €}$	
Proprietà o usufrutto (n=6087)	17,9	56,9	22,7	2,5	100,0
Affitto o subaffitto (n=692)	25,8	54,4	17,3	2,5	100,0
Altro (n=140)	19,3	63,3	15,6	1,8	100,0
Totale (n=6919)	18,8	56,7	22,0	2,5	100,0

Ciò significa che, in termini di percezione di disagio, vi è differenza tra le famiglie che hanno l'abitazione a titolo gratuito e quelle che per cui la casa è di proprietà: il rischio di disagio è di $\exp(1,070) = 2,915$ volte maggiore nelle prime rispetto alle seconde. Invece, il disagio sembra colpire indifferentemente le famiglie che hanno la casa in proprietà rispetto a quelle che vivono in affitto ($p\text{-value} = 0,1139$).

Le variabili che misurano lo stato di salute individuale sembrano incidere nella percezione dello stato di disagio da parte della famiglia. Il logaritmo dell'odds ratio

pari a 2,401 indica che l'odds relativo alla mancanza di autonomia degli invalidi è $\exp(2,401) = 11,034$ volte l'odds relativo all'assenza di invalidità; pertanto, la probabilità di disagio è del 998,9% maggiore in presenza di invalidi non autonomi rispetto all'assenza di invalidi. In questo caso, la modalità presa come riferimento è l'assenza di invalidità; rispetto a tale modalità vengono calcolate le stime. Il valore p è inferiore a 0,0001.

Anche la presenza di invalidi autonomi fa aumentare la percezione di disagio da parte della famiglia, anche se, come è ovvio attendersi, il disagio è maggiore laddove gli invalidi non risultano autonomi. Il rischio di disagio aumenta del 236,4% nel caso in cui in famiglia vi siano invalidi autonomi rispetto al caso in cui tutti i componenti siano sani.

Un risultato analogo si ottiene studiando il disagio in relazione alla presenza di componenti affetti da malattie cronic-degenerative, sebbene, come emerge dai valori delle stime, tale problema sia meno importante rispetto a quello dell'invalidità nel determinare un rischio di disagio per la famiglia. La percezione del disagio aumenta di 7,874 volte in presenza di malati non autonomi e di 2,717 in presenza di malati autonomi, rispetto alla presenza di componenti non affetti da alcuna malattia cronic-degenerativa.

Oltre ai problemi di salute, sembrano minare la stabilità familiare anche quelli di carattere economico. Quelli analizzati riguardano, in particolare, l'eventuale necessità di aiuti monetari per la famiglia nell'ultimo anno e il livello di reddito familiare. Come si può notare, mentre le variabili di carattere sanitario si riferiscono ai singoli componenti, quelle economiche interessano la famiglia nel suo insieme. Questa scelta dipende dal fatto che i problemi economici che colpiscono un solo membro della famiglia, nella maggior parte dei casi, possono essere risolti grazie all'intervento degli altri componenti senza turbare eccessivamente l'equilibrio familiare; pertanto, i veri problemi economici si riscontrano laddove è la famiglia e non il singolo ad averne. D'altra parte, i problemi sanitari, anche se toccano l'intera famiglia, si misurano come deficit dei singoli individui.

Il rischio di disagio è di quasi cinque volte superiore nelle famiglie con reddito inferiore a 1500 euro mensili rispetto a quelle con reddito superiore a 6000 euro

(odds ratio = $1/exp(1,542) = 4,673$) e di oltre quattro volte nelle famiglie che hanno avuto recente bisogno di aiuti monetari rispetto a quelle economicamente autonome (odds ratio = $exp(-1,410) = 4,096$).

Ad un livello del 5%, oltre alle variabili risultate significative all'1%, diventa significativa la differenza tra le famiglie composte da un genitore con figli a carico e quelle di tipo coppia con figli. Il valore p è pari a 0,0214 e l'odds ratio ammonta a $exp(0,524) = 1,689$, mostrando come il rischio di disagio sia del 68,9% maggiore nelle famiglie composte da un genitore con figli rispetto a quelle di tipo coppia con figli.

Diventano significative anche le variabili relative, rispettivamente, alla frequenza con cui i componenti della famiglia pranzano assieme e al numero di automobili possedute dalla famiglia stessa. Con riferimento a quest'ultima variabile, l'odds ratio pari a $1/exp(0,534) = 1,706$ e il valore p a 0,0236 mostrano come il disagio sia del 70,6% maggiore qualora la famiglia non possieda alcuna automobile rispetto al caso in cui ne possieda due o più.

Le famiglie che non possiedono auto si caratterizzano per essere costituite perlopiù da individui anziani (il 65,3% ha un'età media superiore a 70 anni) che vivono soprattutto soli, ma anche in coppia (il 62,1% e il 24,1% sono rispettivamente famiglie di tipo unipersonale o coppia senza figli) (Tab.6.6 e 6.7).

Tab.6.6. *Distribuzione percentuale delle famiglie residenti nel Veneto nel 2004-2005 per numero di automobili e età media dei componenti*

Auto	Età				Totale
	< 30	≥ 30 e < 50	≥ 50 e < 70	≥ 70	
0 (n=260)	2,1	8,4	24,2	65,3	100,0
1 (n=896)	18,1	25,1	35,2	21,6	100,0
2 o più (n=1335)	24,2	54,9	19,3	1,6	100,0
Totale (n=2491)	19,8	39,4	25,6	15,2	100,0

Tab.6.7. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di automobili e tipo di famiglia*

Auto	Tipo di famiglia					Totale
	Unipersonale	Coppia senza figli	Coppia con figli	Un genitore con figli	Altro	
0 (n=260)	62,1	24,1	5,0	5,1	3,7	100,0
1 (n=896)	20,9	34,3	33,5	9,6	1,7	100,0
2 o più (n=1335)	0,0	17,1	77,0	4,9	1,0	100,0
Totale (n=2491)	13,8	24,1	54,0	6,6	1,5	100,0

Tali famiglie vivono in zone isolate di montagna nel 6,2% dei casi (di contro al 2,9% di quelle che possiedono almeno due automobili) e sono colpite soprattutto da problemi sanitari (32,5%) (Tab.6.8 e 6.9).

Tab.6.8. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per numero di automobili e zona di abitazione*

Auto	Età			Totale
	Montagna (oltre i 600 metri)	Collina (da 300 a 600 metri)	Pianura (sotto i 300 metri)	
0 (n=255)	6,2	85,6	8,2	100,0
1 (n=868)	3,8	87,5	8,7	100,0
2 o più (n=1308)	2,9	87,5	9,6	100,0
Totale (n=2431)	3,6	87,3	9,1	100,0

Tab.6.9. *Distribuzione percentuale delle famiglie (con 0 auto) residenti nel Veneto nel 2004-2005 per tipi di problemi presenti al loro interno*

Problemi	
Salute (n=88)	32,5
Economici (n=36)	14,4
Socio-relazionali (n=9)	3,2
Altro (n=2)	1,0

Si sono considerati i risultati specificando un livello di significatività del 10%. Ciò che emerge è la differenza tra le famiglie in cui è presente almeno un componente disoccupato e quelle in cui non ve ne sono, tra le famiglie unipersonali e quelle di tipo coppia con figli, tra quelle che hanno subito furti negli ultimi tre anni e quelle che non li hanno subiti e, infine, tra le famiglie con redditi rispettivamente compreso tra 1500 e 3000 e superiore a 6000 euro mensili.

Paradossalmente, si nota che le famiglie che hanno subito recenti furti sono meno disagiate rispetto alle altre (odds ratio = $\exp(-0,389) = 0,678$). Il rischio di disagio è superiore al 47,5% nelle seconde rispetto alle prime. Questo risultato, apparentemente strano, trova una spiegazione se si analizzano le famiglie rispetto all'aver subito o meno furti e al livello di reddito (Tab. 6.10).

Tab.6.10. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per furti negli ultimi tre anni e livello di reddito*

Furti	Reddito				Totale
	≤ 1500 €	> 1500 e ≤ 3000 €	> 3000 e ≤ 6000 €	> 6000 €	
No (n=6476)	18,8	57,1	21,9	2,2	100,0
Sì (n=443)	17,8	51,7	23,5	7,0	100,0
Totale (n=6919)	18,8	56,7	22,0	2,5	100,0

Chi subisce furti sono perlopiù famiglie economicamente ricche, che non risultano sopraffatte a lungo termine da un evento di questo tipo. Ben il 7,0% delle famiglie che hanno subito un furto ha un reddito superiore a 6000 euro mensili netti, contro il 2,2% delle altre.

L'analisi effettuata sopra è stata svolta escludendo dal modello alcune variabili risultate strategiche nel distinguere le famiglie rispetto al disagio percepito, per vedere se venivano messe in luce altre caratteristiche individuali o familiari correlate al disagio.

In tabella 6.11 si riportano i risultati ottenuti escludendo dal modello la variabile che rileva la presenza di invalidi in famiglia e l'eventuale autonomia.

Tab.6.11. *Stime (log odds ratio) di un modello GEE (non si considera la variabile che rileva la presenza e l'eventuale grado di autonomia degli invalidi) con intervalli di confidenza al 99%, 95% e 90%, struttura di corr. indipendente e legame logit cumulativo*

