INDEX

S No	Question No	PAGE No.	Total Marks	Marks obtained	T. Sign.
1	Question No 1	1-4	10		
2	Question No 2	5-6	16		
3	Question No 3	7-09	7		
4	Question No 4	10-11	7		
5	Question No 5	12-15	10		
6	Question No 6.	16-19	12		
7	Question No 7.	20-22	14		
8	Question No 8.	23-24	14		
9	Question No 9.	25-26	10		

SUBJECT NAME : STATISTICAL TECHNIQUES

Question No 1 State whether the following statement are true and false and also give the reason in support of your answer:

(a) The total number of all possible samples of size 2 without replacement from a population of size 7 is 21.

Solution :- $\boxed{\text{The Statement is TRUE}}$

We have $X = 1, 2, 3, 4, 5, 6, 7$

All possible sample of size 2 are as follows:-

(1,2) (1,3) (1,4) (1,5) (1,6) (1,7)
(2,3) (2,4) (2,5) (2,6) (2,7)
(3,4) (3,5) (3,6) (3,7)
(4,5) (4,6) (4,7)
(5,6) (5,7)
(6,7)

(b) Consecutive 3 radom numbers starting from 8937 by 'Middle square method' are 8937, 8699, 6726.

Solution :- $\boxed{\text{The Statement is TRUE}}$

In this approach, if we wish to create a random number of r digits we take a random number of digit r and square it.

If the square has 2r digit, we take the middle r-digits as the next random number.

If there are less than 2r digits in the square we add zeros to produce 2r digits and then take the middle r digits.

In case of 8937, we square it to get the next number in the series we get 79869969 and take the middle 4 digit 8699.

As a result the next produced random number is 8699.

The following random number in this series are 6726, 2390, 7121, ... and so on.

So, consecutive 3 random number starting from 8937 by "middle square method" are 8937, 8699, 6726.

(c) RBD is suitable in situations where it is not possible to divide the experimental material into a number of homogenous blocks.

Solution :- $\boxed{\text{The statement is FALSE}}$

The RBD is suitable in the situations where it is possible to divide the experimental.

material into a number of homogenous Blocks if it is not feasible to split the experimental material. RBD will not be effective in this situation.

(d) As we increase the sample size, representatives of the population by the sample decreases.

Solution :- The Statement is FALSE

The sample should represent all the characteristic of the population that can be possible only when the selection of items has been done unbaised and each and every unit have an equal probability of chance to be selected in the sample.

The number of units or element which are to be selected in the sample should be sufficient. If the sample size is not sufficient then results cannot be reliable.

The more the sample units in the sample, more reliable results would occur.

(e) In a big hall, there are 50 rows and each row has 60 students. A research scholar select 10 rows randomly and then randomly selects 15 students from each selected row. It is an example of cluster sampling procedure.

Solution :- The Statement is TRUE

Cluster sampling is defined as a sampling method where the researcher creates multiple cluster of people from a population where they are indicative of homogenous characterstics and have an equal chance of being a part of the sample.

QUESTION NO 2 (a) Draw all possible sample of size 2 from the population [2,3,4] and verify that $E(\bar{x}) = \bar{X}$. Also find variance of $\bar{x}$.

Solution In SRSWOR the number of sample is

$$N_{C_n} = {}^3C_2 = 3$$

The samples with their means are as follows.

SNo.	Sample	Mean($\bar{x}$)
1	2,3	2.5
2	2,4	3
3	3,4	3.5
		$\sum_{i=1}^{3} \bar{x}_i = 9$

$$E(\bar{x}) = \frac{1}{{}^NC_n} \sum_{i=1}^{{}^NC_n} \bar{x}_i = \frac{9}{3} = 3$$

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{1}{3}(2+3+4) = 3$$

$$\therefore \boxed{E(\bar{x}) = \bar{X}}$$

Again $V(\bar{x}) = \frac{N-n}{N \cdot n} S^2$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{2}\{(2-3)^2 + (3-3)^2 + (4-3)^2\}$$

$$S^2 = 1$$

$$\therefore \boxed{Var(\bar{x}) = \frac{3-2}{3 \times 2} \times 1 = \frac{1}{6} = 0.166}$$

Question No 2 (b) A sample of 60 students is to be drawn from a population consisting of 600 students belonging to two villages, A and B. The means and standard deviations of their marks are given below:

Villages	Strata sizes (N_i)	Means ($\bar{X}_i$)	Standard deviation (σ_i)
Village A	400	60	20
Village B	200	120	80.

