

## JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

## TEST-3 EXAMINATION JUNE-2016

## B.Tech (BI) IV Semester

COURSE CODE: 10B12MA421

MAX. MARKS: 35

COURSE NAME: BIOSTATISTICS

COURSE CREDITS: 4

MAX. TIME: 2 HRS

*Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means.*

1. We develop a model for estimating heating oil used for a single family home in the month of January based on average temperature and amount of insulation in inches. For the given data set

Oil (Gal)	275. 3	363. 8	164. 3	40. 8	94. 3	230. 9	366. 7	300. 6	237. 8	121. 4	31. 4	203. 5	441. 1	32. 3	52. 5
Temp	40	27	40	73	64	34	9	8	23	63	65	41	21	38	58
Insulation	3	3	10	6	6	6	6	10	10	3	10	6	3	3	10

Obtained the model as  $\hat{Y} = 562 - 5.4x_1 - 20x_2$  where:  $x_1$  = temperature  $^{\circ}F$  and  $x_2$  = attic insulation [inches]

- (i) What could a home owner expect heating oil consumption (in gallons) to be if the outside temperature is  $15^{\circ}F$  when the attic insulation is 10 inches thick?
- (ii) Find total sum square TSS and regression sum square RSS from the model obtained. Find Error sum square ESS and  $R^2$  for the regression model. (1+3=4)
2. The following is the arrangement of defective, d, and non defective, n, pieces produced in the given order by a certain medicine:

n n n n n d d d d n n n n n n n n n d d n n d d d d

Apply Run test for the randomness at 0.05 level of significance. Given  $P[|Z| \leq 1.96] = 0.95$

(3)

3. A clinical study is designed to assess differences in albumin levels in adults following diets with different amounts of protein. Low protein diets are often prescribed for patients with kidney failure. Albumin is the most abundant protein in blood, and its concentration in the serum is measured in grams per deciliter (g/dL). Clinically, serum albumin concentrations are also used to assess whether patients get sufficient protein in their diets. Three diets are compared, ranging from 5% to 15% protein, and the 15% protein diet represents a typical American diet. The albumin levels of participants following each diet are shown below.

5% Protein	3.1	2.6	2.9		
10% Protein	3.8	4.1	2.9	3.4	4.2
15% Protein	4	5.5	5	4.8	

For reference, normal albumin levels are generally between 3.4 and 5.4 g/dL. By inspection, it appears that participants following the 15% protein diet have higher albumin levels than those following the 5% protein diet. The issue is whether this observed difference is statistically significant. Apply normal approximation of relevant nonparametric test and check if there a significant difference in serum albumin levels among subjects on the three different diets at 5% level of significance. Also infer at 1% level of significance.  $\chi^2_{\alpha=5\%,2} = 5.99$ ,  $\chi^2_{\alpha=1\%,2} = 9.21$  (5)

4. Consider a Phase II clinical trial designed to investigate the effectiveness of a new drug to reduce symptoms of asthma in children. A total of  $n = 10$  participants are randomized to receive either the new drug or a placebo. Participants are asked to record the number of episodes of shortness of breath over a 1 week period following receipt of the assigned treatment. The data are shown below.

Placebo	7	5	6	4	12
New Drug	3	6	4	2	1

By inspection, it appears that participants receiving the placebo have more episodes of shortness of breath, but is this statistically significant? Apply normal approximation of relevant nonparametric test and check if there a significant difference in the number of episodes of shortness of breath over a 1 week period in participants receiving the new drug as compared to those receiving the placebo at 5% level of significance. Given  $P[|Z| \leq 1.96] = 0.95$  (5)

5. Consider a random walk between points  $a$  and  $b$  assume that current position is  $h$ ;  $a < h < b$  with probability moving towards right is  $p$ .

- (i) Consider  $h = 0$ ,  $b = L$  and  $a = -2$ , where  $L$  is positive. Write down the probability that the walk eventually reaches  $-1$  rather than  $L$ . Find the limiting value of this probability as  $L \rightarrow \infty$  for  $p = 0.6$
- (ii) For  $p = \frac{2}{3}$ , find the values for absorption probabilities  $w_h$ ,  $u_h$  and mean number of steps  $m_h$  taken until the walk stops for  $a = -1$ ,  $b = 1$ ,  $h = 0$ . (4)

6. Calculate 2-step transition probability  $p_{CR}^{(2)} = P(X_3 = R | X_1 = C)$  for a Markov chain with state space  $S = (R \ C)$  and

$$\text{transition probability matrix } P = \begin{matrix} & \begin{matrix} R & C \end{matrix} \\ \begin{matrix} R \\ C \end{matrix} & \begin{pmatrix} 0.4 & 0.6 \\ 0.25 & 0.75 \end{pmatrix} \end{matrix}$$

- (i) Find  $P(X_3 = R | X_1 = C, X_2 = R)$
- (ii) Find  $P(X_5 = R, X_3 = C, X_1 = C, X_2 = R)$ , where given is  $P(X_1 = C) = 0.5$  (5)

7. The table below is a distance matrix for 5 objects.

	A	B	C	D	F
A	0				
B	0.12	0			
C	0.51	0.25	0		
D	0.28	0.77	0.70	0	
E	0.34	0.61	0.93	0.67	0

Perform hierarchical clustering with complete link and draw the final result by dendrogram. (4)

8. We have 4 medicines as our training data points object and each medicine has 2 attributes. Each attribute represents coordinate of the object.

Object	Attribute 1 (X): weight index	Attribute 2 (Y): pH
Medicine A	1	1
Medicine B	2	1
Medicine C	4	3
Medicine D	5	4

Perform K-Means Clustering and determine which medicines belong to cluster 1 and which medicines belong to the other cluster. Use Euclidean distance to calculate the distance between cluster centroid to each object. (5)

\*\*\*The End\*\*\*