Parametro	Modalità	Stima	Err. Std.	99%–IC		95%–IC		90%–IC		Z	Pr > Z	
Età	< 30	-0,193	0,155	-0,593	0,208	-0,497	0,112	-0,448	0,063	-1,2	0,2150	
Età	≥ 50 e < 70	0,097	0,111	-0,189	-0,382	-0,120	0,314	-0,086	0,279	0,9	0,3831	
Età	≥ 70	0,405	0,157	0,001	0,809	0,097	0,713	0,147	0,663	2,6	0,0099	***
Stato civile	celibe/nubile	0,199	0,157	-0,204	0,603	-0,108	0,506	-0,058	0,457	1,3	0,2031	
Stato civile	separato/divorziato	0,045	0,222	-0,527	0,617	-0,390	0,480	-0,320	0,410	0,2	0,8390	
Stato civile	vedovo/a	0,025	0,176	-0,428	0,477	-0,320	0,369	-0,265	0,314	0,1	0,8889	
Titolo di studio	alto	-0,151	0,081	-0,359	0,057	-0,309	0,007	-0,284	-0,018	-1,8	0,0615	*
Disoccupato	no	-0,760	0,261	-1,433	-0,087	-1,272	-0,248	-1,190	-0,330	-2,9	0,0036	***
Autonomia del malato cronico-degenerativo	no	1,840	0,193	1,344	2,336	1,463	2,218	1,524	2,157	9,6	<,0001	***
Autonomia del malato cronico-degenerativo	sì	0,867	0,126	0,543	1,192	0,620	1,115	0,660	1,075	6,9	<,0001	***
Tipo di famiglia	unipersonale	0,647	0,253	-0,004	1,297	0,152	1,142	0,231	1,062	2,6	0,0104	**
Tipo di famiglia	coppia senza figli	0,018	0,145	-0,355	0,391	-0,266	0,302	-0,220	0,256	0,1	0,9014	
Tipo di famiglia	un gen.con f.a carico	0,487	0,230	-0,105	1,079	0,037	0,937	0,109	0,865	2,1	0,0340	**
Tipo di famiglia	un gen.con f.occupato/i	0,225	0,338	-0,646	1,096	-0,438	0,888	-0,331	0,781	0,7	0,5060	
Tipo di famiglia	altro	0,543	0,459	-0,640	1,726	-0,357	1,443	-0,213	1,298	1,2	0,2372	
Titolo a cui la famiglia occupa l'abitazione	affitto o subaffitto	0,257	0,170	-0,181	0,695	-0,076	0,590	-0,023	0,536	1,5	0,1316	
Titolo a cui la famiglia occupa l'abitazione	altro	1,148	0,335	0,286	2,010	0,492	1,804	0,597	1,699	3,4	0,0006	***
Numero di automobili in famiglia	0	0,537	0,236	-0,072	1,146	0,073	1,000	0,148	0,926	2,3	0,0232	**
Numero di automobili in famiglia	1	-0,105	0,142	-0,472	0,262	-0,384	0,174	-0,339	0,129	-0,7	0,4609	
Pasti assieme	solo fine settimana	0,514	0,235	-0,092	1,120	0,053	0,975	0,127	0,901	2,2	0,0288	**
Pasti assieme	sporadicamente	-0,097	0,268	-0,786	0,593	-0,621	0,428	-0,537	0,343	-0,4	0,7175	
Pasti assieme	mai	1,379	0,310	0,581	2,176	0,772	1,986	0,069	1,888	4,4	<,0001	***
Possibilità per la famiglia di aiuto da parenti	sì	-0,550	0,124	-0,869	-0,231	-0,793	-0,307	-0,754	-0,346	-4,4	<,0001	***
Necessità di aiuti economici per la famiglia	no	-1,432	0,211	-1,976	-0,888	-1,846	-1,018	-1,780	-1,085	-6,8	<,0001	***
Lutti in famiglia	sì	0,002	0,121	-0,311	0,315	-0,236	0,240	-0,198	0,202	0,1	0,9861	
Separazione/divorzio	sì	-0,052	0,455	-1,223	1,120	-0,943	0,839	-0,800	0,696	-0,1	0,9095	
Furti negli ultimi 3 anni	sì	-0,382	0,233	-0,982	0,219	-0,839	0,075	-0,765	0,002	-1,6	0,1016	
Livello di reddito mensile familiare	≤ 6000 e > 3000 €	0,923	0,572	-0,551	2,396	-0,199	2,044	-0,018	1,864	1,6	0,1068	
Livello di reddito mensile familiare	≤ 3000 e > 1500 €	1,129	0,560	-0,314	2,571	0,031	2,226	0,208	2,049	2,0	0,0438	**
Livello di reddito mensile familiare	≤ 1500 €	1,576	0,573	0,100	3,051	0,453	2,698	0,633	2,518	2,7	0,0059	***

Significatività: *** = 1%, ** = 5%, * = 10%

Le differenze rispetto ai risultati nella tabella 6.4 sono evidenziate in grassetto.

Dal confronto, emerge che, ad un livello di significatività dell'1%, la variabile che misura l'età dei componenti riesce a discriminare le famiglie rispetto al disagio. In particolare, il rischio di disagio è di $\exp(0,405) = 1,499$ volte maggiore se l'età è superiore a 70 anni rispetto ad un'età compresa tra i 30 e i 50 anni. La variabile che rileva la presenza e l'autonomia degli invalidi viene ad essere, pertanto, "sostituita" da quella che rileva l'età degli individui con un significato ben preciso come mostra la tabella 6.12.

Tab.6.12. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per invalidità ed età*

Invalidi	Età				Totale
	< 30	≥ 30 e < 50	≥ 50 e < 70	≥ 70	
Non autonomo (n=83)	13,0	10,4	24,9	51,7	100,0
Autonomo (n=65)	16,2	35,3	29,5	19,0	100,0
Non è invalido (n=6771)	30,5	31,4	27,2	10,9	100,0
Totale (n=6919)	30,1	31,2	27,2	11,5	100,0

Oltre la metà degli invalidi non autonomi ha un'età superiore a 70 anni (51,7%). Solo nel 23,4% dei casi l'età risulta inferiore a 50 anni. Ciò sta ad indicare che le caratteristiche individuali che consentono di distinguere le famiglie disagiate da quelle che non si ritengono colpite da problemi di una certa importanza sono la combinazione della mancanza di autonomia causata dall'invalidità e dell'età avanzata.

In tabella 6.13 si riportano i risultati ottenuti escludendo dal modello la variabile che rileva la presenza di malati cronico-degenerativi in famiglia e l'eventuale autonomia.

Tab.6.13. *Stime (log odds ratio) di un modello GEE (non si considera la variabile che rileva la presenza e l'eventuale grado di autonomia dei malati cronico-degenerativi) con intervalli di confidenza al 99%, 95% e 90%, struttura di corr. indipendente e legame logit cumulativo*

Parametro	Modalità	Stima	Err. Std.	99%-IC	95%-IC	90%-IC	Z	Pr > Z				
Età	< 30	-0,149	0,155	-0,550	0,251	-0,454	0,155	-0,405	0,106	-1,0	0,3358	
Età	≥ 50 e < 70	0,161	0,110	-0,122	0,445	-0,054	0,377	-0,020	0,343	1,5	0,1428	
Età	≥ 70	0,607	0,152	0,217	0,998	0,310	0,904	0,358	0,857	4,0	<,0001	***
Stato civile	celibe/nubile	0,112	0,156	-0,291	0,154	-0,194	0,418	0,145	0,369	0,7	0,4742	
Stato civile	separato/divorziato	0,016	0,226	-0,565	0,597	-0,426	0,458	-0,355	0,387	0,1	0,9427	
Stato civile	vedovo/a	-0,015	0,170	-0,453	0,422	-0,348	0,318	-0,295	0,264	-0,1	0,9282	
Titolo di studio	alto	-0,157	0,080	-0,363	0,048	-0,314	-0,001	-0,289	-0,026	-2,0	0,0486	**
Disoccupato	no	-0,804	0,262	-1,478	-0,130	-1,317	-0,291	-1,234	-0,374	-3,1	0,0021	***
Autonomia dell'invalido	no	2,006	0,237	1,396	2,617	1,542	2,471	1,617	2,396	8,5	<,0001	***
Autonomia dell'invalido	sì	0,986	0,275	0,277	1,695	0,446	1,525	0,533	1,438	3,6	0,0003	***
Tipo di famiglia	unipersonale	0,422	0,247	-0,213	1,057	-0,061	0,905	0,016	0,828	1,7	0,0871	*
Tipo di famiglia	coppia senza figli	0,029	0,143	-0,338	0,396	-0,250	0,308	-0,206	0,263	0,2	0,8393	
Tipo di famiglia	un gen.con f.a carico	0,541	0,230	-0,052	1,134	0,090	0,992	0,162	0,920	2,3	0,0188	**
Tipo di famiglia	un gen.con f.occupato/i	0,409	0,315	-0,403	1,221	-0,209	1,027	-0,109	0,928	1,3	0,1942	
Tipo di famiglia	altro	0,473	0,452	-0,692	1,638	-0,413	1,360	-0,271	1,217	1,0	0,2954	
Titolo a cui la famiglia occupa l'abitazione	affitto o subaffitto	0,258	0,172	-0,185	0,700	-0,079	0,595	-0,025	0,540	1,5	0,1338	
Titolo a cui la famiglia occupa l'abitazione	altro	1,103	0,321	0,275	1,931	0,473	1,733	0,575	1,632	3,4	0,0006	***
Numero di automobili in famiglia	0	0,677	0,235	0,071	1,283	0,216	1,138	0,290	1,064	2,9	0,0040	***
Numero di automobili in famiglia	1	-0,088	0,141	-0,450	0,275	-0,364	0,188	-0,319	0,144	-0,6	0,5323	
Pasti assieme	solo fine settimana	0,538	0,230	-0,055	1,132	0,087	0,990	0,159	0,918	2,3	0,0195	**
Pasti assieme	sporadicamente	-0,082	0,259	-0,749	0,585	-0,589	0,425	-0,508	0,344	-0,3	0,7511	
Pasti assieme	mai	1,397	0,317	0,582	2,213	0,777	2,018	0,876	1,918	4,4	<,0001	***
Possibilità per la famiglia di aiuto da parenti	sì	-0,558	0,123	-0,874	-0,241	-0,799	-0,316	-0,760	-0,355	-4,5	<,0001	***
Necessità di aiuti economici per la famiglia	no	1,479	0,207	-2,013	-0,946	-1,885	-1,073	-1,820	-1,139	-7,1	<,0001	***
Lutti in famiglia	sì	-0,016	0,120	-0,326	0,294	-0,252	0,220	-0,214	0,182	-0,1	0,8941	
Separazione/divorzio	sì	0,021	0,437	-1,103	1,146	-0,835	0,877	-0,697	0,739	0,1	0,9614	
Furti negli ultimi 3 anni	sì	-0,393	0,229	-0,983	0,198	-0,842	0,057	-0,770	-0,016	-1,7	0,0867	*
Livello di reddito mensile familiare	≤ 6000 e > 3000 €	0,893	0,596	-0,643	2,429	-0,276	2,062	-0,088	1,874	1,5	0,1344	
Livello di reddito mensile familiare	≤ 3000 e > 1500 €	1,101	0,584	-0,404	2,605	-0,044	2,246	0,140	2,062	1,9	0,0595	*
Livello di reddito mensile familiare	≤ 1500 €	1,563	0,596	0,029	3,097	0,395	2,731	0,583	2,543	2,6	0,0087	***

Significatività: *** = 1%, ** = 5%, * = 10%

Anche in questo caso, come avviene non considerando la variabile relativa all'invalidità, un ruolo strategico è giocato dalla variabile che misura l'età degli individui, con un significato analogo al precedente (Tab.6.14).

Tab.6.14. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per malattie cronico-degenerative ed età*

Malati	Età				Totale
	< 30	≥ 30 e < 50	≥ 50 e < 70	≥ 70	
Non autonomo (n=144)	0,0	6,1	26,8	67,1	100,0
Autonomo (n=505)	4,7	14,8	41,9	38,6	100,0
Non è malato (n=6270)	32,7	33,0	26,1	8,2	100,0
Totale (n=6919)	30,1	31,2	27,2	11,5	100,0

Le persone malate e non autonome sono prevalentemente anziane (il 67,1% ha un'età superiore a 70 anni, mentre nessuna ha un'età inferiore a 30 anni). All'aumentare dell'età aumenta in maniera evidente la percentuale di malati non autonomi e in modo meno marcato anche il numero di malati autonomi.

In tabella 6.15 si riportano i risultati ottenuti escludendo dal modello la variabile che misura il livello di reddito familiare. Come nelle tabelle 6.11 e 6.13, i cambiamenti rispetto alla tabella 6.4 sono riportati in grassetto.

