19 Aprile 2018 - Analisi Esplorativa

Cognome: ....

Nome: ....

Matricola: ....

Tipologia d'esame: ☐ 12 CFU ☐ 15 CFU

Prova scritta - fila A

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

---

Esercizio 1 (13 punti)

Si consideri la seguente matrice dei dati $\underset{n\times p}{X}$ relativi alle 10 aziende più importanti nel mondo (fonte: Forbes Global 2000):

	Vendite (in miliardi di $)	Profitto (in miliardi di $)
Citigroup	108.28	17.05
General Electric	152.36	16.59
American Intl Group	95.04	10.91
Bank of America	65.45	14.14
HSBC Group	62.97	9.52
ExxxonMobil	263.99	25.33
Royal Dutch/Shell	265.19	18.54
BP	285.06	15.73
ING Group	92.01	8.10
Toyota Motor	165.68	11.13

Si risponda alle seguenti domande:

a. Riportare il vettore delle medie $\underset{p\times 1}{\bar{x}}$ e la matrice di varianze/covarianze $\underset{p\times p}{S}$, arrotondando al **secondo decimale**:

$$\underset{p\times 1}{\bar{x}} = \Bigg[\qquad\qquad\qquad\qquad \underset{p\times p}{S} = \Bigg[$$

```
##       [,1]
## [1,] 155.6
## [2,]  14.7
```

```
##          Vendite Profitto
## Vendite  6728.81   273.26
## Profitto  273.26    23.57
```

b. Si riporti la proporzione di varianza spiegata dalla prima componente principale calcolata sulla base di $\underset{p\times p}{S}$:

$$\frac{\lambda_1}{\lambda_1+\lambda_2} = \ldots\ldots$$

dove λ_1 e λ_2 sono gli autovalori di S, arrotondando **alla terza cifra decimale**.

```
## [1] 0.998
```

c. Si calcoli il coefficiente di correlazione tra la j-sima colonna $\underset{n\times 1}{\tilde{x}_j}$ della matrice dei dati centrati $\underset{n\times p}{\tilde{X}}$ e i punteggi $\underset{n\times 1}{y_1}$ della prima componente principale, arrotondando il calcolo al **sesto decimale**:

$$\mathbb{C}\text{orr}(\tilde{x}_1, y_1) = \ldots\ldots, \qquad \mathbb{C}\text{orr}(\tilde{x}_2, y_1) = \ldots\ldots$$

```
##   Vendite
## -0.999998
```

```
##  Profitto
## -0.687407
```

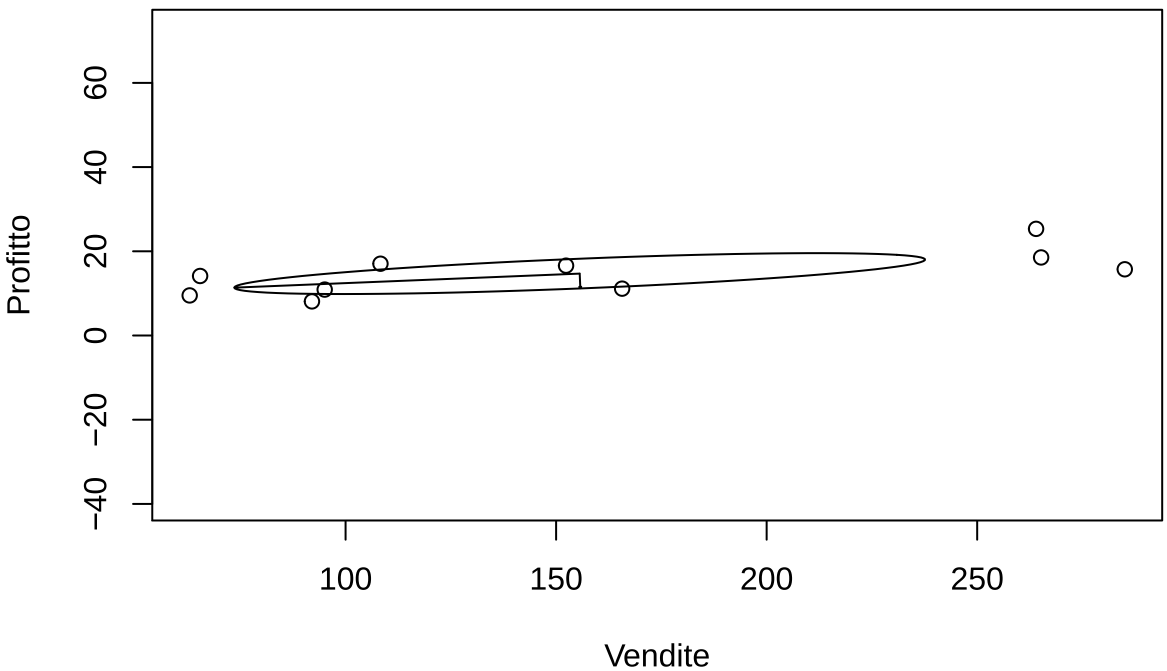
d. Si riporti nel grafico seguente, la lunghezza del semiasse maggiore e quella del semiasse minore dell'ellisse

