

UNIVERSITÀ DEGLI STUDI DI PADOVA  
DIPARTIMENTO DI SCIENZE STATISTICHE  
CORSO DI LAUREA MAGISTRALE IN  
SCIENZE STATISTICHE



## **Indagine Global Monitor sulla salute mentale: un'analisi tramite modelli per dati discreti**

Relatore Prof. Emanuele Aliverti  
Dipartimento di Scienze Statistiche

Laureanda Angela Andrigo  
Matricola 2022771

Anno Accademico 2022/2023



*Alle persone che si prendono cura di me con tanto amore  
Senza di loro, mai sarei arrivata a questo traguardo*



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 L'indagine “<i>Global Monitor</i>”</b>	<b>5</b>
1.1 Descrizione dell'indagine . . . . .	5
1.2 Le variabili d'interesse . . . . .	7
1.2.1 Il disturbo psicologico . . . . .	7
1.2.2 Variabili socio-demografiche . . . . .	8
1.2.3 Gli approcci per la gestione di un disturbo psicologico . . . . .	10
1.3 Operazioni preliminari . . . . .	11
<b>2 Analisi esplorative</b>	<b>13</b>
2.1 Ansia e depressione . . . . .	13
2.2 Metodi per la gestione di ansia e depressione . . . . .	16
<b>3 Modello di regressione logistica</b>	<b>25</b>
3.1 Specificazione del modello . . . . .	25
3.2 Stima e interpretazione dei parametri . . . . .	26
3.3 Selezione del modello . . . . .	28
3.4 Bontà di adattamento . . . . .	31
3.5 Effetti dei fattori socio-demografici . . . . .	32
3.6 Confronto con i risultati in letteratura . . . . .	35
3.7 Modello di regressione logistica con intercetta casuale . . . . .	37
<b>4 Modello di regressione a classi latenti</b>	<b>41</b>
4.1 Specificazione del modello . . . . .	42
4.1.1 Modello a classi latenti . . . . .	42
4.1.2 Modello a classi latenti con inclusione di covariate . . . . .	44
4.2 Stima dei parametri . . . . .	45
4.3 Selezione del modello . . . . .	46
4.4 Bontà di adattamento . . . . .	47
4.5 Descrizione dei profili latenti . . . . .	50
4.6 Effetto dell'età . . . . .	53
4.7 Confronto con modelli alternativi . . . . .	55
4.7.1 Modello di Rasch . . . . .	55
4.7.2 Modello log-lineare . . . . .	57

---

<b>5 Conclusioni</b>	<b>61</b>
<b>Appendice</b>	<b>65</b>
A. 1 Modello congiunto . . . . .	65
A. 2 Modello di regressione a classi latenti . . . . .	66
<b>Bibliografia</b>	<b>69</b>





# Introduzione

L'Organizzazione Mondiale della Sanità (OMS) definisce la salute mentale come uno stato di benessere mentale che consente ad ogni individuo di affrontare le situazioni stressanti della vita, di realizzarsi valorizzando ed esprimendo i propri talenti, apprendere e lavorare in modo proficuo, e contribuire al buon funzionamento e al miglioramento della propria comunità (Saxena et al., 2013). La salute mentale è quindi una componente fondamentale della salute generale di ogni persona.

I disordini mentali sono patologie complesse caratterizzate da “significativi problemi nel pensiero, nella regolazione delle emozioni, o nel comportamento di una persona, che riflettono una disfunzione dei processi psicologici, biologici o dello sviluppo che compongono il funzionamento mentale” (Ministero della Salute, 2022) . Queste patologie possono interferire ed influenzare profondamente la qualità di vita di una persona, la capacità di prendere decisioni e di costruire relazioni con familiari, amici, colleghi e membri della comunità di appartenenza.

Per individuare e riconoscere le diverse categorie diagnostiche, si utilizzano precisi criteri descritti nella quinta edizione del Diagnostic and Statistical Manual of Mental Disorders (DSM-5), pubblicata nel 2013 dall'American Psychiatric Association (APA). Tra le tante tipologie esistenti, i due principali, più conosciuti e diffusi problemi di salute mentale sono ansia e depressione. Il DSM-5 spiega che i disturbi d'ansia hanno in comune aspetti di eccessiva paura e persistente angoscia, con conseguenti disfunzioni comportamentali. I disordini di questo tipo differiscono l'uno dall'altro in base al tipo di oggetti o situazioni che inducono paura o un comportamento evitante. Secondo il DSM-5, il disordine depressivo, noto più semplicemente come depressione, implica un umore depresso, perdita di piacere e disinteresse nelle attività quotidiane per lunghi periodi di tempo. Le cause possono essere molteplici: tra le più frequenti si possono citare l'aver vissuto un abuso, importanti perdite o altri eventi stressanti.

---

“Una persona su otto a livello globale vive una condizione di salute mentale che può avere un impatto sulla sua salute fisica, sul benessere e sul modo in cui si relaziona agli altri” (Ministero della Salute, 2023). Ansia e depressione affliggono attualmente più di 400 milioni di persone di tutte le età nel mondo. Nel 2030 si prevede che questi disturbi mentali rappresenteranno la principale causa di morte e di stati psicologici patologici a livello globale (Clements et al., 2020).

Conoscere l'età dell'esordio di una malattia mentale è fondamentale per garantire assistenza tempestiva e interventi precoci per i soggetti a rischio o con problemi manifesti. I risultati ottenuti dalla meta-analisi condotta da Solmi et al. (2022), su 192 studi, mostrano che l'insorgenza di un disordine mentale avviene prima dei 25 anni quasi per la metà dei pazienti con problemi di salute mentale di varia natura. Tale dato si abbassa fino a 5.5 anni per l'ansia e a 20.5 anni per la depressione, indicando che questi due disturbi psicologici si sviluppano spesso nell'infanzia e nell'età adolescenziale.

Nel “*Mental Health Atlas 2020*”, l'OMS mostra che i traguardi raggiunti dai singoli paesi non sono soddisfacenti rispetto agli obiettivi condivisi nel “*Comprehensive mental health action plan 2013–2030*”, piano d'azione per il miglioramento della salute mentale. Nonostante alcuni progressi e passi in avanti successivi alla pandemia di Covid-19, le istituzioni pubbliche non investono ancora sufficienti risorse per la prevenzione e la riabilitazione nel campo della salute mentale, considerando come prioritarie altre aree della sanità pubblica. La ricerca in questo campo è quindi importante per: promuovere l'importanza del benessere psicologico; aumentare la consapevolezza e sensibilizzare l'opinione pubblica; identificare i fattori protettivi e di rischio per la salute mentale; contribuire allo sviluppo di nuovi strumenti di prevenzione, strategie di intervento, trattamento e recupero a livello individuale, di gruppo e per l'intera popolazione; sperimentare metodi, azioni e approcci per la gestione di queste problematiche.

In questa tesi, si propone un'analisi dei dati raccolti dal “*Global Monitor 2020: Mental Health*”. Si tratta della più ampia indagine sul tema della salute mentale a livello mondiale, condotta nel 2020-2021 da Wellcome Trust, uno tra i più importanti enti benefici a livello globale, in collaborazione con Gallup, società di consulenza americana. Tramite modelli per dati discreti, si vuole caratterizzare il comportamento dei soggetti ansiosi e depressi di 28 paesi, europei e nord-americani, focalizzandosi sulle azioni messe in atto per affrontare questi disturbi mentali. Tra gli obiettivi, si intende valutare come l'esperienza di un disturbo d'ansia o depressione dipenda da alcune caratteristiche socio-demografiche dei singoli individui. Inoltre, si vogliono identificare profili di

soggetti che sono stati ansiosi e depressi, per spiegare le differenti strategie e le scelte comportamentali attuate per ristabilire il proprio benessere mentale.

Nel Capitolo 1, si descrivono i dati utilizzati, si elencano le variabili d'interesse e si giustificano le operazioni preliminari svolte.

Nel Capitolo 2 si presentano alcune analisi esplorative. In particolare, si descrive la prevalenza di ansia e depressione e la scelta dei metodi per affrontare questi disturbi mentali, al variare di alcune caratteristiche socio-demografiche.

Nel Capitolo 3, si adatta un modello di regressione logistica per valutare l'effetto di alcuni fattori socio-demografici (genere, età, reddito, occupazione lavorativa e nazionalità) sull'aver sofferto di ansia o depressione. Dopo aver selezionato il miglior modello in termini di bontà di adattamento, parsimonia e interpretabilità, si mostrano i risultati ottenuti, confrontandoli con quelli di alcuni studi presenti in letteratura. Inoltre, si confronta il modello di regressione logistica con un modello di regressione logistica con intercetta casuale per paese.

Nel Capitolo 4, si stima un modello di regressione a classi latenti per l'identificazione e la caratterizzazione di gruppi tra le osservazioni, in base alle combinazioni dei metodi sperimentati per la gestione di uno stato ansioso o depressivo. Dopo aver selezionato il miglior modello e averne valutato la bontà di adattamento, si descrivono la composizione dei profili latenti individuati e l'effetto delle covariate incluse nel modello. Inoltre, si confronta il modello a classi latenti con due possibili modelli alternativi: il modello di Rasch e il modello log-lineare.

Infine, nel Capitolo 5 si delineano le conclusioni, si riassumono i risultati principali e si riportano alcuni commenti finali sui limiti dello studio e sui possibili sviluppi futuri.



# Capitolo 1

## L'indagine “*Global Monitor*”

### 1.1 Descrizione dell'indagine

I dati utilizzati in questa tesi sono stati raccolti tramite la rilevazione “*Global Monitor 2020: Mental health*”, la più ampia indagine campionaria sul tema della salute mentale a livello mondiale (Clements et al., 2020). I dati sono a libero accesso e disponibili sulla pagina web dedicata al seguente link: <https://wellcome.org/reports/wellcome-global-monitor-mental-health/2020>.

L'indagine è stata condotta da Wellcome Trust, ente di beneficenza indipendente, in collaborazione con Gallup, società americana di analisi e consulenza. Wellcome Trust, società benefica fondata in Inghilterra nel 1939, dopo la morte del farmacista e imprenditore statunitense Sir Henry Wellcome, opera a livello globale finanziando ambiziosi progetti di ricerca in tre ambiti principali: salute mentale, riscaldamento globale e malattie infettive. La sua missione consiste nell'investire numerose risorse per risolvere urgenti sfide che riguardano il benessere psico-fisico di tutti, per un mondo più sano e più felice.

Il primo *Global Monitor* è stato condotto nel 2018. In questa prima indagine sono stati affrontati importanti temi quali: la fiducia che le persone ripongono nella scienza, negli scienziati e nelle informazioni inerenti la salute riportate dalle istituzioni pubbliche; le attitudini circa la sicurezza e l'efficacia dei vaccini.

Il secondo *Global Monitor* è stato condotto tra il 4 agosto 2020 e il 18 febbraio 2021. I temi trattati riguardano la salute mentale, le visioni sul cambiamento climatico, il ruolo della scienza, l'operato dei leader nazionali nel risolvere problemi di salute pubblica e la pandemia di Covid-19. In questa tesi, ci si focalizza sulla sezione dedicata alla salute mentale: “*Global Monitor 2020: Mental health*”.

Migliorare la salute mentale in tutto il mondo rappresenta il fine ultimo del “*Global Monitor 2020: Mental health*”. Il più importante obiettivo è quello di contribuire alla ricerca e allo sviluppo di nuove soluzioni, interventi di prevenzione e trattamenti, per aiutare i soggetti con problemi di salute mentale, nello specifico ansia e depressione. Ansia e depressione sono infatti i due disturbi mentali più comuni e diffusi, affliggono milioni di persone di tutte le età nel mondo e possono influenzare negativamente tutti gli aspetti della vita di un individuo, incluse le relazioni interpersonali nell’ambiente familiare, lavorativo o scolastico.

Altri obiettivi sono: sviluppare una maggiore consapevolezza sui problemi di salute mentale; promuovere interventi di prevenzione, supporto e recupero per i disturbi psicologici, quali ansia e depressione; sensibilizzare e mobilitare le istituzioni governative a sostegno della salute mentale in tutto il mondo.

Inoltre, un’ulteriore finalità è quella di individuare le priorità verso cui dovrebbero orientarsi le decisioni politiche degli stati e delle organizzazioni governative per la tutela della salute pubblica. Attraverso la ricerca si vuole quindi comprendere quali approcci funzionano per gestire i disturbi mentali d’ansia e depressione, per chi, come e perché.

Gli aspetti di maggiore interesse ai fini di questa tesi riguardano: le percezioni sull’importanza della salute mentale per il benessere individuale; i differenti approcci che le persone ansiose e depresse hanno sperimentato per gestire il proprio malessere; le percezioni sull’utilità dei metodi utilizzati.

L’indagine è stata svolta tramite interviste telefoniche su un campione di circa 1000 persone per ciascun paese, in 113 nazioni del mondo.

Ogni partecipante è stato estratto con criterio di scelta casuale dalle liste di coloro che hanno accesso ad un telefono fisso o cellulare, d’età superiore ai 15 anni, secondo una strategia di campionamento volta a costituire un campione statisticamente rappresentativo della popolazione residente a livello nazionale. Gallup ha usato informazioni demografiche disponibili, quali genere, età, livello di istruzione e stato socio-economico, per calcolare un insieme di pesi per ciascun rispondente, per assicurarsi che il campione complessivo rifletta tutti i possibili sottogruppi nella popolazione.

Il questionario utilizzato per la rilevazione è stato sviluppato grazie ad un attento e preciso processo per assicurarne la qualità che include: interviste con ricercatori e *stakeholders*; test preliminari e studi pilota in dieci diversi paesi per verificare che i quesiti siano compresi da tutti in ogni territorio; traduzione in tutte le principali lingue utilizzate nel mondo.

Per un completo ed esauriente approfondimento sulle questioni metodologiche relative a sviluppo, test e traduzione del questionario, gestione delle interviste telefoniche, generazione e selezione del campione, raccolta dei dati, si veda l'Appendice A della guida metodologica dell'indagine.

Si citano nel seguito alcuni principali limiti dell'indagine che comportano la necessità di interpretare con cautela i risultati ottenuti dall'analisi dei dati.

In primo luogo, l'indagine non è stata pensata come un vero e proprio studio epidemiologico: ci si è focalizzati infatti sulle percezioni delle persone rispetto ad ansia e depressione, così come quest'ultime vengono definite nel questionario, ovvero come stati estremi in cui una persona si sente così ansiosa o depressa da non riuscire a continuare le proprie attività quotidiane come farebbe normalmente, per due o più settimane. Non si sono forniti ulteriori dettagli per meglio precisare cosa si debba intendere quando si parla di ansia e depressione.

In secondo luogo, le interviste sono state condotte telefonicamente. Tra gli svantaggi di tale metodologia, rispetto ad altre forme di raccolta dati, si può citare, per esempio, il fatto che qualche partecipante potrebbe essersi sentito imbarazzato nel rispondere a determinate domande e potrebbe non aver quindi dato alcuna risposta.

Inoltre, dato che l'indagine è stata svolta durante la pandemia di Covid-19, non si può sapere con certezza quanto quest'ultima abbia avuto un impatto sui risultati. Le domande sono state però contestualizzate storicamente: i ricercatori sostengono quindi che i risultati possano essere considerati come il riflesso di attitudini ed esperienze a lungo termine.

## 1.2 Le variabili d'interesse

La selezione delle variabili considerate nelle analisi presentate nei capitoli successivi è stata svolta a partire dal questionario per la rilevazione "*Global Monitor 2020: Mental Health*". Nei paragrafi successivi si descrivono le variabili prese in esame.

### 1.2.1 Il disturbo psicologico

La variabile d'interesse principale è:

- **Aver sofferto di ansia o depressione in passato**

Si tratta di una variabile qualitativa con modalità "Sì", "No", "Non lo so/Non rispondo", che permette di identificare i soggetti che sono stati affetti da un disturbo psicologico.

In questo contesto, con il termine disturbo psicologico ci si riferisce senza distinguo soltanto ad ansia e depressione. Più precisamente, nel questionario sopra citato, si definiscono ansia e depressione nei seguenti termini: "*Sei mai stato così ansioso o depresso da non poter continuare le tue regolari attività quotidiane come avresti fatto normalmente per due settimane o più?*".

Per coloro che hanno sofferto di ansia o depressione, si considerano, inoltre, le seguenti variabili:

- **Aver avuto più di un episodio ansioso o depressivo**

Si tratta di una variabile qualitativa con modalità "Sì", "No", "Non lo so/Non rispondo".

Nel dettaglio, il quesito nel questionario associato a questa variabile recita: "*Ti sei sentito ansioso o depresso più di una volta?*".

- **Età al primo episodio ansioso o depressivo**

Si tratta di una variabile qualitativa con modalità "Meno di 13 anni", "13-19 anni", "20-29 anni", "30-39 anni", "40 anni o più".

## 1.2.2 Variabili socio-demografiche

Si considerano le seguenti variabili socio-demografiche:

- **Genere**

Si tratta di una variabile qualitativa con modalità "Maschio" e "Femmina".

- **Livello di istruzione conseguito**

Si tratta di una variabile qualitativa ordinale con modalità "Istruzione primaria", "Istruzione secondaria", "Istruzione terziaria", "Non lo so/Non rispondo".

Si precisa che ogni paese nel mondo ha un proprio modo unico di classificare i livelli di istruzione: l'esatta definizione e organizzazione variano quindi da uno stato all'altro. Per consentire quindi confronti tra paesi, nel questionario si utilizzano queste tre categorie principali. Per "Istruzione primaria" si intende l'equivalente

dell'aver completato circa otto anni di scuola. La categoria "Istruzione secondaria" si riferisce a coloro che hanno completato dai nove ai quindici anni di scuola, ma non hanno ottenuto una laurea universitaria. Infine, hanno conseguito il livello di "Istruzione terziaria" gli individui che hanno studiato per sedici o più anni della loro vita e che hanno quindi ottenuto almeno l'equivalente di una laurea universitaria.

- **Reddito**

Si tratta di una variabile qualitativa ordinale con modalità "Primo quintile", "Secondo quintile", "Terzo quintile", "Quarto quintile", "Quinto quintile".

Il reddito annuale per capita è stato calcolato dividendo il reddito familiare per il numero totale di persone del proprio nucleo familiare. Per i dati di ciascun paese, la distribuzione del reddito annuale per capita è stata successivamente suddivisa in quintili.

- **Arrivare a fine mese**

Si tratta di una variabile qualitativa ordinale con modalità "Confortevolmente", "Cavandosela", "Difficilmente", "Molto difficilmente", "Non lo so", "Non rispondo".

Tale variabile fornisce un'informazione su come il proprio reddito viene percepito soggettivamente.

- **Occupazione lavorativa**

Si tratta di una variabile qualitativa con modalità "Lavoratore dipendente a tempo pieno", "Libero professionista a tempo pieno", "Lavoratore a tempo parziale che non desidera un lavoro a tempo pieno", "Lavoratore a tempo parziale che desidera un lavoro a tempo pieno", "Disoccupato", "Fuori dalla forza lavoro".

Un soggetto viene considerato un "lavoratore dipendente a tempo pieno" se è impiegato presso un datore di lavoro per almeno 30 ore settimanali; un "libero professionista" a tempo pieno se esercita una professione in proprio per almeno 30 ore settimanali; un "lavoratore a tempo parziale" se non lavora per più di 30 ore settimanali; "disoccupato" se afferma di non aver avuto un'occupazione da almeno sette giorni e di aver cercato lavoro nelle ultime quattro settimane essendo in grado di iniziarne uno; "fuori dalla forza lavoro" se il rispondente è inoccupato da almeno sette giorni, non sta cercando un lavoro e non è disponibile ad iniziarne uno (studenti, pensionati, casalinghi, ecc.).

- **Età**

Si tratta di una variabile quantitativa intera con supporto [15, 99].

- **Paese**

Si tratta di una variabile qualitativa con 113 livelli.

### 1.2.3 Gli approcci per la gestione di un disturbo psicologico

In molte indagini, le risposte a domande precedenti determinano se ci sarà o meno una risposta a domande successive: tali quesiti sono noti come domande filtro. La probabilità di una risposta affermativa ad una data domanda è non nulla solo se il partecipante ha risposto affermativamente ad un gruppo di domande che ne precedono altre dal punto di vista logico.

