Name: Enrolment No:	UPES UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online Examination, December 2020

Course: Predictive Modelling | **Semester: III**
Program: M. Tech. (CSE) | **Time : 03 hours**
Course Code: CSDA7002 | **Max. Marks: 100**

SECTION A

1. Each Question will carry 5 Marks
2. Instruction: Complete the statement / Select the correct answer(s)

Q1	The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study: y= -0.0127 + 0.0180x The above equation implies that: a. each beer consumed increases blood alcohol by 1.27% b. on average it takes 1.8 beers to increase blood alcohol content by 1% c. each beer consumed increases blood alcohol by an average of amount of 1.8% d. each beer consumed increases blood alcohol by exactly 0.018	**CO1**
Q2	Regression analysis was applied to return rates of sparrowhawk colonies. Regression analysis was used to study the relationship between return rate (x: % of birds that return to the colony in a given year) and immigration rate (y: % of new adults that join the colony per year). The following regression equation was obtained. y = 31.9 – 0.34x Based on the above estimated regression equation, if the return rate were to decrease by 10% the rate of immigration to the colony would: a. increase by 34% b. increase by 3.4% c. decrease by 0.34% d. decrease by 3.4%	**CO1**
Q3	A fund has a sample R-squared value close to 0.5 and it is doubtlessly offering higher risk adjusted returns with the sample size of 50 for 5 predictors. Find Adjusted R square value. a. 0.164 b. 0.234 c. 0.18 d. 0.423	**CO2**
Q4	The following temperatures were recorded (in F^o) each day for two weeks. 82, 72, 83, 75, 80, 78, 82, 73, 60, 79, 80, 78, 83, 81 What is the mean for this set of data, if the outlier is removed? a. 75 b. 77.6 c. 78.9 d. 79.5	**CO4**
Q5	For a multiple regression model, total sum of square (TSS) = 200 and Error sum of squares (ESS) = 50. The multiple coefficient of determination is a. 0.25 b. 4.00 c. 0.75 d. none of the above	**CO2**

Q6 Suppose the correlation coefficient between height (as measured in feet) versus weight (as measured in pounds) is 0.80. What is the correlation coefficient of height measured in inches versus weight measured in ounces? [12 inches = one foot; 16 ounces = one pound]

a. 0.80
b. 0.40
c. 0.533
d. cannot be determined from information given

CO2

SECTION B

1. Each question will carry 10 marks
2. Instruction: Write short / brief notes

Q7 The table below shows the number of absences, x, in a Calculus course and the final exam grade, y, for 7 students. Find the correlation coefficient and interpret your result.

X	1	0	2	4	3	3
Y	95	90	90	70	80	85

CO2

Q8 For a multiple regression model with 35 observations and 9 independent variables (10 parameters), SSE = 134 and SSR = 289, state and test the null hypothesis that all of the regression parameters are zero at the 0.05 level. Use the following F table for the required significance level.

Critical Values of the F-Distribution: $\alpha = 0.05$

Denom. d.f.	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236

CO3

Q9 What do you mean by multicollinearity? Discuss the method of Variable Inflation Factors (VIF) for detecting multicollinearity.

CO4

Q10	Regression diagnostics consists of autocorrelation, normality and homoscedasticity conditions on the residuals of the regression. Discuss these terms in brief. Assume the following (X,Y) data points: (1,2), (2,5), (3,6), (3,7), (4,8), (5,11) The equation for the best fit line of this data is: $Y = 2X+1$. Apply the Durbin Watson statistic to find autocorrelation of residuals.	**CO4**
Q11	Discuss the method to compute the coefficient of multiple linear regression with the help of an analytical method termed as Normal Equation method. **OR** A jeweler prices diamonds on the basis of quality (with values from 0 to 8, with 8 being flawless and 0 containing numerous imperfections) and color (with values from 1 to 10, with 10 being pure white and 1 being yellow). Based on the price per carat (in thousands) of the following 10 diamonds weighing between 1.0 and 1.5 carats, determine the relationship between quality, color and price.	**CO1**

COLOUR	QUALITY	PRICE
6	5	700
3	6	400
5	8	540
8	1	650
9	3	600
5	4	400
4	0	250
2	6	300
8	7	700
6	4	560

Section C

1. Each Question carries 20 Marks.
2. Instruction: Write long answer.

Q12 The number of officers on duty in a Delhi and the number of robberies for that day are:

Officers	10	15	16	1	4	6	18	12	14	7
Robberies	5	2	1	9	7	8	1	5	3	6

Calculate the regression line for this data, and the residual for the first observation, (10; 5). What percentage of variation is explained by the regression line?

OR

A study involved comparing the per capita income (in thousands) to the number of medical doctors per 10,000 residents. Six small cities in Uttarakhand had the observations:

Per capita income	8.6	9.3	10.1	8.0	8.3	8.7
Doctors	9.6	18.5	20.9	10.2	11.4	13.1

Calculate the regression line for this data. What percentage of variation is explained by the regression line? Predict the number of doctors per 10,000 residents in a town with a per capita income of 8500.

CO3