12 Settembre 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 (9 punti)

Alla matrice di varianze/covarianze relativa a $\underset{n\times p}{X}$ sono associati i seguenti autovalori e autovettori normalizzati: $\lambda_1 = 9$, $\lambda_2 = 6$, $\underset{2\times 1}{v_1} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$ e $\underset{2\times 1}{v_2} = \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}$.

a. Quante sono le colonne di $\underset{n\times p}{X}$? $p = \ldots\ldots$

b. Determinare la matrice di varianze/covarianze $\underset{p\times p}{S} = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$

c. Riportare
   - varianza totale = ... e generalizzata = ...
   - l'indice di variabilità relativo (arrotondare al secondo decimale) = ...

d. Determinare, **arrotondando al secondo decimale**, $\underset{p\times p}{S^{1/2}} = \begin{bmatrix} \ldots\ldots & \ldots\ldots \\ \ldots\ldots & \ldots\ldots \end{bmatrix}$

e. Calcolare la correlazione tra la seconda colonna $\underset{n\times 1}{\tilde{x}_2}$ di $\underset{n\times p}{\tilde{X}}$ e i punteggi $\underset{n\times 1}{y_2}$ della seconda componente principale, **arrotondando al secondo decimale**.

$$= \ldots\ldots\ldots$$

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

```
##      [,1] [,2]
## [1,]  6.6  1.2
## [2,]  1.2  8.4
```

```
## [1] 54
```

```
## [1] 15
```

```
## [1] 0.97
```

```
##      [,1] [,2]
## [1,] 2.56 0.22
## [2,] 0.22 2.89
```

```
## [1] -0.38
```

Esercizio 2 (9 punti)

Si consideri la seguente matrice di distanza:

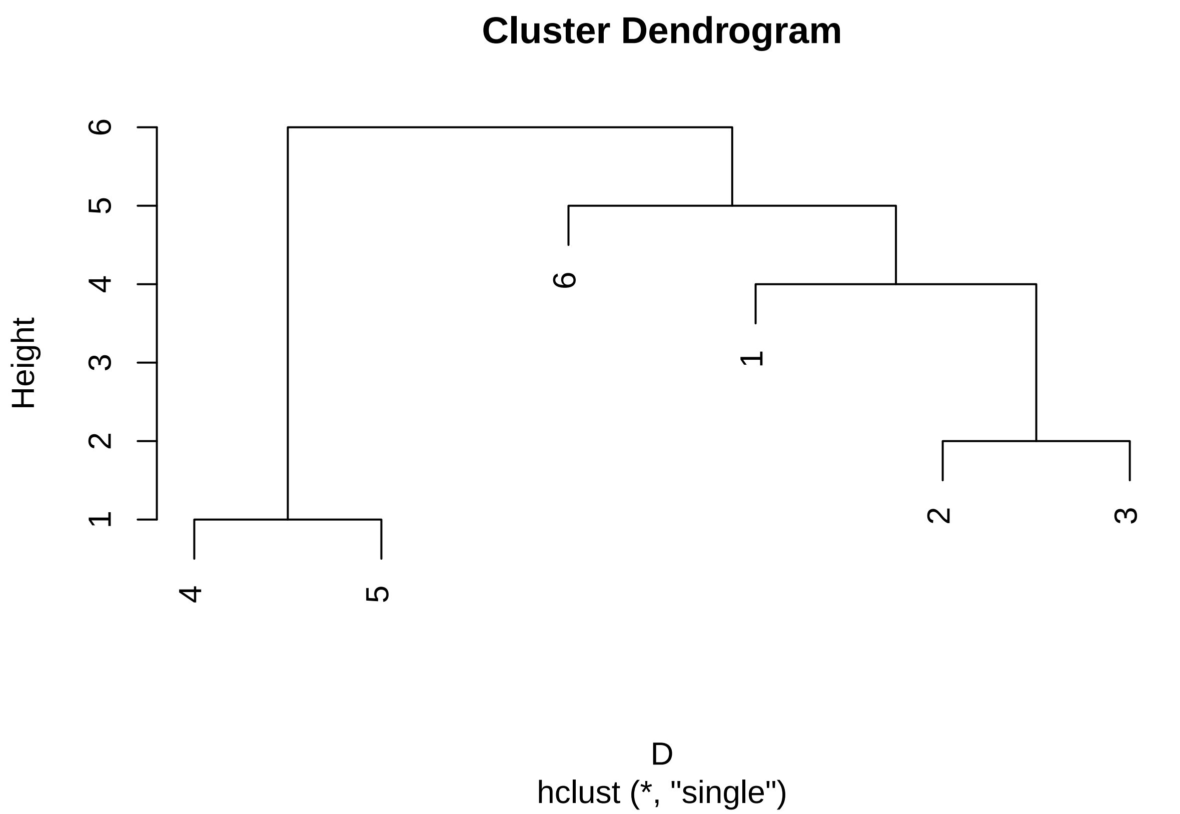
	u_1	u_2	u_3	u_4	u_5	u_6
u_1	0	4	6	12	13	7
u_2	4	0	2	8	9	5
u_3	6	2	0	6	7	5
u_4	12	8	6	0	1	9
u_5	13	9	7	1	0	8
u_6	7	5	5	9	8	0

