

# Statistica Superiore - Relazione I

La sanità pubblica nei paesi dell'Unione Europea

Eleonora Basilico

## 1 Descrizione del problema

La seguente analisi è stata commissionata dalla Commissione Europea. L'obiettivo è quello di confrontare la condizione dei sistemi sanitari dei paesi membri per poter suggerire loro come eventualmente migliorare la propria situazione sanitaria. Così facendo, i paesi membri possono sviluppare policy sanitarie mirate ed efficaci.

I dati considerati fanno riferimento all'anno 2014 e provengono da dodici diverse tabelle fornite dall'Ufficio statistico dell'Unione Europea (link in fondo al documento).

Le osservazioni sono in totale 24 (mancano per povertà di dati Danimarca, Rep. Ceca, Ungheria e Paesi Bassi). Inoltre, ci siamo assicurati di avere un numero sufficiente di paesi appartenenti alle quattro regioni europee (occidentale, meridionale, orientale e settentrionale) al fine di individuare un'eventuale relazione tra la posizione geografica e le condizioni della sanità pubblica.

Abbiamo esaminato i seguenti fattori:

1. Pediatri;
2. Medici di chirurgia generale;
3. Dermatologi;
4. Posti letto negli ospedali;
5. Posti letto nel reparto psichiatria;
6. Dimissioni dagli ospedali;
7. Degenza media;
8. Persone che hanno dichiarato di fare uso di medicinali prescritti;
9. Persone che hanno dichiarato di fare uso di medicinali non prescritti;
10. Risorse economiche dedicate a materiali e servizi sanitari (in milioni di euro);
11. Persone che hanno dichiarato di aver consultato psicologi, psichiatri o psicoterapisti;
12. Persone che hanno dichiarato di fare uso di servizi di assistenza domiciliare.

I dati sono tutti numerici ad eccezione del fattore 5 che è dato per 100.000 abitanti, e dei fattori 8,9,11 e 12 che sono espressi in percentuale<sup>1</sup>. I dati numerici sono stati divisi per il numero di abitanti<sup>2</sup>, mentre i fattori 5,8,9,11 e 12 sono stati riscalati in modo da renderli valori pro capite.

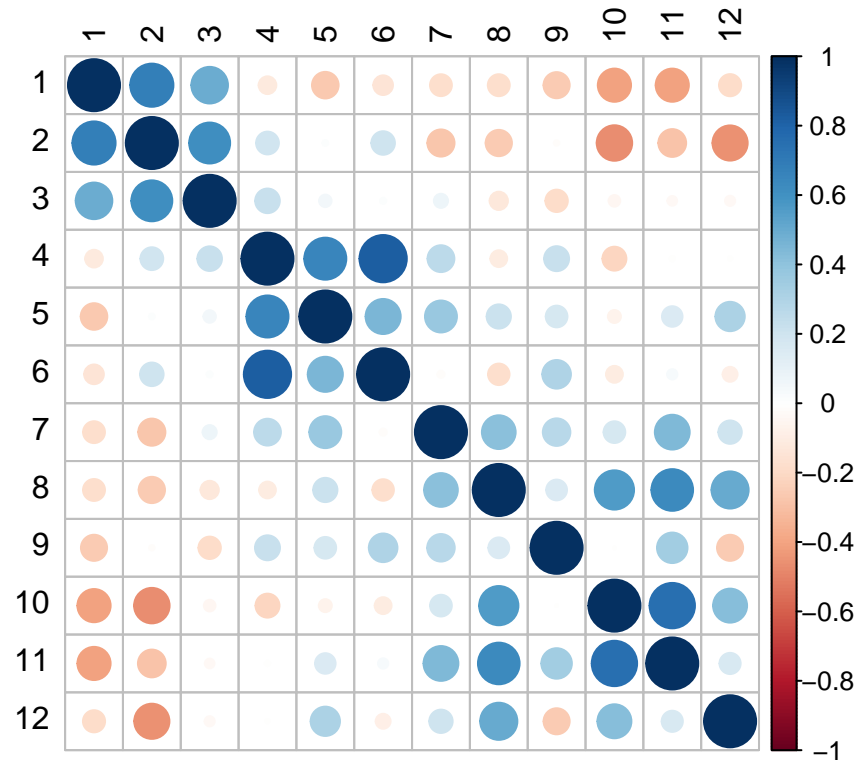
---

<sup>1</sup>Sebbene sia stato fortemente consigliato di scegliere dati non normalizzati, i soli dati numerici a disposizione non consentivano di avere un numero adeguato di osservazioni e di fattori, ed erano, quindi, insufficienti per un'analisi soddisfacente.

