**4 QUESTIONS \* 5 MARKS=20 MARKS**

Q1. The amount of water consumed each day by a healthy adult follows a normal distribution with a mean of 1.52 litres. A sample of 10 adults water consumption in litres is taken and it has a mean of 1.76 litres and S.D of 0.18. Test whether any increase in the consumption of water?

(a) State the null hypothesis and the alternate hypothesis. (1 mark)

(b) Which test is to be performed. (1 mark)

(c) Compute the value of the test statistic. (2 marks)

(d) At the 0.05 significance level, can we conclude that water consumption has increased? (1 mark)

Q2. Now a days the online streaming services like Netflix, Amazon prime, is playing dominant role in the entertainment industry. A public opinion poll surveyed on whether these online streaming services are better than cinema halls? on a sample of 100 random people (classified as teen/Adults/Senior Citizen) with a closed ended question (Yes or No). The details are captured in ‘Online\_straming.xlsx’. dataset Is there any association between age category and voting preference? (Use level of significance as 5%)

A. Frame Null and Alternate Hypothesis ( 1 Mark )

B. State which test to be conducted. ( 1 Mark )

C. Find the test statistic (2 Mark )

D. Interpret the results (1 Mark)

Q3) By the survey conducted by YRBSS the average number of days that students lift the weights in a week is 3 days. In a sample of 100 students from the YRBSS survey of 2013, the average number of days per week that students lifted weights was 2.78 days with a standard deviation of 1.56 days. Assume that the number of days follows normal distribution, test whether there is a difference in the average number of days of overall survey of YRBSS survey and 2013 YRBSS survey?

a) Frame Null and Alternate Hypothesis (2 mark)

b) Compute a 95% confidence interval for number of days that student lift the weights for the 2013 YRBSS survey. (2 mark)

c) Interpret the result (1 mark)

Q4. State wide, the average score on the verbal portion of the college entrance exam is 453, with a standard deviation of 95. A random sample of 137 students at Little Wood Regional High School shows a mean score of 502. Is there a significant difference between the average scores of population and sample?

(a) State the null hypothesis and the alternate hypothesis. (1 mark)

(b) Which test is to be performed. (1 mark)

(c) Compute the value of the test statistic. (2 marks)

(d) interpret the result? (1 mark)

**2 QUESTIONS \* 10 MARKS=20 MARKS**

Q5.(a) In a survey, 200 people were asked to identify their major source of news information; 110

stated that their major source was television news.

i. Construct a 95% confidence interval for the proportion of people in the population

who consider television their major source of news information. (2 marks)

ii. How large a sample would be necessary to estimate the population proportion with a

margin of error of .05 at 95% confidence? (2 marks)

Q5. (b) Phone calls arrive at the rate of 48 per hour at the reservation desk for Regional Airways.

a. Compute the probability of receiving three calls in a 5-minute interval of time. (2marks)

b. Compute the probability of receiving exactly 10 calls in 15 minutes. (2 marks)

c. Compute the probability of receiving not more than 15 calls but not less than 10 calls. (2 marks)

6(a). The below data shows the number of people travelling from Hyderabad to Mumbai and Mumbai to Delhi. Is there evidence to conclude that the number of people travelling from Hyderabad to Mumbai is different from the number of people travelling from Mumbai to Delhi in a week.

Population 1: Hyderabad to Mumbai

n1 = 1200

x1 = 452

s1 = 12

Population 2: Mumbai to Delhi

n2 = 800

x2 = 523

s2 = 18

1.State null and Alternate Hypothesis. (1 mark)

2.find the test statistic and critical value. (3 mark)

3.Interpret the result. (1 mark)

6(b). A survey on amount of time given by people (above 14 years) for sleep was conducted and it is observed that people spend 7.5hrs for sleeping in a day and the corresponding standard deviation is 1 hour.

\* Assume that the time spent in sleeping by people follows a normal distribution.

\* a) What proportion of the people are spending more than 9 hours in sleeping daily? (1.5 mark)

\* b) What proportion of customers are spending less than 6 hours in sleeping daily? (1.5 mark)

\* c) What proportion of customers are spending between 6.5 hours and 8 hours in sleeping daily? (2 mark)

**2 QUESTIONS \* 15 MARKS=20 MARKS**

Q7. Venn’s pizza is famous chain restaurant near university campuses. The management of the chain believes that the monthly sales depends on student population on the campus. The following observations were drawn from 10 randomly chosen campuses.

Restaurant 1 2 3 4 5 6 7 8 9 10

Population in thousands 2 6 8 8 12 16 20 20 22 26

Sales in Rs. lakhs 58 105 88 118 117 137 157 169 149 202

1.Create a data frame and plot the correlation among variables Visualize the linear relationship between Variables (5 marks)

2.Build a regression model to predict sales with population as independent variable & find the intercept and coefficients and interpret the model summary. (5 Marks)

Q8). The data set consists of complete educational details of students right from their schooling to MBA and previous work experience. Salary is the target variable in the data. Perform the following tasks on the data:

i. List out the features (numerical and categorical) and specify what type of tests can be performed for these features and why? (3 Marks)

ii. Frame Null and Alternate Hypothesis for the tests mentioned in question.no.1(3 marks)

iii. Now perform required any four statistical tests on the data to determine what are the significant features in the data with respect to target variable (95% C.I). (6 Marks)

iv. Interpret the above test results. (3 Marks)