




Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II.

doi: 10.18637/jss.v000.i00

Analisi e valutazione del benessere della società: caso studio sugli indici di vita

Carmela Pia Senatore 
Università degli studi di Salerno

1. Definizione degli obiettivi

La seguente analisi si pone degli obiettivi tramite l'utilizzo del database OECD, con particolare attenzione su tre aspetti principali.

- In primo luogo, in questa parte, verrà condotta un'analisi delle serie storiche basata su osservazioni annuali per comprendere l'evoluzione di specifiche variabili nel tempo. Questo consentirà di acquisire una prospettiva longitudinale e identificare eventuali tendenze o cambiamenti significativi.
- Successivamente, nella seconda parte, ci si concentrerà su un anno specifico (il 2017), in cui verrà condotta un'analisi esplorativa e di clustering. Quest'ultima permetterà di ottenere una visione più dettagliata dei pattern e delle relazioni presenti nei dati di quell'anno in particolare.
- Infine, verrà scelta una variabile aleatoria sulla quale svolgere inferenza.

L'obiettivo principale è capire quali elementi hanno un impatto significativo sul livello di benessere e qualità della vita.

2. Descrizione del dataset

Il **Better Life Index** è un database globale che raccoglie dati provenienti da circa 40 Stati in tutto il mondo. L'obiettivo principale del database è coinvolgere e rendere consapevoli i cittadini in un dibattito significativo sulla valutazione del benessere della società e consentire di partecipare attivamente al processo di formulazione delle politiche che influenzano la qualità della vita di tutti. Questo strumento si propone di rendere le persone più informate e partecipi nell'ambito della governance.

Il Better Life Index è suddiviso in 11 sottotemi o macroaree, ognuna delle quali è supportata da uno a tre indicatori. All'interno di ciascuna delle macroaree, viene calcolata una media degli indicatori, ciascuno dei quali ha lo stesso peso nella valutazione. La scelta degli indicatori si basa su criteri statistici rigorosi, come la pertinenza (ovvero quanto un indicatore rifletta in modo adeguato il concetto che si intende misurare), la profondità (la capacità di catturare aspetti rilevanti del benessere), e la rilevanza politica (cioè quanto l'indicatore è utile per le decisioni politiche). Inoltre, la qualità dei dati è un criterio fondamentale nella selezione degli indicatori. Questo include la validità predittiva (la capacità di un indicatore di prevedere il benessere futuro), la copertura (quanto ampiamente l'indicatore può essere applicato a diverse realtà), la tempestività (la disponibilità dei dati in tempo reale o con aggiornamenti regolari), e la comparabilità tra paesi (la possibilità di confrontare dati tra nazioni diverse). Tutti questi criteri sono stabiliti in consultazione con l'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE) e i paesi membri di questa organizzazione. Gli indicatori selezionati sono quindi considerati buone misure dei concetti di benessere, specialmente quando si tratta di condurre un confronto tra paesi. Il Better Life Index mira quindi a fornire un quadro informativo completo e affidabile per valutare il benessere delle società a livello globale, incoraggiando un coinvolgimento significativo dei cittadini nel processo decisionale.

E', quindi, un dataset multivariato che contiene più osservazioni, ciascuna con un set completo di valori per le variabili coinvolte. Questo significa che il dataset è costituito da una matrice

di dati in cui le righe rappresentano le osservazioni e le colonne rappresentano le variabili. Sono presenti dati prevalentemente di tipo quantitativi.

Il dataset include le seguenti macroaree e relative variabili:

- **Housing**

- **Dwellings without basic facilities:** Questo indicatore si riferisce alla percentuale della popolazione che vive in una residenza senza un water destinato esclusivamente all'uso delle famiglie. I water situati all'esterno della residenza non sono da considerare in questa categoria. Inoltre, vengono conteggiati anche i water situati in una stanza in cui è presente anche un'unità doccia o una vasca da bagno. Essenzialmente esprime la percentuale di famiglie che vive senza utilità di tipo primarie.
- **Housing expenditure:** Questo indicatore considera la spesa delle famiglie per l'abitazione e la manutenzione della casa, come definito nel Sistema delle Nazioni Unite (P31CP040: Alloggio, acqua, elettricità, gas e altri combustibili; P31CP050: Arredi, attrezzature per la casa e manutenzione ordinaria della casa). Essa comprende l'affitto effettivo e imputato per l'alloggio, la spesa per la manutenzione e la riparazione dell'abitazione (compresi servizi vari), l'approvvigionamento idrico, l'energia elettrica, il gas e altri combustibili, beni e servizi per la manutenzione ordinaria della casa come percentuale del reddito lordo disponibile delle famiglie.
- **Rooms per person:** si riferisce al numero di stanze (escludendo cucina, dispensa, bagno, gabinetto, garage, studi medici, ufficio, negozio) in un'abitazione diviso per il numero di persone che vivono nell'abitazione.

- **Income**

- **Household net adjusted disposable income:** l'importo massimo che una famiglia può permettersi di consumare senza dover ridurre il proprio patrimonio o aumentare i propri debiti. Si ottiene sommando al reddito lordo delle persone (guadagni, reddito da lavoro autonomo) i trasferimenti sociali in natura che le famiglie ricevono dai governi, e quindi sottraendo le tasse sul reddito e sulla ricchezza.
- **Household net financial wealth:** la ricchezza finanziaria netta composta da: oro monetario, valuta e depositi, titoli diversi dalle azioni, prestiti, azioni e altri strumenti patrimoniali (inclusi i titoli emessi dai fondi d'investimento).

- **Jobs**

- **Employment rate:** Si tratta del numero di persone occupate di età compresa tra 15 e 64 anni rispetto alla popolazione della stessa fascia d'età. Le persone occupate sono coloro che hanno 15 anni o più e che dichiarano di aver lavorato in un'occupazione retribuita.
- **Job security:** l'indicatore rappresenta il numero di dipendenti occupati con una anzianità lavorativa inferiore a 6 mesi rispetto all'occupazione dipendente totale. L'anzianità lavorativa è misurata in base alla durata del tempo durante il quale i lavoratori hanno svolto il loro attuale o principale lavoro o sono stati impiegati presso il loro attuale datore di lavoro.

