JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -3 EXAMINATION- 2023

B.Tech-V Semester (CSE/IT)

COURSE CODE (CREDITS): 20B1WCI531 (2) **MAX. MARKS:** 35

COURSE NAME: Foundation for Data Science and Visualization

COURSE INSTRUCTORS: Dr. Aman Sharma and Dr. Shubham Goel **MAX. TIME:** 2 Hours

Note: (a) *All questions are compulsory.* ***(b)*** *Marks are indicated against each question in square brackets.* ***(c)*** *The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems*

Q1. Write the Output of following codes in R and Python? **[4 Marks, CO-2]**

A)
```
def outer_function(x):
    def inner_function(y):
        return x * y
    return inner_function

result = outer_function(5)
print(result(4))
```

B)
```
def func(a, b, c=1, *args, **kwargs):
    print(a, b, c, args, kwargs)

func(1, 2, 3, 4, 5, key1='value1', key2='value2')
```

C)
```
x <- c(1, 2, 3, 4, 5)
y <- x[x %% 2 == 0]
y[2] <- 6
print(x)
```

D)
```
vec <- c(1, 2, 3, 4, 5)
result <- sapply(vec, function(x) ifelse(x %% 2 == 0, "Even", "Odd"))
print(result)
```

Q2. Explain what is Principal Component Analysis (PCA)? Perform all the steps of PCA with suitable example? **[4 Marks, CO-3]**

Q3. State and prove the normal equations for linear regression in order to obtain slope and intercept based upon the principle of least square error? Use the following data to solve intercept and slope using the above obtained equations **[4.5 Marks, CO-5]**

X	1	3	5	7	1
Y	2	4	6	9	3

Q4. Solve the below mentioned parts for the given matrices? **[5*1.5 Marks, CO-6]**

A. Write a short note on Singular Value decomposition.	Matrix $A = \begin{bmatrix} 4 & 1 & 3 \\ 2 & 0 & 5 \\ 3 & 2 & 6 \end{bmatrix}$,
B. Find the inner product of A and B; also find the outer product of B and C.	Matrix $B = \begin{bmatrix} -1 & 2 & 4 \\ 3 & 5 & 1 \\ 0 & 2 & -3 \end{bmatrix}$, and
C. Evaluate $2\times A+3\times B-C$.	
D. Find Determinant and Rank of matrix *C*.	Matrix $C = \begin{bmatrix} 2 & 0 & 1 \\ 3 & -1 & 2 \\ -2 & 4 & 0 \end{bmatrix}$.
E. Find Inverse of matrix *B*.	

Q5. A) Explain the purpose and significance of using Pandas' library. How does Pandas align with data visualization libraries like Matplotlib and Seaborn? **[2 Marks, CO-4]**

B) Define random variables and differentiate between discrete and continuous types. Explain the roles of probability mass functions (PMFs) and probability density functions (PDFs) in characterizing each type, providing examples to illustrate their applications in statistical analysis. **[3 Marks, CO-5]**

Q6. Is gender independent of education level? A random sample of 395 people was surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey are summarized in the following table: **[5 Marks, CO-5]**

	High School	Bachelors	Masters	PhD
Female	60	54	46	41
Male	40	44	53	57

Critical values of the Chi-square distribution with d degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	d	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

Find out are gender and education level dependent at a 5% level of significance? In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained?

Q7.

(A) The average grade for an exam is 74, and the standard deviation is 7. If 12% of the class is given A grade, and the grades are curved to follow a normal distribution, what is the lowest possible and highest possible value of marks at which A and B can be awarded respectively? Important values $P(Z> 0) = 0.5$, $P(Z<1) = 0.84134$, $P(Z<1.5)=0.93319$, $P(0<Z<1.18)=0.381$, $P(Z<1.4)=0.91924$.

[3 Marks, CO-5]

(B) A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years. Important values $P(Z> 0) = 0.5$, $P(Z<1) = 0.84134$, $P(Z<1.5)=0.93319$, $P(0<Z<1.18)=0.381$, $P(Z<1.4)=0.91924$. **[2 Marks, CO-5]**