ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE
I SEMESTER EXAMINATION, OCTOBER 2018
M.SC IN BIG DATA ANALYTICS
BDA 1118: BASIC STATISTICAL METHODS

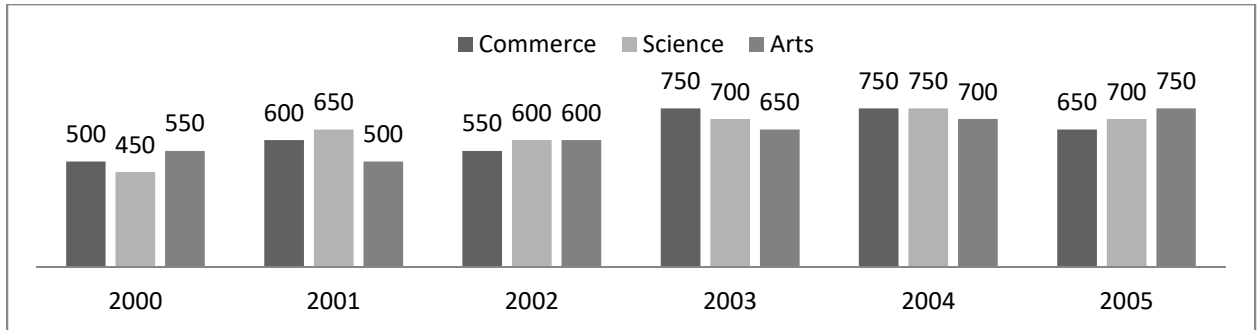**TIME: 2 ½ HRS**   **MAX MARKS 70**

This Question Paper Contains Two Printed Pages and One Part

ANSWER SEVEN QUESTIONS   **7 X10 = 70**

1. Study the following graph carefully to answer the questions that follow: (2 x 5 =10)

Legend: ■ Commerce ■ Science ■ Arts

Year	Commerce	Science	Arts
2000	500	450	550
2001	600	650	500
2002	550	600	600
2003	750	700	650
2004	750	750	700
2005	650	700	750

   a. What is the ratio of the total number of students studying commerce in the year 2000 and 2002 together that of those studying arts in the years 2003 and 2005 together

   b. What is the ratio of total number of students studying Arts, Science, and Commerce in all years together

   c. The number of students studying commerce in the year 2004 forms approximately what per cent of the total number of students studying commerce in all the years together?

   d. What is the total number students studying arts in all the years together?

   e. The number of students studying science in the year 2001 forms what per cent of the total number of students studying all the disciplines together in that year?

2. Differentiate following with examples
   a. Census and Sampling   d) Primary data and Secondary data
   b. Variables and Attributes   e) Interval data and Ratio data
   c. Continuous data and Discrete data   (10)

3. Define qualitative data and quantitative data. Give ten examples of data that you can collect from your day-to-day life. Classify this data as qualitative data or quantitative data   (10)

4.
   a. What is the difference between simple and weighted average? Explain the circumstances under which the latter should be used in preference to the former.
   b. What are the measures of central tendency? Name all measures of central tendency? Explain their importance.   (5+5)

5.
   a. Write briefly about measures of associations.
   b. Life of bulbs produced by two factories A and B are given below:

Length of life	Number of bulbs	
(in hours)	Factory A	Factory B
550-650	10	8
650-750	22	60
750-850	52	24
850-950	20	16
950-1050	16	12

The bulbs of which factory are more consistent from the point of view of length of life? Why? Find the bulb of which factory has more length of life?　　　(4 + 6)

6.

a. Explain briefly about different types of data collection methods
b. What is correlation? What are the different measures? Explain with its importance　　　(4 + 6)

7. The following stem and leaf display contains sample data regarding weight of students:

```
3 |  0  1
4 |  5  8  8
5 |  0  3  4  5  7  8  9
6 |  1  4  7  9
7 |  3  6  9
8 |  0  3  7
```

Key: 3|0 = 30 kgs

Calculate Inter quartile range, 6th Decile, and 78th Percentile. Also obtain the boxplot and give the comments.　　　(10)

8. Mr. James McWhinney, president of Daniel-James Financial Services, believes there is a relationship between the number of client contacts and the dollar amount of sales. To document this assertion, Mr. McWhinney gathered the following sample information. The X column indicates the number of client contacts last month, and the Y column shows the value of sales ($ thousands) last month for each salesperson sampled.

Salesperson	Contacts (X)	Sales (y)
Robert Armstrong	14	24
Jack Bender	12	14
Dorothy Brumley	20	28
Carmen Carella	16	30
Annette Perrault	46	80
Mary Jane Duryee	23	30
David Gwyer	48	90
Harvey Lazik	50	85
Ray Osbeck	55	120
Al Montanaro	50	110

a. Calculate correlation coefficient between contacts and sales and comment
b. Determine the regression equation and interpret
c. Determine the estimated sales if 40 contacts are made. (10)

9.

    a. The following table contains data on the number of complaints received per day at a major retail bank's branches.

No of complaints	0	1	2	3	4	5	6 +
frequency	270	140	92	65	14	5	0

    Test to see whether Poisson distribution is consistent with the data.
    (Critical value = 7.814)

    b. Give one situation in which
        i. The mean is an appropriate measure of central tendency.
        ii. The mean is not an appropriate measure of central tendency but the median is an appropriate measure of central tendency
        iii. The mean is not an appropriate measure of central tendency but the Mode is an appropriate measure of central tendency   (7+3)

10.

    a. A medical trial into the effectiveness of a new medication was carried out. 120 females and 90 males took part in the trial. Out of those people, 70 females responded positively to the medication and 40 males responded negatively to the medication. Construct a contingency table with the given information.
        i. What is the probability that the medicine gives positive results for females?
        ii. What is the probability that the medicine gives negative results for males?
        iii. Was the medication's success independent of gender? Explain
        (Take critical value 0.0039)

    b. In a distribution of 10 observations, the value of mean and standard deviation are given as 20 and 8. By mistake, two values are taken as 2 and 6 instead of 4 and 8. Find out the value of correct mean and variance.   (6+4)