- **Long-term unemployment rate:** questo indicatore si riferisce al numero di persone che sono state disoccupate per un anno o più, espresso come percentuale della forza lavoro (la somma delle persone occupate e delle persone disoccupate).
- **Personal earnings:** l'indicatore si riferisce alla retribuzione annua media per equivalente a tempo pieno di un dipendente, ottenuta dividendo il totale dei salari basati sui conti nazionali per il numero medio di dipendenti nell'economia complessiva, moltiplicato poi per il rapporto tra le ore medie settimanali solitamente lavorate da un dipendente a tempo pieno e le ore settimanali solitamente lavorate da tutti i dipendenti.
- **Community:**
 - **Quality of support network:** esprime una misura del supporto percepito all'interno della rete sociale. Questo indicatore si basa sulla domanda: "Se avessi dei problemi, hai parenti o amici su cui puoi contare per aiutarti ogni volta che ne hai bisogno o no?" e tiene conto delle risposte positive da parte dei partecipanti. In altre parole, misura la percezione delle persone sulla disponibilità di parenti o amici pronti ad aiutarle in caso di necessità.
- **Education**
 - **Educational attainment:** si tratta del numero di individui che hanno completato almeno con successo l'istruzione di livello secondario superiore rispetto alla popolazione compresa tra i 25 e i 64 anni.
 - **Student skills:** Il punteggio medio degli studenti nei test di lettura, matematica e scienze, valutato dal Programma per la Valutazione Internazionale degli Studenti (PISA) dell'OCSE. PISA è un programma di valutazione globale che misura le competenze degli studenti in queste tre aree chiave ed è utilizzato per confrontare le performance degli studenti in tutto il mondo. Il punteggio medio riflette le capacità medie degli studenti in queste materie in un determinato paese o regione.
 - **Years in education:** durata media degli studi in anni.
- **Environment**
 - **Air pollution:** si tratta di una misura della qualità dell'aria nelle aree urbane, con un'enfasi sulle città di dimensioni maggiori, in cui le concentrazioni di PM10 vengono ponderate in base alla popolazione residente. Questo indicatore fornisce una stima della contaminazione atmosferica e del potenziale impatto sulla salute dei residenti nelle aree urbane.
 - **Water quality:** misura il grado di soddisfazione delle persone riguardo alla qualità dell'acqua nell'area in cui vivono, basandosi sulle loro risposte soggettive.
- **Civic engagement**
 - **Consultation on rule-making:** L'indicatore è una media ponderata delle risposte sì/no a varie domande sulla presenza di consultazioni legali da parte dei cittadini e sulla presenza di procedure formali che consentano al pubblico in generale di influenzare la regolamentazione e le azioni governative. L'indicatore descrive l'entità in cui processi formali di consultazione sono integrati in fasi chiave della progettazione di proposte di regolamentazione e quali meccanismi esistono affinché l'esito

di tali consultazioni influenzi la preparazione di leggi primarie e regolamenti subordinati. Le domande su cui si basa l'indicatore riguardano l'esistenza di procedure formali che permettono al pubblico in generale, alle imprese e alle organizzazioni della società civile di influenzare la regolamentazione e le azioni governative, nonché se le opinioni dei cittadini su tali procedure di consultazione siano rese pubbliche.

- **Voter turnout:** la partecipazione elettorale è definita come il rapporto tra il numero di individui che hanno votato durante un'elezione e la popolazione registrata per votare. Poiché le caratteristiche istituzionali dei sistemi di voto variano notevolmente tra i paesi e tra i tipi di elezioni, l'indicatore si riferisce alle elezioni che hanno attratto il maggior numero di elettori in ciascun paese.

- **Health**

- **Life expectancy:** l'aspettativa di vita misura quanti anni in media le persone possono attendersi di vivere in base ai tassi di mortalità specifici per età. Questa misura si riferisce alle persone nate oggi e viene calcolata come una media ponderata dell'aspettativa di vita per gli uomini e le donne.
- **Self-reported health:** si riferisce alla percentuale della popolazione di età superiore a 15 anni che dichiara di avere uno stato di salute "buono". L'Organizzazione Mondiale della Sanità (OMS) raccomanda di utilizzare un sondaggio standard di intervista sulla salute per misurare questo indicatore, formulando la domanda come "Come valuti la tua salute in generale?" con una scala di risposta che include "Molto buona/ Buona/ Sufficiente/ Cattiva/ Molto cattiva".

- **Life Satisfaction:** misura il livello di soddisfazione e benessere generale delle persone in base alla loro percezione della loro situazione di vita rispetto ai loro migliori e peggiori scenari possibili.

- **Safety**

- **Assault rate:** la variabile misura la percentuale di persone che ha subito un'aggressione o una rapina entro l'arco dei 12 mesi precedenti all'indagine.
- **Homicide rate:** numero annuale di omicidi intenzionali registrati dalla polizia.

- **Work-life balance**

- **Employees working very long hours:** proporzione dei dipendenti lavoratori il cui contratto di lavoro prevede 50 ore o più.
- **Time devoted to leisure and personal care:** quantità di minuti (o ore) al giorno che, in media, le persone impiegate a tempo pieno dedicano al tempo libero e alle attività di cura personale.

3. Analisi delle serie storiche

Al fine di comprendere le dinamiche relative alla qualità di vita e benessere sociale, si svolge un'analisi su serie storiche.

L'arco di tempo considerato varia con osservazioni annuali dall'anno 2013 all'anno 2017 per ciascuno stato. Gli stati totali sono 38, le variabili ad essi corredate sono 25. Per facilitare

l'analisi si considerano i seguenti stati: Italia, Stati Uniti, Belgio, Australia, Islanda e Francia. Per ciascuno degli stati in questione è stata creata una tabella in formato **ts**.

La prima analisi effettuata riguarda la qualità dei dati. La presenza di valori mancanti o valori anomali può notevolmente influire sull'analisi. Con la funzione **statsNA** si verifica la presenza di valori mancanti.

