Total No. of Questions : 8]

SEAT No. :

P701

[Total No. of Pages : 4

[5869] - 368

S.E. (Artificial Intelligence and Data Science)

STATISTICS

(2019 Pattern) (Semester - IV) (217528) (Theory)

Time : 2½ Hours] *[Max. Marks : 70*

Instructions to the candidates :

1) *Answer Q.1 or Q.2, Q.3 or Q.4, Q.5 or Q.6 and Q.7 or Q.8.*
2) *Neat diagrams must be drawn wherever necessary.*
3) *Assume suitable data, if necessary.*
4) *Figures to the right indicate full marks.*

Q1) a) Calculate the mean and standard deviation for the following table giving the age distribution of 542 members. **[9]**

Age (in years)	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
No. of members	3	61	132	153	140	51	2

b) In a partially destroyed laboratory, record of an analysis of correlation data, the following results only are legible : Variance of X = 9. Regression equations : 8X–10Y+66=0, 40X–18Y=214. What are : **[9]**

i) the mean values X and Y,

ii) the correlation coefficient between X and Y, and

iii) the standard deviation of Y?

OR

Q2) a) For 10 randomly selected observations the following data were recorded

Observation no :	1	2	3	4	5	6	7	8	9	10
Observation hrs. (X)	1	1	2	2	3	3	4	5	6	7
Additional units (Y)	2	7	7	10	8	12	10	14	11	14

Determine the coefficient of regression and regression equation using the non-linear form $Y = a + b_1 X + b_2 X^2$. **[9]**

P.T.O.

b) Variables X and Y have the joint p.d.f. given by : **[9]**

$$F(x, y) = \frac{1}{3}(x + y); 0 \le x \le 1, 0 \le y \le 2$$

Find

i) r(X, Y),

ii) The two lines of regression, and

iii) The two regression curves for the means

Q3) a) Assume that on an average number out of 15 called between 2pm to 3pm on week days is busy. What is the probability that 6 randomly selected telephone numbers called **[6]**

i) Not more than 3 busy

ii) At least 3 busy

b) If the probability that an individual suffers a bad reaction from certain injection is 0.001. Determine the probability out of 2000 people, by using Poisson's distribution **[5]**

i) Exactly 3

ii) More than 1 will suffer a bad reaction

c) In a Sample of 1000 cases the means of a certain test is 14 and standard deviation is 2.5. Assuming the distribution to be normal find **[6]**

i) How many students scored between 12 & 15.

ii) How many scored below 8.

[Given : A(z = 0.8) = 0.2881, A(z = 0.4) = 0.1554), A(z = 2.4) = 0.4918]

OR

Q4) a) A Random variable X with following probability distribution **[5]**

X	1	2	3	4	5
P (X)	k	$2k$	$2k$	k	$7k^2$

Find

i) k

ii) P ($x \ge 2$)

iii) P ($x < 3$)

iv) P ($2 \le x \le 3$)

v) P ($x \ge 3$)

b) In a continuous distribution density function **[6]**

$f(x) = kx(2 - x)$, $0 < x < 2$.

Find the value of

i) k

ii) Mean

iii) Variance

c) For a normal distribution when mean = 1, standard deviation = 4, find the probabilities of the following intervals : **[6]**

i) $3.43 \le x \le 6.19$ ii) $-1.43 \le x \le 6.19$

[Given : $A(z = 0.81) = 0.2910$, $A(z = 1.73) = 0.4582$]

Q5) a) The following table gives the number of accidents that took place in an industry during various days of the week. Test if accidents are uniformly distributed over the week. **[6]**

Days	Mon	Tue	Web	Thur	Fri	Sat
No. of accidents	14	18	12	11	15	14

Given chi-square$_{0.05,5}$ = 11.09.

b) A normal population has mean 6.8 and standard deviation 1.5. A sample of 400 members gave a mean of 6.75. Is the difference significant? **[6]**

$Z\alpha = 1.96$ at 5% level of significance.

c) Suppose that sweets are sold in packages of fixed weight of contents. The procedure of the packages is interested in testing the average weight of content in packages in 1 kg. Sum of squares of deviations from mean of 12 samples is 0.011967. Using above data should we conclude the average. Given $\overline{X} = 0.9883$, $t_{0.05,11} = 2.201$. **[6]**

OR

Q6) a) A set of five similar coins is tossed 210 times and the result is given in the following table.

No. of heads	0	1	2	3	4	5
Frequency	2	5	20	60	100	31

Use chi-square test to test the hypothesis that data follows a binomial distribution (chi-square = 11.07 at 5% level of significance) **[6]**

b) From the given data below, Intelligence tests of two groups of boys and girls gave the following results. Examine the difference in significance. Given $Z\lambda = 1.96$ at 5% level of significance. **[6]**

	Mean	Standard deviation	Size
Girls	70	10	70
Boys	75	11	110

c) In two independent samples of size 8 and 10, the sum of squares of deviations of sample values from the respective sample means were 84.4 and 102.6. Test whether the difference of variances of the population is significant or not. Given $F_{0.05} = 3.29$ at d.f. (7, 9). **[6]**

Q7) a) If $x \geq 1$ is the critical region for testing $\Theta_0 : \theta = 2$ against the alternative $\theta = 1$ on the basis of the single observation from the population. $f(x, \theta) = \theta\, e^{-\theta x}, 0 \leq x < \infty$, obtain the values of type I, type II error also find power of function. **[8]**

b) State & Prove Neyman-Pearson lemma for testing a simple hypothesis against a simple alternative hypothesis. **[9]**

OR

Q8) a) Write short note on : **[8]**

i) Population and sample

ii) Type I and Type II error

iii) Critical region

iv) Power of test

b) Let $X_1, X_2 \ldots X_n$ be random sample of size n from a normal distribution $N(\mu, \sigma^2)$ where μ and σ^2 both are unknown. Show that LRT used to test $H_0 : \mu = \mu_0$, vs $H_1 : \mu \neq \mu_0$, $0 < \sigma^2 < \infty$ is used t-test. **[9]**

OR

Explain in detail the test for the mean of normal population.

▽▽▽▽