



PROGETTO DI STATISTICA PARTE I

Corso di Laurea Magistrale in Artificial Intelligence and Data Engineering
Cocchella Ludovica | matricola: 532189

Sommario

1. Introduzione al caso di studio e obiettivo dell'analisi	3
1.1 Reperimento dei dati	3
2. Elaborazione dei dati	4
3. Analisi PCA (Principal component analysis).....	5
4. Conclusioni	10

1. Introduzione al caso di studio e obiettivo dell'analisi

In Italia, ogni anno, 1 persona su 4 soffre di un disturbo mentale. Nonostante i numeri, le persone che vivono con disturbi mentali e i loro cari sono di frequente vittime di isolamento e pregiudizi. Pregiudizi che impattano sulla vita di queste persone, sulle loro opportunità sociali, educative e prospettive di lavoro attuali e future.

In Italia, l'Istituto superiore di sanità (Iss) conduce molteplici attività dedicate alla salute mentale. Attività di ricerca, di sorveglianza, di prevenzione e comunicazione che vedono l'impegno di ricercatori e diversi gruppi di lavoro. L'Iss ha adesso l'obiettivo di scoprire quali sono i fattori che più vanno ad influire nello stato di salute mentale delle persone all'interno delle varie regioni italiane. Quest'analisi sarà poi utilizzata a supporto di investimenti di marketing al fine di poter attuare, dove più necessario, campagne di sensibilizzazione a favore di tutti i cittadini e puntare su nuovi investimenti per curare chi sta soffrendo di questo disagio. Le campagne tenderanno a combattere il luogo comune associato ai disturbi mentali, che spesso scoraggia coloro che ne soffrono dal cercare aiuto, e di favorire la fiducia nella professionalità della figura dello psicologo e nella possibilità di potersi far aiutare da questi.

Per questo obiettivo verrà utilizzata la tecnica di analisi statistica multivariata detta *Analisi delle componenti principali (PCA)*.

Per l'implementazione di questa analisi verrà usato il software e il relativo linguaggio R.

1.1 Reperimento dei dati

I dati che troviamo nella tabella, fornita in allegato, sono stati ottenuti dalla banca dati ISTAT, con un attento lavoro di selezione, al fine di costruire una buona tabella per la nostra analisi. La tabella .csv è stata composta seguendo il criterio seguente: l'ISTAT rende disponibili vari dati, sotto forma di tabelle, tra cui quelli riguardanti la sedentarietà, la soddisfazione per la vita e lo stato di salute delle persone. Spiegheremo di seguito, in maniera dettagliata, la procedura che è stata svolta, a partire dalla pagina indirizzata dal link <http://dati.istat.it/>, per estrapolare i dati dei seguenti fattori:

- I.Sedentari: per ricavare i dati relativi a questo fattore, che rappresenta gli individui sedentari, sono state selezionate dall'elenco dei vari temi le seguenti sezioni nel seguente ordine: "Vita quotidiana e opinione dei cittadini", "Sport e amici" e "Sport-regioni e tipo di comune". Dopo aver selezionato le precedenti sezioni viene fornita una tabella dalla quale, tramite l'opzione "Personalizza" è stato selezionato come periodo l'anno 2019, come territorio le regioni italiane e come tipo di dato "non praticano sport, né attività fisica".
- Buona.Salute e A.Disturbi.Nervosi: per ottenere i dati relativi ai seguenti fattori, che rappresentano rispettivamente gli individui che godono di buona salute e gli individui affetti da disturbi nervosi (quali ansia, depressione, psicosi ecc.), sono state selezionate dall'elenco dei vari temi le seguenti sezioni nel seguente ordine: "Salute e sanità", "Stato di salute e consumo di farmaci", "Stato di salute" e "Stato di salute-regioni e tipo di comune". Dopo aver selezionato le sezioni si ottiene una tabella dalla quale, tramite l'opzione "personalizza", è stato selezionato come periodo l'anno 2019, come territorio le regioni italiane e come tipo di dato sono stati selezionati le seguenti colonne "persone in buona salute" e "malati cronici – affetti da disturbi nervosi".
- I.Economica, I.Tempo.Libero, I.Relazioni.Amicali, I.Salute: per ottenere i dati relativi ai seguenti fattori, che rappresentano vari aspetti riguardanti la soddisfazione delle persone per la vita, sono state selezionate dall'elenco dei vari temi le seguenti sezioni nel seguente ordine: "Vita quotidiana e opinione dei cittadini", "Soddisfazione per la vita e il lavoro", "Soddisfazione per alcuni aspetti della vita" e "Soddisfazione per alcuni aspetti – regioni e tipo di comune". Dopo aver selezionato le sezioni si ottiene una tabella dalla quale tramite la funzione "personalizza" selezioniamo come periodo l'anno 2019, come territorio le regioni italiane e come tipo di dato rispettivamente "persone di 14 anni e più per livello di soddisfazione per la situazione economica" e la relativa voce "per niente", "persone di 14 anni e più per livello di soddisfazione per il tempo libero" e la relativa voce "per

