

UNIVERSITÀ DEGLI STUDI DI PADOVA
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in Scienze Statistiche



TESI DI LAUREA

**Internet in Europa:
profilo e segmentazione dei non possessori**

Relatore: Prof. Omar Paccagnella
Dipartimento di Scienze Statistiche

Laureanda: Stefania Maschio
Matricola: 1035583

Anno Accademico 2013/2014

Indice

Introduzione

1	Internet	1
1.1	La nascita di Internet	5
1.2	La diffusione di Internet	7
1.3	Internet e la società	9
1.4	Internet e l'Europa	11
1.5	Obiettivi della tesi	13
2	I dati	15
2.1	Le indagini Eurobarometro	15
2.1.1	Eurobarometro Speciale 335 ed Eurobarometro Speciale 365	16
2.1.2	I questionari	18
2.2	I campioni	19
2.2.1	Caratteristiche dei campioni	20
2.2.2	Caratteristiche dei Paesi	26
2.2.3	Caratteristiche dei sottocampioni	30
2.3	Motivazioni del non possesso di Internet	32
3	I modelli	35
3.1	Modelli per variabili binarie	35
3.2	Modelli multilivello	37
3.2.1	Coefficiente di Correlazione Intraclasse	39
3.2.2	Notazione	40

3.2.3	Modello a intercetta casuale (random intercept model)	40
3.2.4	Modello a pendenza casuale (random slope model)	41
3.2.5	Stima del modello	42
3.3	Modelli a classi latenti	43
3.3.1	Stima	45
3.3.2	Misure di bontà di adattamento	46
3.3.3	Significatività degli effetti	47
3.3.4	Classificazione in classi	47
3.3.5	Estensioni	48
3.4	L'analisi multilivello a classi latenti	48
3.4.1	Modelli multilivello a classi latenti con effetti fissi	49
3.4.2	Modelli multilivello a classi latenti ad effetti casuali	50
3.4.3	Scelta del numero di classi	51
4	Profilo tipico dei non possessori	55
4.1	Stima dei modelli probit	55
4.2	Stima dei modelli probit multilivello	61
4.2.1	Inserimento di variabili di secondo livello	64
4.2.2	Analisi dei residui	66
5	Analisi a classi latenti	71
5.1	Software utilizzati	71
5.2	Analisi a classi latenti con <i>Stata</i>	73
5.2.1	Analisi a classi latenti per il 2009 senza covariate	74
5.2.2	Analisi a classi latenti per il 2011 senza covariate	76
5.3	Analisi a classi latenti con <i>Latent Gold</i>	79
5.3.1	Analisi a classi latenti per il 2009 senza covariate	79
5.3.2	Analisi a classi latenti per il 2011 senza covariate	82
5.3.3	Analisi a classi latenti per il 2009 con le covariate	84
5.3.4	Analisi a classi latenti per il 2011 con le covariate	88
5.3.5	Analisi dei residui bivariati	90
5.3.6	Una sintesi dei principali risultati ottenuti	92
5.3.7	Confronto con <i>Stata</i>	96

5.4	Analisi multilivello a classi latenti	99
5.4.1	Analisi multilivello a classi latenti per il 2009 senza covariate	99
5.4.2	Analisi multilivello a classi latenti per il 2011 senza covariate	105
5.4.3	Analisi multilivello a classi latenti per il 2009 con le covariate	109
5.4.4	Analisi multilivello a classi latenti per il 2011 con le covariate	114
5.4.5	Residui bivariati	117
5.4.6	Procedura a tre passi	118
5.4.7	Una sintesi dei principali risultati ottenuti	118
	Conclusioni	127
	Appendice 1	131
	Appendice 2	135
	Bibliografia	137
	Ringraziamenti	141

Introduzione

Oggi il computer ed Internet svolgono un ruolo importante nel mondo lavorativo, scolastico e nella vita sociale. Internet ha trasformato il nostro modo di comunicare, di scambiare informazioni, di organizzare attività politiche, sociali ed economiche. Ha eliminato le barriere geografiche, riducendo le distanze tra i consumatori e tra quest'ultimi e i produttori. Internet è diventato il driver dell'innovazione e l'infrastruttura principale per la partecipazione delle imprese all'economia mondiale. I suoi effetti positivi non coinvolgono solamente il settore delle comunicazioni elettroniche, ma è considerato uno strumento di prosperità per l'intera economia e per il benessere della società.

Internet permette di migliorare la qualità della vita: è uno strumento che ci consente di comunicare in tempo reale con gli altri, di acquistare beni e servizi online senza spostarci, di lavorare in modo più efficace e di accedere facilmente a molte informazioni. È uno strumento molto utilizzato anche dalle imprese che se ne avvalgono per comunicare con i loro clienti e lo utilizzano come mezzo di marketing per far conoscere i propri prodotti ai potenziali clienti. La conoscenza del web, strumento di comunicazione, di intrattenimento e di business, è ormai indispensabile per restare al passo con i tempi.

Le origini della Rete risalgono alla fine degli anni Cinquanta negli Stati Uniti, ma solo a partire dagli anni Ottanta ha iniziato a diffondersi in altri continenti. Dagli anni Novanta, il numero di persone collegate ad Internet è aumentato costantemente, con una notevole espansione negli anni Duemila. Negli ultimi anni il fenomeno di Internet ha iniziato a

coinvolgere anche i Paesi in via di sviluppo, ma la distanza con i Paesi sviluppati è ancora ampia.

La notevole diffusione della Rete è dovuta a diversi fattori. In primis, lo sviluppo della rete globale è collegato alla diffusione dei computer, diffusione resa possibile dai prezzi sempre più agevoli e da una potenza maggiore dei pc. In secondo luogo, il perfezionamento delle infrastrutture di telecomunicazioni ha reso accessibile a molte famiglie la connessione ad Internet, in particolare grazie allo sviluppo della banda larga. La Commissione Europea utilizza questo termine come sinonimo della rete Internet più veloce di quella attraverso un normale modem analogico. La diffusione della banda larga è vista come un fattore di crescita economica di un Paese in quanto permette di ridurre il “*digital divide*”, ossia il divario tra chi può accedere ad Internet e chi ne è escluso. Questa tecnologia si sta evolvendo ulteriormente e in questi anni si stanno sviluppando la banda ultralarga e le modalità di accesso di nuova generazione (*Next Generation Access-NGA*) per far fronte al traffico dati sempre maggiore.

Un altro fattore che influenza la diffusione di Internet è il web mobile, ossia la possibilità di connettersi alla rete tramite dispositivi mobili. Questo fenomeno è caratterizzato da un’ampia crescita negli ultimi sei anni, grazie alla diffusione di tablet e smartphone, i quali hanno rivoluzionato la nostra vita. Grazie a questi dispositivi di nuova generazione, abbiamo il web a portata di mano e possiamo connetterci alla rete in qualsiasi posto ci troviamo.

Per quanto concerne l’Europa, la maggior parte delle famiglie possiede un accesso domestico ad Internet, in quanto le infrastrutture e tecnologie necessarie sono ormai disponibili ovunque. La Commissione Europea si è posta l’obiettivo di garantire a tutti i cittadini europei la connessione alla banda larga, obiettivo raggiunto nell’ottobre del 2013 grazie alla copertura supplementare fornita dal satellitare. Tuttavia, esiste ancora una percentuale di persone che non possiede una connessione alla Rete a casa. I motivi di tale mancanza sono vari: la poca conoscenza di Internet, i costi troppo elevati, la mancanza di interesse, la poca fiducia nella sicurezza in rete, etc.

A mio parere, il fatto di non poter accedere ad Internet a casa è un grande ostacolo, in quanto limita la possibilità di entrare in contatto con tutte le persone, le imprese, i media e le istituzioni che lo utilizzano come strumento di comunicazione e propaganda. Da

questa riflessione prende spunto la mia tesi, nella quale ho voluto analizzare le caratteristiche delle persone che non possiedono un accesso domestico alla Rete, utilizzando i dati ottenuti dalle indagini Eurobarometro sulle comunicazioni digitali del 2009 e del 2011. Ho utilizzato tali dati per individuare il profilo delle persone che non possiedono l'accesso ad Internet a casa e segmentare tale popolazione sulla base della relativa motivazione. In sostanza, la mia tesi vuole individuare le principali caratteristiche delle persone, residenti nell'Unione Europea, che non possiedono la connessione Internet a casa e il motivo di tale scelta o impossibilità.

Il Capitolo 1 presenta una panoramica della storia e della diffusione di Internet. In particolare, si sofferma sulla situazione europea attuale e sugli obiettivi stabiliti dalla Commissione Europea nell'ambito delle comunicazioni elettroniche. Questo capitolo inoltre descrive in modo più dettagliato gli obiettivi della mia tesi.

Il Capitolo 2 introduce le indagini Eurobarometro, soffermandosi sulle due utilizzate nella mia analisi. Presenta inoltre i dataset impiegati e le analisi esplorative necessarie per avere una buona conoscenza della popolazione di riferimento.

Il Capitolo 3 è dedicato alla presentazione delle principali caratteristiche dei modelli stimati con i dati a disposizione, ossia i modelli probit, multilivello e quelli a classi latenti.

Il Capitolo 4 presenta i risultati ottenuti dalla stima dei modelli probit standard e multilivello, utili a tracciare il profilo dei non possessori di Internet a casa nei due istanti temporali considerati.

Il Capitolo 5 introduce i due software utilizzati nell'analisi a classi latenti, *Stata* e *Latent Gold*, e mostra i risultati ottenuti dalla stima dei modelli senza e con covariate. Un paragrafo è inoltre dedicato ai risultati ottenuti con l'approccio multilivello a classi latenti.

Le conclusioni ripercorrono e confrontano i principali risultati ottenuti dalle diverse analisi.

Capitolo 1

Internet

Internet è la più ampia rete di computer oggi esistente e per questo motivo è definita “la rete delle reti”, “interete”, “rete globale” o “la Rete”.

La sua storia è collegata alla storia contemporanea e la sua nascita risale alla fine degli anni Cinquanta, durante la Guerra Fredda. Negli anni Ottanta le tecnologie alla base di Internet si diffusero in tutto il mondo e nel decennio successivo la sua conoscenza aumentò con la nascita del *World Wide Web*. La rete di Internet si diffuse inizialmente ai Paesi Occidentali ed in seguito a quelli in via di sviluppo, fino a creare la rete globale di computer odierna.

All’inizio la maggior parte degli avvenimenti era nota solo agli addetti ai lavori e per questo motivo spesso si ritiene che la storia di Internet abbia avuto inizio solo negli anni Novanta.

1.1 La nascita di Internet

Nel 1957 l’Unione Sovietica realizzò la messa in orbita dello Sputnik. In risposta a questo evento, il Dipartimento della Difesa degli Stati Uniti istituì l’ARPA (*Advanced Research Projects Agency*), con sede a Washington. Il compito di questo ente era di coordinare e finanziare le attività di ricerca a livello nazionale.

Negli anni successivi, i progressi sovietici in campo spaziale continuarono. Gli Stati Uniti decisero perciò di finanziare maggiormente la ricerca aerea spaziale e costituirono la NASA (*National Aeronautics and Space Administration*), a cui fu trasferito il compito di gestire i programmi spaziali, mentre l'ARPA si concentrò sulla scienza dell'informazione e della comunicazione. Fu così concepita l'idea di sviluppare una rete di computer in grado di comunicare tra loro e di trasferire dati attraverso la normale rete telefonica.

Nel 1969 nacque l'ARPANET, il primo risultato del progetto, che permise la condivisione dei sistemi tra i vari poli universitari. I quattro utenti iniziali furono: l'Università di Los Angeles, l'Università di Santa Barbara, l'Università dello Utah e lo Stanford Research Institute. Ad ogni utente fu consegnato un computer per la trasmissione del traffico dati chiamato IMP (*Interface Message Processor*).

Nel 1971 gli utenti di ARPANET erano qualche centinaio. Il progetto fu caratterizzato da continui sviluppi; pertanto nel 1972 fu indetto il primo congresso internazionale sulle comunicazioni tra computer, durante il quale fu presentato al mondo il progetto. Grazie a questa presentazione, la quantità di traffico in ARPANET aumentò significativamente: l'esperimento iniziale stava diventando realtà.

Tra gli anni Settanta e Ottanta ARPANET raggiunse l'Europa: Norvegia e Inghilterra furono i primi due Paesi europei a connettersi in rete. Le reti di computer si stavano diffondendo sempre più grazie ai costi meno elevati. Le reti create avevano scopi, servizi e protocolli differenti ed erano di diversa dimensione: c'erano quelle ad estensione continentale o intercontinentale (*Wide Area Network* o WAN) e quelle con estensione più limitata (*Local Area Network* o LAN).

All'inizio del 1984 i computer collegati alla rete erano mille. L'infrastruttura di Internet era ormai pronta a essere sfruttata: mancavano un'interfaccia in grado di rendere Internet più semplice agli utenti e una maggiore velocità di trasmissione dei dati.

Nel 1986 anche l'Italia entrò in rete, connettendosi dall'Università di Pisa grazie ai finanziamenti del Dipartimento della Difesa degli Stati Uniti.

Nel 1989 tutti i siti furono trasferiti alla rete NSFnet determinando la fine di ARPANET. Nello stesso anno, la caduta del muro di Berlino determinò la fine della Guerra Fredda

tra URSS e USA. Alla fine degli anni Ottanta gli utenti collegati ad Internet erano dieci mila.

All'inizio dell'ultimo decennio del Novecento furono modificate le politiche di accesso alla rete NSFnet per consentire l'ingresso nella rete anche per fini commerciali. Il numero di utenti continuava ad aumentare notevolmente.

Nel 1991 l'Europa iniziò a collaborare alla struttura di Internet. Nacque così il *World Wide Web* ad opera di Tim Berners Lee del Cern di Ginevra. Grazie a questo sistema fu possibile consultare in modo semplice ed intuitivo le informazioni, i dati, le immagini in Internet. Nacquero anche i primi browser, tra i quali Internet Explorer della Microsoft (1995). Tutto ciò determinò un rapido sviluppo commerciale della Rete con una notevole crescita del numero dei suoi utenti: basti pensare che in tre anni si è passati dai 10 milioni del 1996 ai 200 milioni del 1999.

Il numero di utenti collegati ad Internet ha continuato ad aumentare anche negli anni Duemila, raggiungendo quota 2 miliardi nel 2011.

1.2 La diffusione di Internet

Secondo i dati riportati da "*The Internet Encyclopedia*", il numero di Paesi collegati alla rete è più che triplicato nel corso degli anni Novanta, passando dai 60 del 1993 ai 214 del 2000. Nonostante ciò, Internet ha una distribuzione globale molto asimmetrica.

Nel 2000 i Paesi in via di sviluppo comprendevano l'84% della popolazione mondiale, ma solo il 6% degli utenti globali risiedeva in questi Stati. Il 61,7% degli utenti di Internet viveva nel Nord America e in Europa, rappresentando solo il 20% della popolazione mondiale. Nel maggio del 2002, meno dello 0,1% dei cittadini Africani aveva accesso alla Rete. Tuttavia, iniziavano a registrarsi i primi segnali incoraggianti: nel 2001 gli utenti aumentarono del 30% ed un terzo dei nuovi utenti proveniva dai Paesi in via di sviluppo.

La diffusione di Internet è stata caratterizzata anche dalla comparsa di nuove tecnologie che permettono l'accesso ad Internet, come la banda larga e l'Internet mobile; già nel 2003 si prevedeva che questi mezzi si sarebbero diffusi rapidamente nella seconda metà

del decennio, come dimostrato in seguito dai dati raccolti dall'ITU (*International Telecommunication Union*).

Più precisamente, dal 2005 al 2011 si è registrata una diminuzione delle sottoscrizioni alla telefonia fissa nei Paesi sviluppati e in quelli in via di sviluppo¹ a favore di un aumento delle sottoscrizioni alla telefonia mobile. Nel 2005 il 19,1% della popolazione mondiale possedeva un telefono fisso, a differenza del 17,3% del 2011, percentuale prevista in ulteriore riduzione negli anni successivi. Per quanto concerne il telefono cellulare, la percentuale di sottoscrittori a livello mondiale è passata dal 33,9% del 2005 all'85,5% del 2011 (Figura 1.1).

Nella seconda metà del decennio, si può osservare un graduale aumento degli abbonamenti alla banda larga mobile² e alla banda larga fissa. Nel primo caso i dati sono stati registrati a partire dal 2007, anno in cui la percentuale di persone in tutto il mondo che ha sottoscritto un abbonamento alla banda larga mobile era del 4%. Questa percentuale è poi cresciuta fino a quadruplicarsi nel 2011 raggiungendo il 16,6%. Da notare che, nel 2006, solo lo 0,8% della popolazione dei Paesi in via di sviluppo possedeva un abbonamento alla banda larga mobile, percentuale che è diventata del 20% nel 2011. Dati diversi si sono registrati per gli abbonamenti alla banda larga fissa, la cui percentuale di utilizzatori è aumentata dal 3,4% del 2005 al 9,1% del 2011; probabilmente, sono stati fatti pochi investimenti nella costruzione di tale infrastruttura.

Un altro fattore collegato alla diffusione di Internet è la crescita della percentuale di famiglie che possiedono un computer. Tale percentuale ha subito un aumento negli ultimi anni sia nei Paesi sviluppati (passando dal 55,5% del 2005 al 73,2% del 2011), sia in quelli in via di sviluppo (14,6% nel 2005 e 27,6% nel 2011).

Questi fattori hanno determinato una crescita nella percentuale di famiglie che possiedono un accesso ad Internet, percentuale che è aumentata dal 18,4% del 2005 al 33,6% del 2011. Tale percentuale è quasi raddoppiata nei Paesi sviluppati raggiungendo il 70,2% nel 2011 e più che raddoppiata in quelli in via di sviluppo, raggiungendo il 20,2%.

¹ Per la classificazione dei Paesi sviluppati e in via di sviluppo vedere:
<http://www.itu.int/ITU-D/ict/definitions/regions/index.html>

² La banda larga mobile comprende qualsiasi tipo di connessione Internet ad alta velocità fornita da dispositivi esterni, come chiavette USB 3G, telefoni cellulari con incorporato un dispositivo per trasferimento dati tipo HSPA/UMTS/GPRS.

Di conseguenza, la percentuale di persone che utilizzano Internet è raddoppiata dal 15,8% del 2005 al 32,7% del 2011. Nei Paesi sviluppati si è registrato un aumento dal 50,9% del 2005 al 70,5% del 2001, mentre nei Paesi in via di sviluppo si è passati dal 7,8% al 24,5%.

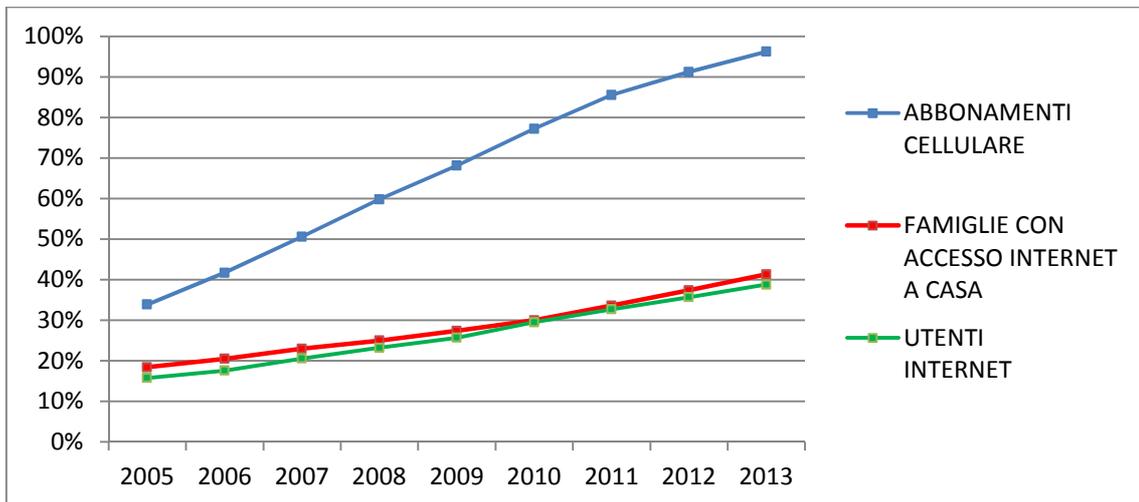


Figura 1.1: Serie storiche della percentuale di abbonamenti cellulare, della percentuale di famiglie con accesso Internet a casa e della percentuale di utenti Internet

Nella prima metà del 2012 gli utenti della rete erano così costituiti: il 44,8% risiedeva in Asia, il 21,5% in Europa, l'11,4% nel Nord America, il 10,4% in America Latina, il 7% in Africa, il 3,7% nel Medio Oriente e l'1% in Oceania³. La percentuale maggiore di utenti si trova ora nel continente asiatico, a differenza dei decenni precedenti. Questo è dovuto principalmente al fatto che l'Asia ospita il 60% della popolazione mondiale.

1.3 Internet e la società

Oggi Internet è uno dei pilastri della nostra società. All'inizio era stato progettato come strumento per mettere in collegamento computer e scambiare dati. In seguito, è diventato uno strumento di comunicazione scritta e di accesso a una quantità elevata di informazioni per poi trasformarsi in un mezzo di comunicazione voce e video e un veicolo po-

³ Fonte dati: "Internet World Stats: Usage and Population Statistics", <http://www.internetworldstats.com/stats.htm>

tente per la pubblicità e la vendita. Internet ha pertanto influenzato molti settori della società come la comunicazione, l'economia e la politica.

Per quanto riguarda la comunicazione, la Rete rende le comunicazioni più facili, veloci e diversificate. È uno strumento che permette di esprimersi e rivolgersi a tutte le persone collegate, diventando quindi un nuovo mezzo di comunicazione: una persona, con il suo computer, si collega con tutto il mondo grazie alla Rete. Consente di accedere a riviste e giornali elettronici, a pagine di cultura e intrattenimento, sostituendo gradualmente i mezzi di comunicazione tradizionali come la posta, il telefono, il fax. Vi è, inoltre, la possibilità di ascoltare la radio in streaming, guardare film, partecipare a dibattiti su forum e conoscere persone nuove.

In ambito economico è impossibile immaginare un mondo senza Internet, anche in quei settori che sembrano essere lontani da questa realtà, come l'agricoltura e l'artigianato. E' diventato un'infrastruttura essenziale per molti Paesi, non solo sviluppati, ma anche in via di sviluppo. Senza la connessione alla Rete l'economia di un Paese non può andare avanti: le banche, i negozi, i fornitori non potrebbero funzionare. Internet è in continua espansione e il suo peso sul PIL è significativo in molti Stati: è un elemento chiave per la crescita. Permette inoltre di creare nuovi posti di lavoro e di modernizzare le attività economiche più tradizionali.

Nelle imprese, Internet permette di modificare e migliorare le relazioni tra l'azienda, i suoi clienti, i suoi fornitori e i suoi partner. Consente di migliorare il funzionamento interno, in particolare le relazioni azienda-impiegati. Nasce così il termine *e-business* che indica l'integrazione all'interno dell'azienda di strumenti basati sulle tecnologie dell'informazione per creare valore per l'impresa stessa, i suoi clienti e i suoi partner (Kotler, 2010). Una faccia dell'e-business è l'*e-commerce* che si occupa dell'utilizzo di un supporto elettronico per la relazione commerciale dell'azienda con dei privati (Belch e Belch, 2003). Internet è utilizzato dalle imprese anche come strumento di propaganda nel marketing, per far conoscere i propri prodotti ed attirare così nuovi clienti. Permette ai consumatori di abbattere le barriere economiche e acquistare un'ampia gamma di prodotti.

Nel mondo politico, Internet permette di coinvolgere i cittadini nella politica del Paese ed informarli in tempo reale. In questo modo, le diverse figure politiche possono fare

conoscere agli elettori le proprie idee e i propri obiettivi e rimanere in contatto con loro, avvalendosi anche di social networks come Facebook e Twitter. In molti Paesi si sta sviluppando la *e-democracy* (Ulrich, 2007), ossia una forma di democrazia diretta in cui vengono utilizzate le moderne tecnologie dell'informazione e della comunicazione per consultare i cittadini e conoscere la loro opinione.

Tuttavia, esiste anche un rovescio della medaglia, dato dai diversi problemi legati ad Internet. In primo luogo, manca una regolamentazione effettiva ufficiale: ciascun utente può condividere quello che vuole, con regole non sempre ben definite. In secondo luogo, la Rete risulta essere vulnerabile e sono presenti molte tipologie di attacco alla sicurezza informatica da parte di pirati informatici con intenzioni malevoli. Sono presenti problemi legati alla privacy, dovuti alle intercettazioni e al filtraggio dei dati e problemi di copyright, in quanto vengono veicolati file protetti da diritti d'autore. Altro problema è quello legato alla tutela dei minori, che possono accedere facilmente alla Rete e a contenuti poco adatti.

1.4 Internet e l'Europa

La diffusione delle tecnologie digitali, in particolare di Internet, è un tema molto importante per la Commissione Europea, che cerca di rendere accessibili tali tecnologie alla maggior parte della popolazione. A questo scopo nel 2010 è stata lanciata l'Agenda Digitale per l'Europa (DAE)⁴, una delle iniziative dell'Europa 2020, che fissa gli obiettivi per una crescita sostenibile dell'Unione Europea da raggiungere entro il 2020. Tale Agenda Digitale permette di sfruttare il potenziale delle tecnologie dell'informazione e della comunicazione per favorire l'innovazione, la crescita economica, il progresso e creare un mercato digitale unico per l'Europa.

L'Agenda Digitale è minata da diversi ostacoli, come la frammentazione dei mercati digitali, l'aumento della criminalità informatica, la mancanza di investimenti e di competenze informatiche. La Commissione Europea si è perciò posta diverse azioni da compiere per far fronte a questi problemi.

⁴ <http://ec.europa.eu/digital-agenda/>

Innanzitutto, la Commissione deve favorire l'accesso ai contenuti online legali, semplificando le procedure di gestione dei diritti d'autore e completare l'area di pagamento unica in euro, in modo da agevolare i pagamenti elettronici. Per far fronte alla poca fiducia nella sicurezza dei pagamenti da parte degli utenti, la Commissione deve rivedere le norme dell'Unione Europea in materia di protezione dei dati e creare un marchio di fiducia UE online come tutela per i consumatori.

Un'altra azione importante è la disposizione di un accesso ad internet veloce e superveloce per tutti e a prezzi competitivi. L'UE deve creare quindi delle reti di accesso di nuova generazione, denominate NGA (*Next Generation Access*), servendosi di fondi europei per effettuare investimenti nella banda larga. Altri investimenti da compiere, sono quelli nelle attività di ricerca e sviluppo nell'ambito delle tecnologie dell'informazione e della comunicazione e quelli necessari al miglioramento delle competenze digitali.

Dal 2010, anno in cui l'Agenda Digitale è stata istituita, ad oggi sono stati raggiunti parecchi miglioramenti in Europa. La banda larga di base (costituita da connessioni xDSL, cavo, FTTP e WiMax) è presente ovunque e le connessioni via satellite sono migliorate e permettono di collegare le aree che non hanno accesso alla banda larga, come le zone rurali. La metà della popolazione è raggiunta dalla banda larga veloce, ma solo il 12% delle aree rurali sono coperte. L'accesso ad Internet sta aumentando sempre più e la connessione tramite telefoni cellulari è passata dal 7% del 2008 al 27% del 2013. Anche le spese per la ricerca hanno subito un lieve aumento.

Il numero di persone che utilizzano Internet è in continua crescita, mentre la percentuale di coloro che non l'hanno mai utilizzato è del 22% nel 2013; il 70% della popolazione lo utilizza regolarmente almeno una volta la settimana. Anche il commercio elettronico è in crescita: il 45% dei cittadini effettua acquisti di beni e servizi online, ma pochi acquistano all'estero.

Tuttavia, la Commissione Europea deve affrontare degli aspetti problematici, in quanto solo il 2% delle famiglie è abbonato alla banda larga ultraveloce, mentre l'obiettivo fissato per il 2020 è del 50%. Inoltre, la metà della popolazione dell'UE non possiede competenze digitali adeguate⁵.

⁵ Fonte dati: "Agenda digitale: valutazione annua dei progressi compiuti nell'UE 2013", http://ec.europa.eu/italia/attualita/primo_piano/informazione/agendadigitale_valutazione2013_it.htm

1.5 Obiettivi della tesi

L'obiettivo della mia tesi è di analizzare la situazione europea a proposito dell'accesso ed utilizzo di Internet. In particolare, vorrei individuare le caratteristiche principali delle persone che non accedono alla Rete a casa e determinare le motivazioni di tale scelta, identificando dei segmenti di individui omogenei al loro interno ed eterogenei tra di loro.

Per fare questo, ho utilizzato i dati delle indagini Eurobarometro sulle comunicazioni digitali del 2009 e 2011, per vedere se e come la situazione è cambiata. Tuttavia, non si tratta di dati longitudinali, in quanto le persone intervistate nei due anni non sono le stesse e, pertanto, le analisi non possono prendere in considerazione gli stessi individui. Tali sondaggi coprono tutti i 27 Paesi membri dell'UE.

La prima parte di analisi che ho eseguito è di tipo esplorativo, per comprendere le caratteristiche socio-economiche degli intervistati, l'utilizzo e l'accesso ad Internet e alla telefonia mobile e fissa. In particolare, mi sono soffermata sull'analisi delle principali caratteristiche dei sottocampioni costituiti solo dai non possessori di Internet nei due istanti temporali.

In seguito, ho utilizzato un modello probit standard per determinare le caratteristiche delle persone che non possiedono Internet a casa nei due diversi istanti temporali e determinarne il profilo tipico. Ho successivamente utilizzato un modello probit multilivello e inserito delle variabili a livello di Paese per analizzare se il fatto di appartenere ad uno Stato, piuttosto che ad un altro, determina delle differenze sulla probabilità di non avere accesso ad Internet a casa e se l'influenza di tali caratteristiche cambia nei due anni.

Infine, ho stimato dei modelli a classi latenti, per suddividere la popolazione di coloro che non possiedono la connessione a casa sulla base della relativa motivazione. Dopo aver identificato il numero di classi per ciascun anno, ho stimato i relativi modelli. Per quest'analisi ho utilizzato in un primo momento il software *Stata*, il quale presenta però delle limitazioni per quanto concerne questo tipo di modelli. Ho così utilizzato il software *Latent Gold*, il quale è specifico per l'analisi a classi latenti. In questo modo ho potuto realizzare anche un confronto tra i due programmi per individuare le analogie e le differenze nei risultati. In seguito, ho utilizzato l'approccio multilivello a classi latenti che permette di suddividere in gruppi i 27 Paesi analizzati.

Capitolo 2

I dati

In questo capitolo spiegherò in cosa consistono le indagini Eurobarometro, soffermandomi in particolare su quelle che ho utilizzato per la mia tesi. In seguito, analizzerò i campioni costituiti dai cittadini europei intervistati nelle due indagini considerate e i sottocampioni formati dalle persone che non possiedono Internet a casa, soffermandomi inoltre su alcune caratteristiche a livello di Paese.

2.1 Le indagini Eurobarometro

Eurobarometro è una famiglia di indagini della Commissione Europea, la cui istituzione risale al 1973. Queste indagini sono molto importanti, in quanto permettono alla Commissione di conoscere tempestivamente l'opinione pubblica negli Stati membri, valutare il proprio operato e prendere decisioni sugli aspetti da migliorare.

Oggi le indagini vengono eseguite ogni anno durante la stagione primaverile ed autunnale ed interessano i 27 Stati membri dell'Unione Europea. Queste ricerche concernono svariati argomenti, tutti di grande interesse per i cittadini europei. Tra i principali segnali: la situazione sociale, la salute, l'ambiente, l'allargamento dell'Unione Europea, l'*information technology* e l'Euro. I dati vengono raccolti tramite sondaggi d'opinione e tramite *focus group*, ossia gruppi di discussione.

Le indagini Eurobarometro sono denominate in modo diverso in base agli argomenti trattati; le principali sono: Eurobarometro Standard, Eurobarometro Speciale, Eurobarometro Flash. Le indagini Eurobarometro Standard analizzano l'opinione dei cittadini europei su argomenti di interesse generale; le relazioni Eurobarometro Speciali si basano su servizi forniti dalla Commissione europea o da altre istituzioni dell'UE; le indagini Eurobarometro Flash sono costituite da questionari rivolti telefonicamente per ottenere risultati in tempi brevi.

2.1.1 Eurobarometro Speciale 335 ed Eurobarometro Speciale 365

Le indagini da cui la mia tesi prende spunto sono l'Eurobarometro Speciale 335 del 2009 e l'Eurobarometro Speciale 365 del 2011. Entrambi affrontano il tema della presenza e dell'utilizzo delle comunicazioni elettroniche nelle famiglie europee. Dal 1998, anno dell'apertura dei mercati delle comunicazioni elettroniche dell'UE, gli operatori e i fornitori di servizi hanno investito in nuove e innovative infrastrutture di reti digitali. Di conseguenza, è cambiato il modo con cui i cittadini europei accedono e utilizzano le reti di comunicazione elettronica.

Gli argomenti trattati nelle due indagini utilizzate in questo lavoro di tesi sono: i differenti tipi di accesso telefonico presenti nella famiglia, la disponibilità di computer, l'accesso ad Internet e la qualità della connessione, la disponibilità di televisori e il modo con cui si ricevono le trasmissioni, la diffusione dei pacchetti di comunicazione (*bundle*), la facilità nel cambiare fornitore di servizi Internet o fornitori di pacchetti, l'accessibilità ai servizi di telefonia mobile e la privacy.

L'indagine del 2009 è stata svolta nei mesi di novembre e dicembre, mentre quella del 2011 è stata svolta nei mesi di febbraio e marzo. I Paesi oggetto dell'analisi sono gli stessi 27 Stati membri dell'UE in entrambe le occasioni. Gli individui intervistati nella prima rilevazione sono stati 26671, mentre quelli della seconda sono stati 26836. Occorre sottolineare che i due campioni sono costituiti da unità statistiche differenti, dunque non è possibile effettuare delle analisi per dati panel.

Il metodo di campionamento utilizzato è multistadio: le unità sono state selezionate con la medesima probabilità di inclusione in ognuno dei 27 Stati. Al primo stadio, le PSU (*Primary Sampling Units*) sono state estratte dalle unità amministrative regionali di ogni Stato, in particolare da campioni stratificati in base al grado di urbanizzazione (in base all'EUROSTAT NUTS II⁶), con probabilità proporzionali alla numerosità e alla densità della popolazione. Allo stadio successivo, da ciascun campione di unità è stato estratto casualmente un elenco di indirizzi. Le interviste sono state in seguito svolte nelle abitazioni degli individui intervistati e nella loro lingua.

I dati pubblicati nel sito di Eurobarometro presentano dei campioni separati per Germania dell'est e Germania dell'Ovest e per Gran Bretagna e Irlanda del Nord. Al fine di interpretare meglio i risultati ho aggregato tali dati in modo da ottenere un campione per la Germania e uno per il Regno Unito. Inoltre, ho eliminato alcune osservazioni che presentavano valori mancanti per l'istruzione e la zona di residenza (il 3,13% nel 2009 e il 5,32% nel 2011) ed ho ottenuto così un campione finale di 25836 osservazioni per il 2009 e 25273 osservazioni per il 2011. La tabella 2.1 riporta le numerosità dei vari campioni nazionali.

	N° osservazioni			N° osservazioni	
	2009	2011		2009	2011
Francia	949	985	Austria	918	907
Belgio	956	1005	Cipro	490	479
Olanda	1004	962	Rep Ceca	1041	926
Germania	1472	1517	Estonia	974	971
Italia	990	952	Ungheria	1015	988
Lussemburgo	481	492	Lettonia	996	1001
Danimarca	991	998	Lituania	1015	954
Irlanda	942	939	Malta	494	481
Regno Unito	1235	1217	Polonia	967	957
Grecia	968	956	Slovacchia	1039	974
Spagna	956	942	Slovenia	989	960
Portogallo	976	904	Bulgaria	994	958
Finlandia	1017	803	Romania	986	1019
Svezia	981	996			

Tabella 2.1: Numerosità dei campioni nazionali

⁶ È una nomenclatura geografica che suddivide il territorio dell'UE in regioni.

2.1.2 I questionari

I questionari utilizzati nelle due indagini sono molto simili (ma non identici) e sono stati rivolti ad un individuo rappresentativo della famiglia che avesse almeno quindici anni di età. Nel seguito presenterò i quesiti considerati nella mia analisi, che hanno la caratteristica di essere stati posti in entrambe le indagini.

La prima parte di domande, uguale nei due questionari, concerne la sfera socio-economica. Alcuni quesiti riguardano l'individuo intervistato come il sesso, lo stato civile, il livello di istruzione, l'accesso e utilizzo di Internet e il possesso di beni di consumo come televisore, auto, computer, lettore dvd e cd. Altre domande riguardano, invece, la famiglia e, in particolare, la zona di residenza, il numero di componenti del nucleo familiare, la presenza di bambini e ragazzi e il livello della società⁷ a cui la famiglia appartiene.

Il secondo gruppo di quesiti riguarda invece il questionario vero e proprio in cui vengono affrontati diversi argomenti legati alle comunicazioni elettroniche. A coloro che possiedono un telefono cellulare viene chiesto l'eventuale sottoscrizione ad un abbonamento cellulare che permette di collegarsi ad Internet. Alle persone che hanno affermato di accedere ad Internet da casa, viene chiesto quale tipo di connessione venga utilizzato: collegamento tramite una linea telefonica standard o linea ISDN, Internet a banda larga tramite collegamento ADSL o XDSL, Internet a banda larga tramite la rete televisiva via cavo, collegamento tramite telefono cellulare, collegamento tramite rete satellitare, collegamento tramite una linea elettrica oppure collegamento tramite una linea a fibra ottica. Infine, viene chiesto se qualcuno della famiglia ha mai pensato di cambiare fornitore di servizi Internet ed eventuali motivazioni.

