**PRACTICE QUESTIONS FOR BOARD EXAM**

1. A computer manager needs to know how efficiency of her new computer program depends on the size of incoming data and how many tables are used to arrange each data set. Efficiency will be measured the number of processed requests per hour. Applying the program to data set of different sizes and number of tables are used, she gets the following results.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Processed requests, Y** | 16 | 26 | 17 | 41 | 50 | 55 | 40 |
| **Data size, (gigabytes), X1** | 15 | 10 | 10 | 8 | 7 | 7 | 6 |
| **Number of tables, X2** | 1 | 2 | 10 | 10 | 20 | 20 | 4 |

The regression equation obtained is Y = 52.7 – 2.87 X1 + 0.85 X2.

Total sum of square = 1452

Sum of square due to regression = 1143.3.

a) Interpret the values of regression coefficients b1 and b2.

b) Test the significance of the regression model at 0.05 level of significance.

c) Is there significant relationship between processed request and number of tables at 0.05 level of significance? Given standard error of b2=0.55.

d) What percentage of variation of processed requests is explained by data size and number of tables?

e) Compute standard error of estimate.

f) Estimate the number of processed requests if data size is 9 gigabytes and number of tables used are 8.

2. In an experiment to determine which of three different missile systems is preferable, the propellant burning rate is measured. The data after coding are given in the table. Use Kruskal-Wallis test (significance level of 0.01) to test the hypothesis that the propellant burning rates are same for three missile systems.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Missile system I** | 22.3 | 16.7 | 22.7 | 19.3 | 18.5 |  |  |
| **Missile system II** | 23.4 | 19.5 | 17.5 | 20.8 | 16.0 | 19.9 |  |
| **Missile system III** | 18.4 | 19.5 | 17.8 | 18.0 | 19.6 | 22.8 | 17.1 |

3. A random sample of students is asked their opinion on proposed core curriculum change. The results are as follows:

|  |  |  |
| --- | --- | --- |
| **Class** | **Options** | |
| **Favoring** | **Opposing** |
| **Freshman** | 125 | 80 |
| **Sophomore** | 60 | 140 |
| **Junior** | 50 | 60 |
| **Senior** | 40 | 55 |

Test the hypothesis that opinion on the change is independent of class standing. Use 0.01 significance level.

4. Define Central limit theorem. The life of a certain brand of an electric bulb may be considered a random variable with mean 1350 hours and standard deviation 55o hours. Using central limit theorem, find the probability that the average life time of 100 bulbs exceeds 1440 hours.

7

5. Define multiple correlation. In a tri-variate distribution X1, X2, and X3, the simple correlation coefficients are given as r12= 0.5, r23=0.6 and r13=0.7. Find

i. partial correlation coefficient between X1 an X2 keeping X3 constant.

ii. multiple correlation coefficient assuming X1 as dependent variable.

8

6. What do you understand by Design of Experiment? Prepare one-way analysis of variance table and carry out the test for the significance of difference in the average yields between different varieties of seed. Given:

Total sum of squares = 258

Sum of square between varieties of seed = 50

Total number of observations = 20

7. Define type I and type II error in testing of hypothesis. It is claimed that Samsung and Huawei mobiles are equally popular in Kathmandu. A random sample of 600 people from Kathmandu showed 350 have Samsung mobile. Test the claim at 5% level of significance.

10

8. Customers of certain Internet service provider connect to the internet at the average rate of 10 new connections per minute. Connections are modeled by binomial counting process.

a. What frame length gives the probability 0.1 of an arrival during given frame?

b. Find the mean and variance for the number of seconds between two consecutive connections.11

9. Write short notes on any two:

a. Difference between parametric and non-parametric test.

b. Required assumptions for linear regression model.

c. Stochastic process.

9. What are the required conditions for error variable in multiple regression analysis? The Internal Revenue Service (IRS) is trying to estimate the monthly amount of unpaid taxes discovered by its auditing division. The IRS estimated this figure on the basis of field auditing labor hours and number of hours of its computers used. The table given below presents these data for the last ten months.

|  |  |  |  |
| --- | --- | --- | --- |
| **Month** | **(X1) Field Audit Labor Hours in 100** | **(X2) Computer Hours in 100** | **(Y) Actual Unpaid Taxes Discovered millions of dollars** |
| Jan | 45 | 16 | 29 |
| Feb | 42 | 14 | 24 |
| Mar | 44 | 15 | 27 |
| April | 45 | 13 | 25 |
| May | 43 | 13 | 26 |
| June | 46 | 14 | 28 |
| Jul | 44 | 16 | 30 |
| Aug | 45 | 16 | 28 |
| Sep | 44 | 15 | 28 |
| Oct | 43 | 15 | 27 |

Given ∑YX1=12005, ∑YX2=4013, ∑X1X2=6485, ∑Y2=7428, ∑X12=19461, ∑X22=2173

i. Develop the estimating equation best describing these data.

ii. Interpret the value of regression coefficients.

iii. Estimate the actual unpaid tax for field if audit labor hours is 4200 and computer hours is 1600 hours.

