B.Sc. (Honours) Examination, 2020
Semester-IV (Back)
Statistics
Course: CC-8A
(Survey Sampling & Indian Official Statistics)
Time: 3 Hours **Full Marks: 40**

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any four** questions

1. What is simple random sampling? In an SRSWR of size n from a finite population of size N, find an unbiased estimator of population mean. Find the variance of this estimator. Discuss the estimation procedure of population proportion. Show that in case of SRSWOR, $var(p) = \frac{N-n}{N-1}\frac{PQ}{n}$. 1+3+2+1+3

2. What is the difference between stratified and cluster sampling? Find the bias and mean square error for the ratio estimator. Discuss the method of optimum allocation for determining the sample size for different strata. 3+2+5

3. What are the different sources of non-sampling error? Mention some advantages of using random number tables. 5+5

4. Write short notes on CSO and NSSO. 5+5

5. Discuss NSC and its Main Functions. Write briefly about Rangarajan commission and its various recommendations. 5+5

6. Write short notes on Indian statistical system, statistical system at the centre, MoSPI legal support for collection of data. 4+3+3

B.Sc. (Honours) Examination, 2020
Semester-IV (Back)
Statistics (Practical)
Course: CC-8B
(Survey Sampling & Indian Official Statistics)
Time: 2 Hours Full Marks: 20

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **all** questions

1. A population have 6 units $1, 3, 5, 7, 9, 11$. Write down all possible sample of size 2 without replacement from the given population. Verify that the sample mean is an unbiased estimator of population mean. Also calculate the sample variances for SRSWR and SRSWOR and verify that $V(\overline{x}_{WR}) \geq V(\overline{x}_{WOR})$. 3+2+2

2. Draw a random sample of size 15 from the exponential distribution with parameter $\lambda = 3$. 7

3. A sample of size 3 is to be selected from a population of 13 units. List all possible sample by $i)$ Linear systematic sampling and $ii)$ Circular systematic sampling. 6

B.Sc. Semester IV Examination 2020

Subject: CC 9 (Linear Models) (Theory)

Full Marks: 40 Time: 3 hours

Answer any four questions:

1. Consider the linear model: $y_1 = \theta_1 + \theta_2 + \epsilon_1$,

   $y_2 = \theta_1 - \theta_2 + \epsilon_1$ and $y_3 = \theta_1 - 2\theta_2 + \epsilon_2$

   where $\epsilon_2 \sim iid\ (0, \sigma_e^2)$

   i) Find at least one unbiased estimator for each of $\theta_1$and θ_2.

   ii) Write down the Linear Model in the matrix form and find the least square estimates of θ_1 and θ_2

   iii) State Gauss-Markoff Theorem in this context. 2+6+2

2. Write down the two-way ANOVA model clearly stating the assumptions. Describe the analysis of two-way lay-out data for r- observations per cell to test the equality of effects of two factors. 10

3. a) Differentiate between ANOVA, ANCOVA and regression models with illustrations. 4

   b) Explain how you would test the significance of a regression line. Take only bivariate case (one independent and one dependent variable). 6

4. Consider the three independent random variables y_1, y_2, y_3 having a common variance σ^2 and expectations $E(y_1) = \beta_1 + \beta_2$, $E(y_2) = \beta_1 + \beta_3$ and $E(y_3) = 2\beta_3 - \beta_2$. Find the Best Linear Unbiased Estimator of the function $\beta_1 - \beta_2 + 2\beta_3$. 10

5. a) Write down the assumptions of linear regression model. Mention some suggestions to deal with situations where assumptions are not followed. 6

   b) How would you judge the fitting of a regression from the residual plot? 4

6. There are four objects ω_1, ω_2 ω_3 and ω_4 whose individual weights are to be determined. They are weighted in the following combinations:

Left pan	Right pan	Weight needed for equilibrium
$\omega_1, \omega_2, \omega_3, \omega_4$	----	20
ω_1, ω_2	ω_3, ω_4	10
ω_1, ω_4	ω_2, ω_4	5
ω_1, ω_4	ω_2, ω_3	1

   Obtain the best estimates of individual weights (ignoring bias). 10

B.Sc. Semester IV Examination 2020

Subject: CC 9 (Linear Models) (Practical)

Full Marks: 25 Time: 2 hours

Answer **all** questions:

1. Four catalysts that may affect the concentration of one component in a three-component liquid mixture are being investigated. The following concentrations were obtained:

Catalysts			
1	2	3	4
58.2	56.3	50.1	52.9
57.2	54.5	54.2	49.9
58.4	57.0	55.4	50.1
55.8	55.3		51.7
54.9			

Do the four catalysts have the same effect on concentration? Give reasons and necessary calculations. 7

2. An experiment is conducted to study the influence of opening temperature and type of face-plate glass in the light output of an oscilloscope tube. The following data are collected

Glass type	Temperature		
	100	125	150
1	580	1090	1397
	568	1087	1380
	570	1085	1386
2	550	1070	1328
	530	1035	1312
	579	1000	1299
3	546	1045	867
	575	1053	904
	599	1066	889

Analyse the data and make suitable interpretations. 10

3. Consider the following data and fi a linear regression equation of y on x:

X	10	12	15	18	20	10	34	21	22	26
y	50	60	69	45	42	50	55	63	70	78

Find the R^2 value and comment. Also plot the residuals of the fitted values and comment on your fitting. 8

B.Sc (Honours) Examination, 2020

Semester- IV

Statistics

Course: CC-10

(Statistical Quality Control)

Time: 3 hours **Full Marks: 40**

Questions are of values as indicated in the margin

Answer **any four** questions

1. A. Describe the following terms:

   $3\,\sigma$ −limits, control charts, UCL, LCL, USL, LSL

   B. Describe, in detail, the construction of control chart for number of defects. 6+4

2. A. Discuss the following concepts in connection with sampling inspection plan:

   Consumer's risk, producer's risk, AOQL, OC curve and ASN curve

   B. For a double sampling plan, obtain the expressions for the OC and ASN functions. 5+5

3. A. Describe the construction of $(\bar{X}, S)$ chart for varying sample size.

   B. Describe the technique of sampling inspection by variables in the normal distribution case. 5+5

4. A. Distinguish between –

   i. Process control and product control

   ii. Sample size and rational subgroup

   B. Describe the construction of control chart for proportion of defectives. 4+6

5. When will you use sequential sampling inspection plan? Compare it with the single inspection plan and double inspection plan in terms of number of samples required to test. 10

