INDIAN STATISTICAL INSTITUTE

B.Stat. (Hons.) III Year: 1995-96

SEMESTRAL - I EXAMINATION

Differential Equations

Date: 13.11.1995 Maximum Marks: 100 Time: $3\frac{1}{2}$

Note: This paper is worth 115 marks. Answer as many questions as you can but the maximum marks that you can score is 100.

1.(a) Find the general solution of the following differential equation in terms of Bessel functions ($x > 0$):

$$x^2y'' - xy' + (x + 1)y = 0.$$

[10]

(b) Directly verify from the series representations of the Bessel functions $J_p(x)$ that

$$\frac{d}{dx}[x^p J_p(x)] = x^p J_{p-1}(x)$$

$$\frac{d}{dx}[x^{-p} J_p(x)] = -x^{-p} J_{p+1}(x).$$

[4]

(c) After first establishing the claim for $J_{\frac{1}{2}}(x)$ and $J_{-\frac{1}{2}}(x)$, use (b) to show that $J_{m+\frac{1}{2}}(x)$ is an elementary function, where m is an integer. (6+4) = [10]

(d) Use (b) to show that the zeros of J_p and J_{p+1} interlace, i.e., there is a zero of J_p between each pair of positive zeros of J_{p+1}, and a zero of J_{p+1} between each pair of positive zeros of J_p.

[6]

2.(a) Express, near $x = 1$, the general solution of the equation

$$(1 - x^2)y'' - xy' + p^2y = 0,$$

where p is a non-negative constant, in terms of the hypergeometric function, and show that the only solutions of this equation whose derivatives are bounded near $x = 1$, are constant multiples of

$$F(p, -p, \frac{1}{2}, \frac{1-x}{2}).$$

Examine, for the above equation, the nature of the point $x = \infty$ as a singularity.

[20]

contd..... 2/-

(b) The differential equation

$$y'' + \lambda x y = 0, \quad y(0) = 0, \ y'(0) = 1$$

can be solved by substituting a power series in λ, $y(x;\lambda) = \sum_{n=0}^{\infty} y_n(x)\lambda^n$ and equating equal powers of λ. Show that this leads to the equations

$$y_0''(x) = 0, \ y_n''(x) = -x y_{n-1}(x), \ n = 1, 2, 3, \ldots .$$

Solve, choosing the constants of integration properly. Show that the series converges for all λ and for all x. [10]

3.(a) Solve for $y_1(t)$ and $y_2(t)$:

$$y_1'(t) + \tfrac{1}{2}[y_1(t) - y_2(t)] = 3$$
$$y_1(t) - 5y_2(t) - 6\int_0^t y_2(s)ds = 4,$$

the initial conditions being $y_1(0) = 4$, $y_2(0) = 0$. [5]

(b) If f is a periodic function of period p, prove that with suitable restrictions on f, the Laplace transform F(s) of f is given by

$$F(s) = \frac{1}{1 - e^{-ps}} \int_0^p e^{-st} f(t)\,dt.$$

Use this to solve, by Laplace transform methods, the differential equation

$$y'' + y = h(t-\pi) - h(t-2\pi), \ y(0) = y'(0) = 0$$

where h(t), Heaviside's unit function, is defined by

$$h(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 . \end{cases}$$ [10]

(c) Solve the following integral equation for y(t):

$$y(t) = \sin ht + e^{-t}\int_0^t e^{x} y(x)\,dx.$$ [5]

4. Solve the following differential equations:

(a) $Y'(t) = AY(t) + Q(t)$,

$$Y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}, \quad Y(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad Q(t) = \begin{bmatrix} 0 \\ -2e^{t} \end{bmatrix},$$

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 1 \end{bmatrix}. \quad \ldots$$

contd..... 3/-

(b) $y' = 1 + y^2 - x^2$, $y(0) = 1$.

(c) $(y^2 - 1)dx + (3x^2 - 2xy)dy = 0$.

(d) $x^2 y'' - xy' + y = x^2 + 2x + 3$.

(6+4+4+6) = [20]

5.(a) Consider the first-order non-linear system

$$Y'(x) = F(x, Y(x)), \quad Y(a) = B$$

where the $\mathbb{R}^n$-valued function $Y(x)$ is defined on a suitable interval $(a-c, a+c)$ in $\mathbb{R}$ and F is an $\mathbb{R}^n$-valued function defined on some suitable set S in $\mathbb{R}^{n+1}$. State carefully, without proof, the conditions on F and S and c which ensure the existence and uniqueness of the solution $Y(x)$ of the system.

[6]

(b) Consider the linear initial-value problem

$$y' + y = 2e^x, \quad y(0) = 1.$$

Find the exact solution Y of this problem. Starting with the initial guess $Y_0(x) = 1$, determine by Picard's method the successive approximations $Y_n(x)$, and find $\lim_n Y_n(x)$.

