

ALMA MATER STUDIORUM – UNIVERSITA' DI BOLOGNA

DIPARTIMENTO DI SCIENZE STATISTICHE

“PAOLO FORTUNATI”

Corso di Laurea in Scienze Statistiche

DETERMINANTI DELLA SPERANZA DI VITA: UN'ANALISI  
GLOBALE DEGLI EFFETTI DI INQUINAMENTO, ABITUDINI DI  
CONSUMO E CONDIZIONI SANITARIE

Utilizzo Statistico di Banche Dati Online

Lucrezia Galli

2025-04-06

Presentata da:  
Lucrezia Galli  
Matricola: 0001113963

Relatore:  
Prof Paolo Verme

APPELLO I

ANNO ACCADEMICO 24 / 25

# 1 Introduction

Analizzare i fattori che influenzano la speranza di vita è essenziale per sviluppare strategie efficaci di sanità pubblica e promuovere il benessere a livello globale. Numerosi studi hanno evidenziato come variabili socioeconomiche, accesso alle cure mediche e abitudini di vita influenzino significativamente la longevità. Tuttavia, l’approccio integrato che considera la complessa interazione di questi fattori è ancora poco esplorato (Esposito et al. 2024). Ricerche recenti hanno mostrato che percorsi socioeconomici svantaggiati possano accelerare l’invecchiamento biologico, ma che interventi mirati sugli stili di vita possano mitigarne gli effetti negativi. Questo studio esplora l’impatto dell’inquinamento atmosferico, del consumo di alcol e tabacco, della mortalità materna, degli incidenti stradali e delle malattie croniche sulla speranza di vita, fornendo un’analisi completa e approfondita.

# 2 Data

Il dataset utilizzato per l’analisi non è stato scaricato direttamente da una singola fonte, ma costruito a partire da diversi dataset disponibili nei database della [World Health Organization \(WHO\)](#),selezionando i dati del 2019 per coerenza temporale. Successivamente, le variabili di interesse sono state aggregate in un unico file Excel e poi importate in R per l’analisi statistica.

Table 1: Descrizione delle variabili chiave del dataset

Variabile	Descrizione	Unita_di_misura
LifeExpatBirth	Speranza di vita alla nascita	Anni
Mortality_30_70_CCDR	Probabilità di morire tra i 30 e i 70 anni a causa di malattie croniche	Tasso (%)
PM25_Annual_Urban	Concentrazione media annua di PM2.5 nelle aree urbane	µg/m <sup>3</sup>
Maternal_Mortality_Ratio	Decessi materni per complicazioni legate alla gravidanza	Decessi per 100.000 nati vivi
Alcohol_Consumption	Consumo medio di alcol della popolazione (>15 anni)	Litri pro capite
Tobacco_Age_Stand	Percentuale di consumatori di tabacco	Percentuale (%)
Pop_Basic_Drinking_Water	Accesso ai servizi idrici di base	Percentuale (%)
Road_traffic_injuries	Mortalità dovuta a incidenti stradali	Tasso per 100.000 abitanti

### 3 Methodology

Per esaminare i fattori che influenzano l'aspettativa di vita alla nascita, è stato adottato un modello di regressione lineare multipla, che stima l'effetto di più variabili esplicative, mantenendo costanti gli altri fattori. Il modello di regressione è il seguente:

$$LifeExp = \beta_0 + \beta_1 \cdot Mort3070 + \beta_2 \cdot PM25 + \beta_3 \cdot MatMort + \beta_4 \cdot AlcCons + \beta_5 \cdot DrinkWater + \beta_6 \cdot RoadInj + \beta_7 \cdot TobUse + \epsilon$$

dove LifeExpBirth rappresenta l'aspettativa di vita alla nascita e le altre variabili sono i fattori esplicativi. Il termine  $\epsilon$  indica l'errore residuo del modello. Il codice utilizzato per questa analisi è stato gestito e versionato tramite GitHub.

Il repository è disponibile al seguente link: [GitHub Repository](#).

### 4 Results

#### 4.1 Effetti delle variabili sulla speranza di vita

Table 2: Risultati del Modello Multiplo

Variable	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	76.364	2.090	36.542	0.000
Mortality_30_70_CCDR	-0.512	0.025	-20.287	0.000
PM25_Annual_Urban	-0.027	0.012	-2.169	0.032
Maternal_Mortality_Ratio	-0.007	0.001	-4.865	0.000
Alcohol_Consumption	-0.031	0.048	-0.647	0.519
Pop_Basic_Drinking_Water	0.104	0.018	5.819	0.000
Road_traffic_injuries	-0.124	0.021	-5.827	0.000
Tobacco_Age_Stand	0.054	0.018	2.957	0.004

```
R2      maxvif
0.9197774 3.1201859
```

