Name: **Enrolment No:**	UPES UNIVERSITY WITH A PURPOSE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2019

Course: Applied Statistical Analysis **Semester: 7th**
Program: B.Tech CSE (BFSI) **Time 03 hrs.**
Course Code: CSIB225 **Max. Marks: 100**
Instructions: All questions are compulsory

SECTION A

S. No.		Marks	CO
Q 1	Three athletes A, B and C are participating in the Olympics. A is twice as likely to win as B and B is twice as likely to win as C. What are the probabilities of their winning?	**4**	**CO1**
Q 2	You got a dataset depicting the popularity of two graphic novels given by a critic which contains three variables. 1) Time of survey (in dd-mm-yy format) 2) Rating of 'Marvel' (in range between 0 to 10) 3) Rating of 'DC' (in range between 0 to 10) The data is collected every day since 1970. You need to graphically represent the data in a chart. What will you use? And why?	**4**	**CO1**
Q 3	Suppose we visited a Edu-fare and the following data is collected by me for the institutes in Europe: **University No.** / **Rating** 1 / 9 2 / 8.5 3 / 8.3 4 / 7.9 5 / 7.6 6 / 7 / 0 Discuss the dataset (which type). Also compute the missing value for University No 6. Explain the methodology you used and why. What is the time complexity for the methodology you chose?	**4**	**CO1**
Q 4	Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?	**4**	**CO3**

<table>
<tr><td colspan="4">SECTION B</td></tr>
<tr><td>Q 5</td><td>Today Vivek marked the attendance of students with respect to time which gradually increased as the lecture progressed (positively correlated). He found the following observations:

<table>
<tr><th>Time Stamp</th><th>No. of Students in class</th></tr>
<tr><td>0</td><td>1</td></tr>
<tr><td>1</td><td>3</td></tr>
<tr><td>2</td><td>4</td></tr>
<tr><td>3</td><td>5</td></tr>
<tr><td>4</td><td>6</td></tr>
<tr><td>5</td><td>9</td></tr>
</table>

Vivek happened to a number freak and wants to find an equation for the observed data. He loves the second degree parabolic equation: $Y = a + bX + cX^2$
Help Vivek to formulate the equation.</td><td>10</td><td>CO2</td></tr>
<tr><td>Q 6</td><td>Suppose X and Y are the two variables having the correlation coefficient 0.85. The following are the values they have:

<table>
<tr><th>X</th><th>Y</th></tr>
<tr><td>10</td><td>40</td></tr>
<tr><td>30</td><td>30</td></tr>
<tr><td>50</td><td>70</td></tr>
<tr><td>60</td><td>80</td></tr>
</table>

If two new variables A and B are obtained by adding 50 to each value of X and 100 to each value of Y respectively, calculate the correlation coefficient between A and B using the above data. Also compare the results.</td><td>10</td><td>CO4</td></tr>
<tr><td>Q 7</td><td>The diameter of an electric cable say X, is assumed to be a continuous random variable with p.d.f.

$F(x) = 6x(1-x); 0 <= x <= 1.$

1) Check that f(x) is a p.d.f.
2) Determine $P(X <= 0.5)$ and $P(X > 0.2)$</td><td>10</td><td>CO4</td></tr>
<tr><td>Q 8</td><td>In a statistical study relating to the prices (in Rs.) of two shares, X and Y, the following two regression lines were found as

$8X – 10Y +70 = 0$
$20X – 9Y – 65 = 0$

The standard deviation of X = 3.
Compute
i) The values of mean of X and mean of Y,
ii) R(X,Y)
iii) Standard deviation of Y.</td><td>10</td><td>CO4</td></tr>
</table>

	OR There are eight coins, seven of which have the same weight and the other one weighs more. In order to find the coin having more weight, a person randomly chooses two coins and puts one coin on each side of a common balance. If these two coins are found to have the same weight, the person then randomly chooses two more coins from the rest and follows the same method as before. The probability that the coin will be identified at the second draw is?		
SECTION-C			
Q 9	a) Metro trains are scheduled every 3 minutes at a certain station. A person comes to the station at a random time and waits for the train. If waiting time follows uniform distribution over the interval (0, 3), then find the probability that he has to wait at least 2 minutes for the train. b) Suppose that on an average 1 customer per minute arrives at a shop. What is the probability that the shopkeeper will wait more than 3 minutes before the first two customers arrive? Assume that waiting time follows gamma distribution.	**20**	**CO3**
Q 10	The variance of a certain dimension article produced by a machine is 7.2 over a long period. A random sample of 20 articles gave a variance 8. Is it justifiable to conclude that variability has increased at 5% level of significance assuming that the measurement of dimension article is normally distributed? For reference:	**20**	**CO5**

Chi-Square (χ^2) Distribution

Degrees of Freedom	Area to the Right of Critical Value 0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892

OR

A new flag of Informatics club is to be designed with 5 vertical strips using some or all of the four colours: green, maroon, red and yellow. In how many ways this can be done so that no two adjacent strips have the same colour?