In questo caso, la domanda filtro è quella relativa all'aver sperimentato o meno un disturbo ansioso o depressivo in un qualche momento della propria vita. A ciascun partecipante che ha risposto "Sì" al quesito precedente, si chiede inoltre quanto segue: *"Quando ti sentivi ansioso o depresso, hai mai messo in atto una delle seguenti azioni per sentirti meglio?"*. Le azioni elencate sono:

- a. Rivolgersi ad un professionista della salute mentale;
- b. Partecipare attivamente ad attività religiose o spirituali o cercare il supporto di una guida religiosa o spirituale;
- c. Parlare con amici o familiari;
- d. Assumere medicinali prescritti da un professionista della cura della salute;
- e. Introdurre nel proprio stile di vita comportamenti più sani, quali attività fisica, miglior riposo e una dieta equilibrata;
- f. Attuare un cambiamento significativo nella propria situazione lavorativa;
- g. Attuare un cambiamento significativo nelle relazioni personali;
- h. Trascorrere del tempo all'aperto o in mezzo alla natura.

Per coloro che hanno sofferto di ansia o depressione in passato, si considerano quindi le seguenti variabili:

- "Aver parlato con un professionista della salute mentale";

- "Aver partecipato ad attività religiose o spirituali";
- "Aver parlato con amici o familiari";
- "Aver assunto specifici medicinali sotto prescrizione medica";
- "Aver ricercato uno stile di vita più sano cambiando alcuni comportamenti";
- "Aver apportato un cambiamento alla propria situazione lavorativa";
- "Aver apportato un cambiamento nelle relazioni personali";
- "Aver trascorso tempo all'aperto o in mezzo alla natura".

Si tratta di 8 variabili qualitative, con modalità "Sì", "No", "Non lo so", che rappresentano gli otto differenti approcci sperimentati o meno dai soggetti ansiosi e depressi per alleviare il proprio malessere.

### 1.3 Operazioni preliminari

Nei paragrafi seguenti, si descrivono alcune operazioni preliminari svolte per rendere i dati pronti per la successiva fase di analisi e si giustificano le scelte di categorizzazione e di ricodifica delle modalità, effettuate per alcune variabili sulla base degli obiettivi prefissati.

Si è deciso di utilizzare i dati relativi ai seguenti 28 paesi: Austria, Belgio, Bulgaria, Cipro, Croazia, Danimarca, Estonia, Finlandia, Francia, Germania, Grecia, Irlanda, Italia, Lettonia, Lituania, Paesi Bassi, Polonia, Portogallo, Repubblica Ceca, Romania, Slovacchia, Slovenia, Spagna, Svezia, Regno Unito, Ungheria, Stati Uniti, Canada. Gli stati elencati rappresentano la quasi totalità dei paesi attualmente membri dell'Unione Europea, a cui sono stati aggiunti Regno Unito, Canada e Stati Uniti. La scelta è ricaduta su questi paesi europei e nord-americani in quanto rientrano nel gruppo dei cosiddetti "paesi sviluppati" con reddito medio-alto, secondo la classificazione proposta dalla Banca Mondiale. Ci si vuole focalizzare quindi su un gruppo di nazioni relativamente omogeneo, con aspetti culturali, economici e politici comuni.

Dato che i soggetti che hanno risposto "Non lo so/non rispondo" al quesito relativo all'aver sofferto di ansia o depressione in passato sono relativamente pochi, si è deciso di considerarli tra coloro che hanno risposto "No", invece che eliminarli. In questo modo si è ottenuta quindi una variabile dicotomica che indica se il singolo soggetto ha dichiarato di essere stato affetto o meno da un disturbo d'ansia o depressione. Si è proceduto in

modo analogo per quanto riguarda l'accorpamento delle modalità della variabile relativa all'aver avuto più di un episodio ansioso o depressivo.

Per quanto riguarda le variabili “Livello di istruzione conseguito” e “Arrivare a fine mese”, sono state eliminate le pochissime unità statistiche che hanno risposto “Non lo so/non rispondo” ai quesiti corrispondenti. Conseguentemente, le modalità sopra-citate sono state eliminate.

Le tre modalità centrali della variabile “Reddito” sono state accorpate in un'unica modalità. Dopo questa operazione, la variabile “Reddito” presenta quindi tre categorie: “Reddito basso”, “Reddito medio”, “Reddito alto”.

Sono state accorpate anche le modalità “Difficilmente” e “Molto difficilmente” della variabile “Arrivare a fine mese”. Dopo questa operazione, la variabile “Arrivare a fine mese” presenta quindi tre categorie: “Confortevolmente”, “Cavandosela”, “Difficilmente”.

Inoltre, considerando la variabile “Occupazione lavorativa”, sono state accorpate le seguenti modalità: “Lavoratore a tempo pieno” e “Libero professionista a tempo pieno” nella modalità “Occupato a tempo pieno”; “Lavoratore a tempo parziale che non desidera un lavoro a tempo pieno” e “Lavoratore a tempo parziale che desidera un lavoro a tempo pieno” nella modalità “Occupato a tempo parziale”.

Infine, per tutte le otto variabili relative agli approcci sperimentati per gestire ansia o depressione, sono state ricodificate le tre modalità presenti, spostando i pochissimi soggetti che hanno risposto “Non lo so” insieme a quelli che hanno risposto “No”, ottenendo quindi otto variabili dicotomiche.

Al termine di queste operazioni preliminari, il dataset di riferimento presenta 27475 osservazioni, circa 1000 per ciascuno dei paesi selezionati, e 18 variabili.

# Capitolo 2

## Analisi esplorative

### 2.1 Ansia e depressione

In questa sezione, si riportano i risultati delle analisi esplorative volte a descrivere il campione in esame e la relazione tra l'aver avuto un disturbo psicologico e i possibili fattori di rischio considerati. In particolare, si vuole fotografare lo stato di salute mentale in Unione Europea, Regno Unito, Canada e Stati Uniti, considerando i disturbi depressivi e i disturbi d'ansia.

La prevalenza di ansia e depressione, stimata sul campione di 27475 soggetti d'età superiore ai 15 anni, residenti nei 28 paesi considerati, è pari a 18%: circa una su cinque persone ha quindi affermato di aver sperimentato, in un qualche momento della propria vita, ansia o depressione, definite come degli stati psicologici debilitanti, al punto da non poter continuare le proprie attività quotidiane normalmente per due o più settimane.

Tra i 4884 soggetti che sono stati affetti da un disturbo mentale, il 72% ha affermato di aver avuto più di un episodio ansioso o depressivo: quasi tre quarti delle persone che sono state ansiose o depresse in passato, si sono sentite in questo modo più di una volta nella loro vita.

La Figura 2.1 riporta le percentuali di persone nel campione che hanno sofferto d'ansia o depressione per la prima volta in cinque fasce d'età. Si può notare che la metà dei soggetti considerati ha avuto la sua prima esperienza prima di compiere 30 anni; solo il 5% è stato ansioso o depresso per la prima volta ad un'età inferiore ai 13 anni.

A livello di singolo paese, dalla Tabella 2.1 si può osservare che gli stati con una più alta prevalenza di soggetti che sono stati ansiosi e depressi almeno una volta nella vita, sono Bulgaria (28%), Francia (27%), Grecia (27%) e Regno Unito (27%), mentre le nazioni con una prevalenza inferiore sono Croazia (11%), Ungheria (11%), Slovenia (9%)

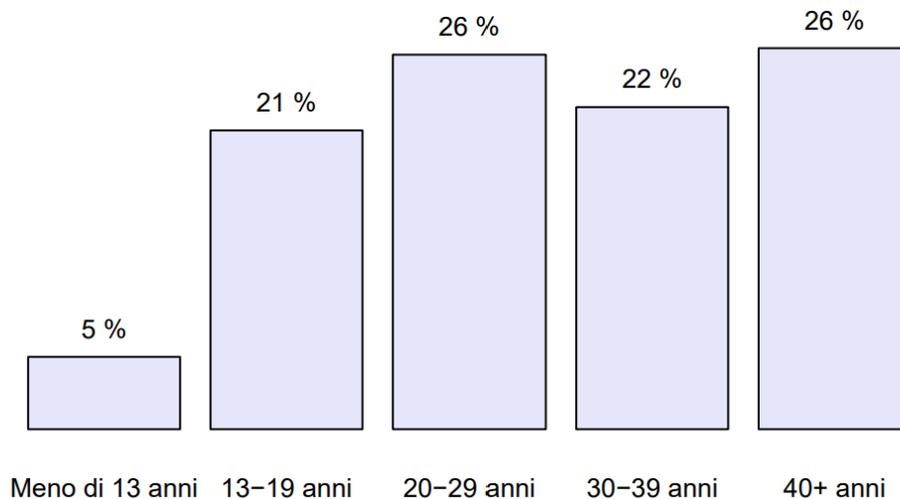


FIGURA 2.1: Età a cui le persone sono state ansiose o depresse per la prima volta.

e Lituania (8%). L'Italia si colloca al di sotto della media con una stima della prevalenza pari a 15%. Si suggerisce di interpretare questi dati con cautela: trarre la conclusione, per esempio, che i lituani sono meno depressi rispetto a tutti gli altri, potrebbe essere fuorviante. Infatti, potrebbero essere presenti differenze culturali tra paesi a livello di definizione di ansia e depressione, o stigmi sociali che inducono le persone a dichiarare il falso.

Le stime delle percentuali di soggetti che hanno dichiarato di aver avuto un disturbo psicologico, nelle quattro fasce d'età considerate (età al momento dell'intervista), sono: 22% per gli under 30, 19% in età 30-49 anni, 17% in età 50-64 anni e 14% per gli over 65. A livello globale, la probabilità di aver sperimentato ansia o depressione sembra quindi decrescere con l'età. In particolare, si osserva una prevalenza massima per coloro che hanno meno di trent'anni e una prevalenza minima per chi ne ha più di 65. In generale, le persone più anziane sembrano meno propense a dichiarare di essere state ansiose o depresse rispetto ai più giovani. Notare ciò è importante in quanto gli adulti hanno avuto più tempo per sperimentare ansia e depressione rispetto agli adolescenti.

Per quanto riguarda la stima della prevalenza di soggetti che hanno sofferto di ansia o depressione per genere, il 20% delle donne ha affermato di aver sperimentato un disturbo psicologico, contro il 15% degli uomini. Dal rapporto tra queste due percentuali, si evidenzia che la prevalenza del disturbo nelle femmine è circa 1.3 volte maggiore rispetto ai maschi. Per quanto riguarda la diffusione di questi disordini mentali nei due sessi, si osserva che il 59% dei soggetti che sono stati ansiosi o depressi è di sesso femminile, il

TABELLA 2.1: Stima della prevalenza di soggetti che hanno sofferto di ansia o depressione per paese in ordine decrescente.

Paese	Stima della prevalenza
Bulgaria	28%
Francia	27%
Grecia	27%
Regno Unito	27%
Portogallo	24%
Canada	21%
Cipro	21%
Spagna	21%
Estonia	19%
Paesi Bassi	19%
Lettonia	18%
Polonia	18%
Svezia	18%
Danimarca	17%
Germania	17%
Irlanda	17%
Belgio	16%
Stati Uniti	16%
Italia	15%
Romania	15%
Austria	14%
Finlandia	14%
Slovacchia	14%
Repubblica Ceca	13%
Croazia	11%
Ungheria	11%
Slovenia	9%
Lituania	8%

41% di sesso maschile.

In generale, i dati evidenziano come ci siano prevalenze maggiori tra i soggetti più svantaggiati, rispetto al livello di educazione, all'occupazione lavorativa, al reddito reale e percepito. Il 22%, 19% e 16% di coloro che hanno conseguito, rispettivamente, un livello di istruzione primaria, secondaria e terziaria, ha dichiarato di aver sofferto di ansia e depressione almeno una volta nella vita. La prevalenza per chi ha una formazione elementare è circa 1.4 volte quella per i soggetti con un diploma universitario.

Il 22%, 18% e 16% di coloro che possiedono, rispettivamente, un reddito basso, medio e alto, ha asserito di aver sofferto di ansia e depressione almeno una volta nella vita. Questo significa che, globalmente, la prevalenza per le persone che si collocano nel primo

quintile della distribuzione del reddito del proprio paese è circa 1.4 volte quella per coloro che appartengono all'ultimo quintile.

Il 32%, 18%, 20% e 16%, rispettivamente, dei disoccupati, inattivi, occupati a tempo parziale e occupati a tempo pieno, ha affermato di aver sofferto di ansia e depressione. Nel dettaglio, i disoccupati che sono stati ansiosi o depressi sono due volte gli occupati a tempo pieno.

Il 29%, 17% e 14% di coloro che ritengono di arrivare a fine mese, rispettivamente, con difficoltà, cavandosela e confortevolmente, ha dichiarato di aver sofferto di ansia e depressione. La prevalenza per chi ritiene di avere problemi economici è circa più di due volte quella per coloro che ritengono di non avere alcuna difficoltà.

Riassumendo, nel campione analizzato, la prevalenza di soggetti che sono stati ansiosi o depressi: decresce con l'età; aumenta del 30% per le donne, rispetto agli uomini; aumenta del 40% per chi ha un'educazione elementare, rispetto a un laureato; aumenta del 40% per chi ha un reddito basso, rispetto ad un reddito alto; è più elevata tra i disoccupati rispetto a inattivi ed occupati; raddoppia tra coloro che arrivano a fine mese con difficoltà, rispetto a chi ci arriva confortevolmente. Questi dati sembrano quindi suggerire che ci sia un'associazione tra: un livello educativo basso e l'aver sperimentato un malessere psicologico; una condizione di svantaggio economico reale o percepito e l'aver sperimentato un disturbo ansioso-depressivo. Inoltre, anche il genere e la condizione lavorativa sembrano essere fattori rilevanti per la salute mentale.

## 2.2 Metodi per la gestione di ansia e depressione

Questa sezione delle analisi esplorative si focalizza su quali e quanti metodi le persone ansiose e depresse hanno sperimentato per cercare di sentirsi meglio e sul ruolo di alcuni fattori socio-demografici nella scelta di tali strategie.

In generale, gli studi statistici sulla disponibilità di servizi per la salute mentale a livello globale, tendono a focalizzarsi sulle strutture e sul personale specializzato in metodi clinici. L'indagine *Global Monitor* ha invece approfondito l'ampia scelta di approcci a cui fanno ricorso le persone per gestire la propria condizione ansiosa o depressiva nella quotidianità.

Dalla Figura 2.2 si può notare che, a livello globale, i tre metodi più sperimentati sono: cercare il supporto di amici e familiari, trascorrere del tempo all'aria aperta e introdurre uno stile di vita più sano. Complessivamente, per cercare di attenuare la propria sofferenza psicologica, l'81% delle persone ha parlato con amici e familiari; il 79% e il 72% dei soggetti ansiosi o depressi ha scelto, rispettivamente, di passare del

tempo in mezzo alla natura e seguire comportamenti più salutari. Partecipare ad attività religiose o spirituali è invece l'approccio meno frequentemente selezionato (25%).

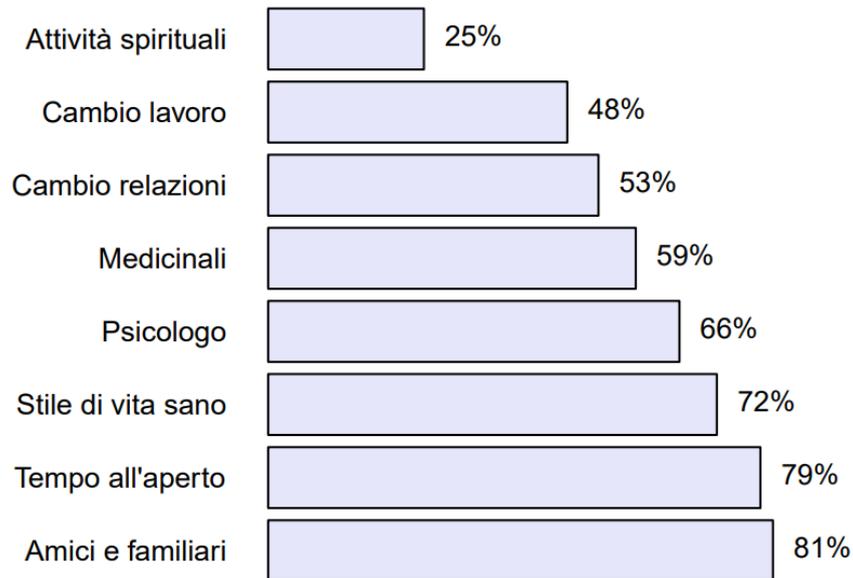


FIGURA 2.2: Percentuali di persone che hanno scelto ciascun metodo in ordine decrescente.

La Figura 2.3 mostra il numero di approcci sperimentati: in media, le persone che sono state ansiose o depresse hanno provato 4.8 delle otto azioni indicate nell'indagine *Global Monitor*. Nella Tabella 2.2 si riportano le 10 combinazioni di coppie di metodi. I seguenti approcci sono maggiormente associati l'uno all'altro: aver parlato con uno psicologo con aver assunto medicinali sotto prescrizione medica, essersi affidati ad amici e familiari con aver trascorso del tempo all'aperto.

Nel seguito, si vuole valutare, a livello descrittivo, se la scelta dei diversi approcci sembra essere legata ad alcuni fattori socio-demografici o meno.

Dalla Figura 2.4 si può notare che, per tutti gli approcci considerati, non sembra evidente una diversa tendenza nella scelta dei metodi sperimentati tra i soggetti di età compresa tra i 30 e i 49 anni e i soggetti tra i 50 e i 64 anni. Inoltre, l'età sembra influenzare maggiormente la scelta dei seguenti comportamenti: parlare con uno psicologo, assumere medicinali, cambiare lavoro e cambiare relazioni.

Rispetto agli altri, i più giovani sembrano meno propensi a ricercare il supporto di uno psicologo, ad assumere medicinali e a cambiare la propria situazione lavorativa. Per

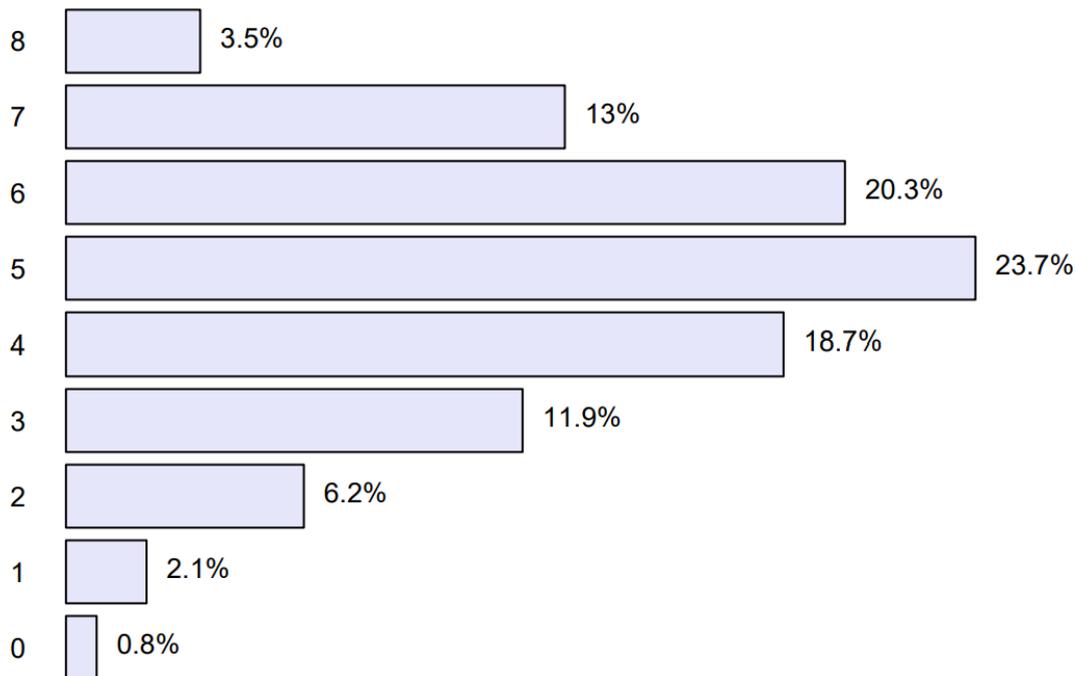


FIGURA 2.3: Numero di approcci sperimentati.

esempio, il 61% degli under 30 ha scelto di rivolgersi ad un professionista della salute mentale, contro il 68% degli appartenenti alla classe d'età 50-64 anni; il 39% degli under 30 ha richiesto la prescrizione di un medicinale, contro il 70% degli over 65.

Rispetto agli altri, i più giovani sembrano più propensi ad attuare un cambiamento significativo nelle proprie relazioni interpersonali. La scelta di questo approccio è circa 1.4 volte più frequente per gli under 30, rispetto agli over 65.