<sup>2</sup>Nelle tabelle corrispondenti ai fattori 1, 2 e 3 mancano i dati relativi alla Slovacchia. Al loro posto è stata inserita la media di tutti gli altri valori presenti nelle singole tabelle già divisi per il numero di abitanti.

## 2 Una prima valutazione dei dati

Per vedere se i dati considerati possono beneficiare dell'analisi delle componenti principali ne esaminiamo la matrice di correlazione:



La matrice di correlazione è sufficientemente variegata: alcuni fattori risultano fortemente correlati, ma solo il 26% delle correlazioni è in valore assoluto maggiore di 0.5. Possiamo, dunque, eseguire la PCA.

Facciamo inoltre le seguenti osservazioni:

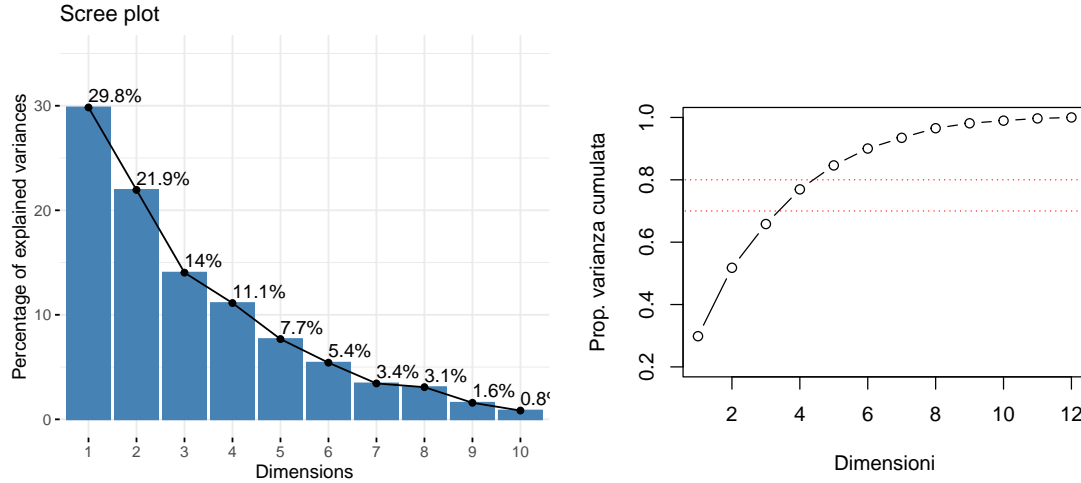
- I fattori 4 (posti letto) e 6 (dimissioni) sono quelli con correlazione positiva più alta, come ci aspettavamo;
- I fattori 1,2 e 3 (pediatri, chirurghi generali e dermatologi) sono piuttosto correlati tra loro in senso positivo, così come i fattori 4 e 5 (posti letto e posti letto nel reparto psichiatria). I primi tre, infatti, sono relativi al numero di medici, mentre il fattore 5 è incluso nel fattore 4;
- I fattori 10 (risorse economiche dedicate ai servizi sanitari) e 11 (persone che hanno consultato psicologi, psichiatri o psicoterapisti) hanno una correlazione di 0.76. Questo fatto è interessante: la consultazione di medici professionisti può essere sinonimo del fatto che da un lato ci sia la possibilità di farlo (e cioè ci sono abbastanza medici e strutture sanitarie apposite), dall'altro che l'importanza della salute personale sia diffusa tra la popolazione. Quanto detto può essere collegato alla quantità di risorse che lo Stato impiega in questo senso;
- Non ci sono correlazioni forti in senso negativo: i due fattori più fortemente negativamente correlati sono 2 e 10 e la loro correlazione è di -0.46.

### 3 Analisi delle componenti principali

Dopo aver standardizzato la tabella di dati (date le forti differenze di scala è necessario), eseguiamo l'analisi delle componenti principali.