Le serie storiche non presentano missing values.

```
[1] "Length of time series:"
[1] 5
[1] "-----"
[1] "Number of Missing Values:"
[1] 0
[1] "-----"
[1] "Percentage of Missing Values:"
[1] "0%"
[1] "-----"
[1] "Number of Gaps:"
[1] 0
[1] "-----"
[1] "Average Gap Size:"
[1] 0
[1] "-----"
[1] "No NAs in the time series."
[1] "-----"

[1] "There are no NAs in the time series"
```

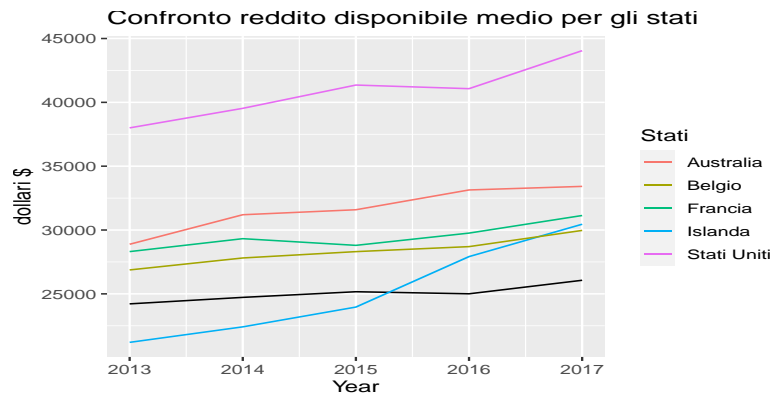
3.1. Analisi descrittiva delle serie storiche

L'attenzione, in questa sezione di analisi, è focalizzata su diverse variabili: reddito imponibile al singolo individuo, aspettativa di vita, tasso di occupazione e tempo dedicato alla cura personale.

Reddito medio individuale

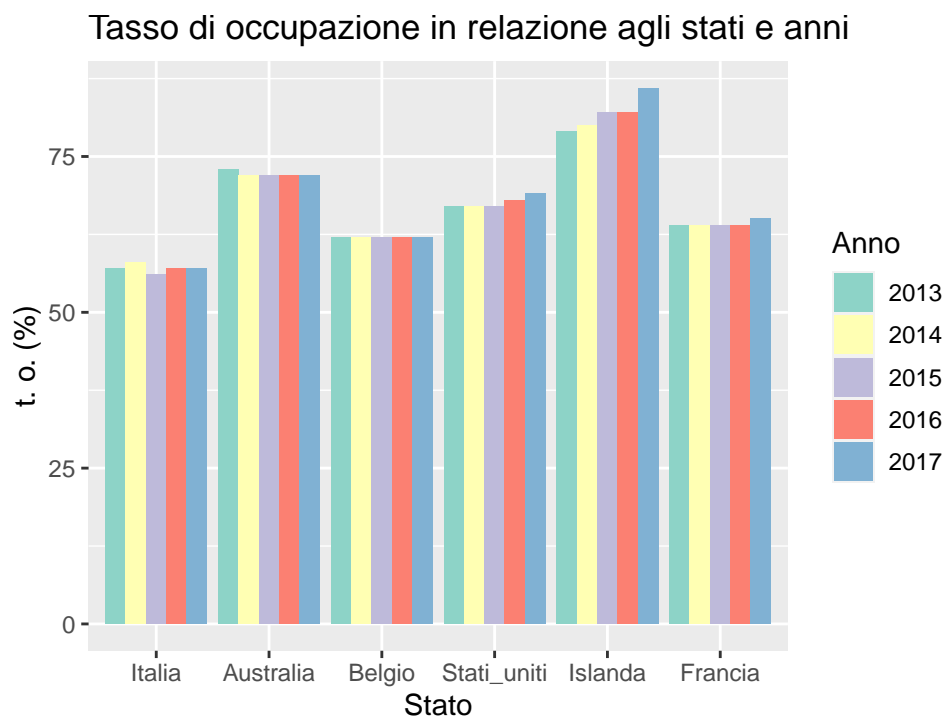
L'analisi descrittiva della serie storica ha come obiettivo la determinazione delle principali componenti: trend, stagionalità e ciclicità, oltre che l'individuazione di irregolarità. Poiché le osservazioni non sono mensili o trimestrali, non sarà possibile catturarne la stagionalità; inoltre non sarà possibile comprenderne la ciclicità in quanto per osservarla sarebbe necessario avere osservazioni almeno di 2 decenni. Tuttavia, dal plot delle osservazioni, è possibile catturare il trend che contraddistingue ogni stato. In media, si osserva un forte trend prettamente crescente, il livello medio di reddito tende a crescere nel tempo; Basti osservare, in linea generale, nell'arco dei 5 anni, il reddito iniziale di partenza di tutti gli stati e il reddito finale registrato nel 2017. Per l'Italia e gli stati uniti, l'anno 2016, ha portato a un leggero decremento del salario medio con aumento nell'anno successivo. In Italia, il reddito medio individuale è aumentato di 2000 dollari, in Islanda di 9000 dollari, negli stati uniti di 6000

dollari, in francia di 4000 dollari, in Australia di 5000 dollari e , infine, in belgio di 3000 dollari. Lo stato per eccellenza i cui livelli medi di reddito superano gli altri sono gli Stati Uniti, seguito dall'Australia. L'islanda, invece, è lo stato in cui salario medio è aumentato di più in 5 anni.