Tab.6.15. *Stime (log odds ratio) di un modello GEE (non si considera la variabile che misura il livello di reddito familiare) con intervalli di confidenza al 99%, 95% e 90%, struttura di corr. indipendente e legame logit cumulativo*

Parametro	Modalità	Stima	Err. Std.	99%–IC		95%–IC		90%–IC		Z	Pr > Z	
Età	< 30	-0,151	0,154	-0,548	0,246	-0,453	0,151	-0,404	0,103	-1,0	0,3280	
Età	≥ 50 e < 70	0,060	0,109	-0,221	0,342	-0,154	0,274	-0,120	0,240	0,5	0,5819	
Età	≥ 70	0,257	0,157	-0,148	0,662	-0,051	0,565	-0,001	0,516	1,6	0,1020	
Stato civile	celibe/nubile	0,126	0,155	-0,273	0,526	-0,177	0,430	-0,129	0,381	0,8	0,4147	
Stato civile	separato/divorziato	0,011	0,223	-0,563	0,586	-0,426	0,448	-0,356	0,378	0,1	0,9600	
Stato civile	vedovo/a	0,182	0,177	-0,273	0,637	-0,164	0,528	-0,109	0,472	1,0	0,3029	
Titolo di studio	alto	-0,225	0,080	-0,433	-0,018	-0,383	-0,067	-0,358	-0,093	-2,8	0,0051	***
Disoccupato	no	-0,912	0,264	-1,591	-0,233	-1,429	-0,396	-1,346	-0,479	-3,5	0,0005	***
Autonomia dell'invalido	no	2,427	0,234	1,824	3,030	1,969	2,886	2,042	2,813	10,4	<,0001	***
Autonomia dell'invalido	sì	1,204	0,271	0,506	1,902	0,673	1,735	0,758	1,650	4,4	<,0001	***
Autonomia del malato cronico-degenerativo	no	2,093	0,193	1,596	2,590	1,715	2,471	1,776	2,410	10,8	<,0001	***
Autonomia del malato cronico-degenerativo	sì	0,995	0,126	0,671	1,320	0,749	1,242	0,788	1,203	7,9	<,0001	***
Tipo di famiglia	unipersonale	0,257	0,239	-0,350	0,872	-0,211	0,725	-0,136	0,650	1,1	0,2821	
Tipo di famiglia	coppia senza figli	0,064	0,138	-0,290	0,419	-0,206	0,334	-0,162	0,291	0,5	0,6410	
Tipo di famiglia	un gen.con f.a carico	0,579	0,225	-0,002	1,160	0,137	1,021	0,208	0,950	2,6	0,0102	**
Tipo di famiglia	un gen.con f.occupato/i	0,282	0,311	-0,815	0,468	-0,661	0,314	-0,229	0,793	0,9	0,3643	
Tipo di famiglia	altro	0,389	0,435	-0,732	1,511	-0,464	1,243	-0,327	1,106	0,9	0,3711	
Titolo a cui la famiglia occupa l'abitazione	affitto o subaffitto	0,258	0,169	-0,178	0,694	-0,074	1,692	-0,021	0,536	1,5	0,1277	
Titolo a cui la famiglia occupa l'abitazione	altro	1,054	0,326	0,215	1,893	0,416	0,590	0,519	1,590	3,2	0,0012	***
Numero di automobili in famiglia	0	0,713	0,226	0,130	1,296	0,270	1,157	0,341	1,085	3,1	0,0016	***
Numero di automobili in famiglia	1	0,048	0,133	-0,295	0,391	-0,213	0,309	-0,171	0,267	0,4	0,7184	
Pasti assieme	solo fine settimana	0,464	0,232	-0,135	1,062	0,008	0,919	0,082	0,846	2,0	0,0459	**
Pasti assieme	sporadicamente	-0,047	0,256	-0,706	0,613	-0,549	0,455	-0,468	0,374	-0,2	0,8552	
Pasti assieme	mai	1,376	0,312	0,572	2,181	0,764	1,988	0,863	1,890	4,4	<,0001	***
Possibilità per la famiglia di aiuto da parenti	sì	-0,540	0,123	-0,856	-0,224	-0,781	-0,300	-0,742	-0,338	-4,4	<,0001	***
Necessità di aiuti economici per la famiglia	no	-1,469	0,206	-1,999	-0,939	-1,873	-1,066	-1,808	-1,131	-7,1	<,0001	***
Lutti in famiglia	sì	0,010	0,120	-0,300	0,320	-0,226	0,246	-0,188	0,208	0,1	0,9337	
Separazione/divorzio	sì	-0,074	0,415	-1,144	0,996	-0,888	0,740	-0,757	0,609	-0,2	0,8584	
Furti negli ultimi 3 anni	sì	-0,355	0,228	-0,942	0,231	-0,801	0,091	-0,730	0,019	-1,6	0,1185	

Significatività: *** = 1%, ** = 5%, * = 10%

Ad un livello di significatività dell'1%, rispetto al caso in cui considero nel modello tutte le variabili elencate nella tabella 6.3, emerge la differenza, nella percezione del disagio, tra gli individui con titolo di studio rispettivamente basso e alto e tra le famiglie che non possiedono automobili e quelle che ne hanno almeno due.

Pertanto, non prendendo in considerazione nell'analisi il livello di reddito familiare, entrano in gioco due variabili che ne fanno, in qualche modo, le veci: è nelle famiglie prive di automobili e a basso titolo di studio che il disagio è avvertito con maggiore intensità. Queste famiglie sono perlopiù famiglie di anziani con redditi da sopravvivenza (Tab.6.6, 6.16, 6.17).

Tab.6.16. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per età e livello di reddito*

Età	Reddito				Totale
	≤ 1500 €	> 1500 e ≤ 3000 €	> 3000 e ≤ 6000 €	> 6000 €	
< 30 (n=2050)	9,8	61,8	25,3	3,1	100,0
≥ 30 e < 50 (n=2135)	11,2	64,4	22,5	1,9	100,0
≥ 50 e < 70 (n=1905)	24,4	49,7	22,5	3,3	100,0
≥ 70 (n=829)	49,0	39,1	10,8	1,0	100,0
Totale (n=6919)	18,8	56,7	22,0	2,5	100,0

Tab.6.17. *Distribuzione percentuale della popolazione residente nel Veneto nel 2004-2005 per età e livello di reddito*

Età	Reddito		Totale
	basso	alto	
< 30 (n=2050)	74,0	26,0	100,0
≥ 30 e < 50 (n=2135)	49,8	50,2	100,0
≥ 50 e < 70 (n=1905)	70,2	29,8	100,0
≥ 70 (n=829)	81,5	18,5	100,0
Totale (n=6919)	66,3	33,7	100,0

In conclusione si può notare come la combinazione di problemi sanitari ed economici sia la causa prima di disagio. Caratteristiche individuali, quali l'età e il titolo di studio, o familiari, quale il numero di automobili, hanno esclusivamente la

funzione di variabili “sostitutive”, rappresentanti la mancanza di un reddito adeguato a far fronte alle necessità quotidiane, a cui si aggiungono deficit sanitari che spesso impediscono alla persona di essere autonoma anche nelle elementari attività.

6.3 I modelli di regressione multilivello

In tutti gli studi che si basano su dati raggruppati in classi, l'utilizzo di un modello di analisi di tipo tradizionale pone dei grossi limiti alle analisi che si possono affrontare.

Di fronte all'esigenza di analizzare dati di questo tipo, anche i modelli lineari multilivello (Bryk e Raudenbush, 1992; Goldstein, 1995; Mason, Wong e Entwisle, 1983) forniscono una possibile soluzione al problema. Essi si prestano allo studio di variabili che appartengono a diversi livelli di gerarchia¹, ricorrendo a tecniche statistiche che tengano conto delle possibili forme di dipendenza.

I modelli lineari multilivello, noti anche come modelli lineari gerarchici, modelli a coefficienti casuali (Rosenberg, 1973) o modelli lineari misti (Goldstein, 1986), sono ampiamente usati negli studi sociologici, psicologici, demografici e nell'analisi di dati spaziali.

Essi sono un'estensione dei modelli di regressione usati per lo studio di dati con struttura gerarchica e, quindi, con struttura di varianza complessa. In presenza di unità statistiche annidate in unità di livello superiore la variabilità complessiva deriva da due fonti: la variabilità entro i gruppi (*within*), legata alle differenze tra unità appartenenti allo stesso gruppo, e quella tra i gruppi (*between*), legata alle differenze tra le medie di gruppo.

All'interno di uno stesso gruppo le unità statistiche non sono usualmente indipendenti, bensì tendono ad essere correlate, essendo più simili tra loro di quanto non lo siano alle unità degli altri gruppi. La violazione dell'ipotesi di indipendenza

¹Si definisce gerarchico un insieme di elementi quando esso è articolato in modo tale che le osservazioni dette di primo livello sono aggregate in vari gruppi che, al loro volta, possono essere aggregati in altri insiemi e ciò può ripetersi per successivi raggruppamenti.

è una delle ragioni principali per cui non vengono usati i tradizionali modelli di regressione.

Si potrebbe scegliere di analizzare i dati a livello individuale o di gruppo, ma entrambe queste scelte darebbero luogo a dei problemi. Nel primo caso, è impossibile fare inferenza sui gruppi, cioè trattare i gruppi osservati come un campione casuale da una popolazione di gruppi. Nel secondo, le variabili si riferiscono al gruppo e non all'individuo, per cui non possono nemmeno concettualmente essere usate per indagare le relazioni a livello di individuo; le relazioni a livello di gruppo (cioè tra le medie di gruppo) sono diverse dalle corrispondenti relazioni a livello individuale (*Ecological Fallacy*).

Nel caso di due livelli di osservazione, il primo caratterizza generalmente gli individui, ovvero le unità elementari, il secondo i gruppi (famiglie, classi, scuole, ...). Le variabili di secondo livello si distinguono in:

- globali: caratteristiche intrinseche delle unità macro (gruppi) che vengono rilevate separatamente e per le quali non esiste la corrispondente misura individuale;
- contestuali: indicatori macro ottenuti per aggregazione delle corrispondenti misure individuali; esprimono la misura collettiva delle caratteristiche del singolo.

I vantaggi dell'analisi multilivello per dati organizzati in cluster sono numerosi:

- elimina la distorsione nella stima degli errori standard dei parametri, che si avrebbe utilizzando un semplice modello di regressione lineare in cui, con una struttura di dati di tipo gerarchico, verrebbero a mancare le ipotesi di omoschedasticità e indipendenza;
- definisce variabili esplicative a livello di gruppo in modo da poter calcolare la variabilità tra gruppi ad ogni livello di aggregazione;
- studia le interazioni tra variabili che fanno parte di diversi livelli di gerarchia;
- modella non solo le medie, ma anche le varianze.