[9]

:bcc:

INDIAN STATISTICAL INSTITUTE

B.Stat. III Year:1995-96

LINEAR STATISTICAL MODELS

Semestral Examination

Date: 15 November 1995 Maximum Marks:100 Time: $3\frac{1}{2}$ hours

Figures in brackets | | indicate marks allotted to the questions.

1. Consider a linear model $(Y, X\beta, \sigma^2 I)$, where Y and β are n- and m-column vectors respectively, X is an $n \times m$ matrix with rank r. Show that it is possible to rewrite the model as $(Y, Z\theta, \sigma^2 I)$ such that (i) Z is of rank r, (ii) θ is of the form $A\beta$ where A is an $r \times m$ matrix and (iii) the estimable parametric functions are the same in both the models. [20]

2. Consider a one-way classification (treatments) model with a single covariate. The covariate is to be used in a quadratic regression. Formulate a suitable analysis of covariance model and explain how to carry out estimation of parameters and test of hypothesis of treatment effects. [You may use results from the general theory of analysis of covariance.] [20]

3. Five toothpastes are being tested for their abrasiveness. The variable (y) of interest is the time (in minutes) until mechanical brushing of a material similar to tooth enamel, exhibits wear. The five toothpastes are the same except for the presence or absence of certain additives. Each toothpaste was tested four times on material assigned at random out of a set of 20. Descriptions of the five toothpastes, summarised data and a preliminary analysis of variance are as follows:

Toothpaste	Additive	Total of 4 observations
1	None	199.0
2	Whitener	197.4
3	Fluoride	211.3
4	Whitener with freshener	186.5
5	Fluoride with freshener	215.8

PLEASE TURN OVER

Analysis of Variance

Source	d.f.	Sum of squares
Toothpastes	4	136.8
Residual	15	13.0

The investigator wants to make the following comparisons:

Additive	vs.	No additive
Whitener	vs.	Fluoride
Whitener	vs.	Whitener with freshener
Fluoride	vs.	Fluoride with freshener

Formulate a suitable linear model for this problem. Set out the hypotheses to be tested. Complete the analysis of variance table to make these comparisons and write your conclusions. [25]

4. In a trial with four drugs, 20 subjects randomly chosen from a population were given each of the four drugs at four points of time (sufficiently separated so that there was no residual effect of one drug when the next one is given) and observations on a response variable were made. Formulate a suitable model for this situation and explain how you will analyse the data to study the differences between the drugs. [15]

5. Describe the Duncan and Newman-Kuels methods of multiple comparisons. [10]

6. Establish the odds ratio properties of the log-linear model denoted by [12][13] in the usual notation. [10]

INDIAN STATISTICAL INSTITUTE

B.STAT.(HONS.) III YEAR: 1995-96

SEMESTRAL - I EXAMINATION

STATISTICAL INFERENCE I

Date: 17.11.95 Maximum Marks:100 Time: 3 hours

Note: Answer all the questions.

1. Under appropitae assumptions derive the Bhattacharya system of lower bounds for the variance of an unbiased estimator of a real-valued function of a real parameter θ based on n iid observations from a distribution with p.d.f / p.m.f $f(x,\theta)$.

Show that the bounds obtained are not less than the information limit. [12+3=15]

2. Let $X_1,\ldots,X_n$ be n iid observations with common density, $f(x,\theta) = e^{-(x-\theta)}, x > \theta$.

Show that min $(X_1, X_2,\ldots,X_n)$ is sufficient for θ. Is it also complete? Justify your answer. Show that $X_1 - X_2$ and min $(X_1,\ldots,X_n)$ are independently distributed. [3+9+4=16]

3. Let $X_1,\ldots,X_n$ be iid with p.d.f $f(x,\theta) = \frac{1}{\pi}\cdot\frac{1}{1+(x-\theta)^2}$ $-\infty < x < \infty, -\infty < \theta < \infty$. Let $F_n(x)$ be the empirical c.d.f i.e, $F_n(x) = (\#\ X_{p_i} \le x)/n$. Show that after suitable standardization $e^{F_n(x)}$ is asymptotically normal as $n\to\infty$. [20]

4. Define monotone likelihood ratio (m.l.r) property. Show that hyper geometric distribution has m.l.r property. Describe the importance of m.l.r property in statistical testing hypothesis problems. [2+6+3=11]

5. Let the distribution of the random variable X belong to a one parameter exponential family of distributions. i.e. $f(x,\theta) = C(\theta)h(x)e^{\theta x}$, $\theta\in\Omega$, an interval of R_1. Then given any test ϕ and $\theta_1\in\Omega$, $\theta_2\in\Omega$ with $\theta_1 < \theta_2$, show that there exists a two sided test ϕ' of the form

$$\phi'(x) = \begin{cases} (x)1, & \text{if } x < x_1 \text{ or } x > x_2 \\ \gamma_i, & \text{if } x = x_i, i = 1,2. \\ 0, & \text{if } x_1 < x < x_2 \end{cases}$$

such that $E_{\theta_i}\phi'(X) = E_{\theta_i}\phi(X)$, $i = 1,2.$ [12]

INDIAN STATISTICAL INSTITUTE

B.Stat. (Hons.) III Year : 1995 - 96

SEMESTRAL - I EXAMINATION

Sample Survey

Date: 20.11.1995 Maximum Marks: 100 Time: $2\frac{1}{2}$ hours

Note: The paper carries 110 marks. You may attempt any part of any question. However, the maximum you can score is 100.

