STAT 1164 Multivariate Analysis

Homework 6

Chapter 15 Cluster Analysis

Wanqian Yao
12531771

Homework 6.

Qianqian Yao
1253177
STAT 7164

Question 1.

$$d_{47} = \sqrt{(-0.158-0.810)^2 + (-0.599+0.134)^2 + (-0.405+0.542)^2 + (0.049-0.943)^2 + (-1.112-0.118)^2}$$

$$= \sqrt{1.568^2 + 0.135^2 + 0.137^2 + 0.894^2 + 1.23^2}$$

$$= \sqrt{6.641415}$$

$$\doteq 2.5318 \qquad 2.578$$

Question 2.

part 1: Calculation processes

part 2: SAS output.

part 3: Summaries of results of 5 methods

Part 1

Calculation process of
- Single Linkage
- Complete Linkage
- Average Linkage
- Centroid Linkage
- Ward's Linkage

Single-Linkage: the distance between two clusters A and B is defined as the minimum distance between an observation in A and an observation in B.

1) 1 2 3 4 5 6 7 8 9

Identify which two clusters have smallest distance,

$\min\{d_{ij}\} = d_{32} = 0.33$

merge 2 3

Now:

2) 1 2,3 4 5 6 7 8 9

$d_{14} = 1.87$
$d_{15} = 3.43$
$d_{16} = 2.58$
$d_{17} = 0.82$
$d_{18} = 2.21$
$d_{19} = 2.60$
$d_{45} = 1.66$
$d_{46} = 2.19$
$d_{47} = 1.54$
$d_{48} = 2.43$
$d_{49} = 3.07$
$d_{56} = 2.80$
$d_{57} = 3.09$
$d_{58} = 3.40$
$d_{59} = 3.80$

$d_{67} = 2.11$
$d_{68} = 1.29$
$d_{69} = 1.71$
$d_{78} = 2.08$
$d_{79} = 2.5118$
$d_{89} = 0.94$

	MAX	AVG
$d(1, 2,3) = \min(d_{12}, d_{13}) = \min(4.01, 4.14) = 4.01$	4.14	4.075
$d(4, 2,3) = \min(d_{42}, d_{43}) = \min(4.49, 4.64) = 4.49$	4.64	4.565
$d(5, 2,3) = \min(d_{52}, d_{53}) = \min(5.24, 5.36) = 5.24$	5.36	5.3
$d(6, 2,3) = \min(d_{62}, d_{63}) = \min(3.48, 3.67) = 3.48$	3.67	3.575
$d(7, 2,3) = \min(d_{72}, d_{73}) = \min(3.97, 4.16) = 3.97$	4.16	4.065
$d(8, 2,3) = \min(d_{82}, d_{83}) = \min(2.91, 3.06) = 2.91$	3.06	2.985
$d(9, 2,3) = \min(d_{92}, d_{93}) = \min(3.49, 3.54) = 3.49$	3.54	3.53

merge 1 7

3) 1,7 2,3 4 5 6 8 9

d_{45} 1.66
d_{46} 2.19
d_{48} 2.73
d_{49} 3.07
d_{56} 2.80
d_{58} 3.40
d_{59} 3.80
d_{68} 1.29
d_{69} 1.71
d_{89} 0.94

$d(4, 1,7) = \min(d_{41}\ d_{47}) = \min(1.87, 1.54) = 1.54$
$d(5, 1,7) = \min(d_{51}, d_{57}) = \min(3.43, 3.09) = 3.09$
$d(6, 1,7) = \min(d_{61}, d_{67}) = \min(2.58, 2.11) = 2.11$
$d(8, 1,7) = \min(d_{81}, d_{87}) = \min(2.21, 2.08) = 2.08$
$d(9, 1,7) = \min(d_{91}, d_{97}) = \min(2.60, 2.5178) = 2.5178$
$d(4, 2,3) = \min(d_{42}, d_{43}) = \min(4.49, 4.64) = 4.49$
$d(5, 2,3) = \min(d_{52}, d_{53}) = \min(5.24, 5.36) = 5.24$
$d(6, 2,3) = \min(d_{62}, d_{63}) = \min(3.48, 3.67) = 3.48$
$d(8, 2,3) = \min(d_{82}, d_{83}) = \min(2.91, 3.06) = 2.91$
$d(9, 2,3) = \min(d_{92}, d_{93}) = \min(3.49, 3.57) = 3.49$
$d(1,7, 2,3) = \min(d_{12}, d_{13}, d_{72}, d_{73}) = \min(4.01, 4.14, 3.97, 4.16) = 3.97$

Max	AVG
1.87	1.705
3.43	3.26
2.58	2.345
4.64	2.145
5.36	2.589
3.67	4.565
3.06	5.3
3.57	3.5775
4.16	2.985
	3.53
	4.07

merge 8 9

4) 1,7 2,3 8,9 4 5 6

$d_{45} = 1.66$
$d_{46} = 2.19$
$d_{56} = 2.80$

$d(4, 1,7) = \min(d_{41}, d_{47}) = \min(1.87, 1.54) = 1.54$
$d(5, 1,7) = \min(d_{51}\ d_{57}) = \min(3.43, 3.09) = 3.09$
$d(6, 1,7) = \min(d_{61}, d_{67}) = \min(2.58, 2.11) = 2.11$
$d(4, 2,3) = \min(d_{42}\ d_{43}) = \min(4.49, 4.64) = 4.49$
$d(5, 2,3) = \min(d_{52}\ d_{53}) = \min(5.24, 5.36) = 5.24$
$d(6, 2,3) = \min(d_{62}, d_{63}) = \min(3.48, 3.67) = 3.48$
$d(4, 8,9) = \min(d_{48}\ d_{49}) = \min(2.73, 3.07) = 2.73$
$d(5, 8,9) = \min(d_{58}\ d_{59}) = \min(3.40, 3.80) = 3.40$
$d(6, 8,9) = \min(d_{68}, d_{69}) = \min(1.29, 1.71) = 1.29$
$d(1,7, 2,3) = \min(d_{12}\ d_{13}\ d_{17}\ d_{73}) = \min(4.01, 4.14, 3.97, 4.16) = 3.97$
$d(1,7, 8,9) = \min(d_{18}\ d_{19}\ d_{78}\ d_{79}) = \min(2.21\ 2.60\ 2.08\ 2.5178) = 2.08$
$d(2,3, 8,9) = \min(d_{28}\ d_{29}\ d_{38}\ d_{39}) = \min(2.91, 3.49, 3.06, 3.57) = 2.91$

Max	AVG
1.87	1.705
3.43	3.26
2.58	2.345
4.64	4.07
5.36	2.367
3.48	3.2575
3.07	
3.80	
1.71	
4.16	
2.60	
3.57	

merge 6, 8,9

AVG
1.705
3.26
2.345
4.565
5.3
3.5775
2.9
3.75
1.5

5) 1,7 2,3 6,8,9 , 4 , 5

d_{45} 1.66

$d(1,7, 2,3) = \min(d_{12}\ d_{13}\ d_{72}\ d_{73}) = \min(4.01, 4.14, 3.97, 4.16) = 3.97$

$d(1,7, 6,8,9) = \min(d_{16}, d_{18}, d_{19}, d_{76}, d_{78}, d_{79}) = \min(2.58, 2.21, 2.60, 2.11, 2.08, 2.578) = 2.08$

$d(1,7, 4) = \min(d_{41}, d_{47}) = \min(1.87, 1.54) = 1.54$

$d(1,7, 5) = \min(d_{51}, d_{57}) = \min(3.43, 3.09) = 3.09$

$d(2,3, 6,8,9) = \min(d_{26}, d_{28}, d_{29}, d_{36}, d_{38}, d_{39}) = \min(3.48, 2.91, 3.49, 3.67, 3.06, 3.54) = 2.91$

$d(2,3, 4) = \min(d_{42}, d_{43}) = \min(4.49, 4.64) = 4.49$

$d(2,3, 5) = \min(d_{52}, d_{53}) = \min(5.24, 5.36) = 5.24$

$d(6,8,9, 4) = \min(d_{46}\ d_{48}\ d_{49}) = \min(2.19, 2.93, 3.07) = 2.19$

$d(6,8,9, 5) = \min(d_{56}, d_{58}, d_{59}) = \min(2.80, 3.70, 3.80) = 2.80$

merge 1,7 4

6) 1,7,4 2,3 6,8,9 5

$d(1,7,4, 2,3) = \min(d_{12}, d_{13}, d_{72}, d_{73}, d_{42}, d_{43}) = 4.01, 4.14, 3.97, 4.16, 4.49, 4.64 = 3.97$

$d(1,7,4, 6,8,9) = \min(d_{16}, d_{18}, d_{19}, d_{76}, d_{78}, d_{79}, d_{46}, d_{48}, d_{49}) =$

$= \min(2.58, 2.21, 2.60, 2.11, 2.08, 2.578, 2.19, 2.93, 3.07) = 2.08$

$d(1,7,4, 5) = \min(d_{51}, d_{57}, d_{54}) = \min(3.43, 3.09, 1.66) = 1.66$

$d(2,3, 6,8,9) = \min(d_{26}, d_{28}, d_{29}, d_{36}, d_{38}, d_{39}) = \min(3.48, 2.91, 3.49, 3.67, 3.06, 3.54) = 2.91$

$d(2,3, 5) = \min(d_{52}, d_{53}) = \min(5.24, 5.36) = 5.24$

$d(6,8,9, 5) = \min(d_{56}, d_{58}\ d_{59}) = \min(2.80, 3.70, 3.80) = 2.80$

merge 1,7,4 5

7) {1, 4, 5, 7} {2, 3} {6, 8, 9}

$d(1,4,5,7, 2,3) = d_{21}\ d_{24}\ d_{25}\ d_{27}\ d_{31}\ d_{34}\ d_{35}\ d_{37}$

$= 4.01, 4.49\ 5.24\ 3.97, 4.14, 4.64, 5.36, 4.16 = 3.97$

$d(1,4,5,7, 6,8,9) = d_{16}\ d_{18}\ d_{19}\ d_{46}\ d_{48}\ d_{49}\ d_{56}\ d_{58}\ d_{59}\ d_{76}\ d_{78}\ d_{79}$

$= 2.58, 2.21, 2.60, 2.19, 2.93, 3.07, 2.80, 3.70, 3.80, 2.11, 2.08, 2.578 = 2.08$

$d(2,3, 6,8,9) = d_{26}\ d_{28}\ d_{29}\ d_{36}\ d_{38}\ d_{39}$

$= 3.48\ 2.91\ 3.49\ 3.67\ 3.06\ 3.54 = 2.91$

Merge 1,4,5,7 6,8,9

8) 1, 4, 5, 6, 7, 8, 9 — 2, 3

d_{31} d_{34} d_{35} d_{36} d_{37} d_{38} d_{39} — 4.14, 4.64, 5.36, 3.67, 4.16, 3.06, 3.57

d_{21} d_{24} d_{25} d_{26} d_{27} d_{28} d_{29} — 4.01 4.49 5.24 3.48 3.97 2.91 3.49

9 Single-Linkage

9	1	2	3	4	5	6	7	8	9	0
8	[2,3]	1	4	5	6	7	8	9		0.33
7	[1,7]	[2,3]	4	5	6	8	9			0.82
6	[8,9]	[1,7]	[2,3]	4	5	6				0.94
5	[6,8,9]	[1,7]	[2,3]	4	5					1.29
4	[1,4,7]	[6,8,9]	[2,3]	5						1.54
3	[1,4,5,7]	[6,8,9]	[2,3]							1.66
2	[1,4,5,6,7,8,9]	[2,3]								2.08
1	[1,2,3,4,5,6,7,8,9]									2.91

Complete Linkage : The distance between two clusters A and B is defined as the maximum distance between a point in A and a point in B.