Alle persone che hanno affermato di non accedere ad Internet da casa viene chiesta l'eventuale intenzione a collegarsi nei sei mesi successivi e le motivazioni per le quali non viene posseduta la connessione a casa. Tali motivazioni includono: la mancanza di conoscenza di Internet, il mancato interesse dei componenti della famiglia, la mancanza di un'infrastruttura per l'accesso alla banda larga, costi troppo elevati, l'accesso ad In-

⁷ In questo ambito si considera il livello economico a cui la famiglia della persona intervistata ritiene di appartenere.

Internet a lavoro o scuola, la preoccupazione per l'accesso a siti non adatti. Sarà su questo ambito che la mia analisi si focalizzerà.

Come già detto in precedenza, le domande non sono esattamente le medesime nelle due indagini Eurobarometro. Tuttavia, a mio parere alcune differenze sono rilevanti.

In primis, nel questionario del 2009 sono presenti delle domande riguardanti il telefono fisso e i telefoni pubblici. Il fatto che questi quesiti siano presenti solo nella prima indagine suggerisce che tali mezzi di comunicazione siano stati superati e sostituiti dalle tecnologie più moderne, come telefoni cellulari ed Internet. Per questo motivo è stato deciso di togliere tali domande nel questionario del 2011.

Altra differenza significativa tra le due indagini è la presenza di un maggior numero di domande riguardanti Internet nel questionario del 2011. La mancanza di tali quesiti nella prima indagine indica che il collegamento Internet tramite telefono cellulare e banda larga si è diffuso in particolare dopo il 2009. Sembra quindi che il telefono fisso e quello pubblico abbiano lasciato posto alla Rete e l'interesse principale della Commissione Europea riguardi ora proprio tale tecnologia che, ormai, è alla base dell'economia, della società, della politica e della vita comune di tutti i giorni.

2.2 I campioni

In questo paragrafo presenterò le principali caratteristiche dei campioni analizzati, facendo dei confronti tra Paesi e utilizzando dei grafici per evidenziare le differenze più importanti. Successivamente, mi soffermerò sui sottocampioni utilizzati nell'analisi a classi latenti, costituiti dalle persone che hanno affermato di non avere l'accesso Internet a casa nei due anni. Le caratteristiche analizzate sono quelle relative al sesso, all'età, all'istruzione, al tipo di comunità in cui gli intervistati vivono, al livello della società a cui appartengono, alla composizione del nucleo familiare e al possesso di alcuni beni di consumo durevole. Come già accennato in precedenza, nell'interpretare i risultati si deve tenere conto del fatto che le persone intervistate sono diverse nei due istanti temporali.

2.2.1 Caratteristiche dei campioni

Considerando il sesso delle persone intervistate, il 54% degli intervistati del 2009 sono donne. Tale percentuale cambia di poco nel 2011, anno in cui le donne sono il 53% del totale degli intervistati. A livello di singolo Stato, si nota che tali proporzioni sono rispettate in tutti i Paesi considerati.

Per quanto riguarda l'età, dalla Figura 2.1 sotto riportata si può osservare che la distribuzione delle osservazioni nelle diverse fasce d'età è pressoché la stessa nei due anni. La fascia d'età che presenta la percentuale maggiore di osservazioni è quella costituita da persone con 65 o più anni. La fascia che, invece, ha il minore numero di intervistati, è quella dei giovani dai 15 ai 24 anni. Le fasce intermedie presentano all'incirca la stessa proporzione di intervistati. In entrambi i dataset l'età media è di 48 anni, quella minima di 15 e la massima è di 98 anni. Dai boxplot (Figura 2.2) si può infine osservare che la mediana coincide con la media.

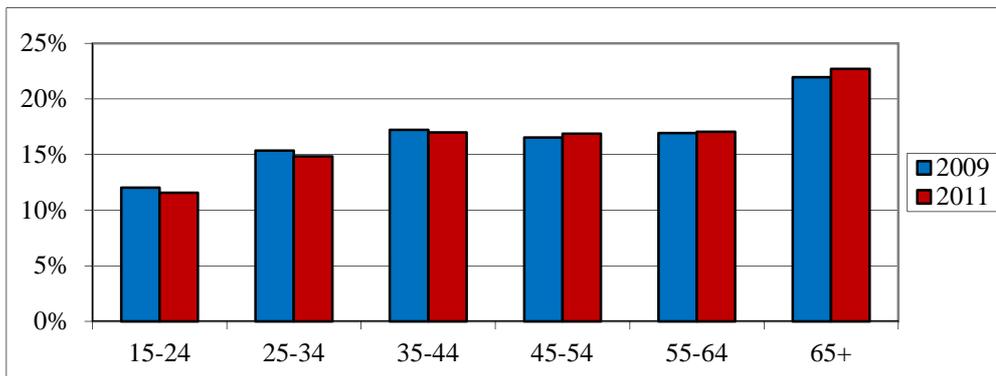


Figura 2.1: Età degli intervistati

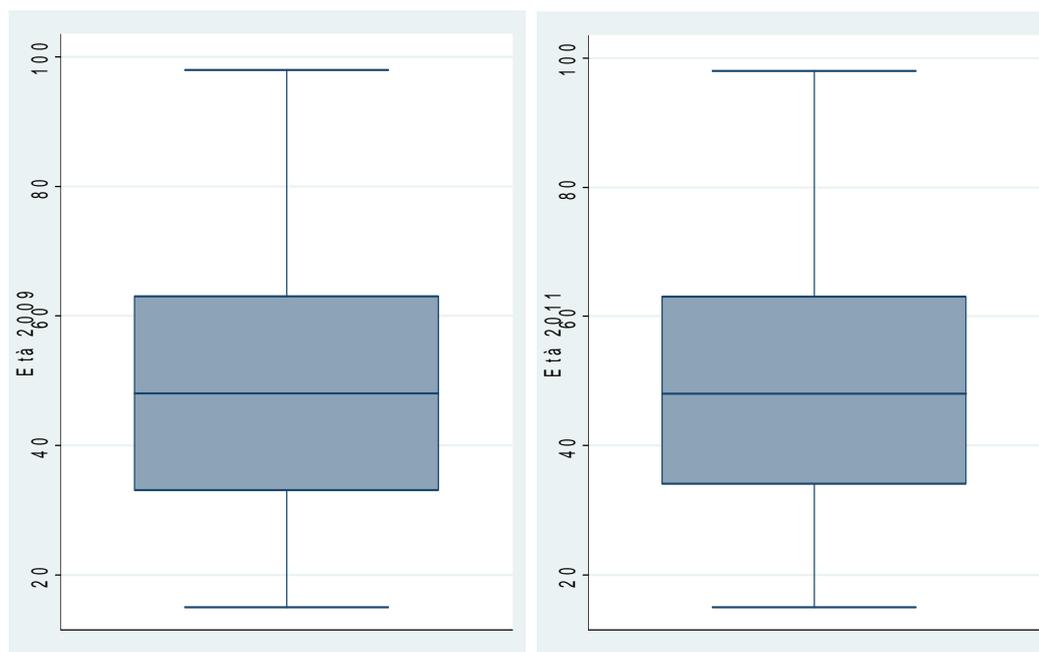


Figura 2.2: Boxplot della variabile Età

Nei questionari di entrambe le rilevazioni non sono presenti domande relative al titolo di studio conseguito. Tuttavia, è stato posto un quesito in cui viene chiesto a quale età l'intervistato ha concluso gli studi. Ho utilizzato tali dati per costruire una variabile che approssimi il livello di istruzione posseduto. In particolare, tale variabile è costituita da quattro categorie: la prima individua coloro che non hanno mai studiato o che hanno terminato gli studi prima dei 7 anni; la seconda identifica le persone che hanno finito gli studi tra i 7 e i 15 anni; la terza individua gli intervistati che hanno concluso gli studi dopo i 15 anni; la quarta identifica coloro che stanno ancora studiando. Inoltre, si deve tenere presente che il sistema scolastico non è organizzato nello stesso modo in tutti gli Stati. Come evidenziato dalla Figura 2.3, tra i Paesi in cui sono presenti persone che non hanno alcun livello di istruzione in entrambe le rilevazioni, segnalo Spagna, Grecia, Portogallo, Bulgaria e Romania. In entrambi i casi, quasi il 70% di queste osservazioni è costituito da persone con un'età pari o maggiore ai 55 anni e quasi il 10% è formato da ragazzi tra i 15 e i 24 anni. I Paesi con una percentuale maggiore di persone che hanno terminato gli studi tra i 7 e i 15 anni sono Portogallo, Grecia, Spagna, Malta. In questi Paesi è presente anche la percentuale minore di persone che hanno concluso gli studi dopo i 15 anni.

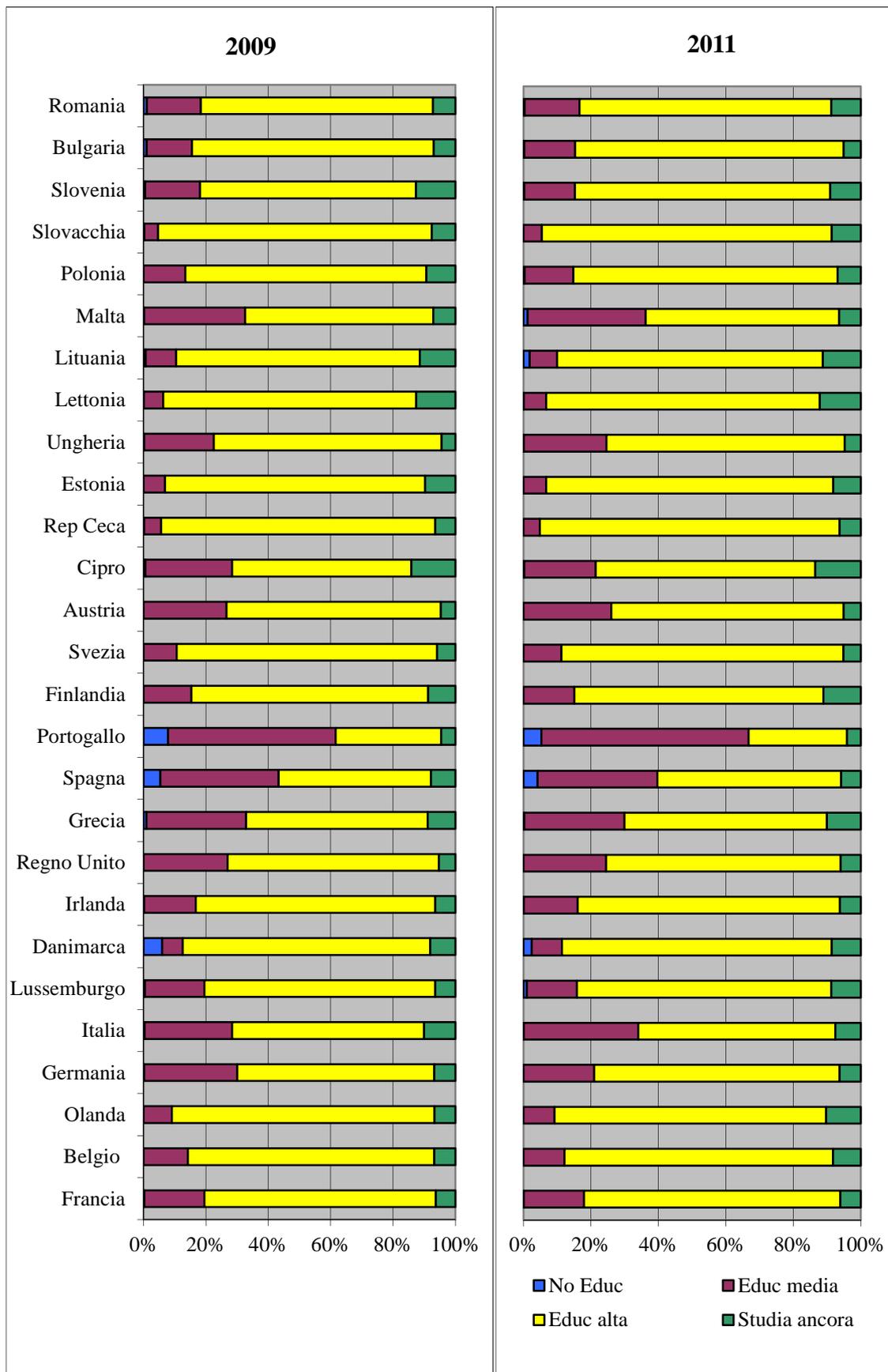


Figura 2.3: Livello di istruzione degli intervistati

Considerando invece la zona di residenza, si nota che gli intervistati sono distribuiti in modo bilanciato tra zone rurali, piccole e medie città e grandi città. Non sono presenti neppure differenze evidenti a livello di Stato.

Per quanto concerne la composizione del nucleo familiare, in entrambi gli anni considerati la percentuale maggiore di intervistati, precisamente il 34%, appartiene ad un nucleo familiare costituito da due persone, mentre il 25% vive in una famiglia con più di 4 componenti. Un particolare evidente è che coloro che vivono in una famiglia numerosa sono soprattutto persone tra i 15 e i 44 anni, mentre gli individui con più di 45 anni tendono a vivere da soli o in coppia. Questo è spiegato dal fatto che le persone con meno di 45 anni vivono in famiglia con figli, mentre le persone con un maggior numero di anni tendenzialmente non hanno più figli che vivono con loro e il nucleo familiare è quindi più ristretto. Inoltre, il 40% delle famiglie numerose risiede in zone rurali.

Sia nel 2009 che nel 2011, il 19% degli intervistati afferma che nella propria famiglia sono presenti bambini con meno di 10 anni e il 12% che ci sono ragazzi tra i 10 e i 14 anni. Gli Stati con un maggior numero di famiglie con bambini sono Irlanda, Lettonia ed Estonia.

La metà del campione è costituita da persone che hanno affermato di appartenere al livello medio della società. Coloro che, invece, hanno risposto di far parte del livello più alto, risiedono in Olanda, Svezia e Finlandia. Al contrario, la maggior parte delle persone che appartengono al livello più basso della società risiedono nei Paesi dell'est Europa, come Bulgaria, Ungheria, Lettonia, Lituania e in Portogallo e sono tendenzialmente persone anziane. Tra il 2009 e il 2011 non sono presenti cambiamenti significativi a differenza di quanto ci si potrebbe aspettare come conseguenza della crisi economica.

Per quanto riguarda i beni di consumo durevoli posseduti, ho preso in considerazione solamente i principali, ossia l'accesso ad Internet a casa, il telefono fisso, il telefono cellulare e l'auto. Ho, inoltre, considerato il possesso di una casa come *proxy* del livello di ricchezza, ipotizzando che una persona che possiede una casa già pagata, sia più benestante di un'altra che vive in una casa in affitto o che paga un mutuo.

Dalla Figura 2.4 si può osservare come il possesso di Internet sia aumentato dal 2009 al 2011, a prova di quanto affermato nel primo capitolo. La descrizione di coloro che non possiedono la connessione alla Rete sarà presentata e discussa nel prossimo paragrafo.

Anche la percentuale di coloro che non possiedono un computer è diminuita tra i due anni, anche se tale percentuale rimane elevata nei Paesi dell'Europa Orientale e in Portogallo. La maggior parte di persone senza un pc ha un'età maggiore ai 55 anni e risiede in zone rurali.

Passando ora al possesso del telefono fisso e del telefono cellulare, la Figura 2.4 mostra come il primo sia in diminuzione a favore del secondo. Le persone che non hanno un telefono fisso a casa risiedono in particolare in Finlandia e nei Paesi dell'Est e hanno meno di 44 anni. Coloro che non hanno un telefono cellulare sono persone che risiedono principalmente in Bulgaria, Romania e Polonia, in zone rurali e sono tipicamente anziane.

Gli intervistati che hanno risposto di non avere una casa di proprietà sono specialmente giovani, mentre quelli che hanno affermato di non avere un'auto sono persone che vivono nei Paesi dell'Europa Orientale, solitamente anziane e di sesso femminile.

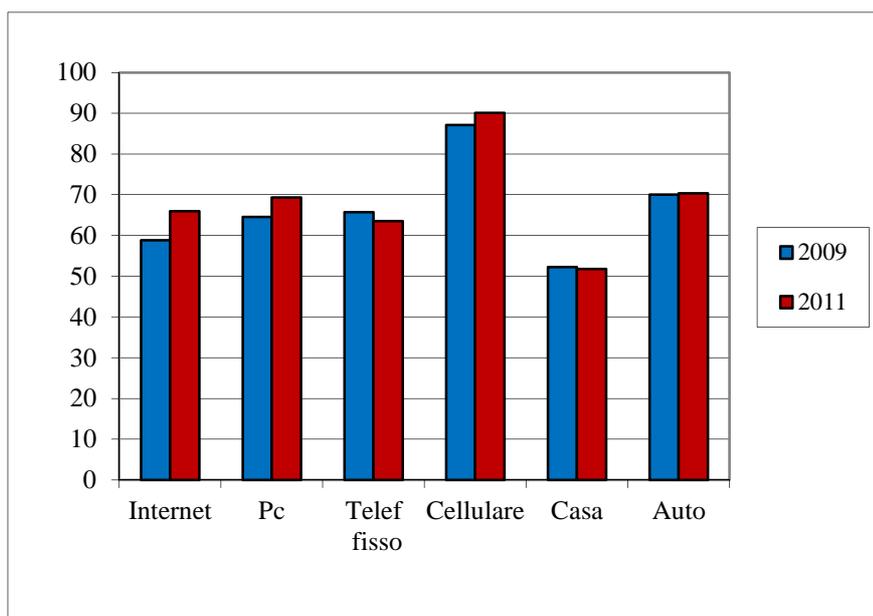


Figura 2.4: Possesso di beni durevoli

Considerando poi l'utilizzo di Internet, una certa crescita è stata registrata tra il 2009 e il 2011. In particolare, mentre nel 2009 la percentuale di intervistati che si connetteva ad Internet a casa era del 57%, nel 2011 è salita al 63%. Il 7% di coloro che hanno la connessione domestica in entrambi gli anni non la utilizza mai; si tratta in particolare di persone anziane. Piccoli aumenti si sono registrati anche nell'utilizzo di Internet in altri luoghi, come scuole e università (21% nel 2009, 22% nel 2011), mentre per quanto con-

cerne l'utilizzo nel posto di lavoro la situazione è rimasta la stessa (28% degli intervistati). Coloro che si collegano ad Internet al lavoro, sono principalmente persone tra i 25 e i 44 anni e persone che hanno concluso gli studi dopo i 15 anni. Al contrario, gli intervistati che si collegano da altri luoghi sono i giovani tra i 15 e i 24 anni.

Analizzando le risposte ad alcune domande dei questionari, emerge inoltre che il 27% degli intervistati del 2009 e il 30% di quelli del 2011, possiede un abbonamento che permette di collegarsi ad Internet dal cellulare. Si tratta in particolare di persone dai 15 ai 44 anni, che hanno concluso gli studi dopo i 15 anni e che appartengono al livello medio della società. I Paesi con la percentuale maggiore di abbonati ad Internet da cellulare sono Svezia, Repubblica Ceca, Slovenia, Lettonia e Gran Bretagna.

Infine, considerando solo coloro che possiedono la connessione Internet a casa e analizzando le modalità di accesso (Figura 2.5), la maggior parte degli intervistati si collega tramite la banda larga (ADSL/XDSL e rete televisiva via cavo): si tratta di persone che risiedono soprattutto nell'Europa Occidentale e Settentrionale. Coloro che, invece, utilizzano la rete telefonica mobile per collegarsi alla Rete, vivono in particolare in Svezia e Finlandia, mentre quelli che utilizzano una linea telefonica standard o una linea ISDN risiedono in Germania (26%) e nei Paesi dell'Europa Orientale. La fibra ottica, chiamata anche banda ultralarga, sembra essere invece ancora poco diffusa. Gli intervistati che hanno affermato di utilizzarla sono concentrati in Svezia, Lituania, Romania, Bulgaria e Slovenia. Sembra perciò che nei Paesi dell'est Europa Internet si stia diffondendo principalmente negli ultimi anni e vengano costruite delle infrastrutture più tecnologiche (la banda ultralarga), le quali affiancano quelle più obsolete (collegamento tramite rete telefonica), senza passare prima per la banda larga. Negli altri Paesi, invece, dove Internet ha iniziato a diffondersi già da anni, le infrastrutture più vecchie sono state sostituite dalla banda larga e quella ultralarga viene ancora poco utilizzata.

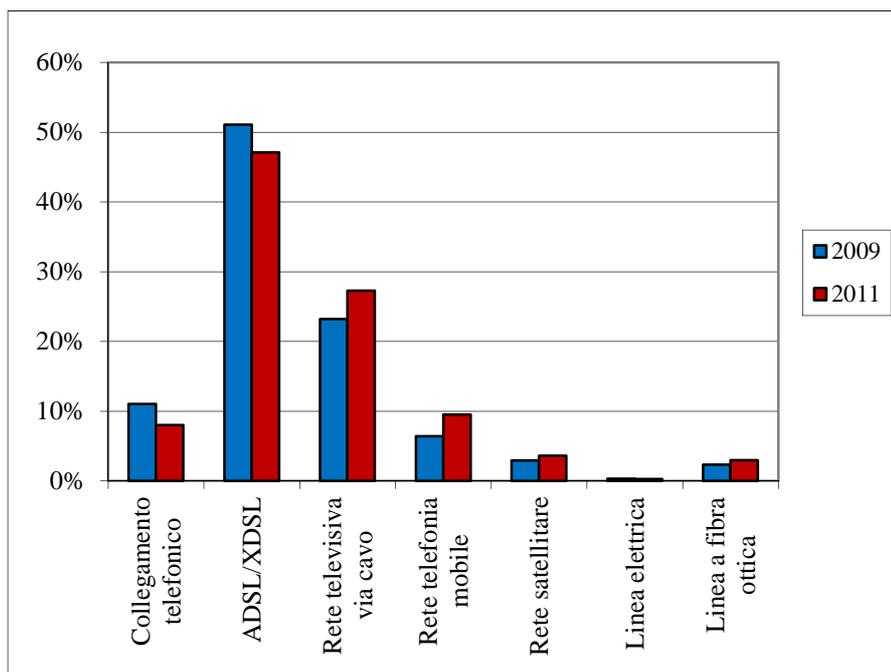


Figura 2.5: Modalità di accesso ad Internet degli intervistati

2.2.2 Caratteristiche dei Paesi

Per l'analisi multilivello ho considerato in aggiunta alcune caratteristiche a livello di Stato: la copertura DSL (*Digital Subscriber Line*), il tasso di disoccupazione e l'HDI (*Human Development Index*).

I dati utilizzati per la copertura DSL sono stati ricavati dal report “*Broadband Coverage in Europe*” dell’Agenda Digitale⁸. La copertura di ogni Paese è stata calcolata come percentuale della popolazione residente che ha accesso alla linea DSL⁹. Gli Stati che permettono a tutta la popolazione di poter accedere ad Internet tramite la linea DSL sono Francia, Belgio, Lussemburgo, Danimarca e Regno Unito, mentre quelli in cui la copertura è minore sono Polonia, Bulgaria e Romania. Tra il 2009 e il 2011 si sono registrati miglioramenti significativi in Irlanda e Romania, mentre negli altri Stati la situazione è rimasta invariata oppure sono stati registrati aumenti di qualche punto percentuale (Figura 2.6).

⁸ http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/broadband_coverage_2010.pdf

⁹ Le linee DSL sono considerate a banda larga, in quanto i dati e la voce sono trasmessi su canali differenti.

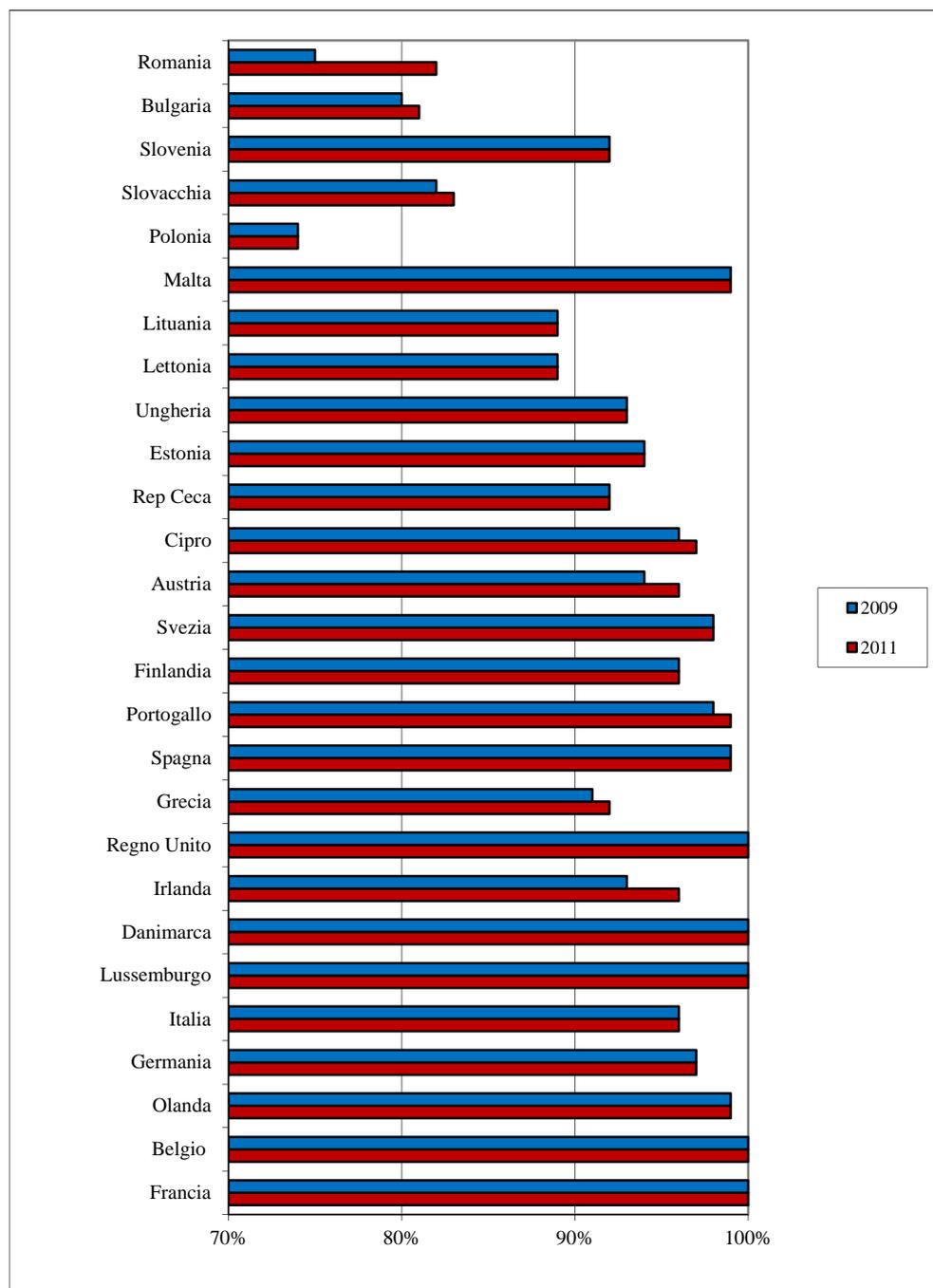


Figura 2.6: Copertura DSL dei 27 Paesi

La seconda variabile che ho considerato è il tasso di disoccupazione e in questo caso i dati provengono da Eurostat¹⁰. Com'è evidente dalla Figura 2.7, gli Stati con il tasso di disoccupazione più elevato sono Spagna e Lettonia, seguiti da Estonia, Slovacchia, Grecia e Irlanda. Gli aumenti più evidenti dal 2009 al 2011 sono quelli di Grecia (dal 10,3% al 15,30%) e Bulgaria (dall'8,10% all'11,2%). Sono state registrate anche delle riduzioni,

¹⁰ http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=une_rt_m&lang=en

in particolare in Lettonia (dal 20,9% al 17,1%) e in Estonia (dal 16,4% al 13,3%).

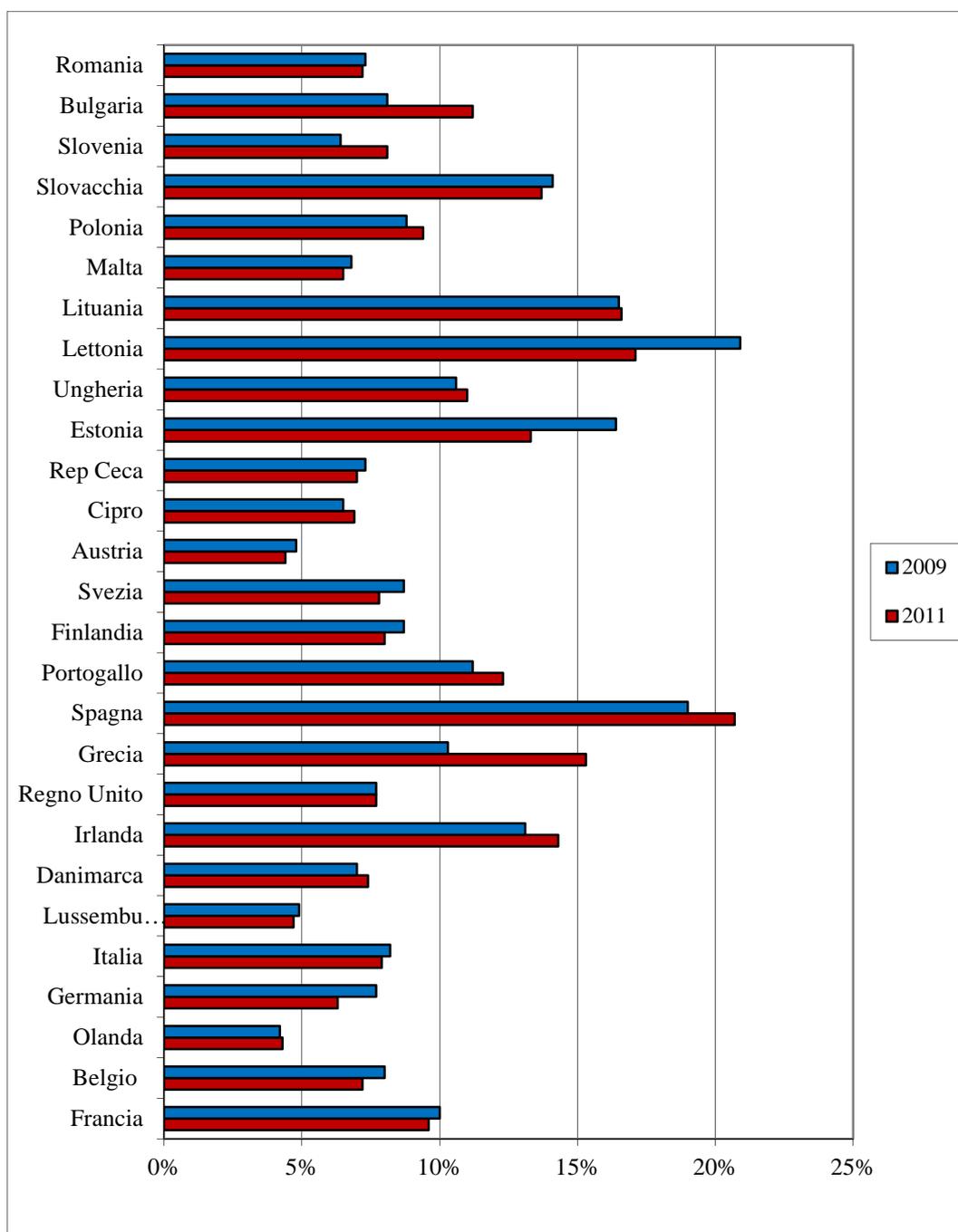


Figura 2.7: Tasso di disoccupazione dei 27 Paesi

La terza variabile che ho utilizzato è l'HDI: si tratta di un indice con valore compreso tra 0 e 1 che, combinando l'aspettativa di vita, il livello di istruzione e il reddito relativi ad ogni Paese, misura il relativo livello di sviluppo. I dati utilizzati provengono dal "Human

Development Report¹¹.

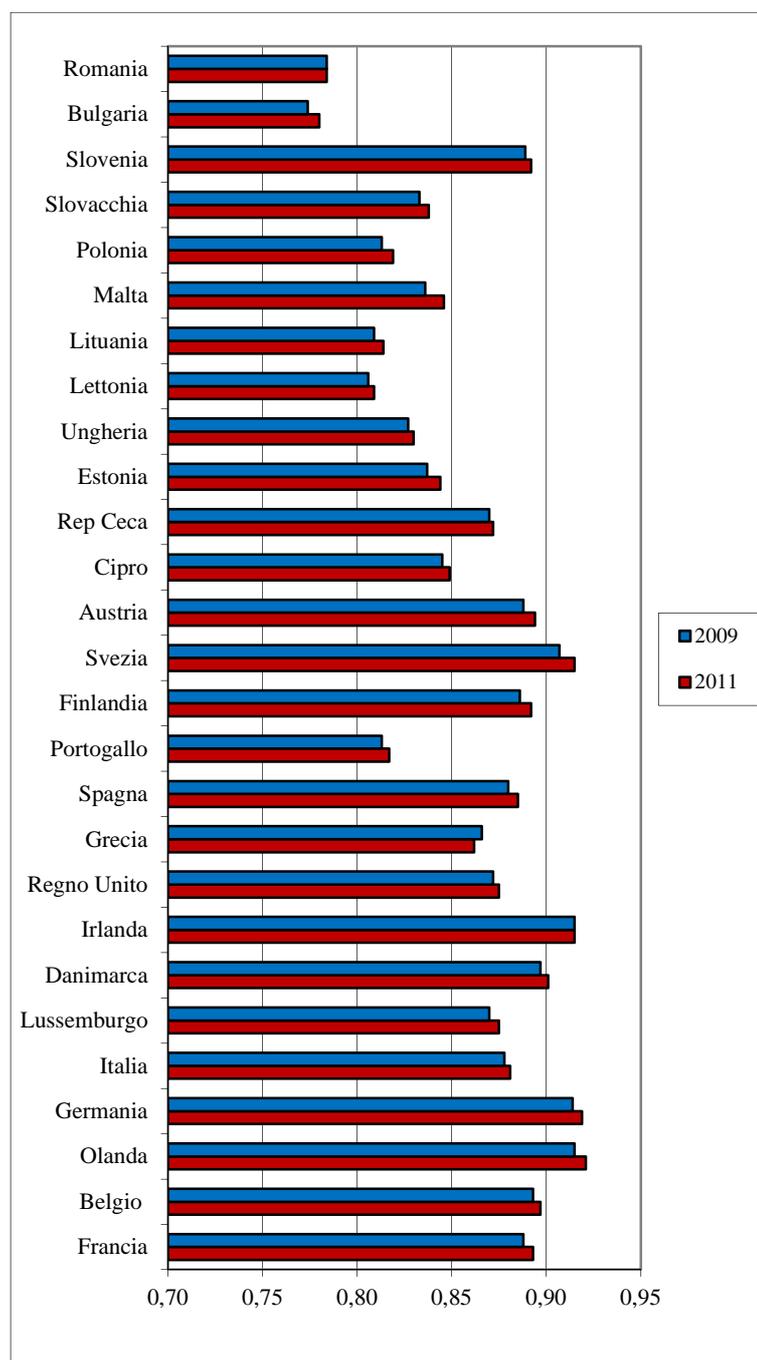


Figura 2.8: HDI dei 27 Paesi

I Paesi con HDI più elevato sono Olanda, Germania, Irlanda e Svezia. Dal 2009 al 2011 sono stati registrati dei piccoli aumenti quasi ovunque tranne in Grecia, dove c'è stata una diminuzione dell'HDI. Probabilmente, questo è dovuto alla forte crisi economica che

¹¹ <http://hdrstats.undp.org/en/indicators/103106.html>

ha colpito questo Stato. Si può, inoltre, osservare come tutti i Paesi dell'Europa Orientale abbiano un HDI al di sotto dello 0,90 (Figura 2.8)

2.2.3 Caratteristiche dei sottocampioni

Nell'analisi a classi latenti, ho considerato solamente le osservazioni relative alle persone intervistate che hanno affermato di non possedere l'accesso alla Rete da casa; si tratta di 10629 osservazioni per il 2009 e 8600 per il 2011 (rispettivamente il 41,14% e il 34,03% delle osservazioni dei campioni iniziali). La Figura 2.9 mostra la percentuale di persone che non possiedono Internet a casa, in ciascun Paese. E' molto evidente che nel 2009 in Portogallo, Spagna, Grecia e negli altri Paesi dell'Europa Orientale, la percentuale di persone che non possiedono la connessione Internet a casa è molto elevata, in particolare Romania e Bulgaria presentano i valori più alti. Al contrario, i Paesi del nord Europa e la maggior parte di quelli occidentali presentano dei valori più bassi, in particolare l'Olanda dove il 90% della popolazione accede ad Internet da casa. Si può osservare un miglioramento della situazione nel 2011, anno in cui la percentuale di persone senza accesso domestico alla Rete è diminuita in tutti gli Stati. Tuttavia, i Paesi elencati in precedenza presentano ancora un'ampia distanza rispetto agli altri. Confrontando tali dati con quelli della copertura DSL evidenziati nel paragrafo 2.2.2, si nota che i Paesi con una copertura minore presentano una proporzione maggiore di persone senza accesso alla Rete: sono quindi rilevanti i problemi legati alle infrastrutture disponibili. Va inoltre osservato che nei Paesi meno vasti è più facile offrire a tutti la possibilità di connettersi da casa e, pertanto, quasi la totalità della popolazione possiede l'accesso domestico a Internet.

