Sample Questions 2

1. 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?
    c. What is the midrange of the data?
    d. Can you find the first and third quartiles of the data?
    e. Give the 5-number summary of the data.
    f. Show a box-plot of the data.
2. Describe the steps involved in data mining when viewed as a process of knowledge discovery.
3. List two data mining functions and compare them.
4. What is a data mining task primitive? How are data mining task primitives useful? List data mining task primitives and explain one of them.
5. Define distributive measure. Explain how a distributive measure can facilitate efficient incremental computation. State whether count is a distributive measure or not. Verify your answer.
6. How do you determine outliers?
7. Normalization by decimal scaling normalizes by moving the decimal point of values of attribute A. The number of decimal points moved depends on the maximum absolute value of A. A value, x, of A is normalized to x' by computing $x' = x/10^j$ where j is the smallest integer such that max ($|x'|$) < 1. Suppose that the recorded values of A are -116, 107, 2006, -1901, and 987. Using decimal scaling, find j and normalize the recorded values of A.
8. Explain the meaning of the sentence "We are drowning in data, but starving for knowledge!" and the differences between data, knowledge, and information.