Per quanto riguarda l'assunzione di medicinali, si nota, invece, una tendenza a scegliere questo metodo crescente con l'età. Infatti, la preferenza degli anziani per questo approccio è 1.8 volte quella dei più giovani.

Per quando concerne il cambiamento nelle relazioni, si osserva una tendenza a scegliere questo metodo decrescente con l'età. Questo comportamento è infatti 0.73 volte meno frequente per gli over 65, rispetto agli under 30.

Prendendo in considerazione i metodi sperimentati per genere, la Figura 2.5 mostra alcune piccole differenze tra maschi e femmine. In particolare, le donne sembrano più propense a parlare con un professionista della salute mentale e ad assumere medicinali rispetto agli uomini; i maschi sembrano più disposti a cambiare la propria condizione lavorativa rispetto alle femmine.

Nella Figura 2.6 si presentano i metodi scelti in base al livello scolastico raggiunto.

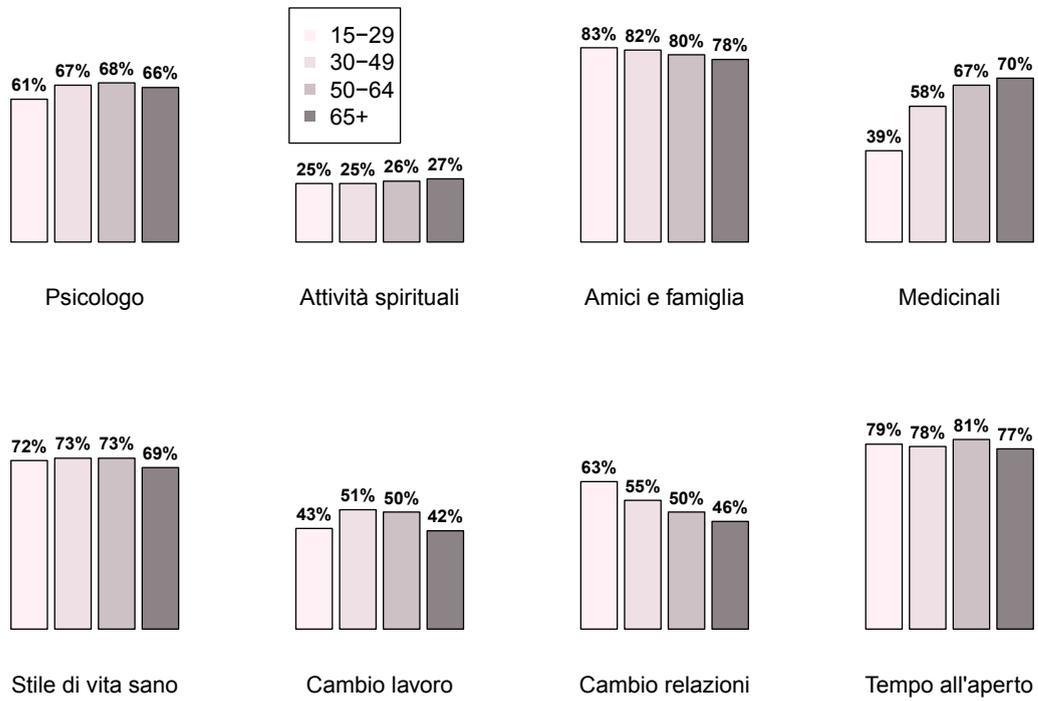


FIGURA 2.4: Metodi scelti per fasce d'età.

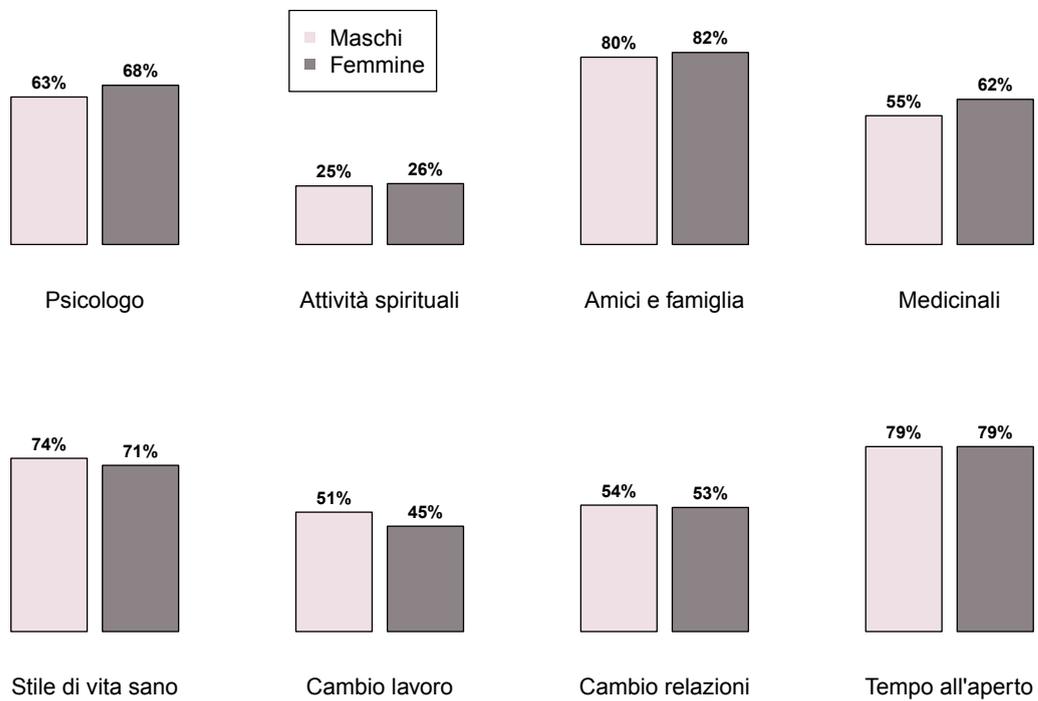


FIGURA 2.5: Metodi scelti per genere.

In confronto a chi ha conseguito un livello di istruzione primaria o secondaria, chi ha ottenuto almeno un diploma universitario sembra più propenso a iniziare una terapia con un analista, chiedere un supporto ad amici e familiari, ricercare uno stile di vita più sano e cambiare lavoro. Per esempio, il 70% dei soggetti con istruzione terziaria ha sperimentato la psicoterapia, contro il 61% di quelli con istruzione primaria; il 76% dei soggetti con istruzione terziaria ha introdotto comportamenti più sani, contro il 67% di quelli con istruzione primaria.

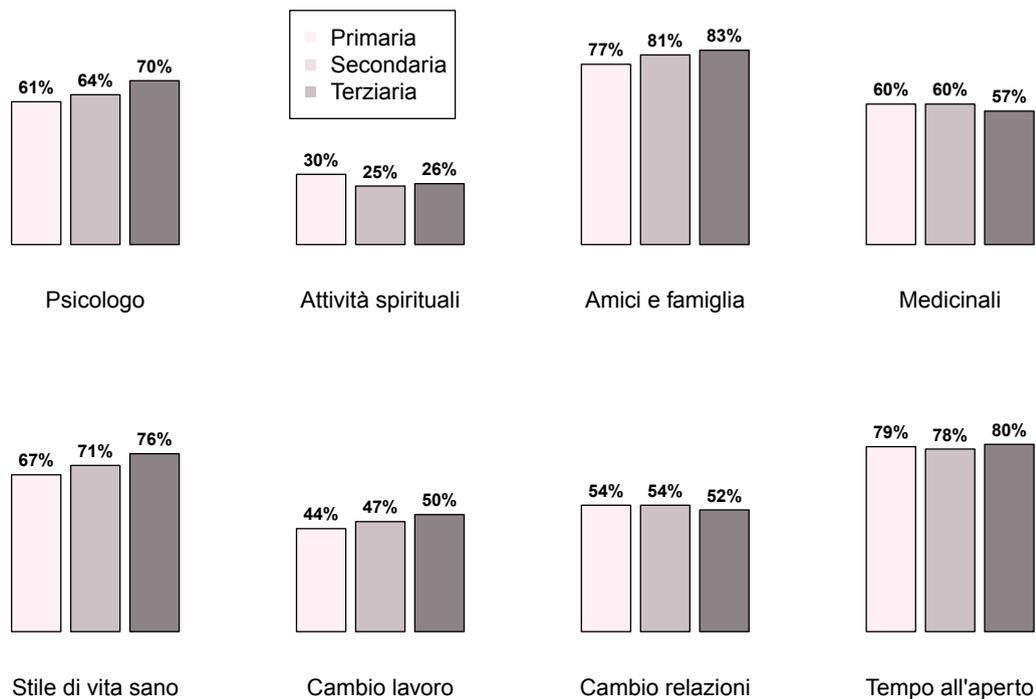


FIGURA 2.6: Metodi scelti per livello di istruzione conseguito.

Nel seguito si riportano i grafici relativi ai metodi scelti per livello di reddito, per tipo di occupazione lavorativa e per percezione soggettiva sulla capacità di arrivare a fine mese. Questi grafici sembrano suggerire l'assenza di un'associazione tra l'aver scelto uno specifico approccio e i tre fattori socio-demografici considerati. Si possono però fare le seguenti osservazioni.

Rispetto a chi ha un reddito medio-basso, chi ha un reddito alto sembra aver cercato più frequentemente il supporto di uno psicologo e aver partecipato meno frequentemente ad attività religiose o spirituali. Infatti, il 70% dei soggetti con reddito alto ha parlato con uno psicoterapeuta, contro il 63% dei soggetti con reddito basso; il 22% dei soggetti

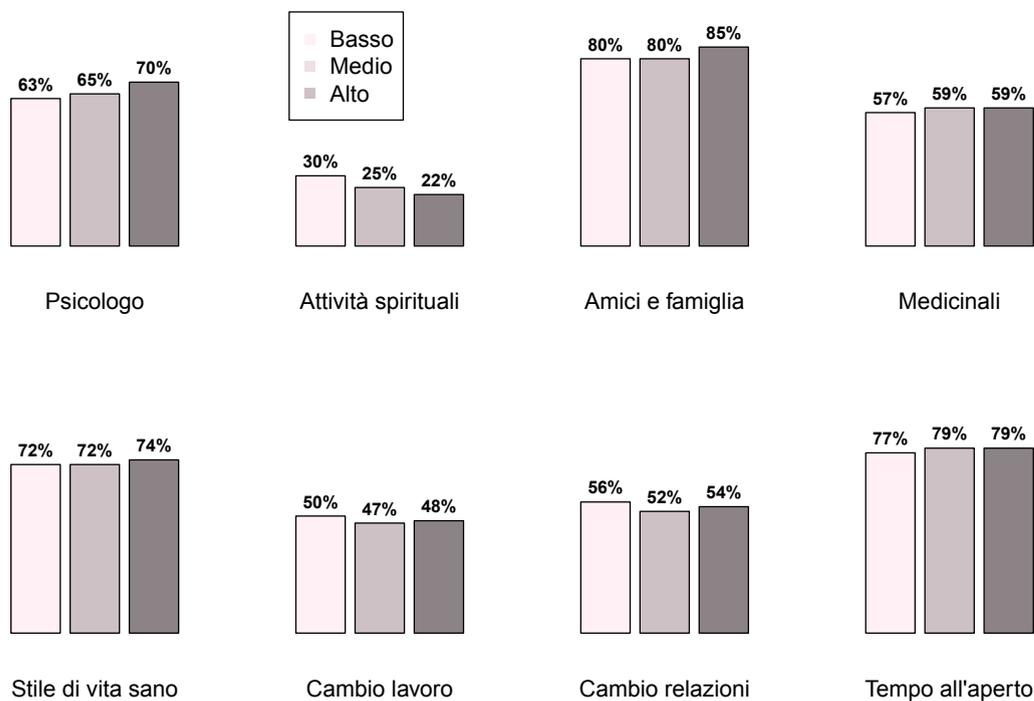


FIGURA 2.7: Metodi scelti per livello di reddito.

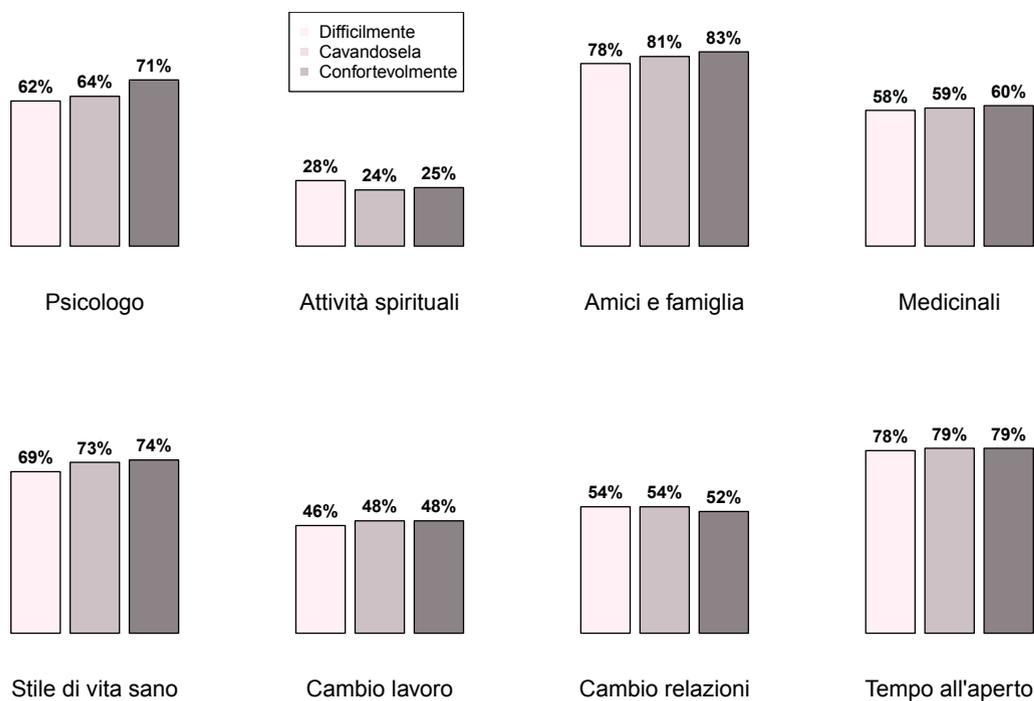


FIGURA 2.8: Metodi scelti per percezione soggettiva sulla capacità di arrivare a fine mese.

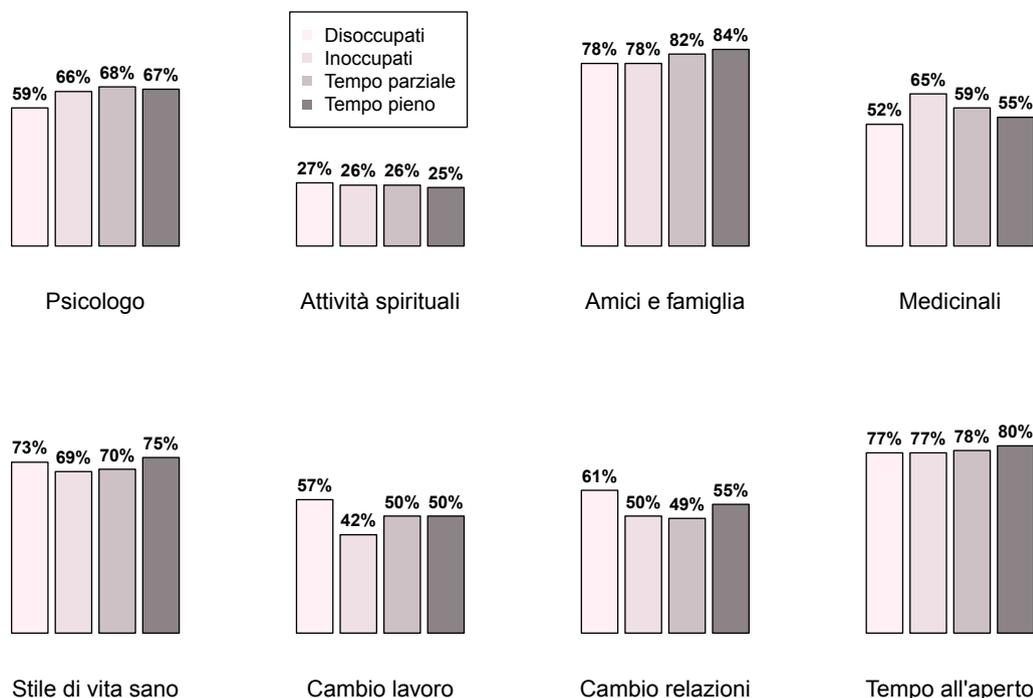


FIGURA 2.9: Metodi scelti per tipo di occupazione lavorativa.

con reddito alto si è rivolto ad una guida religiosa o spirituale, contro il 30% dei soggetti con reddito basso.

Inoltre, rispetto a chi arriva a fine mese con difficoltà o cavandosela, chi ritiene di avere una buona situazione economica sembra essere più propenso a chiedere il supporto di uno psicologo. Il 71% di chi afferma di non avere problemi ad affrontare le spese mensili si è rivolto ad un terapeuta, contro il 62% di coloro che percepiscono il proprio reddito come non sufficiente per giungere serenamente a fine mese.

Infine, il 59% dei disoccupati ha scelto di affidarsi ad un professionista della salute mentale, contro il 68% degli occupati a tempo parziale. Il 52% dei disoccupati ha assunto medicinali sotto prescrizione medica, rispetto al 65% degli inoccupati. Il 57% dei disoccupati ha apportato un cambiamento alla propria situazione lavorativa, in confronto al 42% degli inoccupati.

TABELLA 2.2: Le 10 combinazioni di metodi per gestire ansia o depressione più frequenti.

Psicologo	Attività spirituali	Amici/famiglia	Medicinali	Stile di vita sano	Cambio lavoro	Cambio relazioni	Natura	%
Sì	No	Sì	Sì	Sì	Sì	Sì	Sì	6.1
Sì	No	Sì	Sì	Sì	No	No	Sì	4.7
Sì	No	Sì	Sì	Sì	No	Sì	Sì	4.3
Sì	No	Sì	Sì	Sì	Sì	No	Sì	3.5
Sì	Sì	Sì	Sì	Sì	Sì	Sì	Sì	3.5
No	No	Sì	No	Sì	Sì	Sì	Sì	2.6
Sì	No	Sì	No	Sì	Sì	Sì	Sì	2.5
No	No	Sì	No	Sì	No	No	Sì	2.1
Sì	Sì	Sì	Sì	Sì	No	Sì	Sì	1.9
No	No	Sì	No	Sì	No	Sì	Sì	1.9



# Capitolo 3

## Modello di regressione logistica

In questo capitolo, si considera un modello di regressione logistica, per valutare da un punto di vista inferenziale, se e come l'aver sofferto di ansia o depressione dipenda da alcune caratteristiche socio-demografiche.

Il modello di regressione logistica, o modello logit, appartiene alla classe dei modelli lineari generalizzati (GLM). La famiglia della distribuzione di probabilità è quella binomiale. Il modello logit è utile in tutti i casi in cui la variabile risposta si presenta come una variabile dicotomica discreta, le cui modalità rappresentano due alternative mutualmente esclusive (McCullagh & Nelder, 1989). Per convenzione, la variabile risposta assume valori pari a 0 oppure 1.

### 3.1 Specificazione del modello

Siano  $Y_i$  le variabili risposta che indicano se il soggetto  $i$ -esimo ha sperimentato o meno un disturbo ansioso-depressivo, in un qualche momento della propria vita, dove  $i \in \{1, \dots, n\}$ , con  $n$  pari al totale delle osservazioni nel campione su cui viene stimato il modello.

Sia  $Y_i = 1$  se il soggetto è stato ansioso o depresso;  $Y_i = 0$  altrimenti.

Sia  $y_i$  il valore osservato per  $Y_i$ .

Sia  $x_i$  il vettore dei valori osservati delle covariate per l' $i$ -esima osservazione.

Sia  $\beta$  il vettore dei parametri che devono essere stimati e  $\beta_j$  il coefficiente associato alla  $j$ -esima variabile esplicativa, con  $j \in \{1, \dots, J\}$ .