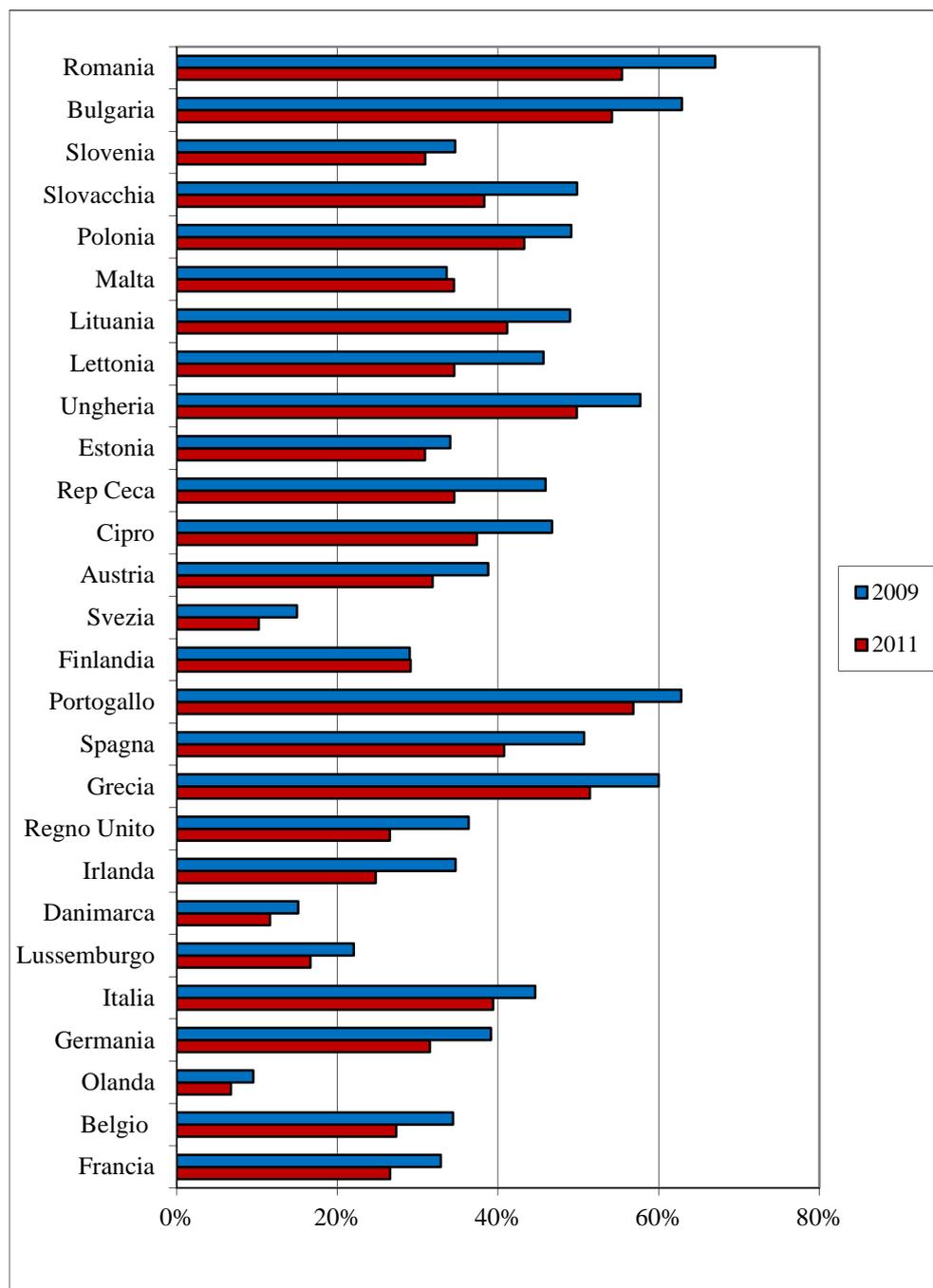


Figura 2.9: Percentuale di intervistati che non possiedono Internet a casa

Analizzando l'età dei sottocampioni, si osserva che la fascia d'età con più osservazioni è quella relativa a 65 e più anni e che, mano a mano che ci si allontana da questa età, la percentuale di intervistati si riduce.

Considerando invece l'istruzione, più della metà di queste persone ha finito gli studi dopo i 15 anni, mentre solo il 2% non ha alcun titolo di studio. Questo risultato è diverso da ciò che mi aspettavo, in quanto a mio parere chi ha finito gli studi in giovane età o

non li ha mai cominciati, non dovrebbe avere le competenze necessarie per utilizzare Internet. Tuttavia, da un'analisi più approfondita emerge che ben l'80% delle persone intervistate senza alcun titolo di studio non ha Internet.

La maggior parte delle persone del sottocampione vive in zone rurali o in piccole città e appartiene al livello basso o medio della società. Considerando la composizione del nucleo familiare, la maggior parte delle persone vive da sola o con un'altra persona. Per quanto riguarda il possesso di una casa propria, il 60% delle osservazioni la possiede. Infine, più dell'80% delle persone senza connessione domestica non possiede nemmeno un computer.

2.3 Motivazioni del non possesso di Internet

La domanda del questionario su cui si concentra la seconda parte della mia analisi è quella in cui viene chiesta la motivazione della mancanza della connessione Internet a casa. Le risposte possibili erano dodici ed ogni persona poteva sceglierne più di una. Il testo della domanda in italiano e le possibili risposte erano i seguenti:

Lei ha detto di non avere l'accesso Internet a casa. Nella lista seguente, quali affermazioni descrivono meglio il motivo per cui la sua famiglia non ha l'accesso ad Internet?

1- Lei o qualcuno della sua famiglia pensa di abbonarsi/collegarsi nei prossimi sei mesi

2- Lei e la sua famiglia non sapete esattamente cos'è Internet

3- Nessuno della sua famiglia è interessato ad Internet

4- La zona dove lei abita non è coperta da una infrastruttura di rete per accesso a banda larga

5- Il costo per l'installazione iniziale della rete a banda larga è troppo elevato

6- Il costo dell'abbonamento mensile ad Internet a banda larga è troppo elevato

7- Il costo per l'acquisto di un PC e di un modem è troppo elevato

8- Il costo dell'abbonamento mensile è troppo elevato

9- I membri della sua famiglia interessati vi accedono al lavoro, a scuola o in qualche altro posto e questo è sufficiente

10- Lei o qualcuno della sua famiglia è preoccupato per la possibilità di accesso a siti con contenuti non adatti

11- Altro

12- Non sa

Dalla Figura 2.10 è evidente che la risposta con la frequenza maggiore è la terza, ossia gli intervistati che non hanno l'accesso Internet a casa perché nessun componente della famiglia è interessato. Tutte le altre risposte presentano delle percentuali molto inferiori e non sembrano esserci molti cambiamenti nei due anni. In particolare, solo l'1,5% ha risposto di essere preoccupato per la possibilità di accedere a siti non adatti.

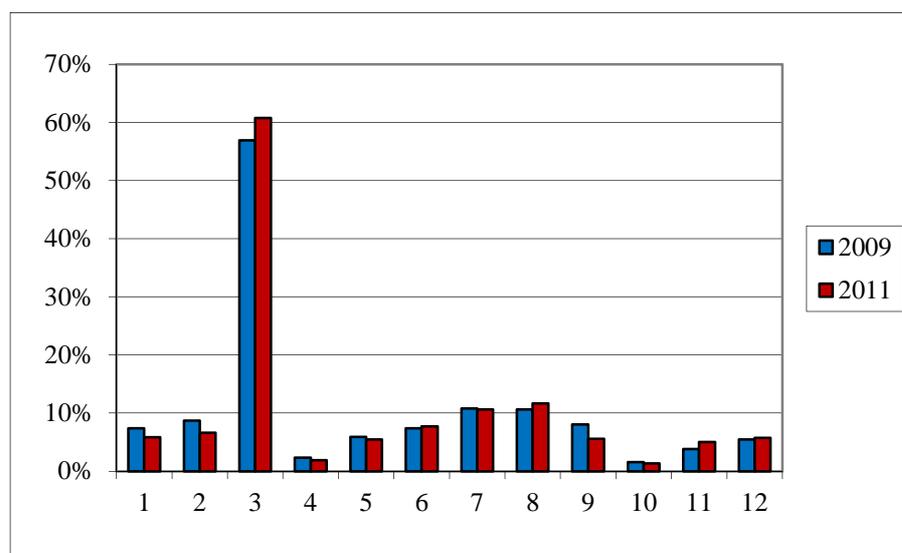


Figura 2.10: Motivazioni date dagli intervistati

Le persone che hanno risposto di voler connettersi ad Internet nei successivi sei mesi sono quelle con un'età compresa tra i 25 e i 44 anni, che risiedono in zone rurali o piccole città e appartengono al livello medio della società. La maggior parte degli intervistati che ha scelto questa prima risposta risiede in Romania, Lituania, Slovacchia e Grecia.

Gli intervistati che hanno affermato di non avere Internet in quanto non sanno esattamente cosa sia sono principalmente donne e persone anziane. Inoltre, più della metà di coloro che hanno scelto tale risposta hanno finito gli studi prima dei 15 anni e il 13% risiede in Romania. Inoltre, la metà delle risposte è stata data da persone che vivono in zone rurali e che appartengono al livello più basso della società.

Le persone che hanno affermato di non essere interessate ad Internet sono per il 50% anziani. Inoltre, sono tendenzialmente le persone che vivono in una zona rurale e coloro che risiedono in Grecia e negli altri Paesi dell'Europa Orientale ad aver scelto tale alter-

nativa.

Gli intervistati che hanno risposto di non possedere la connessione Internet a casa perché risiedono in una zona non coperta dalla banda larga vivono principalmente in Romania, Slovacchia, Slovenia e Polonia.

Gli intervistati che ritengono che il costo iniziale per l'installazione della banda larga sia troppo elevato sono persone che risiedono in particolare in zone rurali, persone che appartengono al livello medio-basso della società e coloro che risiedono in Ungheria, Slovacchia, Romania.

Le persone che hanno affermato di non avere Internet a casa in quanto il costo dell'abbonamento mensile alla banda larga è troppo elevato risiedono principalmente nei Paesi dell'est e, nel 2011, anche in Italia. Inoltre, anche le persone che fanno parte del livello più basso della società hanno scelto tale risposta.

Coloro che ritengono che il costo del pc e del modem ostacoli la possibilità di avere Internet risiedono principalmente in Romania, Bulgaria, Ungheria e Repubblica Ceca e sono persone che appartengono al livello della società più basso. Inoltre, anche le persone che considerano il costo dell'abbonamento mensile troppo elevato vivono nell'Europa Orientale e appartengono alla fascia più disagiata della società.

Gli intervistati che hanno risposto di non avere Internet perché i componenti della famiglia interessati si collegano in altri luoghi sono persone che risiedono principalmente in Slovacchia. Coloro che sono preoccupati per l'accesso a contenuti non adatti non presentano delle caratteristiche particolari.

Infine, coloro che non hanno saputo dare delle risposte precise sono persone anziane e persone dei livelli più bassi della società.

Capitolo 3

I modelli

In questo capitolo presenterò le principali caratteristiche dei modelli stimati nella mia analisi. I primi modelli considerati sono quelli per variabili binarie (in particolare il probit) utilizzati per tracciare il profilo tipico delle persone che non possiedono la connessione Internet a casa. In seguito vengono presentati i modelli multilivello utilizzati in presenza di dati di tipo gerarchico con lo scopo di spiegare la variabilità tra gli Stati o almeno controllare per le differenze tra gli stessi. Infine, sono introdotti i modelli a classi latenti che ho utilizzato per segmentare la popolazione di coloro che non possiedono la connessione Internet da casa, sulla base della relativa motivazione.

3.1 Modelli per variabili binarie

Per individuare le caratteristiche di coloro che non si connettono ad Internet a casa, ho utilizzato un modello per variabili dipendenti binarie. In particolare, ho stimato un modello probit. I modelli probit sono utilizzati quando la variabile risposta da modellare è dicotomica, ossia assume solo due modalità. In questi casi, le modalità della variabile dipendente vengono codificate in “1” e “0”. Nel caso della mia tesi, la variabile binaria considerata è il non possesso dell’accesso Internet da casa. Nel caso di variabili risposta bernoulliane, i modelli che si possono utilizzare sono quello di regressione logistica e il modello *probit*.

Nei modelli a risposta binaria, si è interessati alla probabilità di risposta per vari valori di \mathbf{x}_i (vettore di variabili esplicative $K \times 1$):

$$P(y = 1 | \mathbf{x}_i) = P(y = 1 | x_1, x_2, \dots, x_K)$$

dove y è una variabile binaria che assume il valore 1 se si verifica un determinato evento e 0 altrimenti. Tale variabile è osservata per $i = 1, 2, \dots, N$ individui e si vuole modellare la probabilità di successo in funzione di K variabili esplicative relative a caratteristiche individuali come sesso, età, etc. in modo tale da avere:

$$P(y_i = 1 | \mathbf{x}_i) = F(\mathbf{x}_i, \boldsymbol{\beta})$$

$$P(y_i = 0 | \mathbf{x}_i) = 1 - F(\mathbf{x}_i, \boldsymbol{\beta})$$

con $\boldsymbol{\beta}$ vettore $K \times 1$ di parametri da stimare.

Si consideri un modello a risposta binaria del tipo:

$$P(y_i = 1 | \mathbf{x}_i) = G(\mathbf{x}_i' \boldsymbol{\beta}) \equiv p(\mathbf{x}_i)$$

e si assuma che $0 < G(z) < 1 \forall z \in \mathbb{R}$. Il modello è chiamato *modello indice* (*index model*) (Greene, 2003).

Nella maggior parte dei casi G è una funzione cumulata di probabilità, la cui forma specifica viene derivata dal modello economico corrispondente o scelta per convenienza. Per ricavare il modello indice, si può utilizzare un approccio a variabile latente seguente:

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i \tag{1}$$

in cui y_i^* è la variabile latente mentre si osserva

$$y_i = \begin{cases} 1 & \text{se } y_i^* > 0 \\ 0 & \text{se } y_i^* \leq 0 \end{cases}$$

ossia un indicatore del suo segno. Al posto dello zero può esserci tuttavia qualsiasi altro valore soglia.

Nell'equazione del modello (1) ε_i è un termine d'errore indipendente da \mathbf{x}_i e la sua distribuzione è simmetrica attorno allo zero.

In (1) la direzione degli effetti delle $x_{1i}, x_{2i}, \dots, x_{Ki}$ è data dal segno dei parametri $\beta_1, \beta_2, \dots, \beta_K$. Tuttavia, la variabile latente y_i^* non ha una unità di misura ben definita, perciò la dimensione dei $\beta_1, \beta_2, \dots, \beta_K$ non ha significato. Per questo motivo la dimensione del parametro β_k non può essere interpretata come l'effetto marginale della variabile k -esima (Cappuccio e Orsi, 2005).

Se x_k è una variabile continua, si ha che

$$\frac{\partial p(y_i=1 | \mathbf{x}_i)}{\partial x_{ki}} = g(\mathbf{x}_i' \boldsymbol{\beta}) \cdot \beta_k$$

dove

$$g(z) \equiv \frac{dG(z)}{dz}$$

Se x_k è una variabile binaria, il segno di $\widehat{\beta}_k$ rappresenta il segno dell'effetto marginale di x_{ki} sulla probabilità di successo. Per conoscere la grandezza di tale effetto, occorre stimare (Wooldridge, 2002):

$$G(\widehat{\beta}_1 + \widehat{\beta}_2 x_{2i} + \dots + \widehat{\beta}_k + \dots + \widehat{\beta}_K x_{Ki}) - G(\widehat{\beta}_1 + \widehat{\beta}_2 x_{2i} + \dots + \widehat{\beta}_K x_{Ki})$$

3.2 Modelli multilivello

Dopo aver tracciato i profili delle persone che non possiedono l'accesso Internet a casa, ho stimato dei modelli multilivello per vedere quanta parte della variabilità totale viene spiegata dalla variabilità tra Stati.

I modelli multilivello vengono infatti utilizzati per analizzare dati di tipo gerarchico, ossia dati organizzati su più livelli in cui le unità al livello più basso sono raggruppate in un'unità statistica di livello superiore (gruppi o cluster). Tipici esempi di dati di questo tipo sono quelli relativi agli studenti raggruppati in classi, le quali sono a loro volta raggruppate in scuole oppure i pazienti aggregati in ospedali, a loro volta aggregati in distretti.

I modelli multilivello sono stati sviluppati negli ultimi trent'anni. In precedenza, venivano utilizzati modelli di regressione lineare separati per le unità di ogni livello, tralasciando le relazioni che potevano esistere tra le variabili di un gruppo con quelle di altri gruppi. In questo modo si incorreva in una perdita di informazione chiamata "*ecological fallacy*" (Hox, 2011), in quanto venivano fatte considerazioni solo su un livello senza tenere conto di quelli superiori.

Dagli anni Settanta hanno iniziato a svilupparsi i modelli multilivello, in cui si analizzano contemporaneamente le unità statistiche dei vari livelli. Essi sfruttano infatti la gerarchia di dati e permettono di spiegare l'influenza delle variabili di livello superiore su quelle di livello inferiore (De Leew, Meijer, 2008).

Nei dati di tipo gerarchico i livelli più bassi sono collegati ad un solo livello più alto; esistono quindi delle relazioni uno-a-molti. Le unità del livello inferiore vengono chiamate di primo livello, le quali sono a loro volta raggruppate nelle unità di secondo livello, a loro volta aggregate in quelle di terzo livello e così via. Nella maggior parte dei casi si utilizzano dati relativi a due livelli, per evitare difficoltà di calcolo; si distingue in questi casi tra micro-unità (livello 1) e macro-unità (livello 2).

Le unità che appartengono allo stesso gruppo condividono le stesse influenze specifiche di quel gruppo. Basti pensare al fatto che gli studenti che appartengono alla stessa classe hanno lo stesso insegnante e gli alunni della stessa scuola hanno dei genitori che hanno effettuato la stessa scelta nel decidere la scuola per i figli. Si ottengono così dei gruppi omogenei al loro interno ed eterogenei tra di loro.

I modelli multilivello sono caratterizzati da un'unica variabile risposta, che è sempre una variabile di primo livello, e da una o più variabili esplicative, che possono appartenere a qualunque livello e che permettono di spiegare la variabilità del fenomeno oggetto d'analisi.

Di solito, vengono selezionate le variabili di primo livello che spiegano la variabilità del fenomeno a livello individuale e, successivamente, vengono selezionate anche le variabili di secondo livello, chiamate variabili di contesto, che spiegano la variabilità a livello di gruppo. Queste covariate possono essere variabili che esprimono caratteristiche proprie dei gruppi oppure variabili che esprimono caratteristiche dei gruppi ottenute attraverso la sintesi di caratteristiche a livello individuale, come la media di gruppo di una determinata variabile.

Nella regressione multilivello, l'eterogeneità non osservata è modellata attraverso l'inserimento in un modello di regressione multipla di alcuni effetti casuali. Ci sono due tipi di effetti casuali: intercette casuali e coefficienti casuali. Le prime rappresentano l'eterogeneità non osservata nella risposta globale, mentre i secondi rappresentano l'eterogeneità non osservata negli effetti delle variabili esplicative sulla variabile risposta. In questo modo l'intercetta, ed eventualmente anche i coefficienti, assumono valori diversi tra i gruppi e la curva di regressione assume una posizione diversa nello spazio per ogni gruppo. Questi modelli, a differenza di quelli di regressione classici, presentano quindi più di un termine d'errore nell'equazione del modello, almeno uno per ogni livello di aggregazione.

3.2.1 Coefficiente di Correlazione Intraclasse

Il coefficiente di correlazione intraclasse (ICC) è dato dal rapporto tra la varianza tra i gruppi e la varianza totale. La varianza tra i gruppi viene anche chiamata variabilità between, mentre la varianza totale è data dalla somma della variabilità between con quella within, ossia entro i gruppi.

$$\rho = \frac{\text{variabilità between}}{\text{variabilità between} + \text{variabilità within}}$$

dove $0 \leq \rho \leq 1$.

Questo indice rappresenta la quota di variabilità totale imputabile all'effetto di gruppo ed è utile per capire se è necessario stimare un modello multilivello (Kreft, De Leew, 1999). A tale scopo, si stima un modello senza covariate e si considera il relativo ICC. Se il suo valore è diverso da zero, diventa utile tenere conto della struttura gerarchica del modello e stimare quindi un modello multilivello. Se invece il suo valore è prossimo allo zero, significa che il raggruppamento delle unità non è in grado di determinare effetti positivi sul fenomeno in esame ed è possibile utilizzare dei modelli di regressione ad un solo livello.

Per calcolare le componenti between e within, si deve ricorrere alla centratura delle variabili originali. Sia X_{ij} la variabile esplicativa di primo livello relativa all'individuo i -esimo appartenente al gruppo j -esimo e $\bar{X}_{\cdot j}$ sia la media di gruppo, ossia il valore medio all'interno del gruppo j -esimo di tutte le caratteristiche X_{ij} . Si ottiene:

$$X_{ij} = (X_{ij} - \bar{X}_{\cdot j}) + \bar{X}_{\cdot j}$$

In questa formula, la parte tra parentesi rappresenta la componente within, mentre il secondo addendo rappresenta quella between.

L'indice di correlazione intraclasse è utile anche per capire quali variabili di secondo livello sono più informative nello spiegare la variabilità between. Infatti, se nel modello aggiungiamo una variabile di livello-2 che aiuta a spiegare le differenze tra i gruppi, la variabilità between si riduce, determinando una diminuzione dell'ICC.

In conclusione, l'inserimento nel modello di covariate di primo livello dovrebbe contribuire a spiegare la variabilità individuale, riducendo la componente residua di primo livello. L'inserimento di variabili di secondo livello dovrebbe, invece, spiegare la variabilità tra i gruppi e ridurre la componente di varianza di livello-2. Non è però chiaro il

comportamento per l'altro livello.

3.2.2 Notazione

La notazione utilizzata nella stima dei modelli multilivello è la seguente:

- j indica il generico gruppo ($j = 1, 2, \dots, N$)
- i indica il generico individuo appartenente al gruppo j ($i = 1, 2, \dots, n_j$)
- y_{ij} indica il valore assunto dalla variabile risposta per l'individuo i del gruppo j

Nella trattazione teorica, per semplicità, considererò solo una variabile esplicativa per ogni livello:

- X_{ij} indica il valore assunto dalla variabile esplicativa di primo livello per l'individuo i appartenente al gruppo j
- Z_j indica il valore assunto dalla variabile esplicativa di secondo livello per il gruppo j .

3.2.3 Modello ad intercetta casuale (random intercept model)

Il modello ad intercetta casuale è un modello per dati gerarchici in cui solo l'intercetta varia tra i gruppi.

L'equazione generale del modello ad intercetta casuale è:

$$Y_{ij} = \alpha_j + \beta_j X_{ij} + \varepsilon_{ij} \quad \text{equazione micro}$$

dove:

$$\alpha_j = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$$

$$\beta_j = \gamma_{10} + \gamma_{11} Z_j$$

sono le equazioni macro e α_j sono realizzazioni di variabili casuali; ε_{ij} e u_{0j} sono termini d'errore.

Sostituendo le equazioni macro nella micro, si ottiene:

$$Y_{ij} = \gamma_{00} + \gamma_{01} Z_j + \gamma_{10} X_{ij} + \gamma_{11} Z_j X_{ij} + u_{0j} + \varepsilon_{ij}$$

In questo modo si suppone che l'effetto di ciascuna variabile esplicativa X_{ij} su Y sia lo stesso per tutti i gruppi, mentre solo l'intercetta cambia.

Il modello può quindi essere suddiviso in due parti:

- parte fissa: $\gamma_0 + \gamma_0 X_{ij} + \gamma_1 Z_j + \gamma_1 Z_j X_{ij}$,
- parte casuale: $u_{0j} + \varepsilon_{ij}$

Spesso si assume che i termini d'errore siano indipendenti tra loro e con le covariate. Essi sono così distribuiti:

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$u_{0j} \sim N(0, T_0^2)$$

$$\text{cov}(\varepsilon_{ij}, u_{0j}) = 0$$

La varianza del modello è:

$$\text{Var}(y_{ij} | X_{ij}, Z_j) = \text{Var}(u_{0j}) + \text{Var}(\varepsilon_{ij}) = T_0^2 + \sigma^2$$

Il coefficiente ICC è:

$$\rho = \frac{T_0^2}{T_0^2 + \sigma^2}$$

3.2.4 Modello a pendenza casuale (random slope model)

Il random slope model è un modello per dati gerarchici nel quale i gruppi differiscono casualmente tra loro sia per l'intercetta che per i coefficienti di regressione. Ad ogni gruppo corrisponde, quindi, una retta di regressione che differisce dalle altre sia per l'intercetta sia per la pendenza.

L'equazione del modello è:

$$Y_{ij} = \alpha_j + \beta_j X_{ij} + \varepsilon_{ij}$$

dove:

$$\alpha_j = \gamma_0 + \gamma_1 Z_j + u_{0j}$$

$$\beta_j = \gamma_0 + \gamma_1 Z_j + u_{1j}$$

Sostituendo, si ottiene la seguente equazione del modello in forma compatta:

$$Y_{ij} = \gamma_0 + \gamma_0 X_{ij} + \gamma_1 Z_j + \gamma_1 Z_j X_{ij} + u_{0j} + u_{1j} X_{ij} + \varepsilon_{ij}$$

Il modello può quindi essere suddiviso in:

- parte fissa: $\gamma_0 + \gamma_0 X_{ij} + \gamma_1 Z_j + \gamma_1 Z_j X_{ij}$
- parte casuale: $u_{0j} + u_{1j} X_{ij} + \varepsilon_{ij}$

I termini d'errore sono così distribuiti:

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

$$u_{0j} \sim N(0, T_0^2)$$

$$u_{1j} \sim N(0, T_1^2)$$

$$\text{cov}(u_{0j}, u_{1j}) = \sigma_{u01}$$

$$\varepsilon_{ij} \perp \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim IN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} T_0^2 & \sigma_{u01} \\ \sigma_{u01} & T_1^2 \end{pmatrix} \right)$$

La varianza del modello è:

$$\text{Var}(y_{ij} | X_{ij}, Z_j) = \sigma^2 + T_0^2 + T_1^2 X_{ij}^2 + 2 \sigma_{u01} X_{ij}$$

Le osservazioni non sono tutte indipendenti ed è presente eteroschedasticità. Inoltre, la varianza residua non è costante, ma dipende dal valore delle covariate. Conseguentemente, l'indice ICC perde significato, in quanto la correlazione all'interno dei gruppi è eterogenea e dipende dal valore assunto dalle covariate (Goldstein, 2010).

3.2.5 Stima del modello

Per la stima dei coefficienti di regressione e delle componenti di varianza si possono utilizzare diversi metodi, tra cui la Massima Verosimiglianza (MV), i Minimi Quadrati Generalizzati (GLS) e i metodi Bayesiani. Il metodo della Massima Verosimiglianza permette di ottenere delle stime che godono di alcune proprietà importanti, come la consistenza e l'efficienza. Come detto prima, si assume che il vettore degli errori di primo livello e il vettore degli effetti casuali si distribuiscano come una normale standard.

Per ottenere le stime, si utilizzano dei metodi iterativi. Innanzitutto, il software statistico genera dei valori iniziali per i vari parametri. Dopo un'iterazione, si ottengono le stime dei Minimi Quadrati Generalizzati (GLS). Quando la procedura iterativa converge, si ottengono le stime di Massima Verosimiglianza. Nei casi in cui la procedura non converge o impiega troppo tempo, si possono calcolare le stime GLS.

Per la stima dei modelli multilivello possono essere utilizzati due tipi di Massima Verosimiglianza: quella completa (*Full Maximum Likelihood* - FLM) e quella ristretta (*Restricted Maximum Likelihood* - REML). La differenza è che la Massima Verosimiglianza completa, quando stima le componenti di varianza, considera le stime dei coefficienti di

regressione come quantità note, mentre la Massima Verosimiglianza ristretta le considera come delle stime che portano un certo ammontare di incertezza. Il metodo della REML è più realistico e dovrebbe portare a delle stime migliori, specialmente quando il numero dei gruppi non è ampio. Tuttavia, il metodo della FML ha due vantaggi in più: i calcoli sono più semplici e, dal momento che le stime dei coefficienti sono inserite nella funzione di Massima Verosimiglianza, si può utilizzare il rapporto di verosimiglianza per testare le differenze tra due modelli annidati che differiscono solamente nella parte fissa, ossia nei coefficienti di regressione. Con il metodo REML, invece, in questo modo si possono testare solo le differenze nella parte casuale, ossia nelle componenti di varianza (Hox, 1995).

Nelle mie analisi ho utilizzato il metodo della Massima Verosimiglianza.

3.3 Modelli a classi latenti

Per suddividere la popolazione di coloro che non possiedono la connessione Internet da casa in gruppi (classi), sulla base della motivazione di tale scelta o impossibilità, ho utilizzato i modelli a classi latenti.

Il concetto alla base di tali modelli è che alcuni parametri di un ipotetico modello statistico differiscono tra sottogruppi (classi) non osservati. Questi sottogruppi costituiscono le categorie di una variabile categoriale latente. Pertanto, questi modelli permettono di suddividere una popolazione eterogenea in classi omogenee al loro interno ed eterogenee tra di loro. Le unità statistiche vengono assegnate alle classi su base probabilistica, ossia in base alla probabilità di appartenere a una determinata classe. È inoltre possibile inserire delle covariate per poter descrivere i sottogruppi individuati (Vermunt e Magidson, 2004).

Tali modelli sono stati introdotti per la prima volta nel 1968 da Lazarsfeld e Henry per misurare variabili latenti attitudinali a partire da item dicotomici. Questi modelli potevano essere applicati a dati dicotomici e per questo motivo si diffusero ampiamente, grazie anche ai lavori di Goodman (1974) che estese l'applicazione a variabili nominali e sviluppò l'algoritmo di stima di Massima Verosimiglianza. Ulteriori miglioramenti furono

apportati negli anni successivi, permettendo l'applicazione di tali modelli a indicatori ordinali e continui e l'inserimento di variabili ordinali, nominali e continue.

Il modello a classi latenti che ora presenterò fa riferimento a Vermunt e Magidson (2004) ed è utilizzato nel caso di indicatori categoriali. Gli indicatori sono le variabili osservate o manifeste.

Sia X la variabile latente e Y_l uno degli L indicatori, dove $1 \leq l \leq L$. Sia C il numero di classi latenti e D_l il numero di livelli di Y_l . Una determinata classe latente è indicata con l'indice x , $x = 1, 2, \dots, C$ e un particolare valore di Y_l è indicato da y_l , $y_l = 1, 2, \dots, D_l$. La notazione \mathbf{Y} e \mathbf{y} è utilizzata per indicare un pattern di risposta.

L'idea sottostante qualsiasi modello a classi latenti è che la probabilità di ottenere il pattern di risposta \mathbf{y} , $P(\mathbf{Y} = \mathbf{y})$, è una media pesata delle C probabilità condizionate $P(\mathbf{Y} = \mathbf{y} | X = x)$. Il modello può quindi essere scritto come:

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{x=1}^C P(X = x)P(\mathbf{Y} = \mathbf{y} | X = x)$$

dove $P(X = x)$ indica la probabilità di appartenere alla classe latente x .

Nei modelli a classi latenti viene fatta l'assunzione di *indipendenza locale*, secondo cui le L variabili osservate sono mutualmente indipendenti entro ciascuna classe latente. Tale assunzione può essere formulata come:

$$P(\mathbf{Y} = \mathbf{y} | X = x) = \prod_{l=1}^L P(Y_l = y_l | X = x)$$

dove $P(Y_l = y_l | X = x)$ rappresenta la probabilità di dare una certa risposta y_l alla domanda Y_l , condizionata alla classe di appartenenza x . Tali probabilità, dopo essere state stimate, possono essere confrontate tra le diverse classi in modo da individuare le caratteristiche principali di ognuna di esse e attribuirne un nome.

Combinando le due equazioni appena presentate, si ottiene il seguente modello:

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{x=1}^C P(X = x) \prod_{l=1}^L P(Y_l = y_l | X = x)$$

Uno degli scopi dell'analisi a classi latenti è di suddividere gli individui in segmenti e determinare la probabilità di appartenere alla classe latente x , probabilità che può essere ottenuta da:

$$P(X = x | \mathbf{Y} = \mathbf{y}) = \frac{P(X = x)P(\mathbf{Y} = \mathbf{y} | X = x)}{P(\mathbf{Y} = \mathbf{y})}$$

Come già detto, tale modello si riferisce a indicatori nominali. La probabilità $P(X = x | \mathbf{Y} = \mathbf{y})$ ha una distribuzione logit multinomiale e, come dimostrato da Haberman (1979), il modello a classi latenti può essere specificato come un modello log-lineare.

3.3.1 Stima

Uno degli scopi principali dell'analisi a classi latenti è determinare il minor numero di classi latenti C che riesca a spiegare le relazioni osservate tra le variabili manifeste.

L'analisi inizia con la stima di un modello con una sola classe latente, $C = 1$. Se tale modello non fornisce una stima adeguata dei dati, si stima un secondo modello con $C = 2$ e così via, finché non si perviene al modello più adatto ai dati.

Il metodo di stima principalmente utilizzato per i modelli a classi latenti è quello della Massima Verosimiglianza. Sia I il numero totale dei pattern di risposta, in modo che $I = \prod_{l=1}^L D_l$, e sia i un particolare pattern di risposta. Sia, inoltre, $P(\mathbf{Y} = \mathbf{y}_i)$ la probabilità di osservare il pattern di risposta i e n_i le sue frequenze osservate. La funzione di log-verosimiglianza da massimizzare è:

$$\ln \mathcal{L} = \sum_{i=1}^I n_i \ln P(\mathbf{Y} = \mathbf{y}_i)$$

Solo i pattern che hanno una frequenza diversa da zero contribuiscono alla log-verosimiglianza.

Uno dei problemi che si possono riscontrare nella stima dei modelli a classi latenti è che i parametri potrebbero non essere identificati, anche se i gradi di libertà sono maggiori o uguali a zero. Questo significa che differenti valori dei parametri portano allo stesso punto di massimo nella funzione di log-verosimiglianza e, quindi, le stime non sono uniche.

Un altro problema che si può riscontrare nella stima di questi modelli è la presenza di massimi locali. La funzione di log-verosimiglianza di un modello a classi latenti non è sempre concava; questo significa che l'algoritmo potrebbe convergere ad un massimo

differente a seconda dei valori iniziali generati dal programma di stima. Perciò, il modo migliore di procedere è di stimare il modello con set di valori iniziali casuali differenti: il valore a cui converge la maggior parte dei set può essere considerato la soluzione di MV.

Un ulteriore problema riscontrabile nella stima è quello di ottenere dei valori limiti, ossia delle probabilità uguali a zero o a uno. Per risolvere i problemi legati alla stima dei modelli a classi latenti, si possono utilizzare l'algorithm EM (*Expectation-Maximization*) e l'algorithm NR (*Newton-Raphson*). Il primo è un metodo iterativo molto veloce per la stima di MV quando si hanno dati incompleti; il secondo è una procedura veloce che necessita, tuttavia, di buoni valori iniziali per convergere.

3.3.2 Misure di bontà di adattamento

Per decidere quale tra i modelli stimati sia il migliore, si possono utilizzare diverse misure di bontà di adattamento.

L'approccio più utilizzato è la statistica rapporto di verosimiglianza L^2 , che misura quanto le stime di massima verosimiglianza ottenute per le frequenze attese differiscono dalle corrispondenti frequenze osservate. Tale statistica si distribuisce come un χ^2 :

$$L^2 = 2 \sum_{i=1}^I n_i \ln \frac{n_i}{N \cdot P(\mathbf{Y} = \mathbf{y}_i)}$$

dove N indica la dimensione totale del campione. Il numero di gradi di libertà (df) equivale al numero dei possibili pattern di risposta meno il numero di parametri da stimare.

In un modello a classi latenti non ristretto, si ha:

$$df = \prod_{l=1}^L D_l - C \cdot [1 + \sum_{l=1}^L (D_l - 1)]$$

Quando il numero di variabili osservate considerate nell'analisi, o il numero delle loro categorie, è elevato, o quando si includono nel modello variabili continue, si possono avere dati sparsi. In questo caso, il χ^2 non approssima bene il test L^2 e il p-value associato alla statistica non può essere interpretato. Al posto del p-value asintotico si può invece calcolare il bootstrap p-value, utilizzando un metodo iterativo di stima.

Quando le frequenze attese corrispondono esattamente a quelle osservate, il valore della statistica L^2 è zero e il modello si adatta perfettamente ai dati. Quando invece $L^2 > 0$, esso misura la mancanza di adeguatezza del modello.

Nei casi in cui il modello base (o nullo) con $C = 1$ non si adatta bene ai dati e non spiega l'associazione tra le variabili, il valore della statistica L^2 per il modello base può essere comparato a quello di modelli con più classi come misura dell'adattamento della stima ai dati. Infatti, la percentuale di riduzione di L^2 tra modello nullo e uno con $C > 1$ rappresenta l'associazione totale spiegata da quest'ultimo modello.

Per analizzare la bontà del modello stimato, oltre al rapporto di verosimiglianza si possono utilizzare dei criteri informativi che prendono in considerazione sia la bontà della stima sia la parsimonia del modello. I criteri informativi da sempre più utilizzati sono l'AIC, il BIC e il CAIC; in particolare, gli ultimi due vengono comunemente utilizzati nell'ambito dell'analisi a classi latenti e si preferisce un modello che presenta un valore inferiore della statistica considerata (Magidson, Vermunt, 2004).

3.3.3 Significatività degli effetti

Una volta individuato il modello a classi latenti che si adatta meglio ai dati, si valuta la significatività dei parametri. In particolare, si vuole verificare se gli indicatori Y_l presentano differenze tra le classi. A tale scopo si può utilizzare la statistica ΔL^2 , calcolata come differenza tra L^2 del modello ristretto senza la variabile in esame e L^2 del modello che include tale variabile.

Un altro test che si può utilizzare è quello di Wald, che verifica se i coefficienti di regressione sono uguali tra le classi. Entrambe le statistiche si approssimano sotto H_0 come un χ^2 con $df = (D_l - 1)(C - 1)$, dove D_l indica il numero di categorie dell'indicatore Y_l ; tuttavia, il test di Wald è meno potente di ΔL^2 .