Solution If we regard the villages A and B as representing two different strata then the problem is to draw a stratified random sample of size 30 using technique of proportional allocation. In proportional allocation, we have.

$$\boxed{n_1 = \frac{n}{N} N_1}$$

Therefore, $n_1 = \frac{60}{600} \times 400 = 40$

$$n_2 = \frac{60}{600} \times 200 = 20$$

Thus, the required sample sizes for the villages A and B are 40 and 20 respectively.

Question No 3 To determine the yield rate of wheat in a district of Punjab, 6 groups of 6 plots each were constructed. The data are given in the following table:-

Plot No.	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
1	8	6	18	13	17	12
2	13	5	8	7	15	15
3	11	16	6	13	10	11
4	26	5	10	6	21	17
5	13	16	16	7	20	8
6	31	5	20	2	25	10

Select a cluster sample of 3 clusters from the above data and find its sample mean. Further, explain the procedure of two-stage sampling if we want to draw a sample of 6 plots. Which are the 6 plots in your sample?

Solution In the given data, 6 groups have been formed and we have to draw a sample of 3 groups. Therefore, there will be 20 possible sample of size 3 which may be drawn from the population of size 6.

Let we consider one sample of 3 groups drawn from population of 6 groups is {Group 1, Group 3, Group 5}.

$$\boxed{\bar{x} = \frac{n_1\bar{x}_1 + n_3\bar{x}_3 + n_5\bar{x}_5}{n_1 + n_3 + n_5}}$$

Mean of Group 1

$$\bar{x}_1 = \frac{8+13+11+26+13+31}{6} = \frac{102}{6} = 17$$

Mean of Group 3

$$\bar{x}_3 = \frac{18+8+6+10+16+20}{6} = \frac{78}{6} = 13$$

Mean of Group 5

$$\bar{x}_5 = \frac{17+15+10+21+20+25}{6} = \frac{108}{6} = 18$$

Therefore,

$$\text{Grand Mean } \bar{x} = \frac{6\times17 + 6\times13 + 6\times18}{18}$$

$$= \frac{6\times(17+13+18)}{18} = \frac{6\times48}{18}$$

$$= 16.$$

Hence the sample mean is 16.

TWO - STAGE SAMPLING

Let us select 3 groups as first stage units from the given 6 groups. Let the selected units are Group-1, Group 3 and Group 6. Therefore the first stage sample is.

S.No	Group-1	Group-3	Group-6
1	8	18	12
2	13	8	15
3	11	6	11
4	26	10	17
5	13	16	10
6.	31	20	10

Now, from the selected first stage units, we shall select the 6 second stage units.

These units are selected on the basis of their importance. Let we select 2 units each from these three selected first stage units of size 6.

Let us select 2nd and 4th second stage units from group 1, 1st & 4th unit from Gp-3 - 3rd & 6th

$\boxed{\therefore \text{ Second Stage Sample unit} = \{13, 26, 18, 10, 11, 10\}}$

Question No 4 The following data relate to production in kg of three varieties P, Q, R of wheat.

P:	14	16	18		
Q:	14	13	15	22	
R:	18	16	19	15	20

Is there any significant difference among the three varieties at 5% level of significance?

Solution Null hypothesis

$H_0: \mu_1 = \mu_2 = \mu_3$, i.e there is no difference in the production of these varieties P, Q, R against the alternative.

