



Ramakrishna Mission Vivekananda Educational & Research Institute

Belur Math, Howrah, West Bengal

School of Mathematical Sciences, Department of Data Science

M.Sc. in Big Data Analytic 2019, Mid Semester Exam

Date: 02 Sep 2019

Course : **DA102: Basic Statistics**

Time: 2 hrs

Instructor : Dr. Sudipta Das

Max marks: 50

Student signature and Id:

(Use of calculator is permitted, however, mobile is strictly prohibited)

1. The following data refer to a certain type of chemical impurity measured in parts per million in 25 drinking- water samples randomly collected from different areas of a county.

11	19	24	30	12	20	25	29	15
21	64	31	16	23	25	26	32	17
22	26	35	18	2	18	27		

Calculate the median, mode, coefficient of variation and inter-quartile range for this sample. Check for any outliers. [1+1+2+2+2=8]

2. The grouped data in the following table represent the number of children from birth through the end of the teenage years in a large apartment complex. Find the mean, median and standard deviation for these data:

Class	0-4	5-9	10-14	15-19
Frequency	5	14	15	10

[1+2+2=5]

3. A discrete variable X can take k values $\{x_1, x_2, \dots, x_k\}$ with corresponding frequencies $\{f_1, f_2, \dots, f_k\}$. Prove that the mean absolute deviation about the mean \bar{x} of the variable X , can be calculated as

$$\frac{2}{\sum_{i=1}^k f_i} \left[\bar{x} \sum_{x_i < \bar{x}} f_i - \sum_{x_i < \bar{x}} x_i f_i \right].$$

[5]

4. In two sets of variables x and y with 50 observations each, the following data were observed:

$$\bar{x} = 10, \sigma_x = 3, \bar{y} = 6, \sigma_y = 2 \text{ and } \text{cor}(x, y) = 0.3.$$

But on subsequent verification it was found that one value of $x (= 10)$ and one value of $y (= 6)$ were inaccurate and hence weeded out. With the remaining 49 pairs of values, how is the original value of $\text{cor}(x, y)$ affected? [8]

5. The measure of skewness of a frequency distribution, in terms of its quartiles (Prof Bowley's Coefficient of Skewness) is defined as

$$Sk_B = \frac{3\text{rd Quartile} - 2 \times 2\text{nd Quartile} + 1\text{st Quartile}}{3\text{rd Quartile} - 1\text{st Quartile}}.$$

Prove that $-1 \leq Sk_B \leq 1$.

6. The production of steel in seven different years are reported as follows

Years (X)	1951	1952	1953	1954	1955	1956	1957
Production in tons (Y)	201	263	314	395	427	504	612

From this record calculate the following

- (a) Draw a scatter plot. [2]
 - (b) Fit a curve of the form $Y = ab^X$ [8]
7. Let the population consist of the numbers $\{1, 2, 3, 4, 5\}$. Consider all possible samples consisting of three numbers randomly chosen without replacement from this population.
- (a) Obtain the distribution of the sample mean. [5]
 - (b) Find the mean and variance of the sample mean, also. [2+3=5]

This exam has total 7 questions, for a total of 50 points and 0 bonus points.
