4-Year Undergraduate Program (with Statistics Major) Examination 2025
Semester-IV
Statistics
Course: MJST08 (Statistical Inference I)
Full Marks: 80 **Time: 3 Hours**

Group A

(Answer any seven questions. Each question carries 5 marks.)

1. Give an example in each case to show that an unbiased estimator of a parametric function
   (a) may be worthless
   (b) may not exist. 2.5+2.5

2. State a sufficient condition for consistency of an estimator. Is the condition necessary? Justify with examples. 2+1+2

3. Prove that in sampling of size n from a Cauchy population with location parameter θ and scale parameter 1, the sample mean is not a consistent estimator of θ but the sample median is consistent for the same. 2.5+2.5

4. Explain the notion of sufficient statistic and minimal sufficient statistic with examples. 5

5. State and prove Neyman-Fisher's factorization theorem for the case of discrete variables. 5

6. Define an ancillary statistic. Let X_1, X_2 be a random sample of size 2 from the exponential distribution with p.d.f.
$$f(x,\theta) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}} & \text{if } 0 < x < \infty \\ 0 & \text{otherwise.} \end{cases}$$
Prove that $T = \dfrac{X_1}{X_1 + X_2}$ is ancillary. 5

7. Let X_1, X_2, X_3, X_4 be a random sample from a continuous distribution with p.d.f.
$$f(x,\theta) = \begin{cases} 2\theta^2 x^{-3} & \text{if } 0 < x < \infty \\ 0 & \text{otherwise,} \end{cases}$$
where $\theta > 0$ is an unknown parameter. If the observed values of $x_1 = 15, x_2 = 11, x_3 = 10, x_4 = 17$, find the UMVUE of θ^2. 5

8. Consider two estimators T_1 and T_2 with efficiency e_1, e_2 respectively. Then prove that
$$\sqrt{e_1 e_2} - \sqrt{(1-e_1)(1-e_2)} \leq \rho \leq \sqrt{e_1 e_2} + \sqrt{(1-e_1)(1-e_2)},$$
where ρ denotes the correlation coefficient between T_1 and T_2. 5

9. Let $X_1, X_2, \ldots X_n$ be a random sample from a population with pdf

$$f(x,\theta) = \begin{cases} \frac{1}{2}e^{-\left(\frac{x-2\mu}{\mu}\right)} & \text{if } x > 2\mu \\ 0 & \text{otherwise.} \end{cases}$$

where $-\infty < \mu < \infty$. For estimating μ, consider the estimators

$$T_1 = \frac{\overline{X} - 2}{2}; T_2 = \frac{nX_{(n)} - 2}{2n}$$

Justify whether T_1 and T_2 are consistent or not. 5

10. A random variable X takes the values $0, 1, 2$ with respective probabilities $\frac{\theta}{4N} + \frac{1}{2}\left(1 - \frac{\theta}{N}\right), \frac{\theta}{2N} + \frac{\alpha}{2}\left(1 - \frac{\theta}{N}\right)$ and $\frac{\theta}{4N} + \frac{(1-\alpha)}{2}\left(1 - \frac{\theta}{N}\right)$, where N is a known number and α, θ are unknown parameters. If 75 independent observations on X yielded the values $0, 1, 2$ with frequencies $27, 38, 10$ respectively, estimate θ and α by the method of moments. 5

Group B

(Answer any three questions. Each question carries 15 marks.)

11. Let $X_1, X_2, \ldots X_n$ be a random sample of size n from $U(0, \theta)$ distribution ($\theta > 0$). Find

    (a) an unbiased estimator of θ.

    (b) a consistent estimator of θ.

    (c) a statistic sufficient for θ.

    (d) a statistic complete for θ.

    (e) an estimator of the form $\alpha X_{(n)}$ having the smallest MSE ($X_{(n)}$ being the largest order statistic). 3+3+3+3+3

12. (a) Define uniformly minimum variance unbiased estimator (UMVUE).

    (b) Prove that T^* is UMVUE of $\psi(\theta)$ iff it has zero correlation with all unbiased estimators of zero, for all values of θ in the parametric space.

    (c) Let $X_1, X_2, \ldots X_n$ be a random sample of size n from $Poisson(\lambda)$ distribution. Find the UMVUE of $e^{-\lambda}$. 3+6+6

13. (a) State and prove the Cramer-Rao inequality regarding the lower bound of the variance of an unbiased estimator of a parametric function

    (b) Discuss the case of equality in this inequality.

