SRN □□□□□□□□□□□□

PES UNIVERSITY	**PES University, Bengaluru** (Established under Karnataka Act No. 16 of 2013)	**UE20CS902**

OCT 2022: END SEMESTER ASSESSMENT (ESA)
M TECH DATA SCIENCE AND MACHINE LEARNING_ SEMESTER I

UE20CS902 – Statistical Methods for Decision Making

Time: 3 Hrs	Answer All Questions	Max Marks: 80

INSTRUCTIONS

- All questions are compulsory.
- Section A should be handwritten in the answer script provided.
- Section B and C are coding questions which have to be answered in the system.

Section A -20 Marks

<table>
<tr><td>1</td><td>a)</td><td>Consider the following two samples:
Sample 1: 10, 9, 8, 7, 8, 6, 10, 6
Sample 2: 10, 6, 10, 6, 8, 10, 8, 6
Compare the samples in terms of mean, standard deviation</td><td>2</td></tr>
<tr><td></td><td>b)</td><td>A trading company has eight computers that it uses to trade on the New York Stock Exchange (NYSE). The probability of a computer failing in a day is 0.005, and the computers fail independently. Computers are repaired in the evening and each day is an independent trial. What is the probability that 1 computer will fail in a day?</td><td>2</td></tr>
<tr><td></td><td>c)</td><td>Consider the following ANOVA table, fill in the missing values
<table>
<tr><td>Source</td><td>Sum of Squares</td><td>Degree of Freedom</td><td>MS</td><td>F</td></tr>
<tr><td>Treatment</td><td>2000</td><td></td><td></td><td></td></tr>
<tr><td>Error</td><td></td><td>4</td><td></td><td></td></tr>
<tr><td>Total</td><td>2500</td><td>29</td><td></td><td></td></tr>
</table>
</td><td>2</td></tr>
<tr><td></td><td>d)</td><td>Find the Z score for the value of X =30, where mean = 35, standard deviation is 5</td><td>2</td></tr>
<tr><td></td><td>e)</td><td>List the different types of sampling techniques (any 4 four).</td><td>2</td></tr>
<tr><td>2</td><td>a)</td><td>The sample space of a random experiment is {a, b, c, d, e} with probabilities 0.1, 0.1, 0.2, 0.4, and 0.2, respectively. Let A denote the event {a, b, c}, and let B denote the event {c, d, e}. Determine the following : P(A), P(B)</td><td>2</td></tr>
<tr><td></td><td>b)</td><td>Define Central Limit Theorem.</td><td>2</td></tr>
<tr><td></td><td>c)</td><td>Under what conditions would you prefer using one sample t test.</td><td>2</td></tr>
<tr><td></td><td>d)</td><td>Explain briefly about Chi square goodness of fit test.</td><td>2</td></tr>
<tr><td></td><td>e)</td><td>Define Bayes theorem and provide with an example where Bayes theorem can be applied.</td><td>2</td></tr>
<tr><td colspan="4"><b>SECTION B – 30 MARKS</b></td></tr>
<tr><td>3</td><td>a)</td><td>Fifteen adult males between the ages of 35 and 50 participated in a study to evaluate the effect of diet and exercise on blood cholesterol levels. The total cholesterol was measured in each subject initially and then three months after participating in an aerobic exercise program and switching to a low-fat diet. The data is provided in BC.csv file. Do the data support the claim that low-fat diet and aerobic exercise are of value in producing a mean reduction in blood cholesterol levels? Use alpha =. 0.05. Find the P-value</td><td>6</td></tr>
</table>

SRN

i. State the null hypothesis and the alternate hypothesis. (1 mark)
ii. Which test is to be performed. (1 mark)
iii. Compute test statistics, p value. (3 marks)
iv. At the 0.05 significance level, can we conclude whether program has helped reduction in cholesterol levels ? (1 mark)

b) The compressive strength of samples of cement can be modeled by a normal distribution with a mean of 6000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter. 6

i. What is the probability that a sample's strength is less than 6250 Kg/cm2 (2mark)
ii. What is the probability that a sample's strength is between 5800 and 5900 Kg/cm2 (2 mark)
iii. What strength is exceeded by 95% of the samples (2 mark)

c) An experiment in which shape measurement was determined for several different nozzle types. Interest in this experiment focuses primarily on nozzle type and to determine whether there is difference across various nozzle types in terms of shape. 6

Nozzle						
1	0.78	0.8	0.81	0.75	0.77	0.78
2	0.85	0.85	0.92	0.86	0.81	0.83
3	0.93	0.92	0.95	0.89	0.89	0.83
4	1.14	0.97	0.98	0.88	0.86	0.83
5	0.97	0.86	0.78	0.76	0.76	0.75

i. State the hypothesis and the test to be conducted (1 mark)
ii. Does nozzle type affect shape measurement use a p value (3 mark)
iii. Based on the mean value for each Nozzle type which would you say can play a significant difference (2 mark)

d) Two catalysts are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, catalysts 1 is currently in use, but catalyst 2 is acceptable. Since catalyst is cheaper, it should be adopted, providing it does not change the process yield. A test is run in the pilot plant and results in the data shown table. Is there any difference in mean yields for an α =.05 and assume equal variances 6

Number	1	2	3	4	5	6	7	8
Catlayst1	91.50	94.18	92.18	95.39	91.79	89.07	94.72	89.21
Catalyst2	89.19	90.95	90.46	93.21	97.19	97.04	91.07	92.75

i. State the hypothesis and type of test to be used (2 marks)
ii. Test the hypothesis and conclude (4 marks)

e) The amount of water consumed each day by a healthy adult follows a normal distribution with a mean of 1.52 liters. A sample of 40 adults water consumption in liters is taken and it has a mean of 1.76 liters and S.D of 0.18, now test whether any increase in the consumption of water ? 6

i. State the null hypothesis and the alternate hypothesis. (1 mark)
ii. Which test is to be performed. (1 mark)
iii. Compute test statistics, p value. (3 marks)

		iv. At the 0.05 significance level, can we conclude that water consumption has increased? (1 mark)	
SECTION C – 30 MARKS			
4	a)	Consider the purchases.csv file and answer the following questions i. Provide a summary statistics of all the variables, with respect to sales which variables have high correlation (more than .55), moderate (.4-.55) and low correlation. (3 marks) ii. Provide a histogram for the variable sales and profit, based on histogram and calculation of skewness and kurtosis what would you describe about sales and profit. (4 marks) iii. Plot the histogram for variable sales, profit based on customer segments based on these plotswhich segment provides more sales, profit and less sales, profit (4 marks). iv. Draw a scatter plot for all of the variables, what is your observation and conclusion for the relationship between sales and the other numeric variables (4 marks)	15
	b)	Consider the purchases.csv file and answer the following questions i. Draw a boxplot for sales based on customer segments, what do you observe in terms of outliers, and sales difference for various segments. (3 marks) ii. Conduct a hypothesis to see whether there is a difference in terms of sales for customer segment of home business and corporate. What do you conclude. (4 marks) iii. Conduct an hypothesis to prove whether the sales from small business segment is more than corporate, do you agree or not. (4 marks) iv. Conduct an ANOVA to check whether there is sales difference for products of the following type Paper, Telephones and Communications, Binder and Binder Accessories and Computer Peripherals. (4 marks)	15