Assignment # 4

1) Complete all parts of this problem without the use of a computer to make sure that you understand the details of the clustering algorithms. Consider the following “data” to be clustered as described below.

10 20 40 80 85 121 160 168 195

For each part of the problem, assume that Euclidean distance will be used to measure the distance between the data points.

(a) Use hierarchical agglomerative clustering with single linkage to cluster the data. Draw a dendrogram to illustrate your clustering and include a vertical axis with numerical labels indicating the height of each parental node in the dendrogram.

(b) Repeat part (a) using hierarchical agglomerative clustering with complete linkage.

(c) If two clusters are desired, what data points would be clustered together according to the single linkage method used in part (a)?

(d) If two clusters are desired, what data points would be clustered together according to the complete linkage method used in part (b)?

(e) Use the K-means algorithm with K=3 to cluster the data set. Suppose that the points 160, 168, and 195 were selected as the initial cluster means. Work from these initial values to determine the final clustering for the data. Show your work so that it will be easy to see each step you took to get from the initial values to your final clustering.

(f) verify your results with R.

(g) Show that the clustering produced by the K-means algorithm depends on starting mean values by providing a different set of three starting means that results in a different set of final clusters. You need to provide only the starting means and the final clustering. It is not necessary to show your work

2) Read the paper Fu, Luyang, and Hongyuan Wang, "Estimating Insurance Attrition Using Survival Analysis," Variance 8:1, 2014, pp. 55-72.

http://www.variancejournal.org/issues/08-01/55.pdf answer the following questions

1. 1What is the problem addressed by the paper?
2. Define Survival functions and hazard functions in the context of the problem given in the paper.
3. What is the input data used in the study?How was the data preprocessed?
4. What are the outcomes and conclusions of the experiment? The paper in general?