    (c) Give an example where the Cramer-Rao lower bound is not attained by the variance of an UMVUE. 7+4+4

14. (a) Let $X_1, X_2, \ldots X_n$ be a random sample of size n from $N(\theta, 1)$ distribution, but the values of $X_1, X_2, \ldots X_n$ are unknown. Instead, we only know the number of positive observations. Find the maximum likelihood estimator (MLE) of θ.

(b) Give an example in each case to show that

i. MLE may not be unique

ii. MLE may not be unbiased.

(c) Discuss a situation where the method of maximum likelihood and the method of moments agree in estimating the parameters. 4+(3+3)+5

15. (a) Find the most general form of the distribution for which the sample geometric mean is the MLE of the underlying parameter.

(b) What do you mean by minimum chi-square method of estimation?

(c) Prove that the minimum chi-square estimators are consistent and unique. 5+5+5

– –

4 Year UG Programme (with Statistics Major) Examination 2025
Semester IV
Subject: Statistics
Paper: [MJST09A] Time Series Analysis (Theory)

Full Marks: 60 Time: 3 Hrs.

Group – A (Answer any five questions) $5 \times 2 = 10$

1. Answer the following questions with proper justification.
   (a) Why do values in a time series fluctuate?
   (b) What are the key methods and criteria for choosing an appropriate trend function in time series analysis?
   (c) What is meant by de-seasonalisation? Explain with an example.
   (d) Discuss irregular (random) variations in time series. Why can't they be predicted? How should they be handled in analysis?
   (e) What do you mean by mixed models in a time series? Discuss with examples.
   (f) Why is the multiplicative model more commonly used than the additive model in time series analysis?
   (g) Why we use the link relative method?
   (h) What are the main drawbacks of using the least squares method for curve fitting in time series analysis?

Group – B (Answer any four questions) $5 \times 4 = 20$

2. What are the common methods used to measure trends in time series data? Provide a brief explanation of one such method. 2+3

3. Let $\varepsilon_t, \varepsilon_{t+1}, \ldots$ and ξ be independent random variables with zero mean and unit variance. Given:

$$y_t = a\xi + \varepsilon_t, \quad -\infty < t < \infty$$

   Show that the process is stationary with correlation: $\rho_1 = \rho_2 = \cdots = \frac{a^2}{1+a^2}$ 5

4. Describe the process of fitting a second-degree (parabolic) trend to time series data. What is the general form of the model, and how are the parameters estimated? 5

5. Describe the residual method to measure the cyclic variations of a time series. 5

6. Briefly discuss the method of iterated averages and its usage in time series. 5

7. What do you mean by a correlogram? Briefly discuss its use this in the context of time series. 2+3

Group – C (Answer any three questions) $10 \times 3 = 30$

8. Describe the Modified Exponential, Gompertz, and Logistic growth models in detail. Additionally, explain the method of partial sums used for fitting the Modified Exponential growth curve. 2+2+2+4

9. Explain what is meant by seasonal fluctuations of a time series. Discuss two different methods for determining seasonal fluctuations. Also discuss their merits and demerits. 2+4+4

10. Derive the correlogram (autocorrelation function) of a first-order autoregressive series given by:

$$y_{t+1} = \alpha y_t + \varepsilon_t, \quad |\alpha| < 1$$

Explain why its autocorrelation function exhibits a geometric decay. 8+2

11. Write short notes on: i) Harmonic Analysis ii) General Auto-regression Series. 5+5

4 Year UG Programme (with Statistics Major) Examination 2025
Semester IV
Subject: Statistics
Paper: [MJST09B] Time Series Analysis (Practical)

Full Marks: 20 Time: 2 Hrs.

Answer all questions (*Symbols have their usual meaning*)

1. Given the population figures of India as follows:
   Fit an exponential trend $y = ab^x$ to the above data by the method of least squares and find the trend

Year	1911	1921	1931	1941	1951	1961	1971
Population (in Cr.)	25	25.1	27.9	31.9	36.1	43.9	54.7

   values. Estimate the population in 1981, 2001 and 2011. 10

2. Using ratio to trend method, determine the quarterly seasonal indices for the following data. Also

Year	Q(I)	Q(II)	Q(III)	Q(IV)
1995	30	40	36	34
1996	34	52	50	44
1997	40	58	54	48
1998	54	76	68	62
1999	80	92	86	82

   interpret your result. 7

3. Practical note book & viva-voce. 3

4-Year Undergraduate Program (With Statistics Major) Examination, 2025

Semester- IV

Course: MJST10T

(Demography and Vital Statistics)