In generale, quindi, i modelli multilivello consentono di descrivere la relazione esistente tra la variabile risposta e le variabili individuali e di gruppo, integrando il livello di analisi individuale e quello aggregato. Essi assumono rilevanza per l'analisi di indagini con disegno di campionamento a più stadi: le unità di primo e di secondo stadio sono campioni casuali estratti dalle corrispondenti popolazioni e, pertanto, sarà necessario modellare una componente di varianza ad ogni livello di campionamento e non solo al livello delle osservazioni individuali (Skinner et al., 1989).

6.3.1 Cenni storici

Il problema dell'integrazione tra dimensione micro e macro è da sempre stato centrale nella ricerca sociale. Il conflitto tra l'analisi ecologica e quella individuale non è ancora stato risolto, sebbene si possano segnalare delle tappe importanti nel percorso diretto all'integrazione di queste due dimensioni.

Gli studi condotti fino agli anni sessanta, tuttavia, hanno posto l'attenzione solamente sull'aspetto macro, considerando la dimensione micro come disaggregazione di quella macro. Questa tipicità sottolinea la mancanza di influenza che i ricercatori del tempo attribuivano al contesto più ampio in cui l'individuo trovava il suo spazio. L'analisi ruotava attorno alla definizione delle variabili da utilizzare in modelli di regressione lineare.

Negli anni successivi, gli studiosi iniziarono a considerare questa influenza, concentrando l'attenzione sulla distinzione tra la fonte di variabilità dell'individuo e quella dell'ambiente in cui è inserito. La ricerca sembra invertire tendenza ed interessarsi all'aspetto micro (Courgeau e Baccini, 1998).

Negli anni ottanta si sono introdotti i modelli di regressione multilivello con coefficienti casuali gerarchicamente ordinati. I primi contributi riguardano l'efficienza dei sistemi di insegnamento scolastici (Aitkin et al., 1981); altri contributi si devono a Mason, Wong e Entwisle (1983), Goldstein (1986), Raudenbush e Bryk (1986) e Longford (1987). Il problema del trattamento di osservazioni correlate a causa della struttura gerarchica della popolazione è stato risolto, in particolare grazie allo sviluppo dell'algoritmo EM.

Dalla metà degli anni ottanta, l'analisi multilivello nell'ambito della scienza dell'educazione è consolidata e, a partire dagli anni novanta ha trovato spazio anche in altre discipline (epidemiologia, macroeconomia, sociologia, demografia).

Ad oggi gli approcci proposti per la stima di modelli a coefficienti casuali sono molti, e in continuo sviluppo, soprattutto per la situazione più difficile, cioè nel caso di risposta discreta (dicotomica, politomica o ordinale).

6.3.2 Formulazione dei modelli multilivello

Si disponga di n osservazioni y_1, \dots, y_n , che si vogliono spiegare attraverso p variabili esplicative n -dimensionali $x_{11}, \dots, x_{1p}, x_{21}, \dots, x_{2p}, x_{n1}, \dots, x_{np}$. La generica variabile y_i può essere continua o una variabile dummy, indicatrice delle categorie di appartenenza. Il modello lineare standard assume la forma:

$$y_i = \beta_0 + \sum_{l=1}^p x_{il} \beta_l + \varepsilon_i \quad i = 1, \dots, n$$

dove β_1, \dots, β_p rappresentano i parametri degli effetti fissi, da stimare, e $\varepsilon_1, \dots, \varepsilon_n$ i termini di errore, indipendenti e identicamente distribuiti come una variabile casuale normale standard di media zero e varianza σ_ε^2 .

In forma matriciale, si ha:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (6.16)$$

dove \mathbf{y} rappresenta il vettore di valori osservati y_i , X è la matrice note di variabili esplicative x_{ij} , $\boldsymbol{\beta}$ è il vettore di parametri degli effetti fissi e $\boldsymbol{\epsilon}$ è il vettore non osservato di errori casuali.

Il modello 6.16 è ampiamente utilizzato (Searle, 1971); tuttavia in molte applicazioni l'assunto distributivo di $\boldsymbol{\epsilon}$ è troppo restrittivo. I modelli a effetti misti estendono i modelli lineari generalizzati consentendo una maggiore flessibilità nella specificazione della matrice di covarianza di $\boldsymbol{\epsilon}$.

Per ogni individuo i appartenente al gruppo j , il vettore di risposte y_{ij} è dato da:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij} \quad (6.17)$$

$$i = 1, \dots, n_j$$

$$j = 1, \dots, J$$

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

con x_{ij} variabile esplicativa di primo livello e σ^2 varianza residua di primo livello.

Ad ogni gruppo corrisponde una coppia (β_{0j}, β_{1j}) . Tali coppie sono realizzazioni indipendenti da una distribuzione di probabilità, tipicamente normale bivariata:

$$\begin{pmatrix} \beta_{0j} \\ \beta_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} \gamma_{00} \\ \gamma_{10} \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]$$

β_{0j} e β_{1j} sono, rispettivamente, l'intercetta e il coefficiente angolare del j -esimo gruppo. γ_{00} e γ_{10} rappresentano, rispettivamente, l'intercetta e la pendenza media, τ_{00} e τ_{11} le varianze dell'intercetta e della pendenza e τ_{01} la covarianza tra l'intercetta e la pendenza.

Al secondo livello si aggiungono altre due relazioni funzionali:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + u_{0j} \quad (6.18)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_j + u_{1j} \quad (6.19)$$

con z_j variabile di secondo livello.

Le variabili di secondo livello rappresentano caratteristiche dei gruppi, necessarie a definire un modello per i parametri di primo livello (β_{0j}, β_{1j}) e a ridurre le varianze di secondo livello.

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right]$$

Dalle equazioni 6.17, 6.18 e 6.19 si ottiene il modello lineare gerarchico in forma combinata:

$$y_{ij} = (\gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}z_j + \gamma_{11}z_jx_{ij}) + (u_{1j}x_{ij} + u_{0j} + \epsilon_{ij})$$

In questo modello, il termine di intercetta è γ_{00} , il coefficiente angolare di x_{ij} è γ_{10} e il coefficiente angolare di z_j è γ_{01} . u_{0j} sono componenti erratiche di secondo

livello, indipendenti dalle componenti erratiche di primo livello; u_{1j} sono effetti casuali: i gruppi sono visti come un campione tratto da una popolazione per la quale è possibile definire variabili esplicative che aiutino a spiegare la variabilità tra gruppi. Tale modello combinato è costituito da due parti: la prima parentesi contiene gli effetti fissi, la seconda gli effetti casuali; $\gamma_{11}z_jx_{ij}$ è un termine di interazione.

Pertanto, ad ogni gruppo corrisponde una coppia di parametri con distribuzione normale bivariata rispetto all'intera popolazione dei gruppi. Gli errori associati a questi termini rappresentano la variabilità residua relativa ai coefficienti β_{0j} e β_{1j} , una volta introdotte le variabili di secondo livello z_j .

In base a questo modello, tanto la varianza residua quanto la correlazione tra le osservazioni di uno stesso gruppo dipendono dal valore della variabile indipendente:

$$Var(y_{ij}) = \sigma^2 + \tau_{00} + \tau_{11}x_{ij}^2 + 2\tau_{10}x_{ij}$$

$$Cov(y_{ij}, y_{i'j'}) = \tau_{00} + \tau_{11}x_{ij}x_{i'j'} + \tau_{10}(x_{ij} + x_{i'j'})$$

e si è, pertanto, in presenza di eteroschedasticità tra le risposte (Skrondal, Rabe - Hesketh e Pickles, 2004).

La parte di variabilità spiegata dai residui di secondo livello è data dal coefficiente di correlazione intraclasse ρ , misurato dal rapporto tra la covarianza e la varianza:

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

che misura il grado di omogeneità tra osservazioni appartenenti allo stesso gruppo.

Il nome *modello a effetti misti* deriva dal fatto che esso contiene parametri sia degli effetti fissi (β), sia degli effetti casuali (γ) (Searle, Casella e McCullach, 1992).

In notazione matriciale:

$$\mathbf{y} = X\boldsymbol{\beta} + Z\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

dove Z è una matrice diagonale a blocchi, con i blocchi che corrispondono alle unità di secondo livello e $\boldsymbol{\gamma}$ è il vettore ignoto dei parametri degli effetti casuali.

γ e ϵ sono distribuiti normalmente con media e varianza date da:

$$E \begin{bmatrix} \gamma \\ \epsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Var \begin{bmatrix} \gamma \\ \epsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

La varianza di \mathbf{y} è, pertanto, data da:

$$V = ZGZ' + R$$

È possibile modellare V costruendo la matrice di disegno degli effetti casuali Z e specificando la struttura di covarianza per G e R .

Si tratta di una specificazione generale del modello misto, che è possibile semplificare se Z contiene variabili dummy, G ha una struttura diagonale e $R = \sigma^2 I$, con I matrice identica di dimensione $n_j \times n_j$.

A differenza dei modelli multilivello, i modelli marginali definiscono una formulazione iniziale per la struttura di covarianza e forniscono stime solo per i parametri fissi presenti nel modello, trattando i parametri casuali come “rumore”. Altre limitazioni dei modelli marginali sono riportate in Lindsey e Lambert (1998).

6.3.3 Stima dei modelli multilivello

Si consideri dapprima il modello di regressione lineare semplice:

$$y_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \epsilon_i \quad i = 1, \dots, n$$

Lo stimatore di massima verosimiglianza della varianza residua è dato da:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij}\hat{\beta}_j \right)^2$$

²Il modello lineare generalizzato è un caso particolare con $Z = 0$ e $R = \sigma^2 I$.

Tale stimatore è distorto e $E(\hat{\sigma}^2) = \frac{n-p}{n} \sigma^2$; pertanto si usa generalmente lo stimatore corretto:

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} \hat{\beta}_j \right)^2$$

La stessa distorsione verso il basso si ritrova nei modelli a effetti misti perché non si tiene conto dei gradi di libertà persi nella stima di β .

Per far fronte a questo problema Patterson e Thompson (1971) proposero il cosiddetto metodo REML (*Restricted/Residual Maximum Likelihood*), che porta a stime corrette dei parametri di varianza e covarianza (Skrondal e Rabe - Hesketh, 2004) e ha il vantaggio di trattare dati con valori mancanti casualmente (Rubin, 1976).

La stima di un modello a effetti misti si ottiene minimizzando:

$$(\mathbf{y} - X\beta)'V^{-1}(\mathbf{y} - X\beta)$$

Tuttavia, è necessaria una stima di V e, pertanto, di G e R^3 .

La corrispondente funzione di log-verosimiglianza ha la seguente forma:

$$\begin{aligned} l_R(G, R) &= -\frac{1}{2} \log |V| - \frac{1}{2} \log |X'V^{-1}X| \\ &= -\frac{1}{2} \mathbf{r}'V^{-1}\mathbf{r} - \frac{n-p}{2} \log(2\pi) \end{aligned}$$

con $\mathbf{r} = \mathbf{y} - X(X'V^{-1}X)^{-1}X'V^{-1}\mathbf{y}$ e p rango di X .

Si minimizza -2 volte tale funzione usando un algoritmo di Newton-Raphson.

