M.Sc. Examination, 2022

Semester-I

Statistics

Course: MSC-11

(Linear Models and Distribution Theory)

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. (a). Show that the general solution of the system of homogeneous equations $\boldsymbol{Ax} = \mathbf{0}$ can be expressed as

$$\widetilde{\boldsymbol{x}} = (I - H)\boldsymbol{z}$$

Where $\boldsymbol{z}$ is any arbitrary vector, $\boldsymbol{H} = \boldsymbol{S}^{-}\boldsymbol{S}, \boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X}$

How will you modify this result for the non-homogeneous consistent equations $\boldsymbol{Ax} = \boldsymbol{u}$?

(b) Let $\boldsymbol{l}_1'\boldsymbol{\beta}$ and $\boldsymbol{l}_2'\boldsymbol{\beta}$ are two estimable functions and $\boldsymbol{l}_1'\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{l}_2'\widehat{\boldsymbol{\beta}}$ be their least square estimates respectively. Find $\mathbf{Var}(\boldsymbol{l}_1'\widehat{\boldsymbol{\beta}})$ and $\mathbf{Cov}(\boldsymbol{l}_1'\widehat{\boldsymbol{\beta}}, \boldsymbol{l}_2'\widehat{\boldsymbol{\beta}})$.

6+4

2. (a). Define estimation space and error space. Show that the covariance between any linear function belonging to the error space and any BLUE is zero.

(b). Consider the following linear model:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \qquad i = 1,2; j = 1,2,3$$

Is $\tau_1 - \tau_2$ estimable?

7+3

3. (a) Show that all linear parametric functions, for the linear model $\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{\epsilon}$, in $\boldsymbol{\beta}$ are estimable if and only if X has full rank.

(b). Show that for the liner model $\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{\epsilon}$ a necessary and sufficient condition for a linear parametric function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ to be estimable is that $\boldsymbol{\lambda}'$ is a linear combination of rows of $\boldsymbol{x}'\boldsymbol{x}$.

5+5

4. (a) Show that the conditional minimum of the sum of squares of the residuals $(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})$, in the model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}, \boldsymbol{E}(\boldsymbol{\epsilon}) = 0, V(\boldsymbol{\epsilon}) = \sigma^2 I$, subject to m conditions $\boldsymbol{\Lambda\beta} = \boldsymbol{d}$, where $\boldsymbol{\Lambda\beta}$ are estimable and rank$(\boldsymbol{\Lambda}) = m$, exceeds the error sum of squares (SSE).

(b) Show that $\frac{SSR(\beta)}{\sigma^2}$ follows a non-central χ^2 distribution.

6+4

5. (a) What is general linear hypothesis? When it is called "testable"?

(b) Consider the model $y_i = \theta_i + \epsilon_i, i = 1(1)n$

Where the parameters θ_i are subject to the restriction $\sum_{i=1}^{n} \theta_i = 0$

Write down a test procedure to test the following hypothesis

$H_0: \theta_i = \theta_j \qquad (i \neq j), i, j = 1(1)n$ 3+7

64. (a) State and prove Cochran's theorem on quadratic forms. Give an example where this theorem can be used.

(b) If $\boldsymbol{M} \sim \boldsymbol{W}_p(\boldsymbol{\Sigma}, m)$ and $\boldsymbol{B}$ is a $p \times q$ matrix, then show that

$$\boldsymbol{B}'\boldsymbol{M}\boldsymbol{B} \sim W_q(\boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B}, m)$$ 7+3

7 5. (a) If $\boldsymbol{M} \sim W_p(\boldsymbol{\Sigma}, m), m > p$ then show that the ratio

$\frac{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{M}^{-1}\boldsymbol{a}}$ has the χ^2_{m-p+1}distribution for any fixed p-vector $\boldsymbol{a}$

(b). Prove that the BLUE of any linear combination of estimable parametric function is the linear combination of their BLUEs. 7+3

M.Sc. Examination, 2022
Semester-I
Statistics
Course: MSC-12
(Real Analysis and Measure Theory)

Time: 3 Hours Full Marks: 40

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

NOTE: There are total 6 questions. Answer any 4 questions.

1. (a) Let F be the distribution function on $\Re$ given by:

$$F(x) = \begin{cases} 0, & x < -1 \\ 1+x, & -1 \leq x < 0 \\ 2+x^2, & 0 \leq x < 2 \\ 9, & x \geq 2 \end{cases}$$

Let μ be the Lebesgue-Stieltjes measure corresponding to F. compute the measure of each of the following sets (with explanations).