1. (i) What is stratified sampling ? Discuss different problems associated with stratified sampling.

(ii) A stratified estimator $\hat{\theta}(y)$ for a parameter $\theta(y)$ has the variance of the form as given below (V_{jh}, $j = 1, 2$ not depending on n_h):

$$V(\hat{\theta}(y)) = \sum_h \left(\frac{W_h^2 V_{1h}}{n_h} + V_{2h} \right)$$

(a) Obtain optimum - allocation of total sample size under a linear cost function.

(b) Obtain Neyman type allocation, proportional allocation and equal allocation as particular cases of the optimum allocation obtained in (a) above.

(c) Obtain the variance expressions under Neyman and proportional allocations and compare them.

(5+20) = [25]

2. (i) Define ratio, regression, difference and product estimators for a population mean based on

(a) Simple Random Sampling Without Replacement (SRSWOR) and

(b) Probability Preportional to Size With Replacement (PPSWR) designs.

(ii) Define separate and combined ratio and regression estimators for a population mean in case of stratified simple random sampling without Replacement.

(iii) Derive an approximate expression in large samples for the mean square error of the regression estimator of population mean in PPSWR sampling.

(6+6+8) = [20]

p.t.c.

3. (i) Describe double-sampling technique and define double-sam[illegible] ratio and regression estimators for a population mean.

(ii) Derive large sample approximation to the mean square erro[illegible] of the double-sampling regression estimator. [illegible]

(5+10) = [15]

4. Suppose you want to estimate the total number of house-ho[illegible] living below the 'poverty line' in Bankura district of We[illegible] Bengal. Suggest a suitable unbiased sampling strategy based on a two-stage sampling procedure.

Derive the variance expression for the proposed estimator.

Also obtain an unbiased estimator for the variance of the estimator.

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

5. (i) Describe the Hansen-Hurwitz technique for obtaining an unbiased estimator of the population mean in the presence of non-response.

(ii) Derive an expression for the variance of Hansen-Hurwitz strategy.

(5+10) = [15]

6. Define the Horvitz-Thompson estimator('HTE)for a finite population mean of a variable y. Derive the Yates-Grundy form of the expression of the variance of the HTE estimat[illegible] for a fixed effective size design (you may assume the for[illegible] given by Horvitz-Thompson).

INDIAN STATISTICAL INSTITUTE

B.Stat. (Hons.) III Year: 1995-96

SEMESTRAL - I EXAMINATION

Economics III

Date: 24.11.95 Maximum Marks: 100 Time: 3 hours

Note: Answer any TWO questions from Group A and any TWO questions from Group B. All questions carry equal marks. Use separate answerscripts for Group A and Group B.

GROUP - A

1. Discuss the hypothesis that the balance of payments problem faced by the Indian economy in the early 90's was largely due to the growing fiscal deficit in the 80's.

[25]

2. Examine the economic implications of measures for financial liberalization adopted since 1991.

[25]

3. How far in your opinion are the policies relating to export, import and foreign capital inflow appropriate for steppling up economic growth with viability in the balance of payments ?

[25]

GROUP - B

4. What are the different methods by which inter-regional differences in level of living have been sought to be measured for the India economy ? Describe these methods.

[25]

5. Critically discuss the debate on farm size and productivity in the context of Indian agriculture.

[25]

6. Consider the period prior to the 90's. Write a note on India's industrial growth during this period. What are the major features of the Indian industrialisation programme during this period.

(10+15) = [25]

INDIAN STATISTICAL INSTITUTE

B.STAT (HONS.) III YEAR : 1995-96

SEMESTRAL-I (BACK-PAPER) EXAMINATION

STATISTICAL INFERENCE I

Date:27.12.95 Maximum Marks:100 Time:3½ hours

Note: Answer all the questions.

1. (a) State and prove the factorisation theorem in discrete case.

