

Essential Statistics of Data Science

TUTORIAL-01

Theoretical Assignment

Ques 1.

Suppose that the values for a given set of data are grouped into intervals. The intervals and corresponding frequencies are as follows.

<i>age</i>	<i>frequency</i>
1-5	200
5-15	450
15-20	300
20-50	1500
50-80	700
80-110	44

Compute an *approximate median* value for the data.

Ques 2. 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.

- What is the mean of the data? What is the median?
- What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.).
- What is the midrange of the data?
- Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data?
- Give the five-number summary of the data.
- Show a boxplot of the data.

Ques 3. Suppose a hospital tested the age and body fat data for 18 randomly selected adults with the following result:

Age	23	23	27	27	39	41	47	49	50	52	54	54	56	57	58	58	60	61
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

- Calculate the mean, median, and standard deviation of age and %fat.