


SRN

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

	<p align="center">PES University, Bengaluru (Established under Karnataka Act No. 16 of 2013)</p>	<p align="center">UE20CS902</p>
February 2025: END SEMESTER ASSESSMENT (ESA) M TECH DATA SCIENCE AND ARTIFICIAL INTELLIGENCE_SEMESTER I UE20CS902 – Statistical Methods for Decision Making		
Time: 3 Hrs	<ul style="list-style-type: none"> • All questions are compulsory. • Section A should be handwritten in the answer script provided • Section B and C are coding questions that must be answered in the system. 	Max Marks: 100

Section A : 20 Marks

1	a)	Calculate the mean and standard deviation for the following dataset: 10, 15, 20, 25, 30	2
	b)	Explain the difference between ordinal and interval data scales.	2
	c)	What does ANOVA stand for? State its purpose . Why is it preferred to multiple paired t-tests ?	2
	d)	A group of test scores has a mean of 75 and a standard deviation of 8. If 7 points are added to every score in the group, what will happen to the mean and standard deviation?	2
	e)	Consider two dice are rolled simultaneously. Find the probability that the sum of the two numbers is at least 9.	2
2	a)	In hypothesis testing, if a researcher incorrectly rejects the null hypothesis when it is actually true, what type of error has been committed?	2
	b)	What is the minimum sample size recommended to apply the Central Limit Theorem effectively?	2
	c)	What are the measures of dispersion? Explain them in brief.	2
	d)	A bag contains 5 red marbles, 3 blue marbles, and 2 green marbles. If one marble is drawn at random, what is the probability that it is not blue?	2
	e)	Define Bayes theorem and provide with an example where Bayes theorem can be applied.	2

SECTION B – 40 MARKS

3	a)	A delivery company claims that the average delivery time for their packages is 30 minutes. A sample of 15 deliveries resulted in the following times (in minutes): 28, 32, 29, 35, 31, 30, 27, 33, 29, 30, 34, 28, 31, 29, 32. (i) Perform a hypothesis test at the 0.05 significance level to determine if the mean delivery time differs from 30 minutes. (5 marks) (ii) Compute a 95% confidence interval for the mean delivery time. (3 marks)	8
	b)	A fitness coach wants to test whether a new workout regimen significantly decreases body fat percentage. The body fat percentages of 10 participants before and after the regimen were recorded:	8

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

		<p>Before Workout: [22.5, 24.1, 23.8, 25.2, 24.5, 26.0, 25.5, 23.9, 24.8, 25.1] After Workout: [21.2, 22.8, 22.5, 23.7, 23.1, 24.5, 24.0, 22.4, 23.3, 23.8]</p> <p>(i) Perform a paired t-test at a 0.05 significance level to determine if the workout regimen significantly decreases body fat percentage. (5 marks)</p> <p>(ii) Calculate the mean difference and construct a 95% confidence interval for the mean difference. (3 marks)</p>	
3	c	<p>A scientist wants to test whether the growth rate of a new strain of bacteria differs from a standard rate of 5.2 cm/day. A sample of 12 bacteria colonies was recorded with the following growth rates (in cm/day):</p> <p>5.5, 5.8, 5.3, 5.1, 5.6, 5.4, 5.7, 5.9, 5.2, 5.4, 5.5, 5.6.</p> <p>(i) Conduct a one-sample t-test to determine if the growth rate is significantly different from 5.2 cm/day. (5 marks)</p> <p>(ii) Compute a 95% confidence interval for the mean growth rate. (3 marks)</p>	8
3	d	<p>A researcher wants to determine if there is a significant difference in the average daily water intake (in liters) between two groups of adults: those who exercise regularly and those who do not.</p> <p>Regular Exercisers: [3.1, 2.9, 3.5, 3.2, 3.0, 3.3, 2.8, 3.4, 3.1, 3.2] Non-Exercisers: [2.2, 2.5, 2.3, 2.1, 2.4, 2.0, 2.3, 2.5, 2.2, 2.1]</p> <p>(i) Perform an independent two-sample t-test to determine if regular exercisers consume significantly more water daily than non-exercisers. (5 marks)</p> <p>(ii) Compute a 95% confidence interval for the difference in means. (2 marks)</p>	8
3	e	<p>A study investigates whether there is a relationship between the number of hours of exercise per week and cholesterol levels. The following data shows the number of hours exercised per week and the corresponding cholesterol levels for 12 individuals:</p> <p>Hours Exercised per Week: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12] Cholesterol Levels: [240, 235, 230, 225, 220, 210, 205, 200, 195, 190, 185, 180]</p> <p>(i) Perform a linear regression analysis to determine the relationship between hours of exercise and cholesterol levels. (4 marks)</p> <p>(ii) Predict the cholesterol level for an individual who exercises 6 hours per week. (2 marks)</p> <p>(iii) Test the significance of the regression model using a 0.05 significance level. (2 marks)</p>	8

SECTION C - 40 MARKS

4	<p>a) A producer wants to investigate whether</p> <p>i) There is a significant difference in the preferences for different ice cream flavors among people of different age groups. A survey is conducted, and the following data is obtained: Among participants aged 18-30: Chocolate (45), Vanilla (35), Strawberry (20), Mint (25) Among participants aged 31-50: Chocolate (30), Vanilla (40), Strawberry (25), Mint (25) Among participants aged 51 and above: Chocolate (20), Vanilla (25), Strawberry (30), Mint (25). (10 marks)</p> <p>ii) A call center receives an average of 10 calls per hour. Calculate the probability that the</p>	20
---	--	----

SRN

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

	call center receives exactly 8 calls in a given hour. (5 marks) iii) A researcher wants to conduct a survey to estimate the proportion of people in a population who support a particular political candidate. The researcher wants to achieve a margin of error of 3% with a 95% confidence level. If the researcher expects the proportion of people supporting the candidate to be around 50%, what sample size should be used for the survey? (5 marks)	
4	b). Use the given dataset (studentsperformance.csv) i) Is there any significance in the students final course grades (total score) based on ethnicity? (10 marks) ii) Investigate and find the type of institute that made any impact on the final scores or not? (10 marks)	20