2

10. Physicians depend the laboratory test results when managing the medical problems such as diabetes or epilepsy. In a uniformity test glucose tolerance, three different laboratories were sent nt=5 identical blood samples from a person who had drunk 50 mg. of glucose dissolved in water. The laboratory results are listed below**:(used design of experiment)**

|  |  |  |
| --- | --- | --- |
| **Lab 1** | **Lab 2** | **Lab 3** |
| 12.1 | 9.3 | 10.0 |
| 11.7 | 11.1 | 10.5 |
| 10.9 | 10.7 | 10.1 |
| 10.2 | 10.9 | 11.0 |
| 10.6 | 9.0 | 10.4 |

Do data indicate a difference in the average readings for three laboratories? Use α=.05

3

11. A psychologist wishes to verify that a certain drug increases the reaction time to given stimulus. The following reaction times (in tenth of seconds) were recorded before and after injection of the drug for each of four subjects:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Subject** | | 1 | 2 | 3 | 4 |
| **Reaction Time** | **Before** | 7 | 2 | 12 | 12 |
| **After** | 13 | 3 | 18 | 13 |

Test at 5% level of significance to determine whether the drug significantly increases the reaction time.

12. The following ANOVA summary table was obtained from a multiple regression model with two independent variables:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Source of variation** | **Sum of square** | **Degree of freedom** | **Mean sum of square** | **F-value** |
| **Regression** | 12.62 | 2 | ? | ? |
| **Error** | 0.78 | 12 | ? |
| **Total** | 13.40 | 14 |  |  |

i. Determine the mean sum square due to regression, the mean sum square due to error and F-value.

ii. Test the significance of overall model at 5% level of significance.

iii. Compute coefficient of determination and interpret its value.

iv. Find standard error of estimate.

6

13. Bank of Nepal recorded the sex of first 30 customers who appeared last Monday with notation M M F M M F M F F M M M F F M F F M F F M F F F M F M M M F F. At the 0.005 level of significance, test the randomness of this sequence.

7

14. Social media users use a variety of devices to access social networking; mobile phones are increasingly popular. However, is there a difference in the various age groups in the proportions of social media users who use their mobile phone to access social networking? A study showed the following results for the different age groups.

|  |  |  |  |
| --- | --- | --- | --- |
| **Use mobile phones to access social networking?** | **Age** | | |
| **18-34** | **35-64** | **65+** |
| **Yes** | 60 | 37 | 14 |
| **No** | 40 | 63 | 86 |

At the 0.05 level of significance, is there evidence of a difference among the age groups with respect to use of mobile phones for accessing social networking?

9

15. An effort to estimate the mean amount per customer for dinner at a major Atlanta restaurant, data were collected for a sample of 49 customers and sample mean is found at $24.80. Assume population standard deviation is $5.

a. Compute standard error of mean.

b. Find 95% confidence interval estimate for the population mean.

16. Define Markov chain and its characteristics.11

17. What are the basic concepts of queuing theory? In a super market, the average arrivals rate of customer is 10 per every 30 minutes following Poisson process. The average time taken by the cashier to list and calculate the customers purchase is 2.5 minutes following exponential distribution. What is the probability that queue length exceeds 6? What is the expected time spent by customer in the system?

12

18. Write short notes on the following.

i. Partial and multiple correlation coefficient.

ii. Properties of good estimator.

19. A survey was conducted among 70 students studying B.Sc. CSIT in some colleges randomly. Among them, 50 students secured more than 80% marks in statistics. Compute 99% and 95% confidence intervals for the population proportion of students who secured more than 80% marks in subject statistics, and comment on the results.

20. Apply the Mann-Whitney U test for examining the following knowledge score on IT among two groups of IT workers at a 5% level of significance.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Group A:** | 5 | 8 | 2 | 7 | 6 |
| **Group B:** | 9 | 12 | 4 | 6 |  |

21. A survey was conduct to see the association between hacking status of the email and the type of email account. The survey has reported the following cross tabulation.

|  |  |  |
| --- | --- | --- |
| **Type of e-mail account** | **Hacking status** | |
| **Yes** | **No** |
| **Yahoo** | 60 | 15 |
| **Gmail** | 20 | 120 |

Does the information provide sufficient evidence to conclude that the type email account and the hacking status is associated? Use Chi-square test at 1% level of significance.

22. State the mathematical model for Statistical analysis for m x m LSD for one observation per experimental unit. Also prepare a dummy ANOVA table for this.

23. Every day is generally considered as either sunny or rainy. A sunny day is followed by another sunny day with probability 0.8 whereas a rainy day is followed by a sunny day with probability 0.4. Suppose it rains on Monday. Make forecasts for Tuesday and Wednesday.