$$\underset{1\times p}{(x-\bar{x})'}\underset{p\times p}{S^{-1}}\underset{p\times 1}{(x-\bar{x})} = 1$$

arrotondando **al secondo decimale**.

```
##
## Attaching package: 'ellipse'
```

```
## The following object is masked from 'package:graphics':
##
##     pairs
```



e. Sulla base dei risultati ottenuti nei punti a. - d., ritenete preferibile calcolare le componenti principali sulla base della matrice di correlazione $\underset{p\times p}{R}$? Motivare la risposta.

f. Si ottengano tre gruppi di aziende utilizzando il metodo gerarchico basato sulla matrice delle distanze di Lagrange e il legame completo. Riportare i nomi delle aziende presenti nei tre gruppi:

```
##           Citigroup    General Electric American Intl Group
##                   1                   2                   1
##     Bank of America          HSBC Group         ExxxonMobil
##                   1                   1                   3
##   Royal Dutch/Shell                  BP           ING Group
##                   3                   3                   1
##        Toyota Motor
##                   2
```

Gruppo 1: ..

Gruppo 2: ..

Gruppo 3: ..

g. Si calcoli la distanza di Mahalanobis di ciascuna unità statistica dal baricentro e si riportino i nomi delle aziende che superano il valore 1.4.

```
## [1] "Bank of America" "ExxxonMobil"     "BP"
```

Aziende con $d_M(u_i, \bar{x}) > 1.4$: ..

h. Dimostrare, esplicitando tutti i passaggi e le quantità coinvolte, che $\det(S) = \prod_{j=1}^{p} \lambda_j$, dove λ_j sono gli autovalori di una generica matrice di varianze/covarianze $\underset{p \times p}{S}$

19 Aprile 2018 - Analisi Esplorativa

Cognome: ....

Nome: ....

Matricola: ....

Tipologia d'esame: ☐ 12 CFU ☐ 15 CFU

Prova scritta - fila B

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

---

Esercizio 2 (13 punti)

Si consideri il seguente il modello fattoriale ad 1 fattore:

$$x_1 = 0.9f + u_1$$
$$x_2 = 0.7f + u_2$$
$$x_3 = 0.5f + u_3$$

con le varianze specifiche pari a $\psi_1 = 0.19$, $\psi_2 = 0.51$ e $\psi_3 = 0.75$.

a. Riportare la matrice di correlazione prevista dal modello fattoriale ad 1 fattore:

$$\mathbb{C}\mathrm{orr}(\underset{p\times 1}{x}) = \Lambda\Lambda' + \Psi = \Bigg[$$

```
##      [,1] [,2] [,3]
## [1,] 1.00 0.63 0.45
## [2,] 0.63 1.00 0.35
## [3,] 0.45 0.35 1.00
```

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

c. Riportare i valori delle comunalità:

$$h_1^2 = .. \qquad h_2^2 = .. \qquad h_3^2 = ..$$

```
##      [,1]
## [1,] 0.81
## [2,] 0.49
## [3,] 0.25
```

d. Si supponga di aver osservato $n = 20$ unità statistiche, e di aver calcolato la seguente matrice di correlazione

$$\underset{3\times 3}{R} = \left[\begin{array}{ccc} 1 & 0.5 & 0.4 \\ & 1 & 0.3 \\ & & 1 \end{array}\right].$$

Sulla base dalla matrice di correlazione R, si stimi il modello fattoriale con $k = 1$ fattori utilizzando il metodo della massima verosimiglianza senza effettuare **nessuna rotazione**. Riportare, arrotondando al **secondo decimale**, i seguenti valori

$$\underset{p\times k}{\hat{\Lambda}} = \Bigg[\qquad\qquad \underset{p\times p}{\hat{\Psi}} = \Bigg[$$

```
##      [,1] [,2] [,3]
## [1,]  1.0  0.5  0.4
## [2,]  0.5  1.0  0.3
## [3,]  0.4  0.3  1.0

## [1] 0.82 0.61 0.49

## [1] 0.33 0.62 0.76
```

f. Dimostrare, esplicitando tutti i passaggi e le quantità coinvolte, che nel modello fattoriale a k fattori, $\mathbb{E}(\underset{p\times 1}{x}\underset{1\times k}{f'}) = \underset{p\times k}{\Lambda}$.

g. Dimostrare che la matrice di centramento $\underset{n\times n}{H}$ è idempotente, giustificano tutti i passaggi.

h. Dimostrare, esplicitando tutti i passaggi, e specificando tutte le quantità coinvolte, che la matrice di varianze/covarianze S è semi-definita positiva, esplicitando tutti i passaggi.