niente", "persone di 14 anni e più per livello di soddisfazione per le relazioni amicali" e la relativa voce "per niente" e infine "persone di 14 anni e più per livello di soddisfazione per la salute" e la relativa voce "per niente".

- **Obesi:** per ottenere i dati relativi al seguente fattore sono state selezionate dall'elenco dei vari temi le seguenti sezioni nel seguente ordine: "Salute e sanità", "Stili di vita e fattori di rischio", "Indice di massa corporea e controllo del peso" e "Indice di massa corporea – regioni e tipo di comune". Conseguentemente si ottiene la tabella dalla quale sono state selezionate, tramite la funzione "personalizza", come periodo il periodo relativo al 2019, come territorio le regioni italiane e come tipo di dato "persone di 18 anni e più per indice di massa corporea" e la rispettiva voce "Obesi".

La tabella ottenuta, tramite le operazioni di estrapolazione precedentemente illustrate e l'unione di queste tramite Excel, risulta essere la seguente:

	I.Sedentari	Buona.Salute	I.Economica	I.Tempo.Libero	Obesi	I.Relazioni.Amicali	A.Disturbi.Nervosi	I.Salute
Piemonte	28.8	67.7	9.6	5.3	9.8	3.4	4.7	3.6
Valle d'Aosta / Vallée d'Aoste	23.3	71.3	7.8	4.9	10.2	2.1	4.6	3.5
Liguria	33.3	66.1	9.2	6.1	10.4	3.0	5.4	5.5
Lombardia	25.0	70.3	8.7	5.0	10.0	3.3	3.9	3.3
Trentino Alto Adige / SÄdtirol	12.8	78.7	5.4	4.2	8.3	2.7	4.1	2.5
Provincia Autonoma Bolzano / Bozen	11.5	83.5	4.2	2.8	7.7	1.9	4.1	2.0
Provincia Autonoma Trento	14.1	73.9	6.4	5.6	8.8	3.4	4.2	3.0
Veneto	23.2	70.8	10.6	6.9	11.2	4.3	4.2	4.2
Friuli-Venezia Giulia	25.2	70.6	8.4	6.6	11.1	3.5	4.1	4.0
Emilia-Romagna	28.9	69.2	9.0	5.8	12.3	4.7	5.2	5.6
Toscana	31.5	68.7	8.8	6.1	9.6	3.4	4.8	4.2
Umbria	34.1	68.5	11.4	7.0	11.6	3.0	6.3	4.4
Marche	32.1	67.6	9.1	5.3	11.6	2.9	4.8	4.1
Lazio	39.7	69.1	11.6	6.6	9.3	4.3	4.9	4.7
Abruzzo	35.3	66.6	12.1	7.1	12.7	3.1	5.1	4.0
Molise	53.0	64.1	10.1	6.5	15.0	3.2	4.4	4.2
Campania	51.7	71.4	11.5	6.1	11.7	2.4	4.4	3.2
Puglia	43.7	68.5	12.2	7.0	12.8	3.0	5.5	4.1
Basilicata	51.4	62.8	11.6	7.5	13.7	3.3	5.3	4.8
Calabria	51.1	58.7	16.4	6.9	12.6	2.6	6.6	5.8
Sicilia	55.2	67.4	14.0	6.1	12.6	2.7	5.6	4.3

2. Elaborazione dei dati

La nostra analisi è stata eseguita tramite il supporto del software R e di seguito vengono riportati i comandi necessari al caricamento della tabella sul software.