1) 1 2 3 4 5 6 7 8 9

$\min\{d_{ij}\} = d_{32} = 0.33$

Merge 2 3

2) Now: 1 2,3 4 5 6 7 8 9

Mgerge 1 7 $d_{17} = 0.82$

3) Now: 1,7 2,3 4 5 6 8 9

Merge 8 9 $d_{89} = 0.94$

4) Now:

1,7 2,3 8,9 4 5 6

Merge 4 5 $d_{45} = 1.66$

5) Now:

1,7 2,3 8,9 4,5 6

$d(1,7\ \ 2,3) = \max(d_{12}, d_{13}, d_{72}, d_{73}) = \max(4.01, 4.14, 3.97, 4.16) = 4.16$

$d(1,7\ ,\ 8,9) = \max(d_{18}, d_{19}, d_{78}, d_{79}) = \max(2.21, 2.60\ 2.08, 2.518) = 2.60$

$d(1,7\ ,\ 4,5) = \max(d_{14}, d_{15}, d_{74}, d_{75}) = \max(1.87, 3.43, 1.54, 3.09) = 3.43$

$d(1,7\ ,\ 6) = \max(d_{61}, d_{67}) = \max(2.58, 2.11) = 2.58$

$d(2,3\ \ 8,9) = \max(d_{28}, d_{29}, d_{38}, d_{39}) = \max(2.91, 3.49, 3.06, 3.57) = 3.57$

$d(2,3\ ,\ 4,5) = \max(d_{24}, d_{25}, d_{34}, d_{35}) = \max(4.49, 5.24, 4.64, 5.36) = 5.36$

$d(2,3\ ,\ 6) = \max(d_{62}, d_{63}) = \max(3.48, 3.67) = 3.67$

$d(8,9\ ,\ 4,5) = \max(d_{84}, d_{85}, d_{94}, d_{95}) = \max(2.13, 3.10, 3.07, 3.80) = 3.80$

$d(8,9\ ,\ 6) = \max(d_{68}, d_{69}) = \max(1.29, 1.71) = 1.71$

$d(4,5\ \ 6) = \max(d_{64}, d_{65}) = \max(2.19, 2.80) = 2.80$

Merge 8,9 6 $d = 1.71$

6) 1,7 2,3 6,8,9 4,5

$d(1,7\ \ 2,3) = \max(d_{12}, d_{13}, d_{72}, d_{73}) = \max(4.01, 4.14, 3.97, 4.16) = 4.16$

$d(1,7\ \ 6,8,9) = \max(d_{16}, d_{18}, d_{19}, d_{76}, d_{78}, d_{79}) = \max(2.58, 2.21, 2.60, 2.11, 2.08, 2.578) = 2.60$

$d(1,7\ \ 4,5) = \max(d_{14}\ d_{15}\ d_{74}\ d_{75}) = \max(1.87, 3.43, 1.54, 3.09) = 3.09$

$d(2,3\ \ 6,8,9) = \max(d_{26}, d_{28}, d_{29}, d_{36}, d_{38}, d_{39}) = \max(3.48, 2.91, 3.49, 3.67, 3.06, 3.57) = 3.67$

$d(2,3\ \ 4,5) = \max(d_{24}, d_{25}, d_{34}, d_{35}) = \max(4.49, 5.24, 4.64, 5.36) = 5.36$

$d(6,8,9, 4,5) = \max(d_{46}, d_{48}, d_{49}, d_{56}, d_{58}, d_{59}) = \max(2.19, 2.73, 3.07, 2.80, 3.70, 3.80) = 3.80$

Merge 1,7 6,8,9 $d = 2.60$

7) 1,6,7,8,9 2,3 4,5

$d(1,6,7,8,9,\ 2,3) = \max(d_{21}, d_{26}, d_{27}, d_{28}, d_{29}, d_{31}, d_{36}, d_{37}, d_{38}, d_{39})$
$= \max(4.01, 3.48, 3.97, 2.91, 3.49, 4.14, 3.67, 4.16, 3.06, 3.57) = 4.16$

$d(1,6,7,8,9,\ 4,5) = \max(d_{41}, d_{46}, d_{47}, d_{48}, d_{49}, d_{51}, d_{56}, d_{57}, d_{58}, d_{59})$
$= \max(1.87, 2.19, 1.54, 2.73, 3.07, 3.43, 2.80, 3.09, 3.70, 3.80) = 3.80$

$d(2,3\ \ 4,5) = 5.36$

Merge 1,6,7,8,9 4,5

8) 1,4,5,6,7,8,9 2,3

Max = 5.36

Complete

9										
9	1	2	3	4	5	6	7	8	9	0
8	[2,3]	1	4	5	6	7	8	9		0.33
7	[1,7]	[2,3]	4	5	6	8	9			0.82
6	[8,9]	[1,7]	[2,3]	4	5	6				0.94
5	[8,9]	[1,7]	[2,3]	[4,5]	6					1.66
4	[6,8,9]	[1,7]	[2,3]	[4,5]						1.71
3	[1,6,7,8,9]	[2,3]	[4,5]							2.60
2	[1,4,5,6,7,8,9]	[2,3]								3.80
1	[1,2,3,4,5,6,7,8,9]									5.36

Average Linkage the distance between two clusters A and B is defined as the average of the $n_A n_B$ distances between n_A points in A and the n_B points in B.

1) 1 2 3 4 5 6 7 8 9

$d_{23} = 0.33$

Merge 2 3

2) 1 2,3 4 5 6 7 8 9

Merge 1 7

3) 1,7 2,3 4 5 6 8 9

Merge 8 9

4) 1,7 2,3 8,9 4 5 6

Merge 6 8,9 $d = 1.5$

5) 1,7 2,3 6,8,9 4 5

AVE dis (1.66)

4.07
2.360
1.705
3.26
3.363
4.565
5.3
2.663
3.433

Merge 4 5

6) 1,7 2,3 6,8,9 4,5

AVE
4.235
2.461
2.727
3.363
5.3
3.433

Merge 1,7 6,8,9

7) 1 6,7,8,9 2,3 4,5

12

$d(1,6,7,8,9,2,3) = AVE(d_{21}\ d_{26}, d_{27}, d_{28}, d_{29}, d_{31}\ d_{36}\ d_{37}\ d_{38}\ d_{39})$

$= 4.01 + 4.14 + 3.48 + 3.97 + 2.91 + 3.49 + 3.67 + 4.16 + 3.06 + 3.57$

$= 3.800$

$d(1,6,7,8,9,4,5) = AVE(d_{51}, d_{41}, d_{56}\ d_{57}\ d_{58}\ d_{59}, d_{46}\ d_{47}\ d_{48}\ d_{49}) = 2.911$

Merge 1,6,7,8,9 4,5

8) 1,4,5,6,7,8,9 2,3

AVE (4.14 + 4.64 + 5.36 + 3.67 + 4.16 + 3.06 + 3.57 +
4.01 + 4.49 + 5.24 + 3.48 + 3.97 + 2.91 + 3.49)

= 4.014

g	Average									
9	1	2	3	4	5	6	7	8	9	0
8	[2,3]	1	4	5	6	7	8	9		0.33
7	[1,7]	[2,3]	4	5	6	8	9			0.82
6	[1,7]	[2,3]	[8,9]	4	5	6				0.94
5	[1,7]	[2,3]	[6,8,9]	4	5					1.15
4	[1,7]	[2,3]	[6,8,9]	[4,5]						1.66
3	[1,6,7,8,9]	[2,3]	[4,5]							2.37
2	[1,4,5,6,7,8,9]	[2,3]								2.91
1	[1,2,3,4,5,6,7,8,9]									4.94

Centroid: distance Between A and B is defined as the Euclidean distance between mean vectors (centroids) of the two clusters.

1) 1 2 3 4 5 6 7 8 9

$d_{32} = a33$

merge 2 3

2) 1 2,3 4 5 6 7 8 9

$\overline{y_{23}} = \begin{matrix} 0.894 & -0.275 & 1.727 & -1.605 & 1.264 \\ 0.808 & -0.289 & 1.785 & -1.588 & 1.577 \end{matrix}$

