

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Statistiche

Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Relazione finale

La salute dei giovani e l'impatto del "living arrangement" in Italia

Relatrice
Prof.ssa Fausta Ongaro

Laureando
Elia Sebastiano Pagin
Matricola N° 2010163

Anno accademico 2022/2023

Ai miei cari e ai loro sacrifici.

Abstract

In un paese come l'Italia, dove i giovani solitamente prolungano la loro permanenza nella casa dei genitori più a lungo rispetto ad altri paesi europei, emerge un interrogativo fondamentale: quale impatto ha questa situazione sul loro benessere? Questa ricerca si propone di esaminare se i giovani italiani percepiscano la convivenza con i genitori come un ostacolo rispetto ai loro coetanei europei, oppure se questa realtà non incida sulla loro salute complessiva.

Indice

1	Introduzione	1
1.1	Presentazione del problema	1
1.2	La salute	2
1.3	I living arrangements	3
1.3.1	Le forme di living arrangements in Europa	4
1.3.2	Un approfondimento sui living arrangements in Italia	5
2	Esplorazione preliminare dei dati: il processo di selezione	7
2.1	Caratteristiche del dataset	8
2.2	I punti chiave della ricerca	12
2.2.1	Lo stato di salute	13
2.2.2	Il living arrangement	14
2.2.3	Le caratteristiche individuali	14
2.2.4	Le caratteristiche familiari	17
2.3	Intrecci di benessere: la salute e le sue connessioni	20
2.3.1	Lo stato di salute in rapporto al living arrangement	20
2.3.2	Lo stato di salute in rapporto alle caratteristiche individuali	22
2.3.3	Lo stato di salute in rapporto alle caratteristiche familiari	26
3	Esplorando i dati attraverso modelli	31
3.1	Il modello binomiale	31
3.1.1	Il collegamento tra la salute e i living arrangements	33
3.1.2	L'impatto delle caratteristiche individuali	34
3.1.3	L'impatto delle caratteristiche familiari	36
3.1.4	Le interazioni tra i diversi aspetti	37
3.1.5	La valutazione della bontà dei modelli	40

3.2	Il modello per risposte politomiche	42
3.2.1	Un collegamento alternativo tra la salute e i living arrangements	43
3.2.2	L'effetto delle altre caratteristiche	45
4	Conclusioni	49
A	Creazione <i>Convegn</i>	51
	Riferimenti bibliografici	55

Capitolo 1

Introduzione

1.1 Presentazione del problema

L'obiettivo di questa tesi è indagare la relazione tra lo stato di salute dei giovani in Italia e i living arrangements, con l'obiettivo di comprendere come la salute percepita varia in diverse circostanze.

Ci proponiamo di esaminare come lo stato di salute dei giovani italiani sia influenzato dalla loro convivenza con i genitori. L'Italia è notoriamente uno dei paesi europei in cui i giovani tendono a vivere più a lungo con i propri genitori, e ciò solleva una serie di interrogativi: i giovani italiani godono comunque di buona salute o percepiscono questa condizione come limitante? Quali gruppi di giovani si trovano in condizioni di salute peggiori? Sono maschi o femmine? Appartengono alle fasce d'età dei ventenni o dei trentenni? Sono occupati o disoccupati? Si trovano in situazioni economiche precarie?

Per rispondere a queste domande, inizieremo fornendo le definizioni di salute e di living arrangement, faremo un confronto tra le diverse forme di convivenza esistenti in Europa, con un'attenzione particolare al contesto italiano.

Successivamente, utilizzeremo i dati raccolti dall'indagine Istat del 2016 su "Famiglie, soggetti sociali e ciclo di vita" per condurre le nostre analisi. Dopo aver eseguito una pulizia del dataset iniziale, selezioneremo le variabili più rilevanti rispetto allo scopo della relazione in base alla letteratura esistente, tra cui lo stato di salute, il living arrangement, il genere, l'età e altre variabili di interesse. Condurremo un'analisi univariata di queste variabili al fine di comprendere la composizione del nostro campione e il profilo generale delle distribuzioni. Successivamente,

esamineremo le relazioni di dipendenza tra lo stato di salute e le altre variabili.

Lo stato di salute verrà utilizzato con due diverse codifiche: uno in scala originale Likert con 5 livelli e uno come variabile dicotomica che aggrega le categorie migliori e peggiori. Questa distinzione costituirà il fulcro delle analisi successive.

Utilizzeremo modelli logistici e modelli per risposte politomiche per modellare lo stato di salute, impiegando procedure di *stepwise selection* e *forward selection* per determinare i migliori modelli e interpretare i coefficienti ottenuti.

Infine, forniremo risposte esaustive alle domande iniziali, basate sui risultati delle analisi del campione e dei modelli elaborati.

1.2 La salute

Nel 1948 l'Organizzazione Mondiale della Sanità ha definito la **salute** come «*uno stato di completo benessere fisico, mentale e sociale*» venendo riconosciuta come un diritto inalienabile dell'essere umano senza distinzione di razza, religione, credenza politica, condizione economica o sociale. Si tratta della prima volta che il concetto di salute viene visto come condizione soggettiva dell'individuo e non più come la sola assenza di malattia.

Questa definizione porta ad una nuova visione, dove gli aspetti biologici si fondono in modo completo con gli aspetti psichici e sociali della persona, la salute diventa quindi una condizione dualistica inscindibile, cioè che non permette di distinguere il benessere tra il sé e il corpo, ma che tiene conto del quadro più generale nel quale gli individui esperiscono le proprie vite. Proprio per questo motivo la salute di due individui diversi è molto difficile da comparare, infatti una stessa condizione può essere poco significativa per una persona e risultare invalidante per la vita di un'altra.

Tuttavia, la definizione proposta dall'OMS solleva anche diverse problematiche. In primo luogo, essa comprende una vasta gamma di elementi che determinano il benessere della persona, i quali sono soggettivi e profondamente influenzati dai modelli culturali dominanti. In secondo luogo, data la sua ampiezza, diventa praticamente irrealizzabile per gli Stati garantire un'equa tutela del diritto alla salute per tutti i propri cittadini, considerando infatti le limitate risorse disponibili, i molteplici settori di intervento sociale e i numerosi aspetti sui quali sarebbe necessario intervenire per garantire a tutti uno stato completo di benessere fisico, psicologico e sociale (Nobile et al. [2013](#)).

1.3 I living arrangements

Il termine "**living arrangements**" si riferisce alle diverse forme familiari e alle situazioni abitative in cui un individuo si trova a vivere. Tradizionalmente, il processo di transizione allo stato adulto seguiva un percorso lineare, dopo la permanenza e la maturazione nel nucleo familiare d'origine seguiva il suo abbandono definitivo e la creazione di una nuova famiglia. Tuttavia, negli ultimi anni, questo andamento lineare ha mostrato una complessità crescente, infatti è sempre più frequente che un individuo torni a vivere nel nucleo della famiglia d'origine che aveva abbandonato in precedenza, decidendo di fermarsi o avviando successivamente un nuovo percorso di vita. Questi cambiamenti nei living arrangements sono influenzati da molteplici fattori, come l'economia, la cultura, le tradizioni, l'urbanizzazione e i mutamenti dei valori sociali, riflettendo la crescente diversità delle situazioni abitative nella società moderna.

L'Europa non mostra uniformità nei tempi e nei modi delle diverse fasi familiari del processo di transizione allo stato adulto. In particolare, nei paesi nordici e occidentali, l'indipendenza abitativa è un valore fortemente sostenuto dai giovani, spinti anche da un sistema di welfare pubblico che offre un supporto significativo. Di conseguenza, i giovani di queste regioni tendono a lasciare la casa dei genitori in età più precoce rispetto ai coetanei del sud o dell'est Europa. Per esempio, in Finlandia, l'età media di uscita dalla casa dei genitori è di 20 anni per le femmine e di 22 anni per i maschi, mentre in Italia supera i 27 anni e i 30 anni rispettivamente (Schwanitz e Mulder 2015).

Ad ogni modo, va notato che questo fenomeno non può essere attribuito esclusivamente alle condizioni economiche del paese. Molti studi indicano che influenze culturali e storiche hanno un ruolo significativo. Ad esempio, nelle società mediterranee, la famiglia è spesso considerata un punto di riferimento e di supporto a lungo termine, portando i giovani a rimanere con i genitori per un periodo più lungo prima di intraprendere una vita indipendente (Tosi e Grundy 2018).

Questa diversità di living arrangements e percorsi di vita è il risultato di un complesso intreccio di fattori sociali, culturali ed economici che caratterizzano le diverse regioni europee. Inoltre, la globalizzazione e i cambiamenti nei modelli familiari stanno anche contribuendo a ridefinire le tradizionali fasi familiari di transizione allo stato adulto in tutta Europa.

1.3.1 Le forme di living arrangements in Europa

Le forme dei living arrangements sono influenzate dall'organizzazione e dalle disposizioni di cura sociale, ovvero il welfare. Lungo tutta l'Europa le diversità dei sistemi di tutela sono notevoli.

Nei paesi mediterranei, come Spagna, Italia o Grecia, la cura dei figli o degli anziani è principalmente svolta a livello domestico e ci sono poche leggi che supportano i giovani o i genitori. Al contrario, nei paesi centro settentrionali, come Danimarca, Svezia, Norvegia, Finlandia, Belgio e Francia, la cura dei figli e degli anziani è supportata dallo stato, e i giovani possono contare su diverse agevolazioni e servizi governativi per raggiungere l'autonomia abitativa (Schwanitz e Mulder 2015).

Concentrandoci sulla situazione dei giovani in Europa secondo la transizione allo stato adulto, evidenziamo alcune diversità significative:

1. I giovani tendono a lasciare la casa dei genitori in età più giovane nelle regioni nordiche e occidentali d'Europa, con un'età media che varia circa tra i 20 e i 22 anni, mentre nelle regioni mediterranee e orientali l'età media sale tra i 27 e i 30.
2. Il vivere in coppia è diffuso tra i giovani, ma con differenze significative riguardo all'età in cui avviene. Nel Nord e nell'Ovest Europa, si verifica prima rispetto al Sud ed Est Europa. Inoltre, nelle regioni settentrionali e occidentali, si tratta spesso di convivenza senza matrimonio, mentre in quelle mediterranee e orientali, i giovani tendono a vivere con un partner con cui sono sposati.
3. Nel Nord e nell'Ovest Europa, le dimensioni familiari sono relativamente ridotte in numero di componenti, e il numero di famiglie unipersonali è elevato. Vivere da soli nelle regioni settentrionali è molto comune, soprattutto in giovane età, poiché dopo aver lasciato la casa dei genitori è abituale fare questa esperienza prima di convivere o sposarsi. Nel Sud e nell'Est Europa, invece, le famiglie d'origine sono più numerose e coinvolgono spesso nuclei estesi, nelle regioni più orientali non è raro che all'interno della stessa casa vivano i genitori e i figli sposati con i relativi partner o altri parenti.

1.3.2 Un approfondimento sui living arrangements in Italia

Dopo aver esaminato le notevoli differenze nei modelli di living arrangements in Europa, ci concentriamo sull'Italia. Spesso la letteratura trascura le variazioni regionali, trattando l'Italia come un'entità omogenea in tutto il territorio, ma questa non è la realtà.

La capacità di emanciparsi dai genitori viene spesso associata all'occupazione, poiché è logico presumere che una maggiore occupazione in una determinata area significhi disponibilità di maggiori risorse economiche e facilità nella creazione di nuovi nuclei familiari. Tuttavia, questa logica non sempre si applica, i motivi per cui si rimane in casa con i genitori riflettono anche l'esistenza di contesti economici differenti: nel 2016 la difficoltà nel trovare un lavoro o un lavoro stabile è indicata di più dai giovani del Mezzogiorno (circa 3 su 10), mentre i costi eccessivi per l'affitto o l'acquisto della casa scoraggiano soprattutto i giovani che risiedono nel Nord (circa un terzo). Lo studio è un motivo di permanenza dichiarato soprattutto dai giovani del Centro (Billari e Tomassini 2021).

Non è quindi possibile ricollegare direttamente l'uscita dalla casa dei genitori alla situazione occupazionale, ma è necessario considerare che la decisione di emanciparsi dai genitori è influenzata da diversi fattori, tra cui l'occupazione, ma anche le condizioni economiche generali, i costi immobiliari e persino il percorso di studio intrapreso.

Per comprendere a fondo il modello italiano, è essenziale concentrarsi sul contesto culturale che influenza le relazioni familiari. Come descritto dalla letteratura (Reher 1998), l'Italia fa parte dell'area mediterranea, dove i legami familiari sono solidi e la coesione familiare è elevata, con la famiglia che gode di priorità rispetto all'individuo.

Il modo in cui i genitori italiani si relazionano con i loro figli è influenzato dalla tendenza a enfatizzare la dimensione affettiva, sebbene ciò possa comportare diverse conseguenze, come una ritardata maturità, scarsa autonomia e responsabilità (Rosina, Fraboni e Sabbadini 2003). In risposta a questo atteggiamento dei genitori, i figli reagiscono con fiducia e affetto verso di loro.

Ci sono notevoli differenze regionali nella forza dei legami familiari in Italia. Secondo l'approccio familistico, queste differenze hanno radici storiche profonde. Durante il periodo che va dal Medioevo fino alla metà del diciannovesimo secolo, era comune per i giovani italiani lasciare casa presto, sposarsi e continuare a vivere vicino ai loro genitori, creando un unico sistema familiare chiamato *complex family*.

La vicinanza tra le generazioni implicava un continuo scambio di beni e servizi dai genitori ai figli, contribuendo così a rafforzare i legami familiari (Reher 1998).

Questo influsso culturale ha lasciato una duratura impronta anche nei tempi moderni, poiché le differenze nella prossimità geografica e nel supporto ricevuto riflettono ancora il legame intergenerazionale e hanno un impatto significativo sulla sistemazione dei figli (Dalla Zuanna, Michielin e Bordignon 2007).

Le differenze regionali nella sistemazione dei giovani in Italia possono essere suddivise in tre macroaree, ognuna con caratteristiche distintive (Tomassini e Vignoli 2023):

- Nord: In questa regione, i giovani di entrambi i sessi escono prima relativamente al resto della penisola dalla casa dei genitori e, quando decidono di uscire, optano per esperienze di residenza più moderne, come la convivenza con un partner, un collega o uno studente. Qui il matrimonio avviene più tardi, nel 2019 l'età media al primo matrimonio è di 34 anni, simile al Centro, e il motivo del ritardo è dovuto alla ricerca di stabilità lavorativa e relazionale.
- Sud: Nel Sud, i giovani lasciano la casa dei genitori in età più avanzata rispetto al resto d'Italia. La causa di questa uscita è spesso il matrimonio, infatti per quasi l'80% di essi è il motivo principale, contro il 70% del Nord e il 73% del Centro. Tuttavia, anche dopo il matrimonio, molti giovani continuano a vivere in prossimità della casa in cui sono cresciuti.
- Centro: La parte centrale dell'Italia si trova in una posizione intermedia rispetto alle altre due macroaree, con una tendenza più vicina a quella del Nord. Qui, i modelli familiari tradizionali persistono ancora, ma si stanno diffondendo anche abitudini innovative riguardo alla sistemazione dei giovani.

