

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**ANALISI DELLE CURVE DI DECADIMENTO COGNITIVO
IN UN GRUPPO DI PAZIENTI GERIATRICI**

Relatore: Dott. Andrea Sottosanti
Dipartimento di Medicina

Correlatrice: Dott.ssa Alessandra Coin
UOC Geriatria, Azienda Ospedale – Università Padova

Laureanda: Giulia Palladino
Matricola N. 2062441

Anno Accademico 2023/2024

Indice

Introduzione	ix
1 Declino cognitivo e fisico nell'invecchiamento	1
1.1 Decadimento cognitivo	2
1.2 Decadimento fisico	7
1.3 Strategie di intervento	8
2 Presentazione dei dati	13
2.1 Configurazione sperimentale	13
2.1.1 Descrizione delle variabili fisse	15
2.1.2 Descrizione delle variabili temporali	17
2.2 Pulizia dei dati	22
3 Modelli per studi longitudinali	25
3.1 Modello Lineare a Effetti Misti	25
3.1.1 Specificazione del modello	26
3.2 Stima di massima verosimiglianza per i modelli LME	27
3.2.1 Distribuzioni approssimate	32
3.3 Estensione del modello LME base	33
3.3.1 Strutture di correlazione	34
4 Analisi dei dati	37
4.1 Presentazione del dataset	37
4.1.1 Analisi descrittiva	38
4.2 Applicazione del modello LME	47
4.2.1 Clustering	51
5 Conclusioni e sviluppi futuri	59
5.1 Opportunità	64
Appendice	65
Bibliografia	85

Elenco delle figure

4.1	Distribuzione del MMSE rispetto alle varie visite.	40
4.2	Distribuzione dell'età rispetto alle varie visite.	44
4.3	Evoluzione del punteggio MMSE nel tempo, suddivisa per soggetto.	45
4.4	Andamento del punteggio MMSE nel tempo, analizzato per soggetti e relative diagnosi.	46
4.5	Coefficienti degli effetti fissi con relativi intervalli di confidenza.	49
4.6	Residui standardizzati del modello.	52
4.7	Clustering sugli effetti casuali.	53
4.8	Traiettorie temporali del punteggio MMSE, stimate con il metodo loess e suddivise per cluster distinti.	54
4.9	Albero decisionale per la classificazione dei soggetti nei diversi cluster.	57
A.1	Distribuzione del MMSE in ogni visita (dataset originale).	65
A.2	Distribuzione dell'età in ogni visita (dataset originale).	68
A.3	Frequenze relative delle diagnosi nei cluster.	70
A.4	Frequenze relative dei cluster nelle diagnosi.	70
A.5	Frequenze relative della memantina nei cluster.	71
A.6	Frequenze relative dei cluster nella memantina.	71
A.7	Frequenze relative di altri farmaci nei cluster.	72
A.8	Frequenze relative dei cluster in altri farmaci.	72
A.9	Frequenze relative del genere nei cluster.	73
A.10	Frequenze relative dei cluster nel genere.	73
A.11	Frequenze relative dell'istruzione nei cluster.	74
A.12	Frequenze relative dei cluster nell'istruzione.	74
A.13	Frequenze relative delle malattie psichiatriche nei cluster.	75
A.14	Frequenze relative dei cluster nelle malattie psichiatriche.	75
A.15	Frequenze relative delle malattie cardiovascolari nei cluster.	76
A.16	Frequenze relative dei cluster nelle malattie cardiovascolari.	76
A.17	Frequenze relative delle malattie metaboliche nei cluster.	77
A.18	Frequenze relative dei cluster nelle malattie metaboliche.	77
A.19	Frequenze relative delle malattie gastrointestinali nei cluster.	78

A.20	Frequenze relative dei cluster nelle malattie gastrointestinali.	78
A.21	Frequenze relative delle neoplasie nei cluster.	79
A.22	Frequenze relative dei cluster nelle neoplasie.	79
A.23	Frequenze relative di altre malattie nei cluster.	80
A.24	Frequenze relative dei cluster in altre malattie.	80
A.25	Frequenze relative del fumo nei cluster.	81
A.26	Frequenze relative dei cluster nel fumo.	81
A.27	Frequenze relative dell'alcool nei cluster.	82
A.28	Frequenze relative dei cluster nell'alcool.	82
A.29	Frequenze relative dei traumi cranici nei cluster.	83
A.30	Frequenze relative dei cluster nei traumi cranici.	83
A.31	Frequenze relative dello stato civile nei cluster.	84
A.32	Frequenze relative dei cluster nello stato civile.	84

Elenco delle tabelle

2.1	Numero di pazienti per visita.	14
2.2	Descrizione delle variabili in anagrafica generale.	15
2.3	Elenco delle patologie.	16
2.4	Descrizione dei fattori di rischio.	16
2.5	Descrizione delle variabili relative alla diagnosi e valutazione neuropsicologica.	17
2.6	Descrizione delle variabili cliniche raccolte durante le visite di controllo.	18
2.7	Descrizione delle variabili relative alle terapie prescritte.	20
2.8	Descrizione delle variabili relative a ADL, IADL e valutazioni cognitive.	21
4.1	Panoramica delle variabili e dei soggetti del dataset.	39
4.2	Numero assoluto e % di pazienti in ogni visita	39
4.3	Frequenze assolute e % di memantina nelle varie visite.	41
4.4	Frequenze assolute e % di altri farmaci nelle varie visite.	42
4.5	Frequenze assolute e % per le variabili della sezione APR	42
4.6	Frequenze assolute e % per alcune delle variabili esplicative.	43
4.7	Deviazione standard e correlazione degli effetti casuali con relativi intervalli di confidenza al 95%.	50
4.8	Frequenze assolute e % dei cluster in base agli effetti casuali.	53
A.1	Frequenze assolute e % per alcune delle variabili esplicative.	66
A.2	Frequenze assolute e % di memantina nelle varie visite.	67
A.3	Frequenze assolute e % per le variabili della sezione APR	67
A.4	Tabella con stime, errori standard, DF, t-value e p-value derivanti dal modello LME	69

Introduzione

Il deterioramento cognitivo legato all'invecchiamento è un problema crescente, soprattutto in considerazione dell'aumento dell'aspettativa di vita. Con un numero sempre maggiore di persone che raggiungono età avanzate, è fondamentale identificare terapie che possano almeno rallentare la progressione di questo declino, migliorando così la qualità di vita dei pazienti e alleggerendo il carico sui sistemi sanitari. In questo contesto si inserisce lo studio condotto nel reparto di geriatria dell'ospedale di Padova, avviato nel 2015 e tuttora in corso. Lo studio segue un gruppo di pazienti di età superiore ai sessantacinque anni, caratterizzato da un protocollo flessibile che consente ai pazienti di presentarsi alle visite senza un calendario rigido, purché tra un incontro e l'altro intercorra un minimo di 30 giorni.

Durante la prima visita, sono state raccolte informazioni di base per ciascun paziente (sesso, stato civile, età, livello di istruzione, abitudini come fumo e consumo di alcol, anamnesi patologica remota, ecc.) e sono state assegnate diagnosi specifiche. Le condizioni diagnosticate includono MCI (Mild Cognitive Impairment), AD (Alzheimer's disease), VD (Vascular disease), MD (Mixed Dementia), PD (Parkinson's disease), DLB (Dementia with Lewy Bodies), PSP (Progressive Supranuclear Palsy) e FTD (Frontotemporal Dementia). È stato inoltre rilevato l'eventuale uso di trattamenti preesistenti, quali Ache-I, memantina, antidepressivi, antipsicotici, trazodone e benzodiazepine. La valutazione iniziale ha incluso anche test cognitivi e funzionali, come l'MMSE, ADL e IADL, sulla base dei quali sono stati prescritti eventuali trattamenti. Nei successivi follow-up, i test vengono ripetuti e le terapie modificate a seconda dei risultati e del decorso della malattia.

L'obiettivo di questo elaborato è esplorare i fattori che possono influenzare il deterioramento cognitivo, misurato attraverso il punteggio MMSE e valutare se alcuni trattamenti riescano effettivamente a rallentare la progressione della malattia. Per questo, è stato scelto un modello lineare a effetti misti (LME), considerando come effetti casuali il tempo e l'uso di memantina. Inoltre, si è adottata una struttura di correlazione che tiene conto delle osservazioni su un periodo continuo per ogni paziente.

Successivamente, i pazienti sono stati suddivisi in quattro cluster in base agli effetti casuali individuati: (1) effetto positivo sia del tempo che della memantina sul MMSE, (2) effetto positivo della memantina e negativo del tempo, (3) effetto positivo del tempo e negativo della memantina, (4) effetto negativo di entrambi. Questi cluster sono stati analizzati per identificare come essi si distribuiscano nei vari predittori considerati. Infine, è stato costruito un albero decisionale, il quale ha permesso di individuare le combinazioni di variabili che caratterizzano ciascun cluster, fornendo una visione chiara dei fattori che contribuiscono al decorso del deterioramento cognitivo nei diversi sottogruppi di pazienti.

Capitolo 1

Declino cognitivo e fisico nell'invecchiamento

L'invecchiamento è un processo biologico tramite il quale il corpo umano subisce alterazioni a livello psicologico, fisiologico e, parallelamente, possono emergere peggioramenti nelle funzioni cognitive. Col passare del tempo, le cellule del corpo umano subiscono un deterioramento funzionale, che porta a una diminuzione delle capacità di rigenerazione e adattamento dell'organismo [37]. Più nello specifico si possono riscontrare:

- **Cambiamenti fisici:** con l'avanzare dell'età, il corpo umano subisce diversi cambiamenti fisici significativi [36]. Uno dei più evidenti è la *perdita di massa muscolare e ossea*, conosciuta rispettivamente come sarcopenia e osteoporosi. Queste condizioni portano a una riduzione della forza fisica, aumentando così il rischio di fratture e infortuni. Oltre a questo, vi è una *diminuzione della capacità immunitaria*: il sistema immunitario si indebolisce, rendendo l'organismo più vulnerabile a infezioni e malattie. Infine, anche il funzionamento degli organi vitali inizia a declinare. Organi come il cuore e i reni, ad esempio, perdono parte della loro efficienza, compromettendo funzioni fondamentali come la circolazione del sangue e la digestione.
- **Cambiamenti cognitivi:** con l'avanzare dell'età, si osservano anche cambiamenti significativi a livello cognitivo. Uno degli aspetti più comuni è la *perdita di memoria* e un generale rallentamento nelle funzioni cognitive. Questo processo può influenzare la memoria a breve termine, rendendo più difficile il recupero di informazioni recenti e ostacolando la capacità di apprendere nuove nozioni. Parallelamente, l'invecchiamento comporta un aumento del rischio di sviluppare *malattie neurodegenerative*. Patologie come l'Alzheimer e il Parkinson, ad esempio, diventano più probabili con il passa-

re degli anni, contribuendo ulteriormente al declino delle funzioni cognitive [35].

- **Cambiamenti psicologici e sociali:** i cambiamenti psicologici e sociali che accompagnano l'invecchiamento possono avere un impatto profondo sulla vita degli individui anziani [36]. Tra questi cambiamenti, uno dei più significativi è l'*adattamento emotivo*. Gli anziani possono affrontare variazioni nel loro stato emotivo, sperimentando sentimenti di depressione o ansia. Questi stati possono derivare da diversi fattori, tra cui l'isolamento sociale e la perdita di persone care, che possono creare un senso di solitudine e vulnerabilità. Inoltre, il pensionamento e la conseguente diminuzione delle attività fisiche e sociali possono influenzare la *qualità della vita*. L'uscita dal mondo del lavoro spesso segna una transizione importante, e la mancanza di routine e interazioni quotidiane può contribuire a un senso di vuoto. È quindi cruciale per gli anziani trovare nuove modalità di coinvolgimento sociale e attività che possano promuovere il benessere emotivo, migliorando così la loro qualità della vita.

Dunque, come appena visto, il decadimento fisico, cognitivo e psicologico che si manifesta nelle persone in età avanzata è un processo naturale che coinvolge cambiamenti progressivi nel corpo e nella mente. Tuttavia, questi cambiamenti non si mostrano necessariamente allo stesso modo in tutte le persone, ma sono il risultato di una combinazione di fattori genetici, ambientali e di stile di vita. Di seguito si andranno ad approfondire il decadimento fisico e cognitivo, in quanto risultano i due aspetti di maggiore interesse in questo studio.

1.1 Decadimento cognitivo

Secondo la definizione fornita dall'associazione Americana degli Psicologi (APA) nel 1987, "la demenza o decadimento cognitivo cronico-progressivo, è una malattia del cervello che comporta la compromissione delle funzioni cognitive tale da pregiudicare la possibilità di una vita in autonomia. Ai sintomi cognitivi si associano quasi sempre alterazioni della personalità e del comportamento che variano come entità da individuo a individuo. Inoltre, è presente una progressiva alterazione dello stato funzionale".

Il declino cognitivo negli anziani può essere valutato attraverso diversi test e scale che esaminano le funzioni cognitive come la memoria, l'attenzione, l'orientamento, il linguaggio e le abilità esecutive. Di seguito sono elencati alcuni degli indici più comunemente utilizzati per misurare il declino cognitivo.

Mini Mental State Examination (MMSE)

Il MMSE è uno dei test più utilizzati per valutare lo stato cognitivo globale. Viene spesso adoperato in ambito clinico per identificare disturbi cognitivi legati a condizioni come la demenza, l'Alzheimer e altre patologie neurologiche. Questo test standardizzato è stato ideato da Folstein e colleghi nel 1975 [10] ed è rimasto uno strumento chiave per la sua semplicità e capacità di fornire una visione d'insieme dello stato cognitivo di un individuo in un breve periodo di tempo.

Il MMSE è strutturato per valutare diverse funzioni cognitive: l'orientamento temporale e spaziale, la memoria, l'attenzione, il calcolo, il linguaggio e le capacità visuo-spaziali. Durante l'esame, il soggetto viene sottoposto a una serie di domande e compiti che coprono queste aree. Ad esempio, vengono chieste informazioni sulla data corrente (anno, mese, giorno), sul luogo in cui si trova il paziente e altre domande che richiedono la memoria a breve termine. Ci sono anche esercizi più complessi, come ripetere frasi, seguire comandi specifici, scrivere brevi frasi o copiare disegni semplici.

Il punteggio massimo ottenibile nel MMSE è 30, ma i punteggi inferiori riflettono livelli variabili di deterioramento cognitivo. In generale, un punteggio tra 24 e 30 è considerato nella norma o indicativo di un declino cognitivo lieve, mentre un punteggio tra 18 e 23 suggerisce un deterioramento moderato e un punteggio inferiore a 18 indica un deterioramento cognitivo grave. Va però ricordato che fattori come l'età, il livello di istruzione e la cultura del soggetto possono influenzare il punteggio e vanno quindi presi in considerazione nell'interpretazione dei risultati.

Il MMSE è ampiamente utilizzato non solo come strumento di screening iniziale per la demenza, ma anche per monitorare la progressione di malattie cognitive nel tempo o l'efficacia di interventi terapeutici. Tuttavia, sebbene sia uno strumento utile, non è esaustivo e non può da solo diagnosticare una specifica patologia cognitiva. Per questo, il MMSE è spesso integrato con altri esami neuropsicologici più approfonditi e con valutazioni cliniche più dettagliate.

Clock Drawing Test (CDT)

Il Clock Drawing Test (CDT) rappresenta un metodo rapido e efficace per valutare le abilità cognitive, in particolare le funzioni visuo-spaziali e le funzioni esecutive [11]. Questo test consiste nel chiedere al paziente di disegnare un orologio, includendo tutte le cifre e le lancette che indicano un'ora specifica. La valutazione del disegno avviene attraverso diversi criteri. Innanzitutto, viene esaminato il cerchio dell'orologio: deve essere simmetrico e ben definito per ricevere un punteggio. Successivamente, si considera la distribuzione dei numeri intorno al quadrante: devono

essere posizionati in modo corretto. Anche la correttezza dei numeri è fondamentale; ogni numero deve essere esatto e collocato nei punti appropriati, come ad esempio il 12 in alto e il 6 in basso. Un altro aspetto importante è la correttezza delle lancette, che devono essere chiare e proporzionate, con la lancetta delle ore visibilmente più corta rispetto a quella dei minuti. Inoltre, è cruciale che le lancette indichino l'ora richiesta con precisione. L'organizzazione complessiva del disegno deve risultare coerente e funzionale, e si deve tenere conto anche della presenza di errori o correzioni; ogni errore significativo comporta una penalizzazione nel punteggio.

L'interpretazione dei punteggi è abbastanza chiara: un punteggio di 3 o inferiore indica errori significativi, suggerendo un probabile declino cognitivo severo e la possibilità di demenza. Un punteggio compreso tra 4 e 5 suggerisce errori più gravi che possono indicare un deficit cognitivo moderato, mentre punteggi tra 6 e 8 mostrano errori minori, potenzialmente indicativi di un leggero declino cognitivo. Infine, un punteggio di 9 o 10 indica un disegno sostanzialmente corretto, senza evidenze di problemi cognitivi significativi.

È importante sottolineare che, sebbene il CDT risulti sensibile nel rilevare variazioni lievi nel declino cognitivo e si dimostri abbastanza preciso nella valutazione delle capacità visuospatiali ed esecutive, non può essere considerato uno strumento diagnostico autonomo. È fondamentale integrarlo con ulteriori test per confermare la presenza di deficit cognitivi e per una valutazione più completa delle abilità cognitive del paziente [30].

Geriatric Depression Scale (GDS)

Il GDS è uno strumento di valutazione comunemente utilizzato per misurare i sintomi della depressione negli anziani [13]. Sebbene non sia uno strumento specifico per la valutazione del declino cognitivo, è importante poiché la depressione negli anziani può influire sulle funzioni cognitive e portare a una condizione nota come pseudo-demenza depressiva, che può simulare il declino cognitivo.

Il GDS è un questionario auto-valutativo composto originariamente da 30 domande a cui il paziente risponde "sì" o "no", ma esiste anche una versione abbreviata con 15 domande (GDS-15), più rapida e facile da somministrare. Le domande sono focalizzate su aspetti emotivi, motivazionali e comportamentali, come: tristezza o scoraggiamento, interesse per le attività quotidiane, sensazioni di vuoto o solitudine, mancanza di energia, preoccupazioni per la propria salute. L'interpretazione dei punteggi risulta essere: da 0-4: nessun sintomo depressivo, da 5-8: depressione lieve, da 9-11 depressione moderata, mentre da 12-15: depressione grave.

La depressione negli anziani può spesso confondersi con il declino cognitivo, in quanto entrambe le condizioni possono presentare sintomi come ridotta concentrazione, difficoltà di memoria e rallentamento psicomotorio. La pseudo-demenza depressiva, invece, è una condizione in cui il declino cognitivo sembra presente, ma è in realtà causato da depressione non trattata. Il GDS può aiutare a differenziare la depressione da altre forme di deterioramento cognitivo. Il GDS è spesso utilizzato insieme a strumenti cognitivi come il Mini Mental State Examination (MMSE) per distinguere la depressione dal deterioramento cognitivo.

Behavioral and Psychological Symptoms of Dementia (BPSD)

I BPSD o sintomi comportamentali e psicologici della demenza, si riferiscono a una gamma di sintomi neuropsichiatrici che si manifestano nei pazienti con demenza. Questi sintomi non sono legati ai deficit cognitivi di base, come la perdita di memoria, ma riguardano problemi comportamentali ed emotivi che spesso accompagnano le malattie neurodegenerative, come il morbo di Alzheimer o altre forme di demenza [15].

I sintomi principali del BPSD sono variegati e complessi, influenzando significativamente la vita dei pazienti e di chi si prende cura di loro. Tra questi, *l'agitazione e l'aggressività* si manifestano spesso in modo evidente, con i pazienti che possono apparire irrequieti e talvolta mostrare comportamenti violenti o oppositivi. Molti pazienti con demenza sviluppano sintomi di *ansia e depressione*. Questi stati emotivi possono ulteriormente compromettere la loro capacità di vivere in modo indipendente. Altri sintomi significativi includono *deliri e allucinazioni*, che rappresentano manifestazioni psicotiche. I deliri consistono in false credenze che il paziente tiene con forza, mentre le allucinazioni possono manifestarsi come percezioni distorte della realtà, spesso di tipo visivo o uditivo. Il *vagabondaggio* è un altro comportamento frequentemente osservato, in cui i pazienti si allontanano o si perdono, creando preoccupazione nei loro familiari e caregiver. Infine, la *disinibizione* è un sintomo che implica una riduzione del controllo sugli impulsi, portando i pazienti a comportamenti socialmente inappropriati. Questo può includere azioni che possono essere percepite come offensive o inadeguate, complicando ulteriormente le interazioni sociali e le relazioni con gli altri.

I BPSD hanno un impatto significativo sulla gestione della demenza, poiché contribuiscono al carico assistenziale e riducono la qualità della vita sia dei pazienti che dei loro caregiver. La gestione di questi sintomi richiede spesso una combinazione di interventi farmacologici e non farmacologici, mirati alla riduzione dei sintomi senza sedare eccessivamente il paziente. Una scala dicotomica del BPSD potrebbe assegnare un punteggio 0 per l'assenza del sintomo e 1 per la presenza del sintomo.

Questo metodo è utile per un rapido screening iniziale dei sintomi, ma non fornisce informazioni sulla gravità o sulla frequenza.

Neuropsychiatric Inventory (NPI)

L'NPI è uno strumento clinico utilizzato per valutare una gamma di sintomi neuropsichiatrici comuni nelle persone affette da demenza, come ad esempio l'Alzheimer, il Parkinson, la demenza a corpi di Lewy e altre forme di deterioramento cognitivo. Questi sintomi includono disturbi comportamentali ed emotivi che spesso accompagnano il declino cognitivo e possono influenzare notevolmente la qualità della vita dei pazienti e dei caregiver [7].

L'NPI è composto da un'intervista strutturata condotta con il caregiver del paziente. Valuta 12 domini neuropsichiatrici principali, tra cui: deliri, allucinazioni, agitazione/aggressività, depressione/dismoralità, ansia, euforia/esaltazione, apatia/indifferenza, disinibizione, irritabilità/labilità, comportamenti motori anomali, disturbi del sonno, disturbi dell'appetito e della condotta alimentare. Per ogni dominio, il caregiver risponde a domande che identificano la presenza del sintomo, la sua gravità (da 1 a 3) e la frequenza (da 1 a 4), con un punteggio totale per ciascun dominio calcolato moltiplicando gravità per frequenza. Il punteggio totale dell'NPI può variare da 0 a 144. Valori più alti indicano una presenza più significativa di sintomi neuropsichiatrici e un impatto maggiore sulla vita quotidiana.

Il Neuropsychiatric Inventory (NPI) offre numerosi vantaggi nell'ambito della valutazione dei sintomi neuropsichiatrici. Ad esempio, la capacità di *valutare un'ampia gamma di sintomi*, inclusi quelli psicotici, affettivi e comportamentali, che spesso sfuggono ad altre scale cognitive più tradizionalmente focalizzate su aspetti come la memoria o l'orientamento. Questo lo rende particolarmente utile in contesti clinici, dove è fondamentale avere una visione globale del paziente. Un altro vantaggio significativo dell'NPI è l'*integrazione con il caregiver*, poiché il test viene condotto attraverso un'intervista che coinvolge la persona che assiste il paziente. Questa modalità consente di raccogliere informazioni dettagliate sui sintomi dal punto di vista di qualcuno che vive quotidianamente la realtà del paziente.

Tuttavia, l'NPI presenta anche alcune limitazioni. Una delle principali criticità è la *dipendenza dal caregiver* per le informazioni raccolte. Poiché il test si basa sulle osservazioni e sulle percezioni del caregiver, è possibile che la valutazione sia influenzata dalla soggettività e dal carico emotivo che quest'ultimo può portare. Inoltre, l'NPI si concentra principalmente sui sintomi neuropsichiatrici e *non valuta direttamente le funzioni cognitive*. Pertanto, per ottenere una visione complessiva delle capacità cognitive del paziente, potrebbe essere utile combinarlo con altre scale di valutazione come il Mini-Mental State Examination (MMSE). Questa integrazione permetterebbe di avere un quadro più completo delle condizioni

del paziente, affrontando sia gli aspetti comportamentali ed emotivi sia le funzioni cognitive.

Nel contesto del declino cognitivo l'NPI è particolarmente utile nelle persone con demenza perché i sintomi neuropsichiatrici sono molto comuni nelle fasi moderate e avanzate della malattia e possono includere problemi come l'aggressività, l'agitazione o il comportamento disinibito. Oltre al declino cognitivo, questi sintomi possono influenzare in modo significativo la gestione della malattia, la qualità della vita e il carico del caregiver.

Gli indici GDS, BPSD e l'NPI non misurano in maniera diretta il declino cognitivo globale, piuttosto, ciascuno di questi indici si concentra su aspetti specifici che possono essere associati al declino cognitivo.

1.2 Decadimento fisico

Il declino fisico nell'anziano si riferisce alla progressiva riduzione della riserva fisiologica, che limita la capacità del corpo di rispondere efficacemente a stress e sforzi fisici. Questa riduzione è il risultato dell'interazione tra invecchiamento naturale, malattie croniche e inattività di determinate funzioni o capacità fisiche. Il concetto di *stato funzionale* si riferisce alla capacità di svolgere attività quotidiane. Negli anziani, il declino della funzione fisica si manifesta nella difficoltà a compiere azioni di base (come camminare o vestirsi), evidenziando una perdita di autonomia nella vita di tutti i giorni [38].

Il declino funzionale negli anziani può essere misurato utilizzando una serie di strumenti e scale di valutazione. Questi indici si concentrano su aspetti come le attività della vita quotidiana (ADL), la mobilità e l'autonomia. Di seguito verranno spiegati in dettaglio alcuni degli indici più utilizzati.

Indice delle Attività della Vita Quotidiana (ADL) di Katz

L'indice ADL (Activities of Daily Living) è uno strumento essenziale per valutare il grado di autonomia di un paziente nelle attività quotidiane di base, come fare il bagno, vestirsi, mangiare e usare il bagno. Questo indice è utilizzato principalmente per determinare il livello di assistenza di cui una persona anziana o con disabilità può aver bisogno, specialmente in contesti di cura geriatrica o riabilitativa.

L'indice di Katz [16] è il più conosciuto tra gli strumenti ADL. Questo sistema assegna un punteggio che varia a seconda della capacità del paziente di eseguire sei attività fondamentali: fare il bagno, vestirsi, usare il bagno, trasferirsi (ad esempio, spostarsi dal letto a una sedia), continenza e alimentazione. Ogni attività

è valutata in termini di indipendenza o dipendenza. Un paziente che può svolgere autonomamente tutte queste attività ottiene il punteggio massimo, mentre la necessità di assistenza riduce il punteggio finale.

Questo strumento è estremamente utile non solo per i professionisti sanitari nel determinare il livello di assistenza necessario, ma anche per pianificare interventi riabilitativi e monitorare i progressi nel tempo. Permette di misurare con precisione il declino funzionale legato all'invecchiamento o a malattie croniche e degenerative, consentendo quindi di stabilire priorità di intervento nella cura del paziente.

Scala delle Attività Strumentali della Vita Quotidiana (IADL) di Lawton e Brody

La scala IADL (Instrumental Activities of Daily Living) valuta la capacità di una persona anziana di svolgere attività più complesse rispetto alle ADL (Activities of Daily Living). Queste attività richiedono competenze cognitive, motorie e sociali per interagire con l'ambiente circostante e gestire la vita quotidiana in modo autonomo. Le IADL includono: utilizzare il telefono, fare acquisti, preparare i pasti, gestire i farmaci, mantenere la casa, fare il bucato, utilizzare i trasporti e gestire le finanze. Il punteggio viene assegnato in base alla capacità di eseguire autonomamente ciascuna di queste attività [19].

Questa scala è importante perché permette di valutare non solo la dipendenza fisica del paziente, ma anche la sua autosufficienza e il grado di assistenza necessario per le attività che richiedono organizzazione e interazione con il contesto esterno. Risulta quindi particolarmente utile per determinare se l'anziano può continuare a vivere in modo indipendente o se necessita di supporto o supervisione. La perdita di funzionalità in una o più aree valutate dalla IADL è spesso uno dei primi segnali di declino funzionale nelle persone anziane, il che rende questo strumento fondamentale per l'individuazione precoce di problematiche legate all'autonomia.

1.3 Strategie di intervento

Le strategie di intervento per pazienti geriatrici che soffrono di declino fisico e/o cognitivo mirano a migliorare la qualità della vita, rallentare la progressione del decadimento e ridurre il rischio di complicanze. Questi interventi sono multidisciplinari e comprendono approcci medici, psicologici, sociali e fisici. Di seguito, vengono descritte alcune delle principali strategie terapeutiche.

Interventi farmacologici

Farmaci come gli inibitori della colinesterasi (ad esempio, Donepezil, Rivastigmina) e gli antagonisti del recettore NMDA (ad esempio, Memantina) sono utilizzati nel trattamento della malattia di Alzheimer e altri disturbi cognitivi. Questi farmaci mirano a migliorare la comunicazione tra le cellule cerebrali e rallentare il progressivo declino delle funzioni cognitive.

Gli *inibitori delle colinesterasi* sono farmaci che aumentano i livelli di acetilcolina, un neurotrasmettitore implicato nella memoria e nell'apprendimento, bloccando l'enzima acetilcolinesterasi che la degrada. Sono prescritti nelle fasi lievi-moderate della malattia di Alzheimer e di altre forme di demenza [2]. Essi sono: *donepezil*, indicato per demenza lieve, moderata e grave; *rivastigmina*, usata nella demenza associata a malattia di Alzheimer e a Parkinson e *galantamina*, utile nelle fasi iniziali e intermedie dell'Alzheimer. Questi farmaci possono temporaneamente migliorare o stabilizzare la memoria e la capacità di svolgere attività quotidiane, tuttavia, non fermano il progresso della malattia.

La *memantina* è un farmaco utilizzato principalmente nel trattamento della malattia di Alzheimer di grado moderato-severo. È un antagonista dei recettori NMDA (N-metil-D-aspartato), che sono implicati nella neurotrasmissione glutamatergica. Il glutammato è un neurotrasmettitore eccitatorio nel sistema nervoso centrale, ma un suo eccesso può provocare eccitotossicità e danneggiare le cellule cerebrali. La memantina, bloccando i recettori NMDA, aiuta a proteggere i neuroni dai danni causati da un'eccessiva stimolazione glutamatergica, senza però influenzare negativamente la normale trasmissione sinaptica.

La memantina ha un ruolo fondamentale soprattutto nella gestione dei pazienti geriatrici con demenza, in particolare quelli con malattia di Alzheimer [22]. Le caratteristiche di questo farmaco lo rendono particolarmente adatto ai pazienti anziani per vari motivi: *effetto neuroprotettivo*, riduce il danno neuronale associato alla neurodegenerazione; *miglioramento della funzionalità cognitiva*, sebbene non curi la demenza, può rallentare il declino delle funzioni cognitive, migliorando temporaneamente la qualità della vita e la capacità di svolgere attività quotidiane, inoltre, gode di una *buona tollerabilità*, gli anziani affetti da comorbidità tollerano generalmente bene la memantina, con effetti collaterali relativamente lievi rispetto ad altri farmaci. A volte, la memantina viene anche combinata con farmaci come il donepezil per gestire i sintomi della demenza, poiché hanno meccanismi d'azione differenti e complementari [24].

Gli *antipsicotici* e stabilizzatori dell'umore possono essere usati in alcuni pazienti con demenza, i quali manifestano sintomi comportamentali come agitazione, aggressività, psicosi o depressione. Questi farmaci devono essere utilizzati con cautela, poiché possono aumentare il rischio di eventi avversi, soprattutto negli anziani

[29]. I farmaci più utilizzati in questo settore sono: risperidone o quetiapina (antipsicotici atipici usati per i sintomi comportamentali), valproato o carbamazepina (usati come stabilizzatori dell'umore).

Gli *antidepressivi* possono essere prescritti se la depressione è presente nei pazienti con declino cognitivo, una condizione abbastanza comune che può esacerbare i sintomi cognitivi e comportamentali. I SSRI (inibitori selettivi della ricaptazione della serotonina) come il citalopram o la sertralina sono spesso preferiti poiché hanno un profilo di sicurezza favorevole negli anziani [27]. Infine, altri farmaci per il controllo di dolore cronico, ipertensione, diabete e osteoporosi sono essenziali per gestire le comorbidità che spesso accompagnano il declino fisico nei pazienti anziani [28].

Terapie non farmacologiche

Oltre ai farmaci, le terapie non farmacologiche giocano un ruolo cruciale nel trattamento del declino cognitivo. Questi approcci mirano a migliorare le capacità cognitive residue, ridurre lo stress, e mantenere le capacità funzionali il più a lungo possibile.

La *riabilitazione cognitiva* include una serie di interventi volti a mantenere e migliorare le funzioni cognitive [5], può includere: *training cognitivo* in cui si svolgono attività mentali strutturate che sfidano la memoria, l'attenzione e la risoluzione dei problemi e la *terapia della reminiscenza*, ovvero una tecnica che incoraggia il paziente a ricordare eventi del passato attraverso foto, musica o oggetti familiari, per stimolare la memoria autobiografica.