(i) $\{2\}$

(ii) $\{x : \ |x| + 2x^2 > 1\}$

(iii) $(-1, 0] \cup (1, 2)$

(b) Let $\Omega = \Re^2$. Define A_n as the interior of the circle with radius 1 and center $\left(\frac{(-1)^n}{n}, 0\right)$. Find $\limsup_n A_n$ and $\liminf_n A_n$ with suitable explanations.

[(1 + 2 + 2) + 5 = 10]

2. (a) State Monotone convergence theorem and Dominated convergence theorem. Define Characteristic function. If $X_1, X_2, \cdots, X_n$ are n independent random variables, then show that

$$\phi_{X_1+X_2+\cdots+X_n}(\omega) = \phi_{X_1}(\omega)\phi_{X_2}(\omega)\cdots\phi_{X_n}(\omega).$$

(b) Define open ball and interior point in $\Re^n$. Define open and close sets in n-dimensional Euclidian space $\Re^n$. Prove that, union of finite number of open sets is open. Prove/disprove that the union of infinite number of open sets is always open.

[5 + 5 = 10]

3. (a) Define almost sure convergence, convergence in probability and convergence in distribution. State Borel-Cantelli Lemma. Suppose $f_n \to f$ pointwise. Can we say that $\int f_n d\mu \to \int f d\mu$? Explain with suitable example.

(b) Let μ be a non-negative, finitely additive set function on field F. If $A_1, A_2, \cdots$ are disjoint sets in F and $\bigcup_{n=1}^{\infty} A_n \in F$, show that

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \geq \sum_{n=1}^{\infty} \mu(A_n)$$

[5 + 5 = 10]

CONTINUED

4. (a) State Fatou's lemma, Levy continuity theorem,, and Heine-Borel Theorem.

(b) Determine the radius of convergence and interval of convergence for the following power series:

$$\sum_{n=1}^{\infty} \frac{2^n}{n}(4x-8)^n$$

$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = \frac{\frac{2^{n+1}}{n+1}}{\frac{2^n}{n}}$ $\quad \frac{2^{n+1}}{n+1} \times \frac{n}{2^n}$

2^{n+1-n}

$= \frac{2n}{n+1}$

$= 0$

[5 + 5 = 10]

5. (a) For any characteristic function $\phi(t)$, explain why $|\phi(t)| \leq 1$? Consider the random variable X that has a standard Cauchy distribution. Show that the moment generating function does not exist for this random variable on any real interval with positive length. Does the characteristic function exist for X?

(b) If $X \sim Exponential(\lambda)$, show that

$$\phi_X(\omega) = \frac{\lambda}{\lambda - j\omega}.$$

[5 + 5 = 10]

6. (a) Define adherence point and accumulation point of a set $S \subset \Re^n$. If A is open and B is close, explain why $A - B$ is open.

(b) State Bolzano-Weirstrass theorem and illustrate with a suitable example.

(c) Define Closure of a set. Show that A set S is closed iff $S = \bar{S}$. Define Derived set. Show that, a set is closed iff it contains all its accumulation points.

[3 + 3 + 4 = 10]

$1/R = \lim_{n\to\infty} \left(\frac{2^{n+1}}{n+1} \cdot \frac{n}{2^n} \right)$

$= \lim_{n\to\infty} \left(\frac{2n}{n+1} \right)$

$= \lim_{n\to\infty} \frac{2}{1+1/n} = 2$

converge if $|4x-8| < 1$ $\quad 4x = 9 \quad x = 9/2$

diverge if $|4x-8| > 1$ $\quad 4x = 9 \quad x = 9/2$

at $x = 9/2$

M.Sc. Examination, 2022
Semester-I
Statistics
Course: MSC-13
(Statistical Inference-I)

Time: Three Hours **Full Marks: 40**

Questions are of value as indicated in the margin

Notations have their usual meanings

1. Answer any FOUR questions of the following. 4x5=20

(a) Let $(X_1, X_2, ..., X_n)$ be a random sample from R $(0, \theta)$ distribution with unknown parameter θ. Write $X_{(n)} = \max(X_1, X_2, ..., X_n)$. Then, if $L(\theta)$ denotes the likelihood function, show that, for $\theta > 0$, $L(X_{(n)}) \geq L(\theta)$, and make your comment.

(b) Suppose $X_i; i = 1(1)n$ follows Bernoulli with parameter θ. The prior distribution of θ is Beta with parameters α and β. Find Bayes estimate of θ under squared error loss.