Lindstrom e Bates (1988) forniscono le ragioni per cui si debba preferire l'algoritmo di Newton-Raphson all'EM (*Expectation-Maximum*) descritto in Dempster, Laird e Rubin (1977) e Laird, Lange e Stram (1987).

Uno dei vantaggi risiede nel fatto che la matrice delle derivate seconde valutata nel punto di ottimo, H , è tale per cui, secondo la teoria asintotica di massima

³In molte applicazioni il metodo più appropriato è quello basato sulla verosimiglianza, sfruttando l'assunto di normalità di γ e ϵ (Hartley e Rao, 1967; Patterson e Thompson, 1971; Harville, 1977; Laird e Ware, 1982; Jennrich and Schluchter, 1986)

verosimiglianza (Serfling, 1980), $-2H^{-1}$ è una matrice asintotica di varianze e covarianze dei parametri stimati di G e R . In questo modo si possono costruire test e intervalli di confidenza basati sulla normalità asintotica.

Ottenute le stime di G e R , \hat{G} e \hat{R} , quelle di β e γ si hanno risolvendo (Henderson, 1984):

$$\begin{bmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X'\hat{R}^{-1}\mathbf{y} \\ Z'\hat{R}^{-1}\mathbf{y} \end{bmatrix}$$

da cui:

$$\hat{\beta} = (X'\hat{V}^{-1}X)X'\hat{V}^{-1}\mathbf{y}$$

$$\hat{\gamma} = \hat{G}Z'\hat{V}^{-1}(\mathbf{y} - X\hat{\beta})$$

che ricordano lo stimatore empirico di Bayes (Laird e Ware, 1982).

Tali stime possono essere anche usate per confrontare gli effetti fissi e casuali rispettivamente di diverse unità statistiche e come residui in test diagnostici.

Se G e R sono note, $\hat{\beta}$ è lo stimatore BLUE (*Best Linear Unbiased Estimator*) di β e $\hat{\gamma}$ è lo stimatore BLUP (*Best Linear Unbiased Predictor*) di γ , stimatori ottimi nel senso che minimizzano l'errore quadratico medio.

Tuttavia, G e R sono usualmente ignote; pertanto, la matrice \hat{C} di varianze e covarianze approssimata di $(\hat{\beta} - \beta, \hat{\gamma} - \gamma)$ conterrà le loro stime:

$$\hat{C} = \begin{bmatrix} X'\hat{R}^{-1}X & X'\hat{R}^{-1}Z \\ Z'\hat{R}^{-1}X & Z'\hat{R}^{-1}Z + \hat{G}^{-1} \end{bmatrix}^{-1}$$

\hat{C} tende a sottostimare la vera variabilità campionaria di $(\hat{\beta}, \hat{\gamma})$, non tenendo in considerazione la distorsione nelle stime di G e R (Kackar e Harville, 1984; Kass e Steffly, 1989; Prasad e Rao, 1990).

6.3.4 Applicazione

Obiettivo della seguente analisi è l'individuazione delle caratteristiche familiari e individuali che concorrono a determinare una situazione di disagio per la famiglia.

Il modello proposto per l'analisi è un modello di regressione multilivello in cui il primo livello è costituito dagli individui ($j = 1, \dots, J$) e il secondo dalle famiglie ($i=1, \dots, n_j$).

Il pacchetto statistico utilizzato per l'implementazione di un'analisi multilivello è il SAS.

Per molti anni, la procedura maggiormente utilizzata per lo studio di modelli a effetti fissi è stata la GLM (*General Linear Models*), il cui obiettivo è quello di costruire modelli lineari su dati sperimentali od osservazionali. Una procedura alternativa, di più recente costruzione, è la MIXED, che risulta più potente nella gestione di modelli di analisi della varianza, offrendo la possibilità di definire diversi tipi di assunzioni relativamente ai dati in esame.

La variabile risposta è la percezione di disagio da parte della famiglia, misurata su una scala a quattro livelli (Par.3.2.1), mentre le variabili esplicative sono quelle rappresentate nella tabella 6.3. Si dispone di 6919 individui raggruppati in 2465 famiglie.

Si è tenuta in considerazione la correlazione tra i soggetti appartenenti alla stessa famiglia attraverso l'inclusione di effetti casuali⁴.

L'importanza di un'analisi a effetti casuali risiede nel fatto che consente di fare inferenza sul totale delle famiglie venete, anziché solamente sulle famiglie appartenenti al nostro campione.

L'unico effetto casuale introdotto nel nostro modello è rappresentato dalla variabile identificatrice della famiglia: poiché le famiglie sono state selezionate casualmente, l'effetto della famiglia si considera una variabile casuale di media zero e varianza ignota.

La procedura calcola le variabili dummy associate a ciascuna variabile di classificazione, nel nostro caso la famiglia, inclusa nel modello, al fine di costruire la matrice Z .

La matrice di disegno X è costituita da una colonna di 1 per modellare l'in-

⁴Un effetto casuale è tale in quanto i suoi livelli possono essere sostituiti con altri senza alterare gli obiettivi di ricerca o le conclusioni.

tercetta e da tante altre colonne quanti sono i parametri da stimare, una volta dicotomizzate le variabili incluse nel modello.

Le matrici G e R sono diagonali a blocchi e contengono le componenti di varianza rispettivamente per la variabile che identifica la famiglia e per gli individui facenti parte della medesima famiglia.

Si assume che le osservazioni effettuate su diverse famiglie siano statisticamente indipendenti e che la matrice di varianze e covarianze sia la stessa per ogni famiglia.

La correlazione tra i soggetti appartenenti alla medesima famiglia può essere modellata in diversi modi, potendo anche definire varianze non costanti tra i gruppi.

La scelta della struttura di covarianza è importante perché i test per gli effetti fissi, ossia per le variabili incluse nel modello, sono funzione di questa e i risultati possono essere sbagliati se si specifica una scorretta struttura di covarianza.

Pertanto, si sono poste a confronto tre strutture di covarianza: a componenti di varianza, simmetrica e non strutturata. La struttura a componenti di varianza presuppone $R = \sigma^2 I$. La simmetrica prevede che la correlazione tra tutte le coppie di individui all'interno della stessa famiglia sia costante. La non strutturata la si utilizza quando nessuna struttura di covarianza sembra appropriata e ogni covarianza è stimata individualmente sulla base dei dati osservati; ne segue che l'analisi è meno potente rispetto a quella che fa uso di una specifica struttura di covarianza.

Esistono diverse misure statistiche per valutare l'adeguatezza del modello; le più comuni sono il test rapporto di verosimiglianza e i criteri di Akaike e Schwarz. Nel nostro caso, si è scelta la struttura a componenti di varianza alla quale corrisponde il criterio di informazione di Akaike (AIC) maggiore.

Il metodo utilizzato per la stima delle componenti di varianza, ossia delle varianze dei parametri degli effetti casuali, è quello REML, ossia della massima verosimiglianza ristretta (*residual maximum likelihood*) (Jennrich and Schluchter, 1986).

La funzione da minimizzare è $-2l_R$, ossia -2 volte il logaritmo della verosimiglianza ristretta, e la procedura utilizzata sfrutta, per ottimizzare la funzione di

verosimiglianza, un algoritmo di Newton-Raphson che fa uso delle derivate prime e seconde della funzione obiettivo per trovarne iterativamente il minimo.

Le stime delle componenti di varianza per la variabile identificatrice di famiglia e residua (ϵ_{ij}) sono riportate nella tabella 6.18 con i rispettivi intervalli di confidenza al 5%.

Tab.6.18. *Stime delle componenti di varianza per la variabile identificatrice della famiglia e della componente di varianza residua e relativi intervalli di confidenza al 95%*

Parametro di covarianza	Stima	95%-IC
Famiglia	0,3443	0,3218 - 0,3692
Residua	0,0010	0,0009 - 0,0011

Le stime delle componenti casuali risultano entrambe significative. La significatività di τ_{00} è indice dell'elevato potere esplicativo della variabile di secondo livello (famiglia).

Il coefficiente di correlazione intraclasse è prossimo a uno ($0,3443/(0,3443 + 0.0010)$), mostrando come quasi tutta la variabilità totale sia imputabile ad un effetto dovuto alle famiglie.

I parametri di covarianza sono ciò che distingue i modelli lineari misti dai modelli lineari standard.

Nella tabella 6.19 si riportano i risultati relativi agli effetti fissi; si tratta di test parziali poiché tengono in considerazione tutti gli altri effetti fissi presenti nel modello.

Tab.6.19. *Stima degli effetti fissi di un modello multilivello con struttura di covarianza a componenti di varianza*

Parametro	Modalità	Stima	Err. Std.	t	Pr > Z	
Intercetta		1,918	0,113	16,9	<0,0001	***
Età	< 30	-0,043	0,028	-1,5	0,1267	
Età	≥ 50 e < 70	0,017	0,020	0,9	0,3782	
Età	≥ 70	0,070	0,031	2,3	0,0236	**
Stato civile	celibe/nubile	0,046	0,028	1,6	0,1063	
Stato civile	separato/divorziato	0,025	0,059	0,4	0,6675	
Stato civile	vedovo/a	0,051	0,042	1,2	0,2288	
Titolo di studio	alto	-0,028	0,015	-1,9	0,0600	*
Disoccupato	no	-0,218	0,067	-3,3	0,0011	***
Autonomia dell'invalido	no	0,989	0,067	14,7	<0,0001	***
Autonomia dell'invalido	sì	0,381	0,076	5,0	<0,0001	***
Autonomia del malato cronico-degenerativo	no	0,832	0,052	16,1	<0,0001	***
Autonomia del malato cronico-degenerativo	sì	0,295	0,029	10,1	<0,0001	***
Tipo di famiglia	unipersonale	0,134	0,047	2,9	0,0041	***
Tipo di famiglia	coppia senza figli	0,013	0,022	0,6	0,5432	
Tipo di famiglia	un gen.con f.a carico	0,108	0,037	2,9	0,0033	***
Tipo di famiglia	un gen.con f.occupato/i	0,111	0,031	0,8	0,4379	
Tipo di famiglia	altro	0,091	0,058	1,5	0,1202	
Titolo a cui la famiglia occupa l'abitazione	affitto o subaffitto	0,069	0,023	2,9	0,0033	***
Titolo a cui la famiglia occupa l'abitazione	altro	0,303	0,051	5,9	<0,0001	***
Numero di automobili in famiglia	0	0,224	0,039	5,7	<0,0001	***
Numero di automobili in famiglia	1	0,008	0,017	0,5	0,6410	
Pasti assieme	solo fine settimana	0,078	0,030	2,6	0,0092	***
Pasti assieme	sporadicamente	-0,047	0,037	-1,2	0,2107	
Pasti assieme	mai	0,422	0,047	9,0	<0,0001	***
Possibilità per la famiglia di aiuto da parenti	sì	-0,113	0,016	-7,1	<0,0001	***
Necessità di aiuti economici per la famiglia	no	-0,492	0,030	-16,3	<0,0001	***
Lutti in famiglia	sì	-0,001	0,015	-0,1	0,9519	
Separazione/divorzio	sì	0,007	0,068	0,1	0,9225	
Furti negli ultimi 3 anni	sì	-0,095	0,028	-3,4	0,0007	***
Livello di reddito mensile familiare	≤ 6000 e > 3000 €	0,114	0,046	2,4	0,0141	**
Livello di reddito mensile familiare	≤ 3000 e > 1500 €	0,143	0,045	3,1	0,0016	***
Livello di reddito mensile familiare	≤ 1500 €	0,256	0,049	5,2	<0,0001	***

Significatività: *** = 1%, ** = 5%, * = 10%

Gli errori standard approssimati delle stime dei parametri degli effetti fissi $\hat{\beta}$ sono dati dalla radice quadrata degli elementi diagonali di $(X'\hat{V}^{-1}X)^{-1}$. Lo standard error non elevato indica che le stime dei parametri sono piuttosto affidabili.

