Clustering Analysis

1. What is the purpose of a cluster analysis?
2. What is the difference between agglomerative and divisive clustering?
3. What is the difference between hierarchical and non-hierarchical clustering?
4. How do we determine the appropriate number of clusters? Give two different visualization methods that are used to display the outcome of a cluster analysis.
5. In the context of hierarchical cluster analyses, distinguish between agglomerative clustering and divisive clustering.
6. What is a vertical icicle plot used for? Give a brief description, supporting your answer with sketches.
7. Explain why squared Euclidean Distance may be used in preference to Euclidean Distance.
8. Compute the squared Euclidean distance and the Manhattan (a.k.a. city block) distance between the following points A and B.

$$\mathrm{A} = \{2, 5, 1, 7\}$$

$$\mathrm{B} = \{1, 6, 3, 6\}$$

9. What is a dendrogram? What does does a dendrogram depict?
10. What is two-step clustering analysis?
11. Compare and contrast any three linkage methods.
12. Explain the difference between Ward's method and K-means clustering.
13. Discriminant Analysis is very similar to Clustering Analysis, but one key difference. Explain this difference.
14. What is the difference between a linkage method and a similarity measure? Discuss the purpose of both.
15. Explain the process of k-means clustering, starting with initial cluster allocation. You may work on the basis of a two-cluster solution. Support your answer with several sketches.
16. What is the difference between a linkage method and a distance measure?
17. Compare and contrast any three linkage methods. Support your answer with sketches

18. Interpret the following output for a Clustering Procedure. You may limit your anser to the first six cases.

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	1	2	24.098	0	0	4
2	4	9	46.005	0	0	7
3	5	8	56.649	0	0	5
4	1	6	57.980	1	0	6
5	3	5	88.417	0	3	8
6	1	7	89.622	4	0	7
7	1	4	109.030	6	2	8
8	1	3	169.328	7	5	0

19. Interpret the following output for a Clustering Procedure. You may limit your anser to the first six cases.

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	11	12	.359	0	0	6
2	16	17	.366	0	0	19
3	1	6	.606	0	0	9
4	13	14	.748	0	0	10
5	2	5	.854	0	0	22
6	11	22	.900	1	0	24
7	24	29	1.013	0	0	8
8	8	24	1.031	0	7	10
9	1	31	1.387	3	0	18
10	8	13	1.455	8	4	17
11	34	35	1.470	0	0	14
12	3	7	1.496	0	0	25
13	15	18	1.549	0	0	15
14	30	34	1.569	0	11	28
15	15	32	1.599	13	0	19
16	19	25	1.600	0	0	22
17	8	10	1.632	10	0	18
18	1	8	1.640	9	17	20
19	15	16	1.673	15	2	20
20	1	15	1.722	18	19	21
21	1	23	1.860	20	0	23

20. Compare and contrast k-means clustering and hierarchical clustering in terms of the number of cluster determined.

21. For a 4 cluster solution, Interpret the ANOVA table below.

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Net profit	495,145	3	1419,744	3	,349	,795
Own funds	2878,202	3	2537,200	3	1,134	,460
Assets	842788,443	3	9987,138	3	84,387	,002
Client deposits	634017,636	3	35643,498	3	17,788	,021
Loans	957411,333	3	37401,709	3	25,598	,012