Time: 3 hours **Full Marks: 60**

Questions are of values as indicated in the margin

Answer **any four** questions

1. i. Define a measurement of mortality which is a probability rate. Mention its advantages and disadvantages.
   ii. What is standardized death rate. How would you modify this death rate if, for each age group, number of persons and specific death rates are not available? How this rate can be used to compare mortality situation between two regions? 5+6+4

2. i. What is infant mortality rate? How this rate is computed? What are the disadvantages of this rate? Discuss the importance of this rate keeping in mind the health situation of a region.
   ii. Distinguish between Stable population and stationary population. (2+3+2+4)+4

3. What is a life table? Discuss the interrelationship of al the functions present in a life table. Write down the process of construction of a life table starting from the age specific mortality rates m_x. 2+8+5

4. i. Derive the logistic function to be used for population projection. Discuss various properties of this function.
   ii. Describe any two methods to fit such a Logistic model. 5+10

5. i. Discuss the following measures of fertility:
   General fertility rate, age specific fertility, total fertility rate
   Also mention limitations of these measures.
   ii. Define net reproduction rate (NRR) and gross reproduction rate (GRR). Discuss the cases when NRR$>$ 1, NRR <1 and NRR=1.
   iii. Show that for any community the NRR is necessarily less than GRR.
   5+5+5

6. i. Derive Makeham's graduation formula.
   ii. Discuss method of four points to fit this formula. How this can method be improved using all available observations 5+(6+4)

4-Year Undergraduate Program (With Statistics Major) Examination, 2025

Semester- IV

Course: MJST10P

(Demography and Vital Statistics)

Time: 2 hours **Full Marks: 20**

Questions are of values as indicated in the margin

1. Compute age-specific death rates for the following data set

Age Group	Population	Deaths
0-4	500,000	2,500
5-9	480,000	800
10-14	470,000	600
15-19	460,000	900
20-24	450,000	1,200
25-29	440,000	1,500
30-34	430,000	2,000
35-39	420,000	2,500
40-44	410,000	3,200
45-49	400,000	4,000

4

2. Consider the population data of Italy for last 10 years (in 1000):

Year:	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Population:	60639	60512	60445	60336	60219	60041	59784	59674	59564	59435

Fit a logistic model using Rhodes's method for this data. What will be the approximated population in 2030? 7

3. Complete the following life table

Age(x)	l_x	d_x	q_x	L_x	T_x	e_x
0	100,000	5,000	0.050	97500	?	?
1	95,000	1900	0.020	94050	?	?
2	93200	?	?	92200	88058	?
3	?	1600	?	?	?	?
4	89000	?	?	?	?	?
5	?	3200	?	?	?	?
6	79000	?	?	?	?	?
7	?	?	?	?	?	?
8	58000	?	?	?	?	?
9	?	?	?	?	?	?

6

(symbols having their usual meaning)

4. Practical note book and viva-voce 3

B.Sc. (Honours) Examination 2024
Semester-IV
Statistics
Course: CC-8A (Survey Sampling and Indian Official Statistics (Theory))
Full Marks: 40 **Time: 3 Hours**

(Answer any four questions.)

1. (a) Explain different sources of sampling errors in sample survey.
   (b) What do you mean by a random number series? Mention the advantages in using them.
   (c) Define judgement sampling with an example. 4+4+2

2. (a) Show that under simple random sampling without replacement, the sample proportion p of individuals belonging to a particular characteristic A, is an unbiased estimator of the corresponding population proportion P.
   (b) Find $Var(p)$ and an unbiased estimate of it.
   (c) Suppose there is a population $U = (U_1, U_2, \ldots U_N)$, with unknown variate values $Y_i (i = 1(1)N)$. In order to estimate the population total from a SRSWOR of size n, we use the estimator $t = N\overline{y}$. Now suppose that we have the advance information that the value of Y for the last unit is known and equals Y_N. Consider another estimator $t' = Y_N + (N-1)\overline{y}'$ based on a SRSWOR of size n from the remaining $(N-1)$ units, $\overline{y}'$ being the sample mean. Prove that $Var(t') < Var(t)$. Interpret your result. 2+3+5

3. (a) Mention the problem in estimating the variance of an unbiased estimator of population mean in linear systematic sampling. Explain the role of interpenetrating subsample in this context.
   (b) Show that in the presence of a linear trend in the population, the linear systematic sampling provides a more efficient estimator of the population total than the simple random sampling. 5+5

