INDIAN STATISTICAL INSTITUTE
Final Examination: Semester I (1997-98)
B.STAT-III (1997-98)
THEORY OF STATISTICAL INFERENCE-I
Date: 3 November 1997, Maximum Marks: 100, Duration: $3\frac{1}{2}$ hours

NOTE: Answer as many questions or parts thereof as you please. The whole question paper carries 115 marks. The maximum you can score is 100.

1. $X_1, X_2, X_3, \cdots\cdots, X_n$ are *iid* random variables with $N(\sigma, \sigma^2)$ distribution, $\sigma > 0$. Show that $T(x_1, x_2, x_3, \cdots, x_n) = (\sum_1^n x_i, \sum_1^n x_i^2)$ is minimal sufficient for σ. Show that the family of distributions of T is not complete [10+5]

2. Let $X_1, X_2, X_3, \cdots\cdots, X_n$ be *iid* random variables with common density

$$p_\theta(x) = \begin{cases} e^{-(x-\theta)} & \text{for } x > \theta \\ 0 & \text{otherwise.} \end{cases}$$

where $-\infty < \theta < \infty$. Let $T(x_1, x_2, \cdots, x_n) = Min(x_1, x_2, \cdots, x_n)$.
(a) Show that for any value of θ the random variables $X_1 - T(X_1, X_2, \cdots, X_n)$ and $T(X_1, X_2, \cdots, X_n)$ are independent. Compute $P(X_1 - T > a), a \in R$.
(b) Using (a) and the fact that T is complete and sufficient for θ find the *UMVUE* of $g(\theta) = \int_0^\infty p_\theta(u)\, du$ [10+10]

3. Find the Cramer-Rao lower bound for the variance of an unbiased estimator of $e^{-\theta}$ on the basis of independent observations $X_1, X_2, X_3, \cdots\cdots, X_n$ from the Poisson distribution with mean θ. Find the *UMVUE* of $e^{-\theta}$, compute its variance and check whether the Cramer-Rao lower bound is attained or not. [5+10]

4. State carefully the Neyman-Pearson fundamental lemma. Prove the part of the lemma giving a sufficient condition for a test to be MP level α for testing a simple hypothesis against a simple alternative. [9+6]

5. Define a *Monotone Likelihood Ratio (MLR)* family of densities. [5]

6. Assume that the length of life of an electron tube of a specific make has exponential distribution with expectation θ, $\theta > 0$. A life testing experiment is conducted with n such tubes and r smallest life times $Y_1 < Y_2 < \cdots < Y_r$ are observed.
(a) Write down the joint density of $Y_1, Y_2, \cdots\cdots, Y_r$ and verify that this family of densities is *MLR* in $T(y_1, y_2, \cdots, y_r) = \sum_1^{r-1} y_i + (n - r + 1)y_r$.
(b) Show that $\frac{2T}{\theta}$ has χ^2 distribution with $2r$ degrees of freedom.

(P.T.O)

(c) Write down the *UMP* level α for testing $H : \theta \geq 1000$ against $K : \theta < 1000$ on the basis of $Y_1, Y_2, \cdots\cdots, Y_r$. Find the power of this test when $\theta = 500$, $\alpha = 0.05$ and $r = 4$. [5+5+10]

7. Does there exist a *UMP* level α test for testing $H : \theta = 0$ against $K : \theta \neq 0$ on the basis of n *iid* observations $X_1, X_2, X_3, \cdots\cdots, X_n$ from $N(\theta, 1)$ distribution ? Justify. [8]

8. Define an unbiased test. Show that the UMP unbiased test of size α for testing $H : \sigma^2 = \sigma_0^2$ against $K : \sigma^2 \neq \sigma_0^2$ on the basis of *iid* observations $X_1, X_2, X_3, \cdots\cdots, X_n$ from $N(0, \sigma^2)$ is given by :

$$\textit{Reject } H \textit{ if } \frac{\sum x_i^2}{\sigma_0^2} < C_1 \quad \textit{or} \quad \frac{\sum x_i^2}{\sigma_0^2} > C_2; \textit{ accept } H \textit{ otherwise}$$

where C_1 and C_2 are obtained using the following condition :

$$\int_{C_1}^{C_2} g_n(y)\,dy = \int_{C_1}^{C_2} g_{n+2}(y)\,dy = 1 - \alpha .$$

$g_k(y)$ denoting the χ^2 density with k degrees of freeedom. For $\alpha = 0.05$, $n = 8$, find C_1 and C_2 using the χ^2 – tables. [4+8+5]

INDIAN STATISTICAL INSTITUTE
B.STAT.(HONS.) III YEAR:1997-98
SEMESTRAL-I EXAMINATION
ECONOMICS-III

Date: 7.11.97 Maximum Marks:100 Time: 3 Hours

Note: Answer question 1 and four from the rest. Marks allotted for each question are indicated in brackets.