6. What is a single sampling inspection plan? Find the expression of OC and ASN functions for a single inspection plan. 10

BSC Semester IV Examination, 2020

Statistics

Paper: [CC-10] Statistical Quality Control (Practical)

Full Marks: 20 Time: 2 hours

1. Assume that in the manufacture of 1 kg Mischmetal ingots, the product weight varies with the batch. Below are a number of subsets taken at normal operating conditions, with the weight values given in kg. Construct the (X-bar, R) chart on the basis of these 11 subsets. Measurements are taken sequentially in increasing subset number. Also comment on your findings.

Subset #	Values (kg)
1 (control)	1.02, 1.03, 0.98, 0.99
2 (control)	0.96, 1.01, 1.02, 1.01
3 (control)	0.99, 1.02, 1.03, 0.98
4 (control)	0.96, 0.97, 1.02, 0.98
5 (control)	1.03, 1.04, 0.95, 1.00
6 (control)	0.99, 0.99, 1.00, 0.97
7 (control)	1.02, 0.98, 1.01, 1.02
8 (experimental)	1.02, 0.99, 1.01, 0.99
9 (experimental)	1.01, 0.99, 0.97, 1.03
10 (experimental)	1.02, 0.98, 0.99, 1.00
11 (experimental)	0.98, 0.97, 1.02, 1.03

5

2. Following are the figures for the number of defectives in 16 lots, each containing 2,000 rubber belts:
   341, 225, 322, 280, 306, 337, 305, 356, 402, 216, 264, 126, 409, 193, 326, 280,
   Drawing the control chart for fraction defective, plot the points on it. Comment on the state of control of the process. 5
3. Determine for a sequential sampling plan involving item-by-item inspection, for which $p_0 = 0.01, p_1 = 0.03, \alpha = 0.05$ and $\beta = 0.2$, the acceptance and rejection lines. 6
4. From each lot of 5000 chalks, take 200 at random and inspect them. Accept the lotifthe inspected sample contains at most two defective chalk; otherwise reject it.Suppose a lot has 30 defective items. What is P_a for such a lot? Plot the OC curve for this sampling inspection plan. 4

B.Sc. (Honours) Examination, 2020
Semester-VI
Statistics
Course: CCST-13A
(Design of Experiments)
Time: Three Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any five** questions

1. Discuss basic principles of design of experiments. How do the size and shape of plots and blocks affect the result of a field experiment? 4+4

2. What is a standard Latin Square Design (LSD)? How basic principles of design of experiments are used in LSD? How many different LSD's can be generated from a standard LSD by permuting its rows and columns? When two LSDs are said to be orthogonal? Give an example of two mutually orthogonal LSDs. 2+2+2+1+1

3. How will you estimate the yield of a missing plot in an RBD? Discuss in detail how you will carry out the analysis of an RBD after estimating the yield of missing plot. 3+5

4. Write down the ANCOVA model for RBD with one concomitant variable. Outline a method to judge whether the inclusion of the concomitant variable is worthwhile or not. If worthwhile, give the detailed analysis. 2+3+3

5. (a) What is a factorial experiment? Differentiate it from a single-factor experiment.
(b) How do you obtain the SS due to main effects or interaction effects in a 2^3 experiment? Give only the AVOVA table of a 2^3 experiment conducted in randomized blocks. 3+5

6. Consider (2^4, 2^2) design. Treatment combinations belonging to a block is a, b, cd, abcd. (i) Construct the other three blocks. (ii) Find the confounded effects. (iii) Give the analysis of this experiment. 3+2+3

7. Differentiate the strip-plot experiment from the split-plot experiment. Discuss the layout and analysis of a strip-plot experiment in RBD. 3+5

8. Write short note on any two of the following: 4+4
(i) Confounding in design of experiments.
(ii) Series of experiments.
(iii) Efficiency of LSD compared to RBD.

B.Sc. (Honours) Examination, 2020
Semester-VI
Statistics
Course: CCST-13B (Practical)
(Design of Experiments)
Time: Two Hours Full Marks: 20

Questions are of value as indicated in the margin
Notations have their usual meanings

1. An experiment on sugarcane conducted in 5 randomized blocks gave the following values of number of plants per plot(x) and weight of cane in kg (y). The data on number of plants provide a basis for error control through analysis of covariance. The three treatments are manures-Nitrogen (N), Phosphorus (P) and Potash (K) in appropriate doses.

Block	N		P		K	
	x	y	x	y	x	y
I	40	122	41	81	42	82
II	41	120	50	79	38	80
III	38	138	46	80	54	65
IV	41	121	42	75	40	58
V	39	126	48	83	45	76

Analyze the data. In case of significance, judge the individual treatment differences with a view to finding the best treatment. 7

2. The layout and yields in a manorial experiment using three fertilizers: A: No manure, B: 1.1/2 superphosphate per tree, C: 3 lb. superphosphate per tree in three 3x3 Latin squares are given below:

Square-1			Square-2			Square-3		
B	C	A	C	B	A	A	C	B
41	25	15	28	27	3	11	14	17
A	B	C	A	C	B	B	A	C
22	32	24	4	17	9	24	15	33
C	A	B	B	A	C	C	B	A
20	12	21	22	4	17	22	20	15

Analyze the data and write a report suggesting the best dose of the fertilizer. 8

3. Obtain the layout of a 2^3-factorial experiment in 4 replicates partially confounding all first order and second order interactions. 5

B.Sc. (Honours) Examination, 2020
Semester-VI
Statistics
Course: CC-14A
Multivariate Analysis & Nonparametric Methods

Time: 3 hrs Full Marks:40

Answer any **four** questions of the following.

1. (a) Define p-variate concentration ellipsoid.

   (b) In the context of p-variate linear multiple regression, coefficient of determination is found to be .9. Interpret the result in terms of regression.