$(0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205)$

d_{14}
d_{15}
d_{16}
$d_{17} = 0.82$
d_{18}
d_{19}

$$d(1, \overline{2,3}) = \begin{matrix} -1.214 & -0.950 & -0.586 & -0.043 & -0.571 \\ 0.851 & -0.282 & 1.756 & 1.5965 & 1.4205 \end{matrix}$$

$$= \sqrt{(2.065)^2 + (0.668)^2 + (1.17)^2 + (1.6395)^2 + (1.9915)^2}$$

$$= \sqrt{12.7333815}$$

$$= 3.568$$

d_{45}
d_{46}
d_{47}
d_{48}
d_{49}

$$d(4, \overline{2,3}) = \begin{matrix} -1.159 & 0.817 & -0.612 & 0.396 & -0.961 \\ 0.851 & -0.282 & 1.756 & 1.5965 & 1.4205 \end{matrix}$$

$$\sqrt{(2.01)^2 + (1.099)^2 + (2.428)^2 + (1.2005)^2 + (2.38715)^2}$$

$$= \sqrt{18.284}$$

$$= 4.276$$

d_{56}
d_{57}
d_{58}
d_{59}

$$d(5 \quad \overline{2,3}) = \begin{matrix} -1.048 & 2.198 & -0.503 & 1.285 & -0.800 \\ 0.851 & -0.282 & 1.756 & 1.5965 & 1.4205 \end{matrix}$$

$$\sqrt{(1.899)^2 + (2.48)^2 + (2.259)^2 + (0.3115)^2 + (2.2205)^2}$$

$$= \sqrt{19.887}$$

$$= 4.460$$

d_{67}
d_{68}
d_{69}

$$d(6, \overline{2,3}) = \begin{matrix} 0.905 & 0.429 & -0.435 & 0.428 & -0.408 \\ 0.851 & -0.282 & 1.1756 & 1.5965 & 1.4205 \end{matrix}$$

$$\sqrt{(0.054)^2 + (0.711)^2 + (2.191)^2 + (1.1685)^2 + 1.8285}$$

$$= \sqrt{10.0197}$$

$$= 3.165$$

d_{78}
d_{79}
d_{89}

$$d(7, \overline{2,3}) = -0.758 \quad -0.599 \quad -0.405 \quad 0.049 \quad -1.113$$
$$0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205$$
$$\sqrt{(1.609)^2 + (0.317)^2 + (2.161)^2 + (1.5475)^2 + (2.5335)^2}$$
$$= \sqrt{16.1726695}$$
$$= 4.022$$

$$d(8, \overline{2,3}) = 0.763 \quad -0.599 \quad -0.369 \quad 0.136 \quad 0.301$$
$$0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205$$
$$\sqrt{(0.088)^2 + (0.317)^2 + (2.125)^2 + (1.4605)^2 + (1.1195)^2}$$
$$= \sqrt{8.010}$$
$$= 2.830$$

$$d(9, \overline{2,3}) = 0.810 \quad -0.734 \quad -0.542 \quad 0.943 \quad 0.718$$
$$0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205$$
$$\sqrt{(0.041)^2 + (0.452)^2 + (2.298)^2 + (0.6535)^2 + (0.17025)^2}$$
$$= \sqrt{6.407}$$
$$= 2.531$$ $\overline{y_{2,3}}$ (0.851)

Merge 1, 7

$d(\overline{1,7}, \overline{2,3}) = 4.042$

3) 1,7 2,3 4, 5, 6, 8, 9

$\overline{y}_{1,7} = -1.214 \quad -0.950 \quad -0.586 \quad -0.043 \quad -0.571$

d_{45} d_{46} d_{48} d_{49} d_{56} d_{58} d_{59} d_{68} d_{69} $d_{89} = 0.94$

$-0.758 \quad -0.599 \quad -0.405 \quad 0.049 \quad -1.113$

$(-0.986 \quad -0.7745 \quad -0.4955 \quad 0.006 \quad -0.842)$

$$d(4, \overline{1,7}) = -1.159 \quad 0.817 \quad -0.672 \quad 0.396 \quad -0.967$$
$$-0.986 \quad -0.7745 \quad -0.4955 \quad 0.006 \quad -0.842$$
$$\sqrt{(0.173)^2 + (1.5915)^2 + (0.1765)^2 + (0.390)^2 + (0.125)^2}$$
$$= 1.662$$

$$d(5, \overline{1,7}) = -1.048 \quad 2.198 \quad -0.503 \quad 1.285 \quad -0.800$$
$$-0.986 \quad -0.7745 \quad -0.4955 \quad 0.006 \quad -0.842$$
$$\sqrt{(0.062)^2 + (2.9725)^2 + (0.0075)^2 + (1.279)^2 + (0.042)^2}$$
$$= 3.237$$

$d(4, \overline{2,3}) = 4.216$
$d(5, \overline{2,3}) = 4.460$
$d(6, \overline{2,3}) = 3.165$
$d(8, \overline{2,3}) = 2.830$
$d(9, \overline{2,3}) = 2.531$

$$d(6, \overline{1,7}) = \frac{\begin{matrix} 0.905 & 0.429 & -0.435 & 0.428 & -0.408 \\ -0.986 & -0.7745 & -0.4955 & 0.006 & -0.842 \end{matrix}}{\sqrt{(1.891)^2 + (1.2035)^2 + (0.0605)^2 + (0.422)^2 + (0.434)^2}}$$

$$= 2.323$$

$$d(8, \overline{1,7}) = \frac{\begin{matrix} 0.763 & -0.599 & -0.369 & 0.136 & 0.301 \\ -0.986 & -0.7745 & -0.4955 & 0.006 & -0.842 \end{matrix}}{\sqrt{(1.749)^2 + (0.1755)^2 + (0.1265)^2 + (0.130)^2 + (1.143)^2}}$$

$$= 2.105$$

$$d(9, \overline{1,7}) = \frac{\begin{matrix} 0.810 & -0.734 & -0.542 & 0.943 & 0.718 \\ -0.986 & -0.7745 & -0.4955 & 0.006 & -0.842 \end{matrix}}{\sqrt{(1.796)^2 + (0.0405)^2 + (0.0465)^2 + (0.937)^2 + (1.56)^2}}$$

$$= 2.558$$

Merge 8, 9

4) $\underline{1,7}$ $\underline{2,3}$ $\underline{8,9}$ 4 5 6

d_{45} 1.66
d_{46} 2.19
d_{56} 2.80
$d(4, \overline{1,7}) = 1.662$
$d(5, \overline{1,7}) = 3.237$
$d(6, \overline{1,7}) = 2.323$

$d(4, \overline{2,3}) = 4.246$
$d(5, \overline{2,3}) = 4.460$
$d(6, \overline{2,3}) = 3.165$

$d(\overline{1,7}, \overline{2,3}) = 4.042$
$d(\overline{1,7}, \overline{8,9}) =$
$d(\overline{2,3}, \overline{8,9}) =$

Merge 6, $\underline{8,9}$

$\overline{y_{89}}$ $\begin{matrix} 0.763 & -0.599 & -0.369 & 0.136 & 0.301 \\ 0.810 & -0.734 & -0.542 & 0.943 & 0.718 \end{matrix}$
$(0.7865 \quad -0.6665 \quad -0.4555 \quad 0.5395 \quad 0.5095)$

$$d(4, \overline{8,9}) = \frac{\begin{matrix} -1.159 & 0.817 & -0.672 & 0.396 & -0.967 \\ 0.7865 & -0.6665 & -0.4555 & 0.5395 & 0.5095 \end{matrix}}{\sqrt{(1.9455)^2 + (1.4835)^2 + (0.2165)^2 + (0.1435)^2 + (1.4765)^2}}$$

$$= 2.869$$

$$d(5, \overline{8,9}) = \frac{\begin{matrix} -1.048 & 2.198 & -0.503 & 1.285 & -0.800 \\ 0.7865 & -0.6665 & -0.4555 & 0.5395 & 0.5095 \end{matrix}}{\sqrt{(1.8345)^2 + (2.8645)^2 + (0.0475)^2 + (0.7455)^2 + (1.3095)^2}}$$

$$= 3.7121$$

$$d(6, \overline{8,9}) = \frac{\begin{matrix} 0.905 & 0.429 & -0.435 & 0.428 & -0.408 \\ 0.7865 & -0.6665 & -0.4555 & 0.5395 & 0.5095 \end{matrix}}{\sqrt{(0.1185)^2 + (1.0955)^2 + (0.0205)^2 + (0.1115)^2 + (0.9175)^2}}$$

$$= 1.4438$$

5) 1.7 2.3 6.8.9 4 5

$\overline{y_{689}}$ 0.905 0.429 -0.435 0.428 -0.408
0.763 -0.599 -0.369 0.136 0.301
0.810 -0.734 -0.542 0.943 0.718

(0.826 -0.301 0.449 0.502 0.204)

$d_{45} = 1.66$

$d(4, \overline{1.7}) = 1.662$

$d(5, \overline{1.7}) = 3.239$

$d(4, \overline{2.3}) = 4.216$

$d(5, \overline{2.3}) = 4.460$

$d(4, \overline{689}) = -1.159 \quad 0.817 \quad -0.612 \quad 0.396 \quad -0.961$
$\quad\quad\quad\quad 0.826 \quad -0.301 \quad 0.449 \quad 0.502 \quad 0.204$

$\sqrt{(1.985)^2 + (1.118)^2 + (1.121)^2 + (0.106)^2 + (1.171)^2}$

$= 2.798$

$d(5, \overline{689}) = -1.048 \quad 2.198 \quad -0.503 \quad 1.285 \quad -0.800$
$\quad\quad\quad\quad 0.826 \quad -0.301 \quad 0.449 \quad 0.502 \quad 0.204$

$\sqrt{(1.8174)^2 + (2.499)^2 + (0.952)^2 + (0.783)^2 + (1.004)^2}$

$= 3.505$

$\overline{y_{23}} = (0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205)$

$\overline{y_{17}} = (-0.986 \quad -0.17745 \quad -0.4955 \quad 0.006 \quad -0.842)$

$d(\overline{689}, \overline{23}) = 0.826 \quad -0.301 \quad 0.449 \quad 0.502 \quad 0.204$
$\quad\quad\quad\quad 0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205$

$\sqrt{(0.025)^2 + (0.019)^2 + (1.307)^2 + (1.0945)^2 + (1.2165)^2}$

$= 2.095$

$d(\overline{689}, \overline{17}) = 0.826 \quad -0.301 \quad 0.449 \quad 0.502 \quad 0.204$
$\quad\quad\quad\quad -0.986 \quad -0.17745 \quad -0.4955 \quad 0.006 \quad -0.842$

$\sqrt{(1.812)^2 + (0.47735)^2 + (0.9445)^2 + (0.496)^2 + (1.046)^2}$

$= 2.396$

$d(\overline{23}\ \overline{17}) = \sqrt{(1.8317)^2 + (0.4925)^2 + (2.2515)^2 + (1.5905)^2 + (2.2625)^2}$

$= 4.042$

Merge 1.7 4

6) 1,4 2,3 6,8,9 5

$d(\overline{1,4}, \overline{23}) = 4.042$

$d(\overline{1,4}, \overline{689}) = 2.396$

$d(\overline{23}\ \overline{689}) = 2.095$

$\overline{y}_{45} = \begin{matrix} -1.159 & 0.817 & -0.672 & 0.396 & -0.967 \\ -1.048 & 2.198 & -0.503 & 1.285 & -0.800 \end{matrix}$

$\begin{pmatrix} -1.1035 & 1.5075 & -0.5875 & 0.8405 & -0.8835 \end{pmatrix}$

$d(\overline{45}, \overline{1,4}) = \begin{matrix} -0.986 & -0.7745 & -0.4955 & 0.006 & -0.842 \end{matrix}$

$\sqrt{(0.1175)^2 + (2.282)^2 + (0.092)^2 + (0.8345)^2 + (0.0415)^2}$

$= 2.435$

$d(\overline{45}, \overline{23}) \quad 0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205$