Prima di tutto è necessario impostare la directory da cui importare il file tramite il comando:

```
> setwd()
```

Dopo di che è necessario caricare la tabella tramite il seguente comando:

```
> tab=read.csv2("Cocchella_Ludovica_Tabella.csv", row.names = 1)
```

Per verificare che l'operazione di caricamento della nostra tabella sia andata a buon fine dobbiamo digitare i seguenti comandi che ci restituiscono rispettivamente la struttura della nostra tabella e le prime osservazioni che la compongono:

```
> str(tab)
'data.frame': 22 obs. of 8 variables:
 $ I.Sedentari : num 28.8 23.3 33.3 25 12.8 11.5 14.1 23.2 25.2 28.9 ...
 $ Buona.Salute : num 67.7 71.3 66.1 70.3 78.7 83.5 73.9 70.8 70.6 69.2 ...
 $ I.Economica : num 9.6 7.8 9.2 8.7 5.4 4.2 6.4 10.6 8.4 9 ...
 $ I.Tempo.Libero : num 5.3 4.9 6.1 5 4.2 2.8 5.6 6.9 6.6 5.8 ...
 $ Obesi : num 9.8 10.2 10.4 10 8.3 7.7 8.8 11.2 11.1 12.3 ...
 $ I.Relazioni.Amicali: num 3.4 2.1 3 3.3 2.7 1.9 3.4 4.3 3.5 4.7 ...
 $ A.Disturbi.Nervosi : num 4.7 4.6 5.4 3.9 4.1 4.1 4.2 4.2 4.1 5.2 ...
 $ I.Salute : num 3.6 3.5 5.5 3.3 2.5 2 3 4.2 4 5.6 ...
> head(tab)
      I.Sedentari Buona.Salute I.Economica I.Tempo.Libero Obesi I.Relazioni.Amicali A.Disturbi.Nervosi I.Salute
Piemonte      28.8      67.7      9.6      5.3      9.8      3.4      4.7      3.6
Valle d'Aosta / Vallée d'Aoste 23.3      71.3      7.8      4.9     10.2      2.1      4.6      3.5
Liguria       33.3      66.1      9.2      6.1     10.4      3.0      5.4      5.5
Lombardia     25.0      70.3      8.7      5.0     10.0      3.3      3.9      3.3
Trentino Alto Adige / SÄdtirol 12.8      78.7      5.4      4.2      8.3      2.7      4.1      2.5
Provincia Autonoma Bolzano / Bozen 11.5      83.5      4.2      2.8      7.7      1.9      4.1      2.0
```

3. Analisa PCA (Principal component analysis)

Per raggiungere il nostro obiettivo è stato scelto di effettuare un'analisi PCA. Quest'ultima è una tecnica statistica che, partendo da una tabella, permette la sua interpretazione attraverso un numero ridotto di fattori, che saranno combinazione di quelli originali, limitando il più possibile la perdita di informazioni, semplificandoci così la comprensione del fenomeno in esame. Questo avviene tramite una trasformazione lineare delle variabili che proietta quelle originali in un nuovo sistema cartesiano in cui le nuove variabili vengono proiettate sui nuovi assi seguendo l'ordine decrescente delle loro varianze ovvero la nuova variabile con la varianza maggiore viene proiettata sul primo asse, la seconda per dimensione della varianza sul secondo asse e così via. La riduzione della complessità avviene limitandosi ad analizzare le principali, per varianza, tra le nuove variabili.