Le *terapie occupazionali* sono progettate per aiutare i pazienti a mantenere la capacità di svolgere le attività della vita quotidiana, come vestirsi, cucinare e gestire le proprie finanze. Questi interventi possono anche includere la modifica dell'ambiente domestico per ridurre i rischi e aumentare la sicurezza [12]. Inoltre, risultano essere molto utili anche approcci come le *terapie comportamentali*. Questi cercano di ridurre i sintomi psicologici e comportamentali, come l'agitazione, l'aggressività e l'insonnia, modificando l'ambiente e adattando le risposte ai comportamenti del paziente. L'approccio validante e la terapia di orientamento alla realtà sono esempi di tecniche usate per migliorare la comunicazione e il benessere emotivo del paziente [32].

La *terapia della stimolazione cognitiva* (CST) è una forma di terapia di gruppo che coinvolge attività strutturate per stimolare la memoria, la risoluzione dei problemi e l'interazione sociale. Ha dimostrato di avere effetti benefici su alcune funzioni cognitive e sul benessere generale nei pazienti con declino cognitivo lieve o moderato [31].

L'*attività fisica* regolare ha dimostrato di avere effetti positivi sulle funzioni cognitive e sul benessere complessivo. L'esercizio aerobico, come camminare o nuotare, può migliorare la circolazione cerebrale e promuovere la neuroplasticità [18]. Inoltre, alcuni studi suggeriscono che la *dieta mediterranea*, ricca di antiossidanti, acidi grassi omega-3 e altri nutrienti protettivi, possa essere utile nel ridurre il rischio di declino cognitivo e migliorare la salute cerebrale. Inoltre, la corretta gestione delle carenze vitaminiche, come la vitamina D e la B12, può essere importante nel rallentare il deterioramento cognitivo [23].

Infine, le *tecnologie assistive* possono aiutare i pazienti a ricordare gli eventi quotidiani e mantenere l'indipendenza. Queste includono dispositivi di promemoria elettronici, allarmi per la gestione dei farmaci e strumenti di localizzazione GPS per ridurre il rischio di vagabondaggio [9].

Supporto psicosociale e coinvolgimento della famiglia

La gestione del declino cognitivo richiede spesso il coinvolgimento della famiglia e dei caregiver. Il supporto psicosociale è essenziale sia per il paziente che per chi lo assiste. Offrire educazione, formazione e sostegno ai caregiver può ridurre lo stress familiare e migliorare la qualità dell'assistenza [4].

L'approccio al trattamento del declino cognitivo e fisico è complesso e richiede un intervento integrato che combina terapie farmacologiche e non farmacologiche. La gestione ottimale del paziente affetto da declino cognitivo prevede l'uso di farmaci come la memantina e gli inibitori delle colinesterasi, integrati da approcci non farmacologici come la riabilitazione cognitiva, la stimolazione cognitiva e il supporto psicosociale.

Capitolo 2

Presentazione dei dati

In questo capitolo si discuterà del dataset che ha ispirato il progetto di tesi, ponendo l'attenzione su vari aspetti, tra cui il metodo di raccolta dati, le particolarità dello studio e, soprattutto, le criticità riscontrate. Si spiegheranno nel dettaglio tutte le variabili rilevate e i criteri adottati per la loro selezione. Inoltre, essendo uno studio agli albori, una parte significativa del lavoro ha riguardato la pulizia del dataset, che sarà approfondita nelle sezioni successive. Questi passaggi sono stati cruciali per garantire l'accuratezza e l'affidabilità dei dati raccolti, preparando il dataset alla successiva fase di analisi.

L'obiettivo principale di questo studio è identificare se esistono gruppi di soggetti che, a causa di determinate caratteristiche, mostrano un deterioramento cognitivo più rapido o marcato rispetto ad altri. Se tale ipotesi venisse confermata, sarebbe possibile intervenire con trattamenti immediati e mirati per nuovi pazienti che presentano caratteristiche simili a quelle dei soggetti studiati, migliorando così l'efficacia e la tempestività delle cure fornite.

2.1 Configurazione sperimentale

Come discusso nella sezione precedente, numerose sono le problematiche che colpiscono le persone in età avanzata, non solo a livello fisico, ma anche cognitivo e psicologico. Con l'aumento dell'aspettativa di vita, sempre più individui si trovano a dover affrontare tali complicazioni, spesso in modo simultaneo. Proprio per queste ragioni, i medici del reparto di geriatria dell'ospedale di Padova hanno avviato un complesso disegno sperimentale, focalizzato su uomini e donne di età superiore ai sessantacinque anni, ciascuno con una specifica problematica.

Il progetto in questione è iniziato nel 2015 ed è tuttora in corso, tuttavia i dati trattati in questo elaborato si riferiscono al periodo 2015-2024. Si tratta di uno studio longitudinale, in cui i pazienti vengono seguiti per un periodo prolungato

e, a ogni visita, vengono raccolti per ciascuno di loro un determinato numero di variabili.

La raccolta dei dati è iniziata con una prima visita, denominata "visita conoscitiva". Durante questa fase, sono state rivolte ai pazienti alcune domande standard e sono stati somministrati test per valutare il loro stato di salute psico-fisico. In seguito, si sono svolte ulteriori visite, definite "visite di controllo", in cui sono stati ripetuti gli stessi test effettuati durante la visita precedente, permettendo così di monitorare l'andamento della malattia, sia in termini di peggioramento che di miglioramento. Complessivamente, sono state registrate quattordici visite di controllo.

Una peculiarità di questo studio è che né la prima visita, né le visite di controllo avvengono nello stesso giorno per tutti i pazienti. Ogni paziente ha la possibilità di scegliere la data della visita, purché questa avvenga ad almeno 30 giorni di distanza dalla visita precedente. Inoltre, è responsabilità del paziente presentarsi regolarmente alle visite di controllo. In alcuni casi, i pazienti potrebbero abbandonare lo studio, decidendo di non partecipare più alle visite senza fornire motivazioni specifiche.

La ricerca è iniziata con un campione di 615 pazienti che hanno partecipato alla visita conoscitiva, mentre solo un paziente ha effettuato tutte e quattordici le visite di controllo. La Tabella 2.1 riporta, per ciascuna visita, il numero di soggetti che vi hanno partecipato:

Visita	n(%)	Visita	n(%)
Prima visita	615(100)	VIII controllo	61(9.92)
I controllo	381(61.95)	IX controllo	51(8.29)
II controllo	258(41.95)	X controllo	38(6.18)
III controllo	189(30.73)	XI controllo	29(4.72)
IV controllo	167(27.15)	XII controllo	13(2.11)
V controllo	124(20.16)	XIII controllo	2(0.33)
VI controllo	92(14.96)	XIV controllo	1(0.16)
VII controllo	76(12.36)		

Tabella 2.1: Numero di pazienti per visita.

2.1.1 Descrizione delle variabili fisse

Le variabili che appartengono a questa categoria vengono rilevate una sola volta alla prima visita e si riferiscono a caratteristiche del soggetto che non possono modificarsi nel corso del tempo. Le variabili fisse possono essere distinte in quattro macro-insiemi: anagrafica generale, APR (Anamnesi Patologica Remota), fattori di rischio e diagnostica. Di seguito, verranno analizzate nel dettaglio le variabili appartenenti a ciascuno di questi gruppi.

Anagrafica generale

L'anagrafica generale si riferisce alle informazioni personali di base, necessarie per l'identificazione e la contestualizzazione dell'individuo. Le variabili che rientrano in questo gruppo sono riportate nella Tabella 2.2.

Nome Variabile	Descrizione
Data di nascita	Variabile che indica la data di nascita del paziente in formato AAAA-MM-GG.
Genere	Variabile categoriale che indica il genere del paziente e assume valore 0 se il paziente è uomo, 1 se il paziente è donna.
Stato civile	Variabile categoriale che indica lo stato civile del paziente: 0 per celibe/nubile, 1 per coniugato/a, 2 per divorziato/a, 3 per vedovo/a.
Titolo di studio	Variabile categoriale che indica il titolo di studio del paziente: 0 per nessuno, 1 per elementare, 2 per medie inferiori, 3 per medie superiori, 4 per professionali, 5 per laurea, 6 per master/specializzazione.
Scolarità	Variabile quantitativa discreta che indica il numero di anni di scolarizzazione del paziente, da 1 a 19 anni.
Professione	Variabile categoriale che indica la professione del paziente: 1 per agricoltore, 2 per artigiano, 3 per casalinga, 4 per commerciante, 5 per impiegato, 6 per imprenditore, 7 per insegnante, 8 per professionista, 9 per operaio, 10 per altro.

Tabella 2.2: Descrizione delle variabili in anagrafica generale.

APR (Anamnesi Patologica Remota)

Questo insieme di variabili raccoglie le informazioni riguardanti la storia clinica e le malattie pregresse del paziente. Queste informazioni sono cruciali per comprendere lo stato di salute generale. Tutte queste variabili si trovano nella Tabella 2.3 e sono

di tipo dicotomico, ovvero assumono valore 1 se la malattia è presente, 0 se non è presente.

Nome Variabile	
IPA (Ipertensione Arteriosa Polmonare)	Ictus ischemico
DM (Diabete Mellito)	Ictus emorragico
FA (Fibrillazione Atriale)	Vasculopatia periferica
IMA (Infarto Miocardico Acuto)	Depressione
TIA (Attacco Ischemico Transitorio)	Patologia psichiatrica
Epilessia	Parkinson
Altre patologie neurodegenerative	Neoplasie
Ipertiroidismo	Ipotiroidismo
Patologie gastrointestinali	Altro

Tabella 2.3: Elenco delle patologie.

Fattori di rischio

I fattori di rischio in Tabella 2.4 rappresentano caratteristiche o comportamenti che possono aumentare la probabilità di sviluppare determinate malattie. Le variabili incluse in questo gruppo sono essenziali per la prevenzione e gestione delle patologie.

Nome Variabile	Descrizione
Ipoacusia	Variabile categoriale che assume valore 0 se non è presente, 1 se è presente ed è stata corretta, 2 se è presente, ma non è stata corretta.
Storia di traumi cranici	Variabile dicotomica che assume valore 0 se non è presente, 1 se è presente.
Alcool	Variabile categoriale che assume valore 0 se non si assume alcool, 1 se si assume alcool durante i pasti, 2 se si fa un consumo attivo di alcool, 3 se si è fatto un uso pregresso di alcool.
Fumo	Variabile categoriale che assume valore 0 se non si fuma, 1 se si fuma, 2 se si è un ex fumatore.
Attività fisica	Variabile categoriale che assume valore 0 se non viene praticata, 1 se viene praticata meno di due volte la settimana, 2 se viene praticata più di due volte a settimana.

Tabella 2.4: Descrizione dei fattori di rischio.

Diagnosi

Le variabili diagnostiche riportate in Tabella 2.5 includono i risultati di esami e test clinici che forniscono informazioni dettagliate sullo stato attuale di salute del paziente. Questi dati diagnostici offrono un quadro completo delle condizioni cliniche, contribuendo a identificare eventuali patologie, a valutare l'efficacia dei trattamenti e a monitorare l'evoluzione dello stato di salute nel tempo.

Variabile	Descrizione
Diagnosi	Variabile categoriale che può assumere i seguenti valori: 1 per MCI (Mild Cognitive Impairment), 2 per AD (Alzheimer's disease), 3 per VD (Vascular disease), 4 per MD (Mixed dementia), 5 per PD (Parkinson's disease), 6 per DLB (Dementia with Lewy bodies), 7 per PSP (Progressive Supranuclear Palsy), 8 per FTD (Frontotemporal dementia).
Valutazione NPS	Variabile dicotomica che assume valore 0 se non è stata effettuata, 1 altrimenti.

Tabella 2.5: Descrizione delle variabili relative alla diagnosi e valutazione neuropsicologica.

2.1.2 Descrizione delle variabili temporali

Le variabili che appartengono a questa categoria vengono rilevate sia alla prima visita, ma anche a tutte le successive visite di controllo. Di conseguenza, il loro valore nel corso del tempo può modificarsi. Le variabili temporali possono essere suddivise in tre macro-categorie: raccordo visite, terapia consigliata e test di valutazione fisica e cognitiva. Di seguito, verrà fornita un'analisi dettagliata delle variabili appartenenti a ciascuno di questi gruppi.

Raccordo visite

La Tabella 2.6 comprende tutte le variabili relative alla pianificazione e alla frequenza delle visite mediche. In questo contesto, si analizzano aspetti come l'intervallo tra le visite, la cadenza temporale degli appuntamenti e le terapie in corso al momento della visita. Tali variabili sono fondamentali per comprendere come la tempistica delle visite possa influenzare il decorso della malattia o la risposta alla terapia.

Variabile	Descrizione
Data visita/controllo	Variabile di tipo "Data" che specifica l'anno, il mese e il giorno in cui il paziente ha effettuato la visita.
Tempo trascorso in mesi dalla prima visita	Variabile continua discreta che indica il tempo trascorso dalla prima visita.
Età	Variabile continua rilevata automaticamente che indica l'età del paziente in quella determinata visita.
Stato coabitativo	Variabile categoriale che può assumere i seguenti valori: 0 per solo, 1 per con coniuge, 2 per con figli/parenti, 3 per badante, 4 per istituzionalizzato/a.
Stimolazione cognitiva	Variabile dicotomica che assume valore 1 se la stimolazione è in atto, 0 altrimenti.
Terapia per demenza già in atto	Variabile dicotomica che assume valore 0 se la terapia non è in atto, 1 altrimenti.
Ache-I in atto	Variabile categoriale che assume i seguenti valori: 0 per nessuno, 1 per donepezil, 2 per rivastigmina, 3 per galantamina.
Memantina in atto	Variabile dicotomica che assume valore 1 se il paziente ha assunto la memantina dalla visita precedente, 0 altrimenti.
Antipsicotici in atto	Variabile categoriale che assume i seguenti valori: 0 per nessuno, 1 per aloperidolo, 2 per risperidone, 3 per quetiapina, 4 per olanzapina, 5 per promazina, 6 per altro.
Antidepressivi in atto	Variabile categoriale che assume i seguenti valori: 0 per nessuno, 1 per sertralina, 2 per mirtazapina, 3 per escitalopram/citalopram, 4 per venlafaxina, 5 per duloxetina, 6 per paroxetina, 7 per altro.
Trazodone in atto	Variabile dicotomica che assume valore 1 se il paziente ha assunto il trazodone dalla visita precedente, 0 altrimenti.
Benzodiazepine in atto	Variabile dicotomica che assume valore 1 se il paziente ha assunto il benzodiazepine dalla visita precedente, 0 altrimenti.

Tabella 2.6: Descrizione delle variabili cliniche raccolte durante le visite di controllo.

Terapia consigliata

La Tabella 2.7 include variabili legate alla tempistica dell'inizio, della modifica o dell'interruzione del trattamento terapeutico. In questa categoria rientrano la prescrizione dei farmaci e i cambiamenti nel regime terapeutico. Analizzare queste variabili permette di valutare l'effetto temporale delle terapie consigliate e il loro adattamento nel corso del tempo alle esigenze cliniche dei pazienti.

Variabile	Descrizione
Ache-I prescritto	Variabile categoriale che assume i seguenti valori a seconda della tipologia di terapia Ache-I prescritta: 0 per nessuno, 1 per donepezil, 2 per rivastigmina, 3 per galantamina.
Controindicazioni Ache-I	Variabile categoriale che assume i seguenti valori a seconda delle controindicazioni che possono insorgere a seguito dell'assunzione di Ache-I: 0 per Nessuna controindicazione, 1 per MMSE, 2 per cardiologiche, 3 per VD, 4 per intolleranza gastrointestinale, 5 per intolleranza cardiovascolare, 6 per intolleranza neurologica, 7 per altra intolleranza.
Memantina prescritta	Variabile dicotomica che assume valore 0 se in quella visita la memantina non è stata prescritta, 1 altrimenti.
Controindicazioni memantina	Variabile categoriale che assume i seguenti valori a seconda delle controindicazioni che possono insorgere a seguito dell'assunzione di memantina: 0 per Nessuna controindicazione, 1 per MMSE, 2 per IRC (Insufficienza renale cronica), 3 per altro.
Antipsicotici prescritti	Variabile categoriale che assume i seguenti valori a seconda della tipologia di antipsicotici prescritti: 0 per nessuno, 1 per aloperidolo, 2 per risperidone, 3 per quetiapina, 4 per olanzapina, 5 per promazina, 6 per altro.
Antidepressivi prescritti	Variabile categoriale che assume i seguenti valori a seconda della tipologia di antidepressivi prescritti: 0 per nessuno, 1 per sertralina, 2 per mirtazapina, 3 per escitalopram/citalopram, 4 per venlafaxina, 5 per duloxetina, 6 per paroxetina, 7 per altro.
Trazodone prescritto	Variabile dicotomica che assume valore 0 se in quella visita il trazodone non è stato prescritto, 1 altrimenti.

Variabile	Descrizione
Benzodiazepine prescritte	Variabile dicotomica che assume valore 0 se in quella visita il benzodiazepine non è stato prescritto, 1 altrimenti.
Integratori prescritti	Variabile dicotomica che assume valore 0 se in quella visita non è stato prescritto alcun integratore, 1 altrimenti.

Tabella 2.7: Descrizione delle variabili relative alle terapie prescritte.

Test di valutazione fisica e cognitiva

La Tabella 2.8 raccoglie tutte le variabili relative alla somministrazione dei test diagnostici e valutativi, inclusi test cognitivi ed esami fisici. La testistica è cruciale per monitorare l'evoluzione dello stato di salute del paziente e la risposta alle terapie.

Variabile	Descrizione
ADL (Activity of Daily Living)	Variabile quantitativa discreta con un range da 0 a 6, dove 0 indica che il paziente non è in grado di compiere nessuna delle attività quotidiane e 6 indica che è in grado di compierle tutte.
IADL (Instrumental Activity of Daily Living)	Variabile quantitativa discreta con un range da 0 a 8, dove 0 indica che il paziente non è in grado di compiere nessuna delle attività strumentali della vita quotidiana, mentre 8 indica che è in grado di compierle tutte.
IADL max	Variabile dicotomica assegnata automaticamente che vale 5 se il paziente è un uomo, 8 se è una donna.
IADL %	Variabile quantitativa continua calcolata come $IADL/IADL \text{ max}$.
MMSE orientamento temporale	Variabile quantitativa discreta con un range da 0 a 5, dove 0 indica che il paziente non ha saputo rispondere correttamente a nessuna delle domande.
MMSE orientamento spaziale	Variabile quantitativa discreta con un range da 0 a 5, dove 0 indica che il paziente non ha saputo rispondere correttamente a nessuna delle domande.
MMSE memoria verbale breve	Variabile quantitativa discreta con un range da 0 a 3, dove 0 indica che il paziente non ha saputo rispondere correttamente a nessuna delle domande.

Variabile	Descrizione
MMSE attenzione e calcolo	Variabile quantitativa discreta con un range da 0 a 5, dove 0 indica che il paziente non ha saputo rispondere correttamente a nessuna delle domande.
MMSE richiamo delle tre parole	Variabile quantitativa discreta con un range da 0 a 3, dove 0 indica che il paziente non ha saputo rispondere correttamente a nessuna delle domande.
MMSE linguaggio	Variabile quantitativa discreta con un range da 0 a 3, dove 0 indica che il paziente non ha saputo rispondere correttamente a nessuna delle domande.
MMSE esecuzione di un compito su comando	Variabile quantitativa discreta con un range da 0 a 5, dove 0 indica che il paziente non ha saputo eseguire correttamente nessun comando.
MMSE abilità prassico costruttive	Variabile dicotomica, dove 1 indica che il paziente ha eseguito bene il compito assegnatogli, 0 altrimenti.
MMSE punteggio grezzo	Variabile quantitativa discreta con un range da 0 a 30, calcolata come la somma di tutti gli MMSE precedenti.
MMSE punteggio corretto	Variabile quantitativa continua, calcolata partendo dal MMSE grezzo e aggiustata per Età e Istruzione, secondo un'opportuna tabella.
CDT (Clock Drawing Test)	Variabile quantitativa discreta che prende valori da 0 a 10.
NPI (Neuropsychiatric Inventory)	Variabile quantitativa continua il cui totale è formato da due parti: la presenza/assenza di disturbi (variabili che assumono valore 0 o 1) e comprende: Deliri, Allucinazioni, Agitazione, Depressione, Ansia, Euforia, Apatia, Disinibizione, Attività motoria, Disturbi del sonno, Disturbi dell'appetito, moltiplicata per la gravità di questi. Range da 0 a 68.
BPSD (Behaviour and Psychological Symptoms of Dementia)	Variabile dicotomica che prende valore 1 se sono presenti disturbi legati alla demenza, 0 altrimenti.
GDS (Geriatric Depression Scale)	Variabile quantitativa discreta che prende valori da 0 a 15.

Tabella 2.8: Descrizione delle variabili relative a ADL, IADL e valutazioni cognitive.

Si vede come in questo studio sono stati utilizzati molti degli indici del capitolo 1. In particolare, sono stati ripresi l'indice ADL (Activity of Daily Living) e IADL (Instrumental Activity of Daily Living) per testare lo stato fisico dei pazienti. Per monitorare il deterioramento/miglioramento cognitivo, invece, sono stati utilizzati l'MMSE (Mini-Mental State Examination), l'NPI (Neuropsychiatric Inventory), il BPSD (Behaviour and Psychological Symptoms of Dementia), il CDT (Clock Drawing Test) e il GDS (Geriatric Depression Scale)

Le terapie impiegate in questo studio sono allineate con quanto indicato dalla letteratura scientifica. In particolare, sono stati utilizzati farmaci come la memantina, un antagonista del recettore NMDA, e gli inibitori della colinesterasi (Ache-I) come il donepezil, la rivastigmina e la galantamina. Inoltre, sono stati impiegati il risperidone e la quetiapina come antipsicotici, insieme a vari antidepressivi. Queste scelte terapeutiche riflettono un approccio consolidato nella gestione dei disturbi neurodegenerativi e psichiatrici associati come visto nel paragrafo 1.3.

2.2 Pulizia dei dati

La pulizia dei dati è una fase cruciale nel processo di preparazione del dataset per l'analisi. Questo passaggio ha lo scopo di garantire che i dati siano coerenti, completi e privi di errori, in modo da ottenere risultati affidabili e significativi. Nel contesto della presente ricerca, sono stati adottati diversi approcci per affrontare le problematiche più comuni che emergono nei dati raccolti.

Identificazione e gestione dei valori mancanti

Per ogni paziente si è visto se fosse presente a ciascuna delle 14 visite di controllo. Qualora non fosse stato presente, si sono poste come valori mancanti, ovvero "NA", tutte le variabili relative a quella visita per quel determinato paziente. Si è ritenuto opportuno fare ciò in quanto molti pazienti non presenti a una visita riportavano comunque dei valori per alcune delle variabili.

Alcune variabili come "Valutazione NPS", "GDS" (Geriatric depression Scale), "CDT" (Clock Drawing Test) e "NPI" (Neuropsychiatric Inventory), presentano un'alta percentuale di soggetti a cui non sono stati somministrati i relativi test, per questo si decide di non considerare queste variabili per le analisi.

Alle variabili categoriali: "Terapia per demenza già in atto", "Ache-I in atto", "Ache-I prescritti", "Memantina in atto", "Memantina prescritta", "Antipsicotici in atto", "Antipsicotici prescritti", "Antidepressivi in atto", "Antidepressivi prescritti", "Trazodone in atto", "Trazodone prescritto", "BDZ in atto", "BDZ prescritto", "Integratori prescritti", "Stimolazione cognitiva", è stata aggiunta la

modalità 0 che sta a indicare "No/Nessuno". Di questa categoria fanno parte tutti i pazienti presenti a una determinata visita, ma che riportavano in una o più di queste variabili valore nullo (NA).

Inoltre, alle variabili categoriali: "Controindicazioni Memantina" e "Controindicazioni Ache-I", è stato aggiunto un livello nominato "Nessuna controindicazione", in cui sono stati inseriti tutti i pazienti che erano presenti alla visita, ma riportavano valore nullo (NA).

Per quanto riguarda le variabili quantitative ordinali presenti nella categoria "Test di valutazione fisica e cognitiva", è stata introdotta una nuova modalità denominata "NS", che sta per "Non Somministrato". Questa modalità si riferisce a quei pazienti che, pur avendo effettuato la prima visita o una visita di controllo, non hanno effettuato uno o più dei test previsti per motivi specifici, come ad esempio l'incapacità di eseguirli o altre ragioni particolari.

Controllo di coerenza tra variabili

È stata condotta una revisione incrociata delle variabili per identificare eventuali incoerenze. In particolare, sono state verificate le date relative alle visite, poiché in molti casi i pazienti avevano la stessa data per due visite diverse, oppure le visite risultavano troppo ravvicinate. Inoltre, è stato rilevato che, in alcuni casi, la data della visita successiva era antecedente a quella della visita precedente.

Inoltre, è stata verificata attentamente la coerenza tra la prescrizione di farmaci e i dati relativi alle visite successive. In particolare, se a un paziente è stato prescritto un determinato farmaco (ad esempio, memantina, antidepressivi, ecc.) in una visita, il valore corrispondente nella variabile relativa (ad esempio, "memantina in atto") dovrebbe essere pari a 1 nella visita successiva. Tuttavia, in alcuni casi, questa relazione può non verificarsi, ad esempio quando il paziente ha manifestato effetti avversi al farmaco, con conseguente sospensione della terapia. Queste discrepanze sono state prese in considerazione e corrette dove necessario.

Gestione della variabile d'interesse per il decadimento cognitivo

Infine, è stata condotta una revisione approfondita della variabile di interesse MMSE (Mini-Mental State Examination). In particolare, la variabile "MMSE punteggio grezzo" è stata ricalcolata sommando i punteggi di tutte le singole categorie del MMSE. Successivamente, la variabile "MMSE punteggio corretto" è stata adeguata partendo dal punteggio grezzo, applicando una tabella di correzione specifica basata su età e livello di istruzione del paziente, al fine di ottenere una valutazione più accurata e standardizzata.

Tuttavia, per le analisi è stato deciso di utilizzare il punteggio grezzo del MMSE come variabile di interesse e includere il livello di istruzione e l'età come variabili esplicative. Questa scelta è stata motivata dal fatto che la tabella di correzione, essendo datata, non riesce a eliminare completamente l'influenza dell'istruzione e dell'età sul punteggio, compromettendo così la capacità di depurare del tutto l'effetto di queste variabili.

Come si può intuire da quanto detto fino a ora, questo dataset presenta diverse criticità, le quali hanno influito in modo significativo sulla scelta del metodo di analisi. Una delle particolarità più rilevanti è che il numero di visite effettuate varia da paziente a paziente, rendendo il dataset irregolare per quanto riguarda la raccolta delle osservazioni. Inoltre, l'intervallo di tempo tra una visita e la successiva non è costante e differisce per ciascun individuo, complicando ulteriormente la struttura dei dati.

Un altro aspetto cruciale è che, sebbene alcuni pazienti risultino presenti alla visita programmata, in determinate occasioni non è stato loro somministrato il test MMSE, creando lacune nei dati relativi a questa misura. A ciò si aggiunge il fatto che alcune variabili, come l'assunzione di memantina o l'età, cambiano nel corso del tempo, introducendo una componente di variabilità longitudinale che deve essere adeguatamente considerata nell'analisi. Questi elementi, presi nel loro insieme, hanno reso necessaria l'adozione di un approccio flessibile e robusto, capace di gestire la complessità e la disomogeneità intrinseche ai dati.

Capitolo 3

Modelli per studi longitudinali

Gli studi longitudinali sono progettati per esaminare i cambiamenti nel tempo di una caratteristica misurata ripetutamente per ciascun partecipante. In ambito medico, tali misurazioni possono riguardare parametri come la pressione sanguigna o il livello di colesterolo. A causa della variabilità nella frequenza e nei tempi delle misurazioni tra i partecipanti, i dataset risultano spesso sbilanciati, complicando l'uso di modelli multivariati generali.

Un'alternativa comune è l'uso di modelli a due stadi, che considerano la variabilità individuale tramite *effetti casuali*. Questi modelli non richiedono dati equilibrati e permettono di modellare sia la variazione tra individui che all'interno degli stessi, rendendoli particolarmente adatti per gestire dataset complessi e sbilanciati. Tuttavia, rispetto ai modelli multivariati generali, impongono una struttura di covarianza più rigida, ma offrono stime più efficienti. Questo approccio è utile per studiare gli effetti di variabili come età o trattamenti sui dati longitudinali, facilitando l'interpretazione delle differenze individuali [17].

3.1 Modello Lineare a Effetti Misti

I modelli lineari a effetti misti (LME) sono una generalizzazione dei modelli lineari che combinano sia effetti fissi che effetti casuali. Gli *effetti fissi* rappresentano relazioni che si applicano a tutto il campione o a tutta la popolazione. Sono simili ai coefficienti nei modelli lineari tradizionali e si assumono costanti per tutte le osservazioni. Invece, gli *effetti casuali* rappresentano variabili o coefficienti che variano tra gruppi o unità e vengono introdotti per modellare la variabilità che non può essere spiegata completamente dagli effetti fissi, ma è specifica di ciascun gruppo. I modelli LME si usano quando i dati presentano una struttura gerarchica o raggruppata (ad esempio, misurazioni ripetute su individui), per tener conto delle variazioni sia tra gruppi che all'interno dei gruppi.

3.1.1 Specificazione del modello

Sia \mathbf{y}_i il vettore n_i -dimensionale del punteggio MMSE per il paziente i -esimo, con $i = 1, \dots, 235$, il modello lineare ad effetti misti descritto da Laird e Ware (1982) [17] risulta essere

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i \quad i = 1, \dots, 235, \quad (3.1)$$

con

$$\mathbf{e}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad e \quad \mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}), \quad (3.2)$$

dove, in generale, $\boldsymbol{\beta}$ rappresenta il vettore p -dimensionale degli effetti fissi, \mathbf{b}_i è il vettore q -dimensionale degli effetti casuali, \mathbf{X}_i (di dimensione $n_i \times p$) e \mathbf{Z}_i (di dimensione $n_i \times q$) sono le matrici di regressione note, rispettivamente per gli effetti fissi e per gli effetti casuali. Infine, \mathbf{e}_i è il vettore n_i -dimensionale dell'errore intra-gruppo, che si assume distribuito secondo una Gaussiana sferica. L'ipotesi di base, $\text{Var}(\mathbf{e}_i) = \sigma^2 \mathbf{I}$, può essere rilassata, come verrà mostrato nel paragrafo 3.3.1. In tale paragrafo verrà descritta un'estensione che consente di modellare varianze non costanti, con una particolare struttura di correlazione all'interno del gruppo.

Più nello specifico, la matrice degli effetti fissi \mathbf{X}_i includerà le caratteristiche di base, invarianti nel tempo, del paziente i -esimo, come ad esempio età, sesso o altre informazioni cliniche. La matrice degli effetti casuali \mathbf{Z}_i , invece, sarà di dimensione $n_i \times 2$ poiché verranno inclusi sia il tempo che il trattamento principale considerato nello studio, ovvero la memantina. In particolare, il tempo avrà n_i valori distinti per ciascun paziente, mentre lo stato della memantina potrà variare tra 0 e 1, o viceversa, in qualsiasi istante temporale. Introducendo la memantina e il tempo come effetti casuali, si terrà conto delle variazioni individuali dei pazienti in relazione a questo fattori.

Inoltre, si assume che gli effetti casuali \mathbf{b}_i e gli errori intra-gruppo \mathbf{e}_i siano indipendenti sia tra gruppi diversi, sia all'interno dello stesso gruppo. Poiché si ipotizza che i vettori degli effetti casuali \mathbf{b}_i seguano una distribuzione normale con media $\mathbf{0}$, la loro distribuzione è completamente determinata dalla matrice di varianza-covarianza $\boldsymbol{\Psi}$. Questa matrice deve essere simmetrica e semi-definita positiva, ossia deve avere autovalori non negativi. Tuttavia, si adotterà l'ipotesi più restrittiva che $\boldsymbol{\Psi}$ sia definita positiva, ovvero con autovalori strettamente maggiori di zero.

Per facilitare la stima del modello, si esprimerà spesso la matrice di varianza-covarianza sotto forma di un fattore di precisione relativo, $\boldsymbol{\Delta}$, ovvero

$$\sigma^2 \boldsymbol{\Psi}^{-1} = \boldsymbol{\Delta}^T \boldsymbol{\Delta}. \quad (3.3)$$

L'algoritmo che verrà utilizzato per stimare il modello, adotterà la fattorizzazione log-Cholesky per $\boldsymbol{\Delta}$ [34].

3.2 Stima di massima verosimiglianza per i modelli LME

Sono stati sviluppati diversi metodi per la stima dei parametri nei modelli lineari a effetti misti. Tra questi, ci si concentrerà in particolare sulla stima di massima verosimiglianza ristretta (REML) [25], poiché sarà impiegata per la stima del modello nel capitolo 4. Tuttavia, prima di approfondire la REML, è fondamentale introdurre il concetto generale di stima di massima verosimiglianza, poiché rappresenta la base su cui si costruisce l'approccio REML.