Il modello assume che le  $Y_i$  siano tra loro indipendenti con  $Y_i \sim Bi(1, \pi_i)$ . La distribuzione di  $Y_i$  è detta distribuzione di Bernoulli con parametro  $\pi_i$  e può essere scritta nella forma

$$Pr(Y_i = y_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}. \quad (3.1)$$

Se  $y_i = 1$ , si ottiene

$$Pr((Y_i = 1)) = \pi_i. \quad (3.2)$$

Se  $y_i = 0$ , si ottiene

$$Pr((Y_i = 0)) = 1 - \pi_i. \quad (3.3)$$

Il modello assume inoltre che

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \log\left(\frac{\pi(x_i; \beta)}{1 - \pi(x_i; \beta)}\right) = x_i \beta. \quad (3.4)$$

La probabilità condizionata che il soggetto  $i$ -esimo abbia sofferto di ansia o depressione, date le covariate  $x_i$ , è quindi

$$\pi_i = \pi(x_i; \beta) = Pr(Y_i = 1|x_i) = \frac{\exp(x_i \beta)}{1 + \exp(x_i \beta)}. \quad (3.5)$$

La funzione di verosimiglianza è

$$L(\beta) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} = \prod_{i=1}^n \pi(x_i; \beta)^{y_i} (1 - \pi(x_i; \beta))^{1-y_i}. \quad (3.6)$$

## 3.2 Stima e interpretazione dei parametri

Per stimare i parametri del modello di regressione logistica, si impiega il metodo della massima verosimiglianza. Dato che le equazioni generate dalla massimizzazione della verosimiglianza sono non lineari nei parametri, ovvero non ammettono soluzione esplicita, le stime dei coefficienti si ottengono utilizzando procedure numeriche iterative (McCullagh & Nelder, 1989).

Siano  $\widehat{\beta}_j$  lo stimatore di massima verosimiglianza del  $j$ -esimo parametro e  $\widehat{var}(\beta_j)$  la stima della sua varianza.

Per valutare la significatività dei singoli coefficienti si può testare il sistema d'ipotesi  $H_0 : \beta_j = 0$  contro  $H_1 : \beta_j \neq 0$  utilizzando la statistica test

$$z = \frac{\widehat{\beta}_j - 0}{\sqrt{\widehat{var}(\widehat{\beta}_j)}}. \quad (3.7)$$

Se  $|z| > z_{\alpha/2}$  si rifiuta l'ipotesi nulla al livello di significatività  $\alpha$  prescelto, con  $z_{\alpha/2}$  valore critico della distribuzione normale standard.

Per costruire un intervallo di confidenza per  $\widehat{\beta}_j$  al livello di confidenza  $(1 - \alpha)\%$ , cosiddetto alla Wald, si utilizza la formula

$$\widehat{\beta}_j \pm z_{1-\alpha/2} \sqrt{\widehat{var}(\widehat{\beta}_j)}, \quad (3.8)$$

con  $z_{1-\alpha/2}$  valore critico della distribuzione normale standard.

Il segno di  $\widehat{\beta}_j$  fornisce un'importante informazione a livello interpretativo. Se  $\widehat{\beta}_j > 0$ , allora  $Pr(Y_i = 1|x_i)$  aumenta; viceversa se  $\widehat{\beta}_j < 0$ , allora  $Pr(Y_i = 1|x_i)$  diminuisce.

Per l'interpretazione dei parametri si preferisce spesso ragionare in termini di rapporti tra le quote. La quota, detta *odds*, rappresenta in questa analisi il rapporto tra la probabilità di aver sofferto di ansia o depressione e la probabilità di non averne sofferto. Per l' $i$ -esima unità statistica, la quota si definisce come

$$odds_i = \frac{\pi_i}{1 - \pi_i} = \frac{Pr(Y_i = 1|x_i)}{1 - Pr(Y_i = 1|x_i)} = \frac{Pr(Y_i = 1|x_i)}{Pr(Y_i = 0|x_i)} = \exp(x_i\beta). \quad (3.9)$$

A titolo d'esempio, se si considera una covariata dicotomica  $x_j$ , fissati i valori delle altre variabili esplicative, il rapporto delle quote, detto *odds ratio* (*OR*), per la variabile  $x_j$ , si definisce come

$$OR_{x_j} = \frac{odds(x_j = 1)}{odds(x_j = 0)} = \frac{\frac{Pr(Y_i=1|x_j=1)}{1-Pr(Y_i=1|x_j=1)}}{\frac{Pr(Y_i=1|x_j=0)}{1-Pr(Y_i=1|x_j=0)}} = \exp(\beta_j). \quad (3.10)$$

Se  $OR_{x_j} > 1$ , la quota per  $x_j = 1$  aumenta rispetto alla quota per  $x_j = 0$ ; viceversa, se  $OR_{x_j} < 1$ , la quota per  $x_j = 1$  diminuisce rispetto alla quota per  $x_j = 0$ .

Per costruire un intervallo di confidenza per  $OR_{x_j}$ , ci si limita a calcolare l'esponentiale degli estremi dell'intervallo di confidenza ottenuto con la formula (3.8).

### 3.3 Selezione del modello

La variabile risposta considerata è aver sofferto di ansia o depressione. I dati si trovano nella forma non raggruppata, ovvero il vettore delle risposte rappresenta gli esiti individuali con valori pari a 0, se un soggetto non ha sperimentato un disturbo depressivo o d'ansia, almeno una volta nella vita, oppure 1 se lo ha sperimentato.

Le variabili esplicative di cui si vuole studiare l'effetto sono:

- Genere, con modalità di riferimento “Maschio”;
- Livello di istruzione conseguito, con modalità di riferimento “Istruzione primaria”;
- Arrivare a fine mese (o reddito percepito soggettivamente), con modalità di riferimento “Con difficoltà”;
- Occupazione lavorativa, con modalità di riferimento “Disoccupato”;
- Età;
- Paese, con modalità di riferimento “Italia”.

La scelta delle modalità di riferimento è ricaduta sulle categorie citate, sulla base di considerazioni esplicitate nella Sezione 2.1 delle analisi esplorative. Per una completa descrizione delle variabili elencate si veda il paragrafo 1.2.2.

In primo luogo, si è adattato un modello di regressione logistica con i soli effetti principali. Il modello è stato stimato su tutte le unità statistiche a disposizione ( $n = 27475$ ). Si è inoltre utilizzata una procedura di selezione dei predittori passo a passo in avanti, basata sul criterio dell'AIC, a partire dal modello nullo, ovvero con la sola intercetta. L'AIC si definisce come

$$AIC = -2 \log L + 2k, \quad (3.11)$$

con  $k$  numero di parametri stimati e  $L$  il valore della funzione di verosimiglianza massimizzata.

A seguito di tale procedura, il modello finale, che seleziona tutte le covariate e assume l'assenza di interazioni, è

$$\begin{aligned}
\text{logit}(\pi_i) = x_i\beta = & \alpha + \beta_1g_1 + \beta_2i_1 + \beta_3i_2 + \beta_4x + \beta_5a_1 + \beta_6a_2 + \beta_7l_1 + \beta_8l_2 \\
& + \beta_9l_3 + \beta_{10}s_1 + \beta_{11}s_2 + \beta_{12}s_3 + \beta_{13}s_4 + \beta_{14}s_5 + \beta_{15}s_6 \\
& + \beta_{16}s_7 + \beta_{17}s_8 + \beta_{18}s_9 + \beta_{19}s_{10} + \beta_{20}s_{11} + \beta_{21}s_{12} + \beta_{22}s_{13} \\
& + \beta_{23}s_{14} + \beta_{24}s_{15} + \beta_{25}s_{16} + \beta_{26}s_{17} + \beta_{27}s_{18} + \beta_{28}s_{19} + \beta_{29}s_{20} \\
& + \beta_{30}s_{21} + \beta_{31}s_{22} + \beta_{32}s_{23} + \beta_{33}s_{24} + \beta_{34}s_{25} + \beta_{35}s_{26} + \beta_{36}s_{27},
\end{aligned} \tag{3.12}$$

dove  $x$  indica l'età in anni e

$g_1 = 1$  per femmina, e 0 altrimenti;

$i_1 = 1$  per istruzione secondaria, e 0 altrimenti;

$i_2 = 1$  per istruzione terziaria, e 0 altrimenti;

$a_1 = 1$  per cavarsela, e 0 altrimenti;

$a_2 = 1$  per confortevolmente, e 0 altrimenti;

$l_1 = 1$  per inoccupato, e 0 altrimenti;

$l_2 = 1$  per lavoro a tempo parziale, e 0 altrimenti;

$l_3 = 1$  per lavoro a tempo pieno, e 0 altrimenti.

$s_1 = 1$  per Austria, e 0 altrimenti;

$s_2 = 1$  per Belgio, e 0 altrimenti;

$s_3 = 1$  per Bulgaria, e 0 altrimenti;

$s_4 = 1$  per Canada, e 0 altrimenti;

$s_5 = 1$  per Croazia, e 0 altrimenti;

$s_6 = 1$  per Cipro, e 0 altrimenti;

$s_7 = 1$  per Repubblica Ceca, e 0 altrimenti;

$s_8 = 1$  per Danimarca, e 0 altrimenti;

$s_9 = 1$  per Estonia, e 0 altrimenti;

$s_{10} = 1$  per Finlandia, e 0 altrimenti;

$s_{11} = 1$  per Francia, e 0 altrimenti;

$s_{12} = 1$  per Germania, e 0 altrimenti;

$s_{13} = 1$  per Grecia, e 0 altrimenti;

$s_{14} = 1$  per Ungheria, e 0 altrimenti;

$s_{15} = 1$  per Irlanda, e 0 altrimenti;

$s_{16} = 1$  per Lettonia, e 0 altrimenti;

$s_{17} = 1$  per Lituania, e 0 altrimenti;

$s_{18} = 1$  per Paesi Bassi, e 0 altrimenti;

$s_{19} = 1$  per Polonia, e 0 altrimenti;

$s_{20} = 1$  per Portogallo, e 0 altrimenti;

$s_{21} = 1$  per Romania, e 0 altrimenti;

$s_{22} = 1$  per Slovacchia, e 0 altrimenti;

$s_{23} = 1$  per Slovenia, e 0 altrimenti;

$s_{24} = 1$  per Spagna, e 0 altrimenti;

$s_{25} = 1$  per Svezia, e 0 altrimenti;

$s_{26} = 1$  per Regno Unito, e 0 altrimenti;

$s_{27} = 1$  per Stati Uniti, e 0 altrimenti;

In un secondo momento, si è adattato un modello logit inserendo tra i predittori anche le interazioni del secondo ordine tra le variabili esplicative considerate. Si è utilizzata, anche in questo caso, una procedura di selezione delle covariate passo a passo in avanti, basata sul criterio dell'AIC, a partire dal modello nullo. Il modello finale seleziona tutti gli effetti principali e le seguenti interazioni del secondo ordine:

- Età e Paese;
- Arrivare a fine mese e Paese;
- Genere e Paese;

- Et  e Genere;
- Livello di istruzione e Genere.

Per la scelta tra il primo modello, con i soli effetti principali, e il secondo, con alcune interazioni del secondo ordine, si pu  valutare l'AIC, che per i due modelli   rispettivamente pari a 24465.79 e 24424.02. Si pu  notare che l'AIC diminuisce di pochissimo per il modello che tiene conto anche delle interazioni del secondo ordine, rispetto a quello che non le inserisce. Dato che non si riscontra una differenza evidente in termini di AIC tra l'uno e l'altro modello, si preferisce scegliere quello che tiene conto soltanto degli effetti principali delle variabili selezionate, in quanto pi  parsimonioso e conseguentemente di pi  facile interpretazione. Un'ulteriore motivazione a giustificazione di tale scelta, consiste nel fatto che la maggior parte dei parametri stimati dal modello con le interazioni non risultano statisticamente significativi.

### 3.4 Bont  di adattamento

In questa sezione si valuta la bont  di adattamento del modello di regressione logistica selezionato.

Quando i dati sono organizzati in gruppi, per testare l'ipotesi nulla che il modello presenti un buon adattamento, si utilizzano le statistiche test  $\chi^2$  di Pearson e  $G^2$ , per confrontare valori predetti con valori osservati.

Invece, nel caso di dati non raggruppati, con tanti predittori, tra cui una o pi  variabili continue, Agresti (2019) suggerisce di utilizzare il test di Hosmer e Lemeshow. Infatti, in presenza di una o pi  variabili esplicative quantitative, risulta difficile valutare l'eventuale mancanza di adattamento senza un qualche tipo di raggruppamento. Tuttavia, quando il numero delle covariate cresce, la combinazione delle modalit  delle variabili esplicative pu  produrre una tabella di contingenza con un numero troppo elevato di celle, molte delle quali potrebbero presentare conteggi piccoli (Agresti, 2019). In presenza di tabelle di contingenza sparse, le statistiche test  $\chi^2$  e  $G^2$  non sono appropriate (Van Kollenburg et al., 2015).

Il test di Hosmer e Lemeshow si basa sul partizionamento dei valori previsti e osservati considerando le probabilit  predette. Si procede formando, per esempio, 10 gruppi ciascuno con  $n/10$  osservazioni (Hosmer & Lemeshow, 2000). Nel primo gruppo si collocano le coppie di valori osservati e predetti con probabilit  predette nel primo decile, e cos  via. Tutte le coppie con uguale probabilit  predetta rientrano nello stesso

gruppo. Per ogni gruppo, il valore previsto si calcola come somma delle probabilità predette per uno specifico esito per tutte le osservazioni in quel gruppo.

La statistica test di Hosmer e Lemeshow è

$$G_{HL}^2 = \sum_{j=1}^J \frac{(O_j - E_j)^2}{E_j(1 - E_j/n)} \sim \chi_{g-2}^2, \quad (3.13)$$

con  $g$  numero di gruppi formati,  $J$  numero di covariate inserite nel modello,  $O_j$  valori osservati e  $E_j$  valori predetti.

Per il modello selezionato nel paragrafo precedente, il test è stato applicato con  $g = 10$  gruppi. Si può concludere che la statistica test non mostra evidenza di una mancanza di adattamento ( $p$ -value = 0.42). Inoltre, a partire dalla tabella di confusione, con soglia di discriminazione pari a 0.5, sono state calcolate alcune metriche per la valutazione del modello. In particolare, il modello presenta accuratezza pari a 0.82, specificità pari a 0.98 e precisione pari a 0.79.

### 3.5 Effetti dei fattori socio-demografici

In questa sezione si interpretano i risultati ottenuti adattando il modello di regressione logistica precedentemente selezionato.

Nella Tabella 3.1 si riportano le stime dei rapporti delle quote e i corrispondenti intervalli di confidenza. Per completezza, nella Tabella 3.2 si riportano anche le stime dei parametri, i rispettivi intervalli di confidenza alla Wald al livello 95% e il  $p$ -value corrispondente.

Per quanto riguarda l'interpretazione del modello finale, si precisa che la situazione di riferimento, ovvero quando il predittore lineare è pari alla sola intercetta, è: uomo italiano disoccupato, con livello di istruzione primaria, con reddito percepito basso, che arriva a fine mese con difficoltà.

Per quanto concerne la significatività dei coefficienti stimati, la statistica test di Wald in Eq. (3.7) non mostra evidenza di un effetto dell'istruzione secondaria rispetto all'istruzione primaria e di un effetto di un reddito medio rispetto ad un reddito basso ( $p$ -value > 0.05). Inoltre, non si riscontra evidenza di un effetto sull'aver avuto un disturbo psicologico, rispetto all'Italia, per i seguenti paesi ( $p$ -value > 0.05): Austria, Belgio, Repubblica Ceca, Finlandia, Irlanda, Lettonia, Romania e Slovacchia.

In generale, si osserva che, genere, livello d'istruzione conseguito, età, reddito percepito soggettivamente, occupazione lavorativa e paese, influenzano significativamente

TABELLA 3.1: Stima dei rapporti delle quote con intervalli di confidenza (I.C) per il modello di regressione logistica.

Parametro	Stima <i>odds ratio</i>	I.C.
(Intercetta)	0.81	(0.62, 1.05)
Femmina	1.39	(1.30, 1.49)
Istruzione secondaria	0.92	(0.80, 1.05)
Istruzione terziaria	0.80	(0.69, 0.92)
Età	0.99	(0.98, 0.99)
Reddito medio	1.04	(0.95, 1.14)
Reddito alto	1.15	(1.02, 1.28)
Cavandosela	0.55	(0.50, 0.60)
Confortevolmente	0.42	(0.38, 0.46)
Inoccupato	0.75	(0.65, 0.86)
Occupato a tempo parziale	0.75	(0.63, 0.88)
Occupato a tempo pieno	0.62	(0.54, 0.71)
Austria	1.02	(0.79, 1.31)
Belgio	1.04	(0.81, 1.33)
Bulgaria	1.97	(1.57, 2.48)
Canada	1.87	(1.47, 2.37)
Croazia	0.69	(0.53, 0.91)
Cipro	1.49	(1.17, 1.89)
Repubblica Ceca	0.94	(0.73, 1.22)
Danimarca	1.38	(1.07, 1.76)
Estonia	1.36	(1.07, 1.73)
Finlandia	1.11	(0.86, 1.44)
Francia	2.13	(1.69, 2.68)
Germania	1.35	(1.06, 1.73)
Grecia	1.90	(1.51, 2.40)
Ungheria	0.66	(0.51, 0.87)
Irlanda	1.22	(0.95, 1.56)
Lettonia	1.11	(0.87, 1.41)
Lituania	0.48	(0.35, 0.65)
Paesi Bassi	1.81	(1.42, 2.31)
Polonia	1.32	(1.03, 1.69)
Portogallo	1.75	(1.38, 2.21)
Romania	0.90	(0.70, 1.17)
Slovacchia	0.98	(0.76, 1.26)
Slovenia	0.68	(0.51, 0.90)
Spagna	1.39	(1.10, 1.76)
Svezia	1.72	(1.34, 2.19)
Regno Unito	2.48	(1.97, 3.13)
Stati Uniti	1.39	(1.08, 1.79)

l'aver sofferto di ansia o depressione. Tutti gli *odds ratio* stimati evidenziano un'associazione, più o meno forte, con ciascuno dei fattori analizzati. In particolare, la quota di

TABELLA 3.2: Stime dei parametri per il modello di regressione logistica, con intervalli di confidenza alla Wald al 95% e significatività osservata ( $p$ -value).

Parametro	Stima	I.C. alla Wald al 95%	p_value
(Intercetta)	-0.21	(-0.48, 0.05)	0.11
Femmina	0.33	(0.27, 0.40)	0.00
Istruzione secondaria	-0.09	(-0.22, 0.05)	0.22
Istruzione terziaria	-0.23	(-0.37, -0.08)	0.00
Età	-0.01	(-0.02, -0.01)	0.00
Cavandosela	-0.60	(-0.69, -0.51)	0.00
Confortevolmente	-0.87	(-0.97, -0.78)	0.00
Inoccupato	-0.29	(-0.44, -0.15)	0.00
Occupato a tempo parziale	-0.29	(-0.45, -0.13)	0.00
Occupato a tempo pieno	-0.48	(-0.62, -0.34)	0.00
Austria	0.02	(-0.24, 0.27)	0.91
Belgio	0.04	(-0.21, 0.29)	0.76
Bulgaria	0.68	(0.45, 0.91)	0.00
Canada	0.62	(0.40, 0.88)	0.00
Croazia	-0.37	(-0.64, -0.10)	0.01
Cipro	0.40	(0.16, 0.64)	0.00
Repubblica Ceca	-0.06	(-0.32, 0.20)	0.65
Danimarca	0.32	(0.07, 0.57)	0.01
Estonia	0.31	(0.07, 0.55)	0.01
Finlandia	0.11	(-0.15, 0.36)	0.41
Francia	0.75	(0.53, 0.99)	0.00
Germania	0.30	(0.06, 0.55)	0.02
Grecia	0.64	(0.41, 0.88)	0.00
Ungheria	-0.41	(-0.68, -0.14)	0.00
Irlanda	0.20	(-0.05, 0.45)	0.11
Lettonia	0.10	(-0.14, 0.35)	0.42
Lituania	-0.74	(-1.06, -0.44)	0.00
Paesi Bassi	0.59	(0.35, 0.84)	0.00
Polonia	0.28	(0.03, 0.52)	0.03
Portogallo	0.56	(0.32, 0.79)	0.00
Romania	-0.10	(-0.36, 0.15)	0.44
Slovacchia	-0.02	(-0.28, 0.23)	0.85
Slovenia	-0.38	(-0.67, -0.10)	0.01
Spagna	0.33	(0.09, 0.57)	0.01
Svezia	0.54	(0.30, 0.78)	0.00
Regno Unito	0.91	(0.68, 1.14)	0.00
Stati Uniti	0.33	(0.08, 0.58)	0.01

aver sofferto di ansia o depressione: aumenta del 39% per le donne rispetto agli uomini; diminuisce del 20% per chi ha ottenuto almeno una laurea universitaria, rispetto a chi ha conseguito un livello d'istruzione elementare; si riduce del 45% per chi non ha difficoltà

economiche, rispetto a chi percepisce il proprio reddito come insufficiente; diminuisce, del 25% per un inoccupato o un lavoratore a tempo parziale, del 48% per un lavoratore a tempo pieno, rispetto a un disoccupato.

