Analisi Fattoriale

Contents

Dati Esami

1. Si consideri la seguente matrice di correlazione

```
     Gaelic English History Arithmetic Algebra Geometry
[1,]  1.000   0.439   0.410      0.288   0.329    0.248
[2,]  0.439   1.000   0.351      0.354   0.320    0.329
[3,]  0.410   0.351   1.000      0.164   0.190    0.181
[4,]  0.288   0.354   0.164      1.000   0.595    0.470
[5,]  0.329   0.320   0.190      0.595   1.000    0.464
[6,]  0.248   0.329   0.181      0.470   0.464    1.000
```

relativa ai voti di $n = 220$ studenti maschi nelle materie *Gaelic*, *English*, *History*, *Arithmetic*, *Algebra* e *Geometry* ($p = 6$). Interpretare la matrice di correlazione.

I voti tra le materie sono correlati positivamente, ovvero gli studenti che vanno bene in una specifica materia vanno bene anche nelle altre. Le correlazioni più forti si trovano nel gruppo delle materie matematiche, e in misura leggermente inferiore, nel gruppo delle materie umanistiche, mentre c'è meno correlazione tra i due gruppi di materie.

2. Stimare con il metodo della massima verosimiglianza il modello fattoriale con $k = 2$ fattori (funzione `factanal()`, argomenti `factors=2` e `rotation="none"`), ottenendo

- la stima della matrice dei pesi fattoriali, interpretando i due fattori ottenuti;
- la stima delle comunalità e delle varianze specifiche, interpretando la bontà del modello;
- la differenza tra R e $\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$.

```r
n = 220
# modello fattoriale con 2 fattori senza rotazione
af <- factanal(covmat=R, factors=2, rotation="none", n.obs=n)
af
```

```
Call:
factanal(factors = 2, covmat = R, n.obs = n, rotation = "none")

Uniquenesses:
    Gaelic     English     History Arithmetic     Algebra    Geometry
     0.510       0.594       0.644      0.377       0.431       0.628

Loadings:
     Factor1 Factor2
```

```
[1,]  0.553   0.429
[2,]  0.568   0.288
[3,]  0.392   0.450
[4,]  0.740  -0.273
[5,]  0.724  -0.211
[6,]  0.595  -0.132

               Factor1 Factor2
SS loadings      2.209   0.606
Proportion Var   0.368   0.101
Cumulative Var   0.368   0.469

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 2.33 on 4 degrees of freedom.
The p-value is 0.674
```

```r
# stima dei pesi fattoriali
Lambda <- af$loadings[,]
round(Lambda,3)
```

```
     Factor1 Factor2
[1,]   0.553   0.429
[2,]   0.568   0.288
[3,]   0.392   0.450
[4,]   0.740  -0.273
[5,]   0.724  -0.211
[6,]   0.595  -0.132
```

```r
# primo fattore "andare bene a scuola"
# secondo fattore "math vs non-math"

# stima delle comunalità
h2 <- apply(af$loadings^2,1,sum)
round(h2,3)
```

```
[1] 0.490 0.406 0.356 0.623 0.569 0.372
```

```r
# stima delle varianze specifiche
Psi <- diag(af$uniquenesses)
round(diag(Psi),3)
```

```
[1] 0.510 0.594 0.644 0.377 0.431 0.628
```

```r
# differenza
fit = Lambda%*%t(Lambda) + Psi
round( R - fit, 4)
```

```
      Gaelic English History Arithmetic Algebra Geometry
[1,]  0.0000  0.0011  0.0002    -0.0048  0.0190  -0.0250
[2,]  0.0011  0.0000 -0.0016     0.0120 -0.0303   0.0287
[3,]  0.0002 -0.0016  0.0000    -0.0036  0.0012   0.0068
[4,] -0.0048  0.0120 -0.0036     0.0000  0.0014  -0.0067
[5,]  0.0190 -0.0303  0.0012     0.0014  0.0000   0.0052
[6,] -0.0250  0.0287  0.0068    -0.0067  0.0052   0.0000
```

3. Stimare il modello fattoriale con $k = 2$ fattori eseguendo la rotazione `varimax` della matrice dei pesi fattoriali (argomento `rotation="varimax"`). Confrontare le stime delle varianze specifiche e della matrice dei pesi fattoriali con la soluzione senza rotazione.

```r
# rotazione varimax
af2 <- factanal(covmat = R, factors=2, rotation="varimax", n.obs=n) # default

# le varianze specifiche non cambiano
round(af2$uniquenesses,3)
```

```
    Gaelic     English    History Arithmetic    Algebra   Geometry
     0.510      0.594      0.644      0.377      0.431      0.628
```