Tasso di occupazione

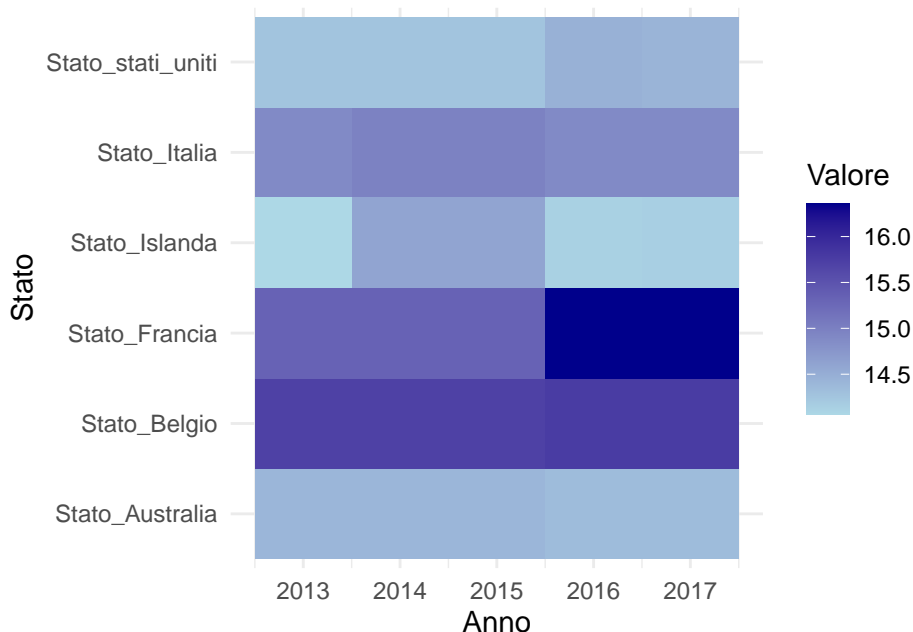
I risultati del grafico suggeriscono che, nel periodo dal 2013 al 2017, il tasso di occupazione è rimasto sostanzialmente stabile in Italia, Francia, Belgio e negli Stati Uniti. Questo significa che il numero di persone occupate o impiegate in questi paesi non ha subito cambiamenti significativi in termini percentuali rispetto alla popolazione totale. D'altra parte, il grafico mostra che l'Islanda ha sperimentato un aumento significativo e rapido del tasso di occupazione nello stesso periodo (+ 5%) . Per cui in Islanda il numero di persone occupate è notevolmente cresciuto rispetto alla popolazione totale nel periodo preso in considerazione.



Tempo dedicato alla cura personale

Chiaramente, tante più le persone lavorano, meno tempo hanno da dedicare ad altre attività come la cura personale o lo svago. La quantità e la qualità del tempo libero sono importanti per il benessere generale e possono procurare ulteriori benefici per la salute fisica e mentale. Un lavoratore a tempo pieno dedica, in media, il 63% della propria giornata, o 15 ore, alla cura personale (mangiare, dormire, ecc.) e al tempo libero (socializzare con amici e familiari, hobby, giochi, utilizzo di computer e televisione, ecc.). Nel contesto dell'analisi condotta sui dati relativi al tempo dedicato alla cura personale nei paesi presi in esame, emergono rilevanti differenze e tendenze tra i vari Stati. In Francia, durante il quinquennio considerato, è stata riscontrata una notevole crescita del tempo dedicato alla cura personale, con una variazione positiva del 6.72% equivalente a un aumento di un'ora. Negli Stati Uniti, la situazione si presenta diversamente. Per i primi tre anni del periodo analizzato, il tempo dedicato alla cura personale è rimasto sostanzialmente costante. Tuttavia, tra il 2013 e il 2017, si è osservato un incremento del 1.19%. Il dato può essere interpretato come un cambiamento graduale nell'approccio alla cura personale. In Islanda, si è invece evidenziata una riduzione del 0.6% nel tempo dedicato alla cura personale nel corso dei cinque anni, nonostante si piazzasse alle migliori posizioni in altri ambiti. Per quanto concerne l'Italia, si può affermare che, in media, il tempo dedicato alla cura personale è rimasto sostanzialmente stabile.

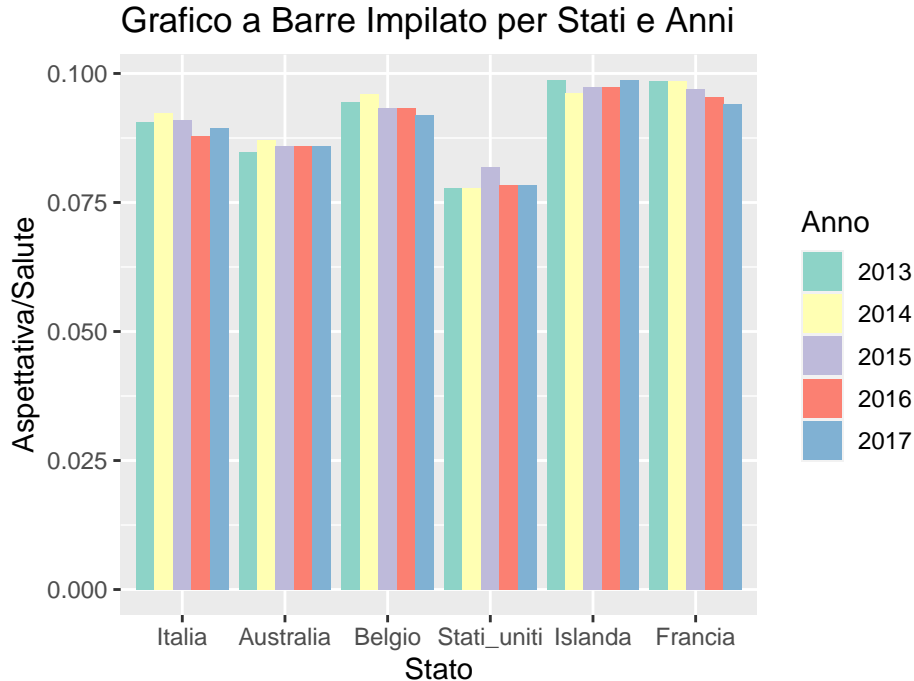
Grafico Heatmap per Tempo Dedicato alla cura



Aspettativa di vita

In maniera sorprendente, l'outcome dell'aspettativa di vita in rapporto alla salute autodichiarata deteriora con l'avanzare del tempo. Il grafico mostra come sebbene le persone stiano vivendo più a lungo, la variazione nella percentuale di persone in buona salute è minore rispetto alla variazione nell'aspettativa di vita. La salute legata alla longevità è diminuita. In parte, questo

è il risultato meccanico del fatto che le persone che vivono più a lungo potrebbero mostrare anche una maggiore incidenza di malattie.