(b) Let $X_1,...,X_n$ be iid with uniform distribution in $[0,\theta]$, $\theta > 0$. Find a sufficient statistic for θ. Is it minimal sufficient? Justify your answer. [10+10=20]

2. Show that the natural parameter space of a one parameter exponential family of distributions is an interval of R_1. Find the m. g. f of a random variable following such a distribution and hence find the mean and variance of the distribution. [7+8=15]

3. State and prove the generalized Neyman Pearson lemma. Under appropiate assumptions find out the form of LMP test for testing $H_0:\theta = \theta_0$ vs $H_1:\theta > \theta_0$ where the random variable X is distributed with p.d.f / p.m.f $f(x,\theta)$. [7+8=15]

4. Show that under appropiate assumptions a consistent m.l. equation estimator for an unknown real parameter θ is asymptotically normal after suitable standardisation as the sample size $n \to \infty$. [12]

5. Derive the Chapman-Robbins inequality and show that the Information inequality follows from it. Explain by an example that the lower bound in the information inequality may not be attained. [10+3+5=18]

6. Let $X_1,...,X_n$ be i i d with p.d.f/p.m.f $f(x,\theta)$. Find the limiting null distribution of the likelihood ratio test statistic for testing $H_0:\theta = \theta_0$ vs $H_1:\theta = \theta_1$ as $n \to \infty$, under appropiate assumptions. [9]

7. Explain the following terms:

(i) consistency of a test, (ii) complete statistic and (iii) monotone likelihood ratio property. [3 x 4 = 12]

INDIAN STATISTICAL INSTITUTE

B.Stat.(Hons.) III Year: 1995-96

Sample Surveys

Semestral-I Backpaper Examination

Date : 28.12.1995 Maximum Marks : 100 Time : $2\frac{1}{2}$ Hours.

The paper carries 115 marks. You may attempt any part of any question. However the maximum you can score is 100.

1.(i) Define ratio, regression, difference and product estimators for population mean $\bar{Y}$ of a variable y in case the population mean of a supplementary variable x is known.

(ii) Consider the strategies $T_0 \equiv$ (Simple Random Sampling Without Replacement(SRSWOR), sample mean), $T_R \equiv$ (SRSWOR, ratio estimator) and $T_p \equiv$ (SRSWOR, product estimator) for $\bar{Y}$. Let c_y and c_x denote coefficients of variation of y and x respectively and ρ_{yx} the correlation coefficient. Which of the above strategies you would suggest to be used for $\bar{Y}$ in the following situations and why ?

(a) $\rho = 0.9$, $c_x \geq 2c_y > 0$ (b) $\rho \geq 0.4$, $c_y \geq 1$, $0 < c_x \leq 0.6$

(c) $\rho \leq -0.3$, $c_y \geq 0.8$, $c_x \leq 0.45$ (d) $\rho > 0.2$, $c_y \geq 0.4$, $c_x \geq 0.3$.

(4+16) = [20]

2. Describe the procedure of cluster sampling and suggest an unbiased estimator for the population mean when clusters are selected by simple random sampling without replacement. Assuming that clusters are of equal size, obtain the expression of the variance of the above estimator in terms of intra cluster correlation coefficient.

(5+20)=[25]

3. Define ratio and regression estimators for population total Y based on unstratified and stratified simple random sampling. Discuss how double sampling procedures may be applied for the use of ratio and regression estimators when the population total of the auxiliary variable is unknown.

[20]

4.(i) Define the Horvitz-Thompson(H-T) estimator of a finite population total.

(ii) Derive the Yates-Grundy form of variance expression of H-T estimator for a fixed effective size design. Prove the results you use connecting the effective sample size and th inclusion probabilities of first two orders. (5+20)=[25]

p.t.o.

5.(i) What do you mean by sampling and non-sampling errors ?

(ii) Indicate, in brief, the sources through which non-sampling errors may arise.

(iii) Describe in brief the steps you would take to control the non-sampling errors while conducting a sample survey.

[5+10+10=25]

INDIAN STATISTICAL INSTITUTE

B.Stat. III Year:1995-96

LINEAR STATISTICAL MODELS

Backpaper Examination

Date: 29 December 1995 Maximum Marks:100 Time: 3 hours

Figures in brackets [ ] indicate marks allotted to the questions.

1. In the model $(Y, X\beta, \sigma^2 I)$ with all elements of β estimable, show that the minimum variance linear unbiased estimator of a linear combination $\ell'\beta$ is given by $\ell'\hat{\beta}$, where $\hat{\beta}$ is the solution to the normal equations. [15]

2. In a two-way classification model with one observation per cell without interaction, write down expressions for sum of squares due to row and column effects and show that these are independently distributed. [20]

3. In studies of the effect of acid rain on the biomass in freshwater lakes, biologists have found that biomass decreases as acid concentration increases. If the lakes have sources of phosphorus, however, biomass increases with an increase in the amount of phosphorus available. In an effort to make a thorough study, researchers took water samples from 18 randomly selected lakes and measured acidity (x_1), available phosphorus (x_2) and population density (y) of a certain species of algal plant. The following statistics were computed (in the usual notation):