Queste distinzioni regionali riflettono le diverse tradizioni culturali, le norme sociali e le opportunità economiche presenti nelle varie parti della penisola italiana (Santarelli e Cottone 2009).

Capitolo 2

Esplorazione preliminare dei dati: il processo di selezione

In questo capitolo viene presentato il dataset oggetto delle analisi, compreso il suo percorso di pulizia e adattamento alle esigenze di questa relazione, le analisi univariate e bivariate.

Vengono utilizzati i dati ricavati dall'indagine Istat del 2016 su *Famiglie, Soggetti sociali e Ciclo di vita*. L'indagine campionaria fa parte del sistema integrato di indagini sociali (Sistema di indagini Multiscopo sulle famiglie), costituisce una delle principali fonti statistiche sulla struttura familiare e sulle caratteristiche sociali della famiglia in Italia. Vengono affrontate diverse tematiche su punti del vivere quotidiano individuale e familiare, gli aspetti più importanti colti durante la rilevazione sono:

- Il ciclo di vita individuale e familiare;
- I rapporti interni alla famiglia;
- La permanenza dei giovani in famiglia e le intenzioni di lasciare la famiglia di origine;
- La mobilità sociale;
- L'economia familiare;
- I percorsi lavorativi e formativi;

- Il rapporto con il mondo del lavoro e la ricerca del lavoro;

Nel 2016 l'indagine è stata svolta a Dicembre su un campione di 32 585 individui di 18 anni e più, distribuiti in 852 comuni italiani di diversa ampiezza demografica. Le interviste sono state effettuate da rilevatori comunali presso l'abitazione dell'individuo campione, con tecnica PAPI (Paper And Pencil Interview), estratto casualmente dalle liste anagrafiche del comune.

2.1 Caratteristiche del dataset

Il dataset che viene utilizzato è di grandi dimensioni e contiene le risposte a tutte le domande del questionario somministrato ad ogni individuo facente parte del campione. Al fine di poter lavorare e ricavare informazioni utili viene fatta una cesura sulla base delle caratteristiche dell'unità statistiche e sulle variabili di interesse. La popolazione di studio è quella *giovane* e secondo la definizione Istat un giovane è un individuo di età compresa dai 15 ai 34 anni. Dato il vincolo della somministrazione del questionario ad un campione di individui di 18 anni e più, è necessario applicare un cambiamento alla definizione posta da Istat sulla popolazione giovane, ridefinendola come «Tutti gli individui di età compresa dai 18 ai 34 anni». In questo modo abbiamo una congruenza tra la definizione di popolazione giovane e la popolazione campionaria utile all'analisi, la cui dimensione è di 5 084 unità statistiche distribuita in modo omogeneo tra le diverse età come si può vedere dalla Figura 2.1.

Il dataset contiene 1359 variabili che nella loro totalità non sono utili ai fini della relazione, è necessario operare una grossa cernita, mantenendo solo le variabili che possono servire, in base alla letteratura e all'utilità nello svolgimento delle analisi. Alla fine del processo di selezione sono state mantenute 11 variabili *core* che si ritengono significative, funzionali allo studio e congruenti alla letteratura.

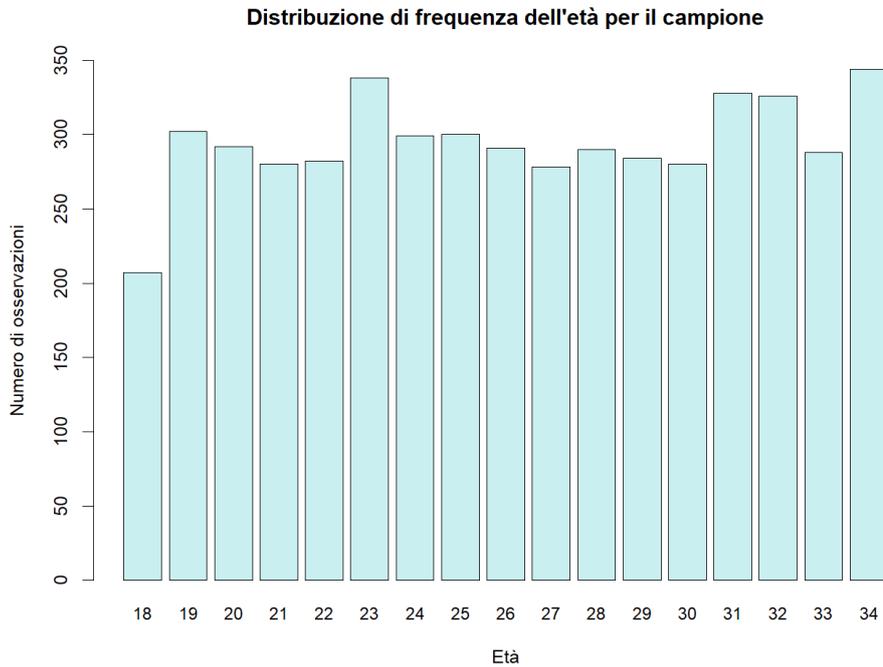


Figura 2.1: Distribuzione di frequenza per età intervistati 18-34 anni.

In base ai diversi ambiti di appartenenza di ognuna, di seguito ne viene riportata una breve descrizione organizzata in 4 gruppi per praticità, il primo contiene la variabile risposta, mentre i rimanenti riguardano le variabili esplicative:

1. Lo stato di salute

- *Salute*: indica lo stato di salute generale dell'intervistato, rilevata come giudizio verbale ordinale (scala Likert) in risposta alla domanda "Come va la sua salute in generale?", composta da 5 modalità ("Molto bene", "Bene", "Né bene, né male", "Male", "Molto male") che si assumono equidistanti una dall'altra;

2. Il living arrangement

- *Convgen*: è un indicatore creato ad hoc per questa relazione a partire da due variabili originarie *madviv2* e *padviv2*. Le due variabili indicavano se la madre o il padre vivessero con l'intervistato ed entrambe presentavano 3 modalità: No, vive altrove; No, è deceduto; Si, vive con me. A questo

punto è stata applicata un ricodifica alle due variabili originarie, unendo le modalità negative e creando due variabili dicotomiche. Dopodiché sono state incrociate e si è ottenuto la variabile *Convgen* che serve ad indicare se un individuo vive senza i genitori, con la madre, con il padre o con entrambi i genitori. A questa è stata applicata un'ulteriore ricodifica che la trasforma in variabile dicotomica, esprimendo il fatto che un individuo vive con almeno un genitore o vive senza i genitori. Per il procedimento approfondito della creazione di questa variabile fare riferimento all'Appendice A.

Questa variabile sarà fondamentale nel corso delle analisi, visto che sarà la discriminante per il confronto tra i diversi gruppi e la variabile proxy per il "living arrangement";

3. Le caratteristiche individuali

- *Sesso, Eta*: caratteristiche proprie dell'individuo che indicano il sesso e l'età in anni compiuti. *Sesso* è una variabile dicotomica che assume valori "Maschio" o "Femmina", mentre *Eta* rilevata come una variabile quantitativa discreta, viene ricodificata in classi d'età, non omogenee in ampiezza, ma che danno un'idea dei diversi stili di vita che si hanno durante la fase giovanile.
- *Cond*: condizione professionale del soggetto, viene fornita l'informazione sul principale stato di occupazione al momento dell'intervista, caratteristica che riguarda la sfera dei percorsi formativi e lavorativi, inizialmente rilevata come variabile qualitativa sconnessa su 8 modalità (Occupato, In cerca di nuova occupazione, In cerca di prima occupazione, Casalinga, Studente, Ritirato dal lavoro, Inabile al lavoro o Altra condizione) e ricodificata in 4 modalità accorpendo alcune specificità non utili alle analisi;
- *Rip*: caratteristica demografica dell'individuo che indica la macroarea di appartenenza secondo le ripartizioni Istat (Nord-Ovest, Nord-Est, Centro, Sud, Isole), ricodificata accorpendo le modalità in "Nord", "Centro" e "Sud", in modo che si abbia una maggiore facilità di interpretazione dei risultati ed eliminando dei dettagli confondenti;

4. Le caratteristiche familiari

- *Titmad, Titpad*: sono caratteristiche dei genitori, cioè il titolo di studio più alto conseguito dalla madre e dal padre, inizialmente rilevato come variabile qualitativa ordinale con 13 modalità e successivamente ricodificata in sole 4 modalità, che mantengono l'ordine originale ed eliminando un livello di specificità troppo elevato. Queste variabili sono utili per spiegare i rapporti interni alla famiglia, che influenzano sin dalla nascita l'individuo e la sua fase di sviluppo, potendo creare dei benefici o anche aspetti negativi;
- *Conmad*: descrive la condizione professionale della madre quando l'intervistato aveva 14 anni, ossia in età adolescenziale, è noto che questo fattore influenzi la stabilità del figlio nella relazione con i genitori sia nel presente ma anche nel futuro, portando a complicazioni o legami più stretti. La variabile è rilevata secondo 10 modalità iniziali e ricodificata in 4 modalità ai fini della relazione;
- *Gensep*: variabile indicatrice che segnala se i genitori dell'intervistato sono separati. La separazione è un evento per la maggior parte dei casi traumatico nei confronti dei figli che può avere grosse ripercussioni nel rapporto con un genitore o con entrambi portandoli ad avere difficili rapporti con essi e il conseguente abbandono della casa nelle età più giovani. La variabile è rilevata secondo due modalità, "Si" e "No", in risposta alla domanda "I suoi genitori hanno mai smesso di vivere insieme, per separazione o divorzio o, se non sposati, per interruzione della relazione affettiva?", con l'aggiunta di una modalità a posteriori, "Non so" per racchiudere anche gli individui che hanno preferito non rispondere alla domanda.
- *Riseco*: è un giudizio da parte dell'intervistato sulle risorse economiche proprie (nel caso viva da solo) o della famiglia negli ultimi 12 mesi, rilevata come variabile qualitativa su scala ordinale secondo 4 modalità, utilizzata come approssimazione dell'economia familiare.

Come ultimo passo prima di iniziare con le analisi, si effettua un controllo sui valori mancanti del dataset appena costruito.

Per quanto riguarda la variabile di interesse *Salute* è necessario per effettuare le analisi che non ci sia nemmeno un valore missing e per verificarlo si analizza la sua distribuzione, riportata nella Tabella 2.1.

Tabella 2.1: Distribuzione di frequenza della variabile risposta *Salute* intervistati 18-34 anni

	<i>Salute</i>
Molto bene	2438
Bene	2265
Né bene, né male	248
Male	43
Molto male	15
NA	75

Si nota che la variabile risposta presenta dei dati mancanti indicati dalla tabella nella riga NA, vengono quindi escluse dal campione queste unità statistiche, della quale non si conosce la risposta alla domanda sulla salute generale.

In riferimento alle altre variabili selezionate nel dataset solo la variabile *Convgen* presenta dei valori mancanti, Tabella 2.2. Visto il numero molto basso di NA(19) pari allo 0.38% dell'intero campione, si decide di escludere anche queste osservazioni

Tabella 2.2: Distribuzione di frequenza della variabile *Convgen* intervistati 18-34 anni

<i>Convgen</i>	n	percent
Vive senza i genitori	1711	34.16
Vive solo con la madre	573	11.44
Vive solo con il padre	141	2.81
Vive con entrambi i genitori	2565	51.21
NA	19	0.38

In questo modo il dataset finale di lavoro sarà composto da 11 variabili e 4 990 osservazioni con nessun valore mancante sia per la variabile risposta che per le altre variabili considerate.

2.2 I punti chiave della ricerca

Nel paragrafo seguente viene effettuata un'analisi descrittiva univariata dei caratteri a disposizione attraverso grafici e tabelle per dare un'idea generale delle caratteristiche del campione e della distribuzione delle variabili prese in considerazione.

2.2.1 Lo stato di salute

L'analisi comincia dalla distribuzione delle risposte alla domanda «Come va in generale la sua salute?», cioè dalla variabile dipendente *Salute*. Come si può vedere dalla Figura 2.2 quasi tutte le risposte sono distribuite tra le prime due modalità, più precisamente il campione risponde "Molto bene" 2425 volte e "Bene" per 2261, vuol dire che queste due contengono il 93.91% di tutte le risposte ed il restante 6.09% si identifica nelle modalità peggiori. Alla luce di questi dati si può dire che lo stato di salute generale percepito dai giovani in Italia, fortunatamente, è molto buono.

Se si considera *Salute* come una variabile dicotomica, raggruppando le prime due modalità in "Bene" e le restanti tre in "Male", le distanze tra le modalità vengono annullate e viene persa l'informazione sull'ordinamento, ma si guadagna in termini di interpretazione, diventa infatti immediato capire se un individuo sta bene o male.

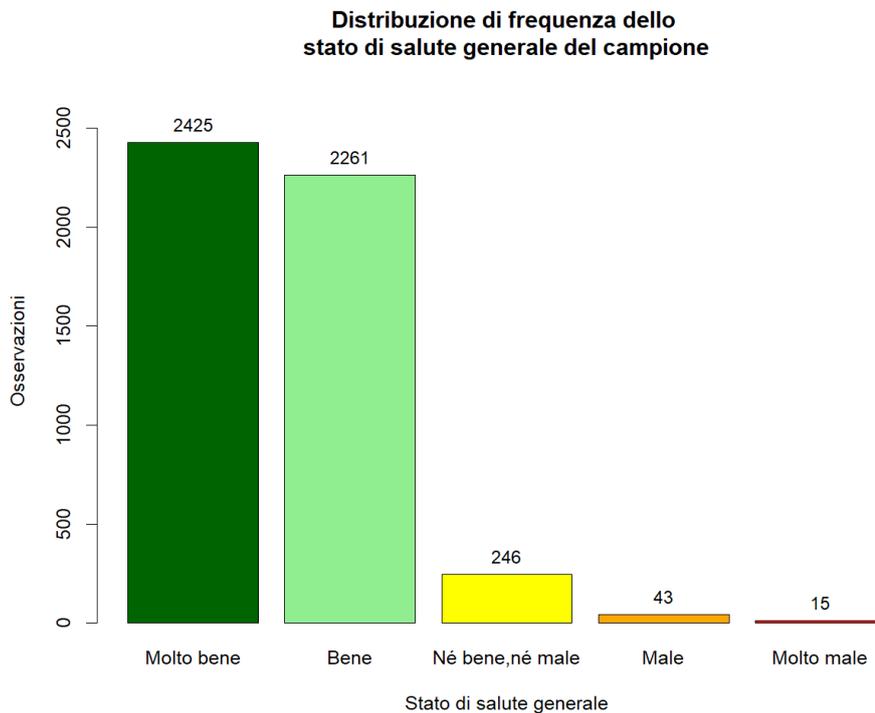


Figura 2.2: Distribuzione di frequenza assoluta dello stato di salute intervistati 18-34 anni.