4. Modelli di previsione Benchmark per l' aspettativa di vita

Per effettuare prevesioni è bene descrivere le supposizioni fatte per prevedere i determinanti dell'aspettativa di vita in buona salute: prevedere le tendenze nei fattori legati allo stile di vita nel tempo è una sfida, poiché tali fattori sono il risultato di una vasta gamma di influenze diverse, che vanno dalla disponibilità, ai cambiamenti culturali, ai fattori geospaziali, ecc. Si fa l'assunzione che le tendenze a lungo termine nei fattori legati allo stile di vita siano costanti. Un metodo di previsione deve tener conto dei fattori dominanti della serie, nel caso particolare Trend, (qualora avessimo osservazioni trimestrali/mensili si sarebbe potuto tenere conto di stagionalità).

In questa sede verranno confrontate le prestazioni di tre semplici modelli di previsione:

1. Average Method.

2. Naive Method.

3. Drift Method.

L' **average method** definisce che le previsioni di tutti i valori futuri sono uguali alla media dei dati storici, per cui si ha lo stesso valore per tutto il periodo di riferimento per cui si vuole fare previsione :

$$\hat{y}_{T+h|T} = \bar{y} = \frac{y_1 + y_2 + \dots + y_T}{T}$$

Per le previsioni con il metodo **naive** imponiamo semplicemente che tutte le previsioni assumano il valore dell'ultima osservazione registrata:

$$\hat{y}_{T+h|T} = y_T$$

A differenza del metodo naive, il **drift method** consente alle previsioni di crescere o decrescere nel tempo. La variazione nel tempo (chiamata drift) è definita come variazione media osservata sulla base delle informazioni passate. La previsione al tempo $T + h$ è data da:

$$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1} \sum_{t=2}^T y_t - y_{t-1} = y_T + h \frac{y_T - y_1}{T-1}$$

Ciò equivale a tracciare una retta tra la prima e l'ultima osservazione e ad estenderla nel futuro.

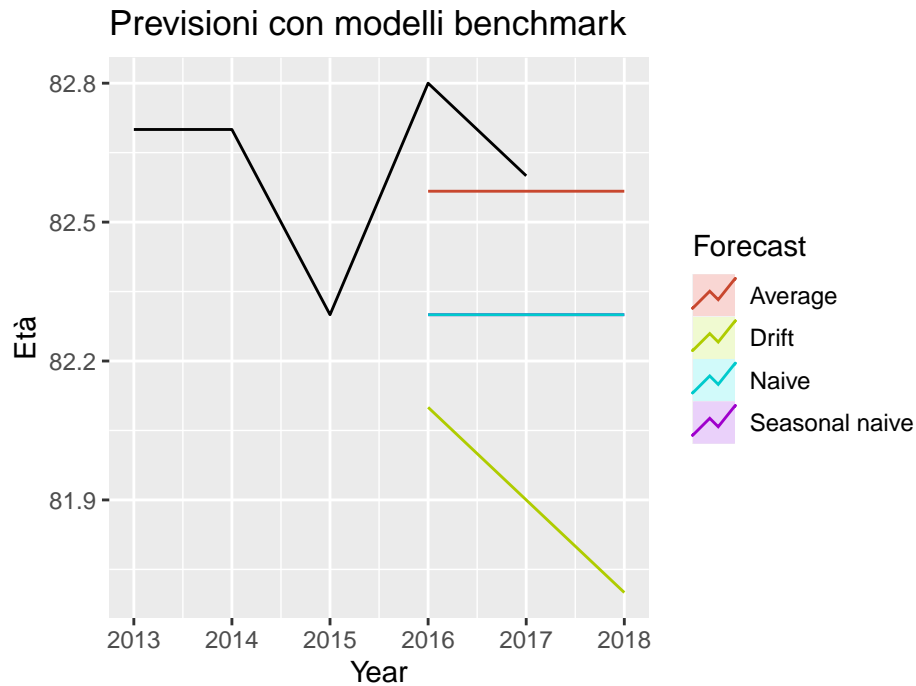
Per misurare la Bontà di Adattamento/Previsione verranno calcolati i seguenti parametri di errore:

- **errore medio assoluto** (mean absolute error: MAE): media aritmetica degli errori presi in valore assoluto;
- **radice errore quadratico medio** (root mean squared error: RMSE): radice quadrata della media aritmetica dei quadrati degli errori;
- **errore medio assoluto percentuale** (mean absolute percentage error: MAPE): media aritmetica degli errori relativi, presi in valore assoluto e moltiplicati per 100. In realtà non si tratta di un indice simmetrico in quanto viene dato un peso maggiore agli errori negativi rispetto agli errori positivi, per questo viene considerato un indice corretto noto come *symmetric mean absolute percentage error*, **SMAPE**.

4.1. Implementazione

Per verificare e comprendere i modelli di previsione, viene effettuata una divisione in **Training set e Test set per l'Italia**. Per far sì che poi successivamente sia possibile confrontare i risultati, gli archi temporali di train e split saranno i seguenti: Il training set conterrà le osservazioni dal 2013 fino al 2015, mentre il test set conterrà le osservazioni dal 2016 al 2017.

Nel grafico, i vari metodi di previsione vengono confrontati e, considerando la linea nera come il test set, emerge una netta differenza tra di essi, soprattutto per quanto riguarda una possibile componente stagionale. Il Drift method sembra essere il metodo che consente di tenere conto della direzione del decremento.



Tuttavia, nessuno dei tre modelli riesce a prevedere con precisione le osservazioni della serie storica in esame:

- il Drift metod , ad esempio, riesce a catturare il trend;
- Il naive individua il livello di andamento della serie assegnando l'ultimo valore registrato;
- l'Average method non coglie nè il trend e nè possibile stagionalità.