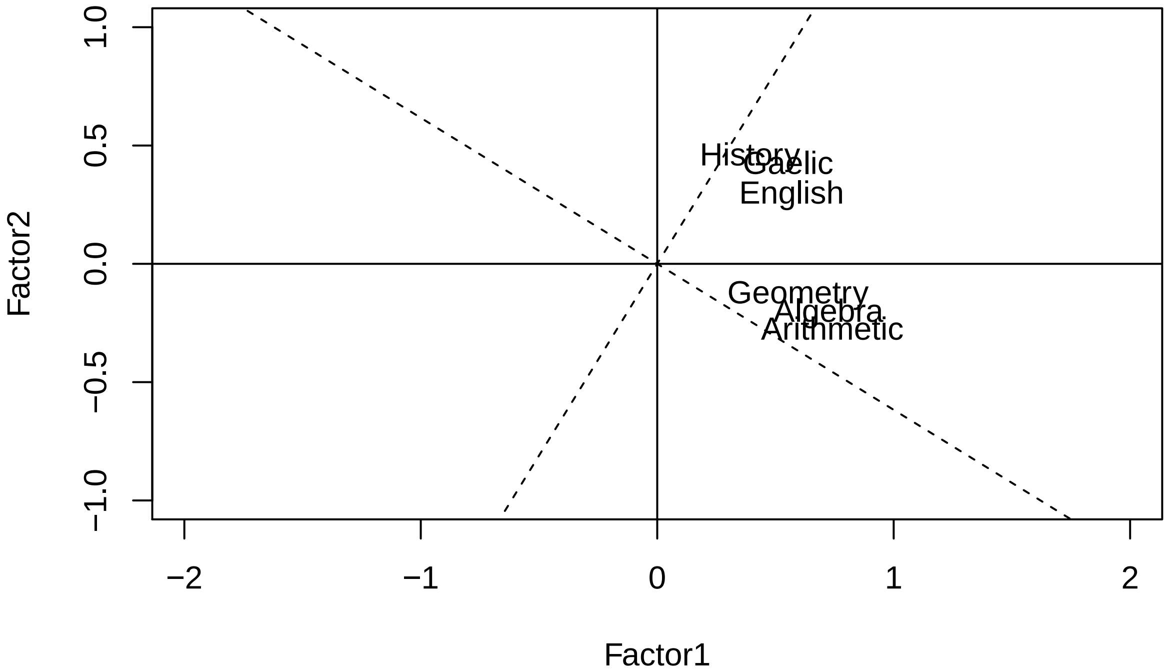
```r
# matrice di rotazione
af2$rotmat
```

```
           [,1]      [,2]
[1,]  0.8420397 0.5394156
[2,] -0.5394156 0.8420397
```

```r
# la matrice dei pesi fattoriali cambia
af2$loadings
```

```

Loadings:
     Factor1 Factor2
[1,] 0.235   0.659
[2,] 0.323   0.549
[3,]         0.590
[4,] 0.771   0.170
[5,] 0.724   0.213
[6,] 0.572   0.210

               Factor1 Factor2
SS loadings      1.612   1.203
Proportion Var   0.269   0.201
Cumulative Var   0.269   0.469
```

```r
# primo fattore "math"
# secondo fattore "non-math"

# rappresentazione grafica della rotazione
plot(af$loadings,xlim=c(-1,1), ylim=c(-1,1), pch="", asp=1)
text(af$loadings, colnames(R))
abline(h=0)
abline(v=0)
af2 <- varimax(af$loadings,normalize=FALSE)
abline(0, af2$rotmat[2,1]/af2$rotmat[1,1], lty=2)
abline(0, af2$rotmat[2,2]/af2$rotmat[1,2], lty=2)
```



3. Stimare il modello fattoriale con $k = 2$ con il metodo dei fattori principali:

- Calcolare la stima iniziale delle comunalità $\hat{h}_i^2$ come

$$\hat{h}_i^2 = 1 - 1/r^{ii}$$

  , dove r^{ii} è l'elemento di posizione (i, i) della matrice R^{-1}
- Ottenere la matrice di correlazione ridotta

$$R^* = R - \hat{\Psi}$$

  dove $\hat{\Psi} = diag(\hat{\psi}_1, \ldots, \hat{\psi}_p)$ con $\hat{\psi}_i = 1 - \hat{h}_i^2$
- Calcolare la stima della matrice dei pesi fattoriali

$$\hat{\Lambda} = V_k L_k^{1/2}$$

  dopo aver ottenuto la decomposizione spettrale di $R^* = VLV'$ e dove $\underset{p \times k}{V_k}$ contiene le prime k colonne di V e $\underset{k \times k}{L_k} = diag(l_1, \ldots, l_k)$ le prime k colonne di L.
- calcolare la differenza tra R e $\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$ e confrontarla con quanto ottenuto dalla stima di massima verosimiglianza.

```
# calcolare R^-1
invR =  solve(R)

# stima iniziale delle comunalità
h2<-1-1/(diag(solve(R)))
round(h2,3)
```

```
[1] 0.300 0.297 0.206 0.420 0.418 0.295
```

```
# stima di Psi
Psi = diag(1-h2)

# matrice di correlazione ridotta
```

```r
Rstar = R - Psi

# stima della matrice dei pesi fattoriali
eigen = eigen(Rstar)
k = 2
Lambda <- eigen$vectors[,1:k] %*% diag(eigen$values[1:k]^{1/2})
Lambda
```

```
           [,1]       [,2]
[1,] -0.5570353 -0.3089276
[2,] -0.5836366 -0.2181057
[3,] -0.4164361 -0.3569268
[4,] -0.6723021  0.2676743
[5,] -0.6763102  0.2412251
[6,] -0.5824034  0.1801402
```

```r
# aggiorno stima delle comunalità
h2 = apply(Lambda^2,1,sum)
Psi = diag(1-h2)

# differenza
fit = Lambda%*%t(Lambda) + Psi
round( R - fit, 4)
```

```
      Gaelic English History Arithmetic Algebra Geometry
[1,]  0.0000  0.0465  0.0678    -0.0038  0.0268  -0.0208
[2,]  0.0465  0.0000  0.0301     0.0200 -0.0221   0.0284
[3,]  0.0678  0.0301  0.0000    -0.0204 -0.0055   0.0028
[4,] -0.0038  0.0200 -0.0204     0.0000  0.0757   0.0302
[5,]  0.0268 -0.0221 -0.0055     0.0757  0.0000   0.0267
[6,] -0.0208  0.0284  0.0028     0.0302  0.0267   0.0000
```

