Cognome:	Nome:	. Matricola:
Γipologia d'esame: □ 12 CFU	□ 15 CFU	

Prova scritta di ASM 12CFU e 15CFU - Modulo Analisi Esplorativa del 11.09.2017

La durata della prova è di 80 minuti. Si svolgano gli esercizi 1, 2 e 3 riportando il risultato dove indicato.

Esercizio 1 (8 punti)

Data la seguente matrice dei dati relativa a 5 unità statistiche,

x	y	z
-2	-1	2
-1	1	-1
0	2	-2
1	0	-1
2	-2	2

arrotondando i calcoli al secondo decimale, si determini:

- a. il vettore delle medie campionarie e la matrice di correlazione R;
- b. la varianza totale di R e l'indice di variabilità relativo;
- c. la varianza delle tre componenti principali basate sulla matrice di correlazione R;
- d. il punteggio relativo alla prima unità statistica per la prima componente principale sapendo che l'autovettore associato al più grande autovalore è (-0.22, 0.71, -0.67)';
- e. il coefficiente di correlazione lineare fra x e la prima componente principale.

```
##
     [,1]
## x
        0
## y
        0
## z
        0
##
              У
## x 1.0 -0.30 0.00
## y -0.3 1.00 -0.93
## z 0.0 -0.93 1.00
## [1] 3
## [1] 0.05
## [1] 1.98 1.00 0.02
## [1] -0.991722
## [1] -0.3095674
```

Esercizio 2 (9 punti)

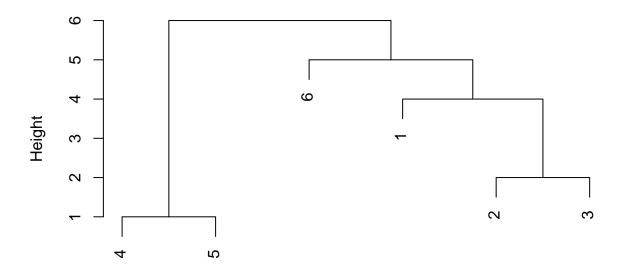
Si consideri la seguente matrice di distanza:

	1	2	3	4	5	6
1	0	4	6	12	13	7
2	4	0	2	8	9	5
3	6	2	0	6	7	5
4	12	8	6	0	1	9
5	13	9	7	1	0	8
6	7	5	5	9	8	0

- a. si determini la sequenza delle partizioni identificata secondo il metodo del legame singolo, riportando le matrici di distanza che si ottengono nei primi 3 passi della procedura;
- b. si disegni il dendrogramma corrispondente al risultato ottenuto al punto precedente, identificando quando è opportuno arrestare la procedura, utilizzando come criterio il valore medio della *silhouette*;
- c. si descriva l'algoritmo delle K-medie.

```
## 1 2 3 4 5
## 2 4
## 3 6 2
## 4 12 8 6
## 5 13 9 7 1
## 6 7 5 5 9 8
```

Cluster Dendrogram



Loading required package: cluster

[1] 0.5932612

[1] 0.4408883

[1] 0.4720238

[1] 0.281746

Esercizio 3 (9 punti)

- a. Si dimostri che la varianza generalizzata calcolata sulla matrice di correlazione coincide con l'indice di variabilità relativo;
- b. si consideri la matrice $A = \begin{bmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{bmatrix}$ simmetrica e a valori reali. Si esprima questa matrice secondo il Teorema di Decomposizione Spettrale;
- c. si dimostri che se le colonne della matrice dei dati centrata \tilde{X} sono linearmente dipendenti (assumendo n>p), allora la varianza generalizzata è pari a 0.