

---

# ANALISI STATISTICA SUI DECESSI PREMATURI DOVUTI AL FUMO PASSIVO

---

Statistica e Analisi dei Dati - UNISA

Cascone Giovanni (0522511934)  
Vitale Ciro (0522501759)

# Contents

<b>1</b>	<b>Introduzione</b>	<b>4</b>
<b>2</b>	<b>Analisi Esplorativa del Dataset</b>	<b>5</b>
2.1	Introduzione al Dataset . . . . .	5
2.2	Misure ed Indici Statistici . . . . .	5
2.2.1	Indici di Centralità . . . . .	5
2.2.2	Indici di Dispersione . . . . .	6
2.3	Quartili . . . . .	8
2.3.1	Boxplot per DPM . . . . .	8
2.3.2	Boxplot per VSL . . . . .	11
<b>3</b>	<b>Statistica Descrittiva Univariata</b>	<b>14</b>
3.1	Funzione di Distribuzione Empirica Continua . . . . .	14
3.1.1	Conclusioni . . . . .	18
3.2	Serie Temporalì . . . . .	18
3.2.1	Serie Temporale - DPM . . . . .	18
3.2.2	Serie Temporale - VSL . . . . .	22
3.2.3	Conclusioni . . . . .	24
3.3	Grafici a Barre . . . . .	24
3.3.1	Conclusioni . . . . .	29
<b>4</b>	<b>Statistica Descrittiva Bivariata</b>	<b>30</b>
4.1	Regressione . . . . .	30
4.1.1	Conclusioni . . . . .	36
<b>5</b>	<b>Analisi dei Cluster</b>	<b>37</b>
5.1	DPM - Metodi Gerarchici . . . . .	37
5.1.1	Metodo del Legame Singolo . . . . .	37
5.1.2	Metodo del Legame Completo . . . . .	39
5.1.3	Metodo del Legame Medio . . . . .	41
5.1.4	Metodo del Centroide . . . . .	43
5.1.5	Metodo della Mediana . . . . .	44
5.2	DPM - Metodi Non Gerarchici . . . . .	46
5.2.1	Metodo kmeans . . . . .	46
5.3	VSL - Metodi Gerarchici . . . . .	47
5.3.1	Metodo del Legame Singolo . . . . .	47
5.3.2	Metodo del Legame Completo . . . . .	49
5.3.3	Metodo del Legame Medio . . . . .	51
5.3.4	Metodo del Centroide . . . . .	53
5.3.5	Metodo della Mediana . . . . .	54
5.4	VSL - Metodi Non Gerarchici . . . . .	56
5.4.1	Metodo kmeans . . . . .	56
5.5	Conclusioni . . . . .	57

<b>6</b>	<b>Inferenza Statistica</b>	<b>58</b>
6.1	Stima Puntuale . . . . .	58
6.1.1	Criterio del chi2 . . . . .	58
6.1.2	Metodo dei Momenti . . . . .	59
6.1.3	Conclusione . . . . .	59
6.2	Intervalli di Confidenza . . . . .	60
6.2.1	Metodo Pivotale . . . . .	60
6.2.2	Conclusioni . . . . .	62
6.3	Confronti tra Popolazioni . . . . .	62
6.3.1	Intervallo di confidenza per $\mu_1 - \mu_2$ con $\sigma_1^2$ e $\sigma_2^2$ note . . . . .	63
6.3.2	Conclusioni . . . . .	63
6.4	Verifica delle Ipotesi . . . . .	63
6.4.1	Conclusioni . . . . .	65
<b>7</b>	<b>Avvenimenti storici</b>	<b>66</b>
7.1	Crisi Finanziaria 2007-2008 . . . . .	66
7.2	Legge Sirchia (Italia) . . . . .	66
<b>8</b>	<b>Conclusioni</b>	<b>67</b>

# 1 Introduzione

Il presente studio vuole approfondire l'*impatto che gli investimenti sulla salute pubblica hanno sulla mortalità*, ponendo l'attenzione sulla **problematica del fumo passivo** tra i *paesi dell'Unione Europea*. Esplorando in profondità i dati raccolti, lo studio, attraverso *varie metodologie analitiche*, si propone di comprendere l'andamento del fenomeno analizzato in relazione agli *investimenti effettuati* dai paesi in politiche di salute pubblica. L'*obiettivo* è determinare se e in che modo le variazioni negli investimenti possano influenzare il numero di decessi. Inoltre, particolare attenzione è rivolta anche alle *normative* e agli *eventi storici*, valutando il loro impatto sul fenomeno in considerazione.

*Infine, questo approccio offre una panoramica generale sulla situazione europea relativa ai decessi attribuibili al fumo passivo e i fattori che li influenzano.*

## 2 Analisi Esplorativa del Dataset

### 2.1 Introduzione al Dataset

Lo studio si concentra sull'analisi di un dataset che cataloga i decessi prematuri attribuibili all'esposizione al fumo passivo (**DPCFP**) in diversi paesi del mondo, coprendo un arco temporale che va dal 1995 al 2018. La ricerca mira a fornire un quadro dettagliato dell'impatto globale del fumo passivo sulla mortalità, utilizzando i dati forniti dall'*OECD (Organisation for Economic Co-operation and Development)*.

Per approfondire lo studio, sono stati selezionati *due indicatori principali*:

- **Numero di Decessi (DPM)**: misura che indica il numero di DPCFP per milione di abitanti, permettendo una valutazione equa dell'impatto del fumo passivo sulla mortalità nei vari paesi, indipendentemente dalle loro dimensioni demografiche. La normalizzazione per milione di abitanti facilita confronti accurati e significativi tra paesi con popolazioni notevolmente diverse.
- **Value of a Statistical Life (VSL)**: misura che indica il valore economico attribuito alla prevenzione di una morte, in questo contesto specifico causata dal fumo passivo. Questo valore è espresso in milioni di USD PPP (Parità di Potere d'Acquisto), utilizzando tassi di conversione che armonizzano il potere d'acquisto di diverse valute, offrendo così una base di confronto più uniforme a livello internazionale.

Entrambi gli indicatori sono espressi da numeri reali. Inoltre, è importante specificare che il dataset **non presenta dati mancanti**, garantendo così l'integrità e la continuità dell'analisi. Il dataset è strutturato in modo da offrire dettagli per anno e per nazione, includendo anche la possibilità di suddividere i dati in base a fasce di età (sotto i 15 anni, tra 15 e 64 anni, oltre 64 anni) e genere (maschile o femminile), permettendo un'esplorazione dettagliata delle varie dinamiche che caratterizzano l'impatto del fumo passivo. È importante notare che, mentre per l'indice DPM è possibile effettuare tali suddivisioni, il valore di VSL rimane costante per ciascuna nazione, non differenziandosi per età o genere.

Nell'ambito della ricerca, è stato deciso di focalizzare lo studio su un sottoinsieme di dati che include esclusivamente i **paesi membri dell'Unione Europea**. Il dataset finale è, quindi, organizzato per nazione, con *27 entità rappresentate*, e per anno, con un arco di *24 anni analizzati*, risultando in un totale di *648 osservazioni paese-Anno*. Per facilitare l'analisi e la manipolazione dei dati, il dataset è stato suddiviso in *due matrici principali distinte*, una per l'indice DPM e una per il VSL. Questa divisione è stata, successivamente, applicata anche per le diverse fasce di età, allo scopo di semplificare ulteriormente l'esame delle dinamiche in gioco.

### 2.2 Misure ed Indici Statistici

#### 2.2.1 Indici di Centralità

Di seguito sono illustrate due tabelle che illustrano gli indici di centralità per entrambe le variabili in esame. Questi indici, fondamentali per l'analisi statistica, includono valore minimo, valore massimo, media e mediana per ciascuna categoria di età selezionata e per ciascun indicatore, offrendo una panoramica quantitativa sull'impatto del fumo passivo e sulla valutazione economica della vita umana nei contesti analizzati.

Attraverso l'analisi di questi indici, emergono informazioni preziose riguardanti l'incidenza del fumo passivo sulla mortalità (DPM) e l'importanza economica assegnata alla vita umana (VSL), evidenziando in particolare quelle categorie di età maggiormente vulnerabili o esposte a rischi più elevati e, inoltre, mettendo in luce le priorità e le aree di maggior bisogno di azione per la tutela della salute pubblica.

DPM	MIN	MAX	MEDIA	MEDIANA
COMPLETO	59.750	600.493	199.785	164.021
< 15 ANNI	0.202	100.264	6.767	2.131
TRA 16 E 64	23.866	285.571	85.229	69.326
> 64 ANNI	202.083	2512.635	901.078	800.962

Table 1: Indici di Centralità (DPM)

VSL	MIN	MAX	MEDIA	MEDIANA
COMPLETO	1.035	9.092	3.442	3.270

Table 2: Indici di Centralità (VSL)

L'esame degli indici di centralità mostra una diversificazione significativa tra le categorie di età per quanto riguarda DPM. In particolare, per quanto riguarda l'indicatore DPM, si osserva che:

- **i dataset risultano sbilanciati verso destra**, mostrando un'asimmetria positiva, poiché le medie presentano valori maggiori delle mediane per ognuna delle fasce di età. In particolare, si denota un rapporto di circa 1.2 per ogni categoria tranne per la categoria <15 ANNI, dove il rapporto è circa 3.2: in questo caso, lo sbilanciamento risulta molto più accentuato.
- **fascia d'età superiore ai 64 anni**: registra il numero più elevato di decessi, con una media di 901.078 DPM, indicando un impatto particolarmente grave del fumo passivo su questa popolazione. Questo dato, confrontato con la media complessiva di 199.785 DPM, sottolinea come l'anzianità sia un fattore di rischio preponderante.
- **fasce d'età inferiore a 15 anni e tra 16 e 64 anni**: registrano valori medi di DPM nettamente inferiori, rispettivamente 6.767 e 85.229, rivelando un rischio minore ma non trascurabile.

Invece, per quanto riguarda VSL:

- **il dataset risulta leggermente sbilanciato verso destra**, poiché presenta una media di 3.442 milioni di USD PPP, con una mediana di 3.270 milioni di USD PPP, suggerendo una distribuzione dei valori relativamente equilibrata, benché con una lieve asimmetria verso destra.

## 2.2.2 Indici di Dispersione

Di seguito sono presentate due tabelle indicanti gli indici di dispersione, essenziali per una comprensione approfondita della variabilità dei dati, includono il Coefficiente di Variazione

(CV), lo Scarto Interquartile, la Varianza e la Deviazione Standard per ciascuna categoria di età e per ogni indicatore considerato. L'attenzione a tali misure di dispersione fornisce un'ulteriore dimensione analitica, permettendo di valutare l'eterogeneità e la variabilità dei dati relativi all'impatto del fumo passivo e alla percezione economica del valore della vita.

Attraverso l'esplorazione di questi indici di dispersione, si ottengono conoscenze critiche relative all'ampiezza della distribuzione e alla variabilità dei dati su DPM e VSL. Per quanto riguarda DPM, risulta:

DPM	CV	Scarto Interquartile	Varianza	Deviazione Std.
COMPLETO	0.529	113.888	11153.037	105.608
< 15 ANNI	2.234	3.976	228.428	15.114
TRA 16 E 64	0.569	50.474	2352.822	48.506
> 64 ANNI	0.524	636.980	223044.927	472.276

Table 3: Indici di Dispersione (DPM)

I valori della seguente tabella indicano rispettivamente:

- **Coefficienti di Variazione (CV):** fornisce una misura relativa della dispersione dei dati rispetto alla media. Notiamo che il CV è particolarmente elevato, 2.234, per la categoria di età inferiore ai 15 anni, indicando una variabilità proporzionalmente maggiore rispetto alle altre categorie di età. Questo suggerisce che, nonostante il numero assoluto di decessi prematuri in questa fascia di età sia relativamente basso, la variabilità tra le diverse popolazioni è notevole. Per le altre categorie, i CV sono più contenuti, tra 0.524 e 0.569, indicando una maggiore uniformità nei dati.
- **Scarto Interquartile:** mostra un'ampia variabilità nei decessi prematuri tra gli over 64, 636.980, significativamente maggiore rispetto alle altre fasce d'età. Questo riflette una grande dispersione nei dati, suggerendo che esiste un'ampia gamma di valori di DPM tra i paesi per questa fascia di età, possibilmente dovuta a differenze nelle politiche sanitarie, nei comportamenti sociali o nelle esposizioni ambientali al fumo passivo.
- **Varianza e Deviazione Standard:** mostrano per gli over 64 anni sono estremamente elevate, rispettivamente 223044.927 e 472.276, confermando l'elevata dispersione dei dati già suggerita dalla scarto interquartile. Per le altre categorie, queste misure di dispersione sono significativamente più basse, riflettendo una minore variabilità nei tassi di mortalità prematura.

Invece, per quanto riguarda VSL:

VSL	CV	Scarto Interquartile	Varianza	Deviazione Std.
COMPLETO	0.412	1.667	2.01	1.418

Table 4: Indici di Dispersione (VSL)

Tali valori indicano:

- **Coefficienti di Variazione (CV):** con valore di 0.412 indica una dispersione relativamente moderata dei dati intorno alla media. Questo suggerisce che, nonostante le differenze economiche e di politiche sanitarie tra i paesi, la valutazione economica della vita umana in termini di prevenzione della mortalità da fumo passivo tende ad essere più uniforme rispetto alla variabilità osservata nei tassi di mortalità prematura (DPM).
- **Scarto Interquartile, Varianza e Deviazione Standard:** tali indici di dispersione, rispettivamente 1.667, 2.01 e 1.418, confermano la presenza di una variabilità moderata. Indicano che, sebbene ci siano delle fluttuazioni nel valore economico attribuito alla prevenzione della mortalità prematura tra i paesi, queste differenze non sono estreme.

## 2.3 Quartili

I **quartili** sono stati calcolati per ogni variabile e fascia di età considerata. L'approccio ritenuto ottimale prevede di effettuare questa operazione per ogni colonna del dataset, offrendo una panoramica della distribuzione annuale, *dal 1995 al 2018*. Questo processo consente di suddividere la distribuzione dei dati in quattro parti uguali, offrendo una visione dettagliata della dispersione e della centralità dei valori all'interno di ciascun gruppo. Attraverso l'identificazione dei quartili, è possibile determinare la mediana, rappresentata dal secondo quartile, nonché lo scarto interquartile, ovvero la differenza tra il terzo e il primo quartile, fornendo così informazioni preziose sull'*eterogeneità dei dati* e sulla loro *tendenza centrale*. Questa metodologia di analisi gioca un ruolo cruciale nell'esplorazione delle caratteristiche fondamentali dei dataset, permettendo di evidenziare pattern, tendenze e potenziali anomalie nei diversi gruppi di età e variabili studiate.

### 2.3.1 Boxplot per DPM

Sono stati presi in esame i grafici dei boxplot relativi alla variabile **DPM**, considerando la fascia di età **COMPLETA**, ma tale andamento si riscontra in maniera analoga anche per le altre fasce di età, seppur ovviamente con valori numerici differenti (vedi Figure 1, 2, 3, 4).

- Per quanto riguarda l'*anno 1995*, si osserva che:
  - Valutando le *distanze tra mediana e il primo quartile*, e tra la *mediana e il terzo quartile*, osserviamo l'equilibrio osservato da queste distanze denota che il **50% dei dati intorno alla mediana** risulta essere **ben distribuito**:
    - \*  $Q3 - Q2 = 62.95$
    - \*  $Q2 - Q1 = 62.133$
  - Tuttavia, la lunghezza del baffo inferiore è nettamente minore rispetto alla lunghezza del baffo superiore, per cui si nota una **forte asimmetria a destra**, o asimmetria superiore, nella distribuzione:
    - \* *estremo baffo superiore* -  $Q3$ : 154.9615
    - \*  $Q1$  - *estremo baffo inferiore*: 65.298
  - È presente, inoltre, anche un **outlier** corrispondente al valore massimo tra i dati, da attribuire in questo caso alla *Bulgaria*.



- Negli *anni successivi al 1995 fino al 1999*, il **trend** sembra essere pressochè **simile**.
- Analizzando gli *anni successivi al 1999*, si denotano **variazioni significative nella distribuzione dei dati**, in particolare che:
  - La *mediana tende sempre più ad avvicinarsi a Q1*, indicando **una minor distribuzione del 50% dei dati centrali attorno alla mediana** nel corso degli anni. Questo graduale avvicinamento porta ad osservare, *nel 2018*, le seguenti distanze:
    - \*  $Q3 - Q2 = 90.719$
    - \*  $Q2 - Q1 = 26.854$
  - Si osserva una **graduale diminuzione** del valore del **baffo inferiore**, indicando una *diminuzione dei valori minimi*, anno dopo anno.
  - Lo stesso trend non si riporta invece per il **baffo superiore**, che presenta infatti un *drastico calo tra il 1995 e il 2000*, procedendo successivamente con un **andamento sinusoidale**.
  - Anche i valori di *Q1 e Q3 tendono a diminuire costantemente ogni anno*, che graficamente si traduce in un **costante spostamento del box verso valori inferiori**, tale trend si verifica *tra il 1995 e il 2015*, in particolare si denota che:
    - \* Q1 diminuisce in media di 2.73 ogni anno.
    - \* Q3 diminuisce in media di 3.19 ogni anno.
  - Dopodiché si verifica un **graduale incremento dei valori fino al 2018**.

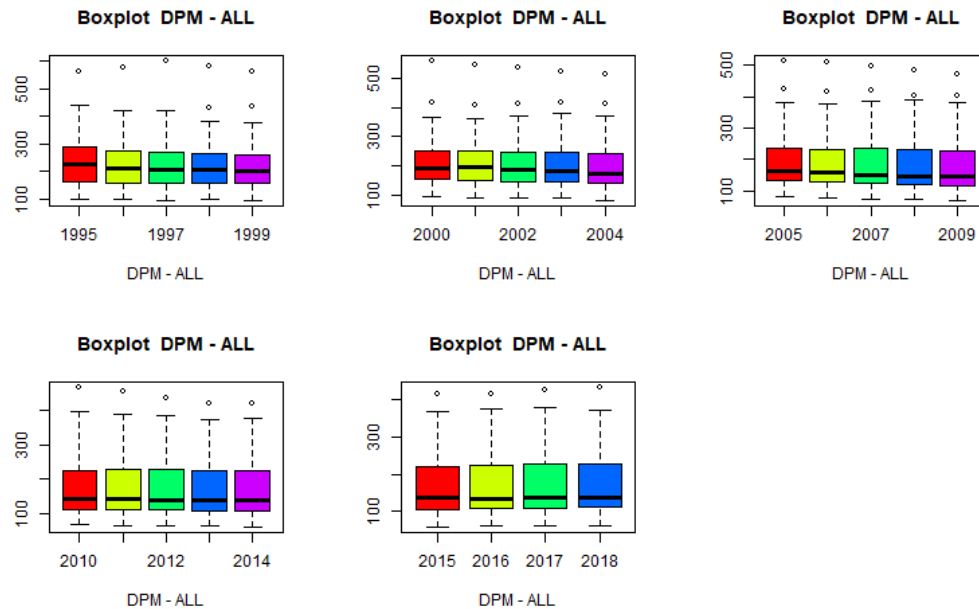


Figure 1: Boxplot per DPM - Fascia di età: COMPLETO

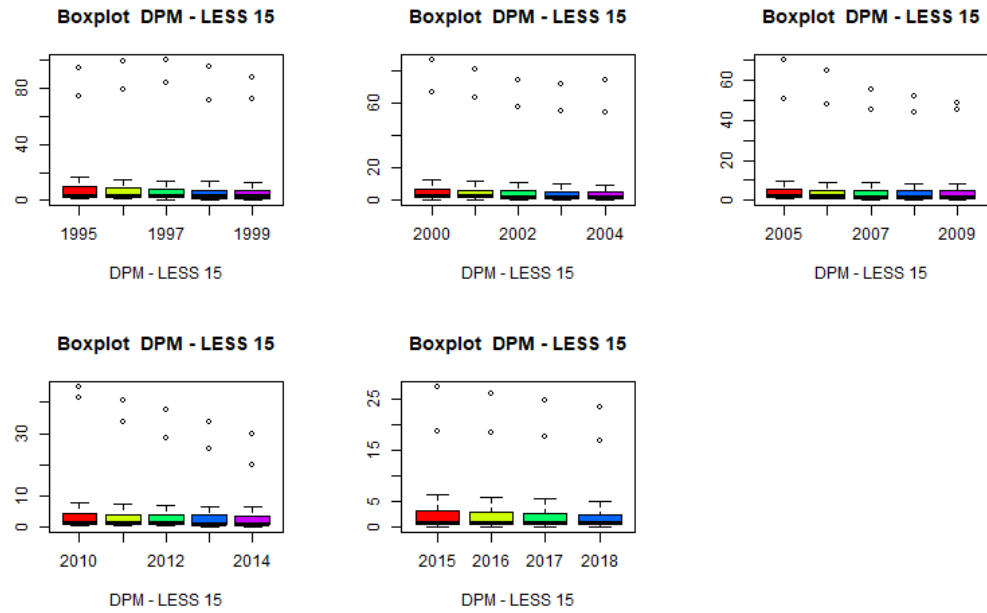


Figure 2: Boxplot per DPM - Fascia di età: <15

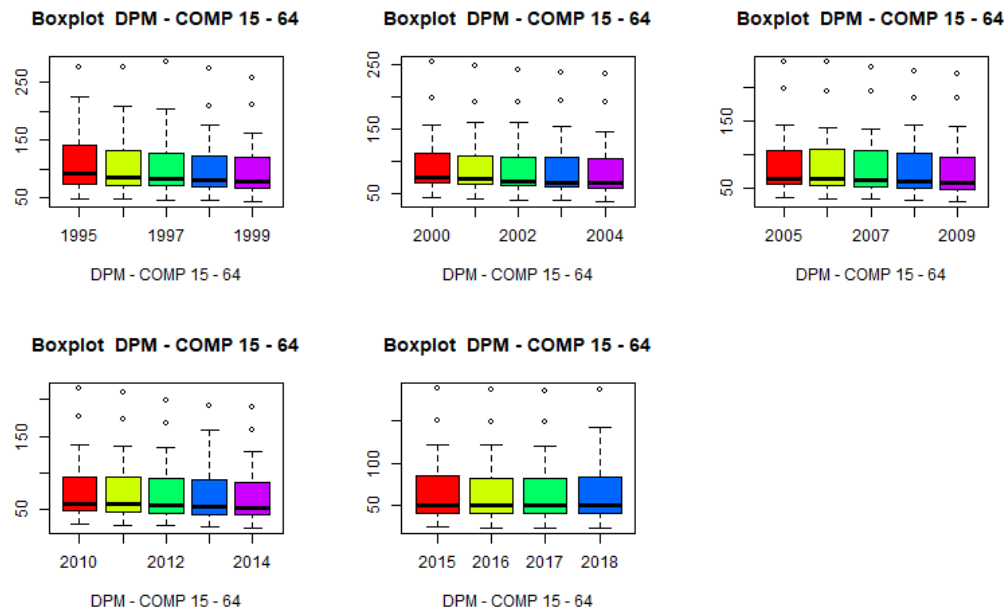


Figure 3: Boxplot per DPM - Fascia di età: TRA 16 E 64

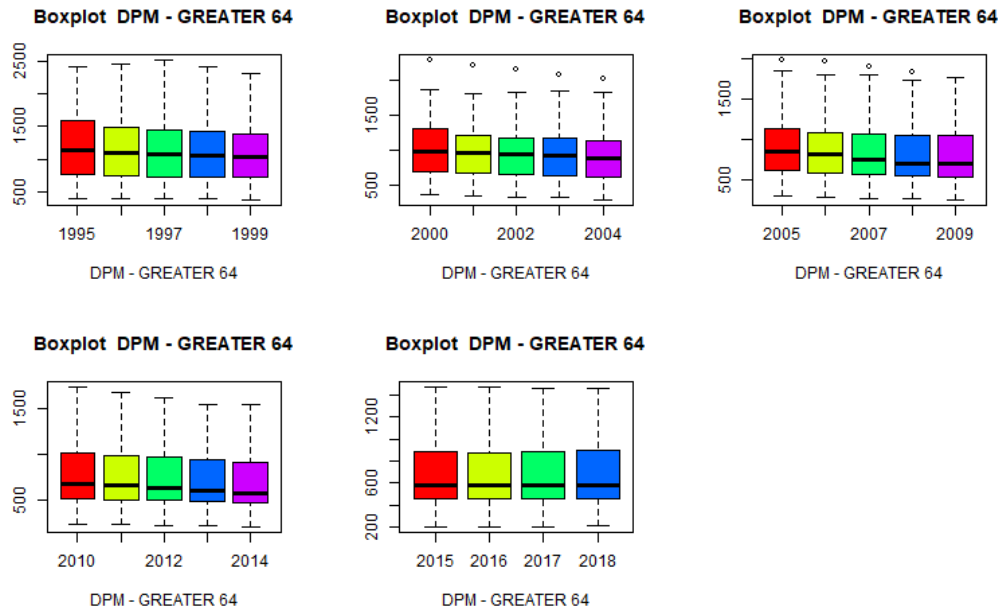


Figure 4: Boxplot per DPM - Fascia di età: > 64 ANNI

### 2.3.2 Boxplot per VSL

Di seguito è riportata l'analisi dei boxplot dedicata alla variabile **VSL**. Come nella sezione precedente, è stato prima osservato l'anno 1995 per poi analizzare l'andamento negli anni a seguire.

