Problem Solving

Assignment 2

Data Mining (CSE4052)

1. Suppose that the data for analysis includes the attribute *age*. The *age* values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
   (a) What is the *mean* of the data? What is the *median*?
   (b) What is the *mode* of the data? Comment on the data's modality (i.e., bimodal, trimodal..)
   (c) What is the *midrange* of the data?
   (d) Can you find (roughly) the first quartile (*Q*1) and the third quartile (*Q*3) of the data? What is the interquartile range?

2. Suppose a group of 12 sales price records has been sorted as follows:
   5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215.

Partition them into three bins by each of the following methods.
(a) equal-frequency (equidepth) partitioning
(b) equal-width partitioning
(c) clustering

3. Use *smoothing by bin means, median, and boundaries* to smooth the following data, using a bin depth of 6.
   Data: 11,13,13,15,15,16,19,20,20,20,21,21,22,23,24,30,40,45,45,45,71,72,73,75

4. Find Q1, Q2, and Q3 for the following data set, and draw a box-and-whisker plot.
   {2,6,7,8,8,11,12,13,14,15,22,23}

5. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):
(a) Compute the Euclidean distance between the two objects.
(b) Compute the Manhattan distance between the two objects.
(c) Compute the Minkowski distance between the two objects, using h = 3.

6. Use the methods below to normalize the following group of data
   200, 300, 400, 600, 1000
 a) min-max normalization by setting min = 0 and max = 1
 b) z-score normalization
 c) z-score normalization using the mean absolute deviation instead of standard deviation
 d) normalization by decimal scaling

7. Compute the pearson correlation of the following data-

Weight (kg)	Length (cm)
3.63	53.1
3.02	49.7
3.82	48.4
3.42	54.2
3.59	54.9
2.87	43.7
3.03	47.2
3.46	45.2
3.36	54.4
3.3	50.4

8. Perform the chi-square test for correlation for the following observation of survey where 256 peopople shared the month of their birth where the expected distribution of moths are evenly distributed.

January	29
February	24
March	22
April	19
May	21
June	18
July	19
August	20
September	23
October	18
November	20
December	23