BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

Work Integrated Learning Programmes Division

Cluster Programme - M. Tech in AI & ML

II Semester , 2023 – 24

Comprehensive Examination (**Regular**)

Course No : AIMLC ZC418
Course Title : Introduction to Statistical Methods
Nature of Exam. : Open Book (Off Line)
Weightage : 40 Marks
Duration : 150 minutes
Date : 28th September ,2024_10 AM

Number of questions:4
Number of Pages: 2

Q. No	Question	Marks
Q.1.a.	A survey shows 50% of all American workers are having workplace retirement plan, 60% have health insurance, and 49% have both benefits. We select a worker at random. (a) What is the probability he has health insurance, if he has a retirement plan? **(b)** What is the probability he has retirement plan, if he has health insurance? **(c)** What is the probability he is not having health insurance given that he has a retirement plan?	**3M**
Q.1.b.	If the chance of running a bus service according to schedule is 0.75, calculate the probability on a day schedule with 10 services: (i) exactly one is late and (ii) at least one is late.	**2M**
Q.1.c	Let the joint probability density function for (X,Y) be $f(x,y) = \frac{x+y}{3}$, $0 < x < 2, 0 < y < 1$; 0, $otherwise$ i) Find the probability P(X > Y) ii) Find the marginal probability density function of X. iii) Find the marginal probability density function of Y. **iv)** Are X and Y independent?	**5M**
Q.2.a.	An agency conducted a survey to understand the trend related to student's preferences in taking admissions **Program** / **Number of Students** B.E(CSE) 320 B.E(AIML) 240 B.E(DSE) 350 B.E(Cyber Security) 200 B.E(IoT) 90 **Total** **1200** Is it reasonable to conclude that there is no preference among the five B E programs offered as mentioned above? (Level of significance = 0.05)	**5M**
Q.2.b.	A company manufactured a product on two machines i.e. Machine A and Machine B. Consider the following information related to sampling of these	**5M**

to test and compare them. Based on this information, can we conclude that the product manufactured by Machine A is superior to Machine B.

Machine	Sample size	Mean	Standard Deviation
Machine A	15	5.5	0.5
Machine B	15	5.1	0.2

Q.3.a. Consider the following time series data **5M**

Month	January	February	March	April	May
Sales(in Lakhs)	50	52	54	55	57

Assuming the forecast for the month of January as 50, fit exponential smoothing models to the above data with a smoothing parameter α = 0.10 and α = 0.70 .Choose the best model between these two and forecast the sales for the month of June.

Q.3.b. Fit 3 – year moving average and 5 – year moving average models to the following data. Use a suitable error function and suggest the optimal model. **5M**

Year	Sales(in crores)
2014	12
2015	16
2016	20
2017	15
2018	16
2019	17
2020	21
2021	18
2022	19

Q.4.a Consider the following data. **7M**

X	10	12	15	13	11	19
Y	20	22	25	18	17	22

i) Comment on the relation between X and Y using co-variance
ii) Comment on the relation between X and Y using coefficient of correlation
iii) Write your inference based on the above.

Q.4.b A project related to prediction is given to two teams A, B and C. **3M**

i) Team A proposed multiple linear regression model(with SSE as loss function) with accuracy of 80%
ii) Team B proposed polynomial regression model(with SSE) with accuracy of 90%
iii) Team C proposed multiple regression model using gradient descent approach with accuracy of 85%

Write your observations on each model if you are asked to select the best model to be used for prediction among A ,B and C