$H_1: \mu_1 \neq \mu_2 \neq \mu_3$

N = Total number of observations = 12

$$\boxed{\text{Correction factor (CF)} = \frac{G^2}{N}}$$

$$CF = \frac{204 \times 204}{12} = 3468$$

$$\boxed{\text{Raw Sum of Squares (RSS)} = \sum\sum y_{ij}^2}$$

$$= 14^2 + 16^2 + 18^2 + 14^2 + 13^2 + 15^2 + 22^2 + 18^2 + 16^2 + 19^2 + 19^2 + 20^2 = 3552$$

$$\boxed{\text{Total Sum of Squares (TSS)} = RSS - CF}$$

$$= 3552 - 3468 = 84$$

$$\boxed{\text{Sum of Squares due to treatment (SST)} = \frac{T_P^2}{3} + \frac{T_Q^2}{4} + \frac{T_R^2}{5} - CF}$$

$$SST = \frac{48 \times 48}{3} + \frac{64 \times 64}{4} + \frac{92 \times 92}{5} - 3468$$

$$\boxed{SST = 16.8}$$

$$\boxed{\text{Sum of Squares due to Error (SSE)} = TSS - SST}$$

$$SSE = 84 - 16.8 = 67.2$$

Mean Sum of Squares due to treatment (MSST) $= \frac{SST}{df} = \frac{16.8}{2} = 8.4$

Mean Sum of Squares due to Error (MSSE) $= \frac{SSE}{df} = \frac{67.2}{9} = 7.467$

Therefore,

$$\boxed{F_{2,9} = \frac{MSST}{MSSE} = \frac{8.4}{7.467} = 1.125}$$

For $v_1 = 2$ and $v_2 = 9$ value of F at 5% level significance is 4.261

$F_{cal} < F_{table\ value}$.

We accept the null hypothesis. There is no significance b/w P, Q and R.

QUESTION No 5 A researcher wants to test four diets A, B, C, D on growth rate in mice. These animals are divided into 3 groups according to their weights. Heaviest 4, next 4 and lightest 4 are put in Block I, Block II and Block III resp. Within each block, one of the diets is given at random to the animals. After 15 days, increase in weight is noted, which is given in the following table:

Blocks	Treatments / Diets			
	A	B	C	D
I	12	8	6	5
II	15	12	9	6
III	14	10	8	5

Perform a two-way ANOVA to test whether the data indicates any significant diff. b/w the four diets due to different blocks.

Solution Null hypothesis are

H_{01}: There is no significant differences b/w mean effects of diets.

H_{02}: There is no significant difference b/w mean effects of different blocks.

Against the alternative hypothesis.

H_{11}: there is significant difference between mean effects of diets.

H_{12}: There is significant difference b/w mean effect of different blocks.

BLOCKS	TREATMENT / DIETS				TOTALS	
	A	B	C	D		
I	12	8	6	5	31	T_1
II	15	12	9	6	42	T_2
III	14	10	8	5	37	T_3
Totals	41 T_1	30 T_2	23 T_3	16 T_4	110	

Squares of observations.

BLOCKS	Treatment / Diets				TOTALS
	A	B	C	D	
I	144	64	36	25	269
II	225	144	81	36	486
III	196	100	64	25	385
Totals	565	308	181	86	1140

Grand Total $= G = \sum\sum y_{ij} = 110$

Correction Factor (CF) $= \frac{G^2}{n} = \frac{(110)^2}{12} = 1008.33$

Raw Sum of Squares (RSS) $= \sum\sum y_{ij}^2 = 1140$

Total Sum Squares (TSS) $=$ RSS $-$ CF

$= 1140 - 1008.33 = 131.66$

Sum of Squares due to Treatments / Diets (SST)

