RV College of Engineering®

Academic Year 2023-2024 (Odd Semester)

USN | 1 | R | V | 2 | 2 | A | I | 0 | 0 | 7 |

Department of Artificial Intelligence and Machine Learning

Course Code: AI235AT
Marks: 50Marks

Date : 22-02-2024
Duration: 90 Mins

CIE 2

Statistics for Data Science

SL. No	Questions	M	BT	CO
1	Consider the following normal curves that the following parameters: $\mu = 1.5, \sigma = 3; \mu = 1.5, \sigma = 6.2; \mu = -2.7, \sigma = 3;$ Which curve has the largest spread? (a) Which curves are centered at the same place? (b) Which curves have the same spread? (c) Which curve is the standard normal curve?	4	3	1
2	For your Design Thinking Laboratory (DTL) course, most of you ended up sending survey forms to your friends and acquaintances. One interesting topic that many of you took up was the survey on Data Structures and Algorithms. Which among the sampling techniques listed below would have helped you better to gain insights for this specific survey and why? *Random Sampling or Non-Probability Sampling* ***Your reason must be logical and not more than 2 sentences***	5	3	1
3	There is a big basket, C, of 100 fruits, the contents of which need to be put into two smaller baskets, say A and B. Known for my laziness, I keep the baskets A and B side by side and flip the contents to C into A and B, ensuring that no fruit falls outside of the two baskets and gets damaged. Each fruit falls independently into basket A with probability 0.5 and into basket B with probability 0.5. Let X be the number of fruits that fall into A and Y be the number of fruits that fall into B. ***(I am not keeping fruits one by one in the baskets. Therefore, the number of fruits that fall into each basket is therefore a random variable!)*** Being a student of the course on statistics for data science, your job is to compute $\text{Var}(X+Y)$.	5	2	3
4	A normal random variable X has an unknown mean μ and standard deviation $\sigma = 2$. If the probability that X exceeds 7.5 is .8023, find μ.	5	3	1
5	Assert or Reject the following statement with proper justification of ***NOT MORE THAN 2 SENTENCES.*** **Statement:** A normal distribution with very small variance tends to have a larger peak and a normal distribution with large variance has a smaller peak.	5	3	2
6	The following table corresponds to the joint probability mass function of two random variables, X and Y. **6 Marks**	6	2	1

	y = 1	y = 2
x = 1	2p	p
x = 2	3p	4p

Department of Artificial Intelligence and Machine Learning

	(a) Obtain the value of p that makes the above table a valid PMF. (b) Find the Cov(X,Y).			
7	You are hiring for a statistician position in your company and two candidates Zark Muckerberg and Melon Rusk, with exactly the same credentials appear for the interview. You are in a fix and hence decide to do a tie-breaker by making the following statement. (*It is surprising that you still remember the concepts you learnt in the Stats for Data Science course!!!*) ***Let X and Y be two random variables such that E[XY] = E[X] E[Y].*** Zarc and Melon, in their eagerness to get the job, make the following statements even before you ask any question. ***Zark Muckerberg: X and Y are independent.*** ***Melon Rusk: X and Y are uncorrelated.*** Based on the above two statements, who will you give the offer to and why? ***Ensure that your answer to WHY must not exceed 2 sentences!***	5	2	3
8	***Gill Bates,*** a well known business tycoon and a Stats Expert, visits RVCE to hire students from specifically AMIL Dept as he got to know that you have done an exclusive course on Stats. To check if you guys have learnt Stats concepts correctly from *Prof. A*, he writes the following matrix on the board and $$C_{XY} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$ and calls it a covariance matrix. Your task is to correctly identify if this is a valid covariance matrix or not, with proper reasoning of ***not more than 2 sentences***. *Your answer is gonna expose **both** Prof A and yourself! Phew!!!*	5	2	1
9	The joint probability density function of (X, Y) is given by $$f_{X,Y}(x,y) = \begin{cases} k(x+y) & 0 < x < 2, 0 < y < 2 \\ 0 & \text{otherwise.} \end{cases}$$ (a) Find the value of the constant k. (b) Find the marginal PDFs of X and Y. (c) Are X and Y independent?	10	3	2

Course Outcomes: After completing the course, the students will be able to	
CO1	Apply the knowledge of statistics in providing solutions to some common real-life and business problems.
CO2	Visualize data better, make logical inferences about the data in real-world scientific/business use cases, and present the analysis results.
CO3	Make inferences about a population from samples through various statistical techniques.
CO4	Use statistical tools to illustrate the principles of data distribution, data sampling, and data visualization.
CO5	Appraise the knowledge of statistics in data science to build a successful career as an AI&ML engineer, work in teams, and communicate their ideas effectively.

M-Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes

	Particulars	CO1	CO2	CO3	CO4	L1	L2	L3	L4	L5	L6
Marks Distribution	Max Marks CIE	24	16	10	-	-	22	28		-	-