1.(a) Define National Income. Discuss the different methods of measuring National Income and show their equivalence. (4+14)

(b) The following data (expressed at current prices and in Rs.crores) relate to the Indian economy for the year 1992-93.

(i)	Net domestic product at factor cost	= 556344
(ii)	Personal Income	= 577748
(iii)	Indirect taxes less subsidies	= 77653
(iv)	Gross domestic capital formation	= 172908
(v)	Private Income	= 593252
(vi)	Net factor income from abroad	= -11408
(vii)	Corporation tax	9200
(viii)	Gross National Product at factor cost	= 616504

Find out from the above data

(A) Net National Product at factor cost

(B) Corporate saving

(C) Depreciation of fixed capital

(D) Gross National Product at market prices

(E) Net domestic capital formation. (10)

2. The economy is at full employment. Now the Govt. wants to change the composition of demand toward investment and away from consumption without, however, allowing aggregate demand to change. What is the required fiscal and monetary policy mix? Use the IS - LM model for your answer. (18)

3. Consider two alternative programmes for contraction. One is the reduction of an investment subsidy; the other is a rise in income tax rate. Use the IS - LM Model to discuss the impact of these alternative policies on income, interest rate and investment. (18)

4. There is 'Hundred Percent Banking' when the cash reserve ratio is unity. Explain the money multiplier process in this situation and hence derive the value of the money multiplier. (18)

contd. ....2.

5. Find out the effect of a reduction in the exogenously given cash reserve ratio in the complete Keynesian Model. (1
6. Find out the effect of wage cut on output, employment, real wage rate, rate of interest and price level in complete Keynesian Model. Explain the mechanism carefully. (1
7. How does does a down ward shift in the demand for money function (i.e. reduction in K) affect output, employment, interest rate and prices in the Classical Model. Explain the mechanism carefully. (

ISS1

INDIAN STATISTICAL INSTITUTE

B.Stat. (Hons.) III Year : 1997 - 98

SEMESTRAL-I EXAMINATION

Geology (Theoretical)

Date: 7.11.1997 Maximum Marks: 100 Time: 3 hours

1.(a) How does the shape of the earth affect its gravity field ? What is meant by gravity anomaly ? How would you account for negative gravity anomaly associated with oceanic trenches?

(2+2+2) = [6]

(b) How is the phenomenon of earthquake explained ? How are seismic records utilized to interpret the structure of the earth ?

(3+4) = [7]

(c) Describe the essential features of earth's magnetic field ? Which rocks are suitable for studying the record of the geomagnetic field ?

(2+1) = [3]

(d) What is the composition of the earth's core ? How is the density variation in the core explained ?

$(2\frac{1}{2}+1\frac{1}{2})$ = [4]

2.(a) Describe the salient features of rock weathering. Name different agents of rock erosion on land. Comment on their relative influence in different latitudinal belts on earth. Is it possible to have weathering and erosion under sea ?

$(2+1\frac{1}{2}+1\frac{1}{2}+1)$ = [6]

(b) How are sedimentary rocks classified ? [3]

(c) What are the different controlling factors of metamorphism ? Which of these factors is predominant in contact metamorphism? How would you differentiate between a hornfels and a regionally metamorphised argillaceous rock ?

$(1\frac{1}{2}+1+1\frac{1}{2})$ = [4]

(d) Briefly describe the concept of metamorphic zones ? Name a metamorphic facies which borders on P , T environment of partial melting.

(3+1) = [4]

(e) Describe the difference between folds and faults in terms of nature of deformation corresponding to these structures.

[3]

p.t.o.

3.(a) How does a plutonic igneous rock differ from a volcanic rock ? Name two plutonic rocks and two volcanic rocks. How does basalt differ from granite in terms of chemical composition ?

(2+2+2) = [6]

(b) How does the sequence of crystallization in a silicate melt representing eutectic system relate to development of graphic texture in granite ?

Describe Bowen's discontinuous reaction series.

