Dhirubhai Ambani Institute of Information and Communication Technology

M.Sc.(Data Science)

Sample Question Paper

- You have up to 90 minutes.
- All questions are compulsory.
- Every item on the test awards 4 points for each correct answer and -1 for an incorrect answer.
- Calculators are allowed.

Part I: Quantitative Aptitude

1. A and B participate in a race of 1 km. A beats B by 28 meters or 7 seconds. How much time does A take to complete the race?
   (a) 4 minutes 3 seconds.
   (b) 4 minutes 10 seconds.
   (c) 3 minutes 57 seconds.
   (d) 4 minutes 17 seconds.

2. How many terms are there in $3, 9, 27, 81, \cdots, 531441$?
   (a) 10
   (b) 11
   (c) 12
   (d) None of the above.

3. What is the missing number in $2, 5, 9, 19, 37, ...$
   (a) 73
   (b) 74
   (c) 75
   (d) None of these.

4. A shopkeeper sells a product at a profit of 20%. If the selling price is $480, what is the cost price of the product?
   (a) $400
   (b) $450
   (c) $500

(d) $600

5. A train travels at a speed of 60 km/hr for the first 30 minutes, then at a speed of 90 km/hr for the next 1 hour and 20 minutes, and finally at a speed of 40 km/hr for the last 45 minutes. What is the total distance covered by the train?

   (a) 90 km

   (b) 120 km

   (c) 180 km

   (d) 210 km

6. If some politicians are corrupt and all corrupt individuals are dishonest, which of the following statements is necessarily true?

   (a) All politicians are dishonest.

   (b) Some dishonest individuals are politicians.

   (c) All corrupt individuals are politicians.

   (d) Some politicians are honest.

7. If A is the brother of B, B is the sister of C, and C is the father of D, then how is A related to D?

   (a) Father

   (b) Brother

   (c) Uncle

   (d) None of the above

8. X and Y are two alloys of copper and iron prepared by mixing metals in the ratio 1:2 and 4:5, respectively. If equal quantities of alloys, e.g., X and Y, are melted to form a third alloy, Z, the ratio of copper to iron in Z is

   (a) 7:11

   (b) 5:7

   (c) 1:3

   (d) None of these

9. A shopkeeper purchased two varieties of rice: 200 KG at Rs. 20/KG and 100 KG at Rs. 50/KG. Being greedy, the shopkeeper mixed the two rice varieties and sold the mixture at a gain of 20%. Find the per KG selling price of the mixed rice.

   (a) 30

   (b) 32

   (c) 36

(d) 40

10. The population of a city X, which is 1,00,000, decreases at 1500/year. The population of city Y, which is 50,000, increases at the rate of 1000/year. After how many years did the population of cities X and Y become equal?

    (a) 20

    (b) 25

    (c) 10

    (d) None of these

11. When dividing a number X by 30, we get 25 as the quotient and 5 as the remainder. On dividing the same number X by 15, what will be the remainder?

    (a) 0

    (b) 5

    (c) 10

    (d) None of these

12. Each worker, e.g., X, Y, and Z, completes a given task in 5, 10, or 20 hours. If these workers work together, how much time is required in hours to complete a given task?

    (a) Less than 2 hours

    (b) 35/12

    (c) 1.5 hours

    (d) 20/7

Part II: Statistics

1. Consider the following data points of a dataset

$$0, 1, 0, 1, 0,$$

the mean, median, and mode of the given dataset are

   (a) 0.4, 0 and 0 respectively.

   (b) 0.4, 0 and 1 respectively.

   (c) 0.4, 1 and 0 respectively.

   (d) 0.5, 0 and 0 respectively

2. For a finite value n, if the standard deviation of the data $1, 2, ..., n$ is K , then the standard deviation of the data $11, 12, ..., 10 + n$ is

   (a) $10K$

   (b) $K + 10$

   (c) K

   (d) $K(K + 1)....(K + 10)$.