$\sqrt{(1.9545)^2 + (1.7895)^2 + (2.3435)^2 + (0.756)^2 + (2.304)^2}$

$= 4.289$

$d(\overline{45}, \overline{689}) \quad 0.826 \quad -0.301 \quad 0.449 \quad 0.502 \quad 0.204$

$\sqrt{(1.9295)^2 + (1.8085)^2 + (1.0365)^2 + (0.3385)^2 + (1.0875)^2}$

$= 3.060$

Merge $\underline{1,4,7}$ $\underline{6\ 8\ 9}$

7) $\underline{1,4,7,6,8,9}$ $\underline{2,3}$ $\underline{5}$

$d(\overline{1,4}, \overline{4,5}) = 2.435$

$\overline{y}_{23689} = \begin{matrix} 0.851 & -0.282 & 1.756 & 1.5965 & 1.4205 \\ 0.826 & -0.301 & 0.449 & 0.502 & 0.204 \end{matrix}$

$\begin{pmatrix} 0.8385 & -0.2915 & 1.1025 & 1.04925 & 0.81225 \end{pmatrix}$

$d(\overline{1,4}, \overline{23689}) \quad -0.986 \quad -0.7745 \quad -0.4955 \quad 0.006 \quad -0.842$

$\sqrt{(1.8245)^2 + (0.4835)^2 + (1.598)^2 + (1.04325)^2 + (1.65425)^2}$

$= 3.153$

Merge $\underline{1,4,7,6,8,9}$ $\underline{5}$

8) $\underline{1,4,5,6,7,8,9}$ $\underline{2,3}$

$d(\overline{4,5}, \overline{23689}) \quad -1.1035 \quad 1.5075 \quad -0.5875 \quad 0.8405 \quad -0.8835$

$\quad 0.8385 \quad -0.2915 \quad 1.1025 \quad 1.04925 \quad 0.81225$

$\sqrt{(1.942)^2 + (1.799)^2 + (1.69)^2 + (0.20875)^2 + (1.695475)^2}$

$= 3.575$

$\overline{1,4}: \quad -0.986 \quad -0.7745 \quad -0.4955 \quad 0.006 \quad -0.842$

$\overline{4,5}: \quad -1.1035 \quad 1.5075 \quad -0.5875 \quad 0.8405 \quad -0.8835$

$\begin{pmatrix} -1.04475 & 0.3665 & -0.5415 & 0.42325 & -0.86275 \end{pmatrix}$

$\begin{pmatrix} 0.8385 & -0.2915 & 1.1025 & 1.04925 & 0.81225 \end{pmatrix}$

$\sqrt{1.88325^2 + 0.658^2 + 1.644^2 + 0.625^2 + 1.675^2}$

$= 3.143$

g	Centroids									
9	1	2	3	4	5	6	7	8	9	0
8	[2,3]	1	4	5	6	7	8	9		0.33
7	[1,1]	[2,3]	4	5	6	8	9			0.82
6	[8,9]	[1,1]	[2,3]	4	5	6				0.94
5	[6,8,9]	[1,1]	[2,3]	4	5					1.438
4	[1,4,7]	[6,8,9]	[2,3]	5						1.66
3	[1,4,6,7,8,9]	[2,3]	5							2.19
2	[1,4,5,6,7,8,9]	[2,3]								2.85
1	[1,2,3,4,5,6,7,8,9]									11.1

Ward's Method

$$D_{AB} = \frac{n_A n_B}{n_A + n_B} (\bar{y}_A - \bar{y}_B)'(\bar{y}_A - \bar{y}_B)$$

1) 1 2 3 4 5 6 7 8 9

Merge 2 3 $d_{23} = 0.233$ → $D_{23} = \frac{1}{2} d_{23}^2 = 0.05445$

2) 1 2,3 4 5 6 7 8 9

$\bar{y}_{23} = (0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205)$

d_{14} $D_{14} = 1.748$
d_{15} $D_{15} = 5.882$
d_{16} $D_{16} = 3.3282$
d_{17} $D_{17} = 0.3362$
d_{18} $D_{18} = 2.44205$
d_{19} $D_{19} = 3.38$

$D(1, \overline{23}) = \frac{1 \times 2}{1+2} \times 3.568^2 = 8.487$
$D(4, \overline{23}) = \frac{1 \times 2}{1+2} \times 4.246^2 = 12.189$
$D(5, \overline{23}) = \frac{1 \times 2}{1+2} \times 4.460^2 = 13.261$
$D(6, \overline{23}) = \frac{1 \times 2}{1+2} \times 3.165^2 = 6.678$
$D(7, \overline{23}) = \frac{1 \times 2}{1+2} \times 4.022^2 = 10.784$
$D(8, \overline{23}) = \frac{1 \times 2}{1+2} \times 2.830^2 = 5.339$
$D(9, \overline{23}) = \frac{1 \times 2}{1+2} \times 2.531^2 = 4.271$

d_{45} $D_{45} = 1.3778$
d_{46} $D_{46} = 2.398$
d_{47} $D_{47} = 1.1858$
d_{48} $D_{48} = 3.726$
d_{49} $D_{49} = 4.712$
d_{56} $D_{56} = 3.92$
d_{57} $D_{57} = 4.774$
d_{58} $D_{58} = 6.845$
d_{59} $D_{59} = 7.22$
d_{67} $D_{67} = 2.226$
d_{68} $D_{68} = 0.832$
d_{69} $D_{69} = 1.462$
d_{78} $D_{78} = 2.1632$
d_{79} $D_{79} = 3.323$
d_{89} $D_{89} = 0.4418$

Merge 1, 7

3) 1,7 2,3 4 5 6 8 9

$\bar{y}_{17} = (-0.986 \quad -0.17145 \quad -0.4955 \quad 0.006 \quad -0.842)$

$D_{45} = 1.3778$
$D_{46} = 2.398$
$D_{48} = 3.726$
$P_{49} = 4.712$
$D_{56} = 3.92$
$D_{58} = 6.845$
$D_{59} = 7.22$
$D_{68} = 0.832$
$D_{69} = 1.462$
$D_{89} = 0.4418$

$D(4, \overline{17}) = \frac{2}{3} \times 1.662^2 = 1.841$
$D(5, \overline{17}) = \frac{2}{3} \times 3.237^2 = 6.985$
$D(6, \overline{17}) = \frac{2}{3} \times 2.323^2 = 3.598$
$D(8, \overline{17}) = \frac{2}{3} \times 2.105^2 = 2.954$
$D(9, \overline{17}) = \frac{2}{3} \times 2.558^2 = 4.362$

$D(\overline{23}, \overline{17}) = \frac{2 \times 2}{2+2} \times 4.042^2 = 16.338$

$D(4, \overline{23}) = \frac{1 \times 2}{1+2} \times 4.246^2 = 12.189$
$D(5, \overline{23}) = \frac{2}{5} \times 4.460^2 = 13.261$
$D(6, \overline{23}) = \frac{2}{5} \times 3.165^2 = 6.678$
$D(8, \overline{23}) = \frac{2}{3} \times 2.830^2 = 5.339$
$D(9, \overline{23}) = \frac{2}{3} \times 2.531^2 = 4.271$

Merge 8 9

4) 1,7 2,3 8,9 4 5 6

$D_{45} = 1.3778$

$D_{46} = 2.398$

$D_{56} = 392$

$D(4, \overline{7}) = 1.841$

$D(5, \overline{7}) = 6.985$

$D(6, \overline{7}) = 3.598$

$D(4, \overline{23}) = 12.189$

$D(5, \overline{23}) = 13.261$

$D(6, \overline{23}) = 6.678$

$D(4, \overline{89}) = \frac{1\times2}{1+2} \cdot 2.869^2 = 5.487$

$D(5, \overline{89}) = \frac{2}{3} \times 3.721^2 = 9.231$

$P(6, \overline{89}) = \frac{2}{3} \times 1.438^2 = 1.379214$

$D(\overline{7}, \overline{23}) = 16.338$

$D(\overline{7}, \overline{89}) = \frac{2\times2}{2+2} \times 2.295^2 = 5.267$

$D(\overline{23}\ \overline{89}) = 6.771$

$\overline{23} * (0.851 \quad -0.282 \quad 1.756 \quad 1.5965 \quad 1.4205)$

$d(\overline{7}, \overline{89}) = -0.986 \quad -0.7145 \quad -0.4955 \quad 0.006 \quad -0.842$

$* 0.7865 \quad -0.6665 \quad -0.4555 \quad 0.5395 \quad 0.5095$

$\sqrt{1.7725^2 + 0.108^2 + 0.04^2 + 0.5335^2 + 1.3515^2}$

$= 2.295$

$d(\overline{23}, \overline{89})$

$\sqrt{0.0645^2 + 0.3845^2 + 2.2115^2 + 1.057^2 + 0.911^2}$

$= 2.644$

Merge 8,9 6

5) 1,7 2,3 6,8,9 4 5

$D(\overline{7}, \overline{23}) = 16.338$

$D(\overline{7}, \overline{89}) = 5.267$

$P(\overline{7}, \overline{45}) = \frac{2\times2}{2+2} \cdot 2.435^2 = 5.929$

$D(\overline{7}, 6) = 3.598$

$D(\overline{23}, \overline{89}) = 6.771$

$D(\overline{23}\ \overline{45}) = 4.289^2 = 18.396$

$D(\overline{23}, 6) = 6.678$

$D(\overline{89}, \overline{45}) = 10.347$

$D(\overline{89}\ 6) = 1.3792$

$D_{45} = 1.37$

Merge 4 5

$d(\overline{89}, \overline{45})$

$0.7865 \quad -0.6665 \quad -0.4555 \quad 0.5395 \quad 0.5095$

$-1.1035 \quad 1.5075 \quad -0.5875 \quad 0.8405 \quad -0.8835$

$\sqrt{1.89^2 + 2.174^2 + 0.132^2 + 0.301^2 + 1.393^2}$

$\sqrt{10.34685}$

3.217.