4. (a) If there are two strata and if ϕ is the ratio of actual $\frac{n_1}{n_2}$ value to the $\frac{n_1}{n_2}$ value obtained by Neyman's allocation, show that whatever be the values of S_1 and S_2, the ratio $\frac{V_{opt}}{V}$ is never less than $\frac{4\phi}{(1+\phi)^2}$, when N_1, N_2 are large. Here V_{opt} is the variance of the usual estimator $\widehat{Y_{st}}$ under Neyman's allocation and V is the variance of $\widehat{Y_{st}}$ under actual allocation. (Notations have their usual meanings)
   (b) Find the optimal sizes of the first and the second stage samples in two stage sampling so that the variance of the unbiased estimator of population mean is minimized subject to a given cost. Assume there are the same number of SSU s within each FSU. 5+5

(P.T.O.)

5. (a) Define ratio estimator for estimating the population mean. Show that it is an biased estimator.

   (b) Define Hartley-Ross' unbiased ratio-type estimator and show that it is unbiased for the population mean. 6+4

6. (a) Who is the head of NSSO? How many divisions does it have? Write down their names. Mention one of its important publications.

   (b) In which year DGCIS wa set up? Mention the area in which they are presently responsible for. Also write one of its important publications.

   (c) Which division under CSO is responsible for the estimation of GDP? Mention one publication related to national income. 4+3+2

B.Sc. (Honours) Examination 2024
Semester-IV
Statistics (Practical)
Course: CC-8B (Survey Sampling and Indian Official System (Practical))
Full Marks: 20 Time: 2 Hours

(1) Draw a random sample of size 10 from a Binomial distribution with parameters $n = 10, p = 0.76$, using a random number table. 5

(2) The following data relate to the number of yearly enrolment in Teachers' training colleges stratified into 3 strata. Obtain the gain in precision due to stratification for estimating the average yearly enrolment per college. 5

Stratum no. (h)	N_h	n_h	$\overline{x_h}$	s_h^2
1	13	9	32.2	2.625
2	18	7	41.638	5.063
3	26	10	19.992	3.549

(3) The following is the population of B.Sc. (Statistics) students of a certain university, together with the data on two variables, viz. Marks obtained in Sample survey course and Marks in Estimation course (both out of 50).

Serial Number	Student Roll no.	Marks in Estimation	Marks in Sample Survey
1	BSC-STAT-01	31	28
2	BSC -STAT-02	26	29
3	BSC -STAT-03	40	42
4	BSC -STAT-04	40	37
5	BSC -STAT-05	36	44
6	BSC -STAT-06	45	45
7	BSC -STAT-07	29	22
8	BSC -STAT-08	21	19
9	BSC -STAT-09	25	21
10	BSC -STAT-10	30	34
11	BSC -STAT-11	39	33
12	BSC -STAT-12	42	43
13	BSC -STAT-13	34	38
14	BSC -STAT-14	30	24
15	BSC -STAT-15	26	28
16	BSC -STAT-16	28	33
17	BSC -STAT-17	45	40
18	BSC -STAT-18	39	43
19	BSC -STAT-19	35	26
20	BSC -STAT-20	41	34

(a) Select 5 students at random from the list.
(b) Estimate the average marks in sample survey by ratio method, using the marks in estimation as auxiliary information, based on your sample. Also find the estimated standard error of your estimator. 2+5

(4) Practical Note Book and Viva-Voce. 3

B.Sc. (Hons.) Examination 2024
Semester IV
Subject: Statistics
Paper: CC-9 [Linear Models (Theory)]

Full Marks: 40 Time: 3 Hrs.

Answer any four questions (*Symbols have their usual meaning*)

1. a) If a linear parametric function $q'\beta$ is estimable then show that $q'\hat{\beta}$ is the best linear unbiased estimator of $q'\beta$. In usual notations, here $\hat{\beta} = (X'X)^{-}X'y$.
   b) It is given that y_1, y_2, y_3 are independent with same variance and $E(y_1) = \beta_1 + \beta_2 = E(y_3)$, $E(y_2) = \beta_1 + \beta_3$. Then show that $\sum_{i=1}^{3} p_i\beta_i$ is estimable if and only if $p_1 = p_2 + p_3$. 5+5