3.3.4 Classificazione in classi

Lo scopo dell'analisi a classi latenti è di classificare gli individui nella classe latente appropriata. A tale scopo si calcola la probabilità a posteriori di appartenere alla classe x ,

dato il pattern di risposta \mathbf{y} ; tale probabilità può essere ottenuta applicando il Teorema di Bayes:

$$\hat{P}(X = x | \mathbf{Y} = \mathbf{y}) = \frac{\hat{P}(X = x) \hat{P}(\mathbf{Y} = \mathbf{y} | X = x)}{\hat{P}(\mathbf{Y} = \mathbf{y})}$$

Le unità vengono in seguito assegnate alla classe con probabilità a posteriori maggiore.

3.3.5 Estensioni

Tra le numerose estensioni del modello a classi latenti tradizionale, quella più importante è l'inserimento di covariate \mathbf{Z} , in modo da ottenere una descrizione più dettagliata della variabile latente e delle classi. Se le covariate sono variabili categoriali, viene utilizzato un modello logit multinomiale per la probabilità di appartenere alla classe x ; tale modello è:

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{Z} = \mathbf{z}) = \sum_{x=1}^C P(X = x | \mathbf{Z} = \mathbf{z}) \prod_{l=1}^L P(Y_l = y_l | X = x)$$

Si possono riscontrare delle situazioni nelle quali un modello con C classi latenti non spieghi adeguatamente i dati, in quanto l'indipendenza locale non è soddisfatta per una o più coppie di indicatori. In questi casi, si possono utilizzare due strategie. La prima consiste nel rilassare l'assunzione di indipendenza locale introducendo effetti diretti tra alcune variabili osservate; la seconda consiste nell'aumentare il numero di variabili latenti invece del numero di classi latenti. Entrambe le strategie portano a dei modelli più parsimoniosi.

3.4 L'analisi multilivello a classi latenti

Una forte limitazione dei modelli appena presentati è l'assunzione che le osservazioni siano indipendenti. Tale assunzione è spesso violata e, per far fronte a tale problema, sono stati sviluppati i modelli multilivello a classi latenti, inizialmente ad opera di Vermunt (2003). Tali modelli, chiamati anche modelli a coefficienti casuali, rilassano l'assunzione di indipendenza.

I modelli multilivello a classi latenti possono essere utilizzati quando si lavora con dati gerarchici, come nel caso della mia tesi, in cui gli individui (unità di primo livello) sono raggruppati in Stati (unità di secondo livello).

In tali modelli, la notazione è sostanzialmente la stessa dei modelli a classi latenti tradizionali; l'unica differenza è l'introduzione degli indici i e j per distinguere le unità dei due livelli. Sia, perciò, X_{ij} una variabile latente con C classi e Y_l una delle L variabili manifeste, con $1 \leq l \leq L$. Sia Y_{ijl} la risposta al quesito Y_l dell'unità i di primo livello, appartenente all'unità j di secondo livello. Il vettore di risposte date dall'unità i -esima del gruppo j -esimo è indicato con \mathbf{Y}_{ij} , mentre \mathbf{y}_{ij} rappresenta un possibile pattern di risposta.

3.4.1 Modelli multilivello a classi latenti con effetti fissi

Il modello può essere formulato nel modo seguente:

$$P(\mathbf{Y}_{ij} = \mathbf{y}_{ij}) = \sum_{x=1}^C P(X_{ij} = x) \prod_{l=1}^L P(Y_{ijl} = y_l | X_{ij} = x)$$

Le due componenti del modello hanno una distribuzione logit multinomiale:

$$P(X_{ij} = x) = \frac{\exp(\gamma_x)}{\sum_{r=1}^C \exp(\gamma_r)}$$

$$P(Y_{ijl} = y_l | X_{ij} = x) = \frac{\exp(\beta_{y_l x}^l)}{\sum_{r=1}^{D_l} \exp(\beta_{rx}^l)}$$

Se l'indice j è presente in entrambe le equazioni, significa che i valori di β e γ dipendono dal gruppo a cui l'individuo i -esimo appartiene. Se l'indice j è presente solo nella prima equazione, significa che la dipendenza dal gruppo coinvolge solo la probabilità di appartenere ad una classe latente. Infine, se l'indice j non compare in nessuna delle due equazioni, si ricade nel modello a classi latenti tradizionale presentato in precedenza.

Come si può notare, nel caso di effetti fissi il numero di parametri da stimare aumenta con l'aumentare del numero dei gruppi e si rischia di ottenere delle stime non stabili: il modello da stimare risulta essere complesso.

3.4.2 Modelli multilivello a classi latenti ad effetti casuali

Per ovviare al problema legato all'approccio ad effetti fissi, si può utilizzare l'approccio ad effetti casuali. In questo modo, invece di stimare i parametri per ogni gruppo, si assume che gli effetti specifici dei gruppi seguano una determinata distribuzione di probabilità, tipicamente normale. In questo modo, in presenza di C classi latenti, si specifica la distribuzione di $C-1$ intercette casuali γ_{xj} e si pone γ_{1j} uguale a zero per l'identificazione.

Si hanno ora due possibilità: la prima è di ipotizzare una distribuzione normale $(C-1)$ -dimensionale, la seconda è di utilizzare una struttura più ristretta $\gamma_{xj} = \gamma_x + \tau_x \cdot u_j$ con $u_j \sim N(0,1)$ e τ_x parametro ignoto da stimare. L'assunzione implicita che viene fatta è che le componenti casuali γ_{xj} siano perfettamente correlate. Nello specifico, lo stesso effetto casuale u_j viene riscaldato per i diversi x in base al valore ignoto di τ_x . Tale approccio viene definito *parametrico* e ha ricevuto molte critiche, in quanto l'assunzione di distribuzione normale per i parametri entro le classi è molto forte e spesso non è rispettata dai dati.

Un'alternativa è rappresentata dall'approccio *non parametrico*, in cui l'ipotesi di distribuzione normale è sostituita con quella di una generica distribuzione discreta meno impegnativa. Tale alternativa comporta la definizione di una variabile latente anche al livello 2. L'idea alla base di tale approccio è che ogni gruppo appartenga a una fra le M classi della variabile latente di secondo livello X_j^g e che le osservazioni entro ciascun gruppo siano indipendenti data l'appartenenza alla classe del gruppo:

$$P(\mathbf{Y}_j = \mathbf{y}_j) = \sum_{x^g=1}^M P(X_j^g = x^g) \prod_{i=1}^{n_j} P(\mathbf{Y}_{ij} = \mathbf{y}_{ij} | X_j^g = x^g)$$

dove \mathbf{y}_j è il vettore di risposte del gruppo j -esimo e X_j^g è la variabile latente discreta di livello 2.

Per la probabilità condizionata $P(X_{ij} = x | X_j^g = x^g)$ si assume, come nel caso precedente, una distribuzione logit multinomiale:

$$P(X_{ij} = x | X_j^g = x^g) = \frac{\exp(\gamma_{xx^g})}{\sum_{r=1}^C \exp(\gamma_{rx^g})}$$

dove $\gamma_{xx^g} = \gamma_x + u_{xx^g}$ e u_{xx^g} ha distribuzione discreta non specificata.

Come per l'approccio parametrico, si ipotizza che il condizionamento al gruppo di appartenenza riguardi solo la probabilità di appartenere ad una determinata classe a livello 1 e non per il pattern di risposta. Questa specificazione corrisponde ad un modello ad intercetta casuale anziché ad un modello a pendenza casuale (Snijders e Bosker, 1999).

Per identificare in modo adeguato i segmenti a livello 1 e 2, è possibile inserire nel modello delle covariate. Sia Z_{j1} una covariata di livello 2 e Z_{ij2} una covariata di livello 1; il modello diventa:

$$P(X_{ij} = x | Z_{1j}, Z_{2j}, X_j^g = x^g) = \frac{\exp(\gamma_{0x^g} + \gamma_{1x} Z_{1j} + \gamma_{2xx^g} Z_{2ij})}{\sum_{r=1}^C \exp(\gamma_{0rx^g} + \gamma_{1r} Z_{1j} + \gamma_{2rx^g} Z_{2ij})}$$

Il modello multilivello a classi latenti non parametrico permette, a differenza di quello tradizionale, di effettuare una duplice segmentazione e determinare a quale classe appartiene ciascun individuo e ciascun gruppo.

3.4.3 Scelta del numero di classi

Il metodo da sempre più utilizzato nell'approccio non parametrico per la scelta del numero di classi a livello 1 e a livello 2 è quello simultaneo. Questo metodo prevede di stimare svariati modelli con combinazioni diverse di C (numero di classi) e M (numero di gruppi) e scegliere alla fine quello che presenta il valore minore del criterio informativo considerato.

Recentemente, è stato sviluppato un lavoro ad opera di Lukočienė, Varriale e Vermunt (2010) che analizza i diversi criteri informativi e ne confronta le relative performance nell'ambito della scelta simultanea del numero di classi e di gruppi. Gli autori presentano, inoltre, un nuovo metodo a tre passi per la selezione del numero di classi a livello 1 e 2 in modo più efficiente.

Lo svantaggio maggiore nella scelta simultanea di C e M è che occorre stimare un elevato numero di modelli; inoltre, non è possibile utilizzare misure diverse nello scegliere il numero di classi e quello dei gruppi. La procedura a tre passi è invece meno intensa dal punto di vista computazionale, tiene in considerazione il fatto che il valore di C può dipendere da quello di M e permette di utilizzare degli indicatori differenti nella scelta di C e di M .

La nuova procedura si compone dunque di tre fasi. Nella prima si determina il numero di classi a livello 1, ignorando la struttura gerarchica dei dati. Nella fase successiva, C viene fissato in base al valore selezionato nel passo precedente e si determina così il numero dei gruppi. Infine, nella terza fase si fissa M sulla base del valore trovato al passo due e si determina nuovamente il numero di classi a livello 1. In questo modo, è possibile utilizzare dei criteri diversi di scelta nelle varie fasi.

Per quanto concerne i diversi criteri informativi l'AIC3 lavora meglio con variabili risposta categoriali, mentre il BIC con quelle continue. Nell'ambito dell'analisi multilivello si riscontrano dei problemi nell'utilizzo di BIC e CAIC, in quanto non è chiaro se, nel loro calcolo, è meglio considerare il numero dei gruppi (K) o il numero di individui (N) o entrambi.

Nella loro analisi, gli autori giungono ad alcuni risultati significativi:

- Nella scelta del numero di classi a livello 1, l'AIC3 è il criterio che funziona meglio, mentre $BIC(N)$ e $CAIC(N)$ tendono a sottostimare il numero di classi. Gli indicatori $BIC(K)$ e $CAIC(K)$ sono allo stesso livello di AIC3, mentre sono migliori rispetto a $BIC(N)$ e $CAIC(N)$.
- Nella scelta del numero di classi a livello 2, AIC3 e $BIC(K)$ sono migliori rispetto a tutte le altre misure. In particolare, AIC3 è leggermente migliore quando c'è una separazione minore tra le classi, ossia quando esse hanno un'ampiezza minore. Anche in questo caso $BIC(N)$ e $CAIC(N)$ tendono a sottostimare il numero di classi rispetto a $BIC(K)$ e $CAIC(K)$. Inoltre, il criterio CAIC è leggermente peggiore del BIC, mentre AIC tende a sovrastimare il numero di classi.
- Nella scelta simultanea del numero di classi a livello 1 e 2, $BIC(K)$ è migliore di AIC3 quando le classi a entrambi i livelli sono ben separate ed è migliore anche rispetto a $BIC(N)$. Infine, considerando la stessa dimensione del campione, i criteri BIC sono migliori di quelli CAIC.
- Nel caso di bassa separazione tra le classi a livello 1, la procedura a tre passi è molto importante. Negli altri casi, essa non porta comunque a dei risultati peggiori.

Oltre a confrontare i diversi criteri informativi e i due diversi metodi, gli autori ricordano che non è importante solo la bontà di adattamento del modello scelto, ma anche

l'interpretabilità dei risultati ottenuti. A volte è perciò preferibile scegliere un numero di classi minore rispetto a quello indicato dal relativo criterio, in modo da ottenere delle classi che possono essere interpretate più facilmente.

Capitolo 4

Profilo tipico dei non possessori

In questo capitolo presenterò i risultati ottenuti dalle stime dei modelli probit con lo scopo di individuare il profilo tipico dei non possessori di Internet a casa nei due istanti temporali considerati. In seguito, utilizzerò l'approccio multilivello per tenere conto della struttura gerarchica dei dati e determinare se esistono differenze nel non possesso di Internet tra gli Stati. Inserirò inoltre delle variabili a livello di Paese e ne analizzerò il contributo nell'apportare informazioni aggiuntive al profilo tipico dei non possessori.

4.1 Stima dei modelli probit

Il primo obiettivo della mia tesi è di determinare quali caratteristiche influenzino maggiormente la probabilità di non possedere Internet a casa, in modo da individuare il profilo tipico dei non possessori. A tale scopo ho stimato innanzitutto un modello probit in cui la variabile dipendente è `No_Internet`, una dummy che vale 1 se la persona intervistata non possiede Internet a casa e 0 altrimenti. Le variabili esplicative inserite nel modello sono:

- Femmina: una dummy che vale 1 se la persona intervistata è di sesso femminile;
- Età ed `Età_quadrato`;
- `Educ0`: una dummy che vale 1 se la persona intervistata non ha mai studiato o ha terminato gli studi prima dei 7 anni (categoria di riferimento per l'insieme di dummy che descrivono il livello d'istruzione dell'intervistato);

- Educ1: una variabile dummy che vale 1 se la persona intervistata ha terminato gli studi tra i 7 e i 15 anni;
- Educ2: una variabile dummy che vale 1 se la persona intervistata ha terminato gli studi dopo i 15 anni;
- Studia_ancora: una variabile dummy che vale 1 se la persona intervistata non ha ancora terminato gli studi;
- Campagna: una dummy che vale 1 se la persona intervistata risiede in una zona rurale (categoria di riferimento per l'insieme di dummy che descrivono la grandezza della città in cui vive l'intervistato);
- Città: una variabile dummy che vale 1 se la persona intervistata risiede in una piccola o media città;
- Grande_Città: una dummy che vale 1 se la persona intervistata risiede in una grande città;
- Livello_Basso: una dummy che vale 1 se la famiglia della persona intervistata appartiene al livello basso della società (categoria di riferimento per l'insieme di dummy che descrivono il livello della società a cui la famiglia dell'intervistato appartiene);
- Livello_Medio: una dummy che vale 1 se la famiglia della persona intervistata appartiene al livello medio della società;
- Livello_Alto: una dummy che vale 1 se la famiglia della persona intervistata appartiene al livello alto della società;
- Bambini: una dummy che vale 1 se nella famiglia della persona intervistata sono presenti bambini con un'età inferiore ai 10 anni;
- Ragazzi: una dummy che vale 1 se nella famiglia della persona intervistata sono presenti ragazzi con un'età compresa tra i 10 e i 14 anni;
- Componenti: una variabile che indica il numero di componenti del nucleo familiare della persona intervistata;
- Auto: una dummy che vale 1 se la persona intervistata possiede un'auto;
- Casa: una dummy che vale 1 se la persona intervistata possiede una casa completamente pagata.

La Tabella 4.1 mostra le stime ottenute per i modelli probit stimati per i due istanti temporali considerati.

4. Profilo tipico dei non possessori

Variabili	2009		2011	
	$\hat{\beta}$	p-value	$\hat{\beta}$	p-value
Femmina	.0162718	0.368	-.0230707	0.233
Età	-2.996143	0.000	-2.803513	0.000
Età_quadrato	4.458369	0.000	4.524816	0.000
Educ1	.1915231	0.063	.0037006	0.979
Educ2	-.4253513	0.000	-.6625513	0.000
Studia_ancora	-.9936416	0.000	-1.133199	0.000
Città	-.2376675	0.000	-.2006911	0.000
Grande_Città	-.3424919	0.000	-.4187321	0.000
Livello_Medio	-.378644	0.000	-.3924649	0.000
Livello_Alto	-.6628078	0.000	-.7150777	0.000
Bambini	.2051206	0.000	.1071129	0.001
Ragazzi	-.0642426	0.042	-.1383562	0.000
Componenti	-.2194687	0.000	-.206008	0.000
Auto	-.7866277	0.000	-.9223933	0.000
Casa	.3183247	0.000	.2797429	0.000
Costante	1.802541	0.000	1.793801	0.000

Tabella 4.1: Stime dei modelli probit per il 2009 e il 2011

Le stime ottenute per l'anno 2009 sono tutte significative al livello del 5% ad eccezione di Femmina e Educ1; quest'ultima è tuttavia significativa al 10%. Nel 2011 invece Educ1 non è più significativa a nessun livello. Questo significa che tutte le variabili considerate influiscono sulla probabilità di non possedere Internet a casa, tranne il sesso dell'intervistato e il fatto di aver terminato gli studi tra i 7 e i 15 anni.

Come spiegato nel Capitolo 3, le stime ottenute possono essere utilizzate solo per determinare la direzione dell'effetto di ciascuna variabile, mentre la grandezza dell'effetto viene determinata calcolando gli effetti marginali che ho riportato nella Tabella 4.2. Occorre ricordare che nell'interpretare la direzione e la dimensione dell'effetto di una variabile si considerano costanti tutte le altre variabili.

4. Profilo tipico dei non possessori

Variabili	2009		2011	
	dy/dx	p-value	dy/dx	p-value
Femmina	.0062971	0.368	-.0080416	0.233
Età	-1.159691	0.000	-.9768629	0.000
Età_quadrato	1.725662	0.000	1.576638	0.000
Educ1	.07498	0.063	.0012902	0.979
Educ2	-.1665396	0.000	-.2432594	0.000
Studia_ancora	-.3104434	0.000	-.2736745	0.000
Città	-.0909759	0.000	-.0689286	0.000
Grande_Città	-.1291684	0.000	-.1376174	0.000
Livello_Medio	-.1455449	0.000	-.1357033	0.000
Livello_Alto	-.240758	0.000	-.2228478	0.000
Bambini	.0803455	0.000	.0379391	0.001
Ragazzi	-.0247063	0.042	-.0468295	0.000
Componenti	-.0849478	0.000	-.0717819	0.000
Auto	-.3047826	0.000	-.337521	0.000
Casa	.1225047	0.000	.0969903	0.000

Tabella 4.2: Effetti marginali dei modelli probit stimati per il 2009 e 2011

Dalla Tabella 4.2 è evidente come, a parte l'età, le variabili con un effetto marginale maggiore sul non possesso di Internet siano quelle relative all'appartenenza al livello alto della società, al possesso di un'auto e al fatto di non aver ancora terminato gli studi. In particolare, queste tre caratteristiche fanno diminuire la probabilità di non aver Internet.

Dalle stime si osserva che la probabilità di non avere Internet a casa tende a diminuire per le persone con età compresa tra 15 e i 31 anni, per poi aumentare e raggiungere i livelli più alti nelle persone anziane. Tale probabilità, a parità di altre condizioni, è maggiore per coloro che hanno finito gli studi tra i 7 e i 15 anni e minore per coloro che hanno terminato gli studi dopo i 15 anni o stanno ancora studiando. Le persone con una maggiore istruzione hanno probabilmente maggiori capacità e conoscenze e quindi è più probabile che siano interessate ad avere Internet a casa, tanto da decidere di installarlo. Chi invece ha terminato gli studi dopo i 15 anni ha una probabilità minore del 16,7% rispetto a chi non ha alcun titolo di studio, e per chi sta ancora studiando è minore del 31%.

La probabilità di non avere la connessione Internet a casa è influenzata anche dalla zona di residenza ed è minore per coloro che vivono in zone urbane, qualunque sia la gran-

dezza della città. Probabilmente, coloro che risiedono in zone non urbane vivono in aree non coperte dalle infrastrutture necessarie. Prendendo come riferimento una persona che risiede in una zona rurale, la probabilità di non avere l'accesso ad Internet a casa per un individuo con le stesse caratteristiche che abita in una piccola o media città è minore del 9,1%; per una persona che risiede in una grande città tale probabilità è minore del 12,9%.

Anche le variabili relative al livello della società influiscono negativamente sulla probabilità di non avere Internet. Infatti, una persona che appartiene al livello medio della società ha una probabilità minore del 14,6% di non avere la connessione domestica, rispetto ad un individuo con le stesse caratteristiche che appartiene al livello basso della società. Se invece l'individuo appartiene al livello più alto, tale probabilità è minore del 24,1%.

La composizione del nucleo familiare influisce sulla probabilità di non avere Internet in due direzioni opposte. Coloro che vivono in un nucleo familiare in cui sono presenti dei bambini hanno una probabilità maggiore dell'8%, rispetto a chi non vive con dei bambini, *ceteris paribus*; dall'altra parte la probabilità in esame è minore del 2,5% per chi vive in una famiglia con dei ragazzi. Probabilmente, il fatto di avere dei bambini comporta delle spese maggiori e un minore interesse nell'aver l'accesso ad Internet. In generale, con l'aumentare del numero dei componenti della famiglia la probabilità di non possedere Internet diminuisce dell'8,5%: più persone ci sono, più è probabile che qualcuna di esse decida di sottoscrivere un abbonamento Internet.

Il segno positivo della variabile Bambini sembra discordare con quello negativo della variabile Componenti, in quanto un nucleo familiare con dei bambini comprende un numero maggiore di persone; ci si aspetta, pertanto, che l'effetto delle due variabili vada nella stessa direzione. Tuttavia, analizzando il campione si nota che il 60% delle famiglie con più di due componenti non ha nemmeno un bambino e sono invece composte da ragazzi o persone adulte, che potrebbero avere maggiori necessità ad avere la connessione domestica; questo spiega il segno negativo della variabile relativa al numero dei componenti del nucleo familiare.

Infine, una persona che possiede un'auto ha una probabilità minore del 30,5% di non avere l'accesso ad Internet da casa, rispetto ad un individuo con le stesse caratteristiche che non possiede una macchina. Chi possiede una casa ha invece una probabilità mag-

giore del 12,3% di non avere Internet, rispetto a chi non possiede una casa di proprietà. Questo risultato è diverso da quello atteso, in quanto si presume che una persona che non deve pagare un affitto o un mutuo abbia una maggiore disponibilità economica e decida quindi di sottoscrivere un abbonamento Internet. Tuttavia, va ricordato che il possesso di una casa di proprietà è solo una *proxy* del livello di ricchezza.

Considerando ora le stime per l'anno 2011, si può innanzitutto notare come i segni dei coefficienti stimati siano gli stessi del 2009; tuttavia, la variabile Educ1 non è più significativa a nessun livello. Un altro elemento che cambia è il punto di svolta dell'età, ora in corrispondenza dei 34 anni. Le variabili considerate determinano, pertanto, un effetto con la medesima direzione sulla probabilità di non avere l'accesso Internet a casa nei due istanti temporali analizzati. Inoltre, per quanto riguarda gli effetti marginali delle stime per il secondo istante temporale, sono evidenti dei cambiamenti solo in alcuni valori. In particolare, come già detto, l'effetto della variabile Educ1 non più è significativo a nessun livello e cambiano gli effetti marginali delle altre dummy relative all'istruzione. Infatti, nel 2011 chi ha terminato gli studi dopo i 15 anni ha, a parità di altre condizioni, una probabilità inferiore del 24,3% di non possedere Internet a casa, rispetto a chi non ha mai studiato (a differenza di un valore pari a 16,7% ottenuto dalle stime per il 2009).

Dalle stime ottenute si può determinare il profilo tipico delle persone che non hanno Internet a casa, profilo che rimane sostanzialmente lo stesso nei due istanti temporali analizzati. In particolare, coloro che non hanno la connessione domestica sono principalmente persone anziane, che hanno terminato gli studi tra i 7 e i 15 anni, vivono in una zona rurale e appartengono al livello basso della società. Sono persone che vivono in una famiglia in cui ci sono dei bambini, ma non dei ragazzi e formata da pochi componenti. Infine, sono persone che non hanno un'auto, ma hanno una casa. Il profilo tipico del 2011 subisce solo un piccolo cambiamento: non si tratta più di persone che hanno terminato gli studi tra i 7 e i 15 anni, ma tendenzialmente di persone che non hanno alcun titolo di studio.

I profili tipici individuati corrispondono a quelli attesi. Una persona anziana e con un basso livello di istruzione è più probabile non abbia alcun motivo per avere Internet a casa. Una persona che vive in una zona non urbana e che appartiene al livello basso della società è più probabile non abbia la possibilità economica per l'installazione di Internet o abiti in una zona non coperta dalle infrastrutture necessarie. Un individuo che vive da

solo o in coppia è più probabile non abbia Internet a casa, specialmente se si tratta di una persona anziana.

Per valutare la bontà dei modelli probit si può utilizzare lo pseudo R^2 , un indicatore che vale 0 se tutte le stime dei coefficienti valgono zero (modello nullo) e vale 1 se il modello genera dei valori previsti che coincidono esattamente con quelli osservati. Nel mio caso, lo pseudo R^2 per l'anno 2009 è pari a 0,26 e quello per il 2011 è pari a 0,31. Considerando il fatto che solitamente i modelli probit presentano dei buoni adattamenti ai dati anche in presenza di valori di pseudo R^2 non molto elevati (a differenza di quanto accade nei modelli di regressione lineare), l'adattamento dei modelli stimati ai dati sembra quindi discreto; è inoltre evidente un miglioramento dal 2009 al 2011. Sembra pertanto che nel primo anno ci siano delle caratteristiche non osservabili che potrebbero spiegare la probabilità di non possedere Internet a casa.

4.2 Stima dei modelli probit multilivello

Con lo scopo di determinare se esistono diversità nel non possesso di Internet a casa a livello di Paesi, ho stimato dei modelli probit multilivello. Essi infatti permettono di tenere in considerazione la struttura gerarchica dei dati. In particolare, ho stimato dei *random intercept model*, assumendo che solamente l'intercetta vari tra i gruppi. Le variabili utilizzate sono le stesse che ho considerato nelle stime del modello probit; le stime sono riportate nella Tabella 4.3.

4. Profilo tipico dei non possessori

Variabili	2009			2011		
	$\hat{\beta}$	p-value	dy/dx	$\hat{\beta}$	p-value	dy/dx
Femmina	.0095361	0.616	.00366	-.0221781	0.274	-.0075491
Età	-3.627605	0.000	-1.392452	-3.502903	0.000	-1.191911
Età_quadrato	6.008186	0.000	2.306235	6.130004	0.000	2.085818
Educ1	-.0740997	0.538	-.0282485	-.0864083	0.581	-.0289383
Educ2	-.6137337	0.000	-.2387515	-.6615487	0.000	-.2389248
Studia_ancora	-1.120998	0.000	-.3295511	-1.058576	0.000	-.2529021
Città	-.2770607	0.000	-.104801	-.2365215	0.000	-.0790081
Grande_Città	-.5079613	0.000	-.1860742	-.5684616	0.000	-.1771214
Livello_Medio	-.3168748	0.000	-.1210106	-.3399294	0.000	-.1149597
Livello_Alto	-.5154701	0.000	-.1884857	-.5360632	0.000	-.1676414
Bambini	.2995081	0.000	.1170372	.2251261	0.000	.0793609
Ragazzi	-.0134213	0.686	-.0051444	-.081529	0.030	-.0272467
Componenti	-.3263536	0.000	-.1252704	-.3159393	0.000	-.1075026
Auto	-.6780649	0.000	-.2628571	-.8690236	0.000	-.3132562
Casa	-.0373971	0.095	-.0143574	-.0226421	0.338	-.0077059
Costante	2.207612	0.000		2.013045	0.000	
σ^2	.5694182			.5561134		
ICC	24.5%			23.6%		

Tabella 4.3: Stime dei modelli probit multilivello ed effetti marginali (2009 e 2011)

Considerando le stime per l'anno 2009, si osserva che il numero di variabili non significative è maggiore rispetto al modello probit standard. I coefficienti non significativi a nessun livello sono quelli di Femmina, Educ1 e Ragazzi; la variabile Casa è ai limiti della significatività. I segni dei coefficienti stimati sono i medesimi del modello probit standard, ad eccezione della variabile Casa. Le variabili con un effetto marginale maggiore sono di nuovo quelle relative al fatto di non aver ancora concluso gli studi, all'appartenenza al livello alto della società e al possesso di un'auto. Ad esse si aggiunge il fatto di abitare in una grande città che ha un effetto maggiore sul non possesso di Internet rispetto al modello probit standard, mentre l'appartenenza al livello alto della società ha un effetto più ridotto.

Dalle stime si nota che la probabilità di non avere la connessione alla Rete da casa tende a diminuire dai 15 ai 30 anni, per poi crescere e raggiungere il livello più alto nelle persone anziane.

Coloro che hanno una probabilità minore di non avere Internet a casa sono le persone che hanno terminato gli studi dopo i 15 anni o che stanno ancora studiando. La probabilità per una persona che ha terminato gli studi dopo i 15 anni è minore del 23,9% rispetto a quella di un individuo con le stesse caratteristiche ma senza alcun titolo di studio. Per una persona che sta ancora studiando la probabilità è minore del 33% circa.

Il fatto di risiedere in una zona urbana influenza negativamente la probabilità di non avere Internet a casa. Infatti, tenendo fisse le altre variabili, coloro che vivono in una piccola o media città hanno una probabilità inferiore del 10,5% di non possedere Internet a casa, rispetto a chi vive in zone rurali. Tale probabilità per chi vive in una grande città è inferiore del 18,6%.

La probabilità di non avere Internet diminuisce con l'aumentare del livello sociale a cui la persona appartiene, considerando fisse le altre variabili. Tale diminuzione è del 12,1% per le persone che appartengono al livello medio della società e del 18,9% per coloro che fanno parte del livello alto.

Per quanto riguarda la composizione del nucleo familiare, gli effetti delle variabili vanno di nuovo in due direzioni diverse. Il fatto di avere dei bambini nella propria famiglia, tenendo fisse le altre variabili, aumenta la probabilità di non avere l'accesso domestico alla Rete mentre al contrario, all'aumentare dei componenti del nucleo familiare la probabilità di non avere la connessione da casa diminuisce.

Infine, una persona che possiede un'auto ha una probabilità minore del 26,3% di non avere la connessione a casa, rispetto ad un'altra con le stesse caratteristiche che non ha un veicolo. Un individuo con una casa di proprietà completamente pagata ha una probabilità minore dell'1,44%. Si nota una differenza rispetto al modello probit standard, in cui l'effetto del possesso di una casa sulla probabilità di non avere Internet era significativo e positivo; tale risultato può essere dovuto al fatto di non aver considerato la struttura gerarchica nel modello precedente. Al contrario, quando considero la struttura gerarchica l'effetto diventa negativo, in quanto la distribuzione delle persone che possiedono una casa nei vari Stati è diversa.

Considerando ora le stime per l'anno 2011, si nota che le variabili non significative sono le stesse del 2009 ad eccezione della variabile Ragazzi che è significativa al 5% e di Casa che non è significativa a nessun livello.

I profili tipici delle persone che non hanno Internet a casa individuati con la stima dei modelli multilivello presentano solo qualche leggera differenza rispetto a quelli determinati con la stima dei modelli probit standard. Nel 2009, le persone senza l'accesso Internet a casa sono principalmente anziani, senza alcun titolo di studio, che vivono in zone rurali e appartengono al livello basso della società. Sono persone che vivono in una famiglia poco numerosa e con dei bambini, senza un'auto e una casa di proprietà. Nel 2011 alle caratteristiche tipiche appena presentate si aggiunge l'assenza di ragazzi nel nucleo familiare, mentre il possesso di una casa non ha alcuna rilevanza.

Per quanto concerne la variabilità tra gli Stati, si nota che l'ICC è pari al 24,5% nel 2009 e al 23,6% nel 2011. Questo indica che una parte non trascurabile della variabilità delle osservazioni è dovuta alle differenze tra i Paesi e pertanto è opportuno utilizzare i modelli multilivello e considerare la struttura gerarchica dei dati.

4.2.1 Inserimento di variabili di secondo livello

Per cercare di spiegare la variabilità tra Paesi ho inserito nel modello multilivello alcune variabili di secondo livello: la copertura DSL, il tasso di disoccupazione e l'HDI. I risultati sono riportati nella Tabella 4.4.

4. Profilo tipico dei non possessori

Variabili	2009		2011	
	$\hat{\beta}$	p-value	$\hat{\beta}$	p-value
Femmina	.009589	0.614	-.0220678	0.277
Età	-3.630751	0.000	-3.505455	0.000
Età_quadrato	6.011687	0.000	6.13382	0.000
Educ1	-.072745	0.545	-.0846754	0.588
Educ2	-.6144821	0.000	-.6613382	0.000
Studia_ancora	-1.121828	0.000	-1.058189	0.000
Città	-.2769576	0.000	-.2364811	0.000
Grande_Città	-.508006	0.000	-.56861	0.000
Livello_Medio	-.3157664	0.000	-.3385992	0.000
Livello_Alto	-.5140965	0.000	-.533902	0.000
Bambini	.2991521	0.000	.2248628	0.000
Ragazzi	-.0131794	0.691	-.0812854	0.031
Componenti	-.3265663	0.000	-.3161386	0.000
Auto	-.6754644	0.000	-.866024	0.000
Casa	-.0392912	0.080	-.0248441	0.293
Copertura_DSL	-2.636232	0.082	-1.433695	0.406
Disoccupazione	1.778007	0.379	2.451091	0.255
HDI	-4.999058	0.082	-5.600432	0.060
Costante	8.786918	0.000	7.951442	0.000
σ^2	.4155766		.4326808	
ICC	14.7%		15.8%	

Tabella 4.4: Stime dei modelli probit multilivello con variabili di livello-2

Dalle stime riportate (Tabella 4.4) si osserva che le variabili significative a livello di individuo sono le stesse del modello precedente e il loro effetto sulla variabile dipendente è uguale. Per quanto riguarda le variabili a livello di Paese, nel 2009 solo la copertura DSL e l'HDI sono significativi al livello del 10%. L'effetto di tali caratteristiche sulla probabilità di non possedere Internet a casa è negativo. Infatti, una persona che risiede in uno Stato con una copertura DSL maggiore ha una probabilità minore di non avere la connessione domestica rispetto ad una persona con le medesime caratteristiche che risiede in un Paese con una copertura più bassa. Anche coloro che abitano in uno Stato con HDI più elevato hanno una probabilità minore di non avere Internet. Al contrario, il fatto di risiedere in uno Stato con tasso di disoccupazione minore non ha alcun effetto sulla probabilità in esame. L'ICC mostra che, nonostante la non elevata significatività, le va-

riabili inserite aiutano a spiegare la variabilità nel non possesso di Internet a casa esistenti tra i Paesi: il suo valore è diminuito infatti di quasi 10 pp.

Per quanto concerne il 2011, solo il coefficiente relativo all'HDI è significativo al 10% e ha un effetto negativo sulla probabilità di non avere la connessione Internet a casa. Il fatto che le altre variabili a livello di Paese non siano significative indica che nel 2011 le differenze nella probabilità di non avere Internet tra persone che appartengono a Stati diversi non sono più dovute a differenze nella copertura DSL. Sembra quindi che le infrastrutture Internet abbiano avuto un significativo miglioramento, come dimostrato anche nelle precedenti analisi esplorative, e non siano più motivo di diversità tra gli Stati. L'ICC è diminuito di quasi 8 pp; pertanto le variabili di secondo livello inserite spiegano meglio la variabilità tra gli Stati nel 2009, piuttosto che nel 2011.

Allo stesso tempo però, questi risultati indicano che è ancora presente una buona proporzione di variabilità totale ascrivibile a variabilità tra Paesi (mentre senza variabili di secondo livello c'era un ICC maggiore nel 2009).

Con l'inserimento delle variabili di secondo livello vengono aggiunte due nuove caratteristiche al profilo tipico delle persone che non possiedono Internet a casa nel 2009: esse risiedono in particolare in Paesi con copertura DSL e HDI minori. Per il 2011, solo il fatto di vivere in uno Stato con HDI più basso va aggiunto al profilo tipico, mentre la copertura DSL non determina differenze tra gli Stati.

4.2.2 Analisi dei residui

Per studiare ulteriormente la variabilità tra gli Stati si possono analizzare i residui a livello 2 che stimano l'efficacia dei vari gruppi. Per fare ciò si può utilizzare l'approccio suggerito da Goldstein e Healy (1995). Usando una notazione analoga a quella vista nel Capitolo 3 dove sono stati introdotti i modelli random intercept, dato $u_{0j} \sim N(0, T_0^2)$ il termine d'errore a livello di gruppo, si considerano le relative stime \hat{u}_{0j} e sfruttando gli *standard error* delle stime si costruiscono degli intervalli di confidenza al 5% di livello di significatività. Se gli intervalli di confidenza di due residui non si sovrappongono, i due residui sono statisticamente differenti e gli effetti dei due gruppi sono statisticamente diversi. Si possono così confrontare i diversi Paesi per vedere quali hanno una proba-

bilità maggiore che i loro cittadini non abbiano Internet a casa nei due istanti temporali. Le Figure 4.1 e 4.2 riportano gli intervalli di confidenza per i *value-added residuals*, ossia i residui aggiustati per le caratteristiche delle unità che compongono il gruppo (cioè i valori dopo la stima del modello probit multilivello riportato nella Tabella 4.3).