   $\bar{y} = 1400$; $\bar{x}_1 = 2100$; $\bar{x}_2 = 760$

   Corrected sum of squares and products matrix of (y, x_1, x_2):

$$\begin{bmatrix} 14.400 & -3000 & 2100 \\ & 1600 & 900 \\ & & 3600 \end{bmatrix}$$

   $\hat{\beta}_1 = -2.563$; $\hat{\beta}_2 = 1.224$; $s^2 = 276$

PLEASE TURN OVER

(a) Write down the equation of the least-squares plane. [5]

(b) Test $\beta_1 = \beta_2 = 0$. [20]

(c) If acidity is increased one unit and phosphorus held fixed, what is the effect on population density? [5]

4. Explain, with a suitable example, why the 'interaction mean square' is sometimes a more suitable denominator than the residual mean square in an analysis of variance F-ratio. [15]

5. Consider three-dimensional frequency tables. Describe various hierarchical log-linear models and state their respective odds ratio properties. [20]

INDIAN STATISTICAL INSTITUTE

B.Stat. (Hons.) II and III Year: 1995-96

BACK-PAPER SEMESTRAL-I EXAMINATION

Geology

Date: 29.12.1995 Maximum Marks: 100 Time: 3 hours

Note: Attempt any FIVE. Answers should be brief, to the point.

1. Can fossils be used by human beings for their benefit ? Describe one such use which you think to be quite important.

   What is a coprolite and a gastrolith ?

   (4+12+4) = [20]

2. Describe the role of water in the weathering of rock.

   Can you work out what happens because of erosion in a granitic terrain in a somewhat humid climate ? What would be the expected topography if the same terrain is subjected to weathering in an extremely arid climate ?

   (8+9+3) = [20]

3. Discuss the composition of the primeval atmosphere and the origin of the free oxygen in that.

   (10+10) = [20]

4. What is meant by the "Proterozoic rocks" ? If someone says that he has located coal beds in a Proterozoic sequence of rocks, would you agree ? Justify your answer. What is the interpretation of the presence of the Ediacaran fauna in the Proterozoic sequence ?

   (5+5+10) = [20]

5.(a) What is a fold and its axis ? Define the dip and the strike of a bed.

   (b) What is meant by "sorting" of terrigenous sediments ? How does the fabric of a sedimentary deposit help in understanding the sedimentary processes ?

   (10+10) = [20]

6. Describe the basic ideas basing on which the theory of radioactivity has been built up. In how many ways the age of a rock may be determined ?

   Describe the K-Ar method employed in the age determination of a marine sedimentary rock.

   (6+6+8) = [20]

p.t.o.

7. What are the uses of the P- and S - wave study in an earthquake ? What is the Low Velocity Zone and where does it occur ?

   Describe in short the physical characteristics of the earth's core.

   (6+6+8) = [20]

8. Can a mineral be a crystal as well as a crystalline substance

   How does a mineral form ? Name <u>four</u> rock-forming minerals along with their respective chemical compositions. Give an example of an amorphous mineral and state its usefulness.

   (6+5+6+3) = [20

:bcc:

INDIAN STATISTICAL INSTITUTE

B.Stat.(Hons.) III Year : 1995-96

Stochastic Processes

Semestral-II Examination

Date : 22.4.1996 Maximum Marks : 60 Time : 3 Hrs.

1. Each question carries 12 marks.
2. Write neatly, justify all your answers.

1. Consider the Markov Chain with states 0,1,2,3,4,5 and transition matrix as given below :

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(a) What are the Communicating classes ?

(b) Compute all eigen values of modulus 1 for this matrix and also their geometric multiplicities

[4+8]

2.(a) We have an irreducible markov chain with state space as the set of integers and transition matrix (p_{ij}). If the chain is transient show that there is a non constant bounded solution of the system of equations

$$\sum_j p_{ij} u_j = u_i \text{ for all } i \neq 0.$$

(b) Use the above criterion to show that the simple symmetric random walk in one dimension is recurrent.

[6+6]

3. At Allahabad Bank, customers arrive at a Poisson rate of 6 per hour between 10 A.M. and 11 A.M. It increases steadily from 6 per hour at 11 A.M. to 16 per hour at 1 P.M. From then on it steadily decreases to 0 per hour at 2 P.M.

(a) What is the average number of customers entering the bank per day ?

(b) Given that 20 customers entered between 10 A.M. and 11 A.M. What is the conditional probability that 2 customers enter between 1 P.M. and 2 P.M. ?

[8+4]

p.t.o.

4.(a) Define standard Brownian motion $(B_t)_{t\geq 0}$.

(b) Fix $t_1 < t < t_2$. Using your definition above calculate the joint density of (B_{t_1}, B_t, B_{t_2})

(c) Given $B_{t_1} = a$ and $B_{t_2} = b$ calculate the conditional density of B_t; its mean and variance.

