CSCI 5080 Assignment 2

Use Microsoft **WORD** to write your answer and submit it to the Dropbox in D2L.
Please indicate **how much time** you spend on each problem.

1. (15 points) Given the following data (in increasing order) for the attribute *age*: 11, 13, 15, 17, 19, 21, 21, 23, 23, 23, 23, 25, 27, 30, 33, 33, 33, 33, 36, 36, 38, 40, 46, 48, 54, use *smoothing by bin* means to smooth the above data, using a bin depth of 5. Illustrate your steps.

2. (30 points) Use the methods below to *normalize* the following group of data:

100, 200, 400, 700, 1100

(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

3. (20 points) Using the data for *age* given in Question 1, answer the following:

(a) Use min-max normalization to transform the value 25 for age onto the range [0.0, 1.0].
(b) Use z-score normalization to transform the value 25 for age. You may use Excel to calculate the standard deviation.
(c) Use normalization by decimal scaling to transform the value 25 for age.
(d) Comment on which method you would prefer to use for the given data, giving reasons as to why.

4. (20 points) Using the data for *age* and *body fat* given in the following table, calculate their *correlation coefficient* using Eq. 3.4 and 3.5. Illustrate your steps. Are these two attributes positively or negatively correlated?

Check the example of computing the *covariance* is on page 98. You may reuse the standard deviation values calculated in Question 2 of Assignment 1.

age	20	22	25	25	36	40	45	48	49
%fat	8.4	25.3	7.6	18.8	27.5	24.6	28.1	28.8	30.2

age	51	53	53	57	58	59	60	61	62
%fat	32.7	40.2	29.8	32.3	30.7	33.9	40.1	33.1	36.4

5. (15 points) Using the data for *age* given in Question 1, plot an equal-width histogram of width 5.