2.2.2 Il living arrangement

Per l'analisi di *Convgen*, ossia la variabile che indica il living arrangement dell'intervistato, si guarda la distribuzione di frequenza, Tabella 2.3, che presenta 4 modalità distinte derivate dall'intersezione delle due variabili originali, vedere Appendice A. Dalle distribuzioni si evidenzia che il 34.29% del campione vive già autonomamente, mentre il restante 65.71% vive per la maggior parte con entrambi i genitori (51.4%), se invece vive con un solo genitore, in prevalenza lo fa con la madre (11.48%) rispetto al padre (2.83%).

Tabella 2.3: Intervistati 18-34 anni per living arrangement, valori assoluti e distribuzione percentuale

<i>Convgen</i>	n	%
Vive senza i genitori	1711	34.29
Vive solo con la madre	573	11.48
Vive solo con il padre	141	2.83
Vive con entrambi i genitori	2565	51.40

2.2.3 Le caratteristiche individuali

Per quanto riguarda le caratteristiche proprie dell'individuo, si propongono due grafici, la Figura 2.3 mostra la distribuzione di frequenza, all'interno della barra assoluta e al di sopra percentuale, del campione rispetto alla variabile *Sesso*. La Figura 2.4 presenta la distribuzione di frequenza, allo stesso modo di quella precedente, per la variabile *Eta* ricodificata in tre classi non omogenee in ampiezza ma che ben rappresentano le diverse fasi di vita di un individuo, infatti nella prima classe "18-24" si troveranno molti individui che possono essere ancora in età scolare e che abitano verosimilmente con i genitori, nella classe "25-29" si avranno molti individui che entrano nel mondo del lavoro e che creano legami affettivi esterni al nucleo familiare, infine nell'ultima classe "30-34" presumibilmente si avrà la fase dell'autonomia abitativa e formazione di nuovi nuclei.

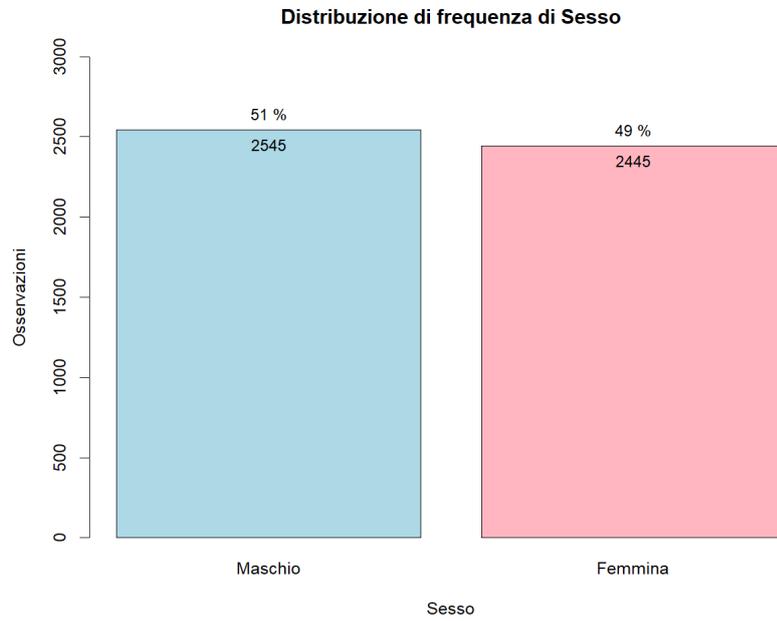


Figura 2.3: Intervistati 18-34 anni per genere, valori assoluti e distribuzione percentuale.

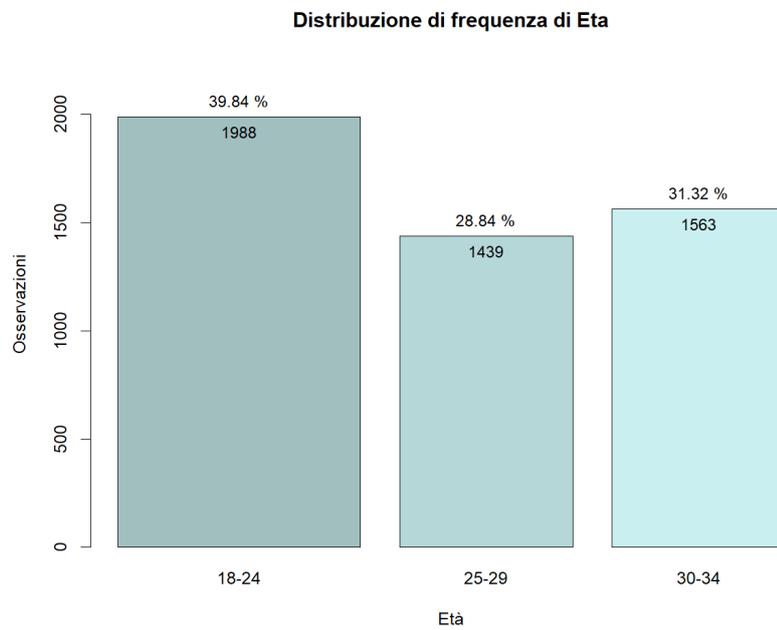


Figura 2.4: Intervistati 18-34 anni per classi d'età, valori assoluti e distribuzione percentuale.

I risultati mostrano che il campione è equidistribuito rispetto alla variabile *Sesso*, mentre la variabile *Eta* presenta numerosità non omogenee rispetto alle tre classi d'età, con la prima classe che racchiude il 39.84% dell'intero campione, questo perché l'ampiezza della classe non è uguale alle altre due.

In riferimento alle caratteristiche del soggetto nell'ambito formazione e lavoro si analizza la variabile *Cond*, che indica la condizione professionale al momento dell'intervista, ricodificata per questa analisi in 4 modalità, utili ad una immediata comprensione dello stato lavorativo. I risultati della distribuzione di frequenza sono visibili in Tabella 2.4. Nella variabile originale era tenuta in considerazione anche la condizione di casalinga che è stata accorpata in "Altra condizione" perché poco presente nelle classi d'età inferiori e negli individui di sesso maschile. Si nota che il 51.62% del campione è occupato e il 25.17% si trova invece nella condizione di "Studente", il resto del campione è distribuito nelle restanti 2 modalità.

Tabella 2.4: Intervistati 18-34 anni per condizione professionale, valori assoluti e distribuzione percentuale

<i>Cond</i>	n	%
Occupato	2576	51.62
In cerca di lavoro	637	12.77
Studente	1256	25.17
Altra condizione	521	10.44

Per analizzare le caratteristiche demografiche degli individui si utilizza la variabile *Rip*, ossia la ripartizione geografica inizialmente codificata da Istat in 5 suddivisioni e successivamente ricodificata per questa analisi in 3 macroaree, in modo che si abbia una maggiore facilità di interpretazione dei risultati. Le 3 ripartizioni geografiche ottenute sono "Nord", "Centro" e "Sud", dove Nord comprende Nord-Ovest e Nord-Est mentre Sud comprende Sud ed Isole, rimane invariato il Centro.

Tabella 2.5: Intervistati 18-34 anni per macroarea di appartenenza, valori assoluti e distribuzione percentuale

<i>Rip</i>	n	%
Nord	2064	41.36
Centro	806	16.15
Sud	2120	42.48

Dalla Tabella 2.5 si riscontra che "Nord" e "Sud" hanno all'incirca la stessa numerosità, mentre il "Centro" rappresenta solo il 16.15% del campione.

2.2.4 Le caratteristiche familiari

Di seguito vengono analizzate le caratteristiche della famiglia d'origine del soggetto intervistato, si usano inizialmente come variabili *Titmad* e *Titpad*, ossia il titolo di studio più alto conseguito dai genitori. È stata apportata una ricodifica rispetto alle variabili originali, che contavano ben 13 modalità ordinali secondo la classificazione Istat dei titoli di studio, accorpate poi in 4 modalità sempre ordinali. Dalle Tabelle 2.6 e 2.7, si evidenzia che la distribuzione di frequenza dei titoli di studio dei genitori è molto simile tra madre e padre, con i padri che hanno in maggior numero un titolo di studio alto, 10.12%, rispetto alle madri 8.10%, inversamente alla modalità "Titolo medio" in cui le madri hanno una maggiore frequenza dei padri, se invece si considera la modalità "Non so", per i padri si ha una frequenza quasi doppia rispetto a quella delle madri.

Tabella 2.6: Intervistati 18-34 anni per titolo di studio della madre, valori assoluti e distribuzione percentuale

<i>Titmad</i>	n	%
Titolo alto	404	8.10
Titolo medio	1981	39.70
Titolo basso	2529	50.68
Non so	76	1.52

Tabella 2.7: Intervistati 18-34 anni per titolo di studio del padre, valori assoluti e distribuzione percentuale

<i>Titpad</i>	n	%
Titolo alto	505	10.12
Titolo medio	1762	35.31
Titolo basso	2595	52.00
Non so	128	2.57

Dalla letteratura è noto che il titolo dei genitori nella maggior parte dei casi è uguale o simile, per verificare questo si crea una tabella a doppia entrata con le frequenze assolute per le variabili *Titmad* e *Titpad*, dopodiché se ne misura la concordanza attraverso l'indice τ_b di Kendall. Questo indice misura la concordanza della stessa variabile ordinale espressa da due osservatori diversi, applicando una forzatura visto che la modalità "Non so" non rispetta l'ordine dei titoli di studio, ma ai fini dell'analisi si ritiene sia inferiore a "Nessun titolo".

Dalla Tabella 2.8 si può notare come la diagonale principale contenga la maggior parte delle osservazioni, tenendo conto dei ranghi ripetuti l'indice τ_b di Kendall restituisce un valore pari a 0.57, con un intervallo di confidenza di livello 0.95 pari

Tabella 2.8: Intervistati 18-34 anni per titolo di studio dei genitori, distribuzione congiunta in valori assoluti

<i>Titmad/Titpad</i>	Titolo alto	Titolo medio	Titolo basso	Non so
Titolo alto	229	140	33	2
Titolo medio	238	1176	535	32
Titolo basso	34	441	2022	32
Non so	4	5	5	62

a (0.55,0.59), il che significa che i titoli di studio dei genitori hanno una relazione monotona positiva e una buona concordanza, ossia all'aumentare della modalità di uno aumenta anche l'altro. Si può concludere che i genitori tendono ad avere lo stesso titolo di studio, per questo viene mantenuto solo il titolo del padre, *Titpad* per le analisi successive.

Continuando ad analizzare le caratteristiche familiari si propone la Tabella 2.9 che mostra la distribuzione di frequenza assoluta e percentuale del campione rispetto alla condizione professionale della madre quando l'intervistato aveva 14 anni, ossia era in età adolescenziale. Si osserva che più di metà del campione aveva la madre occupata, 54.51%, mentre quasi tutta la restante metà aveva la madre casalinga, 41.30%, solo una piccola parte, circa un 4%, aveva la madre nella condizione di disoccupata (in cerca di lavoro o di primo lavoro) o in altra condizione (risultante dall'accorpamento delle modalità originali: Studentessa, Inabile al lavoro, Ritirata dal lavoro, Deceduta, Altra condizione).

Tabella 2.9: Intervistati 18-34 anni per condizione professionale della madre in età adolescenziale, valori assoluti e distribuzione percentuale

<i>Conmad</i>	n	%
Occupata	2720	54.51
In cerca di lavoro	51	1.02
Casalinga	2061	41.30
Altra condizione	158	3.17

Un evento molto significativo nella vita dei figli è la stabilità della relazione dei genitori e quando viene a mancare posso presentarsi grandi problemi nel rapporto genitori-figli, per questo viene analizzata la variabile *Gensep* che tiene conto se i genitori dell'intervistato sono separati o meno. Nella Tabella 2.10, che presenta le

distribuzioni delle frequenze assolute e percentuali della condizione della relazione dei genitori, si evidenzia che l'84.35% del campione non ha i genitori separati, mentre il 14.61% ha i genitori separati. Una piccola percentuale del campione, l'1.04% non è a conoscenza della condizione della relazione tra i genitori.

Tabella 2.10: Intervistati 18-34 anni per condizione della relazione dei genitori, valori assoluti e distribuzione percentuale

<i>Gensep</i>	n	%
No	4209	0.8435
Sì	729	0.1461
Non so	52	0.0104

Per approfondire la questione dell'economia familiare si analizza la variabile *Riseco*, ovvero un giudizio da parte dell'intervistato sulle risorse economiche negli ultimi 12 mesi della famiglia d'origine, se vive con i genitori, o la propria nel caso l'intervistato sia già uscito di casa, esprimibile attraverso 3 modalità.

Dalla Tabella 2.11 si vede che il 70.92% dei rispondenti pensa di avere a disposizione buone risorse economiche, il 26.65% invece esprime un giudizio negativo, ritenendo che le risorse economiche nell'ultimo anno siano state inadeguate, il 2.63% del campione non ha espresso nessun giudizio.

Tabella 2.11: Intervistati 18-34 anni per giudizio sulle condizione economica, valori assoluti e distribuzione percentuale

<i>Riseco</i>	n	%
Ottime o adeguate	3539	70.92
Scarse o insufficienti	1320	26.45
Non so	131	2.63

2.3 Intrecci di benessere: la salute e le sue connessioni

In questo paragrafo si analizzano le relazioni tra la variabile risposta *Salute*, nella codifica dicotomica e multivariata, rispetto a tutte le altre 10 variabili core presenti nel dataset. Lo stato di salute viene codificato in due modi diversi, un formato breve (dicotomico: *Salute_d*) che rende subito interpretabile il risultato dato la sua natura dualistica ed esclusiva, infatti un individuo in questo modo può stare o bene o male, mentre nel formato esteso (multivariato: *Salute_m*), un individuo può presentare una serie di stati di salute più dettagliata, ovvero con 5 diverse modalità ordinali che vanno dal molto bene al molto male, con la distanza tra questi stati assunta equivalente. Vengono utilizzate queste due codifiche per ottenere in primo luogo, grazie allo stato di salute in formato dicotomico una visione d'insieme del fenomeno, con una successiva analisi poi, grazie al formato lungo, si mettono in luce dettagli meno evidenti del fenomeno.