- Dunque, per l'anno 1995 si osserva che:
  - Valutando le *distanze tra la mediana e il primo quartile*, e tra la *mediana e il terzo quartile*, osserviamo:
    - \*  $Q3 - Q2 = 1,107$
    - \*  $Q2 - Q1 = 0.901$

Nel caso dell'indicatore VSL, quindi, i *valori sono in generale molto bassi rispetto a DPM*, data la differente unità di misura che descrive i due indicatori. Pertanto, nonostante la differenza tra i due risultati sia relativamente bassa, può essere considerata comunque significativa. Dunque si osserva che **il 50% dei dati intorno alla mediana risulta sbilanciato**, anche se leggermente.

- Proprio come accadeva per la variabile DPM, anche in questo caso la *lunghezza del baffo inferiore è nettamente minore rispetto a quella del baffo superiore*, notando un'**accentuata asimmetria superiore** nella distribuzione dei dati. In particolare:
  - \* *estremo baffo superiore - Q3*: 2.972
  - \* *Q1 - estremo baffo inferiore*: 0.518

È, però, importante osservare che questa notevole differenza nelle lunghezze dei due baffi è causata principalmente dalla presenza di un valore molto elevato che

però **non è considerato outlier** nell'anno 1995, il Lussemburgo, con un valore di VSL pari a 6.533. Volendo infatti considerare questo valore un outlier (cosa che verrà fatta negli anni successivi), si può osservare come la **distribuzione risulterebbe decisamente più simmetrica**:

\* *estremo baffo superiore* - Q3: 0.532

\* Q1 - *estremo baffo inferiore*: 0.498

- Analizzando i grafici relativi agli *anni successivi al 1995*, si possono notare **miglioramenti generali**, si osserva in particolare che:

- *A partire dal 2000 la mediana*, in seguito ad un leggero allontanamento dal primo quartile, subisce un *progressivo avvicinamento a Q1* che si traduce in una **minor distribuzione dei dati centrali attorno alla mediana**, rispetto all'anno di partenza.
- Di conseguenza, i valori osservati nel 2018 per le *distanze tra la mediana e i quartili Q1 e Q3* sono i seguenti:
  - \*  $Q3 - Q2 = 0.857$
  - \*  $Q2 - Q1 = 0.465$
- Sulla base dell'*analisi dei valori statistici dei boxplot (boxplot\$stats)*, si osserva un *graduale incremento generale dei valori*, includendo quindi gli estremi dei baffi, i limiti della scatola e la mediana. Questo riflette un **incremento della spesa pubblica** per contrastare il problema della mortalità analizzato in questo studio.
- Mentre il trend appena descritto si verifica nella maggioranza degli anni presi in esame, possono essere osservate delle **significative deviazioni nell'andamento** per gli *anni 2009 e 2010*. Infatti, questi due anni si caratterizzano per un'improvviso calo, questo decremento può essere probabilmente attribuito alla *crisi finanziaria* che ha caratterizzato il periodo storico appena precedente (2007/2008) (vedi Sezione 7.1). *A partire dal 2011*, si osserva una **ripresa del trend incrementale**, seppur inizialmente meno marcata, evidenziando comunque una tendenza a lungo termine verso un aumento della spesa pubblica per contrastare la mortalità causata dal fumo passivo.

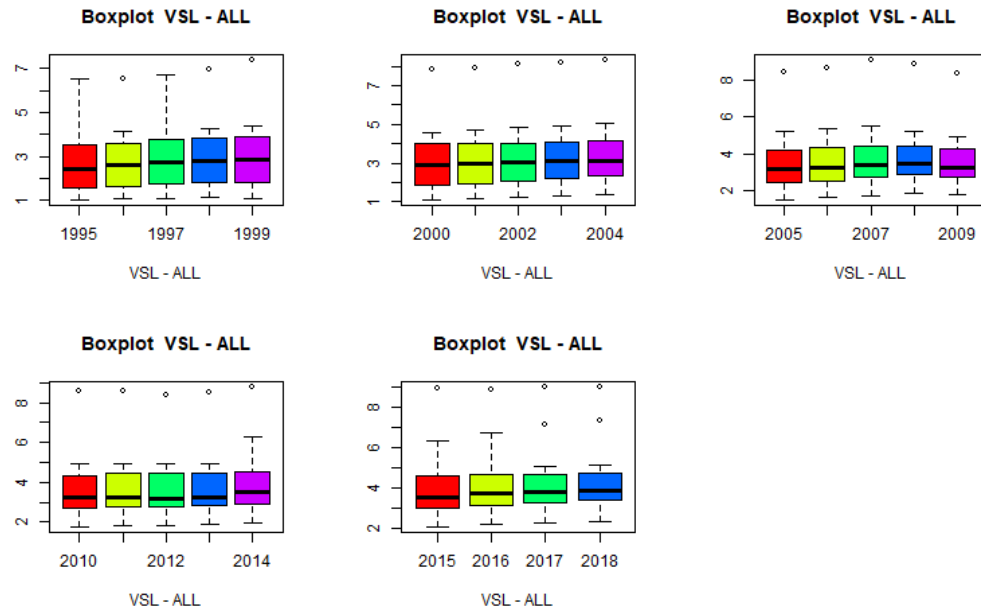


Figure 5: Boxplot per VSL

### 3 Statistica Descrittiva Univariata

La **statistica descrittiva univariata** costituisce un pilastro essenziale dell'analisi statistica, concentrando l'attenzione sull'analisi e l'interpretazione delle caratteristiche di una singola variabile quantitativa all'interno di un insieme di dati. La sua applicazione permette di indagare in maniera efficace le proprietà e la distribuzione della variabile, offrendo un fondamento empirico solido per le successive analisi e deduzioni.

#### 3.1 Funzione di Distribuzione Empirica Continua

Nell'ambito dell'analisi statistica dei fenomeni quantitativi continui, sono stati categorizzati i dati di ogni paese in cinque gruppi attraverso una metodologia basata sui quantili. Tale approccio garantisce che ogni gruppo rappresenti circa il 20% del totale delle osservazioni. Queste categorie sono state delineate per abbracciare l'intera gamma di dati, estendendosi dal valore minimo al valore massimo registrato. La procedura successiva ha comportato il calcolo della **Funzione di Distribuzione Empirica Continua (FdDC)** per ciascun paese e indicatore. I grafici illustrano come la percentuale cumulativa di osservazioni si distribuisca attraverso il range delle variabili, offrendo una rappresentazione continua della distribuzione dei dati che prescinde dalla suddivisione in classi discrete. (vedi Figure 6, 7, 8, 9, 10, 11, 12). La FdDC è stata calcolata per evidenziare come le *osservazioni si accumulino progressivamente lungo il range delle variabili*, definita nel seguente modo:

$$F(x) = \begin{cases} 0, & \text{se } x < \min \\ 0.2, & \text{se } \min \leq x < \min + \frac{\max - \min}{5} \\ 0.4, & \text{se } \min + \frac{\max - \min}{5} \leq x < \min + \frac{2(\max - \min)}{5} \\ 0.6, & \text{se } \min + \frac{2(\max - \min)}{5} \leq x < \min + \frac{3(\max - \min)}{5} \\ 0.8, & \text{se } \min + \frac{3(\max - \min)}{5} \leq x < \min + \frac{4(\max - \min)}{5} \\ 1, & \text{se } x \geq \min + \frac{4(\max - \min)}{5} \text{ (ovvero se } x \geq \max) \end{cases}$$

L'analisi grafica della FdDC ha evidenziato variazioni significative nella pendenza, svelando differenze nella densità di dati tra le varie nazioni. Per rendere l'analisi più gestibile e informativa, le *osservazioni che seguono sono state formulate raggruppando i paesi in categorie basate su andamenti simili*. Questo approccio consente di identificare modelli di distribuzione comuni, evitando un'analisi eccessivamente frammentata e laboriosa che avrebbe comportato l'esame dettagliato di ogni singolo paese, che sarebbe risultato un processo ripetitivo.

Per l'indicatore **DPM**, la maggior parte dei grafici evidenzia una pendenza costante, un indizio di distribuzione uniforme dei valori lungo gli intervalli, particolarmente perfetta in nazioni come il *Danimarca, Germania e Lussemburgo*. Tuttavia, si osservano deviazioni significative in paesi come *Belgio, Croazia e Cipro* dove le peculiarità dei grafici di distribuzione emergono chiaramente. In particolare:

- *Belgio, Croazia, Estonia, Italia, Malta, Polonia, Portogallo, Rep. Ceca, Ungheria*: mostrano **variazioni di pendenza**, caratterizzati da tratti che alternano inclinazioni diverse; questo comportamento indica una distribuzione dei dati che non è uniforme, con intervalli di densità variabile, quindi, i dati si concentrano maggiormente in alcuni

segmenti rispetto ad altri, riflettendo una certa eterogeneità nella distribuzione dei valori.

- *Grecia, Lituania*: iniziano con una **pendenza molto ripida**, indicando una forte concentrazione di dati nei primi intervalli, successivamente, la pendenza **diventa costante**, suggerendo che, superato il picco iniziale di densità, i valori si distribuiscono più uniformemente; questo comportamento riflette una distribuzione dei dati che ha una marcata prevalenza di valori bassi, seguita da una distribuzione equilibrata dei valori residui.
- *Cipro, Francia, Lettonia, Romania, Slovacchia*: presentano tratti con **differenze notevoli di pendenza**, segnalando variazioni significative nella concentrazione dei dati lungo gli intervalli; queste differenze indicano che ci sono segmenti dove i dati sono estremamente concentrati, intervallati da aree con minore densità di valori. Questo modello suggerisce la presenza di caratteristiche distintive nella distribuzione dei valori, come picchi di alta densità seguiti da aree di rarefazione.
- *Austria, Bulgaria, Finlandia, Irlanda, Paesi Bassi, Slovenia, Spagna*: presentano una **pendenza costante**, indicativa di una distribuzione uniforme dei dati lungo tutti gli intervalli considerati; questa uniformità suggerisce che i valori si distribuiscono in modo equilibrato attraverso l'intero range, senza particolari concentrazioni o vuoti di dati.
- *Danimarca, Germania, Lussemburgo, Svezia*: mostrano una **pendenza perfettamente costante** nei loro grafici di distribuzione, segnalando la più uniforme distribuzione dei valori tra tutti i paesi analizzati; la perfetta costanza della pendenza indica che i dati sono distribuiti in maniera estremamente equilibrata, senza alcuna variazione significativa di densità dei dati attraverso l'intero spettro dei valori.

Per quanto riguarda **VSL**, è stata evidenziata un'interessante varietà nella distribuzione dei valori tra i paesi, la maggior parte dei paesi, come *Austria, Belgio e Croazia*, presentano grafici con pendenze che variano a tratti, suggerendo una distribuzione dei valori di VSL che non è costantemente uniforme, ma caratterizzata da specifiche concentrazioni dei dati in determinati intervalli. Questa osservazione indica che, sebbene alcuni paesi mantengano una distribuzione perfettamente uniforme, come *Lettonia, Germania e Italia*, dei valori di VSL, esistono significative deviazioni in termini di densità di dati, riflettendo la complessità e la diversità delle distribuzioni di valore nei diversi paesi.

- *Austria, Belgio, Croazia, Danimarca, Estonia, Francia, Lituania, Paesi Bassi, Portogallo, Rep. Ceca, Romania, Slovacchia, Slovenia, Spagna*: presentano **variazioni di pendenza** a tratti, questo comportamento suggerisce una distribuzione dei dati che varia lungo i diversi intervalli, con segmenti di maggiore o minore densità di valori.
- *Finlandia*: mostra una **pendenza inizialmente costante**, che **diventa ripida** verso la fine, evidenziando una concentrazione significativa di dati negli ultimi intervalli, a differenza degli intervalli iniziali.
- *Irlanda, Malta*: presentano una **iniziale pendenza molto ripida** nei loro grafici, indicando una forte concentrazione di dati nei primi intervalli; tale comportamento

suggerisce che un numero significativo di valori si aggira intorno al limite inferiore del range, seguito da una concentrazione più scarsa dei dati negli intervalli successivi.

- *Lussemburgo*: presenta una **iniziale pendenza piatta** che, successivamente, **diventa ripida**; questo modello di distribuzione indica che i dati sono inizialmente scarsi nei primi intervalli, con un'impennata nella densità di valori verso la fine del range, ciò può riflettere una concentrazione di valori verso il limite superiore.
- *Bulgaria, Cipro, Grecia, Polonia, Svezia, Ungheria*: mostrano una **pendenza costante**, indicando una distribuzione uniforme dei dati lungo tutto il range di valori, evidenziando che i valori sono distribuiti abbastanza equamente, senza particolari concentrazioni o vuoti.
- *Lettonia, Germania, Italia*: presentano una **pendenza perfettamente costante**, segnalando la distribuzione più uniforme dei valori tra tutti i paesi analizzati; ciò indica una distribuzione ideale e equilibrata dei dati, senza variazioni significative nella densità di valori attraverso l'intero spettro considerato.

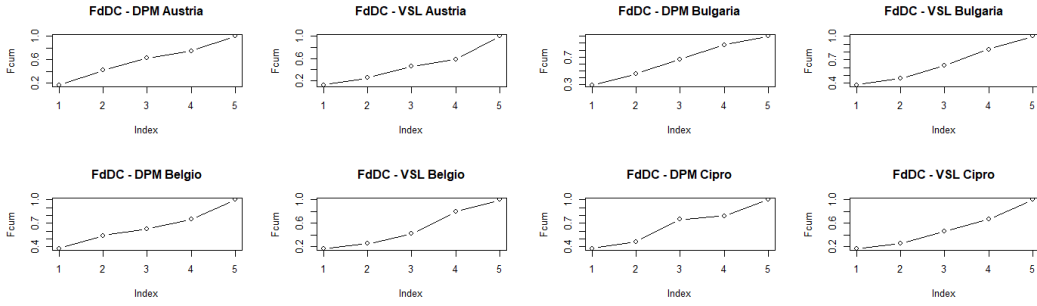


Figure 6: FdDC - Austria, Belgio, Bulgaria, Cipro

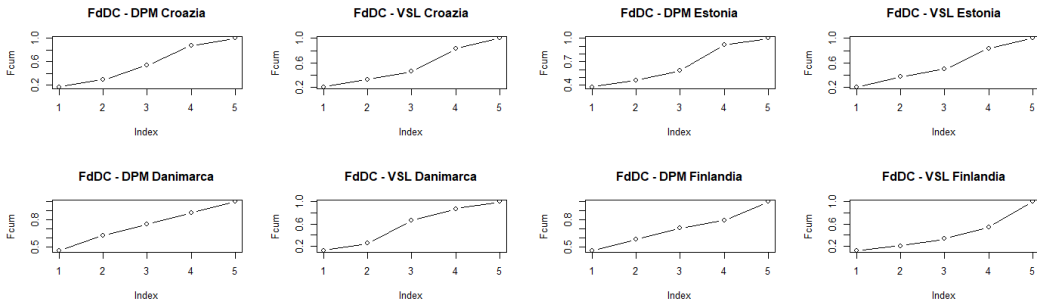


Figure 7: FdDC - Croazia, Danimarca, Estonia, Finlandia



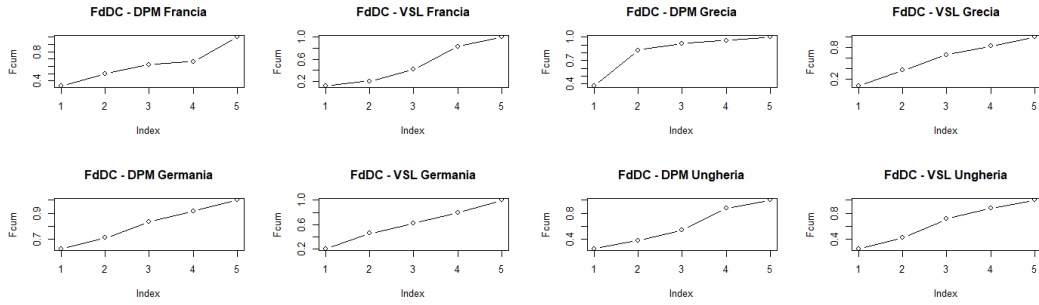


Figure 8: FdDC - Francia, Germania, Grecia, Ungheria

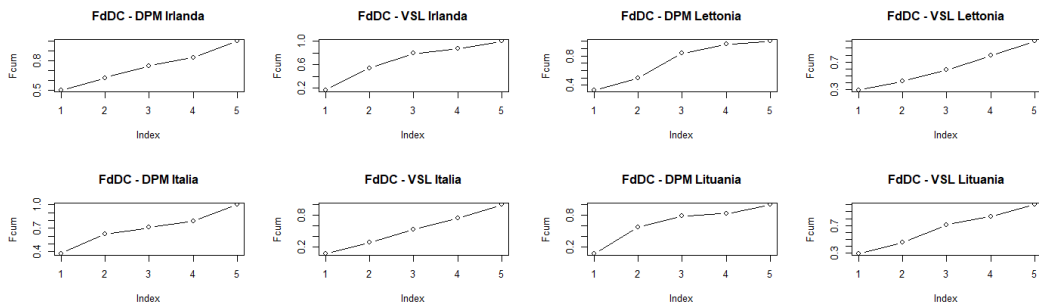


Figure 9: FdDC - Irlanda, Italia, Lettonia, Lituania

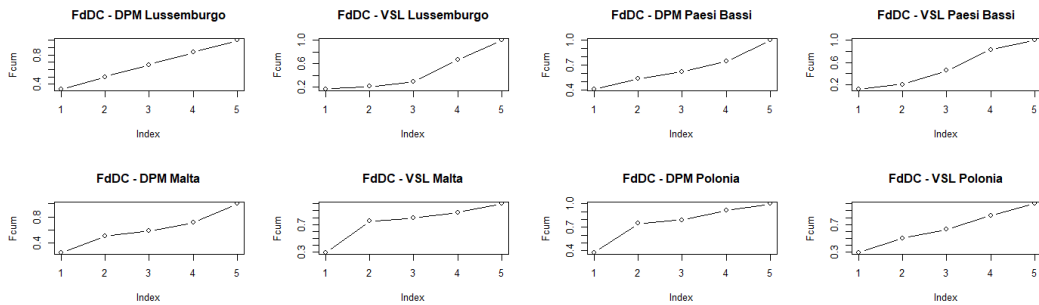


Figure 10: FdDC - Lussemburgo, Malta, Paesi Bassi, Polonia

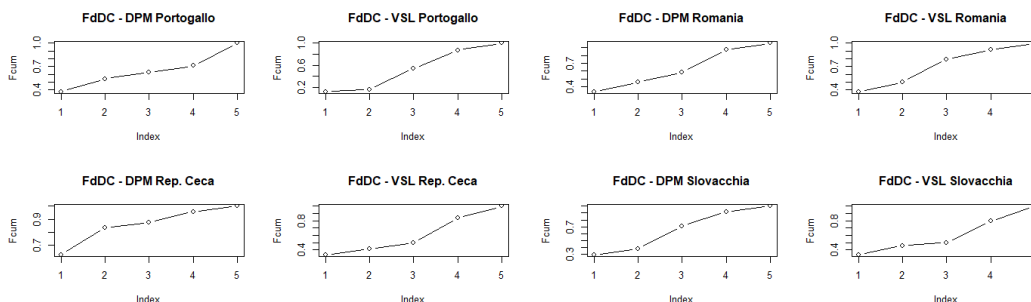


Figure 11: FdDC - Portogallo, Rep. Ceca, Romania, Slovacchia

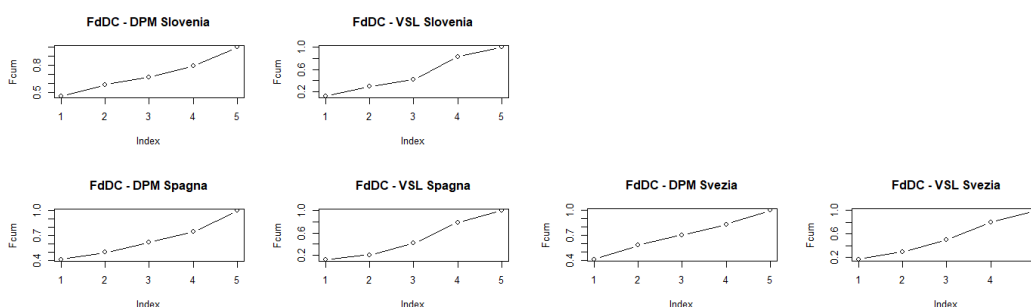


Figure 12: FdDC - Slovenia, Spagna, Svezia

### 3.1.1 Conclusioni

In conclusione, l'esame combinato di DPM e VSL, relativo a FdDC, evidenzia l'**importanza di strategie personalizzate** e mirate per affrontare il fumo passivo e valutare il valore della vita, **tenendo conto delle specificità nazionali** all'interno dell'UE.

## 3.2 Serie Temporal

Di seguito è stata analizzata l'evoluzione dei due indicatori DPM e VSL, con l'intento di tracciare le tendenze e le variazioni di questi ultimi nell'arco temporale che va dal 1995 al 2018.

### 3.2.1 Serie Temporale - DPM

Il primo grafico su cui è posta l'attenzione descrive il trend temporale per l'indicatore **DPM**, considerando tutte le categoria di età (vedi Figure 13).

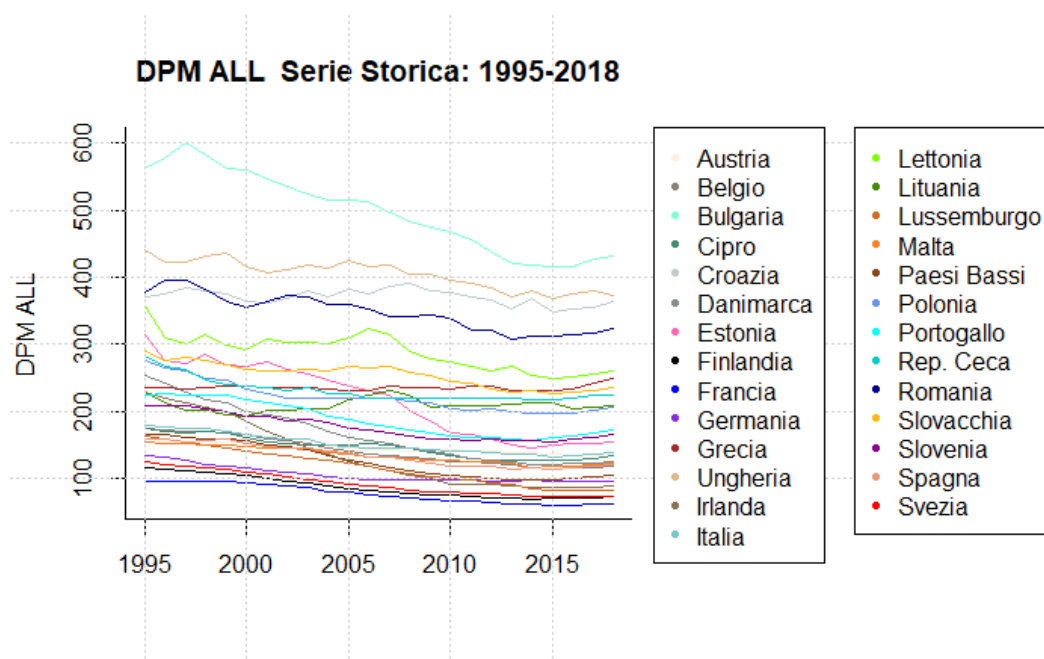


Figure 13: Serie Temporale - DPM Completo

In questo grafico si osserva come i **trend** non siano **analoghi tra i vari paesi** presi in esame, anzi *in alcuni casi i trend sono completamente opposti* o comunque presentano grandi differenze. In base a questi differenti trend, è possibile suddividere i paesi in diverse macro-categorie per l'indicatore DPM, su cui sono state tratte le considerazioni in maniera collettiva:

- **Bulgaria:** rappresentata nel grafico dalla linea posizionata più in alto, indicando **valori significativamente superiori** rispetto agli altri paesi, in particolare:

	1995	2018	Calo
Bulgaria	562.212	432	23.69%

- **Ungheria, Croazia, Romania:** presentavano **valori di DPM elevati nel 1995** e hanno avuto un **calo** molto leggero **negli anni successivi**, in particolare:

	1995	2018	Calo
Ungheria	440.9	371.77	15.68%
Croazia	370.22	363.15	1.91%
Romania	377.26	322.24	14.58%

- **Austria, Grecia, Lituania:** presentavano un **valore medio o basso di DPM nel 1995** riscontrando **successivamente un calo leggero**, quasi assente o addirittura un aumento nel corso degli anni, in particolare:

	1995	2018	Calo
Austria	154.02	134.88	12.43%
Grecia	234.47	249.86	-6.56%
Lituania	227.59	208.42	8.43%

- *Tutti gli altri paesi:* partono da un valore medio o basso hanno riportato **elevate percentuali di calo della mortalità**, con valori compresi circa tra il 20% e il 60%.

	1995	2018	Calo		1995	2018	Calo
Belgio	174.76	117.97	32.5%	Lussemburgo	162.16	82.17	49.33%
Cipro	174.87	133.62	23.59%	Malta	153.12	122.47	20.02%
Danimarca	252.85	124.33	50.83%	Paesi Bassi	166.08	105.22	36.65%
Estonia	313.07	153.88	50.85%	Polonia	275.01	205.2	25.38%
Finlandia	115.79	72.7	37.21%	Portogallo	222.99	173.07	22.38%
Francia	95.56	62.27	34.84%	Rep. Ceca	281.57	223.6	20.59%
Germania	134.71	96.68	28.23%	Slovacchia	290.31	234.73	19.14%
Irlanda	229.02	89.22	61.04%	Slovenia	209.26	165.49	20.92%
Italia	179.27	138.45	22.77%	Spagna	159.56	121.06	24.127%
Lettonia	355.98	259.91	26.99%	Svezia	123.96	73.64	40.6%

Per la fascia di età <15 anni (vedi Figure 14), il **trend** risulta essere **molto positivo**, *costantemente decrescente*. Ciò risulta in particolar modo evidente per Romania e Bulgaria, gli outliers di questo dataset, ma lo stesso andamento è riportato anche da tutti gli altri paesi con percentuali di calo molto elevate.