4.2. Valutazione dell'accuratezza

Dalla valutazione del fitting sul training set emerge che fra i metodi il migliore è il Random Walk con Drift perchè ritorna un valore di fit più basso.

	RMSE	MAE	MPE	MAPE
Average	0.1885618	0.1777778	-0.0005223987	0.2154895
Naive	0.2828427	0.2000000	-0.2430133657	0.2430134
Random Walk con Drift	0.2000000	0.2000000	-0.0005876986	0.2424257

5. Modelli di regressione

5.1. Modello con trend

Poichè la serie storica risulta essere contraddistinta da trend è possibile stimare un modello in cui l'unico regressore è questa componente. La possibile applicazione vale esplicitamente

e espressamente per le serie storiche con andamento prettamente crescente o decrescente, mentre non rimane valida per serie che mostrano molti cambi di direzione.

Il primo modello specificato comprende solo il trend. Il trend è non statisticamente significativo, a un livello di α pari a 0.05. Tuttavia, per ogni anno che passa l'aspettativa media di vita diminuisce di 0.20 anni.

```
Call:
tslm(formula = ts_train ~ trend)

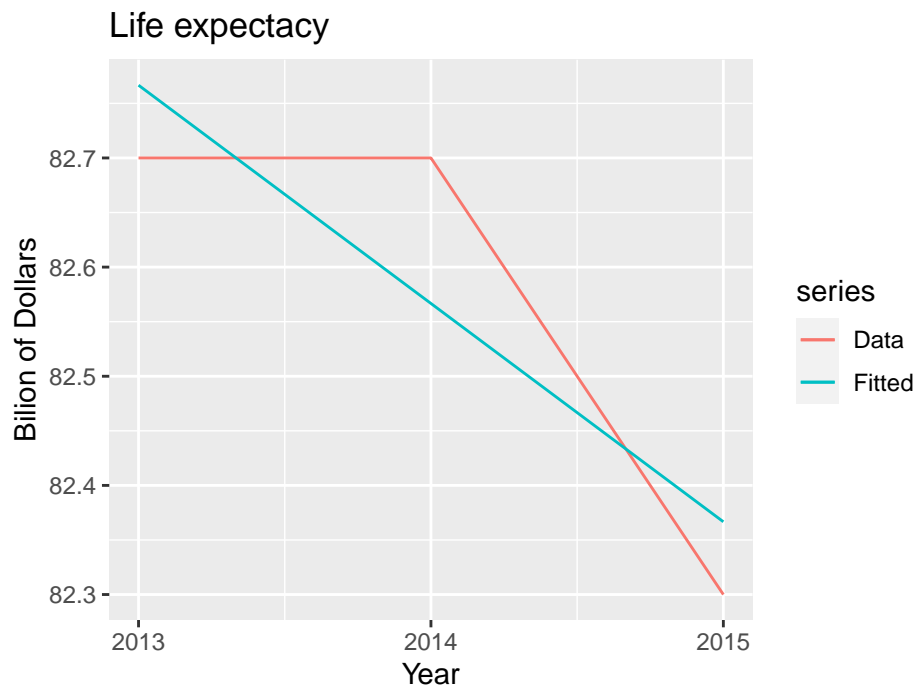
Residuals:
[1] -0.06667  0.13333 -0.06667

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  82.9667      0.2494 332.607  0.00191 **
trend        -0.2000      0.1155  -1.732  0.33333
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1633 on 1 degrees of freedom
Multiple R-squared:  0.75, Adjusted R-squared:  0.5
F-statistic:      3 on 1 and 1 DF,  p-value: 0.3333
```

Confronto grafico tra aspettativa di vita osservata e prevista

Il modello stimato con sola componente di trend non fitta in maniera ottimale i dati, sottostima l'osservazione dell'anno 2014 e sovrastima quelle del 2013 e 2015.

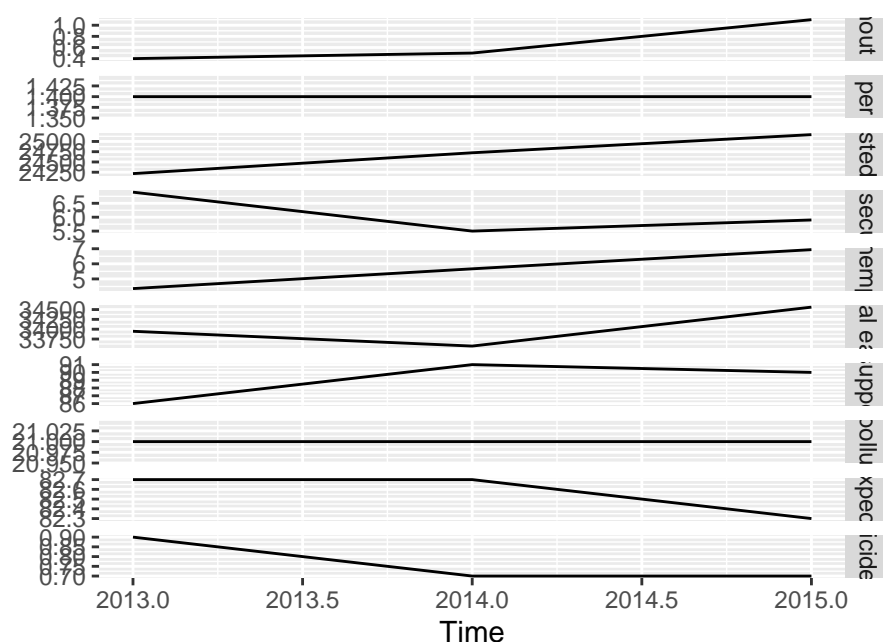


5.2. Modello di regressione con altre variabili

Il modello di regressione multipla permette di tener conto delle relazioni presenti tra la variabile dipendente e quelle indipendenti. In particolar modo consente di indagare le associazioni e l'influenza delle covariate sul valore medio della variabile di risposta. In questo lavoro verranno osservate le relazioni tra variabili attraverso l'analisi di un'aggregazioni di serie storiche composte dalle seguenti variabili:

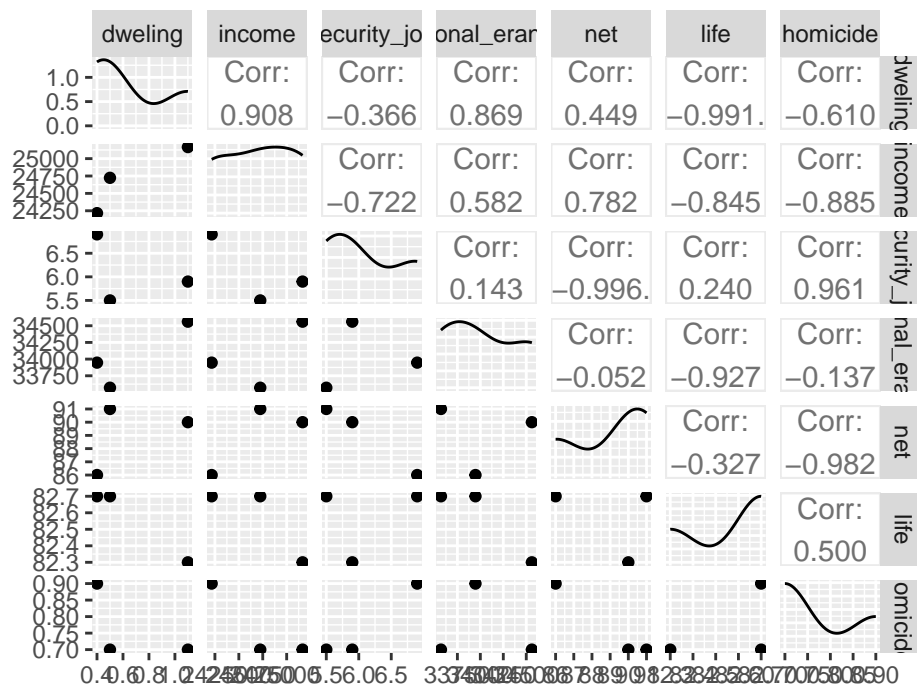
- Percentuale di servizi sanitari;
- Camere per persona;
- Reddito netto disponibile;
- Insicurezza del mercato di lavoro
- Guadagno personale;
- Qualità del supporto di rete

Il plot della serie multivariata è un grafico che rappresenta l'andamento delle variabili nel tempo. In questo tipo di grafico, ogni variabile della serie è rappresentata su un asse (l'asse y), mentre il tempo è rappresentato sull'altro asse (l'asse x). Le diverse variabili della serie sono rappresentate come linee, ognuna delle quali mostra l'andamento della variabile specifica nel tempo. Il grafico è utile per analizzare e confrontare l'andamento delle diverse variabili della serie, e per individuare eventuali relazioni tra di esse. Ad esempio, **dwelling e income** mostrano un andamento simile nel tempo, ciò può indicare una possibile correlazione tra di esse. Al contrario, **tasso di omicidi e aspettativa di vita** mostrano un andamento opposto nel tempo, ciò può indicare una relazione inversa tra di esse.



Un'ulteriore analisi preliminare alla costruzione del modello di regressione è la rappresentazione, tramite una matrice, delle relazioni tra le variabili. In particolare, la diagonale principale è rappresentata dalla visualizzazione della distribuzione della variabile in questione. La parte destra superiore alla diagonale principale, riporta i valori delle correlazioni. La variabile *life*, che sarà dipendente nel modello, risulta essere fortemente correlata negativamente con dwelling, income e personal earningd, debolmente correlata positivamente con security of market. La forza di tali correlazioni è mostrata dal coefficiente che si trova nella prima riga.

Un problema comune con le serie storiche è quello di fare l'errore di implementare regressioni spurie: ovvero, mettere in relazione delle serie storiche che risultano essere correlate ma che in realtà misurano fenomeni completamente differenti. Non è questo il caso in quanto le variabili sono state scelte con cura: la presenza di un numero maggiore di famiglie che vivono in situazioni disastrose potrebbe far diminuire l'aspettativa di vita così come l'aumento del reddito medio percepito potrebbe far aumentare l'aspettativa di vita.



Modello con DWELLING: Il modello considerato è quello contenente il regressore relativo alla presenza in percentuale per l'Italia di famiglie che vivono in condizioni disagiate con non accesso ai servizi sanitari primari. Il summary del modello permette di avere una colonna dedicata ai coefficienti stimati. Fissato un livello alpha pari a 0.10:

- La variabile DWELLING presenta un coefficiente stimato pari a -0.60, tenendo conto dell'effetto marginale della presenza di famiglie senza servizi sanitari. A seguito di una variazione di questa percentuale ci si aspetta un decremento dell'aspettativa di vita di 0.60 anni in media; I residui variano in un range che va da -0.02 a 0.004. Media e mediana approssimativamente coincidono il che potrebbe essere indice di una distribuzione simmetrica. Confrontando i residui di questo modello con quello relativo al solo trend in cui il campo di variazione è più ampio permette di comprendere che la scelta di un regressore che effettivamente impatta sulla variabile dipendente offre la possibilità di fare stime e previsioni vicine a quelli reali. L R^2 quadro corretto, con valore uguale a 0.98 è indice di un buon adattamento ai dati. La decisione di affidare i risultati all' R^2 corretto piuttosto che all' r^2 sta nel fatto che a differenza del normale R^2 corretto non esplode dal crescere delle variabili oltre che aumenta al aumentare dei regressori anche se le variabili non hanno nessun impatto, perché i regressori vengono calcolati con l'obiettivo di minimizzare gli errori e quindi tende verso 1. Un altro indice utile per la comprensione dell'adattamento ai dati è il residual standard error, questo pari a 0.04 implica che il modello potrebbe essere ottimo in quanto i valori mediamente si discostano poco dalla retta di regressione. Il test sull'insieme dei coefficienti, rappresentando dalla statistica Test F a cui si affianca il valore del p-value, in cui l'ipotesi H_0 equivale a considerare tutti i coefficienti del modello sono statisticamente uguali a 0 contro H_1 in cui almeno un coefficiente è statisticamente diverso da 0 per un livello di alpha prossimo a 0.10, aiuta nella comprensione del modello. Se accettassimo H_0 , il modello con solo intercetta

funziona meglio rispetto al modello con i regressori. In questo caso il p-value prossimo allo 0 indica che il regressore ha un impatto sulla variabile dipendente.

