

PCA

E.Chiola & P.Iezzi

21/9/2020

Dati economici su Paesi EU

Caricamento dataset e trasformazione variabili in numeriche

```
library(readr)
dati_economici <- read_delim("C:/Users/Patrizio/Desktop/TUTTO/Ud'A/CLEBA/SUPERATE/DATA SCIENCE/4 progetto/PAESI/PAESI.csv",
                             delim=";", as_is=T, guess_max=1000000)
View(dati_economici)
attach(dati_economici)
```

Si pulisce il dataset da "N/A" presenti e si crea una nuova matrice eliminando la prima colonna contenente i nomi dei Paesi. Nel comando rownames viene indicato il nome da apporre alle righe.

```
dati_economici1<-na.omit(dati_economici) #elimino gli NA
matrix0<- data.matrix(dati_economici1) #creo una nuova matrice
matrix1<-matrix0[,2:ncol(matrix0)] #creo nuova matrice senza la colonna 1
rownames(matrix1)<-dati_economici1$Paese #indico il nome delle righe richiamando la variabile (che non ho più in attach)
```

Principal component analysis

Calcolo Loadings e deviazione standard spiegata

```
dati_economici_pca<- prcomp(matrix1, center = TRUE, scale. = TRUE) #principal components
dati_economici_pca #mostriamo i loadings
```

```
## Standard deviations (1, ..., p=6):
## [1] 1.6403000 1.2444061 1.0068171 0.7126454 0.4717960 0.1293598
##
## Rotation (n x k) = (6 x 6):
##
##      PC1      PC2      PC3      PC4
## Popolazione (Millioni)  0.4720544 -0.48424353 -0.1017114 0.18527202
## PIL (Miliardi)          0.4575327 -0.51535278 -0.1099477 0.09135679
## Tasso di crescita (%)   -0.4075218 -0.07362841 -0.5002922 0.75516533
## Disoccupazione (%)       0.3783315 0.50961548 0.1821236 0.43569235
## Inflazione (%)          -0.1902583 -0.26639628 0.8324310 0.40105294
## Debito Pubblico (% del PIL) 0.4716159 0.40475136 -0.0341128 0.19074135
##
##      PC5      PC6
## Popolazione (Millioni) -0.07203346 0.70200646
## PIL (Miliardi)         -0.05169944 -0.70849698
```

```
## Tasso di crescita (%)      0.07941132 -0.04039534
## Disoccupazione (%)         -0.60929912 -0.05399169
## Inflazione (%)             0.19650273 -0.02089692
## Debito Pubblico (% del PIL) 0.75892914 -0.01534252
```