Mantenendo l'ordine delle analisi univariate, si parte ad analizzare la dipendenza tra lo stato di salute e il living arrangement, passando successivamente alle caratteristiche individuali ed infine a quelle familiari, le analisi si effettuano attraverso tabelle e grafici, verificando l'indipendenza.

2.3.1 Lo stato di salute in rapporto al living arrangement

Per valutare la relazione tra lo stato di salute e il living arrangement, entrambe in formato dicotomico, si fa riferimento alla distribuzione congiunta e condizionata della salute rispetto alla convivenza con i genitori, Tabella 2.12. Si nota che le distribuzioni condizionate sono molto simili alla marginale, infatti il test del chi-quadrato di Pearson restituisce un p-value pari a 0.126, è possibile quindi rifiutare l'ipotesi di associazione delle due variabili al 10%.

Se si considera invece lo stato di salute in formato esteso i risultati cambiano molto, avendo a disposizione più informazioni è possibile individuare l'andamento di fondo del fenomeno. Nella Figura 2.5 è riportato il grafico della distribuzione percentuale condizionata di *Salute_m*, forma multivariata, rispetto al living arrangement in formato dicotomico, da cui è possibile vedere come chi vive con i genitori e sta bene, la maggior parte appartiene alla categoria "Molto bene", stessa cosa se si considera chi sta male, chi vive con i genitori sta "meno" male, ossia appartiene

a categorie migliori. Anche il test sull'indipendenza ci restituisce lo stesso quadro, con un p-value osservato di $1.702 \cdot 10^{-11}$, cioè l'ipotesi nulla di indipendenza viene rifiutata.

Tabella 2.12: Stato di salute percepito, in formato breve, intervistati 18-34 anni per living arrangement, valori assoluti e distribuzione percentuale condizionata

<i>Convgen/Salute_d</i>	Bene	Male	Totale
Vive senza i genitori	1594 93.16%	117 6.84%	1711 100%
Vive con almeno un genitore	3092 94.30%	187 5.70%	3279 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: 0.126

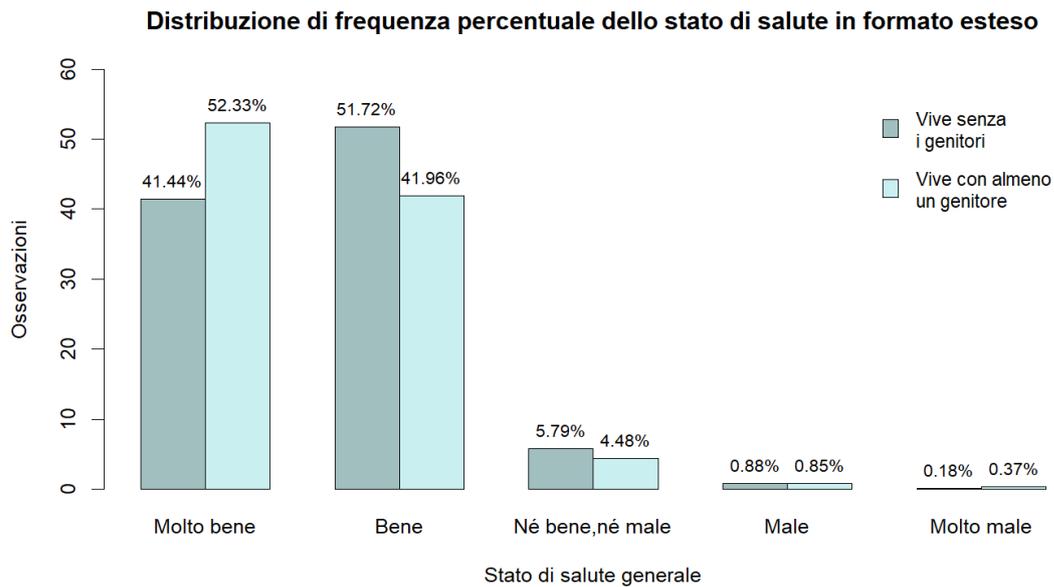


Figura 2.5: Stato di salute percepito, in formato esteso, intervistati 18-34 anni per living arrangement, valori percentuali.

Si ottengono due risultati opposti in base alla codifica della variabile *Salute*, questo perché il test utilizzato è molto sensibile alle numerosità delle celle, per ora

non viene esclusa nessuna delle due ipotesi, ma verrà verificata in seguito durante le analisi del capitolo successivo.

2.3.2 Lo stato di salute in rapporto alle caratteristiche individuali

Con lo stato di salute in formato breve, Tabella 2.13 si nota che è non possibile accettare l'ipotesi nulla di indipendenza tra le variabili *Salute_d* e *Sesso* ad un livello del 5%. Bisogna prestare particolare attenzione al fatto che il test sull'indipendenza è molto sensibile alla numerosità di ogni cella.

Tabella 2.13: Stato di salute percepito, in formato breve, intervistati 18-34 anni per genere, valori assoluti e distribuzione percentuale condizionata

<i>Sesso/Salute_d</i>	Bene	Male	Totale
Maschio	2411 94.73%	134 5.27%	2545 100%
Femmina	2275 93.05%	170 6.95%	2445 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: 0.01499

Nella Figura 2.6 viene mostrata la distribuzione condizionata della variabile *Salute_m*, in formato esteso, rispetto al genere dell'individuo, con le frequenze assolute al di sopra delle barre. Come si può vedere le distribuzioni sembrano simili, ma con leggere differenze tra maschi e femmine, apparentemente i maschi stanno meglio delle femmine, evidenza riscontrata anche con il formato dicotomico. Applicando il test dell'indipendenza il p-value osservato è $1.8 \cdot 10^{-4}$, ossia è possibile rifiutare l'ipotesi nulla di indipendenza delle due variabili, potendo così ipotizzare una associazione tra le due variabili in esame.

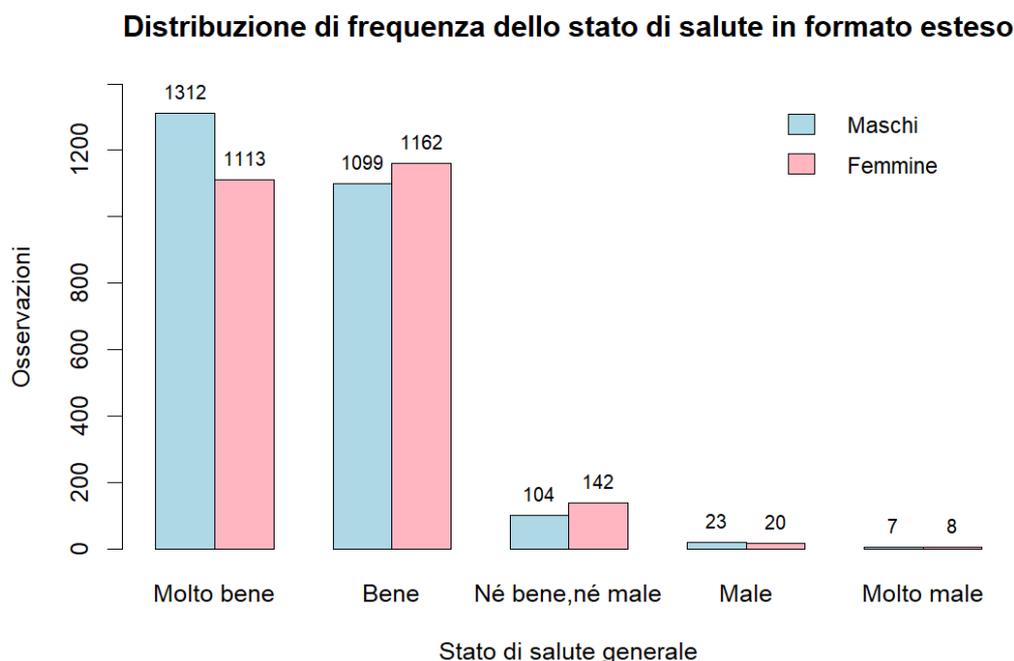


Figura 2.6: Stato di salute percepito, in formato esteso, intervistati 18-34 anni per genere, valori assoluti.

Si analizza la relazione tra le classi d'età, variabile *Eta*, con lo stato di salute generale, *Salute*.

La Tabella 2.14 mostra la distribuzione di frequenza congiunta e condizionata dello stato di salute in formato breve rispetto alle classi d'età, viene applicato poi il test di indipendenza di Pearson che restituisce un p-value pari a $9.6 \cdot 10^{-4}$, cioè viene rifiutata l'ipotesi nulla, quindi le due variabili sono associate. Si nota che all'aumentare dell'età, ossia spostandosi da una classe d'età minore ad una maggiore, il livello di benessere cala e il malessere aumenta, la direzione è univoca.

Utilizzando la variabile *Salute_m* in formato esteso Figura 2.7, otteniamo lo stesso risultato dello stato di salute in formato breve, con una grande diminuzione del p-value ($2.97 \cdot 10^{-12}$) del test sull'indipendenza, come era ovvio aspettarsi.

Tabella 2.14: Stato di salute percepito, in formato breve, intervistati 18-34 anni per classi d'età, valori assoluti e distribuzione percentuale condizionata

<i>Eta/Salute_d</i>	Bene	Male	Totale
18-24	1891 95.12%	97 4.88%	1988 100%
25-29	1355 94.16%	84 5.84%	1439 100%
30-34	1440 92.13%	123 7.87%	1563 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: $9.6 \cdot 10^{-4}$

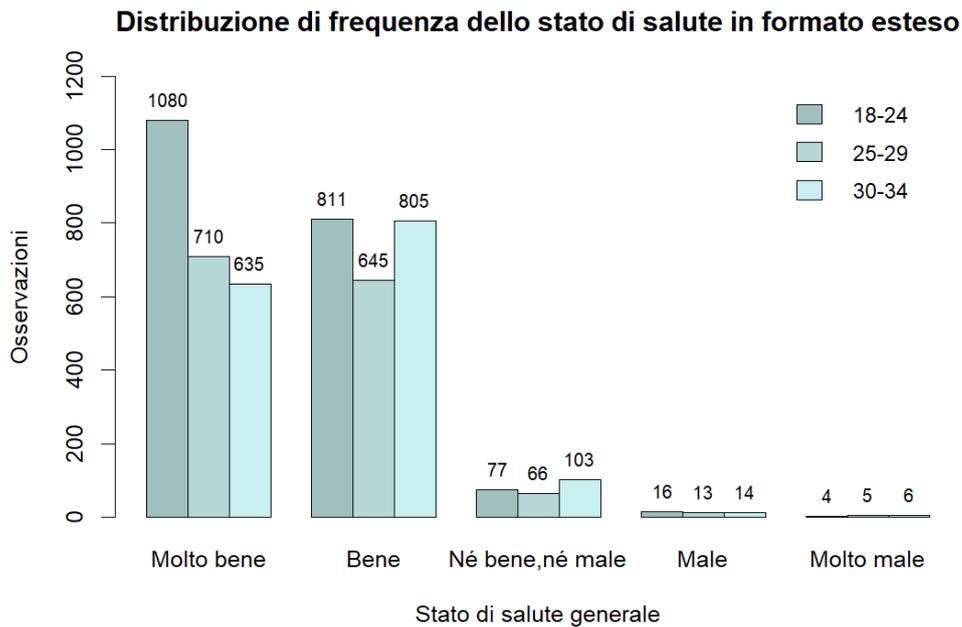


Figura 2.7: Stato di salute percepito, in formato esteso, intervistati 18-34 anni per classi d'età, valori assoluti.

Esaminando la relazione tra la condizione professionale e lo stato di salute in formato breve, Tabella 2.15, è evidente come alcune condizioni professionali influenzino molto la salute generale di un individuo, infatti chi appartiene alla categoria "In cerca di lavoro" (per definizione della statistica economica sono compresi i disoccupati e le persone in cerca di primo lavoro), sta peggio rispetto agli altri, ma un risultato sorprendente si ottiene dagli individui appartenenti ad "Altra condizione". Questa categoria comprende persone che non studiano e non cercano lavoro per loro volontà o a causa di altri problemi, oltre ad una piccola porzione di casalinghe, vi rientrano anche i cosiddetti *NEET*, individui che non studiano, non lavorano e non ricevono formazione. Tra gli appartenenti alla categoria analizzata, il 12.86%, non sta bene, che se comparato alla proporzione di campione totale che sta male, il 6.09%, è più che doppia.

Il test di indipendenza, come è ovvio aspettarsi dopo aver visto i dati precedenti, rifiuta l'ipotesi nulla, con un p-value osservato praticamente pari a 0. Anche per la forma estesa dello stato di salute si ottengono risultati simili a quelli della forma dicotomica, si rifiuta l'ipotesi nulla di indipendenza con un p-value osservato pari a 0.

Tabella 2.15: Stato di salute percepito, in formato breve, intervistati 18-34 anni per condizione professionale, valori assoluti e distribuzione percentuale condizionata

<i>Cond/Salute_d</i>	Bene	Male	Totale
Occupato	2435 94.53%	141 5.47%	2576 100%
In cerca di lavoro	588 92.31%	49 7.69%	637 100%
Studente	1209 96.26%	47 3.74%	1256 100%
Altra condizione	454 87.14%	67 12.86%	521 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: 0

Approfondendo la relazione tra la ripartizione geografica e lo stato di salute in formato breve, si evidenzia che non c'è alcuna associazione, Tabella 2.16, le percentuali di riga non variano a seconda della macroarea di appartenenza e il p-value osservato del test sull'indipendenza è molto alto, pari a 0.665, il che porta ad accettare l'ipotesi nulla a qualsiasi usale livello di significatività. Anche per il formato esteso otteniamo un risultato analogo a quello appena evidenziato.

Tabella 2.16: Stato di salute percepito, in formato breve, intervistati 18-34 anni per macroarea di appartenenza, valori assoluti e distribuzione percentuale condizionata

<i>Rip/Salute_d</i>	Bene	Male	Totale
Nord	1941 94.04%	123 5.96%	2064 100%
Centro	761 94.42%	45 5.58%	806 100%
Sud	1984 93.58%	136 6.42%	2120 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: 0.665

Riassumendo, sia per il formato breve sia per il formato esteso dello stato di salute, sembra che il genere, l'età e la condizione professionale abbiano un impatto tra le caratteristiche individuali. Per quanto riguarda la ripartizione geografica, non si osserva alcuna associazione indipendentemente dalla modalità di codifica utilizzata.

2.3.3 Lo stato di salute in rapporto alle caratteristiche familiari

Viene approfondita in seguito la relazione tra lo stato di salute, nelle due codifiche, rispetto alle caratteristiche familiari.