4. Si riporti il p-value del test rapporto di verosimiglianza per verificare l'ipotesi nulla H_0 : '$k = 1$ fattore è sufficiente', e si concluda ad un livello di significatività del 5% se è opportuno utilizzare il modello fattoriale ad 1 fattore.

```r
factanal(covmat=R, factors=1, n.obs = n)$PVAL
```

```
   objective
4.528823e-08
```

5. Si calcolino le statistiche test rapporto di verosimiglianza

$$T = n \log(|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|/|R|)$$

e

$$T_{Bartlett} = ((n-1) - (2p + 4k + 5)/6) \log(|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}|/|R|)$$

e il corrispondente p-value. Infine, determinare c per la regione critica $[c, +\infty)$ al livello $\alpha = 5\%$.

```r
k=1
p=6
af = factanal(covmat=R, factors=k, n.obs = n)
Lambda = af$loadings[,]
Psi = diag(af$uniqueness)
fit = Lambda %*% t(Lambda) + Psi
t = n*log(det(fit)/det(R))
t
```

```
[1] 53.08106
```

```
gdl = ((p-k)^2 - p - k)/2
pchisq(t, lower.tail=FALSE, df=gdl)
```

```
[1] 2.821254e-08
```

```
# regione critica livello 5%
alpha=0.05
c= qchisq(1-alpha, df=gdl)
c
```

```
[1] 16.91898
```

Dati Open/Closed Book Examination Data

1. Caricare il dati `scor` presenti nella libreria `bootstrap` e calcolare la matrice di correlazione $\underset{p\times p}{R}$.

```
library(bootstrap)

# carico i dati:
data(scor)
X <- scor
n <- nrow(X)
p <- ncol(X)

# matrice di correlazione
R <- cor(X)
round(R, 2)
```

```
     mec  vec  alg  ana  sta
mec 1.00 0.55 0.55 0.41 0.39
vec 0.55 1.00 0.61 0.49 0.44
alg 0.55 0.61 1.00 0.71 0.66
ana 0.41 0.49 0.71 1.00 0.61
sta 0.39 0.44 0.66 0.61 1.00
```

Metodo dei fattori principali

2. Calcolare la stima iniziale delle comunalità $\hat{h}_i^2$ come

$$\hat{h}_i^2 = \max_{j\neq i} |r_{ij}|$$

per $i = 1, \ldots, p$, dove r_{ij} è l'elemento di posizione (i, j) della matrice R. Ottenere la matrice di correlazione ridotta

$$R^* = R - \hat{\Psi}$$

dove $\hat{\Psi} = diag(\hat{\psi}_1, \ldots, \hat{\psi}_p)$ con $\hat{\psi}_i = 1 - \hat{h}_i^2$.

```
# sostituisco 1 sulla diagonale di R con 0
R0 <- R - diag(rep(1,p))

# calcolo la stima iniziale della comunalità
h2 <- apply(abs(R0), 2, max)
h2
```

```
      mec       vec       alg       ana       sta 
0.5534052 0.6096447 0.7108059 0.7108059 0.6647357 
```

```r
# calcolo la stima di Psi
Psi = diag(1-h2)
Psi
```

```
          [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.4465948 0.0000000 0.0000000 0.0000000 0.0000000
[2,] 0.0000000 0.3903553 0.0000000 0.0000000 0.0000000
[3,] 0.0000000 0.0000000 0.2891941 0.0000000 0.0000000
[4,] 0.0000000 0.0000000 0.0000000 0.2891941 0.0000000
[5,] 0.0000000 0.0000000 0.0000000 0.0000000 0.3352643
```

```r
# calcolo la matrice di correlazione ridotta R*
Rstar <- R - Psi
round(Rstar,2)
```

```
     mec  vec  alg  ana  sta
mec 0.55 0.55 0.55 0.41 0.39
vec 0.55 0.61 0.61 0.49 0.44
alg 0.55 0.61 0.71 0.71 0.66
ana 0.41 0.49 0.71 0.71 0.61
sta 0.39 0.44 0.66 0.61 0.66
```

3. Considerando il modello fattoriale con $k = 2$ fattori

- calcolare la stima della matrice dei pesi fattoriali

$$\hat{\Lambda} = V_k L_k^{1/2}$$

dopo aver ottenuto la decomposizione spettrale di $R^* = VLV'$ e dove $\underset{p\times k}{V_k}$ contiene le prime k colonne di V e $\underset{k\times k}{L_k} = diag(l_1, \dots, l_k)$.

- aggiornare la stima delle comunalità

$$\hat{h}_i^2 = \sum_{j=1}^{k} \hat{\lambda}_{ij}^2$$

dove $\hat{\lambda}_{ij}$ è l'elemento di posizione (i, j) della matrice $\hat{\Lambda}$ ottenuta al passo precedente