$$= \frac{T_1^2}{3} + \frac{T_2^2}{3} + \frac{T_3^2}{3} + \frac{T_4^2}{3} - CF$$

$$SST = \frac{1}{3}[1681 + 900 + 529 + 256] - 1008.33$$

$$= 113.67.$$

Sum of Squares due to Block (SSB) $= \frac{1}{4}(T_1^2 + T_2^2 + T_3^2) - CF$

$$= \frac{1}{4}[259 + 486 + 385] - 1008.33$$

$$SSB = 15.167$$

Sum of Squares due to Error (SSE) $=$ TSS $-$ SST $-$ SSB

$$= 131.6667 - 113.6667 - 15.16667$$

$$= 2.833$$

Mean sum of Squares due to treatment (MSST)

$$= \frac{SST}{df} = \frac{113.6667}{3} = 37.8889$$

Mean Sum of Squares due to Blocks (MSSB)

$$= \frac{SSB}{df} = \frac{15.1667}{2} = 7.58335$$

Mean Sum of Squares due to Errors (MSSE)

$$= \frac{SSE}{df} = \frac{2.8333}{6} = 0.4722.$$

Source of Variation	SS	DF	MSS	F
B/w Treatments Diets	113.6667	3	$\frac{113.6667}{3} = 37.888$	$F_1 = \frac{37.88}{0.4722} = 80.239$
Between Blocks	15.1667	2	$\frac{15.1667}{2} = 7.5833$	$F_2 = \frac{7.583}{0.47} = 160.896$
Due to Error.	2.8333	6	$\frac{2.8333}{6} = 0.4722$	
	131.6667	11		

Tabulated F at 5% level of significance with (3,6) degree of freedom is 4.76 and tabulated F at 5% level of significance with (2,6) DOF is 5.14

F_1 > tabulated F, we reject H_{01}

F_2 > tabulated F, we reject H_{02}

Question No 6 In the following data, two values are missing. Estimate these values by Yates method and analyse the data by suitable technique.

Treatments	BLOCKS		
	I	II	III
A	12	14	12
B	10	y	8
C	x	15	10

Solution We convert the two missing plot problems into one missing plot problem, for which we take the average of the values of I block in which x is missing. This average is $\frac{10+12}{2}=11$

Thus, the estimate of x is taken to be $x_1=11$ and it is inserted in place of x.

Treatments	Blocks			Total
	I	II	III	
A	12	14	12	$T_A=38$
B	10	y	8	$T_B=18+y$
C	11	15	10	$T_C=36$
Total	$B_1=33$	$B_2=29+y$	$B_3=30$	$G=92+y$

Thus, from the above table we get

$p=3$, $q=3$, $B_2'=29$, $T_B'=18$, $G=92$.

Applying the missing estimation formula

$$\hat{y} = \frac{pT_1 + qB_1 - G}{(q-1)(p-1)} = \frac{3\times18 + 3\times29 - 92}{4} \simeq 12$$

Now the estimated value of y is taken to be $y_1 = 12$ and it is inserted in place of y and the following table of totals is formed by taking x unknown.

Treatments	Blocks			Total
	I	II	III	
A	12	14	12	$T_A = 38$
B	10	12	8	$T_B = 30$
C	x	15	10	$T_C = 25+x$
Totals.	$B_1 = 22+x$	$B_2 = 41$	$B_3 = 30$	$G = 93+x$

Thus, from the above table we get

$p=3$ $q=3$ $B_1 = 22$, $T_C = 25$, $G = 93$

Again applying the missing estimation formula

$$\hat{x} = \frac{3\times25 + 3\times22 - 93}{4}$$

$$x_2 = 12$$

Again using $x_2 = 12$, we estimate the second estimate of y, i.e y_2 for which

$$B_2 = 29, \; T_B = 18, \; G = 92$$

$$\hat{y} = \frac{3 \times 18 + 3 \times 29 - 93}{4}$$

$$\hat{y} = 12$$

Thus, we take the estimated values of $\hat{x} = 12$ and $\hat{y} = 12$. Inserting both the estimated value of x and y

Treatments	Blocks			Total
	I	II	III	
A	12	14	12	$T_A = 38$
B	10	12	8	$T_B = 30$
C	12	15	10	$T_C = 37$
Total	$B_1 = 34$	$B_2 = 41$	$B_3 = 30$	$G = 105$

Correction Factor (CF) $= \frac{(105)^2}{9} = 1225$

Raw Sum of Squares (RSS) = 1261

Total Sum of Squares (TSS) = 36

Treatment Sum of Squares = 12.67

Block sum of Squares (SSB) = 20.67

Error Sum of Squares (SSE) = TSS - SST

= 2.66.