(c) Let $(X_1, X_2, ..., X_n)$ be a random sample of size n drawn from Bernoulli population with parameter π. Find an UMVUE of $g(\pi) = 1 + n\pi + \frac{n(n-1)}{2}\pi^2$.

(d) Describe Kendall's τ and derive its association with U-statistic.

(e) Describe squared error loss function, absolute error loss function and all-or-nothing loss function. What are the Bayes estimates in these cases? Comment on Bayes estimate of mean of normal distribution.

(f) Define U-statistic of degree m. Show that U-statistic is an unbiased estimator of population variance.

2. Answer any TWO questions of the following.

(a) Describe maximum likelihood method of parameter estimation. State its properties. Suppose $(X_1, X_2, ..., X_n)$ is a random sample of size n from a distribution with cdf

$$F(x \mid \alpha, \beta) = 0 \quad if\ x < 0$$

$$= \left(\frac{x}{\beta}\right)^{\alpha} \quad if\ 0 \leq x \leq \beta$$

$$= 1 \quad if\ x > \beta$$

where the parameters $\alpha(>0)$ and $\beta(>0)$ are unknown. Obtain the maximum likelihood estimators of α and β. 3+3+4

(b) Find the mean of U-statistic. Derive the limiting form of variance of it. State the result regarding the asymptotic distribution of U-statistic with conditions, if any. 3+4+3

(c) Derive Bayes estimate of a real parametric function $\gamma(\theta)$ under squared error loss. Let $X_1, X_2, \ldots, X_n$ be samples drawn from a normal distribution with mean μ and variance σ^2, where μ is unknown. The prior distribution of μ is assumed to be normal with mean γ and variance η^2. Derive the Bayesian point estimate of μ under the quadratic (squared error) loss function. 4+6

M.Sc. Examination 2022
Semester I
Statistics
Course: MSC-14
(Sample Survey)

Time: 3 Hrs.

Full Marks: 40

Answer any four questions.

1. (a) Explain the rationale behind the randomized response technique.
   (b) Describe how you will estimate the proportions of the people of West Bengal belonging to different political parties by unrelated questionnaire method, where the unrelated question has multiple answers. 4+6

2. (a) Explain how the double sampling technique can be applied for the ratio estimation of the population total.
   (b) Find the approximate expressions for the bias and the MSE of the estimator of the population total under this sampling scheme.
   (c) Also find the condition under which the scheme outperforms simple random sampling. 2+6+2

3. (a) Derive the expressions of the first and second order inclusions probabilities in PPSWR (N, n) and SRSWR (N, n) designs.
   (b) Also find the expected values of the effective sample sizes in these two cases separately. You should derive the necessary results. 4+6

4. (a) Distinguish between informative and non-informative sampling designs.
   (b) Define homogeneously linear unbiased estimator (HLUE). Prove that among the class of HLUEs, no one exists with uniformly minimum variance. 2+(2+6)

5. (a) What do you mean by two stage sampling? Compare its relative merits and demerits with that of simple random sampling.
   (b) Suggest an unbiased estimator of the population total under this sampling scheme and find its variance.
   (c) Describe how can you obtain the optimal sizes of the first and the second stage samples subject to a given cost. (You may assume that each first stage unit contains the same number of second stage units). 3+3+4

6. (a) Define Desh Raj's estimator. Prove that it is unbiased for the population total under PPSWOR. What is the drawback of this estimator and how can you overcome it?
   (b) Define Midzuno's sampling scheme. Find the first order inclusion probabilities under this scheme. (2+2+2)+(2+2)

M.Sc Semester I Examination, 2022

Statistics

MSC-15(Practical)

Time: Four Hours | Full Marks: 40

One may use computer, if necessary

1. Following data represent a random sample of size from the Cauchy population with the probability density function $f(x,\theta)=\frac{1}{\pi}\cdot\frac{1}{1+(x-\theta)^2};-\infty<x,\theta<\infty$. Find out the MLE of θ. The observations are 3.8807, 2.9957, 5.2043, 4.9893, 2.7468, 4.9557, 4.9367, 3.9649, and 3.1674. 7

2. Consider the problem of point estimation of θ in $N(\theta,1)$. Given that θ belongs to [-1, 1]. On the basis of a sample of size n, the following estimator has been defined.

$$T = -1 \text{ if } \bar{X} < -1$$
$$= \bar{X} \text{ if } -1 \le \bar{X} \le 1$$
$$= 1 \text{ if } \bar{X} > 1,$$

$\bar{X}$ being sample mean. Assuming (i) squared error loss and (ii) absolute error loss draw the risk curve of $\bar{X}$ and T over the range $\theta \in [-1,1]$ on the same graph paper and comment. Take n=15. 8

3. (i) Suppose that the two observations on each of three treatments are as follows.