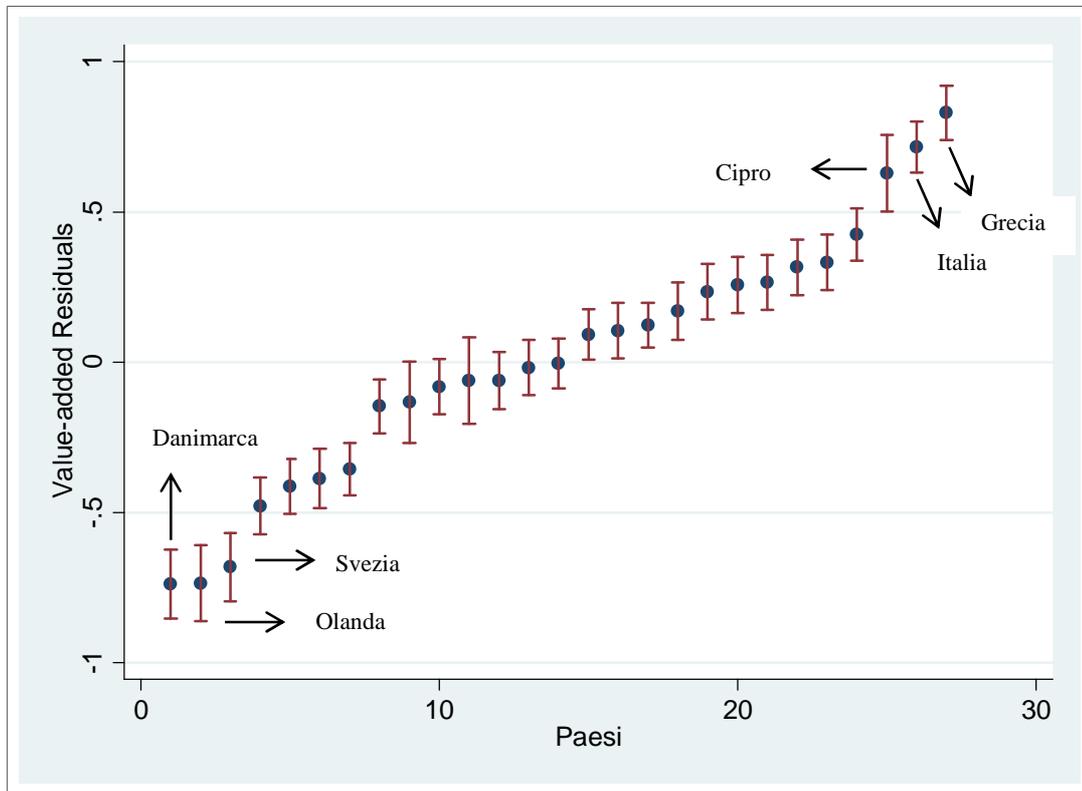


Figura 4.1: Intervalli di confidenza per i residui a livello 2 (2009)

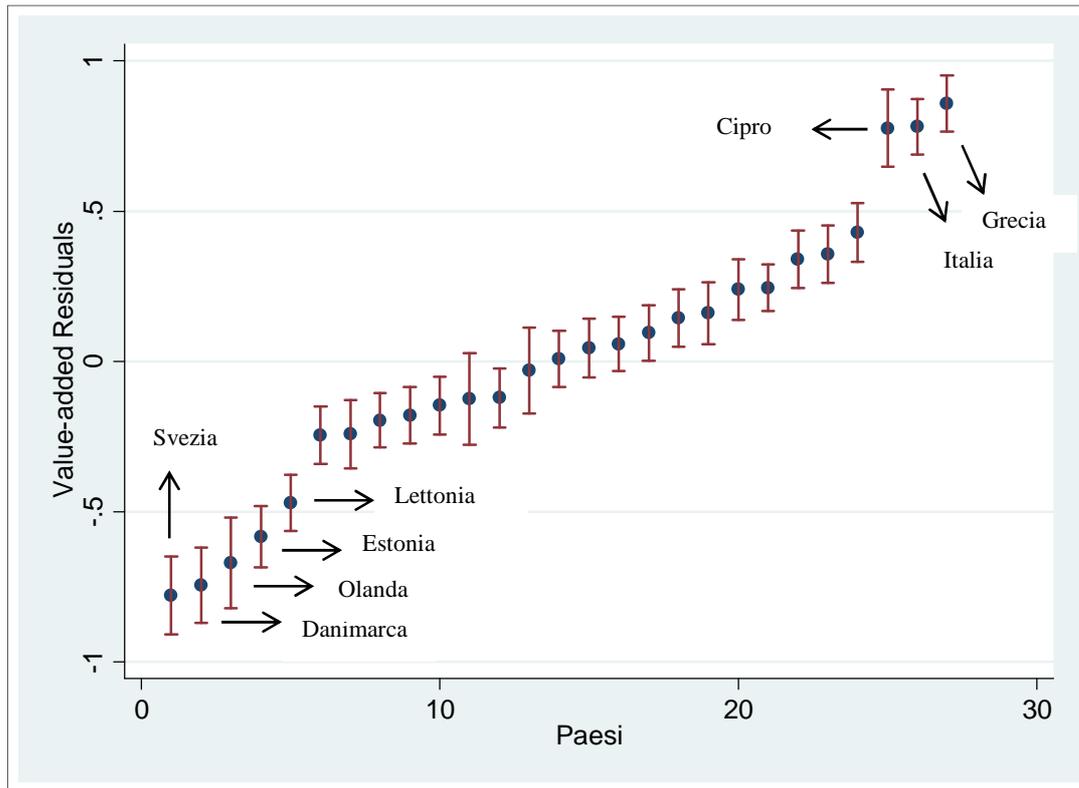


Figura 4.2: Intervalli di confidenza per i residui a livello 2 (2011)

Dal grafico dei residui del 2009 si può osservare che a parità di altre condizioni, i cittadini europei con una maggiore probabilità di non avere Internet a casa risiedono in Grecia, Italia e Cipro, mentre quelli con una probabilità minore risiedono in Danimarca, Olanda e Svezia.

Per quanto riguarda il 2011, gli Stati i cui cittadini hanno una probabilità più bassa di non avere Internet a casa sono ancora Svezia, Danimarca e Olanda, anche se l'ordine è leggermente diverso, a cui si aggiungono Estonia e Lettonia. I Paesi in cui abitano le persone con una probabilità più elevata di non avere Internet a casa sono nuovamente Cipro, Italia e Grecia.

I risultati ottenuti sono diversi rispetto a quelli determinati con le analisi esplorative. In particolare, nel Capitolo 2 era emerso che i Paesi con una percentuale minore di persone che non possiedono la connessione Internet a casa sono in entrambi gli istanti temporali Olanda, Danimarca, Svezia; al contrario, quelli con una percentuale maggiore sono Grecia, Portogallo, Romania e Bulgaria. Occorre però ricordare che i residui che ho conside-

rato sono aggiustati per le caratteristiche delle unità che compongono i gruppi dei vari campioni.

È inoltre evidente la sovrapposizione dei residui, in particolare nella parte centrale di ciascun grafico. Questo indica che gli effetti dei vari Paesi i cui residui si sovrappongono non sono statisticamente diversi. Si possono così confrontare solamente gli effetti di Paesi i cui residui sono nettamente separati. Tuttavia, i residui centrali sono statisticamente diversi da quelli più bassi e da quelli più alti; sembra ci siano dei gruppi composti da Stati che non differiscono per quanto riguarda la probabilità di non avere Internet.

Con la successiva analisi a classi latenti sarà possibile determinare il motivo per cui le persone non hanno la connessione domestica e, con l'approccio multilivello, segmentare i diversi Paesi sulla base di tale motivazione.

Capitolo 5

Analisi a classi latenti

In questo capitolo presenterò i risultati ottenuti con l'analisi a classi latenti utilizzata per segmentare i non possessori di Internet a casa, sulla base della motivazione di questa mancanza. Inizialmente presenterò i risultati un po' deludenti ottenuti con il software *Stata*, per poi esaminare quelli ottenuti con *Latent Gold*. Per quanto concerne l'analisi a classi latenti effettuata con *Stata*, ho stimato dei modelli senza covariate e senza considerare la struttura gerarchica dei dati. Con *Latent Gold* è stato invece possibile inserire delle covariate e utilizzare l'approccio multilivello. Ricordo che le osservazioni utilizzate sono quelle del sottocampione costituito da coloro che non possiedono la connessione Internet a casa.

5.1 Software utilizzati

Il principale programma statistico che ho utilizzato per le mie analisi è *Stata 12*¹². In seguito, non soddisfatta dei risultati ottenuti per quanto riguarda i modelli a classi latenti, ho utilizzato il software *Latent Gold 5.0*¹³ e confrontato i risultati ottenuti con i due programmi.

¹² <http://www.stata.com>

¹³ <http://statisticalinnovations.com/products/latentgold.html>

Stata è un programma statistico molto utilizzato in discipline come economia, sociologia, psicologia, biostatistica ed epidemiologia. Permette di applicare un vasto insieme di metodi statistici e consente di creare, modificare ed elaborare dataset molto ampi. È un programma semplice da utilizzare, anche per gli utenti meno esperti, grazie ad una grafica molto intuitiva. Permette di svolgere analisi complete, con l'utilizzo di numerosi tipi di grafici e il calcolo di statistiche riassuntive. Consente, inoltre, la stima di una vasta gamma di modelli, come modelli lineari, modelli per dati longitudinali, modelli multilivello e ad effetti misti, modelli per variabili binarie, di conteggio e limitate, modelli lineari generalizzati, modelli per serie storiche, modelli fattoriali, etc.

Uno dei limiti che però ho riscontrato riguarda l'analisi a classi latenti, in quanto non esiste una procedura che permetta di stimare in modo semplice e veloce una tale classe di modelli. È stato creato un modulo, chiamato GLLAMM (Rabe-Hesketh et al. 2004), che permette di stimare modelli multilivello con variabili latenti per diversi tipi di variabili risposta, ma, in presenza di dataset molto ampi, risulta essere davvero molto lento e piuttosto complicato.

Recentemente, è stato sviluppato un plugin¹⁴ per *Stata* con lo scopo di stimare dei modelli a classi latenti anche per dati gerarchici. Tale plugin è molto facile da utilizzare e richiede poco tempo. Tuttavia, se nel modello vengono inserite alcune covariate il programma non mostra in output gli indicatori utili per la scelta del numero ottimale di classi latenti e spesso presenta delle difficoltà computazionali. Inoltre, pur permettendo di considerare la struttura gerarchica dei dati, non inserisce una variabile latente anche a livello di Paese e pertanto non permette di suddividere gli Stati considerati in gruppi, ma solo di ottenere stime separate per ciascuno di essi.

Al contrario, *Latent Gold* è un software statistico dedicato all'analisi a classi latenti tradizionale e multilivello. Il progetto alla base del programma è iniziato nel 1998 in Olanda e la prima versione del software, la 2.0, è stata rilasciata nel 2000. *Latent Gold* presenta un'interfaccia molto intuitiva e necessita di poco tempo per la stima dei modelli. Inoltre, possono essere utilizzati diversi tipi di variabili, vengono mostrati in output diversi indicatori utili per la scelta del numero di classi e i residui bivariati per verificare la presenza di dipendenza locale. Esistono due versioni del programma: una base e una

¹⁴ Latent Class Analysis Stata Plugin: <http://methodology.psu.edu/downloads/lcastata>

avanzata; quest'ultima permette di stimare modelli con variabili a classi latenti continue e modelli per dati gerarchici.

5.2 **Analisi a classi latenti con *Stata***

Per l'analisi a classi latenti ho utilizzato come indicatori 9 variabili dicotomiche relative alla motivazione del non possesso di Internet a casa:

- y1: intenzione ad abbonarsi nei prossimi sei mesi
- y2: poca conoscenza di Internet
- y3: mancanza di interesse
- y4: mancanza della copertura a banda larga
- y5: costo per l'installazione iniziale troppo elevato
- y6: costo eccessivo per l'abbonamento mensile
- y7: costo per l'acquisto di un pc e di un modem troppo elevato
- y8: accesso ad Internet in altri luoghi (scuola, lavoro, biblioteca, etc.)
- y9: preoccupazione per l'accesso a siti non adatti.

Come detto all'inizio del capitolo, *Stata* presenta dei limiti per quanto concerne l'analisi a classi latenti. Per questo motivo ho stimato inizialmente dei modelli con un numero di classi latenti da 1 a 8, senza alcuna covariata e senza considerare la struttura gerarchica dei dati. L'output del plugin LCA mostra solamente alcuni indici (AIC, BIC e CAIC) che ho riportato nelle successive tabelle di sintesi. Vengono inoltre evidenziati il valore dell'entropy-based R^2 , il numero dei parametri stimati e i gradi di libertà di ciascun modello.

5.2.1 Analisi a classi latenti per il 2009 senza covariate

N classi	BIC (LL)	CAIC (LL)	AIC (LL)	R2 Entr	N param	df
1	9050,5441	9059,5441	8985,1020	1,0000	9	502
2	5302,0410	5321,0410	5163,8855	0,7697	19	492
3	2095,8743	2124,8743	1885,0054	0,8846	29	482
4	1517,6386	1556,6386	1234,0562	0,9062	39	472
5	1540,6731	1589,6731	1184,3774	0,9015	49	462
6	1295,7604	1354,7604	866,75124	0,8270	59	452
7	1386,4835	1455,4835	884,76092	0,6855	69	442

Tabella 5.1: Statistiche di sintesi dei modelli a classi latenti senza covariate secondo *Stata* (2009)

Osservando la Tabella 5.1 che riporta le statistiche riassuntive per il 2009, si nota come tutti i tre indicatori (BIC CAIC e AIC) suggeriscano di utilizzare un modello con sei classi latenti. Un'altra misura utile è l'entropy-based R^2 che indica quanto bene il modello riesce a prevedere l'appartenenza alle varie classi a partire dalle risposte osservate. Un valore pari a 0 indica che non c'è separazione tra le classi individuate, mentre un valore pari a 1 indica che c'è una perfetta separazione. Il valore assunto dalla statistica nel modello a sei classi latenti è abbastanza elevato, anche se nel modello con 5 classi è maggiore. Ho tuttavia deciso di considerare il modello con sei classi latenti, come indicato dai tre criteri informativi.

Descrizione e interpretazione delle classi

La Tabella 5.2 mostra l'ampiezza delle sei classi individuate e le probabilità di risposta condizionate alla classe di appartenenza. Confrontando tali probabilità, è possibile attribuire un nome e dare una interpretazione a ciascuna classe.

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6
Ampiezza	1,42%	7,36%	50,11%	10,35%	12,62%	18,14%
y1	0,01%	99,92%	0,00%	0,00%	0,00%	0,00%
y2	6,25%	0,00%	6,12%	9,46%	0,08%	25,19%
y3	38,87%	0,01%	99,99%	50,20%	4,75%	2,51%
y4	6,91%	0,00%	0,31%	2,61%	3,10%	7,73%
y5	97,49%	0,00%	0,20%	16,71%	18,15%	2,03%
y6	99,23%	0,00%	0,04%	53,09%	60,83%	0,20%
y7	85,67%	0,00%	0,37%	49,25%	33,52%	0,26%
y8	10,64%	0,00%	0,48%	2,16%	18,27%	28,20%
y9	5,78%	0,00%	0,30%	2,16%	3,16%	3,78%

Tabella 5.2: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello a classi latenti senza covariate secondo *Stata* (2009)

La classe 1 è quella meno numerosa: comprende solo l'1,4% delle persone che hanno affermato di non possedere Internet a casa nel 2009. Questa classe è costituita in particolare da coloro che ritengono troppo elevato il costo per l'installazione iniziale della rete a banda larga, quello per l'abbonamento mensile e quello per l'acquisto di un pc o di un modem. È costituita da persone con una probabilità elevata di essere interessate ad avere Internet a casa, anche se vi è una probabilità non trascurabile di mancanza d'interesse a questo prodotto. Questo segmento è pertanto quello dei “non possessori senza la possibilità economica per l'installazione e utilizzo di Internet”.

La seconda classe, che comprende il 7,4% dei non possessori, è costituita in particolare da coloro che al momento dell'indagine non possiedono Internet, ma sono intenzionati ad abbonarsi nei sei mesi successivi. Questo è il segmento degli “intenzionati ad abbonarsi”.

La classe 3 è quella più numerosa: comprende circa la metà degli intervistati senza Internet ed è costituita da coloro che presentano una probabilità più elevata di non essere interessati ad avere la connessione Internet a casa. Pertanto, questo segmento raggruppa i “non interessati”.

La classe 4 è formata dal 10,4% dei non possessori e presenta probabilità condizionate importanti per vari indicatori. In particolare, questa classe è caratterizzata da non possessori con una probabilità rilevante di riscontrare difficoltà nel pagamento dell'abbonamento mensile e nell'acquisto di un computer e/o di un modem. Inoltre, la

probabilità relativa alla mancanza di interesse per il prodotto in esame è pari al 50% circa. Si tratta della classe dei “non possessori senza la possibilità economica per l’abbonamento mensile e l’acquisto di un pc e/o modem”.

La quinta classe è costituita dal 12,6% dei non possessori e anche in questo caso la probabilità più elevata è quella relativa all’indicatore y6 (costo eccessivo dell’abbonamento mensile). Questa classe è costituita prevalentemente da persone interessate ad Internet ma che non lo possiedono in quanto l’abbonamento mensile è troppo elevato. Rispetto alla classe precedente vi è una probabilità meno elevata di appartenere a questa classe a causa dell’eccessivo costo di pc e/o modem. Tale segmento è quello degli “interessati, ma senza la possibilità economica per l’abbonamento mensile e l’acquisto di un pc e/o modem” e si differenzia dalla classe 4 soprattutto per la maggiore probabilità che i componenti siano interessati al possesso di Internet.

La classe 6 è la seconda per numerosità: contiene il 18,1% degli intervistati senza Internet. Questa classe presenta probabilità condizionate basse per tutti gli indicatori; tuttavia, le più elevate sono quelle relative a y2 e y8. Questo segmento è costituito prevalentemente da persone che non hanno dato una motivazione specifica e da coloro che hanno risposto di non conoscere Internet o di accedervi da altri luoghi. Si tratta perciò della classe dei “non possessori senza una motivazione precisa”.

5.2.2 Analisi a classi latenti per il 2011 senza covariate

N classi	BIC (LL)	CAIC (LL)	AIC (LL)	R ² Entr	N param	df
1	6391,8306	6400,8306	6328,2949	1,0000	9	502
2	2491,1772	2510,1772	2357,0463	0,8214	19	492
3	1130,8156	1159,8156	926,0896	0,8637	29	482
4	1160,0886	1199,0886	884,7674	0,8943	39	472
5	941,9690	990,9690	596,0527	0,8804	49	462
6	972,02655	1031,0265	555,51502	0,8877	59	452
7	1057,5545	1126,5545	570,44777	0,7145	69	442

Tabella 5.3: Statistiche di sintesi dei modelli a classi latenti senza covariate secondo *Stata* (2011)

Dalla Tabella 5.3 si nota che i criteri BIC e CAIC suggeriscono un modello a 5 classi latenti, a differenza dell'indice AIC che suggerisce un modello con 6 classi. Per quanto riguarda l'entropy-based R^2 , la differenza tra i due modelli è minima. Ho quindi deciso di considerare l'indice BIC, che è quello più utilizzato in queste analisi, e scegliere il modello con 5 classi latenti.

Descrizione e interpretazione delle classi

	CL 1	CL 2	CL 3	CL 4	CL 5
Ampiezza	1,13%	5,37%	56,20%	17,25%	20,06%
y1	13,07%	99,62%	0,00%	0,00%	0,22%
y2	24,70%	0,01%	5,22%	0,06%	17,03%
y3	54,12%	1,88%	99,99%	20,38%	1,70%
y4	9,49%	0,56%	0,17%	3,21%	5,54%
y5	34,19%	0,36%	0,53%	24,79%	2,41%
y6	90,80%	1,55%	1,17%	74,51%	5,77%
y7	64,20%	0,53%	0,92%	45,36%	7,67%
y8	21,81%	1,84%	0,45%	4,75%	20,70%
y9	23,58%	0,69%	0,36%	0,33%	3,90%

Tabella 5.4: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello a classi latenti senza covariate secondo *Stata* (2011)

La classe 1 è quella meno numerosa: contiene solo l'1,1% degli intervistati del sottocampione. È costituita da persone con una probabilità più alta di ritenere troppo elevato l'abbonamento mensile ad Internet e il costo per l'acquisto del pc e modem necessari. Tuttavia, si nota anche che la probabilità condizionata relativa alla mancanza di interesse ad avere Internet a casa è di poco superiore al 50%. Questo segmento è quello dei "non possessori senza la possibilità economica per l'installazione e l'utilizzo di Internet".

La classe 2 è costituita dal 5,4% delle persone considerate ed è formata da coloro che presentano una probabilità più alta di essere in attesa di abbonarsi. Pertanto, questo è il segmento degli "intenzionati ad abbonarsi".

La terza classe è quella più numerosa: comprende oltre la metà dei non possessori ed è costituita principalmente dalle persone che non sono interessate ad avere Internet a casa. È la classe dei "non interessati".

La classe 4 comprende il 17,3% degli intervistati senza Internet a casa ed è costituita soprattutto dalle persone che considerano troppo elevato l'abbonamento mensile ad Internet e il costo per l'acquisto di un pc e/o modem. A differenza della prima classe, la classe 4 è caratterizzata da persone con una probabilità minore di non essere interessate ad Internet. Si tratta pertanto della classe degli "interessati, ma senza la possibilità economica per l'installazione e l'utilizzo di Internet".

La quinta classe è la seconda per numerosità: raccoglie circa il 20% dei non possessori. Le probabilità condizionate relative a tale classe sono tutte basse. Pertanto, essa è costituita da coloro che non hanno dato una motivazione precisa al fatto di non avere Internet a casa. Tuttavia, si notano delle probabilità condizionate un po' più elevate per y2 e y8. Questo segmento è quello dei "non possessori senza una motivazione precisa".

Confronto con il 2009

Dai risultati ottenuti emerge come in entrambi gli anni analizzati sia possibile suddividere i non possessori in classi aventi delle "etichette" uguali.

In particolare, in entrambi gli istanti temporali analizzati la classe più ampia è quella dei "non interessati", classe che nel 2011 comprende una maggiore percentuale di non possessori. La seconda classe per numerosità è quella costituita dai "non possessori senza una precisa motivazione", mentre le classi con numerosità più esigua sono quelle dei "non possessori senza la disponibilità economica per l'abbonamento mensile e l'acquisto di un pc e/o modem", e gli "intenzionati ad abbonarsi". Si può notare come la classe costituita dalle persone in attesa di installare Internet abbia un'ampiezza minore nel 2011. Tale cambiamento potrebbe essere collegato alla maggiore diffusione di Internet raggiunta dopo il 2009; pertanto, chi era interessato ad Internet e non aveva problemi economici o di altro tipo, l'ha installato prima del 2011. Questa è solo un'ipotesi dal momento che le persone intervistate nei due istanti temporali non sono le stesse.

Nei due modelli stimati sono presenti delle differenze per quanto riguarda le difficoltà economiche riscontrate dalle persone interessate e non interessate ad avere l'accesso a Internet a casa. Tuttavia, da entrambi emerge come le spese da sostenere per l'installazione iniziale, per acquistare gli strumenti necessari e per l'abbonamento mensile siano considerate proibitive.

Per ottenere maggiori informazioni sulle caratteristiche delle diverse classi, ho aggiunto come covariate le stesse variabili considerate nella stima dei modelli probit. Tuttavia, *Stata* riscontra dei problemi nel calcolo della log-verosimiglianza e non mostra in output nessun indicatore utile per la scelta del numero di classi; per questi motivi ho deciso di utilizzare il software *Latent Gold* per queste ulteriori analisi.

5.3 Analisi a classi latenti con *Latent Gold*

Come nel paragrafo precedente, ho stimato inizialmente per ciascun anno dei modelli a classi latenti senza covariate e senza considerare la struttura gerarchica dei dati, con un numero di classi compreso tra 1 e 8. Le Tabelle 5.5 e 5.7 riportano gli stessi indicatori di bontà di adattamento del modello del paragrafo precedente, con l'aggiunta della statistica rapporto di verosimiglianza L^2 e il criterio AIC3 al posto di AIC.

5.3.1 Analisi a classi latenti per il 2009 senza covariate

N classi	BIC (LL)	CAIC (LL)	AIC3 (LL)	L^2	p value	R ² Entr	N param	df
1	57341,466	57350,466	57285,024	8967,102	1,7e-1527	1,0000	9	502
2	52807,955	52826,955	52688,8	4340,8773	2,3e-606	0,8183	19	492
3	50386,787	50415,787	50204,918	1826,9955	2,30e-155	0,8770	29	482
4	49527,419	49566,419	49282,837	874,9144	1,70e-26	0,8626	39	472
5	49167,021	49216,021	48859,725	421,8026	0,91	0,8748	49	462
6	49117,53	49176,53	48747,52	279,5985	1	0,8552	59	452
7	49175,526	49244,526	48742,804	244,8815	1	0,8642	69	442
8	49198,203	49277,203	48702,767	174,845	1	0,8051	79	432

Tabella 5.5: Statistiche di sintesi per il modello a classi latenti senza covariate secondo *Latent Gold* (2009)

I criteri BIC e AIC riportati nella Tabella 5.5 suggeriscono di utilizzare un modello con 6 classi latenti, mentre il criterio AIC3 assume valore minore con 8 classi latenti. Per evitare problemi di interpretabilità dovuti ad un'ampiezza esigua della classi, ho deciso di considerare il modello con 6 classi, che presenta in ogni caso un AIC3 di poco supe-

riore. Un'altra statistica da prendere in considerazione è l'entropy-based R^2 che è maggiore nel modello con 6 classi rispetto a quello con 8 classi.

Le stime ottenute per i parametri relativi agli indicatori sono tutte significative al livello del 5%. Per quanto concerne le intercette, esse sono significative al 5% ad eccezione di quella relativa all'indicatore y9, significativa al 10%, e quella di y3 che non è significativa a nessun livello. Tuttavia, i coefficienti più importanti sono quelli relativi all'effetto della variabile latente sugli indicatori e pertanto il fatto che due intercette non siano significative non causa dei problemi alla bontà del modello.

Per quanto riguarda la bontà di adattamento del modello, l' R^2 relativo all'indicatore y4 è pari a 0,06 e quello relativo a y9 è pari a 0,03; pertanto il modello non spiega bene tali variabili. Questo problema potrebbe essere causato dalla bassa frequenza di risposte registrata da queste due opzioni di scelta.

Descrizione e interpretazione delle classi

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6
Ampiezza	48,79%	18,03%	12,88%	7,78%	7,38%	5,14%
y1	0,00%	0,00%	0,00%	0,00%	99,71%	0,00%
y2	0,65%	0,04%	5,02%	99,42%	0,00%	0,03%
y3	99,98%	4,65%	31,72%	39,79%	0,01%	2,15%
y4	0,28%	10,05%	2,10%	1,25%	0,00%	0,00%
y5	0,36%	7,87%	32,64%	1,01%	0,00%	0,02%
y6	0,50%	20,04%	81,26%	1,13%	0,00%	4,49%
y7	1,61%	13,97%	55,93%	1,76%	0,00%	2,66%
y8	0,41%	7,99%	9,34%	1,01%	0,00%	99,42%
y9	0,28%	5,36%	2,74%	0,46%	0,00%	0,95%

Tabella 5.6: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello senza covariate secondo *Latent Gold* (2009)

Nonostante la mancanza di covariate nel modello, è possibile determinare ex-post le principali caratteristiche delle persone che compongono le diverse classi identificate analizzando alcune delle variabili presentate nel Capitolo 2.

Sulla base dei risultati della Tabella 5.6, la classe 1 è quella più numerosa e comprende quasi il 50% delle persone che non hanno Internet a casa. È caratterizzata da non posses-

sori che presentano una probabilità elevata di non essere interessati alla connessione da casa e corrisponde pertanto alla classe dei “non interessati”. Sono tendenzialmente persone con più di 50 anni di età, che hanno terminato gli studi tra i 7 e i 15 anni o successivamente e vivono in una zona rurale. Si tratta di persone che vivono soprattutto in famiglie composte da due componenti, appartengono al livello medio-basso della società e possiedono una casa.

La classe 2 è la seconda per numerosità e contiene circa il 18% dei non possessori. È caratterizzata da probabilità basse per tutti gli indicatori ed è pertanto costituita dai “non possessori senza una precisa motivazione”. Si nota che tale segmento, rispetto agli altri, ha una probabilità condizionata maggiore per y4 e y9. Sembra che le persone che hanno risposto di vivere in una zona non coperta dalla banda larga e coloro che sono preoccupati per l’accesso a siti non adatti siano inclusi in questo segmento. I componenti di questo segmento sono principalmente giovani con meno di 24 anni che hanno terminato gli studi dopo i 15 anni. Si tratta di persone che abitano tendenzialmente in zone rurali o in piccole o medie città (aspetto che potrebbe spiegare la mancata copertura della banda larga), vivono in famiglie numerose e appartengono al livello medio-basso della società.

La terza classe comprende circa il 13% delle persone intervistate senza Internet a casa e si caratterizza per l’elevata probabilità di riscontrare difficoltà economiche. Pertanto, questa classe corrisponde a quella dei “non possessori senza la possibilità economica per l’installazione e utilizzo di Internet”. Coloro che rientrano in tale segmento sono principalmente giovani con meno di 24 anni, che hanno terminato gli studi dopo i 15 anni e risiedono in zone rurali. Per quanto concerne la composizione del nucleo familiare, si tratta di persone che vivono da sole o con un’altra persona. Infine, quasi la metà dei componenti appartiene al livello basso della società.

La classe 4, che comprende il 7,8% dei non possessori, è costituita soprattutto dalle persone che non sanno esattamente cosa sia Internet. Questo segmento, non individuato con *Stata*, può essere definito come quello dei “non conoscitori di Internet”. Le persone che appartengono a questo segmento hanno tendenzialmente più di 60 anni e hanno terminato gli studi tra i 7 e i 15 anni. Più della metà dei componenti abita in una zona rurale e quasi la totalità vive da solo o con un’altra persona. Infine, sono persone che appartengono al livello basso della società, che non possiedono un’auto, ma una casa.

La classe 5 raggruppa il 7,4% delle persone del sottocampione ed è costituita da non possessori con una probabilità più elevata di voler abbonarsi a Internet nei mesi successivi all'indagine. Questa classe corrisponde a quella degli "intenzionati ad abbonarsi". I componenti sono soprattutto giovani con meno di 24 anni, che hanno terminato gli studi dopo i 15 anni. Si tratta di persone che abitano in particolare in zone rurali e in piccole o medie città, vivono da sole o in famiglie numerose, appartengono al livello medio della società e possiedono un'auto.

L'ultima classe è la meno numerosa e comprende circa il 5% dei non possessori. È costituita da persone con un'elevata probabilità di accedere ad Internet in altri luoghi, come scuole, biblioteche, etc. Questo segmento, non individuato con *Stata*, può essere definito come quello degli "aventi accesso in altri luoghi" ed è costituito quindi da persone che utilizzano già Internet, ma non hanno la necessità e l'interesse di utilizzarlo a casa. Coloro che rientrano in questa classe hanno in particolare meno di 40 anni e hanno terminato gli studi dopo i 15 anni. Sono principalmente persone che abitano in zone non urbane e in piccole o medie città, vivono da sole o in famiglie numerose e appartengono al livello medio della società. Si tratta soprattutto di persone che possiedono un'auto e una casa.

5.3.2 Analisi a classi latenti per il 2011 senza covariate

N classi	BIC (LL)	CAIC (LL)	AIC3 (LL)	L ²	p value	R ² Entr	N param	df
1	43163,067	43172,067	43108,532	6310,2949	1,0e-988	1,0000	9	502
2	39262,408	39281,408	39147,277	2319,0403	5,50e-234	0,8210	19	492
3	37902,049	37931,049	37726,323	868,0866	1,70e-24	0,8488	29	482
4	37690,192	37729,192	37453,871	565,6343	0,0019	0,8352	39	472
5	37626,54	37675,54	37280,624	411,3873	0,96	0,8458	49	462
6	37623,28	37682,28	37265,77	317,5287	1	0,7883	59	452
7	37652,345	37721,345	37234,239	256,0018	1	0,8067	69	442
8	37703,512	37782,512	37224,81	216,5729	1	0,7255	79	432

Tabella 5.7: Statistiche di sintesi per il modello a classi latenti senza covariate secondo *Latent Gold* (2011)

Anche per le analisi del 2011 il criterio AIC3 (Tabella 5.7) suggerisce un modello con 8 classi latenti, mentre BIC e CAIC suggeriscono un modello con 6 classi latenti.

L'entropy-based R^2 è ancora una volta maggiore nel modello con 6 classi rispetto a quello con 8 classi. Per questi motivi e per evitare problemi legati a classi poco numerose, ho deciso di utilizzare il modello con 6 classi.

Le stime dei coefficienti relativi agli indicatori sono tutte significative, mentre le inter-cette di y_3 e y_8 non lo sono. L' R^2 associato ai diversi indicatori mostra che il modello non spiega bene le variabili y_4 ($R^2 = 0,04$) e y_9 ($R^2 = 0,02$), proprio come riscontrato nel modello per il 2009.

Descrizione e interpretazione delle classi

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6
Ampiezza	53,77%	20,21%	11,49%	5,41%	4,82%	4,29%
y1	0,00%	0,26%	1,15%	99,14%	4,10%	2,82%
y2	5,21%	1,43%	28,76%	0,02%	1,06%	4,48%
y3	99,96%	25,90%	3,85%	1,85%	4,70%	23,35%
y4	0,08%	3,00%	7,82%	0,65%	1,44%	5,65%
y5	0,37%	12,76%	2,10%	0,40%	1,16%	55,17%
y6	0,12%	53,31%	2,65%	1,62%	8,61%	96,26%
y7	0,29%	30,94%	5,38%	0,78%	5,07%	77,36%
y8	0,27%	0,01%	0,05%	0,02%	97,96%	16,22%
y9	0,31%	1,30%	5,02%	0,95%	2,70%	3,73%

Tabella 5.8: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello senza covariate secondo *Latent Gold* (2011)

La classe 1 comprende oltre la metà delle persone che non hanno Internet a casa ed è quella dei “non interessati” (Tabella 5.8). Si tratta principalmente di persone con più di 55 anni d'età, che hanno terminato gli studi tra i 7 e i 15 anni o successivamente, risiedono in zone rurali e in piccole o medie città. La maggior parte dei componenti di questa classe vive da solo o con un'altra persona, appartiene al livello medio-basso della società e possiede una casa.

La classe 2, che comprende il 20,2% dei non possessori, è quella dei “non possessori senza la possibilità economica per l'abbonamento mensile e l'acquisto di un pc e/o modem”. È costituita tendenzialmente da giovani con meno di 27 anni, che hanno concluso gli studi dopo i 15 anni. Si tratta in particolare di persone che risiedono in zone rurali e

in piccole o medie città, vivono da sole o con un'altra persona e appartengono al livello basso della società.

La terza classe raggruppa l'11,5% degli intervistati senza Internet ed è quella dei "non possessori senza una motivazione precisa". In particolare, le probabilità più elevate sono quelle relative a coloro che non conoscono Internet e a coloro che vivono in zone non coperte dalla banda larga. Questo segmento è costituito principalmente da giovani con meno di 27 anni, che hanno finito gli studi dopo i 15 anni, risiedono in zone rurali e in piccole o medie città, vivono da soli o con un'altra persona e appartengono al livello medio-basso della società.

La classe 4 comprende il 5,4% delle persone del sottocampione in esame ed è quella degli "intenzionati ad abbonarsi". Sono principalmente giovani con meno di 27 anni, che hanno terminato gli studi dopo i 15 anni. Si tratta di persone che risiedono in una zona rurale e in una piccola o media città, vivono famiglie composte da due componenti e appartengono al livello medio della società.

La quinta classe, che comprende il 4,8% dei non possessori di Internet, è quella degli "aventi accesso in altri luoghi". Sono tendenzialmente giovani con meno di 27 anni che hanno terminato gli studi dopo i 15 anni. Circa il 45% dei componenti abita in una piccola o media città e il 32,9% in una zona non urbana. Sono persone che appartengono principalmente ad una famiglia costituita da 2 componenti, fanno parte del livello basso della società e possiedono un'auto.

L'ultima classe è quella dei "non possessori senza la possibilità economica per l'installazione e l'utilizzo di Internet". È costituita in particolare da giovani con meno di 27 anni, che hanno terminato gli studi dopo i 15 anni e risiedono in una zona rurale. Sono tendenzialmente persone che vivono da sole o con un'altra persona, appartengono al livello basso della società e solo il 37,9% possiede un'auto.

5.3.3 Analisi a classi latenti per il 2009 con le covariate

Per determinare l'effetto di alcune covariate sulla predizione dei segmenti ho inserito nei modelli le medesime variabili utilizzate nei modelli probit standard, ad eccezione di

Età_quad. Ho stimato dei modelli con un numero di classi compreso tra 1 e 8; la Tabella 5.9 mostra i risultati ottenuti.

N classi	BIC (LL)	CAIC (LL)	AIC3 (LL)	L ²	p value	R ² Entr	N param	df
1	57341,47	57350,47	57285,02	51594,66	6,7e-5256	1,000	9	10620
2	50790,92	50825,92	50571,43	44803,06	1,1e-4114	0,761	35	10594
3	48475,99	48536,99	48093,44	42247,07	1,5e-3702	0,813	61	10568
4	47226,51	47313,51	46680,91	40756,54	6,3e-3469	0,828	87	10542
5	46414,67	46527,67	45706,06	39703,64	1,1e-3307	0,797	113	10516
6	46416,25	46555,25	45544,53	39464,16	5,7e-3277	0,788	139	10490
7	46352,29	46517,29	45317,52	39159,15	7,5e-3236	0,793	165	10464
8	46504,75	46695,75	45306,93	39070,56	3,3e-3229	0,686	191	10438

Tabella 5.9: Statistiche di sintesi per il modello a classi latenti con le covariate secondo *Latent Gold* (2009)

Come si nota dalla Tabella 5.9, i criteri BIC e CAIC portano a scegliere un modello con 7 classi latenti, mentre il criterio AIC3 preferisce il modello con 8 classi. Tuttavia, il valore dell'entropy-based R² è maggiore nel modello con 7 classi rispetto a quello con 8. Il modello con una classe in meno sembra avere un adattamento migliore, considerando anche i problemi di interpretabilità dovuti all'esigua ampiezza delle classi del modello meno parsimonioso.

Tutti i parametri relativi agli indicatori sono significativi al 5%, anche se il valore dell'R² continua a rimanere basso per y4 e y9. I parametri delle covariate sono significativi al 5%, ad eccezione della variabile relativa al sesso che non è significativa a nessun livello. Le variabili inserite sono pertanto utili nel predire le classi individuate.