La quota stimata di aver sofferto di ansia o depressione, fermo restando il valore delle altre variabili esplicative, è inferiore rispetto all'Italia per i seguenti paesi: Croazia, Slovenia, Ungheria, Lituania. La quota stimata è invece maggiore per i seguenti paesi: Polonia, Germania, Estonia, Danimarca, Spagna, Stati Uniti, Cipro, Svezia, Portogallo, Paesi Bassi, Canada, Grecia, Bulgaria, Francia, Regno Unito.

Il genere e la condizione educativa sembrano quindi giocare un ruolo importante per il benessere mentale. In particolare, la propensione ad aver avuto un disturbo ansioso-depressivo aumenta se il soggetto è di sesso femminile e diminuisce se il livello d'istruzione conseguito è quello terziario.

Contrariamente alle aspettative, all'aumentare dell'età, la probabilità di aver sofferto di ansia o depressione si riduce. Ciò sembra contro-intuitivo, in quanto una persona più adulta ha avuto concretamente più tempo per sperimentare un disordine mentale. Una possibile spiegazione potrebbe essere che i meno giovani siano semplicemente meno propensi a dichiarare di essere stati ansiosi o depressi, il che non significa che non lo siano mai stati.

Se un soggetto arriva a fine mese cavandosela è meno probabile che abbia avuto un disturbo psicologico, ancor meno probabile se il soggetto ritiene di non avere difficoltà economiche.

La probabilità di essere stato ansioso o depresso, si riduce in ugual misura per un inoccupato e per un lavoratore a tempo parziale, rispetto a un disoccupato, e si riduce maggiormente per un lavoratore a tempo pieno.

Riassumendo, questi risultati sottolineano che, per migliorare la salute mentale della collettività, è importante investire in interventi a favore dei soggetti più a rischio, ovvero i giovani, le donne, i meno istruiti, i disoccupati, coloro che ritengono di non avere una disponibilità economica sufficiente per affrontare le spese mensili e hanno quindi un basso reddito. I programmi di promozione della salute mentale futuri, dovrebbero considerare strategie che abbiano come *target* i gruppi con queste caratteristiche.

### **3.6 Confronto con i risultati in letteratura**

In questa sezione si intende confrontare i risultati presentati nella sezione precedente con quelli di alcune ricerche in letteratura. L'obiettivo è quello di comprendere in quale

direzione si orientano i risultati ottenuti tramite l'adattamento del modello di regressione logistica.

In generale, gli studi presenti in letteratura mostrano che le donne hanno più probabilità di soffrire di ansia o depressione rispetto agli uomini (Ghirini & Vichi, 2018). Inoltre, i disturbi ansioso-depressivi si associano a condizioni di svantaggio sociale ed economico: coloro che hanno un basso livello di istruzione o un basso reddito riferiscono più spesso problemi d'ansia o depressione (Ghirini & Vichi, 2018). In aggiunta, le ricerche confermano che i disoccupati hanno più probabilità di sperimentare uno stato mentale patologico rispetto ai coetanei occupati (Ghirini & Vichi, 2018).

In sintesi, l'evidenza che emerge dagli studi passati conferma che l'esclusione sociale e la deprivazione materiale sono i più importanti fattori che determinano una cattiva salute mentale (Dreger et al., 2016).

Lehtinen et al. (2005) hanno analizzato il benessere mentale, valutato tramite l'Indice di energia e vitalità, in 11 paesi europei, con i dati dell'Eurobarometro 2002. In generale, è stata individuata un'associazione statisticamente significativa con il genere, il reddito familiare, la situazione occupazionale e l'età. Si è riscontrata una peggiore salute mentale per le donne, per i pensionati e per coloro che appartengono al quartile più basso della distribuzione del reddito nazionale. Inoltre, si è affermato che il benessere mentale diminuisce con l'età.

Schütte et al. (2014) hanno analizzato i dati dell'Indagine europea sulla qualità della vita (EQLS) del 2007. Questi ricercatori hanno esplorato le differenze in termini di benessere psicologico, misurato tramite l'indice di salute mentale WHO-5, tra individui con livelli di istruzione differenti, in 31 paesi europei.

Il WHO-5 misura alcune dimensioni principali del benessere psicologico, quali ansia, depressione e perdita di controllo comportamentale ed emozionale. Tale indice varia tra 0 e 100: a valori crescenti corrisponde un migliore stato di salute mentale, punteggi inferiori a 50 indicano un rischio di depressione.

I risultati ottenuti da Schütte et al. (2014) mostrano che in quasi tutte le nazioni considerate nell'analisi, la prevalenza di una cattiva salute mentale risulta più elevata nei gruppi con livello educativo più basso.

Dreger et al. (2016) hanno studiato i dati del 2012 dell'*European Quality of Life Survey* per 26 paesi europei. I risultati ottenuti mostrano una prevalenza di un buono stato di salute mentale inferiore tra le donne rispetto agli uomini, con  $OR = 0.76$  e intervallo di confidenza al 95% (0.71, 0.81). Inoltre, è emerso che gli individui di sesso femminile sono più propensi a riportare sentimenti di tristezza, ansia, depressione, perdita di fiducia in se stessi, incapacità di superare ostacoli e problemi quotidiani

(Dreger et al., 2016).

Il *Gender Equality Index 2021* mostra che, tra gli individui maggiorenni in Unione Europea, il WHO-5 è più alto per gli uomini rispetto alle donne: il punteggio medio è pari a 66 per i maschi e 62 per le femmine (EIGE, 2021). Inoltre, uomini e donne con un livello di reddito superiore presentano un maggiore benessere psicologico. Il livello di salute mentale varia tra i paesi europei: è più alto in Irlanda e Danimarca, con indice WHO-5 pari a 70 punti e più basso in Croazia e Italia, con circa 60 punti (EIGE, 2021). Si riscontra anche una differenza in media di circa 7 punti tra coloro che hanno un livello educativo più basso e coloro che hanno conseguito un livello di istruzione più elevato. Infine, si evidenzia che gli occupati godono di un benessere mentale superiore di almeno 6 punti rispetto ai disoccupati.

In conclusione, i risultati ottenuti analizzando i dati del *Global Monitor: Mental health*, tramite un modello di regressione logistica, sono coerenti con quelli presenti in letteratura. Un solo risultato sembra però andare in direzione opposta: contrariamente a quanto si riscontra in letteratura, la probabilità di aver sofferto di ansia o depressione decresce con l'età. Questo risultato può essere però interpretato in termini di propensione a dichiarare di aver convissuto con un problema di salute mentale.

### 3.7 Modello di regressione logistica con intercetta casuale

I modelli misti lineari generalizzati (*Generalized Linear Mixed Model* - GLMM) rappresentano un'estensione dei GLM e consentono di inserire nel predittore lineare sia effetti casuali che effetti fissi (Agresti, 2013). I GLMM si utilizzano, per esempio, quando i dati sono organizzati su più livelli, quando si vuole gestire una variabile qualitativa con tante modalità o in presenza di misure ripetute. L'inserimento di un effetto casuale nel modello permette di tenere conto della correlazione positiva tra le osservazioni all'interno di uno stesso gruppo.

Come principale modello alternativo al modello di regressione logistica, presentato nelle sezioni precedenti, si è adattato, quindi, un modello di regressione logistica a effetti misti con intercetta casuale per paese. I dati analizzati presentano, infatti, una struttura multilivello (individui al primo livello e stati al secondo) e la variabile Paese presenta 28 categorie.

Sia  $Y_{pi}$  la variabile risposta per il soggetto  $i$  nel paese  $p$ , con  $Y_{pi} = 1$  se il soggetto  $i$  ha affermato di essere stato ansioso o depresso;  $Y_{pi} = 0$  altrimenti. Sia  $u_p$  l'effetto casuale

per il gruppo  $p$ , con  $p \in \{1, \dots, P\}$ ,  $P = 28$ . Sia  $x_{pi}$  il vettore dei valori osservati delle covariate per l' $i$ -esima osservazione, con  $i \in \{1, \dots, n\}$ ,  $n = 27475$ . Siano  $\beta$  il vettore dei parametri stimati e  $\beta_j$  il coefficiente associato alla  $j$ -esima variabile esplicativa, con  $j \in \{1, \dots, J\}$ .

Il modello assume che

$$\text{logit}[Pr(Y_{pi} = 1|u_p)] = x_{pi}\beta + u_p, \quad (3.14)$$

con  $u_p \sim N(0, \sigma^2)$ ,  $u_p$  indipendenti e  $\sigma$  sconosciuta. Quando  $\sigma = 0$ , il modello (3.14) si semplifica in un modello di regressione logistica ordinario, dove tutte le osservazioni si considerano come indipendenti l'una dall'altra.

Per valutare la complessità del modello, è stata utilizzata una procedura passo a passo all'indietro, basata sul criterio dell'AIC, a partire dal modello con tutte le interazioni del secondo ordine. Il modello finale include gli effetti fissi di Genere, Età, Reddito percepito, Livello di istruzione e Occupazione lavorativa.

Per quanto riguarda la bontà di adattamento del modello a intercetta casuale, il test di Hosmer e Lemeshow, applicato con  $g = 10$  gruppi, non mostra evidenza di mancanza di adattamento ( $p$ -value = 0.44). L'accuratezza è pari a 0.82, la specificità è pari a 0.98 e la precisione è pari a 0.79. I valori di queste metriche coincidono con quelli calcolati per il modello di regressione logistica. Inoltre, l'AIC per il modello con intercetta casuale è pari a 24525, mentre quello per il modello di regressione logistica è pari a 24465.79. Si può concludere, quindi, che non è presente un'evidente differenza in termini di bontà di adattamento tra i due modelli.

Per quanto concerne l'interpretazione dei risultati ottenuti tramite il modello con intercetta casuale, la stima dell'effetto casuale  $u_p$  è pari a 0.1382.  $\hat{u}_p$  rappresenta la variabilità stimata nell'intercetta in scala logit. L'effetto casuale aggiusta l'intercetta, ma non modifica gli effetti fissi. La stima della deviazione standard dell'intercetta casuale riassume il grado di eterogeneità della popolazione analizzata ed è pari a 0.3718.

Per quanto riguarda l'inferenza sugli effetti fissi, in Tabella 3.3 si riportano le stime dei coefficienti, con intervalli di confidenza al livello 95% e significatività osservata, mentre in Tabella 3.4 si riportano gli *odds ratio* e i corrispondenti intervalli di confidenza, calcolati esponenziando gli estremi degli I.C. per le stime dei parametri. Si può concludere che anche le stime dei parametri e dei rapporti delle quote per i due modelli coincidono quasi perfettamente.

TABELLA 3.3: Stime dei parametri per il modello di regressione logistica con intercetta casuale, con intervalli di confidenza alla Wald al 95% e significatività osservata ( $p$ -value).

Parametro	Stima	I.C. alla Wald	$p$ -value
Intercetta	-0.67	(-0.91, -0.44)	0.00
Femmina	0.33	(0.26, 0.39)	0.00
Istruzione secondaria	-0.09	(-0.22, 0.05)	0.22
Istruzione terziaria	-0.22	(-0.37, -0.08)	0.00
Età	-0.23	(-0.27, -0.20)	0.00
Cavandosela	-0.60	(-0.69, -0.51)	0.00
Confortevolmente	-0.87	(-0.96, -0.78)	0.00
Inoccupato	-0.29	(-0.44, -0.15)	0.00
Occupato a tempo parziale	-0.29	(-0.45, -0.13)	0.00
Occupato a tempo pieno	-0.48	(-0.62, -0.35)	0.00

TABELLA 3.4: Stima dei rapporti delle quote con intervalli di confidenza (I.C.) per il modello di regressione logistica con intercetta casuale.

Parametro	Stima <i>odds ratio</i>	I.C.
Intercetta	0.51	(0.40, 0.64)
Femmina	1.39	(1.30, 1.48)
Istruzione secondaria	0.92	(0.80, 1.05)
Istruzione terziaria	0.80	(0.69, 0.92)
Età	0.79	(0.76, 0.82)
Cavandosela	0.55	(0.50, 0.60)
Confortevolmente	0.42	(0.38, 0.46)
Inoccupato	0.74	(0.64, 0.86)
Occupato a tempo parziale	0.75	(0.63, 0.88)
Occupato a tempo pieno	0.62	(0.54, 0.71)



# Capitolo 4

## Modello di regressione a classi latenti

L'analisi delle classi latenti (*Latent Class Analysis - LCA*), nota anche come analisi della struttura latente, è una tecnica statistica per l'identificazione e la caratterizzazione di gruppi tra le osservazioni. Si tratta di un approccio che permette di cogliere l'eterogeneità tra gruppi e nei gruppi. La LCA si utilizza frequentemente quando le variabili osservate, dette anche “manifeste”, rappresentano molteplici indicatori di uno stesso concetto latente e si presentano come una serie di variabili risposta categoriali.

Questa tecnica trova un'ampia applicazione in molti ambiti differenti, quali la ricerca clinica, il marketing, le scienze sociali e biomediche (Van Kollenburg et al., 2015). Inoltre, è diventata sempre più popolare tra i ricercatori in psicologia, per studiare, per esempio, la struttura dei disordini psichiatrici.

Per esempio, Petersen et al. (2019) hanno applicato l'analisi delle classi latenti per individuare profili di bambini con sintomi di salute mentale comuni. Hagenaars (1998) ha spiegato cinque indicatori dicotomici relativi alla vita politica, ovvero approvazione di proteste, livello ideologico, partecipazione politica, attenzione ai bisogni e approvazione della repressione, attraverso due profili latenti (classe liberale e classe conservatrice). De France & Hollenstein (2021) hanno individuato tre profili di adolescenti che differiscono uno dall'altro per l'uso di sei diverse strategie di regolazione delle emozioni, mostrando che il tipo di emozione e la sua intensità sono associati con la strategia utilizzata.

Il modello base, detto modello a classi latenti, considera una tabella di contingenza osservata come una mistura finita di tabelle non osservate, generate sotto una struttura di indipendenza condizionata alle categorie di una variabile latente. Ogni soggetto viene

quindi classificato in uno specifico profilo in base ad una certa probabilità; condizionatamente alla classe latente, si hanno quindi delle aspettative su come quell'individuo risponderà ad ogni variabile in modo indipendente dalle altre. Le osservazioni con combinazioni simili dei valori delle variabili risposta tenderanno a far parte dello stesso gruppo. Il modello più generale, detto modello di regressione a classi latenti (*Latent Class Regression* - LCR), permette, inoltre, di stimare gli effetti delle covariate nel prevedere la probabilità di appartenenza di un'osservazione a ciascuna classe latente.

Per i dati analizzati in questa tesi, sono presenti otto variabili dicotomiche manifeste: quest'ultime rappresentano otto differenti approcci per la gestione di ansia e depressione, che 4884 soggetti hanno sperimentato o meno quando erano ansiosi o depressi. L'obiettivo è quello di modellare l'accordo tra questi metodi, esaminando come gli individui possono essere divisi in gruppi a seconda della combinazione di azioni scelte.

Per descrivere i comportamenti delle persone ansiose o depresse, si assume quindi che le persone affette da un disturbo mentale possano essere divise in  $Q$  profili ideali, i quali esprimono le differenti attitudini rispetto ad un malessere psicologico. Questi profili ideali corrispondono a diverse scelte comportamentali: per esempio, alcuni hanno provato tanti metodi per ristabilire il proprio benessere mentale; altri hanno sperimentato una specifica combinazione di approcci curativi.

## 4.1 Specificazione del modello

### 4.1.1 Modello a classi latenti

Il modello a classi latenti più semplice assume l'esistenza di una variabile latente categoriale, tale che le variabili risposta osservate siano condizionatamente indipendenti, data quella variabile (Lewis & Linzer, 2011). Il modello LC assume inoltre che  $n$  osservazioni possano essere partizionate in  $Q$  profili latenti, i quali formano le categorie della variabile latente discreta (Van Kollenburg et al., 2015). All'interno di ogni gruppo, si assume quindi che le risposte alle variabili osservate siano indipendenti l'una dall'altra (assunzione di indipendenza locale).

Siano  $Y_j$ , con  $j \in \{1, \dots, J\}$ , le variabili risposta categoriali osservate che rappresentano gli otto differenti approcci per affrontare ansia e depressione, ciascuna con due possibili esiti, per ogni soggetto. Sia quindi  $K = 2$  il numero di categorie per ogni  $Y_j$ .

Sia  $Y_j = 1$  se il soggetto  $i$  ha sperimentato l'approccio  $j$  e  $Y_j = 0$  altrimenti, dove  $i \in \{1, \dots, n\}$ , con  $n$  pari al totale delle osservazioni nel campione su cui viene stimato il modello.

La variabili osservate sono:

- $Y_1$  = Aver parlato con un professionista della salute mentale;
- $Y_2$  = Aver partecipato ad attività religiose o spirituali;
- $Y_3$  = Aver cercato il supporto di amici o familiari;
- $Y_4$  = Aver assunto specifici medicinali sotto prescrizione medica;
- $Y_5$  = Aver migliorato il proprio stile di vita con comportamenti più sani;
- $Y_6$  = Aver apportato un cambiamento nella propria situazione lavorativa;
- $Y_7$  = Aver apportato un cambiamento nella propria vita relazionale;
- $Y_8$  = Aver trascorso tempo all'aperto o in mezzo alla natura.

Sia  $Z$  la variabile latente discreta, con  $Z_i \in \{1, \dots, Q\}$ , con  $Q$  numero totale di classi latenti, tale che per ogni possibile combinazione degli esiti delle variabili risposta e per ogni categoria  $q$  di  $Z$ ,

$$Pr(Y_1 = y_1, \dots, Y_8 = y_8 | Z = q) = Pr(Y_1 = y_1 | Z = q) \cdots Pr(Y_8 = y_8 | Z = q). \quad (4.1)$$

I parametri stimati dal modello sono le probabilità di classificazione nelle classi latenti  $Pr(Z = q)$ , ovvero la proporzione di osservazioni appartenenti a ciascun gruppo, e le probabilità condizionate degli esiti per ogni  $Y_j$  in ogni classe latente  $q$ , ovvero  $Pr(Y_j = y_j | Z = q)$ . Il numero di parametri stimati è  $Q + (Q - 1)$ .

La tabella di contingenza a otto entrate ( $Y_1, \dots, Y_8$ ) è osservata, mentre quella a nove entrate, che considera anche la variabile latente, non è osservata (Agresti, 2013). Per la tabella osservata, il modello assume una distribuzione multinomiale per le  $K^8 = 2^8 = 256$  celle, ovvero le possibili combinazioni dei livelli delle otto variabili dicotomiche. La probabilità di ogni cella è

$$\pi_{y_1, \dots, y_8} = \sum_{q=1}^Q Pr(Y_1 = y_1, \dots, Y_8 = y_8 | Z = q) Pr(Z = q). \quad (4.2)$$

Questo modello implica che il modello loglineare indicato con  $(Y_1Z, \dots, Y_8Z)$  sia valido per la tabella di contingenza non osservata e assume che le  $Y_j$  siano mutualmente indipendenti in ogni categoria di  $Z$ . Il modello con  $q = 1$  classi latenti è equivalente al modello loglineare di mutua indipendenza  $(Y_1, Y_2, \dots, Y_8)$ , ovvero con i soli effetti principali.

Sia  $\pi_{jqk}$  la probabilità condizionata che  $Y_j = k$ ,  $k \in \{0, 1\}$ , per un'osservazione nella classe  $q$ , dove  $\sum_{k=1}^2 \pi_{jqk} = 1$ . Sia  $y_{ijk} = 1$  se  $Y_j = k$ ;  $y_{ijk} = 0$  altrimenti. Sia  $p_q$  la probabilità di appartenenza alla  $q$ -esima classe, con  $\sum_{q=1}^Q p_q = 1$ .

Assumendo indipendenza tra le  $Y_j$  entro ogni gruppo  $q$ , la probabilità di osservare una specifica combinazione degli esiti delle  $J$  variabili manifeste, per un individuo  $i$  nella classe  $q$ , è

$$f(Y_i; \pi_q) = \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}}. \quad (4.3)$$

La funzione di probabilità congiunta è

$$Pr(Y_i | \pi, p) = \sum_{q=1}^Q p_q \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}}. \quad (4.4)$$

La probabilità di osservare una specifica combinazione è quindi una mistura di distribuzioni multinomiali con pesi pari a  $p_q$  (Van Kollenburg et al., 2015).

