**12 Febbraio 2018 - Analisi Esplorativa**

Cognome: ...............................................

Nome: ...................................................

Matricola: ...............................................

Tipologia d'esame: ☐ 12 CFU ☐ 15 CFU

---

**Prova scritta**

*Si svolgano gli esercizi riportando il risultato dove indicato. Durata: 60 minuti*

---

**Esercizio 1 (5 punti)**

Sapendo che

- la prima riga di $\underset{n\times 3}{X}$ è pari a $\underset{1\times 3}{x_1'} = (9,\ 12,\ 22)$ e il baricentro è pari a $\underset{1\times 3}{\bar{x}'} = (8.05,\ 4.1,\ 13.74)$
- la seconda colonna di $\underset{3\times 3}{V}$ è $v_2 = (0.74, -0.67, -0.08)'$, dove $\underset{3\times 3}{V}$ è la matrice degli autovettori di $\underset{3\times 3}{S}$
- $s_{11} = 89.9$ e il secondo autovalore di $\underset{3\times 3}{S}$ è $\lambda_2 = 1$

si risponda alle seguenti domande:

a. Riportare il valore del primo elemento di $\underset{n\times 1}{y_2} = \tilde{X}v_2$, arrotondando al **secondo decimale** : .........
   dove $\tilde{X}$ è la matrice dei dati centrati

```
xbar = matrix(c(8.05,4.1,13.74), ncol=1)
u1 = matrix(c(9,12,22), nrow=1)
v2 = matrix(c(0.74,-0.67,-0.08), ncol=1)
y12 = (u1-t(xbar)) %*% v2
round(y12, 2)
```

```
##        [,1]
## [1,] -5.25
```

b. Si calcoli il coefficiente di correlazione tra la prima colonna $\underset{n\times 1}{\tilde{x}_1}$ di $\underset{n\times 3}{\tilde{X}}$ e i punteggi $\underset{n\times 1}{y_2}$ della seconda componente principale, arrotondando il calcolo al **secondo decimale**:

$$\mathbb{C}\mathrm{orr}(\tilde{x}_1, y_2) = \ldots\ldots$$

c. Si calcoli $d_\infty(x_1, \bar{x}) = \ldots\ldots$

```
# b
s11 = 89.9
lambda2=1
round( v2[1] * sqrt(lambda2) / sqrt(s11) , 2)
```

```
## [1] 0.08
```

```
# c
max(abs(u1-t(xbar)))
```

```
## [1] 8.26
```

**Esercizio 2 (9 punti)**

Si consideri la seguente matrice di correlazione $\underset{3\times 3}{R} = \begin{bmatrix} 1 & 0.83 & 0.78 \\ & 1 & 0.67 \\ & & 1 \end{bmatrix}$.

a. Sulla base di $R$, si consideri il modello fattoriale ad un fattore $\underset{p\times 1}{x} = \underset{p\times 1}{\Lambda}\underset{1\times 1}{f} + \underset{p\times 1}{u}$ dove $\Lambda' = (\lambda_1, \lambda_2, \lambda_3)$. Nel *metodo di stima naive*, partendo da $R = \Lambda\Lambda' + \Psi$, si ottiene il seguente sistema di equazioni $\hat{\lambda}_1\hat{\lambda}_2 = 0.83, \hat{\lambda}_3\hat{\lambda}_2 = 0.67, \hat{\lambda}_1\hat{\lambda}_3 = 0.78, \hat{\psi}_i = 1 - \hat{\lambda}_i^2, i = 1, 2, 3$. Risolvendo il sistema di equazioni, si ottiene $\hat{\lambda}_1^2 = r_{12}r_{13}/r_{23}$, dove $r_{ij}$ è l'elemento di $R$ in posizione $(i, j)$. Si riportino, arrotondando al **secondo decimale**:

$$\hat{\lambda}_1 = \ldots\ldots\ldots, \hat{\lambda}_2 = \ldots\ldots\ldots, \hat{\lambda}_3 = \ldots\ldots\ldots, \hat{\psi}_1 = \ldots\ldots\ldots, \hat{\psi}_2 = \ldots\ldots\ldots, \hat{\psi}_3 = \ldots\ldots\ldots$$

b. Il numero di parametri corrispondenti al modello fattoriale ad un fattore (senza vincoli) è pari a .........

c. Sia $z_1' = (1, 2, 3)$ la prima riga della matrice dei dati standardizzati $\underset{n\times p}{Z}$. Riportare, arrotondando al **secondo decimale**, la stima del punteggio fattoriale con il *metodo di Bartlett*

$$\hat{f}_1 = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Psi}^{-1}z_1 = \ldots\ldots\ldots$$