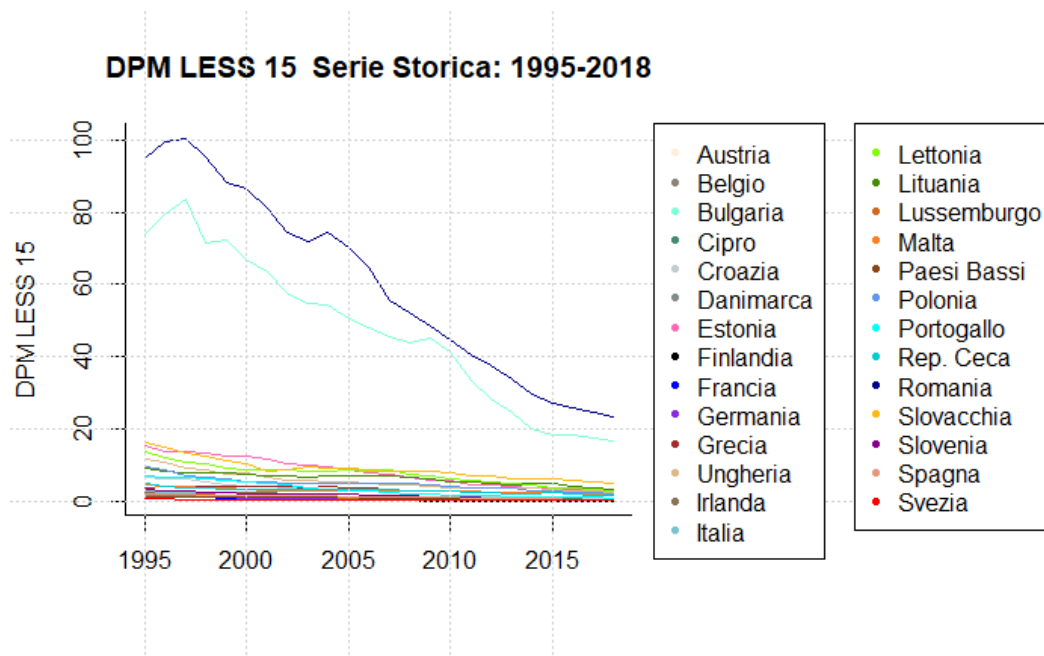


Figure 14: Serie Temporale - DPM <15 anni

Quindi, in generale, per questa fascia di età si denota un andamento nel complesso analogo tra i vari paesi, con cali particolarmente pronunciati, suggerendo un non trascurabile impegno nella prevenzione e trattamento della problematica per la fascia infantile. In particolare, si osserva che le percentuali di calo oscillano tra il 43% e l'86%, con una media del 63% di decessi in meno dal 1995 al 2018.

Per quanto riguarda le fasce di età TRA 16 E 64 anni e >64 anni si osserva **andamento decrescente** analogo (vedi Figure 15, 16). Infatti, nonostante i valori reali decisamente distanti gli uni dagli altri, difatti si può notare come l'asse Y su cui sono riportati i valori di DPM siano decuplicati nel caso di >64 anni, le percentuali di calo e i trend per i singoli paesi risultano seguire la stessa scia per queste due differenti categorie.

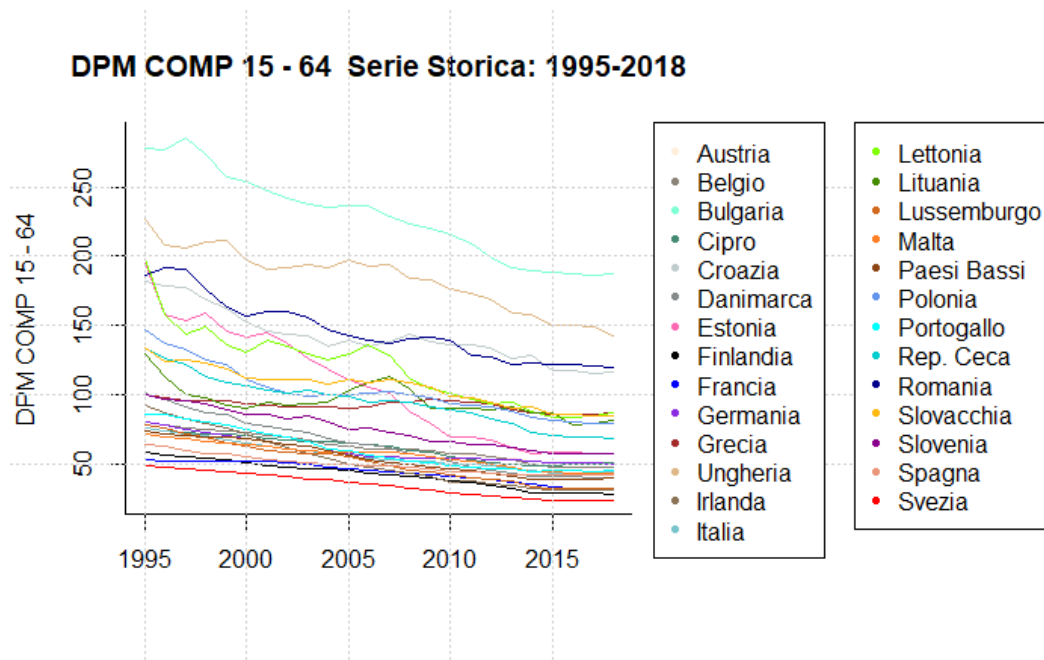


Figure 15: Serie Temporale - DPM 15-64 anni

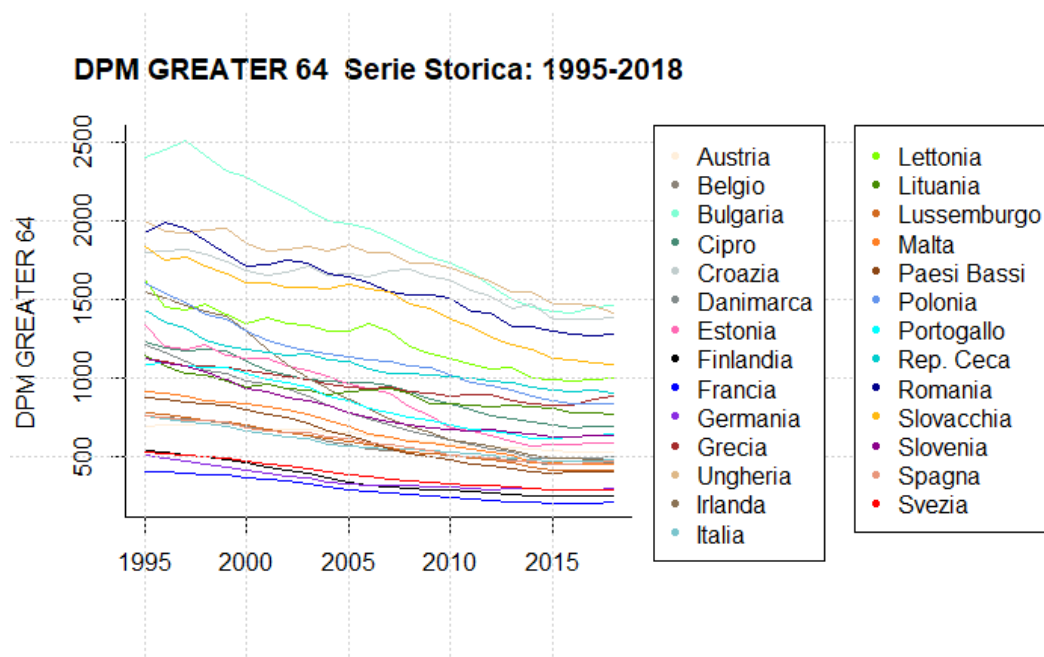


Figure 16: Serie Temporale - DPM >64 anni

Ponendo dunque un focus sulla fascia di età TRA 16 E 64 anni, risulta che il trend vede percentuali di calo in media comprese tra il 30% e il 50% come si può notare dal grafico. Solo un paese si distacca da questo range in negativo, ovvero la *Grecia* con una calo solamente del 13%, mentre emergono in positivo *Danimarca*, *Estonia*, *Irlanda* e *Lussemburgo* con percentuali che oscillano tra il 58% e il 70%.

Dunque, in generale, si riporta per l'indicatore DPM un notevole **trend di declino**, traducibile in un graduale ma significativo miglioramento dal punto di vista della mortalità, nella totalità dei paesi europei presi in esame. Si denotano però andamenti molto differenti tra i diversi paesi e ciò potrebbe essere dettato dalle differenze che li caratterizzano, sia dal punto di vista sociale che legislativo.

### 3.2.2 Serie Temporale - VSL

Osservando i trend analizzati per l'indicatore DPM, ci si aspetta di osservare un andamento crescente per l'indicatore **VSL**, prestando particolare attenzione però alle nazioni che hanno avuto cali dei decessi lievi o addirittura negativi, ma anche a quei paesi che hanno visto cali importanti come *Irlanda*, *Estonia* o *Danimarca*. Di seguito, è mostrato il grafico che mostra l'andamento temporale per quanto riguarda i valori dell'indicatore VSL (vedi Figure 17).

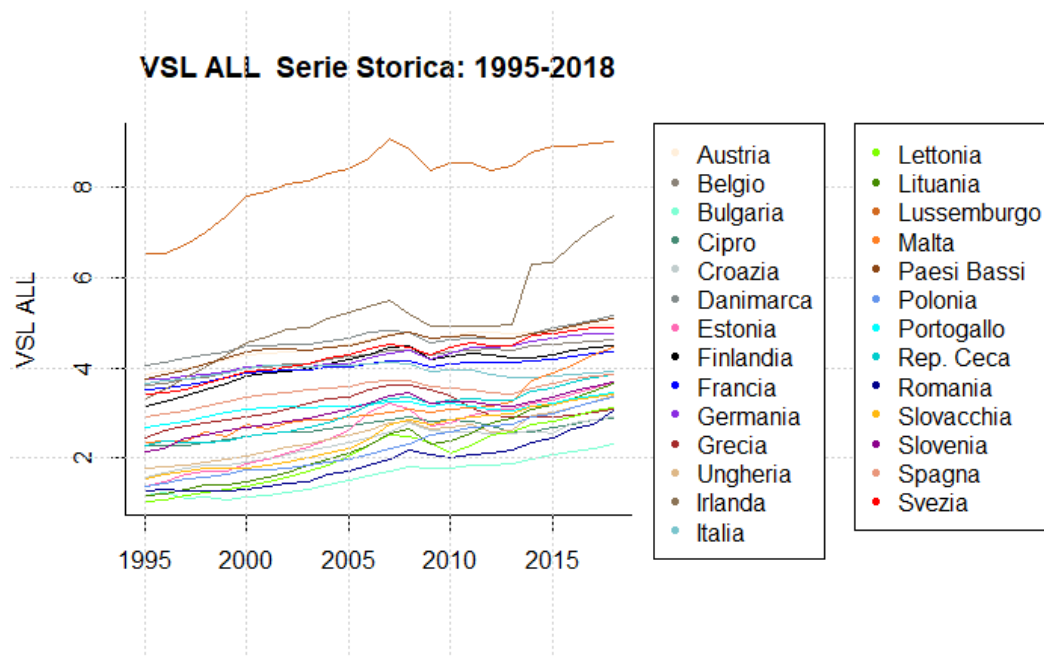


Figure 17: Serie Temporale - VSL

Quindi, risulta che:

- *Lussemburgo*: ottiene particolare risalto in questo grafico, essendo l'unico **outlier** dell'intero dataset. Infatti, proprio come ci si poteva aspettare dai risultati ottenuti nell'analisi del *trend storico per l'indicatore DPM*, che vedeva Lussemburgo tra i paesi con il maggior calo, esso presenta un **importante incremento della spesa pubblica** per affrontare tale mortalità. Infatti presenta un incremento del 27.5%, nonostante partisse da un valore già molto elevato rispetto ad altri.
- *Irlanda*: presenta una **particolare impennata del valore dal 2013** in poi, in seguito al decremento di investimento caratterizzato dal precedente periodo di crisi finanziaria (vedi Sezione 7.1). Più in generale, l'intera serie storica vede un incremento di VSL per l'Irlanda del 54.9%.
- *Italia*: risulta il paese con **minor crescita** da questo punto di vista, nonostante presenti valori comunque positivi rispetto agli indici statistici che descrivono questo dataset, con un incremento solamente del 6.47%.
- *Bulgaria, Croazia, Estonia, Ungheria, Lettonia, Lituania, Malta, Polonia, Repubblica Ceca, Romania, Slovacchia*: presentano un **andamento piuttosto positivo**, ciò denota come grande attenzione a questa problematica è stata prestata, in particolar modo, dall'*Est Europa*, con percentuali che vanno dal 40.5% al 66.9%. Inoltre, su questo aspetto, si può anche notare come i paesi che hanno **incrementato maggiormente il VSL** siano proprio quelli che **presentavano i risultati più gravi sull'indicatore DPM**.

### 3.2.3 Conclusioni

In conclusione, l'analisi delle serie storiche per entrambi gli indicatori evidenzia un **trend generale nettamente positivo** per la quasi totalità dei paesi inclusi nel dataset, con andamenti che seguono in larga misura una traiettoria analoga per tutti. In particolare, si nota come il **numero di decessi** sembri essere **inversamente correlato all'indicatore VSL**, indicando che l'incremento costante di quest'ultimo ha avuto un *impatto significativo e positivo* sul ridimensionamento del DPM, portando a un graduale calo nel tempo. Inoltre, emerge chiaramente che **i paesi che originariamente presentavano situazioni più critiche**, in termini di mortalità, come evidenziato dal DPM, **hanno investito maggiormente** per affrontare questa problematica. Questa tendenza è particolarmente evidente per i paesi dell'Europa dell'Est. Infine, un altro aspetto rilevante emerso dall'analisi è l'impatto significativo che la crisi economico-finanziaria del 2007-2008 (vedi Sezione 7.1) ha avuto sui valori dell'indicatore VSL, evidenziato dalle interruzioni drastiche delle tendenze di crescita nei valori di VSL durante gli anni successivi a questo evento.

### 3.3 Grafici a Barre

In seguito alla precedente analisi delle serie temporali (vedi Sezione 3.2) relative agli indicatori DPM e VSL, che ha evidenziato tendenze significative e variazioni rilevanti tra i diversi paesi e fasce di età nell'arco temporale considerato, ci si appresta ora a esaminare i grafici a barre dedicati specificamente alla variabile DPM suddivisa per fasce di età. Questi grafici hanno lo scopo di offrire una rappresentazione alternativa e complementare dei dati, facilitando una lettura immediata delle differenze tra le diverse categorie demografiche.

Le conclusioni emerse dall'analisi delle serie temporali, che hanno rivelato trend distinti tra i paesi e una marcata variazione della mortalità in relazione alle diverse fasce di età, presentando in generale **trend positivi** per entrambi gli indicatori, troveranno conferma anche nei grafici a barre. Pertanto, in questa sezione, si opterà per non ripetere le conclusioni già discusse, ma piuttosto per presentare i grafici a barre come un ulteriore strumento di visualizzazione che rafforza e illustra in modo diretto le osservazioni precedentemente fatte (vedi Figure 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44).