6) 2,3 6,8,9 4,5 1,7

D(2̄3̄, 8̄9̄) = 6.771

D(2̄3̄ 4̄5̄) = 18.396

D(2̄3̄, 1̄6̄7̄) = 15.640

D(8̄9̄, 4̄5̄) = 10.347

D(8̄9̄, 1̄6̄7̄) = $\frac{2\times3}{2+3}$ × = 3.604

D(4̄5̄, 1̄6̄7̄) = 5.551

$\bar{y}_{167}$ -1.214 -0.950 -0.586 -0.042 -0.571

0.905 0.429 -0.435 0.428 -0.408

-0.758 -0.599 -0.405 0.049 -1.113

(-0.356 -0.373 -0.475 0.145 -0.697)

d(8̄9̄, 1̄6̄7̄) (0.17865 -0.6665 -0.4555 0.5395 0.5095)

$\sqrt{1.1425^2 + 0.2935^2 + 0.0195^2 + 0.3945^2 + 1.2065^2}$

$= \sqrt{3.00310}$

d(2̄3̄, 1̄6̄7̄)

0.851 -0.282 1.756 1.5965 1.4205

-0.356 -0.373 -0.475 0.145 -0.697

$\sqrt{1.207^2 + 0.091^2 + 2.231^2 + 1.4515^2 + 2.1175^2}$

d(4̄5̄, 1̄6̄7̄)

-1.1035 1.5075 -0.5875 0.8405 -0.8835

-0.356 -0.373 -0.475 0.145 -0.697

$\sqrt{0.7475^2 + 1.8805^2 + 0.1125^2 + 0.6955^2 + 0.1865^2}$

7) Merge 8,9 1,6,7

2,3 4,5 8,9,1,6,7

D(2̄3̄, 4̄5̄) = 18.396

D(2̄3̄, 8̄9̄1̄6̄7̄) = $\frac{2\times5}{2+5}$? = 13.230

D(4̄5̄, 8̄9̄1̄6̄7̄) = 9.622

$\bar{y}_{16789}$ = -0.356 -0.373 -0.475 0.145 -0.697

89 0.7865 -0.6665 -0.4555 0.5395 0.5095

→ (0.21525 -0.51975 -0.46525 0.34225 -0.09375)

d(2̄3̄, -) 0.851 -0.282 1.756 1.5965 1.4205

$\sqrt{0.63575^2 + 0.23775^2 + 2.22125^2 + 1.25425^2 + 1.51425^2}$

d(4̄5̄, → -1.1035 1.5075 -0.5875 0.8405 -0.8835

$\sqrt{1.31875^2 + 2.02725^2 + 0.12225^2 + 0.49825^2 + 0.78975^2}$

Merge 4,5, 8,9,1,6,7

8) 2,3 , 1,4,5,6,7,8,9

D(2̄3̄, 1̄4̄5̄6̄7̄8̄9̄) $\frac{2\times4}{2+4}$ × □ = 18.891

$\bar{y}_{45.16789}$

→ -0.444 0.494 -0.526 0.591 -0.489

23 0.851 -0.282 1.756 1.5965 1.4205

$1.295^2 + 0.1976^2 + 2.282^2 + 1.0055^2 + 1.9095^2$

g	wards									
9	1	2	3	4	5	6	7	8	9	0
8	[2,3]	1	4	5	6	7	8	9		0.05445
7	[1,7]	[2,3]	4	5	6	8	9			0.3362
6	[8,9]	[1,7]	[2,3]	4	5	6				0.4418
5	[6,8,9]	[1,7]	[2,3]	4	5					1.3418
4	[6,8,9]	[1,7]	[2,3]	[4,5]						1.38
3	[1,6,7,8,9]	[2,3]	[4,5]							5.82
2	[1,4,5,6,7,8,9]	[2,3]								8.85
1	[1,2,3,4,5,6,7,8,9]									21.14

The SAS System

Obs	bullet	antimony	copper	arsenic	bismuth	silver
1	1	4123.33	152.667	68.33	112.000	36.6667
2	2	28215.00	250.667	1213.33	16.000	66.0000
3	3	27239.00	248.667	1242.33	17.000	71.0000
4	4	4746.67	409.333	26.00	139.000	30.3333
5	5	6020.00	610.000	109.67	193.667	33.0000
6	6	28348.00	353.000	143.33	141.000	39.2667
7	7	9330.00	203.667	158.00	117.667	28.0000
8	8	26725.00	203.667	175.67	123.000	50.6000
9	9	27255.33	184.000	90.00	172.667	57.2667

The SAS System

Obs	bullet	antimony	copper	arsenic	bismuth	silver
1	1	-1.21390	-0.94959	-0.58618	-0.04337	-0.57098
2	2	0.89354	-0.27506	1.72673	-1.60455	1.26432
3	3	0.80817	-0.28883	1.78531	-1.58829	1.57716
4	4	-1.15937	0.81703	-0.67169	0.39571	-0.96724
5	5	-1.04798	2.19822	-0.50267	1.28472	-0.80040
6	6	0.90518	0.42929	-0.43468	0.42824	-0.40831
7	7	-0.75844	-0.59856	-0.40505	0.04879	-1.11323
8	8	0.76320	-0.59856	-0.36936	0.13552	0.30079
9	9	0.80960	-0.73393	-0.54241	0.94322	0.71790

The SAS System

The CLUSTER Procedure
Single Linkage Cluster Analysis

Eigenvalues of the Covariance Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	3.27389429	2.39801141	0.6548	0.6548
2	0.87588288	0.17076057	0.1752	0.8300
3	0.70512231	0.57428630	0.1410	0.9710
4	0.13083601	0.11657150	0.0262	0.9971
5	0.01426451		0.0029	1.0000

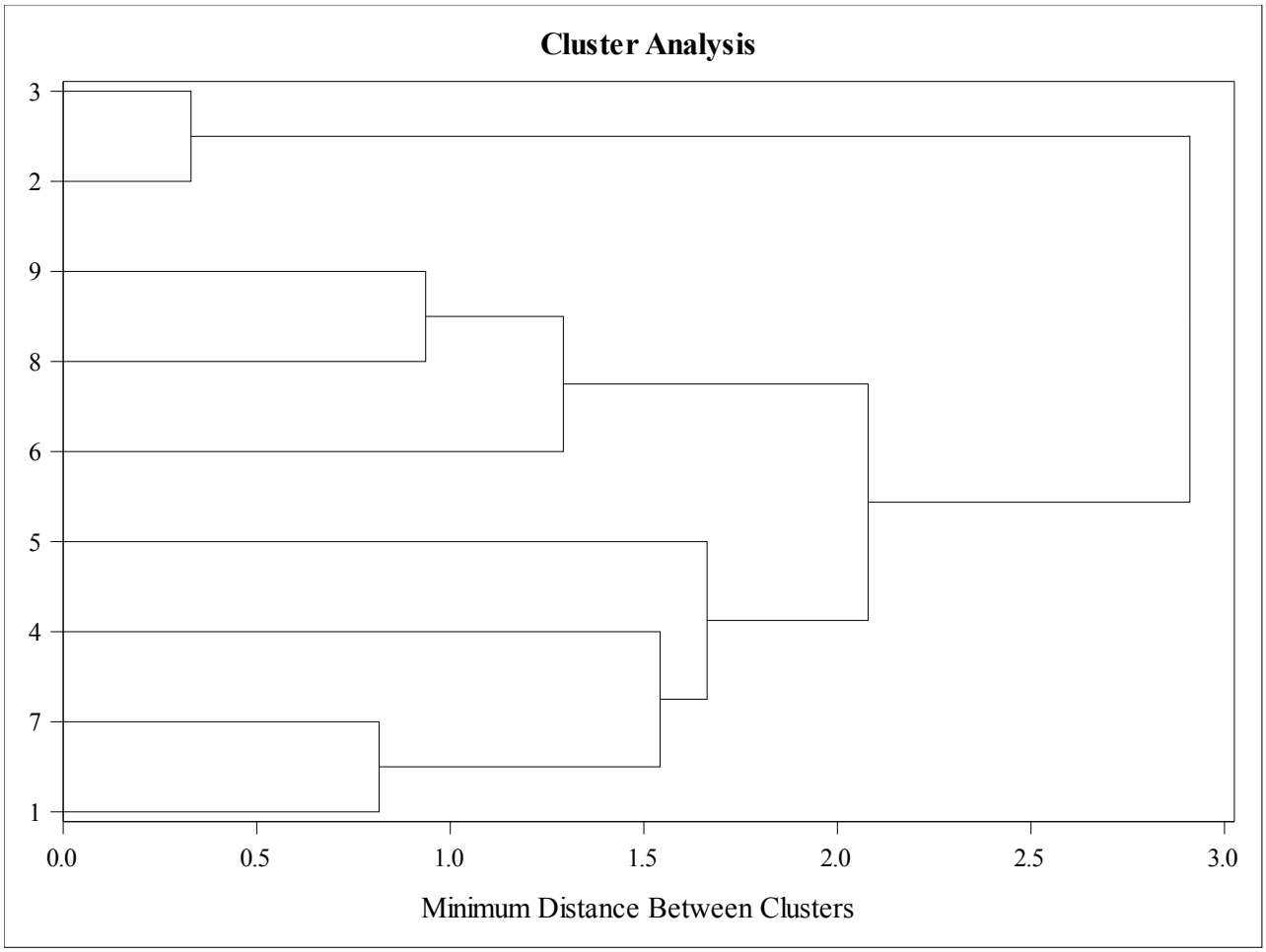
Root-Mean-Square Total-Sample Standard Deviation	1

Single Linkage

Cluster History					
Number of Clusters	**Clusters Joined**		**Freq**	**Min Dist**	**Tie**
8	2	3	2	0.3302	
7	1	7	2	0.8161	
6	8	9	2	0.9364	
5	6	CL6	3	1.2921	
4	CL7	4	3	1.5419	
3	CL4	5	4	1.6634	
2	CL3	CL5	7	2.0793	
1	CL2	CL8	9	2.9106	

The SAS System

The CLUSTER Procedure
Single Linkage Cluster Analysis



The SAS System

The CLUSTER Procedure
Complete Linkage Cluster Analysis

Eigenvalues of the Covariance Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	3.27389429	2.39801141	0.6548	0.6548
2	0.87588288	0.17076057	0.1752	0.8300
3	0.70512231	0.57428630	0.1410	0.9710
4	0.13083601	0.11657150	0.0262	0.9971
5	0.01426451		0.0029	1.0000