La funzione di verosimiglianza LME a un singolo livello

Si considera innanzitutto il modello (3.1) con un singolo livello di effetti casuali. I parametri del modello sono $\boldsymbol{\beta}$, σ^2 e quelli che determinano $\boldsymbol{\Delta}$. Si indica con $\boldsymbol{\theta}$ l'insieme dei parametri non vincolati che definiscono $\boldsymbol{\Delta}$. La funzione di verosimiglianza associata al modello (3.1) rappresenta la densità di probabilità dei dati condizionatamente ai parametri, ma viene trattata come una funzione dei parametri. Si ha

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) = p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) \quad (3.4)$$

dove L è la verosimiglianza, p è la densità di probabilità e \mathbf{y} è l'intero vettore di risposta N -dimensionale, $N = \sum_{i=1}^M n_i$. Poiché gli effetti casuali non osservabili \mathbf{b}_i , con $i = 1, \dots, M$, fanno parte del modello, per ottenere la densità marginale dei dati è necessario integrare la densità condizionata dei dati, data la presenza degli effetti casuali, rispetto alla densità marginale degli effetti casuali. Si può utilizzare l'indipendenza di \mathbf{b}_i e \mathbf{e}_i per esprimere la verosimiglianza come

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) &= \prod_{i=1}^M p(\mathbf{y}_i | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) \\ &= \prod_{i=1}^M \int p(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) p(\mathbf{b}_i | \boldsymbol{\theta}, \sigma^2) d\mathbf{b}_i, \end{aligned} \quad (3.5)$$

dove la densità condizionale di \mathbf{y}_i è una normale multivariata del tipo

$$p(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) = \frac{\exp(-\|\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_i \mathbf{b}_i\|^2 / 2\sigma^2)}{(2\pi\sigma^2)^{n_i/2}} \quad (3.6)$$

e la densità marginale di \mathbf{b}_i è anch'essa una normale multivariata del tipo

$$p(\mathbf{b}_i | \boldsymbol{\theta}, \sigma^2) = \frac{\exp(-\mathbf{b}_i^T \boldsymbol{\Psi}^{-1} \mathbf{b}_i)}{(2\pi)^{q/2} \sqrt{|\boldsymbol{\Psi}|}} = \frac{\exp(-\|\boldsymbol{\Delta} \mathbf{b}_i\|^2 / 2\sigma^2)}{(2\pi\sigma^2)^{q/2} \text{abs}|\boldsymbol{\Delta}|^{-1}}, \quad (3.7)$$

dove $|\mathbf{A}|$ denota il determinante della matrice \mathbf{A} e abs rappresenta il valore assoluto. Sostituendo le formule (3.6) e (3.7) in (3.5), la verosimiglianza diventa

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) &= \prod_{i=1}^M \frac{\text{abs}|\boldsymbol{\Delta}|}{(2\pi\sigma^2)^{n_i/2}} \int \exp \left[-\frac{(\|\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\|^2 + \|\boldsymbol{\Delta}\mathbf{b}_i\|^2)/2\sigma^2}{(2\pi\sigma^2)^{q/2}} \right] d\mathbf{b}_i \\ &= \prod_{i=1}^M \frac{\text{abs}|\boldsymbol{\Delta}|}{(2\pi\sigma^2)^{n_i/2}} \int \frac{\exp \left(-\|\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i\boldsymbol{\beta} - \tilde{\mathbf{Z}}_i\mathbf{b}_i\|^2/2\sigma^2 \right)}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i, \end{aligned} \quad (3.8)$$

dove

$$\tilde{\mathbf{y}}_i = \begin{bmatrix} \mathbf{y}_i \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\mathbf{X}}_i = \begin{bmatrix} \mathbf{X}_i \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\mathbf{Z}}_i = \begin{bmatrix} \mathbf{Z}_i \\ \boldsymbol{\Delta} \end{bmatrix} \quad (3.9)$$

rappresentano rispettivamente i vettori di dati aumentati e le matrici del modello. Questo metodo, che modifica il contributo della distribuzione marginale degli effetti casuali aggiungendo righe extra per la risposta e le matrici di disegno, è noto come approccio *pseudo-dati* perché simula l'effetto della distribuzione marginale inserendo osservazioni fittizie.

Valutazione della verosimiglianza attraverso la decomposizione QR

Come prima cosa si prende una decomposizione ortogonale-triangolare della matrice del modello aumentato (3.9), ovvero

$$\tilde{\mathbf{Z}}_i = \mathbf{Q}_{(i)} \begin{bmatrix} \mathbf{R}_{11(i)} \\ \mathbf{0} \end{bmatrix} \quad \tilde{\mathbf{X}}_i = \mathbf{Q}_{(i)} \begin{bmatrix} \mathbf{R}_{10(i)} \\ \mathbf{R}_{00(i)} \end{bmatrix} \quad \tilde{\mathbf{y}}_i = \mathbf{Q}_{(i)} \begin{bmatrix} \mathbf{c}_{1(i)} \\ \mathbf{c}_{0(i)} \end{bmatrix}$$

dove $\mathbf{Q}_{(i)}$ è $(n_i + q) \times (n_i + q)$, $\mathbf{R}_{11(i)}$ è $q \times q$, $\mathbf{R}_{10(i)}$ è $q \times p$, la matrice $\mathbf{R}_{00(i)}$ è $n_i \times p$, il vettore $\mathbf{c}_{1(i)}$ è q -dimensionale e il vettore $\mathbf{c}_{0(i)}$ è n_i -dimensionale. Formando un'altra decomposizione ortogonale-triangolare

$$\begin{bmatrix} \mathbf{R}_{00(1)} & \mathbf{c}_{0(1)} \\ \vdots & \vdots \\ \mathbf{R}_{00(M)} & \mathbf{c}_{0(M)} \end{bmatrix} = \mathbf{Q}_0 \begin{bmatrix} \mathbf{R}_{00} & \mathbf{c}_0 \\ \mathbf{0} & \mathbf{c}_{-1} \end{bmatrix} \quad (3.10)$$

e dopo vari passaggi algebrici, la verosimiglianza (3.8) si può ora riscrivere come

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \mathbf{y}) &= (2\pi\sigma^2)^{-N/2} \exp \left(\frac{\|\mathbf{c}_{-1}\|^2 + \|\mathbf{c}_0 - \mathbf{R}_{00}\boldsymbol{\beta}\|^2}{-2\sigma^2} \right) \prod_{i=1}^M \text{abs} \left(\frac{|\boldsymbol{\Delta}|}{|\mathbf{R}_{11(i)}|} \right). \end{aligned} \quad (3.11)$$

Per un dato $\boldsymbol{\theta}$, i valori di $\boldsymbol{\beta}$ e σ^2 che massimizzano 3.11 sono

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \mathbf{R}_{00}^{-1} \mathbf{c}_0 \quad \text{e} \quad \widehat{\sigma}^2(\boldsymbol{\theta}) = \frac{\|\mathbf{c}_{-1}\|^2}{N} \quad (3.12)$$

i quali forniscono la seguente log-verosimiglianza profilo

$$\begin{aligned} \ell(\boldsymbol{\theta} \mid \mathbf{y}) &= \log L(\boldsymbol{\theta} \mid \mathbf{y}) \\ &= \frac{N}{2} [\log N - \log(2\pi) - 1] - N \log \|\mathbf{c}_{-1}\| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right). \end{aligned} \quad (3.13)$$

La log-verosimiglianza profilo (3.13) è massimizzata rispetto a $\boldsymbol{\theta}$, producendo la stima di massima verosimiglianza $\widehat{\boldsymbol{\theta}}$. Le stime di massima verosimiglianza $\widehat{\boldsymbol{\beta}}$ e $\widehat{\sigma}^2$ sono ottenute sostituendo $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ in (3.12). Sebbene tecnicamente gli effetti casuali \mathbf{b}_i non sono parametri del modello statistico, in alcuni aspetti si comportano come tali e spesso si desidera stimare i loro valori. I valori condizionati degli effetti casuali, sono i *Best Linear Unbiased Predictors* (BLUPs) dei \mathbf{b}_i , per $i = 1, \dots, M$ e possono essere calcolati utilizzando le matrici derivanti dalle decomposizioni ortogonale-triangolari, come

$$\widehat{\mathbf{b}}_i(\boldsymbol{\theta}) = \mathbf{R}_{11(i)}^{-1} \left(\mathbf{c}_{1(i)} - \mathbf{R}_{10(i)} \widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) \right). \quad (3.14)$$

In pratica, il vettore incognito $\boldsymbol{\theta}$ viene sostituito dalla sua stima di massima verosimiglianza $\widehat{\boldsymbol{\theta}}$, producendo i BLUP stimati $\widehat{\mathbf{b}}_i(\widehat{\boldsymbol{\theta}})$.

Stima della verosimiglianza ristretta (REML)

Le stime di massima verosimiglianza delle *componenti di varianza*, come σ^2 , tendono a sottostimare questi parametri. Molti analisti preferiscono le stime di massima verosimiglianza ristrette per queste quantità. Esistono diversi modi per definire il criterio di stima REML, una definizione che fornisce una forma computazionale conveniente [17] è

$$L_R(\boldsymbol{\theta}, \sigma^2 \mid \mathbf{y}) = \int L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 \mid \mathbf{y}) d\boldsymbol{\beta},$$

che, all'interno di un quadro bayesiano, corrisponde all'assunzione di una distribuzione a priori localmente uniforme per gli effetti fissi $\boldsymbol{\beta}$ e alla loro integrazione fuori dalla verosimiglianza. Utilizzando (3.11), si ottiene la log-verosimiglianza ristretta

$$\begin{aligned} \ell_R(\boldsymbol{\theta}, \sigma^2 \mid \mathbf{y}) &= \log L_R(\boldsymbol{\theta}, \sigma^2 \mid \mathbf{y}) \\ &= -\frac{N-p}{2} \log(2\pi\sigma^2) - \frac{\|\mathbf{c}_{-1}\|^2}{2\sigma^2} - \log \text{abs} |\mathbf{R}_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right). \end{aligned}$$

Questo produce la stima condizionata $\hat{\sigma}_R^2(\theta) = \|\mathbf{c}_{-1}\|^2 / (N - p)$ per σ^2 , dalla quale si ottiene la log-verosimiglianza profilo ristretta

$$\begin{aligned} \ell_R(\boldsymbol{\theta} | \mathbf{y}) &= \ell_R(\boldsymbol{\theta}, \hat{\sigma}_R^2(\boldsymbol{\theta}) | \mathbf{y}) \\ &= \text{const} - (N - p) \log \|\mathbf{c}_{-1}\| - \log \text{abs} |\mathbf{R}_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|\mathbf{R}_{11(i)}|} \right). \end{aligned} \quad (3.15)$$

Le componenti della log-verosimiglianza profilo ristretta in (3.15) sono simili a quelle della log-verosimiglianza ristretta (3.13), eccetto per il fatto che il logaritmo della norma del vettore residuo ha un moltiplicatore diverso e c'è un termine aggiuntivo dato da $\log \text{abs} |\mathbf{R}_{00}| = \log \left| \sum_{i=1}^M \mathbf{X}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i \right| / 2$.

La stima dei parametri di massima verosimiglianza ristretta (REML) viene ottenuta ottimizzando la log-verosimiglianza profilo ristretta (3.15) rispetto a $\boldsymbol{\theta}$, utilizzando poi la stima risultante $\hat{\boldsymbol{\theta}}_R$ per ottenere la stima REML di σ^2 , cioè $\hat{\sigma}_R^2(\hat{\boldsymbol{\theta}}_R)$. Analogamente, i BLUP (Best Linear Unbiased Predictors) stimati tramite REML degli effetti casuali vengono ottenuti sostituendo $\boldsymbol{\theta}$ con $\hat{\boldsymbol{\theta}}_R$ in (3.14).

In un certo senso, parlare della stima REML di $\boldsymbol{\beta}$ confonde la definizione del criterio REML. Il criterio REML dipende infatti solo da $\boldsymbol{\theta}$ e σ . Tuttavia, è utile, e forse anche ragionevole, valutare la migliore ipotesi su $\boldsymbol{\beta}$ a partire da (3.12), una volta determinato $\hat{\boldsymbol{\theta}}_R$ utilizzando il criterio REML.

La funzione di verosimiglianza ristretta non è invariante rispetto alle riparametrizzazioni uno-a-uno degli effetti fissi (ad esempio, una modifica nei contrasti che rappresentano una variabile categoriale). La modifica delle matrici \mathbf{X}_i comporta una variazione di $\log |\det(\mathbf{R}_{00})|$ e una corrispondente variazione di $\ell_R(\boldsymbol{\theta} | \mathbf{y})$. Di conseguenza, modelli LME con diverse strutture di effetti fissi, adattati utilizzando REML, non possono essere confrontati sulla base delle loro verosimiglianze ristrette. In particolare, i test del rapporto di verosimiglianza non sono validi in tali circostanze.

Algoritmi di ottimizzazione

L'ottimizzazione della log-verosimiglianza profilo o della log-verosimiglianza profilo ristretta di un modello LME viene generalmente realizzata tramite iterazioni EM (Expectation-Maximization) o iterazioni Newton-Raphson [17] [20] [21].

L'algoritmo EM [8] è un metodo iterativo molto utilizzato per la stima della verosimiglianza nei modelli con dati incompleti. Le iterazioni EM per il modello LME si basano sul trattamento degli effetti casuali \mathbf{b}_i , con $i = 1, \dots, M$, come dati non osservati. Durante l'iterazione, si utilizza il vettore corrente dei parametri di varianza-covarianza, $\boldsymbol{\theta}^{(w)}$, per valutare la distribuzione di $\mathbf{b} | \mathbf{y}$ e derivare il

valore atteso della log-verosimiglianza per un nuovo valore di θ , data questa distribuzione condizionale. Questo passaggio è noto come passaggio E . Il passaggio M consiste nel massimizzare questo valore atteso rispetto a θ per ottenere $\theta^{(w+1)}$. Ogni iterazione dell'algoritmo EM produce un miglioramento della funzione di log-verosimiglianza, sebbene tale miglioramento possa essere piccolo. Implementazioni efficienti dell'algoritmo EM per i modelli LME sono descritte in [1].

L'algoritmo di *Newton-Raphson* [34] è una delle tecniche di ottimizzazione più comunemente utilizzate. Esso si basa su un'espansione di primo ordine della funzione di punteggio attorno alla stima corrente $\theta^{(w)}$, per calcolare la stima successiva $\theta^{(w+1)}$. Ogni iterazione di Newton-Raphson richiede il calcolo della funzione di punteggio e della sua derivata, ossia la matrice hessiana della log-verosimiglianza. In condizioni generali, normalmente soddisfatte nella pratica, l'algoritmo di Newton-Raphson converge quadraticamente. Poiché il calcolo della matrice hessiana in ogni iterazione può risultare computazionalmente costoso, talvolta vengono utilizzate approssimazioni più semplici e veloci da calcolare, che danno origine agli algoritmi Quasi-Newton.

Ogni algoritmo di ottimizzazione iterativa richiede valori iniziali per i parametri. Poiché sia la log-verosimiglianza profilo, che quella profilo ristretta possono essere espresse come funzioni dei parametri θ , è necessario fornire valori iniziali per θ quando si esegue l'ottimizzazione iterativa nei modelli LME. Questi valori iniziali possono essere ottenuti da un adattamento precedente su dati simili o derivati dai dati correnti. Una procedura generale per derivare i valori iniziali di θ dai dati, durante la fase di adattamento, è descritta in [1] ed è implementata nella funzione `lme`.

Le singole iterazioni dell'algoritmo EM sono computazionalmente rapide e semplici. Sebbene le iterazioni EM generalmente spingano i parametri rapidamente verso una regione prossima all'ottimo, il progresso verso l'ottimo stesso tende a rallentare quando si è già vicini ad esso. Le iterazioni di Newton-Raphson, al contrario, sono più intense dal punto di vista computazionale rispetto a quelle EM e possono risultare instabili se distanti dall'ottimo, ma convergono molto rapidamente una volta vicini all'ottimo. Pertanto, è consigliabile un approccio ibrido che inizi con un valore iniziale $\theta^{(0)}$, esegua un numero moderato di iterazioni EM, e poi passi alle iterazioni di Newton-Raphson. In pratica, le iterazioni EM possono essere viste come un perfezionamento delle stime iniziali prima di avviare la routine di ottimizzazione più approfondita.

La funzione `lme` che si utilizzerà per stimare il modello descritto nel capitolo 4, implementa questo schema ibrido: inizialmente calcola le stime dei parametri θ , quindi utilizza diverse iterazioni EM per avvicinarsi all'ottimo e infine passa alle iterazioni di Newton-Raphson per completare la convergenza. Inoltre, è stato

necessario aumentare il numero di iterazioni dei metodi di Newton-Raphson e dell'algoritmo EM a causa di problemi di convergenza.

3.2.1 Distribuzioni approssimate

L'inferenza sui parametri di un modello lineare a effetti misti richiede distribuzioni approssimate per le stime di massima verosimiglianza ristretta, derivate da risultati asintotici. Si dimostra che, in determinate condizioni di regolarità, le stime REML nel modello LME generale descritto in 3.1.1 sono coerenti e asintoticamente normali. La matrice di varianza-covarianza approssimata per le stime di massima verosimiglianza è data dall'inverso della matrice d'informazione [6] corrispondente alla funzione di log-verosimiglianza ristretta $\ell_R = \ell_R(\boldsymbol{\beta}, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q, \sigma^2)$. Dato che

$$\mathbb{E} \left[\frac{\partial^2 \ell_R}{\partial \boldsymbol{\beta} \partial \boldsymbol{\theta}_q^\top} \right] = 0, \quad q = 1, \dots, Q \quad \text{e} \quad \mathbb{E} \left[\frac{\partial^2 \ell_R}{\partial \boldsymbol{\beta} \partial \sigma^2} \right] = 0,$$

la matrice d'informazione associata a un modello lineare a effetti misti con Q livelli di raggruppamento è diagonale a blocchi. Di conseguenza, le stime di massima verosimiglianza ristretta degli effetti fissi $\boldsymbol{\beta}$ sono asintoticamente non correlate con le stime di massima verosimiglianza ristretta dei parametri casuali $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q$ e di σ^2 . Le distribuzioni approssimate delle stime REML in un modello LME con Q livelli di raggruppamento sono le seguenti

$$\begin{aligned} \hat{\boldsymbol{\beta}} &\sim \mathcal{N} \left(\boldsymbol{\beta}, \sigma^2 \left[\mathbf{R}_{00}^{(-1)} \mathbf{R}_{00}^{(-T)} \right] \right), \\ \begin{bmatrix} \hat{\boldsymbol{\theta}}_1 \\ \vdots \\ \hat{\boldsymbol{\theta}}_Q \\ \log \hat{\sigma} \end{bmatrix} &\sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_Q \\ \log \sigma \end{bmatrix}, \mathcal{I}^{-1}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q, \sigma) \right), \\ \mathcal{I}(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q, \sigma) &= - \begin{bmatrix} \frac{\partial^2 \ell_R}{\partial \boldsymbol{\theta}_1^2} & \cdots & \frac{\partial^2 \ell_R}{\partial \log \sigma \partial \boldsymbol{\theta}_1^\top} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell_R}{\partial \boldsymbol{\theta}_1 \partial \log \sigma} & \cdots & \frac{\partial^2 \ell_R}{\partial \log \sigma^2} \end{bmatrix}, \end{aligned} \quad (3.16)$$

dove $\ell_R = \ell_R(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_Q, \sigma^2)$ denota la funzione di log-verosimiglianza profilo ristretta sugli effetti fissi, \mathcal{I} denota la matrice di informazione empirica e \mathbf{R}_{00} è definito come in (3.10). Si utilizza $\log \sigma$ al posto di σ^2 in (3.16) per ottenere una parametrizzazione non vincolata per la quale l'approssimazione normale tende a essere più accurata. Le distribuzioni approssimate per le stime di massima verosimiglianza ristretta sono utilizzate per produrre test d'ipotesi e intervalli di confidenza per i parametri del modello LME.

Intervalli di confidenza

Gli intervalli di confidenza approssimati per le componenti di varianza-covarianza e per gli effetti fissi, sono ottenuti utilizzando le distribuzioni approssimate per le stime REML. Siano df_j i gradi di libertà del denominatore per il t -test condizionato corrispondente al j -esimo effetto fisso basato su $\hat{\beta}$, un intervallo di confidenza approssimato di livello $1 - \alpha$ per β_j è

$$\hat{\beta}_j \pm t_{df_j}(1 - \alpha/2)\hat{\sigma}_R\sqrt{[\mathbf{R}_{00}^{-1}\mathbf{R}_{00}^{-T}]_{jj}},$$

dove $t_{df_j}(1 - \alpha/2)$ denota il quantile $1 - \alpha/2$ della distribuzione t con df_j gradi di libertà e $\hat{\sigma}_R$ denota la stima REML di σ . La matrice \mathbf{R}_{00} è valutata al valore stimato di $\boldsymbol{\theta}$. Gli intervalli di confidenza per la deviazione standard all'interno dei gruppi σ sono ottenuti dalla distribuzione approssimata in (3.16). Indicando con $[\mathcal{I}^{-1}]_{\sigma\sigma}$ l'ultimo elemento diagonale della matrice informativa empirica inversa definita in (3.16), un intervallo di confidenza approssimato di livello $1 - \alpha$ per σ è

$$\left[\hat{\sigma} \exp\left(-z(1 - \alpha/2)\sqrt{[\mathcal{I}^{-1}]_{\sigma\sigma}}\right), \hat{\sigma} \exp\left(z(1 - \alpha/2)\sqrt{[\mathcal{I}^{-1}]_{\sigma\sigma}}\right) \right],$$

dove $z(1 - \alpha/2)$ denota il quantile $1 - \alpha/2$ della distribuzione normale standard.

3.3 Estensione del modello LME base

Il modello lineare a effetti misti a singolo livello di base (3.1) assume che gli errori all'interno del gruppo \mathbf{e}_i siano vettori casuali indipendenti $\mathcal{N}(0, \sigma^2 I)$. Il modello lineare a effetti misti a singolo livello esteso rilassa questa assunzione consentendo errori eteroschedastici e correlati all'interno del gruppo, espressi come

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, \quad i = 1, \dots, M, \quad (3.17)$$

dove

$$\mathbf{b}_i \sim \mathcal{N}(0, \boldsymbol{\Psi}), \quad \text{e } \mathbf{e}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{\Lambda}_i), \quad i = 1, \dots, M,$$

dove le $\boldsymbol{\Lambda}_i$ sono matrici definite positive parametrizzate da un insieme fisso, generalmente ridotto, di parametri $\boldsymbol{\lambda}$. Anche in questo modello si assume che gli errori intra-gruppo \mathbf{e}_i siano indipendenti per diversi i e indipendenti dagli effetti casuali \mathbf{b}_i . Sia

$$\begin{aligned} \mathbf{y}_i^* &= \left(\boldsymbol{\Lambda}_i^{-1/2}\right)^T \mathbf{y}_i, & \mathbf{e}_i^* &= \left(\boldsymbol{\Lambda}_i^{-1/2}\right)^T \mathbf{e}_i, \\ \mathbf{X}_i^* &= \left(\boldsymbol{\Lambda}_i^{-1/2}\right)^T \mathbf{X}_i, & \mathbf{Z}_i^* &= \left(\boldsymbol{\Lambda}_i^{-1/2}\right)^T \mathbf{Z}_i, \end{aligned}$$

$$\mathbf{e}_i^* \sim \mathcal{N} \left(\left[\left(\boldsymbol{\Lambda}_i^{-1/2} \right)^T \mathbf{0}, \sigma^2 \left(\boldsymbol{\Lambda}_i^{-1/2} \right)^T \boldsymbol{\Lambda}_i \boldsymbol{\Lambda}_i^{-1/2} \right] \right) = \mathcal{N} \left(\mathbf{0}, \sigma^2 \mathbf{I} \right),$$

si può riscrivere la (3.17) come

$$\mathbf{y}_i^* = \mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{Z}_i^* \mathbf{b}_i + \mathbf{e}_i^*, \quad i = 1, \dots, M,$$

$$\mathbf{b}_i \sim \mathcal{N} \left(\mathbf{0}, \boldsymbol{\Psi} \right), \quad \mathbf{e}_i^* \sim \mathcal{N} \left(\mathbf{0}, \sigma^2 \mathbf{I} \right), \quad i = 1, \dots, M.$$

Quindi, \mathbf{y}_i^* è descritto da un modello lineare a effetti misti base. Poiché il differenziale della trasformazione lineare $\mathbf{y}_i^* = (\boldsymbol{\Lambda}_i^{-1/2})^T \mathbf{y}_i$ è semplicemente $d\mathbf{y}_i^* = \left| \boldsymbol{\Lambda}_i^{-1/2} \right| d\mathbf{y}_i$, la funzione di verosimiglianza ristretta corrispondente al modello esteso (3.17) è definita integrando gli effetti fissi dalla verosimiglianza

$$L_R(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda} | \mathbf{y}) = \int L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda} | \mathbf{y}) d\boldsymbol{\beta} = L_R(\boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda} | \mathbf{y}^*) \prod_{i=1}^M \left| \boldsymbol{\Lambda}_i^{-1/2} \right|.$$

La funzione $L_R(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda} | \mathbf{y}^*)$ corrisponde a una funzione di verosimiglianza ristretta di un modello LME base. Pertanto, tutti i risultati presentati nella sezione 3.2 possono essere utilizzati per ottenere una rappresentazione numericamente efficiente della log-verosimiglianza profilo ristretta.

3.3.1 Strutture di correlazione

Le strutture di correlazione sono utilizzate per modellare la dipendenza tra le osservazioni. Nel contesto dei modelli a effetti misti, vengono utilizzate per modellare la dipendenza tra gli errori all'interno dei gruppi. Storicamente, le strutture di correlazione sono state sviluppate per due principali classi di dati: dati temporali (time-series) e dati spaziali. I primi sono generalmente associati a osservazioni indicizzate da una variabile temporale a valori interi, mentre i secondi si riferiscono principalmente a osservazioni indicizzate da un vettore di posizione spaziale bidimensionale, con valori nel piano reale.

Modelli autoregressivi

Questa famiglia di strutture di correlazione, descritta in dettaglio in [3], include diverse classi di modelli lineari stazionari: modelli autoregressivi, modelli a media mobile e una combinazione di modelli autoregressivi e a media mobile. I modelli di correlazione autoregressivi a media mobile assumono che i dati siano osservati a punti temporali interi e, per semplicità, si utilizza la notazione e_t per un'osservazione presa al tempo t . La distanza, o lag, tra due osservazioni e_t e e_s è data da $|t - s|$. Il lag-1 si riferisce quindi a osservazioni distanti un'unità temporale, e così via.

I modelli autoregressivi esprimono l'osservazione corrente come una funzione lineare delle osservazioni precedenti più un termine di rumore omoschedastico, a_t , centrato in 0 ($E[a_t] = 0$) e assunto indipendente dalle osservazioni precedenti:

$$t = \phi_1 t_{-1} + \cdots + \phi_p t_{-p} + a_t. \quad (3.18)$$

Il numero di osservazioni passate incluse nel modello lineare (3.18), p , è chiamato l'ordine del modello autoregressivo, che viene indicato come $AR(p)$. Ci sono p parametri di correlazione in un modello $AR(p)$, dati da $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)$.

Il modello $AR(1)$ è il più semplice modello autoregressivo. La sua funzione di correlazione decresce esponenzialmente in valore assoluto con il ritardo.

$$h(k, \phi) = \phi^k, \quad k = 0, 1, \dots \quad (3.19)$$

L'unico parametro di correlazione, ϕ , rappresenta la correlazione al ritardo 1 e assume valori compresi tra -1 e 1 . Il modello $AR(1)$ è una delle poche strutture di correlazione seriale che può essere generalizzata a misurazioni in tempo continuo. Si definisce dunque la funzione di correlazione $AR(1)$ in tempo continuo [14], denotata $CAR(1)$, come

$$h(s, \phi) = \phi^s, \quad s \geq 0, \phi \geq 0. \quad (3.20)$$

Si noti che l'unico parametro di correlazione ϕ in (3.20) deve essere non negativo. Questa struttura di correlazione sarà utilizzata nel capitolo successivo per tenere conto della dipendenza temporale continua tra le osservazioni.

Capitolo 4

Analisi dei dati

In questa sezione verranno presentati i risultati derivanti dall'applicazione del modello descritto nel capitolo 3 ai dati presentati nel capitolo 2. Prima di procedere con l'analisi, è opportuno delineare quali sono i vincoli imposti al dataset, ovvero è fondamentale chiarire quali interventi sono stati effettuati per garantire la qualità e l'affidabilità dei dati utilizzati. Successivamente, verranno presentate le variabili considerate per la stima del modello, descrivendo le loro caratteristiche principali e la loro rilevanza per l'analisi.

Una volta definite le variabili, si procederà a esaminare come queste si distribuiscono nel campione nei vari periodi temporali, attraverso un'analisi univariata. Infine, verrà utilizzato un modello a effetti misti (LME) per analizzare le relazioni tra queste variabili, tenendo conto anche degli effetti casuali che potrebbero influenzare i risultati. Al termine dell'applicazione del modello, si procederà con un'analisi di clustering, utilizzando i risultati ottenuti per raggruppare i dati in categorie. Questa fase finale consentirà di identificare eventuali schemi o gruppi omogenei all'interno del campione, fornendo ulteriori spunti per l'interpretazione dei risultati.

4.1 Presentazione del dataset

Per ottenere risultati più stabili e affidabili, sono stati applicati specifici criteri di inclusione al campione in esame. In particolare, sono stati selezionati solo i pazienti che avevano effettuato un numero di visite (inclusa la prima visita basale) pari o superiore a tre. Inoltre, per evitare problematiche durante l'adattamento del modello, sono state escluse le osservazioni con valore nullo di MMSE alla visita basale, così come sono stati esclusi i soggetti che presentavano valore nullo in una delle variabili esplicative prese in considerazione.

Per i soggetti che presentavano un valore nullo del MMSE in una delle visite di controllo, è stata eliminata solo quella specifica visita, passando direttamente a quella successiva, come se non fosse mai stata effettuata. Inoltre, per garantire una numerosità campionaria sufficiente, sono stati inclusi solo i soggetti con una delle prime quattro diagnosi: MCI (Mild Cognitive Impairment), AD (Alzheimer's disease), VD (Vascular disease), MD (Mix dementia).

Dopo aver implementato tutti gli accorgimenti e le modifiche finora descritti, il dataset include ora 235 soggetti. Le variabili esplicative considerate per spiegare la variabile risposta, il punteggio MMSE, sono le seguenti: tempo, diagnosi, trattamento con memantina, altri farmaci, genere, età, livello di istruzione, traumi, consumo di alcool, fumo, stato civile, e una serie di condizioni cliniche preesistenti (APR), tra cui ipertensione arteriosa (IPA), diabete mellito (DM), fibrillazione atriale (FA), infarto miocardico acuto (IMA), attacchi ischemici transitori (TIA), ictus ischemico, ictus emorragico, vasculopatia, depressione, patologie psichiatriche, morbo di Parkinson, epilessia, altre patologie neurologiche, ipotiroidismo, ipertiroidismo, patologie gastrointestinali, neoplasie, e altre patologie. È inclusa anche una variabile identificativa del soggetto per garantire una corretta specificazione del modello ad effetti misti (LME).

La variabile "Altri farmaci" assume valore 1 per ciascun paziente, in ogni visita, se è in corso almeno una delle seguenti terapie: terapia per demenza, Ache-I, antipsicotici, antidepressivi, trazodone o benzodiazepine. Tra le variabili esplicative considerate, solo memantina, altri farmaci ed età variano nel tempo, mentre tutte le altre rimangono costanti. La variabile "Tempo", invece, è stata definita come il numero di giorni trascorsi fra la visita basale e ciascuna visita di controllo. La Tabella 4.1 mostra i dati relativi ai primi quattro soggetti del dataset, riportando i valori delle prime otto variabili associate ai primi quattro individui.

4.1.1 Analisi descrittiva

L'analisi descrittiva fornisce una panoramica iniziale sulle caratteristiche principali di ciascuna variabile considerata nel modello. Questo tipo di analisi permette di esplorare la distribuzione, la variabilità e le proprietà di ogni variabile. Attraverso statistiche sintetiche e rappresentazioni grafiche, si ottengono informazioni preliminari che guidano e supportano l'interpretazione dei risultati successivi, evidenziando le caratteristiche fondamentali dei dati raccolti.

Nel processo di riduzione del dataset iniziale, alcune categorie di variabili hanno subito una drastica diminuzione delle osservazioni, tanto che alcune modalità con poche occorrenze nel dataset originale sono state completamente eliminate nel dataset ridotto. Di conseguenza, è stato necessario riformulare le variabili esplicative che, nel dataset ristretto, presentano modalità con scarse osservazioni.

id	mmse	tempo	diagnosi	memantina	genere	età	istruzione
1	24	0	MD	0	0	81	2
1	20	189	MD	0	0	82	2
1	24	359	MD	0	0	82	2
1	23	695	MD	0	0	83	2
2	24	0	MCI	0	1	85	1
2	24	49	MCI	0	1	85	1
2	27	294	MCI	0	1	86	1
3	25	0	AD	0	1	84	1
3	26	210	AD	0	1	84	1
3	21	392	AD	0	1	85	1
4	20	0	MD	1	1	79	4
4	20	360	MD	1	0	80	4
4	19	532	MD	1	0	80	4
4	22	890	MD	1	0	81	4
4	23	1051	MD	1	0	82	4
4	20	1232	MD	1	0	82	4
4	19	1411	MD	1	0	83	4
4	21	1596	MD	1	0	83	4
4	22	1782	MD	1	0	84	4

Tabella 4.1: Panoramica delle variabili e dei soggetti del dataset.