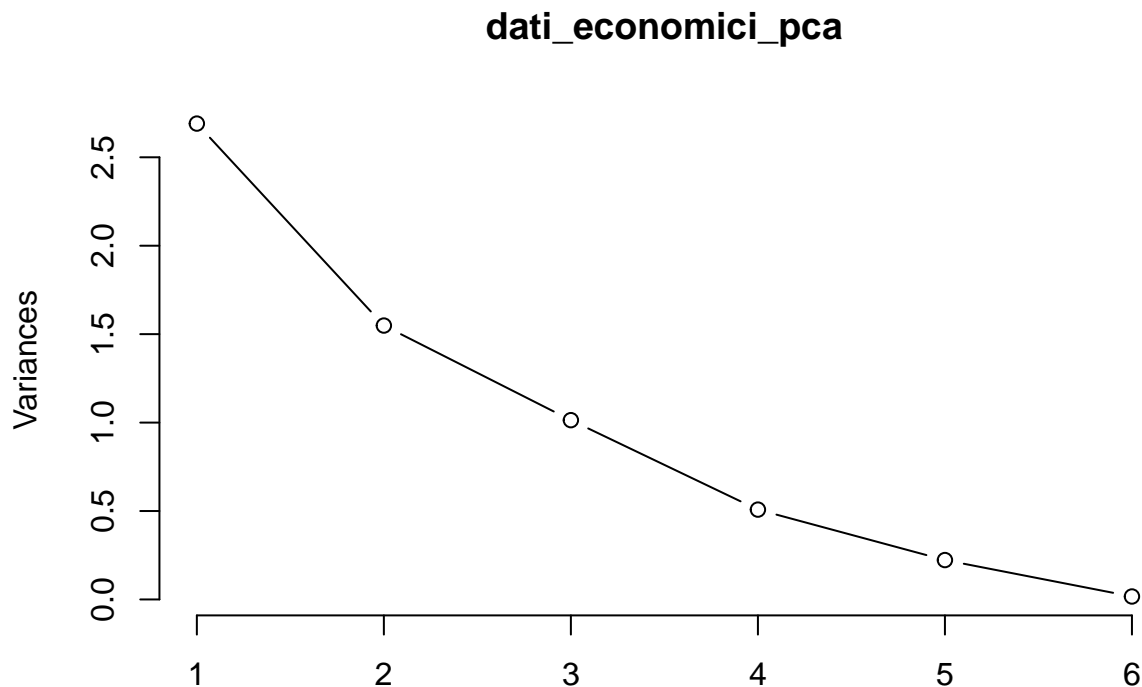
Si osserva che i loadings della PC1 pongono un peso su PIL, POPOLAZIONE e DEBITO. Il loading della PC2, avendo già battezzato PIL nella PC1, dà più peso a DISOCCUPAZIONE e DEBITO PUBBLICO consentendo così di concludere che queste due variabili siano direttamente proporzionali.

```
summary(dati_economici_pca)
```

```
## Importance of components:
##              PC1    PC2    PC3    PC4    PC5    PC6
## Standard deviation  1.6403 1.2444 1.0068 0.71265 0.4718 0.12936
## Proportion of Variance 0.4484 0.2581 0.1689 0.08464 0.0371 0.00279
## Cumulative Proportion 0.4484 0.7065 0.8755 0.96011 0.9972 1.00000
```

Ponendo una PVE (proportion of variance explained) accettabile pari o superiore a 80%, si dovranno considerare per l'eventuale regressione le prima tre componenti principali (PC1,PC2,PC3) Tabella indicativa della percentuale di varianza spiegata.

```
screeplot(dati_economici_pca, type = c("lines")) #elbow non visibile
```



```
options(scipen = 20) #annotazione scientifica a partire da 20 cifre
```

Scores

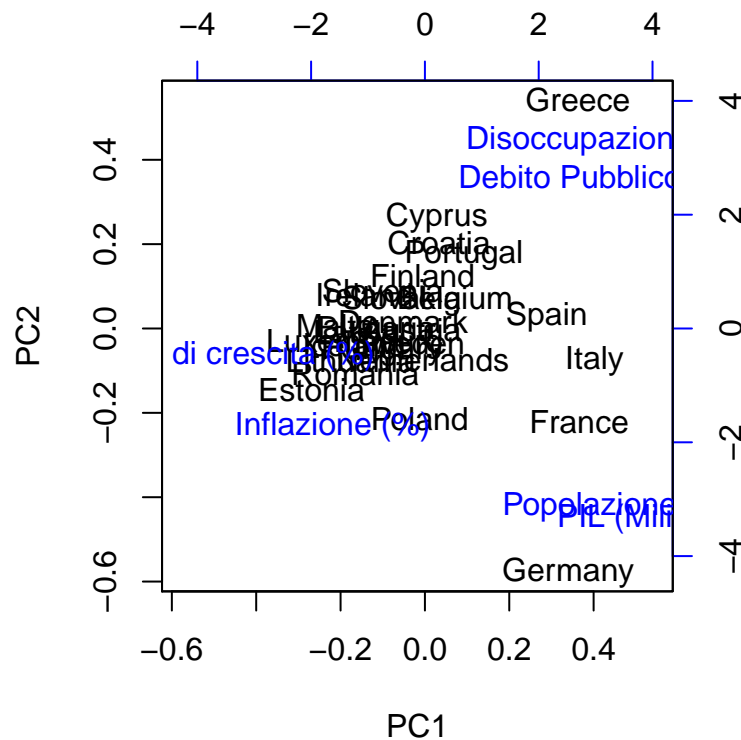
```
scores<- dati_economici_pca$x
scores
```