2. a) Briefly discuss influential point with an example.
   b) Let $q'\beta$ be an estimable function of β in the model $y = X\beta + \epsilon$ with $Var(y) = \sigma^2 I$. Let $\hat{\beta}$ be any solution to the normal equations $(X'X)\hat{\beta} = X'y$ and let r be any solution to $(X'X)r = q$. Then show that: i) $Var(r'X'y) = \sigma^2 r'q$, ii) $Var(q'\hat{\beta}) = \sigma^2 q'(X'X)^{-}q$, iii) $Var(q'\hat{\beta})$ is unique, that is, invariant to the choice of r or, $(X'X)^{-}$. 2+(2+3+3)

3. Write down detailed analysis, assumption and hypothesis testing of a one way ANOVA fixed effects model. 10

4. a) Consider the model $y_{ij} = \mu + \tau_i + \epsilon_{ij},\ i = 1, 2;\ j = 1, 2, 3$. Then find X matrix, $\hat{\beta}$.
   b) Consider a multiple linear regression model of the form $y = X\beta + \epsilon$ with the assumption $\epsilon \sim N_n(0, \sigma^2 I)$. Derive the maximum likelihood estimators of β and σ^2. 5+5

5. a) In usual notations show that $H, (I - H)$ are both symmetric and idempotent.
   b) What do you mean by multicollinearity? What are the primary sources? How to deal with multicollinearity? (2+2)+(2+2+2)

6. a) In which situation we use ANOVA and ANCOVA? Explain in detail with an example.
   b) Write short notes on: Standardized Residuals, R Student, Projection Matrix. 4+(2+2+2)

B.Sc. (Hons.) Examination 2024
Semester IV
Subject: Statistics
Paper: CC-9B [Practical on Linear Models]

Full Marks: 20 Time: 2 Hrs.

Answer all questions

1. Find first four Legendre orthogonal polynomials. 4

2. Three sides (S_1, S_2, S_3) of an equilateral triangle were measured by five pupils $(P_1, P_2, P_3, P_4, P_5)$ with the following results: 8

	P_1	P_2	P_3	P_4	P_5
S_1	5.44	5.41	5.43	5.42	5.43
S_2	5.43	5.41	5.42	5.43	5.44
S_3	5.45	5.43	5.43	5.43	5.44

Analyze the data to see if there is any significant difference between the measurement of the pupils or, between the sides of the triangle.

3. For the ANOVA model: $y_{ij} = \mu + \tau_i + \varepsilon_{ij},\ i, j = 1, 2, 3$. Find the rank of $(X'X)$. 5

4. Practical note book and viva-voce. 3

B.Sc (Honours) Examination, 2024

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. Distinguish between:
   i) Process control and Product Control
   ii) Control limits and Specification limits

   B. Describe, in detail, the construction of control chart for range. How would you modify the control limits if the process is found to be out of control? 4+6

2. Describe the construction of chart for mean assuming unknown variance. Discuss both the control charts by estimating the unknown variance by sample standard deviation and sample range. 10

3. A. When a system is said to be out of control? If all observations are within the control limits can a system still be out of control?

   B. How would you modify a control chart if sample sizes are varying? Give an example. 4+6

4. A. What is Shewart's control chart? Discuss the choice of 3-σ limit in this context.

   B. Describe the construction of control chart for number of defectives. 5+5

5. A. Define the following

   LTPD, AOQ, AOQ, ASN and ATI

   B. For a double sampling inspection plan, find expressions for all the terms mentioned in (A) assuming a hypergeometric distribution. 3+7

6. A. Define Consumer's risk and producer's risk.

   B. Assuming a binomial distribution find the expression for the producer's risk, consumer's risk, AOQ and ASN functions for a single sampling inspection plan. 3+7

BSC Semester IV Examination, 2024

Statistics

Paper: [CC-10 B] 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

6

2. Following are the figures for the number of defectives in 16 lots, each containing 2,000 rubber belts:
   (556-your roll number), 225, 322, 280, 306, 337, 305, 356, 402, 216, 264, 126, 409, 193, 326, (280+your roll number)
   Drawing the control chart for fraction defective, plot the points on it. Comment on the state of control of the process. 6

3. From each lot of 5000 chalks, take 200 at random and inspect them. Accept the lot if the inspected sample contains at most two defective chalks; otherwise reject it. Suppose a lot has 30 defective items. What is the acceptance probability for such a lot? Plot the OC curve for this sampling inspection plan. 5

4. Practical note book and Viva-voce 3

B.Sc. (Hons.) Examination 2024
Semester IV
Subject: Statistics
Paper: [SECC-2] Statistical Data Analysis Using Software Packages

Full Marks: 25 Time: 2 Hrs.