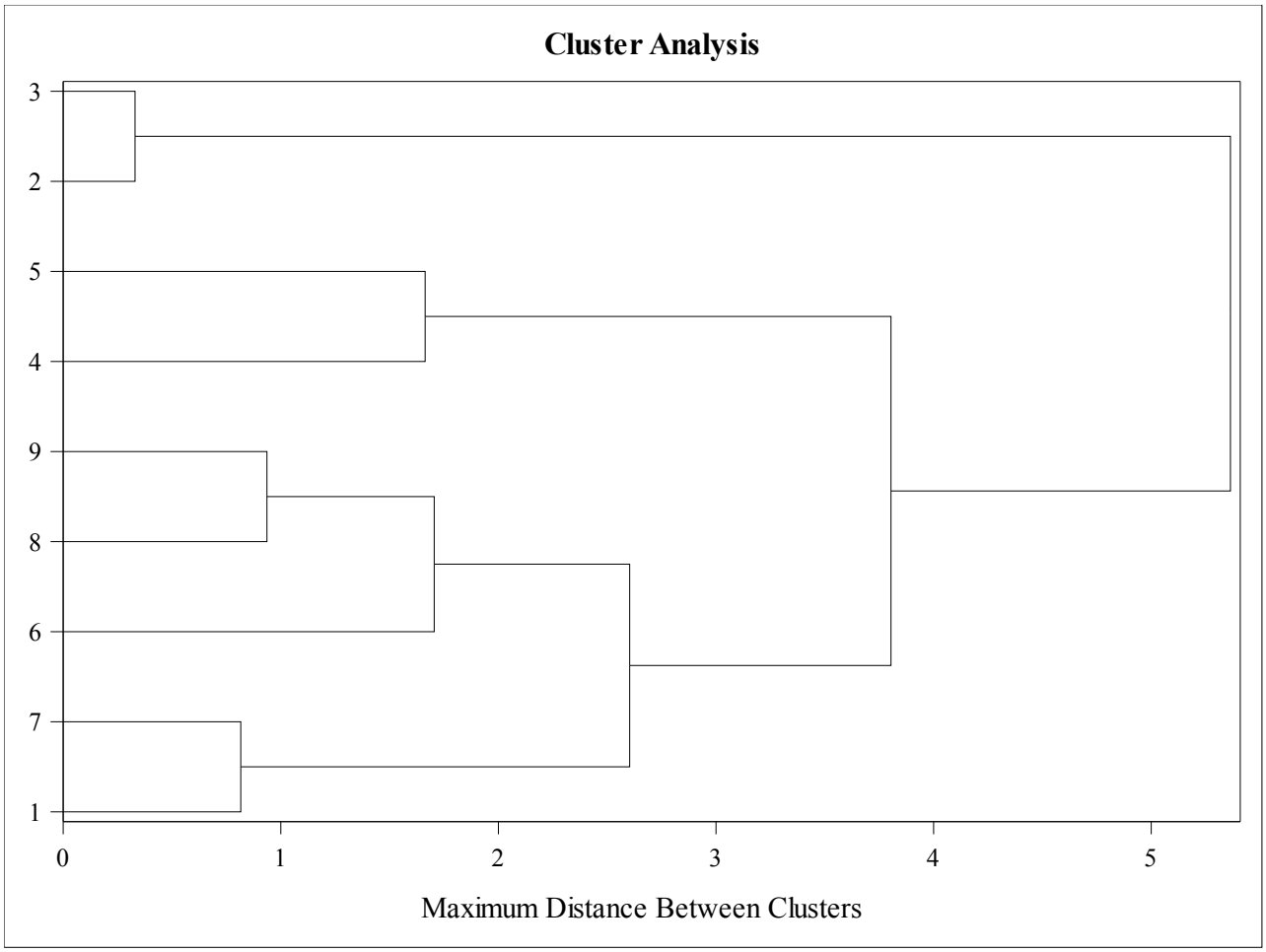
Root-Mean-Square Total-Sample Standard Deviation	1

Cluster History					
Number of Clusters	**Clusters Joined**		**Freq**	**Maximum Distance**	**Tie**
8	2	3	2	0.3302	
7	1	7	2	0.8161	
6	8	9	2	0.9364	
5	4	5	2	1.6634	
4	6	CL6	3	1.7051	
3	CL7	CL4	5	2.6034	
2	CL3	CL5	7	3.8041	
1	CL2	CL8	9	5.364	

Complete Linkage

The SAS System

The CLUSTER Procedure
Complete Linkage Cluster Analysis



The SAS System

The CLUSTER Procedure
Average Linkage Cluster Analysis

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	3.27389429	2.39801141	0.6548	0.6548
2	0.87588288	0.17076057	0.1752	0.8300
3	0.70512231	0.57428630	0.1410	0.9710
4	0.13083601	0.11657150	0.0262	0.9971
5	0.01426451		0.0029	1.0000

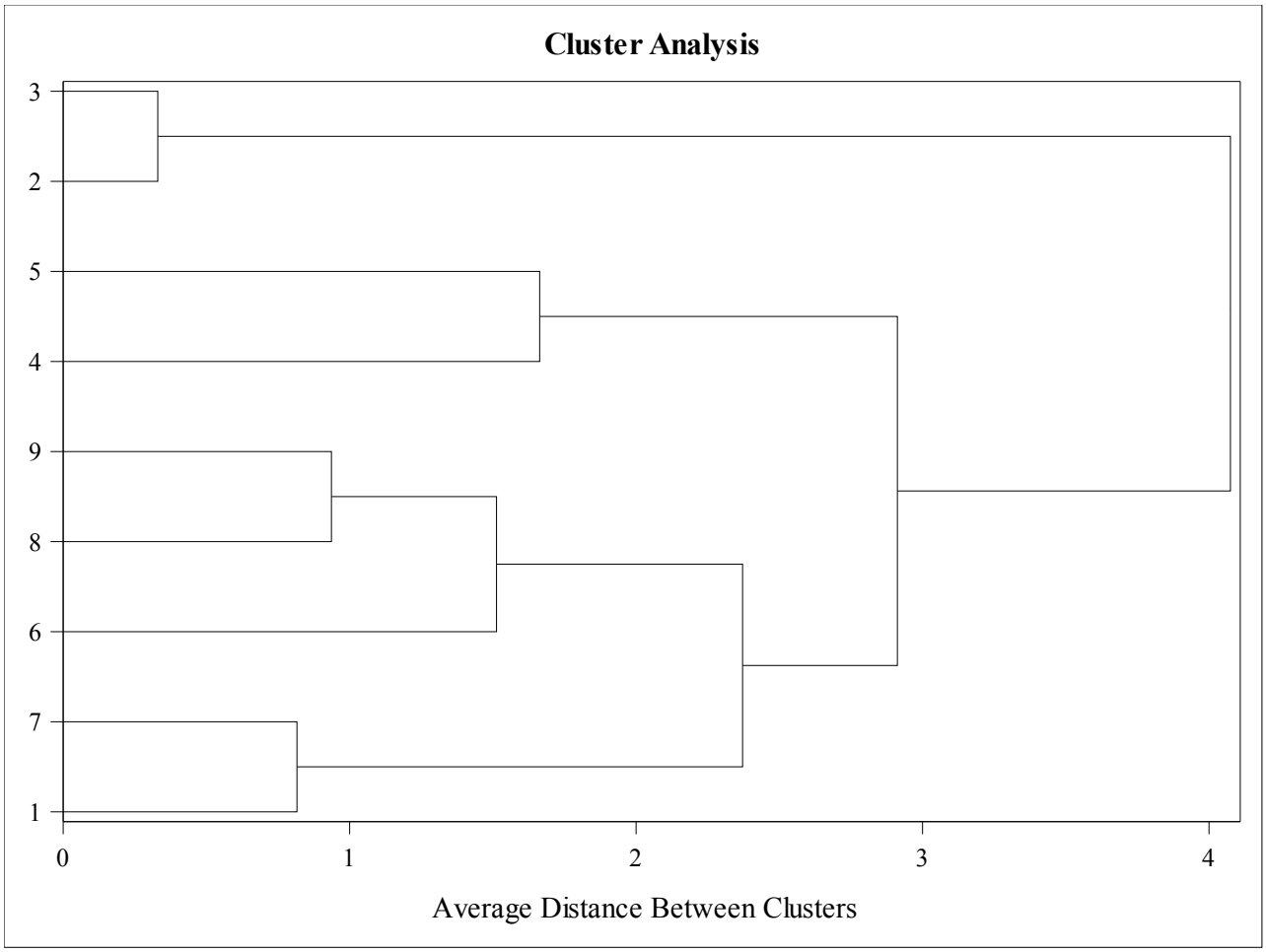
Root-Mean-Square Total-Sample Standard Deviation	1

Cluster History					
Number of Clusters	Clusters Joined		Freq	RMS Distance	Tie
8	2	3	2	0.3302	
7	1	7	2	0.8161	
6	8	9	2	0.9364	
5	6	CL6	3	1.5127	
4	4	5	2	1.6634	
3	CL7	CL5	5	2.3717	
2	CL3	CL4	7	2.9122	
1	CL2	CL8	9	4.0747	

Average Linkage

The SAS System

The CLUSTER Procedure
Average Linkage Cluster Analysis



The SAS System

The CLUSTER Procedure
Centroid Hierarchical Cluster Analysis

Eigenvalues of the Covariance Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	3.27389429	2.39801141	0.6548	0.6548
2	0.87588288	0.17076057	0.1752	0.8300
3	0.70512231	0.57428630	0.1410	0.9710
4	0.13083601	0.11657150	0.0262	0.9971
5	0.01426451		0.0029	1.0000