Oltre alle stime e agli errori standard approssimati, si riporta la statistica t , calcolata come rapporto tra la stima e l'errore standard.

Le stime degli effetti fissi per le variabili incluse nel modello mostrano che, ad un livello di significatività del 99%, le caratteristiche che determinano una situazione di rischio di disagio familiare sono la presenza di componenti disoccupati, la presenza di componenti invalidi o affetti da malattie cronico-degenerative e la loro mancanza di autonomia, l'essere la famiglia di tipo unipersonale, l'abitazione in affitto o a titolo gratuito, la mancanza di automobili, l'abitudine a non condividere i pasti, l'impossibilità di contare sull'aiuto da parte di parenti, la necessità di recenti aiuti economici, il non accadimento di furti e il basso livello di reddito.

Ad un livello del 95% risultano a rischio di disagio anche le famiglie con componenti anziani (ultra settantenni) e ad un livello del 99% quelle con componenti con titolo di studio basso (diploma di scuola media superiore - 2/3 anni - o inferiore).

I risultati sono molto simili a quelli ottenuti sfruttando i modelli marginali (Tab. 6.4) e ne costituiscono, pertanto, una conferma. Le differenze sono evidenziate nella tabella 6.4 in grassetto e di seguito si vanno ad analizzare le più salienti.

Ad un livello di significatività del 99%, l'analisi che sfrutta i modelli multilivello evidenzia una significativa differenza, in termini di percezione di disagio, tra le famiglie che vivono in affitto e quelle che hanno l'abitazione in proprietà (p-value = 0,0033). Il disagio è avvertito solamente nelle prime. Nella precedente analisi, che fa uso dei modelli marginali, tale differenza non emerge, ma risulta significativa solamente quella tra le famiglie che vivono in proprietà e quelle che occupano l'abitazione secondo un altro titolo (prevalentemente gratuito).

Ad un livello del 95% si possono discriminare le famiglie che si ritengono colpite da disagio da quelle che non avvertono problemi di una certa importanza rispetto all'età dei componenti: il disagio è maggiore nelle famiglie con componenti oltre i 70 anni d'età (p-value = 0,0236).

La variabile che misura il livello di reddito familiare assume una certa importanza nell'analisi: con una significatività del 99% (p-value < 0,0001 e 0,0016 rispettivamente) la percezione del disagio è avvertita con maggior frequenza nelle famiglie con reddito inferiore a 1500 euro mensili e in quelle con reddito compreso tra 1500 e 3000 euro rispetto a quelle con reddito superiore a 6000 euro. Invece, la differenza, in termini di disagio, tra le famiglie con reddito compreso tra 3000 e 6000 euro e quelle con reddito superiore è significativa al 95%. Nell'analisi con i modelli marginali la variabile che misura il reddito non discrimina le famiglie in modo così dettagliato, bensì mette in risalto, con una significatività del 99%, solamente il maggior disagio avvertito dalle famiglie con reddito inferiore a 1500 euro rispetto a quelle con reddito superiore a 6000 e, ad una significatività del 10% il maggior disagio per quelle con reddito compreso tra 1500 e 3000 euro.

Come osservato dalle analisi effettuate nei precedenti capitoli, l'effetto del reddito sul disagio si può spiegare solo prendendo in considerazione altre variabili legate alle condizioni sanitarie dei componenti, alle caratteristiche strutturali della famiglia, alla presenza di disoccupati.

Non si ritiene opportuno commentare ulteriormente i risultati perché si rischierebbe di duplicare quelli ottenuti con lo studio dei modelli marginali o, analizzando le differenze tra le due analisi, introdurre delle forzature dal momento che le differenze sono sottili e particolareggianti.

Capitolo 7

Conclusioni

Le famiglie venete sono di piccola dimensione: in media formate da due (28,2%) o tre (26,6%) componenti; quelle unipersonali (13,9%), in rapido aumento negli ultimi anni, sono costituite principalmente da ultrasessantacinquenni (61,8% delle famiglie unipersonali), donne nel 70,5% dei casi, in grande maggioranza vedovi/e (56,1%). Le famiglie intervistate sono formate per il 51,7% da coppie con figli e per il 23,3% da coppie senza figli: oltre il 70% delle famiglie è, pertanto, composto da una coppia, indice del fatto che la famiglia di tipo tradizionale è ancora una realtà molto forte nel nostro paese.

La famiglia svolge un ruolo tutt'altro che trascurabile nella determinazione delle condizioni sociali, economiche, relazionali in cui versano gli individui. Essa può, dunque, agire come unità all'interno della quale si attenuano, fino a scomparire, le difficoltà vissute da qualcuno dei suoi membri (in tal caso essa rappresenta una risorsa per gli individui) o come luogo in cui si accentua la fragilità cui sono esposti i singoli, essendo essa stessa fonte di emarginazione e deprivazione.

Il 12,8% delle famiglie venete ha problemi di salute, il 12,0% economici e il 4,0% socio-relazionali. Il 74,6% non si ritiene colpito da problemi generatori di disagio, il 17,9% si ritiene colpito da problemi risolvibili con le proprie forze, il 5,4% da problemi risolvibili con l'aiuto esterno alla famiglia e il restante 2,1% da

gravi problemi non risolvibili, nemmeno con l'intervento dei servizi.

Le famiglie più esposte a serie difficoltà sono quelle nelle quali sono presenti componenti con problemi di salute e quelle in condizione di non autosufficienza economica, caratteristiche che le rendono più deboli e vulnerabili anche sul piano sociale. Ha disagio grave il 30,9% delle famiglie con invalidi al suo interno e il 14,3% di quelle con reddito inferiore a 1500 euro netti mensili.

Spesso le disgrazie si accompagnano e diversi problemi si presentano contemporaneamente nella stessa famiglia: una famiglia povera, rispetto ad una benestante, ha più difficoltà a sostenere le spese sanitarie e, pertanto, a risolvere anche i problemi di salute meno gravi. Le famiglie con problemi economici e sanitari hanno problemi gravi nel 21,9% dei casi: si tratta spesso di famiglie di anziani non autonomi e con redditi da sopravvivenza; quelli che vivono soli sono donne nel 78,2% dei casi, vedove, che percepiscono la sola pensione di anzianità.

Una tipologia di famiglia a rischio di disagio è anche quella costituita da giovani che lasciano la famiglia di origine per formarne una propria, non ancora autonomi dal punto di vista economico.

Altre situazioni problematiche, che tuttavia si possono collocare in una zona grigia, e non nera, di disagio, ossia in una zona di affanno più che di disagio vero e proprio, sono quelle a cui la famiglia riesce a far fronte con le proprie forze.

Si tratta ad esempio delle famiglie in cui è presente un genitore con figli a carico. Le famiglie monogenitoriali sono colpite da disagio grave o estremo nel 5,2% dei casi, rispetto ad un 2,5% in quelle in cui, *ceteris paribus*, la composizione della famiglia è di diversa natura. Il 27,4% delle prime riesce a risolvere i problemi che si presentano loro. Queste famiglie hanno perlopiù problemi socio-relazionali (23,6%) ed economici (21,4%).

Ancora, mentre il lutto di un figlio è, per i genitori coinvolti, motivo di disagio estremo a cui nessuno può porre rimedio, il lutto di un parente anziano, malato terminale, è dapprima un disagio estremo perché richiede tempi di cura ai familiari, poi un affanno e, successivamente, una sorta di liberazione della famiglia da un periodo frustrante di assistenza.

Eventi critici quali sfratti, furti o incidenti si sono rivelati poco o per nulla

significativi nella determinazione del disagio.

Il ritratto della famiglia a rischio di disagio è stato tracciato grazie ad un'approfondita analisi di segmentazione che ha voluto studiare nel dettaglio le diverse forme e le diverse intensità di disagio da cui la famiglia può essere colpita. A tale analisi sono state di sostegno altre specifiche per lo studio di dati con struttura gerarchica.

Si è cercata una visione globale del fenomeno: il livello di vita degli individui all'interno di una famiglia è condizionato da vari aspetti, tra cui le capacità economiche, il livello di istruzione, lo stato di salute, le condizioni abitative e altri fattori rilevanti.

L'uso di diversi metodi applicati agli stessi dati trova la sua giustificazione nel tentativo di dare un valore fondato e preciso ai risultati: le caratteristiche delle famiglie venete a rischio di disagio messe in rilievo contemporaneamente da ciascun metodo sono, verosimilmente, quelle che più discriminano le famiglie rispetto al disagio percepito.

La prima tecnica impiegata per la ricerca dei "gruppi a rischio" è stata quella dell'analisi di segmentazione, che ha avuto l'obiettivo di suddividere il campione di famiglie in gruppi di unità, sulla base della combinazione di caratteristiche che davano il contributo maggiore all'interpretazione del fenomeno del disagio.

La difficoltà riscontrata nella sua applicazione, e in particolare nella scelta del criterio sulla base del quale suddividere le unità, è stata causata dalla particolare distribuzione delle frequenze della percezione del disagio, fenomeno caratterizzato da gradazioni progressive di gravità, con una forte concentrazione di unità in corrispondenza dell'assenza del fenomeno stesso e un decadimento molto rapido in corrispondenza delle massime intensità di disagio.

Le proposte metodologiche, impiegate nella ricerca del criterio di segmentazione che fosse in grado di creare coppie di gruppi tra loro massimamente differenti, fanno riferimento a quattro approcci diversi di analisi asimmetrica, basati, rispettivamente sulla massimizzazione della distanza distributiva, della differenza tra le mediane condizionate, dei rischi relativi e della significatività mediante soluzione non

parametrica. Gli ultimi due si sono rilevati adatti allo studio del nostro fenomeno.

L'indicatore di dipendenza basato sui rischi relativi fa ricorso ai modelli a odds proporzionali, che riescono ad evidenziare la differenza tra due distribuzioni anche nei casi di frequenze prossime a zero per le categorie di disagio estremo. La soluzione non parametrica è basata sulla metodologia dei test di permutazione; nonostante la sua funzionalità, essa fornisce solo una misura di significatività e, per questo motivo, si è preferito il precedente approccio: l'analisi di segmentazione è stata condotta usando, come variabile criterio, una trasformata del disagio in un odds proporzionale.