[3+3+6]

5. A barber shop across the campus has 5 chairs and 3 barbers working. Potential customers arrive at the rate of 6 per hour, however a potential customer leaves if no chair is empty. Each barber works at the rate of 2 customers per hour.

(a) State your assumptions clearly and formulate the model.

(b) Calculate the average number of customers in the shop (assuming steady state)

(c) What proportion of time the shop is loosing customers

[4+4+4]

INDIAN STATISTICAL INSTITUTE
B.Stat.(Hons.) III Year:1995–96
STATISTICS COMPREHENSIVE
SEMESTRAL EXAMINATION

Date:24 April 1996 Maximum Marks:100 Time:$3\frac{1}{2}$ hours

Figures in brackets [ ] *indicate marks allotted to the questions.*
You may use books, notes, photocopied material, etc. that you care to bring; but these will not be allowed to be passed on to others.

1. Airlines are allowed to 'overbook' (that is, book more passengers than available seats) as a safeguard against cancellations. If it so happens that more passengers than available seats actually report for a flight, the airline is expected to take care of passengers who could not be accommodated in the flight, until they are accommodated in subsequent flights.

   Suppose that an airline allows $n \geq 200$ passengers to book on a flight with maximum capacity of 200. Suppose that
   (i) each passenger who flies brings the airline revenue of Rs. 1000;
   (ii) each passenger for whom there is no place in the flight costs the airline an amount of Rs. 1500 (that is, a revenue of Rs. -1500);
   (iii) passengers have a probability 0.05 of cancelling their flight independently of one another.

   Let X denote the number of passengers reporting for a flight.

   (a) Stating any further assumptions that you may use, write down expressions for $\Pr(X = x)$, for the revenue $R_n(x)$ at $X = x$, and the expected revenue R_n, for given n. [7]

PLEASE TURN OVER

(b) It is required to find the optimal value of n, that is, the value of n that maximises the expected revenue R_n. Towards this goal,

i. sketch the shape of this revenue function R_n over n. [5]

ii. show how R_n can be computed by computing it at $n = 208$ to a reasonable degree of approximation. [18]

2. It has been observed that the distribution of first digits of physical constants is far from discrete uniform and is more concentrated at smaller values. For instance, the first digits of the areas in km^2 of 40 European countries in 1988 are distributed as follows:

1	2	3	4	5	6	7	8	9
10	7	6	6	3	2	2	1	3

(a) Examine if the following *Benford's* law fits this distribution: $\Pr(d) = \log_{10}(1 + \frac{1}{d})$. [5]

(b) Derive Benford's law as follows:
Start with the premise that the observed quantity θ has a density $p(\theta) \propto \frac{1}{\theta}$.

i. Show that a restriction of θ to the interval (a, b), where $a = 10^k, b = 10^{k+1}$ has density $1/(\theta \log_e 10)$. [5]

ii. Obtain the probability that θ has first digit d as $\log_{10}(1+\frac{1}{d})$. [10]

3. (a) Formulate clearly the Generalised Linear Model (GLIM). [5]

(b) Formulate clearly the log-linear model for multiway frequency data. [5]

(c) Examine if the log-linear model is an example of GLIM. [10]

CONTINUED

4. In a sample survey, from 200 households (*hh*) distributed as follows in four villages, it is desired to select a simple random sample of 20 *hh*'s with replacement.

Village	1	2	3	4	Total
Number of *hh*'s	50	30	40	80	200

Examine each of the following three schemes and explain, with reasons, if it is suitable for this purpose:

(a) Select 5 *hh*'s from Village 1, 3 from Village 2, 4 from Village 3 and 8 from Village 4, using simple random sampling with replacement.

(b) Number the households 1 to 200. Select a number at random from 1 to 10. Let it be I. Then select *hh*'s with numbers

$$I + 10n : n = 0, 1, 2, \ldots, 19.$$

(c) Select a village with probability proportional to the number of *hh*'s and from the selected village select one *hh* at random. Repeat this procedure 20 times.

[20]

5. Write a short (about a page) account of the significant contributions to the development of Statistics of one of the following:
Bayes, Fisher, Mahalanobis, Tukey [10]

contd. ....2.

INDIAN STATISTICAL INSTITUTE
B.STAT.(HONS.) III YEAR: 1995-96
DESIGN OF EXPERIMENTS
SEMESTRAL-II EXAMINATION

Date: 26.4.96 Maximum Marks:100 Time: 3 Hours

Note: Answer Question no.1 and any THREE out of the rest. Marks allotted to a question are indicated in brackets [ ] at the end. Submit your Practical Records, if not already done, to the Course Instructor on or before 30.4.96. These records carry 10 marks.