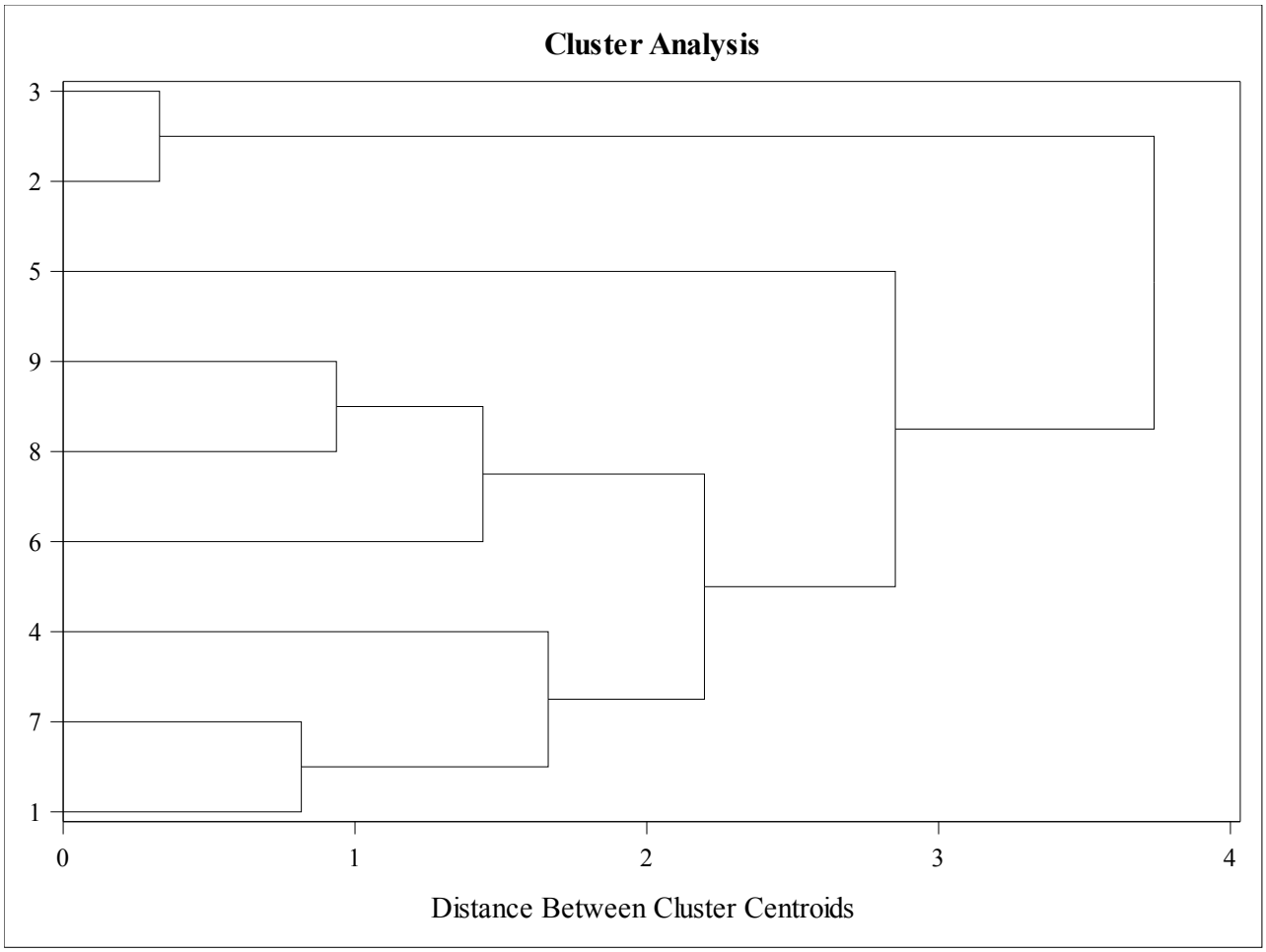
Root-Mean-Square Total-Sample Standard Deviation	1

Centroid Hierarchical

Cluster History					
Number of Clusters	Clusters Joined		Freq	Centroid Distance	Tie
8	2	3	2	0.3302	
7	1	7	2	0.8161	
6	8	9	2	0.9364	
5	6	CL6	3	1.4385	
4	CL7	4	3	1.6621	
3	CL4	CL5	6	2.1975	
2	CL3	5	7	2.8517	
1	CL2	CL8	9	3.7383	

The SAS System

The CLUSTER Procedure
Centroid Hierarchical Cluster Analysis



The SAS System

The CLUSTER Procedure
Ward's Minimum Variance Cluster Analysis

Eigenvalues of the Covariance Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	3.27389429	2.39801141	0.6548	0.6548
2	0.87588288	0.17076057	0.1752	0.8300
3	0.70512231	0.57428630	0.1410	0.9710
4	0.13083601	0.11657150	0.0262	0.9971
5	0.01426451		0.0029	1.0000

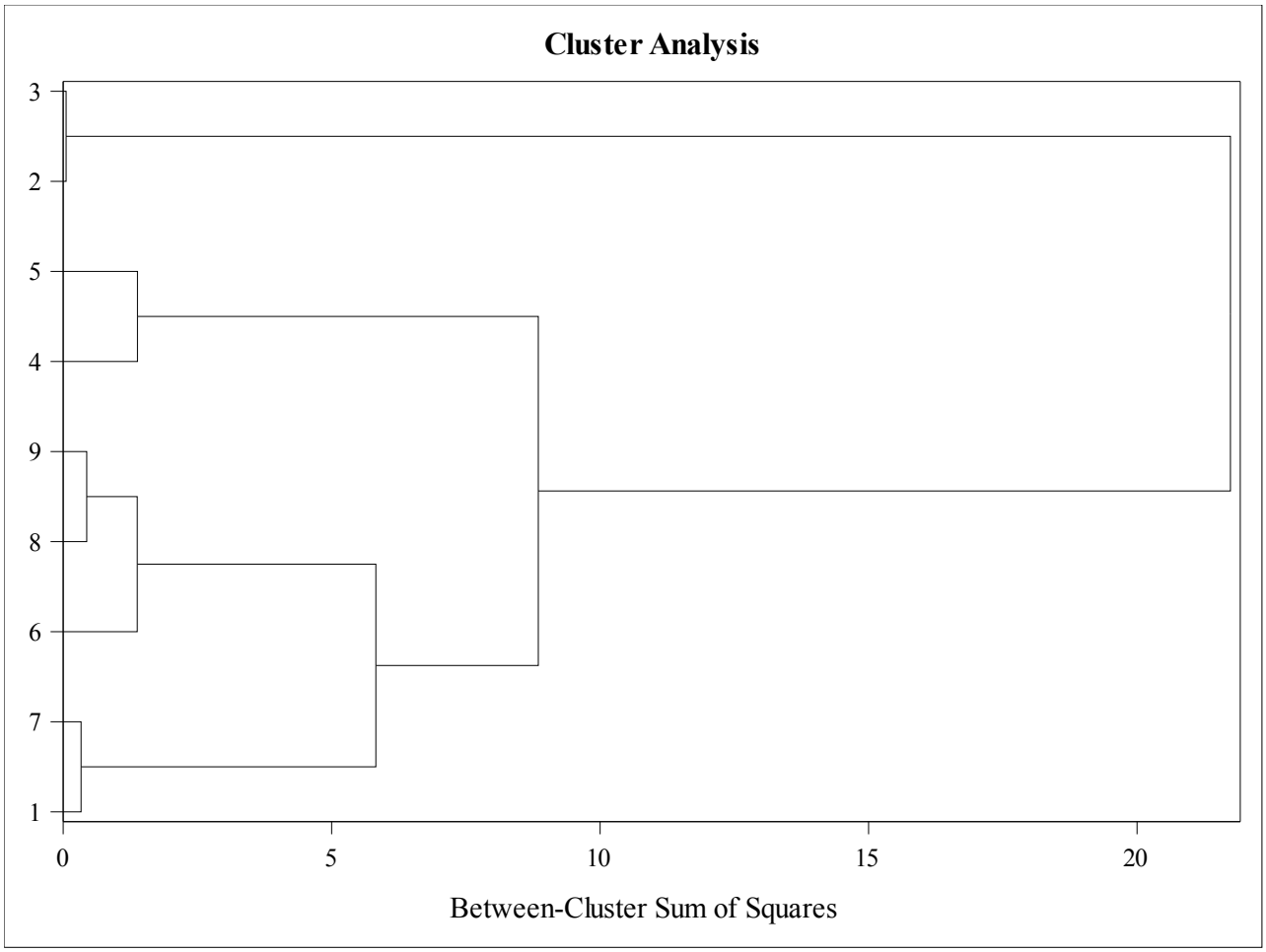
Root-Mean-Square Total-Sample Standard Deviation	1

ward's

Cluster History							
Number of Clusters	**Clusters Joined**		**Freq**	**Semipartial R-Square**	**R-Square**	**Between Cluster Sum of Squares**	**Tie**
8	2	3	2	0.0014	.999	0.0545	
7	1	7	2	0.0083	.990	0.333	
6	8	9	2	0.0110	.979	0.4384	
5	6	CL6	3	0.0345	.945	1.3795	
4	4	5	2	0.0346	.910	1.3834	
3	CL7	CL5	5	0.1456	.765	5.8232	
2	CL3	CL4	7	0.2212	.543	8.8489	
1	CL2	CL8	9	0.5435	.000	21.739	

The SAS System

The CLUSTER Procedure
Ward's Minimum Variance Cluster Analysis



part 3: Summaries of Results

The Results are checked by SAS code output.

See Attachments for SAS code.

Part 3

Single-Linkage

g	Clusters	Distance
9	[1] [2] [3] [4] [5] [6] [7] [8] [9]	0
8	[1] [2,3] [4] [5] [6] [7] [8] [9]	0.33
7	[1,7] [2,3] [4] [5] [6] [8] [9]	0.82
6	[1,7] [2,3] [8,9] [4] [5] [6]	0.94
5	[1,7] [2,3] [6,8,9] [4] [5]	1.29
4	[1,4,7] [2,3] [6,8,9] [5]	1.54
3	[1,4,5,7] [2,3] [6,8,9]	1.66
2	[1,4,5,6,7,8,9] [2,3]	2.08
1	[1,2,3,4,5,6,7,8,9]	2.91

✓ code checked with correct.

Complete Linkage

g	Clusters	Distance
9	[1] [2] [3] [4] [5] [6] [7] [8] [9]	0
8	[1] [2,3] [4] [5] [6] [7] [8] [9]	0.33
7	[1,7] [2,3] [4] [5] [6] [8] [9]	0.82
6	[1,7] [2,3] [8,9] [4] [5] [6]	0.94
5	[1,7] [2,3] [8,9] [4,5] [6]	1.66
4	[1,7] [2,3] [6,8,9] [4,5]	1.71
3	[1,6,7,8,9] [2,3] [4,5]	2.60
2	[1,4,5,6,7,8,9] [2,3]	3.80
1	[1,2,3,4,5,6,7,8,9]	5.36

✓ checked with SAS code output correct.

