

**UNIVERSITY OF PETROLEUM AND ENERGY STUDIES**

**End Semester Examination, December 2021**

**Course: Applied Statistics**

**Semester : III**

**Program: B.Tech CS- AI&ML**

**Duration : 03 hrs.**

**Course Code: CSBA2010**

**Max. Marks: 100**

**Instructions: All Questions are compulsory (Internal choices are given in questions, wherever applicable)**

**SECTION A**

**(Scan and upload)**

**(5Qx 4M = 20 Marks)**

		<b>Marks</b>	<b>COs</b>																																																		
<b>Q 1</b>	Cite at least two examples from your own experience illustrating the application of statistics	<b>4</b>	<b>CO1</b>																																																		
<b>Q 2</b>	<p>The total scores X obtained by 50 students in a psychology test of 100 marks are given below.</p> <table border="1" data-bbox="193 730 796 889"> <tr><td>75</td><td>89</td><td>66</td><td>52</td><td>90</td><td>68</td><td>83</td><td>94</td><td>77</td><td>60</td></tr> <tr><td>38</td><td>47</td><td>87</td><td>65</td><td>97</td><td>49</td><td>65</td><td>70</td><td>73</td><td>81</td></tr> <tr><td>85</td><td>77</td><td>83</td><td>56</td><td>63</td><td>79</td><td>69</td><td>82</td><td>84</td><td>70</td></tr> <tr><td>62</td><td>75</td><td>29</td><td>88</td><td>74</td><td>37</td><td>81</td><td>76</td><td>74</td><td>63</td></tr> <tr><td>69</td><td>73</td><td>91</td><td>87</td><td>76</td><td>58</td><td>63</td><td>60</td><td>71</td><td>82</td></tr> </table> <p>Answer the following question on the basis of the data given above.</p> <p>a) Is the random variable X = Score of a student, discrete or continuous? What are the minimum and maximum scores?</p> <p>b) Using the classes 20 - 29, 30 - 39, 40 - 49, . . . and 90 - 99 draw up the frequency distribution of X.</p> <p>c) What percentage of the students score above the pass marks of 50?</p> <p>d) How many of the students score between 50 and 79?</p>	75	89	66	52	90	68	83	94	77	60	38	47	87	65	97	49	65	70	73	81	85	77	83	56	63	79	69	82	84	70	62	75	29	88	74	37	81	76	74	63	69	73	91	87	76	58	63	60	71	82	<b>1 x 4= 4</b>	<b>CO1</b>
75	89	66	52	90	68	83	94	77	60																																												
38	47	87	65	97	49	65	70	73	81																																												
85	77	83	56	63	79	69	82	84	70																																												
62	75	29	88	74	37	81	76	74	63																																												
69	73	91	87	76	58	63	60	71	82																																												
<b>Q 3</b>	<p>Mention any four applications of linear regression technique.</p> <p style="text-align: center;"><b>Or</b></p> <p>A coin is tossed 4 times , what is the probability of getting ,</p> <p>a) Exactly 2 heads</p> <p>b) At least 2 heads</p>	<b>4</b>	<b>CO3</b>																																																		
<b>Q 4</b>	<p>Given the following sample of 20 numbers:</p> <p>12 41 48 58 14 43 50 59 15 45</p> <p>52 72 18 45 54 78 41 47 56 79</p> <p>a) Compute the mean, the variance, and the standard deviation.</p> <p>b) If the largest value in the above set of number is changed to 500 to what degree are the mean and variance affected by the change?</p>	<b>3+1=4</b>	<b>CO1</b>																																																		
<b>Q 5</b>	What is hypothesis testing? What are the Type I and Type II errors in hypothesis testing?	<b>4</b>	<b>CO2</b>																																																		

**P.T.O**

**SECTION B**  
(Scan and upload)

(4Qx10M = 40 Marks)

<b>Q 1</b>	Elucidate the role of statistics in research and development and mention the steps of research process	<b>10</b>	<b>CO1</b>																
<b>Q 2</b>	The mean height of 500 students is 151cm and the standard deviation is 15 cm. Assuming that the heights are normally distributed, find how many students have heights between 120 and 155 cm.	<b>10</b>	<b>CO2</b>																
<b>Q 3</b>	A company manufactures phone batteries with an average life of 2 years. An engineer believes this value to be less, using sample of 10 batteries he measures the average life span to be 1.8 years with a standard deviation of 1.5. a) State the null & alternate hypothesis. b) At 99% confidence level is there enough confidence to discard the null hypothesis <b>Or</b> What are the different methods of measuring central tendency of any dataset? Mention each with suitable example.	<b>10</b>	<b>CO2</b>																
<b>Q 4</b>	Calculate the coefficient of correlation between X and Y  <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tr> <td style="padding: 2px;">X</td> <td style="padding: 2px;">23</td> <td style="padding: 2px;">27</td> <td style="padding: 2px;">28</td> <td style="padding: 2px;">28</td> <td style="padding: 2px;">29</td> <td style="padding: 2px;">30</td> <td style="padding: 2px;">31</td> </tr> <tr> <td style="padding: 2px;">Y</td> <td style="padding: 2px;">18</td> <td style="padding: 2px;">20</td> <td style="padding: 2px;">20</td> <td style="padding: 2px;">27</td> <td style="padding: 2px;">21</td> <td style="padding: 2px;">29</td> <td style="padding: 2px;">27</td> </tr> </table>	X	23	27	28	28	29	30	31	Y	18	20	20	27	21	29	27	<b>10</b>	<b>CO3</b>
X	23	27	28	28	29	30	31												
Y	18	20	20	27	21	29	27												