$\underline{Group - A}$ $5 \times 2 = 10$

$\underline{Group - A\ (Answer\ any\ ten\ questions)}$ $10 \times 1 = 10$

1. Answer the following questions with proper justification.
   (a) In python, how do you find $exp(3) + tan^{-1}(10)$?
   (b) Which function do we use to find bar diagram plot in python?
   (c) Round off π upto 4 decimal point.
   (d) Which python library enable us to use dataframe?
   (e) Create a list of three objects. It input six positive numbers and output median, variance and coefficient of variation as elements of the list.
   (f) Create a user defined function that input x, y from user and generates the value $xy^2 + \cos(x^2 + e^y)$.
   (g) Write down the python code to find the raw moments upto order 4 for a data of length 8 of your own choice.
   (h) Write down the algorithm to check for a prime number.
   (i) Generate a sqeuence from 50 to 1 with increment -0.2.
   (j) Write down two different plot style in python.
   (k) Write down a measure of kurtosis. Create a hypothetical data of length 10 and compute your mentioned measure using python.
   (l) Write down the python code to draw the graph of $y = \tan(x^2 + \sin(x))$ for $x \in [-5, 5]$.
   (m) How to install a package 'PPPP' in python?

$\underline{Group - B\ (Answer\ any\ three\ questions)}$ $3 \times 5 = 15$

2. Write down the python code to find the range, variance, mean deviation about median, inter-quartile range and coefficient of variation for the data (x): 26 25 32 63 33 68 72 87 36 39 54 74 24 25 33 57 44 78 63 75 45 35. 5

3. Suppose we have a bivariate data of 17 students, Marks in School (x): 146 135 159 136 161 166 165 160 136 176 145 144 153 158 145 126 170 and Marks in College (y): 369 333 417 301 425 372 438 415 393 349 306 380 338 450 326 381 359. Write down the python code to plot the scatter diagram and correlation matrix. Also write down the code for linear regression of y on x and x on y. 5

4. Write down the python code to plot the multiple graphs from $N(2, \sigma = 2)$, $N(2, \sigma = 4)$ and $N(4, \sigma = 4)$ distributions. 5

5. Suppose there are 150 people numbered from 1 to 150. Use circular systematic sampling to select 21 lucky people for prize money. Explain the process briefly together with the python code. 5

6. Generate 500 random samples from χ^2_3 and $Poisson(3)$. Also draw the respective histograms. 5

B.Sc. (Honours) Examination 2023
Semester-IV
Statistics
Course: CC-8A (Survey Sampling and Indian Official System (Theory))
Full Marks: 40 **Time: 3 Hours**

(Question 1 is compulsory. Answer any three questions from the rest.)

1. Answer the following questions. 2×5

   (a) How many divisions does the CSO have? Mention any two of them.

   (b) What are the two wings of MoSPI?

   (c) Mention two important publications relating to the industrial statistics.

   (d) Define index of industrial production.

   (e) Write down the full forms of DGCIS and DAMI.

2. (a) Describe the basic principles of sample survey.

   (b) Explain the advantages of sample survey over complete enumeration.

   4+6

3. Consider a simple random sample of size n drawn without replacement from a population of size N.

   (a) Find the first and second order inclusion probabilities.

   (b) Let f_α denote the frequency of the αth member of the population in the sample $(1, 2, \ldots N)$. Find

      i. $E(f_\alpha)$
      ii. $Var(f_\alpha)$
      iii. $Cov(f_\alpha, f_\beta), \alpha \neq \beta$.

   (c) Hence, or otherwise, find the variance of the sample mean.

   3+3+4

4. (a) If there are two strata and if ϕ is the ratio of actual $\frac{n_1}{n_2}$ value to the $\frac{n_1}{n_2}$ value obtained by Neyman's allocation, show that whatever be the values of S_1 and S_2, the ratio $\frac{V_{opt}}{V}$ is never less than $\frac{4\phi}{(1+\phi)^2}$, when N_1, N_2 are large. Here V_{opt} is the variance of the usual estimator $\widehat{Y_{st}}$ under Neyman's allocation and V is the variance of $\widehat{Y_{st}}$ under actual allocation. (Notations have their usual meanings)

   (b) Find the expression of the estimated gain due to stratification over simple random sampling.

   5+5

5. (a) Why is the systematic sampling called a mixed sampling scheme? Suggest an unbiased estimator of the population total under this sampling scheme and find its variance.

   (b) What is the problem in estimating the variance of the unbiased estimator of the population total? Explain how can you overcome the problem.

