JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- 2026

B.Tech-IV Semester (CSE/IT)

COURSE CODE (CREDITS): 25BIWCI431 (2) MAX. MARKS: 15

COURSE NAME: Data Analytics USING R & PYTHON

COURSE INSTRUCTOR: ATA, RVS, AYS MAX. TIME: 1 Hour

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

(c) Use of calculators is allowed

Q.No	Question	CO	Marks
Q1	A hospital dataset is structured such that each record corresponds to one patient, and the stored fields include *Patient ID, Age, Blood Pressure, Severity Level (Low / Medium / High)*, and a sequence of sensor readings *(120, 122, 125, 130)*. A trainee analyst makes the following claims: (a) *"Blood Pressure is an element because its value differs across patients."* State whether the claim is correct and why. (b) *"The sequence 120, 122, 125, 130 directly explains the patient's medical condition."* Classify what these numbers represent and what is still required. (c) *"Since severity levels are ordered, their average severity can be meaningfully calculated."* Accept or reject this claim with reference to measurement scale. (d) *"Age is an independent variable here, but it may not be so in another analysis."* In one sentence, comment on the validity of this statement.	1	**(1+1.5+1.5+1)**
Q2	What will be the output of the following code? If the code is incorrect, write the correct version. a: import pandas as pd data = {'A':[1,2,3], 'B':[4,5,6]} df = pd.DataFrame(data) print(df.mode()) c: x <- c(2,4,6,8) median[x]	2	5

14

	b	import matplotlib.pyplot as plt x = [1,2,3,4] y = [2,4,6,8] plt.plot(x, y) plt.xlabel("X-axis") plt.ylabel("Y-axis") plt.title("Line Plot") plt.show()	d	library(ggplot2) df <- data.frame(x=c(1,2,3), y=c(2,4,6)) ggplot(df, aes(x, y)) + gom_line()
			e	import pandas as pd data = [1,2,3,4] df = pd.DataFrame(data) print(df.mean)

Q3	The average marks for the Data Analytics subject in a university are claimed to be 65. Two professors have different opinions regarding this claim. Professor A believes that the average marks are not equal to 65. Professor B believes that the average marks are greater than 65. To examine these claims, a random sample of 15 students is selected. The sample shows: Mean marks = 68 Standard deviation = 5 Assume that the marks are normally distributed and the population standard deviation is unknown. (a) State the null hypothesis and alternative hypothesis for Professor A's claim. (b) State the null hypothesis and alternative hypothesis for Professor B's claim. (c) At the 90% confidence level, test both claims using an appropriate t-test and state which hypothesis is selected in each case. ALL THE BEST	1	5

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
[redacted]											
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073