

BCA404 DATA ANALYSIS AND STATISTICAL TECHNIQUES

Duration: 2 Hours

Total Marks: 50

INSTRUCTIONS:

- I. Figures to the right indicate maximum marks
- II. Start each answer on a fresh page.
- III. Non scientific, non programmable calculator allowed.
- IV. Graph paper will be provided on request

1. Answer the following

[5x2=10 marks]

A. Match the following

- | | |
|------------------------|---|
| i. Arithmetic mean | a. $l + \left(\frac{N}{2} - c\right) h/f$ |
| ii. Meadian | b. $\frac{Q3-Q1}{2}$ |
| iii. Weighted mean | c. $\sqrt{\frac{\sum fx^2}{N} - x^2}$ |
| iv. Quartile deviation | d. largest value-smallest value |
| v. Range | e. $\frac{\sum xw}{\sum w}$ |

B. Fill in the blanks

- i. For an impossible event $n(E) = \underline{\hspace{2cm}}$
- ii. If E is an event of sample space S and E' is the event that E does not happen then $P(E') = \underline{\hspace{2cm}}$
- iii. If A and B are two independent events such that $P(A) \neq 0$ and $P(B) \neq 0$ then $P(A' \cap B') = \underline{\hspace{2cm}}$
- iv. In a Poisson distribution if $P(2) = P(3)$, then the mean of the distribution is $\underline{\hspace{2cm}}$
- v. Two events A and B of the sample space are said to be exhaustive if $\underline{\hspace{2cm}}$

2. Answer any two of the following

[10 marks]

- A. Find all the three quartiles of the for the following data of marks out of 100 students obtained by 80 students

marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
frequency	4	9	14	25	18	6	2	1	1	0

- B. Write down the Apriori Algorithm in Data Mining

- C. There are 5 main political parties. The following table gives the number of votes in lacs received by each one in 2 elections. Find the Karl Pearson's coefficient of correlation

party	I	II	III	IV	V
Past election	5	3	3	4	5
Present election	7	4	2	2	5

3. Answer any two of the following

[10 marks]

- A. A D.T.P operator claims that she can type a regular text at an average speed of 100 words per minute. In 36 trials, her average speed was 95 words per minute with deviation of 10 words. Is her claim justified? Use 1% l.o.s. (at 1% l.o.s $z=2.58$)
- B. In a certain city 10000 electric bulbs were used for street lighting. The average life of bulbs is 1000 hours with a standard deviation of 200 hours. Find the number of bulbs expected to fail
- In first 800 hours
 - Between 800 and 1200 hours

Z	1	2	3
P	0.3413	0.4772	0.4987

- C. Write a short note on

- Stratified sampling
- Simple random sampling

4. Answer any two of the following

[10 marks]

- A. Draw a less than and more than ogive for the data given below representing the distribution of balance amounts in bank accounts on March 2018

Amount	5000-6000	6000-7000	7000-8000	8000-9000	9000-10000
No. of bank accounts	25	42	55	35	11

- B. For the bivariate data, find

- Coefficient of regression
- Coefficient of correlation
- Equations of line of regression

x	3	4	9	10	12	13
y	6	3	1	2	2	1

- C. The average number of customers who appear at the counter of a bank in one minute is 2. Find the probability that in a given minute
- No customer appears
 - Almost 2 customers appear
(take $e^{-2}=0.135$)

5. Answer any two of the following

[10 marks]

A. Calculate the mean deviation from the mean for the following data

Class interval	5-15	15-25	25-35	35-45	45-55	55-65
frequency	2	5	12	15	12	4

B. Two regression equations are given below

- Identify the regression lines and state the regression coefficient
- Find coefficient of correlation
- Mean of x and y

$$6x+5y=50 \text{ and } 10x+3y=62$$

C. A random sample of size 500 has sample proportion $p=0.15$. can we say that it is drawn from a population with proportion $P=0.2$ at 5% l.o.s? (at 5% l.o.s $z=1.96$)