3. The mean of the following grouped data is

Marks	No. of Students
10-12	6
12-14	8
14-16	5
16-18	7
18-20	4

   (a) 14.67

   (b) 12.54

   (c) 13.67

   (d) 15.27

4. Consider the following dataset What will be the correlation between X and Y?

X	Y
1	1
2	1
3	1
4	1
5	1

   (a) 1

   (b) 0.5

   (c) 0

   (d) -1

5. Two cards are drawn randomly from an ordinary deck of 52 cards. Find the the probability that both are hearts.

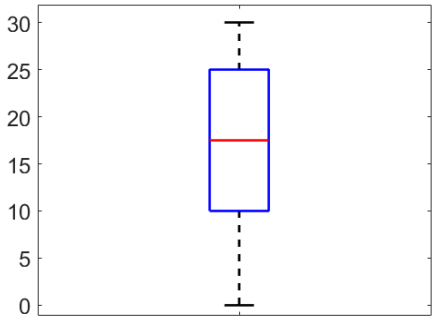
   (a) $\frac{3}{51}$

   (b) $\frac{13}{102}$

   (c) $\frac{2}{13}$

   (d) $\frac{1}{4}$

6. The inter-quartile range for the dataset represented by the following boxplot is –



   (a) 15

   (b) 10

   (c) 25

   (d) 17.5

7. The expected fraction of data points which lie in the inter-quartile range of the boxplot is

   (a) 0.25

   (b) 0.60

   (c) 0.75

   (d) 0.50

8. For a given continuous random variable X, let $F(X)$ be the Cumulative Distribution Function. Consider the following statements :-

   (A) F should be continuous function.

   (B) F should be monotonically non-decreasing function.

   (C) F should necessarily be a logarithmic function.

   Which group of the above statements is always true ?

   (a) A and B

   (b) B only

(c) All

(d) A and C

9. For a given continuous random variable X, let $f(X)$ be the density function. Consider the following statements :-

   (A) $f(x)$ should be always in [0,1].

   (B) $P(a \leq X \leq b)$ can be obtained by evaluating $\int_a^b f(x)$.

   (C) $\int_{-\infty}^{\infty} f(x)$ is 1.

   Which group of the above statements is always true ?

   (a) A and C

   (b) B only

   (c) All

   (d) B and C

10. Let X be a random variable with parameter $n = 100$ and $p = 0.2$, then mean and variance of X are

    (a) 20 and 16

    (b) 16 and 10

    (c) 20 and 24

    (d) None of above

11. Let X be the random variable which models the waiting time until an accident at a manufacturing plant and follows the exponential distribution , then for given a and $b > 0$, which of the following is correct ?

    (a) $P(X > a) = P(X > a + b|X > b)$

    (b) $P(X > a) > P(X > a + b|X > b)$

    (c) $P(X > a) < P(X > a + b|X > b)$

    (d) None of above

12. Let the random variable X $\sim N(\mu, \sigma)$, then which of them are correct for random variable $Y = aX + b$?

    (a) $Y \sim N(a\mu + b, a^2\sigma^2)$

    (b) $Y \sim N(a\mu + b, a^2\sigma + b^2)$

    (c) $Y \sim N(a^2\mu + b^2, a\sigma^2)$

    (d) $Y \sim N(a\mu, a\sigma^2)$

Part III: Mathematics

1. Suppose $f : [1, 2] \to \mathbb{R}$ is a continuous function. Then which of the following can be true?

   (a) $f[1, 2] = (-10, 10000)$.

   (b) $f[1, 2] = [-10, 10000)$.

   (c) $f[1, 2] = (-10, 10000]$.

   (d) $f[1, 2] = [-10, 10000]$.

2. Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined by

$$f(x) = \begin{cases} x^2 \sin(\frac{1}{x}), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0. \end{cases}$$

   Then which of the following is true?

   (a) $f'(0)$ exists and it is equal to 0.

   (b) $f'(0)$ exists and it is equal to 1.