Descrizione e interpretazione delle classi

Nella Tabella 5.10 ho riportato le probabilità di risposta condizionate alla classe di appartenenza; le probabilità condizionate ricavate per le covariate sono riportate in Appendice 1.

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
Ampiezza	48,36%	16,01%	11,51%	9,61%	7,37%	5,29%	1,85%
y1	0,00%	0,00%	0,00%	0,00%	99,88%	0,00%	0,01%
y2	0,00%	1,54%	67,56%	0,40%	0,00%	10,30%	5,46%
y3	96,85%	3,39%	42,47%	6,99%	0,01%	72,39%	7,51%
y4	0,19%	1,52%	0,18%	0,00%	0,00%	3,04%	97,29%
y5	0,32%	22,83%	0,73%	0,49%	0,00%	34,55%	6,09%
y6	1,10%	60,19%	1,10%	4,22%	0,00%	72,18%	6,36%
y7	1,26%	38,15%	0,74%	2,99%	0,00%	67,80%	5,51%
y8	0,03%	10,96%	0,76%	61,04%	0,00%	3,31%	7,48%
y9	0,27%	3,38%	0,90%	6,11%	0,00%	2,98%	1,18%

Tabella 5.10: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello con le covariate secondo *Latent Gold (2009)*

La classe 1, quella più ampia, è costituita dai “non interessati” (Tabella 5.10). Questa classe è composta da persone che presentano probabilità maggiori di avere un’età superiore ai 49 anni, aver terminato gli studi tra i 7 e i 15 anni o successivamente, risiedere in zone rurali e vivere in famiglie composte da due componenti. Per quanto concerne il livello della società, i componenti di tale segmento presentano una probabilità maggiore del 50% di appartenere al livello medio della società e una probabilità di circa il 33,6% di appartenere al livello più basso.

La classe 2, che comprende il 16% circa dei non possessori, è quella degli “interessati, ma senza la possibilità economica per l’installazione e l’utilizzo di Internet”. Tale classe presenta probabilità più elevate per quanto concerne i giovani con meno di 24 anni, che hanno terminato gli studi dopo i 15 anni, abitano in aree rurali e in piccole o medie città, vivono da soli o in famiglie numerose. Inoltre, la probabilità di appartenere al livello basso della società è del 44,1% e quella relativa al livello medio è del 42,8%.

La classe 3 comprende l’11,5% delle persone del sottocampione ed è quella dei “non conoscitori di Internet”. I componenti di tale segmento presentano delle probabilità maggiori di avere più di 59 anni d’età, aver terminato gli studi tra i 7 e i 15 anni, vivere in zone rurali da sole o con un’altra persona e appartenere al livello basso della società. Rilevante è anche la probabilità relativa al possesso di una casa di proprietà.

La classe 4 è costituita dagli “aventi accesso in altri luoghi”. I componenti di questa classe hanno una probabilità pari al 32,8% di avere meno di 24 anni e pari al 30% circa

di avere un'età compresa tra i 24 e i 39 anni. Per quanto riguarda il livello di studi, la probabilità di avere terminato gli studi dopo i 15 anni è pari all'82,6%. Probabilità importanti si registrano anche per quanto concerne il fatto di risiedere in zone non urbane e in piccole o medie città, vivere con un'altra persona, appartenere al livello medio della società e possedere una casa di proprietà.

La classe 5, formata dal 7,4% degli intervistati senza Internet, è quella degli "intenzionati ad abbonarsi". I componenti di tale segmento si caratterizzano per una probabilità più alta di avere un'età inferiore ai 24 anni, mentre quella di aver terminato gli studi dopo i 15 anni è pari al 77,4%. Le persone di tale segmento presentano inoltre probabilità rilevanti di risiedere in zone rurali e in piccole o medie città, vivere in famiglie numerose, appartenere al livello medio della società e possedere un'auto, ma non una casa.

La classe 6 comprende il 5,3% delle unità ed è quella dei "non possessori senza la possibilità economica per l'installazione e utilizzo di Internet". Le probabilità maggiori sono quelle relative ai giovani con meno di 24 anni, che hanno terminato gli studi dopo i 15 anni, risiedono in zone rurali e in piccole o medie città, vivono da soli o con un'altra persona e appartengono al livello basso della società. Tale segmento differisce dal secondo in quanto la probabilità di includere persone non interessate ad Internet è maggiore.

L'ultima classe, la meno ampia, comprende meno del 2% degli intervistati senza Internet ed è costituita prevalentemente da coloro che sono interessati ad avere Internet a casa, ma vivono in una zona non coperta dalla banda larga. Questa classe, non individuata in precedenza, è quella degli "interessati, ma senza la copertura alla banda larga". I componenti di questo segmento presentano probabilità maggiori di avere meno di 24 anni e aver finito gli studi dopo i 15 anni. Si tratta di persone con probabilità più alte di abitare in una zona non urbana, vivere in famiglie numerose e appartenere al livello medio della società. I componenti di tale segmento presentano inoltre probabilità elevate di possedere una casa e un'auto.

5.3.4 Analisi a classi latenti per il 2011 con le covariate

N classi	BIC (LL)	CAIC (LL)	AIC3 (LL)	L^2	p value	R^2 Entr	N param	df
1	43163,07	43172,07	43108,53	39079,50	5,9e-3798	1,0000	9	8591
2	37546,43	37581,43	37334,35	33227,32	2,3e-2837	0,7664	35	8565
3	36204,72	36265,72	35835,09	31650,06	1,1e-2592	0,7755	61	8539
4	35771,13	35858,13	35243,95	30980,92	2,9e-2494	0,7648	87	8513
5	35680,61	35793,61	34995,89	30654,86	3,0e-2450	0,7856	113	8487
6	35735,58	35874,58	34893,31	30474,28	3,7e-2429	0,7879	139	8461
7	35778,95	35943,95	34779,13	30297,10	1,3e-2408	0,7874	165	8435
8	35895,24	36086,24	34737,87	30162,84	8,8e-2395	0,7837	191	8409

Tabella 5.11: Statistiche di sintesi per il modello a classi latenti con le covariate secondo *Latent Gold* (2011)

Dalla Tabella 5.11 si nota che gli indicatori BIC e CAIC suggeriscono di utilizzare un modello con 5 classi latenti, al contrario di AIC3 che preferisce un modello con 8 classi. Tuttavia, il valore dell'entropy-based R^2 è di poco maggiore nel modello con un numero inferiore di classi. Ho pertanto scelto il modello con 5 classi.

I parametri relativi agli indicatori sono tutti significativi; tuttavia, il modello non spiega bene y_4 e y_9 , come indicato dal loro R^2 che assume valori molto bassi. I coefficienti delle covariate sono tutti significativi al 5%, ad eccezione di quello della variabile Femmina.

Descrizione e interpretazione delle classi

Nella Tabella 5.12 ho riportato le probabilità di risposta condizionate alla classe di appartenenza; le probabilità condizionate ricavate per le covariate sono riportate in Appendice 1.

	CL 1	CL 2	CL 3	CL 4	CL 5
Ampiezza	54,43%	19,96%	12,03%	8,52%	5,05%
y1	0,07%	1,05%	4,95%	0,02%	99,53%
y2	0,10%	2,86%	2,97%	66,19%	0,31%
y3	95,18%	20,88%	2,88%	51,32%	1,18%
y4	0,04%	4,07%	8,21%	0,71%	0,28%
y5	0,37%	24,63%	2,46%	0,56%	0,03%
y6	1,72%	66,69%	11,45%	1,19%	0,96%
y7	0,87%	47,61%	5,16%	0,39%	0,11%
y8	0,40%	4,34%	36,76%	0,72%	0,03%
y9	0,40%	1,79%	5,33%	1,31%	0,44%

Tabella 5.12: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello con le covariate secondo *Latent Gold* (2011)

Sulla base dei risultati della Tabella 5.12, la classe 1, la più ampia, raggruppa circa il 54% di coloro che non hanno Internet ed è quella dei “non interessati”. I componenti di tale segmento hanno maggiori probabilità di avere più di 52 anni d’età ed aver concluso gli studi tra i 7 e i 15 anni o successivamente. Probabilità importanti si registrano anche per coloro che abitano in aree rurali e in piccole o medie città, vivono con un’altra persona o da sole, appartengono al livello medio della società e possiedono una casa.

La classe 2, che comprende il 20% circa dei non possessori, è quella dei “non possessori senza la possibilità economica per l’installazione e utilizzo di Internet”. Le probabilità maggiori sono quelle relative ai giovani con meno di 27 anni, che hanno terminato gli studi dopo i 15 anni e abitano in zone non urbane e in piccole o medie città. I componenti di questa classe si caratterizzano inoltre per le probabilità più alte di vivere da soli o con un’altra persona, appartenere al livello basso della società e non possedere un’auto.

La classe 3 comprende il 12% circa delle unità del sottocampione ed è quella dei “non possessori senza una motivazione precisa”. Tuttavia, si nota che le probabilità condizionate relative agli indicatori y4 (mancanza delle infrastrutture) e y8 (preoccupazione per l’accesso a siti non adatti) sono le più elevate. Questa classe è pertanto quella dei “non possessori senza una motivazione precisa”. I suoi componenti presentano delle probabilità più elevate di avere un’età inferiore ai 27 anni e aver terminato gli studi dopo i 15 anni. Essi presentano inoltre una probabilità del 41,8% di vivere in una piccola o media città e del 36,6% di abitare in un’area rurale. Infine, si tratta di non possessori con una

maggior probabilità di vivere con un'altra persona, appartenere al livello medio della società e possedere un'auto.

La classe 4 comprende l'8,5% degli intervistati senza la connessione. Questo segmento è pertanto quello dei "non conoscitori di Internet" e le persone che lo compongono presentano delle probabilità maggiori di avere un'età superiore ai 60 anni, aver finito gli studi tra i 7 e i 15 anni, vivere in un'area rurale da sole o in coppia. I componenti di questo segmento presentano una probabilità del 46,7% di appartenere al livello basso della società e del 41,4% di appartenere al livello medio. La probabilità relativa al possesso di una casa è del 74,6% mentre quella di possedere un'auto è bassa.

La classe 5 è la meno numerosa, comprende il 5% circa dei non possessori ed è costituita principalmente da coloro che non hanno Internet a casa, ma hanno intenzione di abbonarsi nei sei mesi successivi all'indagine. Questa classe è pertanto quella degli "intenzionati ad abbonarsi". I componenti di tale classe presentano probabilità maggiori di avere meno di 27 anni, aver terminato gli studi dopo i 15 anni, risiedere in zone non urbane e in piccole o medie città, vivere con un'altra persona e possedere un'auto.

5.3.5 Analisi dei residui bivariati

Per verificare l'ipotesi di indipendenza locale si possono utilizzare le stime dei residui bivariati. Questa ipotesi prevede che la probabilità di dare una certa risposta y_1 alla domanda Y_1 , condizionatamente alla classe di appartenenza x , sia indipendente dalla probabilità condizionata ad x di dare la risposta y_2 alla domanda Y_2 . Se i residui bivariati non superano il valore soglia l'assunzione di indipendenza locale è rispettata.

In Appendice1 sono riportate le tabelle contenenti i residui bivariati dei modelli presentati nei precedenti paragrafi. Come si può notare, in ciascun modello sembra esserci una correlazione non trascurabile tra almeno una coppia di indicatori. Ho pertanto stimato i diversi modelli inserendo in ognuno l'effetto diretto tra gli indicatori che presentavano il residuo bivariato più elevato. Nella Tabella 5.13 ho riportato i risultati ottenuti.

5. Analisi a classi latenti

Modello senza covariate (2009)								
N classi	BIC (LL)	CAIC (LL)	AIC3 (LL)	L ²	p value	R ² Entr	N param	df
6 cl- effetto diretto y4-y7	49094,514	49154,514	48718,233	247,3107	1	0,8625	60	451
Modello senza covariate (2011)								
N classi	BIC (LL)	CAIC (LL)	AIC3 (LL)	L ²	p value	R ² Entr	N param	df
6 cl- effetto diretto y1-y9	37614,812	37674,812	37251,241	300,004	1	0,7935	60	451
Modello con covariate (2009)								
N classi	BIC (LL)	CAIC (LL)	AIC3 (LL)	L ²	p value	R ² Entr	N param	df
7 cl- effetto diretto y8-y9	46449,056	46615,056	45408,014	39246,648	4,6e-3250	0,7345	166	10463
Modello con covariate (2011)								
N classi	BIC (LL)	CAIC (LL)	AIC3 (LL)	L ²	p value	R ² Entr	N param	df
5 cl- effetto diretto y4-y8	35943,5229	36057,5229	35252,738	30908,71	1,8e-2490	0,6778	114	8486

Tabella 5.13: Statistiche di sintesi dei modelli con gli effetti diretti

Come si può notare, i modelli senza covariate che comprendono l'effetto diretto tra gli indicatori presentano dei valori delle statistiche di poco inferiori rispetto ai modelli senza effetti specifici (Tabelle 5.5 e 5.7). Inoltre, i risultati (in termini di stime dei parametri e di caratterizzazione delle classi latenti) che si ottengono con questi modelli sono sostanzialmente gli stessi di quelli ottenuti in precedenza.

Al contrario, i modelli con le covariate che comprendono gli effetti diretti si adattano peggio ai dati rispetto a quelli senza l'effetto specifico. Analizzando le risposte date ho notato che in entrambi gli anni quasi il 10% delle persone non ha scelto nessuna motivazione riguardo al non possesso di Internet, mentre coloro che hanno scelto due motivazioni sono circa il 17% per ogni anno. Tale percentuale non è molto elevata e potrebbe incidere sulla stima dei residui in quanto, non avendo molta informazione nel campione, le stime dei residui potrebbero non essere adeguate.

Dai risultati ottenuti ho pertanto deciso di mantenere i modelli stimati in precedenza senza l'introduzione di effetti diretti.

5.3.6 Una sintesi dei principali risultati ottenuti

Le maggiori differenze presenti tra le classi emerse dalla stima dei vari modelli sono dovute soprattutto alla diversa suddivisione delle persone che hanno difficoltà a pagare l'installazione iniziale, l'acquisto del computer o del modem e l'abbonamento mensile. In particolare, nei modelli con poche classi tutte le persone, interessate o meno ad aver la connessione Internet a casa, sono raggruppate nello stesso segmento senza distinzione tra i costi che causano i maggiori disagi economici. Con l'aumentare del numero di classi si distinguono le persone interessate ad Internet da quelle che non lo sono e coloro che trovano elevati tutti i costi o solo quelli per l'abbonamento mensile e l'acquisto del computer.

Altre differenze si riscontrano negli indicatori relativi alle risposte che hanno registrato una frequenza minore. In particolare, nei modelli con un numero inferiore di classi tutte le motivazioni marginali (ossia quelle con una probabilità minore di essere scelte) sono riunite nella classe dei "non possessori senza una motivazione precisa"; all'aumentare del numero di segmenti si inizia a distinguere tra coloro che accedono ad Internet in altri luoghi, coloro che non lo conoscono, coloro che vivono in zone non coperte dalla banda larga, etc.

Come prevedibile, da queste osservazioni emerge pertanto che i modelli con un maggior numero di classi distinguono in modo più dettagliato le motivazioni che portano gli intervistati dei due sottocampioni a non avere l'accesso ad Internet a casa. Questo non significa che necessariamente si deve scegliere il modello con il numero più alto di segmenti, ma occorre considerare anche l'ampiezza e l'interpretabilità delle classi. Un numero elevato di classi comporta un'ampiezza esigua delle stesse e difficoltà ad attribuire delle etichette che le rappresentino.

Di seguito ho riportato una tabella riassuntiva (Tabella 5.14) che presenta le classi individuate con i diversi modelli e la loro ampiezza, al fine di fare dei confronti tra i due istanti temporali considerati e tra i modelli senza e con covariate.

Senza covariate	
2009	2011
Non interessati (48,78%)	Non interessati (53,77%)
Non possessori senza una motivazione precisa (18,03%)	Non possessori senza la possibilità economica per l'abbonamento mensile e l'acquisto di un pc e/o modem (20,21%)
Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (12,88%)	Non possessori senza una motivazione precisa (11,49%)
Non conoscitori di Internet (7,78%)	Intenzionati ad abbonarsi (5,41%)
Intenzionati ad abbonarsi (7,38%)	Aventi accesso in altri luoghi (4,82%)
Aventi accesso in altri luoghi (5,14%)	Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (4,29%)
Con covariate	
2009	2011
Non interessati (48,36%)	Non interessati (54,43%)
Interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet (16,01%)	Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (19,96%)
Non conoscitori di Internet (11,51%)	Non possessori senza una motivazione precisa (12,03%)
Aventi accesso in altri luoghi (9,61%)	Non conoscitori di Internet (8,52%)
Intenzionati ad abbonarsi (7,37%)	Intenzionati ad abbonarsi (5,05%)
Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (5,29%)	
Interessati, ma senza la copertura alla banda larga (1,85%)	

Tabella 5.14: Classi individuate con *Latent Gold*

Considerando i modelli senza covariate, si nota che la classe più vasta è sempre quella dei “non interessati”, più ampia nel 2011 rispetto al 2009. Questo sta a indicare che nel 2011 la maggior parte delle persone che non possiedono Internet ha come motivazione principale il fatto di non essere interessata, ancor più che nel 2009. Tuttavia, occorre ricordare che nel secondo istante temporale il numero di non possessori è minore.

Nel secondo istante temporale è presente la classe dei “non possessori senza la possibilità economica per l'abbonamento mensile e l'acquisto di un pc e/o modem”, mentre quella che comprende nel 2009 coloro con una maggiore probabilità di ritenere troppo elevato il costo per l'installazione iniziale è meno ampia. Nel 2011 c'è pertanto una separazione tra coloro che ritengono troppo elevati i costi dell'abbonamento e dei computer e

coloro che ritengono eccessivo anche il costo per l'installazione. Al contrario, nel 2009 le persone che evidenziano probabilità più alte di riscontrare un qualsiasi tipo di difficoltà economica sono riunite nella classe dei "non possessori senza la possibilità economica per l'installazione e utilizzo di Internet". Dai modelli senza covariate emerge inoltre che i problemi economici sono più evidenti nel 2011, probabilmente come conseguenza della crisi economica.

Altra differenza evidente è la mancanza nel 2011 della classe costituita principalmente da coloro che non sanno cosa sia Internet, classe che tuttavia ricompare nei modelli con le covariate. Pertanto, la mancanza di tale classe nel modello senza covariate non sembra indicare una maggior diffusione della conoscenza di Internet.

Confrontando i modelli con le covariate, si nota come il numero di classi individuate nel 2009 sia maggiore. Il primo segmento coincide come il solito ed è più ampio nel 2011 che nel 2009.

La seconda classe del 2011 raggruppa al suo interno le persone interessate e non interessate che presentano una probabilità più elevata di riscontrare delle difficoltà nel pagare tutte le spese necessarie per poter installare e utilizzare Internet. Questa classe corrisponde al secondo e al sesto segmento del 2009 ed è meno numerosa nel 2011, a differenza dei modelli senza covariate. Inoltre, è evidente come la classe delle persone che non conoscono Internet sia diventata meno numerosa nel 2011, anche se comprende ancora una percentuale importante di unità.

Nel 2009 non è stata individuata la classe composta da coloro che non hanno dato una motivazione precisa; tuttavia, sono stati individuati dei segmenti aggiuntivi che nel 2011 sembrano rientrare nella classe di coloro che non hanno indicato un motivo specifico. Tali classi sono quella delle persone che si connettono ad Internet da altri luoghi e quella delle persone che vivono in una zona non coperta dalla banda larga. Queste motivazioni sembrano avere un'importanza minore nel 2011, segnalando un possibile miglioramento nelle infrastrutture e nei costi per accedere a Internet.

Considerando ora i due modelli stimati per il 2009, si nota come in quello con le covariate non sia presente la classe di coloro che non hanno saputo dare una motivazione precisa al fatto di non avere Internet a casa. Al contrario, sono presenti delle classi nuove: quella delle persone che risiedono in una zona non coperta dalla banda larga e quella degli interessati con problemi economici. Sembra quindi che la classe formata da coloro

che non hanno dato una motivazione precisa con l'inserimento delle covariate sia stata suddivisa tra le altre e per questo anche il segmento delle persone che accedono da altri luoghi sia diventato più ampio. Inoltre, il modello con le covariate identifica una classe a sé stante per coloro che vorrebbero avere Internet a casa, ma non riescono a sostenere i costi conseguenti.

Il modello con le covariate porta delle informazioni aggiuntive, utili per comprendere le motivazioni per cui le persone intervistate non possiedono Internet a casa e rappresentare in modo più dettagliato la realtà dei non possessori.

Per quanto concerne il 2011, nel modello senza covariate sono state individuate 6 classi e in quello con le covariate ne sono state individuate 5. Il modello con le covariate riunisce in un unico segmento le persone con una probabilità più alta di riscontrare un qualsiasi tipo di difficoltà economica, mentre il modello senza covariate distingue tra chi ritiene eccessivi solo i costi per l'abbonamento e per l'acquisto di un computer e coloro che ritengono elevato anche il costo per l'installazione iniziale.

Nel modello con le covariate la classe delle persone intenzionate ad abbonarsi è meno ampia. Esso individua inoltre la classe costituita da coloro che non conoscono Internet, mentre non identifica quella delle persone che accedono in altri luoghi. Nel modello con le covariate, che ha un numero di classi minore, le persone che accedono ad Internet in altri luoghi rientrano nella classe di quelli senza una motivazione precisa.

Tra il 2009 e il 2011 sembrano esserci stati alcuni interessanti cambiamenti. Innanzitutto, emerge come la motivazione principale del non possesso di Internet a casa sia il mancato interesse nell'aver la connessione domestica, motivazione ancora più forte nel 2011. La Commissione Europea potrebbe pianificare degli interventi atti a diffondere maggiormente la conoscenza dei vantaggi dovuti all'utilizzo di Internet. Tuttavia, essendo il segmento dei non interessati costituito anche da persone anziane e con un livello di istruzione medio-basso è difficile pensare che esse possano cambiare il loro modo di pensare. La Commissione potrebbe pertanto focalizzarsi sulle persone con un'istruzione più elevata e un'età meno avanzata.

Un problema da affrontare è invece quello del costo degli abbonamenti mensili e dei computer e modem, problema riscontrato in particolare dai giovani. La Commissione Europea potrebbe quindi rivedere i costi che le imprese di telecomunicazioni attuano per

far sì che il costo mensile non sia più un ostacolo all'installazione di Internet. Potrebbe inoltre fornire degli incentivi per l'acquisto di computer e modem alle famiglie più disagiate. Considerando il costo per l'installazione iniziale della Rete, esso sembra avere un peso marginale sul fatto di non possedere la connessione a casa.

Un altro ostacolo da superare è quello legato alla poca conoscenza di Internet, impensabile ai giorni nostri. Tuttavia, si tratta in particolare di anziani con un basso livello di istruzione, che vivono in aree non urbane e appartengono al livello più basso della società. In questo caso nessuna soluzione sembra utile in quanto è difficile che le persone anziane cambino i loro usi, non possedendo nemmeno le conoscenze necessarie.

Per quanto concerne gli intenzionati ad abbonarsi a Internet nei mesi successivi all'indagine, sarebbe utile verificare se queste persone hanno effettivamente sottoscritto un abbonamento. In caso negativo, si potrebbero individuare le motivazioni della mancata installazione della connessione da casa.

Infine, emergono i gruppi composti da coloro che accedono ad Internet da altri luoghi e coloro che non sono coperti dalle infrastrutture della banda larga. Chi accede in altri luoghi, in particolare giovani che studiano o lavorano, utilizza in ogni caso Internet ed è a conoscenza dei vantaggi che tale tecnologia porta. Il problema dovuto alla mancanza delle infrastrutture necessarie in alcune aree è invece già stato risolto dalla Commissione Europea nell'ottobre 2013, quando è stata raggiunta la copertura del 100%.

5.3.7 Confronto con *Stata*

La tabella 5.15 riporta le classi individuate con la stima dei modelli a classi latenti senza covariate utilizzando *Stata*.

Senza covariate	
2009	2011
Non interessati (50,11%)	Non interessati (56,20%)
Non possessori senza una motivazione precisa (18,14%)	Non possessori senza una motivazione precisa (20,06%)
Interessati, ma senza la possibilità economica per l'abbonamento mensile e l'acquisto di un pc e/o modem (12,62%)	Interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet (17,25%)
Non possessori senza la possibilità economica per l'abbonamento mensile e l'acquisto di un pc e/o modem (10,35%)	Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (11,33%)
Intenzionati ad abbonarsi (7,36%)	Intenzionati ad abbonarsi (5,37%)
Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (1,42%)	

Tabella 5.15: Classi individuate con *Stata*

Per quanto riguarda il 2009, il numero di classi individuate con l'analisi a classi latenti effettuata con *Latent Gold* coincide con quello ottenuto con *Stata*. Tuttavia, le classi identificate non sono le stesse. Con *Latent Gold* sono state individuate due classi nuove: i "non conoscitori di Internet" e gli "aventi accesso in altri luoghi". Le altre classi sono simili a quelle individuate con *Stata*; ci sono delle piccole differenze per quanto concerne i problemi legati ai costi troppo elevati per l'abbonamento mensile, per l'installazione iniziale e l'acquisto di un pc e di un modem. Per quanto riguarda l'ampiezza delle classi, sono evidenti delle piccole differenze in particolare nella classe dei "non interessati", che nella seconda analisi (quella con *Latent Gold*) è meno ampia.

Considerando invece il 2011, il numero di classi latenti individuate dai due software è diverso. Mentre *Stata* ha identificato 5 classi latenti, *Latent Gold* ne ha individuate 6 e sono presenti piccole differenze anche nelle caratterizzazioni delle altre classi. La classe più numerosa è sempre quella dei non interessati; la classe aggiuntiva determinata con la seconda analisi è quella delle persone che accedono a Internet in altri luoghi. Anche in questo caso le classi individuate differiscono per quanto riguarda le difficoltà economiche riscontrate dalle persone intervistate. In particolare, *Stata* suddivide le classi principalmente basandosi sui problemi economici, senza considerare le motivazioni che presentano una probabilità minore. Per questo, le classi individuate con *Stata* sono più difficili da interpretare.

Il numero di classi individuate con i due software coincide quindi nel 2009, ma non nel 2011. Ho pertanto stimato per il secondo istante temporale un modello con 5 classi utilizzando *Latent Gold*, con lo scopo di analizzare ulteriormente le differenze presenti tra i due programmi. I risultati sono riportati nella Tabella 5.16.

	CL 1	CL 2	CL 3	CL 4	CL 5
Ampiezza	55,58%	19,66%	15,13%	5,30%	4,30%
y1	0,00%	1,30%	1,22%	99,02%	4,25%
y2	5,17%	2,03%	21,93%	0,01%	1,09%
y3	99,95%	22,98%	2,78%	1,83%	3,91%
y4	0,14%	3,53%	6,81%	0,61%	1,41%
y5	0,47%	24,20%	2,74%	0,27%	0,33%
y6	0,86%	73,59%	4,08%	0,90%	3,92%
y7	0,79%	44,65%	8,36%	0,34%	3,24%
y8	0,27%	6,02%	0,02%	0,02%	97,66%
y9	0,34%	1,97%	4,18%	0,74%	2,38%

Tabella 5.16: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello a 5 classi latenti senza covariate secondo *Latet Gold* (2011)

La prima classe comprende il 55,6% dei non possessori di Internet ed è quella dei "non interessati (Tabella 5.16). La classe 2, costituita dal 19,7% delle persone che non possiedono la connessione a casa, è quella dei "non possessori, senza la possibilità economica per l'installazione e utilizzo di Internet". La classe 3, che raggruppa il 15,1% delle persone del sottocampione, è quella dei "non possessori senza una motivazione precisa" (tuttavia, le probabilità condizionate relative agli indicatori y2, y4 e y8 sono rilevanti). La classe 4 comprende il 5,3% dei non possessori ed è quella degli "intenzionati ad abbonarsi". Infine, l'ultima classe che comprende il 4,3% delle persone del sottocampione è quella degli "aventi accesso in altri luoghi".

Si nota pertanto che anche utilizzando lo stesso numero di classi, i due software individuano dei segmenti con caratteristiche differenti. In particolare, *Latent Gold* oltre ad offrire numerosi strumenti di analisi individua delle classi più facili da interpretare e descrivere, che rappresentano in modo più dettagliato le motivazioni dei non possessori di Internet a casa.

5.4 Analisi multilivello a classi latenti

Per tenere conto della struttura gerarchica dei dati, nell'ultima parte di questo lavoro di tesi ho stimato dei modelli multilivello a classi latenti. In questo modo, oltre ad identificare i segmenti che raggruppano le persone che non hanno Internet si individuano dei gruppi che includono diversi Paesi o un singolo Paese. All'inizio ho stimato dei modelli senza covariate per poi proseguire con la stima di modelli con le covariate. Ho inoltre utilizzato due metodi diversi per determinare il numero di classi ai diversi livelli: inizialmente ho scelto in modo simultaneo il numero di classi a livello 1 e 2; in seguito ho utilizzato il metodo a 3 passi presentato alla fine del Capitolo 3, metodo che ha dato risultati difficili da interpretare.

5.4.1 Analisi multilivello a classi latenti per il 2009 senza covariate

Per scegliere il numero di segmenti a livello di individuo e di Paese ho stimato dei modelli con diverse combinazioni di classi a livello 1 e 2. Ho deciso di utilizzare il criterio BIC nella scelta del numero di classi in quanto gli altri criteri portavano a delle classi poco numerose e difficili da interpretare. La Tabella 5.17 riassume i risultati ottenuti.

N classi	N gruppi								
	1	2	3	4	5	6	7	8	9
1	57341,5	57350,7	57360,0	57369,3	57378,6	57387,8	57397,1	57406,4	57415,6
2	52808,0	52534,1	52474,4	52476,0	52494,3	52512,4	52530,9	52549,9	52568,0
3	50386,8	50099,7	50039,8	50008,7	50021,5	50017,2	50040,7	50063,5	50096,4
4	49527,4	49227,8	49170,6	49163,7	49162,8	49160,4	49140,0	49167,8	49216,5
5	49167,0	48854,3	48745,3	48769,3	48692,9	48693,7	48749,3	48773,7	48769,2
6	49187,9	48812,7	48689,3	48692,8	48529,7	48521,3	48546,6	48871,4	48739,6
7	49135,4	48830,6	48670,8	48602,1	48517,7	48658,1	48609,1	48711,8	48607,3
8	49198,2	48886,9	48781,2	48679,7	48610,7	48552,3	48702,6	48638,0	48676,6

Tabella 5.17: Statistiche di sintesi del modello multilivello senza covariate secondo *Latent Gold* (2009)

Il criterio BIC suggerisce di utilizzare un modello con 7 classi a livello 1 e 5 classi a livello 2. I parametri che misurano l'effetto delle classi a livello 1 sugli indicatori sono tutti significativi al livello del 5%.

Per quanto concerne l'adattamento del modello, il valore dell'entropy-based R^2 relativo alle classi a livello 1 è pari a 0,80, mentre quello relativo alle classi a livello 2 è dello 0,99. Tuttavia, il valore dell' R^2 associato ad y_4 e y_9 è basso (rispettivamente 0,10 e 0,03), come riscontrato nelle stime degli analoghi modelli non multilivello.

Descrizione e interpretazione delle classi a livello 1

La tabella 5.18 riporta le probabilità di risposta condizionate all'appartenenza alle varie classi a livello 1.

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
Ampiezza	50,90%	13,43%	11,42%	7,56%	6,15%	6,00%	4,55%
y1	0,00%	0,00%	0,00%	99,79%	0,00%	0,00%	0,00%
y2	3,13%	1,77%	0,01%	0,00%	9,93%	99,31%	0,01%
y3	99,98%	9,73%	8,77%	0,01%	49,84%	28,35%	0,19%
y4	0,06%	14,08%	0,02%	0,00%	4,55%	1,03%	0,20%
y5	0,25%	5,58%	22,12%	0,00%	35,15%	1,22%	0,01%
y6	1,08%	7,16%	68,69%	0,00%	73,95%	1,55%	0,95%
y7	1,08%	9,85%	33,24%	0,00%	73,70%	0,59%	2,20%
y8	0,52%	7,13%	13,41%	0,00%	5,13%	1,04%	99,70%
y9	0,29%	6,02%	2,62%	0,00%	3,25%	0,25%	0,91%

Tabella 5.18: Probabilità condizionate $P(Y_1 = 1 | X = x)$

La classe 1, che comprende circa la metà delle persone senza Internet a casa, è quella dei “non interessati” (Tabella 5.18). Questa classe è costituita principalmente da persone con più di 55 anni che hanno terminato gli studi tra i 7 e i 15 anni o successivamente. Il 42% circa dei componenti risiede in una zona rurale e il 33,7% in una piccola o media città. Si tratta soprattutto di non possessori che vivono da soli o con un'altra persona, appartengono a livello medio-basso della società e possiedono una casa di proprietà.

La classe 2 è costituita dal 13,4% dei non possessori. Questo segmento è quello “non possessori senza una motivazione precisa”; si nota come la probabilità condizionata più alta sia quella relativa ad y_4 , ossia al fatto di vivere in una zona non coperta dalla banda

larga. Si tratta principalmente di giovani con meno di 24 anni, che hanno terminato gli studi dopo i 15 anni. Il 44,2% dei componenti vive in una zona rurale. Sono in particolare persone che vivono da sole o con un'altra persona e appartengono al livello medio-basso della società.

La terza classe, costituita dall'11,4% delle persone senza la connessione domestica, è quella degli "interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet". Questo segmento comprende tendenzialmente giovani con meno di 24 anni, che hanno concluso gli studi dopo i 15 anni e risiedono in una zona non urbana o in una piccola o media città. Si tratta per lo più di persone che vivono in famiglie numerose e che appartengono al livello medio-basso della società.

La classe 4, che comprende il 7,6% dei non possessori, è quella degli "intenzionati ad abbonarsi". Essa è costituita principalmente da giovani che hanno concluso gli studi dopo i 15 anni. Si tratta in particolare di persone che risiedono in zone non urbane e in piccole o medie città, vivono da sole, appartengono al livello medio della società e possiedono un'auto.

La quinta classe, costituita dal 6,3% dei non possessori, è quella dei "non possessori senza la possibilità economica per l'installazione e utilizzo di Internet". Questo segmento comprende in particolare giovani con meno di 24 anni, che hanno finito gli studi dopo i 15 anni. Il 42,4% risiede in una zona non urbana. Sono persone che vivono da sole o con un'altra persona e appartengono al livello basso della società.

La classe 6, poco meno numerosa della precedente, è quella dei "non conoscitori di Internet". I suoi componenti sono principalmente anziani, che hanno terminato gli studi tra i 7 e i 15 anni e abitano in una zona rurale. Si tratta di persone che vivono da sole o con un'altra persona, appartengono al livello basso della società e possiedono una casa. Solo il 31,8% possiede un'auto.

Infine, l'ultima classe comprende il 4,6% delle osservazioni ed è quella degli "aventi accesso in altri luoghi". Questo segmento è costituito principalmente da persone con meno di 40 anni, che hanno terminato gli studi dopo i 15 anni. Il 38,6% dei componenti abita in una zona rurale e il 35,4% risiede in una piccola o media città. Sono persone che vivono con un'altra persona, appartengono al livello medio della società, possiedono un'auto e una casa.

Descrizione e interpretazione delle classi a livello 2

La Tabella 5.19 riporta le classi in cui sono stati suddivisi i 27 Paesi europei considerati sulla base dell'empirical Bayes modal prediction. La Tabella 5.20 presenta invece le probabilità di appartenere alla classe latente x^g a livello 2 data l'appartenenza alla classe latente x a livello 1.

Gruppi	Dimensione	Paesi inclusi
1	36,26%	Danimarca, Francia, Germania, Irlanda, Lussemburgo, Olanda, Polonia, Regno Unito, Slovenia, Svezia
2	22,20%	Cipro, Finlandia, Grecia, Malta, Spagna
3	18,67%	Austria, Belgio, Lituania, Portogallo, Rep Ceca, Ungheria
4	18,58%	Bulgaria, Estonia, Italia, Lettonia, Slovacchia
5	4,29%	Romania

Tabella 5.19: Dimensione e composizione delle classi a livello 2

	Gruppo 1	Gruppo 2	Gruppo 3	Gruppo 4	Gruppo 5
Classe 1	55,63%	60,25%	49,32%	38,03%	25,06%
Classe 2	16,60%	6,97%	4,50%	18,67%	36,35%
Classe 3	7,26%	11,85%	17,20%	15,51%	1,46%
Classe 4	9,11%	5,75%	7,10%	7,69%	5,16%
Classe 5	5,28%	1,05%	10,88%	7,04%	15,32%
Classe 6	3,66%	9,63%	5,61%	4,53%	15,21%
Classe 7	2,46%	4,50%	5,39%	8,54%	1,44%

Tabella 5.20: Probabilità di appartenere alla classe latente a livello 1, data l'appartenenza alla classe latente a livello 2

Il primo gruppo di Paesi è il più numeroso e comprende 10 Stati; si tratta principalmente di Paesi che appartengono all'Europa Occidentale e Settentrionale (Danimarca, Francia, Germania, Irlanda, Lussemburgo, Olanda, Regno Unito, Svezia), ad eccezione di Polonia e Slovenia. I non possessori che risiedono in questi Paesi presentano una probabilità più alta di appartenere alla classe a livello 1 dei "non interessati" e a quella dei "non possessori senza una motivazione precisa" (Tabella 5.20). I Paesi dell'Europa Occidentale e Settentrionale sono dei Paesi sviluppati in cui Internet ha già raggiunto un alto livello di diffusione. Pertanto, le persone che non hanno la connessione a casa sono principalmente persone di una certa età che non sono interessate ad avere e utilizzare Internet oppure persone che sono in attesa di installarlo. In particolare, Germania e Danimar-

ca si distinguono per un'elevata probabilità di persone non interessate; al contrario, Francia, Olanda, Regno Unito e Svezia registrano una probabilità elevata di intenzionati ad abbonarsi. Anche Polonia e Slovenia rientrano in questo segmento. Dalle analisi esplorative è emerso come questi due Paesi siano tra quelli con una copertura a banda larga più bassa e specialmente la Polonia è il Paese con la copertura minore. Per questo motivo questi due Stati hanno registrato una maggiore frequenza di risposte affermativo all'indicatore y4 durante l'indagine.