- aggiornare la stima della matrice di correlazione ridotta

$$R^* = R - \hat{\Psi}$$

dove $\hat{\Psi} = diag(\hat{\psi}_1, \dots, \hat{\psi}_p)$ con $\hat{\psi}_i = 1 - \hat{h}_i^2$ dove $\hat{h}_i^2$ è la stima ottenuta al passo precedente

```r
# decomposizione spettrale di R*
eigen <- eigen(Rstar)

#  k=2
k = 2

# stima di Lambda
Lambda <- eigen$vectors[,1:k] %*% diag(eigen$values[1:k]^{1/2})
Lambda
```

```
           [,1]       [,2]
[1,] -0.6459890  0.35354561
```

```
[2,] -0.7125528  0.30300833
[3,] -0.8639382 -0.05134392
[4,] -0.7864129 -0.24856797
[5,] -0.7419269 -0.27558100
```

```
# nuova stima comunalità
h2.new = apply(Lambda^2, 1, sum)
h2.new
```

```
[1] 0.5422963 0.5995455 0.7490254 0.6802313 0.6264004
```

```
# nuova stima di Psi
Psi.new <- diag(1-h2.new)
Psi.new
```

```
          [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.4577037 0.0000000 0.0000000 0.0000000 0.0000000
[2,] 0.0000000 0.4004545 0.0000000 0.0000000 0.0000000
[3,] 0.0000000 0.0000000 0.2509746 0.0000000 0.0000000
[4,] 0.0000000 0.0000000 0.0000000 0.3197687 0.0000000
[5,] 0.0000000 0.0000000 0.0000000 0.0000000 0.3735996
```

```
# nuova stima di R*
Rstar.new = R - Psi.new
Rstar.new
```

```
          mec       vec       alg       ana       sta
mec 0.5422963 0.5534052 0.5467511 0.4093920 0.3890993
vec 0.5534052 0.5995455 0.6096447 0.4850813 0.4364487
alg 0.5467511 0.6096447 0.7490254 0.7108059 0.6647357
ana 0.4093920 0.4850813 0.7108059 0.6802313 0.6071743
sta 0.3890993 0.4364487 0.6647357 0.6071743 0.6264004
```

4. Iterare per 100 volte la procedura descritta al punto 3, ottenendo le stime finali per $\hat{\Lambda}$ e $\hat{\Psi}$. Calcolare la differenza tra R e $\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$ e commentare.

```
for (i in 1:100){
  h2 <- apply(Lambda^2, 1, sum)
  Rstar <- R0 + diag(h2)
  eigen <- eigen(Rstar)
  Lambda <- eigen$vectors[,1:k] %*% diag(eigen$values[1:k]^{1/2})
}

# stima finale per Lambda
Lambda
```

```
           [,1]        [,2]
[1,] -0.6427483  0.34238878
[2,] -0.7081925  0.28687387
[3,] -0.8966056 -0.08718706
[4,] -0.7710248 -0.23504626
[5,] -0.7179803 -0.22818566
```

```
# stima finale per le comunalità
h2 <- apply(Lambda^2, 1, sum)
h2
```

```
[1] 0.5303554 0.5838332 0.8115032 0.6497260 0.5675644
```

```r
# stima finale per Psi
Psi = diag(1-h2)

# differenza
fit = Lambda%*%t(Lambda) + Psi
round( R - fit, 4)
```

```
        mec     vec     alg     ana     sta
mec  0.0000  0.0000  0.0003 -0.0057  0.0057
vec  0.0000  0.0000 -0.0003  0.0065 -0.0066
alg  0.0003 -0.0003  0.0000 -0.0010  0.0011
ana -0.0057  0.0065 -0.0010  0.0000  0.0000
sta  0.0057 -0.0066  0.0011  0.0000  0.0000
```

Stima di Massima Verosimiglianza

5. Utilizzando la funzione `factanal()`, stimare con il metodo della massima verosimiglianza il modello fattoriale con $k = 2$ fattori (argomento `factors=2`) senza eseguire la rotazione della matrice dei pesi fattoriali (argomento `rotation="none"`). Visualizzare la stima della matrice dei pesi fattoriali, delle comunalità e delle varianze specifiche, e calcolare la differenza tra R e $\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$.

```r
# modello fattoriale con 2 fattori senza rotazione
af <- factanal(X, factors=2, rotation="none")

af
```

```

Call:
factanal(x = X, factors = 2, rotation = "none")

Uniquenesses:
  mec   vec   alg   ana   sta 
0.466 0.419 0.189 0.352 0.431 

Loadings:
    Factor1 Factor2
mec  0.628   0.373 
vec  0.695   0.312 
alg  0.899         
ana  0.780  -0.201 
sta  0.727  -0.200 

               Factor1 Factor2
SS loadings      2.824   0.319
Proportion Var   0.565   0.064
Cumulative Var   0.565   0.629

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 0.07 on 1 degree of freedom.
The p-value is 0.785 
```

```r
# stima dei pesi fattoriali
Lambda <- af$loadings[,]
Lambda
```

```
      Factor1    Factor2
mec 0.6283935  0.3731279
vec 0.6953763  0.3120836
alg 0.8994080 -0.0499577
ana 0.7796021 -0.2010654
sta 0.7273443 -0.1998705
```

```r
# stima delle comunalità
h2 <- apply(af$loadings^2,1,sum)
h2
```

```
      mec       vec       alg       ana       sta
0.5341029 0.5809445 0.8114306 0.6482067 0.5689780
```

```r
# stima delle varianze specifiche
Psi <- diag(af$uniquenesses)
Psi
```

```
          [,1]      [,2]      [,3]      [,4]     [,5]
[1,] 0.4658969 0.0000000 0.0000000 0.0000000 0.000000
[2,] 0.0000000 0.4190553 0.0000000 0.0000000 0.000000
[3,] 0.0000000 0.0000000 0.1885694 0.0000000 0.000000
[4,] 0.0000000 0.0000000 0.0000000 0.3517932 0.000000
[5,] 0.0000000 0.0000000 0.0000000 0.0000000 0.431021
```

```r
# differenza
fit = Lambda%*%t(Lambda) + Psi
round( R - fit, 4)
```

```
        mec     vec    alg     ana     sta
mec  0.0000  0.0000  2e-04 -0.0055  0.0066
vec  0.0000  0.0000 -2e-04  0.0057 -0.0070
alg  0.0002 -0.0002  0e+00 -0.0004  0.0006
ana -0.0055  0.0057 -4e-04  0.0000 -0.0001
sta  0.0066 -0.0070  6e-04 -0.0001  0.0000
```