Analizzando la relazione tra lo stato di salute in formato breve e il titolo di studio del padre, si evidenzia che c'è una forma di associazione tra le due variabili, infatti il p-value osservato del test di indipendenza è piccolo, $1.6 \cdot 10^{-4}$, il che porta a rifiutare l'ipotesi nulla di indipendenza a qualsiasi livello di significatività usuale. Si

nota anche dalla Tabella 2.17 che la relazione è quasi monotona decrescente, ossia al diminuire del grado del titolo di studio aumenta il malessere, ma la modalità "Titolo alto" ha una percentuale maggiore di individui che stanno male rispetto a chi ha il padre con titolo medio, per le restanti modalità invece il malessere aumenta.

Tabella 2.17: Stato di salute percepito, in formato breve, intervistati 18-34 anni per titolo di studio del padre, valori assoluti e distribuzione percentuale condizionata

<i>Titpad/Salute_d</i>	Bene	Male	Totale
Titolo alto	478 94.65%	27 5.35%	505 100%
Titolo medio	1685 95.63%	77 4.37%	1762 100%
Titolo basso	2409 92.83%	186 7.17%	2595 100%
Non so	114 89.06%	14 10.94%	128 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: $1.6 \cdot 10^{-4}$

Questo fenomeno dal punto di vista logico non trova molto senso, visto che un titolo di studio maggiore dei genitori dovrebbe essere sinonimo di maggiore benessere per l'intera famiglia. Una spiegazione plausibile potrebbe essere che la numerosità campionaria ridotta porti ad un effetto distorsivo nei dati. Gli stessi risultati si ottengono anche considerando lo stato di salute in formato esteso.

Un'altra caratteristica familiare di cui è stato tenuto conto è la condizione professionale della madre in età adolescenziale dell'intervistato. Nella Tabella 2.18 è presente la distribuzione condizionata dello stato di salute in formato breve rispetto alla condizione professionale della madre quando l'intervistato aveva 14 anni.

Si nota che le percentuali di riga non sono tutte uguali tra di loro, ma è presente una leggera variazione tra le diverse modalità di *Conmad*, con un aumento della percentuale del malessere per chi ha la madre disoccupata o senza lavoro, il p-value del test di indipendenza risulta 0.01831, il che vuol dire che ad un livello di significatività del 5% è possibile rifiutare l'ipotesi nulla di indipendenza delle due

variabili. Il risultato è ulteriormente confermato quando lo stato di salute è utilizzato nella sua codifica estesa, con un p-value del test di indipendenza pari a 0.03488.

Tabella 2.18: Stato di salute percepito, in formato breve, intervistati 18-34 anni per condizione professionale della madre in età adolescenziale, valori assoluti e distribuzione percentuale condizionata

<i>Conmad/Salute_d</i>	Bene	Male	Totale
Occupata	2569 94.45%	151 5.55%	2720 100%
In cerca di lavoro	44 86.27%	7 13.73%	51 100%
Casalinga	1930 93.63%	131 6.36%	2061 100%
Altra condizione	143 90.51%	15 9.49%	158 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: 0.01831

Viene analizzata anche l'associazione con la stabilità della relazione tra i genitori, attraverso la variabile indicatrice della separazione, ossia si vuole valutare l'effetto della separazione sullo stato di salute del figlio. La Tabella 2.19 evidenzia che gli individui che hanno i genitori separati stanno leggermente peggio rispetto a chi non è separato, ma questa differenza non è significativa visto il p-value del test di indipendenza pari a 0.2918, ossia si accetta l'ipotesi di indipendenza tra le due variabili dello stato di salute e la separazione dei genitori. Il p-value del test di indipendenza quando la variabile dello stato di salute è codificata in formato esteso aumenta ancora di più con un valore pari a 0.9156, confermando i risultati già ottenuti.

Tabella 2.19: Stato di salute percepito, in formato breve, intervistati 18-34 anni per stato di separazione dei genitori, valori assoluti e distribuzione percentuale condizionata

<i>Gensep/Salute_d</i>	Bene	Male	Totale
No	3961 94.11%	248 5.89%	4209 100%
Sì	678 93.00%	51 7.00%	729 100%
Non so	47 90.38%	5 9.62%	52 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: 0.2918

Si può immaginare a priori che esista una dipendenza tra le risorse economiche e lo stato di salute. Oltre alle cause note in letteratura, anche il buon senso fa pensare che chi ha maggiori disponibilità economiche possa permettersi cure migliori e uno stile di vita più appagante.

Tabella 2.20: Stato di salute percepito, in formato breve, intervistati 18-34 anni per giudizio sulle risorse economiche, valori assoluti e distribuzione percentuale condizionata

<i>Riseco/Salute_d</i>	Bene	Male	Totale
Ottime o adeguate	3379 95.48%	160 4.52%	3539 100%
Scarse o insufficienti	1181 89.47%	139 10.53%	1320 100%
Non so	126 96.18%	5 3.82%	131 100%
Totale	4686 93.91%	304 6.09%	4990 100%

* p-value test di indipendenza: $3.62 \cdot 10^{-14}$

Nella Tabella 2.20 è evidente che ci sia una dipendenza tra queste due variabili, infatti la percentuale di chi sta "Male" tra chi giudica le risorse economiche familiari

non sufficienti è più che doppia rispetto a chi giudica le risorse economiche familiari buone. Anche il test del chi-quadrato di Pearson riporta un p-value prossimo a 0, ossia viene rifiutata l'ipotesi nulla di indipendenza. Risultati simili si ottengono con la codifica dello stato di salute in formato esteso.

Ricapitolando lo stato di salute è influenzato da alcune caratteristiche familiari come il titolo di studio del padre, la condizione professionale della madre in età adolescenziale e le condizioni economiche della famiglia, mentre non sembra avere alcuna influenza lo stato di separazione dei genitori. Questi risultati sono validi per entrambe le codifiche della variabile riguardante lo stato di salute.

Capitolo 3

Esplorando i dati attraverso modelli

In questo capitolo vengono effettuate le analisi dei dati attraverso la creazione di modelli aderenti allo scopo della relazione per ottenere i risultati cercati. Per fare questo si sfrutta la variabile risposta, *Salute*, nella sua doppia codifica, ossia in formato dicotomico attraverso un modello binomiale e in formato multivariato attraverso un modello per risposte politomiche.

3.1 Il modello binomiale

Si considera la variabile risposta *Salute* in formato dicotomico, ossia con le modalità "Bene" e "Male", da notare che nella seconda modalità è considerata anche la condizione neutrale (né bene, né male), si assume che la modalità di riferimento sia "Bene", cioè corrisponde allo 0, ovvero all'insuccesso della variabile dicotomica. Si può quindi pensare che ogni risposta sia una variabile binomiale elementare $Y_i \sim Bi(1, \pi_i)$ dove $\mu_i = \pi_i$ è la probabilità dell' i -esimo soggetto di stare male. Viene definita anche $g(\mu_i) = \eta_i = x_i\beta$ funzione di legame da specificare.

Poiché la media degli Y_i assume valori nell'intervallo $(0,1)$, come funzione di legame $g(\cdot)$ si sceglie una funzione $g : [0,1] \rightarrow \mathbb{R}$ monotona crescente. A questa categoria appartengono tutte le funzioni inverse di funzioni di ripartizione di variabili casuali continue con supporto in \mathbb{R} . Se infatti $F(\cdot)$ è la funzione di ripartizione di variabili casuali continue con supporto \mathbb{R} , la sua funzione inversa $F^{-1}(\mu)$ ha come

dominio $[0,1]$, come codominio \mathbb{R} ed è monotona crescente. Il modello assume allora che per una tale $F(\cdot)$,

$$g(\mu_i) = F^{-1}(\mu_i) = x_i\beta,$$

ossia

$$\mu_i = \pi_i = F(x_i\beta).$$

La funzione di legame scelta per l'analisi è la "logistica" o "logit":

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = x_i\beta,$$

con inversa

$$\mu_i = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}}.$$

Il modello lineare generalizzato corrispondente a tale specificazione è detto di regressione logistica o, modello logit; la distribuzione con funzione di ripartizione $F(z) = e^z/(1 + e^z)$, $z \in \mathbb{R}$, è infatti la distribuzione logistica (Salvan, Sartori e Pace 2020).

Viene scelta questa funzione di legame per le sue proprietà interpretative nei risultati, la funzione di legame logistica permette di interpretare il predittore lineare $\eta_i = \log(\mu_i/(1 - \mu_i)) = \log(\pi_i/(1 - \pi_i))$ in termini di logaritmo della quota (log-odds). La quota $\pi_i/(1 - \pi_i)$ è il rapporto tra la probabilità di successo e insuccesso, che moltiplicata per 100 esprime la quota di successi su 100 insuccessi. Di conseguenza il generico coefficiente β_r esprime l'effetto sul logaritmo della quota di un incremento unitario di x_{ir} , fermo restando il valore delle ulteriori esplicative del modello.

Per la selezione del modello si usano, in parte, delle procedure automatiche chiamate *forward selection* e *stepwise selection*. La selezione in avanti (*forward*) include via via nel modello le variabili esplicative che presentano il contributo parziale più significativo, ossia il livello di significatività osservato più piccolo per il test di nullità del coefficiente di regressione corrispondente e minore livello di α osservato. La selezione *stepwise* è una variante della selezione in avanti perché quando viene inclusa una nuova variabile esplicativa nel modello, si eliminano quelle variabili già nel modello il cui contributo parziale sia divenuto non significativo, ad un livello prefissato, dopo il nuovo inserimento. Queste procedure vengono usate non in modo canonico, infatti potrebbero portare all'eliminazione di variabili che invece sono fondamentali per l'ottenimento dei risultati, quindi verranno seguite in linea

generale, ma abbandonate nel momento in cui vadano ad intaccare la buona uscita del modello.

3.1.1 Il collegamento tra la salute e i living arrangements

Come prima fase di modellazione viene mantenuto l'intero campione e sullo stato di salute viene regredita la variabile del living arrangement, mantenendo anche l'intercetta, in questo modo è possibile studiare il puro effetto della convivenza con i genitori sulla salute dei giovani, si farà riferimento a questo modello chiamandolo **Modello1**.

Il **Modello1** può essere definito come segue:

1. $Y_i \sim Bi(1, \pi_i)$ con $\pi_i = Pr(Y_i = 1) = E(Y_i)$
2. $logit(\pi_i) = \log(\pi_i/(1 - \pi_i)) = \eta_i$
3. $\eta_i = \beta_0 + \beta_1 x_{i1}$
4. $x_{i1} = \begin{cases} 1, & \text{se } x_{i1} = \text{Vive con almeno un genitore} \\ 0, & \text{se } x_{i1} = \text{Vive senza i genitori} \end{cases} \quad (\text{Convgen_dico})$

Il coefficiente riguardante la modalità "Vive con almeno un genitore" della variabile living arrangement, visibile in Tabella 3.1, esprime l'incremento o il decremento, per la variazione unitaria della variabile associata che per la natura della stessa può assumere solo valori 0 o 1, non presente o presente. Il risultato che si ottiene evidenzia che chi vive con almeno un genitore ($x_{i1} = 1$), ha una quota pari a $e^{-0.19364} = 0.8240$ di stare male rispetto a chi vive senza alcun genitore ($x_{i1} = 0$), visto che è minore di 1 significa che starà "meno" male. Tuttavia questo coefficiente non è significativo, il valore del p-value osservato del test di nullità del parametro è pari a 0.112, maggiore del 5%, non è possibile quindi concludere che il living arrangement abbia effetto sullo stato di salute.

Tabella 3.1: Regressione logistica che prevede lo stato di salute, basata sul living arrangement

Variabile	Modalità	Stima	Rapporto di quota	p-value
	Intercetta	-2.6118		$< 2 \cdot 10^{-16}$ ***
Living arrangement (<i>Convgen_dico</i>)	Vive senza i genitori (Rif)			
	Vive con almeno un genitore	-0.1936	0.8240	0.112

Nota: Codici di significatività: 0 '***', 0.001 '**', 0.01 '*', 0.1 ' . ' .

3.1.2 L'impatto delle caratteristiche individuali

Come descritto in precedenza, si passa ad inserire nel modello le variabili riguardanti le caratteristiche individuali, chiameremo il modello risultante **Modello2**.

Le variabili significative sono l'età degli individui, codificata in classi, e la condizione professionale, mentre il genere e la ripartizione geografica non risultano significativi, tuttavia il sesso viene mantenuto perché non si esclude possa avere un legame con il living arrangement viste le analisi bivariate, Sezione 2.3.2.

Guardando i coefficienti si può notare che in primo luogo il living arrangement cambia segno rispetto al **Modello1**, ma comunque rimane non significativo, il valore del p-value ottenuto dal test di nullità del parametro quasi triplica rispetto al precedente, non è possibile ottenere alcuna informazione su chi sta meglio o peggio relativamente a questa condizione, ma è possibile dire che non ha effetto sullo stato di salute, in base ai nostri dati.

Esaminando i coefficienti delle variabili riguardanti le caratteristiche individuali, si vede che essere femmina sembra avere effetti negativi sulla salute, ma il risultato non è significativo, quindi lo stato di salute non è influenzato dal sesso. L'età di un individuo invece è significativa rispetto allo stato di salute, infatti applicando un test di significatività congiunta alla variabile *Eta* il p-value restituito è pari a 0.044, ossia significativo ad un livello del 5%, nello specifico appartenere alla categoria 18-24 o 25-29 non ha alcuna influenza sullo stato di salute, ricordandosi però della presenza dell'intercetta, perché una è la modalità base, ossia il suo coefficiente è pari a 0, mentre l'altra poco più grande, 0.0788 non è significativamente diversa dalla base. Abbiamo invece che appartenere alla categoria 30-34, mantenendo tutte le altre variabili fissate, risulta in un peggioramento della salute, con una quota

pari a $e^{0.3879} = 1.4739$ vale a dire che appartenere a questa categoria aumenta del 47.39% la possibilità di stare male.

Per la condizione professionale, anch'essa congiuntamente significativa (p-value test di nullità congiunta $1.56 \cdot 10^{-8}$), abbiamo che lo studente sta meglio, dato il segno negativo del coefficiente corrispondente, rispetto a tutti gli altri, ma non è possibile confermare il risultato dato che il coefficiente non risulta significativamente diverso da 0, ovvero diverso dalla condizione di occupato. Si può invece dire che chi è disoccupato o in altra condizione, a parità delle altre variabili, sta peggio, entrambi i coefficienti sono positivi e significativi, guardando le quote, chi è disoccupato ha una quota pari a 1.4765, chi si trova in altra condizione ha una quota ancora più alta, 2.4474, ossia ogni 100 individui che stanno bene 245 stanno male se appartengono a questa categoria, i risultati sono disponibili in Tabella 3.2.