La funzione di log-verosimiglianza è

$$\log L = \sum_{i=1}^N \log \sum_{q=1}^Q p_q \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}}. \quad (4.5)$$

La probabilità a posteriori che un soggetto appartenga a ciascuna classe, condizionatamente ai valori osservati delle variabili manifeste, usando la formula di Bayes, è

$$\widehat{Pr}(q | Y_i) = \frac{\widehat{p}_q f(Y_i; \widehat{\pi}_q)}{\sum_{q=1}^Q \widehat{p}_q f(Y_i; \widehat{\pi}_q)}. \quad (4.6)$$

### 4.1.2 Modello a classi latenti con inclusione di covariate

Il modello a classi latenti più generale viene denominato “modello di regressione a classi latenti” e permette di includere una o più covariate per prevedere l'appartenenza di un soggetto a ciascuna classe latente (Lewis & Linzer, 2011).

A differenza del modello più semplice precedentemente presentato, le probabilità di appartenere ad una specifica classe latente non sono uguali per tutte le osservazioni, ma variano da individuo a individuo, in funzione del valore delle variabili concomitanti. L'inclusione delle covariate avviene attraverso gli effetti di tali covariate sulle probabilità  $p_q$ .

Sia  $p_{qi}$  la probabilità di appartenenza alla classe  $q$  per l'individuo  $i$ , con  $\sum_{q=1}^Q p_{qi} = 1$  per ogni individuo. Siano  $x_i$  il vettore osservato delle variabili esplicative per l' $i$ -esimo individuo e  $\beta_q$  il vettore di coefficienti corrispondenti alla  $q$ -esima classe. Dato che in questo caso il primo gruppo è fissato come riferimento,  $\beta_1 = 0$ .

Il modello LCR assume che

$$\begin{aligned} \log(p_{2i}/p_{1i}) &= x_i\beta_2 \\ \log(p_{3i}/p_{1i}) &= x_i\beta_3 \\ &\vdots \\ \log(p_{Qi}/p_{1i}) &= x_i\beta_Q. \end{aligned} \tag{4.7}$$

Questi log-rapporti possono essere convertiti in probabilità con la formula

$$p_{qi} = p_q(x_i; \beta) = Pr(Z_i = q|x_i) = \frac{\exp(x_i\beta_q)}{\sum_{q=1}^Q \exp(x_i\beta_q)}. \tag{4.8}$$

Il contributo alla verosimiglianza per il soggetto  $i$  è

$$\sum_{q=1}^Q p_q(x_i; \beta) \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}}. \tag{4.9}$$

La funzione di log-verosimiglianza è

$$\log L = \sum_{i=1}^N \log \sum_{q=1}^Q p_q(x_i; \beta) \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}}. \tag{4.10}$$

Il numero di parametri stimati è  $Q + (S+1)(Q-1)$ , con  $S$  numero totale di covariate inserite nel modello.

Le probabilità a posteriori di appartenenza a ciascuna classe latente sono

$$\widehat{Pr}(q|Y_i) = \frac{\widehat{p}_q(x_i; \widehat{\beta}) f(Y_i; \widehat{\pi}_q)}{\sum_{q=1}^Q \widehat{p}_q(x_i; \widehat{\beta}) f(Y_i; \widehat{\pi}_q)}. \tag{4.11}$$

## 4.2 Stima dei parametri

Per massimizzare la funzione di log-verosimiglianza del modello e calcolare le stime di massima verosimiglianza dei parametri, viene impiegato l'algoritmo *Expectation-Maximization* (EM) (Lewis & Linzer, 2011). La funzione di log-verosimiglianza in Eq. (4.10) viene massimizzata rispetto ai parametri  $p_q$ ,  $\pi_{jqk}$  e  $\beta$ .

Per evitare di incorrere in un massimo locale, invece che globale, soprattutto quando il numero delle classi latenti aumenta, si fa ripartire l'algoritmo più volte con valori dei parametri iniziali differenti, scelti casualmente.

### 4.3 Selezione del modello

In questa sezione si presentano alcuni strumenti utili per il confronto tra modelli e si precisa quale procedura è stata seguita per selezionare il modello finale.

Per determinare un numero appropriato di classi latenti, si può procedere prendendo in considerazione motivazioni di carattere teorico, nel caso in cui ve ne siano, oppure scegliendo il modello migliore in termini di bontà di adattamento.

L'approccio più comune consiste nel partire da un modello con  $Q = 1$  e successivamente incrementare  $Q$  di un'unità alla volta, fino a che non si identifica un buon modello. All'aumentare del numero di classi latenti, l'adattamento del modello tende a migliorare, incontrando però il rischio di cadere nel sovra-adattamento e stimando un maggior numero di parametri.

I due criteri più utilizzati per il confronto tra modelli sono il *Bayesian information criterion* (BIC), anche detto criterio di Schwarz, e l'*Akaike information criterion* (AIC). BIC e AIC penalizzano la funzione di log-verosimiglianza con una funzione del numero di parametri che devono essere stimati. Rispetto ad AIC, BIC implica una più elevata penalità per i modelli sovra-parametrizzati. Il modello migliore è quello che minimizza l'uno, o l'altro criterio, o entrambi. Il criterio di Schwarz è così definito:

$$BIC = -2 \log L + k \log n, \quad (4.12)$$

con  $L$  valore della funzione di verosimiglianza massimizzata e  $k$  numero di parametri stimati.

Si è specificato quindi un modello di regressione a classi latenti. Il numero di unità statistiche è pari a 4884, ovvero il numero di soggetti nel campione che hanno affermato di aver sofferto d'ansia o depressione. Le possibili covariate sono le seguenti: Genere, Livello di istruzione conseguito, Età, Reddito, Occupazione lavorativa, Arrivare a fine mese e Paese. Con una procedura di selezione dei predittori passo a passo in avanti basata sul criterio BIC, partendo dal modello senza covariate, condizionatamente ad un fissato numero di classi latenti, sono stati ottenuti sei modelli.

Nella Tabella 4.1, per ciascun modello, si riportano il valore del criterio d'informazione BIC e le covariate incluse. Confrontando tali modelli, sempre utilizzando il criterio

BIC, il miglior modello risulta essere quello con cinque classi latenti che include come unica covariata soltanto l'età. Si può notare, infatti, che all'aumentare del numero di classi latenti, BIC diminuisce fino a 46136.75 con cinque classi e aumenta poi a 46151.32 con sei classi.

La differenza tra il BIC del modello a cinque classi (46136.75) e il BIC del modello a quattro classi (46148.72) è però minima. Inoltre, entrambi i modelli includono la stessa covariata (l'età). In aggiunta, da una prima analisi dei due modelli, l'interpretazione dei profili latenti risulta più chiara per il modello con quattro classi. Per questi motivi, si è selezionato quest'ultimo modello come più appropriato.

TABELLA 4.1: BIC e covariate incluse nei sei migliori modelli a classi latenti.

Numero classi	BIC	Covariate selezionate
1	47741.21	Nessuna
2	46636.28	Paese, Età, Genere, Istruzione
3	46167.55	Età, Paese, Genere
4	46148.72	Età
5	46136.75	Età
6	46151.32	Età, Genere

## 4.4 Bontà di adattamento

La bontà di adattamento di un modello a classi latenti si valuta prettamente usando le statistiche  $\chi^2$  e  $G^2$  (Van Kollenburg et al., 2015). Tali statistiche permettono di valutare il disaccordo tra le frequenze osservate  $n_c$  e le frequenze attese sotto il modello  $e_c$ , con  $C$  numero di combinazioni possibili delle modalità delle  $Y_j$ . Un modo per determinare l'eventuale mancanza di adattamento è quello di confrontare il valore della statistica test con la sua distribuzione sotto l'ipotesi nulla di uguaglianza tra le frequenze attese e osservate.

La statistica  $\chi^2$  si definisce come

$$\chi^2 = \sum_{c=1}^C \frac{(n_c - e_c)^2}{e_c}. \quad (4.13)$$

La statistica  $G^2$  si definisce come

$$G^2 = 2 \sum_{c=1}^C n_c \left[ \ln \left( \frac{n_c}{e_c} \right) \right]. \quad (4.14)$$

Queste due statistiche non sono però appropriate quando la tabella di contingenza analizzata è sparsa, ovvero quando la numerosità campionaria non è sufficientemente elevata, il numero delle celle nella tabella è grande, una porzione delle frequenze attese è piccola, o sono presenti tante celle vuote. In questi casi, le assunzioni per le distribuzioni asintotiche di queste statistiche non reggono: infatti, la statistica  $X^2$  tende ad essere molto grande, comportando un  $p$ -value pari a 0, mentre la statistica  $G^2$  tende ad essere piccola, implicando un  $p$ -value pari a 1 (Van Kollenburg et al., 2015).

Van Kollenburg et al. (2015) propongono tre statistiche utili in caso di tabelle di contingenza sparse: l'Indice di Dissimilarità, i Residui Bivariati e i Residui Bivariati Totali.

L'Indice di Dissimilarità (DI) indica quanto differiscono le frequenze stimate da quelle osservate. Il valore di questa statistica è pari alla proporzione di osservazioni che dovrebbe essere spostate in un'altra cella per ottenere un adattamento perfetto. Tale indice si definisce come

$$DI = \frac{\sum_{c=1}^C |n_c - e_c|}{2n}. \quad (4.15)$$

I Residui Bivariati (BVR) possono essere usati per determinare violazioni dell'assunzione di indipendenza locale, quantificando l'associazione residua tra coppie di variabili osservate. Siano  $r$  la risposta alla variabile  $j$  e  $r'$  la risposta alla variabile  $j'$ . Siano  $n_{rr'}$  e  $e_{rr'}$  rispettivamente la frequenza osservata e la frequenza attesa nella tabella di contingenza a due entrate per le variabili  $j$  e  $j'$ . Sia  $R_j$  il numero di modalità della variabile  $j$ .

La statistica BVR si definisce come

$$BVR_{jj'} = \sum_{r=1}^{R_j} \sum_{r'=1}^{R_{j'}} \frac{(n_{rr'} - e_{rr'})^2}{e_{rr'}}. \quad (4.16)$$

Infine, la statistica TBVR si ottiene sommando le statistiche BVR per tutte le coppie di variabili in questo modo:

$$TBVR = \sum_{j=1}^{J-1} \sum_{j'=j+1}^J BVR_{jj'}. \quad (4.17)$$

Un'importante vantaggio delle statistiche BVR e TBVR è che sono molto meno sensibili alla presenza di sparsità rispetto alle altre statistiche citate.

Considerando il modello a classi latenti selezionato, la tabella di contingenza osservata conta 256 celle, di cui circa un terzo presentano conteggi osservati e predetti inferiori

a 5. Sulla base delle considerazioni precedenti, nelle circostanze descritte, le statistiche  $X^2$  e  $G^2$  non risultano quindi adeguate e affidabili, poiché le assunzioni non vengono rispettate e l'approssimazione chi-quadratica risulta povera.

Per valutare la bontà di adattamento del modello finale, sono state quindi calcolate le seguenti statistiche: DI, BVR e TBVR. L'indice di dissimilarità per il modello con quattro classi latenti e l'età come covariata è pari a 0.09: questo significa che solo il 9% delle osservazioni dovrebbe essere spostato in una cella differente per ottenere un perfetto adattamento. Il TBVR è pari a 101.2555.

Nella Tabella 4.2 si riportano i valori dei residui bivariati per ciascuna coppia di variabili osservate. Per alcuni, la regola del pollice prevede che valori superiori a 3.84 indichino cattiva bontà di adattamento; altri ritengono, invece, che siano valori maggiori di 1 a segnalare un cattivo adattamento. La seconda linea guida porta a conclusioni più conservative. Utilizzando il criterio meno conservativo, si nota una violazione dell'assunzione di indipendenza locale per le seguenti coppie di variabili: A-C, C-F, C-H, E-H, F-G.

TABELLA 4.2: Residui Bivariati, (BVR) per tutte le coppie di variabili osservate, per il modello a classi latenti.

	A	B	C	D	E	F	G
B	1.93						
C	4.72	1.49					
D	0.17	0.69	0.46				
E	0.22	2.27	1.67	0.17			
F	0.65	3.10	6.41	0.14	0.19		
G	0.36	0.43	4.28	0.11	1.71	40.21	
H	0.69	2.55	17.28	0.17	4.71	1.62	2.66

TABELLA 4.3: Accuratezza, moltiplicata per 100, per ciascuna variabile, per il modello a classi latenti.

Variabile	Accuratezza
A	71
B	60
C	79
D	99
E	86
F	64
G	66
H	79

Inoltre, per confrontare valori osservati e predetti, è stata calcolata anche l'accuratezza marginalmente per ciascuna variabile osservata. I valori di questa metrica sono riportati nella Tabella 4.3.

Complessivamente, si può affermare che il modello presenta un adattamento soddisfacente.

## 4.5 Descrizione dei profili latenti

In questa sezione, si descrivono i risultati ottenuti con riferimento al modello con quattro classi latenti che include l'età come unica covariata.

Dalla Tabella 4.4, si può notare che le stime delle probabilità di appartenenza a ciascun gruppo, in ordine crescente, sono:  $\hat{p}_1 = 0.14$ ,  $\hat{p}_2 = 0.19$ ,  $\hat{p}_3 = 0.22$  e  $\hat{p}_4 = 0.45$ . Si osserva quindi che la classe 1 e la classe 4 sono rispettivamente la meno e la più rappresentativa della popolazione analizzata.

TABELLA 4.4: Stime dei parametri  $p_q$  con intervalli di confidenza (I.C.) alla Wald al 95%, per il modello a classi latenti.

Classe	$\hat{p}_q$	I.C.
1	0.14	(0.10, 0.18)
2	0.19	(0.15, 0.24)
3	0.22	(0.18, 0.27)
4	0.45	(0.40, 0.50)

Per quanto riguarda la composizione dei quattro profili latenti, la Figura 4.1 mostra le stime, moltiplicate per 100, dei parametri  $p_q$  specifici per ogni gruppo.

TABELLA 4.5: Legenda per la Figura 4.1.

A	Aver parlato con un professionista della salute mentale
B	Aver partecipato ad attività religiose o spirituali
C	Aver parlato con amici o familiari
D	Aver assunto medicinali sotto prescrizione medica
E	Aver introdotto comportamenti più sani
F	Aver cambiato la propria situazione lavorativa
G	Aver modificato la propria situazione relazionale
H	Aver trascorso tempo all'aperto o in mezzo alla natura

Innanzitutto, si può notare che aver chiesto supporto ad amici e familiari (C) e aver trascorso tempo in mezzo alla natura (H) sono metodi comuni a tutti e quattro i profili, con probabilità crescenti nei quattro gruppi. Infatti, la probabilità di aver sperimentato

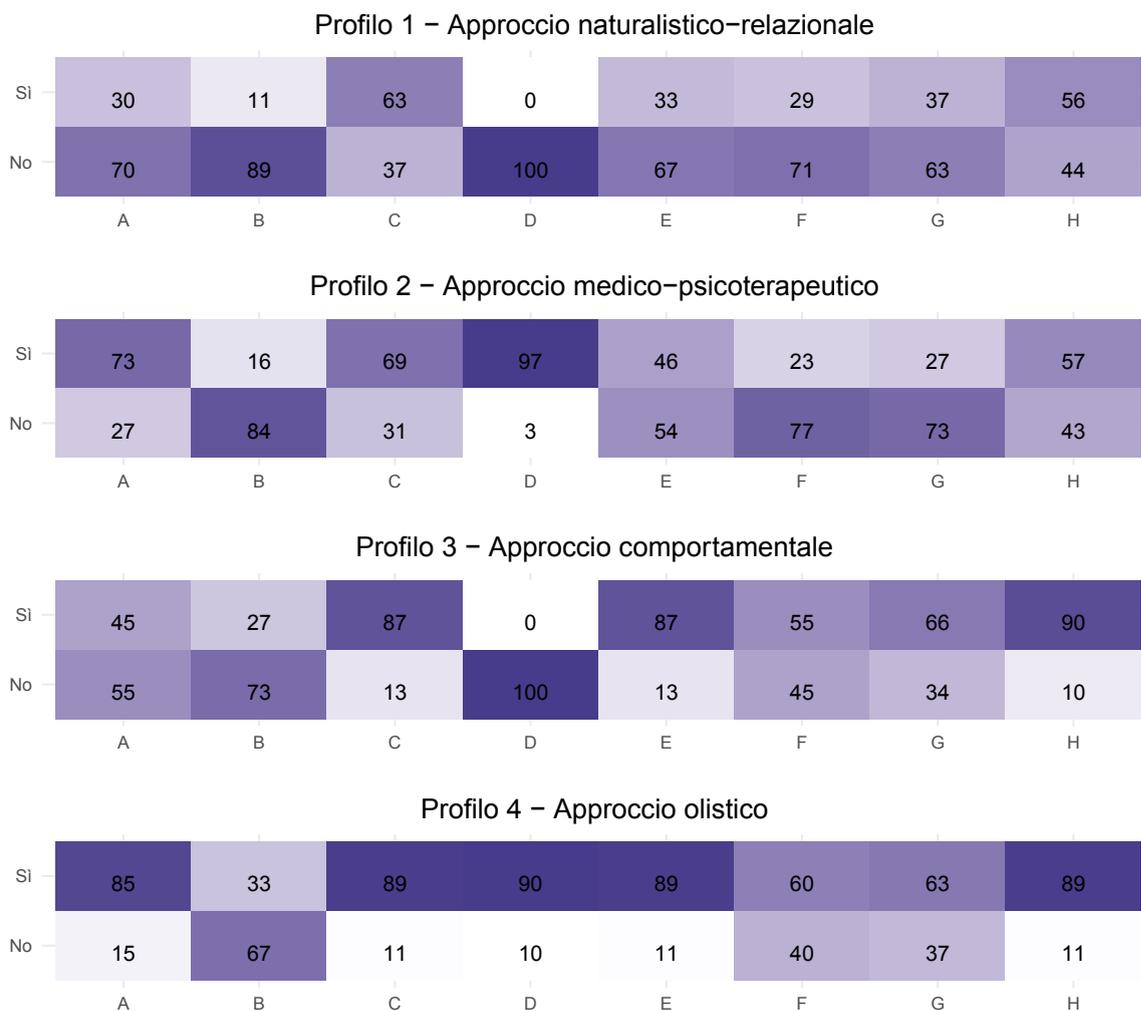


FIGURA 4.1: Stime dei parametri  $\pi_q$  specifici per ciascun profilo, per il modello a classi latenti. L'intensità del colore delle celle varia in accordo con i valori delle probabilità stimate: sfumature più chiare corrispondono a valori più piccoli. Per ogni metodo da A ad H, il numero in ciascuna cella rappresenta la stima delle probabilità condizionata a ciascuna classe, moltiplicata per 100. Nella Tabella 4.5 si riporta una legenda utile per la lettura e l'interpretazione di questa figura.

C passa da 63% nel profilo 1 a 89% nel profilo 4; la probabilità di aver scelto H passa da 56% nel profilo 1 a 90% nel profilo 3. Questo non stupisce, in quanto dalle analisi esplorative si è evidenziato che questi due metodi sono quelli complessivamente più sperimentati. Inoltre, si osserva che i soggetti hanno partecipato ad attività spirituali (B) con una probabilità pari o inferiore al 33% in tutti e quattro i gruppi. Anche questo non sorprende, in quanto le analisi esplorative hanno mostrato che si tratta dell'approccio generalmente meno scelto.

Il primo profilo è composto dagli individui che, quando erano ansiosi o depressi, hanno scelto quasi esclusivamente un approccio naturalistico-relazionale per recuperare

il proprio benessere psicologico. I soggetti in questo gruppo, infatti, hanno chiesto supporto ad amici e familiari (C) con una probabilità del 63% e hanno trascorso tempo all'aperto o in mezzo alla natura (H) con una probabilità del 56%. Con probabilità pari a circa 30% hanno seguito un percorso terapeutico con un professionista della salute mentale (A), hanno introdotto comportamenti più sani (E), hanno cambiato lavoro (F) o hanno modificato la propria situazione relazionale (G). Inoltre, hanno partecipato ad attività religiose o spirituali (B) con probabilità dell' 11% e non hanno mai assunto farmaci sotto prescrizione medica (D). Nel proseguo di questa tesi ci si riferirà a questo gruppo come "naturalistico-relazionale".