6. Stimare il modello fattoriale con $k = 2$ fattori eseguendo la rotazione `varimax` della matrice dei pesi fattoriali (argomento `rotation="varimax"`). Confrontare le stime delle varianze specifiche e della matrice dei pesi fattoriali con la soluzione senza rotazione.

```r
# rotazione varimax
af2 <- factanal(X, factors=2, rotation="varimax") # default

# le varianze specifiche non cambiano
af2$uniquenesses
```

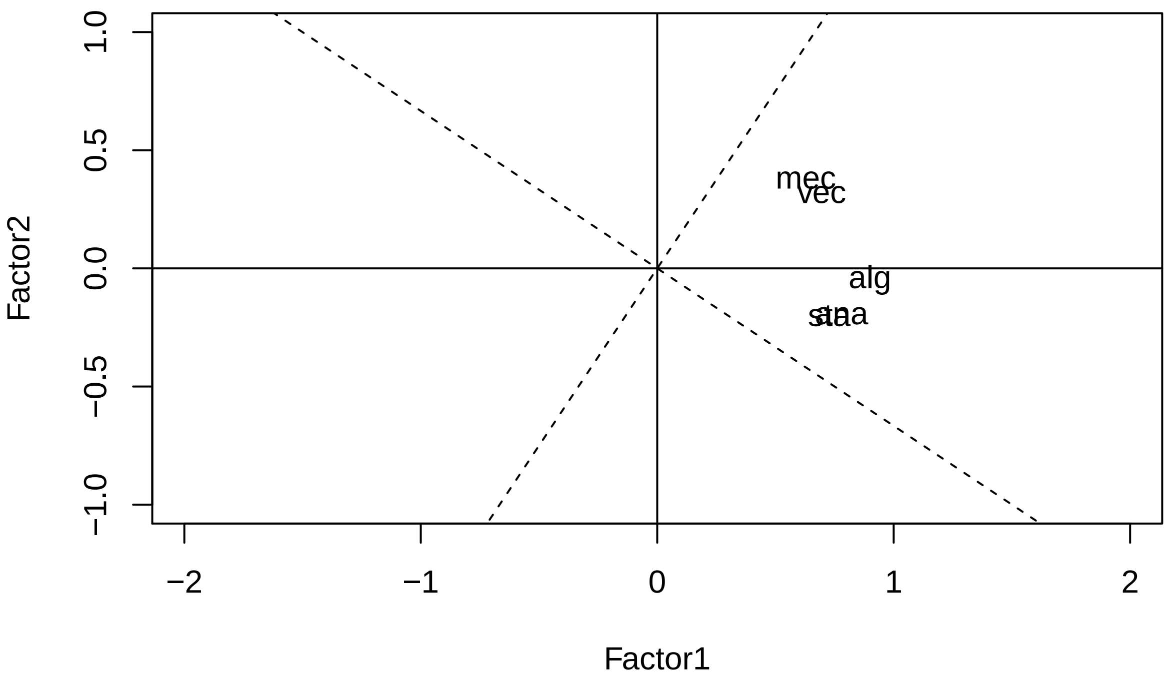
```
      mec       vec       alg       ana       sta
0.4658969 0.4190553 0.1885694 0.3517932 0.4310210
```

```r
# la matrice dei pesi fattoriali cambia
af2$loadings
```

```

Loadings:
    Factor1 Factor2
mec 0.265   0.681
vec 0.356   0.674
alg 0.740   0.514
ana 0.738   0.322
```

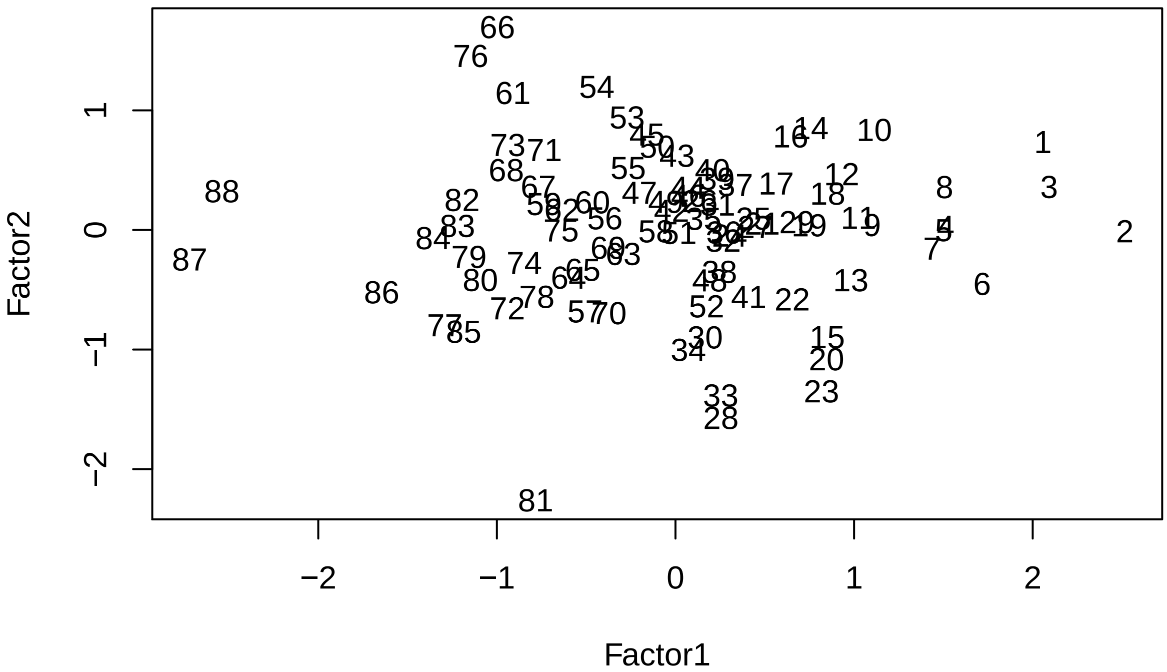
```
sta 0.696   0.290

               Factor1 Factor2
SS loadings      1.774   1.370
Proportion Var   0.355   0.274
Cumulative Var   0.355   0.629
```

```r
# primo fattore "being good at school"
# secondo fattore "ability in closed book vs open book exams"

# rappresentazione grafica della rotazione
plot(af$loadings,xlim=c(-1,1), ylim=c(-1,1), pch="", asp=1)
text(af$loadings, names(X))
abline(h=0)
abline(v=0)

af2 <- varimax(af$loadings,normalize=FALSE)
abline(0, af2$rotmat[2,1]/af2$rotmat[1,1], lty=2)
abline(0, af2$rotmat[2,2]/af2$rotmat[1,2], lty=2)
```



