1. **ESEMPIO DI PCA**

(nota: sono valori assoluti, non percentualizzati al numeri di abitanti e sono espressi in quintali)

A <- read.table ('clipboard', header=TRUE)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | cereali | tuberi | ortaggi | frutta | agrumi | vite | olivo |
| Abru | 3590855 | 1751220 | 5449426 | 708934 | 560 | 3365847 | 1294269 |
| Basi | 4363879 | 62500 | 3708553 | 1641121 | 1559906 | 390346 | 364225 |
| Cala | 1739823 | 662162 | 5519088 | 1452920 | 14683931 | 741594 | 10077833 |
| Camp | 3924660 | 3003977 | 10458024 | 7590865 | 704345 | 2496557 | 2405586 |
| EmRo | 28225968 | 2270305 | 23975740 | 15426392 | 0 | 8595036 | 69929 |
| FrVG | 9025287 | 205466 | 168715 | 394817 | 0 | 1548909 | 2274 |
| Lazi | 4623582 | 733404 | 6841368 | 2694107 | 71986 | 2642408 | 1716086 |
| Ligu | 18120 | 94820 | 230370 | 41257 | 4880 | 127165 | 180400 |
| Lomb | 32935561 | 290328 | 6959450 | 912570 | 0 | 1876437 | 45364 |
| Marc | 6804551 | 137124 | 1947801 | 256951 | 0 | 0 | 231802 |
| Moli | 2080439 | 136160 | 919430 | 215050 | 0 | 457307 | 449975 |
| Piem | 20721193 | 475128 | 2508111 | 4382869 | 0 | 3876660 | 824 |
| Pugl | 10020854 | 797250 | 28828207 | 2052407 | 2782495 | 18336000 | 12139400 |
| Sard | 1113011 | 469177 | 3242480 | 543111 | 853664 | 1329089 | 498888 |
| Sici | 8802313 | 2236681 | 11053669 | 3467058 | 18483078 | 10624445 | 3361878 |
| Tosc | 5384898 | 1167367 | 3288339 | 828962 | 955 | 4023110 | 1174819 |
| TrAA | 20535 | 219900 | 113826 | 15099290 | 0 | 1631673 | 10500 |
| Umbr | 7091903 | 77400 | 1725350 | 91299 | 0 | 1296170 | 409179 |
| ValA | 1665 | 22000 | 0 | 39450 | 0 | 45000 | 0 |
| Vene | 31121694 | 1233404 | 5441741 | 4449690 | 0 | 11464375 | 77865 |

PCA = princomp(A)

biplot(PCA)

Standardizziamo.

AA = matrix(nrow=20,ncol=7)

for ( i in 1:7) {

AA[,i] = ( A[,i] - mean(A[,i]) ) / ( sd(A[,i]) )

}

PCAA = princomp(AA)

biplot(PCAA)

- - - - - - - - - - -

AA =A

for ( i in 1:7) {

AA[,i] = ( A[,i] - mean(A[,i]) ) / ( sd(A[,i]) )

}

PCAA = princomp(AA)

biplot(PCAA)

Ora funziona!

**Esercizio**. I) normalizzare la tabella al numero di abitanti. Si può usare questo vettore che offre una misura approssimata della popolazione regionale:

pop<-c(1.3,0.6,1.9,6,4.4,1.2,5.6,1.6,9.9,1.5,0.3,4.4,4.1,1.6,5,3.7,1,0.9,0.1,4.9)

ii) ripetere l’analisi PCA evidenziando alcune differenze nei risultati.

1. **ARRICCHIMENTO GRAFICO**

> summary(PCAA)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7

St dev 1.693183 1.2520744 0.9412341 0.77244868 0.65850498 0.44655767 0.3159122

Prop Var 0.431108 0.2357429 0.1332213 0.08972586 0.06520734 0.02998703 0.0150076

Cum Pro 0.431108 0.6668509 0.8000722 0.88979803 0.95500537 0.98499240 1.0000000

plot(c(0, 0.431108, 0.6668509, 0.8000722, 0.88979803, 0.95500537, 0.98499240, 1.0000000),type="b")

varianza.cumulativa=c(0, 0.431108, 0.6668509, 0.8000722, 0.88979803, 0.95500537, 0.98499240, 1.0000000)

plot(varianza.cumulativa,type="b")

1. **PUNTEGGI E CLASSIFICHE**

Per fare la singola classifica lungo una componente principale, es. la seconda (di solito si fa rispetto alla prima, farlo per esercizio), eseguiamo:

Q = cov(AA)

e2 = eigen(Q) $vector[,2]

e2

[1] -0.48466279 -0.14118643 -0.03382209 -0.44328597 0.50798079 -0.01338213 0.53786469

> sum(e2^2)

1

Ora bisogna moltiplicare le righe della tavola AA (cioè gli individui, le regioni) per e2

> AA%\*%e2

Errore in AA %\*% e2 : requires numeric/complex matrix/vector arguments

Il problema è che A è una tavola, non una matrice e gli usuali comandi di moltiplicazione vettoriale non funzionano. Bisogna tradurla in matrice. Basta introdurre:

AAA = as.matrix(AA)

AAA%\*%e2

Rifacciamo tutto con e1:

e1 = eigen(Q) $vector[,1]

AAA%\*%e1

Il risultato non ci piace per via dei segni scambiati:

e1 = eigen(Q) $vector[,1]

class = -AAA%\*%e1

class

Vediamo infine la classifica:

class[order(-class[,1]),]

trovare in rete comandi migliori per ordinare.

1. **TUTTI I PUNTEGGI**

predict(PCAA)

(troppi dati)

class = -predict(PCAA)[,1]

class

sort(class)

E’ più semplice!

1. **LOADINGS**

PCAA$loadings

Esercizio: provare con l’altra tabella.

1. **ESPLORAZIONE DI TUTTI I PIANI**

C<-predict(PCAA)

i=1; j=2

plot(C[,c(i,j)],type="n",asp=1)

text(C[,c(i,j)],labels=as.character(row.names(AA)))

Questo caso ora è in scala. Provare a ragionare piano per piano.

Mettiamo le esplorazioni insieme al biplot

par(mfrow=c(1,2))

i=2; j=3

biplot(PCAA)

plot(C[,c(i,j)],type="n",asp=1)

text(C[,c(i,j)],labels=as.character(row.names(AA)))

1. **ESPLORAZIONE A COLORI**

**Esercizio**: arricchire la tabella di una colonna di colori (usare i numeri corrispondenti).

Poi usare comandi del tipo

i=1; j=2

plot(C[,c(i,j)],type="p",asp=1)

text(C[,c(i,j)],labels=as.character(row.names(A)),col=A.col[,6])

1. **ESPLORAZIONE TRIDIMENSIONALE**

Bisogna innanzi tutto scaricare il package scatterplot3D. Si veda l’Appendice qui sotto per le istruzioni.

require(scatterplot3d)

dopo il quale funzionano i suoi comandi. Ecco un esempio:

scatterplot3d(C[,1], C[,2], C[,3])

scatterplot3d(C[,1], C[,2], C[,3], pch=row.names(A) )

Ecco un’altra opzione grafica

scatterplot3d(C[,1], C[,2], C[,3])

scatterplot3d(C[,1], C[,2], C[,3], pch=row.names(AA), type="h")

Provare con scatterplot3d(C[,1], C[,2], C[,3], text=row.names(A) )

1. **DA FARE PER LA VOLTA SUCCESSIVA**

Bisogna scaricare i pacchetti rgl, misc3d, plot3D, plot3Drgl. Una volta scaricati, si provi coi comandi

require(rgl)

require(misc3d)

require(plot3D)

require(plot3Drgl)

se l’installazione è avvenuta correttamente.

1. **APPENDICE: SCARICAMENTO PACCHETTI AGGIUNTIVI**

Primo modo (ad es. per scatterplot3D): dopo averlo trovato alla pag.

<http://cran.r-project.org/web/packages/scatterplot3d/index.html>

si esegue il download. Ad es. chi ha Windows può scaricare

|  |  |
| --- | --- |
| Windows binaries: | r-release: [scatterplot3d\_0.3-35.zip](http://cran.r-project.org/bin/windows/contrib/3.1/scatterplot3d_0.3-35.zip) |

Poi, si inserisce l’intera cartella (decompressa, ma forse questo non è necessario) in library. Chiudere R e ripartire (può darsi che questa operazione non sia necessaria). A questo punto è apribile.

Secondo modo: usare la tendina “Pacchetti” dalla pagina di R, da cui si può o installare da zip locali (saltando quindi metà delle operazioni precedenti) oppure usare da subito “installa pacchetti” che reca ad un indirizzo a scelta, e. Padova, dove poi si trovano tutti i pacchetti.