Tabella 3.2: Regressione logistica che prevede lo stato di salute, basata su living arrangement e caratteristiche individuali

Variabile	Modalità	Stima	Rapporto di quota	p-value
	Intercetta	-3.2189		$< 2 \cdot 10^{-16}$ ***
Living arrangement (<i>Convgen_dico</i>)	Vive senza i genitori (Rif)			
	Vive con almeno un genitore	0.1555	1.1682	0.2789
Sesso dell'individuo (<i>Sesso</i>)	Maschio (Rif)			
	Femmina	0.1781	1.1949	0.1556
Classi d'età (<i>Eta</i>)	18-24 (Rif)			
	25-29	0.0788	1.0820	0.6356
	30-34	0.3879	1.4739	0.0234 *
Condizione professionale (<i>Cond</i>)	Occupato (Rif)			
	In cerca di lavoro	0.3897	1.4765	0.0271 *
	Studente	-0.2982	0.7422	0.1279
	Altra condizione	0.8950	2.4474	$3.87 \cdot 10^{-8}$ ***

Nota: Codici di significatività: 0 '***', 0.001 '**', 0.01 '*', 0.1 ' . ' .

Ricapitolando, dal **Modello2** otteniamo un risultato opposto al **Modello1** per il living arrangement, ma entrambi non sono significativi. L'aggiunta delle caratteristiche individuali conferma in parte i risultati ottenuti dalla analisi bivariate della Sezione 2.3.3, ossia la ripartizione geografica non influenza lo stato di salute,

mentre l'età e la condizione professionale sì. Per quanto riguarda il genere si hanno due risultati opposti, visto che nelle analisi bivariate risultava associato allo stato di salute percepito, mentre nella modellazione non risulta significativo, per ora non escludiamo nessuna delle due ipotesi, continuando con le analisi cercheremo di dare una risposta.

3.1.3 L'impatto delle caratteristiche familiari

Partendo dal **Modello2** si vanno ad aggiungere le variabili riguardanti le caratteristiche familiari, si farà riferimento a questo modello chiamandolo **Modello3**, risultati disponibili in Tabella 3.3.

Applicando il test di significatività congiunta si nota che le variabili familiari significative sono le risorse economiche e il titolo di studio del padre, mentre la condizione professionale della madre in età adolescenziale e la separazione dei genitori non risulta significativa, anche in questo caso si confermano parzialmente i risultati delle analisi bivariate, Sezione 2.3.3.

Partendo dal living arrangement, è possibile vedere che è significativo al 10%, quindi c'è differenza nello stato di salute tra chi vive con i genitori e chi vive senza, ma facendo attenzione che se viene mantenuto il convenzionale livello di significatività $\alpha = 5\%$ il risultato appena trovato non è più valido. Il segno del coefficiente è positivo, questo indica che chi vive con i genitori sta peggio di chi vive senza i genitori fissando tutti gli altri fattori, calcolando la quota otteniamo $e^{0.2506} = 1.2848$ cioè chi vive con i genitori ha il 28.48% in più di possibilità di stare male rispetto a chi vive senza i genitori.

Per le caratteristiche individuali, cambiano leggermente i valori dei coefficienti delle variabili riguardanti il sesso e le classi d'età ma le conclusioni sono uguali a quelle per il **Modello2**, per la condizione professionale si ha invece che i disoccupati non risultano più significativamente diversi dagli occupati o dagli studenti per quanto riguarda lo stato di salute, mentre rimane significativamente diverso da zero "Altra condizione" portando sempre ad un grosso peggioramento della salute.

Passando ora alle caratteristiche familiari si ha che il titolo di studio, mantenendo un livello di significatività del 5%, non ha alcun effetto sullo stato di salute percepito dai giovani, contrariamente a quanto riscontrato nelle analisi bivariate. Come da aspettative invece, le risorse economiche influenzano di molto lo stato di salute dei giovani, infatti chi ritiene le risorse economiche siano scarse o inadeguate ha un aumento di quota del 116.08% di stare male rispetto a chi non si trova in questa

condizione, a parità di tutte le altre variabili. Chi non ha risposto alla domanda non sembra essere significativamente diverso da chi ritiene le risorse siano adeguate.

Tabella 3.3: Regressione logistica che prevede lo stato di salute, basata su living arrangement, caratteristiche individuali e familiari

Variabile	Modalità	Stima	Rapporto di quota	p-value
	Intercetta	-3.3561		$< 2 \cdot 10^{-16}$ ***
Living arrangement (<i>Convgen_dico</i>)	Vive senza i genitori (Rif)			
	Vive con almeno un genitore	0.2506	1.2848	0.0821 .
Sesso dell'individuo (<i>Sesso</i>)	Maschio (Rif)			
	Femmina	0.1823	1.1999	0.1485
Classi d'età (<i>Eta</i>)	18-24 (Rif)			
	25-29	0.0813	1.0847	0.6267
	30-34	0.3977	1.4884	0.0206 *
Condizione professionale (<i>Cond</i>)	Occupato (Rif)			
	In cerca di lavoro	0.1601	1.1737	0.3748
	Studente	-0.2415	0.7855	0.2273
	Altra condizione	0.7690	2.1575	$3.38 \cdot 10^{-6}$ ***
Titolo di studio del padre (<i>Titpad</i>)	Titolo alto (Rif)			
	Titolo medio	-0.4067	0.6658	0.0823 .
	Titolo basso	-0.0698	0.9325	0.7524
	Non so	0.2957	1.3441	0.4101
Giudizio risorse economiche (<i>Riseco</i>)	Ottime o adeguate (Rif)			
	Scarse o insufficienti	0.7705	2.1609	$1.7 \cdot 10^{-9}$ ***
	Non so	-0.2265	0.7973	0.6272

Nota: Codici di significatività: 0 '***', 0.001 '**', 0.01 '*', 0.1 ' . ' .

3.1.4 Le interazioni tra i diversi aspetti

Partendo dal **Modello3**, il modello completo con le variabili core significative, si passa ad analizzare le interazioni tra il living arrangement e le altre covariate per vedere se c'è qualche effetto specifico della convivenza con i genitori per qualche sottogruppo di giovani, creando il modello finale chiamato **Modello4**.

Utilizzando allo stesso modo la procedura di selezione *stepwise*, si cercano quali interazioni risultano significative e se sono molteplici quali hanno un livello di significatività osservato minore.

Con i nostri dati l'unica interazione significativa tra il living arrangement e le altre variabili core è quella con la condizione professionale, Tabella 3.4, significa che solo ad un gruppo specifico di giovani, fermo restando tutte le altre variabili il living arrangement ha un effetto significativo sullo stato di salute.

Descrivendo brevemente i coefficienti risultanti dal modello è possibile vedere come quello del living arrangement, senza interazioni, è praticamente pari a zero e il p-value associato si aggira attorno all'80%, così come la condizione professionale non è più significativa, ma questo è dovuto all'inserimento dell'interazione tra le due. Il coefficiente del sesso è significativo (molto vicino al 5%), indicando una differenza tra maschi e femmine, essendo positivo vuol dire che le femmine hanno una quota maggiore di stare male rispetto ai maschi del 26.68%. L'interpretazione e la significatività dei coefficienti riguardanti le caratteristiche familiari non cambia rispetto al **Modello3**.

Passando all'interpretazione dell'interazione è possibile dire che le modalità base sono tutte quelle composte da chi vive senza i genitori o è occupato, quindi oltre a questi abbiamo altri tre diversi gruppi, due dei quali non hanno alcuna differenza rispetto ai base e sono chi vive con almeno un genitore ed è in cerca di lavoro o è uno studente, infatti i p-value derivanti dai test di nullità dei parametri risultato maggiori del 5% cioè non significativi. L'ultimo gruppo rimanente, quello di chi vive con i genitori e si trova in altra condizione professionale ha un coefficiente molto alto e significativo, da questo deriva che la loro quota aumenta del 334.48% rispetto a chi non si trova in questo gruppo, fermo restando tutte le altre variabili. La quota di questo gruppo è molto alta, per capirne meglio le cause è utile sviscerare cosa si intende per "Altra condizione", qui infatti possiamo trovare le casalinghe, i NEET e altri soggetti come gli inabili al lavoro o particolari condizioni. Da questo è possibile dedurre che molti di questi individui non siano in grado di avere una propria autonomia abitativa, anche se desiderata, a causa della mancanza di opportunità lavorative, economiche o di capacità proprie e questo porta il loro stato di salute ad essere peggiore rispetto a tutto il resto della popolazione.

Tabella 3.4: Regressione logistica che prevede lo stato di salute, basata su living arrangement, caratteristiche individuali e familiari con interazioni

Variabile	Modalità	Stima	Rapporto di quota	p-value
	Intercetta	-3.3090		$< 2 \cdot 10^{-16}$ ***
Living arrangement (<i>Convgen_dico</i>)	Vive senza i genitori (Rif)			
	Vive con almeno un genitore	0.0504	1.0517	0.7826
Sesso dell'individuo (<i>Sesso</i>)	Maschio (Rif)			
	Femmina	0.2365	1.2668	0.0595 .
Classi d'età (<i>Eta</i>)	18-24 (Rif)			
	25-29	0.0812	1.0846	0.6279
	30-34	0.4132	1.5116	0.0161 *
Condizione professionale (<i>Cond</i>)	Occupato (Rif)			
	In cerca di lavoro	0.4103	1.5072	0.1695
	Studente	0.3627	1.4373	0.4585
	Altra condizione	0.1241	1.1321	0.6251
Titolo di studio del padre (<i>Titpad</i>)	Titolo alto (Rif)			
	Titolo medio	-0.4016	0.6692	0.0872 .
	Titolo basso	-0.0324	0.9681	0.8839
	Non so	0.3640	1.4390	0.3126
Giudizio risorse economiche (<i>Risecco</i>)	Ottime o adeguate (Rif)			
	Scarse o insufficienti	0.7539	2.1253	$4.64 \cdot 10^{-9}$ ***
	Non so	-0.1693	0.8442	0.7168
Living arrangement * Condizione professionale (<i>Convgen_dico:Cond</i>)	Vive senza i genitori o occupato (Rif)			
	Vive con almeno un genitore e in cerca di lavoro	-0.2829	0.7536	0.4421
	Vive con almeno un genitore e studente	-0.556	0.5732	0.2876
	Vive con almeno un genitore e altra condizione	1.2071	3.3438	0.0002 ***

Nota: Codici di significatività: 0 '***', 0.001 '**', 0.01 '*', 0.1 ' . ' .

3.1.5 La valutazione della bontà dei modelli

Per valutare i modelli appena creati vengono utilizzati alcuni metodi tradizionali come:

- Il criterio di informazione di Akaike(AIC), fornisce una misura della qualità di un modello statistico tenendo conto sia della bontà di adattamento che della complessità del modello, permette il confronto tra modelli ed è definita come la log-verosimiglianza penalizzata moltiplicata per -2 ;
- Il criterio di informazione bayesiano(BIC), simile ad AIC ma introduce una penalizzazione per il numero di parametri del modello;
- La curva ROC(*Receiver Operating Characteristic*), è un sistema di valutazione grafico per classificatori binari, esprime il tasso dei veri positivi in funzione del tasso dei falsi positivi. L'area sotto tale curva (AUC , *Area Under the Curve*) coincide con l'indice di concordanza e dà l'idea della capacità predittiva del modello.

Dalla Tabella 3.5 e dalla Figura 3.1 si vede come il **Modello1** sia peggiore in termini di AIC , BIC e di capacità predittiva con AUC pari a 0.522 molto vicino allo 0.5, cioè incapace di prevedere. Con l'aggiunta delle caratteristiche individuali il **Modello2** migliora abbastanza con una diminuzione dei criteri di informazione e con l'aumento della capacità predittiva. Il **Modello3**, quello con l'aggiunta delle caratteristiche familiari, porta anch'esso ad un miglioramento del AIC , BIC e AUC , infatti questo modello ha il numero maggiore di variabili significative, utili a spiegare lo stato di salute. Il **Modello4** aggiunge l'interazione tra il living arrangement e la condizione professionale, portando ad un miglioramento del AIC ma al peggioramento del BIC , infatti quest'ultimo penalizza il modello per aver aggiunto più parametri rispetto al beneficio portato dall'aggiunta di quest'ultimi. La capacità predittiva migliora leggermente.

In conclusione il **Modello4** è il migliore per quanto riguarda la capacità predittiva, ma anche il **Modello3**, dato il basso valore del BIC , non è da scartare per la sua facilità interpretativa vista l'assenza di interazione tra parametri.

Tabella 3.5: Valori AIC e BIC, modelli logistici stimati

Modello	AIC	BIC
Modello1	2291.9	2304.9
Modello2	2247.5	2299.6
Modello3	2208.0	2292.7
Modello4	2194.9	2299.1

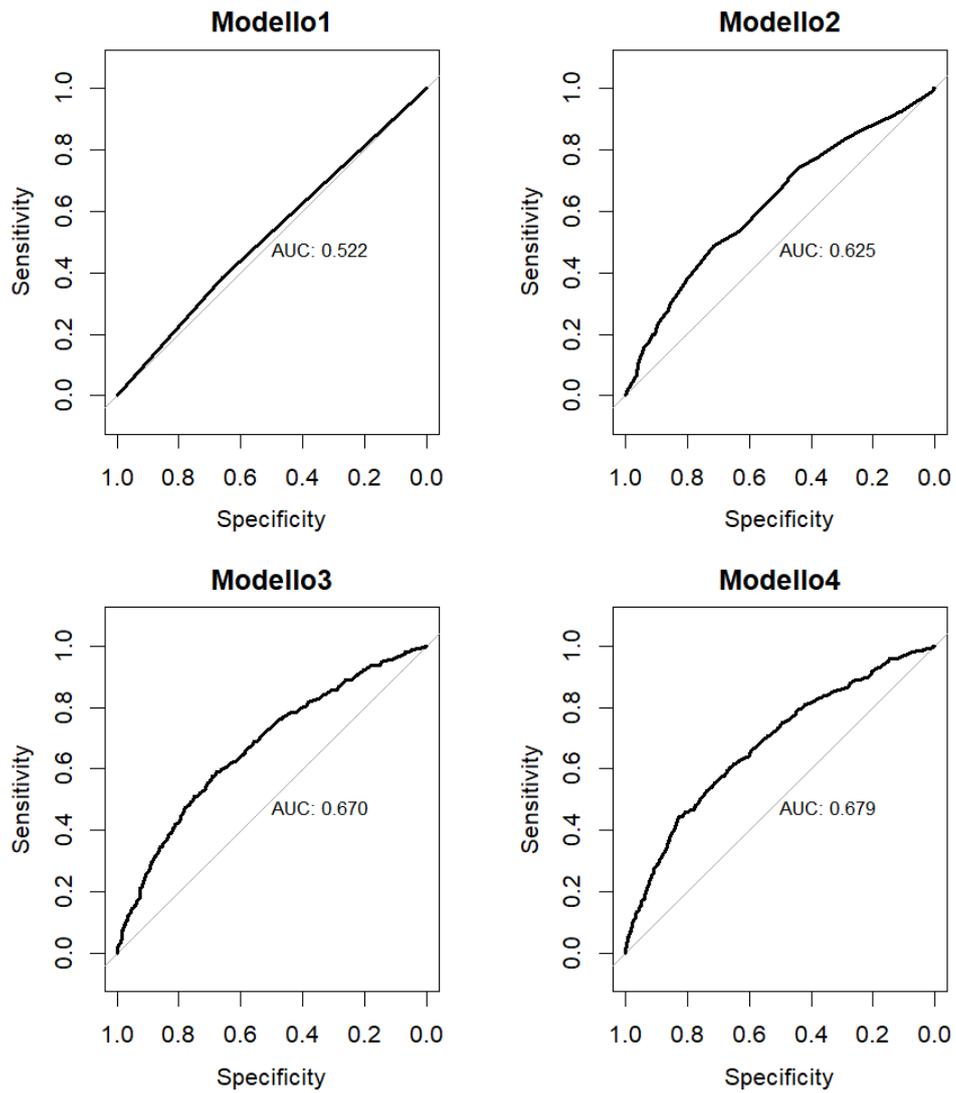


Figura 3.1: Curve ROC e valori AUC, modelli logistici stimati.