7. Stimare i punteggi fattoriali con il metodo di Thompson e rappresentarli graficamente.

```r
# thompson
punt.t <- factanal(X, factors=2, rotation="none", scores="regression")

# plot dei punteggi fattoriali
plot(punt.t$scores,pch="")
text(punt.t$scores, labels=c(1:88))
```




```r
scor[66,]
```

```
   mec vec alg ana sta
66  59  53  37  22  19
```

```r
scor[81,]
```

```
   mec vec alg ana sta
81   3   9  51  47  40
```

```r
scor[2,]
```

```
  mec vec alg ana sta
2  63  78  80  70  81
```

```r
scor[87,]
```

```
   mec vec alg ana sta
87   5  26  15  20  20
```

Speranza di vita

I dati nella seguente Tabella mostrano l'aspettativa di vita in anni per paese, età e sesso. I dati provengono da Key Tz e Flieger (1971) e riguardano le aspettative di vita negli anni '60.

Per importare i dati:

```r
"life" <-
  structure(.Data = list(c(63., 34., 38., 59., 56., 62., 50., 65., 56., 69., 65., 64., 56., 60., 61., 4
                          64., 67., 61., 68., 67., 65., 59., 58., 57.)
                        , c(51., 29., 30., 42., 38., 44., 39., 44., 46., 47., 48., 50., 44., 44., 45.,
                          43., 45., 40., 46., 45., 46., 43., 44., 46.)
                        , c(30., 13., 17., 20., 18., 24., 20., 22., 24., 24., 26., 28., 25., 22., 22.,
                          21., 23., 21., 23., 23., 24., 23., 24., 28.)
                        , c(13., 5., 7., 6., 7., 7., 7., 7., 11., 8., 9., 11., 10., 6., 8., 9., 6., 8.
```

```
                        8., 9., 10., 9., 9.)
                    , c(67., 38., 38., 64., 62., 69., 55., 72., 63., 75., 68., 66., 61., 65., 65.,
                        68., 74., 67., 75., 74., 71., 66., 62., 60.)
                    , c(54., 32., 34., 46., 46., 50., 43., 50., 54., 53., 50., 51., 48., 45., 49.,
                        47., 51., 46., 52., 51., 51., 49., 47., 49.)
                    , c(34., 17., 20., 25., 25., 28., 23., 27., 33., 29., 27., 29., 27., 25., 27.,
                        24., 28., 25., 29., 28., 28., 27., 25., 28.)
                    , c(15., 6., 7., 8., 10., 14., 8., 9., 19., 10., 10., 11., 12., 9., 10., 8., 7
                        10., 10., 10., 12., 10., 11.)
  )
  , class = "data.frame"
  , names = c("m0", "m25", "m50", "m75", "w0", "w25", "w50", "w75")
  , row.names = c("Algeria", "Cameroon", "Madagascar", "Mauritius", "Reunion", "Seychelles", "South Afr
                  "Tunisia", "Canada", "Costa Rica", "Dominican Rep", "El Salvador", "Greenland", "Grena
                  "Honduras", "Jamaica", "Mexico", "Nicaragua", "Panama", "Trinidad(62)", "Trinidad (67)
                  "United States (66)", "United States (NW66)", "United States (W66)", "United States (
                  "Chile", "Columbia", "Ecuador")
  )
knitr::kable(life, "html")
```

m0

m25

m50

m75

w0

w25

w50

w75

Algeria

63

51

30

13

67

54

34

15

Cameroon

34

29

13

5

38

32

17

6

Madagascar

38

30

17

7

38

34

20

7

Mauritius

59

42

20

6

64

46

25

8

Reunion

56

38

18

7

62

46

25

10

Seychelles

62

44

24

7

69

50

28

14

South Africa(C)

50

39

20

7

55

43

23

8

South Africa(W)

65

44

22

7

72

50

27

9

Tunisia

56

46

24

11

63

54

33

19

Canada

69

47

24

8

75

53

29

10

Costa Rica

65

48

26

9

68

50

27

10

Dominican Rep

64

50

28

11

66

51

29

11

El Salvador

56

44

25

10

61

48

27

12

Greenland

60

44

22

6

65

45

25

9

Grenada

61

45

22

8

65

49

27

10

Guatemala

49

40

22

9

51

41

23

8

Honduras

59

42

22

6

61

43

22

7

Jamaica

63

44

23

8

67

48

26

9

Mexico

59

44

24

8

63

46

25

8

Nicaragua

65

48

28

14

68

51

29

13

Panama

65

48

26

9

67

49

27

10

Trinidad(62)