   (c) Suppose a group of students have the average height μ_0. A researcher, on the basis of 10 students of that class, claims that the average height is smaller than μ_0. Propose a nonparametric test to justify his claim.

   3+3+4

2. (a) Suppose you have a data set of size 9. Discuss a test procedure briefly to check if the data comes from standard normal variable.

   (b) What do you mean by run created by two alphabets?

   (c) When do you use Run test? Briefly state the rejection rules for run test under different alternatives.

   3+2+2+3

3. (a) Suppose you have a data set of size 8. You want to test if the central value of the population, the data coming from, is 20. Under this null hypothesis, deduce the expectation and variance of Wilcoxon signed rank test statistic.

   (b) Let X be a 4×1 component vector following multinomial distribution (n, p_1, p_2, p_3). Show that the regression equation of X_1 on X_2, X_3, X_4 is linear. Also check conditional variance is homoscedastic.

   5+3+2

4. (a) Let $X_p \sim N_p(\mu, \Sigma)$. Let us partition $\mathbf{X}_{p\times 1} = \begin{pmatrix} \mathbf{X}_{(1)q\times 1} \\ \mathbf{X}_{(2)p-q\times 1} \end{pmatrix}$ and $\mu = \begin{pmatrix} \mu_{(1)q\times 1} \\ \mu_{(1)p-q\times 1} \end{pmatrix}$ and $\Sigma_{p\times p} = \begin{pmatrix} \Sigma_{11q\times q} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Show that the conditional distribution of $X_{(2)}$ given $X_{(1)}$ follow a multivariate normal of order $p - q$.

   (b) Let $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3 \left[(1, -1, 3)', \Sigma = \begin{pmatrix} 40 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{pmatrix} \right]$. Check if X_1 and $X_1 + 3X_2 - 2X_3$ are independent.

6+4

5. (a) Establish the relation ${\rho_{1.23\cdots p}}^2 = (1 - \frac{1}{\sigma_{11}\sigma^{11}})^{1/2}$ where $\rho_{1.23\cdots p}$ being the population correlation coefficient, σ^{11} being the $(1,1)$ element of Σ^{-1} and σ_{11} being the $(1,1)$ element of Σ.

   (b) Let $r_{12.3} = 0$, $r_{13.2} = 0$ and $r_{23.1} = 0$. Does it imply $r_{1.23} = 0$? Explain.

   (c) Show that for given r_{12} and r_{13}, r_{23} must lie in the range $r_{12}r_{13} \pm (1 - r_{12}^2 - r_{13}^2 + r_{12}^2 r_{13}^2)^{1/2}$.

5+2+3

6. (a) Show that in multiple regression theory, the residual variance $s^2_{1.23\cdots p} = (1 - r^2_{1.23\cdots p})s_1^2$ where $r_{1.23\cdots p}$ being the multiple correlation.

   (b) Suppose you want to test whether distribution of X is stochastically larger than that of Y with respect to a location parameter θ. Write the null and alternative hypothesis in terms of distribution functions. Also propose a test statistic for testing it.

5+3+2

B.Sc. (Honours) Examination, 2020
Semester-VI
Statistics
Course: CC-14B
Multivariate Analysis & Nonparametric Methods
Time: 2 hrs Full Marks:20

Answer **all** the questions. Tables are attached at the bottom of the questions.

1. let X_1, X_2 and X_3 are three correlated variables where $s_1 = s.d(X_1) = 1, s.d.(X_2) = 1.3, s_3 = s.d(X_3) = 1.9$ and $r_{12} = .37, r_{13} = -.641$ and $r_{23} = -.736$. Compute $r_{23.1}$. If $X_4 = X_1 + X_2$, then obtain $r_{42.3}$.

2+3

2. $\mathbf{X}_{3\times1} \sim N_3(\mu, \Sigma)$ where $\mu = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{pmatrix}$.
   (a) Which one of the following combination of random variable(s) i s/are independent?
      i. X_1, X_2
      ii. X_1 & $X_1 + 3X_2 - 2X_3$
      iii. X_1& X_3
   (b) Find out the conditional distribution of X_1 given $X_2 = 2$ and $X_3 = 4$.

3+2

3. The data below represent earnings(in dollars) for a random sample of seven common stocks listed in the new York stock exchange.
   1.68, 3.35, 2.50, 6.23, 3.24, 4.29, 2.99. Check if these data can be regarded as a random sample from a normal distribution with $\mu = 3$ and $\sigma = 1$.

5

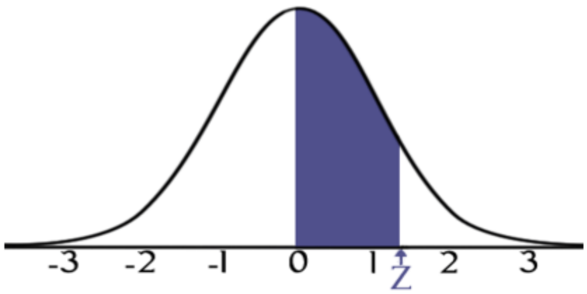
4. Consider the following combination of upward and downward price changes
   $+, +, -, -, +, -, +, -, -, -, +, +$
   where $+$: upward price change and $-$: downward price change. Does it approve that price change (upward/downward) happen in random manner?