1.(a) The following table gives the yields (lbs/plot) of ear corn (y) and the number of plants in the plot (x) for five different varieties A,B,C,D,E, in 2 randomised blocks.

Blocks		Treatments A	B	C	D	E
I	x	18	16	18	14	15
	y	8.6	7.5	6.7	6.5	8.2
II	x	14	16	15	16	19
	y	6.5	7.6	8.2	8.3	4.6
III	x	16	15	15	17	16
	y	7.1	7.5	5.7	7.5	9.6

Test if the variety yields differ significantly or not, adjusting for variation in x. Use $\alpha = 0.05$. [21]

(b) Practical Records. [10]

2.(a) Describe the missing plot technique. Prove that the error sum of squares obtained by the technique is the valid error for the incomplete data.

(b) Suppose in a randomised block design for v treatments in b blocks, two observations are lost, affecting two different blocks and two different treatments. Use the missing plot technique to get the "estimates" of the missing cells, and the variances of various estimated elementary treatment contrasts. [(5+9)+(6+3) = 23]

3.(a) Show that the maximum number of mutually orthogonal latin squares (MOLS) of order s is less than or equal to (s - 1). Also prove that all the (s - 1) MOLS of order s exist whenever s is a prime or power of a prime.

(b) Develop the simultaneous confidence intervals for multiple comparison purposes, using the methods of (i) Bonferroni, (ii) Scheffe' and (iii) Tukey. [(3+8)+(3x4) = 23]

contd. ....2.

4. (a) Give a balanced confounding scheme for a $(2^5, 2^3)$ factorial design. Construct one replication of this design. Give the analysis of variance for the suggested design, indicating clearly how the various sums of squares are to be computed.

(b) The following is a block of a replication of a $(2^7, 2^4)$ desig Determine all the confounded effects in this replication. (bcdefg, abcd, cdg, bdf, bce, acdef, abdeg, abcfg, de, cf, bg, adfg, aceg, abef, a, efg). [(5+4+9)+5 = 23]

5. (a) An experiment to study the effects of irrigation (I) at four levels was originally planned according to a 4 x 4 latin square arrangement. Subsequently, it was decided to include in the study two more factors, namely, N and P, the nitrogeneous and the phosphatic fertilizers, each at three levels. Consequently, a split plot design was thought appropriate, with more importance attached to the factors N and P. Give a rough sketch of the layout plan for the suggested design. Also give the analysis of variance for this design, indicating clearly (i) the underlying model, and (ii) the expressions for the various sums of squares involved.

(b) If the subplots were arranged into strips for operational convenience, how would your model and the analysis of variance change? Give the details. [(3+2+10)+8 = 2

INDIAN STATISTICAL INSTITUTE

B.Stat. (Hons.) III Year : 1995-96

Statistical Inference II

Semestral-II Examination

Date : 30.4.1996 Maximum Marks : 100 Time : $3\frac{1}{2}$ Hrs.

Group A is closed notes and Group B is open notes. You can answer Group B only after you submit the answerscript of Group A. Two separate answerscripts should be used. No books are allowed.

GROUP A

Answer as many as you like. The maximum you can score is 55.

1. Compute the expectation and variance of a simple linear rank statistics and the covariance between two simple linear rank statistics. Hence find out the variance of Spearman's rank correlation coefficient.

   Consider the two-sample problem where the populations sampled may be assumed to be normal. Suppose the original observations are lost but the ranks of the observations are preserved. Suggest an appropriate testing procedure for the equality of the two populations against the location alternative. [12+5=17]

2. Explain Hajek's projection principle and its use in the deviation of the asymptotic normality of a U-statistic.

   With i.i.d. observations having a finite and positive variance, consider the sample Gini coefficient

   $$G = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} |X_i - X_j|.$$ Derive the asymptotic normality of G.

   [12+6 = 18]

3. Suppose $X_1, \ldots, X_n$ are iid $N(\theta,1)$ and consider testing $H_0: \theta = 0$ against $H_1: \theta > 0$. Find the asymptotic relative efficiency of the sign test with respect to the appropriate parametric test.

   [12]

4. Show that there is no nonnegative unbiased estimator of the density. Explain the histogram and kernel method for density estimation. With informal calculations, calculate the optimal rate of convergence and the optimal order of the bandwidth.

   [6+6+6=18]

p.t.o.

GROUP 3 Max.Marks : 45

1.(a) Briefly discuss how one constructs an SPRT for composite hypotheses by using a statistic whose distribution is completely specified by two hypotheses.

(b) Illustrate the above principle in the following case. X_i's are i.i.d. $N(\mu,\sigma^2)$. To test $H_1 : \sigma^2=1$ vs. $H_2:\sigma^2=c$ where $c>1$.

(c) For the special case considered in (b), construct a sequence of i.i.d. random variables $Y_2,Y_3,\ldots$ which are $N(0,\sigma^2)$ and $\sum_{i=1}^{1} (X_j-\bar{X}_1)^2 = \sum_{j=2} Y_j^2$. Hence show that Wald theory for simple hypotheses are applicable here.