Nelle sezioni successive verrà descritto in dettaglio il processo di riformulazione di queste variabili. Le analisi esplorative delle variabili elencate in precedenza, basate sul dataset originale, sono disponibili in Appendice 5.1. A seguito della riduzione del dataset, il numero di soggetti per ciascuna visita è diminuito rispetto a quanto riportato nella Tabella 2.1. Il numero aggiornato delle visite è riportato nella Tabella 4.2.

Visita	n(%)	Visita	n(%)
Prima visita	235(100)		
I controllo	235(100)	VII controllo	67(28.51)
II controllo	235(100)	VIII controllo	54(22.98)
III controllo	177(75.32)	IX controllo	42(17.87)
IV controllo	156(66.38)	X controllo	34(14.47)
V controllo	114(48.51)	XI controllo	23(9.79)
VI controllo	84(35.74)	XII controllo	9(3.83)

Tabella 4.2: Numero assoluto e % di pazienti in ogni visita

MMSE (Mini-Mental State Examination)

Come descritto nel capitolo 2, si è scelto di utilizzare il punteggio grezzo del MMSE (Mini-Mental State Examination) come variabile risposta, includendo invece Età e Istruzione come variabili esplicative, piuttosto che utilizzare il MMSE corretto per tali variabili. Il punteggio di MMSE, che varia tra 0 e 30, consente di misurare il livello di funzione cognitiva di ciascun paziente. Dato che si tratta di uno studio longitudinale, il punteggio di ciascun soggetto può cambiare da una visita all'altra.

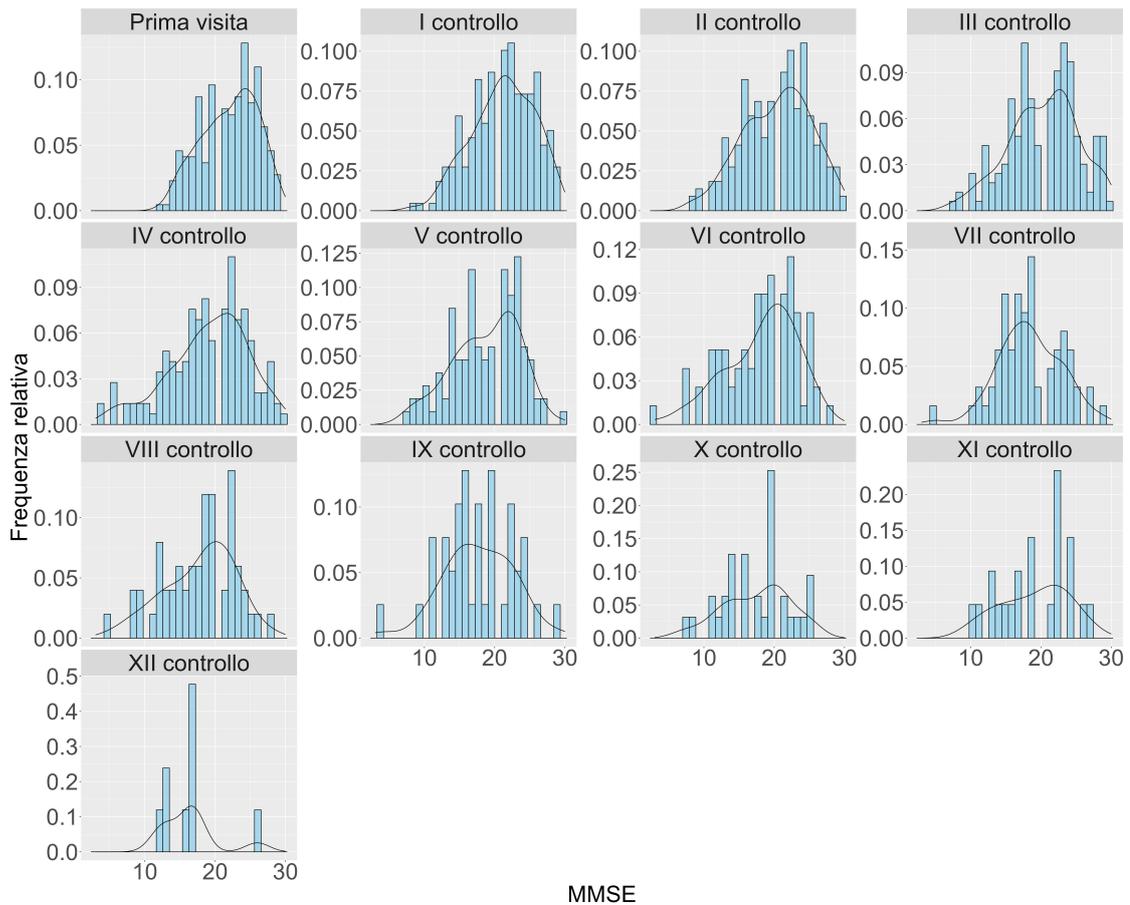


Figura 4.1: Distribuzione del MMSE rispetto alle varie visite.

La Figura 4.1 illustra la distribuzione del MMSE all'interno di ogni campione e per ciascuna visita. Ciascun istogramma riporta, sull'asse delle ascisse, i punteggi del MMSE, mentre sull'asse delle ordinate è rappresentata la frequenza relativa di ciascun valore all'interno del campione. Nelle prime visite, la distribuzione appare leggermente asimmetrica verso valori alti, mentre dalla terza visita di controllo in poi tende a diventare abbastanza simmetrica, presentando più mode. Le ultime

visite mostrano un comportamento instabile, a causa del limitato numero di osservazioni. La linea nera presente in ogni istogramma rappresenta la stima della funzione di densità della probabilità dei punteggi del MMSE. Questa linea, ottenuta attraverso tecniche di lisciatura come il Kernel Density Estimation, offre una rappresentazione fluida della distribuzione dei punteggi, facilitando l'osservazione di tendenze, picchi e asimmetrie.

Memantina in atto

Si è deciso di introdurre la variabile "Memantina in atto" separatamente rispetto alle altre terapie descritte nel Capitolo 2 e sintetizzate nella variabile Altri farmaci. Questa scelta è motivata dal fatto che, come evidenziato nel Capitolo 1, diversi studi hanno dimostrato il ruolo fondamentale della memantina nella gestione dei pazienti geriatrici affetti da demenza, soprattutto nei casi di malattia di Alzheimer. Dalla Tabella 4.3 si evince chiaramente come la percentuale di soggetti che fanno uso di memantina cresce con il susseguirsi delle visite.

Visite	No (%)	Si (%)
Prima visita	227 (96.6%)	8 (3.4%)
I visita controllo	201 (86.5%)	34 (14.5%)
II visita controllo	172 (73.2%)	63 (26.8%)
III visita controllo	122 (68.9%)	55 (31.1%)
IV visita controllo	101 (64.7%)	55 (35.3%)
V visita controllo	72 (63.2%)	42 (36.8%)
VI visita controllo	46 (54.8%)	38 (45.2%)
VII visita controllo	39 (58.2%)	28 (41.8%)
VIII visita controllo	29 (58.2%)	25 (41.8%)
IX visita controllo	21 (50%)	21 (50%)
X visita controllo	18 (52.9%)	16 (47.1%)
XI visita controllo	16 (65.2%)	8 (34.8%)
XII visita controllo	5 (55.6%)	4 (44.4%)

Tabella 4.3: Frequenze assolute e % di memantina nelle varie visite.

Altri farmaci in atto

La variabile "Altri farmaci in atto", assume valore 1 se, al momento della visita, il paziente sta sostenendo una delle seguenti terapie: terapia per demenza, Ache-I, antipsicotici, antidepressivi, trazodone o benzodiazepine. Dalla Tabella 4.4 risulta abbastanza evidente come, a seguito della prima visita conoscitiva, alla maggior parte dei pazienti sia stato prescritto almeno uno dei farmaci appena elencati.

Visite	No (%)	Si (%)
Prima visita	119 (50.6%)	116 (49.4%)
I visita controllo	76 (32.3%)	159 (67.7%)
II visita controllo	44 (18.7%)	191 (81.3%)
III visita controllo	18 (10.2%)	159 (89.8%)
IV visita controllo	15 (9.6%)	141 (90.4%)
V visita controllo	5 (4.4%)	109 (95.6%)
VI visita controllo	2 (2.4%)	82 (97.6%)
VII visita controllo	3 (4.5%)	64 (95.5%)
VIII visita controllo	2 (3.7%)	52 (96.3%)
IX visita controllo	2 (4.8%)	40 (95.2%)
X visita controllo	0 (0%)	34 (100%)
XI visita controllo	1 (4.3%)	22 (95.7%)
XII visita controllo	1 (1.1%)	8 (88.9%)

Tabella 4.4: Frequenze assolute e % di altri farmaci nelle varie visite.

Variabili esplicative fisse

Dalla Tabella A.3 in appendice emerge che molte variabili APR presentano una percentuale elevata di individui con un esito negativo della malattia. Si decide dunque di accorparle in macro-categorie: "Malattie cardiovascolari" (ipertensione arteriosa polmonare, infarto miocardico acuto, fibrillazione atriale, attacco ischemico transitorio, ictus ischemico, ictus emorragico, vasculopatia periferica); "Malattie metaboliche" (diabete mellito, ipotiroidismo, ipertiroidismo); "Malattie neurologiche" (parkinson, epilessia, altre pat. neurologiche); "Malattie psichiatriche" (depressione, patologia psichiatrica); "Malattie gastrointestinali"; "Neoplasie"; "Altre".

Variabile APR	Freq. Assente (%)	Freq. Presente (%)
Altre patologie	40 (17%)	195 (83%)
Cardiovascolari	47 (20%)	188 (80%)
Gastrointestinali	176 (75%)	59 (25%)
Metaboliche	168 (71%)	67 (29%)
Neoplasie	186 (79%)	49 (21%)
Psichiatriche	171 (73%)	64 (27%)

Tabella 4.5: Frequenze assolute e % per le variabili della sezione APR

Tuttavia, la variabile relativa alle malattie neurologiche risulta fortemente sbilanciata, motivo per cui si è deciso di rimuoverla dal dataset. La Tabella 4.5 evidenzia

come le malattie cardiovascolari e altre patologie rappresentino le categorie con il maggior numero di soggetti affetti.

Variabile	Freq. assolute	Freq. relative
Diagnosi		
MCI (Mild Cognitive Impairment)	35	15%
AD (Alzheimer's disease)	43	19%
VD (Vascular disease)	38	16%
MD (Mix dementia)	119	50%
Genere		
Uomo	84	36%
Donna	151	64%
Titolo di studio		
Nessuno/Elementari	153	65%
Medie/Superiori/Laurea	82	35%
Stato civile		
Coniugato/a	144	61%
Libero/a	91	39%
Traumi cranici		
No	204	87%
Si	31	13%
Alcool		
No	191	81%
Si	44	19%
Fumo		
No	188	80%
Si	47	20%

Tabella 4.6: Frequenze assolute e % per alcune delle variabili esplicative.

Come descritto in precedenza, alcune modalità delle variabili esplicative sono state raggruppate per consentire una maggiore stabilità dei risultati, le frequenze delle variabili fisse, con le modalità originali calcolate sul dataset senza restrizioni si trovano nella Tabella A.1 dell'appendice. Nella Tabella 4.6 si nota che la diagnosi più comune è quella di demenza di tipo misto, con una percentuale del 50%. A seguire, si trovano le diagnosi di Alzheimer (AD), vascolare (VD) e deterioramento cognitivo lieve (MCI). Le donne predominano nel campione, rappresentando il 64% dei partecipanti. Inoltre, emerge che una parte consistente dei soggetti, pari al 65%, non ha superato la scuola elementare. Il 61% dei pazienti risulta essere sposato, mentre l'87% dichiara di non aver subito traumi cranici fino all'inizio dello studio.

Infine, la maggior parte del campione, ovvero l'80%, riferisce di non fumare né bere alcolici.

Età

Gli istogrammi mostrati in Figura 4.2 riportano, sull'asse delle ascisse, l'età, mentre sull'asse delle ordinate è rappresentata la frequenza relativa di ciascun valore all'interno del campione. Si osserva chiaramente come l'età tenda ad aumentare nel tempo, spostandosi progressivamente verso valori sempre più elevati. Questo spostamento verso destra nella distribuzione evidenzia come, nel corso del tempo, ci sia una predominanza di soggetti più anziani, suggerendo possibili implicazioni relative alla salute della popolazione considerata. Anche in questo la linea nera presente in ogni istogramma rappresenta la stima della funzione di densità della probabilità dei punteggi del MMSE.

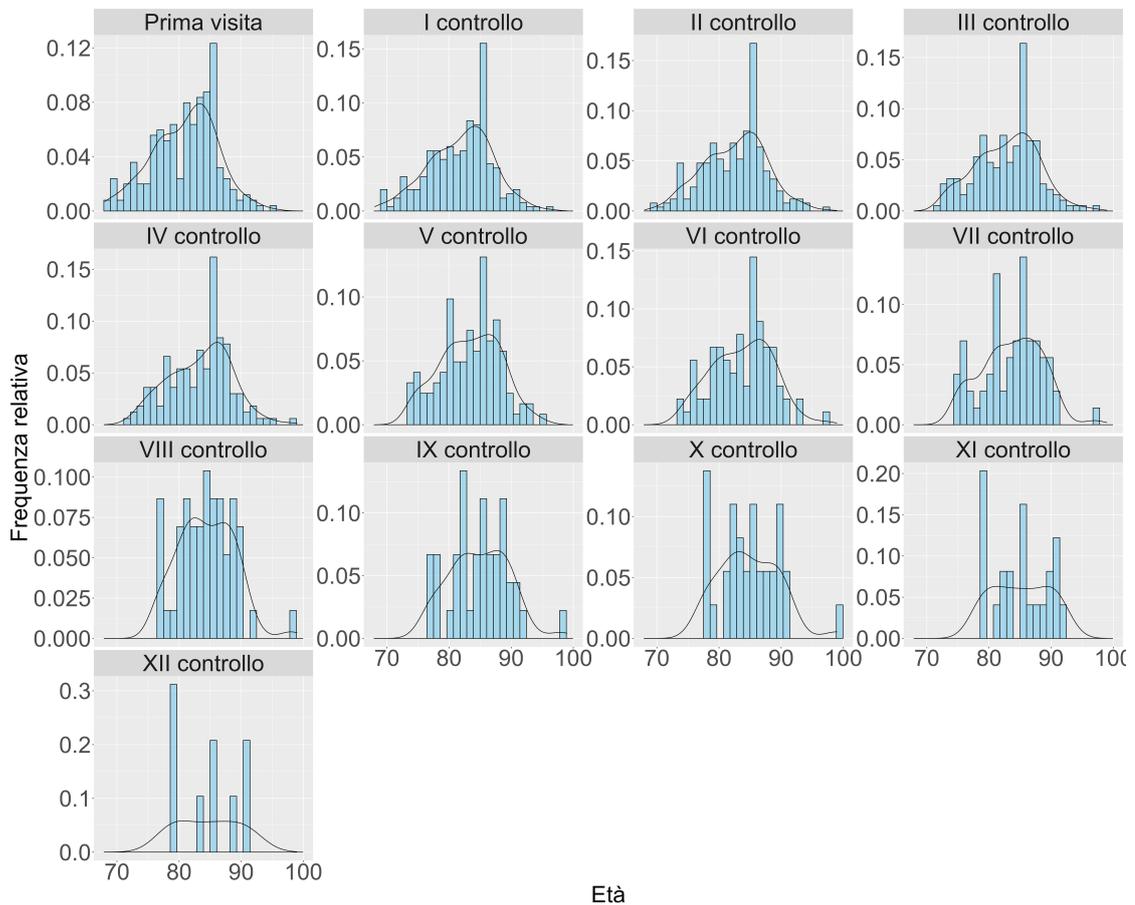


Figura 4.2: Distribuzione dell'età rispetto alle varie visite.

Variazione del MMSE nel tempo

La Figura 4.3 mostra sull'asse delle ascisse il tempo, espresso in giorni dalla prima visita e sull'asse delle ordinate il punteggio MMSE, illustrando come questo vari progressivamente con il passare del tempo. Ogni linea rappresenta un soggetto distinto, mentre la linea nera centrale rappresenta la curva generale stimata tramite loess e la rispettiva banda di variabilità. Questa rappresentazione offre una visione dettagliata delle dinamiche del test nel corso delle diverse visite, permettendo di osservare trend, fluttuazioni o cambiamenti nei punteggi dei soggetti nel tempo. In questo modo, è possibile acquisire una comprensione più approfondita dell'evoluzione della funzione cognitiva nel campione analizzato.

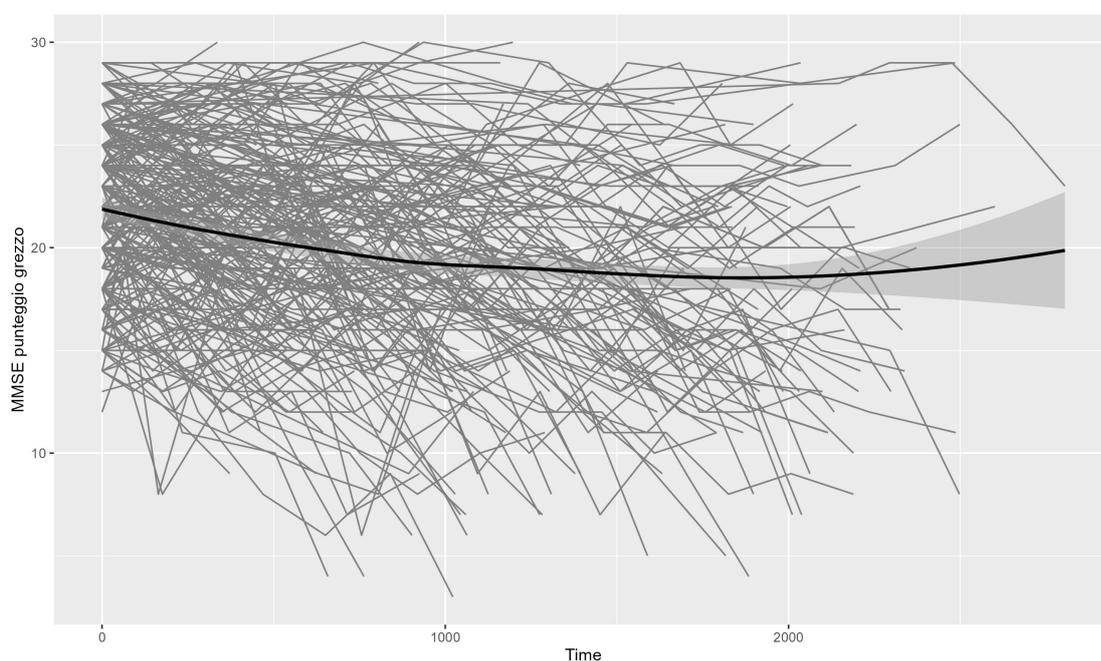


Figura 4.3: Evoluzione del punteggio MMSE nel tempo, suddivisa per soggetto.

In aggiunta, il punteggio MMSE è rappresentato nel tempo, mettendo in risalto le variazioni legate alla diagnosi e ai soggetti, per esaminare le potenziali differenze tra le varie categorie diagnostiche. Attraverso questa analisi, sarà possibile ottenere informazioni preziose sulle variazioni del funzionamento cognitivo in relazione alle diverse condizioni cliniche, contribuendo a una comprensione più approfondita dell'impatto delle diagnosi sul decorso del punteggio MMSE.

Oltre a mettere in evidenza un numero significativamente maggiore di individui nel gruppo con diagnosi di MD (Demenza Mista), la Figura 4.4 rivela che i soggetti con diagnosi di MCI (Deterioramento Cognitivo Lieve) tendono a registrare valori

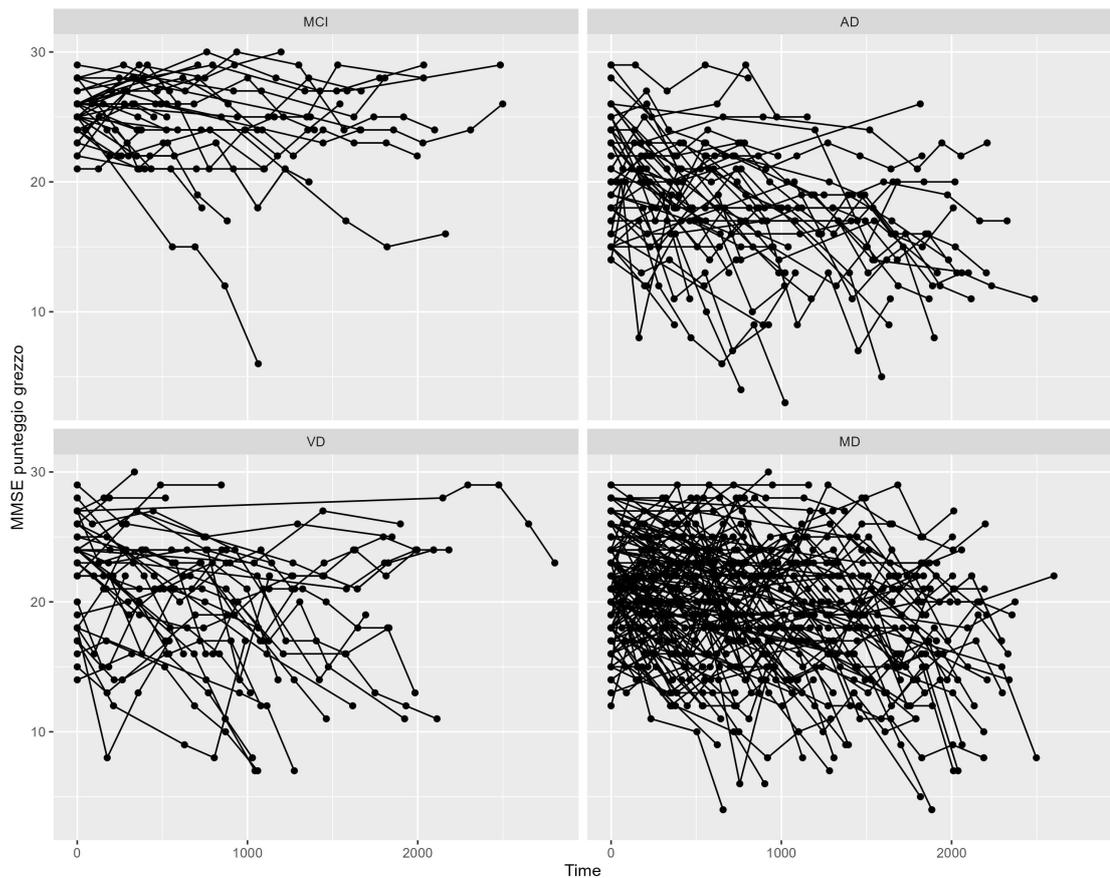


Figura 4.4: Andamento del punteggio MMSE nel tempo, analizzato per soggetti e relative diagnosi.

di MMSE più elevati, i quali rimangono relativamente stabili nel tempo, a parte alcune eccezioni. Al contrario, le altre diagnosi, pur partendo da un valore iniziale di MMSE più basso, mostrano un marcato peggioramento nel punteggio nel corso del tempo. Questa distinzione suggerisce che le diverse categorie diagnostiche presentano andamenti distintivi nel funzionamento cognitivo, con implicazioni significative per la comprensione del decorso della malattia e la gestione clinica dei pazienti.

4.2 Applicazione del modello LME

A seguito delle modifiche precedentemente apportate alle variabili esplicative, il dataset ora è composto da 19 variabili esplicative. A questo dataset è stato applicato un modello a effetti misti tramite il comando *lme* presente nel software R [26]. Come descritto anche nel capitolo 3, è necessario specificare nel modello LME i predittori che influenzano la variabile risposta in modo fisso (con un coefficiente unico per tutti i soggetti) e quelli che la influenzano in modo variabile (con coefficienti diversi per ciascun soggetto). In questo caso, si ha:

- **Effetti fissi:** tempo, diagnosi, memantina, altri farmaci, genere, età, istruzione, traumi cranici, alcool, fumo, stato civile, APR: Malattie cardiovascolari, APR: Malattie metaboliche, APR: Malattie psichiatriche, APR: Malattie gastrointestinali, APR: Neoplasie; APR: Altre patologie;
- **Effetti casuali:** tempo e memantina.

La variabile indicatrice del tempo è stata riscalata per evitare problemi di convergenza del modello. Inoltre, è stata specificata una struttura di correlazione autoregressiva di ordine 1 per i dati, adattata per gestire misure di tempo continue, come visto nel paragrafo 3.3. Utilizzando il comando *corCAR1(form = tempo / id)* all'interno della funzione *lme*, è possibile modellare, per ciascun soggetto, la dipendenza dell'osservazione successiva da quella precedente, con una correlazione che diminuisce progressivamente nel tempo. In questo modo, ogni soggetto risulta avere una propria matrice di correlazione.

La Figura 4.5 mostra i coefficienti degli effetti fissi derivanti dall'adattamento del modello ai dati, mentre in appendice si trova la Tabella A.4 con i coefficienti, gli standard error, *t*-test e *p*-value. In particolare, si osserva come il valore dell'*intercetta* (39.53) rappresenta il punteggio medio di base del MMSE quando tutte le variabili predittive sono a zero. È significativamente diverso da zero ($p < 0.001$), indicando che c'è un punteggio di partenza significativamente alto nel modello. Il coefficiente di -1.79 per la variabile *tempo* suggerisce che, in media, il punteggio dell'outcome diminuisce di 1.79 unità per ogni incremento di deviazione standard nella variabile tempo (circa 648 giorni). Questo effetto è altamente significativo ($p < 0.001$), evidenziando un calo sostanziale nel punteggio con il passare del tempo.

La *diagnosi AD* ha un coefficiente di -4.45 ($p < 0.001$), indicando che, rispetto al gruppo di riferimento (MCI), i soggetti con diagnosi di Alzheimer (AD) tendono ad avere un punteggio medio inferiore di 4.45 unità. *Diagnosi VD* (demenza vascolare) e *diagnosi MD* (demenza mista) hanno anche coefficienti negativi (-2.95 e -3.77, rispettivamente), entrambi significativi, suggerendo punteggi mediamente inferiori

in queste diagnosi rispetto al gruppo di riferimento. Il coefficiente per la variabile *memantina* (-0.47) indica un piccolo effetto negativo sul punteggio. Tuttavia, questo effetto non è significativo ($p = 0.12$), suggerendo che l'uso di memantina potrebbe non avere un impatto consistente sul punteggio del MMSE in questo campione. I coefficienti per le variabili *altri farmaci* (0.20) e *genere* (-1.41) non risultano statisticamente significativi ($p = 0.43$ e $p = 0.05$, rispettivamente), il che indica che queste variabili non hanno un effetto chiaro e consistente sul punteggio nel modello.

L'*età* ha un effetto negativo significativo (-0.20; $p < 0.001$), il che suggerisce che all'aumentare dell'età, il punteggio medio diminuisce, anche tenendo conto di altre variabili. Il coefficiente positivo di 1.36 ($p = 0.07$) per l'*istruzione* superiore suggerisce che un livello di istruzione più alto è associato a un punteggio MMSE tendenzialmente più elevato, anche se questo effetto è marginalmente significativo. I coefficienti per il consumo di *alcool* (-1.03; $p = 0.08$) e di *fumo* (-0.60; $p = 0.30$) non risultano significativamente associati al punteggio dell'outcome. Questo potrebbe indicare che, in questo modello e campione, queste abitudini di vita non hanno un impatto chiaro sui punteggi. Il coefficiente di 0.66 per lo *stato civile* libero/a non è significativo ($p = 0.52$), suggerendo che lo stato civile non ha un effetto evidente sull'outcome.

Tra le variabili anamnestiche, solo *APR: metaboliche* (con un coefficiente di 1.23; $p = 0.01$) risulta significativa, suggerendo che la presenza di patologie metaboliche potrebbe essere associata a punteggi MMSE più alti. Le altre categorie anamnestiche, come APR psichiatriche, cardiovascolari, gastrointestinali, neoplasie e altre patologie, non risultano significativamente associate al punteggio dell'outcome. Le *interazioni* tra genere e istruzione e tra genere e stato civile, mostrano effetti molto piccoli e non significativi ($p = 0.94$ e $p = 0.65$, rispettivamente). Tuttavia, le interazioni sono state incluse nel modello poiché, senza di esse, l'effetto del genere risulterebbe significativo. Infatti, diversi studi hanno dimostrato che, quando si considerano interazioni come quelle appena descritte, il declino cognitivo nelle donne non è significativamente diverso da quello negli uomini. Quindi, l'inserimento di queste interazioni permette di modellare meglio la reale dinamica del decadimento cognitivo, evitando di attribuire al genere un effetto che in realtà potrebbe essere influenzato da altri fattori.

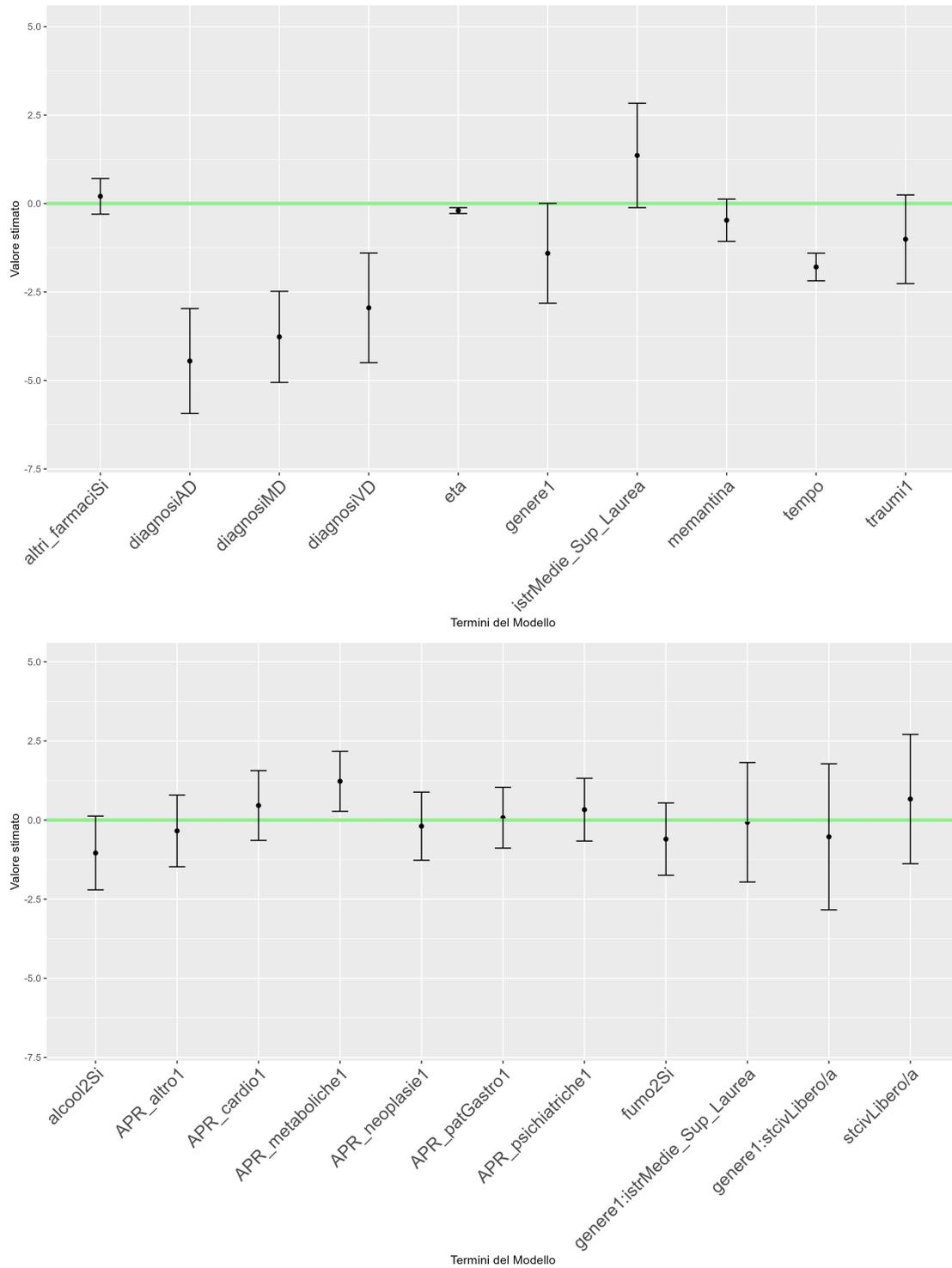


Figura 4.5: Coefficienti degli effetti fissi con relativi intervalli di confidenza.

La Tabella 4.7 descrive la struttura di varianza-covarianza degli effetti casuali nel modello, parametrizzata secondo l'approccio Log-Cholesky per garantire che la matrice di varianza-covarianza sia definita positiva. In particolare, in questo modello Lineare Misto (LME), si sta cercando di capire come i fattori "tempo" e "memantina" influiscano sulla variabile risposta MMSE, considerando però che questi effetti possano variare per ciascun individuo, identificato dal fattore "id".