5

alpha values							
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
5	--	--	--	--	--	0	2
6	--	--	--	--	0	2	3
7	--	--	--	0	2	3	5
8	--	--	0	2	3	5	8
9	--	0	1	3	5	8	10
10	--	1	3	5	8	10	14
11	0	3	5	8	10	13	17
12	1	5	7	10	13	17	21
13	2	7	9	13	17	21	26
14	4	9	12	17	21	25	31
15	6	12	15	20	25	30	36
16	8	15	19	25	29	35	42
17	11	19	23	29	34	41	48
18	14	23	27	34	40	47	55
19	18	27	32	39	46	53	62
20	21	32	37	45	52	60	69
21	25	37	42	51	58	67	77
22	30	42	48	57	65	75	86
23	35	48	54	64	73	83	94
24	40	54	61	72	81	91	104
25	45	60	68	79	89	100	113
26	51	67	75	87	98	110	124
27	57	74	83	96	107	119	134

alpha values							
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
28	64	82	91	105	116	130	145
29	71	90	100	114	126	140	157
30	78	98	109	124	137	151	169
31	86	107	118	134	147	163	181
32	94	116	128	144	159	175	194
33	102	126	138	155	170	187	207
34	111	136	148	167	182	200	221
35	120	146	159	178	195	213	235
36	130	157	171	191	208	227	250
37	140	168	182	203	221	241	265
38	150	180	194	216	235	256	281
39	161	192	207	230	249	271	297
40	172	204	220	244	264	286	313
41	183	217	233	258	279	302	330
42	195	230	247	273	294	319	348
43	207	244	261	288	310	336	365
44	220	258	276	303	327	353	384
45	233	272	291	319	343	371	402
46	246	287	307	336	361	389	422
47	260	302	322	353	378	407	441
48	274	318	339	370	396	426	462
49	289	334	355	388	415	446	482
50	304	350	373	406	434	466	503

tables



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.

	alpha values						
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
5	--	--	--	--	--	0	2
6	--	--	--	--	0	2	3
7	--	--	--	0	2	3	5
8	--	--	0	2	3	5	8
9	--	0	1	3	5	8	10
10	--	1	3	5	8	10	14
11	0	3	5	8	10	13	17
12	1	5	7	10	13	17	21
13	2	7	9	13	17	21	26
14	4	9	12	17	21	25	31
15	6	12	15	20	25	30	36
16	8	15	19	25	29	35	42
17	11	19	23	29	34	41	48
18	14	23	27	34	40	47	55
19	18	27	32	39	46	53	62
20	21	32	37	45	52	60	69
21	25	37	42	51	58	67	77
22	30	42	48	57	65	75	86
23	35	48	54	64	73	83	94
24	40	54	61	72	81	91	104
25	45	60	68	79	89	100	113
26	51	67	75	87	98	110	124
27	57	74	83	96	107	119	134

	alpha values						
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
28	64	82	91	105	116	130	145
29	71	90	100	114	126	140	157
30	78	98	109	124	137	151	169
31	86	107	118	134	147	163	181
32	94	116	128	144	159	175	194
33	102	126	138	155	170	187	207
34	111	136	148	167	182	200	221
35	120	146	159	178	195	213	235
36	130	157	171	191	208	227	250
37	140	168	182	203	221	241	265
38	150	180	194	216	235	256	281
39	161	192	207	230	249	271	297
40	172	204	220	244	264	286	313
41	183	217	233	258	279	302	330
42	195	230	247	273	294	319	348
43	207	244	261	288	310	336	365
44	220	258	276	303	327	353	384
45	233	272	291	319	343	371	402
46	246	287	307	336	361	389	422
47	260	302	322	353	378	407	441
48	274	318	339	370	396	426	462
49	289	334	355	388	415	446	482
50	304	350	373	406	434	466	503

Figure 1: Wilcoxon signed rank statistic for one sample

Kolmogorov-Smirnov Test
Critical Values

SAMPLE SIZE (N)	LEVEL OF SIGNIFICANCE FOR D = MAXIMUM [$F_0(X)$ - $S_n(X)$]				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.210	.220	.240	.270	.320
30	.190	.200	.220	.240	.290
35	.180	.190	.210	.230	.270
OVER 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

Figure 2: Kolmogorov Smirnov Table

TABLE G

Critical values of *r* in the runs test*

Given in the tables are various critical values of *r* for values of *m* and *n* less than or equal to 20. For the one-sample runs test, any observed value of *r* which is less than or equal to the smaller value, or is greater than or equal to the larger value in a pair is significant at the $\alpha = .05$ level.

m \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2	2	2	2	2	2	2	2	2
											–	–	–	–	–	–	–	–	–
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
					–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
				9	9	–	–	–	–	–	–	–	–	–	–	–	–	–	–
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
			9	10	10	11	11	–	–	–	–	–	–	–	–	–	–	–	–
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
		–	9	10	11	12	12	13	13	13	13	–	–	–	–	–	–	–	–
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
		–	–	11	12	13	13	14	14	14	14	15	15	15	–	–	–	–	–
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
		–	–	11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
		–	–	–	13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
		–	–	–	13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
		–	–	–	13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
	–	–	–	–	13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
	–	–	–	–	–	15	16	17	18	19	19	20	20	21	21	22	22	23	23
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
	–	–	–	–	–	15	16	17	18	19	20	20	21	22	22	23	23	23	24
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
	–	–	–	–	–	15	16	18	18	19	20	21	22	22	23	23	24	24	25
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
	–	–	–	–	–	–	17	18	19	20	21	21	22	23	23	24	25	25	25
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
	–	–	–	–	–	–	17	18	19	20	21	22	23	23	24	25	25	26	26
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
	–	–	–	–	–	–	17	18	19	20	21	22	23	24	25	25	26	26	27
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
	–	–	–	–	–	–	17	18	20	21	22	23	23	24	25	26	26	27	27
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14
	–	–	–	–	–	–	17	18	20	21	22	23	24	25	25	26	27	27	28

* Adapted from Swed, and Eisenhart, C. (1943). Tables for testing randomness of grouping in a sequence of alternatives. *Annals of Mathematical Statistics*, **14**, 83–86, with the kind permission of the authors and publisher.

table.pdf

B.Sc. (Honours) Examination, 2020
Semester-VI
Statistics
Course: DSE-3
(Operations Research)
Time: 3 Hours **Full Marks: 40**

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any four** questions

1. What is a convex set? What are its extreme points? Show that the intersection of two convex sets is also a convex set. What can you say about the union of two convex sets? Prove that in E^2, the set $X = \{(x, y) : \ x^2 + y^2 \leq 7\}$ is a convex set. 1+1+3+1+4