L'analisi di segmentazione è stata condotta, in un primo momento, considerando una scala a quattro modalità per la misura del disagio e, successivamente, dicotomizzando tale variabile in modo da individuare le famiglie con e senza disagio.

La dicotomizzazione non distingue le famiglie che si trovano in una zona nera di disagio da quelle che si trovano in una zona grigia, di affanno, a differenza di quanto avviene considerando una scala di misura più dettagliata. Dal confronto delle due analisi è possibile notare che l'affanno non è l'anticamera del disagio, bensì individua altre tipologie di famiglie con caratteristiche loro proprie.

Le variabili messe in rilievo da entrambe le analisi di segmentazione sono quelle che rilevano la presenza di invalidi, l'autonomia dei malati cronico-degenerativi, il livello di reddito, il titolo a cui la famiglia occupa l'abitazione e il tipo di famiglia.

Le variabili messe in rilievo dalle analisi di segmentazioni, dicotomica e politomica, condotte attraverso il pacchetto statistico SAS sono circa le stesse: nell'analisi dicotomica entra anche quella che rileva l'età media dei componenti, mentre non entrano quelle che rilevano la condivisione dei pasti da parte dei membri della famiglia e la possibilità di aiuti da parte di parenti, rispettivamente.

Altre caratteristiche familiari che sono causa di disagio, evidenziate solo dall'analisi di segmentazione dicotomica condotta con il software LAID-OUT *proxy* di quelle comuni alle due analisi precedenti.

Un output singolare è fornito dall'analisi neurale, che ha avuto come obiettivo quello di confrontare la distribuzione osservata del disagio con quella prevista da

alcuni dataset simulati, al fine di confermare o smentire i risultati dell'analisi di segmentazione.

Tutti i confronti tra ciascuna coppia di modalità di tutte le variabili mostrano una significativa differenza tra le distribuzioni del disagio. Questo risultato sorprende non poco: probabilmente la rete costruita è troppo sensibile a piccole variazioni nella distribuzione del disagio, non riuscendo, in tal modo, a mettere in evidenza le famiglie che si trovano in una situazione di reale disagio.

Si tratta di una debolezza del metodo applicato ai dati a nostra disposizione e, non ottenendo risultati migliori neppure modificando qualche parametro della rete, si è proceduto con lo studio di modelli per dati correlati.

I modelli applicati sono i marginali e i multilivello, adatti alla struttura gerarchica di cui si dispone: individui raggruppati in famiglie.

Rispetto all'analisi di segmentazione, questi modelli forniscono risultati più puntigliosi: a diversi livelli di significatività (1%, 5% e 10%), essi pongono a confronto tutte le possibili combinazioni di coppie di ciascuna variabile. Inoltre, per la ricerca delle caratteristiche che possono essere fonte di disagio familiare, vengono presi in considerazione aspetti, non solo familiari, ma anche individuali.

Le conclusioni dei due metodi di analisi per dati correlati sono circa le medesime. Le poche differenze sono legate al fatto che, nei modelli marginali e multilivello, alcune variabili risultano significative a livelli diversi.

Nella sostanza, i risultati non differiscono neppure da quelli ottenuti via segmentazione e ne rappresentano, pertanto, una conferma.

La conoscenza delle tipologie di famiglie più a rischio di disagio permetteranno alle istituzioni di progettare risposte adeguate ed interventi efficaci di politiche familiari.

I risultati ottenuti si configurano con una duplice valenza: da un lato, rispondono ad un immediato bisogno di conoscenze finalizzato al miglioramento delle iniziative e delle politiche di intervento (pubblico e privato); dall'altro, forniscono le linee metodologiche ed un primo contributo per realizzare un'azione di costante monitoraggio del fenomeno del disagio, sia replicando l'esperienza di indagine nella

stessa regione del Veneto, sia allargandola all'intero paese.

Oltre alle caratteristiche che possono essere fonte di disagio, si sono individuati i fattori che possono prevenirlo.

La giovane età dei componenti può comportare qualche difficoltà, soprattutto di carattere economico, ma spesso costituisce un punto di forza, che facilita il superamento dei problemi: il 4,0% delle famiglie con componenti di età inferiore a 30 anni avverte disagio di una certa importanza, contro il 18,3% di quelle con età superiore a 70 anni.

Un'altra caratteristica di tipo strutturale che preserva dal rischio di disagio è l'essere la famiglia composta da una coppia di genitori con figli. La presenza di un bambino in casa abbassa drasticamente il rischio di disagio: il 94,9% di queste famiglie non avverte alcun tipo di problema o ha problemi risolvibili al proprio interno. I figli rappresentano, pertanto, una risorsa che la famiglia riesce a gestire con le proprie forze o con l'aiuto di una rete parentale solida su cui fare affidamento.

Per concludere, si ritiene opportuno esprimere un commento sul rapporto tra le famiglie e i servizi: la rete secondaria di aiuti, quella rappresentata dalle istituzioni, è poco utilizzata, mentre ha un notevole peso la rete primaria, costituita da genitori, figli, fratelli/sorelle, a cui la famiglia si rivolge in caso di bisogno.

Una bassa percentuale di famiglie venete cerca nell'ambito dei servizi offerti dalle istituzioni l'aiuto per far fronte ai propri problemi: l'1,4% fa regolarmente ricorso al consultorio familiare, il 2,4% ai servizi sociali del comune e ben il 21,9% ai servizi socio-sanitari dell'ULSS.

È alta la percentuale di famiglie che non sa se esistano servizi nelle vicinanze della propria residenza (46,7%), problema solo apparente dal momento che, in caso di reale necessità, la famiglia si adopera per avere informazioni in merito.

Sebbene nel nostro paese la famiglia sia sempre stata, per i suoi componenti, il punto di riferimento fondamentale in caso di bisogno, è bene, da parte delle istituzioni, pubblicizzare i servizi offerti e svolgere una campagna di sensibilizzazione rivolta soprattutto alle famiglie colpite da disagio affinché possano sollevarsi, almeno in parte, dal peso dei problemi che le sovrasta.

Bibliografia

- [1] Accornero A., Carmignani F. (1986) *I paradossi della disoccupazione*. Il Mulino, Bologna
- [2] Agresti A. (1984) *Analysis of Ordinal Categorical Data*. John Wiley and Sons, New York
- [3] Agresti A. (2002) *Categorical Data Analysis*. Wiley-interscience, Hoboken
- [4] Aitkin M., Anderson D., Hinde J. (1981) Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society, Part A*, **144**, 148-161
- [5] Angeniol B. (1988) *Self-organizing features maps and the travelling salesman problem*. Neural Network, vol.1
- [6] Armitage P. (1971) *Statistical Methods in Medical Research*. Oxford, Blackwell
- [7] Azzalini A., Scarpa B. (2004) *Analisi dei dati e data mining*. Springer, New York
- [8] Barbagli M. (1990) *Provando e riprovando. Matrimonio, famiglia e divorzio in Italia e in altri paesi occidentali*. Il Mulino, Bologna
- [9] Beresford B. (1994) Resources and strategies: how parents cope with the care of a disabled child. *Journal of Child Psychology and Psychiatry*, **35**, **1**, 171-209
- [10] Bishop C.M. (1995) *Neural network for pattern recognition*. Clarendon Press, Oxford

- [11] Breda C., Fabbris L. (2002) Il rischio di disagio nelle famiglie di Padova. In: Puggioni G. (a cura di) *Modelli e metodi per l'analisi di rischi sociali e sanitari*. Cleup, Padova, 161-184
- [12] Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984) *Classification and Regression Trees*. Wadsworth and Brooks, Pacific Grove, CA
- [13] Brodley C.E., Smyth P. (1997) Applying classification algorithm in practice. *Statistics and Computing*, **7**, 45-56
- [14] Bryk A.S., Raudenbush S.W. (1992) *Hierarchical Linear Models*. Newbury Park, California: Sage
- [15] Capiluppi C. (2000) Il modulo di scheduling CATI del sistema CAPTOR. In: Fabbris L. (a cura di) *CAPTOR 2000: qualità della didattica e sistemi di computer-assisted*. Cleup, Padova, 189-193
- [16] CART (1998) *Salford Systems*. San Diego, California, USA
- [17] Cavalli A., Galland O. (1995) *Youth in Europe*. London, Pinter
- [18] Cioni E. (1997) Il sistema di parentela. In: Barbagli M. Saraceno C. (a cura di) *Lo stato delle famiglie in Italia*. Il Mulino, Bologna, 151-162
- [19] Ciucci R. (2001) *Adolescenza lunga: il caso della famiglia*. Regione Toscana, Livorno
- [20] Clayton D.G. (1974) Some Odds Ratio Statistics for the Analysis of Ordered Categorical Data. *Biometrika*, **61**, 525-531
- [21] Clark L.A., Pergibon D. (1992) Tree-based models. In: J.M. Chambers, T.J. Hastie (eds) *Statistical models in S*. Wadsworth and Brooks, Pacific Grove, California, USA, 377-420
- [22] Courgeau D., Baccaini B.(1998), *Multilvel analysis in the social sciences*, Population 10

- [23] Cox D.R. (1970) *The Analysis of Binary Data*. Chapman and Hall, London
- [24] Cremaschi M. (1996) Casa e relazioni sociali. In *Indagine sociale lomabrda*. Milano, Irer
- [25] Dempster A.P., Laird N.M., Rubin D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, Methodological*, **39**, 1-22
- [26] Donati P., Di Nicola P. (1996) *Lineamenti di Sociologia della famiglia. Un approccio relazionale all'indagine sociologica*. NIS, Roma
- [27] Diggle et al. (1994) *Analysis of longitudinal data*. Oxford University Press (Oxford)
- [28] Endrizzi I., Fabbris L., Martini M.C. (2002) Il turnover nello stato di disagio delle famiglie di Padova. In: L. Fabbris (a cura di) *LAID-OUT: scoprire i rischi con l'analisi di segmentazione*. Cleup, Padova, 191-207
- [29] Endrizzi I., Fabbris L., Schievano C. (2002) Gruppi di famiglie padovane a rischio di disagio mediante analisi di segmentazione. In: L. Fabbris (a cura di) *LAID-OUT: scoprire i rischi con l'analisi di segmentazione*. Cleup, Padova, 209-232
- [30] Esping-Andersen G. (1990) *The Three World of Welfare Capitalism*. Oxford, Polity
- [31] Esping-Andersen G. (1995) *Il welfare state senza il lavoro. l'ascesa del familismo nelle politiche sociali dell'Europa continentale*. In: "Stato e mercato", XLV, 3, 347-380
- [32] Fabbris L. (1997) *Statistica multivariata, analisi esplorativa dei dati*. McGraw-Hill, Milano
- [33] Fabbris L. (1989) *L'indagine campionaria: metodi, disegni e tecniche di campionamento*. NIS, Roma