How would you explain that olivine and quartz rarely occur together in a plutonic igneous rock ?

(3+2+2) = [7]

(c) How does the continental crust differ from the oceanic crust in terms of mineralogical composition ? Which rock(s) is (are, representative of the upper mantle ? State the range of SiO_2 content in these rocks.

(3+1+1) = [5]

(d) What is Mohorovicic discontinuity ? Does it have uniform depth below continents and oceans ?

(1+1) = [2]

4.(a) State the primary observation that led to the first proposition of continental drift. Describe the main geological observations in support of continental drift.

(1+3) = [4]

(b) Describe the nature of linear magnetic anomaly pattern on ocean floor. How would you account for its development ?

(2+2) = [4]

(c) How would you explain the association of oceanic trenches and Andean type mountain chains in terms of plate tectonic theory ?

$[2\frac{1}{2}]$

(d) How does a transform fault differ from ordinary strike-slip fault ? What would be the expected depth range of earthquake focii associated with a transform fault that offsets an ocean ridge ?

$(1\frac{1}{2}+1) = [2\frac{1}{2}]$

(e) Describe the mode of occurence of principal lignite and bituminous coal deposits of India.

[2]

contd..... 3/-

(f) How does petroleum occur in nature ? Which of the following is most suitable for exploratory drilling for hydrocarbon deposits — (1) garnetiferous mica schist horizon of Singhbhum, (2) Gondwana sedimentary rocks of Godavari valley, (3) Tertiary strata of onshore Bengal Basin. Give reasons for your answer. How is water table defined ? (2+2+1) = [5]

5.(a) What is the difference between a body fossil and an ichno-fossil ? [2]

(b) What are the utilities of studying fossils ? [6]

(c) What is the fundamental lithostratigraphic unit ? Define with example. [2]

(d) State the law of uniformitarianism. [2]

(e) How would you differentiate between an angular unconformity and a disconformity ? [3]

(f) Describe the principle of radiometric dating ? [5]

6.(a) Describe the major changes that occurred in the biosphere at the begining of the Phanerozoic. [3]

(b) What are the main divisions of the Mesozoic ? [2]

(c) When did the first bird appear on the earth ? [1]

(d) At which stage of the Quaternary did worldwide glaciation take place ? [1]

(e) What are the major points of Darwin's theory of natural selection ? [6]

(f) What is the main difference between microevolution and macroevolution ? [4]

(g) What is adaptive radiation ? [3]

:bcc:

INDIAN STATISTICAL INSTITUTE

B.Stat. (Hons.) III Year: 1997 - 98

SEMESTRAL I EXAMINATION

Introduction to Sociology and Sociometry

Note: Answer any five questions with at least one question from each group. All the questions carry equal marks.

Date: 7.11.1997 Maximum Marks: 100 Time: 3 hours

QUESTIONS

Group - A

1. Do you consider it important to study Comte's ideas in studying sociology? Discuss two reasons in support of your statement.

2. How does Durkheim explain suicide in a society? Indicate if any other explanation of suicide is possible using Marx's or Weber's idea.

Group - B

3. What is social stratification? What are the two main attributes of social stratification of Indian society? Discuss their differences.

4. State the major substantive approaches in academic sociology. Which one will you prefer to study social change in India and why?

Group - C

5. What is a family? Define kinship and discuss how does it differ from a family.

6. Discuss briefly whether it is important to study family and kinship in studying society in India?

contd. .....?.

Group - D

7. Suppose you want to study change in rural culture like, whether pattern of marriage is changing among the rural youth, whether way of life of a tribal group has changed after migration to a village, etc.

Discuss, mentioning your topic how you will plan your study specially referring to the following:

(a) two alternative hypotheses you will like to use;

(b) how you propose to collect your data, from whom, and prepare a list of a few questions to illustrate what you will ask to gather your data; and

(c) how you will test which of the two alternative explanations is more acceptable according to data.

8. A group consists of 10 (ten) friends, namely A, B, C, D, E, F, W, X, Y, Z. Usually A invites anybody except Z on any occasion, but D invites B, C, F and Y. E invites C and W but not A. F invites all others. W invites X, Y, and Z. Y and Z do not invite any body.

How will you measure reciprocity within this group? Obtain the value of this measure and discuss whether this group is highly reciprocative.