2. Briefly discuss the different phases of operations research. 10

3. a) Show that in a balanced transportation problem having m origins and n destinations $(m, n \geq 2)$, the exact number of basic variables is $m + n - 1$.
b) Find the optimal assignment and the optimal assignment and corresponding cost matrix $\begin{bmatrix} -14 & -7 & -22 & -11 & -6 \\ -18 & -22 & -14 & -15 & -9 \\ -18 & -12 & -9 & -12 & -12 \\ -10 & -22 & -15 & -22 & -8 \\ -16 & -16 & -14 & -10 & -10 \end{bmatrix}$. 5+5

4. a) What do you mean by dominance property in game theory? Solve the 2×4 game graphically $\begin{bmatrix} 2 & 2 & 3 & -1 \\ 4 & 3 & 2 & 6 \end{bmatrix}$.
b) Solve the LPP by algebraic method: Maximize, $z = 5x_1 + 2x_2 + 2x_3$. Subject to, $x_1 + 2x_2 - 2x_3 \leq 30$, $x_1 + 3x_2 + x_3 \leq 36$, $x_1, x_2 \geq 0$. (2+4)+4

5. a) Show that every two person zero sum game problem can be converted into a linear programming problem.
b) Briefly discuss about the inventory control problems. 6+4

6. a) Formulate an EOQ model with non-uniform demand rate and infinite replenishment rate.
b) A company has a demand of $24,000$ units/year and it can produce 3000 such item per month. The cost of one setup is 400 and the holding cost/unit/month is $Rs.$ 0.20. Find the optimum lot size and the total cost per year, assuming the cost of 1 unit as $Rs.$ 4.00. Also find the maximum inventory and manufacturing time. 5+5

B.Sc. (Honours) Examination, 2020
Semester-VI
Statistics (Practical)
Course: DSE-3B
(Practical on Operations Research)
Time: 2 Hours **Full Marks: 20**

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **all** questions

1. Reduce the following game to a 2×2 game and then solve it.

$$\begin{bmatrix} 4 & 1 & 2 \\ 2 & 3 & 5 \\ 3 & 2 & 1 \end{bmatrix}.$$ 6

2. Obtain the initial basic feasible solution to the following transportation problem by matrix minima method then find out an optimal solution and corresponding cost of the transportation.

$$\begin{bmatrix} & D_1 & D_2 & D_3 & D_4 & a_i \\ O_1 & 5 & 4 & 6 & 14 & 15 \\ O_2 & 2 & 9 & 8 & 6 & 4 \\ O_3 & 6 & 11 & 7 & 13 & 8 \\ b_j & 9 & 7 & 5 & 6 & 27 \end{bmatrix}$$ 7

3. Solve the following linear programming problem by simplex method.

$Minimize\ Z = 4x_1 + 3x_2 + 5x_3,$
Subject to,
$2x_1 + 3x_2 + 2x_3 \leq 440,$
$4x_1 + 3x_3 \leq 470,$
$2x_1 + 5x_2 \leq 430,$
$x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0.$ 7

Ref.
No.**VB/EXAM/REG/06/UG/STAT/SemIV/18043/2020**

This question paper contains two pages

B.Sc.

Semester IV Examination, 2020

Statistics

SECC-2 Statistical Data Analysis Using R

Time: Two Hours Full Marks: 25

Answer any **five** questions.

1. Describe different data structures in R.

2. a) What would be the output of the following R code

   log(-2, base=10)

   b) Let A be a data set already saved in R. Write the code to find the number of missing observations in A.

3. A data set in excel file contains 100 observations on few numeric columns and few qualitative variables. As usual each column is a variable. Write R code for the followings:

   a) Read the data in object name 'data' when the excel file is saved as .csv.

   b) Read the data in object name 'data2' when the excel file is saved as .txt

   c) Find the nature of each columns of the data set.

   d) Find the dimension of the dataset

   e) To view the last 10 observations of the data set.

4. What are the steps to build and evaluate a linear regression model in **R**?

5. a) How to install and make use of one package in R (Assume name of package is new_package)

   b) Mention names of two packages in R which you have used.

6. Write down the output of the following code:

```
fruits<-c("apple","orange","pomegranate")
for(i in fruits){
  print(i)
}
```

7. Let us consider the following object x and y in R

x <-c(1,2,3,4,5,6,7,8,9, 10, 11)
y <-c(1.5,3.4, 6, 6.4, 4, 6.7, 8.1, 2, 5.9, 4.7, 9)

Write the R code to compute the mean, variance standard deviations of x and y. Also write the code for the scatter plot of the variables x, y with a proper heading of the plot.

Supplementary Undergraduate Examination-2020
Semester-II(CBCS)
Subject- Statistics
Generic Elective Course - (STAT-GE-2)
Introductory Probability

Time-3 hours

Full Marks-40

Answer Question no. 1 and answer any 3 from the rest. All questions carry equal credits.

Q1. Answer all the following:

i. You own an unusual die. Three faces are marked with the letter "X," two faces with the letter "Y," and one face with the letter "Z." What is the probability that at least one of the first two rolls is a "Y"?

(A) $\frac{1}{6}$ (B) $\frac{2}{3}$ (C) $\frac{1}{3}$ (D) $\frac{5}{9}$

ii. You roll two six-sided dice. What is the probability that the sum is 6 given that one die shows a 4?

(A) $\frac{2}{12}$ (B) $\frac{2}{11}$ (C) $\frac{11}{36}$ (D) $\frac{2}{36}$

iii.Which of the following statements is incorrect?

(A) The histogram of a binomial distribution with p = 0.5 is always symmetric no matter the value of n, the number of trials.
(B) The histogram of a binomial distribution with p = 0.1 is unimodal.
(C) The histogram of a binomial distribution with p = 0.1 is skewed to the left.
(D) The histogram of a binomial distribution with p = 0.01 looks more and more symmetric, the larger the value of n.

iv. Suppose E and F are independent events with P(E) = 0.4 and P(E and F) = 0.15. Which of the following is a true statement?

(A) P(F) = 0.4
(B) P(F) = 0.6
(C) P(E or F) = 0.375
(D) P(E or F) = 0.625

v. A dentist compiles the following summary data from interviews of her patients. 55% floss once a day, and these patients have a 0.04 probability of a cavity each year. 30% floss twice a day, and these patients have a 0.01 probability of a cavity each year. 15% don't floss, and these patients have a 0.10 probability of a cavity each year. What is the probability that one of the dentist's patients flosses and has a cavity for any given year?

