



# INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

## Mid-Autumn Semester Examination 2022-23

Date of Examination: **23.09.2022**

Session: **FN**

Duration: 02 hours

Full Marks: 50

Subject No. : **CS 61061** Subject: **Data Analytics**

Computer Science & Engineering Department

- 1) Non-programmable calculator may be allowed.
- 2) Statistical tables may be allowed.

Special instructions:

- Answer to all questions.
  - All symbols in the question, if not mentioned explicitly bear their usual meanings.
  - You may make reasonable assumptions, if any.
- 

### Part-A

1. Multiple options are given against each question below. More than one option may be correct. You have to select all the correct option(s). Choosing a wrong option will not fetch any credit. Otherwise, credit will be given on prorata basis.  
No negative marking for selecting any wrong option. [10×1=10]

- i. Which of the following category according to NOIR classification does not allow the calculation of median as a measure of central tendency?

- (a) Nominal.
- (b) Ordinal.
- (c) Interval scale.
- (d) Ratio scale.

- ii. Which of the following can be considered to remove outliers in data?

- (a) Q1 quartile measure.
- (b) Median measure.
- (c) IQR (Inter Quartile Range) measure.
- (d) Box plot.

- iii. Which of the following statement is true according to the Central Limit theorem?

- (a) Population's mean can be inferred from a sample's mean for any population.
- (b) Population's variance can be inferred from the sample's variance for any population.
- (c) Population's mean can be inferred from a sample's mean for a sample chosen at random.
- (d) Population's variance can be inferred from the sample's variance for a population provided that the size of sample is large and the population data is normally distributed.

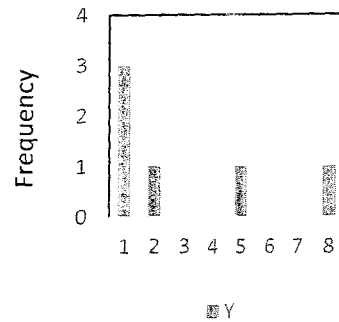
- iv. Which of the following probability distribution function(s) is (are) applicable to discrete random variables?

- (a) Gaussian distribution
- (b) Poisson distribution
- (c) Weibull distribution
- (d) Chi-square distribution

v. Which of the following is NOT true about a Bernoulli process?

- (a) Each trial has two or more outcomes.
- (b) Outcomes are mutually exclusive.
- (c) Trials are independent to each other.
- (d) A trial may be repeated for a number of times.

vi. A distribution of data is shown in the figure given below.



Select the mean and median of the sample from the option given below?

- (a) Mean(Y) = 2.0, Median(Y) = 2.5
- (b) Mean(Y) = 2.5, Median(Y) = 2.0
- (c) Mean(Y) = 3.0, Median(Y) = 1.5
- (d) Mean(Y) = 3.5, Median(Y) = 1.0

vii. A probability mass function  $f(x)$  for a discrete random variable is given below.

$x$	0	1	2
$f(x)$	0.64	0.32	0.04

Select the variance of the data from the option given below?

- (a) 0.64
- (b) 0.32
- (c) 0.20
- (d) 0.04

viii. A sample follows normal distribution with mean  $\mu = 0$  and variance  $\sigma^2 = 1$ . What is  $P(X=2)$ ?

- (a) 0.1468
- (b) 0.1568
- (c) 0.1668
- (d) 0.1768

ix. A data distribution in discrete domain satisfies the Poisson distribution. The mean arrival rate is 16 over a fixed period of time, say  $t = 1$ . Which of the following is true about the mean and standard deviation of the data?

- (a) Mean = 16, Standard deviation = 16
- (b) Mean = 16, Standard deviation = 4
- (c) Mean = 4, Standard deviation = 16
- (d) Mean = 4, Standard deviation = 4

x. In the following Table, Column A lists some sampling distributions, whereas Column B lists the name of sampling distributions. All symbols bear their usual meanings.

Column A		Column B	
(A)	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	(W)	Normal distribution
(B)	$\frac{\bar{X} - \mu}{S/\sqrt{n}}$	(X)	Chi-squared distribution
(C)	$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$	(Y)	t-distribution
(D)	$\frac{(n-1)S^2}{\sigma^2}$	(Z)	F distribution

Some matchings from Column A and Column B are given below. Select the correct matching?

- (a) (A)-(Z), (B)-(X), (C)-(Y), (D)-(Z)
- (b) (A)-(X), (B)-(Z), (C)-(W), (D)-(W)
- (c) (A)-(Y), (B)-(W), (C)-(X), (D)-(Y)
- (d) (A)-(W), (B)-(Y), (C)-(Z), (D)-(X)

### Part-B

2. Passengers drop by a busy store at an average rate of  $\lambda=4$  per minute. If the number of passengers dropping by the store obeys a Poisson distribution, what is the approximate probability that 16 passengers drop by the store in a particular 4-minute period? [5]

3. The following is the distribution of marks obtained by 109 students in a class. Find the Geometric Mean (GM). [10]

Marks	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	36-40
No. of students	6	10	18	30	15	12	10	6	2

4. The following sample was taken from a normally distributed population.  
3, 4, 5, 5, 6, 6, 6, 7, 7, 9, 10, 11, 12, 12, 13, 13, 13, 14, 15

Compute the confidence interval estimation on the population mean. Assume 95% confidence interval. [5]

5. Given a sample drawn randomly which is shown below:

8.08	7.71	7.89	7.72
8.00	7.90	7.77	7.81
8.33	7.67	7.79	7.79
7.94	7.84	8.17	7.87

Test the hypothesis that the population variance is 0.01 at 5% level of confidence.

[10]

6. Let the random variable X represents the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
f(x)	0.51	0.38	0.10	0.01

- (a) Find mean  $\mu$   
 (b) Find variance  $\sigma^2$   
 (c) Find coefficient of variation, CV

[3+4+3]

---\*---