Il secondo profilo è composto dagli individui che, quando erano ansiosi o depressi, insieme ad un approccio naturalistico-relazionale, hanno scelto una strategia medico-psicoterapeutica. In particolare, i soggetti in questo gruppo hanno fatto ricorso a farmaci prescritti da un medico (D) con probabilità pari a 97% e si sono rivolti a un professionista della salute mentale (A) con probabilità pari a 73%. In confronto al primo gruppo, una frazione di individui circa uguale tra i due gruppi ha chiesto un aiuto ad amici e familiari (C) e ha trascorso tempo all'aperto (H). Ci si riferirà a questo gruppo come "medico-psicoterapeutico".

Il terzo profilo è composto dagli individui che, quando erano ansiosi o depressi, insieme ad un approccio naturalistico-relazionale, hanno sperimentato una strategia di tipo comportamentale, ovvero hanno modificato o introdotto alcuni atteggiamenti nella propria vita personale, lavorativa e relazionale. Nello specifico, in questo profilo, la probabilità di aver introdotto comportamenti più sani (E) è 87%, in confronto a 33% per l'approccio naturalistico-relazionale e 46% per quello medico-psicoterapeutico. In aggiunta, la probabilità di cambiare lavoro (F) è 55%, in confronto a 29% per il gruppo naturalistico-relazionale e 23% per quello medico-psicoterapeutico; la probabilità di modificare la propria vita relazionale (G) è 66%, in confronto a 37% per la strategia naturalistico-relazionale e 27% per quella medico-psicoterapeutica. Ci si riferirà a questo gruppo come "comportamentale".

Infine, il quarto profilo è caratterizzato da individui che, quando erano ansiosi o depressi, hanno sperimentato un'ampia combinazione di metodi per ristabilire il proprio benessere mentale. La probabilità è circa pari a 90% per aver chiesto il supporto di uno psicologo (A), essersi rivolti ad amici e familiari (C), aver assunto medicinali (D), aver introdotto comportamenti più sani (E) e aver trascorso tempo in mezzo alla natura (H). Inoltre, i soggetti appartenenti a questo gruppo hanno cambiato la propria situazione lavorativa (F) con probabilità pari a 60% e hanno apportato un cambiamento significativo nella propria vita relazionale (G) con probabilità pari a 63%. Ci si riferirà a questo

gruppo come “olistico”, dal greco “tutto”/“totale”, ovvero un gruppo in cui i soggetti tendono ad aver scelto tutti i metodi disponibili.

Riassumendo, i quattro profili latenti possono essere interpretati in termini di quali metodi per la gestione di ansia e depressione sono prevalenti all’interno di uno specifico gruppo e quale strategia contraddistingue un gruppo da tutti gli altri. Alcuni comportamenti sono comuni a tutti e quattro i profili: per esempio, la probabilità di aver chiesto il supporto di amici e familiari e di aver trascorso del tempo all’aperto è superiore al 50% in tutti e quattro i gruppi, anche se in alcuni tra questi la percentuale cresce maggiormente. Invece, la probabilità di aver partecipato ad attività religiose e spirituali è inferiore al 50% in tutte e quattro le classi. Altri comportamenti variano invece attraverso i profili considerati: per esempio, hanno assunto farmaci prescritti da un medico soltanto i soggetti che hanno seguito un approccio medico-psicoterapeutico o un approccio olistico.

## 4.6 Effetto dell’età

Per quanto riguarda l’effetto dell’età sulle probabilità di appartenenza a ciascun profilo latente, nella Tabella 4.6, si riportano le stime dei coefficienti  $\beta$ , tutte significative al livello 0.05.

TABELLA 4.6: Stime e corrispondenti intervalli di confidenza alla Wald al 95% per  $\beta$ , per il modello a classi latenti.

Parametro	Stima	I.C. alla Wald
$\beta_{21}$	-0.98	(-1.56, -0.41)
$\beta_{22}$	0.03	(0.02, 0.04)
$\beta_{31}$	1.40	(0.88, 1.92)
$\beta_{32}$	-0.02	(-0.03, -0.01)
$\beta_{41}$	0.62	(0.19, 1.04)
$\beta_{42}$	0.01	(0.00, 0.02)

Prendendo come riferimento la prima classe latente, i coefficienti stimati per l’età sono i seguenti:  $\hat{\beta}_{22} = 0.03$ ,  $\hat{\beta}_{32} = -0.02$ ,  $\hat{\beta}_{42} = 0.01$ . I parametri  $\beta_{21}$ ,  $\beta_{31}$  e  $\beta_{41}$  rappresentano invece l’intercetta per ciascuna classe.

La Figura 4.2 mostra la probabilità di appartenenza a ciascuna delle quattro classi latenti in funzione dell’età. Si può notare che un soggetto d’età compresa tra i 15 e i 23 anni, è più probabile abbia seguito l’approccio comportamentale (verde) quando era ansioso o depresso. Inoltre, è evidente che, a partire dai 23 anni, il profilo olistico

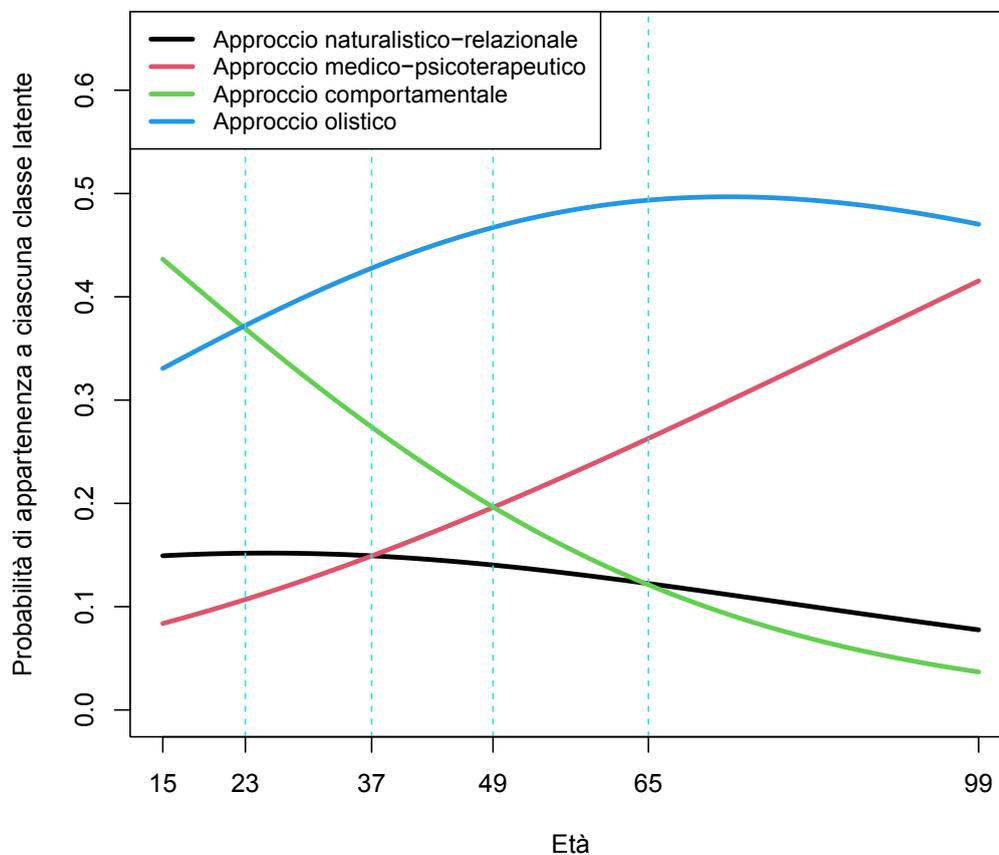


FIGURA 4.2: Probabilità di appartenenza ai quattro profili latenti in funzione dell'età.

(blu) è quello più rappresentativo. La probabilità di appartenere al profilo medico-psicoterapeutico (rosso) cresce all'aumentare dell'età, mentre la probabilità di far parte del gruppo comportamentale (verde) decresce con l'età.

Tralasciando il gruppo olistico, i soggetti d'età compresa tra i 23 e i 49 anni tendono ad appartenere di più al profilo comportamentale; dopo i 49 anni la probabilità è più elevata per l'appartenenza al profilo medico-psicoterapeutico.

In particolare, dai 49 anni in poi la probabilità di appartenere al profilo medico-psicoterapeutico (rosso) continua a crescere rispetto a quella per il profilo comportamentale (verde). Gli adulti hanno quindi una maggiore tendenza ad affidarsi a metodi che implicano il supporto di un professionista della salute mentale, quale un medico o uno psicologo, per esempio per intraprendere una terapia o richiedere una prescrizione per specifici medicinali come gli antidepressivi.

Tra i 15 e i 37 anni il gruppo meno rappresentativo è quello medico-psicoterapeutico

(rosso), tra i 37 e i 65 anni è quello naturalistico-relazionale (nero), mentre per gli over 65 l'approccio meno probabile è quello comportamentale (verde).

Riassumendo, i giovani tra i 15 e i 23 anni tendono di più ad appartenere al profilo comportamentale e quindi a introdurre cambiamenti concreti in ambito relazionale e personale. Tralasciando il gruppo olistico, che è il più rappresentativo a tutte le età dopo i 23 anni, gli over 50 tendono di più a sperimentare l'approccio medico-psicoterapeutico rispetto a quello comportamentale.

## 4.7 Confronto con modelli alternativi

Si presentano nel seguito due possibili modelli alternativi al modello di regressione a classi latenti stimato e descritto nelle sezioni precedenti: il modello di Rasch e il modello log-lineare.

### 4.7.1 Modello di Rasch

Il modello a classi latenti e il modello di Rasch cercano di spiegare la relazione tra un tratto latente, ovvero una caratteristica non osservata, e le sue manifestazioni, per esempio dei risultati o delle risposte osservate. L'obiettivo di entrambi è quindi quello di “spiegare come varia la probabilità di osservare un certo *pattern* di risposte in funzione del tratto latente misurato” (Bacci, 2006). La principale differenza tra i due modelli è che: nel primo, la variabile latente si assume discreta, mentre nel secondo continua.

Il modello di Rasch (RM) è stato sviluppato negli anni 50 del secolo scorso dal matematico danese Georg Rasch, come strumento per testare i risultati scolastici dei bambini. L'originario ambito di applicazione è quindi quello della ricerca psicologica, per “la valutazione dell'abilità di un gruppo di studenti a cui è sottoposto un test costituito da problemi di difficoltà variabile” (Bacci, 2006).

Un RM assume che il tratto latente (la conoscenza o l'abilità di uno studente, l'attitudine di un insegnante, ecc.) e gli *item* siano organizzati in un continuum non osservabile e che la risposta di un soggetto ad un *item* sia funzione della differenza tra le sue abilità e la difficoltà dell'*item* (Bacci, 2006).

Sia  $J$  il numero di *item* e  $I$  il numero di individui. Siano  $\beta_j$ ,  $j \in \{1, \dots, J\}$ , i parametri di “difficoltà” degli *item* e  $\theta_i$ ,  $i \in \{1, \dots, I\}$ , i parametri di “abilità” degli individui. I primi indicano il livello di difficoltà di un *item*, per esempio una domanda in un questionario a cui si può rispondere “Sì” o “No”, in termini di quanto è probabile che un individuo risponda correttamente. I parametri di “abilità” forniscono un'informazione

su quanto l'attributo latente che si vuole misurare, per esempio la conoscenza, è presente in ciascun individuo.

Il modello di Rasch, nel caso in cui gli *item* siano variabili dicotomiche, assume la forma

$$g_j(\theta_i) = Pr(Y_{ij} = y_{ij} | \theta_i, \beta_j) = \frac{\exp[x_{ij}(\theta_i - \beta_j)]}{1 + \exp(\theta_i - \beta_j)} \quad (4.18)$$

dove  $g_j(\theta_i)$  è la curva caratteristica del  $j$ -esimo *item* in funzione del tratto latente presente nel soggetto  $i$ ,  $y_{ij}$  è la risposta dell'individuo  $i$  alla domanda  $j$ , con  $y_{ij} = 0, 1$ .

Come principale approccio alternativo al modello a classi latenti, si è quindi considerato il modello di Rasch. Gli *item* sono gli otto approcci per la gestione di ansia e depressione, mentre gli individui sono i 4884 soggetti che hanno avuto un disturbo psicologico.

Per il modello di Rasch, AIC e BIC sono rispettivamente pari a 46795.07 e 46840.55, mentre per il modello a classi latenti sono invece pari a 45716.24 e 46148.72. Per verificare che il modello a classi latenti sia effettivamente migliore rispetto al modello di Rasch, sono stati calcolati anche l'Indice di Dissimilarità, pari a 0.9, e i Residui Bivariati riportati nella Tabella 4.7. I valori di queste statistiche confermano il pessimo adattamento del modello di Rasch, in generale e rispetto al modello di regressione a classi latenti. Infatti, solo sei combinazioni delle 256 possibili presentano una frequenza attesa non nulla.

TABELLA 4.7: Residui Bivariati (BVR) per tutte le coppie di variabili osservate per il modello di Rasch.

	A	B	C	D	E	F	G
B	2448.09						
C	5501.00	954.76					
D	4783.21	3195.92	4027.95				
E	28980.63	984.64	4947.55	8464.62			
F	8109.73	5739.61	2565.12	14137.57	3719.55		
G	12142.12	4199.57	3330.26	50414.02	5953.02	11620.83	
H	7682.56	1303.20	14164.78	6000.89	5239.86	2783.56	3629.04

Eliminando successivamente gli *item* per cui si riscontra un peggiore adattamento, non si giunge comunque ad un buon modello. Anche la ricerca di sottogruppi della popolazione, omogenei rispetto a una o più variabili di aggregazione (es. genere, età), rispetto ai quali il modello presenti un buon adattamento, non ha portato a buoni risultati.

### 4.7.2 Modello log-lineare

I modelli log-lineari appartengono alla famiglia dei modelli lineari generalizzati e si utilizzano comunemente per la modellazione di dati di conteggio organizzati in tabelle di contingenza, per descrivere gli schemi di indipendenza, associazione o interazione tra due o più variabili categoriali (Salvan et al., 2020). Questi modelli considerano congiuntamente tutte le variabili osservate come variabili risposta, modellando le frequenze in ogni cella per ciascuna combinazione dei livelli delle variabili considerate.

Siano  $A, B, C, D, E, F, G$  e  $H$  otto fattori, con  $j = 2$  livelli ciascuno, con  $a, b, c, d, e, f, g, h \in \{0, 1\}$ . Si consideri una tabella di contingenza a otto entrate con  $j^8 = 2^8 = 256$  celle.

Il modello log-lineare assume che le frequenze osservate  $y_{abcdefgh}$  siano realizzazioni di variabili casuali  $Y_{abcdefgh} \sim P(\mu_{abcdefgh})$ , con  $\mu_{abcdefgh}$  frequenze attese per la specifica combinazione  $abcdefgh$ . Il modello log-lineare più semplice, detto di mutua indipendenza, considera i soli effetti principali, corrisponde ad un modello a classi latenti con una sola classe e ha la seguente forma

$$\log(\mu_{abcdefgh}) = \lambda + \lambda_a^A + \lambda_b^B + \lambda_c^C + \lambda_d^D + \lambda_e^E + \lambda_f^F + \lambda_g^G + \lambda_h^H. \quad (4.19)$$

Come valida alternativa al modello a classi latenti, si è quindi adattato un modello log-lineare ai dati relativi ai 4884 soggetti che hanno sofferto d'ansia o depressione.

Per valutare la complessità del modello, è stata utilizzata una procedura passo a passo in avanti, a partire dal modello nullo, basata sul criterio dell'AIC. La scelta è ricaduta su questa procedura, in quanto più rapida e parsimoniosa rispetto ad una procedura passo a passo all'indietro. Il modello finale include: tutti gli effetti principali; le interazioni del secondo ordine tra le coppie di variabili H-E, A-D, C-H, B-E, C-E, E-A, C-A, E-G, H-G, B-D, G-F, E-F, B-F, H-F, A-F, D-G, E-D, C-B, B-G, A-G, H-B, C-G; le interazioni del terzo ordine tra A-G-F, H-B-E, C-H-G, C-A-G.

AIC per il modello finale è pari a 1282, mentre BIC è uguale a 1509.505; l'Indice di Dissimilarità è pari a 0.06 e il TBVR è uguale a 4.45. Inoltre, la Tabella 4.8 mostra i Residui Bivariati calcolati per tutte le coppie di variabili. I valori molto bassi non sorprendono, in quanto un modello log-lineare con effetti del secondo ordine cattura in modo perfetto le marginali bivariate corrispondenti. Fidarsi di queste metriche potrebbe quindi essere fuorviante: per questo motivo si decide di attribuire un po' più d'importanza ai criteri di informazione.

AIC e BIC indicano che il modello log-lineare è migliore, in termini di bontà di adattamento, rispetto al modello a classi latenti. Quest'ultimo, a differenza del primo,

consente però di individuare e caratterizzare gruppi di soggetti in base alle diverse strategie perseguite per alleviare il proprio stato di malessere psicologico e di includere covariate in modo semplice. Inoltre, quando sono presenti tante interazioni, tra cui alcune del terzo ordine, l'interpretazione dei risultati di un modello log-lineare diventa più complessa.

TABELLA 4.8: Residui Bivariati (BVR), per tutte le coppie di variabili osservate, per il modello log-lineare.

	A	B	C	D	E	F	G
B	0.29						
C	0.00	0.00					
D	0.00	0.00	0.16				
E	0.00	0.00	0.00	0.00			
F	0.00	0.00	0.33	0.15	0.00		
G	0.00	0.00	1.81	0.00	0.00	0.00	
H	0.58	0.00	0.00	0.89	0.00	0.00	0.00

Nonostante presenti dei limiti, il modello log-lineare sembra essere migliore rispetto al modello di regressione a classi latenti. Per questo, nel seguito si commentano brevemente alcune tra le interazioni più importanti.

L'interpretazione dei parametri di un modello log-lineare si effettua considerando i termini di grado massimo. Per esempio, in un modello che presenta tutte le interazioni del secondo ordine, le stime degli effetti principali sono irrilevanti. Un'interazione del terzo ordine, in un modello che non presenta interazioni di grado superiore, consente all'associazione tra qualsiasi coppia di quelle tre variabili, di variare attraverso i livelli della terza variabile, fissati i valori di tutte le altre variabili non coinvolte nell'interazione a tre. Per quanto riguarda le interazioni a due, non è appropriato cercare di interpretare quelle che presentano coppie di variabili già coinvolte in un'interazione a tre. In generale,

TABELLA 4.9: Legenda per la Tabella 4.10.

A	Aver parlato con un professionista della salute mentale
B	Aver partecipato ad attività religiose o spirituali
C	Aver parlato con amici o familiari
D	Aver assunto medicinali sotto prescrizione medica
E	Aver introdotto comportamenti più sani
F	Aver cambiato la propria situazione lavorativa
G	Aver modificato la propria situazione relazionale
H	Aver trascorso tempo all'aperto o in mezzo alla natura

un'interazione a due implica che quelle due variabili sono condizionatamente dipendenti: l'*odds ratio* è lo stesso ad ogni combinazione dei livelli delle rimanenti variabili.

TABELLA 4.10: Stime dei parametri, errori standard e significatività osservata, per il modello log-lineare. In Tabella 4.9 si riporta un'utile legenda.

	Stima	Errore Standard	<i>p</i> -value
Intercetta	3.40	0.10	0.00
C	0.17	0.11	0.13
H	-0.36	0.10	0.00
B	-2.33	0.16	0.00
E	-1.08	0.11	0.00
A	-0.83	0.11	0.00
D	-0.79	0.08	0.00
G	-0.95	0.15	0.00
F	-1.36	0.10	0.00
H:E	0.99	0.08	0.00
A:D	1.67	0.07	0.00
C:H	0.93	0.11	0.00
B:E	0.82	0.17	0.00
C:E	0.57	0.08	0.00
E:A	0.31	0.07	0.00
C:A	0.21	0.11	0.05
E:G	0.41	0.07	0.00
H:G	0.65	0.14	0.00
B:D	0.25	0.07	0.00
G:F	0.87	0.10	0.00
E:F	0.49	0.07	0.00
B:F	0.28	0.07	0.00
H:F	0.29	0.07	0.00
A:F	0.36	0.09	0.00
D:G	-0.26	0.06	0.00
E:D	0.22	0.07	0.00
C:B	0.22	0.09	0.02
B:G	0.16	0.07	0.02
A:G	0.03	0.15	0.86
H:B	0.39	0.16	0.01
C:G	0.27	0.16	0.10
A:G:F	-0.28	0.12	0.02
H:B:E	-0.37	0.19	0.05
C:H:G	-0.48	0.16	0.00
C:A:G	0.29	0.15	0.05

Con riferimento alla Tabella 4.10, l'interazione a tre più significativa è quella tra i metodi C, H e G. Questo significa, per esempio, che l'associazione tra aver parlato con

amici e familiari (C) e aver trascorso del tempo all'aperto (H) varia tra chi ha modificato la propria situazione relazionale (G) e chi non lo ha fatto.

