Data Mining & Predictive Analytics (CSE 4859)

Quiz 1

Date: September 26, 2025 **Section: 21**

Maximum Marks: 15 **Time: 50 minutes**

Note: *Use of mobile phones is strictly prohibited during the test. A non-programmable calculator may be used for computations, if needed.*

---

1. What is Data Mining? Discuss how data mining can help in improving the business in various sectors. [2]

2. Suppose that a hospital tested the age and body fat data for some randomly selected adults with the following results: [2+3+2]

age	23	23	27	27	39	41	47	49	50
% fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2

   (a) Calculate the mean, median, and standard deviation of age and % fat.

   (b) Calculate the correlation coefficient between these two attributes. Compute their covariance.

   (c) Transform the above numeric attributes using the z-score normalization.

3. If the median of a distribution given below is 28.5, then find the value of x and y. [2]

class	0-10	10-20	20-30	30-40	40-50	50-60	Total
frequency	5	x	20	15	y	5	60

4. Suppose a list of sales price records is given follows: [2]

$$T = \Big\{5,\ 204,\ 210,\ 8,\ 10,\ 11,\ 55,\ 65,\ 72,\ 92,\ 35,\ 50,\ 108,\ 150,\ 187,\ 13,\ 15,\ 215\Big\}.$$

   Partition them into 4 bins and apply **smoothing using bin-means** by each of the following methods:

   (a) equal-frequency partitioning,

   (b) equal-width partitioning.

5. Find the interquartile-range and create a boxplot representation of the following data: [2]

$$X = \Big\{11, 8, 7, 2, 6, 14, 22, 15, 15, 23, 13, 8, 12\Big\}.$$