	Stima	Intervallo 95%
sd(Intercept)	3.95	(3.46 - 4.51)
sd(tempo)	2.08	(1.73 - 2.49)
sd(memantina)	1.76	(0.99 - 3.12)
cor(Intercept, tempo)	0.70	(0.56 - 0.80)
cor(Intercept, memantina)	-0.35	(-0.66 - 0.05)
cor(tempo, memantina)	-0.12	(-0.61 - 0.43)

Tabella 4.7: Deviazione standard e correlazione degli effetti casuali con relativi intervalli di confidenza al 95%.

In particolare, si osserva come l'*intercetta* (ovvero il punto di partenza) varia notevolmente tra gli individui, con una deviazione standard di circa 3.95 (I.C. 95%: 3.46-4.51). Questo valore conferma una variabilità significativa dei livelli di partenza tra gli individui, con una stima piuttosto precisa grazie all'ampiezza ridotta dell'intervallo di confidenza. Tale variabilità può riflettere differenze intrinseche nei livelli iniziali della variabile di risposta per ciascun individuo. L'effetto casuale del *tempo* ha una deviazione standard stimata di 2.08 (I.C. 95%: 1.73-2.50). Questa variabilità suggerisce che l'effetto del tempo non è uniforme per tutti gli individui; alcuni potrebbero mostrare cambiamenti più marcati nel tempo rispetto ad altri. Anche in questo caso, l'intervallo di confidenza indica una stima precisa della variabilità. La deviazione standard dell'effetto casuale di *memantina* è 1.76, con un intervallo di confidenza più ampio (I.C. 95%: 0.99-3.13). Questo suggerisce una variabilità nell'effetto della memantina tra gli individui, ma anche una maggiore incertezza nella stima di questa variabilità rispetto a alla variabile tempo e all'*intercetta*. La variabilità più ampia potrebbe essere dovuta a risposte individuali diverse al trattamento con memantina.

La *correlazione* stimata tra l'*intercetta* e tempo è di 0.71 (I.C. 95%: 0.57-0.81), indicando una relazione positiva e piuttosto consistente. Questo significa che gli individui con livelli iniziali più alti tendono ad avere un effetto temporale più pronunciato, ovvero, l'evoluzione nel tempo potrebbe essere più intensa per chi parte da valori maggiori. La correlazione tra l'*intercetta* e memantina è stimata a -0.35, tuttavia l'intervallo di confidenza risulta abbastanza ampio (I.C. 95%: -0.66-0.05) e include anche lo zero. Questo suggerisce che non è possibile concludere con

sicurezza se esista una relazione significativa tra il livello di partenza e l'effetto di memantina a livello individuale. La correlazione tra tempo e memantina è stimata a -0.12 (I.C. 95%: $-0.61-0.43$) con un intervallo di confidenza che include lo zero, indicando che non vi è una relazione ben definita tra questi due effetti. Ciò implica che l'effetto di tempo e l'effetto di memantina sembrano essere indipendenti tra loro, o che non sono presenti abbastanza dati per trarre una conclusione chiara sulla loro interazione.

L'errore standard stimato all'interno dei gruppi ha una stima puntuale di 2.22 (I.C. 95%: 2.09-2.36). Questo errore standard rappresenta la variabilità residua a livello individuale, ossia la parte della variabilità che non è spiegata né dagli effetti fissi né dagli effetti casuali. L'intervallo di confidenza abbastanza stretto attorno alla stima suggerisce che l'errore standard è stato stimato con una certa precisione. Questo valore conferma che c'è una quantità di variabilità all'interno dei gruppi che il modello non spiega, il che è normale in analisi di questo tipo. Tuttavia, un errore standard di questa entità, rispetto alle altre deviazioni standard stimate, indica che il modello riesce comunque a catturare una buona parte della variabilità tra individui attraverso gli effetti casuali di "tempo" e "memantina".

La struttura *corCAR1* è stata utilizzata per modellare la correlazione nel tempo per i dati continui. Questo tipo di struttura di correlazione è particolarmente utile per dati longitudinali o sequenziali, dove le osservazioni successive nel tempo tendono a essere correlate. In questo caso, il parametro di correlazione autoregressiva ϕ ha una stima puntuale di 0.0013, con un intervallo di confidenza al 95% che va da 0.0001 a 0.0128. Un valore basso di ϕ , con un intervallo di confidenza che si avvicina allo zero, indica una correlazione temporale molto debole tra le osservazioni adiacenti. In altre parole, le misurazioni prese a tempi vicini sembrano avere una correlazione molto bassa, suggerendo che ogni misurazione è quasi indipendente da quelle precedenti. La debole correlazione temporale potrebbe essere dovuta alla natura della variabile di risposta o alle caratteristiche del campione. Potrebbe significare che le osservazioni successive non seguono un trend temporale ben definito, oppure che gli effetti dei predittori "tempo" e "memantina" catturano già gran parte della variazione temporale, lasciando poca correlazione residua tra misurazioni vicine. L'immagine 4.6 mostra i residui del modello per ogni soggetto al variare del tempo.

4.2.1 Clustering

Per esplorare nuove modalità di analisi, si è deciso di effettuare un clustering utilizzando gli effetti casuali del modello lineare a effetti misti. Questo approccio ha permesso di identificare raggruppamenti naturali nei dati, basandosi sulle variazioni individuali catturate dagli effetti casuali del modello. In un modello LME,

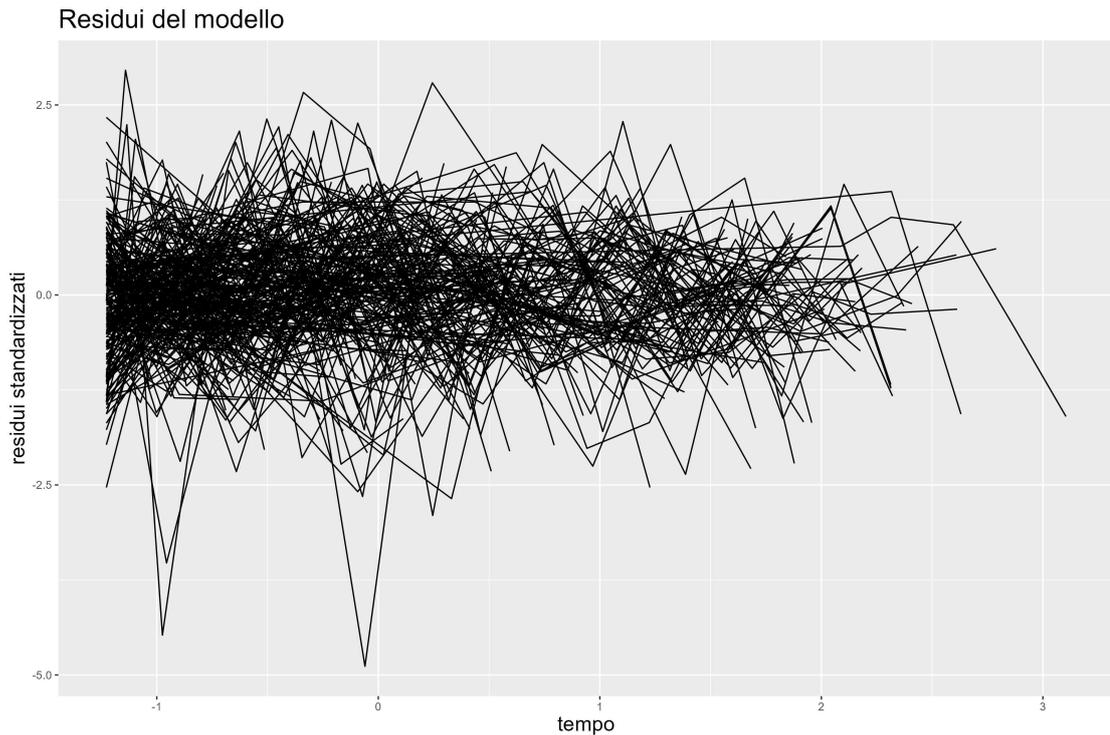


Figura 4.6: Residui standardizzati del modello.

i dati vengono modellati come combinazione di effetti fissi, che rappresentano relazioni sistematiche valide per l'intero campione, ed effetti casuali, che catturano le variazioni specifiche di ciascun individuo. Tramite il comando *ranef* della libreria *nlme*, si ottengono gli effetti casuali stimati per ogni soggetto. Questi effetti rappresentano le deviazioni specifiche dei singoli soggetti rispetto all'effetto fisso complessivo. Analizzando tali deviazioni, è possibile comprendere come ciascun gruppo si discosta dalla media e raggruppare elementi simili in cluster.

In pratica, il clustering degli effetti casuali può aiutare a identificare sottogruppi all'interno dei dati che mostrano comportamenti simili. In questo caso, poiché sia la memantina sia il tempo presentavano valori sia positivi che negativi, si è deciso di raggruppare i soggetti in quattro cluster distinti:

- soggetti con effetto positivo sia della memantina che del tempo,
- soggetti con effetto positivo della memantina e negativo del tempo,
- soggetti con effetto negativo della memantina e positivo del tempo,
- soggetti con effetti negativi sia della memantina che del tempo.

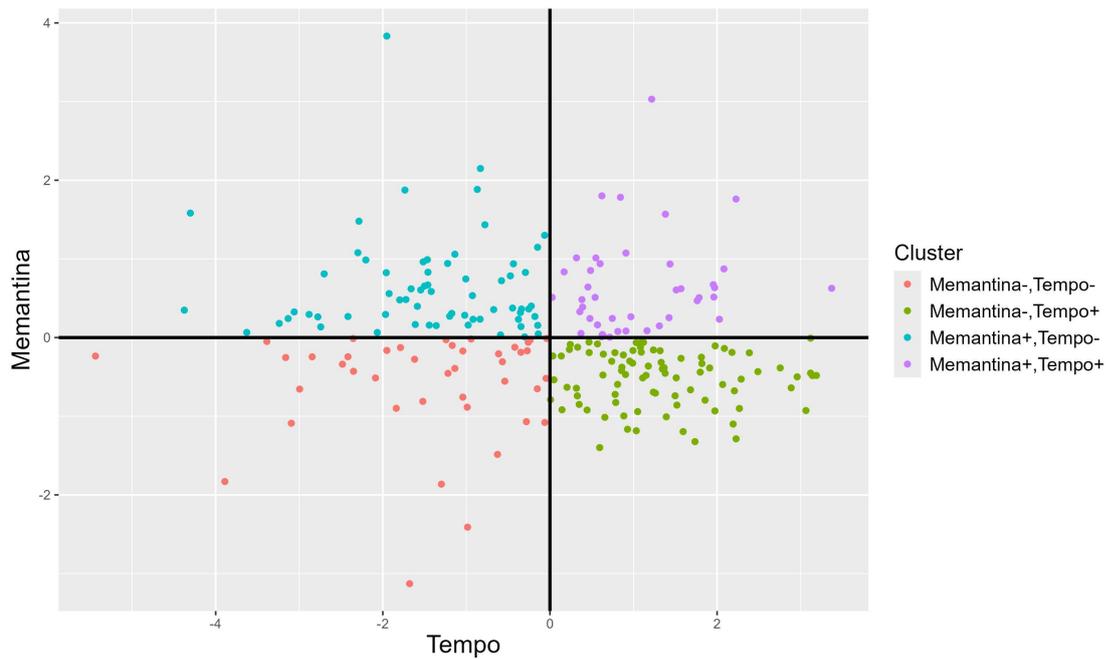


Figura 4.7: Clustering sugli effetti casuali.

La Figura 4.7 mostra gli effetti casuali delle variabili del "Tempo" (in ascissa) e della "Memantina" (in ordinata) per ciascun soggetto, suddiviso in base al cluster assegnato. Le frequenze dei cluster all'interno del campione sono riportate nella Tabella 4.8.

	M-, T-	M-, T+	M+, T-	M+, T+
Frequenza	42 (18%)	85 (36%)	66 (28%)	42 (18%)

Tabella 4.8: Frequenze assolute e % dei cluster in base agli effetti casuali.

La Figura 4.8, invece, mostra l'andamento del punteggio MMSE nel tempo, con linee rappresentative dei singoli soggetti suddivisi nei quattro cluster. Il grafico evidenzia chiaramente i trend distintivi dei cluster, con l'ombreggiatura grigia che rappresenta la banda di variabilità attorno alle curve loess di ciascun gruppo. In particolare, i soggetti con effetti negativi sia della memantina che del tempo, nonostante partano da un punteggio medio di MMSE elevato, subiscono un drastico decadimento nel corso del tempo. Al contrario, coloro che mostrano effetti positivi da entrambi i fattori tendono a mantenere una stabilità nel tempo. Inoltre, si evidenzia che i soggetti con un effetto positivo del tempo, ma negativo della memantina, presentano una risposta migliore rispetto a coloro che hanno un effetto positivo della memantina, ma negativo del tempo.

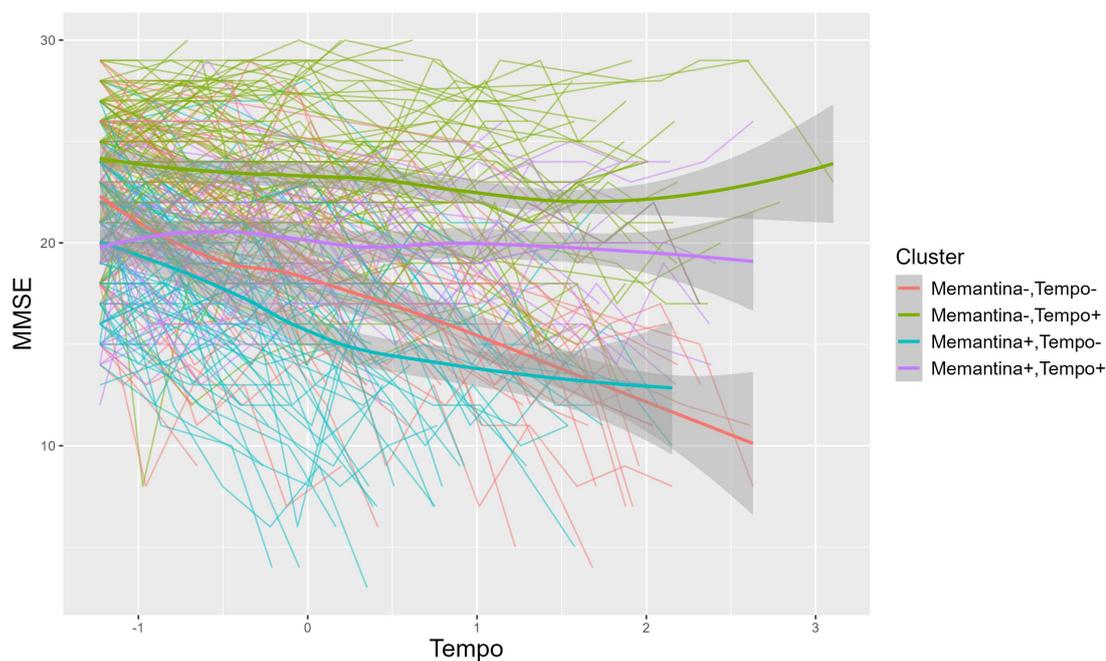


Figura 4.8: Traiettorie temporali del punteggio MMSE, stimate con il metodo loess e suddivise per cluster distinti.

In appendice, sono riportate le figure che illustrano la distribuzione delle variabili esplicative, tra cui diagnosi, memantina, altri farmaci, genere, titolo di studio, patologie psichiatriche, malattie cardiovascolari, malattie metaboliche, malattie gastrointestinali, neoplasie, altre malattie, consumo di alcool, fumo, traumi cranici e stato civile, all'interno dei diversi cluster. Nella stessa sezione, sono presentati anche i grafici che evidenziano la proporzione dei cluster in ciascuna modalità delle variabili esplicative precedentemente menzionate.

Per calcolare queste frequenze relative, è stato necessario adattare il dataset longitudinale utilizzato precedentemente per stimare il modello. Le *variabili fisse*, ovvero quelle che non cambiano nel tempo (come il genere, il livello di istruzione e alcune condizioni cliniche di base), sono state mantenute tali, prendendo direttamente il valore fisso per ogni soggetto. Per le variabili che invece potevano variare nel corso del tempo, come l'assunzione di memantina e altri farmaci, è stato adottato un approccio di aggregazione. In particolare, per la variabile *memantina*, è stato impostato il valore a 1 se il soggetto ha assunto il farmaco almeno una volta durante il periodo di osservazione, mentre è stato impostato a 0 se il soggetto non ha mai assunto memantina. Lo stesso criterio è stato applicato anche per la variabile *altri farmaci*, impostando il valore a 1 se il farmaco è stato assunto almeno una volta e a 0 in caso contrario. Per le variabili continue che possono variare nel

tempo, come l'età e il *tempo* di trattamento, è stata calcolata la media di tutte le osservazioni disponibili per ogni soggetto. In questo modo, si ottiene un valore medio che rappresenta in modo riassuntivo l'andamento di ciascun soggetto nel periodo di osservazione.

Ad esempio, la Figura A.3 in appendice, mostra le frequenze relative di ciascuna diagnosi, in quattro diversi cluster. Il primo cluster "Memantina-, Tempo-" presenta una distribuzione dominata dalla diagnosi di MD (47.6%), seguita da AD (31%). Le diagnosi di MCI e VD sono meno rappresentate, con percentuali rispettivamente del 7.1% e 14.3%. Nel secondo cluster "Memantina-, Tempo+", la diagnosi di MD è ancora prevalente (48.2%), ma è presente anche una quota significativa di soggetti con diagnosi di MCI (22.4%). In questo cluster, AD e VD sono rappresentate in modo simile, rispettivamente al 15.3% e 14.1%. Il terzo cluster "Memantina+, Tempo-" mostra una netta predominanza di diagnosi di MD (54.5%), seguita da una distribuzione equamente divisa tra AD e VD (18.2% per entrambe). In questo gruppo, la percentuale di MCI è piuttosto bassa (9.1%). Nel quarto cluster "Memantina+, Tempo+", la diagnosi di MD continua a essere dominante (52.4%), mentre le altre diagnosi sono distribuite in modo più equo, con MCI e VD presenti entrambe al 16.7% e AD al 14.3%. Nel complesso, si può osservare che la diagnosi di MD è largamente prevalente in tutti i cluster, in particolare nei cluster "Memantina-, Tempo-" e "Memantina+, Tempo-", dove supera rispettivamente il 47% e il 54%. La AD mostra una certa prevalenza nel cluster "Memantina-, Tempo-", mentre MCI e VD hanno una rappresentanza relativamente minore in tutti i cluster, con valori più elevati in specifici cluster (ad esempio, MCI nel cluster Memantina-, Tempo+).

La Figura A.5 in appendice, mostra la distribuzione della somministrazione di memantina nei quattro cluster. Nel primo cluster "Memantina-, Tempo-", la maggioranza dei pazienti (71.4%) è trattata con memantina, mentre il 28.6% non riceve il farmaco. Questo suggerisce che, nonostante l'ampio uso di memantina, il declino cognitivo legato al tempo prevale comunque. Nel cluster "Memantina-, Tempo+", si osserva una netta predominanza di soggetti non trattati con memantina (82.4%). Al contrario, nei cluster in cui le persone rispondono positivamente al trattamento, ossia "Memantina+, Tempo-" e "Memantina+, Tempo+", la percentuale di soggetti trattati è simile, rispettivamente 59.1% e 57.1%. Questo suggerisce che, indipendentemente dal progresso del declino cognitivo nel tempo, il trattamento con memantina viene mantenuto, evidenziando una stabilità nell'uso del farmaco quando risulta efficace. Tuttavia, risulta più interessante, per alcuni aspetti, valutare le proporzioni dei cluster all'interno di ciascuna variabile esplicativa. Questo aspetto sarà approfondito e commentato nel capitolo successivo.

Albero di classificazione

Per esplorare le interazioni tra le variabili e comprendere meglio i fattori che conducono ai quattro cluster, è stato costruito un albero decisionale (Figura 4.9) tramite il comando *rpart* dell'omonima libreria [33] presente nel software R. Questo è stato stimato tramite il dataset appena descritto, ovvero con una riga per ogni osservazione. Esso rappresenta un modello di classificazione volto a predire i cluster, suddivisi secondo l'effetto del tempo e memantina sul punteggio del MMSE, in base a variabili come l'uso della memantina, il tempo di trattamento, il genere, il livello di istruzione, la diagnosi, l'età e la presenza di patologie metaboliche. La struttura dell'albero inizia con una distinzione fondamentale: l'uso o meno della memantina. Questo primo livello divide i pazienti in due gruppi principali e indirizza l'analisi verso rami distinti a seconda che la memantina venga assunta o meno.

Per i pazienti che non assumono memantina, l'albero considera successivamente il tempo medio di trattamento (indicato come tempo2), se questo valore è inferiore a 554, il modello introduce ulteriori criteri, focalizzandosi sul genere del paziente. Nelle pazienti di sesso femminile, si osserva poi la presenza di patologie metaboliche, che portano a predire cluster distinti: le pazienti senza queste patologie vengono associate al cluster "Memantina+, Tempo-", mentre quelle con patologie metaboliche sono classificate in "Memantina-, Tempo+". I pazienti di genere maschile, invece, vengono classificati principalmente nel cluster "Memantina-, Tempo+". Quando il tempo medio di trattamento è maggiore o uguale a 554, il modello si orienta invece verso il cluster "Memantina-, Tempo+", suggerendo che una durata più lunga del trattamento in pazienti che non usano memantina tende a essere associata a questo gruppo.

Per i pazienti che assumono memantina, l'albero introduce un'altra variabile cruciale: il livello di istruzione. Chi ha un livello di istruzione medio, superiore o universitario viene classificato in base alla diagnosi, che distingue i pazienti in diversi gruppi clinici. Nei pazienti con diagnosi di VD o MD, il tempo medio di trattamento diventa il criterio successivo: per valori maggiori a 1010, si associa il cluster "Memantina-, Tempo-". Per chi ha valori più bassi, l'albero introduce un'ulteriore suddivisione sempre sulla base del tempo medio di trattamento: pazienti con tempo medio pari o maggiore a 652, vengono inseriti nel cluster "Memantina+, Tempo-", se il tempo risulta minore di 652, si associa il cluster "Memantina-, Tempo-". Nei pazienti con diagnosi di MCI o AD, il cluster associato è "Memantina-, Tempo-".

Se i pazienti con memantina hanno invece un basso livello di istruzione (elementare o nessuna istruzione), l'albero passa a considerare l'età. I pazienti più anziani, di età pari o superiore a 80 anni, sono ulteriormente distinti in base alla soglia dei 90 anni: coloro che superano questa età sono associati principalmente al cluster "Me-

mantina+, Tempo+”, mentre i più giovani all’interno dello stesso gruppo educativo tendono a essere associati al cluster “Memantina+, Tempo-”. Per i pazienti con età inferiore a 80 anni, invece, l’albero tende a prevedere il cluster “Memantina+, Tempo+”. All’interno dei nodi foglia viene visualizzato anche il tasso di classificazione nel nodo, espresso come numero di classificazioni corrette sul numero di osservazioni totali nel nodo.

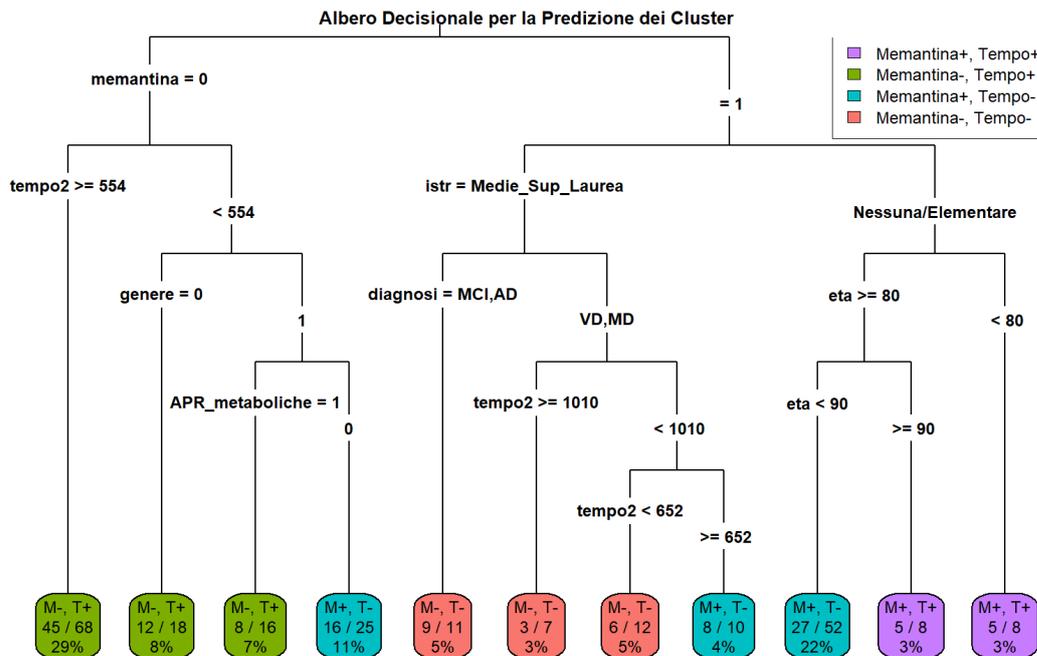


Figura 4.9: Albero decisionale per la classificazione dei soggetti nei diversi cluster.

In conclusione, si può dire che le donne senza patologie metaboliche e con un tempo di osservazione inferiore a 544 giorni, potrebbero trarre beneficio dall’introduzione della memantina, che finora non è stata loro somministrata. Tuttavia, questo beneficio potrebbe non perdurare nel lungo periodo, poiché in questo gruppo il tempo tende a influire negativamente sul punteggio MMSE. Per i pazienti che già assumono memantina e hanno un basso livello di istruzione, la combinazione di memantina e tempo mostra generalmente un effetto positivo, tranne nei casi di età compresa tra 80 e 90 anni: qui, nonostante un iniziale effetto positivo dato dalla memantina, il progredire del tempo porta comunque a un declino cognitivo. Infine, nei pazienti con istruzione superiore, si osserva prevalentemente un effetto negativo della memantina e del tempo, a eccezione di coloro con diagnosi di MD o VD e un tempo di osservazione medio tra 652 e 1010 giorni, i quali sembrano beneficiare di un effetto positivo di entrambi i fattori.

Capitolo 5

Conclusioni e sviluppi futuri

Al termine di questo lavoro di ricerca, è possibile riflettere sui risultati raggiunti e sull'importanza delle analisi svolte in relazione agli obiettivi prefissati. L'obiettivo principale dello studio era comprendere le dinamiche e i fattori che influenzano il deterioramento cognitivo in età avanzata, osservando come questo processo possa variare in funzione di specifiche caratteristiche cliniche e demografiche dei pazienti. Tale analisi ha permesso di evidenziare schemi di declino cognitivo distintivi e di individuare possibili indicatori predittivi di progressione, contribuendo a delineare prospettive utili per la gestione e il supporto di questa popolazione.

Per l'analisi delle curve di decadimento cognitivo, si è utilizzato un dataset di 615 pazienti geriatrici con età maggiore di sessantacinque anni, nel quale sono state raccolte variabili sia statiche (come stato sociale, sesso, titolo di studio, professione, consumo di alcol e tabacco, attività fisica, traumi cranici, anamnesi patologica remota e diagnosi iniziale) sia dinamiche, rilevate a ogni visita (tra cui terapie in atto, nuove terapie prescritte, punteggi MMSE, ADL e IADL). Questo dataset presenta alcune particolarità: il numero di visite effettuate dai pazienti varia e ciascuno può scegliere liberamente quando presentarsi, purché tra una visita e l'altra intercorra un periodo minimo di 30 giorni. Inoltre, non sempre tutti i test previsti sono stati eseguiti in ciascuna visita, generando una serie di dati con misurazioni mancanti.

Queste caratteristiche hanno richiesto l'applicazione di modelli per dati longitudinali, che consentissero di tenere conto della variabilità interindividuale, della frequenza irregolare delle visite e delle misurazioni mancanti. In particolare, si è fatto ricorso ai modelli ad effetti misti lineari (LME), che permettono di analizzare la progressione del decadimento cognitivo nel tempo incorporando le variabili fisse e variabili a livello di singolo paziente, per ottenere stime robuste anche in presenza di dati non uniformemente distribuiti.

Prima di applicare il modello LME, è stato necessario apportare alcune modifiche al dataset originale per renderlo adatto alla fase di analisi. Come variabile di interesse per misurare il decadimento cognitivo di ogni paziente, si è scelto di utilizzare il punteggio grezzo del MMSE (Mini Mental State Examination), che rappresenta la risposta da modellare. Si sono poi selezionate diverse variabili esplicative per analizzare l'evoluzione del punteggio MMSE nel tempo. La variabile *tempo* è stata calcolata come il numero di giorni trascorsi dalla visita basale di ciascun paziente, impostando il giorno zero come riferimento della prima visita. Tra le altre variabili incluse nell'analisi, si trova la *diagnosi* iniziale e l'assunzione di farmaci, in particolare la *memantina*, registrata come assunta o non assunta, e una variabile aggiuntiva costruita ad hoc per indicare la presenza o meno di *altri farmaci* assunti durante la visita.

Per approfondire ulteriormente il ruolo dei fattori socio-demografici e di salute sul declino cognitivo, sono state prese in considerazione altre caratteristiche stabili nel tempo. Tra queste vi sono il *genere*, il *titolo di studio*, l'*età* e lo *stato civile*. Inoltre, per indagare il potenziale impatto di eventi traumatici o di abitudini di vita sul MMSE, sono stati considerati i *traumi cranici*, l'assunzione di *alcool* e il *fumo*. Una parte rilevante del modello è stata dedicata alle patologie pregresse, catalogate nei gruppi: *malattie cardiovascolari*, *malattie metaboliche*, *malattie psichiatriche*, *malattie gastrointestinali*, *neoplasie* e, infine, un gruppo residuale di *altre patologie*.

Per garantire una maggiore omogeneità dei dati, si è deciso di includere nel modello soltanto i pazienti con una diagnosi iniziale specifica, selezionando coloro che presentavano una delle seguenti condizioni: Mild Cognitive Impairment (MCI), demenza vascolare (VD), malattia di Alzheimer (AD) o demenza mista (MD). Inoltre, per garantire che i dati fossero sufficientemente ricchi e rappresentativi dell'evoluzione nel tempo, si è stabilito come criterio di inclusione che ciascun paziente avesse effettuato almeno tre visite, inclusa la visita basale. Questo requisito permette di osservare l'andamento del decadimento cognitivo con un numero minimo di punti temporali, essenziale per l'analisi longitudinale.

È stato imposto il criterio di esclusione per i pazienti con un valore mancante di MMSE alla visita basale, poiché senza questo dato iniziale sarebbe stato impossibile definire un punto di partenza per valutare il decadimento cognitivo nel tempo. Tuttavia, in pratica, dopo aver applicato le precedenti restrizioni (inclusione di pazienti con diagnosi MCI, VD, AD, MD e almeno tre visite), è emerso che nessuno dei pazienti rimanenti presentava un valore nullo di MMSE alla prima visita. Di conseguenza, pur avendo introdotto questo criterio per ragioni metodologiche, non è stato necessario escludere ulteriori pazienti. Mentre, nel caso in cui un paziente avesse un valore mancante di MMSE in una delle visite di follow-up, è stata eliminata solo quella visita specifica dal dataset, mantenendo comunque i dati delle altre visite.

Il modello LME è stato dunque applicato a un dataset finale composto da 235 pazienti e un totale di 19 variabili. Per gestire al meglio la variabilità tra individui, sono stati introdotti come *effetti casuali* sia il tempo sia l'assunzione della memantina, catturando così le differenze individuali nell'evoluzione del decadimento cognitivo e nella risposta al trattamento farmacologico. Inoltre, per tenere conto della natura continua e irregolare dei tempi di osservazione, è stata imposta una struttura di correlazione di tipo autoregressivo di primo ordine (corCAR1). Questa scelta è risultata fondamentale per gestire il fatto che l'intervallo di tempo tra le visite variava in modo continuo tra un paziente e l'altro.

L'*analisi dei coefficienti fissi* del modello LME mette in evidenza risultati significativi riguardo alla progressione del deterioramento cognitivo nei pazienti. Il coefficiente per la variabile "tempo" standardizzata indica che, con l'aumento del tempo di 648 giorni (pari a una deviazione standard), il punteggio MMSE tende a ridursi di circa 1.79 punti. Questo dato suggerisce che pazienti seguiti per periodi più lunghi mostrano, in media, un peggioramento cognitivo maggiore. Inoltre, i pazienti affetti da Alzheimer presentano, in media, punteggi MMSE inferiori di 4.45 unità rispetto al gruppo di riferimento composto da soggetti con Mild Cognitive Impairment. Anche coloro con demenza vascolare e demenza mista registrano punteggi significativamente più bassi (rispettivamente -2.95 e -3.77). L'età ha un effetto negativo, con una diminuzione media di 0.20 punti MMSE per ogni anno in più. D'altra parte, la presenza di patologie metaboliche è associata a un punteggio MMSE leggermente più alto (1.23), suggerendo che questi pazienti potrebbero avere una condizione cognitiva migliore rispetto agli altri.