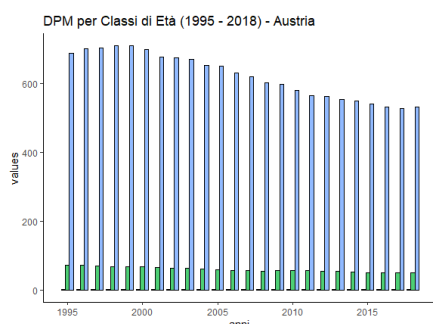


Figure 18: Grafico a barre - DPM Austria

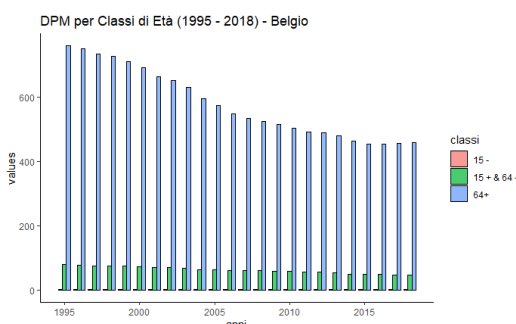


Figure 19: Grafico a barre - DPM Belgio



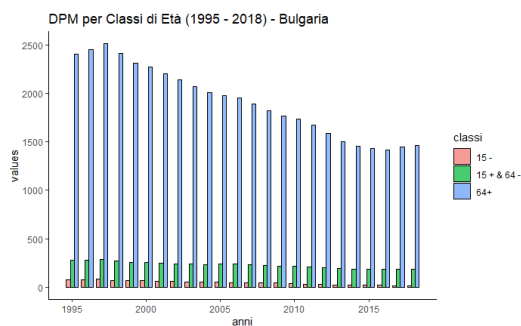


Figure 20: Grafico a barre - DPM Bulgaria

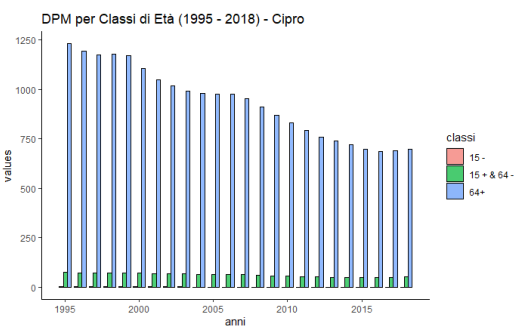


Figure 21: Grafico a barre - DPM Cipro

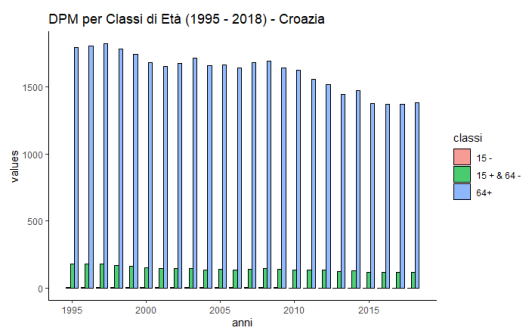


Figure 22: Grafico a barre - DPM Croazia

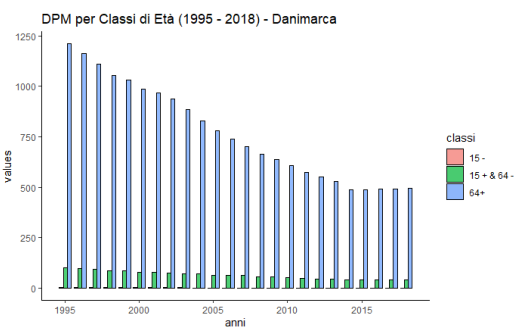


Figure 23: Grafico a barre - DPM Danimarca

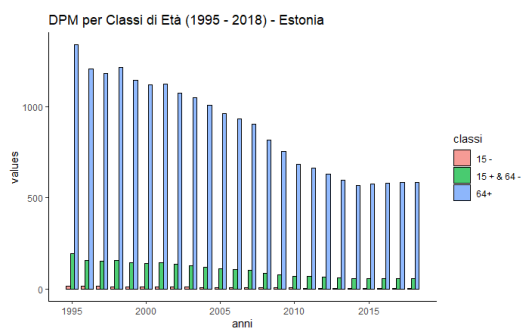


Figure 24: Grafico a barre - DPM Estonia

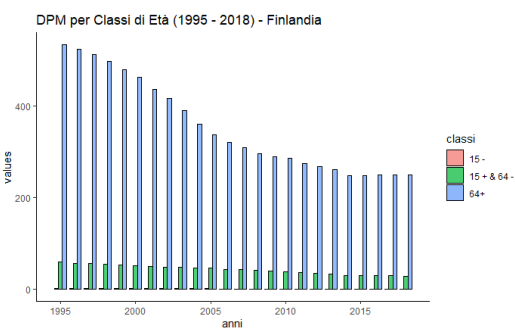


Figure 25: Grafico a barre - DPM Finlandia

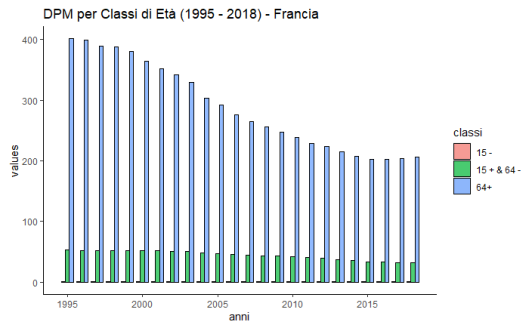


Figure 26: Grafico a barre - DPM Francia

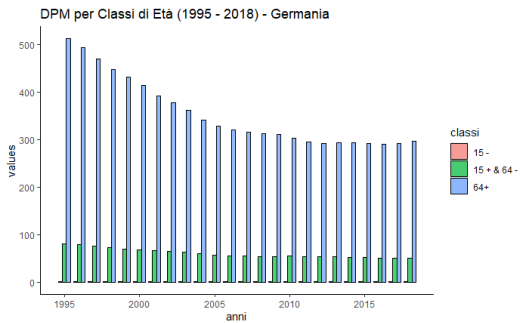


Figure 27: Grafico a barre - DPM Germania

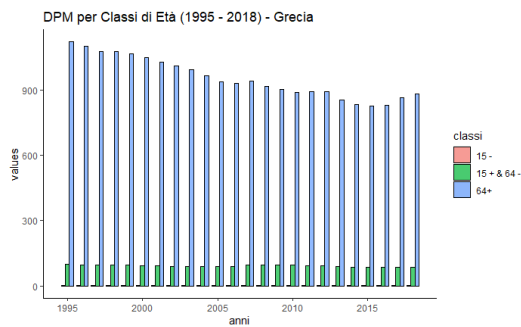


Figure 28: Grafico a barre - DPM Grecia

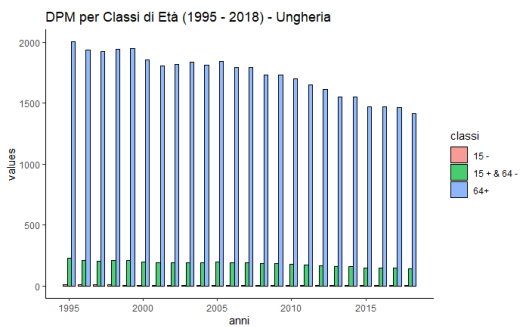


Figure 29: Grafico a barre - DPM Ungheria

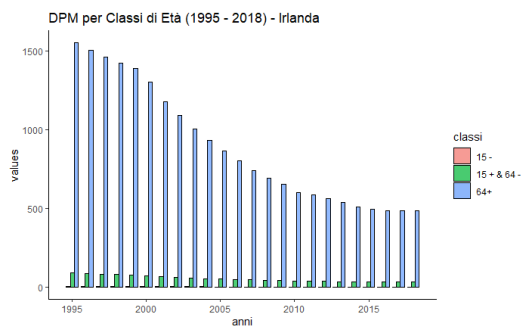


Figure 30: Grafico a barre - DPM Irlanda

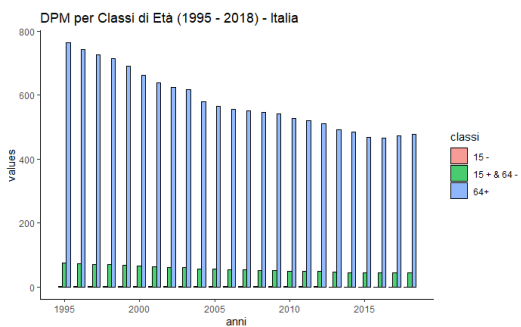


Figure 31: Grafico a barre - DPM Italia

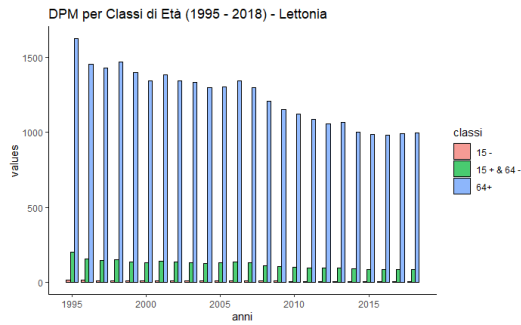


Figure 32: Grafico a barre - DPM Lettonia

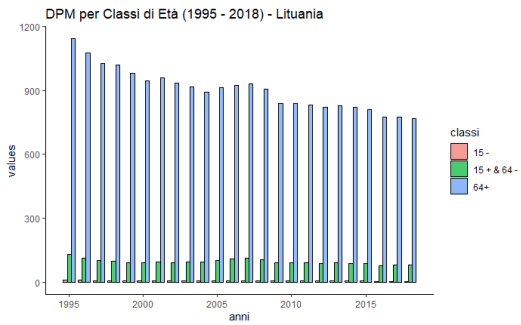


Figure 33: Grafico a barre - DPM Lituania

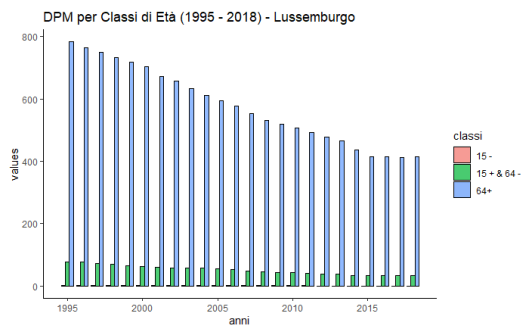


Figure 34: Grafico a barre - DPM Lussemburgo

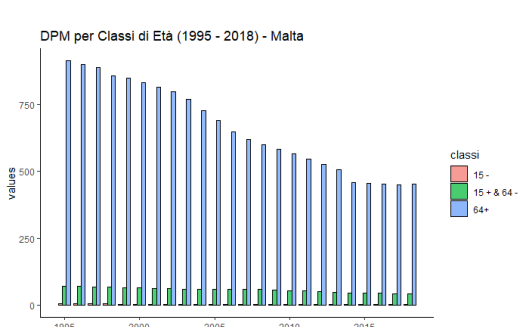


Figure 35: Grafico a barre - DPM Malta

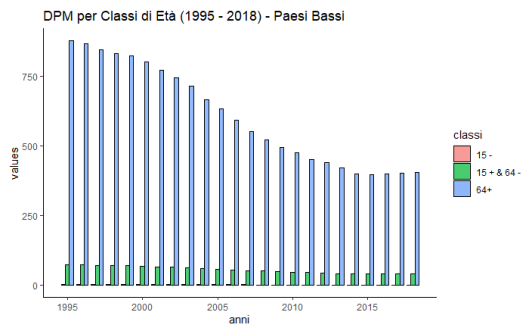


Figure 36: Grafico a barre - DPM Paesi Bassi

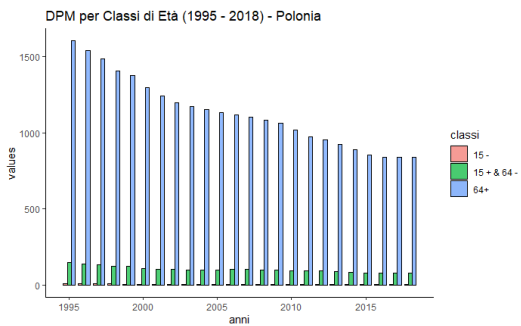


Figure 37: Grafico a barre - DPM Polonia

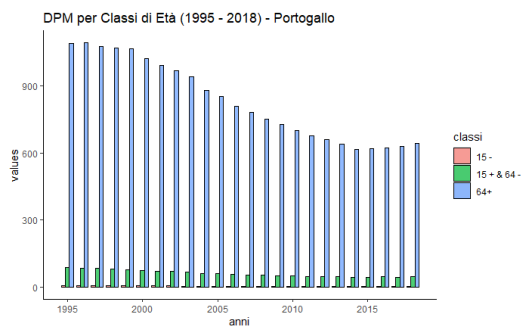


Figure 38: Grafico a barre - DPM Portogallo

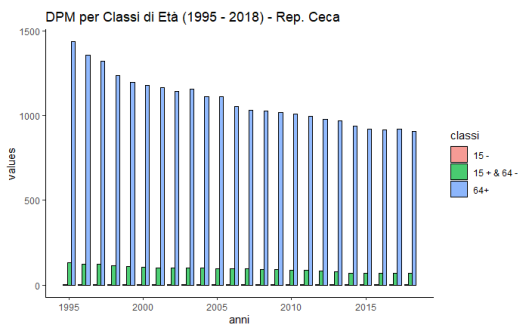


Figure 39: Grafico a barre - DPM Repubblica Ceca

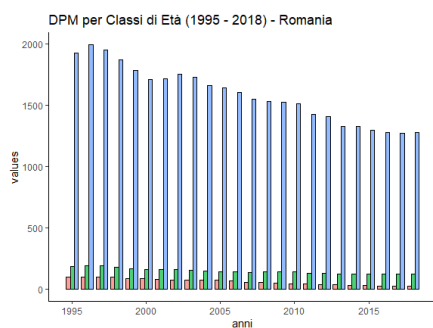


Figure 40: Grafico a barre - DPM Romania

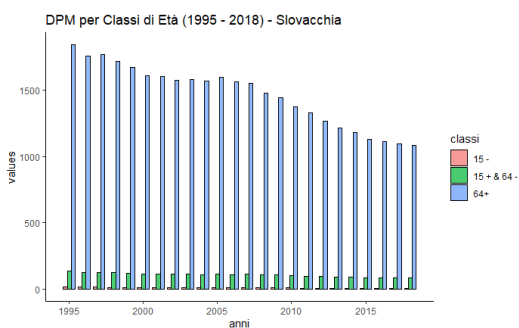


Figure 41: Grafico a barre - DPM Slovacchia

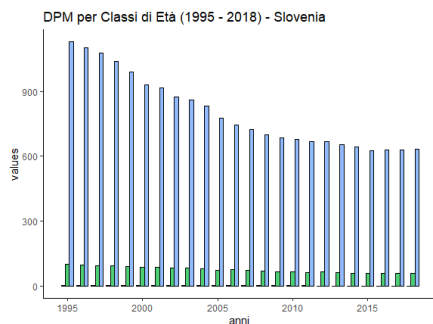


Figure 42: Grafico a barre - DPM Slovenia

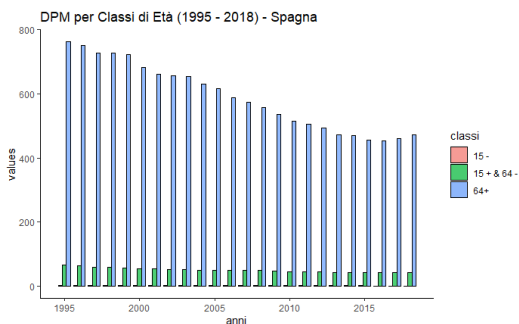


Figure 43: Grafico a barre - DPM Spagna

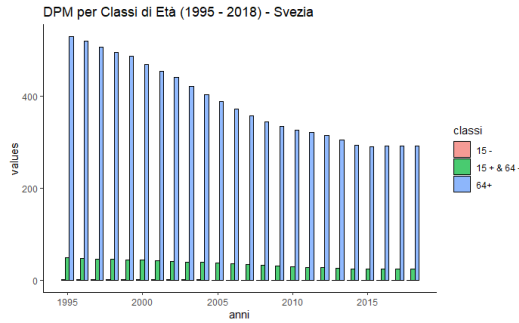


Figure 44: Grafico a barre - DPM Svezia

### 3.3.1 Conclusioni

Attraverso questa presentazione grafica,   stato possibile constatare come le tendenze individuate attraverso l'analisi temporale (vedi Sezione 3.2) si manifestino concretamente nelle variazioni della mortalit  per et , **confermando l'efficacia degli interventi e delle politiche adottate** nei vari paesi per affrontare questa problematica. I grafici a barre, dunque, si integrano nell'analisi complessiva come una dimostrazione visiva delle dinamiche gi  identificate, offrendo una prospettiva immediata sull'andamento della mortalit  nelle diverse fasce di et .

## 4 Statistica Descrittiva Bivariata

La **statistica descrittiva bivariata** rappresenta un campo fondamentale dell'analisi statistica, focalizzandosi sull'esame e l'interpretazione delle *relazioni esistenti tra due variabili quantitative* all'interno di un dataset. La sua applicazione, in questo contesto, consente di esplorare efficacemente le complesse relazioni tra DPM e VSL, fornendo una solida base empirica per le successive interpretazioni e conclusioni. La scelta di adottare *tecniche di regressione* si fonda sulla necessità di comprendere non solo l'esistenza ma anche l'intensità e la specificità delle correlazioni tra le variabili di interesse.

### 4.1 Regressione

La **regressione** emerge come uno strumento cruciale per investigare le dinamiche intercorrenti tra i due indicatori analizzati: i decessi per milione di abitanti (DPM) e il valore statistico della vita (VSL). Attraverso l'utilizzo di tali tecniche, si mira a decifrare la complessità delle relazioni tra queste variabili, mirando a una comprensione più profonda di *come variazioni nel VSL possano influenzare*, o essere associate a, *cambiamenti in DPM*, difatti nella seguente analisi è stato posto l'indicatore **DPM come variabile dipendente**, quindi valore che si intende predire e l'indicatore **VSL come variabile indipendente**.

Le relazioni sono state illustrate mediante *scatterplot* con relative *curve di regressione* e *diagrammi dei residui* per ciascun paese analizzato (vedi Figure 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71), impiegati per valutare le relazioni tra i due indicatori, permettendo di osservare visivamente sia l'andamento generale dei dati sia le specifiche deviazioni dal modello di regressione utilizzato. Per ogni paese, sono stati calcolati i **coefficienti di correlazione e di determinazione** per una serie di modelli di regressione: *lineare semplice, quadratico, esponenziale, logaritmico e semilogaritmico*, che misurano rispettivamente la forza e la direzione della relazione, e quanto bene il modello si adatta ai dati. Questa molteplicità di approcci modellistici è stata adottata con l'**intento di individuare il modello che meglio si adatta ai dati**, ovvero quello che offre la migliore interpretazione della relazione tra DPM e VSL, che per la quasi totalità dei casi risulta essere quello **quadratico**, tranne per quanto riguarda la Lettonia che risulta essere quello **semilogaritmico** (vedi Tabella 6). In particolare, da questi analisi si evince che:

- *Austria, Belgio, Bulgaria, Germania, Malta, Polonia, Svezia*: presentano un *coefficiente di determinazione maggiore di 0.9*, rappresentando una **correlazione (negativa) perfetta** tra gli indicatori.
- *Danimarca, Estonia, Finlandia, Francia, Irlanda, Lussemburgo, Paesi Bassi, Rep. Ceca, Romania, Slovacchia, Slovenia*: presentano un *coefficiente di determinazione compreso tra 0.7 e 0.9*, indicando una **correlazione (negativa) molto forte** tra gli indicatori.
- *Cipro, Ungheria, Lettonia, Spagna*: presentano un *coefficiente di determinazione compreso tra 0.5 e 0.7*, indicando una **forte correlazione (negativa)** tra gli indicatori.
- *Portogallo*: presentano un *coefficiente di determinazione compreso tra 0.3 e 0.5*, rappresentando una **correlazione (negativa) moderata** tra gli indicatori.

- *Italia*: presentano un *coefficiente di determinazione compreso tra 0.1 e 0.3*, indicando una **correlazione (negativa) debole** tra gli indicatori.
- *Croazia, Grecia, Lituania*: presentano un *coefficiente di determinazione compreso tra 0 e 0.1*, rappresentando una **correlazione (negativa) assente** tra gli indicatori.

Tali valori evidenziano, nonostante le specificità e le eccezioni rilevate per alcuni stati, la maggioranza dei paesi analizzati mostra un'**efficacia notevole dei modelli di regressione adottati**. In particolare, *per il 67% dei paesi* risulta che il coefficiente di determinazione supera il 0.7, indicando una *correlazione (negativa) molto forte* tra gli indicatori DPM e VSL. Questo suggerisce che, nella grande maggioranza dei casi, i modelli di regressione forniscono un'adeguata rappresentazione della relazione tra le variabili in esame, rivelando una forte capacità predittiva e di adattamento ai dati. La consistenza di questi risultati positivi sottolinea la **forte correlazione tra DPM e VSL**.

Inoltre, la presenza di **pattern nei diagrammi dei residui** per *Portogallo, Italia, Cipro, Croazia, Grecia e Lituania* conferma, ulteriormente, l'evidenza che i modelli di regressione adottati non siano i più adeguati per delineare con precisione la relazione tra DPM e VSL per questi paesi. Questo fenomeno, in combinazione con i valori relativamente bassi dei rispettivi coefficienti di determinazione per questi paesi, sottolinea una **non bontà dell'adattamento** del modello alle dinamiche specifiche dei dati.

	best_model	c_cor	lin	quad	exp	semilog	log
Austria	<b>quad</b>	-0.95583	0.913616	0.95827	0.95110	0.907177	0.891355
Belgio	<b>quad</b>	-0.96944	0.939827	0.94910	0.94312	0.927025	0.916474
Bulgaria	<b>quad</b>	-0.94974	0.902007	0.91444	0.83993	0.903172	0.907101
Cipro	<b>quad</b>	-0.73141	0.534967	0.62233	0.49303	0.500934	0.514137
Croazia	<b>quad</b>	-0.23627	0.055824	0.06922	0.06449	0.058827	0.054650
Danimarca	<b>quad</b>	-0.88802	0.788591	0.84637	0.69157	0.766891	0.776820
Estonia	<b>quad</b>	-0.90682	0.822328	0.82242	0.75180	0.789957	0.763102
Finlandia	<b>quad</b>	-0.93558	0.875325	0.87832	0.85702	0.842070	0.831480
Francia	<b>quad</b>	-0.90150	0.812714	0.84220	0.83414	0.792350	0.779961
Germania	<b>quad</b>	-0.86699	0.751682	0.95934	0.63286	0.772965	0.798328
Grecia	<b>quad</b>	-0.06304	0.003975	0.02480	0.00697	0.004084	0.003048
Ungheria	<b>quad</b>	-0.82410	0.679154	0.68041	0.65151	0.676270	0.662848
Irlanda	<b>quad</b>	-0.79732	0.635733	0.78899	0.26793	0.635821	0.686798
Italia	<b>quad</b>	-0.24189	0.058511	0.21737	0.04697	0.043554	0.046437
Lettonia	<b>semilog</b>	-0.75675	0.572678	0.58095	0.57233	0.588740	0.559929
Lituania	<b>quad</b>	0.20548	0.042223	0.07680	0.01250	0.047561	0.051402
Lussemburgo	<b>quad</b>	-0.89142	0.794643	0.81559	0.77357	0.733462	0.716500
Malta	<b>quad</b>	-0.85927	0.738346	0.92892	0.49425	0.748150	0.809075
Paesi Bassi	<b>quad</b>	-0.90330	0.815963	0.81648	0.77929	0.792471	0.785325
Polonia	<b>quad</b>	-0.86608	0.750103	0.91641	0.55474	0.777959	0.849038
Portogallo	<b>quad</b>	-0.69087	0.477301	0.48437	0.45072	0.452015	0.455263

Table 5: Confronto coefficienti di correlazione e determinazione per i vari Modelli di Regressione

	best_model	c_cor	lin	quad	exp	semilog	log
Rep. Ceca	<b>quad</b>	-0.79346	0.629586	0.83523	0.47344	0.652136	0.698443
Romania	<b>quad</b>	-0.87357	0.763129	0.82512	0.58668	0.766593	0.801107
Slovacchia	<b>quad</b>	-0.89036	0.792742	0.79466	0.77607	0.792613	0.779287
Slovenia	<b>quad</b>	-0.91919	0.844928	0.85721	0.75083	0.829795	0.833811
Spagna	<b>quad</b>	-0.75644	0.572205	0.57595	0.54599	0.540978	0.542838
Svezia	<b>quad</b>	-0.97335	0.947424	0.95169	0.87334	0.941346	0.938632

Table 6: Confronto coefficienti di correlazione e determinazione per i vari Modelli di Regressione

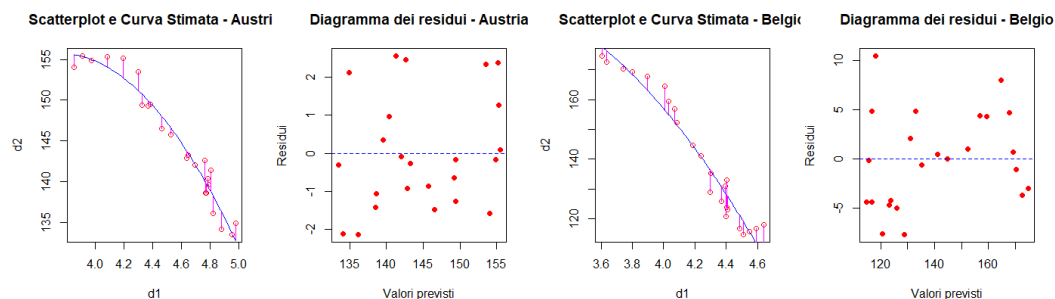


Figure 45: Grafici di Regressione - Austria

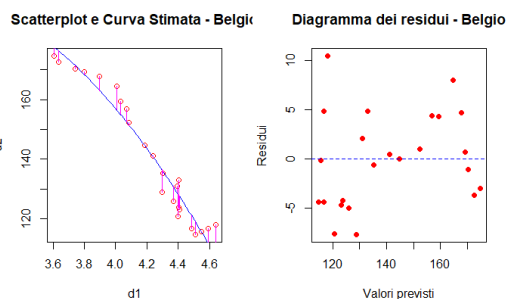


Figure 46: Grafici di Regressione - Belgio

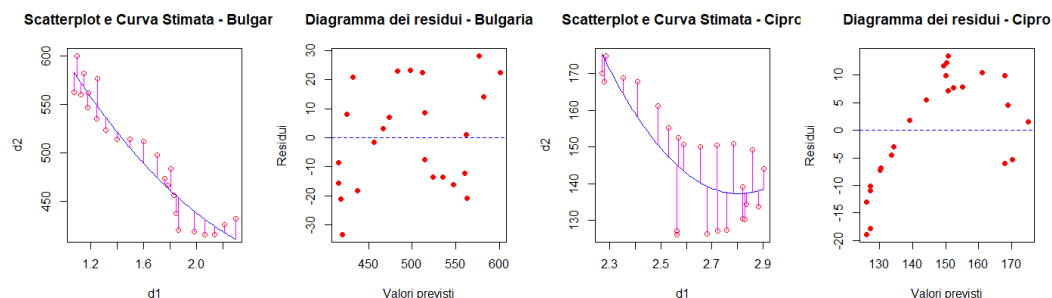


Figure 47: Grafici di Regressione - Bulgaria

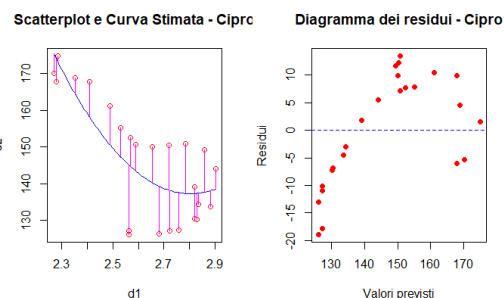


Figure 48: Grafici di Regressione - Cipro

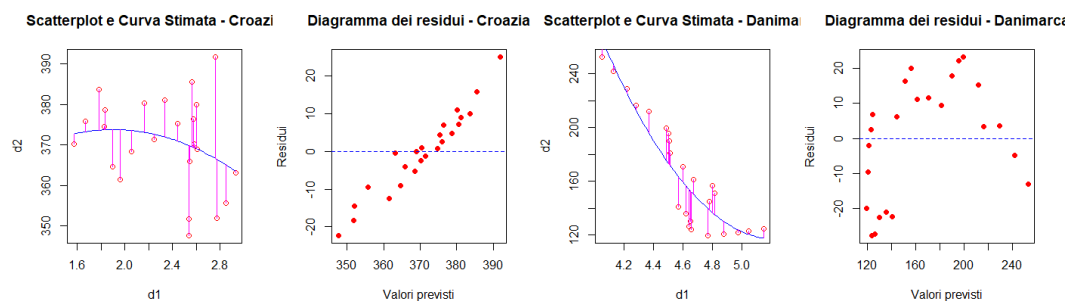


Figure 49: Grafici di Regressione - Croazia

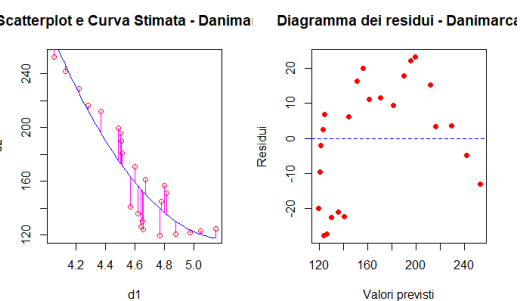


Figure 50: Grafici di Regressione - Danimarca



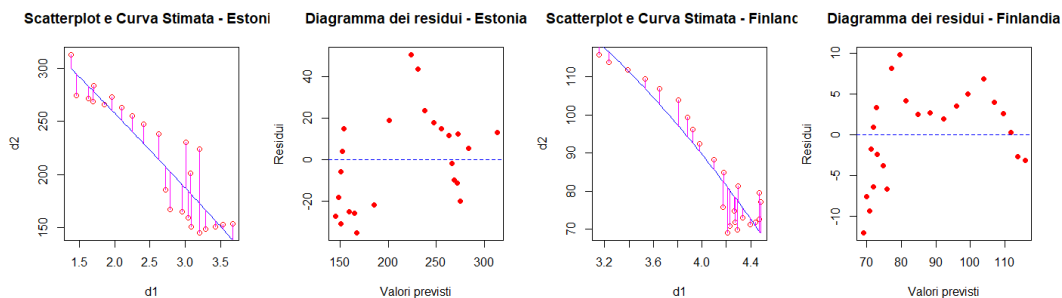


Figure 51: Grafici di Regressione - Estonia Figure 52: Grafici di Regressione - Finlandia

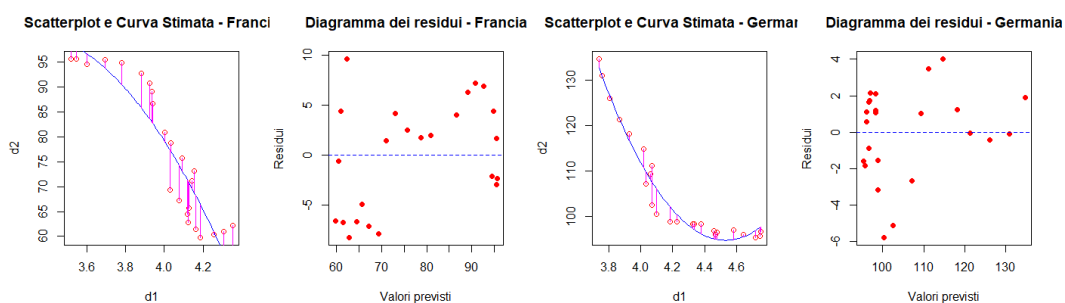


Figure 53: Grafici di Regressione - Francia Figure 54: Grafici di Regressione - Germania

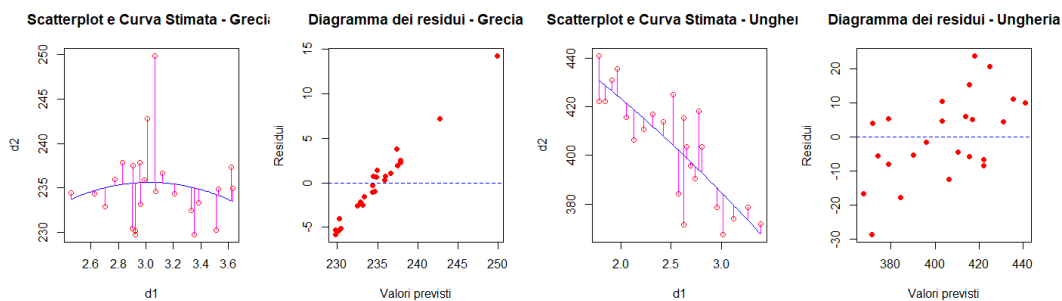


Figure 55: Grafici di Regressione - Grecia Figure 56: Grafici di Regressione - Ungheria

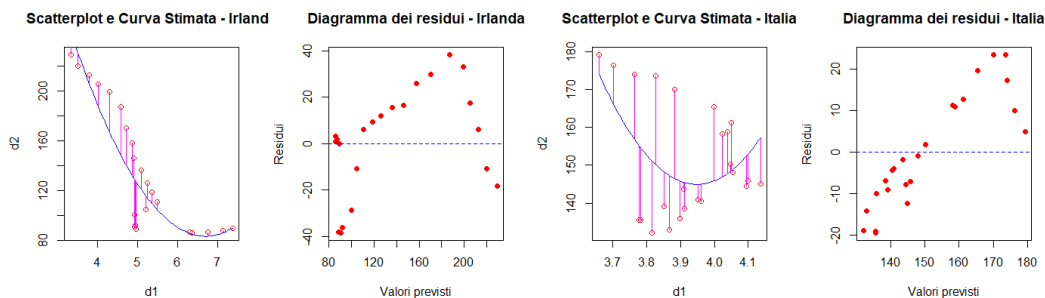


Figure 57: Grafici di Regressione - Irlanda

Figure 58: Grafici di Regressione - Italia

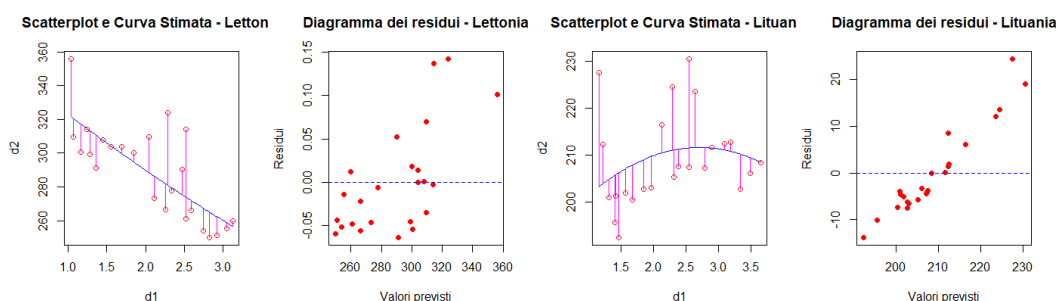


Figure 59: Grafici di Regressione - Lettonia

Figure 60: Grafici di Regressione - Lituania

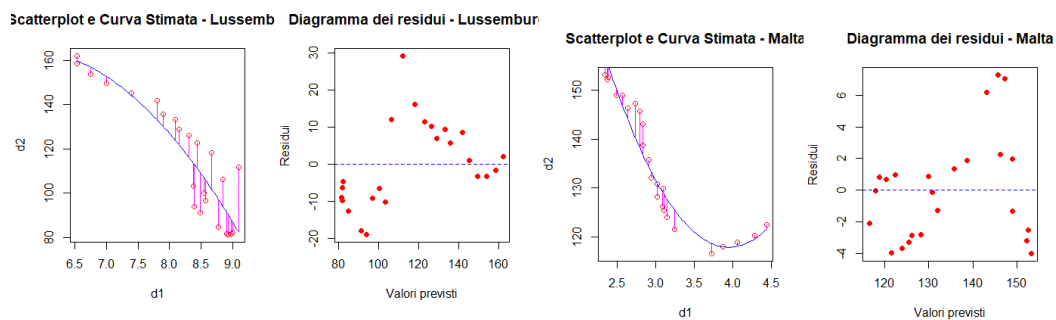


Figure 61: Grafici di Regressione - Lussemburgo

Figure 62: Grafici di Regressione - Malta

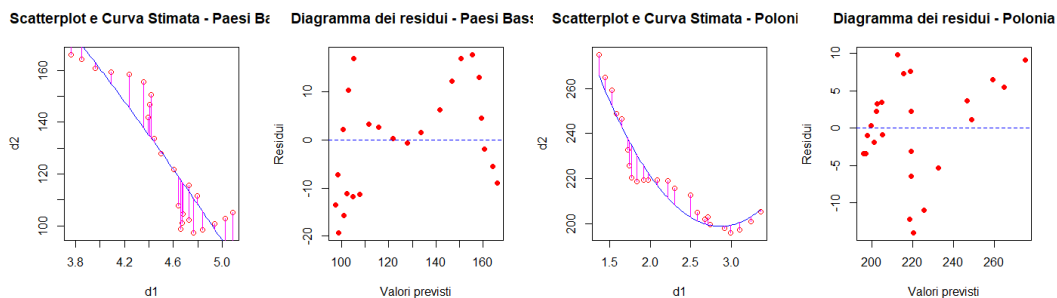


Figure 63: Grafici di Regressione - Paesi Bassi Figure 64: Grafici di Regressione - Polonia