Average Linkage

g	Clusters	Distance
9	[1] [2] [3] [4] [5] [6] [7] [8] [9]	
8	[1] [2,3] [4] [5] [6] [7] [8] [9]	0.33
7	[1,7] [2,3] [4] [5] [6] [8] [9]	0.82
6	[1,7] [2,3] [8,9] [4] [5] [6]	0.94
5	[1,7] [2,3] [6,8,9] [4] [5]	1.5
4	[1,7] [2,3] [6,8,9] [4,5]	1.66
3	[1,7,6,8,9] [2,3] [4,5]	2.37
2	[1,4,5,6,7,8,9] [2,3]	2.91
1	[1,2,3,4,5,6,7,8,9]	4.014

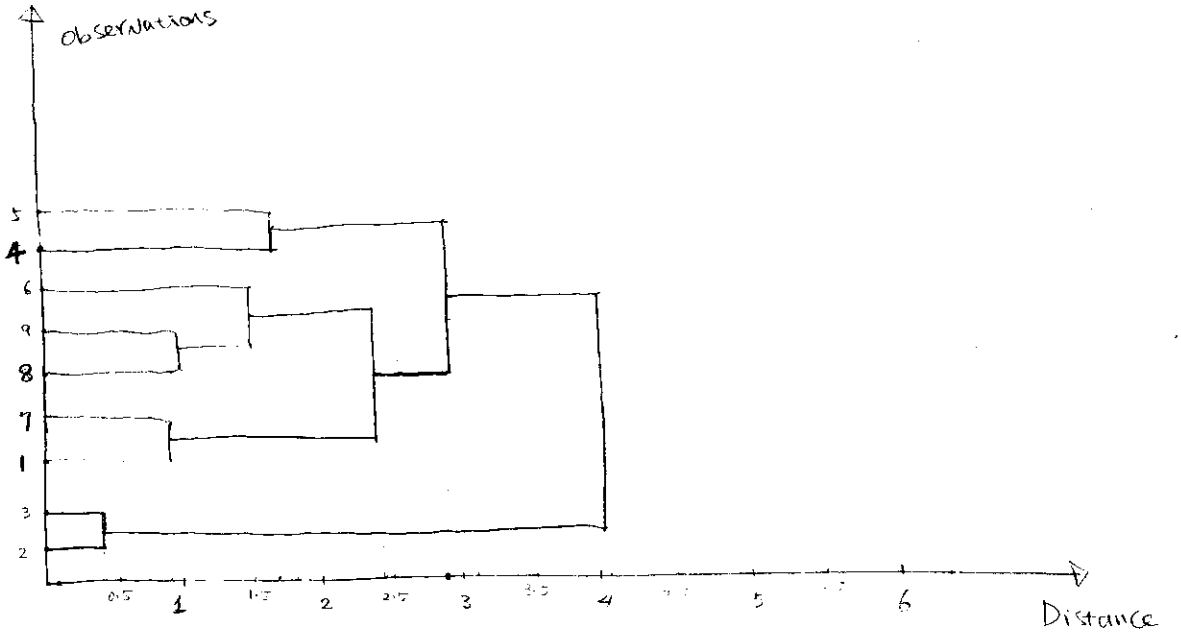
Centroid

9	clusters	
9	[1] [2] [3] [4] [5] [6] [7] [8] [9]	0
8	[1] [2,3] [4] [5] [6] [7] [8] [9]	0.33
7	[1,7] [2,3] [4] [5] [6] [8] [9]	0.82
6	[1,7] [2,3] [8,9] [4] [5] [6]	0.94
5	[1,7] [2,3] [6,8,9] [4] [5]	1.438
4	[1,4,7] [2,3] [6,8,9] [5]	1.66
3	[1,4,7,6,8,9] [2,3] [5]	2.19
2	[1,4,5,6,8,9] [2,3]	2.85
1	[1,2,3,4,5,6,7,8,9]	3.143

Ward's

g		Distance
9	[1] [2] [3] [4] [5] [6] [7] [8] [9]	0
8	[2 3] [1] [4] [5] [6] [7] [8] [9]	0.05445
7	[1, 7] [2, 3] [4] [5] [6] [8] [9]	0.3362
6	[1, 7] [2, 3] [8, 9] [4] [5] [6]	0.4418
5	[1, 7] [2, 3] [6, 8, 9] [4] [5]	1.3778
4	[1, 7] [2, 3] [6, 8, 9] [4, 5]	1.38
3	[1, 6, 7, 8, 9] [2, 3] [4, 5]	5.87
2	[1, 4, 5, 6, 7, 8, 9] [2, 3]	8.85
1	[1, 2, 3, 4, 5, 6, 7, 8, 9]	21.74

3. Dendrogram of clusters Analysis using Average Linkage.



4. $g=3$

Single Linkage : [1,4,5,7] [2,3] [6,8,9]

Complete Linkage : [1,6,7,8,9] [2,3] [4,5]

Average : [1,6,7,8,9] [2,3] [4,5]

Centroid : [1,4,6,7,8,9] [2,3] [5]

Ward's : [1,6,7,8,9] [2,3] [4,5]

5. $g=4$

Single Linkage : [1,4,7] [2,3] [6,8,9] [5]

Complet Linkage : [1,7] [2,3] [6,8,9] [4,5]

Average : [1,7] [2,3] [6,8,9] [4,5]

Centroid : [1,4,7] [2,3] [6,8,9] [5]

Ward's : [1,7] [2,3] [6,8,9] [4,5]

6. Yes. There are similarities.

when $g=4$, the four clusters are almost identical by five methods.

when $g=3$, [2,3] is the common cluster by all the methods.

7.

$d_{05} = 5.24$

$d_{29} = 3.49$

$d_{59} = 3.80$

CODE: radius = 5.24/2
maxC = 2
Iteration = 10

See Attached SAS code and output.

The SAS System

Obs	bullet	antimony	copper	arsenic	bismuth	silver
1	1	4123.33	152.667	68.33	112.000	36.6667
2	2	28215.00	250.667	1213.33	16.000	66.0000
3	3	27239.00	248.667	1242.33	17.000	71.0000
4	4	4746.67	409.333	26.00	139.000	30.3333
5	5	6020.00	610.000	109.67	193.667	33.0000
6	6	28348.00	353.000	143.33	141.000	39.2667
7	7	9330.00	203.667	158.00	117.667	28.0000
8	8	26725.00	203.667	175.67	123.000	50.6000
9	9	27255.33	184.000	90.00	172.667	57.2667

The SAS System

Obs	bullet	antimony	copper	arsenic	bismuth	silver
1	1	-1.21390	-0.94959	-0.58618	-0.04337	-0.57098
2	2	0.89354	-0.27506	1.72673	-1.60455	1.26432
3	3	0.80817	-0.28883	1.78531	-1.58829	1.57716
4	4	-1.15937	0.81703	-0.67169	0.39571	-0.96724
5	5	-1.04798	2.19822	-0.50267	1.28472	-0.80040
6	6	0.90518	0.42929	-0.43468	0.42824	-0.40831
7	7	-0.75844	-0.59856	-0.40505	0.04879	-1.11323
8	8	0.76320	-0.59856	-0.36936	0.13552	0.30079
9	9	0.80960	-0.73393	-0.54241	0.94322	0.71790

The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=2.62 Maxclusters=3 Maxiter=10 Converge=0.02

Initial Seeds					
Cluster	**antimony**	**copper**	**arsenic**	**bismuth**	**silver**
1	0.809595115	-0.733929622	-0.542407884	0.943216287	0.717902023
2	0.893543053	-0.275062623	1.726726969	-1.604547505	1.264319933
3	-1.047983203	2.198215839	-0.502674339	1.284724324	-0.800396844

Minimum Distance Between Initial Seeds =	3.486582

Iteration History				
		Relative Change in Cluster Seeds		
Iteration	**Criterion**	**1**	**2**	**3**
1	0.6681	0.3898	0.0474	0.2385
2	0.4573	0	0	0

Convergence criterion is satisfied.

Criterion Based on Final Seeds =	0.4573

Cluster Summary						
Cluster	**Frequency**	**RMS Std Deviation**	**Maximum Distance from Seed to Observation**	**Radius Exceeded**	**Nearest Cluster**	**Distance Between Cluster Centroids**
1	5	0.6314	1.4835		3	2.4888
2	2	0.1044	0.1651		1	3.4395
3	2	0.5260	0.8317		1	2.4888

Statistics for Variables				
Variable	**Total STD**	**Within STD**	**R-Square**	**RSQ/(1-RSQ)**
antimony	1.00000	0.82308	0.491906	0.968138
copper	1.00000	0.59067	0.738334	2.821665
arsenic	1.00000	0.09156	0.993712	158.037440
bismuth	1.00000	0.41503	0.870812	6.740643

The SAS System

The FASTCLUS Procedure
Replace=FULL Radius=2.62 Maxclusters=3 Maxiter=10 Converge=0.02

Statistics for Variables				
Variable	**Total STD**	**Within STD**	**R-Square**	**RSQ/(1-RSQ)**
silver	1.00000	0.60140	0.728737	2.686457
OVER-ALL	1.00000	0.56012	0.764700	3.249894

Pseudo F Statistic =	9.75

Approximate Expected Over-All R-Squared =	.

Cubic Clustering Criterion =	.

WARNING: The two values above are invalid for correlated variables.

Cluster Means					
Cluster	**antimony**	**copper**	**arsenic**	**bismuth**	**silver**
1	0.101128466	-0.490271087	-0.467534292	0.302478908	-0.214766834
2	0.850854844	-0.281945594	1.756017076	-1.596416361	1.420737871
3	-1.103676008	1.507623312	-0.587181347	0.840219090	-0.883820787

Cluster Standard Deviations					
Cluster	**antimony**	**copper**	**arsenic**	**bismuth**	**silver**
1	1.006840085	0.533712877	0.092606602	0.399438318	0.725819059
2	0.060370245	0.009733990	0.041422466	0.011499174	0.221208369
3	0.078761521	0.976645318	0.119510956	0.628625331	0.117979271

The SAS System

Obs	bullet	CLUSTER	DISTANCE
1	8	1	0.86797
2	6	1	1.24356
3	7	1	1.27518
4	9	1	1.35916
5	1	1	1.48352
6	2	2	0.16511
7	3	2	0.16511
8	4	3	0.83169
9	5	3	0.83169

```
ods rtf file='hw6_sas_f23.rtf';

/*Read in original data set*/
data bullets0;
  infile 'bullet.csv' firstobs=2 dsd;
  input bullet antimony copper arsenic bismuth silver;
run;

/*Print original data set*/
proc print data=bullets0;
run;

/*Standardize data set*/
proc standard data=bullets0 out=bullets mean=0 std=1;
var antimony copper arsenic bismuth silver;
run;

/*Print standardized data set to make sure it matches data table on homework assignment*/
proc print data=bullets;
run;

/*Insert SAS code for Homework 6 here*/
/* Single Linkage */
proc cluster data=bullets outtree=qsingle method=single nonorm;
var antimony copper arsenic bismuth silver;
id bullet;
run;

/* Complete Linkage */
proc cluster data=bullets outtree=qcomplete method=complete nonorm;
var antimony copper arsenic bismuth silver;
id bullet;
run;

/* Average Linkage */
proc cluster data=bullets outtree=qaverage method=average nonorm;
var antimony copper arsenic bismuth silver;
id bullet;
run;

/* Centroid Method */
proc cluster data=bullets outtree=qcentroid method=centroid nonorm;
var antimony copper arsenic bismuth silver;
id bullet;
run;

/* Ward's Method */
proc cluster data=bullets outtree=qward method=ward nonorm;
var antimony copper arsenic bismuth silver;
id bullet;
run;


/* K- means clutering radius=6 3 cluters*/
proc fastclus data=bullets radius=2.62 maxc=3 replace=full maxiter=10 out=Clus_OUT;
var antimony copper arsenic bismuth silver;
id bullet;
run;

proc sort data=Clus_OUT;
by cluster distance;
run;
```

```
proc print data=Clus_OUT;
var bullet cluster distance;
run;


ods rtf close;
```