64

63

21

7

68

47

25

9

Trinidad (67)

64

43

21

6

68

47

24

8

United States (66)

67

45

23

8

74

51

28

10

United States (NW66)

61

40

21

10

67

46

25

11

United States (W66)

68

46

23

8

75

52

29

10

United States (67)

67

45

23

8

74

51

28

10

Argentina

65

46

24

9

71

51

28

10

Chile

59

43

23

10

66

49

27

12

Columbia

58

44

24

9

62

47

25

10

Ecuador

57

46

28

9

60

49

28

11

1. Per iniziare, useremo il test per il numero di fattori incorporati nell'approccio alla massima verosimiglianza

```
round(sapply(1:3, function(f)
factanal(life, factors = f, method ="mle")$PVAL
),4)
```

```
objective objective objective
   0.0000    0.0000    0.4578
```

Questi risultati suggeriscono che una soluzione a tre fattori potrebbe essere adeguata per tenere conto per le covarianze osservate nei dati, sebbene debba essere ricordato che, con solo 31 paesi, l'uso di un risultato del test asintotico potrebbe essere piuttosto sospetto.

La soluzione a tre fattori è la seguente (notare che la soluzione è quella risultante da una soluzione varimax. il valore predefinito per la funzione factanal()):

```
factanal(life, factors = 3, method ="mle")
```

```

Call:
factanal(x = life, factors = 3, method = "mle")

Uniquenesses:
   m0   m25   m50   m75    w0   w25   w50   w75
0.005 0.362 0.066 0.288 0.005 0.011 0.020 0.146

Loadings:
    Factor1 Factor2 Factor3
m0  0.964   0.122   0.226
m25 0.646   0.169   0.438
m50 0.430   0.354   0.790
m75         0.525   0.656
w0  0.970   0.217
w25 0.764   0.556   0.310
w50 0.536   0.729   0.401
w75 0.156   0.867   0.280
```

```
                Factor1 Factor2 Factor3
SS loadings      3.375   2.082   1.640
Proportion Var   0.422   0.260   0.205
Cumulative Var   0.422   0.682   0.887

Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 6.73 on 7 degrees of freedom.
The p-value is 0.458
```

Esaminando i pesi stimati dei fattori, vediamo che

- il primo fattore è dominato dall'aspettativa di vita alla nascita (sia maschi che femmine); potrebbe essere etichettato *forza vitale alla nascita*
- il secondo fattore secondo fattore misura sostanzialmente l'aspettativa di vita per le donne anziane; potrebbe essere etichettato *forza vitale per le donne anziane*
- il terzo fattore ha pesi più elevati per le aspettative di vita degli uomini di età compresa tra 50 e 75 anni: potrebbe essere etichettato *forza vitale per gli uomini anziani*

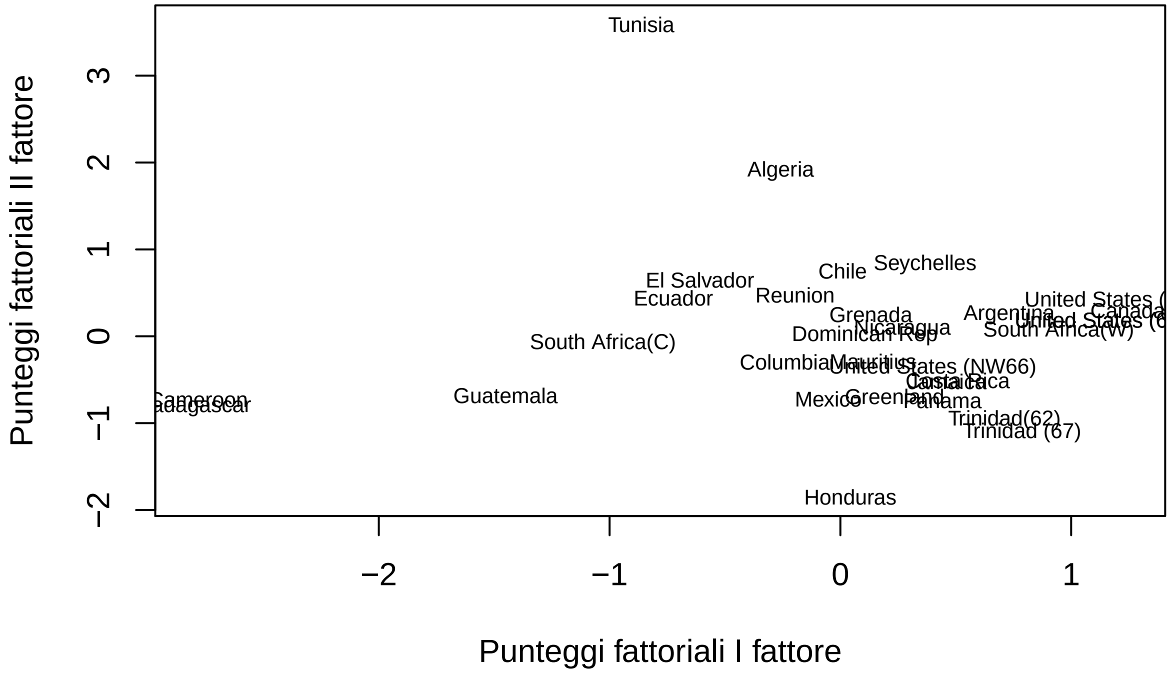
Le stime dei punteggi fattoriali secondo il metodo di Thomson sono

```r
scores <- factanal(life, factors = 3, method = "mle", scores = "regression")$scores
scores
```

```
                         Factor1     Factor2     Factor3
Algeria              -0.258062561  1.90095771  1.91581631
Cameroon             -2.782495791 -0.72340014 -1.84772224
Madagascar           -2.806428187 -0.81158820 -0.01210318
Mauritius             0.141004934 -0.29028454 -0.85862443
Reunion              -0.196352142  0.47429917 -1.55046466
Seychelles            0.367371307  0.82902375 -0.55214085
South Africa(C)      -1.028567629 -0.08065792 -0.65421971
South Africa(W)       0.946193522  0.06400408 -0.91995289
Tunisia              -0.862493550  3.59177195 -0.36442148
Canada                1.245304248  0.29564122 -0.27342781
Costa Rica            0.508736247 -0.50500435  1.01328707
Dominican Rep         0.106044085  0.01111171  1.83871599
El Salvador          -0.608155779  0.65100820  0.48836431
Greenland             0.235114220 -0.69123901 -0.38558654
Grenada               0.132008172  0.25241049 -0.15220645
Guatemala            -1.450336359 -0.67765804  0.65911906
Honduras              0.043253249 -1.85175707  0.30633182
Jamaica               0.462124701 -0.51918493  0.08032855
Mexico               -0.052332675 -0.72020002  0.44417800
Nicaragua             0.268974443  0.08407227  1.70568388
Panama                0.442333434 -0.73778272  1.25218728
Trinidad(62)          0.711367053 -0.95989475 -0.21545329
Trinidad (67)         0.787286051 -1.10729029 -0.51958264
United States (66)    1.128331259  0.16389896 -0.68177046
United States (NW66)  0.400058903 -0.36230253 -0.74299137
United States (W66)   1.214345385  0.40877239 -0.69225320
United States (67)    1.128331259  0.16389896 -0.68177046
Argentina             0.731344988  0.24811968 -0.12817725
Chile                 0.009751528  0.75222637 -0.49198911
Columbia             -0.240602517 -0.29543613  0.42919600
Ecuador              -0.723451797  0.44246371  1.59164974
```