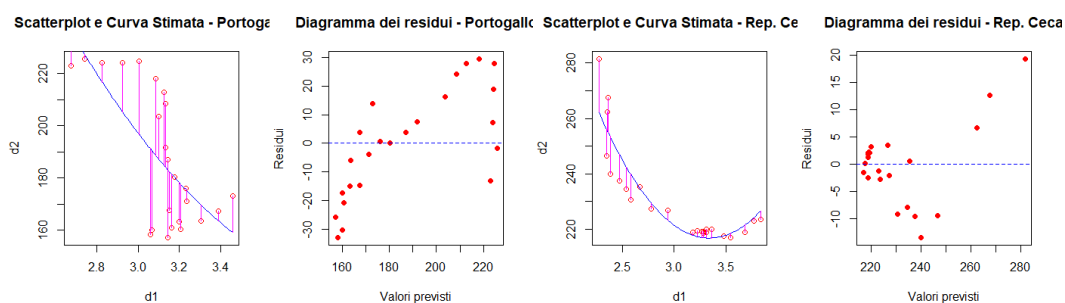


Figure 65: Grafici di Regressione - Portogallo Figure 66: Grafici di Regressione - Rep. Ceca

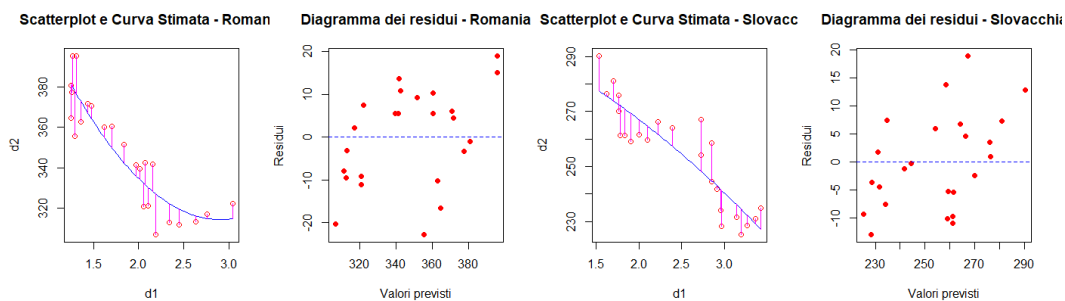


Figure 67: Grafici di Regressione - Romania Figure 68: Grafici di Regressione - Slovacchia

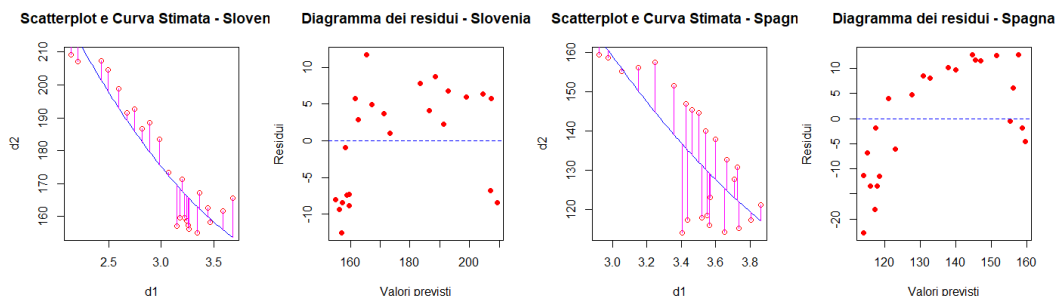


Figure 69: Grafici di Regressione - Slovenia      Figure 70: Grafici di Regressione - Spagna

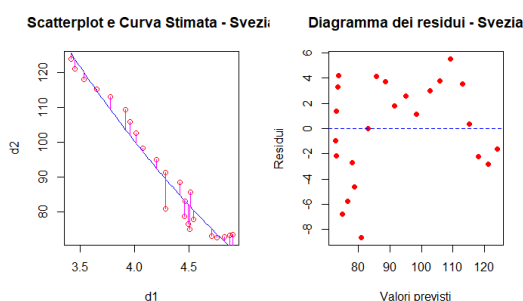


Figure 71: Grafici di Regressione - Svezia

#### 4.1.1 Conclusioni

In conclusione, l'analisi di regressione ha evidenziato una correlazione significativa tra DPM e VSL, con il *modello di regressione non lineare quadratico* che si adatta meglio ai dati nella quasi totalità dei casi. In particolare, i risultati evidenziano che per il 67% dei paesi analizzati il coefficiente di determinazione supera il 0.7, indicando una **correlazione negativa molto forte tra DPM e VSL**. Tuttavia, per alcuni paesi questa forte correlazione non si verifica, e i pattern nei diagrammi dei residui suggeriscono che altri modelli potrebbero essere più adeguati a descrivere la relazione tra queste variabili in tali contesti. In generale, questa correlazione negativa tra VSL e DPM suggerisce che *migliori condizioni economiche o investimenti nella salute e sicurezza possono portare a una riduzione dei tassi di mortalità*, riflettendo l'importanza del VSL come indicatore nel miglioramento delle condizioni di vita e della salute pubblica.

## 5 Analisi dei Cluster

L'**analisi dei Cluster** mira a identificare gruppi omogenei di paesi sulla base delle variabili quantitative esaminate. Questa fase potrebbe indurre all'*identificazione di interessanti legami* tra i paesi appartenenti ad uno stesso gruppo, rilevando potenzialmente pattern non emersi fino a questo momento e dinamiche più complesse che regolano il fenomeno analizzato.

L'approccio adottato è stato caratterizzato da una procedura ciclica che ha coinvolto l'applicazione di diversi metodi di clustering. Privilegiando l'utilizzo della **metrica Euclidea** tra le altre, per la definizione della *matrice delle distanze*, sono stati esaminati diversi esiti ottenuti utilizzando un numero variabile di cluster, compreso tra 2 e 10, al fine di determinare la quantità ottimale di cluster per ciascun metodo, mediante l'analisi di Screeplot e grafici di non omogeneità.

### 5.1 DPM - Metodi Gerarchici

#### 5.1.1 Metodo del Legame Singolo

L'applicazione del **Metodo del Legame Singolo** alla matrice delle distanze di DPM ha permesso un'analisi approfondita della distribuzione dei DPM tra i paesi dell'Unione Europea. L'osservazione degli Screeplot (vedi Figura 73) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 74, 75), hanno determinato che la suddivisione ottimale dei dati si ottiene con **4 cluster** (vedi Figura 76). Questa metodologia ha evidenziato i paesi con i tassi di mortalità per fumo passivo più critici, isolando due cluster composti unicamente da un singolo paese: Bulgaria (C1), che rappresenta l'*outlier* principale del dataset, e Ungheria (C2), con il secondo più alto valore di DPM. Questa distinzione enfatizza l'importanza di considerare gli estremi nella valutazione delle politiche di salute pubblica. Un terzo cluster raggruppa Croazia e Romania (C3), rispettivamente il terzo e il quarto paese per DPM, sottolineando una fascia di rischio intermedia. Infine, l'ultimo cluster (C4) aggrega la maggior parte degli altri paesi, caratterizzati da tassi di DPM relativamente più bassi.

In conclusione, l'utilizzo del Metodo del Legame Singolo ha rivelato aspetti importanti riguardanti la distribuzione del rischio associato al fumo passivo nell'Unione Europea, permettendo un'identificazione precisa dei paesi con i valori più critici di mortalità. Questo metodo può essere sfruttato per concentrare gli sforzi su quei paesi che registrano i tassi più elevati di DPM, offrendo al contempo una base metodologica solida per future indagini e l'elaborazione di politiche pubbliche più incisive.

Partizione	Paesi
C1	Bulgaria
C2	Ungheria
C3	Croazia, Romania
C4	Lettonia, Estonia, Danimarca, Portogallo, Slovenia, Francia, Finlandia, Svezia, Lituania, Slovacchia, Grecia, Polonia, Rep. Ceca, Germania, Lussemburgo, Paesi Bassi, Irlanda, Austria, Cipro, Italia, Belgio, Malta, Spagna

Table 7: Partizioni ottenute con il Metodo del Legame Singolo - DPM

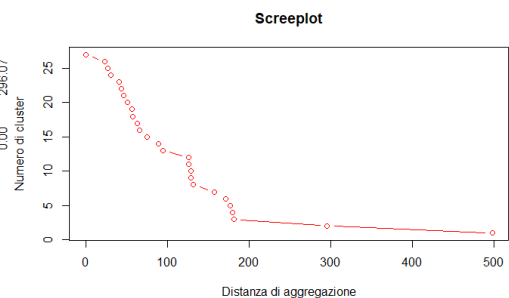
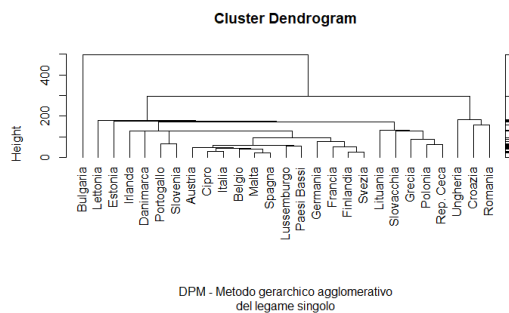


Figure 72: DPM - Legame Singolo - Dendrogramma

Figure 73: DPM - Legame Singolo - Screeplot

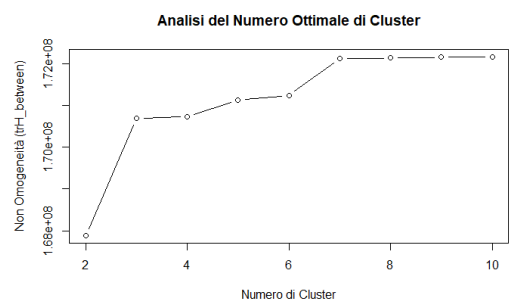
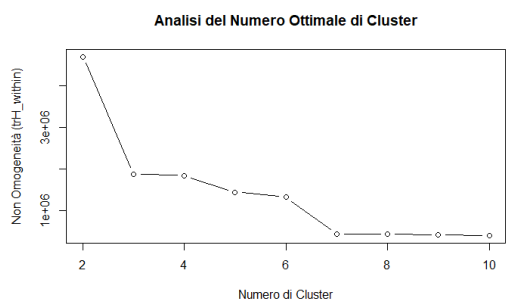


Figure 74: DPM - Legame Singolo - trH within

Figure 75: DPM - Legame Singolo - trH between

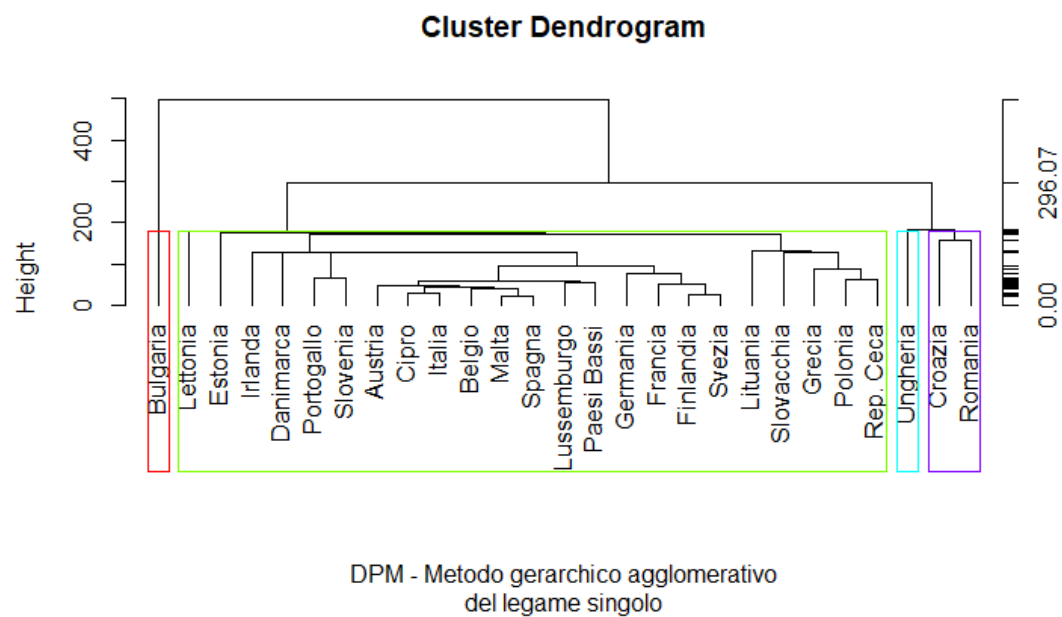


Figure 76: DPM - Legame Singolo - 4 Cluster

### 5.1.2 Metodo del Legame Completo

Nell'analisi dei DPM utilizzando il **Metodo del Legame Completo**, approccio che enfatizza l'omogeneità all'interno dei cluster a scapito della differenziazione tra i gruppi, l'analisi di Screeplot (vedi Figure 78) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 79, 80) ha portato alla determinazione di **6 cluster** come configurazione ottimale (vedi Figura 81). Anche questo approccio ha permesso di approfondire la comprensione sulla distribuzione dei decessi attribuibili al fumo passivo tra i paesi dell'Unione Europea, seppur con una visione leggermente diversa. Questo metodo ha nuovamente identificato la Bulgaria come un outlier significativo, assegnandola a un cluster isolato (C1) a causa della sua distanza marcata da ogni altro paese. Interessante notare come, a differenza del metodo del Legame Singolo, Ungheria, Croazia e Romania siano stati raggruppati insieme (C2), riflettendo una coerenza nelle elevate mortalità che li distingue dagli altri. I restanti quattro cluster (C3, C4, C5, C6) articolano i paesi rimanenti in gruppi più omogenei, dimostrando la capacità del metodo del legame completo di individuare sottogruppi con caratteristiche simili in termini di DPM, privilegiando così l'omogeneità interna.

In conclusione, è stata rivelata una struttura di clustering che pone un'enfasi distinta sull'omogeneità interna, conducendo all'identificazione di sei cluster distinti. Il raggruppamento in cluster più definiti e omogenei fornisce una base solida per ulteriori indagini e per l'elaborazione di politiche sanitarie mirate, con l'obiettivo di affrontare in modo efficace l'impatto del fumo passivo sulla salute pubblica in Europa. Questi risultati sottolineano l'importanza di approcci metodologici diversificati a seconda dei paesi.

Partizione	Paesi
C1	Bulgaria
C2	Ungheria, Croazia, Romania
C3	Francia, Finlandia, Svezia, Germania, Lussemburgo, Paesi Bassi, Irlanda, Austria, Cipro, Italia, Belgio, Malta, Spagna
C4	Lettonia, Slovacchia
C5	Lituania, Grecia, Polonia, Rep. Ceca
C6	Estonia, Danimarca, Portogallo, Slovenia

Table 8: Partizioni ottenute con il Metodo del Legame Completo - DPM

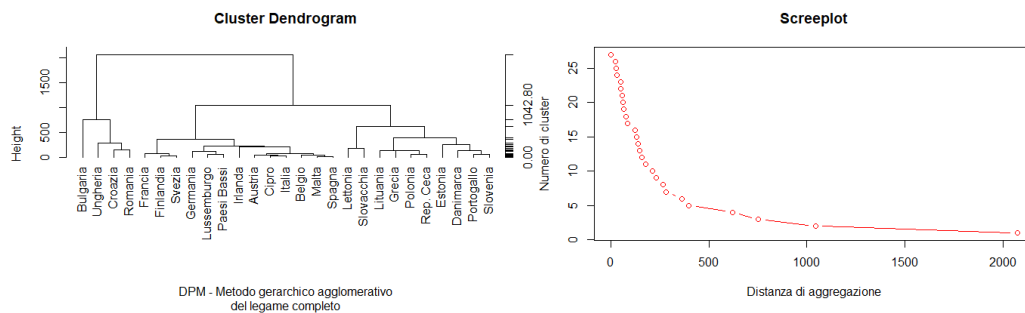


Figure 77: DPM - Legame Completo - Den-  
dogramma

Figure 78: DPM - Legame Completo -  
Screeplot

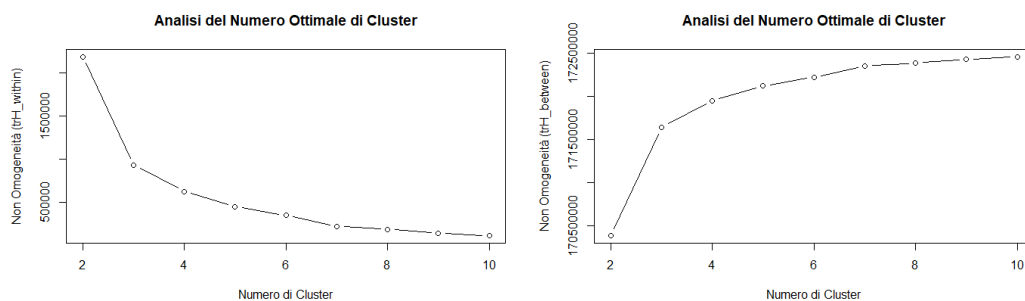


Figure 79: DPM - Legame Completo - trH  
within

Figure 80: DPM - Legame Completo - trH  
between

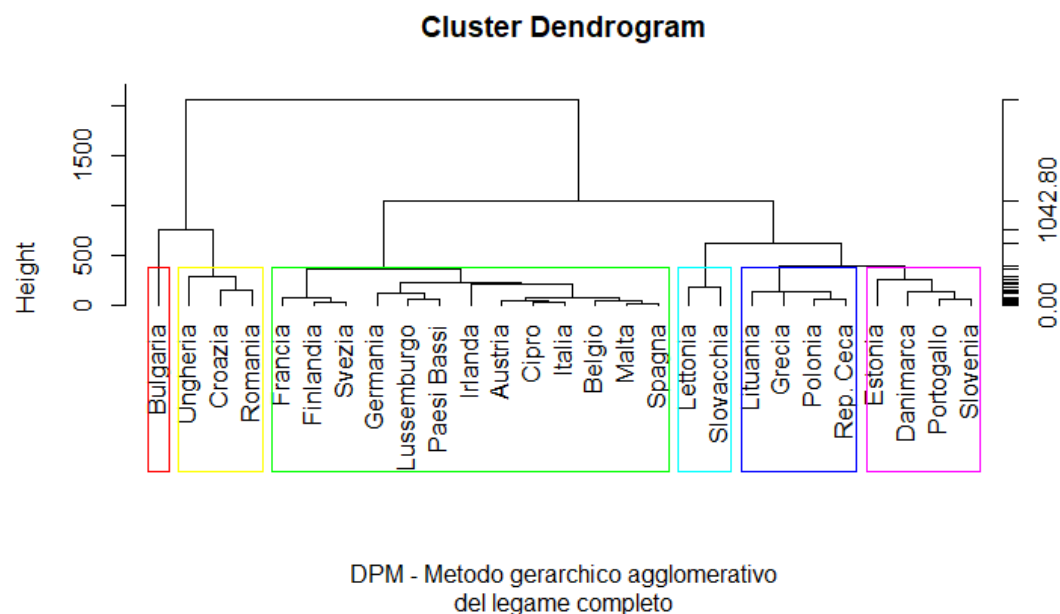


Figure 81: DPM - Legame Completo - 6 Cluster



### 5.1.3 Metodo del Legame Medio

L'impiego del **Metodo del Legame Medio** per DPM ha offerto una prospettiva equilibrata tra l'aggregazione dei cluster. Questa metodologia ha permesso di analizzare le distanze medie tra tutti i possibili accoppiamenti di cluster, offrendo così un compromesso tra sensibilità agli outlier e tendenza alla formazione di cluster più coesi e rappresentativi. Attraverso l'analisi dello Screeplot (vedi Figura 83) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 84, 85), è stata identificata una configurazione ottimale di **5 cluster** (vedi Figura 86). Tale segmentazione cluster ha rivelato una struttura interessante, con la Lettonia emergente come un cluster indipendente (C2), sottolineando la sua posizione distinta nel contesto dei DPM all'interno dell'UE. Allo stesso modo, si conferma la presenza di un cluster esclusivo per la Bulgaria (C4) e di un altro che raggruppa Ungheria, Croazia e Romania (C5), coerentemente con quanto osservato in precedenza con altri metodi di clustering. Questa analisi ha inoltre permesso di distinguere due ulteriori gruppi di paesi in base ai loro valori di DPM: uno collocato nella fascia medio-alta (C3) e l'altro nella fascia bassa della distribuzione (C1).

In conclusione, è stata evidenziata una disposizione di cluster che bilancia efficacemente l'omogeneità interna e l'eterogeneità esterna. Identificando cinque distinti cluster, questa metodologia ha non solo confermato l'unicità di alcuni paesi come la Bulgaria e la Lettonia, ma ha anche facilitato una suddivisione più sfumata e informativa degli altri paesi. Infine, la capacità di questo metodo di mediare tra estremi e fornire una visione complessiva equilibrata rende i suoi risultati particolarmente utili per orientare le politiche sanitarie e le iniziative preventive in tutta l'Unione Europea.