Prima di tutto osserviamo il problema dal punto di vista delle correlazioni:

```
> round(cor(tab),2)
```

	I.Sedentari	Buona.Salute	I.Economica	I.Tempo.Libero	Obesi	I.Relazioni.Amicali	A.Disturbi.Nervosi	I.Salute
I.Sedentari	1.00	-0.74	0.76	0.66	0.74	-0.03	0.57	0.48
Buona.Salute	-0.74	1.00	-0.82	-0.77	-0.58	-0.26	-0.67	-0.81
I.Economica	0.76	-0.82	1.00	0.73	0.41	0.17	0.74	0.75
I.Tempo.Libero	0.66	-0.77	0.73	1.00	0.63	0.44	0.54	0.63
Obesi	0.74	-0.58	0.41	0.63	1.00	0.05	0.35	0.26
I.Relazioni.Amicali	-0.03	-0.26	0.17	0.44	0.05	1.00	-0.02	0.48
A.Disturbi.Nervosi	0.57	-0.67	0.74	0.54	0.35	-0.02	1.00	0.70
I.Salute	0.48	-0.81	0.75	0.63	0.26	0.48	0.70	1.00

Possiamo notare che siamo di fronte ad una tabella di valori abbastanza correlati fra loro.

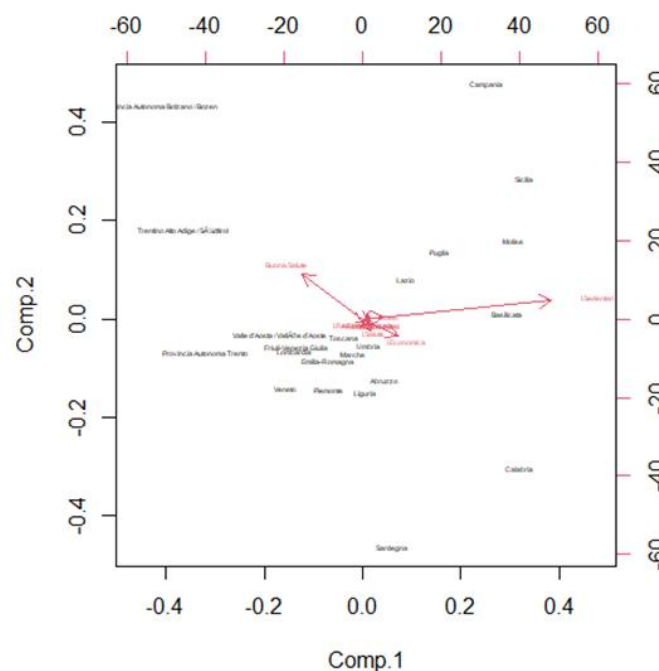
Implementiamo adesso la PCA dei dati a disposizioni tramite il comando:

```
> tab.pca=princomp(tab)
```

Guardiamo il risultato graficamente tramite il comando:

```
> biplot(tab.pca,cex=0.3)
```

Il comando precedente ci fornisce una rappresentazione bidimensionale degli individui sul piano principale (piano formato dalle prime due componenti). Il grafico che abbiamo ottenuto risulta essere il seguente:



La prima questione che dobbiamo risolvere, guardando il biplot, è se è necessario o meno standardizzare la nostra tabella. Possiamo notare in maniera abbastanza evidente la presenza di alcuni fattori che predominano su gli altri a causa dei differenti ordini di grandezza. Quindi, al fine di comprendere al meglio i nostri dati risulta indispensabile esaminare l'analisi delle componenti principali sulla tabella standardizzata. Ripetiamo dunque i comandi precedenti utilizzando la tabella suddetta:

```
> tab.st.pca = princomp(scale(tab))  
> biplot(tab.st.pca, cex=0.6)
```

Con il comando precedente otteniamo una rappresentazione della variabilità dei fattori sul piano principale in cui, come vedremo nel grafico che analizzeremo successivamente, le frecce risultano essere all'incirca della stessa lunghezza.

Per verificare se la PCA sui nostri dati risulta essere soddisfacente è necessario eseguire il comando `summary(tab.st.pca)` e guardare l'analisi delle componenti principali a livello di varianze al fine di vedere se il modello costruito riesce a rappresentare in maniera soddisfacente il fenomeno in questione.