Considerando le interazioni a due, è più probabile aver parlato con uno psicologo (A) per chi ha assunto medicinali (D) rispetto a chi non li ha assunti e per chi ha cambiato la propria situazione lavorativa (F) rispetto a chi non lo ha fatto. Inoltre, è più probabile aver partecipato ad attività spirituali (B) per chi ha richiesto la prescrizione di un medicinale specifico (D), per chi ha modificato la propria situazione lavorativa (F) e quella relazionale (G). Se un soggetto ha modificato la propria situazione lavorativa (F) è più probabile che abbia trascorso anche del tempo in mezzo alla natura (H); coloro che hanno apportato un cambiamento nelle relazioni interpersonali hanno assunto medicinali (D) con una probabilità maggiore rispetto a chi ha lasciato inalterate le proprie cerchie sociali. In aggiunta, è più probabile aver introdotto comportamenti più sani (E) per chi ha intrapreso un percorso psicoterapeutico (A) rispetto a chi non si è rivolto ad un professionista della salute mentale, per chi ha assunto medicinali (D) rispetto a chi non ne ha richiesto la somministrazione, per chi ha modificato la propria situazione relazionale (G) e lavorativa (F).

# Capitolo 5

## Conclusioni

Nella presente tesi è stata proposta un'analisi dei dati del “Global Monitor 2020: Mental Health”, indagine sulla salute mentale a livello mondiale, tramite modelli per dati discreti. In particolare, è stato analizzato un campione di 27475 soggetti, d'età compresa tra i 15 e i 99 anni, residenti in 28 paesi europei e nord-americani. Tra questi, 4884 intervistati hanno dichiarato di essere stati ansiosi o depressi. In questa sede, ansia e depressione sono state intese come stati psicologici debilitanti, al punto da non poter continuare le proprie attività quotidiane normalmente, per due o più settimane.

Per valutare come l'esperienza di un disturbo d'ansia o depressione dipenda da alcune caratteristiche socio-demografiche, è stato applicato un modello di regressione logistica. Per caratterizzare il comportamento dei soggetti intervistati, in termini di strategie messe in atto per ristabilire il proprio benessere mentale, per profilare la popolazione e indagare il ruolo di alcuni fattori socio-demografici nella scelta dei metodi per la gestione di un disagio psicologico, è stato applicato un modello di regressione a classi latenti. Inoltre, ogni modello è stato messo a confronto con tecniche alternative, coerenti con l'obiettivo specifico.

Le analisi esplorative hanno suggerito la possibile presenza di associazione tra l'aver sofferto d'ansia o depressione e aver conseguito un livello di istruzione primaria, essere disoccupato, avere un reddito basso, essere femmina, arrivare a fine mese con difficoltà e avere meno di 30 anni. Per quanto riguarda i metodi sperimentati per la gestione di ansia e depressione, dalle analisi descrittive è emerso che la scelta di alcuni comportamenti sembra essere legata all'età, soprattutto parlare con uno psicologo, assumere medicinali e modificare la propria situazione lavorativa o relazionale. Il ruolo degli altri fattori socio-demografici non sembra invece particolarmente rilevante.

Il modello di regressione logistica ha mostrato un effetto significativo dei fattori socio-demografici considerati sull'aver sperimentato un disturbo ansioso-depressivo. In particolare, i risultati ottenuti, coerentemente con quelli in letteratura, indicano un'associazione tra: un livello educativo basso e l'aver provato un malessere psicologico; una condizione di svantaggio economico reale o percepito e l'aver sofferto di un disturbo ansioso-depressivo. Inoltre, anche il genere e la condizione lavorativa sembrano essere fattori rilevanti per la salute mentale: la probabilità di aver sofferto di ansia o depressione aumenta per le donne e per i disoccupati. In aggiunta, la propensione a dichiarare di aver avuto un disordine mentale diminuisce con l'avanzare dell'età.

Il gruppo dei paesi dove è più probabile essere stati ansiosi o depressi, rispetto all'Italia, include le seguenti nazioni: Francia, Bulgaria, Grecia, Canada, Paesi bassi, Portogallo, Svezia, Cipro, Stati Uniti, Spagna, Danimarca, Estonia, Germania e Polonia. La probabilità di aver sofferto di ansia o depressione diminuisce, invece, in Croazia, Slovenia, Ungheria e Lituania, rispetto all'Italia.

Data la presenza di una struttura multivello e di una variabile qualitativa con tante modalità (paese), come principale alternativa al modello di regressione logistica, si è adattato un modello di regressione logistica a effetti misti, con intercetta casuale per stato. I due modelli sono stati valutati sulla base del criterio AIC, del test di Hosmer e Lemeshow e di alcune metriche, quali accuratezza, specificità e precisione. Non si è riscontrata un'evidente differenza tra i due modelli in termini di bontà di adattamento. Inoltre, le stime dei parametri coincidono e i risultati si orientano quindi nella stessa direzione.

Tramite il modello di regressione a quattro classi latenti, che include l'età come unica covariata, sono stati identificati gruppi di soggetti, sulla base della probabilità di sperimentare una certa combinazione di metodi per la gestione di uno stato ansioso-depressivo. Per quanto riguarda la composizione delle quattro classi latenti, analizzando le stime dei parametri specifici per ogni gruppo, è stata proposta la seguente classificazione.

Il primo profilo (naturalistico-relazionale), rappresenta il 14% della popolazione ed è composto dagli individui che, quando erano ansiosi o depressi, hanno scelto quasi esclusivamente un approccio naturalistico-relazionale per recuperare il proprio benessere psicologico. Il secondo profilo (medico-psicoterapeutico), rappresenta il 19% della

popolazione ed è composto dagli individui che, insieme ad un approccio naturalistico-relazionale, hanno adottato una strategia medico-psicoterapeutica. Il terzo profilo (comportamentale), rappresenta il 22% della popolazione ed è composto dagli individui che, insieme ad un approccio naturalistico-relazionale, hanno sperimentato una strategia di tipo comportamentale, ovvero hanno modificato la propria condizione lavorativa e relazionale o hanno introdotto comportamenti più sani nella propria vita personale. Il quarto profilo (olistico), rappresenta il 45% della popolazione ed è caratterizzato da individui che hanno sperimentato un'ampia combinazione di metodi per ristabilire il proprio benessere mentale.

Per quanto riguarda l'effetto dell'età sulla probabilità di appartenere ad una specifica classe latente, si è osservato quanto segue. In sintesi, i giovani tra i 15 e i 23 anni tendono di più a far parte del profilo comportamentale; tralasciando il gruppo olistico, che è il più rappresentativo a tutte le età dopo i 23 anni, gli adulti con più di 50 anni tendono di più a sperimentare l'approccio medico-psicoterapeutico, rispetto a quello comportamentale.

Infine, il modello di regressione a classi latenti è stato messo a confronto con due principali modelli alternativi: il modello di Rasch e il modello log-lineare. I tre modelli sono stati valutati sulla base dei criteri d'informazione AIC e BIC e tramite le statistiche DI, BVR e TBVR: il modello di Rasch presenta un pessimo adattamento, mentre il modello log-lineare sembra migliore, in termini di bontà di adattamento, rispetto a quello a classi latenti.

In conclusione, genere, livello d'istruzione, reddito, occupazione lavorativa, età e paese, hanno un effetto sull'aver sperimentato una condizione di salute mentale invalidante. Sulla base dei risultati ottenuti, i programmi di promozione della salute mentale futuri, nei paesi analizzati, dovrebbero considerare strategie che abbiano come *target* i soggetti più fragili, in particolare, i meno istruiti, i disoccupati e coloro che vivono in condizioni di difficoltà economica. Riassumendo, i quattro profili latenti individuati, si contraddistinguono per la scelta delle seguenti strategie di gestione di uno stato ansioso-depressivo: approccio naturalistico-relazionale, approccio comportamentale, approccio medico-psicoterapeutico e approccio olistico.

Per quanto riguarda i limiti del presente lavoro, si precisa che l'indagine *Global Monitor* non può essere considerata come un vero e proprio studio epidemiologico, in quanto non sono state valutate delle diagnosi, ma soltanto le percezioni delle persone rispetto ad ansia e depressione, così come queste vengono definite nel questionario.

Inoltre, si suggerisce di interpretare con cautela i dati relativi al ruolo dei singoli paesi:

trarre la conclusione, per esempio, che i francesi sono più depressi rispetto agli italiani, potrebbe essere fuorviante. Infatti, potrebbero essere presenti differenze culturali tra paesi a livello di definizione di ansia e depressione, o stigmi sociali che inducono le persone a dichiarare il falso.

Infine, il modello log-lineare è migliore rispetto al modello a classi latenti, quindi dovrebbe essere preferibile, ma non permette di profilare la popolazione e di includere covariate.

Per quanto riguarda i possibili sviluppi del presente lavoro, si propone quanto segue. In primo luogo, per ottenere delle informazioni aggiuntive, si potrebbe meglio analizzare e interpretare il modello log-lineare, dato che l'adattamento ai dati è migliore rispetto a quello del modello a classi latenti. In secondo luogo, si potrebbe proporre un'ulteriore analisi, considerando anche le variabili relative alla percezione sull'utilità di ciascun metodo per la gestione di ansia e depressione, in quanto se ne potrebbero ricavare informazioni utili e interessanti. In terzo luogo, si potrebbe ricercare una tecnica alternativa a quella considerata in Appendice, per proporre un modello congiunto, combinando il modello di regressione logistica con quello a classi latenti.

# Appendice

## A. 1 Modello congiunto

Nel corso di questo lavoro di tesi, ci si è chiesti se fosse possibile combinare il modello di regressione logistica con il modello di regressione a classi latenti, proponendo quindi un modello congiunto per dati in questionari con domande filtro, ovvero per dati sequenziali.

In particolare, la domanda filtro a cui si fa riferimento è quella relativa all'aver sofferto d'ansia o depressione. Solo i soggetti che hanno risposto affermativamente a tale quesito, hanno poi risposto anche alle domande relative ai metodi sperimentati per la gestione del disturbo psicologico.

L'obiettivo che ci si è posti è stato quello di modellare la probabilità congiunta di aver sofferto di ansia o depressione e di aver sperimentato una specifica combinazione delle variabili che rappresentano otto differenti strategie per affrontare il malessere psicologico. Si può però facilmente dimostrare che modellare la probabilità congiunta equivale a modellare la probabilità condizionata.

Di seguito si riporta una breve dimostrazione.

Sia  $Y_{0i}$ ,  $i \in \{1, \dots, n\}$ , la variabile risposta del modello di regressione logistica, ovvero aver sofferto d'ansia o depressione, con  $Y_{0i} = 1$  se il soggetto ne ha sofferto;  $Y_{0i} = 0$  altrimenti. Siano  $Y_{1i}, \dots, Y_{8i}$  le variabili osservate del modello di regressione a classi latenti, con  $Y_{ij} = 1$  se il soggetto  $i$  ha sperimentato il metodo  $j$ ;  $Y_{ij} = 0$  altrimenti.

La probabilità congiunta è

$$Pr(Y_{i0} = y_{i0}, Y_{i1} = y_1, \dots, Y_{i8} = y_8) = Pr(Y_{i1} = y_1, \dots, Y_{i8} = y_8 | Y_{i0} = 1) \times Pr(Y_{i0} = 1). \quad (\text{A. 1})$$

Si ponga

$$w_i = Pr(Y_{i0} = 1) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i}. \quad (\text{A. 2})$$

La funzione di verosimiglianza è

$$\begin{aligned}
 L &= \prod_{i=1}^n \sum_{q=1}^Q p_q \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}} \times \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \\
 &= \prod_{i=1}^n \sum_{q=1}^Q p_q \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}} \times w_i.
 \end{aligned} \tag{A. 3}$$

La funzione di log-verosimiglianza è

$$\begin{aligned}
 \log L &= \sum_{i=1}^n \log \left[ \sum_{q=1}^Q p_q \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}} \times \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \right] \\
 &= \sum_{i=1}^n \log \left[ \sum_{q=1}^Q p_q \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}} \times w_i \right] \\
 &= \sum_{i=1}^n \left[ \log \left( \sum_{q=1}^Q p_q \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}} \right) + \log(w_i) \right].
 \end{aligned} \tag{A. 4}$$

Massimizzando la funzione di log-verosimiglianza rispetto a un qualsiasi parametro,  $w_i$  scompare. Per esempio,

$$\frac{\partial \log L}{\partial p_q} = \frac{\prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}}}{\sum_{q=1}^Q p_q \prod_{j=1}^8 \prod_{k=1}^2 (\pi_{jqk})^{y_{ijk}}}. \tag{A. 5}$$

Di conseguenza,  $w_i$  non influenza la stima dei parametri e non ha quindi senso proporre un modello congiunto di questo tipo.

## A. 2 Modello di regressione a classi latenti

Nella Tabella A. 1 si riportano le stime delle probabilità condizionate alla classe di appartenenza, per gli approcci da A ad H, per l'esito "aver sperimentato l'approccio  $j$ ", con i rispettivi intervalli di confidenza alla Wald al 95%.

TABELLA A. 1: Probabilità condizionate alla classe per i metodi da A ad H, con intervalli di confidenza alla Wald al 95%, per il modello a classi latenti.

Metodo	Classe $q$	$\widehat{Pr}(Y_i = 1 q)$	I.C. alla Wald
A	1	0.30	(0.21, 0.38)
	2	0.73	(0.69, 0.77)
	3	0.45	(0.38, 0.52)
	4	0.85	(0.83, 0.87)
B	1	0.11	(0.08, 0.15)
	2	0.16	(0.12, 0.19)
	3	0.27	(0.23, 0.30)
	4	0.33	(0.31, 0.36)
C	1	0.63	(0.58, 0.68)
	2	0.69	(0.64, 0.73)
	3	0.87	(0.84, 0.90)
	4	0.89	(0.87, 0.91)
D	1	0.00	(0.00, 0.00)
	2	0.97	(0.85, 1.09)
	3	0.00	(0.00, 0.00)
	4	0.90	(0.83, 0.97)
E	1	0.33	(0.25, 0.41)
	2	0.46	(0.40, 0.52)
	3	0.87	(0.82, 0.91)
	4	0.89	(0.86, 0.91)
F	1	0.29	(0.24, 0.34)
	2	0.23	(0.18, 0.28)
	3	0.55	(0.51, 0.59)
	4	0.60	(0.57, 0.63)
G	1	0.37	(0.31, 0.43)
	2	0.27	(0.22, 0.32)
	3	0.66	(0.62, 0.71)
	4	0.63	(0.60, 0.66)
H	1	0.56	(0.50, 0.62)
	2	0.57	(0.52, 0.62)
	3	0.90	(0.87, 0.93)
	4	0.89	(0.87, 0.91)



# Bibliografia

- AGRESTI, A. (2013). *Categorical Data Analysis Third Edition*. John Wiley & Sons, Inc., Hoboken, New Jersey.
- AGRESTI, A. (2019). *An Introduction to Categorical Data Analysis Third Edition*. John Wiley & Sons Inc. Doi: 10.1002/0470114754.
- ALIVERTI, E. & RUSSO, M. (2022). Dynamic modeling of the italians' attitude towards covid-19. *Statistics in Medicine* **41**. Doi: 10.1002/sim.9560.
- APA (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. American Psychiatric Association, Arlington, VA.
- AZZALINI, A. & SCARPA, B. (2012). *Data Analysis and Data Mining: An Introduction*. Oxford University Press Inc.
- BACCI, S. (2006). I modelli di rasch nella valutazione della didattica universitaria. *Statistica Applicata* **18**.
- BAKER, F. & KIM, S. (2004). *Item Response Theory: Parameter Estimation Techniques Second Edition*. CRC Press, 1992.
- CLEMENTS, L., GIBBONS, P., GREENWOOD, E., GORMAN, C. & HICKMAN, M. (2020). *The role of science in mental health. Insights from the Wellcome Global Monitor*.
- DE FRANCE, K. & HOLLENSTEIN, T. (2021). Emotion regulation strategy use and success during adolescence: Assessing the role of context. *Journal of Research on Adolescence* **32**. Doi: 10.1111/jora.12672.
- DREGER, S., GERLINGER, T. & BOLTE, G. (2016). Gender inequalities in mental well-being in 26 european countries: Do welfare regimes matter? *The European Journal of Public Health* **26**, ckw074. Doi: 10.1093/eurpub/ckw074.

- EIGE (2021). Gender equality index 2021 health Doi: 10.2839/035225.
- GHIRINI, S. & VICHI, M. (2018). Salute mentale [rapporto osservasalute 2018] Doi: 10.13140/RG.2.2.23386.67524.
- HAGENAARS, J. A. (1998). Categorical causal modeling: Latent class analysis and directed log-linear models with latent variables. *Sociological Methods & Research* **26**, 436–486. Doi: 10.1177/0049124198026004002.
- HOSMER, D. W. & LEMESHOW, S. (2000). *Assessing the Fit of the Model*, chap. 5. John Wiley & Sons, Ltd, pp. 143–202.
- LEHTINEN, V., SOHLMAN, B. & KOVESH-MASFETY, V. (2005). Level of positive mental health in the european union: Results from the eurobarometer 2002 survey. *Clinical practice and epidemiology in mental health : CP & EMH* **1**, 9. Doi: 10.1186/1745-0179-1-9.
- LEWIS, J. & LINZER, D. (2011). polca: An r package for polytomous variable latent class analysis. *Journal of Statistical Software* **42**. Doi: 10.18637/jss.v042.i10.
- MAIR, P., RUSCH, T., HATZINGER, R., MAIER, M. J. & DEBELAK, R. (2023). *Package ‘eRm’*. R package version 1.0.4.
- MCCULLAGH, P. & NELDER, J. (1989). *Generalized Linear Models. 2nd Edition*. Chapman and Hall, London. Doi: 10.1007/978-1-4899-3242-6.
- MS (2022). Che cos'è la salute mentale .
- MURPHY, J., HOUSTON, J. & SHEVLIN, M. (2012). Principles and applications of latent class analysis in psychological research. *The Irish Journal of Psychology* **28**, 87–96. Doi: 10.1080/03033910.2007.10446251.
- PETERSEN, K., QUALTER, P. & HUMPHREY, N. (2019). The application of latent class analysis for investigating population child mental health: A systematic review. *Frontiers in Psychology* **10**. Doi: 10.3389/fpsyg.2019.01214.
- SALVAN, A., SARTORI, N. & PACE, L. (2020). *Modelli lineari generalizzati*. Springer Verlag. Doi: 10.1007/978-88-470-4002-1.2.
- SAXENA, S., FUNK, M. & CHISHOLM, D. (2013). World health assembly adopts comprehensive mental health action plan 2013–2020. *Lancet* **381**, 1970–1. Doi: 10.1016/S0140-6736(13)61139-3.

- SCHÜTTE, S., CHASTANG, J.-F., PARENT-THIRION, A., VERMEYLEN, G. & NIEDHAMMER, I. (2014). Social inequalities in psychological well-being: A european comparison. *Community mental health journal* **50**. Doi: 10.1007/s10597-014-9725-8.
- SOLMI, M., RADUA, J., OLIVOLA, M., CROCE, E., SOARDO, L., SALAZAR DE PABLO, G., SHIN, J. I., KIRKBRIDE, J., JONES, P., KIM, J. H., KIM, J. Y., CARVALHO, A., SEEMAN, M., CORRELL, C. & FUSAR-POLI, P. (2022). Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol Psychiatry*. **27**. Doi: 10.1038/s41380-021-01161-7.
- VAN KOLLENBURG, G., MULDER, J. & VERMUNT, J. (2015). Assessing model fit in latent class analysis when asymptotics do not hold. *Methodology* **11**, 65–79.
- WHO (2021a). *Depression and Other Common Mental Health Disorders: Global Health Estimate*. World Health Organization. WHO/MSD/MER/2017.2.
- WHO (2021b). *Mental Health ATLAS 2020*. World Health Organization. ISBN: 978-92-4-003670-3.