- [34] Facchini C. (1997) Gli anziani e la solidarietà tra generazioni. In: Barbagli M. Saraceno C. (a cura di) *Lo stato delle famiglie in Italia*. Il Mulino, Bologna, 151-162
- [35] Finch J., Mason J. (1993) *Negotiating family, responsibilities*. Routledge e Kegan Paul, London
- [36] Fisher W.D. (1958) On grouping for maximum homogeneity. *Jasa*, **53**, 789-798
- [37] Fleiss J.L. (1981) *Statistical Methods for Rates and Proportions*. John Wiley and Sons, New York
- [38] Freeman L. C. (1965) *Elementary Applied Statistics*. Wiley, New York
- [39] Gart J.J. (1962) On the combination of relative risks. *Biometrics*, **18**, 601-610
- [40] Gart J.J. (1970) Point and interval estimation of the common odds ratio in the combination of tables with fixed marginals. *Biometrika* **57**, 471-475
- [41] Gini C. (1912) Variabilità e mutabilità, reprinted in *Memorie di metodologica statistica* (1955). Pizzetti E., Salverini T., Roma
- [42] Goldstein H. (1986) Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, **73**, 43-56
- [43] Goldstein H. (1995) *Multilevel Statistical Models*. Kendall's Library of Statistics, London
- [44] Goodman L., Kruskal W. (1980) Measures of Association for Cross Classification. *Journal of American Statistical Association*, **49**, 265-269
- [45] Grossberg S. *Studies of the Mind and Brain*. Dordrecht, Holland: Reidel Press
- [46] Hartley H.O., Rao J.N.K. (1967) Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, **54**, 93-108

- [47] Harville D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320-338
- [48] Hastie T.J., Tibshirani R.J. (1990) *Generalized Additive Models*. Chapman & Hall/CRC
- [49] Hebb D. (1949) *The Organization of Behavior*. Wiley, New York
- [50] Hepplen S., Brown A. (1989) *How Parents Cope: Caring for a Child who has a Handicap*. Ilford: Barnardoso
- [51] Henderson C.R. (1984) ANOVA, MIVQUE, REML, and ML algorithms for estimation of variances and covariances. *Statistics: An Appraisal*, 257-280
- [52] Hopfield J. (1982) Neural Network and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, **79**, 2554-2558
- [53] Hosmer D.W., Lemeshow S. (2000) *Applied Logistic Regression*. Wiley, New York
- [54] ISTAT (1998) *Indagine multiscopo sulle famiglie*, Istat, Roma
- [55] ISTAT (2001) *Rapporto annuale. La situazione del Paese nel 2001*, Istat, Roma
- [56] ISTAT (2002) *Indagine multiscopo sulle famiglie*, Istat, Roma
- [57] ISTAT (2003) *Indagine multiscopo sulle famiglie*, Istat, Roma
- [58] ISTAT (2005) *Indagine multiscopo sulle famiglie*, Istat, Roma
- [59] James W., Stein C. (1961) Estimation with Quadratic Loss. In Neyman J. (Ed.). *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. **1**, 361-379. Berkeley: University of California Press

- [60] Jennrich R.I., Schluchter M.D. (1986) Unbalanced Repeated-Measures Models with Structured Covariance Matrices. *Biometrics*, **42**, 805-820
- [61] Kacker R.N., Harville D.A. (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, **79**, 853-862
- [62] Kass R.E., Steffey D. (1989) Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, **84**, 717-726
- [63] Laird N.M., Lange N., Stram D. (1987) Maximum likelihood computations with repeated measures: Application of the EM algorithm. *Journal of the American Statistical Association*, **82**, 97-105
- [64] Laird N.M., Ware J.H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974
- [65] Liang K.Y., Zeger S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, **73**, 13-22
- [66] Lindley D.V., Smith A.F.M. (1972) Bayes Estimates for the Linear Model. *Journal of the Royal Statistical Society*, **B 34**, 1-41
- [67] Lindsey J.K., Lambert P. (1998) On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, **17**, 447-469
- [68] Lindstrom M.J., Bates D.M. (1988) Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, **83**, 1014-1022
- [69] Litwak E. (1960) Occupational mobility and extended family cohesion. *American Sociological review*, **XXV**, **1**, 9-21

- [70] Longford N.T. (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, **74**, 817-827
- [71] Maggioni G. (1997) Le separazioni e i divorzi. In: Barbagli M. Saraceno C. (a cura di) *Lo stato delle famiglie in Italia*. Il Mulino, Bologna, 151-162
- [72] Mantel N., Haenszel W. (1959) Statistical Aspects of the Analysis of Data from Retrospective Studies. *Journal of National Cancer Institute*, **22**, 719-748
- [73] Mason W.M., Wong G.Y., Entwisle B. (1983) Contextual Analysis Through the Multilevel Linear Model. *Sociological Methodology 1983-84*, Jossey-Bass, San Francisco, 72-103
- [74] Mastrovita S., Siciliani I. (2002) Un'analisi del rischio di permanenza in situazioni di disagio economico mediante segmentazione: un'applicazione a dati ECHP 1994-1998. In: Fabbris L. (a cura di) *LAID-OUT: scoprire i rischi con l'analisi di segmentazione*. Cleup, Padova, 233-255
- [75] McCubbin H.I., Dahl B., Lester G., Ross B. (1975) The returned prisoner of war: factors in family reintegration. *Journal of Marriage and The Family*, **38**, 461-471
- [76] McCubbin H.I. et al. (1976) Coping repertoires of wives adapting to prolonged war-induced separations. *Journal of Marriage and The Family*, **39**, 263-278
- [77] McCullagh P. (1980) Regression models for ordinal data. *Journal of Royal Statistical Society*, **42**, 109-142
- [78] McCullagh P., Nelder J.A. (1989) *Generalized Linear Models*. Chapman & Hall, London
- [79] McCulloch W., Pitts W. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, **5**, 115-133

- [80] McLean R.A., Sanders W.R. (1988) Approximating degrees of freedom for standard errors in mixed linear models. *ASA Proceedings of the Statistical Computing section*, 50-59
- [81] Methodological SRL (2001) NPC Test Version 2.0. Treviso, Italy
- [82] Minsky M.L. (1954) *Theory of Neural Analog Reinforcement Systems and its Application to the Brain Model Problem*. PhD thesis, Princeton
- [83] Minsky M.L., Papert S. (1969) *Perceptron*. Cambridge, Mass: MIT Press
- [84] Muton A. (1990) Job relocation, stress and the family. *Journal of Organizational Behavior*, **2**, 5-24
- [85] Patterson H.D., Thompson R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545-554
- [86] Pearson K. (1900) On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Magazine*, **50**, 157-175
- [87] Pesarin F. (2001) *Multivariate Permutation Tests with Applications in Biostatistics*. Wiley, Chichester
- [88] Plackett R.L. (1965) A class of bivariate distributions. *Journal of American Statistical Association*, **60**, 516-522
- [89] Prasad N.G.N., Rao J.N.K. (1990) The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, **85**, 163-171
- [90] Prentice R.L. (1988) Correlated Binary Regression With Covariates Specific to Each Binary Observation. *Biometrics*, **44**, 1033-1048
- [91] Raudenbush S.W., Bryk A.S. (1986) A Hierarchical Model for Studying School Effects. *Sociology of Education*, **59**, 1-17

- [92] Ripley B.D. (1996) *Pattern Recognition and neural network*. Cambridge University Press, Cambridge
- [93] Rosenberg B. (1973) Linear Regression with Randomly Dispersed Parameters. *Biometrika*, **60**, 61-75
- [94] Rosenblatt F. (1958) The perceptron: A probabilistic Model for information storage and organization in the brain. *Psychological Review*, **65**, 386-408
- [95] Rubin D.B. (1976) Inference and missing data. *Biometrika*, **63**, 581-590
- [96] Sabbadini L.L., palomba R. (1994) *Tempi diversi. L'uso del tempo di uomini e donne nell'Italia di oggi*. Roma, Istat/Commissione nazionale per la Parità e le Pari Opportunità tra Uomo e Donna, Presidenza del Consiglio dei Ministri
- [97] Santini A. (1997) La fecondità. In: Barbagli M. Saraceno C. (a cura di) *Lo stato delle famiglie in Italia*. Il Mulino, Bologna, 113-121
- [98] Saraceno C. (1997) Le politiche per la famiglia. In: Barbagli M. Saraceno C. (a cura di) *Lo stato delle famiglie in Italia*. Il Mulino, Bologna, 151-162
- [99] Saraceno C. (2003) *Mutamenti della famiglia e politiche sociali in Italia*. Il Mulino, Bologna
- [100] SAS Institute Inc. Cary, NC, U.S.A., *SAS System*. Version 9.0.
- [101] Schievano C. (2002) LAID-OUT.1: un programma per l'analisi di segmentazione binaria con riferimento ad una variabile dicotomica trasformata in logit. In Puggioni G. (a cura di) *Modelli e metodi per l'analisi di rischi sociali e sanitari*. Cleup, Padova, 21-36
- [102] Searle S.R. (1971) A Biometrics Invited Paper. Topics in variance component estimation *Biometrics*, **27**, 1-76
- [103] Searle S., Casella G., McCulloch C. (1992). *Variance Components*. Wiley, New York

- [104] Selvin S. (1991) *Statistical Analysis of Epidemiologic Data*. Oxford University, New York
- [105] Serfling R.J. (1980) *Approximation theorems of mathematical statistics*. John Wiley and Sons, New York, Chichester
- [106] Shannon C.E., Weaver W. (1949) *The mathematical theory of communication*. Urbana, University of Illinois Press
- [107] Skinner C.J., Holt D., Smith T.M.F. (1989) *Analysis of complex surveys*. Chichester: Wiley
- [108] Skrondal A., Rabe-Hesketh S., Pickles A. (2004) Generalized multilevel structural equation modelling *Psychometrika*, **69** (2), 167-190
- [109] Stella S.P. (1997) I giovani in famiglia. In: Barbagli M. Saraceno C. (a cura di) *Lo stato delle famiglie in Italia*. Il Mulino, Bologna, 151-162
- [110] Therneau T.M., Atkinson E.J. (1997) *An introduction to recursive partitioning using the RPART routines*. Technical Report Mayo Foundation
- [111] Valtolina G. (2002) *Famiglia e disabilità*. Franco Angeli, Milano
- [112] Venables W.N., Ripley B.D. (1997) *Modern applied statistics with S-PLUS*. Springer, New York
- [113] Wedderburn R.W.M. (1974) Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method. *Biometrika*, **61**, 439-447
- [114] Widrow B., Hoff M.E. (1960) Adaptive switching circuits. In: *IRE Convention Record. New York, Part 4*, 96-104
- [115] Widrow B., Stern D. (1985) *Adaptive Signal Processing*. Prentice-Hall, New York
- [116] White H.A. (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, **48**, 817-838

- [117] Zanatta A.L. (1997) Le famiglie con un solo genitore. In: Barbagli M. Saraceno C. (a cura di) *Lo stato delle famiglie in Italia*. Il Mulino, Bologna, 51-162