Ramakrishna Mission Vivekananda University

Belur Math, Howrah, West Bengal

School of Mathematical Sciences, Department of Data Science

M.Sc. in Big Data Analytic 2018, Mid Semester Exam   Date: 10 Sep 2018

Course : **DA102: Basic Statistics**   Time: 2 hrs

Instructor : Dr. Sudipta Das   *Max marks: 40*

Student signature and Id:

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

1. The following data give the time in months from hire to promotion to manager for a random sample of 25 software engineers from all software engineers employed by a large telecommunications firm.

5	7	229	453	12	14	18	14	14
483	22	21	25	23	24	34	37	34
49	64	47	67	69	192	125		

   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, variance, standard deviation for these data:

Class	0-3	4-7	8-11	12-15	16-19
Frequency	7	4	19	12	8

     [1+1+1+1=4]

3. An wedding planner company named *Band Baja Barat* needs to have a wedding planning questionnaire is so that it knows exactly what its client expect on her/his wedding day. Please design a suitable questionnaire for the above-mentioned purpose on behalf of *Band Baja Barat*. Note that, a typical wedding planning questionnaire will include questions in a logical manner that ask the bride and groom about what they plan for their special day. This includes what they might expect out of their guests and service personnel such as catering staff, DJ, etc.   [6]

4. Let X, Y and Z be three uncorrelated variables having same standard deviation. What will be the correlation coefficient between $X + Y$ and $Y + Z$?   [4]

5. For the variables x and y the equations of two regression lines are $4x - 5y + 33 = 0$ and $20x - 9y = 107$. What is the estimated value of y, at x=10?   [4]

P.T.O.

6. Use the method of least squares to fit a straight line $(y = a + bx)$ to the accompanying data points. Give the estimates of a and b. Plot the points and sketch the fitted least-squares line. The observed data values are given in the following table.

x	-1	0	2	-2	5	6	8	11	12	-3
y	-5	-4	2	-7	6	9	13	21	20	-9

[8]

7. Suppose you have n observations:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n),$$

on two variables x and y, and you have fitted a linear regression $y = a + bx$ by the method of least squares. Denote the 'estimated' value of y at x_i by $\hat{y}_i$ and the residual $y - \hat{y}_i$ by e_i. Show that $Cor(\hat{y}, e) = 0$. [6]

This exam has total 7 questions, for a total of 40 points and 0 bonus points.

———————————————————