L'*analisi degli effetti casuali* suggerisce una variabilità considerevole nella progressione del deterioramento cognitivo. L'effetto casuale del tempo, con una deviazione standard di 2.08, indica che alcuni pazienti sperimentano un declino più marcato rispetto ad altri. Anche la variabilità nella risposta alla memantina è significativa (deviazione standard di 1.76), sebbene con un grado di incertezza maggiore. La correlazione tra l'intercetta e il tempo è di 0.71, indicando che i pazienti con punteggi iniziali più elevati tendono a mostrare un declino più accentuato, mentre le correlazioni tra intercetta e memantina e tra tempo e memantina non risultano significative. L'errore standard all'interno dei gruppi è di 2.22, rivelando una certa variabilità non spiegata dal modello. Infine, la struttura di correlazione autoregressiva (corCAR1) presenta un parametro di correlazione ϕ molto debole, indicando una correlazione temporale quasi inesistente tra osservazioni consecutive.

Infine, è stata eseguito un *clustering* dei pazienti a seconda degli effetti casuali di tempo e memantina sulla variabile risposta MMSE. In particolare i soggetti sono stati suddivisi in quattro gruppi: chi ha un effetto positivo sia della memantina che del tempo, chi ha un effetto positivo del tempo e negativo della memantina, chi ha un effetto negativo del tempo e positivo della memantina e chi ha un effetto

negativo di entrambi i fattori. Valutando la variabile risposta nel tempo suddivisa in questi quattro gruppi (Figura 4.8), si vede come i soggetti con effetti negativi sia della memantina che del tempo, nonostante partano da un punteggio medio di MMSE elevato, subiscano un drastico decadimento nel corso del tempo. Al contrario, coloro che mostrano effetti positivi da entrambi i fattori tendono a mantenere una stabilità nel tempo. Inoltre, si evidenzia come i soggetti con un effetto positivo del tempo, ma negativo della memantina, partono da un punteggio di MMSE più elevato e tendono a rimanere stabili nel tempo, rispetto a coloro che hanno un effetto positivo della memantina, ma negativo del tempo.

Nell'analisi della suddivisione dei quattro cluster rispetto ai diversi predittori del modello, emerge un quadro interessante. Dalla Figura A.4 in appendice, risulta evidente che i pazienti con una diagnosi di MCI hanno una probabilità significativamente maggiore di essere collocati nel gruppo caratterizzato da un effetto negativo della memantina ma positivo del tempo. Questo potrebbe indicare che, nonostante la memantina non offra benefici tangibili a questi pazienti, il passare del tempo consenta loro di adattarsi meglio o di stabilizzare le loro capacità cognitive. Al contrario, i pazienti affetti da Alzheimer mostrano una tendenza a presentare effetti negativi sia della memantina che del tempo. Questa osservazione suggerisce che per i soggetti con AD, la terapia con memantina non solo potrebbe non essere efficace, ma il tempo stesso può contribuire a un deterioramento cognitivo ulteriore, evidenziando la complessità della gestione della malattia. Per quanto riguarda i pazienti con diagnosi di demenza vascolare e demenza mista, i gruppi più prevalenti comprendono coloro che sperimentano un effetto negativo della memantina, ma positivo del tempo e un effetto positivo della memantina, ma negativo del tempo.

Tra coloro che hanno assunto memantina almeno una volta durante lo studio (Figura A.6 dell'appendice), il gruppo prevalente è quello caratterizzato da un effetto positivo del farmaco ma negativo del tempo. Questo suggerisce che, sebbene la memantina produca benefici immediati, questi pazienti continuano a mostrare un declino cognitivo legato al passare del tempo. Inoltre, i pazienti che riportano nell'anamnesi patologica una malattia metabolica (Figura A.18 dell'appendice) hanno una probabilità maggiore di rientrare nel gruppo con un effetto negativo della memantina, ma positivo del tempo. Questa stessa tendenza si osserva anche nei soggetti che dichiarano di avere una storia di fumo o consumo di alcool (Figure rispettivamente A.26 e A.28 dell'appendice). Questi fattori possono influire sulla risposta al trattamento, portando a un impatto ridotto della memantina e, allo stesso tempo, a un effetto temporale che, per quanto limitato, è positivo rispetto al deterioramento previsto.

In conclusione, si osserva che i pazienti con un effetto positivo del tempo tendono a presentare un decadimento cognitivo minore, indipendentemente dall'effetto positivo o negativo della memantina. Questo suggerisce che il fattore tempo possa avere un'influenza sul decorso del punteggio MMSE più rilevante rispetto al trattamento farmacologico. Analizzando le variabili incluse nel modello, emerge che la percentuale più elevata di pazienti con un effetto positivo del tempo (indipendentemente dall'assunzione di memantina) si riscontra tra quelli con diagnosi di Mild Cognitive Impairment (MCI), pari al 74,3%, seguita da percentuali più basse nei casi di demenza vascolare (VD) e demenza mista (MD). Chi soffre di Alzheimer, quindi, sembra avere un declino cognitivo maggiore. Inoltre, gli uomini sembrano beneficiare di un effetto positivo del tempo in misura maggiore rispetto alle donne, così come i pazienti con malattie metaboliche rispetto a quelli che ne sono privi. Condizioni come il fumo e il consumo di alcol non sembrano influire significativamente sul declino cognitivo poiché, in entrambi i casi, la maggior parte dei pazienti che ne fanno uso mostrano un effetto positivo del tempo.

L'*albero decisionale* (Figura 4.9) ha evidenziato che la variabile memantina è la più influente nel discriminare tra i gruppi. In particolare, le donne che non hanno mai assunto memantina durante le visite, con una durata media di osservazione inferiore a 554 giorni e senza patologie metaboliche, tendono a mostrare un effetto positivo della memantina sul punteggio MMSE, ma un effetto negativo del tempo, indicando che per loro la terapia potrebbe offrire benefici iniziali che non si mantengono nel lungo periodo. Per i pazienti che hanno assunto memantina, il livello di istruzione emerge come un fattore discriminante. Tra coloro con un basso livello di istruzione (nessuna o elementare), i pazienti con età compresa tra 80 e 90 anni rientrano principalmente nel cluster in cui la memantina ha un effetto positivo ma il tempo ha un effetto negativo. D'altra parte, i pazienti di età superiore a 90 anni o inferiore a 80 mostrano tendenzialmente effetti positivi sia della memantina che del tempo, suggerendo che la memantina possa stabilizzare o rallentare il declino cognitivo nelle fasce di età più giovani e più anziane. Infine, per i pazienti con livelli di istruzione più alti (scuola media, superiore o laurea), si osserva una tendenza a un effetto negativo sia della memantina che del tempo. Un'eccezione si verifica per i pazienti con diagnosi di demenza mista (MD) o demenza vascolare (VD) e con una durata media di osservazione compresa tra 652 e 1010 giorni, che tendono a beneficiare di un effetto positivo della memantina e del tempo.

5.1 Opportunità

Nel contesto di questo elaborato, il dataset presenta effettivamente delle limitazioni che influenzano le analisi possibili. L'indisponibilità del punteggio MMSE per tutti i pazienti in tutte le visite e il numero limitato di pazienti con visite di follow-up dopo la visita basale, rendono le analisi longitudinali meno potenti. La presenza di molteplici valori nulli in alcune variabili, come l'NPI e il CDT, limita l'uso di questi test nella modellizzazione. Infine, il fatto che il tempo tra una visita e l'altra sia continuo e variabile per ciascun soggetto, complica l'applicazione di modelli temporali classici, restringendo il campo a metodi che gestiscono tempi irregolari.

D'altra parte, il dataset offre anche numerose opportunità. Potrebbero essere aggiunte variabili relative a terapie farmacologiche specifiche, come inibitori delle colinesterasi (Ache-I), benzodiazepine, antidepressivi, ecc., che, considerate singolarmente, potrebbero fornire una comprensione più dettagliata sugli effetti delle terapie. Anche l'analisi di altri indici di decadimento funzionale, come ADL e IADL, potrebbe arricchire le conclusioni, aiutando a costruire un quadro più completo della progressione della patologia nei pazienti.

Dal punto di vista metodologico, è possibile esplorare nuovi approcci. L'ampliamento della verosimiglianza di processi a tempo continuo, come il modello di Ornstein-Uhlenbeck, potrebbe essere particolarmente utile per modellare l'evoluzione della malattia nel tempo, adattandolo alle peculiarità di questo dataset. Un'altra via interessante potrebbe essere l'applicazione di tecniche avanzate di clustering, come metodi di clustering basati su traiettorie o modelli mistura bayesiani, per raggruppare i pazienti in base a schemi di progressione comuni. In conclusione, pur partendo da un dataset con alcune limitazioni, l'integrazione di nuove variabili, l'indagine di indici alternativi e l'adozione di metodologie avanzate potrebbero potenzialmente migliorare la qualità delle analisi e offrire una visione più approfondita delle dinamiche della malattia nel tempo.

Appendice

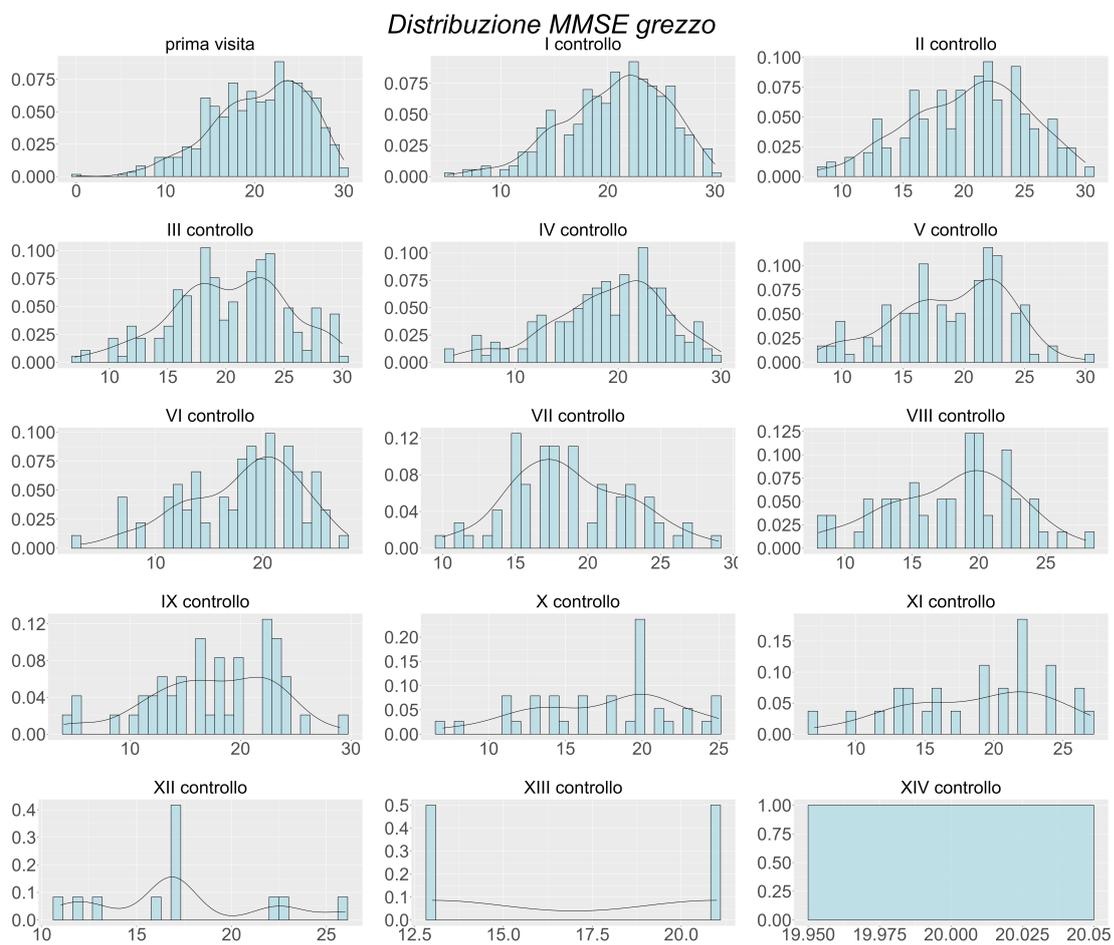


Figura A.1: Distribuzione del MMSE in ogni visita (dataset originale).

Variabile	Freq. assolute	Freq. relative
Diagnosi		
MCI (Mild Cognitive Impairment)	94	15%
AD (Alzheimer's disease)	114	19%
VD (Vascular disease)	119	19%
MD (Mix dementia)	262	43%
PD (Parkinson's disease)	6	1%
DLB (Dementia with Lewy bodies)	11	2%
PSP (Progressive Supranuclear Palsy)	3	0%
FTD (Frontotemporal dementia)	6	1%
Genere		
Uomo	234	38%
Donna	381	62%
Titolo di studio		
Nessuno	13	2%
Elementari	395	64%
Medie inferiori	105	18%
Medie superiori	49	8%
Professionali	25	4%
Laurea	26	4%
NA	2	0%
Stato civile		
Celibe/nubile	22	4%
Coniugato/a	345	56%
Divorziato/a	10	2%
Vedovo/a	235	38%
NA	3	0%
Traumi cranici		
No	524	85%
Si	88	15%
NA	3	0%
Alcool		
No	539	88%
Si (ai pasti)	69	11%
Consumo attivo	3	0.5%
Consumo pregresso	2	0.25%
NA	2	0.25%
Fumo		
No	488	79%
Attivo	32	5%
Ex fumatore	93	15%
NA	2	1%

Tabella A.1: Frequenze assolute e % per alcune delle variabili esplicative.

Memantina in atto	No (%)	Si (%)	Non presenti
Prima visita	596 (96.9%)	19 (3.1%)	0
I visita controllo	328 (86.1%)	53 (13.9%)	234
II visita controllo	189 (73.3%)	69 (26.7%)	357
III visita controllo	132 (69.8%)	57 (30.2%)	426
IV visita controllo	110 (65.9%)	57 (34.1%)	448
V visita controllo	81 (65.3%)	43 (34.7%)	491
VI visita controllo	53 (57.6%)	39 (42.4%)	523
VII visita controllo	43 (56.6%)	33 (43.4%)	539
VIII visita controllo	35 (57.4%)	26 (42.6%)	554
IX visita controllo	22 (43.1%)	29 (56.9%)	564
X visita controllo	20 (52.6%)	18 (47.4%)	577
XI visita controllo	16 (55.2%)	13 (44.8%)	586
XII visita controllo	6 (46.2%)	7 (53.8%)	602
XIII visita controllo	1 (50.0%)	1 (50.0%)	613
XIV visita controllo	1 (100.0%)	0 (0.0%)	614

Tabella A.2: Frequenze assolute e % di memantina nelle varie visite.

Variabile APR	Freq. 0 (%)	Freq. 1 (%)
ALTRE PAT. NEURO.	610 (99.19)	5 (0.81)
ALTRO	101 (16.42)	514 (83.58)
DEPRESSIONE	474 (77.07)	141 (22.93)
DM	488 (79.35)	127 (20.65)
EPILESSIA	608 (98.86)	7 (1.14)
FA	508 (82.60)	107 (17.40)
ICTUS EMORRAGICO	604 (98.21)	11 (1.79)
ICTUS ISCHEMICO	576 (93.66)	39 (6.34)
IMA	551 (89.59)	64 (10.41)
IPA	421 (68.46)	194 (31.54)
IPERTIROIDISMO	600 (97.56)	15 (2.44)
IPOTIROIDISMO	563 (91.54)	52 (8.46)
NEOPLASIE	492 (80.00)	123 (20.00)
PARKINSON	607 (98.70)	8 (1.30)
PAT. GASTRO.	475 (77.24)	140 (22.76)
PAT. PSICHIATRICA	581 (94.47)	34 (5.53)
TIA	591 (96.10)	24 (3.90)
VASCULOPATIA P.	531 (86.34)	84 (13.66)

Tabella A.3: Frequenze assolute e % per le variabili della sezione APR

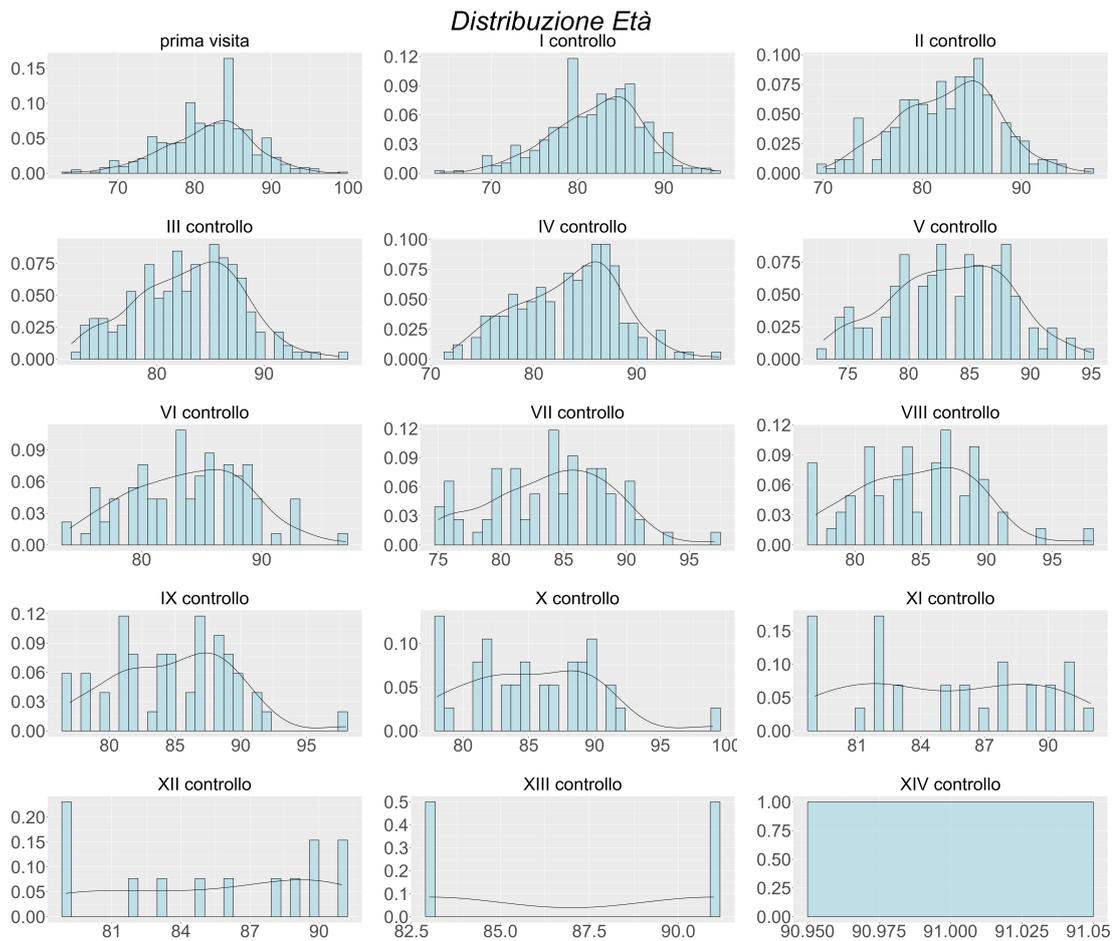


Figura A.2: Distribuzione dell'età in ogni visita (dataset originale).

Coefficients fissi del modello LME

Variabile	Valore	Std.Error	DF	t-value	p-value
(Intercept)	39.53	3.71	1226	10.66	0.00
tempo	-1.79	0.20	1226	-9.01	0.00
diagnosi: AD	-4.45	0.75	217	-5.92	0.00
diagnosi: VD	-2.95	0.79	217	-3.75	0.00
diagnosi: MD	-3.77	0.65	217	-5.77	0.00
memantina	-0.47	0.30	1226	-1.55	0.12
altri farmaci Si	0.20	0.26	1226	0.79	0.43
genere: donna	-1.41	0.72	217	-1.96	0.05
eta	-0.20	0.04	1226	-4.67	0.00
istruzione: Medie,Sup,Laurea	1.36	0.75	217	1.82	0.07
traumi: Si	-1.09	0.63	217	-1.59	0.11
alcool: Si	-1.03	0.59	217	-1.76	0.08
fumo: Si	-0.60	0.58	217	-1.04	0.30
stato civile: Libero/a	0.66	1.04	217	0.64	0.52
APR psichiatriche: Si	0.33	0.50	217	0.65	0.51
APR cardio: Si	0.46	0.56	217	0.82	0.41
APR metaboliche: Si	1.23	0.48	217	2.55	0.01
APR Gastro: Si	0.08	0.55	217	0.16	0.88
APR neoplasie: Si	-0.19	0.55	217	-0.35	0.72
APR altro: Si	-0.34	0.57	217	-0.59	0.56
donna*Medie,Sup,Laurea	-0.07	0.96	217	-0.07	0.94
donna*Libero/a	-0.53	1.70	217	-0.45	0.65

Tabella A.4: Tabella con stime, errori standard, DF, t-value e p-value derivanti dal modello LME

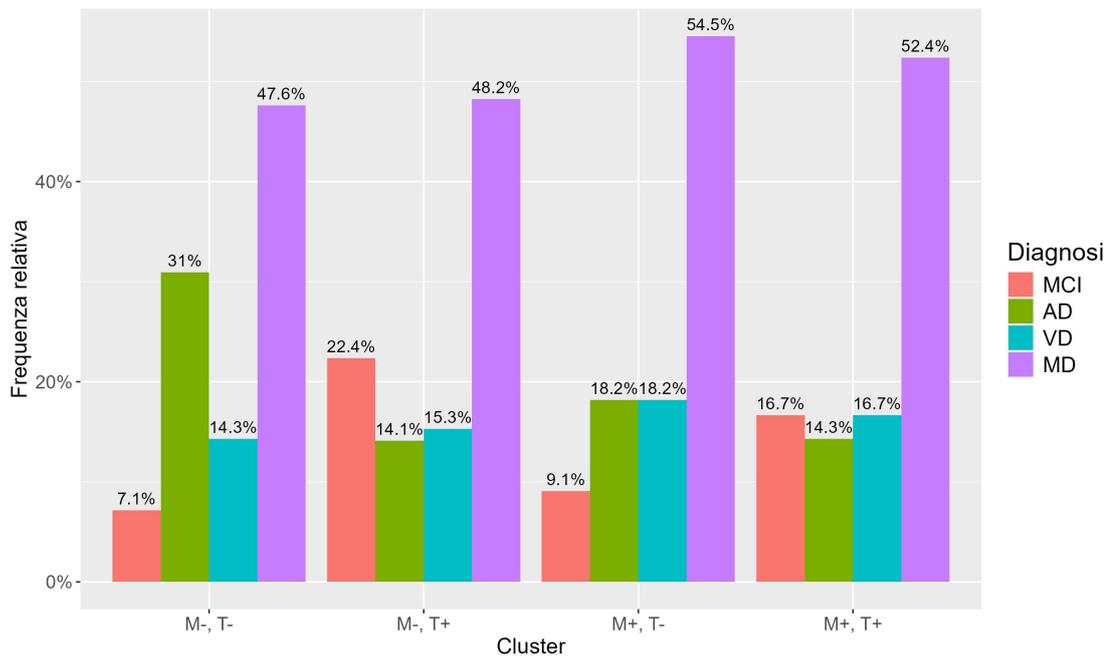


Figura A.3: Frequenze relative delle diagnosi nei cluster.

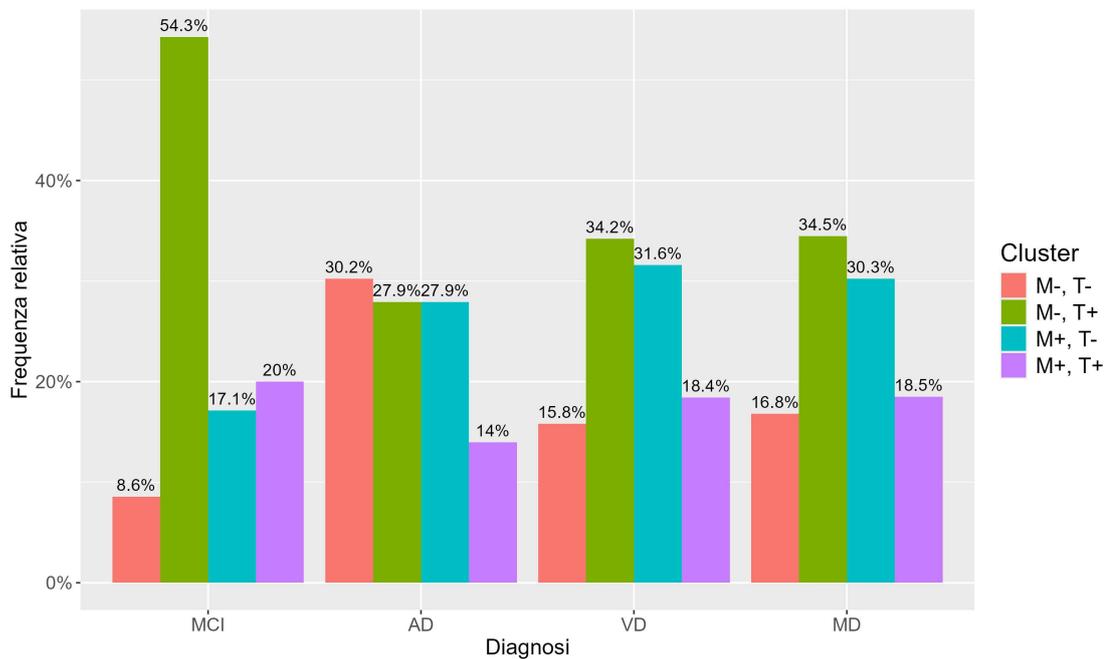


Figura A.4: Frequenze relative dei cluster nelle diagnosi.

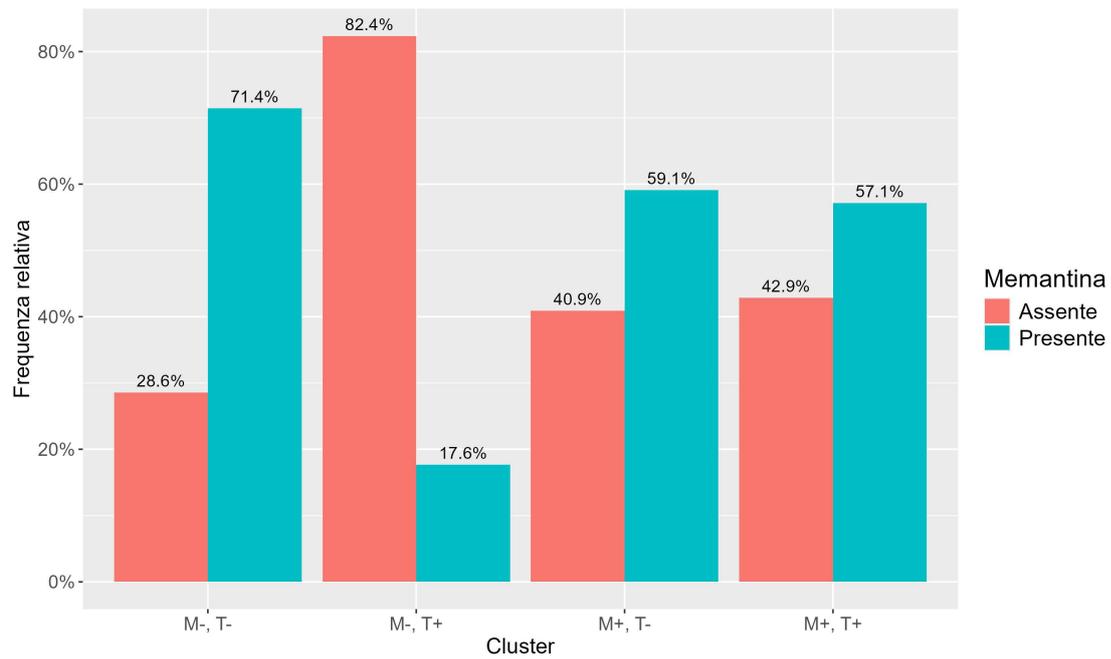


Figura A.5: Frequenze relative della memantina nei cluster.

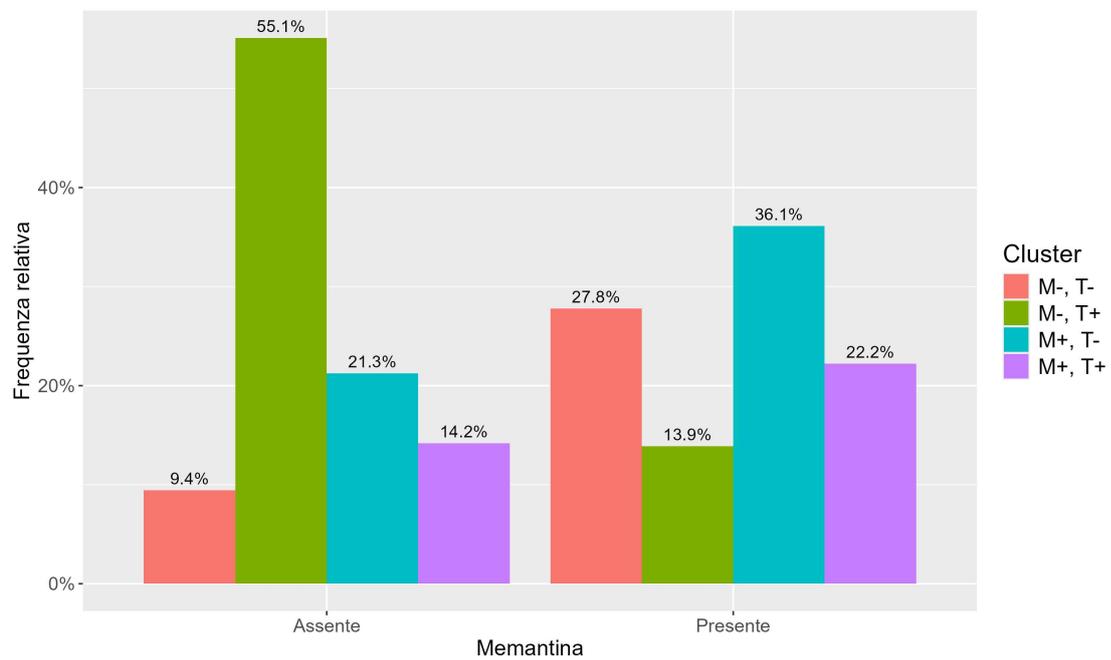


Figura A.6: Frequenze relative dei cluster nella memantina.

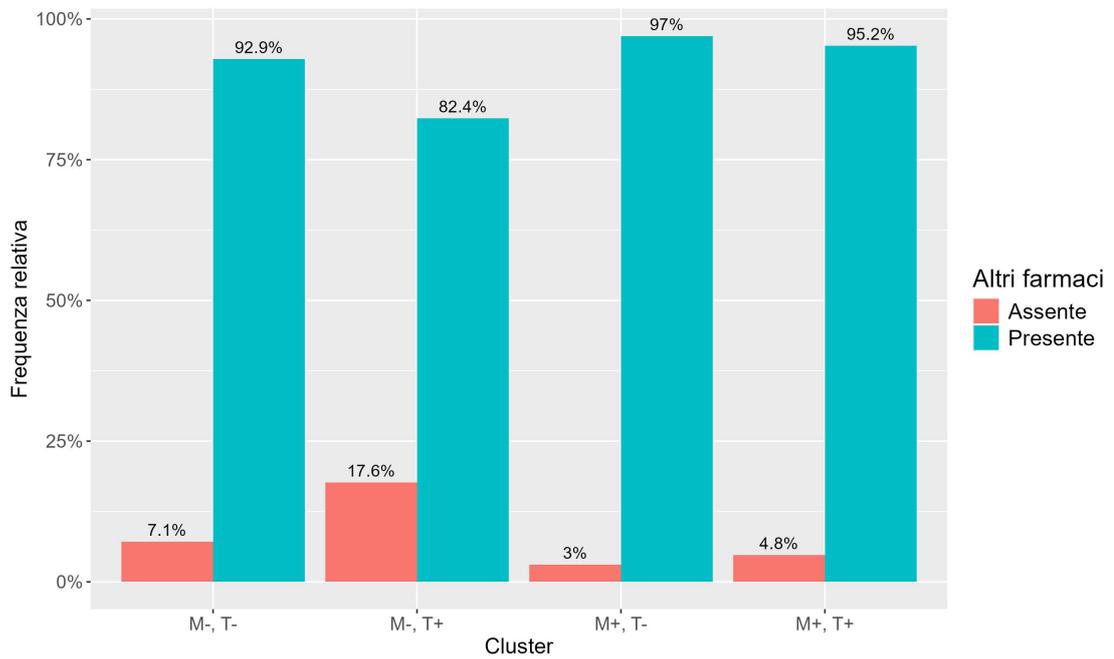


Figura A.7: Frequenze relative di altri farmaci nei cluster.