Taking $c=2$ $\alpha=\beta=0.01$, calculate $? \; {}_{\sigma^2=3}$ (SPRT rejects H_1).

[Hint for (c) : choose an orthogonal transformation Y_2((X_1,X_2)), Y_3(of (X_1,X_2,X_3)) etc. recursively).

[6+6+8+8 = 30]

OR

2.(a) For the secretary selection problem with N=5, write down explicitly the optimal stopping time and compare its performance with the asymptotically optimal stopping ti discussed in the class.

Suppose the relative ranks for an interviewee at the interview (if all the five interviewers were taken) are (1,1,2,2,2). What are the absolute ranks here ? At whic interview will the optimal stopping time stop ?

(b) A fair coin is tossed repeatedly. Let $X_n=1$ denote a hea at the nth toss, and $X_n = -1$ a tail. Let the reward sequence be Y_n where

$$Y_n = \left[\prod_{i=1}^{n} (X_i + 1) \right]. \frac{2n}{n+1}$$

and consider the optimal stopping problem.

(i) Show that

$$E(Y_{n+1} | X_1=1,\ldots,X_n=1) = \frac{1}{2} \cdot \frac{n+1}{n+2} \cdot 2^{n+2} > Y_n$$ and hen it $X_1=1,\ldots,X_n=1$, it would seem profitable to conti What would be the consequence of adopting this strategy ?

Contd.....

2.(b) (ii) Comment on the implication of this example for the principle dynamic programming.

[10+5+10+2+3 = 30]

3. Let X_i's be i.i.d. $P\{X_i = 1\} = P\{X_i = -1\} = \frac{1}{2}$. Let $s_m = X_1 + \ldots + X_m$. Let n be the first m for which $s_m = \pm 10$. Show that $n < \infty$ with probability one and calculate E(n).

[8+7 = 15]

OR

4. Let X_i's be i.i.d. $B(1,\omega)$. Let $H_1: \omega = \frac{1}{4}$, $H_2: \omega = \frac{3}{4}$ and $\alpha = \beta = 0.05$. Set up a suitable SPRT and calculate $E_\omega(n)$ for $\omega = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}$. Indicate how the graph for $E_\omega(n)$ will look.

[3+9+3 = 15]

INDIAN STATISTICAL INSTITUTE
B.STAT(HONS.) III YEAR:1995-96
ECONOMICS IV
SEMESTRAL-II EXAMINATION

Date: 2.5.96 Maximum marks: 100 Time: 3 Hrs.

Answer any five questions

1. Consider the following regression model:

$$y=\beta_1 x_1+\beta_2 x_2+u$$

where u's are i.i.d. $N(0,\sigma^2)$, and the x's are non-stochastic. Assume that all variables are measured as deviations from sample means.

a) It is desired to test the hypothesis that $\beta_1=\beta_2$.

Carefully describe, step by step, how you would perform the test on a given set of data. Be explicit and give algebraic expressions, whenever possible.

b) Assuming that $\beta_1=\beta_2$ is accepted, indicate how will you estimate β_1 and β_2 from a given set of data ? (16 + 4=20)

2. Consider the following regressionmodel

$$y_t=\beta x_t+u_t, \quad t=1,2,\ldots,T;$$

$$E(u_t)=0,\ E(u_t^2)=\sigma^2 x_t^2 \text{ and } E(u_t u_s)=0,\ t\neq s.$$

Prove that the OLS estimator of β is unbaised but inefficient and that the formula for its estimated variance yields a downward biased estimate of the true variance.

(2 + 6 + 12= 20)

3. Describe DW test for autocorrelation of the random disturbance term of a linear regression model. Derive the bounds of probability distribution of the DW statistic under the assumption that $\rho = 0$.

(20)

4. Explain the problem of identification in simultaneous equations model. Derive the rank and order conditions of identification under exclussion restrictions.

(6 + 14= 20)

p.t.o.

5. Explain the least A Variance Ratio (LVR) method of estimation of a single equation of a simultaneous equations system. Show that it is a member of k-class estimator.

(8 + 12= 20)

6. Formulate a general recursive model of simultaneous equations. Show that a recursive model is always identified and its OLS estimators are unbiased and consistent.

(4 +. 16= 20)

7. Give an example of a simultaneous equation model where restrictions on the covariance of disturbance terms help in identifying an equation. How will you consistently estimate the parameters of such a model ?

(5 + 15= 20)

8. Write short notes on any two of the following :-

a) Indirect least squares method ;

b) Tests for heteroscedasticity ;

c) Adjusted R^2 ;

d) Cochrane-Orcutt estimation procedure of an equation with a serially correlated error.

(10 + 10= 20