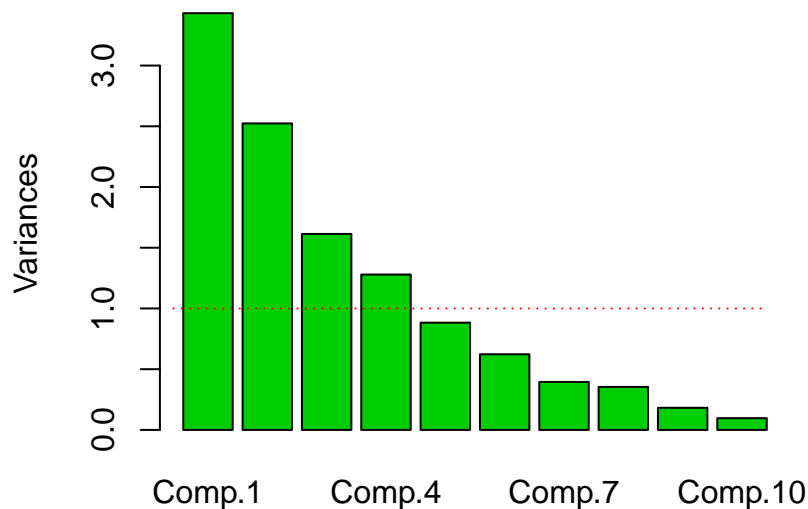
#### 3.1 Riduzione della dimensione del problema

Al fine di ridurre dimensionalmente il problema, osserviamo per prima cosa i grafici della varianza spiegata e della varianza cumulata rispetto alle componenti principali:



I risultati sono incoraggianti: le prime quattro componenti principali catturano quasi il 77% di varianza spiegata mentre le prime cinque quasi l'85%. Ci chiediamo, dunque, se conviene concentrarci sulle prime quattro oppure sulle prime cinque componenti. In entrambi i casi otteniamo una notevole riduzione dimensionale.

Gli autovalori associati alle componenti principali possono essere utili nel determinare il numero di componenti da tenere. Ricordiamo che, se i dati sono standardizzati, le componenti con autovalori strettamente maggiori di 1 sono quelle che catturano più varianza di quella catturata da una variabile originale, e pertanto sono quelle da tenere. Nel seguente grafico vediamo il valore degli autovalori associati alle singole componenti principali:



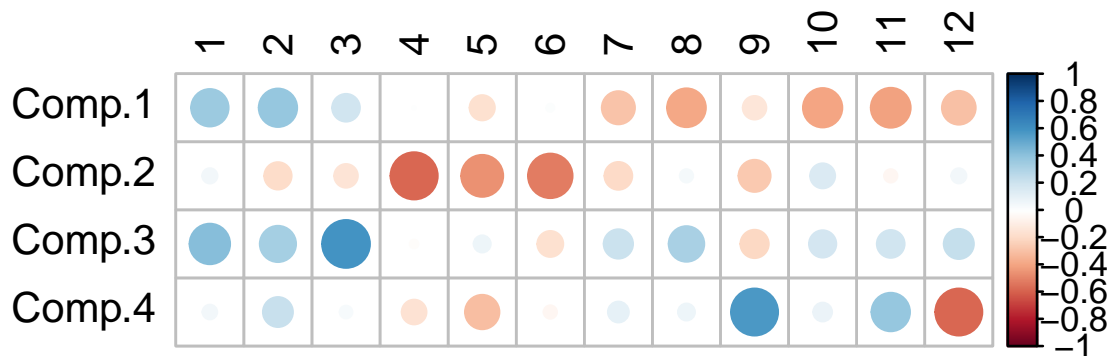
Per quanto osservato in precedenza, poichè solo i primi quattro autovalori sono  $> 1$ , possiamo concentrare la nostra analisi sulle prime quattro componenti principali: l'autovalore

associato alla quarta componente non è di molto  $> 1$  ma lo manteniamo comunque al fine di una migliore interpretazione dei dati.

