***B.Sc. CSIT Third Semester Statistics-II Practical Problems***

1. The time (in minutes) spent by 10 randomly selected customers using internet in a cybercafé are as follows: 35, 20, 30, 45, 60, 40, 65, 40, 25, 50. Can you say average time spent by customers is more than 30 minutes at 5% level of significance? Also find the 90% confidence interval for population mean and show the linkage between testing of hypothesis and confidence interval. **(t-test for single mean)**
2. A quality controller wishes to determine whether there is a difference in outcome between two different tools of software I and II. The following data shows the outcome of two different tools. Can the controller conclude that a difference exists? **(t test for two independent samples)**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Software I | 24.0 | 16.7 | 22.8 | 19.8 | 18.9 | 22.5 | 31.0 | 18.6 | 17.2 |
| Software II | 23.2 | 19.8 | 18.1 | 17.6 | 20.2 | 19.8 | 17.4 | 19.2 |

1. Marks of 8 students before and after tuition is given below:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Before Tuition | 50 | 54 | 52 | 53 | 48 | 51 | 53 | 54 |
| After Tuition | 54 | 57 | 54 | 55 | 52 | 56 | 56 | 55 |

Can you conclude that tuition has benefited the students? **(Paired t test)**

1. A sample of size 150 gave mean 47.5 and standard deviation 12. Can you regarded from the population of mean 50 at 5% level of significance? Also find 99% confidence limits for the population mean. **(Z test single mean)**
2. Test whether two sample means are significantly different if they are selected from population with standard deviation 840 and 920 respectively with following score values**. (z test double mean)**



1. In 30 toss of a coin the following sequence of heads(H) and Tails(T) is obtained.

H T T H T H H H T H H T T H T H T H H T H T T H T H H T H T

Test at 0.05 level of significance whether the sequence is random. **(Run test)**

1. Modern email servers and anti-spam ﬁlters attempt to identify spam emails and direct them to a junk folder. There are various ways to detect spam, and research still continues. In this regard, an information security oﬃcer tries to conﬁrm that the chance for an email to be spam depends on whether it contains images or not. The following data were collected on n = 1000 random email messages.

|  |  |  |  |
| --- | --- | --- | --- |
| Obs. | With images | No images | Total |
| Spam | 160 | 240 | 400 |
| No spam | 140 | 460 | 600 |
| Total | 300 | 700 | 1000 |

Test it at 5% level of significance by using **chi-square test.**

1. Following are the scores obtained by trainees in 3 different categories. Test whether 3 categories have performed equally by using **Kruskal Wallis H test.**



1. The C programming papers were marked by two teachers A and B. The final marks were recorded as follows:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Teacher A | 73 | 89 | 82 | 43 | 80 | 73 | 66 | 45 | 93 | 36 | 77 | 60 |
| Teacher B | 88 | 78 | 91 | 48 | 85 | 74 | 77 | 31 | 78 | 62 | 76 | 77 |

**Using median test** at 5% level of significance to determine if the marks distributions of two teachers differ significantly.

1. Test the hypothesis of no difference between the ages of male and female employees of a certain company, using the **Mann-Whitney U test** for the sample data below. Use α = 0.1

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Male | 35 | 43 | 26 | 44 | 40 | 42 | 33 | 38 | 25 | 26 |
| Female | 30 | 41 | 34 | 31 | 36 | 32 | 25 | 47 | 28 | 24 |

1. From the following data:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X1 | 11 | 10 | 6 | 10 | 8 | 9 | 10 | 7 | 11 | 8 |
| X2 | 60 | 67 | 53 | 56 | 64 | 57 | 71 | 58 | 67 | 57 |
| X3 | 57 | 55 | 49 | 52 | 57 | 48 | 59 | 50 | 62 | 51 |

Determine all partial correlation coefficients, multiple correlation coefficients.

1. A researcher wanted to examine how the lifetime of a personal computer which is used by children is affected by the time (in hours) spends by the children per day to play games and the available random-access memory (RAM) measured in megabytes (MB) of a used computer. The data is provided in the following table.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Lifetime(years) | 5 | 1 | 7 | 2 | 3 | 4 | 6 |
| Play time(hours/day) | 2 | 8 | 1 | 5 | 6 | 3 | 2 |
| RAM (in GB) | 8 | 2 | 6 | 3 | 2 | 4 | 7 |

1. Fit the multiple regression model.
2. Predict the lifetime of a computer which have play time 4 hours per day and RAM is 5 GB.
3. How much variation in life time of computer is explained by Play time and RAM?
4. Compute standard error of the estimate.
5. Test the significance of regression coefficients and overall fit of the regression equation.
6. Conduct the residual analysis.
7. The yield of treatments in different plots are as shown in the following plots. Carry out analysis. (CRD)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t4 1401 | t3 2536 | t3 2459 | t1 2537 | t3 2827 | t1 2069 |
| t2 2211 | t1 1797 | t4  1170 | t4 1516 | t4 2104 | t3 2385 |
| t2 3366 | t1 2104 | t2 2591 | t3 2460 | t4 1077 | t2  2544 |

1. The following table gives the result of the experiment on four varieties of a crop in 5 blocks of plot.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Block I | Block II | Block III | Block IV | Block V |
| A 32 | B 33 | D 30 | A 35 | C 36 |
| B 34 | C 34 | C 35 | C 32 | D 29 |
| C 31 | A 34 | B 36 | B 37 | A 37 |
| D 29 | D 26 | A 33 | D 28 | B 35 |

Analyze the above result to test whether there is significant difference between yields of four varieties. (RBD)

1. The following is the table is the 5×5 Latin square design for data taken from a manurial experiment with sugarcane. The five treatments were applied

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Row | Column | | | | |
| I | II | III | IV | V |
| I | A 52.5 | E 46.3 | D 44.1 | C 48.1 | B 40.9 |
| II | D 44.2 | B 42.9 | A 51.3 | E 49.3 | C 32.6 |
| III | B 49.1 | A 47.3 | C 38.1 | D 41.0 | E 47.2 |
| IV | C 43.2 | D 42.5 | E 67.2 | B 55.1 | A 45.3 |
| V | E 47.0 | C 43.2 | B 46.7 | A 46.0 | D 43.2 |

Analyze the above data to find if there are any treatments effects. (LSD)

1. The number of disease infected tomato plants in 10 different plots of equal size are given below.Test whether the disease infected plants are uniformly distributed over the entire area use kolmogorov smirnov test.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Plot no. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| No. of infected plants | 8 | 10 | 9 | 12 | 15 | 7 | 5 | 12 | 13 | 9 |