Source of Variation	DF	SS	MSS	Variation Ratio	
				Calculated	Tabulated
Treatments	3-1=2	12.67	6.34	4.77	9.55
Blocks	3-1=2	20.67	10.34	7.77	9.55
Error	4-2=2	2.66	1.33		
Total	8-2=6				

In case of both treatments and blocks, calculated value of F is less than tabulated value of F at 5% level of significance thus treatment and block means are not significantly different.

Question No7 Identify the design in the following table and then carry out the analysis:

Column	Row			
	I	II	III	IV
I	A8	C18	B11	D8
II	C16	B 10	D7	A4
III	B12	D10	A6	C20
IV	D10	A9	C28	B16

Solution

Col. / Row.	I	II	III	IV	Row Total
I	A 8	C 18	B 11	D 8	45
II	C 16	B 10	D 7	A 4	37
III	B 12	D 10	A 6	C 20	48
IV	D 10	A 9	C 28	B 16	63
Column Total	46	47	52	48	193

Correction Factor (CF) $= \frac{(193)^2}{16} = 2328.06$

Raw Sum of Squares (RSS) $= (8)^2 + (16)^2 + .. + 20^2 + 16^2$
$= 2895$

Total Sum of Squares (TSS) $= 2895 - 2328.06$
$= 566.94$

Row Sum of Squares (SSR) $= \frac{45^2 + 37^2 + 40^2 + 63^2}{4} - CF$
$= \frac{2025 + 1369 + 2304 + 3969}{4} - 2328.06$
$= 88.69$

Column Sum of Squares (SSC) $= \frac{46^2 + 47^2 + 52^2 + 48^2}{4} - CF$
$= 5.19$

Treatment Sum of Squares (SST) $= \frac{729 + 2401 + 6724 + 1225}{4} - 2328.06$
$= 441.69$

Error Sum of Squares (SSE) $= TSS - SSR - SSC - SST$
$= 566.94 - 88.69 - 5.19 - 441.69$
$= 31.37$

ANOVA TABLE

Source of Variation	DF	SS	MSS	Variance Ratio		Conclusion
				Cal.	Tab.	
Rows	4-1=3	88.69	29.56	4.71	5.41	Insignificant
Columns	4-1=3	5.19	1.73	0.28	5.41	Insignificant
Treat.	4-1=3	441.69	147.23	23.48	5.41	Significant
Error	6-1=5	31.37	6.27			
Total.	18-1 =14					

Since for treatment effect Calculated value of F is greater than the tabulated value of F at 5%.

$$SE = \sqrt{\frac{2MSSE}{m}} = \sqrt{\frac{2 \times 6.27}{4}} = 1.77.$$

Critical diff. $CD = SE \times t_{\alpha/2} = 1.77 \times 2.571 = 4.55$

Treatment means

$\bar{A} = \frac{27}{4} = 6.75$, $\bar{B} = \frac{49}{4} = 12.25$, $\bar{C} = \frac{82}{4} = 20.5$, $\bar{D} = \frac{35}{4} = 8.75$

Pair of treatments	Difference	CD	Inference.
A, B	$\lvert\bar{A}-\bar{B}\rvert = 05.50$	4.55	Significant
A, C	$\lvert\bar{A}-\bar{C}\rvert = 13.75$	4.55	Significant
A, D	$\lvert\bar{A}-\bar{D}\rvert = 0.2.00$	4.55	Insignificant
B, C	$\lvert\bar{B}-\bar{C}\rvert = 08.25$	4.55	Significant
B, D	$\lvert\bar{B}-\bar{D}\rvert = 0.3.50$	4.55	Insignificant
C, D	$\lvert\bar{C}-\bar{D}\rvert = 11.75$	4.55	Insignificant