Possiamo usare i punteggi per interpretare i dati

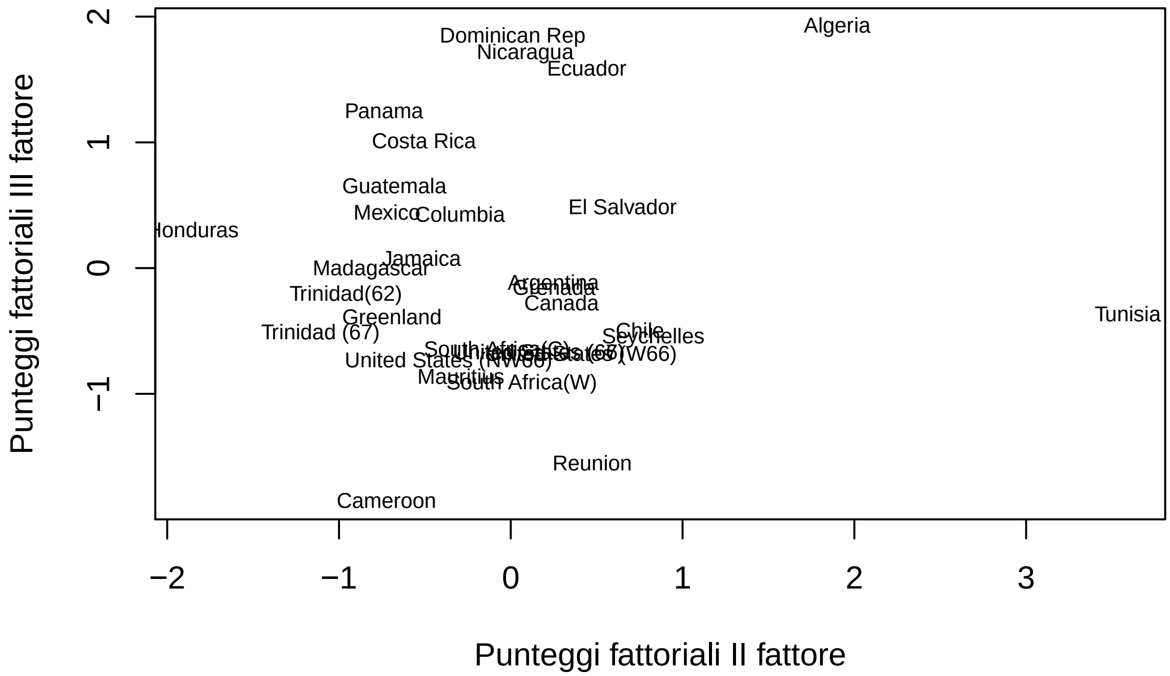
```
plot(scores[,1], scores[,2], type='n', xlab='Punteggi fattoriali I fattore', ylab='Punteggi fattoriali I
text(scores[,1], scores[,2], labels = row.names(life), cex = 0.7 )
```



Vediamo che il Madagascar e il Camerun hanno i valori più bassi per il primo fattore, mentre gli Stati Uniti e il Canada hanno valori elevati.

Il secondo fattore misura sostanzialmente l'aspettativa di vita per le donne anziane. Tunisia ha un punteggio elevato, mentre Honduras basso per questo fattore.

```
plot(scores[,2], scores[,3], type='n', xlab='Punteggi fattoriali II fattore', ylab='Punteggi fattoriali
text(scores[,2], scores[,3], labels = row.names(life), cex = 0.7 )
```



Il terzo fattore riflette (principalmente) l'aspettativa di vita per gli uomini più anziani. Il Camerun ha il punteggio più basso per questo fattore, mentre L'Algeria ha il punteggio più alto.