<b>Partizione</b>	<b>Paesi</b>
C1	Germania, Francia, Finlandia, Svezia, Danimarca Portogallo, Slovenia, Irlanda, Lussemburgo, Paesi Bassi Belgio, Malta, Spagna, Austria, Cipro, Italia
C2	Lettonia
C3	Estonia, Slovacchia, Lituania, Grecia, Polonia, Rep. Ceca
C4	Bulgaria
C5	Ungheria, Croazia, Romania

Table 9: Partizioni ottenute con il Metodo del Legame Medio - DPM

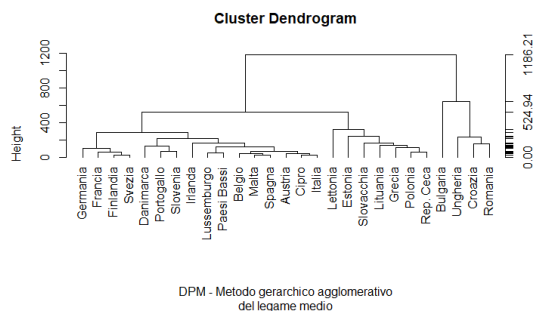


Figure 82: DPM - Legame Medio - Dendrogramma

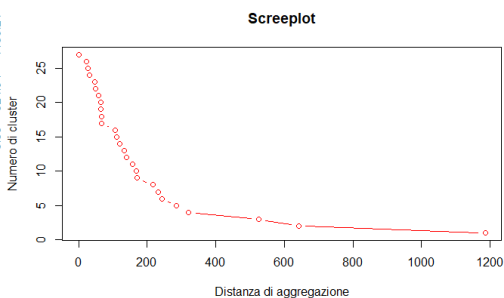


Figure 83: DPM - Legame Medio - Screeplot

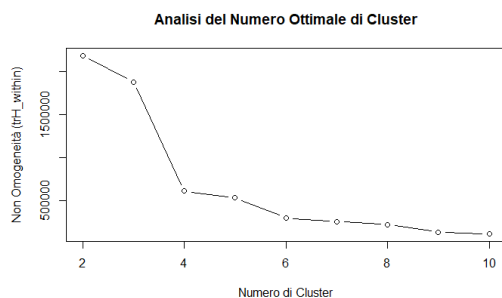


Figure 84: DPM - Legame Medio - trH within

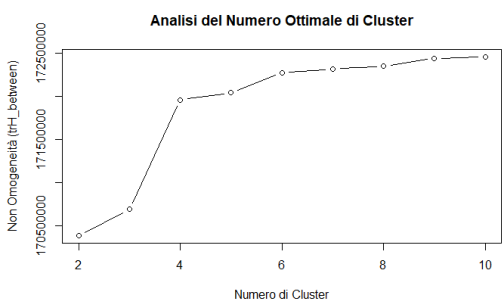


Figure 85: DPM - Legame Medio - trH between

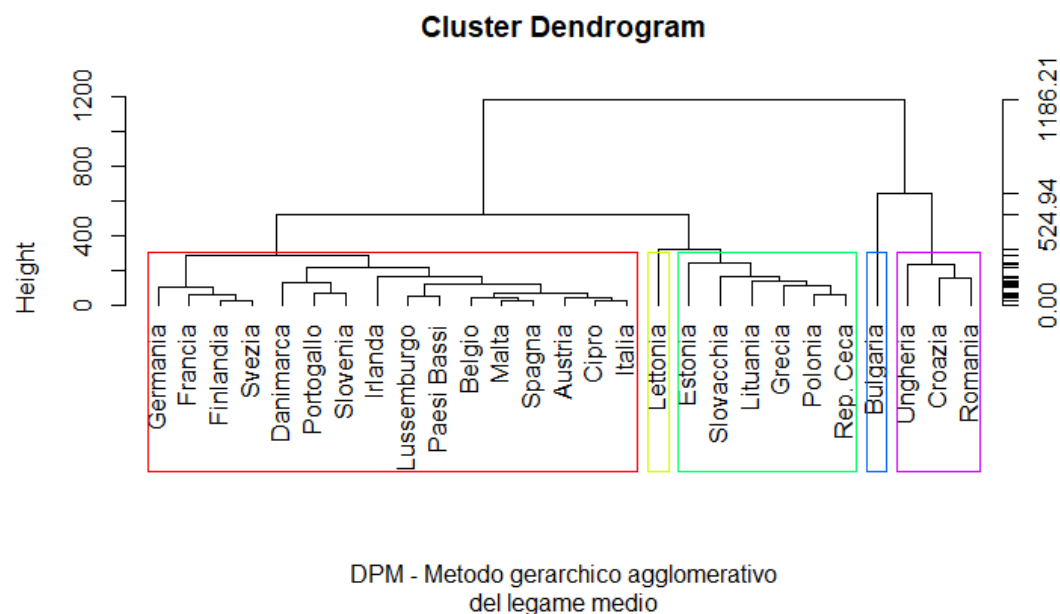


Figure 86: DPM - Legame Medio - 5 Cluster

### 5.1.4 Metodo del Centroide

L'applicazione del **Metodo del Centroide**, che considera il punto centrale di un cluster come riferimento per l'aggregazione, con il rischio di dare luogo a fenomeni gravitazionali in cui gruppi più grandi tendono ad attrarre quelli più piccoli, ci ha permesso di effettuare un'ulteriore analisi dell'indicatore. Attraverso l'analisi dello Screeplot (vedi Figura 88) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 89, 90), è stata identificata una configurazione ottimale di **5 cluster** (vedi Figura 91). Tale suddivisione, esattamente identica a quella ottenuta con il Metodo del Legame Medio, suggerisce una convergenza tra i centri dei cluster che potrebbe riflettere una distribuzione omogenea e bilanciata dei dati analizzati.

In conclusione, questo metodo ha rivelato una struttura di cluster coerente con i risultati ottenuti attraverso il Metodo del Legame Medio. La **replicabilità dei cluster** tra diversi metodi sottolinea potrebbe indicare una robustezza nella segmentazione dei paesi dell'UE basata sull'indicatore DPM, offrendo spunti preziosi per interpretazioni più profonde e per la formulazione di politiche sanitarie mirate in base alle differenze nazionali.

Partizione	Paesi
C1	Germania, Francia, Finlandia, Svezia, Danimarca Portogallo, Slovenia, Irlanda, Lussemburgo, Paesi Bassi Belgio, Malta, Spagna, Austria, Cipro, Italia
C2	Lettonia
C3	Estonia, Slovacchia, Lituania, Grecia, Polonia, Rep. Ceca
C4	Bulgaria
C5	Ungheria, Croazia, Romania

Table 10: Partizioni ottenute con il Metodo del Centroide - DPM

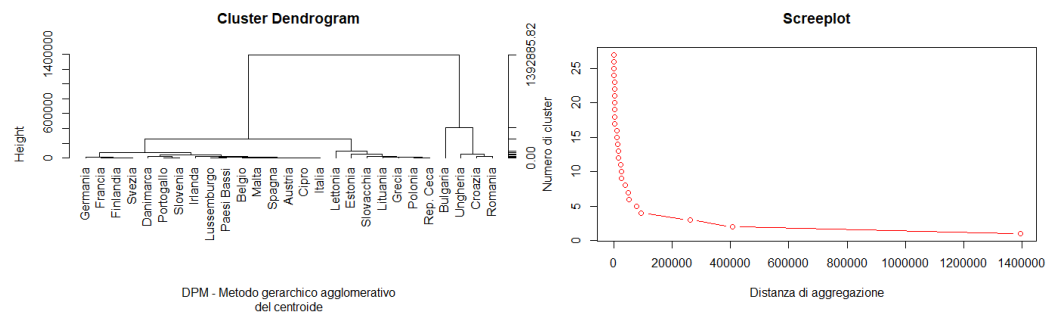


Figure 87: DPM - Metodo Centroide - Den-Dendrogram

Figure 88: DPM - Metodo Centroide - Screeplot

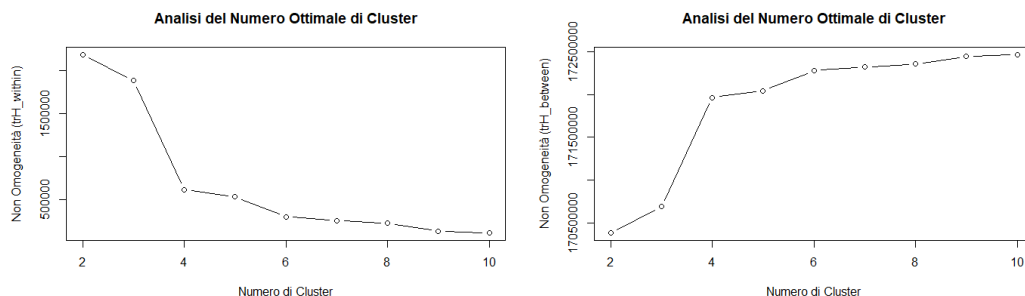


Figure 89: DPM - Metodo Centroide - trH within  
Figure 90: DPM - Metodo Centroide - trH between

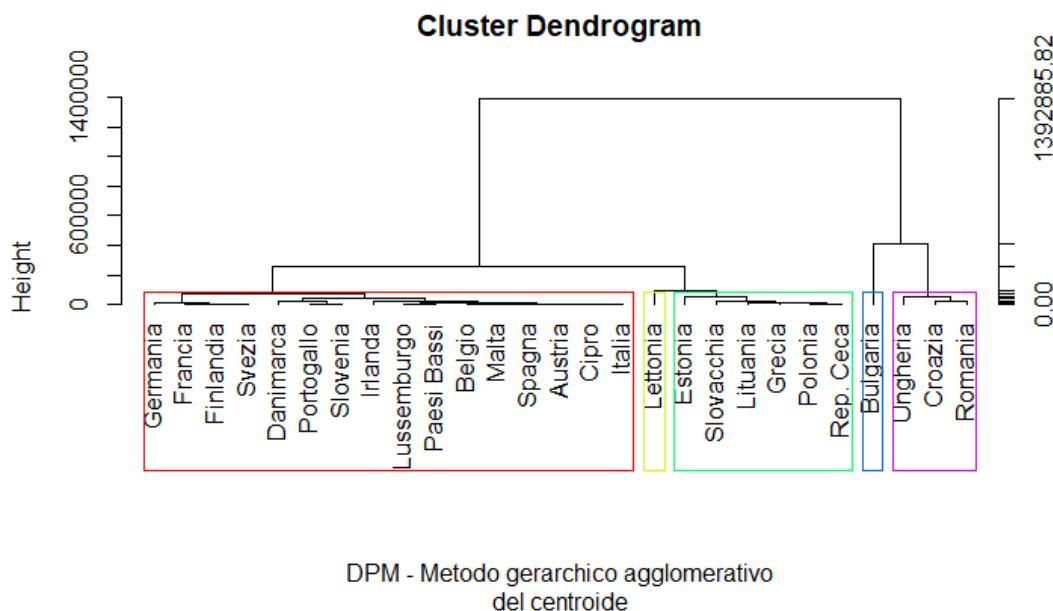


Figure 91: DPM - Metodo Centroide - 5 Cluster

### 5.1.5 Metodo della Mediana

Il Metodo della Mediana ha fornito una nuova prospettiva nell'analisi dei DPM, enfatizzando un approccio centrato sul valore mediano per definire i centri dei cluster. Anche questo metodo, può dare origine a catene tra gli individui. Attraverso l'analisi dello Screeplot (vedi Figura 93) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 94, 95), è stata identificata una configurazione ottimale di **5 cluster** (vedi Figura 96). Interessante notare come, anche con questo metodo, la configurazione ottimale sia stata di 5 cluster, con una particolare attenzione ai paesi con i valori di DPM più bassi, quali Francia, Finlandia e Svezia (C3), che formano un cluster distinto. Questa suddivisione sottolinea una specifica configurazione dei dati che evidenzia i paesi con migliori performance in termini di controllo del fumo passivo.

In conclusione, l'applicazione del Metodo della Mediana nell'analisi dei DPM ha confermato la tendenza a formare una configurazione ottimale di 5 cluster, con risultati che mettono in evidenza sia i paesi con i tassi di DPM più bassi sia quelli con valori significativamente più alti. Questi risultati enfatizzano, ulteriormente, l'importanza di un'analisi approfondita e diversificata per ogni gruppo di paesi.

Partizione	Paesi
C1	Germania, Danimarca, Portogallo, Slovenia Irlanda, Lussemburgo, Paesi Bassi, Belgio, Malta, Spagna, Austria, Cipro, Italia
C2	Francia, Finlandia, Svezia
C3	Lettonia, Estonia, Slovacchia, Lituania, Grecia, Polonia, Rep. Ceca
C4	Bulgaria
C5	Ungheria, Croazia, Romania

Table 11: Partizioni ottenute con il Metodo della Mediana - DPM

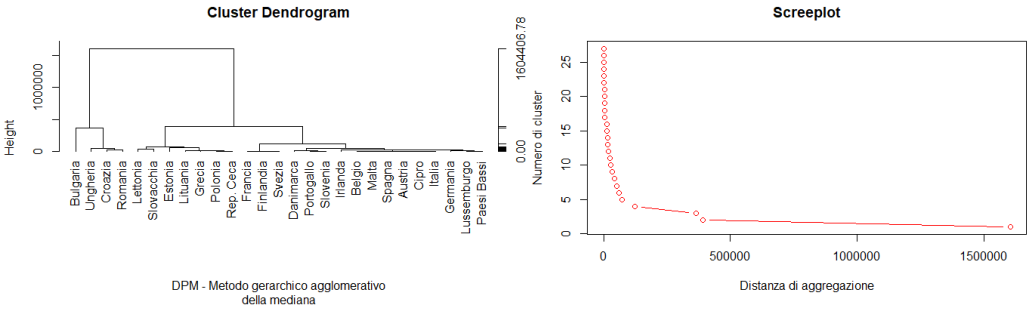


Figure 92: DPM - Metodo Mediana - Dendo-gramma Figure 93: DPM - Metodo Mediana - Screeplot

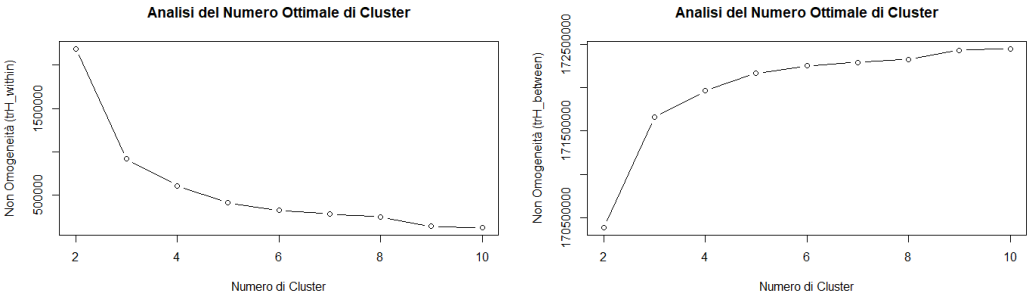


Figure 94: DPM - Metodo Mediana - trH within Figure 95: DPM - Metodo Mediana - trH between

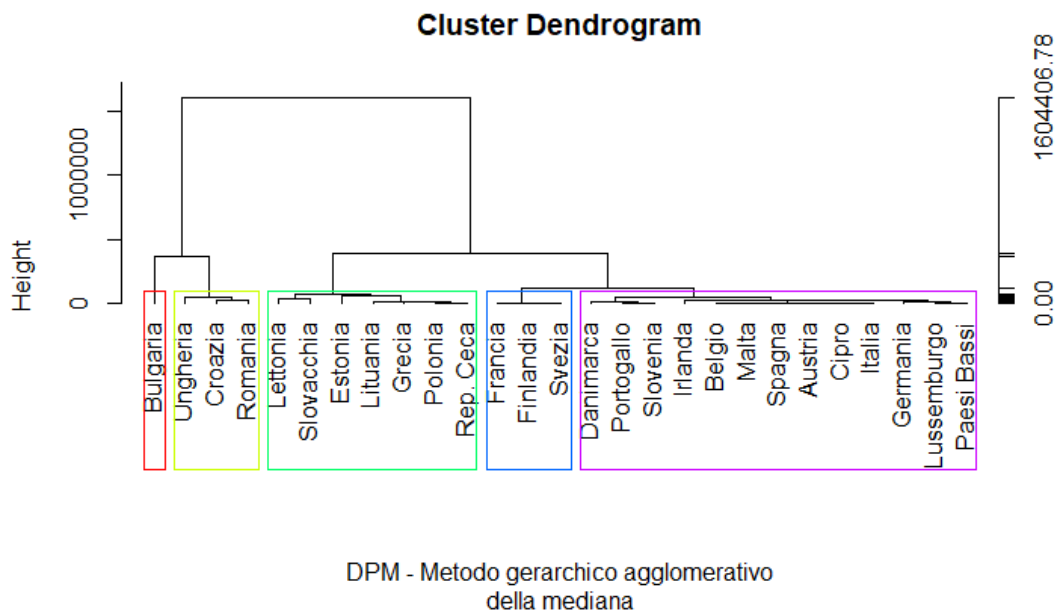


Figure 96: DPM - Metodo Mediana - 5 Cluster

## 5.2 DPM - Metodi Non Gerarchici

### 5.2.1 Metodo kmeans

Basandoci sui risultati ottenuti dalle analisi precedenti, abbiamo condotto l'analisi con il **Metodo del kmeans**, noto per la sua efficienza e la capacità di minimizzare la varianza interna ai cluster, sperimentando con un numero di *cluster variabile da 4 a 6*. Questo processo di valutazione ha avuto l'obiettivo di determinare la configurazione ottimale che potesse al meglio rappresentare le dinamiche dei dati analizzati. Tra le configurazioni testate, la suddivisione in **5 cluster** si è confermata come la scelta ottimale (vedi Figure 97), per la sua capacità di mantenere una coerenza con le suddivisioni precedentemente identificate e per esplorare in modo efficace come questa metodologia potesse differenziare o confermare le aggregazioni dei paesi.

In conclusione, la scelta di procedere con 5 cluster, dopo aver valutato configurazioni da 4 a 6, si è rivelata ideale anche con il Metodo del kmeans. Questa configurazione ottimale non solo conferma le analisi precedenti ma fornisce anche una base solida per interpretazioni più approfondite e per l'elaborazione di politiche sanitarie mirate.

Partizione	Paesi
C1	Austria, Belgio, Cipro, Danimarca, Irlanda, Italia, Malta, Paesi Bassi, Portogallo, Slovenia, Spagna
C2	Finlandia, Francia, Germania, Lussemburgo, Svezia
C3	Estonia, Grecia, Lettonia, Lituania, Polonia, Rep. Ceca, Slovacchia
C4	Bulgaria
C5	Ungheria, Croazia, Romania

Table 12: Partizioni ottenute con il Metodo kmeans - DPM

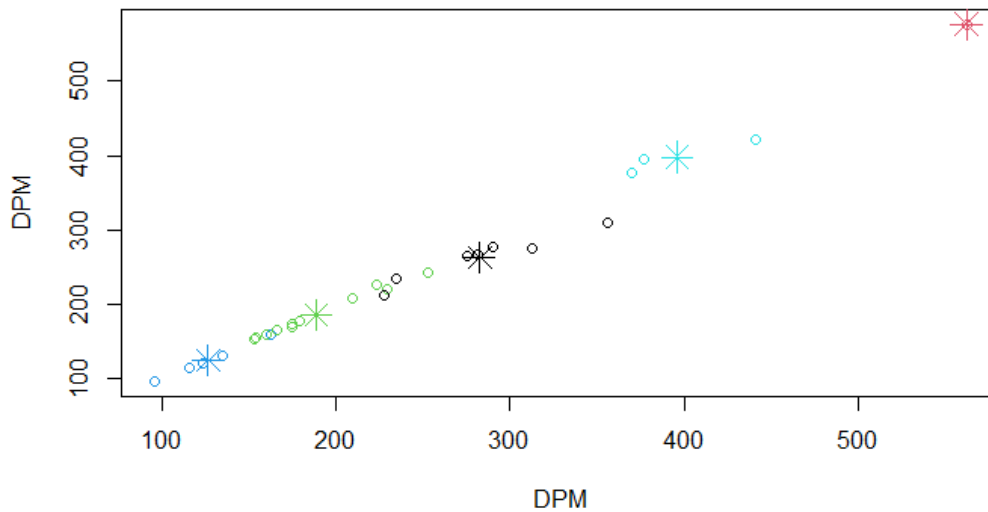


Figure 97: DPM - Metodo kmeans - 5 Cluster

## 5.3 VSL - Metodi Gerarchici

### 5.3.1 Metodo del Legame Singolo

Il **Metodo del Legame Singolo**, che considera la minima distanza tra gli elementi di cluster differenti, ha portato all'analisi di Screeplot (vedi Figure 99) e Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 101, 102), rivelando che la suddivisione ideale sia in **4 cluster** (vedi Figure 100). Quest'analisi ha evidenziato le variazioni significative nei valori di VSL, distinguendo in modo particolare Lussemburgo (C1) e Irlanda (C2), i due paesi con i valori più elevati di VSL, assegnandoli a cluster separati e isolati, successivamente, sono presenti i paesi raggruppati in valori medio-alti (C3) e in medio-bassi (C4). Questa suddivisione ha offerto una base per indagare le divergenze nell'attribuzione del valore VSL tra i paesi dell'Unione Europea.

In conclusione, quest'analisi, ha fornito una panoramica chiara e dettagliata delle differenze di VSL tra i paesi dell'UE. Questi risultati forniscono una base importante per ulteriori discussioni e analisi sulle politiche di valutazione del rischio e sulla predisposizione di strategie di intervento mirate.

Partizione	Paesi
C1	Lussemburgo
C2	Irlanda
C3	Danimarca, Austria, Paesi Bassi, Italia, Svezia Belgio, Germania, Finlandia, Francia
C4	Cipro, Bulgaria, Romania, Croazia, Estonia, Ungheria, Slovacchia, Lettonia, Lituania, Polonia Spagna, Grecia, Portogallo, Malta, Rep. Ceca, Slovenia

Table 13: Partizioni ottenute con il Metodo del Legame Singolo - DPM

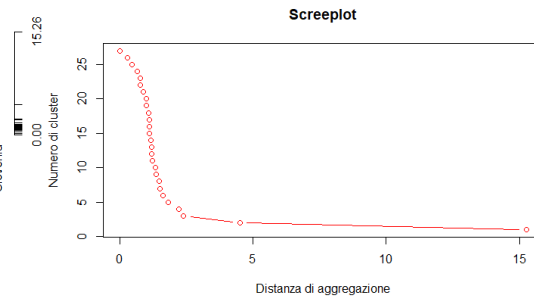
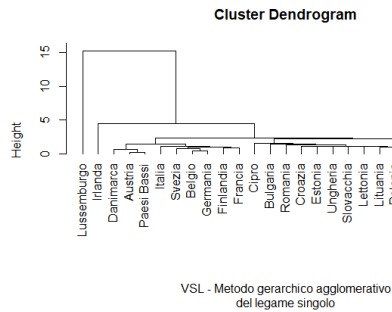


Figure 98: VSL - Legame Singolo - Dendrogramma

Figure 99: VSL - Legame Singolo - Screeplot



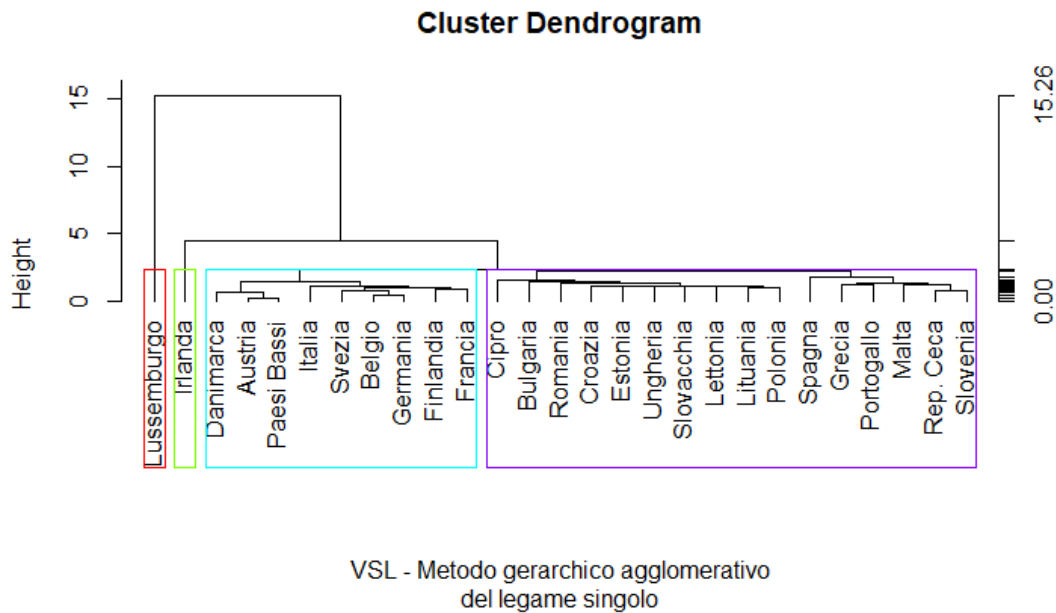


Figure 100: VSL - Legame Singolo - 4 Cluster

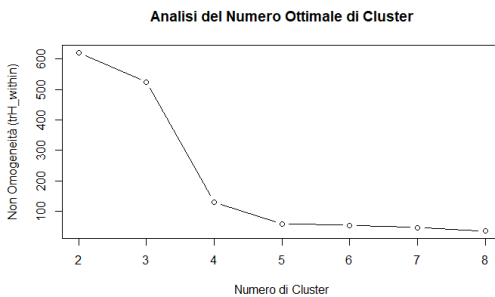


Figure 101: VSL - Legame Singolo - trH within

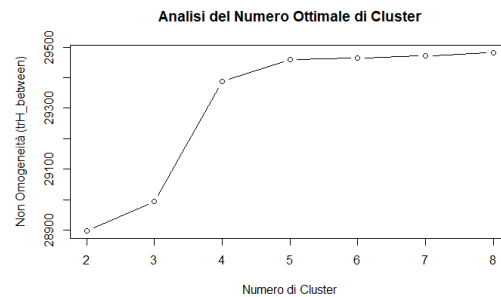


Figure 102: VSL - Legame Singolo - trH between

### 5.3.2 Metodo del Legame Completo

Nell'analisi VSL utilizzando il **Metodo del Legame Completo**, approccio che enfatizza l'omogeneità all'interno dei cluster a scapito della differenziazione tra i gruppi, l'analisi di Screeplot (vedi Figure 104) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 106, 107) ha portato alla determinazione di **5 cluster** come configurazione ottimale (vedi Figura 105). Tale suddivisione presenta i primi due cluster (C1, C2) che si distinguono per valori elevati di VSL, Inoltre, è presente la fascia alta (C3). I restanti due cluster (C4, C5), raggruppano i paesi nelle fasce di valore medio-bassa e bassa di VSL, offrendo una panoramica dettagliata su come i paesi valutino il VSL.

In conclusione, il Metodo del Legame Completo ha rivelato un pattern di distribuzione del VSL tra i paesi dell'Unione Europea che evidenzia differenze marcate nella valutazione del VSL, suggerendo strategie mirate per i vari paesi.