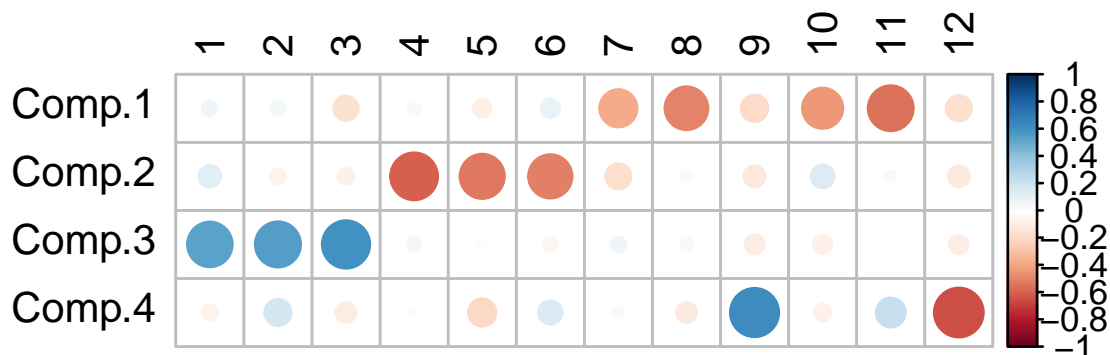
Notiamo, inoltre, che i primi due autovalori sono abbastanza più grandi rispetto ai successivi, infatti le prime due componenti catturano insieme quasi il 52% di varianza spiegata sul 77% catturato dalle prime quattro.

### 3.2 Interpretazione delle componenti principali

Cerchiamo ora di interpretare le prime quattro componenti principali. A tal fine, studiamo la matrice dei loadings ristretta alle prime quattro componenti:



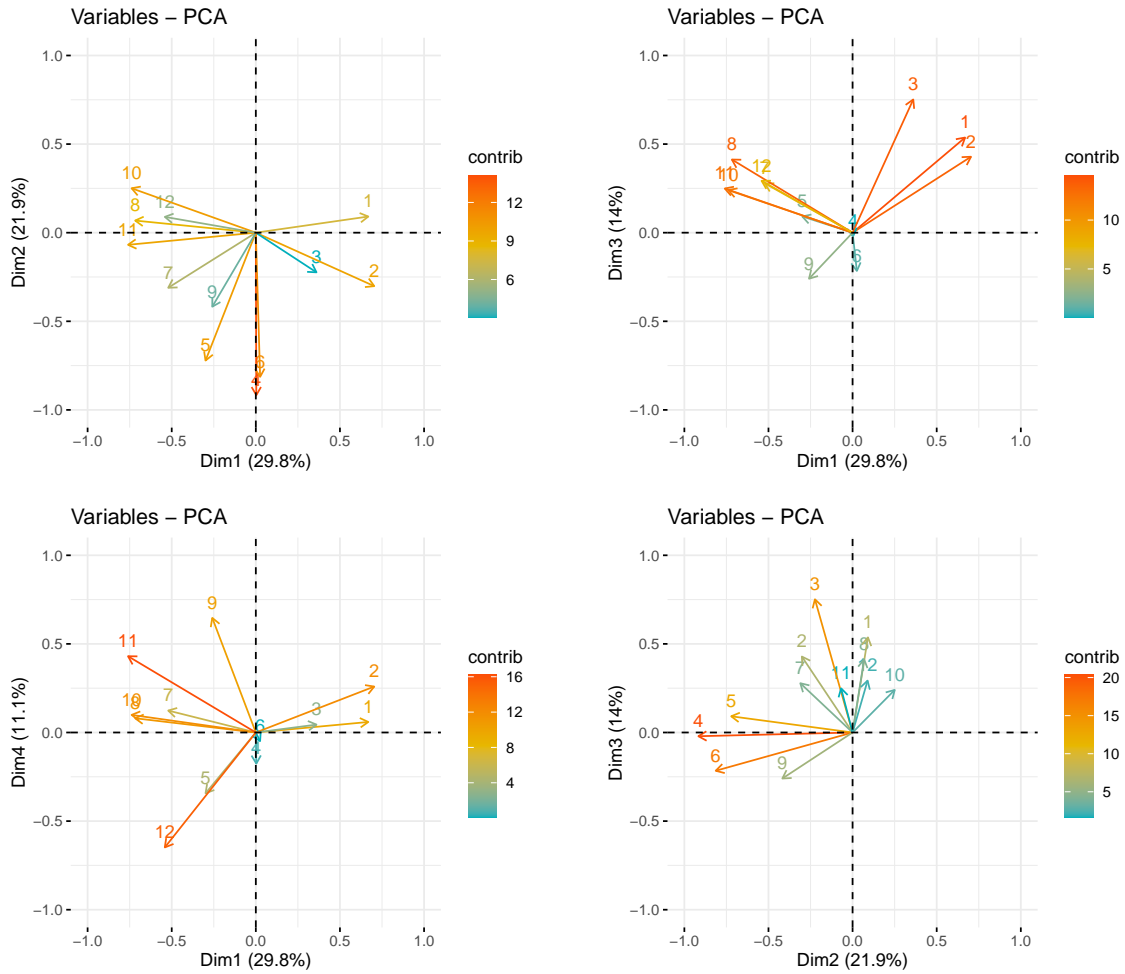
La rotazione varimax rende le associazioni dei singoli fattori alle componenti principali più nette:



Possiamo fare le seguenti associazioni:

- Comp.1: Dipende negativamente in maniera più marcata dai fattori 7, 8, 10 e 11 e meno marcata dai fattori 9 e 12. Questi fanno riferimento ai seguenti aspetti: fondi destinati alla sanità, uso di medicinali e di servizi sanitari, consultazione di medici professionisti. Possiamo, quindi, interpretare più in generale la prima componente principale come "Servizi sanitari";
- Comp.2: Dipende fortemente in senso negativo dai fattori 4,5 e 6. Interpretiamo la seconda componente principale come "Capienza delle strutture ospedaliere";
- Comp.3: Contribuiscono fortemente in senso positivo alla terza componente i fattori 1, 2 e 3. Possiamo interpretarla in generale come "Numero di medici per abitante";
- Comp.4: La quarta ed ultima componente dipende fortemente dai fattori 9 e 12 in senso positivo e negativo rispettivamente. In un certo senso essa ci permette di distinguere coloro che preferiscono assistenza sanitaria da parte di professionisti da coloro che, invece, preferiscono assumere farmaci senza consultare medici.

Per avere ulteriore conferma della bontà delle nostre associazioni, visualizziamo graficamente la posizione di ciascun fattore sui piani principali che sono risultati più significativi:



Il parametro *contrib* ci permette di distinguere quei fattori che contribuiscono maggiormente alle due componenti del piano principale considerato: quanto più una variabile contribuisce alle componenti tanto più il corrispondente valore di *contrib* è alto.

Esso è definito nel seguente modo: fissato un fattore  $i$ , se  $PC_1$  e  $PC_2$  sono due componenti principali, e  $\lambda_1$  e  $\lambda_2$  sono gli autovalori ad essi associati, allora

$$contrib(i) = \frac{(C_1 \times \lambda_1 + C_2 \times \lambda_2)}{\lambda_1 + \lambda_2},$$

dove  $C_1$  e  $C_2$  sono rispettivamente i contributi del fattore  $i$  alle componenti  $PC_1$  e  $PC_2$ <sup>3</sup>.