(A) 0.04 + 0.01
(B) 0.04 + 0.01 + 0.10
(C) (0.55 + 0.30)(0.04 + 0.01)
(D) (0.55)(0.04) + (0.30)(0.01)

vi. Suppose two events, S and T, have the nonzero probabilities p and q, respectively. Which of the following is impossible?

(A) $p + q > 1$
(B) $p – q < 0$
(C) $pq > 1$
(D) S and T are both independent and mutually exclusive.

vii. According to a college's records, 46 percent of its students live in the dorms, 65 percent have meal contracts, and 33 percent do both. For a randomly selected student, what is the probability that he/she either lives in a dorm or has a meal contract, but not both?

(A) 0.13 (B) 0.22 (C) 0.32 (D) 0.45

viii. 85 percent of people wear seat belts. The probability of serious injury in an accident is 8 percent for those wearing seat belts and 36 percent for those not wearing seat belts. What is the probability of serious injury in an accident?

(A) 0.054 (B) 0.068 (C) 0.122 (D) 0.318

ix. At many colleges, many nonmath majors choose to take at least one math/computer science course. Suppose 16 percent take computer science, 53 percent take statistics, and 62 percent take at least one of these two offerings. What is the probability a randomly chosen nonmath major takes both a computer science class and a statistics class?

(A) (0.16)(0.53) (B) 0.16 + 0.53 – 0.62 (C) 0.62 – (0.16)(0.53) (D) (0.62 – 0.16)(0.62 – 0.53)

x. A research firm is successful in contacting 65 percent of the households randomly selected for telephone surveys. What is the probability that the firm is successful in contacting a household given that it was unsuccessful in contacting the previous two households?

(A) 0.65 (B) $(0.35)^2(0.65)$ (C) $3(0.35)^2(0.65)$ (D) $1-(0.35)^2(0.65)$

Q2. Compute the mode of the binomial distribution with parameters n and p.

Q3. Derive the limiting form of binomial distribution when n is large and p is very small.

Q4. Define normal distribution. Derive its mean and standard deviation. Using these mean and standard deviation give the standard normal random variable.

Q5. State atleast seven charactaristic of normal distribution. State the law of large number.

Q6. State and prove Baye's probability theorem.

Supplementary Undergraduate Examination-2020
Semester-II(CBCS)
Subject- Statistics
Generic Elective Course - (STAT-GE-2)(Practical)
Introductory Probability

Time-2 hours

Full Marks-20

Answer all the questions. All questions carry equal marks.

Q1. Seven coins are tossed and number of heads noted.The experiment is repeated 128 times and the following distribution is obtained:

No.of heads	0	1	2	3	4	5	6	7	Total
Frequencies	7	6	19	35	30	23	7	1	128

Fit a Binomial dtstribution assuming:

(i) The coin is unbaised,
(ii) The nature of the coin is not known.
(iii) Probability of a head for four coins is 0·5 and for the remaining three coins is 0·45.

Q2. Obtain the equation of the normal curve that may be
tted the following data:

Class	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100
Frequencies	3	21	150	335	326	135	26	4

Also obtain the expected normal frequencies.

M.Sc. Examination, 2020
Semester-II
Statistics
Course: MSc21
Inference II

Time: 3 hrs Full Marks:40

Answer any **four** questions of the following.

1. (a) State and prove Neyman Pearson fundamental Lemma.
   (b) Let there be a test $H_0 : X \sim f_0$ against $H_1 : X \sim f_1$ where $f_0 = \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}x^2)$ and $f_1 = exp(-x)$. Construct a Neyman Pearson test.

   6+4

2. (a) Find the mean and variance of one sample general linear rank statistic.
   (b) Define one sample Wilcoxon Signed rank test statistic and describe in which context will it be used?

   6+4

3. (a) Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where σ is unknown. Construct a test $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu > \mu_0$.
   (b) What do we mean by an unbiased test? Prove that Most powerful test is unbiased.

   5+5

4. (a) Establish that Kolmogorov-Smirnov one sample test statistic is distribution free.
   (b) Define monotone likelihood ratio property of a family of distribution. Show that generalized exponential family obeys MLR property.

   5+5

5. Write short note on any two of the following.
   (a) Mann Whitney U statistic
   (b) Size of a test
   (c) Kruskal Wallis Test

   5+5

M.Sc.
Semester II Examination, 2020
Statistics

MSC-22 (Applied Multivariate Analysis)

Time: Three Hours Full Marks: 40

Answer any **four** questions.

1. a) Define Principal Components and write its objectives.

   b) Derive the first two Principal Components of a p-variate random vector. 4+6

2. a) Clearly stating the assumptions write down the factor orthogonal model with m-common factors.

   b) Show that how the factor model approximates the variance-covariance matrix of the multivariate data. 5+5

3. a) Does factor model solution always exist? Answer with reasons.

   b) Differentiate between discriminant analysis and classification. Give examples of both the techniques. 5+5

4. a) Derive the discriminant rule by minimizing the Expected Cost of Misclassification (ECM) assuming all relevant information are given.

   b) Explain the decision rule algorithm for the purpose of classification.

5. a) Explain the concept of canonical correlation and derive the first pair of canonical variables and find their correlation.

   b) Write down the agglomerative hierarchical clustering method. 6+4

6. a) Explain the concept of clustering and give some real-life examples where this method is applied.

   b) Differentiate Single linkage, Complete linkage, and Average linkage methods of clustering.

   c) What is a dendrogram? 3+5+2

Visva Bharati University
M.Sc. Semester II Back Paper Examination 2020
Subject: Statistics
Paper: MSC-23 (Sample Survey)

Full Marks: 40 Time: 3 Hrs.

Answer any four questions.

1. (a) State and prove a necessary and sufficient condition for the existence of an unbiased estimator of the population total.

   (b) Show that among the class of HLUEs, no one exists with uniformly minimum variance.

   5+5=10

2. (a) Suggest an unbiased estimator of the population total under PPSWR sampling scheme. Find its variance.