   (c) $f'(0)$ does not exist.

   (d) $f'(0)$ exists and it is equal to 2.

3. Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined by

$$f(x) = \begin{cases} \alpha, & \text{if } x < 0 \\ ax^2 - bx + c, & \text{if } x \geq 0. \end{cases}$$

   Then f is continuous at 0 if

   (a) $\alpha = a$.

   (b) $\alpha = b$.

   (c) $\alpha = c$.

   (d) None of the above.

4. Let $T : M_{n\times n}(\mathbb{R}) \to M_{n\times n}(\mathbb{R})$ be a linear transformation defined by $T(A) = A^t$, where A^t is the transpose of the matrix A. Then which of the following is true?

   (a) 0 is an eigenvalue of T.

   (b) 1 is an eigenvalue of T.

   (c) 2 is an eigenvalue of T.

   (d) 3 is an eigenvalue of T.

5. Which of the following differential equation is exact?

   (a) $(x^2 + y)dx + (y^2 + x)dy = 0$.

   (b) $(x^2 - y)dx + (y^2 + x)dy = 0$.

(c) $(x^2 + y)dx + (y^2 - x)dy = 0$.

(d) $(x + y)^2dx + (x - y)^2dy = 0$.

6. What is the minimum value of the function $f(x, y) = x^2 + 2y^2 - 2x - 4y$?

   (a) 0

   (b) -2

   (c) -1

   (d) -3

7. Let $f(x, y) = \frac{xy^2}{x^2+y^4}$. Then which of the following is true?

   (a) $\lim_{(x,y)\to(0,0)} f(x, y) = 0$.

   (b) $\lim_{(x,y)\to(0,0)} f(x, y) = 1$.

   (c) $\lim_{(x,y)\to(0,0)} f(x, y) = -1$.

   (d) $\lim_{(x,y)\to(0,0)} f(x, y)$ does not exist.

8. Suppose A is a 100×100 non-zero matrix with real entries such that $A^{50} = 0$. Then which of the following is true?

   (a) A is invertible.

   (b) A is diagonalizable.

   (c) $A^{100} = 0$.

   (d) A has a non-zero eigenvalue.

9. Consider the matrix $M = \begin{pmatrix} 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$. Let n be the number of linearly independent eigen vectors of M. Then which of the following is true?

   (a) $n = 1$.

   (b) $n = 2$.

   (c) $n = 3$.

   (d) $n = 4$.

10. Suppose f is a continuous function and satisfies the equation $\int_0^x f(t)dt = x^2 + \sin x$ for all x. Then what is the value of $f(\pi/2)$?

    (a) $\frac{\pi}{2}$

    (b) π

    (c) $\frac{\pi}{4}$

(d) $\frac{\pi^2}{4} + 1$

11. Which of the following is true?

(a) The series $\sum_{n=1}^{\infty} \frac{n}{2^n}$ is divergent and the sequence $\{\frac{1}{\sqrt{n}}\}$ is convergent.

(b) The series $\sum_{n=1}^{\infty} \frac{n}{2^n}$ is convergent and the sequence $\{\frac{1}{\sqrt{n}}\}$ is convergent.

(c) The series $\sum_{n=1}^{\infty} \frac{n}{2^n}$ is divergent and the sequence $\{\frac{1}{\sqrt{n}}\}$ is divergent.

(d) The series $\sum_{n=1}^{\infty} \frac{n}{2^n}$ is convergent and the sequence $\{\frac{1}{\sqrt{n}}\}$ is divergent.

12. Let $P = \{2, 3, 5, 7, 11, 13, \cdots, \}$ be the set of prime numbers. Let R be the radius of convergence of the power series $f(x) = \sum_{p \in P} x^p$. Then which of the following is true?

(a) $R = 0$.

(b) $R = 1$.

(c) $R = 2$.

(d) $R = 3$.

Part IV: Computer Science

1. Which of the following is not a correct example of an Abstract Data Type with respect to the C or C++ programming languages?