Il secondo gruppo comprende 5 Paesi. Si tratta principalmente di Stati dell'Europa Mediterranea come Grecia, Spagna, Malta e Cipro. Tuttavia, questo gruppo comprende anche la Finlandia, Paese del nord Europa dove Internet è più diffuso. I Paesi che compongono questo gruppo comprendono quegli individui che hanno la più alta probabilità di appartenere alla classe a livello 1 dei "non interessati". Tuttavia, questo gruppo si distingue dal primo in quanto i non possessori che vi sono riuniti presentano una maggiore probabilità di appartenere alla classe degli "interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet". In particolare, per quanto riguarda i costi da sostenere, Cipro è tra i Paesi che risentono maggiormente delle spese di installazione e, insieme alla Spagna, dell'abbonamento mensile.

Il terzo gruppo comprende 6 Stati. Si tratta di Paesi dell'Europa Occidentale come Belgio, dell'Europa Mediterranea come il Portogallo, dell'Europa Centrale come l'Austria e dell'Europa Orientale come Lituania, Repubblica Ceca e Ungheria. Sembra strano che il Belgio, Paese con la copertura del 100% e un elevato HDI, possa trovarsi in questo gruppo. In questi Paesi la probabilità condizionata maggiore è quella relativa alla classe a livello 1 dei "non interessati". Tuttavia, i non possessori che risiedono in questi Stati presentano una probabilità rilevante di appartenere ai segmenti con maggiori difficoltà economiche (classi 3 e 5 a livello 1). In particolare, in Belgio, Austria, Repubblica Ceca e Ungheria sono evidenti problemi legati ai costi dell'installazione iniziale; in Austria, Repubblica Ceca e Ungheria sono inoltre chiare le difficoltà legate all'acquisto di un computer e/o di un modem; in tutti i Paesi inclusi si registrano difficoltà economiche dovute dagli abbonamenti mensili. Questo gruppo sembra quello in cui risiedono i non possessori con maggiore probabilità di riscontrare problemi economici.

Il gruppo 4 è costituito da 5 Paesi. Si tratta di Paesi dell'Europa Orientale (Bulgaria, Estonia, Lettonia, Slovacchia) ad eccezione dell'Italia. Gli individui che risiedono in

questi Stati presentano, come nei gruppi precedenti, una probabilità maggiore di appartenere al segmento dei “non interessati”, ma rilevanti sono anche la probabilità di appartenere alle classi a livello 1 dei “non possessori senza una motivazione precisa” e degli “interessati, ma senza la possibilità economica per l’installazione e utilizzo di Internet”. Questi Stati comprendono le persone con una maggiore probabilità di vivere in una zona non coperta dalle infrastrutture necessarie. In particolare, Slovacchia e Bulgaria sono tra gli Stati con la copertura DSL più bassa, come emerso dalle analisi esplorative del Capitolo 2. Inoltre, i non possessori che risiedono in Slovacchia, Lettonia ed Estonia ritengono troppo elevati i costi per gli abbonamenti mensili, l’installazione iniziale e l’acquisto del computer e/o del modem. L’Italia si caratterizza principalmente per le alte probabilità relative all’accesso in altri luoghi e ai costi eccessivi degli abbonamenti.

Infine, l’ultimo gruppo è costituito da coloro che risiedono in Romania. Queste persone presentano una probabilità più elevata di appartenere alla classe a livello 1 dei “non possessori senza una motivazione precisa”. Probabilità elevate si registrano anche per quanto riguarda la classe 1, la 5 e la 6. Dunque, le persone che risiedono in Romania hanno una probabilità più elevata di riscontrare problemi relativi alla scarsità delle infrastrutture necessarie, problemi economici per acquistare gli strumenti necessari all’utilizzo di Internet e per pagare l’abbonamento mensile. Inoltre, coloro che risiedono in questo Paese presentano una probabilità rilevante di non conoscere Internet.

5.4.2 Analisi multilivello a classi latenti per il 2011 senza covariate

N classi	N gruppi								
	1	2	3	4	5	6	7	8	9
1	43163,1	43172,1	43181,2	43190,2	43199,3	43208,4	43217,4	43226,5	43235,5
2	39262,4	39116,6	39106,2	39120,7	39138,2	39156,1	39174,3	39192,4	39210,5
3	37902,0	37722,1	37673,7	37651,2	37651,7	37661,8	37686,7	37712,4	37726,3
4	37690,2	37526,1	37528,8	37525,8	37521,1	37540,7	37543,0	37544,6	37576,1
5	37630,3	37609,3	37525,1	37380,6	37508,3	37600,5	37372,3	37395,7	37489,5
6	37623,3	37492,4	37381,9	37414,7	37428,5	37389,7	37504,0	37640,6	37480,0
7	37668,0	37507,0	37472,1	37385,2	37360,1	37483,5	37409,4	37565,2	37624,4
8	37702,7	37522,7	37445,5	37495,1	37502,5	37535,4	37472,4	37545,9	37657,3

Tabella 5.21: Statistiche di sintesi del modello multilivello senza covariate secondo *Latent Gold* (2011)

Anche in questo caso il criterio BIC suggerisce di utilizzare un modello con 7 classi a livello 1 e 5 classi a livello 2 (Tabella 5.21). I coefficienti che misurano l'effetto delle classi a livello 1 sugli indicatori sono tutti significativi al 5%, ad eccezione di quello relativo ad y_2 che è significativo al 10%.

Il modello presenta un valore dell'entropy-based R^2 pari a 0,83 per le classi a livello 1 e a 0,97 per quelle a livello 2, valore leggermente inferiore rispetto al modello per il 2009. Gli R^2 relativi a y_4 e y_9 sono rispettivamente pari a 0,06 e 0,03.

Descrizione e interpretazione delle classi a livello 1

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
Ampiezza	54,21%	16,96%	11,03%	6,06%	4,91%	3,61%	3,22%
y1	0,04%	4,20%	0,46%	0,00%	99,22%	3,41%	0,01%
y2	0,01%	4,83%	0,00%	95,87%	0,01%	13,32%	0,00%
y3	99,93%	0,38%	18,57%	54,33%	1,53%	53,08%	1,82%
y4	0,16%	7,06%	0,45%	0,50%	0,13%	13,99%	1,11%
y5	0,39%	4,93%	25,67%	0,40%	0,08%	37,08%	0,66%
y6	1,35%	17,21%	79,86%	0,24%	0,07%	70,05%	3,35%
y7	1,01%	12,00%	53,08%	0,47%	0,02%	45,76%	2,44%
y8	0,45%	3,71%	6,52%	0,85%	2,44%	6,36%	99,15%
y9	0,38%	4,57%	0,67%	0,00%	0,35%	6,25%	1,50%

Tabella 5.22: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello multilivello senza covariate secondo *Latent Gold* (2011)

Sulla base dei risultati della Tabella 5.22, la prima classe comprende oltre la metà degli individui senza Internet a casa ed è quella dei “non interessati”. Si tratta in particolare di persone con più di 55 anni, che hanno terminato gli studi tra i 7 e i 15 anni o successivamente. I componenti di questa classe risiedono prevalentemente in un’area rurale oppure in una piccola o media città, vivono da soli o con un’altra persona, appartengono al livello medio-basso della società e possiedono una casa di proprietà.

La classe 2 raggruppa il 17% circa delle unità del sottocampione ed quella dei “non possessori senza una motivazione precisa”. Essa è costituita principalmente da giovani con meno di 27 anni, che hanno terminato gli studi dopo i 15 anni. Si tratta soprattutto di persone che risiedono in zone rurali e in piccole o medie città, vivono da sole o con un’altra persona e appartengono al livello medio-basso della società.

La classe 3, che comprende l’11% circa dei non possessori, corrisponde a quella degli “interessati, ma senza la possibilità economica per l’installazione e utilizzo di Internet”. Questo segmento è costituito in particolare da giovani con meno di 27 anni, che hanno terminato gli studi dopo i 15 anni. I componenti di questa classe risiedono principalmente in zone rurali e in piccole o medie città e vivono da sole o con un’altra persona. Inoltre, più del 50% dei componenti appartiene al livello basso della società.

La classe 4 comprende il 6,1% delle persone che non hanno la connessione domestica e corrisponde al segmento dei “non conoscitori di Internet”. Si tratta principalmente di anziani, che hanno terminato gli studi tra i 7 e i 15 anni. Questo segmento è costituito per il 51,4% da persone che risiedono in una zona rurale. Sono soprattutto persone che vivono in una famiglia composta da uno e due componenti, appartengono al livello basso della società, possiedono una casa, ma non un’auto.

La classe 5 raggruppa il 5% circa dei non possessori ed è quella degli “intenzionati ad abbonarsi”. Questo segmento è costituito prevalentemente da giovani con meno di 27 anni, che hanno terminato gli studi dopo i 15 anni. Si tratta prevalentemente di persone che risiedono in zone rurali e in piccole o medie città, vivono con un’altra persona e appartengono al livello medio della società.

La classe 6 comprende il 3,6% delle persone senza Internet a casa e corrisponde a quella dei “non possessori senza la possibilità economica per l’installazione e l’utilizzo di Internet”. Si tratta principalmente di giovani con meno di 27 anni, che hanno concluso gli studi dopo i 15 anni. Sono persone che vivono in una zona non urbana, fanno parte di una famiglia composta da due componenti e appartengono al livello medio-basso della società. Inoltre, solo il 26,8% dei componenti possiede un’auto.

L’ultima classe, poco meno numerosa della precedente, è quella degli “aventi accesso in altri luoghi”. Questo segmento è costituito in particolare da persone con meno di 40 anni, che hanno terminato gli studi dopo i 15 anni. Il 44,5% dei componenti abita in una piccola o media città e il 33% in un’area rurale. Si tratta principalmente di persone che vivono con un’altra persona o in una famiglia costituita da 4 o più componenti, appartengono al livello medio della società e possiedono un’auto.

Descrizione e interpretazione delle classi a livello 2

Gruppi	Dimensione	Paesi inclusi
1	39,80%	Austria, Bulgaria, Cipro, Francia, Germania, Irlanda, Olanda, Polonia, Portogallo, Regno Unito, Svezia
2	22,14%	Estonia, Grecia, Italia, Lettonia, Lituania, Slovacchia
3	19,22%	Finlandia, Lussemburgo, Malta, Slovenia, Spagna
4	10,98%	Belgio, Danimarca, Romania
5	7,86%	Rep Ceca, Ungheria

Tabella 5.23: Dimensione e composizione delle classi a livello 2

	Gruppo 1	Gruppo 2	Gruppo 3	Gruppo 4	Gruppo 5
Classe 1	60,31%	48,29%	58,65%	45,22%	41,71%
Classe 2	17,34%	17,44%	15,36%	24,33%	7,35%
Classe 3	10,71%	15,93%	3,86%	0,33%	31,31%
Classe 4	3,82%	3,87%	12,67%	8,49%	4,03%
Classe 5	3,99%	6,88%	4,77%	4,97%	4,28%
Classe 6	1,90%	0,01%	3,31%	15,92%	5,91%
Classe 7	1,93%	7,58%	1,39%	0,74%	5,41%

Tabella 5.24: Probabilità di appartenere alla classe latente a livello 1, data l'appartenenza alla classe latente a livello 2

Come si nota dalla Tabella 5.23, il primo gruppo è costituito da 11 Stati. Alcuni si trovano nell'Europa Occidentale come Francia, Olanda, Germania, Regno Unito, Irlanda, Austria, altri appartengono all'Europa Orientale come Polonia e Bulgaria, altri si trovano nell'Europa Mediterranea (Cipro e Portogallo) e settentrionale (Svezia). Questi Paesi comprendono i non possessori con la più alta probabilità di appartenere alla prima classe a livello 1, ossia a quella dei “non interessati”. Rilevante è anche la probabilità di appartenere al segmento dei “non possessori senza una motivazione precisa” (Tabella 5.24).

Il secondo gruppo comprende le repubbliche baltiche (Estonia, Lettonia e Lituania), Italia, Grecia e Slovacchia. Le persone che vivono in questi Paesi presentano una probabilità più elevata di appartenere alla classe dei “non interessati”, ma importanti sono anche le probabilità di appartenere alle classi dei “non possessori senza una motivazione precisa” e degli “interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet”. In particolare, Estonia, Lettonia, Lituania e Slovacchia si caratterizzano per la mancanza delle infrastrutture necessarie in alcune aree e i costi eccessivi per

l'abbonamento. Estonia e Slovacchia presentano inoltre le probabilità maggiori di riscontrare costi eccessivi per l'installazione iniziale e l'acquisto di un computer o di un modem. La Grecia risente principalmente dei costi dell'abbonamento mensile, mentre i non possessori che risiedono in Italia hanno una maggiore probabilità di riscontrare difficoltà nel pagare l'abbonamento mensile e di accedere ad Internet in altri luoghi.

Il terzo gruppo comprende Finlandia, Lussemburgo, Malta, Slovenia e Spagna. Le persone senza Internet che risiedono in questi Paesi presentano delle probabilità elevate relative alle classi 1 e 2 a livello 1 ("non interessati" e "non possessori senza una motivazione precisa"). Tuttavia, gli individui compresi in questo gruppo presentano una probabilità condizionata alta anche per quanto concerne il segmento dei "non conoscitori" e questo vale in particolare per coloro che risiedono in Spagna e Malta.

Il quarto gruppo è costituito da Belgio, Danimarca e Romania. Le probabilità condizionate maggiori per le persone che risiedono nei Paesi compresi in questo gruppo sono quelle relative all'appartenenza alle classi a livello 1 dei "non interessati", dei "non possessori senza una motivazione precisa" e dei "non possessori senza la possibilità economica per l'installazione e utilizzo di Internet". Infatti, questi Paesi si differenziano per l'elevata probabilità di riscontrare difficoltà economiche.

Il gruppo 5 è costituito da Repubblica Ceca e Ungheria. Le persone che risiedono in questi Paesi hanno una probabilità elevata di appartenere al segmento dei "non interessati" e a quello degli "interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet". Infatti, questi Paesi comprendono i non possessori con una maggiore probabilità di riscontrare problemi legati al pagamento dell'abbonamento mensile e all'acquisto di un computer e/o di un modem.

5.4.3 Analisi multilivello a classi latenti per il 2009 con le covariate

Nella stima dei modelli con le covariate ho utilizzato di nuovo il metodo di scelta simultanea del numero di classi a livello 1 e 2. Ho stimato dei modelli con varie combinazioni di classi e gruppi. In Tabella 5.25 ho riportato i risultati per i modelli con tre o più classi

a livello 1. Come covariate ho utilizzato alcune variabili di livello 1 particolarmente rappresentative: Femmina, Età, Educ1, Educ2, Studia_ancora, Città e Grande_Città.

N classi	N gruppi						
	1	2	3	4	5	6	7
3	48645,4	48538,1	48544,0	48504,9	48516,8	48505,4	48537,7
4	47476,6	47263,6	47197,9	47147,3	47093,1	47103,0	47128,3
5	46654,8	46411,8	46345,4	46267,7	46263,4	46260,5	46279,1
6	46547,1	46311,9	46270,2	46219,3	46250,6	46237,7	46237,5
7	46480,1	46244,4	46124,6	46097,0	46057,9	46174,7	46191,6
8	46563,6	46301,9	46144,4	46150,7	46117,1	46071,7	46175,8

Tabella 5.25: Statistiche di sintesi del modello multilivello con le covariate secondo *Latent Gold* (2009)

Il criterio BIC suggerisce nuovamente di utilizzare un modello con 7 classi a livello 1 e 5 classi a livello 2. I coefficienti che misurano l'effetto delle classi a livello 1 sugli indicatori sono tutti significativi all'1%, ad eccezione di quello relativo ad y_2 che è significativo al 5%. Infine, i parametri che misurano l'effetto delle covariate sulle classi a livello 1 sono significativi all'1%, ad eccezione di quello della variabile che rappresenta il sesso della persona, che non è significativo a nessun livello.

Il modello presenta un'entropy-based R^2 pari a 0,78 per le classi a livello 1 e pari a 0,99 per quelle a livello 2. Come riscontrato negli altri modelli, i valori dell' R^2 relativi a y_4 e y_9 sono bassi (0,12 e 0,03).

Descrizione e interpretazione delle classi a livello 1

Nella Tabella 5.26 ho riportato le probabilità di risposta condizionate alla classe di appartenenza; le probabilità condizionate ricavate per le covariate sono riportate in Appendice 2.

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
Ampiezza	46,99%	13,37%	10,87%	10,32%	7,68%	5,63%	5,15%
y1	0,00%	0,00%	0,00%	0,00%	99,91%	0,00%	0,00%
y2	0,54%	0,49%	4,98%	69,58%	0,00%	10,29%	0,00%
y3	99,97%	3,98%	13,40%	41,48%	0,01%	64,16%	4,26%
y4	0,00%	1,58%	16,80%	0,46%	0,00%	3,07%	0,04%
y5	0,20%	23,57%	5,07%	0,63%	0,00%	32,43%	0,01%
y6	1,02%	66,59%	6,47%	1,04%	0,00%	71,10%	0,44%
y7	0,91%	37,51%	8,67%	0,43%	0,00%	69,06%	1,49%
y8	0,11%	13,66%	5,88%	0,81%	0,00%	4,49%	99,85%
y9	0,20%	3,16%	6,73%	0,52%	0,00%	2,66%	1,15%

Tabella 5.26: Probabilità condizionate $P(Y_i = 1 | X = x)$ per il modello multilivello con le covariate secondo Latent Gold (2011)

La classe 1 comprende il 47% circa delle persone che non hanno Internet a casa ed è quella dei “non interessati”. I componenti di questo segmento hanno delle probabilità più alte di avere un’età maggiore ai 50 anni, aver terminato gli studi tra i 7 e i 15 anni o successivamente e risiedere in una zona rurale e in una piccola o media città.

La classe 2 comprende il 13,4% dei non possessori e rappresenta il segmento degli “interessati, ma senza la possibilità economica per l’installazione e utilizzo di Internet”. Le probabilità condizionate più elevate si registrano per quanto concerne i giovani con meno di 24 anni, che hanno terminato gli studi dopo i 15 anni e vivono in zone non urbane e in piccole o medie città.

La classe 3 raggruppa al suo interno il 10,9% di coloro che non hanno Internet a casa. Questo segmento è quello dei “non possessori senza una precisa motivazione”, che racchiude al suo interno le persone con una probabilità più alta di vivere in una zona non coperta dalla banda larga e di essere preoccupate per l’accesso a siti non adatti. I componenti di questa classe presentano inoltre delle probabilità più elevate di avere un’età inferiore ai 40 anni, aver terminato gli studi dopo i 15 anni e di abitare in zone rurali.

La classe 4 comprende il 10,3% dei non possessori e rappresenta il segmento dei “non conoscitori di Internet”. I componenti di questo segmento registrano delle probabilità maggiori di essere anziani, aver terminato gli studi tra i 7 e i 15 anni e risiedere in una zona non urbana.

La classe 5 raggruppa al suo interno il 7,7% delle persone che non hanno Internet a casa ed è quella degli “intenzionati ad abbonarsi”. Questo segmento è costituito da persone con probabilità maggiori di avere un’età inferiore di 24 anni, aver concluso gli studi dopo i 15 anni e vivere in zone rurali e in piccole o medie città.

La classe 6 racchiude il 5,6% dei non possessori e corrisponde al segmento dei “non possessori senza la possibilità economica per l’installazione e utilizzo di Internet”. Questa classe è costituita da persone con probabilità più elevate di avere un’età compresa tra i 25 e i 39 anni, aver terminato gli studi dopo i 15 anni e vivere in una zona rurale.

L’ultima classe è costituita dal 5,2% delle persone che non possiedono Internet ed è quella degli “aventi accesso in altri luoghi”. Si tratta di persone che presentano probabilità più alte di avere un’età inferiore ai 40 anni, aver terminato gli studi dopo i 15 anni e abitare in una zona non urbana.

Descrizione e interpretazione delle classi a livello 2

Gruppi	Dimensione	Paesi inclusi
1	32,69%	Cipro, Finlandia, Grecia, Irlanda, Lussemburgo, Malta, Polonia, Portogallo, Spagna
2	25,80%	Francia, Germania, Italia, Olanda, Regno Unito, Slovenia, Svezia
3	25,73%	Austria, Belgio, Danimarca, Lituania, Rep Ceca, Slovacchia, Ungheria
4	11,49%	Bulgaria, Estonia, Lettonia
5	4,29%	Romania

Tabella 5.27: Dimensione e composizione delle classi a livello 2

	Gruppo 1	Gruppo 2	Gruppo 3	Gruppo 4	Gruppo 5
Classe 1	55,31%	48,11%	43,90%	36,98%	22,17%
Classe 2	11,83%	11,65%	17,31%	15,76%	5,47%
Classe 3	8,13%	16,42%	3,82%	16,59%	25,29%
Classe 4	11,60%	3,97%	10,55%	13,38%	29,28%
Classe 5	6,77%	10,69%	7,60%	4,96%	4,20%
Classe 6	2,12%	3,66%	10,35%	7,18%	11,77%
Classe 7	4,25%	5,50%	6,49%	5,16%	1,82%

Tabella 5.28: Probabilità di appartenere alla classe latente a livello 1, data l’appartenenza alla classe latente a livello 2

Dalla Tabella 5.27 si osserva che il primo gruppo è quello più ampio e comprende 9 Stati. Si tratta di Paesi dell'Europa Occidentale e Settentrionale (Finlandia, Irlanda, Lussemburgo), Paesi dell'Europa Mediterranea (Cipro, Grecia, Malta, Portogallo e Spagna) e dell'Europa Orientale (Polonia). Coloro che risiedono in questi Stati presentano la probabilità più alta di appartenere alla prima classe a livello 1, ossia a quella dei “non interessati” (Tabella 5.28). Inoltre, i non possessori che risiedono in questi Paesi presentano probabilità più elevate di appartenere alle classi degli “interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet” e dei “non conoscitori di Internet”. In particolare, i residenti di Finlandia, Cipro, Grecia e Malta presentano delle probabilità elevate di non essere interessati ad avere Internet; quelli di Portogallo, Grecia, Finlandia, Cipro e Malta presentano probabilità più alte di non conoscere Internet. Considerando i problemi legati ai costi, i non possessori che vivono in Irlanda, Cipro, Spagna e Portogallo registrano le probabilità maggiori di riscontrare difficoltà nei pagamenti.

Il secondo gruppo è costituito da 7 Stati, i quali appartengono all'Europa Occidentale e Settentrionale ad eccezione dell'Italia e della Slovenia. Le persone che vivono in questi Stati presentano delle probabilità maggiori di appartenere alle classi a livello 1 dei “non interessati”, dei “non possessori senza una motivazione precisa” e degli “intenzionati ad abbonarsi”. In particolare, le persone che risiedono in ciascuno di questi Paesi hanno una probabilità elevata di essere intenzionate ad abbonarsi a Internet, mentre quelle che vivono in Slovenia hanno una probabilità alta di abitare in zone non coperte dalle infrastrutture necessarie.

Il gruppo 3 comprende al suo interno 7 Stati. Si tratta principalmente di Paesi dell'Europa Orientale, come Lituania, Repubblica Ceca, Slovacchia e Ungheria; tuttavia, sono presenti anche Paesi dell'Europa Occidentale e Centrale, come Belgio, Austria e Danimarca. Le probabilità condizionate più elevate che caratterizzano le persone che risiedono in questi Paesi sono quelle relative alla classe dei “non interessati” e degli “interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet”.

Il quarto gruppo è costituito da tre Paesi dell'Europa Orientale: Bulgaria, Estonia e Lettonia. I non possessori che risiedono in questi Stati presentano probabilità condizionate più alte per diverse classi. In particolare, oltre alla classe dei non interessati, le persone che vivono negli Stati di questo gruppo hanno probabilità più elevate di riscontrare diffi-

coltà nei pagamenti (classe 2), di non conoscere Internet (classe 4) e di vivere in zone non coperte dalle infrastrutture necessarie (classe 3).

L'ultimo gruppo è costituito da un unico Paese: la Romania. Le persone che vivono in questo Stato hanno una probabilità più alta di appartenere alla classe dei "non conoscitori di Internet" e al segmento dei "non possessori senza una precisa motivazione". Dalle analisi esplorative è infatti emerso che in Romania i problemi principali sono la mancanza di infrastrutture necessarie, la poca conoscenza di Internet, probabilmente dovuta anche alle caratteristiche socio-economiche, e i problemi economici legati ai costi da sostenere per installare Internet.

5.4.4 Analisi multilivello a classi latenti per il 2011 con le covariate

N classi	N gruppi						
	1	2	3	4	5	6	7
3	36467,1	36324,4	36281,0	36262,0	36271,5	36290,8	36311,8
4	36027,8	35891,1	35834,6	35809,4	35798,1	35813,6	35835,7
5	35882,7	35762,5	35737,0	35704,7	35754,4	35854,8	35739,7
6	35881,0	35796,9	35757,7	35754,5	35725,7	35806,1	35737,9
7	35993,1	35862,8	35829,1	35832,6	35754,3	35815,5	35806,6
8	35966,1	35864,4	35840,1	35773,3	35790,8	35895,4	35934,6

Tabella 5.29: Statistiche di sintesi del modello multilivello con le covariate secondo *Latent Gold* (2011)

Il criterio BIC suggerisce di utilizzare un modello con 5 classi a livello 1 e 4 classi a livello 2 (Tabella 5.29). I coefficienti che misurano l'effetto delle classi a livello 1 sugli indicatori sono tutti significativi all'1%. Infine, i parametri che misurano l'effetto delle covariate sulle classi a livello 1 sono tutti significativi al 5%, ad eccezione di quello della variabile che rappresenta il sesso della persona e di Educ1, che non sono significativi a nessun livello.

Il valore dell'entropy-based R^2 è pari a 0,77 per le classi a livello 1 e a 0,95 per quelle a livello 2. Inoltre, gli indicatori y4 e y9 presentano dei valori molto bassi dell' R^2 , rispettivamente 0,03 e 0,01.

Descrizione e interpretazione delle classi a livello 1

Le probabilità condizionate relative alle covariate sono riportate in Appendice2, mentre la Tabella 5.30 mostra quelle relative ai diversi indicatori.

	CL 1	CL 2	CL 3	CL 4	CL 5
Ampiezza	51,71%	21,43%	11,83%	10,81%	4,22%
y1	0,10%	0,95%	48,05%	0,00%	4,58%
y2	0,00%	3,40%	1,57%	51,06%	0,00%
y3	98,13%	20,57%	1,02%	51,22%	2,08%
y4	0,03%	4,81%	5,98%	0,69%	1,48%
y5	0,35%	23,12%	0,70%	0,49%	0,84%
y6	1,45%	65,36%	1,72%	1,07%	6,21%
y7	0,76%	45,03%	1,08%	0,13%	3,49%
y8	0,40%	4,94%	0,03%	0,94%	99,40%
y9	0,36%	2,13%	3,86%	1,36%	2,59%

Tabella 5.30: Probabilità condizionate $P(Y_1 = 1 | X = x)$ per il modello multilivello con le covariate secondo *Latent Gold* (2011)

La classe 1 comprende oltre la metà delle persone senza la connessione a casa e corrisponde al segmento dei “non interessati”. Si tratta di persone che presentano probabilità più elevate di avere un’età superiore ai 52 anni, aver terminato gli studi tra i 7 e i 15 anni o successivamente e abitare in zone non urbane e in piccole o medie città.

La classe 2 raggruppa il 21,4% dei non possessori ed è quella dei “non possessori senza la possibilità economica per l’installazione e utilizzo di Internet”. I componenti di questo segmento presentano delle probabilità più alte di avere un’età inferiore ai 27 anni, aver terminato gli studi dopo i 15 anni e vivere in zone rurali e in piccole o medie città.

La classe 3 comprende l’11,8% dei non possessori. Questo segmento è quello dei “non possessori senza una motivazione precisa”, i cui componenti presentano una probabilità maggiore di essere intenzionati ad abbonarsi e vivere in zone non coperte dalle infrastrutture necessarie. Le persone riunite in questa classe presentano delle probabilità più alte di avere un’età inferiore ai 28 anni, aver terminato gli studi dopo i 15 anni e vivere in aree rurali e in piccole o medie città.

La classe 4 è costituita dal 10,8% delle persone senza Internet e corrisponde al segmento dei “non conoscitori di Internet”. I componenti di questa classe presentano delle pro-

babilità più alte di avere un'età maggiore ai 60 anni, avere un livello d'istruzione medio-basso e vivere in zone rurali.

L'ultima classe raggruppa poco più del 4% dei non possessori ed è quella degli "aventi accesso in altri luoghi". Questo segmento è costituito da persone che presentano probabilità più alte di avere un'età inferiore ai 27 anni, aver terminato gli studi dopo i 15 anni e vivere in una piccola o media città.

Descrizione e interpretazione delle classi a livello 2

Gruppi	Dimensione	Paesi inclusi
1	49,56%	Austria, Danimarca, Finlandia, Francia, Germania, Grecia, Irlanda, Italia, Lituania, Lussemburgo, Olanda, Polonia, Slovenia, Svezia
2	23,18%	Belgio, Estonia, Lettonia, Rep Ceca, Slovacchia, Ungheria
3	15,66%	Bulgaria, Malta, Romania, Spagna
4	11,61%	Cipro, Portogallo, Regno Unito

Tabella 5.31: Dimensione e composizione delle classi a livello 2

	Gruppo 1	Gruppo 2	Gruppo 3	Gruppo 4
Classe 1	53,79%	44,18%	45,81%	65,87%
Classe 2	16,81%	33,66%	20,83%	17,52%
Classe 3	15,70%	8,12%	9,51%	5,84%
Classe 4	8,89%	8,64%	21,80%	8,51%
Classe 5	4,82%	5,41%	2,06%	2,26%

Tabella 5.32: Probabilità di appartenere alla classe latente a livello 1, data l'appartenenza alla classe latente a livello 2

Il primo gruppo comprende quasi la metà delle unità ed è costituito da 14 Stati. Si tratta principalmente di Paesi dell'Europa Occidentale e Settentrionale. Sono presenti anche Paesi dell'Europa Mediterranea (Italia e Spagna) e dell'Europa Orientale (Lituania, Polonia e Slovenia) (Tabella 5.31). Coloro che vivono in questi Stati presentano probabilità più alte di appartenere alle classi a livello 1 dei non interessati, a quella dei non possessori che riscontrano difficoltà economiche e a quella di coloro che non hanno dato una motivazione precisa (Tabella 5.32). Questo gruppo riunisce al suo interno Paesi molto diversi tra loro dal punto di vista culturale ed economico. Le persone che risiedono in Lussemburgo, Grecia, Austria, Lituania e Italia sono caratterizzati da una maggiore pro-

babilità di essere intenzionate ad abbonarsi nei mesi successivi all'indagine; i residenti di Svezia, Finlandia e Danimarca (tutti Paesi del nord) registrano una probabilità elevata di essere preoccupati per l'accesso a siti non adatti. Per quanto riguarda le difficoltà economiche riscontrate, i non possessori che risiedono in Austria e Grecia hanno probabilità più alte di riscontrare problemi legati ai costi da sostenere.

Il secondo gruppo comprende 6 Stati: Belgio, Estonia, Lettonia, Repubblica Ceca e Ungheria. Sono Stati dell'Europa Orientale ad eccezione del Belgio. Le persone che vivono in questi Paesi hanno maggiori probabilità di appartenere alle classi dei "non interessati" e a quella dei "non possessori senza la possibilità economica per l'installazione e utilizzo di Internet". Dunque, le persone che abitano nei Paesi che appartengono a questo gruppo risentono dei costi eccessivi che si devono sostenere per avere Internet a casa.

Il terzo gruppo è costituito da due Stati dell'Europa Mediterranea (Spagna e Malta) e due Stati dell'est (Bulgaria e Romania). In questi Paesi le probabilità maggiori sono quelle del segmento dei non interessati, dei non possessori con difficoltà economiche e di coloro che non conoscono Internet.

L'ultimo gruppo è formato da Cipro, Portogallo e Regno Unito. In questi Paesi risiedono le persone con una probabilità più alta di non essere interessate ad avere Internet a casa e di riscontrare difficoltà economiche.

5.4.5 Residui bivariati

Nei modelli multilivello a classi latenti i residui bivariati sono difficili da interpretare. Infatti, quando non si considera la struttura gerarchica dei dati essi indicano l'associazione residua tra i diversi indicatori a livello di individuo. Al contrario, quando si considera la struttura gerarchica dei dati i residui bivariati misurano l'associazione residua tra i diversi indicatori a livello totale. Per questo motivo e visti i risultati ottenuti dall'inserimento degli effetti diretti nei modelli non multilivello, ho mantenuto i modelli stimati.

5.4.6 Procedura a tre passi

Alla fine del Capitolo 3 ho presentato una procedura alternativa per la scelta del numero di classi ai diversi livelli: la procedura a tre passi. Tale metodo permette di utilizzare dei criteri differenti nella scelta del numero di classi ai vari livelli. Ho provato a utilizzare tale procedura, ottenendo dei risultati difficili da interpretare. Infatti, in tutti i modelli tale procedura suggerisce di scegliere 8 classi a livello 1; in questo modo si ottengono delle classi con un'ampiezza esigua a cui è difficile dare un'interpretazione economica. Ho pertanto ritenuto migliori i risultati ottenuti con il metodo della scelta simultanea del numero di classi a livello 1 e 2.

5.4.7 Una sintesi dei principali risultati ottenuti

I risultati ottenuti con l'approccio multilivello non sono molto soddisfacenti, in particolare per il 2011. I diversi gruppi individuati si distinguono tra loro solo per qualche piccola differenza e non sempre la suddivisione degli Stati sembra appropriata. Tuttavia, non posso fare confronti adeguati con i modelli multilivello del Capitolo 4, perché le popolazioni di riferimento sono diverse. Da questi ultimi era infatti emerso come tra i Paesi analizzati fossero presenti delle differenze nella probabilità di non avere Internet a casa. Dai modelli a classi latenti emerge invece come tra i diversi Stati non esistano differenze importanti nelle motivazioni del non possesso di questo prodotto.

Come evidenziato dalle tabelle precedenti, tutti i segmenti individuati a livello 2 si caratterizzano per la predominanza di non possessori con un'elevata probabilità di non essere interessati ad avere l'accesso Internet a casa. La mancanza d'interesse è pertanto la motivazione principale del non possesso della connessione a casa, motivazione comune a tutti i 27 Paesi europei analizzati. Inoltre, tutti i gruppi individuati si caratterizzano per una probabilità rilevante di riscontrare difficoltà legate ai costi da sostenere. Pertanto, le differenze più evidenti tra i gruppi si registrano nelle probabilità relative alle classi a livello 1 con un'ampiezza minore. La Tabella 5.33 riassume le classi individuate a livello 1 e 2 dai modelli multilivello stimati.

Senza covariate**Classi a livello 1**

2009	2011
Non interessati (50,90%)	Non interessati (54,21%)
Non possessori senza una motivazione precisa (13,43%)	Non possessori senza una motivazione precisa (16,96%)
Interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet (11,42%)	Interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet (11,03%)
Intenzionati ad abbonarsi (7,56%)	Non conoscitori di Internet (6,06%)
Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (6,25%)	Intenzionati ad abbonarsi (4,91%)
Non conoscitori di Internet (6,00%)	Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (3,61%)
Aventi accesso in altri luoghi (4,55%)	Aventi accesso in altri luoghi (3,22%)

Classi a livello 2

2009	2011
Danimarca, Francia, Germania, Irlanda, Lussemburgo, Olanda, Polonia, Regno Unito, Slovenia, Svezia	Austria, Bulgaria, Cipro, Francia, Germania, Irlanda, Olanda, Polonia, Portogallo, Regno Unito, Svezia
Cipro, Finlandia, Grecia, Malta, Spagna	Estonia, Grecia, Italia, Lettonia, Lituania, Slovacchia
Austria, Belgio, Lituania, Portogallo, Rep Ceca, Ungheria	Finlandia, Lussemburgo, Malta, Slovenia, Spagna
Bulgaria, Estonia, Italia, Lettonia, Slovacchia Romania	Belgio, Danimarca, Romania Rep Ceca, Ungheria

Con le covariate**Classi a livello 1**

2009	2011
Non interessati (46,99%)	Non interessati (51,71%)
Interessati, ma senza la possibilità economica per l'installazione e utilizzo di Internet (13,37%)	Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (21,43%)
Non possessori senza una motivazione precisa (10,87%)	Non possessori senza una motivazione precisa (11,83%)
Non conoscitori di Internet (10,39%)	Non conoscitori di Internet (10,81%)
Intenzionati ad abbonarsi (7,68%)	Aventi accesso in altri luoghi (4,22%)
Non possessori senza la possibilità economica per l'installazione e utilizzo di Internet (5,63%)	
Aventi accesso in altri luoghi (5,15%)	

Classi a livello 2

2009	2011
-------------	-------------

5. Analisi a classi latenti

Cipro, Finlandia, Grecia, Irlanda, Lussemburgo, Malta, Polonia, Portogallo, Spagna	Austria, Danimarca, Finlandia, Francia, Germania, Grecia, Irlanda, Italia, Lituania, Lussemburgo, Olanda, Polonia, Slovenia, Svezia
Francia, Germania, Italia, Olanda, Regno Unito, Slovenia, Svezia	Belgio, Estonia, Lettonia, Rep Ceca, Slovacchia, Ungheria
Austria, Belgio, Danimarca, Lituania, Rep Ceca, Slovacchia, Ungheria	Bulgaria, Malta, Romania, Spagna
Bulgaria, Estonia, Lettonia Romania	Cipro, Portogallo, Regno Unito

Tabella 5.33: Classi individuate con l'approccio multilivello

Considerando i modelli senza le covariate, le classi a livello 1 individuate nel 2009 e nel 2011 sono le stesse; sono evidenti solo alcune differenze nelle loro dimensioni. In particolare, la classe degli "intenzionati ad abbonarsi" è meno ampia nel 2011 e questo può indicare che chi voleva abbonarsi a Internet l'ha già fatto prima di tale anno. La classe delle persone che non hanno dato una motivazione precisa è invece più ampia nel secondo istante temporale, mentre quella che comprende i non possessori, interessati e non, che ritengono troppo elevati i costi da sostenere è meno ampia.