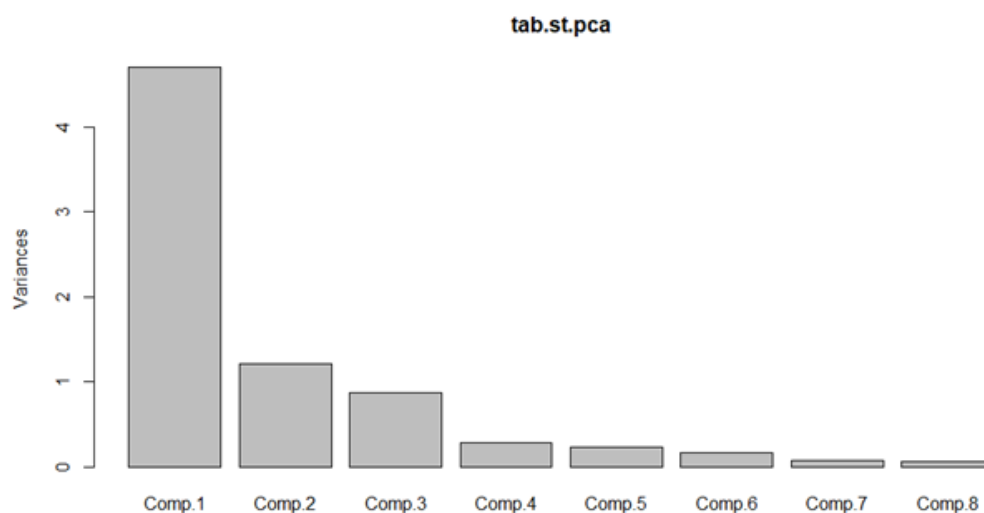
```
> summary(tab.st.pca)  
Importance of components:  


|                        | Comp.1    | Comp.2    | Comp.3    | Comp.4     | Comp.5     | Comp.6     | Comp.7    | Comp.8      |
|------------------------|-----------|-----------|-----------|------------|------------|------------|-----------|-------------|
| Standard deviation     | 2.1685729 | 1.1051364 | 0.9347829 | 0.53841580 | 0.48747184 | 0.41298244 | 0.2764738 | 0.252976400 |
| Proportion of Variance | 0.6158309 | 0.1599356 | 0.1144287 | 0.03796199 | 0.03111806 | 0.02233452 | 0.0100097 | 0.008380567 |
| Cumulative Proportion  | 0.6158309 | 0.7757665 | 0.8901952 | 0.92815716 | 0.95927521 | 0.98160973 | 0.9916194 | 1.000000000 |


```

Nella figura precedente, osservando la voce "Cumulative Proportion", possiamo notare come le prime due componenti spieghino circa il 78% del fenomeno e con la terza componente si arriva all'89%. L'obiettivo è quello di raggiungere una soglia di circa 80% usando meno componenti possibili. Quindi, possiamo dire che analizzando solo le prime due componenti, o eventualmente le prime tre, non perdiamo troppa informazione e quindi riusciamo a semplificare in maniera notevole la situazione.

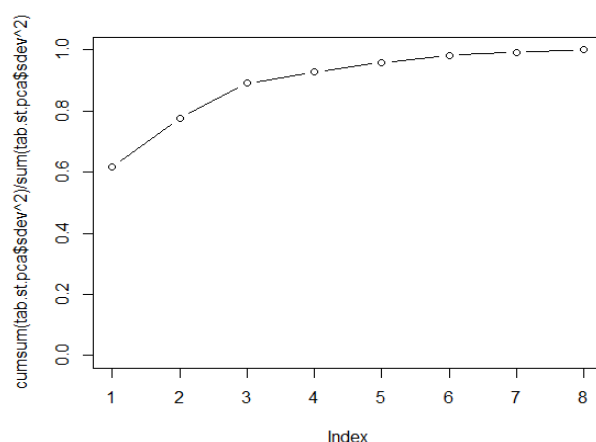
Vediamo anche graficamente come si distribuisce la varianza nelle componenti tramite l'istogramma ottenuto con il comando `plot(tab.st.pca)`:



Dal grafico ottenuto vediamo come le prime componenti (in particolare la prima) sovrastano tutte le altre, dunque, è presumibile che il problema permetta una riduzione dimensionale.