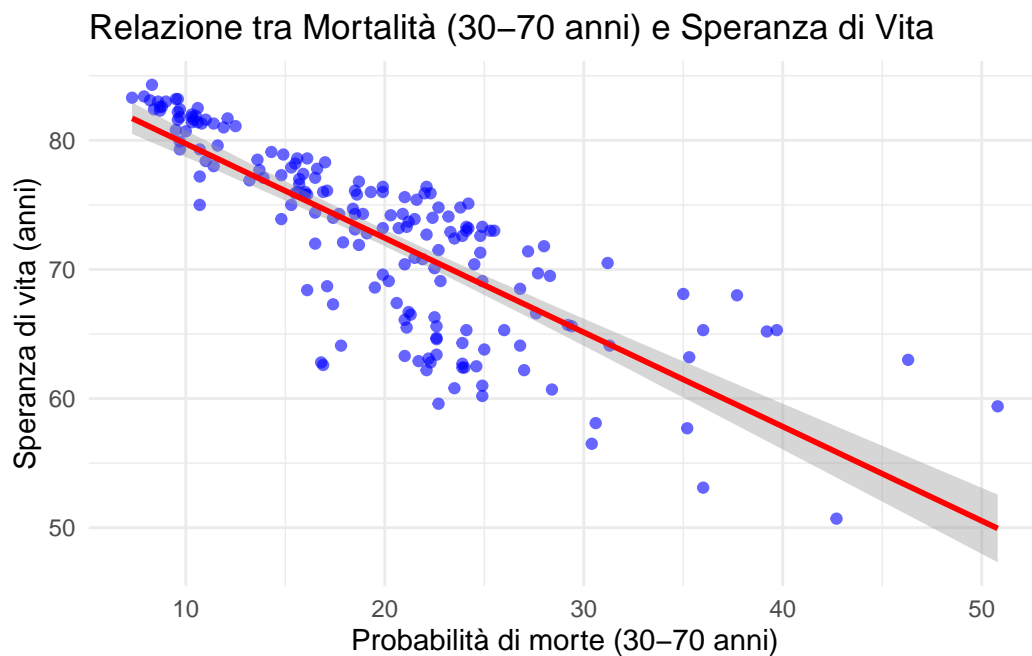
Il modello di regressione ha un  $R^2$  di 0.9198, indicando che il 91.98% della variabilità della speranza di vita è spiegata dalle variabili incluse. La mortalità tra i 30 e i 70 anni (-0.512,  $p < 0.001$ ), la mortalità materna (-0.007,  $p < 0.001$ ), l'inquinamento atmosferico (-0.027,  $p = 0.032$ ) e gli incidenti stradali (-0.124,  $p < 0.001$ ) hanno un impatto negativo significativo. L'accesso ai servizi idrici (+0.104,  $p < 0.001$ ) è associato a un aumento della speranza di vita. Il consumo di tabacco (+0.054,  $p = 0.004$ ) mostra un effetto positivo inaspettato, mentre l'alcol non è significativo (-0.031,  $p = 0.519$ ). Nonostante l' $R^2$  elevato, i VIF (tutti inferiori a 5) confermano l'assenza di multicollinearità e la

stabilità delle stime del modello, evidenziando l'impatto delle condizioni sanitarie e ambientali sulla longevità.

#### 4.2 Relazione tra mortalità prematura e speranza di vita

Nel modello lineare che ho stimato per analizzare i fattori che influenzano la speranza di vita alla nascita, la variabile più significativa è risultata `Mortality_30_70_CDDR`, ovvero la probabilità di morte tra i 30 e i 70 anni per cause croniche. Questo è stato determinato tramite l'analisi dei p-value dei coefficienti del modello. Per esplorare meglio la relazione tra `Mortality_30_70_CDDR` e `LifeExpatBirth`, ho creato un grafico a dispersione con una retta di regressione.

```
`geom_smooth()` using formula = 'y ~ x'
```

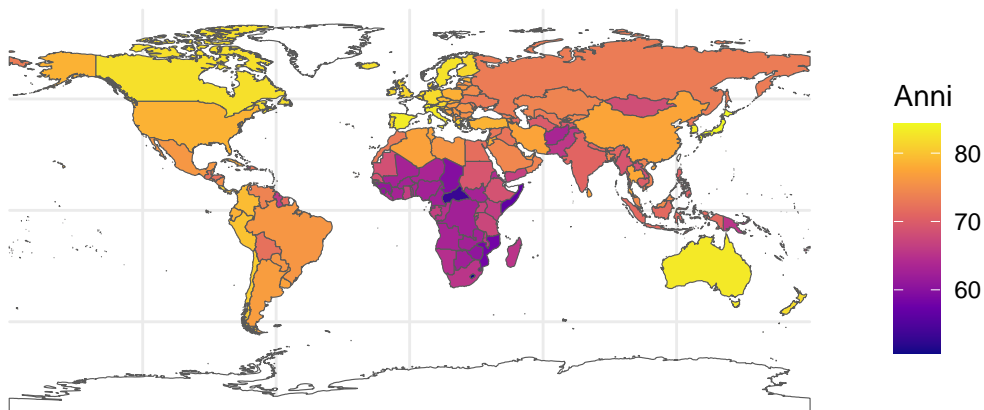


Il grafico mostra chiaramente una relazione negativa: all'aumentare della probabilità di morte tra i 30 e i 70 anni, la speranza di vita alla nascita diminuisce. Questo è coerente con l'aspettativa che una maggiore mortalità in questa fascia d'età riduca la durata media della vita. La retta di regressione (in rosso) conferma questa tendenza negativa, mentre l'area grigia rappresenta l'incertezza nelle stime. La distribuzione dei punti suggerisce una relazione piuttosto forte tra le due variabili, confermando l'importanza di `Mortality_30_70_CDDR` come predittore chiave della speranza di vita.