3.2 Il modello per risposte politomiche

Oltre al formato dicotomico dello stato di salute, trattato nella sezione precedente, è disponibile anche il formato esteso e da questo è possibile ottenere informazioni molto più dettagliate sfruttando a pieno la natura della variabile. Viene ampliata la scala di misura, passando da due a cinque modalità ordinali e per ognuna di queste il modello per risposte politomiche fornisce una stima, così facendo è possibile misurare la salute soggettiva in modo diverso e potrebbero emergere nuovi risultati diversi.

Quando la variabile risposta è qualitativa su scala ordinale, risulta importante valutare l'effetto di variabili esplicative, quantitative o qualitative ordinali, sull'ordinamento delle distribuzioni della risposta.

Come per una variabile continua, anche per una risposta qualitativa ordinale si può definire l'ordinamento stocastico tramite la funzione di ripartizione, sia Y_i la variabile casuale che descrive la risposta dell' i -esimo soggetto, $i = 1, \dots, n$, con modalità ordinate: $1, 2, \dots, j, \dots, c$. Se, per $k \neq i$,

$$Pr(Y_i \leq j) \geq Pr(Y_k \leq j), \text{ per ogni } j = 1, \dots, c,$$

con la disuguaglianza stretta per almeno un j in $1, \dots, c$, allora la distribuzione della variabile casuale Y_i è stocasticamente più piccola di quella della variabile casuale Y_k , se invece valgono le disuguaglianze in senso opposto, allora la distribuzione di Y_i è stocasticamente più grande di quella di Y_k .

Con variabili risposta qualitative su scala ordinale risulta dunque naturale formulare modelli per le probabilità cumulate

$$Pr(Y_i \leq j) = \pi_{i1} + \dots + \pi_{ij}, \quad j = 1, \dots, c,$$

con $\pi_{ij} = Pr(Y_i = j)$, $i = 1, \dots, n$, $j = 1, \dots, c$.

Si dice *logit cumulato* la trasformazione logit della probabilità cumulata

$$\begin{aligned} \text{logit}[Pr(Y_i \leq j)] &= \log \frac{Pr(Y_i \leq j)}{1 - Pr(Y_i \leq j)} = \log \frac{Pr(Y_i \leq j)}{1 - Pr(Y_i > j)} \\ &= \log \frac{\pi_{i1} + \dots + \pi_{ij}}{\pi_{i,j+1} + \dots + \pi_{ic}}, \quad j = 1, \dots, c - 1. \end{aligned}$$

Il logit cumulato è il logaritmo della probabilità che la risposta assuma valori minori o uguali a un certo livello rapportata alla probabilità che assuma valori

maggiori dello stesso livello. Quindi, il logit cumulato è anche detto *logaritmo della quota cumulata*.

Un modello parsimonioso è il modello di regressione per logit cumulati (McCullagh 1980). Esso assume che

$$\text{logit}[Pr(Y_i \leq j)] = \alpha_j + x_i\beta, \quad j = 1, \dots, c-1, \quad \text{con } \alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{c-1}.$$

Dove gli α_j rappresentano i parametri di intercetta e il vettore β contiene i coefficienti di regressione rispetto alle ulteriori variabili esplicative, assunte costanti al variare di j .

Se vale il modello definito, allora:

$$\log \frac{Pr(Y_i \leq j|x_i = \mathbf{u})/Pr(Y_i > j|x_i = \mathbf{u})}{Pr(Y_i \leq j|x_i = \mathbf{v})/Pr(Y_i > j|x_i = \mathbf{v})} = (\mathbf{u} - \mathbf{v})\beta.$$

Il logaritmo del rapporto delle quote, detto in questo contesto *rapporto delle quote cumulate*, è lineare rispetto alla differenza tra \mathbf{u} e \mathbf{v} con β indipendente da j . In particolare, se l' r -esima variabile passa da x_{ir} a $x_{ir} + 1$, fermo restando il valore delle restanti variabili esplicative, la quota cumulata $Pr(Y_i \leq j)/Pr(Y_i > j)$ risulta moltiplicata per e^{β_r} , costante in j . Per questa proprietà di effetto costante su tutte le probabilità cumulate, il modello è anche detto con *quote proporzionali* (Salvan, Sartori e Pace 2020).

3.2.1 Un collegamento alternativo tra la salute e i living arrangements

Il primo modello stimato, **Modello1b**, è simile a quello della Sezione 3.1.1, con la differenza che non è presente una sola intercetta, ma ben 4. In questo modo si può scegliere una diversa intercetta in base al livello di cui vogliamo conoscere la probabilità di verificarsi, facendo attenzione che non è possibile conoscere la probabilità di ogni singolo stato di salute, ma solo la possibilità di appartenere ad una categoria inferiore rispetto ad una superiore, proprio per questo le intercette sono 4.

Il parametro stimato del living arrangement si interpreta all'incirca nello stesso modo del modello logistico, ossia facendone la trasformata esponenziale questo esprime la quota di appartenere ad una modalità più alta, ovvero stare più male rispetto a chi non appartiene alla stessa categoria, fermo restando il valore delle altre variabili.

Visti i risultati, Tabella 3.6, possiamo dire che chi vive con un genitore sta meglio di chi vive da solo data la quota pari a 0.6656 e l'effetto è significativo.

Tabella 3.6: Regressione logistica cumulata con quote proporzionali che prevede lo stato di salute, basata sul living arrangement

Variabile	Modalità	Stima	Rapporto di quota	p-value
Intercette	Molto bene Bene	-0.3231		$1.32 \cdot 10^{-11}$ ***
	Bene Né bene,né male	2.4865		$< 2 \cdot 10^{-16}$ ***
	Né bene,né male Male	4.1953		$< 2 \cdot 10^{-16}$ ***
	Male Molto male	5.5563		$< 2 \cdot 10^{-16}$ ***
Living arrangement (<i>Convgen_dico</i>)	Vive senza i genitori (Rif)			
	Vive con almeno un genitore	-0.4070	0.6656	$2.6 \cdot 10^{-12}$ ***

Nota: Codici di significatività: 0 '***', 0.001 '**', 0.01 '*', 0.1 ' . ' .

Sfruttando a pieno le potenzialità del modello è possibile fare una previsione della distribuzione di frequenza percentuale, per poi confrontarla con quella osservata del campione, Tabella 3.7. I valori sono molto simili, soprattutto nelle modalità con numerosità più alta ovvero le prime tre, mentre per le ultime due, avendo una numerosità molto bassa è più difficile avere una stima precisa, tuttavia il modello sembra adattarsi bene ai dati.

Tabella 3.7: Valori osservati e valori predetti dal modello di regressione logistica cumulata per la salute basata sul living arrangement

	Stato di Salute/ Living arrangement	Molto bene	Bene	Né bene,né male	Male	Molto male
Valori osservati	Vive senza i genitori	0.4144	0.5172	0.0579	0.0088	0.0018
	Vive con almeno un genitore	0.5233	0.4196	0.0448	0.0085	0.0037
Valori predetti	Vive senza i genitori	0.4199	0.5033	0.0620	0.0110	0.0038
	Vive con almeno un genitore	0.5209	0.4266	0.0425	0.0074	0.0026

Sempre grazie a questa modellazione è facile trovare le probabilità cumulate e le quote di appartenere ad uno stato di salute o a quello inferiore, data la tipologia di living arrangement. Utilizzando la formula inversa della definizione del modello $\text{logit}[Pr(Y_i \leq j)] = \alpha_j + x_i\beta$ ossia $Pr(Y_i \leq j) = e^{\alpha_j + x_i\beta} / (1 + e^{\alpha_j + x_i\beta})$ si trovano

le probabilità cumulate e successivamente le quote tramite $\text{logit}[Pr(Y_i \leq j)]/(1 + \text{logit}[Pr(Y_i \leq j)])$, Tabella 3.8.

Dai dati si evidenzia che circa il 90% del campione sta "Bene" o "Molto bene" sia se vive con i genitori sia se vive da solo, dalle quote però è ben visibile il puro effetto del living arrangement, per chi vive da solo la quota è molto più alta nelle categorie peggiori dello stato di salute, rispetto a chi vive con almeno un genitore.

Tabella 3.8: Valori probabilità cumulate e rapporto delle quote cumulate della regressione logistica cumulata per la salute basata sul living arrangement

	Stato di Salute/ Living arrangement	Molto bene Bene	Bene Né bene,né male	Né bene,né male Male	Male Molto male
Probabilità cumulate	Vive senza i genitori	0.4199	0.9232	0.9852	0.9962
	Vive con almeno un genitore	0.3252	0.8889	0.9779	0.9942
Rapporto delle quote cumulate	Vive senza i genitori	0.7239	12.0194	66.3735	258.8496
	Vive con almeno un genitore	0.4819	8.0010	44.1828	172.3084

3.2.2 L'effetto delle altre caratteristiche

Dopo aver analizzato l'effetto puro del living arrangement sullo stato di salute, introduciamo le altre variabili core nel modello sempre utilizzando le strategie *forward selection* e *stepwise selection*, i risultati vengono presentati in un'unica tabella, Tabella 3.9.

Come secondo modello, **Modello2b**, si aggiungono le caratteristiche individuali al **Modello1b** e grazie al test per la significatività congiunta si evidenzia che il genere, le classi d'età e la condizione professionale sono significative, mentre la ripartizione geografica non lo è, in linea ai risultati delle analisi bivariate.

Con l'aggiunta delle caratteristiche individuali il living arrangement non è più significativo, ossia non possiamo dire che c'è una differenza tra chi vive con i genitori e chi senza. Per quanto riguarda i segni dei coefficienti invece abbiamo che essere femmina porta ad un peggioramento significativo dello stato di salute, fermo restando le altre variabili, rispetto ad essere maschio e anche appartenere alla classe d'età 30-34 porta ad un peggioramento dello stato di salute percepito. Per la condizione professionale si può dire che chi è studente sta meglio rispetto a chi è occupato, per i disoccupati non c'è un peggioramento significativo, mentre come per il modello logistico, **Modello2**, chi appartiene alla categoria "Altra condizione"

riscontra un peggioramento dello stato di salute, con una quota pari a 1.46 rispetto agli occupati.

Nel **Modello3b** con l'aggiunta delle caratteristiche familiari, rispetto alle variabili già presenti nel modello precedente non c'è cambio di segno o di significatività, quindi valgono gli stessi risultati. Il titolo di studio del padre evidenzia un andamento crescente del malessere con un decremento delle modalità, ossia ad un titolo minore corrisponde uno stato di salute del figlio peggiore, questo è vero per le modalità "Titolo basso" e "Non so", mentre non è possibile dire che i figli dei padri con titolo medio stanno peggio dei figli di quelli con titolo alto in base ai nostri dati. Per le risorse economiche abbiamo un peggioramento significativo della salute se le risorse sono ritenute scarse o inadeguate, mentre il coefficiente della modalità "Non so" non è statisticamente diverso da chi ha buone risorse economiche. Confrontando i risultati con il **Modello3** logistico non si trovano grosse differenze.

Come ultimo passaggio vengono studiate le interazioni e come per il **Modello4** binomiale, l'unica interazione significativa è quella tra il living arrangement e la condizione professionale. I segni e le significatività dei coefficienti delle variabili già inserite precedentemente nel modello non cambiano, a parte quelli per la condizione professionale che diventano non significativi a causa dell'introduzione dell'interazione.

Come il suo corrispettivo binomiale, il **Modello4b** evidenzia che chi vive con i genitori e si trova nella condizione di disoccupato o studente percepisce uno stato di salute diverso da chi vive senza i genitori o è occupato. Il gruppo di giovani che invece sta peggio rispetto a tutti gli altri, fermo restando le altre variabili, è quello di chi vive con i genitori e si trova in altra condizione, ossia nel gruppo sono comprese le casalinghe, i NEET e altri soggetti inabili al lavoro o in particolari condizioni non note.