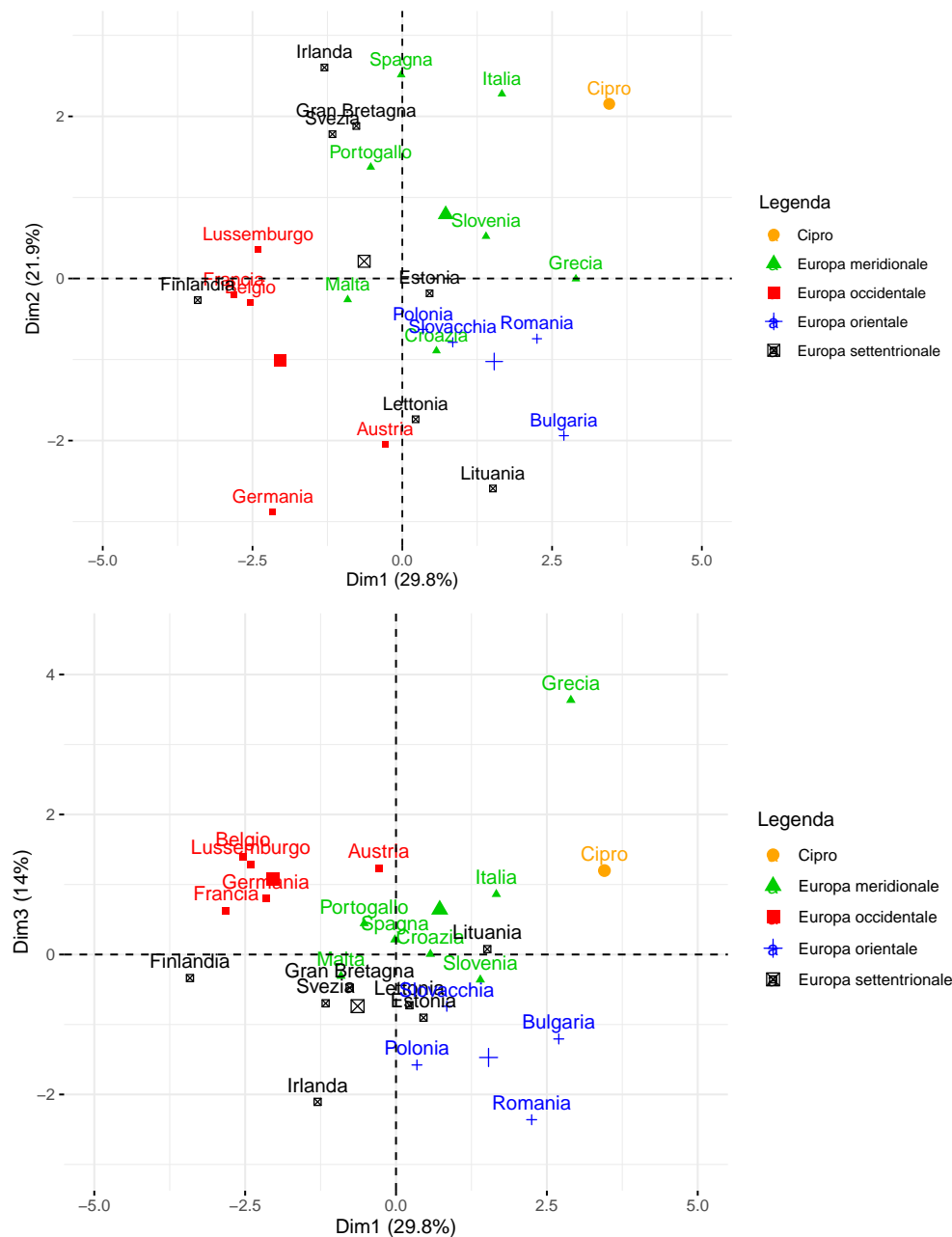
I grafici rispettano le associazioni fatte in precedenza. Notiamo, però, che questi si basano sulla sola matrice dei loadings, pertanto le associazioni alle componenti principali risultano meno nette rispetto a quelle che si hanno dopo aver applicato la rotazione varimax. In particolare, osserviamo che:

- I fattori 4, 5 e 6 sono senza dubbio da associare alla seconda componente principale: le frecce che li rappresentano sono sempre allineate alla seconda componente tranne nel secondo e nel terzo grafico in cui, però, la loro rappresentazione non è significativa;
- La quarta componente principale non può essere tralasciata, infatti è l'unica a racchiudere in modo significativo i fattori 9 e 12.

<sup>3</sup>Il contributo dei primi  $n$  fattori alle componenti può essere estratto con il comando `head(var$contrib,n)`, dove `var=get_pca_var(pca)`.

### 3.3 Posizione degli Stati nei piani principali

Visualizziamo graficamente come si collocano gli Stati nel piano principale generato dalle componenti 1,2 e 1,3 rispettivamente. Scegliamo questi due piani da un lato perchè sono più interessanti, dall'altro perchè sono più in linea con l'obiettivo della nostra analisi.



Come riportato in legenda, abbiamo diviso gli Stati in 5 gruppi in base alla suddivisione dell'Europa data dall'ONU. Cipro è isolato dagli altri in quanto sebbene faccia parte dell'UE è uno Stato asiatico.

Basandoci sulla matrice dei loadings possiamo interpretare la posizione degli Stati nei due piani principali nel seguente modo:

- La prima componente principale rappresenta i "servizi sanitari": quanto più ci si sposta verso sinistra nei grafici tanto più i servizi sanitari sono migliori;
- La seconda componente principale rappresenta la "capienza delle strutture ospedaliere": questo valore è tanto più alto quanto più si va in giù nel grafico;

- La terza componente principale rappresenta il "numero di medici per abitante": sono tanti di più quanto più si va in alto nel grafico.