##	PC1	PC2	PC3	PC4	PC5
## Austria	-0.26384966	-0.02452782	0.69085641	-0.33431227	0.67241473
## Belgium	0.60915106	0.42064049	1.10390594	-0.54264220	0.79812826
## Bulgaria	-0.98010210	-0.03252865	-0.43443289	-0.73956811	-0.70094797
## Croatia	0.28715058	1.32395122	-0.09566657	-0.39084809	-0.42303027
## Cyprus	0.24098206	1.70558063	-0.85253843	-0.02530791	-0.01054289
## Denmark	-0.43986073	0.09939778	-0.11407156	-1.48133481	-0.48069469
## Estonia	-2.29121238	-0.93845759	1.76580028	0.93003883	-0.24398874
## Finland	-0.04628796	0.80712837	-0.58860214	-0.78205861	-0.44564691
## France	3.11711366	-1.42325224	-0.35799192	-0.15289271	-0.15164982
## Germany	2.89348158	-3.74025193	-0.50985229	-0.06593292	0.06269590
## Greece	3.09199228	3.50988214	0.65241991	0.63716486	-0.01855525
## Hungary	-0.86518049	-0.33998839	0.51756662	0.09122490	0.79779317
## Iceland	-1.43358804	-0.25665158	0.06169610	-0.78939684	0.31462403
## Ireland	-1.22187041	0.51241992	-2.70238291	1.01974101	0.13323622
## Italy	3.42870821	-0.49396323	-0.03583217	0.03577422	0.31016231
## Latvia	-1.39559791	-0.04454077	1.16923782	0.79135867	-0.38020677
## Lithuania	-1.49985435	-0.49368273	2.12760071	0.70817991	0.04656089
## Luxembourg	-1.42131086	-0.23234736	0.53057639	-0.48222369	-0.37536336
## Malta	-1.81409624	0.03805858	-1.33952063	0.55351208	0.39195303
## Netherlands	-0.04396656	-0.48335302	-0.39774887	-0.75751931	0.08135212
## Poland	-0.10500123	-1.38595532	-0.25225180	0.46775205	0.12730202
## Portugal	0.78876738	1.12920377	0.25322147	-0.13025057	0.93436417
## Romania	-1.40701272	-0.68723229	-1.54769043	0.96005076	-0.11738034
## Slovakia	-0.47718348	0.48214224	-0.20375300	-0.46191218	-0.44327258
## Slovenia	-0.85671630	0.57983665	-0.65858275	0.26250100	0.39168238
## Spain	2.46073966	0.19049558	0.60844194	1.43715294	-0.91248433
## Sweden	-0.35539505	-0.22200443	0.60959478	-0.75825100	-0.35850531
##	PC6				
## Austria	-0.0466018721				
## Belgium	-0.0322390315				
## Bulgaria	0.1235776360				
## Croatia	0.0004703589				
## Cyprus	-0.0740857142				
## Denmark	0.0263789035				
## Estonia	-0.0662969122				
## Finland	-0.0126050275				
## France	0.0189022697				
## Germany	-0.3049917043				
## Greece	-0.0698665864				
## Hungary	0.0796765841				
## Iceland	0.0274467228				
## Ireland	-0.1990372991				
## Italy	0.2255517721				
## Latvia	-0.0708756137				
## Lithuania	-0.0547298129				

```
## Luxembourg -0.0287553200
## Malta -0.0500804645
## Netherlands -0.0858120316
## Poland 0.3742175526
## Portugal 0.0471527524
## Romania 0.2100014627
## Slovakia 0.0180727297
## Slovenia -0.0421842243
## Spain 0.0303497244
## Sweden -0.0436368545
```

Grafico di confronto PC1, PC2 e relativi scores

```
biplot(dati_economici_pca, col=c("black", "blue"), arrow.len=0, var.axes=FALSE)
```



Osservando il grafico si nota come Popolazione e PIL hanno lo stesso peso sia in PC1 sia in PC2. All'esatto centro del grafico c'è un evidente cluster.

La Grecia è molto vicina a disoccupazione e debito pubblico, come evidenziato dal punteggio elevato (score) sulla PC2, e diagonalmente molto distante dalla variabile tasso di crescita. Si conclude che la Grecia sia una nazione in evidente crisi.