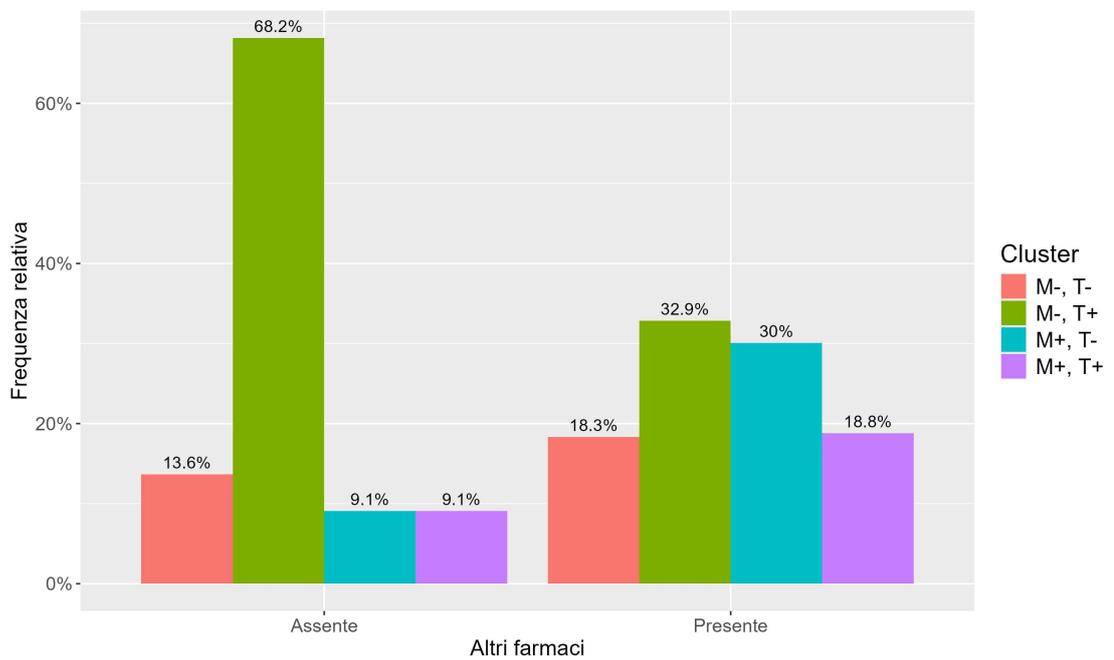


Figura A.8: Frequenze relative dei cluster in altri farmaci.

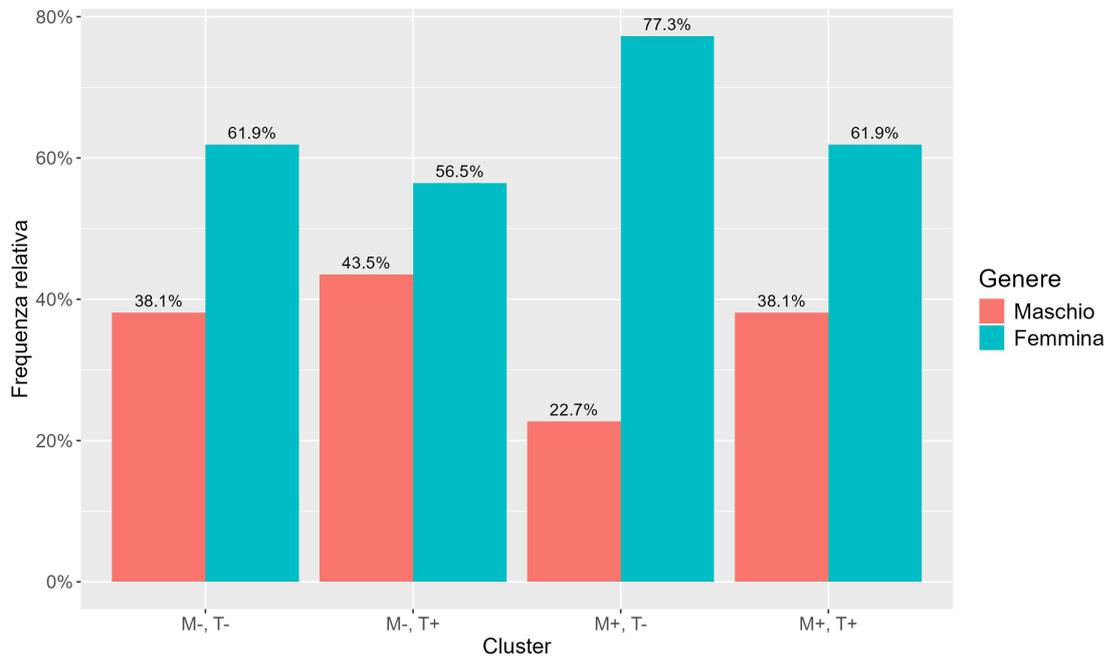


Figura A.9: Frequenze relative del genere nei cluster.

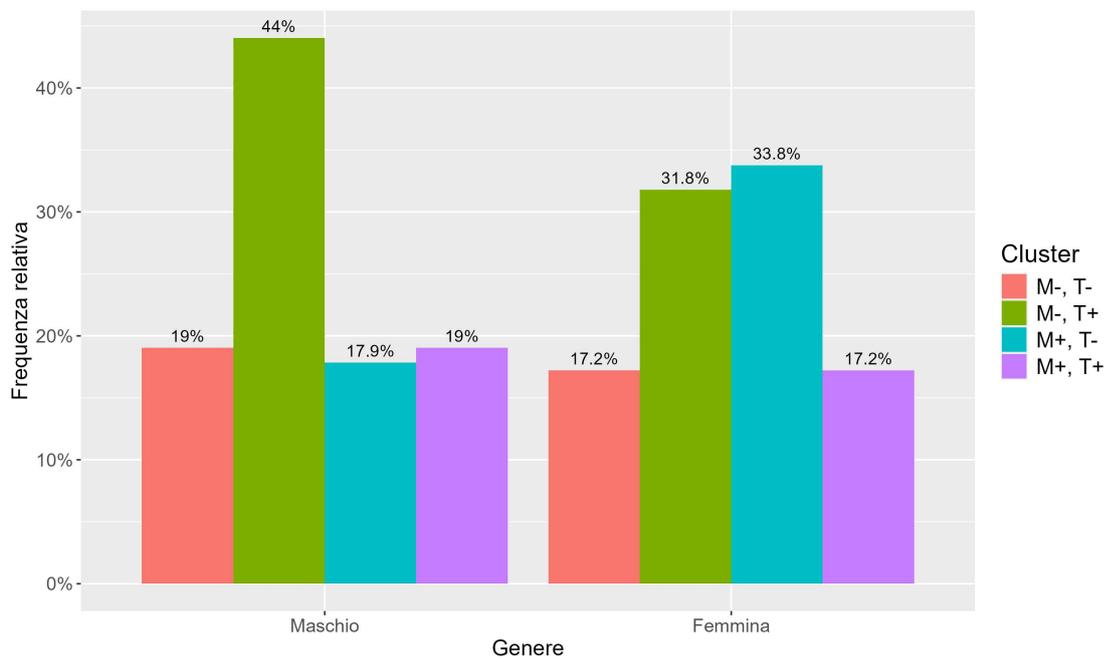


Figura A.10: Frequenze relative dei cluster nel genere.

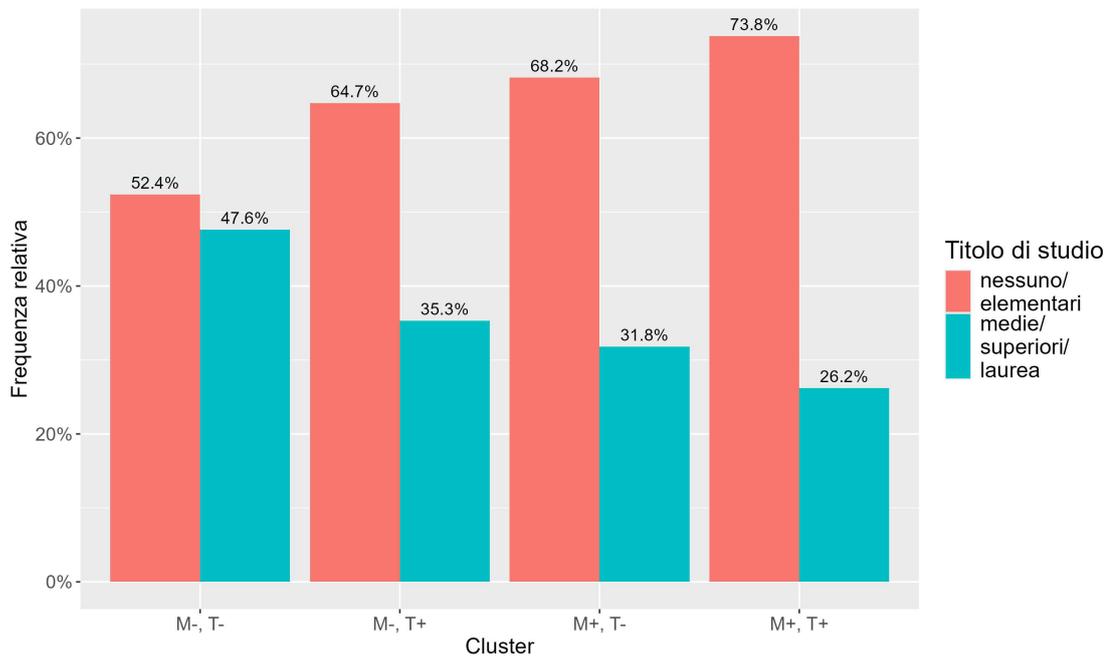


Figura A.11: Frequenze relative dell'istruzione nei cluster.

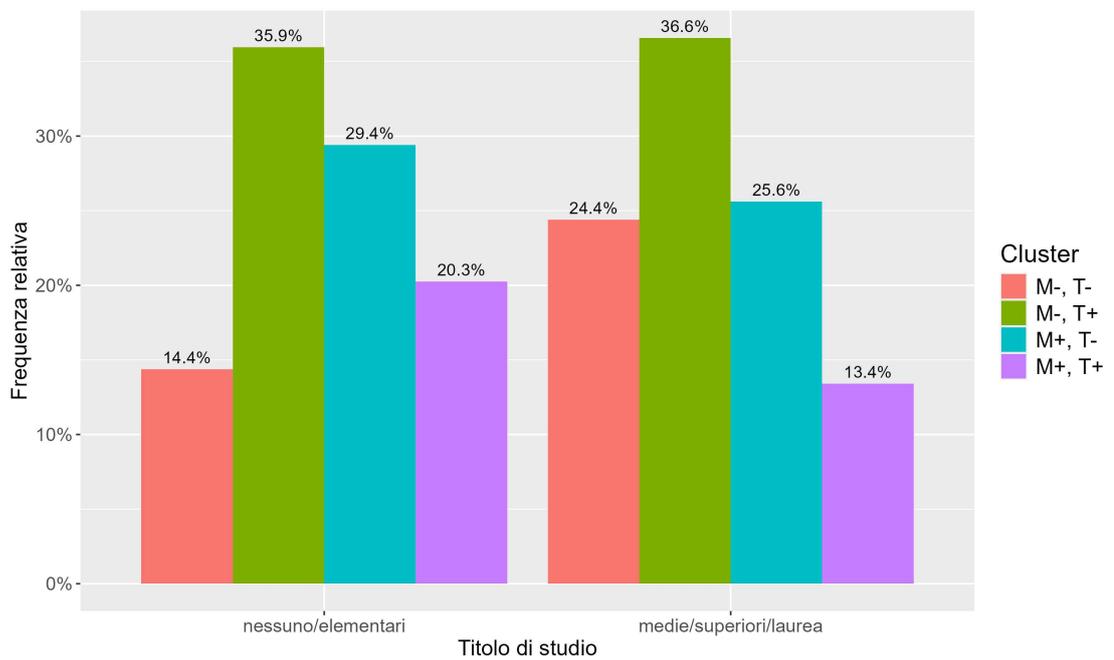


Figura A.12: Frequenze relative dei cluster nell'istruzione.

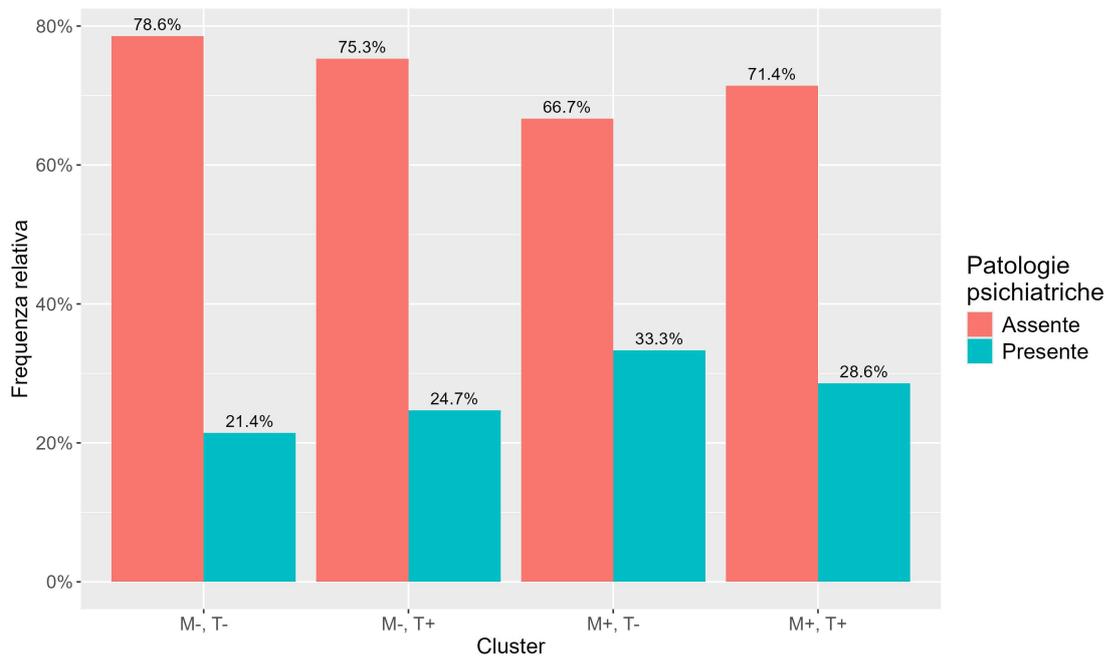


Figura A.13: Frequenze relative delle malattie psichiatriche nei cluster.

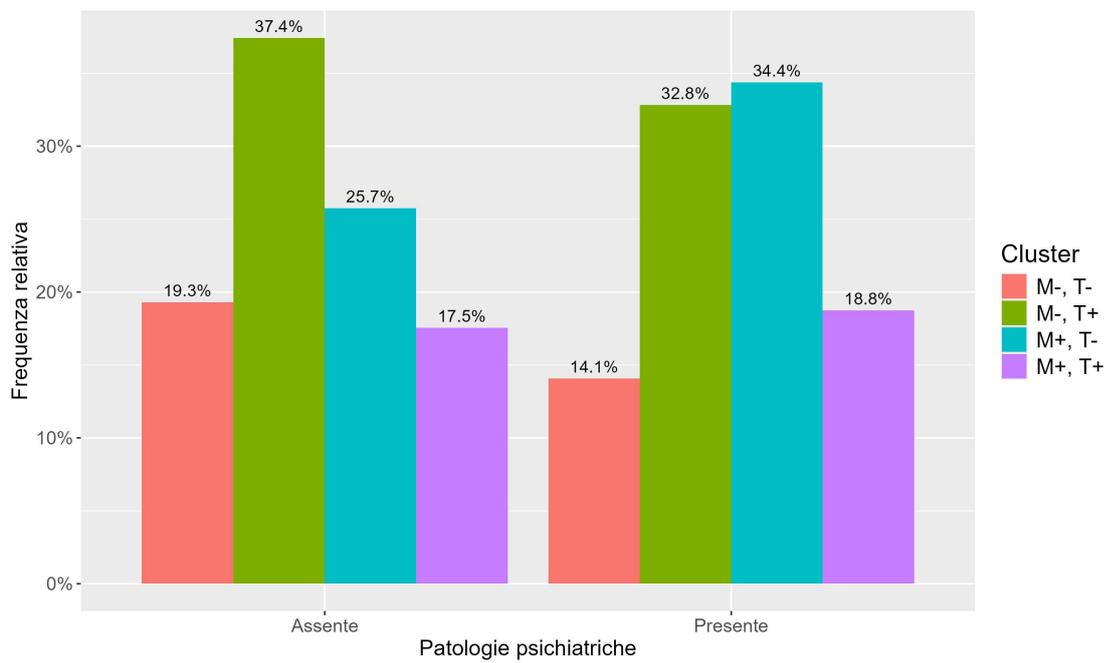


Figura A.14: Frequenze relative dei cluster nelle malattie psichiatriche.

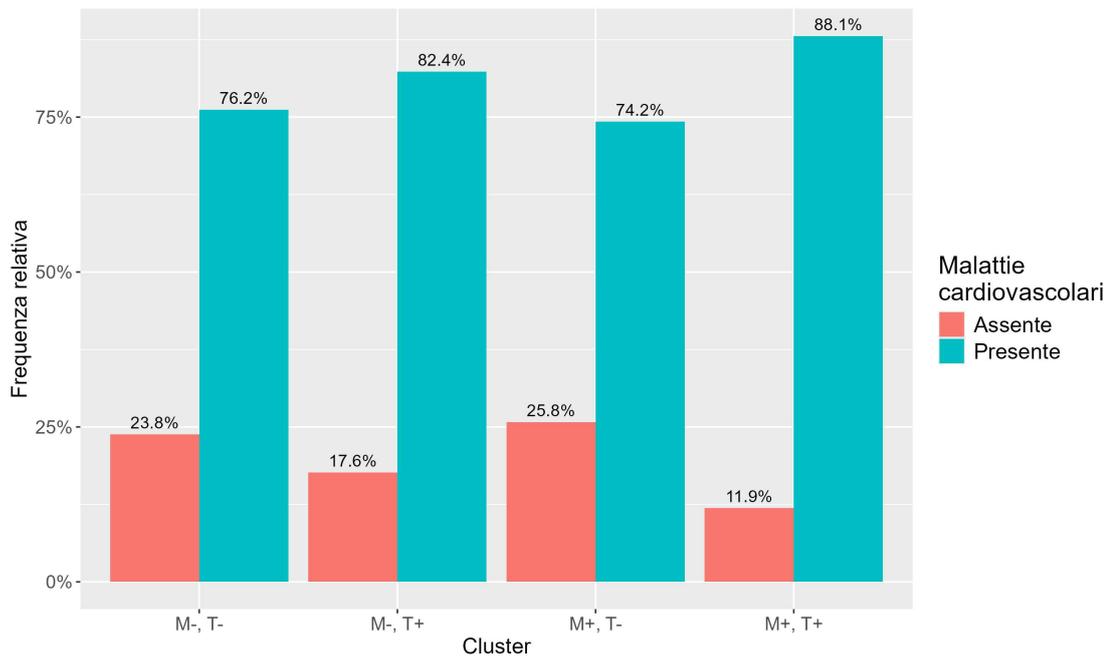


Figura A.15: Frequenze relative delle malattie cardiovascolari nei cluster.

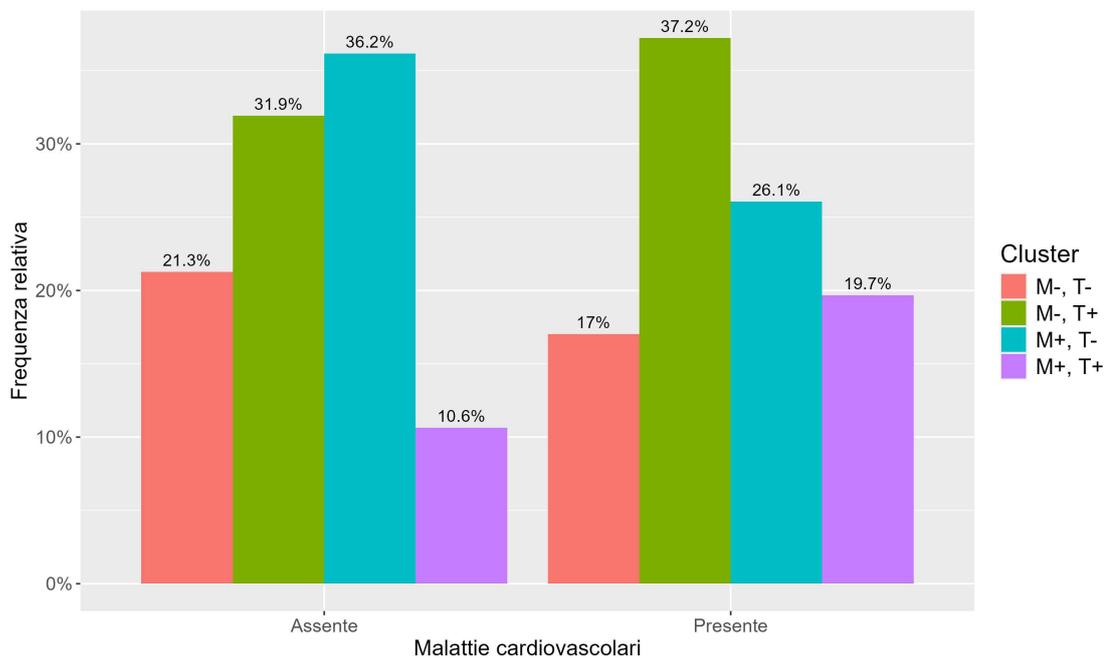


Figura A.16: Frequenze relative dei cluster nelle malattie cardiovascolari.

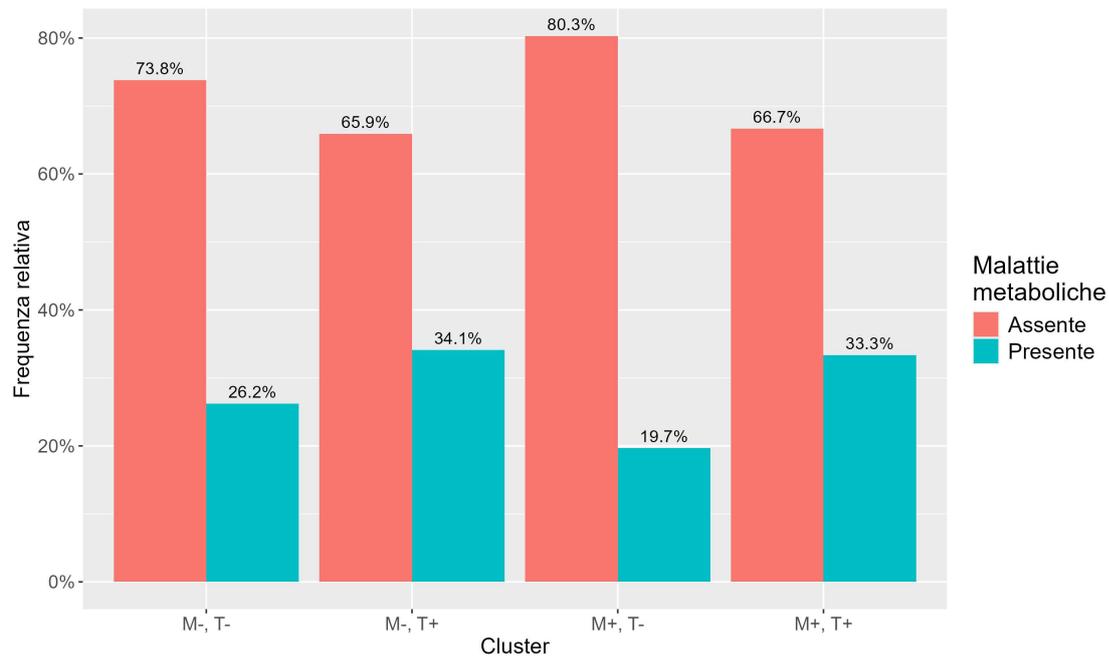


Figura A.17: Frequenze relative delle malattie metaboliche nei cluster.

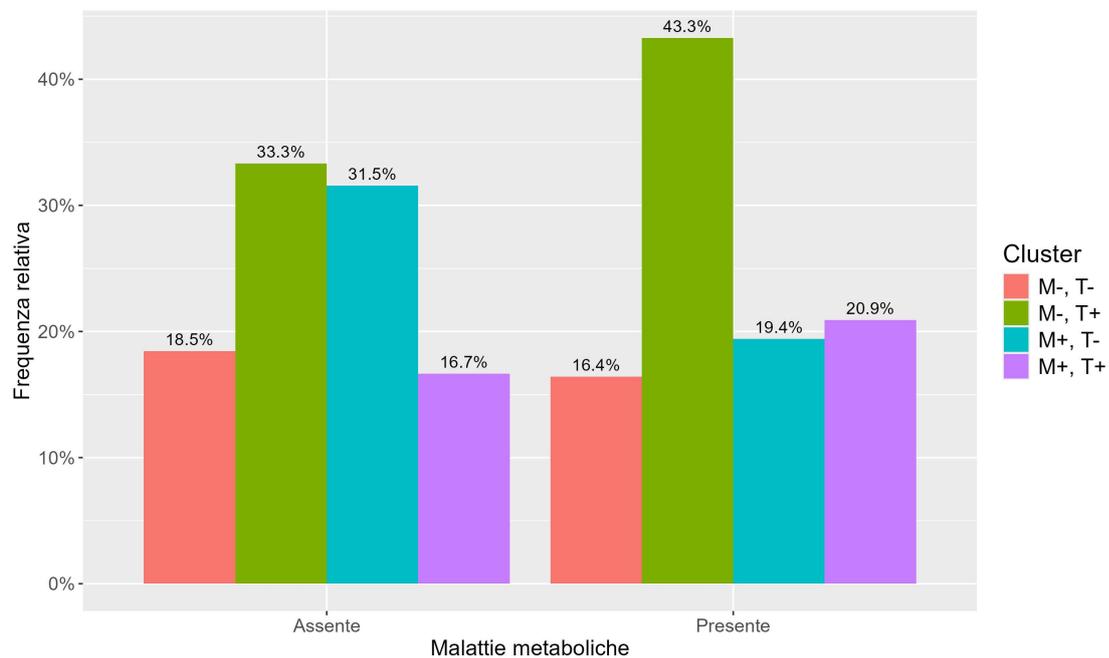


Figura A.18: Frequenze relative dei cluster nelle malattie metaboliche.

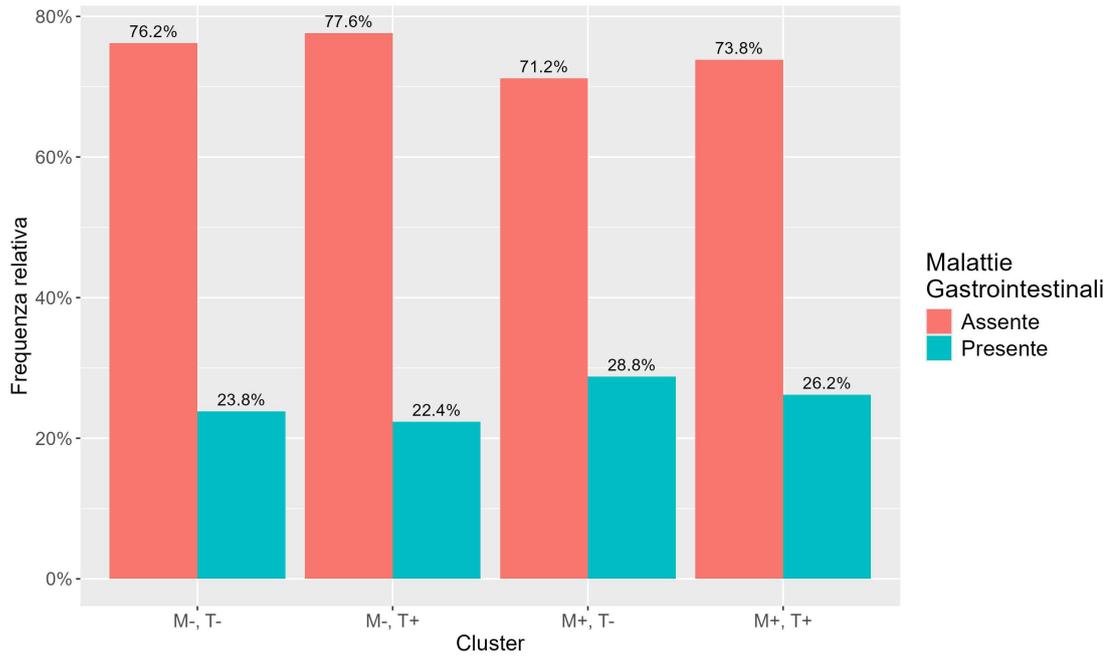


Figura A.19: Frequenze relative delle malattie gastrointestinali nei cluster.

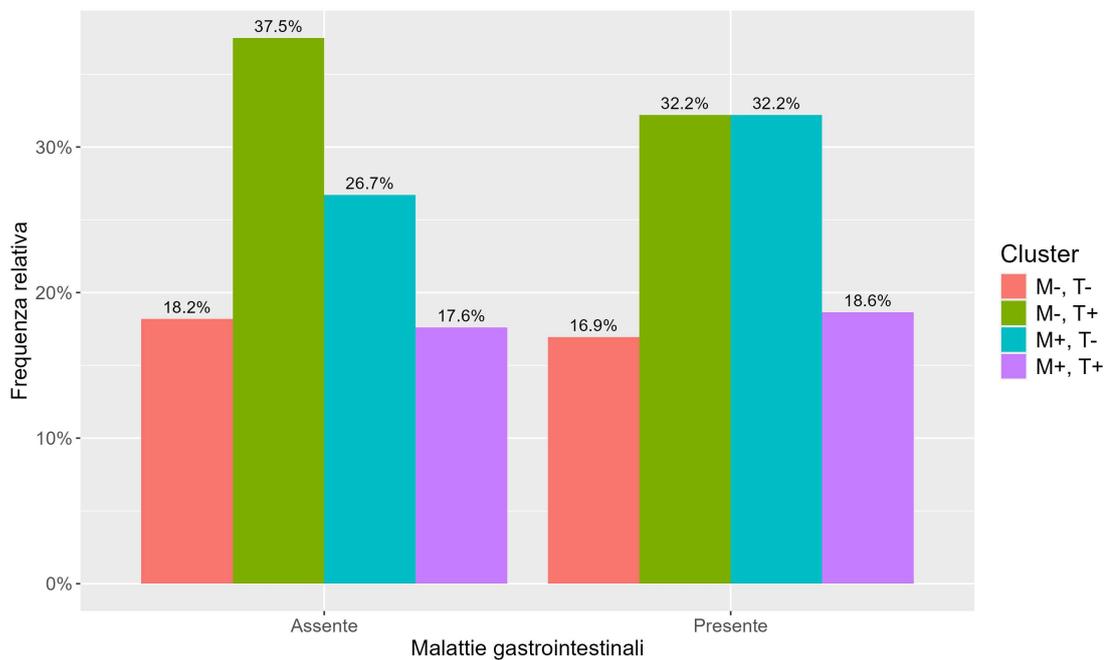


Figura A.20: Frequenze relative dei cluster nelle malattie gastrointestinali.

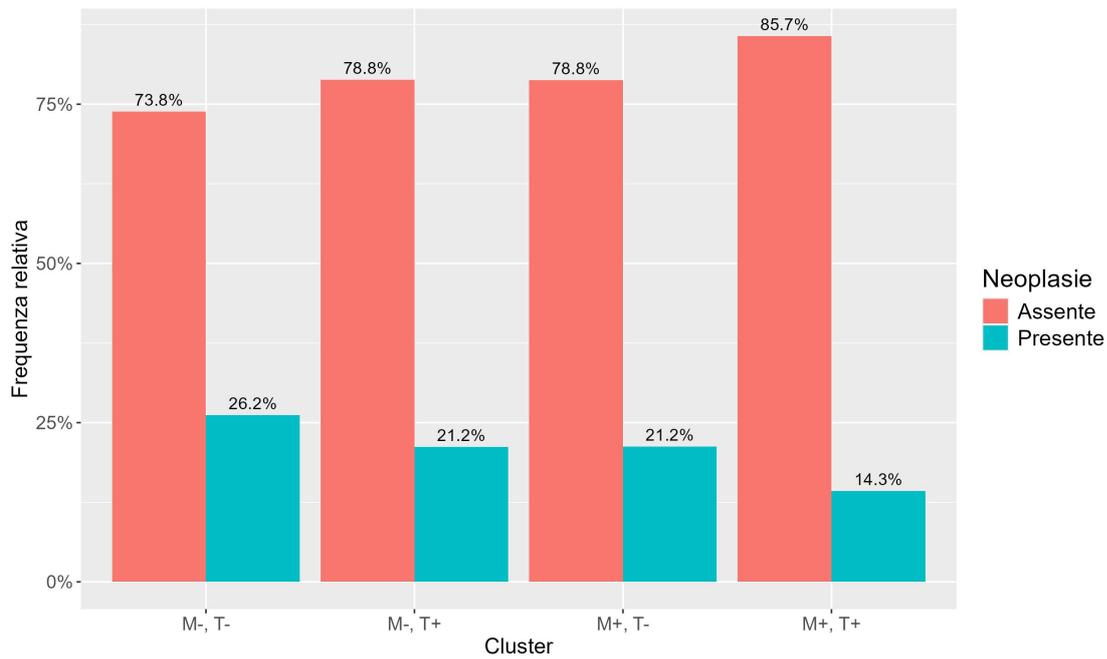


Figura A.21: Frequenze relative delle neoplasie nei cluster.