Vediamo il grafico delle proporzioni di varianza cumulata tramite il comando:

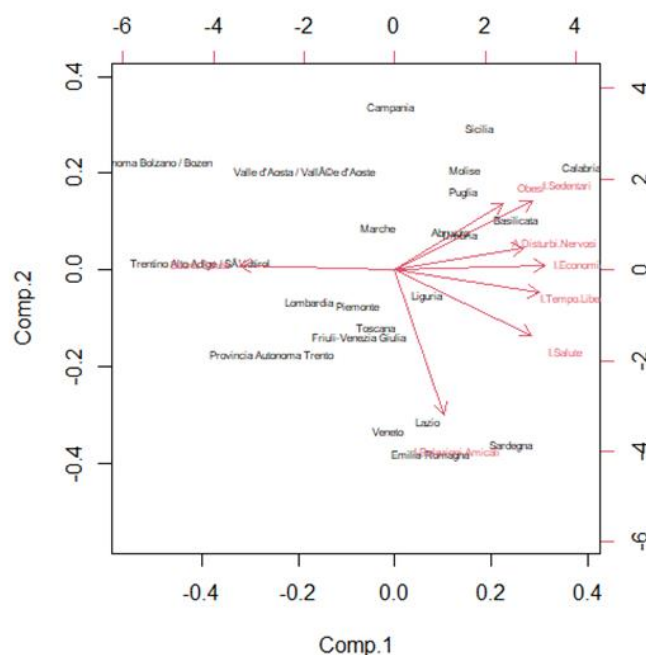
```
> plot(cumsum(tab.st.pca$sdev^2)/sum(tab.st.pca$sdev^2),type="b",ylim=c(0,1))
```



Dal grafico ottenuto vediamo come già con la seconda componente si arriva alla soglia dello 80% che viene superata ampiamente con la terza componente. Quindi, possiamo vedere come le prime componenti portino ad una varianza ampia e le ultime componenti abbiano meno varianza, infatti, vediamo una curva che tende ad appiattirsi verso uno.

Dai grafici ottenuti possiamo dunque confermare le considerazioni fatte prima analizzando il summary.

Il passo successivo consiste nell'associare i fattori d'ingresso alle componenti principali e di interpretare quest'ultime. Ciò lo facciamo guardando sia il biplot che la matrice dei loadings. Di seguito riportiamo dunque il biplot raffigurante il piano principale, i cui assi sono formati dalla componente 1 e 2:



Possiamo fare una prima analisi osservando il grafico. Da quest'ultimo possiamo notare le frecce rosse e vediamo come il fattore "Buona Salute" e "I. Economica" siano completamente allineati alla componente 1 mentre gli altri fattori contribuiscono ad entrambi le componenti anche se intuitivamente possiamo notare

che risultano più inclinate verso la prima componente (ad eccezione del fattore I.Relazioni.Amicali tutti gli altri sembrano contribuire maggiormente alla componente 1 rispetto che alla 2).

Al fine della nostra interpretazione, oltre che all'inclinazione e alla direzione delle frecce, risulta interessante osservare anche il verso delle stesse. Si evidenzia infatti che relativamente alla componente 1 il fattore "Buona Salute" ha verso opposto rispetto ai fattori che caratterizzano i disturbi nervosi. Infatti, come era prevedibile ad una Buona salute corrisponde l'assenza di disturbi nervosi che vanno a peggiorare la salute mentale delle persone.