Treatments		
t_1	t_2	t_3
8	5	12
6	3	14

Assuming the linear model to be $y = X\beta + \epsilon$, where $\beta' = (\mu, t_1, t_2, t_3)$, find the BLU estimates of the following functions:

a. μ

b. $t_1 - t_2$

c. $2\mu + t_1 + t_2$

d. $\mu + \frac{t_1+t_2+t_3}{3}$

e. $t_1 + t_2$

(ii) Would you like to reject the null hypothesis $t_1 - t_2 = 7$ at 5% level?

8+6

4. Consider the following data on perspiration in 10 healthy females measured in terms of sweating rate (X_1) along with $X_2 = Na\ content$ and $X_3 = Ka\ content.$

No.	X_1	X_2	X_3
1	3.7	48.5	9.3
2	5.7	65.1	8
3	3.8	47.2	10.9
4	3.2	53.2	12
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14
8	7.2	33.1	7.6
9	6.7	47.4	8.5
10	5.4	54.1	11.3

Assuming $X \sim N(\mu, \Sigma)$, perform a test procedure to test the following hypothesis:

$$H_0: \mu = \begin{pmatrix} 4 \\ 50 \\ 10 \end{pmatrix}$$

4

5. Practical Note Book and Viva-Voce 5

****

qf (α, df1, df2)

M.Sc. Examination 2022
Semester I
Statistics (Practical)
Course: MSC-16

Full Marks: 40 **Time: 4 Hours**

1. At an experimental station, there are 80 fields sown with wheat. Each field was divided into 16 plots of equal size ($1/16^{th}$ hectare). Out of 80 fields, 8 were selected by SRSWOR. From each selected field, 4 plots were chosen by SRSWOR. The yields in kg/plot are given below:

Selected Field	1	2	3	4	5	6	7	8
Plots	Yield							
1	4.32	4.16	3.06	4.00	4.12	4.08	5.16	4.20
2	4.84	4.36	4.24	4.84	4.68	3.96	4.24	4.66
3	3.96	3.50	4.76	4.32	3.46	3.42	4.96	3.64
4	4.04	5.00	3.12	3.72	4.02	3.08	3.84	5.00

(i) Estimate the yield of wheat per hectare for the experimental station along with its standard error.

(ii) How can one estimate obtained from a simple random sample of 32 plots be compared with the estimate obtained in (i)?

(iii) Obtain optimum n and m under cost function $5n + 2mn = 100$, where n and m respectively stand for the number of first stage units drawn and the number of second stage units drawn from each sampled FSU by SRSWOR. 5+3+4

2. Consider the following population data.

Unit Number	x-value	y-value
1	7	32
2	11	41
3	4	25
4	10	67

Compare the performance of Desh Raj's estimator with Horvitz-Thompson estimator assuming a PPSWOR sample of size 2 is drawn. 8

3. A survey on 32 household was conducted and Warner's randomized response technique (Related Question method with $\pi = 0.61$) was applied among the heads of the households to ask about the habit of underpaying the income tax. The actual amount of income tax underpaid is given in the following table.

$\sum_{i=1}^{n} \frac{y_i^2}{p_i^2/\delta_i} - t^2_{RHC}$

Household		Response	Amount underpaid
Serial No.	Size	Yes(1)/No(0)	
1	3	1	2300
2	2	1	17000
3	5	0	5568
4	1	1	1304
5	3	0	0
6	2	1	0
7	4	0	711
8	3	0	1203
9	2	1	9874
10	4	1	2200
11	4	1	0
12	7	0	12000
13	2	1	1807
14	3	1	1400
15	4	0	708
16	4	1	1500
17	5	1	0
18	2	0	1100
19	1	0	1825
20	4	1	0
21	5	1	1407
22	3	1	342
23	2	1	645
24	4	1	0
25	3	0	713
26	5	0	1822
27	2	1	0
28	3	1	1623
29	5	1	1108
30	6	0	365
31	2	1	0
32	3	1	1409

(i) Take a sample of 4 households using Rao-Hartley-Cochran's sampling scheme.

(ii) Estimate the total amount of income tax underpaid by these 32 households under the above scheme. Also provide an unbiased variance estimate.

(iii) Use the sample to estimate the proportion underpaying the income tax.

(iv) Now take a sample of 4 distinct households using Lahiri's method. Provide an estimate of the total amount of income tax underpaid under the sampling scheme.

5+3+2+5

4. Practical Note Book and Viva-Voce. 5