Partizione	Paesi
C1	Lussemburgo
C2	Irlanda
C3	Danimarca, Austria, Paesi Bassi, Italia, Svezia Belgio, Germania, Finlandia, Francia
C4	Malta, Rep. Ceca, Slovenia, Spagna, Grecia, Portogallo
C5	Cipro, Bulgaria, Romania, Croazia, Estonia, Ungheria, Slovacchia, Lettonia, Lituania, Polonia

Table 14: Partizioni ottenute con il Metodo del Legame Completo - DPM

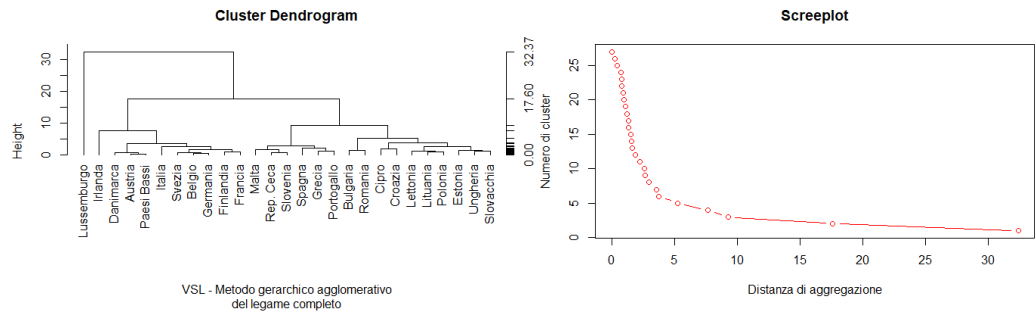


Figure 103: VSL - Legame Completo - Den-Figure 104: VSL - Legame Completo - dendrogram Screeplot

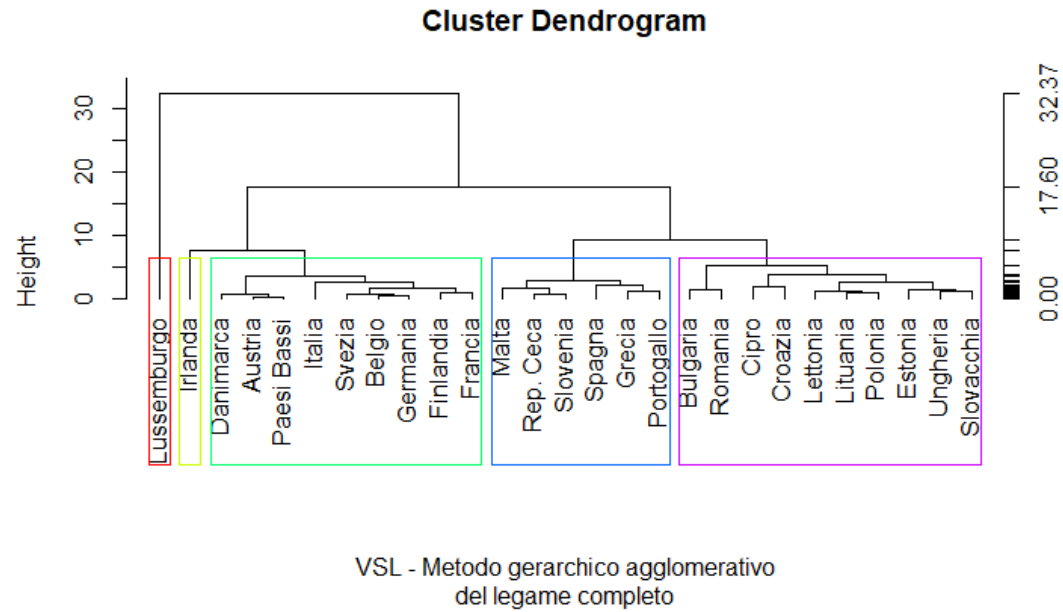


Figure 105: VSL - Legame Completo - 5 Cluster

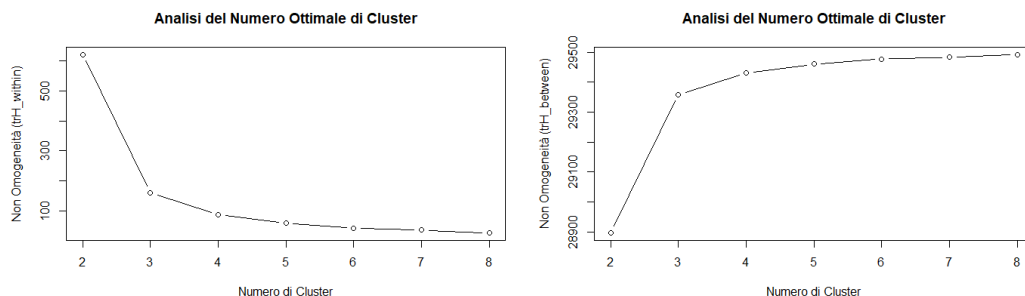


Figure 106: VSL - Legame Completo - trH within      Figure 107: VSL - Legame Completo - trH between

### 5.3.3 Metodo del Legame Medio

L'impiego del **Metodo del Legame Medio** per DPM ha offerto una prospettiva equilibrata tra l'aggregazione dei cluster. Questa metodologia ha permesso di analizzare le distanze medie tra tutti i possibili accoppiamenti di cluster, offrendo così un compromesso tra sensibilità agli outlier e tendenza alla formazione di cluster più coesi e rappresentativi. Attraverso l'analisi dello Screeplot (vedi Figura 109) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 111, 112), è stata identificata una configurazione ottimale di **4 cluster** (vedi Figura 110). Con il metodo del legame medio si è ottenuto un risultato esattamente analogo a quello ottenuto con il metodo del legame singolo, seppur con aggregazioni avvenute a livelli differenti dovute alle differenze negli algoritmi e nei calcoli delle distanze. Nel complesso però, sono stati ottenuti analoghi risultati portando ai medesimi raggruppamenti.

Partizione	Paesi
C1	Lussemburgo
C2	Malta, Rep. Ceca, Slovenia, Spagna, Grecia, Portogallo, Bulgaria, Romania, Cipro, Lettonia, Lituania Polonia, Croazia, Estonia, Ungheria, Slovacchia
C3	Irlanda
C4	Danimarca, Austria, Paesi Bassi, Italia, Svezia Belgio, Germania, Finlandia, Francia

Table 15: Partizioni ottenute con il Metodo del Legame Medio - DPM

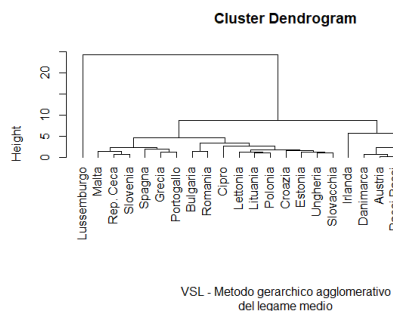


Figure 108: VSL - Legame Medio - Dendrogramma

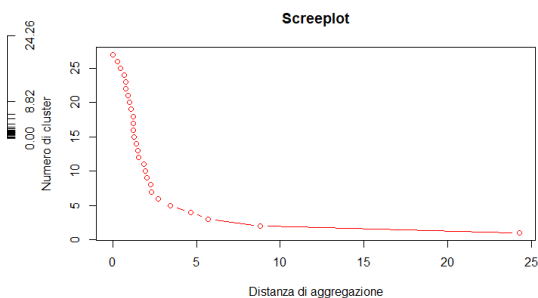


Figure 109: VSL - Legame Medio - Screeplot

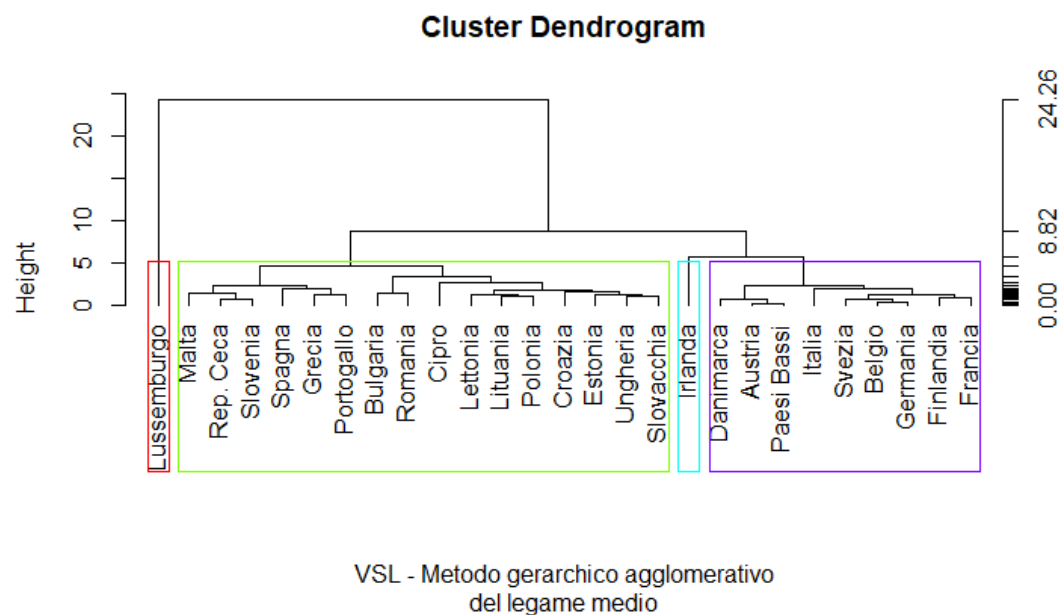


Figure 110: VSL - Legame Medio - 4 Cluster

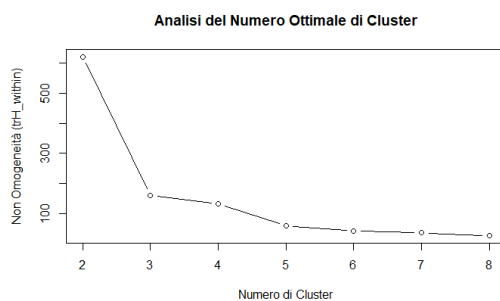


Figure 111: VSL - Legame Medio - trH within

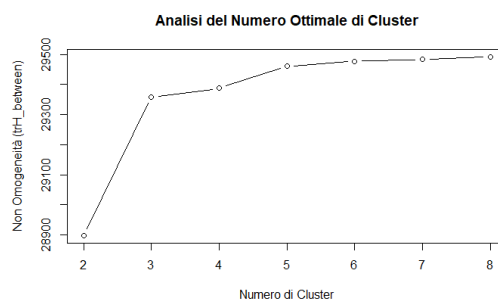


Figure 112: VSL - Legame Medio - trH between

### 5.3.4 Metodo del Centroide

L'applicazione del **Metodo del Centroide**, che considera il punto centrale di un cluster come riferimento per l'aggregazione, con il rischio di dare luogo a fenomeni gravitazionali in cui gruppi più grandi tendono ad attrarre quelli più piccoli, ci ha permesso di effettuare un'ulteriore analisi dell'indicatore. Attraverso l'analisi dello Screeplot (vedi Figura 114) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 116, 117), è stata identificata una configurazione ottimale di **4 cluster** (vedi Figura 115). Tale suddivisione, esattamente identica a quella ottenuta con i Metodi del Legame Singolo e del Legame Medio, suggerisce una convergenza tra i centri dei cluster che potrebbe riflettere una distribuzione omogenea e bilanciata dei dati analizzati.

In conclusione, questo metodo ha rivelato una struttura di cluster coerente con i risultati ottenuti attraverso i suddetti metodi. La **replicabilità dei cluster** tra diversi metodi sottolinea potrebbe indicare una robustezza nella segmentazione dei paesi dell'UE basata sull'indicatore VSL, offrendo spunti preziosi per interpretazioni più profonde e per la formulazione di politiche di intervento mirate in base alle differenze nazionali.

Partizione	Paesi
C1	Lussemburgo
C2	Malta, Rep. Ceca, Slovenia, Spagna, Grecia, Portogallo, Bulgaria, Romania, Cipro, Lettonia, Lituania Polonia, Croazia, Estonia, Ungheria, Slovacchia
C3	Irlanda
C4	Danimarca, Austria, Paesi Bassi, Italia, Svezia Belgio, Germania, Finlandia, Francia

Table 16: Partizioni ottenute con il Metodo del Centroide - DPM

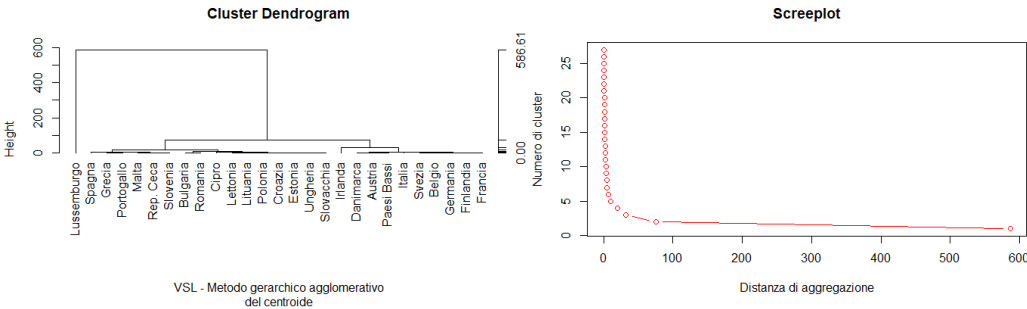


Figure 113: VSL - Metodo Centroide - Den-drogramma

Figure 114: VSL - Metodo Centroide - Screeplot

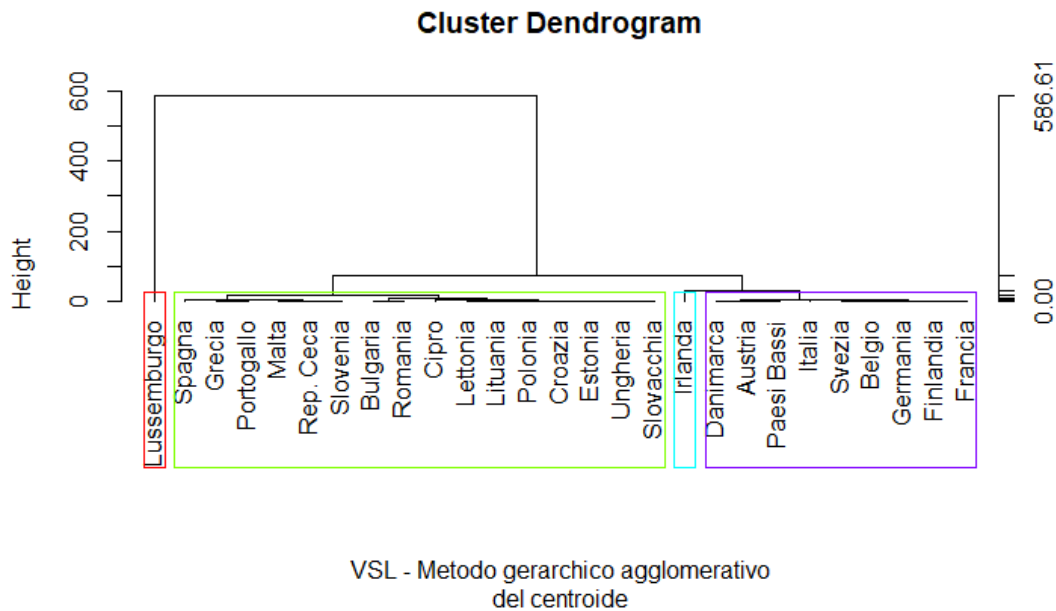


Figure 115: VSL - Metodo Centroide - 4 Cluster

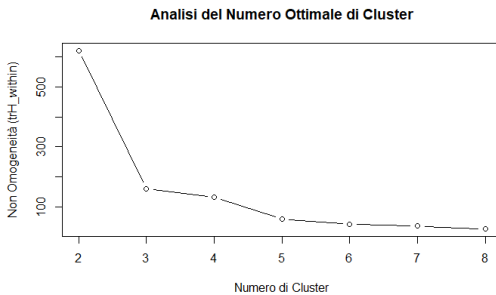


Figure 116: VSL - Metodo Centroide - trH  
within

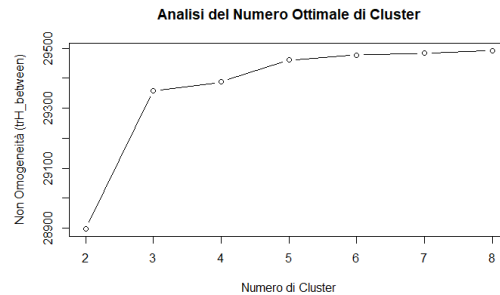


Figure 117: VSL - Metodo Centroide - trH  
between

### 5.3.5 Metodo della Mediana

Il Metodo della Mediana ha fornito una nuova prospettiva nell'analisi dei VSL, enfatizzando un approccio centrato sul valore mediano per definire i centri dei cluster, anche questo metodo può dare origine a catene tra gli individui. Attraverso l'analisi dello Screeplot (vedi Figura 119) e delle Matrici di Non Omogeneità, sia **within** che **between** (vedi Figure 121, 122), è stata identificata una configurazione ottimale di **4 cluster** (vedi Figura 120). Interessante notare come, anche con questo metodo, la configurazione ottimale individuata sia identica ai Metodi del Legame Singolo, Medio e Centroide.

In conclusione, l'applicazione del Metodo della Mediana nell'analisi del VSL ha confermato la tendenza a formare una configurazione ottimale di 4 cluster. Questi risultati enfatizzano, ulteriormente, l'importanza di un'analisi approfondita e diversificata per ogni gruppo di paesi.

Partizione	Paesi
C1	Lussemburgo
C2	Bulgaria, Romania, Malta, Rep. Ceca, Slovenia, Spagna, Grecia, Portogallo, Cipro, Lettonia, Lituania, Polonia, Croazia, Estonia, Ungheria, Slovacchia
C3	Irlanda
C4	Danimarca, Austria, Paesi Bassi, Italia, Svezia, Belgio, Germania, Finlandia, Francia

Table 17: Partizioni ottenute con il Metodo della Mediana - DPM

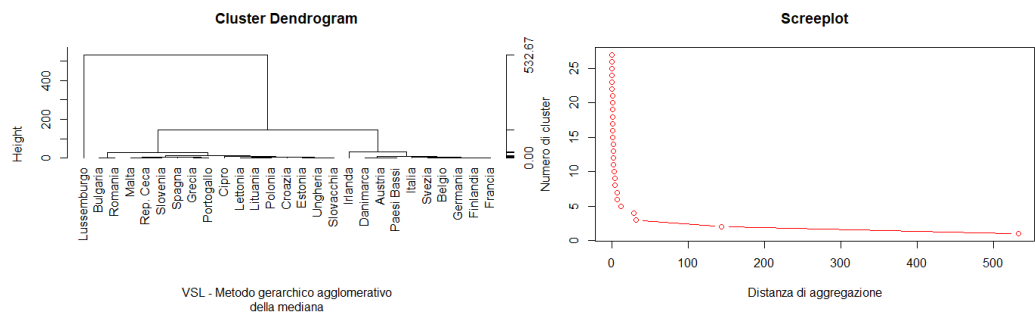


Figure 118: VSL - Metodo Mediana - Dendo-gramma      Figure 119: VSL - Metodo Mediana - Screeplot

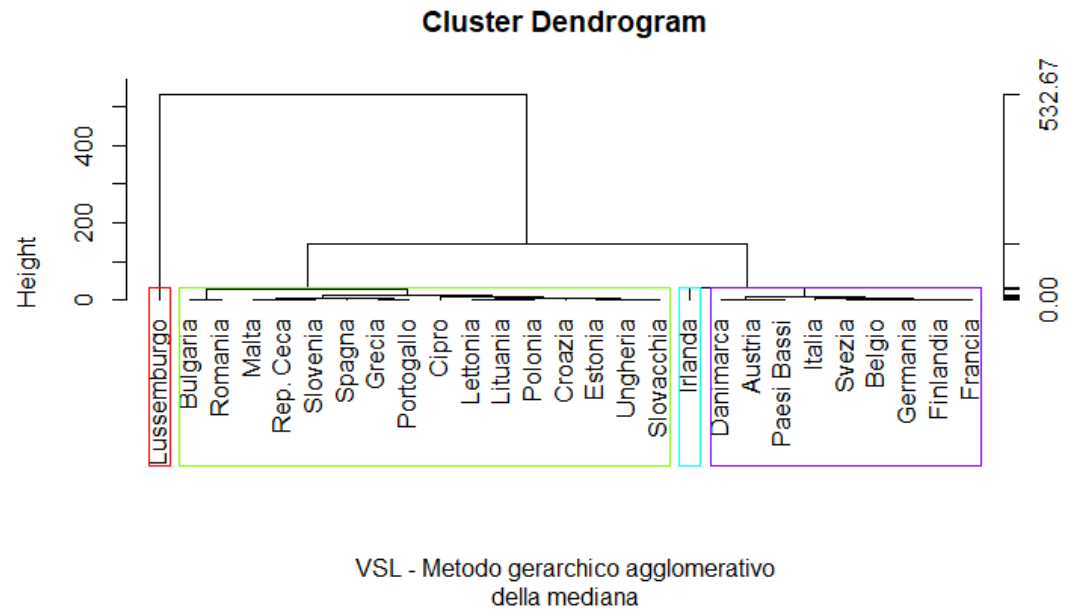


Figure 120: VSL - Metodo Mediana - 4 Cluster

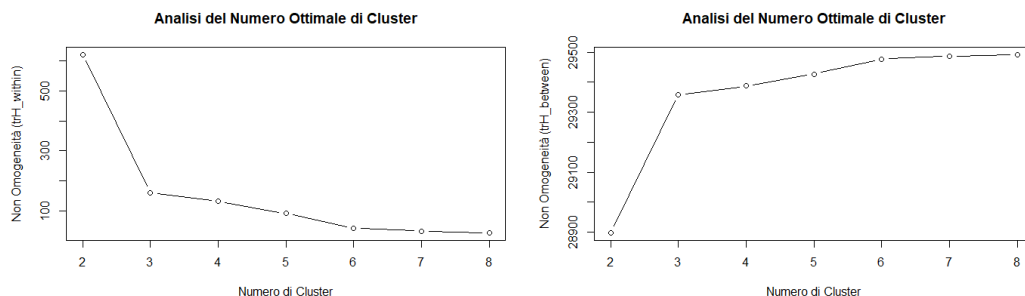


Figure 121: VSL - Metodo Mediana - trH within Figure 122: VSL - Metodo Mediana - trH between

## 5.4 VSL - Metodi Non Gerarchici

### 5.4.1 Metodo kmeans

Basandoci sui risultati ottenuti dalle analisi precedenti, abbiamo condotto l'analisi con il **Metodo del kmeans**, noto per la sua efficienza e la capacità di minimizzare la varianza interna ai cluster, sperimentando con un numero di *cluster variabile da 4 a 6*. Questo processo di valutazione ha avuto l'obiettivo di determinare la configurazione ottimale che potesse al meglio rappresentare le dinamiche dei dati analizzati. Tra le configurazioni testate, la suddivisione in **4 cluster** si è confermata come la scelta ottimale (vedi Figure 123), per la sua capacità di mantenere una coerenza con le suddivisioni precedentemente identificate e per esplorare in modo efficace come questa metodologia potesse differenziare o confermare le aggregazioni dei paesi.

In conclusione, la scelta di procedere con 4 cluster, dopo aver valutato configurazioni da 3 a 5, si è rivelata ideale anche con il Metodo del kmeans. Questa configurazione ottimale non solo conferma le analisi precedenti ma fornisce anche una base solida per interpretazioni più approfondite e per l'elaborazione di politiche sanitarie mirate.

Partizione	Paesi
C1 (3)	Lussemburgo
C2 (2)	Cipro, Grecia, Malta, Portogallo, Rep. Ceca, Slovenia, Spagna
C3 (1)	Austria, Belgio, Danimarca, Finlandia, Francia, Germania, Irlanda, Italia, Paesi Bassi, Svezia
C4 (4)	Bulgaria, Croazia, Estonia, Ungheria, Lettonia, Lituania, Polonia, Romania, Slovacchia

Table 18: Partizioni ottenute con il Metodo kmeans - DPM



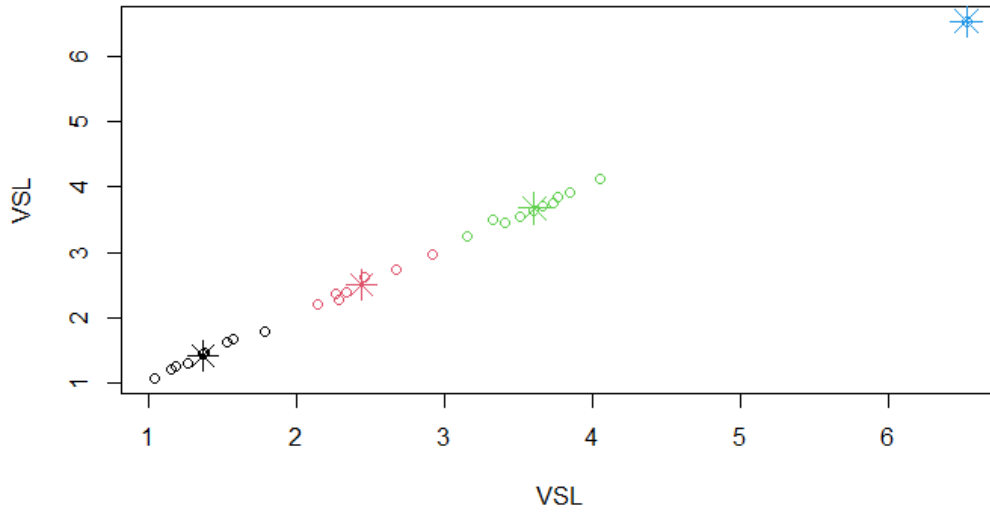


Figure 123: VSL - Metodo kmeans - 4 Cluster

## 5.5 Conclusioni

In conclusione, è stata evidenziata, dalla coerenza dei cluster identificati, il beneficio di combinare diversi approcci di clustering, sia gerarchici che non gerarchici, al fine di ottenere una comprensione olistica e dettagliata sugli indicatori in esame. Dopo aver attentamente analizzato i risultati ottenuti applicando i vari metodi, si è giunti alla conclusione che il numero ideale di per **DPM è 5** e per **VSL è 4**. Inoltre, l'identificazione di cluster specifici per paesi con performance sia ottimali sia meno favorevoli, nel contesto di DPM e VSL, sottolinea l'importanza di strategie di intervento mirate e basate su evidenze concrete.