Tabella 3.9: Regressione logistica cumulata con quote proporzionali che prevede lo stato di salute, basata su living arrangement, caratteristiche individuali e familiari con interazioni

Variabile	Modalità	Modello2b	Modello3b	Modello4b
Intercette	Molto bene Bene	0.0271 (0.7614)	0.3277 (0.0106 *)	0.3171 (0.0150 *)
	Bene Né bene,né male	2.8665 ($< 2 \cdot 10^{-16}$ ***)	3.2025 ($< 2 \cdot 10^{-16}$ ***)	3.1957 ($< 2 \cdot 10^{-16}$ ***)
	Né bene,né male Male	4.5806 ($< 2 \cdot 10^{-16}$ ***)	4.9234 ($< 2 \cdot 10^{-16}$ ***)	4.9189 ($< 2 \cdot 10^{-16}$ ***)
	Male Molto male	5.9436 ($< 2 \cdot 10^{-16}$ ***)	6.2886 ($< 2 \cdot 10^{-16}$ ***)	6.2853 ($< 2 \cdot 10^{-16}$ ***)
Living arrangement (<i>Convgen_dico</i>)	Vive senza i genitori (Rif)			
	Vive con almeno un genitore	-0.1319 (0.554 .)	-0.0699 (0.3147)	-0.1213 (0.1410)
Sesso dell'individuo (<i>Sesso</i>)	Maschio (Rif)			
	Femmina	0.1974 (0.0006 ***)	0.2001 (0.0005 ***)	0.2152 (0.0002 ***)
Classi d'età (<i>Eta</i>)	18-24 (Rif)			
	25-29	0.0359 (0.6325)	0.0537 (0.4768)	0.0589 (0.4362)
	30-34	0.3007 (0.0003 ***)	0.3246 ($9.12 \cdot 10^{-5}$ ***)	0.3317 ($6.53 \cdot 10^{-5}$ ***)
Condizione professionale (<i>Cond</i>)	Occupato (Rif)			
	In cerca di lavoro	0.0710 (0.4203)	-0.0731 (0.4187)	0.0610 (0.7272)
	Studente	-0.3015 (0.0002 ***)	-0.2399 (0.0036 **)	-0.1957 (0.4102)
	Altra condizione	0.3754 ($9.2 \cdot 10^{-5}$ ***)	0.2933 (0.0025 **)	0.0382 (0.7715)
Titolo di studio del padre (<i>Titpad</i>)	Titolo alto (Rif)			
	Titolo medio		0.0284 (0.7806)	0.0303 (0.7665)
	Titolo basso		0.2061 (0.0402 *)	0.2161 (0.0317 *)
	Non so		0.6543 (0.0010 ***)	0.6771 (0.0007 ***)
Giudizio risorse economiche (<i>Riseco</i>)	Ottime o adeguate (Rif)			
	Scarse o insufficienti		0.4973 ($5.67 \cdot 10^{-14}$ ***)	0.4908 ($1.35 \cdot 10^{-13}$ ***)
	Non so		-0.1297 (0.4676)	-0.1140 (0.5231)
Living arrangement * Condizione professionale (<i>Convgen_dico : Cond</i>)	Vive senza i genitori o occupato (Rif)			
	Vive con almeno un genitore e in cerca di lavoro			-0.1545 (0.4433)
	Vive con almeno un genitore e studente			-0.0205 (0.9340)
	Vive con almeno un genitore e in altra condizione			0.5841 (0.0022 **)

Note: Codici di significatività: 0 '***', 0.001 '**', 0.01 '*', 0.1 ' . ' .

Sono riportati i valori delle quote cumulate e tra parentesi i p-value del test di nullità del parametro.

Per valutare l'adeguatezza dei modelli, utilizziamo i criteri di informazione AIC e BIC, i quali sono elencati nella Tabella 3.10. Sebbene l'aggiunta di nuove variabili e interazioni porti il **Modello4b** ad avere il valore più basso di AIC, notiamo che il BIC non è il più basso tra tutti i modelli considerati.

In linea con quanto concluso per i modelli binomiali, in base ai nostri scopi possiamo scegliere se mantenere il modello completo o optare per il **Modello3b**, considerandolo più parsimonioso, poiché ha un AIC abbastanza simile e un BIC inferiore.

Tabella 3.10: Valori AIC e BIC, modelli logistici cumulati stimati

Modello	AIC	BIC
Modello1b	9104.1	9136.7
Modello2b	9030.2	9101.9
Modello3b	8951.9	9056.1
Modello4b	8947.3	9071.1

Capitolo 4

Conclusioni

La stesura di questo elaborato ha permesso di mettere a confronto lo stato di salute percepito e i living arrangements dei giovani in Italia, trovando alcuni risultati interessanti e non noti a priori.

Dopo un'accurata operazione di pulizia del dataset, in cui si sono organizzate meglio le informazioni e si è risolto il problema dei dati mancanti, le analisi univariate hanno dato una preliminare immagine delle variabili in gioco. Le analisi bivariate hanno invece dato un'idea marginale della relazione tra esplicative variabile d'interesse, ovvero lo stato di salute relativo ad ogni soggetto.

Durante queste analisi è emerso che i fattori più rilevanti sono il living arrangement, il genere, l'età, la condizione professionale, il titolo di studio del padre, la condizione professionale della madre quando l'intervistato aveva 14 anni e il giudizio delle risorse economiche. Si è infatti trovato che questi fattori influenzano in modo significativo lo stato di salute di un individuo, per esempio essere disoccupato o l'avanzare dell'età peggiora la percezione dello stato di salute.

Questi risultati sono chiaramente delle osservazioni marginali e per un'analisi più approfondita sono stati utilizzati modelli di regressione. Inizialmente trattando lo stato di salute in formato dicotomico è stato utilizzato un modello binomiale con legame logistico, poi mantenendo la codifica originaria dello stato di salute, su scala Likert, è stato utilizzato un modello per risposte politomiche logit cumulato con quote proporzionali.

Dalle modellazioni è emerso che le variabili esplicative più significative sono le stesse trovate nelle analisi preliminari, a parte la condizione professionale della madre in età adolescenziale, e che queste influenzano in modo simile lo stato di

salute, indipendentemente dalla codifica della variabile risposta.

La domanda iniziale era se, nonostante il ritardo nell'indipendenza dalla casa dei genitori dei giovani italiani, essi godessero di buona salute. Fortunatamente, la risposta a questa interrogativo è affermativa: la maggior parte dei giovani dichiara di essere in ottima o in buona salute, ma purtroppo alcune circostanze portano a un deterioramento della percezione dello stato di salute. Essere di genere femminile, superare i trent'anni o vivere in condizioni economiche precarie comporta un leggero peggioramento del proprio stato di salute generale percepito, risultato evidenziato in entrambe le modellazioni. Inoltre, il livello di istruzione del padre ha un'influenza significativa sullo stato di salute; infatti, minore è il grado di istruzione del padre, peggiore è la salute del figlio, quest'ultima correlazione non è emersa dal modello logistico ma è stata confermata attraverso il modello per le risposte politomiche.

Per quanto concerne il living arrangement non sembra avere un impatto diretto sullo stato di salute. Tuttavia, in entrambe le modellazioni, è emersa un'interazione significativa tra il living arrangement e la condizione professionale, comprendendo i soggetti che vivono con almeno un genitore e sono collocati in "Altra condizione", ovvero tutti gli individui che non sono occupati, disoccupati o studenti. In questo piccolo gruppo, formato da 232 intervistati (circa il 4.6% dell'intero campione), riscontriamo un peggioramento rilevante dello stato di salute percepito. Le cause potrebbero essere molteplici visto che a questo gruppo appartengono gli inabili al lavoro, i ritirati dal lavoro e altri soggetti come i NEET, ad ogni modo non viene esclusa la possibilità che la convivenza con i genitori possa essere una probabile causa del deterioramento della percezione dello stato di salute.

Appendice A

Creazione *Convegn*

Ai fini dell'analisi è necessario creare un indicatore che sia in grado di distinguere puntualmente se un individuo al momento della compilazione del questionario vive con o senza i genitori. Per fare questo partiamo da due variabili presenti nel dataset originario, *madviv2* e *padviv2*, che sono le variabili contenenti le risposte alla domanda «I suoi genitori vivono con Lei?», con tre diverse modalità di risposta, di seguito viene riportato il tracciato record riguardante le due variabili e il codice necessario per l'acquisizione dei dati in R.

Tabella A.1: Tracciato record delle variabili *madviv2* e *padviv2*

Posizione	Lunghezza	Nome/Modalità	Descrizione
519	1	<i>madviv2</i> 1 = No, vive altrove 2 = No, è deceduto 3 = Sì, vive con me	Madre vivente con intervistato
520	1	<i>padviv2</i> 1 = No, vive altrove 2 = No, è deceduto 3 = Sì, vive con me	Padre vivente con intervistato

```
1 library(tidyverse)#serve per il comando "recode"  
2  
3 #--- Creazione variabile Convegn ----  
4 #Seleziono le due variabili originali madviv2 e padviv2  
5 madviv2=dataset2[,519]  
6 padviv2=dataset2[,520]
```

```
7 n_oss=length(madviv2) #Numero di osservazioni
```

Codice A.1: Caricamento dei dati

Per la creazione dell'indicatore non è necessario distinguere il motivo per il quale l'intervistato non vive con i genitori, vengono quindi unificate le prime due modalità e si crea così una variabile dicotomica.

```
1 #Ricodifica delle variabili in dicotomiche
2 madviv2_dico=recode(as.factor(madviv2),
3                     "1" = "No",
4                     "2" = "No",
5                     "3" = "Si")
6 padviv2_dico=recode(as.factor(padviv2),
7                     "1" = "No",
8                     "2" = "No",
9                     "3" = "Si")
10
11 table(madviv2_dico, padviv2_dico, useNA = "ifany")
```

Codice A.2: Ricodifica per ottenere due variabili dicotomiche

Tabella A.2: Distribuzione di frequenza congiunta delle variabili dicotomizzate *madviv2* e *padviv2*, valori assoluti

	padviv2		
madviv2	No	Si	NA
No	1711	141	4
Si	573	2565	8
NA	0	1	6

Per creare la nuova variabile è necessario inicializzarla come un oggetto vuoto e di lunghezza pari al numero di osservazioni, dopodiché, attraverso un ciclo *for* che scorre tutte le osservazioni e degli *ifelse* che controllano le modalità di ogni osservazione è possibile verificare gli incroci della Tabella A.1 creando le 4 modalità della variabile *Convgen*, che sono "Vive senza genitori", "Vive solo con la madre", "Vive solo con il padre" e "Vive con entrambi i genitori".

```
1 #Creo la variabile Convgen di lunghezza pari alle
  osservazioni ma vuota
```

```

2 Convgen=rep(NA,n_oss)
3
4 #Combinazione delle modalita' per ottenere Convgen
5 for (i in 1:n_oss) {
6   #non vive ne' con la madre ne' con il padre
7   ifelse(madviv2_dico[i]=="No" & padviv2_dico[i]=="No",
8         (Convgen[i]="Vive senza genitori"), "")
9   #vive solo con la madre
10  ifelse(madviv2_dico[i]=="Si" & padviv2_dico[i]=="No",
11        (Convgen[i]="Vive solo con la madre"), "")
12  #vive solo con il padre
13  ifelse(madviv2_dico[i]=="No" & padviv2_dico[i]=="Si",
14        (Convgen[i]="Vive solo con il padre"), "")
15  #vive con entrambi i genitori
16  ifelse(madviv2_dico[i]=="Si" & padviv2_dico[i]=="Si",
17        (Convgen[i]="Vive con entrambi i genitori"), "")
18 }
19
20 table(Convgen, useNA="ifany")

```

Codice A.3: Inizializzazione e creazione della variabile *Convgen*

Tabella A.3: Distribuzione di frequenza della convivenza con i genitori, valori assoluti

	<i>Convgen</i>
Vive senza genitori	1711
Vive solo con la madre	573
Vive solo con il padre	141
Vive con entrambi i genitori	2565
NA	19

Alla fine del processo possiamo verificare la corrispondenza tra l'incrocio delle due variabili originali Tabella A.2 con la frequenza delle modalità della nuova variabile *Convgen* Tabella A.3.

Nello svolgimento delle analisi è stata utilizzata un'ulteriore codifica della variabile *Convgen*, quella binomiale, utilizzata per capire se un intervistato convive con almeno un genitore.

A partire dalla variabile *Convgen* viene applicata una ricodifica, accorpando le ultime tre modalità in una nuova modalità chiamata "Vive con almeno un genitore", ossia che risponde in modo affermativo alla domanda "Vive con almeno un genitore?", e la prima modalità viene ricodificata in "Vive senza genitori", che risponde in modo negativo alla domanda precedente. Le nuove numerosità sono osservabili in Tabella A.4

```

1 Convgen_dico=recode(Convgen,
2   "Vive senza i genitori"="Vive senza i genitori",
3   "Vive solo con la madre"="Vive con almeno un genitore",
4   "Vive solo con il padre"="Vive con almeno un genitore",
5   "Vive con entrambi i genitori"="Vive con almeno un
   genitore")
6 table(Convgen_dico, useNA = "ifany")

```

Codice A.4: Ricodifica di *Convgen* in formato dicotomico

Tabella A.4: Distribuzione di frequenza della convivenza con i genitori in formato dicotomico, valori assoluti

	<i>Convgen_dico</i>
Vive senza genitori	1711
Vive con almeno un genitore	3279
NA	19

Bibliografia

- Billari, Francesco C e Cecilia Tomassini (2021). *Rapporto sulla popolazione: L'Italia e le sfide della demografia*. Società editrice il Mulino.
- Dalla Zuanna, Gianpiero, Francesca Michielin e Fabio Bordignon (2007). «Proximity between adult children and their mothers: A European comparison». In: *Workshop "Changes in leaving arrangements and family relationships in the context of strong family ties. Southern Europe and Eastern Asia: trends, causes and consequences"*, Rostock.
- McCullagh, Peter (1980). «Regression models for ordinal data». In: *Journal of the Royal Statistical Society: Series B (Methodological)* 42.2, pp. 109–127.
- Nobile, M et al. (2013). «Riflessioni sulla definizione di salute dell'OMS». In: *9. edizione del Seminario Nazionale di Sociologia del diritto, Capraia*.
- Özdemir, Erhan, Terry Ward e Eszter Zolyomi (2014). «The effect of the crisis on young people's ability to live independently». In: *Research note 5/2014, European Commission*.
- Reher, David Sven (1998). «Family Ties in Western Europe: Persistent Contrasts». In: *Population and Development Review* 24.2, pp. 203–234. ISSN: 00987921, 17284457. URL: <http://www.jstor.org/stable/2807972>.
- Rosina, A, R Fraboni e LL Sabbadini (2003). «Diventare donne e uomini in Italia». In: *Genere e demografia. Bologna: Il Mulino*, pp. 173–192.
- Salvan, A., N. Sartori e L. Pace (2020). *Modelli Lineari Generalizzati*. UNITEXT. Springer Milan. ISBN: 9788847040021. URL: <https://books.google.it/books?id=vAn2DwAAQBAJ>.
- Santarelli, Elisabetta e Francesco Cottone (2009). «Leaving home, family support and intergenerational ties in Italy: Some regional differences». In: *Demographic Research* 21, pp. 1–22.

BIBLIOGRAFIA

- Schwanitz, Katrin e Clara H Mulder (2015). «Living arrangements of young adults in Europe». In: *Comparative Population Studies* 40.4.
- Sironi, Emiliano e Alessandro Rosina (2016). «Leaving the parental home in Italy during the economic crisis». In: *Genus* 71.2-3.
- Tomassini, Cecilia e Daniele Vignoli (2023). *Rapporto sulla popolazione Le famiglie in Italia. Forme, ostacoli, sfide*. Società editrice il Mulino.
- Tosi, Marco e Emily Grundy (2018). «Returns home by children and changes in parents' well-being in Europe». In: *Social Science & Medicine* 200, pp. 99–106.