TE Sem V (C scheme) R-2019 " AI & DS " Nov-Dec 2022

2/12/2022

[Time: 3 Hours]

[Marks:80]

N.B. 1. Question No. 1 is compulsory.
2. Attempt any three questions out of remaining five.
3. All questions carry equal marks
4. Assume Suitable data, if required and state it clearly.

Attempt any four:

Q.1
a) Find the standard deviation of the average temperatures recorded over a five-day period last winter: 19, 21, 18, 24, 12?   20

b) X is a normally distributed variable with mean $\mu = 30$ and standard deviation $\sigma = 4$. Find:
   i) P (x < 40),   ii) P (30 < x < 35)?

c) Discuss Boot strapping vs. re-sampling

d) The school principal wants to test if it is true what teachers say – that high school juniors use the computer an average 3.2 hours a day. What are our null and alternative hypotheses?

e) What do you mean by correlation and regression? Explain with example

Q.2 a) Find the value of the correlation coefficient from the data given in the following table: 10

SUBJECT	AGE (X)	GLUCOSE LEVEL(Y)
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

b) Explain briefly why ANOVA is used? Solve using One-way ANOVA   10

OBSERVATIONS	A	B	C
1	25	31	24
2	30	39	30
3	36	38	28
4	38	42	25
5	31	35	28

method:

F701C2E2C5BE928BFEF5A43EA75CF0F8

Q.3 a) Explain type 1 & type 2 error in detail.
(ii) What is the use of scatter plot and box plot?

b) In a manufacturing unit, four teams of operators were randomly selected and 10 sent to four different facilities for machining techniques training. After the 10 training, the supervisor conducted the exam and recorded the test scores. At 95% confidence level does the scores are same in all four facilities?
(Hint: Use Kruskal–Wallis test)

Facility 1	Facility 2	Facility 3	Facility 4
88	77	71	52
82	76	56	65
86	84	64	68
87	59	51	81

Q.4 a) If the sample mean and expected mean value of the marks obtained by 15 10 students in a class test is 290 and 300 respectively. What is the t-score if the standard deviation of the marks is 50?

b) Find out what is the relation between the GPA of a class of students and the 10 number of hours of study and the height of the student

GPA	Height	Study Hours
2.9	66	7
3.16	57	7
3.62	64.5	6
2	62	7
3.45	69.5	8
2.8	65	9
3.63	63	6
2.81	68	5
3.33	59.5	4
2.75	64	10
3.86	69	7

Q.5 a) A farmer is trying out a planting technique that he hopes will increase the 10 yield on his pea plants. The average number of pods on one of his pea plants is 145 pods with a standard deviation of 100 pods. This year, after trying his new planting technique, he takes a random sample of his plants and finds the average number of pods to be 147. He wonders whether this is a statistically significant increase. What are his hypotheses and the test statistic? Use a 0.05 significance level.

b) Find the simple linear regression equation that fits the given data and 10 coefficient of determination:

Hour	Temp
2	21
4	27
6	29
8	86
10	86
12	92

14523

F701C2E2C5BE928BFEF5A43EA75CF0F8

Q.6 a) An agent sells life insurance policies to five equally aged, healthy people. 10
According to recent data, the probability of a person living in these
conditions for 30 years or more is 2/3. Calculate the probability that after 30
years if
  i. All five people are still living.
  ii. At least three people are still living.
  iii. Exactly two people are still living. (Hint: Binomial Distribution)

b) Write short notes on (any two) 10
  i. Confidence Interval
  ii. Central Limit Theorem
  iii. Standard Error

_____