a. Si rappresenti con il dendrogramma la sequenza di partizioni ottenuta con l'algoritmo gerarchico agglomerativo basato sul legame singolo, riportando sull'asse delle ordinate le distanze corrispondenti.

b. Si riporti il valore medio della *silhouette* (**arrotondando al secondo decimale**) per i raggruppamenti in $K = 2, 3, 4$ e 5 gruppi. Si identifichi il numero di gruppi ottimale secondo questo criterio.

N.ro di gruppi K	Valore medio della silhouette
2	...
3	...
4	...
5	...

N.ro di gruppi ottimale? $K^* = \ldots\ldots$




```
## [1] 0.59
```

```
## [1] 0.44
```

```
## [1] 0.47
```

```
## [1] 0.28
```

Si consideri la seguente matrice di distanza:

	u_1	u_2	u_3	u_4	u_5
u_1	0				
u_2	9	0			
u_3	3	7	0		
u_4	6	5	9	0	
u_5	11	10	2	8	0

c. Se utilizziamo un algoritmo gerarchico agglomerativo, le unità u_3 e u_5 vengono messe assieme nel gruppo (u_3, u_5). Aggiornare la matrice delle distanze utilizzando i metodi del legame singolo e completo:

Singolo	(u_3, u_5)	u_1	u_2	u_4
(u_3, u_5)	0			
u_1	……	0		
u_2	……	……	0	
u_4	……	……	……	0

Completo	(u_3, u_5)	u_1	u_2	u_4
(u_3, u_5)	0			
u_1	……	0		
u_2	……	……	0	
u_4	……	……	……	0

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`. 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.

a. Si riporti la stima dei pesi fattoriali (arrotondando al **primo decimale**):

$$\underset{p \times k}{\hat{\Lambda}} = \begin{bmatrix} \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots \end{bmatrix}$$

b. Calcolare, arrotondando al **quarto decimale**, il valore della statistiche test

$$T = n \log\left(\frac{\det(\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})}{\det(R)}\right) = \ldots$$

c. Si descriva il modello fattoriale con k fattori in forma matriciale, specificando tutte le assunzioni

d. Data una generica matrice $\underset{n \times p}{X}$ con vettore delle medie $\underset{p \times 1}{\bar{x}}$ e matrice di varianze/covarianze $\underset{p \times p}{S}$, si riporti la definizione della distanza di Mahalanobis $d_M(u_i, \bar{x})$ tra l'i-sima unità statistica $\underset{1 \times p}{u_i'}$ e il baricentro $\underset{1 \times p}{\bar{x}'}$.

$d_M(u_i, \bar{x}) =$

```
##      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
```

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