Al fine di poter confermare e supportare le nostre associazioni andiamo ad approfondirle tramite la matrice dei loadings:

```
> loadings(tab.st.pca)
```

```
Loadings:
      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
I.Sedentari    0.374  0.364  0.164  0.406  0.118  0.484  0.515  0.148
Buona.Salute  -0.421         -0.246  0.496 -0.257  0.134 -0.644 -0.172
I.Economica    0.407         -0.246  0.496 -0.257  0.134 -0.644 -0.172
I.Tempo.Libero 0.391 -0.124  0.288 -0.660 -0.420  0.329 -0.165
Obesi          0.295  0.354  0.597 -0.433  0.215         -0.367 -0.237
I.Relazioni.Amicali 0.135 -0.769  0.358         0.409 -0.104  0.288
A.Disturbi.Nervosi 0.349  0.117 -0.504 -0.616 -0.250  0.279         0.300
I.Salute       0.371 -0.347 -0.301         0.449         0.243 -0.617

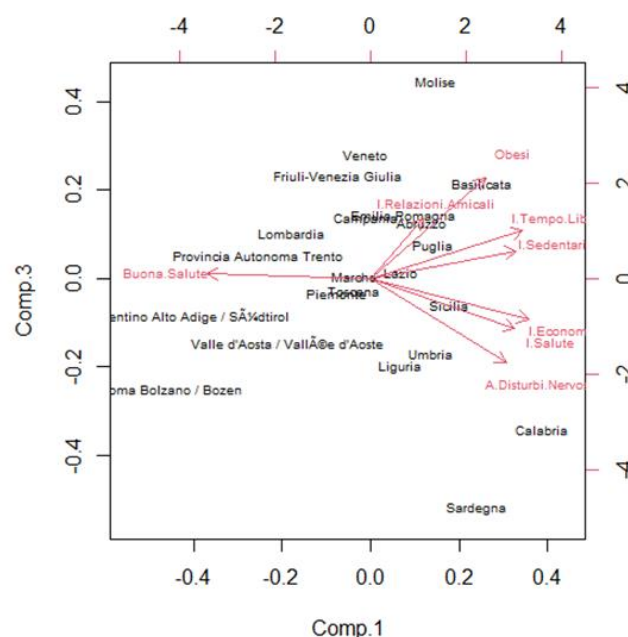
      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
SS loadings  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var 0.125  0.125  0.125  0.125  0.125  0.125  0.125  0.125
Cumulative Var 0.125  0.250  0.375  0.500  0.625  0.750  0.875  1.000
```

La matrice dei loadings non ci dà delle informazioni aggiuntive rilevanti. Infatti, possiamo notare che le osservazioni che abbiamo fatto sul biplot sono confermate dai valori della matrice dei loadings, sia per quanto riguarda i contributi che i versi.

Può risultare opportuno e interessante andare ad esplorare altri piani principali visto che la componente 3 spiega circa il 12% dell'informazione. Analizziamo così il piano composto dalle componenti 1 e 3 dato che la componente 1 porta il 61% dell'informazione.

I comandi per visualizzare il piano, composto dalla componente 1 e 3, è il seguente:

```
> biplot(tab.st.pca, choices = c(1,3), cex=0.6)
```



Da questo grafico possiamo notare come il fattore “Obesi” incida maggiormente sulla componente 3. Infatti, nel grafico raffigurante il piano principale (composto dalle componenti 1 e 2) possiamo notare come la lunghezza della freccia corrispondente al fattore “Obesi” sia leggermente più corta rispetto alle altre.

Il grafico ottenuto risulta meno interpretabile rispetto al biplot delle componenti 1 e 2, quindi non risulta molto significativo per l’interpretazione e la comprensione del nostro fenomeno. Prenderemo dunque in considerazione solo l’analisi delle prime due componenti, in particolare della prima componente principale.

Per confermare l’associazione dei fattori alle nuove componenti potrebbe convenire fare delle rotazioni delle componenti principali. L’idea delle rotazioni è quella di massimizzare le varianze dei coefficienti. La rotazione avviene tramite i seguenti comandi:

```
> tab.st.ld = loadings(tab.st.pca)
> varimax(tab.st.ld[,1:3])
$loadings

Loadings:
               Comp.1 Comp.2 Comp.3
I.Sedentari      0.154  0.131  0.508
Buona.Salute    -0.344  0.126 -0.212
I.Economica      0.466
I.Tempo.Libero   0.130 -0.344  0.341
Obesi           -0.179          0.731
I.Relazioni.Amicali      -0.851
A.Disturbi.Nervosi   0.577  0.230
I.Salute          0.505 -0.262 -0.158

               Comp.1 Comp.2 Comp.3
ss loadings    1.000  1.000  1.000
Proportion Var 0.125  0.125  0.125
Cumulative Var 0.125  0.250  0.375

$rotmat
      [,1]      [,2]      [,3]
[1,] 0.77066938 -0.2772255 0.5737724
[2,] -0.08364354 0.8486105 0.5223639
[3,] -0.63172182 -0.4505622 0.6308100
```

Possiamo dunque confermare che le associazioni che abbiamo fatto, guardando sia il biplot che la matrice dei loadings, risultano essere confermate. Infatti il fattore “I.Relazioni.Amicali” risulta essere associato alla seconda componente principale, il fattore “Obesi” alla terza componente e i restanti fattori alla prima componente principale.