d. Si spieghi cos'è un caso di Heywood.

```r
R=matrix(c(1,.83,.78, .83,1,.67, .78, .67,1), byrow=T,ncol=3)
# a
lambda1 = sqrt( R[1,2]*R[1,3]/R[2,3] )
lambda2 = sqrt( R[1,2]*R[2,3]/R[1,3] )
lambda3 = sqrt( R[1,3]*R[2,3]/R[1,2] )
round(lambda1,2)
```

```
## [1] 0.98
```

```r
round(lambda2,2)
```

```
## [1] 0.84
```

```r
round(lambda3,2)
```

```
## [1] 0.79
```

```r
psi1 = 1-lambda1^2
psi2 = 1-lambda2^2
psi3 = 1-lambda3^3
round(psi1,2)
```

```
## [1] 0.03
```

```r
round(psi2,2)
```

```
## [1] 0.29
```

```r
round(psi3,2)
```

```
## [1] 0.5
```

```r
# c
z1 = matrix(c(1,2,3),ncol=1)
hatLambda = matrix(c(lambda1,lambda2,lambda3),ncol=1)
hatPsi = diag(c(psi1,psi2,psi3))
invPsi = diag(1/c(psi1,psi2,psi3))
```

```r
hatf1B = solve(t(hatLambda)%*%invPsi%*%hatLambda) %*% t(hatLambda) %*% invPsi %*% z1
round(hatf1B,2)
```

```
##      [,1]
## [1,] 1.23
```

**Esercizio 3 (8 punti)**

Un gruppo di $n = 112$ individui si è sottoposto a $p = 6$ prove di abilità e intelligenza. Caricare la matrice di *varianze/covarianze* $S$ `ability.cov` presente nella libreria `datasets` e si risponda alle seguenti domande:

a. Sulla base dalla matrice di *correlazione* $R$, si stimi il modello fattoriale con $k = 2$ fattori utilizzando il metodo della massima verosimiglianza senza effettuare **nessuna rotazione**. Riportare, arrotondando al **terzo decimale** il valore delle statistiche test

$$T = n\log\left(\frac{\det(\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})}{\det(R)}\right) = \dots \qquad T_{Bartlett} = [(n-1)-(2p+4k+5)/6]\log\left(\frac{\det(\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})}{\det(R)}\right) = \dots$$

```r
# a
rm(list=ls())
n = 112
S = ability.cov$cov
D = diag(diag(S)^(-1/2))
R = D %*% S %*%D
af <- factanal(covmat=R, factors=2, n.obs = n, rotation = "none")
Lambda = af$loadings[,]
Psi = diag(af$uniquenesses)
Tstat = n*log(det(Lambda %*% t(Lambda) + Psi )/det(R))
round(Tstat,3)
```

```
## [1] 6.399
```

```r
p = ncol(R)
k = 2
Tbart = ( (n-1) - (2*p + 4*k + 5)/6 )*log(det(Lambda %*% t(Lambda) + Psi )/det(R))
round(Tbart,3)
```

```
## [1] 6.104
```

```r
#round(af$STATISTIC,3)
```

b. Si riportino, arrotondando al **primo decimale**, i seguenti valori

$$\underset{p\times k}{\hat{\Lambda}} = \begin{bmatrix} \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \end{bmatrix} \qquad \underset{p\times k}{\hat{\Lambda}}^{*} = \underset{p\times k}{\hat{\Lambda}}\underset{k\times k}{A} = \begin{bmatrix} \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \\ \dots\dots & \dots\dots \end{bmatrix}$$

dove $\underset{p\times k}{\hat{\Lambda}}$ è la stima dei pesi fattoriali del modello descritto al punto a. e $\underset{2\times 2}{A} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$ con $\phi = 0$ è una matrice ortogonale che produce una rotazione oraria dei pesi fattoriali $\hat{\Lambda}$.

```r
# b
round(Lambda,1)
```

```
##      Factor1 Factor2
## [1,]     0.6     0.4
```

```
## [2,]     0.3     0.5
## [3,]     0.5     0.7
## [4,]     0.3     0.4
## [5,]     1.0    -0.1
## [6,]     0.8     0.0
```

```r
phi = 0
A = matrix(c(cos(phi),sin(phi),-sin(phi),cos(phi)), byrow=TRUE, ncol=2)
Lambdastar = Lambda %*% A
round(Lambdastar,1)
```

```
##      [,1] [,2]
## [1,]  0.6  0.4
## [2,]  0.3  0.5
## [3,]  0.5  0.7
## [4,]  0.3  0.4
## [5,]  1.0 -0.1
## [6,]  0.8  0.0
```

c. Riportare, arrotondando al **terzo decimale**, il $p$-value

$$\Pr(\chi^2_{\frac{1}{2}[(p-k)^2-p-k]} > t) = \ldots\ldots$$

dove $t$ è il valore osservato della statistica test $T_{Bartlett}$ calcolata al punto a.

d. Definire l'ipotesi nulla $H_0$ e ipotesi alternativa $H_1$ del test considerato al punto a.

$H_0$ :                                        $H_1$ :

```r
# c
# round(af$PVAL,3)
p=6
k=2
df = (1/2)* ( (p-k)^2 -p -k )
round(pchisq(Tbart, df, lower.tail = F),3)
```

```
## [1] 0.192
```

**Esercizio 4 (4 punti)**

Dimostrare, esplicitando tutti i passaggi e le quantità coinvolte, che

a. se esiste un $\underset{p\times 1}{c} \neq 0$ tale che $\underset{n\times p}{\tilde{X}}\underset{p\times 1}{c} = \underset{n\times 1}{0}$, allora $\det(S) = 0$ dove $S = \frac{1}{n}\tilde{X}'\tilde{X}$.

b. nel modello fattoriale a $k$ fattori, $\mathbb{E}(\underset{p\times 1}{x}\underset{1\times k}{f'}) = \underset{p\times k}{\Lambda}$