Call:

```
tslm(formula = life ~ dweling, data = aus_dat)
```

Residuals:

```
[1] -0.027907  0.032558 -0.004651
```

Coefficients:

```
              Estimate Std. Error  t value Pr(>|t|)
(Intercept) 82.96977      0.05920 1401.525 0.000454 ***
dweling     -0.60465      0.08056  -7.506 0.084323 .
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.04313 on 1 degrees of freedom

Multiple R-squared: 0.9826, Adjusted R-squared: 0.9651

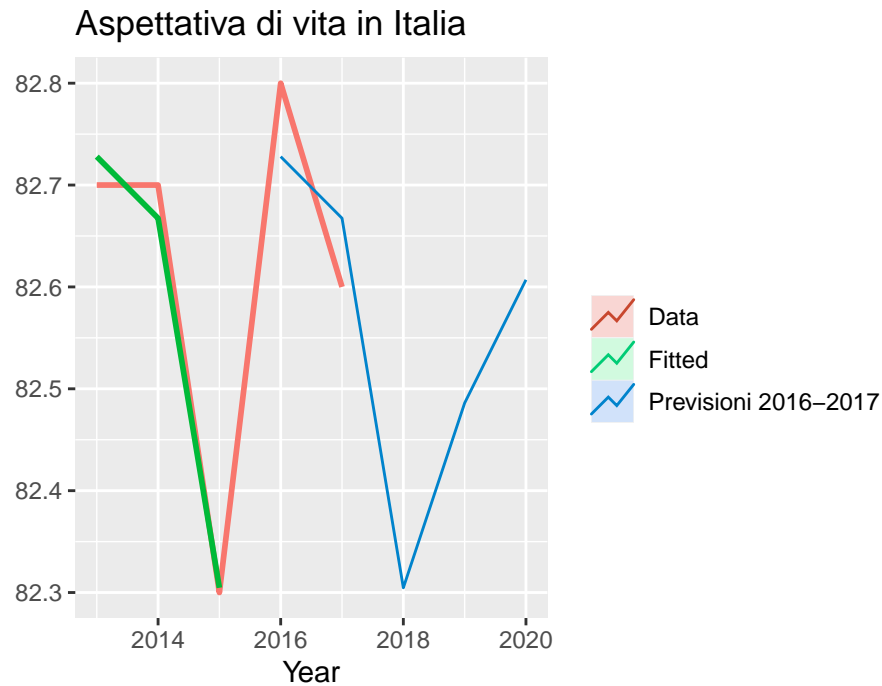
F-statistic: 56.33 on 1 and 1 DF, p-value: 0.08432

5.3. Previsioni con il modello di regressione

Il training set viene utilizzato per addestrare il modello: ovvero, il modello impara dalle osservazioni nel training set e cerca di trovare i pattern o le relazioni tra le variabili che possano aiutarlo a fare previsioni su nuovi dati. Una volta che il modello è stato addestrato, viene testato utilizzando il test set. Il test set rappresenta una porzione separata del dataset complessivo, che non è stata utilizzata per addestrare il modello. Il modello utilizza i dati del test set per fare previsioni e valutare la sua capacità di generalizzazione, ovvero di fare previsioni accurate su nuovi dati. Per cui dopo aver stimato e addestrato il modello sul training set, è possibile utilizzare il test set per fare **previsioni ex-post**. La previsione ex-post implica l'utilizzo di dati per i predittori che sono stati osservati e che non sono stati utilizzati per la stima del modello.

	RMSE	MAE	MAPE
Training set	0.02490291	0.02170543	0.02625509
Test set	0.06980619	0.06976744	0.08435881

Il modello riesce a stimare abbastanza bene i valori nel campione di training mentre mostra una leggera difficoltà nelle previsioni. In particolare, rispetto al test set, sovrastima pesantemente i valori. La problematica principale è legata a : 1) Poca abbondanza di osservazioni che non permette di aggregare più features altrimenti il modello diventa troppo complesso e non riesce a convergere; 2) Osservazioni annuali e non trimestrali/mensili: questo permette al modello al massimo di catturare trend ma non stagionalità.



6. Conclusioni

L'analisi dettagliata condotta sui dati relativi all'indice *Better Life* rivela una panoramica complessa della situazione nei paesi membri. Questo indice, pur fornendo un quadro generale del benessere, cela una serie di sfumature cruciali. I dati indicano che ciò che è valido in media non si applica uniformemente a ogni nazione e mostra variazioni sostanziali all'interno delle diverse fasce della popolazione. Un esempio di questa variazione è rappresentato dall'aspettativa di vita, spesso considerata un segnale positivo. Tuttavia, in alcuni paesi, questa crescita si sta stabilizzando, fungendo da campanello d'allarme. Nel periodo compreso tra il 2013 e il 2017, numerosi paesi hanno registrato peggioramenti significativi in indicatori chiave come l'accessibilità all'alloggio, la povertà relativa al reddito e gli aspetti sociali. Le previsioni relative all'Italia indicano una graduale diminuzione di questo indice nel tempo. Queste problematiche non colpiscono soltanto un numero limitato di nazioni, bensì coinvolgono un considerevole numero di paesi membri. Inoltre, i maggiori progressi nel benessere si sono verificati prevalentemente nei paesi che originariamente presentavano un livello di benessere più basso. È interessante notare che i miglioramenti nel benessere non sono sempre strettamente correlati alla crescita del PIL o agli indici economici. Esistono situazioni in cui alcuni paesi registrano un aumento del benessere senza necessariamente seguire una crescita economica significativa.

Affiliation:

Carmela Pia Senatore

Università degli studi di Salerno

Matricola: 0522501721

E-mail: c.senatore50@studenti.unisa.it