## 6 Inferenza Statistica

In questo capitolo, si vuole applicare l'**Inferenza Statistica** per ottenere informazioni sui parametri non noti della popolazione, analizzando un campione estratto dalla stessa. Per applicare questi concetti, è stato prelevato un campione dal dataset contenente gli indici di DPM, in particolare, è stato selezionato il paese *'Italia'*, per i quali sono stati estratti i dati corrispondenti a tutti gli anni disponibili, *fino al 2019*.

### 6.1 Stima Puntuale

La **stima puntuale** consente di stimare il valore di un parametro non noto della popolazione, a partire da un campione osservato. Inizialmente, occorre determinare il tipo di popolazione in maniera tale da poter applicare il procedimento corretto in relazione alla tipologia di *distribuzione*. Per ottenere questa informazione, è stato utilizzato il *criterio del chi quadrato*. In questo contesto, possiamo utilizzare questo criterio per determinare se i dati di DPM, per l'Italia, seguono una *distribuzione normale o un'altra distribuzione specifica*.

#### 6.1.1 Criterio del chi2

Utilizzando il **criterio del chi-quadrato (chi2)** sul campione estratto dal dataset DPM, si vuole inizialmente verificare (o rifiutare) l'ipotesi che la distribuzione sia di tipo **normale**. Le ipotesi di partenza sono così definite:

- **Popolazione di tipo Normale:**  $H_0$
- **Popolazione di tipo Non Normale:**  $H_1$

Il campione preso in esame presenta i seguenti indici statistici:

- **Media campionaria:** 150.43
- **Deviazione Standard Campionaria:** 15.26
- **Dimensione del Campione:** 25

Applicando il test chi-quadrato di misura  $\alpha = 0.05$ , si desidera verificare se la popolazione da cui proviene il campione può essere descritta da una variabile aleatoria  $X$  di *densità normale*. Suddividendo l'insieme dei possibili valori che possono essere assunti da tale variabile aleatoria  $X$  in 5 sottoinsiemi denominati  $I_1, \dots, I_5$ , le probabilità che  $X$  assuma un valore appartenente a  $I_i$  (con  $i = 1, \dots, 5$ ) saranno  $p_i = 0.2$ . In tale maniera, è soddisfatta la condizione:  $\min(np_1, \dots, np_r) \geq 5$ , essendo  $np_i = 25 \cdot 0.2 = 5 \geq 5$ .

Dunque gli intervalli sono definiti come segue:

Intervallo	Lim. Sinistro	Lim. Destro	Elementi
I1	$-\infty$	137.5863	6
I2	137.5863	146.5660	8
I3	146.5660	154.3005	2
I4	154.3005	163.2802	3
I5	163.2802	$+\infty$	6
			25

Table 19: Intervalli  $I_i$

La colonna *Elementi* (vedi Tabella 19) indica quanti elementi ricadono nel singolo intervallo: la somma totale equivale ovviamente a 25, ovvero il numero di elementi che compongono il campione considerato. Successivamente, avendo definito gli intervalli, è possibile effettuare il **calcolo di chi2**: il risultato ottenuto è  $\chi^2 = 4.8$

Siccome la distribuzione normale ha, per definizione, due parametri non noti quali  $\mu$  e  $\sigma^2$ , si è posto  $k = 2$  e  $\alpha = 0.05$ , e definito il numero di gradi di libertà ottenuto da  $r - k - 1 = 2$ , si ha:

- $\chi_{\alpha/2, r-k-1}^2 = 7.38$
- $\chi_{1-\alpha/2, r-k-1}^2 = 0.05$

Essendo  $0.05 < \chi^2 < 7.38$ , è possibile affermare che l'ipotesi  $H_0$  per cui la popolazione è di tipo normale può essere accettata. Questa verifica ci permette di poter applicare il corretto procedimento per l'applicazione delle stime puntuali, considerando la popolazione analizzata come **popolazione normale**.

### 6.1.2 Metodo dei Momenti

Attraverso il **Metodo dei Momenti**, applicato su una popolazione normale in base a quanto appurato nella sezione precedente, è possibile determinare gli **stimatori** dei parametri  $\mu$  e  $\sigma^2$ . Tale metodo definisce come stimatore del valore medio  $\mu$  la media campionaria, mentre come stimatore della varianza  $\sigma^2$  la variabile aleatoria  $\frac{(n-1) \cdot S^2}{n}$ , da cui si ricavano i seguenti risultati:

Stimatore	Valore
$\mu$	150.43
$\sigma^2$	223.69

Table 20: Stimatori secondo il Metodo dei Momenti

### 6.1.3 Conclusione

In conclusione, è stato appurata la densità normale della popolazione in esame e, successivamente, **stimati i parametri non noti**  $\mu = 150.43$  e  $\sigma^2 = 223.69$ , che indicano rispettivamente valore medio e varianza della popolazione.

## 6.2 Intervalli di Confidenza

Avendo determinato nella precedente fase un singolo valore per stimare i parametri non noti della popolazione, si intende adesso estendere questa stima sostituendo i valori con gli **intervalli di fiducia**, entro i quali i parametri non noti dovranno essere compresi con un determinato **coefficiente di confidenza**. Posto nuovamente  $\alpha = 0.05$ , si seleziona un grado di fiducia pari a  $1 - \alpha = 0.95$ , valore che definisce un *ottimo compromesso tra precisione e affidabilità delle stime*.

### 6.2.1 Metodo Pivotal

Attraverso il **Metodo Pivotal** si intende determinare gli intervalli di confidenza per la popolazione analizzata, che è stato già determinato essere di tipo normale.

In particolare, si intende coprire tutte le possibili casistiche, quali:

- (i) Determinazione degli intervalli di fiducia di grado  $1 - \alpha$  per  $\mu$  in caso di varianza  $\sigma^2$  della popolazione normale nota.
- (ii) Determinazione degli intervalli di fiducia di grado  $1 - \alpha$  per  $\mu$  in caso di varianza  $\sigma^2$  della popolazione normale non nota.
- (iii) Determinazione degli intervalli di fiducia di grado  $1 - \alpha$  per  $\sigma^2$  in caso di valore medio  $\mu$  della popolazione normale noto.
- (iv) Determinazione degli intervalli di fiducia di grado  $1 - \alpha$  per  $\sigma^2$  in caso di valore medio  $\mu$  della popolazione normale non noto.

#### (i) Intervalli di fiducia di grado $1 - \alpha$ per $\mu$ con $\sigma^2$ nota

In questa sezione l'obiettivo è definire l'intervallo di confidenza di grado  $1 - \alpha$  per il valore medio,  $\mu$ , nel caso in cui la varianza  $\sigma^2$  sia nota. La definizione degli intervalli in questo caso è definita come:

- $\underline{C}_n = \overline{X}_n - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
- $\overline{C}_n = \overline{X}_n + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

che rappresentano rispettivamente il limite inferiore e quello superiore dell'intervallo che si intende definire.

Siccome è stato scelto un  $\alpha = 0.05$ , allora  $\frac{\alpha}{2} = 0.025$ . Attraverso il comando `qnorm(1 -  $\alpha/2$ )` si ottiene il valore di  $z_{\alpha/2} = 1.96$ . Sostituendo tale valore nelle formule appena viste, e avendo nota sia la media campionaria, che la varianza  $\sigma$ , che il numero di elementi totali, si ottiene il seguente intervallo:

- $\underline{C}_n = 144.45$
- $\overline{C}_n = 156.42$

con un *coefficiente di confidenza pari a 0.95*.

Si denota dunque che la stima dell'intervallo di confidenza di grado  $1 - \alpha = 0.95$  per il numero di decessi in Italia è  $(144.45, 156.42)$ . Inoltre, si osserva che la media del campione, 150.43, è compresa in questo intervallo.

**(ii) Intervalli di fiducia di grado  $1 - \alpha$  per  $\mu$  con  $\sigma^2$  non nota**

In questa sezione l'obiettivo è definire l'intervallo di confidenza di grado  $1 - \alpha$  per il valore medio,  $\mu$ , nel caso in cui la varianza  $\sigma^2$  non sia nota.

La definizione degli intervalli in questo caso è definita come:

- $\underline{C}_n = \overline{X}_n - t_{\alpha/2, n-1} \cdot \frac{S_n}{\sqrt{n}}$
- $\overline{C}_n = \overline{X}_n + t_{\alpha/2, n-1} \cdot \frac{S_n}{\sqrt{n}}$

i quali rappresentano rispettivamente il limite inferiore e quello superiore dell'intervallo che si intende definire. Siccome è stato posto  $\alpha = 0.05$ , allora  $\frac{\alpha}{2} = 0.025$ . Attraverso il comando `qt(1 -  $\alpha/2$ ,  $n - 1$ )`, si ottiene il valore di  $t_{\alpha/2, n-1} = 2.064$ . Sostituendo tale valore nelle formule appena viste, e avendo noti media campionaria, deviazione standard campionaria  $S_n$  e il numero di elementi  $n$ , si ottiene il seguente intervallo:

- $\underline{C}_n = 144.13$
- $\overline{C}_n = 156.73$

con un *coefficiente di confidenza pari a 0.95*.

Si denota dunque che la stima dell'intervallo di confidenza di grado  $1 - \alpha = 0.95$  per il numero di decessi in Italia è  $(144.13, 156.73)$  e che, anche in questo caso, la media del campione, 150.43, è compresa in questo intervallo.

**(iii) Intervalli di fiducia di grado  $1 - \alpha$  per  $\sigma^2$  con  $\mu$  noto**

In questa sezione l'obiettivo è definire l'intervallo di confidenza di grado  $1 - \alpha$  per la varianza,  $\sigma^2$ , nel caso in cui il valore medio  $\mu$  sia noto.

La definizione degli intervalli in questo caso è definita come:

- $\underline{C}_n = \frac{(n-1) \cdot S_n^2 + n(\overline{X}_n - \mu)^2}{\chi_{\alpha/2, n}^2}$
- $\overline{C}_n = \frac{(n-1) \cdot S_n^2 + n(\overline{X}_n - \mu)^2}{\chi_{1-\alpha/2, n}^2}$

dove  $\underline{C}_n$  e  $\overline{C}_n$  rappresentano rispettivamente il limite inferiore e quello superiore dell'intervallo che si intende definire. Siccome è stato scelto un  $\alpha = 0.05$ , allora  $\frac{\alpha}{2} = 0.025$ . Attraverso i comandi `qchisq( $\alpha/2$ ,  $df = n$ )` e `qchisq( $1 - \alpha/2$ ,  $df = n$ )`, si ottengono rispettivamente i valori  $\chi_{\alpha/2, n}^2$  e  $\chi_{1-\alpha/2, n}^2$ . Sostituendo tali valore nelle formule sopracitate, e avendo noto il valore medio  $\mu$ , la varianza campionaria  $\sigma^2$  e il numero di elementi  $n$ , si ottiene il seguente intervallo:

- $\underline{C}_n = 137.78$
- $\overline{C}_n = 426.85$

con un *coefficiente di confidenza pari a 0.95*.

Si denota dunque che la stima dell'intervallo di confidenza di grado  $1 - \alpha = 0.95$  per il numero di decessi in Italia è  $(137.78, 426.85)$ . Si osserva che anche in questo caso la media del campione, 150.43, è compresa in questo intervallo.

**(iv) Intervalli di fiducia di grado  $1 - \alpha$  per  $\sigma^2$  con  $\mu$  non noto**

In questa sezione l'obiettivo è definire l'intervallo di confidenza di grado  $1 - \alpha$  per la varianza,  $\sigma^2$ , nel caso in cui il valore medio  $\mu$  sia non noto.

La definizione degli intervalli in questo caso è definita come:

- $\underline{C}_n = \frac{(n-1)*S_n^2}{\chi_{\alpha/2, n-1}^2}$
- $\overline{C}_n = \frac{(n-1)*S_n^2}{\chi_{1-\alpha/2, n-1}^2}$

che rappresentano rispettivamente il limite inferiore e quello superiore dell'intervallo che si intende definire. Siccome è stato scelto  $\alpha = 0.05$ , allora  $\frac{\alpha}{2} = 0.025$ . Attraverso i comandi `qchisq( $\alpha/2$ ,  $df = n-1$ )` e `qchisq( $1-\alpha/2$ ,  $df = n-1$ )`, si ottengono i valori rispettivamente di  $\chi_{\alpha/2, n-1}^2$  e  $\chi_{1-\alpha/2, n-1}^2$ . Sostituendo tali valore nelle formule sopracitate, e avendo nota sia la media campionaria, che la varianza campionaria  $\sigma^2$ , che il numero di elementi  $n$ , si ottiene il seguente intervallo:

- $\underline{C}_n = 142.06$
- $\overline{C}_n = 450.94$

con un *coefficiente di confidenza pari a 0.95*.

Si denota dunque che la stima dell'intervallo di confidenza di grado  $1 - \alpha = 0.95$  per il numero di decessi in Italia è  $(142.06, 450.94)$ . Si osserva che anche in questo caso la media del campione, 150.43, è compresa nell'intervallo.

## 6.2.2 Conclusioni

In conclusione, attraverso il metodo Pivotal e coprendo tutte le possibili situazioni, è stato possibile determinare gli intervalli di fiducia per entrambi i parametri, con un coefficiente di fiducia pari a 0.95, ed è risultato che la **media campionaria è sempre compresa negli intervalli**.

## 6.3 Confronti tra Popolazioni

Nella fase del **confronto tra due popolazioni**, l'obiettivo è quello di effettuare un confronto tra la popolazione appena analizzata, contenente i valori di DPM in Italia e un campione contenente i valori di DPM in Italia *dal 2005 in poi*. La scelta di effettuare questo confronto nasce dal fatto che nel 2005, appunto in Italia, viene emanata un'importante legge antifumo, la **Legge Sirchia** (vedi Sezione 7.2), che ha imposto il divieto di fumo nei locali pubblici al chiuso, inclusi uffici, ristoranti e bar. Di conseguenza, effettuare un confronto tra il numero generale di decessi e quello relativo al periodo storico successivo al 2005 può portare a osservazioni importanti e interessanti sull'impatto di leggi più severe rispetto all'indicatore DPM.

Per far ciò, è innanzitutto necessario verificare che anche questa seconda popolazione, denominata  $Y$ , sia normale. È quindi stato applicato, anche su questa popolazione, il criterio del **chi-quadrato**. Utilizzando tale criterio, sul campione contenente i valori di DPM in Italia dal 2005 in poi, si vuole verificare (o rifiutare) l'ipotesi che la distribuzione sia di tipo **normale**. Le ipotesi di partenza sono così definite:

- **Popolazione Y di tipo Normale:**  $H_0$
- **Popolazione Y di tipo Non Normale:**  $H_1$

Il campione preso in esame presenta i seguenti indici statistici:

- **Media campionaria:** 138.94
- **Deviazione Standard Campionaria:** 4.60
- **Dimensione del Campione:** 14

Applicando il test chi2 di misura  $\alpha = 0.05$ , in maniera del tutto analoga al test effettuato con la popolazione 'Italia' completa, si osserva che l'ipotesi  $H_0$  per cui la *popolazione è di tipo normale* può essere accettata. Appurato che entrambe le popolazioni X e Y prese in esame presentano una distribuzione normale, è possibile effettuarne correttamente il confronto.

### 6.3.1 Intervallo di confidenza per $\mu_1 - \mu_2$ con $\sigma_1^2$ e $\sigma_2^2$ note

Determinare gli intervalli di fiducia per  $\mu_1 - \mu_2$  in caso di varianze note, significa trovare il limite inferiore e superiore dell'intervallo, dove questi limiti sono per definizione così definiti:

$$\overline{x}_{n_1} - \overline{y}_{n_2} - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \overline{x}_{n_1} - \overline{y}_{n_2} + z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

dove  $\overline{x}_{n_1}$  e  $\overline{y}_{n_2}$  corrispondono alle medie campionarie dei due campioni da confrontare, X e Y. Applicando, in R, la funzione `qnorm()` sul valore  $1 - \alpha/2$  si ottiene il valore associato a  $z_{\alpha/2} = z_{0.025} = 1.96$ . Utilizzando tale valore all'interno delle formule appena descritte, è determinata la differenza tra il numero medio di morti nelle due differenti popolazioni X e Y, ovvero  $\mu_1$  e  $\mu_2$ . Questi limiti sono:

- $\overline{x}_{n_1} - \overline{y}_{n_2} - z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 0.085$
- $\overline{x}_{n_1} - \overline{y}_{n_2} + z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 19.63$

Si osserva che sia il limite superiore che quello inferiore sono positivi, il che permette di dedurre che il numero di decessi nell'arco temporale completo è mediamente maggiore rispetto a quelle registrate dal 2005 in poi, anno in cui si registra l'entrata in vigore della legge Sirchia (vedi Sezione 7.2).

### 6.3.2 Conclusioni

In conclusione, sono stati riscontrati limiti positivi per gli intervalli, che posso essere tradotti in una diminuzione del numero di decessi causati dal fumo passivo a partire dal 2005 rispetto alla popolazione generale. Questo potrebbe essere indice di effetti positivi dell'applicazione della legge Sirchia, pur non essendo probabilmente l'unico fattore incidente in questo calo di valori, confermando la molteplicità di fattori che incidono sull'indicatore DPM.

## 6.4 Verifica delle Ipotesi

Nella fase finale di **verifica delle ipotesi**, si vuole verificare che il valore medio di DPM *dal 2005 in poi* sia inferiore a 167 circa, valore della media in Italia di DPM nel periodo precedente al 2005, per *verificare l'impatto positivo* dell'entrata in vigore della rigida Legge Sirchia (vedi Sezione 7.2).

Le ipotesi sono così definite:

- **Ipotesi Nulla:**  $H_0 : \mu \geq \mu_0$ , il numero medio di decessi per milione di abitanti dal 2005 in poi è maggiore o uguale a 167, quindi la Legge non ha inciso positivamente.
- **Ipotesi Alternativa:**  $H_1 : \mu < \mu_0$ , il numero medio di decessi per milione di abitanti dal 2005 in poi è inferiore a 167, ovvero la Legge ha inciso positivamente.

Per effettuare il test di queste ipotesi composite è stato utilizzato il **test unilaterale destro**  $\psi$ , il quale prevede (vedi Figure 124):

- si accetti  $H_0$  se:  $\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > -z_\alpha$
- si rifiuti  $H_0$  se:  $\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} < -z_\alpha$

### Densità normale standard

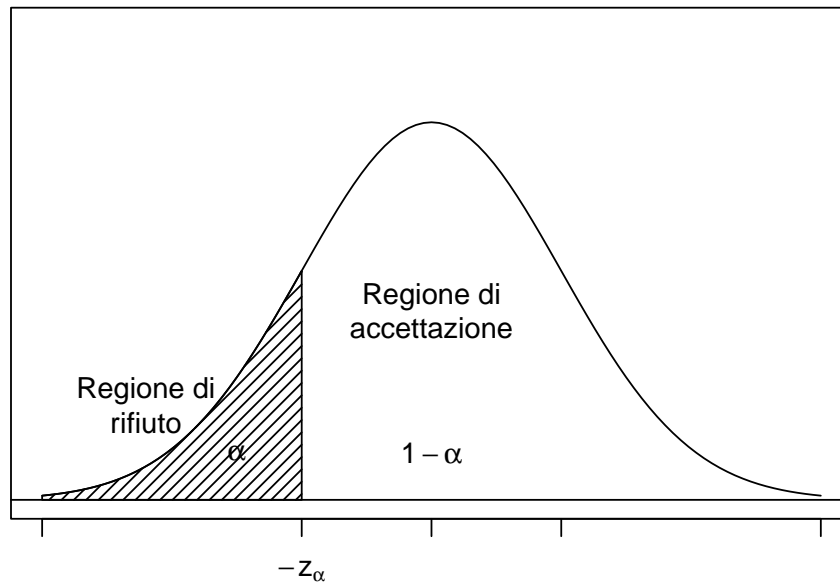


Figure 124: Densità normale standard e regioni di accettazione e di rifiuto dell'ipotesi nulla per il test unilaterale destro

Si pone, nuovamente,  $\alpha = 0.05$  e, quindi, secondo quanto definito  $\mu_0 = 167$  e  $\sigma = 10$ . Di seguito, risulta che  $-z_\alpha = -1.65$  e  $z = -10.50$ . Risulta che:  $z < -z_\alpha$ , quindi la statistica del test osservata cade nella regione di rifiuto dell'ipotesi nulla, indicando che i dati forniscono prove sufficienti per **rifiutare  $H_0$  a favore di  $H_1$** .

Per avere una conferma del risultato ottenuto, è possibile verificare anche con il **test del pvalue**, che per il test unilaterale destro è definito come segue:  $pvalue = P(Z_n \leq z)$ , dove  $z$  è la *stima della statistica del test*. Risulta che  $pvalue = 4.42 \cdot 10^{-26}$ . Questo valore estremamente basso conferma ulteriormente il **rifiuto dell'ipotesi nulla**, poiché è molto inferiore al livello di significatività  $\alpha = 0.05$  ( $pvalue < \alpha$ ).



#### 6.4.1 Conclusioni

In conclusione, sia l'analisi della statistica del test  $z$  che il calcolo del *pvalue* portano alla stessa decisione: **rifiutare l'ipotesi nulla a favore dell'ipotesi alternativa**. Quindi, basandosi sui dati analizzati, esiste una significativa evidenza statistica che il valore di **DPM è diminuito dopo l'introduzione della legge antifumo nel 2005**, scendendo al di sotto del valore medio di 167 decessi osservato negli anni precedenti al 2005. Questo risultato dimostra non solo l'efficacia delle politiche di salute pubblica e dell'aumento dell'indicatore VSL, per cui l'Italia non presenta un incremento notevole negli anni come mostrato precedentemente, ma anche l'**importanza dell'adozione di leggi specifiche** per combattere fenomeni che hanno un impatto negativo sulla società. Pertanto, oltre all'incremento del VSL, risulta essere molto utile combattere questi fenomeni attraverso l'implementazione di leggi mirate, confermando la molteplicità di fattori che incidono sull'indicatore DPM e l'essenziale ruolo che la legislazione gioca nella promozione del benessere collettivo e nella riduzione dei rischi per la salute pubblica.

## 7 Avvenimenti storici

### 7.1 Crisi Finanziaria 2007-2008

La **crisi finanziaria del 2007-2008**, parte della Grande Recessione, fu scatenata dalla deflazione delle bolle dei prezzi, inclusa la bolla immobiliare americana, e dalle perdite significative delle istituzioni finanziarie. Questa crisi portò a una *crisi di liquidità e solvibilità per banche e stati*, una scarsità di credito per le imprese, il crollo dei prezzi di borsa, e il fallimento di numerosi istituti finanziari. Gli interventi statali furono necessari per prevenire una crisi sistemica, causando crisi del debito pubblico e una recessione globale, con un calo del PIL globale del 2,2% nel 2009.

### 7.2 Legge Sirchia (Italia)

La **Legge Sirchia n. 3 (art. 51)** del 2003, o 'Tutela della Salute dei non fumatori', entrata in vigore il 10 gennaio 2005, ha previsto il divieto di fumo luoghi pubblici chiusi, inclusi uffici, ristoranti e bar, prevedendo specifiche aree riservate ai fumatori solo se dotate di adeguati sistemi di ventilazione. L'obiettivo era quello di proteggere la salute dei non fumatori, *riducendo la loro esposizione al fumo passivo*.

## 8 Conclusioni

Lo studio offre una panoramica completa sull'*impatto delle politiche di salute pubblica e delle condizioni economiche sulla mortalità per fumo passivo*, con un focus specifico sui paesi dell'Unione Europea. Esso ha messo in luce l'importanza delle **strategie personalizzate e mirate** per combattere efficacemente la mortalità per fumo passivo, enfatizzando la necessità di considerare le **specificità nazionali**. La **correlazione fortemente negativa tra DPM e VSL** indica che gli investimenti in salute e sicurezza, e, in generale, condizioni economiche in miglioramento hanno contribuito a una riduzione significativa dei tassi di mortalità. La **tendenza positiva**, osservata nella quasi totalità dei paesi analizzati, conferma l'efficacia degli interventi e delle politiche implementate, con un'attenzione particolare *ai paesi dell'Europa dell'Est*, che hanno dimostrato un impegno considerevole nel migliorare le proprie situazioni a partire da basi critiche. Inoltre, l'analisi ha evidenziato che eventi esterni, come la *crisi economica e finanziaria del 2007-2008* (vedi Sezione 7.1), hanno avuto un **impatto negativo** rilevante sugli indicatori di salute pubblica, sottolineando come i fattori macroeconomici possano influenzare direttamente la qualità della vita e la salute. In particolare, per quanto riguarda l'Italia, l'entrata in vigore di una *rigida legge antifumo nel 2005* (vedi Sezione 7.2) si è rivelata un elemento significativo nella riduzione dei decessi dovuti al fumo passivo, confermando l'efficacia di politiche precise e mirate nel contrastare fenomeni nocivi per la salute pubblica. Questo evidenzia il ruolo cruciale di un **approccio legislativo nella salute pubblica**, dimostrando che l'introduzione di leggi specifiche può avere un impatto diretto e quantificabile sulla società.

*In conclusione, il progetto evidenzia come l'integrazione di interventi mirati, investimenti in salute e sicurezza, e l'implementazione di politiche legislative specifiche possano determinare miglioramenti significativi nella salute pubblica e nella qualità della vita.*