Per completare la nostra analisi è possibile avere un’idea numerica andando a visualizzare i punteggi di ciascun individuo (le regioni) rispetto alle componenti principali per poi poter stilare una classifica. Visto che una parte consistente del fenomeno viene spiegata dalla prima componente, estrapoliamo i punteggi degli individui relativi ad essa per vedere come le diverse regioni si collocano lungo l’asse della prima componente. Per ottenere ciò è necessario eseguire i seguenti comandi:

```
> ranking=round(predict(tab.st.pca)[,1],2)
> sort(ranking)
```

Provincia Autonoma Bolzano / Bozen	Trentino Alto Adige / S��d��tiro��	Provincia Autonoma Trento	Valle d'Aosta / Vall��e d'Aoste
-5.58	-4.09	-2.60	-1.88
Lombardia	Piemonte	Friuli-Venezia Giulia	Toscana
-1.79	-0.78	-0.74	-0.39
Marche	Veneto	Campania	Liguria
-0.35	-0.13	-0.08	0.69
Lazio	Emilia-Romagna	Abruzzo	Umbria
0.71	0.75	1.17	1.37
Puglia	Molise	Sicilia	Sardegna
1.46	1.50	1.81	2.44
Basilicata	Calabria		
2.57	3.95		

4. Conclusioni

Dall'analisi che è stata svolta possiamo affermare di aver raggiunto in maniera abbastanza soddisfacente l'obiettivo della PCA. Infatti, è stato possibile ridurre il numero di fattori senza perdere troppa informazione riuscendo così a semplificare notevolmente la comprensione del fenomeno in esame (si evidenzia infatti come con solo due componenti arriviamo a spiegare circa l'80% del fenomeno).

Per quanto riguarda invece l'interpretazione delle componenti possiamo dire che, guardando i biplot e i loadings, la prima componente ci fornisce una sorta di classificazione delle regioni visualizzando lungo la sua direzione lo stato di salute mentale generale che le caratterizza. Quindi le regioni caratterizzate da un buon livello di salute mentale avranno punteggi nel ranking più bassi mentre quelle in cui individui presentano problemi di salute mentale sono caratterizzate da punteggi più alti. Per quanto riguarda invece la seconda componente possiamo notare come questa risulti effettivamente di più difficile comprensione portando quindi ad una maggiore difficoltà nell'interpretazione.

Nonostante che da un punto di vista statistico risulti alquanto insoddisfacente comprendere il fenomeno basandosi esclusivamente sulla prima componente, nel nostro caso però questa analisi può essere ritenuta comunque utile. Infatti, tramite la prima componente si riesce ad ottenere una sorta di indice riguardo allo stato di salute mentale generale delle varie regioni italiane che rappresenta un buon punto di partenza da fornire ad un team di persone che avranno così una visione generale della situazione che gli permetterà di semplificare notevolmente la scelta su dove focalizzare inizialmente e principalmente la propria attenzione.

Pertanto, l'analisi aveva come obiettivo quello di individuare le regioni nelle quali prevale il maggior numero di individui affetti da problemi di salute mentale al fine quindi di intervenire con maggior urgenza verso queste regioni. I risultati mostrano come le regioni più critiche, per quanto riguarda la presenza di problemi di salute mentale, siano: Calabria, Basilicata, Sardegna e Sicilia. Risulta quindi urgente investire in una forte campagna di intervento e sensibilizzazione in queste regioni.

I risultati della nostra analisi sono confermati dal Focus sul Disagio Mentale dell'Osservatorio Nazionale per la Salute nelle Regioni del Policlinico Universitario Gemelli IRCCS di Roma, reso noto alla vigilia della Giornata Mondiale per la Salute Mentale. La depressione è più diffusa nel Centro e Sud del paese (in particolare in Umbria, 9,5% della popolazione, e Sardegna 7,3%, contro il 2,8% della popolazione in Trentino-Alto Adige e il 4,3% in Lombardia) e colpisce le persone più vulnerabili sul fronte socio-economico.