POST GRADUATE DIPLOMA IN APPLIED STATISTICS (PGDAST)

Term-End Examination

December, 2019

MST-005 : STATISTICAL TECHNIQUES

Time : 3 hours   *Maximum Marks : 50*

Note : (i) *Question no. 1 is compulsory.*

(ii) *Attempt any four questions from the remaining questions no. 2 to 7.*

(iii) *Use of Scientific Calculator (non-programmable) is allowed.*

(iv) *Use of Formulae and Statistical Tables Booklet for PGDAST is allowed.*

(v) *Symbols have their usual meanings.*

1.   State whether the following statements are True or False. Give reason in support of your answers.   2x5=10

   (a)   The four Pseudo Random Numbers (PRN) using Linear Congruential Generator (LCG) Method :

   $x_i = (1573 x_{i-1} + 19) \mod (10^3)$ are $x_1 = 836$, $x_2 = 47$, $x_3 = 950$ and $x_4 = 379$ where $x_0 = 429$

   (b)   If a sample of size 25 is drawn from a population of size 1024 using simple random sampling technique, then probability that a specified unit is selected in the sample is $24/1023$.

   (c)   In design of experiments, the experimental material is divided into a number of groups or blocks which are so formed that they are within heterogeneous and between homogeneous.

   (d)   In 2^3 factorial experiment, there will be 6 treatment combinations.

   (e)   In one way ANOVA, there is one response variable whereas in two way ANOVA there are two response variables.

2. (a) Explain the procedure of selecting a sample from a population using systematic 3+2
sampling technique. Write the units of your sample if a systematic random sample
of 1000 people from a population of 10,000 is drawn under the assumption that
first randomly selected unit is 4.

(b) Explain the procedure of selecting a sample using stratified random sampling 2+3
technique. Now suppose you work in a company of 1000 employees and want to
find out how they are saving for their retirement. The following information is
collected on age of the employees and no. of employees are recorded for different
age group.

Age (in years)	:	20 - 29	30 - 39	40 - 49	50 - 59	60+
Total no. of employees :		160	220	240	200	180

It is given that ideal sample size for this study is 50. Determine required stratified
random sample size for each stratum using proportional allocation.

3. (a) An investigator is interested in finding the level of knowledge about the history of 7
India of 4 different schools in a city. A test is given to 5, 6, 7, 6 students of 8th
class of the 4 schools. Their scores out of 10 are given below :

School I (S_1)	8	6	7	5	9		
School II (S_2)	6	4	6	5	6	7	
School III (S_3)	6	5	5	6	7	8	5
School IV (S_4)	5	6	6	7	6	7	

Test the equality of average scores of the 4 schools at 5% level of significance.

(b) Discuss two-stage sampling with an example. 3

4. Analyse the data of a 2^3 factorial experiment conducted using a RBD with three 10
replications. The three factors were fertilizers viz. Nitrogen (N), Phosphorus (P) and
Potassium (K). The purpose of the experiment is to determine the effect of different
kinds of fertilizers on potato crop yield. The yields under 8 treatment combinations for
each of the three randomised blocks are given below :

Block - I

npk	(l)	k	np	p	n	nk	pk
450	101	265	373	312	106	291	391

Block - II

p	nk	k	np	(l)	npk	pk	n
324	306	272	338	106	449	407	89

Block - III

p	npk	nk	(l)	n	k	pk	np
323	471	334	87	128	279	423	324

Analyse the design at 5% level of significance.

5. (a) Describe an algorithm using IPT for generating a variate from Logistic distribution, whose p.d.f. is given by :   5

$$f(x) = \frac{e^{-x}}{\left(1 + e^{-x}\right)^2} \, , - \infty < x < \infty$$

Take $U = 0.3$ to generate a variate x from the above distribution.

(b) A train is expected to arrive at a station at 8 : 00 AM. However it has been   5 observed that it reaches at the station between 7 : 55 AM to 8 : 05 AM and the time is uniformly distributed between the above interval. Using the following U (0, 1) random numbers, simulate arrival time of the train for ten days.

U (0, 1) :   0.579   0.052   0.312   0.307   0.645

            0.945   0.645   0.956   0.394   0.110

            0.854   0.876

6. (a) A restaurant wants to estimate the average employees satisfaction with their job **10** (the scale is from 1 to 7). They have 120 restaurants. A stratify random sample of 10 restaurants using proportional allocation. Then they use simple random sampling in each stratum restaurant. The data are given as follows :

Restaurant	Employee Satisfaction
1	5, 7, 6, 5, 4, 7, 6, 6, 4, 5
2	7, 7, 7, 6, 5, 4, 7, 7, 6, 6
3	5, 6, 5, 6, 4, 5, 6, 5, 4, 5, 4, 6, 5, 6
4	6, 5, 7, 6, 7, 6, 5, 7, 5, 7, 6, 5, 7, 6
5	4, 5, 4, 5, 5, 6, 5, 4, 4, 4
6	5, 7, 6, 7, 4, 3, 1, 5, 4, 6, 4, 5
7	7, 6, 7, 7, 6, 6, 5, 7
8	6, 6, 5, 4, 6, 7, 5, 5, 7, 6, 5
9	7, 6, 5, 4, 6, 5, 7, 4, 3, 6, 5, 7
10	7, 6, 6, 5, 7, 3, 5, 4, 5

On the basis of this data what is the estimate of the average employees satisfaction and its variance ?

(b) Explain the middle square method to generate the random numbers with example. **5**

7. In the following design, the letters A, B, C and D represent 4 varieties of wheat; the **10** rows represent 4 different fertilizers; and the columns represent 4 different years. The data on the yields for the 4 varieties of wheat are measured in kilograms per plot. Under the assumption, that various sources of variation don't interact, test at $\alpha = 0.05$, the hypothesis that there is no difference in the

   (i) average yields of the 4 varieties of wheat,
   (ii) fertilizers and
   (iii) years :

Fertilizers \ Year	2001	2002	2003	2004
1	A 70	B 75	C 68	D 81
2	D 66	A 59	B 55	C 63
3	C 59	D 66	A 39	B 42
4	B 41	C 57	D 39	A 55