Date:

Registration number:

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU-27
BCA(Data Analytics) - I SEMESTER
SEMESTER EXAMINATION - OCTOBER 2021
(Examination conducted in January-March 2022)

BCADA 1221: EXPLORATORY DATA ANALYSIS USING EXCEL

TIME: 3 HOURS **MAX MARKS: 100**

This Paper contain SIX printed pages and THREE parts

PART A

ANSWER ALL QUESTIONS (MCQs) **20 X 1 = 20**

1. If the mean is 5, median is 6 then mode is:
   (a) 10
   (b) 8.6
   (c) 11
   (d) 8

2. Which of the following relations among the location parameters does not hold?
   (a) Q_2=median
   (b) P_{50}=median
   (c) D_5=median
   (d) D_6=median

3. The standard deviation of a set of 50 observations is 8. If each observation is multiplied by 2, then the new value of standard deviation will be:
   (a) 4
   (b) 8
   (c) 16
   (d) 20

4. What is the probability of getting an odd number in tossing a dice?
   (a)1/6
   (b) 1/3
   (c)1/2
   (d) 1/4

5. All normal distributions are_____

   (a) bell-shaped

   (b) symmetrical

   (c) defined by its parameters μ and standard deviation

   (d) all of the above

6. The standard deviation of the binomial distribution is:
   (a) np

(b) $\sqrt{np}$
(c) npq
(d) $\sqrt{npq}$

7. For a normal distribution if μ=30, then its mode value is:
(a) 1.5
(b) 30
(c) 60
(d) none of these

8. If P(AUB)= P(A), then events A and B are
(a) mutually exclusive
(b) dependent
(c) independent
(d) none of these

9. Which of the following is a relative measure of dispersion?
(a) standard deviation
(b) variance
(c) coefficient of variation
(d) all of the above

10. Which of the following measures of dispersion is least affected by extreme values of observations in a data set?
(a) range
(b) quartile deviation
(c) mean absolute deviation
(d) standard deviation

11. If the relationship between variables x and y is linear, then the points on the scatter diagram
(a) will fall exactly on a straight line
(b) will fall on a curve
(c) must represent population parameters
(d) are best represented by a straight line

12. If the relationship between x and y is positive, as the variable y decreases, variable x
(a) increases
(b) decreases
(c) remains same
(d) changes directly

13. Of the following measurement levels, which is the required level for the valid calculation of the Pearson correlation coefficient
(a) nominal
(b) ordinal
(c) interval
(d) ratio

14. Two regression lines are perpendicular to each other when
(a) r = 0
(b) r = 1/3
(c) r = -1/2
(d) r=±1

15. If the first and third quartiles are 22.16 and 56.36 respectively, then the quartile deviation is
(a) 17.1
(b) 34.2
(c) 51.3
(d) 30.3

16. Two dice are thrown. Find the chance of obtaining a sum of at least 9 points?

(a) 3/18

(b) 3/20

(c) 5/18

(d) 1/20

17. Which of the following is a demerit of Arithmetic Mean?

(a) A.M provides a good basis for comparison

(b) The mean is unduly affected by extreme values

(c) It is easy to determine and understand A.M

(d) It can be used for further algebraic treatment

18. The A.M of the values of a variable x is 25. If each value is increased by 5, what will be the new mean?

(a) 25

(b) 30

(c) 20

(d) 23

19. If the average of 7,9, 12, x,5, 4 and 11 is 9 then x is

(a) 13

(b) 14

(c) 15

(d) 8

20. What is the mode of the following series ____

7 9 11 7 7 5 9 13

(a) 9

(b) 7

(c) 11

(d) 5

ANSWER ANY TEN QUESTIONS **10 x 5 = 50**

21.The following data is the annual income (in $1,000s of U.S. dollars) taken from ten randomly chosen students in the Hopkins Internet based MPH program:

3, 37, 102, 34, 12, 111, 56, 72, 17 ,33

a) Calculate the sample mean income.
b) Calculate the sample median income.
c) Calculate the sample standard deviation of these incomes.