   (b) Estimate the gain in precision if we use PPSWR instead of SRSWR for estimating the population total. You should prove the necessary results.

   (2+2)+6=10

3. (a) Define Desh Raj's estimator . Show that it is unbiased for the population total.

   (b) Find the unbiased estimator of its variance.

   (c) Mention a drawback of the estimator and the way of improvement.

   3+4+3=10

4. (a) Distinguish between informative and non-informative sampling design.

   (b) For a binary variable Y, prove that $P\left[\overline{y} - \dfrac{3}{4n} \leq \overline{Y} \leq \overline{y} - \dfrac{3}{4n}\right] \geq \dfrac{8}{9}$, where the symbols have their usual meanings..

   2+8=10

5. (a) Explain the problem you will encounter while asking people directly about their belongings to a dichotomous sensitive characteristic.

   (b) Discuss how will you estimate the proportion of individuals belonging to different stigmatizing versions of a polytomous qualitative characteristic using unrelated questionnaire method, where the unrelated questions have 'Yes/No' type response.

   4+6=10

6. (a) Compare the relative merits and demerits of cluster sampling and two-stage sampling.

   (b) Suggest an unbiased estimator of population total under two-stage sampling. Find its variance and an unbiased estimator of the variance,

   2+(1+7)=10

7. Given an unbiased estimator t of the population total Y based on a sample s drawn according to a design p, describe a procedure to construct an estimator having variance smaller than that of t.

   10

M.Sc. Examination, 2020

Semester-II

Statistics

Course: MSc 25 (Practical)

Time: 4 hrs Full Marks:40

1. Construct a test for $H_0 : \sigma^2 = 3$ against $H_1 : \sigma^2 > 3$ when a sample of size 10 is drawn from a $N(2, \sigma^2)$. The sample observations are given below.
   3.2,4.1,1.9,.8,0,1.6,6.9,9.2,6.2,7.0.
   The level of significance you can choose .05. Is it a UMP test?

   4+2

2. Propose a test for testing $\lambda = 1.5$ against $\lambda = 3$ in the context of Poisson distribution where the sample observations are 1,1,5,6,8,4,0,0,3,3. Choose $\alpha = .1$.

   4

3. The data below represent earnings(in dollars) for a random sample of seven common stocks listed in the new York stock exchange.
   1.68, 3.35, 2.50, 6.23, 3.24, 4.29, 2.99. Check if these data can be regarded as a random sample from a normal distribution with $\mu = 3$ and $\sigma = 2$.

   5

Visva-Bharati University
M. Sc. Semester-II Back Paper Examination, 2020
Subject: Statistics (Practical)
Paper: MSC-26

Time: Four hours | Full Marks: 40

---

1. Obtain the C matrix of the following block design 5

Block 1	1 3 5
Block 2	2 4
Block 3	3 5 7
Block 4	4 6
Block 5	1 5 7
Block 6	2 6
Block 7	8 9

2. Consider the following incomplete block design with $v = 7$ treatments, 1, 2, , 7 and $b = 7$ blocks. The blocks are

Block 1	1 2 3
Block 2	1 2 4
Block 3	1 3 4
Block 4	2 3 4
Block 5	5 6
Block 6	5 7
Block 7	6 7

Show that the design is disconnected. 5

3. Consider an incomplete block design with $v = 5$ treatments, $b = 8$ blocks and block contents as follows:

Block 1	1 2 3
Block 2	1 2 4
Block 3	1 3 4
Block 4	2 3 4
Block 5	1 1 5
Block 6	2 2 5
Block 7	3 3 5
Block 8	4 4 5

Examine whether this design is variance-balanced. 5

4. Construct a Hadamard matrix of order 8 starting from a Hadamard matrix of order 2. Use this matrix to construct a BIBD. What are the parameters of this design? Based on the incidence matrix of the design, verify the statement - Each pair of treatments appears in blocks = the

number of treatments common between every pair of distinct blocks. 5

5. The following is the population of M.Sc. (Statistics) students of a certain univer- sity, together with the data on two variables, viz. Marks in Sample survey and Marks in Estimation (both out of 50).

Serial Number	Student Name	Marks in Sample Survey	Marks in Estimation
1	Ajoy	30	45
2	Bikram	24	32
3	Buddhadev	40	47
4	Anushree	36	40
5	Payel	31	25
6	Ramen	40	30
7	Ritika	34	29
8	Rima	28	43
9	Somnath	25	27
10	Riddhi	31	44
11	Sulagna	43	31
12	Surya	48	46
13	Dipanjan	32	34
14	Bishal	43	49
15	Sudip	27	39
16	Molay	40	22
17	Arnoneel	34	28
18	Arindam	36	33

(a) Select 3 students (distinct) from this list using simple random sampling. Estimate the mean marks of all students in Sample Survey, using their marks in Estimation as an auxiliary information. Provide an estimate of the variance of your estimator.

(b) Next select 3 students (distinct) from this list using PPS sampling and estimate the mean marks of all students in Sample Survey. Provide an estimate of variance of your estimator.

(c) Now form 3 groups, each containing 6 students using SRSWOR. From each group, select one student by PPS sampling. Under this scheme, provide an estimate of the mean marks of all students in Sample Survey.

(d) Compare the estimate obtained in these three cases with the true mean. 6 + 6 +5 + 3 = 20

x———— Best of Luck————x

M.Sc. Examination, 2020
Semester-IV
Statistics
Course: MSC-41(MSS-4)
(Demography)
Time: Three Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any five** questions

1. (a) Assuming an exponential distribution for time to first conception and choosing an appropriate prior for the parameter involved show that the resulting distribution is decreasing failure rate.
(b) Starting from suitable assumptions derive the distribution of time to the nth child birth. 4+4

2. Derive Lotka's integral equation for birth function and hence explain a method of finding out the solution. 5+3

3. (a) Distinguish between curtate expectation e_x and complete expectation of life e_x^0 at age x and find approximate relation between them.
(b) Write down the probability generating function of Life table values $l_1, l_2, ..., l_w$ given the cohort l_0. Discuss its application. 2+6