Question No 8 (b) Generate a complete cycle for the LCG given below: $x_i = (5x_{i-1} + 3) \bmod 16$ with $x_0 = 8$. A man tosses an unbaised coin ten times. Using the first ten random numbers generated above, obtain a sequence of heads and tails by taking Head (H) as $u \geq 0.5$.

Solution We have given LCG

$$x_i = (5x_{i-1} + 3) \bmod 16, \text{ with } x_0 = 5$$

which gives 10 random number is

12, 15, 14, 09, 00, 03, 02, 13, 04, 07

We also have $u_i = x_i/16$

Therefore,

u_i	0.75	0.94	0.87	0.56	0.00	0.19	0.12	0.81	0.25	0.44
x_i	H	H	H	H	T	T	T	H	T	T.

Question No 8 (a) The distribution function of Pareto distribution is given by

$$f(x) = 1 - \left(\frac{k}{x}\right)^a, \quad a > 0, \; 0 < k \leq x$$

Given a $U \sim U(0,1)$, generate a random number from the above distribution, when $a = 2$ and $k = 1$. Suppose $U = 0.5$, then find x.

Solution Equate

$$U = 1 - \left(\frac{k}{x}\right)^a$$

$$x = \frac{k}{(1-u)^{1/a}}$$

when $u = 0.5$ then

$$x = \frac{1}{(1-0.5)^{1/2}} = \frac{1}{\sqrt{0.5}}$$

$$= \frac{1}{0.707} = 1.414$$

Hence, $x = 1.414$ is the generated value.

Question No 4 Times between successive crashes of a computer system were generated for a 6-month period and are given in increasing order as follows (time in hours):

1	10	20	30	40	52	63	70	80	90	100
102	130	140	190	210	266	310	530	590	640	1340

The parameter $a = 0.00435$, mean $= \frac{1}{\alpha} = 230$ hours
Use the Kolmogorov-Smirnov test to examine the goodness of fit of exponential distribution.

Solution

For the exponential distribution

$$P(X \leq x) = F_0(x) = 1 - e^{-\alpha x}$$

$$= 1 - e^{-0.00435x}$$

The following table gives the required calculations $F_n(x) = x/22$

S.No.	1	2	3	4	5	6	7	8	9	10	11	12
X	1	10	20	30	40	52	63	70	80	90	100	102
$F_n(x)$	0.045	0.091	0.136	0.182	0.227	0.273	0.318	0.364	0.409	0.454	0.500	0.545
$F_0(x)$	0.009	0.051	0.032	0.122	0.160	0.202	0.240	0.262	0.294	0.324	0.353	0.386
$\lvert F_n(x)-F_0(x)\rvert$	0.036	0.040	0.104	0.060	0.067	0.071	0.078	0.102	0.115	0.130	0.147	0.187

S.No.	13	14	15	16	17	18	19	20	21	22
X	130	140	190	210	266	310	530	590	640	1340
$F_n(x)$	0.545	0.591	0.682	0.727	0.773	0.818	0.864	0.909	0.954	1.000
$F_0(x)$	0.358	0.452	0.562	0.594	0.686	0.740	0.900	0.923	0.938	0.997
$\lvert F_n(x)-F_0(x)\rvert$	0.187	0.139	0.180	0.170	0.180	0.087	0.070	0.036	0.016	0.003

$$D_n = \sup_x \lvert F_0(x) - F_n(x) \rvert$$

$$= 0.187$$

and $d_{\alpha,n} = 0.05 = 0.28$

D_n observed is smaller than tabulated value of 0.28 at $\alpha = 0.05$. Hence, we do not reject the null hypothesis.