(d) When is it appropriate to use the arithmetic mean as opposed to the median?

22. Given is the number of boys in families with 4 children

x	Prob(X=x)
0	1/16
1	1/4
2	3/8
3	1/4
4	1/16

(a) What is the expected value of X? What does it mean?

(b) What is the standard deviation of X?

23. Define a Poisson distribution. Evaluate the probability of 2 lymphocytes out of 10 white blood cells if the probability of any one cell being a lymphocyte is 0.2.

24. Demonstrate the idea of the level of measurement. Differentiate between categorical and continuous variables.

25. Sketch the shape of a normal distribution, a positively skewed distribution and a negatively skewed distribution.

26.. What are the major limitations of Statistics? Explain with suitable examples.

27. The following table gives the marks of 58 students in Statistics. Calculate the average marks of this group.

No. of Students	**Marks**
0-10	4
10-20	8
20-30	11
30-40	15
40-50	12
50-60	6
60-70	2
Total	**58**

28. What are dependent and independent events? For two events A and B, let P(A)=0.4,

P(A+B) =0. 58 and P(B)= 0.3. Check whether A and B are independent or not.

29. Differentiate between correlation and regression

30. A fair coin is tossed thrice. Find the probability distribution of the number of heads obtained.

31. When can Poisson distribution be a reasonable approximation to the binomial? Give some applications of Poisson distribution

32. A problem in Statistics is given to five students A, B, C, D, E. Their chances of solving it are ½, 1/3, 1/4, 1/5 and 1/6 respectively. What is the probability that the problem will be solved?

PART C

ANSWER ANY THREE QUESTIONS **3 X 10 = 30**

33. Prove that the correlation coefficient is independent of the change of origin and scale.

Compute correlation coefficient for the data given below:

X	1	2	3	4	5	6	7	8	9
Y	9	8	10	12	11	13	14	16	15

34.Explain the concept of regression. Why are there two lines of regression in a bivariate distribution?

The following table gives information on years of education of 10 farmers and annual yield per acre:

Years of education(x)	0	2	4	6	8	10	12
Annual yield per acre(y)	4	4	6	10	12	9	11

Find the regression equation of yield per acre on education.

35.What is conditional probability? Explain with an example.

The probability that a regularly scheduled train leaves the station on time is

P(D) =0.80, the probability that it arrives on time is P(A)=0.82; the probability that it departs and arrives on time is P(D∩A) =0.75. Find the probability that a train (i) arrives on time given that it departed on time and (ii) departed on time given that it has arrived on time.

36.State and prove the Addition Theorem of Probability.

A meteorologist has forecast the probability of rains on Monday, Tuesday and Wednesday as follows:

Monday: 0.60
Tuesday: 0.50

Wednesday: 0.30

Assuming that the weather from day to day is independent, what is the probability that it will rain at least once in these three days?

37.Define hyper geometric distribution. Mention two areas where hypergeometric distribution can be applied.

A taxi cab company has 12 Ambassadors and 8 Fiats. If 5 of these Taxi cabs are in the shop for repairs and Ambassador is as likely to be in for repairs as a Fiat. What is the probability that

(i) 3 of them are Ambassadors and 2 are fiats?
(ii) At least 3 of them are Ambassadors?
(iii) All 5 of them are of the same make?

38. Discuss the relationship between correlation and regression.

The following data are about sales and advertisement expenditures of a firm:

	Sales (in crores of Rs)	Advertisement expenditures (in crore of Rs)
Mean	40	6
S.D	10	1.5

Coefficient of correlation r=0.9

(a)Estimate the likely sales for a proposed advertisement expenditure of Rs 10 crore

(b) What would be the advertisement expenditure if the firm fixes a sales target of 60 crores of rupees?