4. How one can adjust age data? Discuss the uses of (i) Whipple's index, (ii) Meyer's blended index and (iii) Age dependency ratio. 2+2+2+2

5. What is social mobility? Explain the special situations of 'perfect mobility' and 'perfect immobility'. Discuss some possible measures of social mobility when the socio-economic categories can be ordered in some sense. 2+2+4

6. (a) Give the Logistic equation for population growth and interpret the parameters involved.
(b) Describe a method for fitting this equation to observed population data. 5+3

7. Differentiate between the Gross Reproduction Rate (GRR) and the Net Reproduction Rate (NRR), and interpret the situations when for a society (i) GRR=NRR and (ii) NRR=1. 5+3

8. Discuss the development of Makeham's graduation formula. Outline a method for estimating its parameters using all the available observations. 4+4

M.Sc. Examination 2020
Semester-IV
Statistics
Course: MSC-42 (Survival Analysis)

Time: 3 hours **Full Marks:40**

Questions are of values as indicated in the margin
Answer **any four** questions

1. Define hazard function, cumulative hazard function, survival function, median survival time and establish their inter-relationships whenever exists. For a two-parameter Weibull distribution, find the expressions of all these functions. 4+6
2. Describe life-table estimator of survival function. How does it differ from Kaplan-Meier (KM) estimator? Show that KM estimator is an unbiased estimator of the survival function. Find an expression of the variance of the estimator. 2+2+3+3
3. Define Nelson-Aalen estimator for the cumulative hazard function
$$\Lambda(t) = \int_0^t \lambda(u)du$$
Show that the estimator is approximately unbiased. Find out the variance of the estimator. Also find out an unbiased estimator of$Var(\widehat{\Lambda}(t))$. 2+2+3+3
4. A. Describe in details how to construct likelihood for right censored data. How will you modify the likelihood, if data is left censored?
B. Suppose the underlying survival time T is from an exponential distribution with parameter λ and we have right censored data. Find the $100(1-\alpha)\%$ confidence interval forλ. (4+2)+4
5. Suppose we want to compare more than two survival functions. Define weighted log rank test statistic in this regard. How does it differ from log rank test? Find mean and variance of the statistic. Also derive the asymptotic distribution of the statistic. 2+1+4+3
6. A. What is a proportional hazard model? Why is it called proportional? Why is it considered as semi parametric model?
B. What is partial likelihood? Describe the inferential procedure to estimate parameters involved in a Cox proportional hazard model using partial likelihood. (2+1+1)+(1+5)

M.Sc Semester IV Examination, 2020

Statistics

MSC-43(Practical)

Time: Four Hours   Full Marks: 40

One may use Computer Laboratory, if necessary.

1. The following table gives the population of Philippines by single years of age for the range 10 to 89 for the year 1960. Calculate preference indices for terminal digits by Myers Blended method. Hence compute a summary index of age preference. 10

Age(in years)	Number	Age(in years)	Number	Age(in years)	Number
10	841356	37	242462	64	34381
11	581400	38	316210	65	102440
12	796786	39	225207	66	26445
13	619293	40	434156	67	35311
14	596592	41	128632	68	40711
15	565714	42	217881	69	20921
16	566942	43	169167	70	136771
17	538891	44	151142	71	13000
18	651318	45	319118	72	28017
19	491441	46	160329	73	16662
20	565801	47	160855	74	14490
21	494895	48	237287	75	50558
22	515823	49	155094	76	15010
23	456892	50	313636	77	11878
24	425212	51	78534	78	23353
25	522203	52	128935	79	9212
26	358549	53	93279	80	73791
27	376221	54	95715	81	5532
28	395766	55	163093	82	9331
29	300610	56	87754	83	5653
30	535924	57	71828	84	5089
31	333086	58	93049	85	18604
32	318481	59	72206	86	4803
33	246260	60	275436	87	5617
34	233700	61	31299	88	4388
35	401936	62	49634	89	4000
36	242659	63	40154		

2. Given the following one generation transition probability matrix, calculated two different measures of mobility in a society with an initial distribution =(0.4,0.3,0.1,0.2) among four social classes

0.24	0.26	0.26	0.24
0.10	0.50	0.10	0.30
0.10	0.15	0.70	0.05
0.10	0.25	0.25	0.40

Interpret the results obtained. 5

[Turn Over]

3. Supply the missing entries in the following table: 5

x	l_x	d_x	$1000q_x$	L_x	T_x	e^0_x
10	86104	_	0.72	_	_	_
11	_	_	0.68	_	_	_
12	_	_	0.74	_	22, 63, 125	_

4. Consider a study involving 50 participants. The data are shown below.

Interval in Years	**Number Alive at Beginning of Interval**	**Number of Deaths During Interval**	**Number Censored**
0-3	50	2	5
4-7	43	1	2
8-11	40	1	4
12-15	35	1	3
16-19	31	1	4

Find the life table estimate of the survival function. 5

5. Consider the following data to compare two combination treatments in patients with advanced lung cancer. Twenty participants withlung cancer are randomly assigned to receive chemotherapy before surgery or chemotherapy after surgery. The primary outcome is death and participants are followed for up to 48 months (4 years) following enrollment into the trial. The experiences of participants in each arm of the trial are shown below.

Chemotherapy Before Surgery			**Chemotherapy After Surgery**	
Month of Death	Month of Last Contact (Censored)		Month of Death	Month of Last Contact (Censored)
8	8		33	48
12	32		28	48
26	20		41	25
14	40			37
21				48
27				25
				43

(a) Find Kaplan-Meier estimators for both the treatment arms and plot them together.

(b) Perform log-rank test to test the null hypothesis

H_0: *Two survival curves are same vs.* H_1: *Two survival curves are different*

4+6

6. Fit a Cox proportional hazard model for the following data:

Time: 9 13 13 18 23 28 31 34 45 48 161 5 5 8 8 12 16 23 27 30 33 43 45

Status: 1 1 0 1 1 0 1 1 0 1 0 1 1 1 1 1 0 1 1 1 1 1 1

X: M MMMMMMMMMM NM NMNMNMNMNMNMNMNMNMNMNM(M: Maintained; NM: Non-maintained)

Comment on the result. 5

****