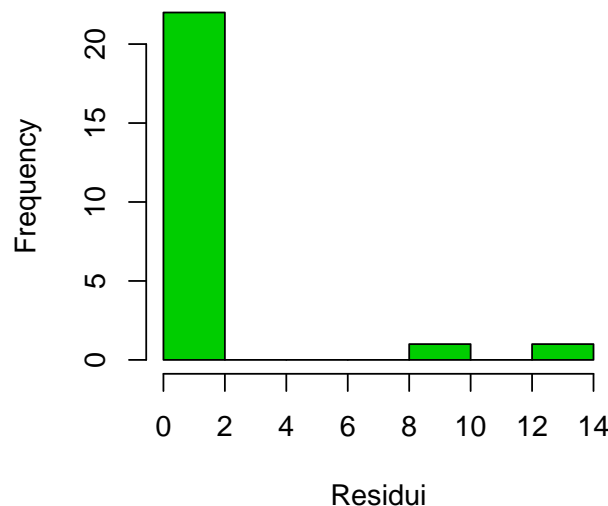
Notiamo che la posizione degli Stati rispetto alle prime tre componenti principali ci permette di cogliere le differenze che sussistono tra le varie regioni Europee in merito alle condizioni della sanità. In particolare:

- Gli Stati dell'Europa occidentale si collocano tutti nella parte sinistra dei grafici, pertanto godono di sistemi sanitari buoni. Inoltre, sono tra i paesi UE con più alto numero di medici per abitante;
- L'Europa settentrionale comprende Stati che rispetto alla prima componente si collocano nella parte centrale dei grafici, alcuni alla destra dello zero (Estonia, Lettonia, Lituania), altri alla sinistra dello zero (Gran Bretagna, Svezia, Irlanda e Finlandia). In particolare, la Finlandia risulta essere il paese UE con i migliori servizi sanitari;
- Anche l'Europa meridionale comprende alcuni Stati che si collocano alla destra e altri alla sinistra dello zero. Notiamo, inoltre, che in generale questi paesi non hanno una grande capienza ospedaliera, ma hanno un buon numero di medici per abitante. Spicca la Grecia che, nonostante non goda di buoni servizi sanitari, risulta essere il paese UE con più medici per abitante;
- Gli Stati dell'Europa orientale e Cipro si collocano nella parte destra dei grafici. Inoltre, mentre Cipro risulta avere pochi posti letto e molti medici per abitante, i restanti Stati hanno un buon numero di posti letto per abitante (spicca in questo senso la Bulgaria) e pochi medici per abitante.

Notiamo, inoltre, i seguenti fatti: l'Irlanda è agli ultimi posti sia rispetto alla seconda che alla terza componente principale; la Spagna non ha una buona capienza ospedaliera; Bulgaria, Romania, Grecia, Italia e Cipro sono in generale i paesi con i peggiori servizi sanitari. Inoltre, i primi due sono tra i paesi con meno medici e più posti letto per abitante, mentre gli ultimi due sono tra i paesi con meno capienza ospedaliera e più medici per abitante. Questo ci permette di dire che medici e posti letto non si implicano a vicenda: più medici non significa necessariamente più posti letto e viceversa.

### 3.4 Stabilità del modello

Concludiamo la nostra analisi esaminando la stabilità del modello: togliamo una osservazione alla volta prima di eseguire la PCA, ne prevediamo la collocazione sul nuovo piano principale e infine calcoliamo quanto questa si discosta dalla sua posizione effettiva.



La stabilità del modello è soddisfacente: i valori dei residui sono bassi, ad eccezione di Spagna e Bulgaria che presentano un residuo piuttosto alto rispetto a quello delle altre nazioni. Che la Bulgaria sia un'osservazione significativa potrebbe essere dovuto al fatto che essa realizza il massimo nel fattore 6 e il minimo nel fattore 7. Ciò, tuttavia, necessita sicuramente di uno studio più approfondito.

## 4 Conclusioni

L'analisi delle componenti principali risulta efficace, infatti ci permette di ridurre notevolmente la dimensione del problema. Abbiamo individuato quattro sole componenti principali, tutte di facile interpretazione, che catturano ben il 77% di varianza spiegata. Questo ci consente di trarre delle conclusioni solide.