Le classi a livello 2 sono in entrambi i casi 5, ma cambia la suddivisione dei Paesi e le risposte che caratterizzano i diversi gruppi. Nel 2009 i Paesi più sviluppati dell'Europa 27, ad eccezione di Belgio, Italia e Finlandia, condividono lo stesso segmento con Polonia e Slovenia e si caratterizzano per la probabilità maggiore di includere persone non interessate. Nel modello del 2011 i Paesi più sviluppati sono invece raggruppati con quelli che hanno raggiunto uno sviluppo minore e di cui non condividono le caratteristiche culturali ed economiche, come la Bulgaria; tuttavia, essi si caratterizzano sempre per l'elevata probabilità di non interessati.

Il Belgio, Paese con un elevato HDI e una copertura DSL del 100%, rientra in entrambi gli anni nel gruppo i cui non possessori si distinguono per la disponibilità economica insufficiente per l'installazione e il successivo utilizzo di Internet. La Finlandia, pur essendo caratterizzata dalla mancanza di interesse come gli altri Paesi dell'Europa Occidentale e Settentrionale, è aggregata con gli Stati dell'area mediterranea in quanto i non possessori che vi risiedono hanno una probabilità maggiore di non conoscere Internet. Per quanto concerne l'Italia, in entrambi gli istanti temporali il nostro Paese condivide lo stesso gruppo con i Paesi dell'Europa Orientale, gruppo caratterizzato da non possessori che registrano una probabilità maggiore di riscontrare delle difficoltà nel pagamento dell'abbonamento mensile e di accedere ad Internet in altri luoghi. Altra differenza si no-

ta per la Danimarca, che nel 2009 è riunita con gli altri Paesi sviluppati, mentre nel 2011 si trova nel gruppo caratterizzato da maggiori difficoltà economiche a causa dei costi eccessivi degli abbonamenti.

Altra differenza evidente tra i due istanti temporali è che la Romania nel 2009 costituisce un gruppo a sé che si differenzia per la presenza di non possessori con una probabilità maggiore di non conoscere Internet e di accedere in altri luoghi. Nel 2011, invece, si caratterizza in particolare per le difficoltà economiche riscontrate nel pagare l'abbonamento e acquistare il pc e/o il modem.

Un altro caso particolare è quello della Bulgaria, Paese dell'Europa Orientale con una bassa copertura DSL. Nel 2009 questo Stato è raggruppato con gli altri della stessa area in cui si registrano evidenti difficoltà economiche, mentre nel 2011 si trova nello stesso gruppo dei Paesi occidentali.

Per quanto riguarda i modelli con le covariate, nel 2011 le classi a livello 1 sono in numero inferiore rispetto al 2009. In particolare, non è presente la classe degli "interessati senza la possibilità economica per l'installazione e utilizzo di Internet", mentre quella delle persone in attesa di abbonarsi rientra probabilmente in quella di coloro che non hanno dato una motivazione specifica. A parte questo, le altre classi individuate sono le stesse del 2009 con qualche differenza nell'ampiezza.

Anche le classi a livello 2 sono in numero minore rispetto all'istante temporale precedente. Si nota immediatamente la presenza di un gruppo che raggruppa quasi tutti i Paesi dell'Europa Occidentale e Settentrionale insieme a Paesi dell'est e dell'area mediterranea. Questo gruppo comprende gli Stati in cui risiedono le persone con una maggiore probabilità di non essere interessate ad avere Internet a casa e di non aver indicato una motivazione precisa. Il secondo gruppo del 2011 può essere paragonato al terzo del 2009: entrambi si caratterizzano per la presenza di non possessori che registrano delle probabilità maggiori in riferimento alla mancanza d'interesse e alle difficoltà riscontrate nel pagamento delle spese necessarie. Il gruppo che nel 2009 è costituito solo dalla Romania, nel 2011 comprende anche Bulgaria, Malta e Spagna; le persone che risiedono in questi Paesi sono caratterizzate da una probabilità maggiore di riscontrare problemi economici e di non conoscere Internet.

Considerando invece i due modelli stimati per il 2009, le classi a livello 1 individuate con i modelli senza e con covariate sono le stesse. Si notano solo delle piccole differenze

nelle loro dimensioni; in particolare, la classe dei “non conoscitori” è più ampia nel modello con le covariate.

Maggiori diversità si riscontrano per quanto riguarda le classi a livello 2. I gruppi più simili nei due modelli sono gli ultimi tre. In particolare, è evidente come nei Paesi dell’Europa Orientale siano più diffusi i problemi legati ai costi da sostenere per avere Internet a casa, la mancanza di infrastrutture e, nel caso della Romania, la mancanza di conoscenza.

Il primo gruppo del modello senza covariate, costituito principalmente da Paesi dell’Europa Occidentale, sembra corrispondere al secondo gruppo del modello con le covariate. I Paesi compresi non presentano problemi particolari, se non la mancanza di interesse tipica delle persone di una certa età.

Una differenza evidente è che, mentre nel modello senza le covariate l’Italia si trova nel medesimo gruppo di alcuni Paesi dell’Europa Orientale, nel modello con le covariate è unita agli altri Stati più sviluppati. Il modello senza covariate oltre alla presenza di persone intenzionate ad abbonarsi, mette in evidenza i costi eccessivi degli abbonamenti; al contrario, quello con le covariate sottolinea principalmente la presenza di non possessori con una elevata probabilità di essere in attesa di abbonarsi.

Per quanto concerne i modelli stimati per il 2011, il numero di classi a livello 1 individuate dal modello con le covariate è minore rispetto a quello senza covariate; in particolare, sono state “perse” le classi degli “intenzionati ad abbonarsi” e degli “interessati, ma senza la possibilità economica per l’installazione e utilizzo di Internet”. Il segmento più ampio è sempre quello dei “non interessati”; tuttavia, quello delle persone senza una precisa motivazione è meno esteso nel modello con le covariate. Inoltre, quelli dei “non conoscitori di Internet” e degli “aventi accesso in altri luoghi” sono più numerosi. Sembra pertanto che le persone che appartenevano alle classi aggiuntive del modello senza covariate, nel modello con le covariate siano stati suddivisi tra gli altri segmenti. In particolare, le persone interessate che evidenziano difficoltà economiche sono state unite con le persone non interessate con le stesse difficoltà.

Conseguentemente, anche le classi individuate a livello 2 sono differenti. Il gruppo più esteso è sempre quello costituito per la maggior parte da Paesi dell’Europa Occidentale e Settentrionale, i cui abitanti presentano probabilità più elevate di non essere interessati e di non avere indicato nessuna motivazione precisa. Per quanto riguarda il Belgio e altri

Paesi dell'Europa Orientale, entrambi i modelli evidenziano difficoltà economiche per il pagamento delle spese da sostenere.

Con l'approccio multilivello a classi latenti è stato possibile suddividere i 27 Paesi analizzati in gruppi, sulla base della motivazione per la quale i non possessori che vi risiedono non hanno Internet a casa. Il numero di gruppi identificato è limitato e varia da 4 a 5, indicando una certa omogeneità di comportamento a livello di gruppo. Risulta inoltre che i modelli stimati per il 2011 presentano un adattamento peggiore rispetto a quelli del 2009.

Dalla stima dei modelli con e senza covariate si sono ottenuti dei gruppi che presentano differenze negli Stati inclusi. Tuttavia, i risultati di fondo sono gli stessi. In particolare, i motivi principali del non possesso di Internet comuni ai residenti di tutti gli Stati sono l'assenza d'interesse, dovuta all'età, alla mancanza delle capacità necessarie e ancor più forte nel 2011, e le difficoltà economiche legate ai costi da sostenere evidenziate dai giovani. Pertanto, le caratteristiche che distinguono maggiormente i vari segmenti riguardano le motivazioni marginali.

Dalle analisi effettuate emerge che i Paesi che si trovano geograficamente vicini e che condividono le stesse caratteristiche economiche, sociali e culturali, hanno una situazione simile per quanto concerne il non possesso di Internet a casa delle persone che vi risiedono; esistono tuttavia delle eccezioni.

Innanzitutto, i Paesi dell'Europa Occidentale e Settentrionale hanno raggiunto uno sviluppo economico notevole e presentano una copertura a banda larga totale o quasi. I non possessori di Internet che risiedono in questi Stati sono in particolare persone di una certa età, con un livello d'istruzione medio-basso e che non sono interessate ad avere la connessione da casa. Tuttavia, si riscontrano anche non possessori con un alto livello di istruzione che potrebbero rappresentare il target di una campagna europea d'informazione su tutte le possibilità offerte da Internet. All'interno di questi Stati, sono state però riscontrate delle eccezioni.

In primis il Belgio, Paese con la copertura totale e un *Human Development Index* tra i più alti d'Europa, i cui residenti hanno una probabilità rilevante di trovare eccessivi i costi che devono sostenere in particolare per l'installazione della rete e l'abbonamento mensile; anche l'Austria presenta problemi legati alle spese da sostenere. Altra eccezio-

ne è rappresentata dalle persone che risiedono in Finlandia, le quali presentano una maggiore probabilità di non conoscere Internet.

Considerando invece i Paesi dell'area mediterranea, essi presentano una situazione differente rispetto a quelli dell'Europa Occidentale e settentrionale per quanto concerne Internet. Dalle analisi esplorative è infatti emerso che questi Paesi presentano delle percentuali maggiori di non possessori, soprattutto Grecia e Portogallo.

In particolare, i non possessori che risiedono in Italia, soprattutto i giovani, trovano eccessivi i costi per gli abbonamenti mensili; tuttavia, le persone che risiedono in questo Stato sono caratterizzate da un'elevata probabilità di essere in attesa di abbonarsi e di accedere alla Rete da altri luoghi. Dalle analisi esplorative è emerso che l'Italia presenta una copertura più bassa rispetto agli altri Paesi sviluppati, copertura che non è aumentata dal 2009 al 2011.

Gli altri Paesi, ossia Spagna, Grecia, Portogallo, Cipro e Malta, oltre a presentare difficoltà economiche dovute principalmente agli abbonamenti, si caratterizzano per la presenza di persone con una maggiore probabilità di non conoscere Internet. Si tratta di persone anziane, con un basso livello di istruzione e che risiedono in zone rurali.

I Paesi dell'Europa Orientale presentano tre problemi principali. Da una parte il basso livello di istruzione delle persone che risiedono in queste zone determina una minore conoscenza di Internet. Dall'altra, il livello di sviluppo di questi Paesi causa maggiori difficoltà nel sostenimento dei costi necessari, difficoltà che colpiscono in particolare i giovani. Infine, la mancanza delle infrastrutture necessarie è un problema che ricade sulle persone che vorrebbero installare Internet a casa.

Lo Stato che nel 2009 si trovava nella situazione peggiore è la Romania, in cui i problemi economici erano affiancati dalla poca conoscenza e dalla mancanza di infrastrutture. Tuttavia, la situazione è migliorata nel 2011, anche grazie all'aumento della copertura DSL, che rimane tuttavia tra le più basse d'Europa.

Anche la Polonia presenta una copertura molto bassa che, a differenza di quella della Romania, non è aumentata dal 2009. Tuttavia, questo Stato è stato raggruppato insieme ad altri dell'Europa Occidentale e Settentrionale data l'ampia probabilità che caratterizza le persone non interessate a Internet. L'analisi a classi latenti sembra pertanto mettere in secondo piano il problema principale di questo Paese.

Ungheria e Repubblica Ceca sono i Paesi i cui residenti presentano le difficoltà economiche maggiori dovute ai costi eccessivi di computer e modem, dell'installazione iniziale e dell'abbonamento. Anche le Repubbliche baltiche (Estonia, Lettonia ed Estonia), Slovenia, Slovacchia e Bulgaria presentano problemi legati ai costi ma si riscontrano inoltre difficoltà dovute alla scarsità delle infrastrutture necessarie.

Dato il grosso problema dovuto ai costi eccessivi da sostenere per installare e poter utilizzare Internet a casa, la Commissione Europea potrebbe dare maggiore sostegno alle persone più disagiate attraverso degli incentivi. Questo intervento deve tuttavia essere accompagnato da una campagna che informi della possibilità di ricevere aiuti economici. Al contrario, il problema delle infrastrutture è già stato risolto nel 2013, ma anche in questo caso sarebbe utile informare i cittadini europei del raggiungimento della copertura totale.

Conclusioni

Internet è diventato l'infrastruttura principale che ci permette di comunicare, informarci, acquistare prodotti e servizi, partecipare alla vita politica del Paese, eliminando le barriere geografiche e migliorando la qualità della vita. Internet è uno strumento molto caro alla Commissione Europea, che vuole rendere accessibile la Rete alla maggior parte dei cittadini. Infatti, anche se Internet oggi è ampiamente diffuso, esistono ancora degli Stati in cui il numero di cittadini che vi accedono è limitato.

In questa tesi ho analizzato i dati raccolti con le indagini Eurobarometro Speciale 335 e 365 di novembre 2009 e febbraio 2011, relative alla diffusione delle tecnologie digitali in Europa, in particolare di Internet. Nella mia analisi ho considerato oltre alle caratteristiche demografiche degli intervistati e il possesso di beni durevoli tra cui Internet, il motivo del non possesso della connessione a casa. In questo modo ho potuto raggiungere i miei obiettivi iniziali, che riguardavano l'identificazione del profilo tipico dei non possessori e la loro segmentazione sulla base della motivazione.

Risultati importanti sono emersi dalle analisi esplorative iniziali, in cui ho considerato anche delle caratteristiche a livello di Paese tra cui la copertura DSL. In particolare, è emerso che i Paesi più sviluppati, ossia quelli dell'Europa Occidentale e Settentrionale e alcuni Paesi dell'area mediterranea (Italia, Grecia e Portogallo) hanno una copertura maggiore rispetto agli altri. Al contrario, Polonia, Bulgaria, Romania e Slovacchia sono gli Stati con la copertura DSL più bassa.

Le analisi dei sottocampioni dei non possessori hanno evidenziato il *digital divide* presente tra i Paesi più sviluppati e quelli con un livello di sviluppo minore. Inoltre, hanno

aiutato a definire le principali caratteristiche dei non possessori che sono state poi confermate dai risultati delle stime dei modelli probit.

Con la stima dei modelli probit univariati ho determinato il profilo tipico dei non possessori che rimane pressoché lo stesso nei due istanti temporali. Si tratta principalmente di persone anziane, che hanno terminato gli studi tra i 7 e i 15 anni, vivono in una zona rurale e appartengono al livello basso della società. Sono soprattutto persone che vivono in una famiglia poco numerosa, non hanno un'auto, ma possiedono una casa di proprietà.

Utilizzando l'approccio multilivello è stato possibile evidenziare alcune caratteristiche ulteriori: si tratta in particolare di persone che vivono in Stati con copertura DSL e HDI bassi. Tuttavia, nel 2011 la copertura DSL del Paese di residenza non ha però alcun impatto sul non possesso di Internet a casa. Questo suggerisce che nel secondo istante temporale la presenza delle infrastrutture non è più motivo di differenza tra gli Stati.

Con l'analisi a classi latenti è stato possibile segmentare i non possessori sulla base della loro motivazione. Ho utilizzato 9 variabili dicotomiche come indicatori, ognuna relativa ad un preciso motivo. I risultati ottenuti con il plugin LCA di *Stata* non sono stati molto soddisfacenti in quanto non rappresentano bene la realtà dei sottocampioni, soffermandosi principalmente sulle questioni economiche e tralasciando motivazioni marginali.

Con l'utilizzo di *Latent Gold* si sono ottenuti dei risultati migliori, in particolare con l'inserimento delle covariate che ha permesso di descrivere in maniera più dettagliata la realtà dei non possessori. I risultati ottenuti dalla stima dei modelli con e senza covariate sono leggermente diversi, ma i risultati principali sono gli stessi. Innanzitutto, il segmento più numeroso è quello dei "non interessati", costituito prevalentemente da persone di una certa età che risiedono in zone rurali. Un altro segmento identificato è quello dei "non possessori senza una motivazione precisa", segmento che riunisce soprattutto coloro che vivono in zone non coperte dalla banda larga, accedono ad Internet in altri luoghi o sono preoccupati per l'accesso a siti non adatti. A seconda dei modelli stimati questo segmento è rimasto intatto o suddiviso in classi specifiche in base alle motivazioni, come accaduto per il 2009. In particolare, sono stati individuati i segmenti dei "non conoscitori", che comprende in particolare le persone anziane, con un basso livello di istruzione che risiedono in zone rurali, degli "aventi accesso in altri luoghi" e degli "interessati ma senza la copertura alla banda larga", costituiti per lo più da giovani che hanno finito gli studi da poco e vivono in zone rurali. Un'altra classe individuata è quella relativa alle

difficoltà economiche riscontrate dai non possessori, in particolare giovani che appartengono al livello basso della società. Infine, è stato individuato il segmento degli “intenzionati ad abbonarsi”, costituito ancora da giovani, che risulta essere prevedibilmente più ampio nel 2009.

Con l’approccio a classi latenti multilivello è stato possibile suddividere i Paesi in gruppi. I risultati sono stati poco soddisfacenti, in particolare per il 2011. Il numero di gruppi determinato è limitato in tutti i modelli, indicando che le differenze maggiori si hanno a livello di individui e non di Paese. Inoltre, tutti i gruppi individuati si caratterizzano per la presenza di persone con una alta probabilità di non essere interessate e di riscontrare difficoltà economiche. Le caratteristiche distintive dei vari gruppi si hanno per quanto riguarda le motivazioni di minore importanza. In particolare è emerso che i Paesi dell’Europa Occidentale e Settentrionale, con qualche eccezione, sono caratterizzati da persone con un’elevata probabilità di rispondere che non sono interessate ad avere Internet e in attesa di abbonarsi; i non possessori che risiedono nei Paesi del Mediterraneo sono caratterizzati da una probabilità più alta di non conoscere Internet e di considerare troppo elevati i costi da sostenere; coloro che vivono nei Paesi dell’Europa Orientale presentano maggiori probabilità di abitare in zone non coperte dalle infrastrutture, riscontrare difficoltà economiche e non conoscere Internet.

A fronte dei risultati ottenuti si può concludere che la Commissione Europea dovrebbe focalizzarsi principalmente sul problema dei costi troppo elevati riscontrato dai giovani e ancor più evidente nei Paesi dell’Europa Orientale. Il futuro è rappresentato dai giovani e se non viene data loro la possibilità di connettersi a Internet non potranno restare al passo con i tempi e sfruttare al massimo tutto ciò che la tecnologia offre, restando al di fuori della vita sociale, politica ed economica.

Il problema della mancanza di interesse, pur essendo la questione più diffusa, non è così preoccupante, in quanto i non interessati sono persone di una certa età che difficilmente cambierebbero i loro usi. Ci si potrebbe focalizzare invece sui “non interessati” con un livello d’istruzione superiore, per spronarli a utilizzare la tecnologia disponibile e non restare al di fuori dell’evoluzione tecnologica.

Per quanto riguarda invece i problemi che ho definito come marginali, essi caratterizzano principalmente i non possessori dell’Europa Orientale, ma non costituiscono degli ostacoli eccessivi. Infatti, il problema legato alla scarsità di infrastrutture è già stato risolto con il raggiungimento della copertura del 100% ottenuta grazie al satellitare. Oc-

corre, tuttavia, informare i cittadini europei dell'obiettivo raggiunto e della possibilità che viene loro offerta. Per quanto concerne coloro che accedono alla Rete da altri luoghi, si tratta di persone con un livello d'istruzione elevato che hanno le capacità necessarie per navigare nel web e vi accedono già; si tratta quindi di persone che fanno già parte dell'era di Internet. Al contrario, coloro che non sanno esattamente cosa sia Internet sono persone anziane che non hanno le capacità necessarie per navigare e sfruttare le possibilità che questa tecnologia offre.

Infine, le persone che hanno risposto di essere intenzionate ad abbonarsi nei mesi successivi all'indagine non rappresentano un problema, a meno che non abbiano ancora sottoscritto un abbonamento per il sopraggiungere di altre cause.

Informazioni maggiori si sarebbero ottenute se i campioni delle due indagini fossero stati gli stessi, in quanto si sarebbe potuto controllare l'evolversi della situazione e i cambiamenti avvenuti nel corso dei due anni considerati.

Appendice 1

Probabilità condizionate $P(Z = z | X = x)$ per il modello a 7 classi latenti del 2009.

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
Età							
1-23	9,25%	42,59%	3,43%	32,80%	45,90%	17,03%	40,64%
24-39	14,75%	28,19%	8,15%	29,98%	29,79%	24,03%	23,25%
40-49	22,18%	15,80%	17,16%	20,79%	12,79%	24,32%	20,64%
50-58	27,00%	9,23%	29,96%	10,65%	7,29%	20,54%	11,21%
59-83	26,82%	4,19%	41,29%	5,78%	4,22%	14,08%	4,26%
Educ1							
No	57,97%	81,46%	44,84%	87,85%	85,81%	76,59%	80,59%
Sì	42,03%	18,54%	55,16%	12,15%	14,19%	23,41%	19,41%
Educ2							
No	44,25%	24,68%	61,42%	17,36%	22,63%	26,51%	27,67%
Sì	55,75%	75,32%	38,58%	82,64%	77,37%	73,49%	72,33%
Studia_ancora							
No	99,43%	94,40%	99,99%	95,57%	92,07%	99,34%	91,74%
Sì	0,57%	5,60%	0,01%	4,43%	7,93%	0,66%	8,26%
Femmina							
No	44,08%	42,45%	39,88%	44,85%	44,38%	43,52%	43,38%
Sì	55,92%	57,55%	60,12%	55,15%	55,62%	56,48%	56,62%
Città							
No	65,93%	64,20%	73,65%	65,99%	60,23%	68,23%	81,39%
Sì	34,07%	35,80%	26,35%	34,01%	39,77%	31,77%	18,61%
Grande_Città							
No	75,58%	71,51%	82,03%	71,54%	75,83%	73,77%	96,80%
Sì	24,42%	28,49%	17,97%	28,46%	24,17%	26,23%	3,20%
Bambini							
No	93,98%	74,76%	97,44%	82,98%	73,79%	87,84%	65,58%
Sì	6,02%	25,24%	2,56%	17,02%	26,21%	12,16%	34,42%
Ragazzi							
No	97,21%	86,05%	97,69%	88,58%	85,17%	95,09%	85,99%
Sì	2,79%	13,95%	2,31%	11,42%	14,83%	4,91%	14,01%

Appendice I

Componenti							
1	37,45%	22,69%	44,06%	21,32%	18,67%	35,41%	10,80%
2	45,81%	29,29%	46,01%	34,92%	34,15%	41,83%	29,14%
3	9,57%	20,92%	6,02%	19,87%	20,46%	11,31%	20,82%
4+	7,17%	27,10%	3,91%	23,89%	26,72%	11,45%	39,24%
Liv_medio							
No	49,07%	57,21%	61,69%	51,20%	50,51%	65,54%	52,53%
Sì	50,93%	42,79%	38,31%	48,80%	49,49%	34,46%	47,47%
Liv_Alto							
No	84,55%	86,90%	88,02%	69,86%	76,21%	84,60%	77,93%
Sì	15,45%	13,10%	11,98%	30,14%	23,79%	15,40%	22,07%
Auto							
No	49,21%	51,86%	68,48%	32,20%	32,10%	55,21%	29,04%
Sì	50,79%	48,14%	31,52%	67,80%	67,90%	44,79%	70,96%
Casa							
No	33,22%	52,80%	28,46%	38,49%	53,20%	34,81%	37,78%
Sì	66,78%	47,20%	71,54%	61,51%	46,80%	65,19%	62,22%

Probabilità condizionate $P(Z = z | X = x)$ per il modello a 5 classi latenti del 2011.

	CL 1	CL 2	CL 3	CL 4	CL 5
Età					
1-27	8,00%	36,81%	41,31%	5,83%	47,52%
28-42	16,23%	28,57%	28,39%	9,62%	23,14%
43-51	23,04%	16,98%	14,66%	18,72%	14,51%
52-59	25,52%	11,53%	8,95%	23,66%	8,15%
60-83	27,21%	6,12%	6,69%	42,17%	6,67%
Educl					
No	55,62%	73,75%	85,27%	41,77%	83,26%
Sì	44,38%	26,25%	14,73%	58,23%	16,74%
Educ2					
No	46,20%	29,56%	23,01%	63,55%	24,57%
Sì	53,80%	70,44%	76,99%	36,45%	75,43%
Studia_ancora					
No	99,83%	97,16%	91,77%	99,69%	92,64%
Sì	0,17%	2,84%	8,23%	0,31%	7,36%
Femmina					
No	44,19%	43,46%	47,32%	40,33%	46,03%
Sì	55,81%	56,54%	52,68%	59,67%	53,97%
Città					
No	61,07%	64,02%	58,18%	67,26%	60,13%
Sì	38,93%	35,98%	41,82%	32,74%	39,87%
Grande_Città					
No	78,25%	75,54%	78,38%	83,62%	77,37%
Sì	21,75%	24,46%	21,62%	16,38%	22,63%
Bambini					
No	95,90%	79,74%	83,74%	95,55%	79,34%
Sì	4,10%	20,26%	16,26%	4,45%	20,66%
Ragazzi					
No	98,28%	88,59%	88,11%	99,11%	90,12%
Sì	1,72%	11,41%	11,89%	0,89%	9,88%

Appendice I

Componenti					
1	39,49%	26,96%	23,70%	47,48%	23,52%
2	48,34%	33,21%	31,55%	43,72%	32,70%
3	7,32%	16,85%	19,56%	5,62%	22,17%
4+	4,84%	22,98%	25,19%	3,18%	21,61%
Liv_medio					
No	52,46%	62,60%	50,03%	58,61%	47,22%
Sì	47,54%	37,40%	49,97%	41,39%	52,78%
Liv_Alto					
No	85,47%	91,16%	75,83%	88,09%	79,33%
Sì	14,53%	8,84%	24,17%	11,91%	20,67%
Auto					
No	52,37%	63,40%	38,25%	70,76%	39,76%
Sì	47,63%	36,60%	61,75%	29,24%	60,24%
Casa					
No	31,54%	47,93%	47,41%	25,38%	46,57%
Sì	68,46%	52,07%	52,59%	74,62%	53,43%

Residui bivariati per i modelli stimati con *Latent Gold*.

Modello senza covariate (2009)									
Indicatori	y1	y2	y3	y4	y5	y6	y7	y8	y9
y1	.								
y2	0,0161	.							
y3	0,0505	0,3823	.						
y4	0,0033	1,2734	0,0758	.					
y5	0,0089	0,6514	0,0233	0	.				
y6	0,0225	0,0352	0,2967	0	0,3735	.			
y7	0,0162	0,1949	0,6483	4,5333	0,8124	0,0027	.		
y8	0,0181	0,2499	3,6836	0,1653	0,2142	0,547	0,0284	.	
y9	0,0022	1,7993	0,3599	2,1622	0,0405	1,6435	3,6003	0,332	.
Modello senza covariate (2011)									
Indicatori	y1	y2	y3	y4	y5	y6	y7	y8	y9
y1	.								
y2	0,001	.							
y3	0,016	1,0414	.						
y4	0,2726	1,3447	0,0876	.					
y5	0,0421	0,2758	0,1085	0,073	.				
y6	0,085	0,0128	0,0174	0,0064	0,0204	.			
y7	0,1552	0,6729	0,0024	0,0229	0,3253	0	.		
y8	0,3847	0,2027	0,1507	0,3448	0,4567	0,0088	0,0779	.	
y9	4,8346	3,722	2,4439	1,2627	2,1688	0,021	0,5517	1,974	.
Modello con covariate (2009)									
Indicatori	y1	y2	y3	y4	y5	y6	y7	y8	y9
y1	.								
y2	0,0127	.							
y3	0,0415	0,0074	.						
y4	0,0109	0,277	0,0266	.					
y5	0,0083	0,7361	0,0513	2,0556	.				
y6	0,0193	0,0047	0,0132	0,048	4,2273	.			
y7	0,015	0,1608	0,0291	1,7183	0,1313	0,121	.		

Appendice I

y8	0,0131	0,1339	0,1134	0,0039	0,7802	1,0439	0,9074	.	.
y9	0,0022	0,2062	0,0298	1,4944	0,3779	0,2261	2,7021	20,228	.
Modello con covariate (2011)									
Indicatori	y1	y2	y3	y4	y5	y6	y7	y8	y9
y1
y2	0,0793
y3	0,0146	0,6822
y4	0,0032	16,1643	0,3653
y5	0,8004	0,1389	1,9026	0,1703
y6	0,038	0,4558	0,4056	0,8203	8,2113
y7	1,6209	1,0411	0,6043	0,1006	3,7407	0,9912	.	.	.
y8	0,0881	3,9888	0,399	21,3499	1,5578	0,037	2,0535	.	.
y9	7,7725	0,0008	0,2511	1,5247	0,5163	0,0039	2,7011	0,831	.

Appendice 2

Probabilità condizionate $P(Z = z | X = x)$ per il modello a 7 classi e 5 gruppi del 2009.

	CL 1	CL 2	CL 3	CL 4	CL 5	CL 6	CL 7
Età							
1-23	9,07%	45,59%	27,95%	2,30%	45,91%	15,83%	35,67%
24-39	14,69%	28,42%	23,31%	7,26%	29,79%	24,72%	32,46%
40-49	21,97%	14,83%	21,81%	16,43%	12,79%	24,83%	20,53%
50-58	27,06%	7,84%	16,92%	30,34%	7,29%	20,72%	7,54%
59-83	27,22%	3,32%	10,01%	43,67%	4,22%	13,91%	3,79%
Educ1							
No	57,60%	82,19%	77,70%	43,10%	85,81%	77,14%	89,38%
Sì	42,40%	17,81%	22,30%	56,90%	14,19%	22,86%	10,62%
Educ2							
No	44,63%	23,70%	27,99%	63,17%	22,63%	26,44%	15,70%
Sì	55,37%	76,30%	72,01%	36,83%	77,37%	73,56%	84,30%
Studia_ancora							
No	99,47%	94,14%	95,89%	100,00%	92,07%	99,23%	95,18%
Sì	0,53%	5,86%	4,11%	0,00%	7,93%	0,77%	4,82%
Femmina							
No	43,96%	42,46%	45,94%	39,89%	44,38%	42,61%	41,81%
Sì	56,04%	57,54%	54,06%	60,11%	55,62%	57,39%	58,19%
Città							
No	65,89%	62,62%	71,65%	73,10%	60,23%	68,45%	65,13%
Sì	34,11%	37,38%	28,35%	26,90%	39,77%	31,55%	34,87%
Grande_Città							
No	75,64%	70,89%	77,59%	81,76%	75,83%	72,72%	73,87%
Sì	24,36%	29,11%	22,41%	18,24%	24,17%	27,28%	26,13%

Probabilità condizionate $P(Z = z | X = x)$ per il modello a 5 classi e 4 gruppi del 2011.

	CL 1	CL 2	CL 3	CL 4	CL 5
Età					
1 - 27	8,26%	36,84%	43,77%	4,91%	40,74%
28 - 42	16,40%	28,35%	24,43%	10,06%	33,12%
43 - 51	23,08%	17,06%	14,76%	19,12%	14,82%
52 - 59	25,42%	11,71%	9,21%	24,61%	7,03%
60 - 83	26,84%	6,06%	7,83%	41,29%	4,28%
Educ1					
No	55,83%	74,24%	83,72%	43,65%	87,82%
Sì	44,17%	25,76%	16,28%	56,35%	12,18%
Educ2					
No	45,94%	29,44%	24,78%	61,13%	19,13%
Sì	54,06%	70,56%	75,22%	38,87%	80,87%
Studia ancora					
No	99,81%	96,70%	92,18%	99,61%	93,06%
Sì	0,19%	3,30%	7,82%	0,39%	6,94%
Femmina					
No	44,08%	44,14%	46,14%	41,37%	47,97%
Sì	55,92%	55,86%	53,86%	58,63%	52,03%
Città					
No	60,99%	63,04%	61,24%	66,09%	54,99%
Sì	39,01%	36,96%	38,76%	33,91%	45,01%
Grande_Città					
No	78,19%	75,78%	77,38%	83,37%	78,05%
Sì	21,81%	24,22%	22,62%	16,63%	21,95%

Bibliografia

- Belch G. E., Belch M. A., *Advertising and Promotion: an integrated marketing communications perspective*, McGraw-Hill/Irwin: Burr Ridge, 2003
- Bidgoli H., *The Internet Encyclopedia*, Volume 2, John Wiley & Sons, 2003
- Cappuccio N., Orsi R., *Econometria*, Il Mulino, 2005
- Commissione Europea, *Report: E-communications Household Survey*, TNS Opinion & Social, Ottobre 2010
- Commissione Europea, *Report: E-communications Household Survey*, TNS Opinion & Social, Luglio 2011
- Commissione Europea, *Report: E-communications Household Survey*, TNS Opinion & Social, 2013
- De Leeuw J., Meijer E., *Handbook of Multilevel Analysis*, Springer, 2008
- GESIS, *Eurobarometer 72.5*, GESIS – Leibniz Institute for the Social Sciences, 2013
- Goldstein H., Healy M. J. R., The graphical presentation of a collection of means, *Journal of the Royal Statistical Society*, 158: 175-177, 1995
- Goldstein H., *Multilevel Statistical Models*, John Wiley & Sons, 2010
- Goodman L.A., The Analysis of systems of qualitative variables when some of the variables are unobservable, *American Journal of Sociology*, 79: 1179-1259, 1974
- Greene W. H., *Econometric Analysis*, Person, 2003
- Haberman S. J., *Analysis of Qualitative Data, Vol. 2, New Developments*, Academic Press, 1979

- Hox J.J., *Applied Multilevel Analysis*, TT-Publikaties, 1995
- Hox J. J., *Multilevel Analysis: Techniques and Applications*, The Psychology Press, 2010
- Kreft I., De Leeuw, *Introducing Multilevel Modeling*, Sage Publications Ltd, 1999
- Kotler P., *Il Marketing dalla A alla Z*, John Wiley & Sons, 2010
- Lanza S. T., Dziak J. J., Huang L., Wagner A. T., Collins L. M., *LCA Stata Plugin Users' Guide*, The Pennsylvania State University, 2013
- Lazarsfeld P.F., Henry N. W., *Latent structure Analysis*, Houghton Mifflin, 1968
- Lukočienė O., Varriale R., Vermunt J.K., The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis, *Social Methodology*, 40 (1): 247-283, 2010
- McCutcheon A. L., *Latent Class Analysis*, Sage Publications, 1987
- Ulrich M., *eDemocracy: Public Online Engagement by the European Commission*, VDM Verlag
- Vermunt J.K., Magidson J., Latent Class Models for Classification, *Computational statistics and data analysis*, 41: 531-537, 2003
- Vermunt J. K., Applications of Latent Class Analysis in Social Science Research, *Symbolic and quantitative approaches to reasoning with uncertainty proceeding*, 2711: 22-36, 2003
- Vermunt J.K., Magidson J., Latent Class Analysis, *The Sage Encyclopedia of Social Sciences Research Methods*, 549-553, 2004
- Vermunt J.K., Multilevel Latent Class Models, *Social Methodology*, 33: 213-239, 2003
- Vermunt J. K., Magidson J., *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Statistical Innovations Inc., 2005
- Vermunt J. K., Magidson J., *Latent Gold 4.0 User's Guide*, Statistical Innovations Inc., 2005
- Vermunt J. K., Magidson J., *Latent Gold 5.0 Upgrade Manual*, Statistical Innovations Inc., 2013
- Wooldridge J.M., *Econometrics analysis of cross section and panel data*, The MIT

Press, 2002

Siti di riferimento

<http://www.itu.int/ITU-D/ict/definitions/regions/index.htm>

<http://www.internetworldstats.com/stats.htm>

<http://ec.europa.eu/digital-agenda/>

http://ec.europa.eu/italia/attualita/primo_piano/informazione/agendadigitale_valutazione_2013_it.htm

http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/broadband_coverage_2010.pdf

http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=une_rt_m&lang=en

<http://hdrstats.undp.org/en/indicators/103106.html>

Ringraziamenti

Alla fine del mio percorso di studi vorrei ringraziare tutti coloro che mi sono stati vicini e che, in un modo o nell'altro, mi hanno dato una mano in questi anni.

Grazie ai miei genitori che mi hanno permesso di studiare in questi anni e per tutti gli insegnamenti e i consigli che mi hanno dato.

Grazie ad Andrea, Claudia & C. per la costante presenza e per tutto ciò che avete fatto da quando ero piccola.

Un ringraziamento particolare a Marco per l'enorme pazienza dimostrata in questi due anni e più, per avermi sempre supportata (e sopportata) e avere sempre creduto in me. Grazie per avermi spronata fin dall'inizio a dare il massimo e per tutto quello che fai nella vita di tutti i giorni.

Grazie ad Arianna, Cristina, Elisa e Luana per tutte le avventure passate e le esperienze fatte, perché insieme siamo cresciute trovando sempre tempo per la nostra amicizia. Grazie anche ai rispettivi boys che sanno sempre come portare allegria alle serate trascorse insieme.

Grazie a Stefania, Alice ed Elena per tutti i giorni trascorsi insieme all'università, le ore di studio e i risultati che siamo riuscite ad ottenere.

Un ringraziamento speciale al prof Omar Paccagnella per la grande pazienza e disponibilità con cui mi ha accompagnato in questo lavoro di tesi.

Infine, grazie a tutte le altre persone che ho conosciuto e che mi stanno vicine, persone con cui trascorro momenti indimenticabili e che si ricordano di me.

