

Tribhuvan University  
Institute of Science and Technology  
Model Question

Bachelor Level / Third Semester / Science

**Computer Science and Information Technology(STA210)**

((TU CSIT) Statistics II)

Candidates are required to give their answers in their own words as far as practicable.

The figures in the margin indicate full marks.

Full marks: 60

Pass marks: 24

Time: 3 hours

---

**Group A**

Attempt any Two questions. (2 x 10 = 20)

1. Suppose a population of 4 computers with their lifetimes 3, 5, 7 & 9 years. Comment on the population distribution. Assuming that you sample with replacement, select all possible samples of  $n = 2$ , and construct sampling distribution of mean and compare the population distribution and sampling distribution of mean. Compare population mean versus mean of all sample means, and population variance versus variance of sample means and comment on them with the support of theoretical consideration if any.
2. A computer manager is keenly interested to know how efficiency of her new computer program depends on the size of incoming data and data structure. Efficiency will be measured by the number of processed requests per hour. Data structure may be measured on how many tables were used to arrange each data set. All the information was put together as follows.

Data size(gigabytes)	6	7	7	8	10	10	15
Number of tables	4	20	20	10	10	2	1
Processed requests	40	55	50	41	17	26	16

Identify which one is dependent variable? Fit the appropriate multiple regression model and provide problem specific interpretations of the fitted regression coefficients.

3. State and explain the mathematical model for randomized complete block design. Explain all the steps to be adopted to carry out the analysis, and finally prepare the ANOVA table.

**Group B**

Attempt any Eight questions. (8 x 5 = 40)

4. In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a sample standard deviation of 9.2. At the 1% level of significance, do these data provide considerable evidence that the mean number of concurrent users is greater than 35? Draw your conclusion based on your result.
5. A sample of 250 items from lot A contains 10 defective items, and a sample of 300 items from lot B is found to contain 18 defective items. At a significance level  $\alpha = 0.05$ , is there a significant difference between the quality of the two lots?
6. 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.

Spam status	Image containing status		Total
	With images	No images	
Spam	160	240	400
No spam	140	460	600
Total	300	700	1000

Assess whether being spam and containing images are independent factors at 1% level of significance.

7. Two computer makers, A and B, compete for a certain market. Their users rank the quality of computers on a 4-point scale as "Not satisfied", "Satisfied", "Good quality", and "Excellent quality, will recommend to others. The following counts were observed:

Computer maker	Not satisfied	Satisfied	Good quality	Excellent quality
A	20	40	70	20
B	10	30	40	20

Is there a significant difference in customer satisfaction of the computers produced by A and by B using Mann-Whitney U test at 5% level of significance.

8. Define queuing systems with suitable examples. Also explain the main components of queuing systems in brief.

9. In some town, each day is either sunny or rainy. A sunny day is followed by another sunny day with probability 0.7, whereas a rainy day is followed by a sunny day with probability 0.4. Weather conditions in this problem represent a homogeneous Markov chain with 2 states: state 1 = "sunny" and state 2 = "rainy." Transition probability matrix of sunny and rainy days is given below.

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$$

Compute the probability of sunny days and rainy days using the steady-state equation for this Markov chain.

10. Consider a completely randomized design with 4 treatments with 7 observations in each. For the ANOVA summary table below, fill in all the missing results. Also indicate your statistical decision.

Source	Degrees of freedom	Sum of Squares	Mean Sum of Squares	F-ratio
Treatments	?	SSA =?	70	F =?
Error	?	SSE = 590	?	
Total	?	SST =?		

11. Following are the scores obtained by 10 university staffs on the computer proficiency skills before training and after training. It was assumed that the proficiency of computer skills is expected to be increased after training.

Staffs	Score	
1	Before training	After training
2	50	55
3	30	40
4	15	30
5	22	30
6	34	36
7	45	45
8	40	41
9	10	30
10	26	40

Test at 5% level of significance whether the training is effective to improve the computer proficiency skills applying appropriate statistical test. Assume that the given score follows normal distribution.

12. Write short notes on the following.

a) Concept of Latin Square Design

b) Multiple correlation