(2+3)+(2+3)

6. (a) Explain how the double sampling technique can be implemented in ratio method of estimation of the population mean. Find approximate expressions of the bias and MSE of the corresponding estimator.

   (b) Mathematically explain why cluster sampling is usually less efficient than uni-stage sampling.

(2+5)+3

B.Sc. (Honours) Examination, 2023
Semester-IV
Statistics (Practical)
Course: CC-8B (Survey Sampling and Indian Official System (Practical))
Full Marks: 20 Time: 2 Hours

(1) Draw a random sample of size 10 from Binomial $(n = 15, p = 0.67)$ distribution. (5)

(2) The following data relate to the number of yearly enrolment in Teacher's Training College stratified into 3 strata. Obtain the gain in precision due to stratification for estimating the average yearly enrolment per college. (5)

Stratum No. (i)	N_i	n_i	$\bar{y}_i$	s_i^2
1	13	9	32.2	2.625
2	18	7	41.368	5.063
3	26	10	19.992	3.549

(3) A random sample of 185 girls from 9248 girls in an area showed that 19 had some nutritional deficiency. Estimate the proportion of nutritionally deficient girls and the standard error of your estimate. (3)

(4) The following figures relate to a study of a variable Y (in kg.) together with an auxiliary variable X(in ft.):

Population Size $(N) = 12908, \bar{Y} = 782.5, \bar{X} = 782.5, S_y^2 = 45.387, S_x^2 = 39.228, S_{xy} = 36.116$
First-phase sample size$(m) = 1528$, sample mean $(\overline{x_m}) = 85.7$
Second-phase sample size$(n) = 100$, sample means $(\overline{x_n} = 86.99, \overline{y_n} = 769.68)$

Find an estimate of the population mean of Y by ratio method and the variance of the estimator. (4)

(5) Practical Note Book and Viva-voce. (3)

B.Sc. (Hons.) Examination 2023
Semester IV
Subject: Statistics
Paper: CC-9 [Linear Models (Theory)]

Full Marks: 40 Time: 3 Hrs.

Answer any four questions (*Symbols have their usual meaning*)

1. a) Show that a linear parametric function $q'\beta$ is estimable if and only if $q' = q'H$ with $H = (X'X)^{-}X'X$.
   b) It is given that y_1, y_2, y_3, y_4 are independent with same variance and $E(y_1) = \beta_1 - \beta_2 + \beta_3$, $E(y_2) = \beta_1 + \beta_2 - \beta_4$, $E(y_3) = \beta_1 - \beta_2 + \beta_4$ and $E(y_4) = \beta_2 + \beta_3 + \beta_4$. Now check whether all parametric functions are estimable or, not. 4+6

2. a) What do you mean by the null space of a matrix A? Is it same to the Nullity? What is the relation between rank and nullity of A?
   b) Show that $[X'_{(i)}X_{(i)}]^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}x_i x_i'(X'X)^{-1}}{1-h_{ii}}$. Here the matrix $[X'_{(i)}X_{(i)}]$ is obtained from $[X'X]$ by removing i^{th} row of X, x_i' is the i^{th} row of X and $h_{ii} = x_i'(X'X)^{-1}x_i$ is the i^{th} diagonal element of hat matrix, $X_{(i)}$ is the matrix obtained from X by removing i^{th} row. (1+1+1)+7

3. Write down detailed analysis, assumption and hypothesis testing of a two way ANOVA with one observation per cell. 10

4. a) Briefly write down some applications of regression analysis.
   b) What do you mean by regression model through origin? What is the OLS estimator of its parameter? Is it unbiased? Find variance of this estimator. Also find an estimator of σ^2. 3+7

5. a) In usual notations for simple linear regression model: $y = \beta_0 + \beta_1 x + \varepsilon$, show that the OLS estimators of β_0, β_1 are BLUE. Also show that $\sum_{i=1}^{n} \hat{y}_i e_i = 0$.
   b) In usual notations for multiple linear regression, show that $V(\widehat{y}_{n\times 1}) = \sigma^2 H$. (5+2)+3

6. a) Write down the basic differences among Regression, ANOVA and ANCOVA. Cite one example in each category.
   b) Write short notes on: Partial regression plot, Normal probability plots. 5+(2.5+2.5)

B.Sc. (Hons.) Examination 2023
Semester IV
Subject: Statistics
Paper: CC-9B [Practical on Linear Models]

Full Marks: 20 Time: 2 Hrs.