#### 4.3 Distribuzione della speranza di vita per continente

Per esplorare le differenze nella speranza di vita alla nascita tra i vari continenti, è stata realizzata una mappa tematica. Questa visualizzazione consente di apprezzare chiaramente le disparità tra le diverse aree geografiche, offrendo una rappresentazione immediata dei valori e facilitando il confronto tra le regioni del mondo.

##### Speranza di vita alla nascita nel mondo

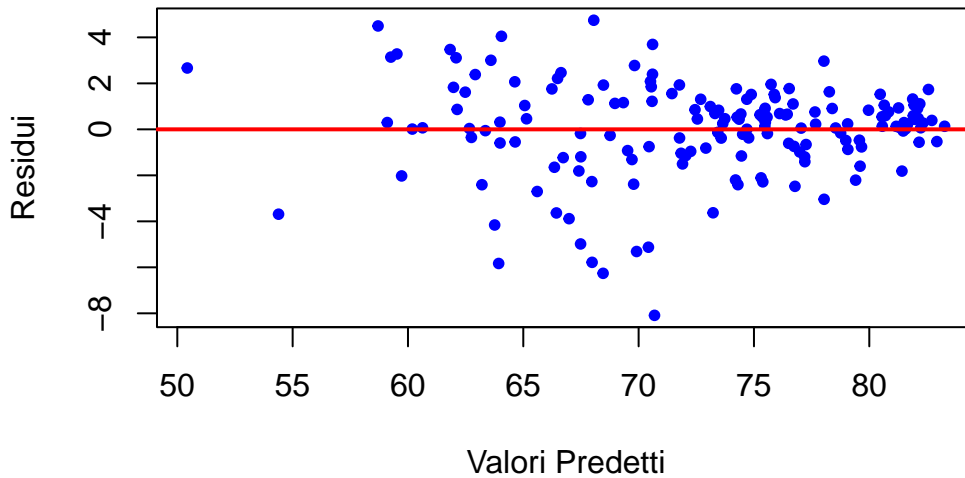


La mappa offre una rappresentazione visiva immediata delle disuguaglianze nella speranza di vita alla nascita tra i continenti. Le aree con valori più bassi, come l’Africa subsahariana, sono evidenziate con tonalità di viola scuro, mentre le regioni con speranze di vita più alte, come Nord America, Europa e Australia, sono contrassegnate da colori più caldi come il giallo. Questo contrasto cromatico rende facile identificare le disparità globali, facilitando l’interpretazione delle differenze regionali nella speranza di vita.

#### 4.4 Visualizzazione dei risultati

Il grafico “Residui vs. Valori Predetti” è uno strumento fondamentale per valutare l’adeguatezza del modello di regressione lineare. Nel grafico, i residui sono rappresentati rispetto ai valori predetti dal modello, consentendo di osservare eventuali schemi sistematici o irregolarità.

**Figura 1: Residui vs. Valori Predetti**



Osservando il grafico, i residui si distribuiscono in modo abbastanza casuale intorno alla linea orizzontale a zero, senza pattern evidenti. Questo supporta l'ipotesi di omoschedasticità e indica che il modello non esclude informazioni rilevanti. Tuttavia, si nota una leggera concentrazione di residui positivi per valori predetti più alti, suggerendo una possibile sottostima della speranza di vita in questi casi. Nonostante questa lieve distorsione, il modello appare complessivamente adeguato.

## 5 Conclusions

Questo studio ha esaminato in modo integrato l'impatto di fattori ambientali, sanitari e comportamentali sulla speranza di vita, colmando una lacuna nella letteratura. I risultati mostrano che la mortalità prematura, l'inquinamento atmosferico e gli incidenti stradali hanno un effetto negativo significativo, mentre l'accesso ai servizi idrici di base è associato a una maggiore longevità. Il modello adottato ha spiegato gran parte della variabilità osservata, confermando la necessità di strategie di sanità pubblica mirate per ridurre i principali fattori di rischio e migliorare il benessere globale.

## 6 References

World Health Organization. (2025). Global Health Observatory data: Health workforce statistics. World Health Organization. Recuperato da <https://apps.who.int/gho/data/node.main?showonly=HWF>

Esposito, Simona, Marialaura Bonaccio, Augusto Di Castelnuovo, Emilia Ruggiero, Mariarosaria Persichillo, Sara Magnacca, Amalia De Curtis, et al. 2024. “Life-Course Socioeconomic Trajectories and Biological Aging: The Importance of Lifestyles and Physical Wellbeing.” *Nutrients* 16 (19): 3353. <https://doi.org/10.3390/nu16193353>.