   (a) Integer

   (b) Array

   (c) String

   (d) Date

2. Consider the following code snippet for a search algorithm:

```
int CustomeSearch(int a[], int len, int x) {
    int mid, low = 0;
    int high = len-1;
    while (low <= high) {
        mid = (low+high) / 2;
        if (x == a[mid])
            return mid;
        else if (x > a[mid])
            low = mid+1;
        else
            high = mid-1;
    }
    return -1;
}
```

   After k^{th} iteration, how many elements of the given array will remain alive (i.e., still to be searched)?

   (a) $\frac{2^k}{n}$

   (b) $\frac{n}{k}$

   (c) $\frac{n}{2^k}$

   (d) $\frac{k}{n}$

3. Consider two strings S_1 and S_2 each having k_1 and k_2 ($k_1 \leq k_2$) characters respectively. Which of the following data structures should you use (in addition to storing S_1&S_2 themselves) to minimize the worst-case time complexity for counting the unique **un**common characters?

   (a) Auxiliary Queue

   (b) Auxiliary Stack

   (c) Auxiliary 1-D Array

   (d) Auxiliary 2-D Array

4. Suppose an array $A[-15 \ldots 64]$ is stored in a memory. Assume that the word size for each element is 2. How many number of elements are there in the array? Also, if one word of the memory equals 2 bytes, how much memory is required to store the entire array?

(a) 40, 160 bytes

(b) 80, 160 bytes

(c) 40, 80 bytes

(d) 80, 40 bytes

5. An array is filled with marks of k students. An algorithm has to be designed for the m^{th} highest marks ($m < k$). Select the data structure that would lead to the minimum time-complexity:

(a) Auxiliary array

(b) Auxiliary queue

(c) Auxiliary stack

(d) No need of additional data structure

6. Which of the following statements is true regarding virtual inheritance in C++?

(a) It is used to create multiple instances of a base class in a derived class.

(b) It resolves the diamond problem by allowing multiple inheritance without creating ambiguity.

(c) It is achieved by declaring the base class constructor as virtual.

(d) It leads to increased performance due to reduced overhead in managing shared base class instances.

7. Which of the following statements regarding the scope of local variables in C is incorrect?

(a) Local variables are accessible only within the function in which they are declared.

(b) Local variables can be accessed from any function within the same file.

(c) Local variables are not accessible outside the block in which they are declared.

(d) Local variables may have the same name as global variables without causing conflicts.

8. Which of the following sorting algorithms has a worst-case time complexity of $O(n^2)$ but a space complexity of $O(1)$?

(a) Merge Sort

(b) Quick Sort

(c) Heap Sort

(d) Insertion Sort

9. What is the output of the following code snippet?

```
#include <stdio.h>
int main() {
    int x = 5;
    printf("%d", ++x * x++);
    return 0;
}
```

(a) 30

(b) 38

(c) 42

(d) Compiler Error

10. In C++, which of the following is true regarding the initialization of base class members in a derived class constructor?

(a) Base class members are automatically initialized by the derived class constructor.

(b) Base class members are explicitly initialized using the base class constructor in the member initialization list of the derived class constructor.

(c) Base class members are initialized using the derived class constructor body.

(d) Base class members are initialized using the default constructor of the base class.

11. What is the output of the following code snippet?

```
#include <stdio.h>
void foo(int x) {
    printf("%d-", x);
    if (x > 0)
        foo(--x - 1);
}
int main() {
    foo(3);
    return 0;
}
```

(a) 3 2 1 -1

(b) 0 1 2 3

(c) 3 1 0

(d) 3 1 -1

12. Consider a 2D Array filled up with marks and roll numbers of N students. If the marks can range from $0 : m$ then what will be the **best-case** space complexity of storing the marks.

(a) $\Omega(mN)$

(b) $\Omega(m + N)$

(c) $\Omega(m^2)$

(d) $\Omega(N^2)$