Answer all questions

1. Fit a simple linear regression model on the following data and discuss the lac of fit in detail. 10

x	1	1	2	3.3	3.3	4	4	4	4.7
y	10.84	9.3	16.35	22.88	24.35	24.56	25.86	29.16	24.59
x	5	5.6	5.6	5.6	6	6	6.5	6.9	-
y	22.25	25.9	27.2	25.61	25.45	26.56	21.03	21.46	-

2. Four specified persons (P_1, P_2, P_3, P_4) determine the moisture content of samples of a powder. Each person taking a sample from each of six given consignments $(C_1, C_2, C_3, C_4, C_5, C_6)$. The assessments are given below: 7

	C_1	C_2	C_3	C_4	C_5	C_6
P_1	9	10	9	10	11	11
P_2	12	11	9	11	10	10
P_3	11	10	10	12	11	10
P_4	12	13	11	14	12	10

Analyze the data to see if there is any significant differene between the consignment or, between the person.

3. Practical note book and viva-voce. 3

B.Sc (Honours) Examination, 2023

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:

   3σ –limits, control charts, UCL, LCL, Upper Specification Limit, Lower Specification Limit

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

2. A. When a system is said to be out of control?

   B. How would you modify a control chart if sample sizes are varying? Give an example. 3+7

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. A. Define the following

   LTPD, AOQ, AOQ, ASN and ATI

   B. For a single sampling inspection plan, find expressions for all the terms mentioned in (A). 3+7

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

BSC Semester IV Examination, 2020

Statistics

Paper: [CC-10B] 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, S) 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. a. Suppose a tyre supplier ships tyres in lots of size 400 to the buyer. A single sampling plan with n = 15 and c = 0 is being used for the lot inspection. The supplier and the buyer's quality control inspector decide that AQL = 0.01 and LTPD = 0.10. Compute the producer's risk and consumer's risk for this single sampling plan.
   b. Suppose the rejected lots are screened and all defective tyres are replaced by non-defective tyres. Construct the AQO curve for this plan. 4+3

4. Practical note book and viva voce. 3

B.Sc. (Hons.) Examination 2023
Semester IV
Subject: Statistics
Paper: [SECC-2] Statistical Data Analysis Using Software Packages

Full Marks: 25 Time: 2 Hrs.

$\underline{Group - A}$ $5 \times 2 = 10$

$\underline{Group - A\ (Answer\ any\ ten\ questions)}$ $10 \times 1 = 10$

1. Answer the following questions with proper justification.
   (a) In python, how do you find $\log(3) + \log_2(3)$?
   (b) Which function do we use to find box plot in Python?
   (c) Round off $\sqrt[3]{5}$ upto 4 decimal point.
   (d) Which python library enable us to use pie diagram?
   (e) Create a list of three objects. It input four positive numbers and output AM, GM and HM as elements of the list.
   (f) Create a user defined function that input x, y from user and generates the value $xy + \sin(x + y)$.
   (g) Input five names and sort them in descending order.
   (h) How to install a package 'PPP' in Python?
   (i) Create a vector x with elements $0, 0.4, 0.8, 1.2, .., 8$.
   (j) Write down two different line style plot in python.
   (k) Write down a measure of skewness. Create a hypothetical data of length 8 and compute your mentioned measure using python.
   (l) Using python draw the graph of $y = \cos(e^x)$ for $x \in [-5, 5]$.
   (m) In python, how can you merge two data frames?

$\underline{Group - B\ (Answer\ any\ three\ questions)}$ $3 \times 5 = 15$

2. Using Python, find the raw and central moments up to order 5 for the data (x): 26 34 38 14 24 37 32 26 11 34 18 38 38 21 20 29 39 13 12 24. 5

3. Fit a regression line based on the bivariate data of 20 students, Marks in School (x): 126 157 153 156 152 135 145 132 153 143 165 132 180 161 170 176 165 163 157 180 and Marks in College (y): 292 367 361 366 343 326 331 307 343 322 363 296 412 377 394 384 370 365 349 403. Fit a regression line Y on X and display the summary statistics. 5

4. Using Python, plot the multiple graphs of Chi-square distributions χ^2_3, χ^2_6 and χ^2_{10}. 5

5. Suppose there are 10 plants named: Areca Palm, English Ivy, Indian Basil, Spider Plant, Snake plant, Weeping Fig, Azalea, Dracaena, Aloe Vera and Small Coconut. Select six of these by using SRSWR and SRSWOR. Explain the process briefly together with the python code. 5

6. Generate 1000 random samples from N(0,1) and U[2,5]. Also draw the respective histograms. 5