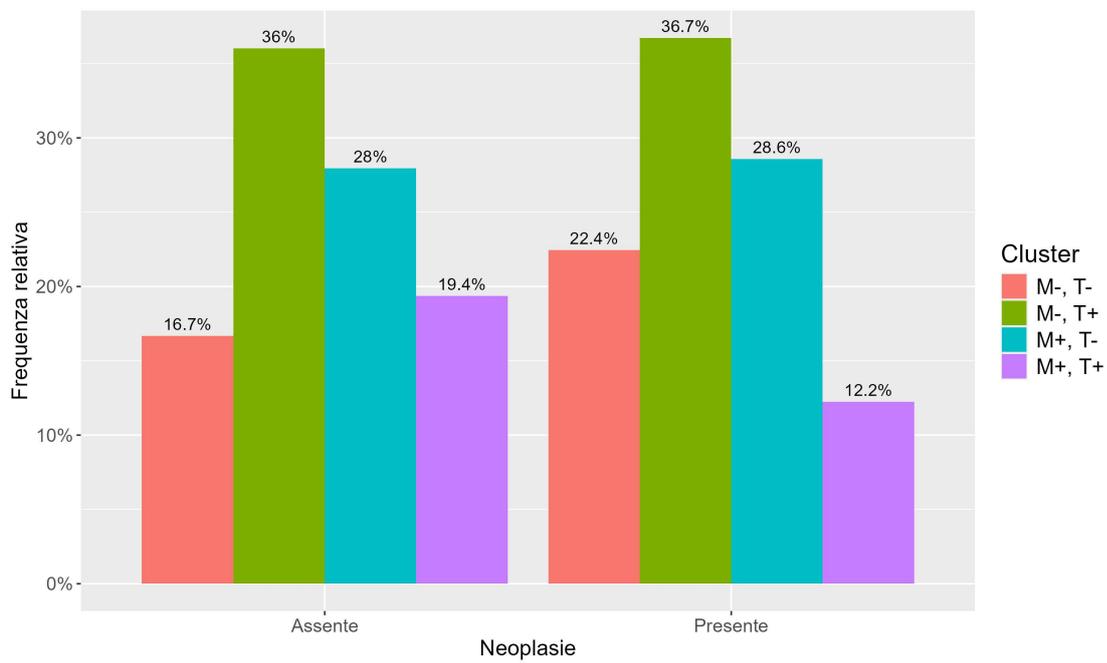


Figura A.22: Frequenze relative dei cluster nelle neoplasie.

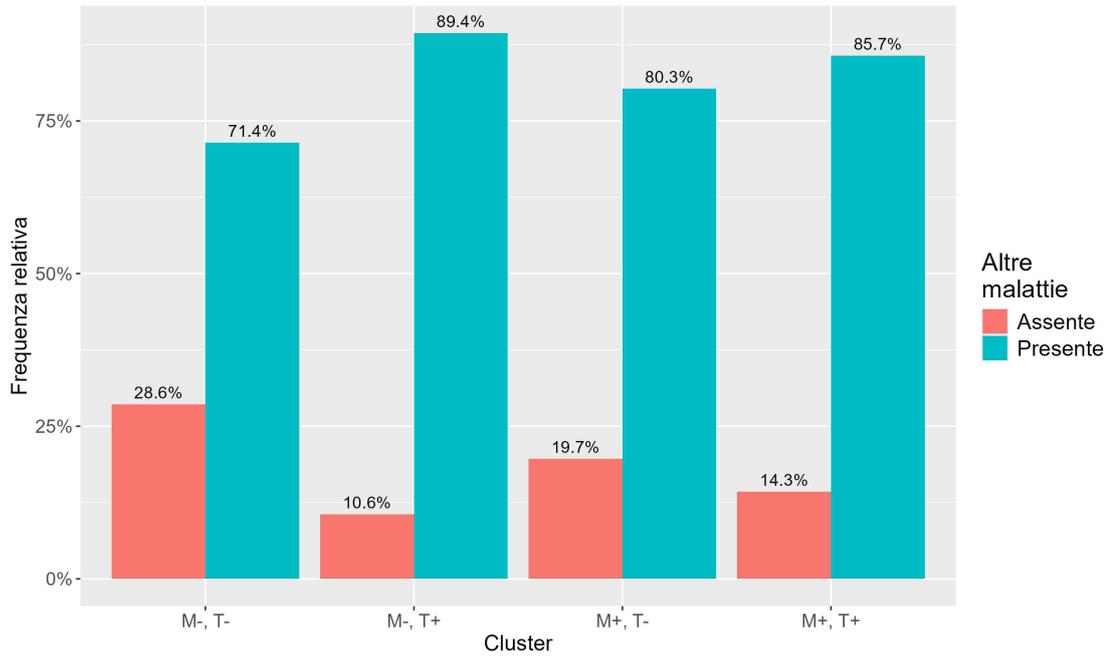


Figura A.23: Frequenze relative di altre malattie nei cluster.

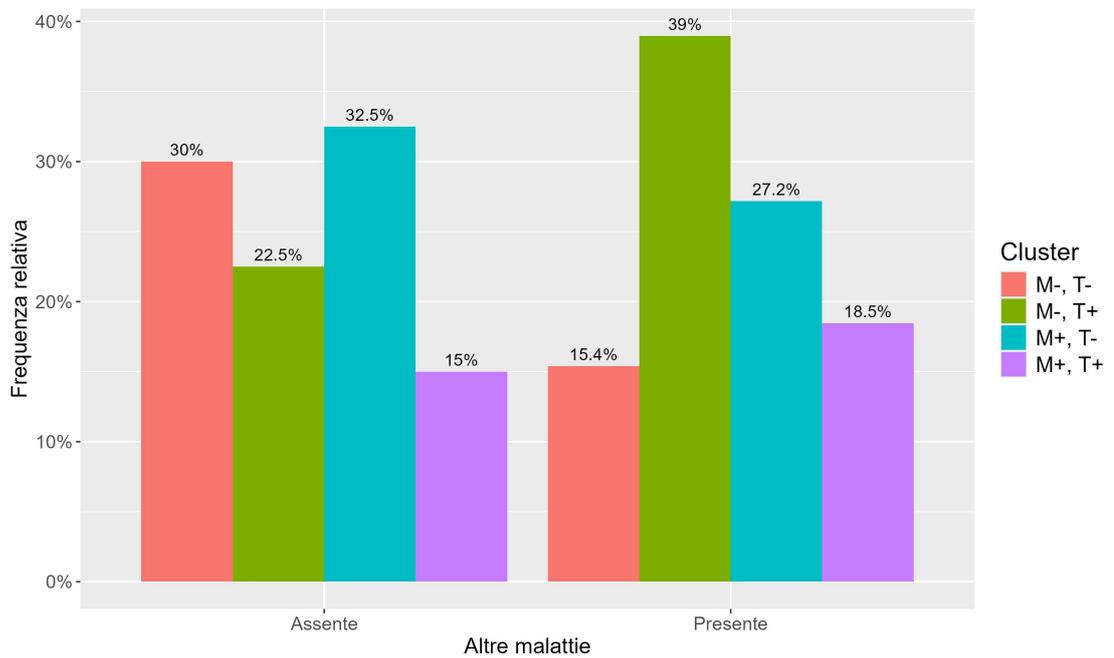


Figura A.24: Frequenze relative dei cluster in altre malattie.

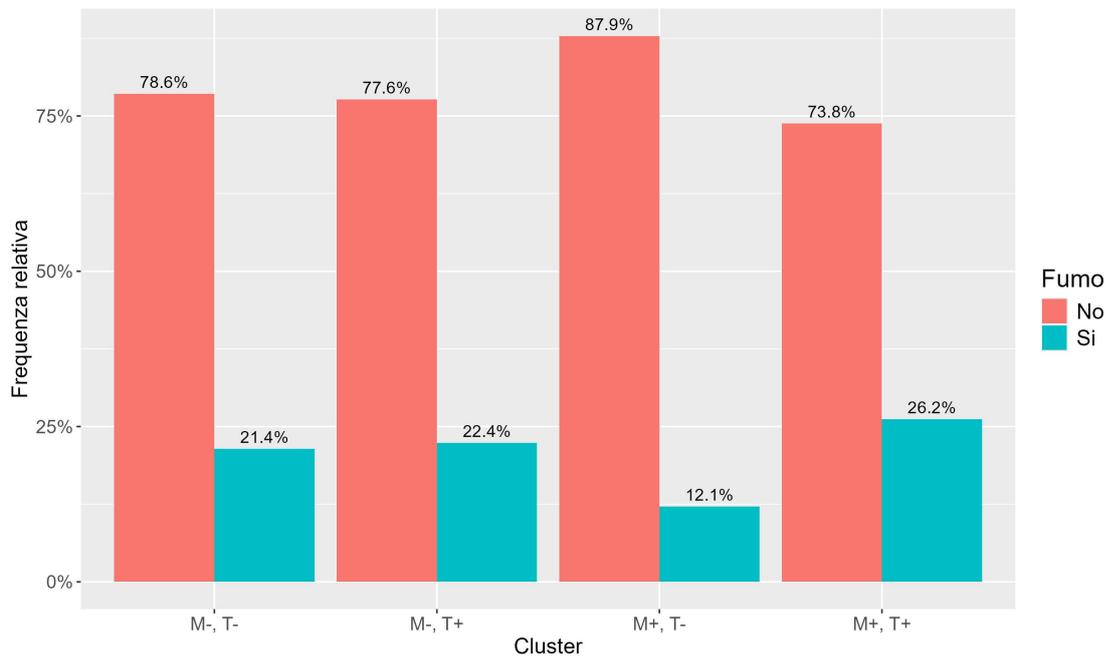


Figura A.25: Frequenze relative del fumo nei cluster.

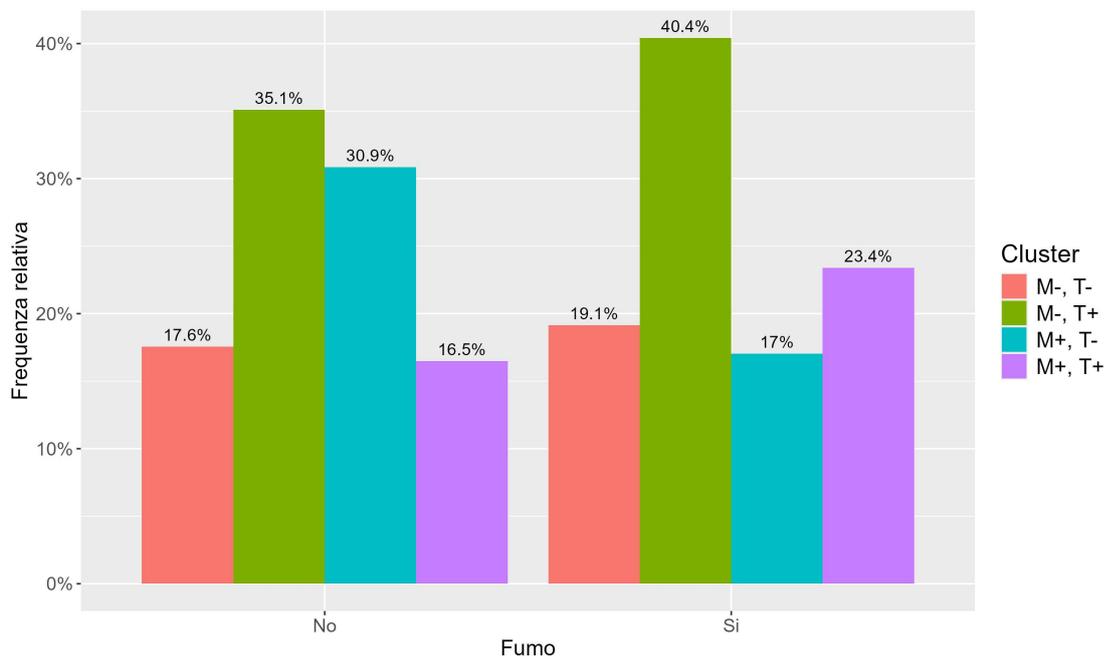


Figura A.26: Frequenze relative dei cluster nel fumo.

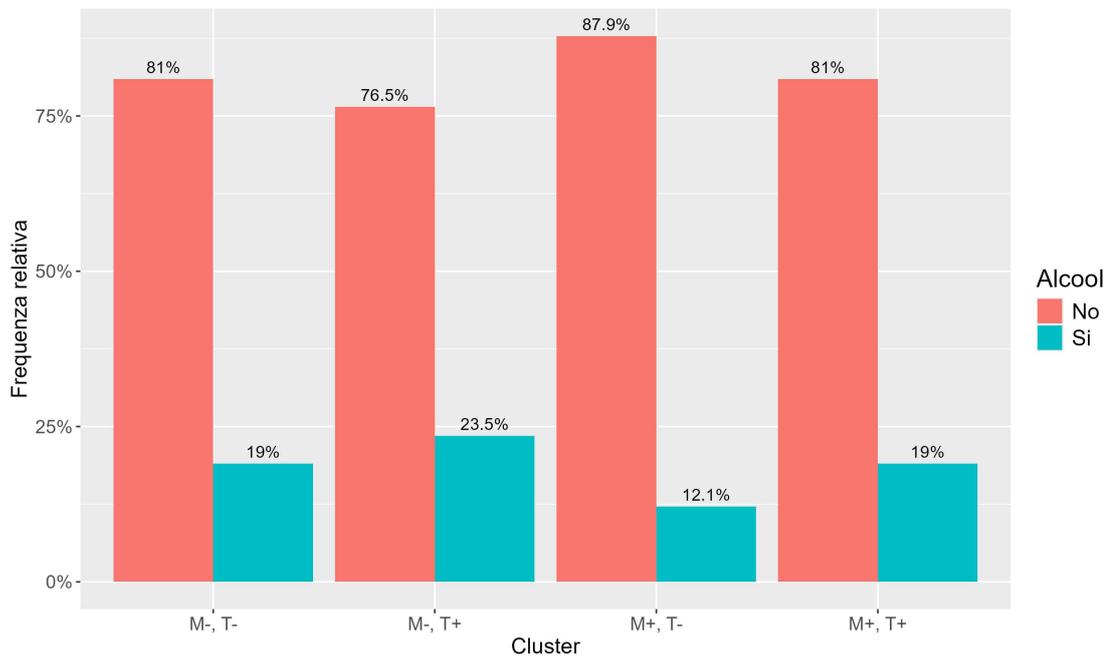


Figura A.27: Frequenze relative dell'alcool nei cluster.

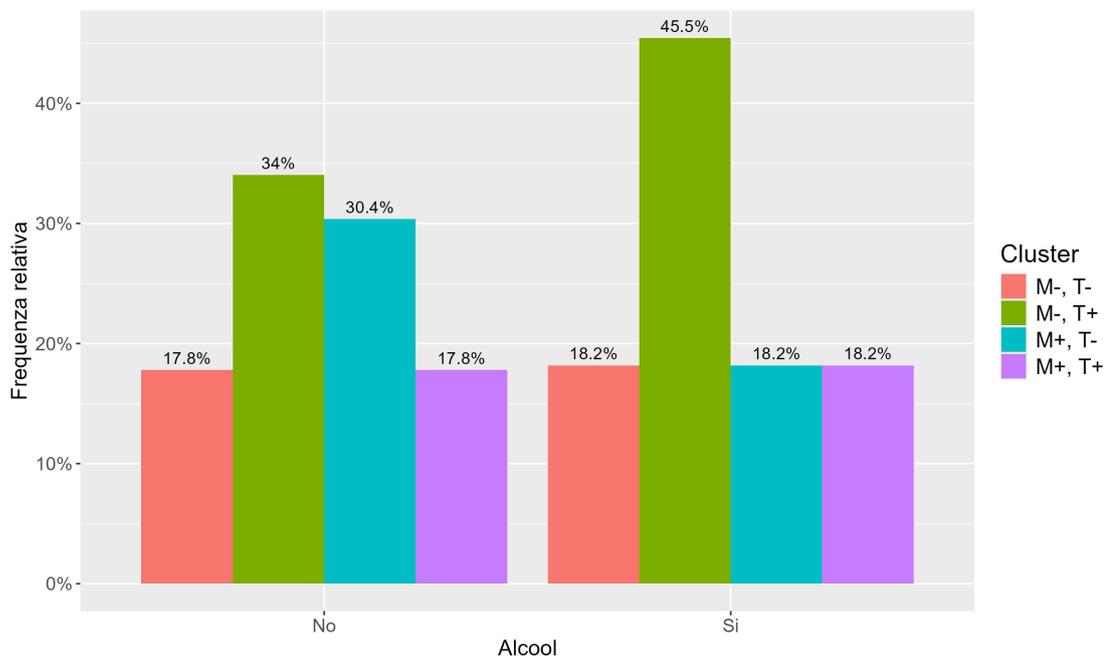


Figura A.28: Frequenze relative dei cluster nell'alcool.

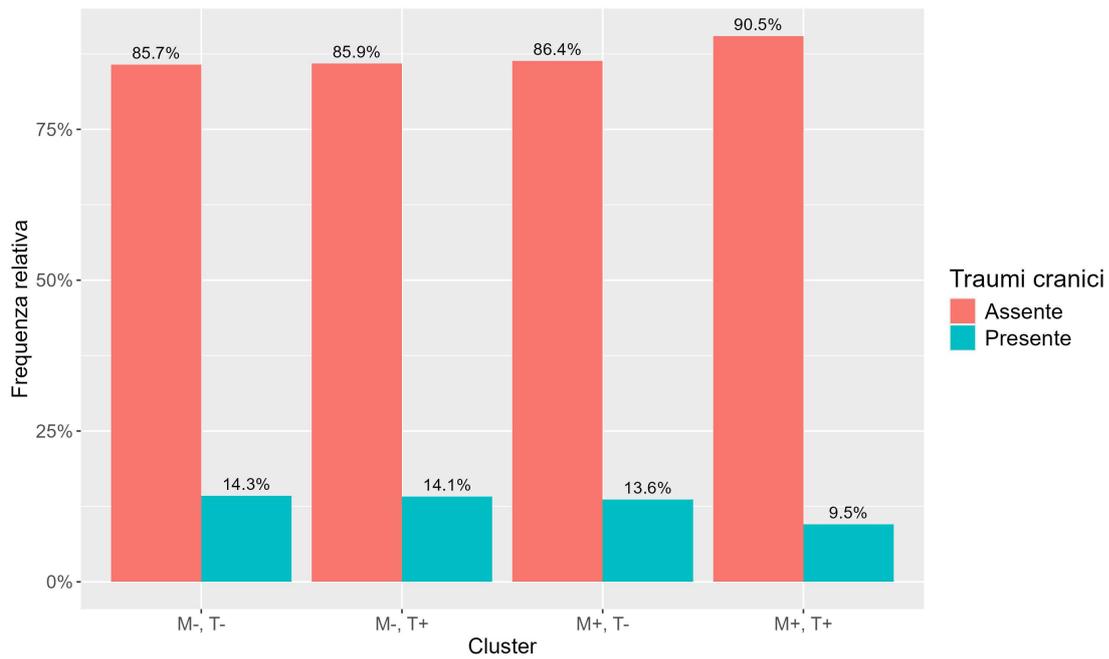


Figura A.29: Frequenze relative dei traumi cranici nei cluster.

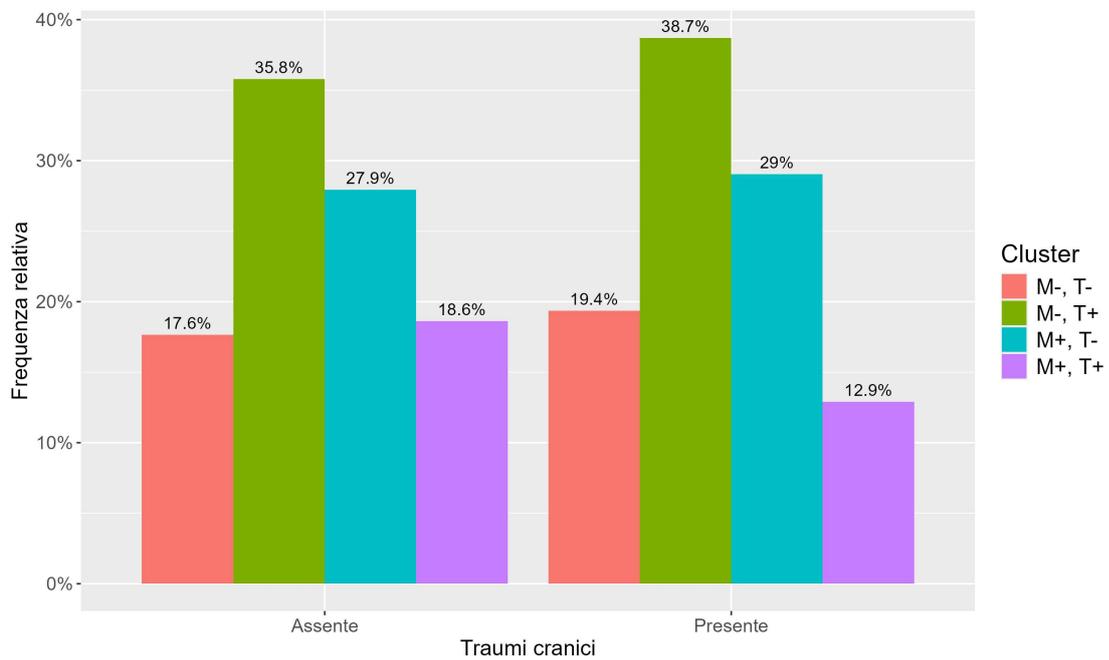


Figura A.30: Frequenze relative dei cluster nei traumi cranici.

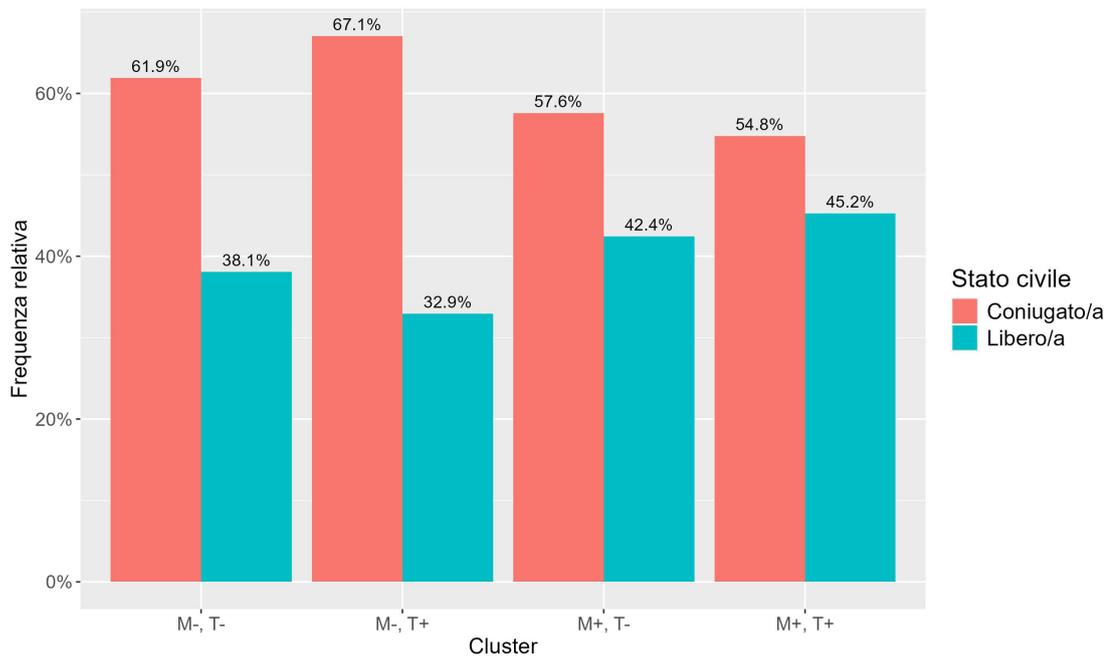


Figura A.31: Frequenze relative dello stato civile nei cluster.

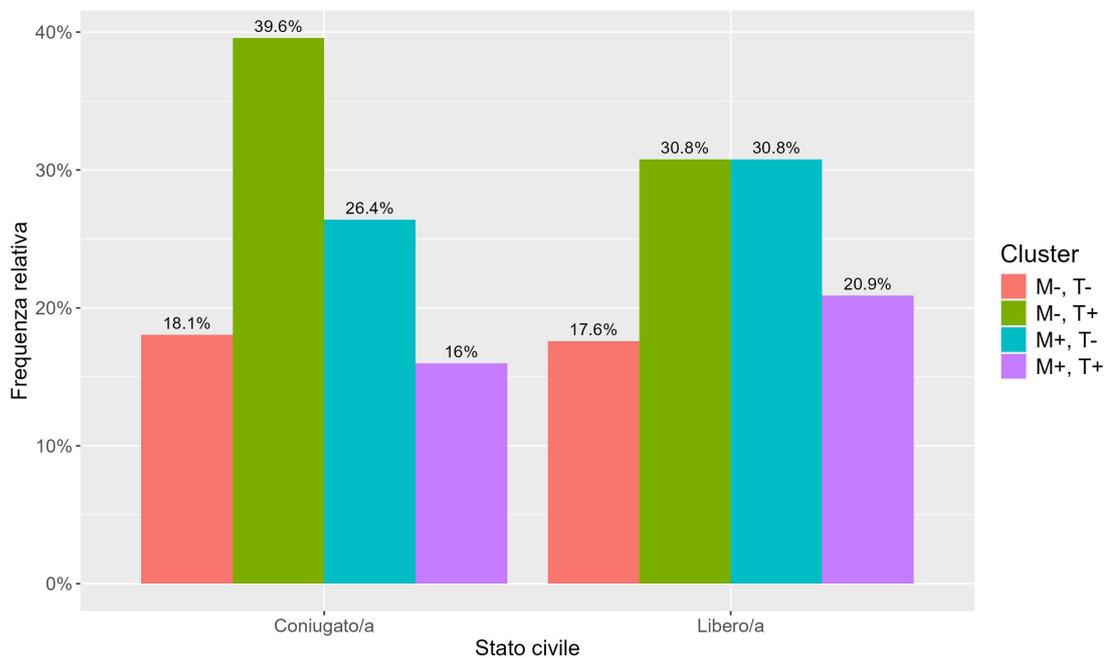


Figura A.32: Frequenze relative dei cluster nello stato civile.

Bibliografia

- [1] Douglas M Bates e José C Pinheiro. «Computational methods for multilevel modelling». In: *University of Wisconsin, Madison, WI* (1998), pp. 1–29.
- [2] Jacqueline S Birks e Richard J Harvey. «Donepezil for dementia due to Alzheimer’s disease». In: *Cochrane Database of systematic reviews* 6 (2018).
- [3] George EP Box et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [4] Henry Brodaty, Alisa Green e Annette Koschera. «Meta-analysis of psychosocial interventions for caregivers of people with dementia». In: *Journal of the American Geriatrics Society* 51.5 (2003), pp. 657–664.
- [5] Linda Clare et al. «Cognitive rehabilitation and cognitive training for early-stage Alzheimer’s disease and vascular dementia». In: *Cochrane Database Syst Rev* 4 (2003), p. CD003260.
- [6] David Roxbee Cox e David Victor Hinkley. *Theoretical statistics*. Chapman e Hall, London, 1974.
- [7] Jeffrey L Cummings et al. «The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia». In: *Neurology* 44.12 (1994), pp. 2308–2308.
- [8] Arthur P Dempster, Nan M Laird e Donald B Rubin. «Maximum likelihood from incomplete data via the EM algorithm». In: *Journal of the royal statistical society: series B (methodological)* 39.1 (1977), pp. 1–22.
- [9] Richard Fleming e Shima Sum. «Empirical studies on the effectiveness of assistive technology in the care of people with dementia: a systematic review». In: *Journal of Assistive Technologies* 8.1 (2014), pp. 14–34.
- [10] Marshal F Folstein, Susan E Folstein e Paul R McHugh. «“Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician». In: *Journal of psychiatric research* 12.3 (1975), pp. 189–198.
- [11] M Freedman. *Clock drawing: A neuropsychological analysis*. Oxford University Press, 1994.

- [12] Maud JL Graff et al. «Community based occupational therapy for patients with dementia and their care givers: randomised controlled trial». In: *Bmj* 333.7580 (2006), p. 1196.
- [13] Yesavage Ja. «Development and validation of a geriatric depression screening scale; A preliminary report». In: *J Psychiatr res* 39 (1983), pp. 37–49.
- [14] Richard H Jones. *Longitudinal data with serial correlation: a state-space approach*. Chapman e Hall/CRC, 2018.
- [15] Helen C Kales, Laura N Gitlin e Constantine G Lyketsos. «Assessment and management of behavioral and psychological symptoms of dementia». In: *Bmj* 350 (2015).
- [16] Sidney Katz et al. «Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function». In: *jama* 185.12 (1963), pp. 914–919.
- [17] Nan M Laird e James H Ware. «Random-effects models for longitudinal data». In: *Biometrics* (1982), pp. 963–974.
- [18] Eric B Larson et al. «Exercise is associated with reduced risk for incident dementia among persons 65 years of age and older». In: *Annals of internal medicine* 144.2 (2006), pp. 73–81.
- [19] M Lawton, E Brody e Utilisateurs Médecin. «Instrumental activities of daily living (IADL)». In: *gerontologist* 9 (1969), pp. 179–186.
- [20] Mary J Lindstrom e Douglas M Bates. «Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data». In: *Journal of the American Statistical Association* 83.404 (1988), pp. 1014–1022.
- [21] NT Longford. «Regression analysis of multilevel data with measurement error». In: *British Journal of Mathematical and Statistical Psychology* 46.2 (1993), pp. 301–311.
- [22] Rupert McShane, Almudena Areosa Sastre e Neda Minakaran. «Memantine for dementia». In: *Cochrane database of systematic reviews* 2 (2006).
- [23] Martha Clare Morris et al. «MIND diet associated with reduced incidence of Alzheimer’s disease». In: *Alzheimer’s & Dementia* 11.9 (2015), pp. 1007–1014.
- [24] CG Parsons, W Danysz e G Quack. «Memantine is a clinically well tolerated N-methyl-D-aspartate (NMDA) receptor antagonist—a review of preclinical data». In: *Neuropharmacology* 38.6 (1999), pp. 735–767.
- [25] José Pinheiro e Douglas Bates. *Mixed-effects models in S and S-PLUS*. Springer science & business media, 2006.

- [26] José Pinheiro, Douglas Bates e R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-164. 2023. URL: <https://CRAN.R-project.org/package=nlme>.
- [27] Bruce G Pollock et al. «Comparison of citalopram, perphenazine, and placebo for the acute treatment of psychosis and behavioral disturbances in hospitalized, demented patients». In: *American Journal of Psychiatry* 159.3 (2002), pp. 460–465.
- [28] Chengxuan Qiu, Bengt Winblad e Laura Fratiglioni. «The age-dependent relation of blood pressure to cognitive function and dementia». In: *The Lancet Neurology* 4.8 (2005), pp. 487–499.
- [29] Lon S Schneider et al. «Effectiveness of atypical antipsychotic drugs in patients with Alzheimer’s disease». In: *New England Journal of Medicine* 355.15 (2006), pp. 1525–1538.
- [30] Kenneth I Shulman. «Clock-drawing: is it the ideal cognitive screening test?». In: *International journal of geriatric psychiatry* 15.6 (2000), pp. 548–561.
- [31] Aimee Spector et al. «Efficacy of an evidence-based cognitive stimulation therapy programme for people with dementia: randomised controlled trial». In: *The British Journal of Psychiatry* 183.3 (2003), pp. 248–254.
- [32] Linda Teri et al. «Behavioral treatment of depression in dementia patients: a controlled clinical trial». In: *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 52.4 (1997), P159–P166.
- [33] Terry Therneau e Beth Atkinson. *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1.23. 2023. URL: <https://CRAN.R-project.org/package=rpart>.
- [34] Ronald Aaron Thisted. *Elements of statistical computing*. Chapman e Hall, London, 1988.

Sitografia

- [35] Francesca Fiorillo. *Decadimento cognitivo e anziani: quando il cervello perde i colpi*. Accessed: 2024-11-04. 2019. URL: <https://psiche.santagostino.it/decadimento-cognitivo-anziani/>.
- [36] La Mente è Meravigliosa. *Fasi della vecchiaia: cambiamenti fisici e psicologici*. Accessed: 2024-11-04. 2024. URL: <https://lamenteemeravigliosa.it/fasi-vecchiaia-cambiamenti-fisici-psicologici/>.
- [37] Medicomunicare Magazine. *Invecchiamento del Corpo Umano: Cambiamenti Drammatici a Due Specifiche Età*. Accessed: 2024-11-03. 2024. URL: <https://www.medicomunicare.it/2024/08/26/invecchiamento-del-corpo-umano-cambiamenti-drammatici-a-due-specifiche-eta-secondo-uno-studio/>.
- [38] Valore in RSA. *Declino funzionale: definizione e strumenti di valutazione*. Accessed: 2024-11-04. 2024. URL: <https://www.valoreinrsa.it/strumenti-di-lavoro/declino-funzionale-e-cognitivo/157-valutazione-declino-funzionale-cognitivo-rsa-toscana.html>.