Lo studio della collocazione dei paesi UE nei piani principali generati dalle componenti 1,2 e 1,3 ci permette di cogliere le differenze che sussistono tra le varie regioni Europee riguardo la condizione della sanità pubblica. Abbiamo visto, in particolare, che c'è una notevole disparità tra Europa Occidentale ed Europa Orientale. Questo può essere sicuramente utile per lo sviluppo di policy sanitarie mirate a ridurre tali differenze e rendere così più omogenea l'offerta sanitaria per i cittadini europei.

Infine, abbiamo visto anche quali sono gli esatti paesi UE che si trovano agli ultimi posti rispetto a servizi sanitari, numero di medici per abitante e capienza ospedaliera. Ciò può essere utile alla Commissione Europea per capire quali paesi hanno bisogno di maggiore aiuto e sotto quale aspetto della sanità pubblica bisognerebbe aiutarli.



## 5 Fonti

I dati utilizzati provengono dalle seguenti tabelle reperibili sul sito <https://ec.europa.eu/eurostat/web/main/data/database> (l'anno considerato è il 2014):

- Physicians by medical speciality (General paediatricians): [https://ec.europa.eu/eurostat/databrowser/view/HLTH\\_RS\\_SPEC\\_\\_custom\\_1594060/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/HLTH_RS_SPEC__custom_1594060/default/table?lang=en)
- Physicians by medical speciality (General surgery): [https://ec.europa.eu/eurostat/databrowser/view/HLTH\\_RS\\_SPEC\\_\\_custom\\_1594076/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/HLTH_RS_SPEC__custom_1594076/default/table?lang=en)
- Physicians by medical speciality (Dermatology): [https://ec.europa.eu/eurostat/databrowser/view/HLTH\\_RS\\_SPEC\\_\\_custom\\_1594089/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/HLTH_RS_SPEC__custom_1594089/default/table?lang=en)
- Hospital beds by type of care (Available beds in hospitals): [https://ec.europa.eu/eurostat/databrowser/view/HLTH\\_RS\\_BDS\\_\\_custom\\_1594094/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/HLTH_RS_BDS__custom_1594094/default/table?lang=en)
- Psychiatric care beds in hospitals: <https://ec.europa.eu/eurostat/databrowser/view/tps00047/default/table?lang=en>
- Hospital discharges and length of stay for inpatient and curative care (Discharges): [https://ec.europa.eu/eurostat/databrowser/view/HLTH\\_CO\\_DISCHLS\\_\\_custom\\_1594114/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/HLTH_CO_DISCHLS__custom_1594114/default/table?lang=en)
- Hospital discharges and length of stay for inpatient and curative care (In-patient average length of stay (in days)): [https://ec.europa.eu/eurostat/databrowser/view/HLTH\\_CO\\_DISCHLS\\_\\_custom\\_1594150/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/HLTH_CO_DISCHLS__custom_1594150/default/table?lang=en)
- Self-reported use of prescribed medicines by sex, age and educational attainment level: [https://ec.europa.eu/eurostat/databrowser/view/hlth\\_ehis\\_md1e/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_md1e/default/table?lang=en)
- Self-reported use of non-prescribed medicines by sex, age and degree of urbanisation: [https://ec.europa.eu/eurostat/databrowser/view/hlth\\_ehis\\_md2u/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_md2u/default/table?lang=en)
- Total health care expenditure: <https://ec.europa.eu/eurostat/databrowser/view/tps00207/default/table?lang=en>
- Self-reported consultation of mental healthcare or rehabilitative care professionals by sex, age and educational attainment level: [https://ec.europa.eu/eurostat/databrowser/view/HLTH\\_EHIS\\_AM6E\\_\\_custom\\_1594190/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/HLTH_EHIS_AM6E__custom_1594190/default/table?lang=en)
- Self-reported use of home care services by sex, age and educational attainment level: [https://ec.europa.eu/eurostat/databrowser/view/hlth\\_ehis\\_am7e/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_am7e/default/table?lang=en)
- Population on 1 January by age and sex: [https://ec.europa.eu/eurostat/databrowser/view/demo\\_pjan/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/demo_pjan/default/table?lang=en)

NOTA: Ho notato che alcuni di questi link non si aprono su Microsoft Edge, ma potrebbe essere colpa del mio computer. Su Google Chrome si aprono tutti.