**SECTION-C**  
(Scan and upload)

(2Qx 20M= 40 Marks)

<b>Q 1</b>	Elucidate the following with examples ( <b>Attempt any 4</b> ) a. Decision Trees b. Neural Networks c. Cluster Analysis d. Factor Analysis e. Classification	<b>4Q x 5=20</b>	<b>CO4</b>																																	
<b>Q 2</b>	Sales data of 10 months for a coffee house situated near a prime location of a city comprising the number of customers (in hundreds) and monthly sales (in Thousand Rupees) are given below:  <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th style="padding: 2px;">S. No</th> <th style="padding: 2px;">No. of Customers (in hundreds)</th> <th style="padding: 2px;">Monthly Sales (in thousand Rs.)</th> </tr> </thead> <tbody> <tr><td style="padding: 2px;">1</td><td style="padding: 2px;">6.0</td><td style="padding: 2px;">01</td></tr> <tr><td style="padding: 2px;">2</td><td style="padding: 2px;">6.1</td><td style="padding: 2px;">06</td></tr> <tr><td style="padding: 2px;">3</td><td style="padding: 2px;">6.2</td><td style="padding: 2px;">08</td></tr> <tr><td style="padding: 2px;">4</td><td style="padding: 2px;">6.3</td><td style="padding: 2px;">10</td></tr> <tr><td style="padding: 2px;">5</td><td style="padding: 2px;">6.5</td><td style="padding: 2px;">11</td></tr> <tr><td style="padding: 2px;">6</td><td style="padding: 2px;">7.1</td><td style="padding: 2px;">20</td></tr> <tr><td style="padding: 2px;">7</td><td style="padding: 2px;">7.6</td><td style="padding: 2px;">21</td></tr> <tr><td style="padding: 2px;">8</td><td style="padding: 2px;">7.8</td><td style="padding: 2px;">22</td></tr> <tr><td style="padding: 2px;">9</td><td style="padding: 2px;">8.0</td><td style="padding: 2px;">23</td></tr> <tr><td style="padding: 2px;">10</td><td style="padding: 2px;">8.1</td><td style="padding: 2px;">25</td></tr> </tbody> </table> <p style="margin-top: 10px;">Find the simple linear regression equation that fits the given data.</p> <p style="text-align: center;"><b>Or</b></p>	S. No	No. of Customers (in hundreds)	Monthly Sales (in thousand Rs.)	1	6.0	01	2	6.1	06	3	6.2	08	4	6.3	10	5	6.5	11	6	7.1	20	7	7.6	21	8	7.8	22	9	8.0	23	10	8.1	25	<b>20</b>	<b>CO3</b>
S. No	No. of Customers (in hundreds)	Monthly Sales (in thousand Rs.)																																		
1	6.0	01																																		
2	6.1	06																																		
3	6.2	08																																		
4	6.3	10																																		
5	6.5	11																																		
6	7.1	20																																		
7	7.6	21																																		
8	7.8	22																																		
9	8.0	23																																		
10	8.1	25																																		

The Variance of certain dimension article produces by a machine is 7.2 over a long period. A random sample of 20 articles gave 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: t- table and Chi- Square distribution table are given below respectively.

**Significance level =  $\alpha$**

Degrees of Freedom	.005 (1-tail)	.01 (1-tail)	.025 (1-tail)	.05 (1-tail)	.10 (1-tail)	.25 (1-tail)
	.01 (2-tails)	.02 (2-tails)	.05 (2-tails)	.10 (2-tails)	.20 (2-tails)	.50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	2.776	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711
8	3.355	2.896	2.306	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697
12	3.054	2.681	2.179	1.782	1.356	.696
13	3.012	2.650	2.160	1.771	1.350	.694
14	2.977	2.625	2.145	1.761	1.345	.692
15	2.947	2.602	2.132	1.753	1.341	.691
16	2.921	2.584	2.120	1.746	1.337	.690
17	2.898	2.567	2.110	1.740	1.333	.689
18	2.878	2.552	2.101	1.734	1.330	.688
19	2.861	2.540	2.093	1.729	1.328	.688
20	2.845	2.528	2.086	1.725	1.325	.687
21	2.831	2.518	2.080	1.721	1.323	.686
22	2.819	2.508	2.074	1.717	1.321	.686
23	2.807	2.500	2.069	1.714	1.320	.685
24	2.797	2.492	2.064	1.711	1.318	.685
25	2.788	2.485	2.060	1.708	1.316	.684
26	2.779	2.479	2.056	1.706	1.315	.684
27	2.771	2.473	2.052	1.703	1.314	.684
28	2.763	2.467	2.048	1.701	1.313	.683
29	2.756	2.462	2.045	1.699	1.311	.683
Large	2.575	2.327	1.960	1.645	1.282	.675

Chi-Square ( $\chi^2$ ) Distribution

Area to the Right of Critical Value

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