9. Define participant - observation and structured - questionnaire techniques of data collection. Which one of these two will be best applicable for studying:

(a) evaluation of impact of an irrigation project;

(b) studying inter-caste conflict and tension and,

(c) estimating percentage of literates in a city.

INDIAN STATISTICAL INSTITUTE
B.STAT.(HONS.) III YEAR:1997-98
SEMESTRAL-I EXAMINATION
SAMPLE SURVEYS

Date:10.11.97 Maximum Marks:100 Time: 3 Hours

Note: Answer all questions

1.(a) Distinguish between sampling and non-sampling errors in a sample survey. List down the various sources of non-sampling errors at the field operations stage of a survey.

(b) If from a simple random sample of n units drawn without replacement, a random sub-sample of n' units is selected without replacement, duplicated and added to the original sample, prove that the mean t' based on the (n+n') units is an unbiased estimator of the population mean. Also show that the variance of t' is greater than the variance of the mean based on the original sample of n units by an approximate factor $(1+3n'n^{-1})(1+n'n^{-1})^{-2}$. For what value of n'/n does the relative loss in efficiency attain its maximum value?

(4+4)+(3+6+3)=(20)

2.(a) What do you understand by the term 'Intra class correlation coefficient'? A population consists of 14 clusters of size 6 each. Find the bounds for the intracluster correlation coefficient among the elements of the cluster.

(b) Show that the relative efficiency of sampling a cluster of M units, compared to direct samoling is given by $1/\{1+(M-1)\rho\}$, where the intra class correlation coefficient, $\rho=1-M\sigma_w^2/(M-1)\sigma^2$, σ_w^2 and σ^2 denoting the wittin and total variances respectively.

(c) Suppose that n clusters are selected at random and without replacement from a population of N clusters of sizes $M_i, i=1,2,\ldots N$. Let $\bar{Y}_i=\sum_{j=1}^{M_i} Y_{ij}/M_i$, be the i th cluster mean, where Y_{ij} is the value taken by the study variable y on the j th unit of the i th cluster, $j=1,2,\ldots,M_i$, $i=1,2,\ldots N$.

Write down the conventional unbiased estimator for $\bar{Y}=\sum_1^N M_i\bar{Y}_i \mid \sum_1^N M_i$, and its variance. (3+2)+6+(2+3) =(16)

3.(a) Explain what you understand by 'combined and separate regression estimators' in stratified sampling. Compare the two estimators based on the (large sample) variance criterion.

contd. .....2.

3.(b) Describe the role of multistage sampling as a compromise between direct sampling and cluster sampling.

(c) Develop the Midzuno-Sen sampling scheme to make the estimator $t = \sum_{i=1}^{n} y_i \Big/ \sum_{i=1}^{n} x_i$ unbiased for the ratio $R = \sum_{1}^{N} Y_i \Big/ \sum_{1}^{N} X_i$. Obtain the probability of inclusion of i th unit π_i, for this scheme.

(4+8)+5+(3+4)=(24)

4.(a) What are the advantages of stratified sampling?

(b) The table below presents the summary data for all the 112 villages in a certain tehsil of M.P. The villages are stratified by size of their agricultural area into 3 strata as shown in col.(2) of the table.

The number N_i of villages in different strata are given in col.(3). Col.(4) gives the number of villages n_i selected from the i th stratum and in col.(5) the sampling design used in the i th stratum is described. The yield of jute for the selected villages and the auxiliary information where available are given in col.(6) of the table:

Stratum (1)	Sizes of the village in acres of agr. area (2)	N_i (3)	n_i (4)	Sampling Design (5)	Yield of jute y_{ij} (6)	Auxiliary information (agr. area) x_{ij}
I	≤40	51	6	Probability Proportional to agricultural area with replacement	75 101 5 78 79 45	25 39 2 26 26 16
II	41-80	38	4	Circular systematic sampling	indep. subsample 1 (y_{ij}) 247 238 359 125	indep. subsample2 (y_{ij}) 256 214 368 141
III	>80	23	2	Simple random sampling without replacement	y_{ij} 427 326	

It is also known that the total agricultural area of villages in the first stratum is 1968 acres.

(i) Estimate the average yield of jute in the tehsil.

(ii) Obtain an estimate of the variance of your estimate in (i) above.

contd. .....3.

4.(c) An experienced teacher makes a guess of the scores x_i in an examination for each of the 200 students of his class based on past performance. He obtains an average score of $\bar{X}=58$. For a simple random sample (without replacement) of 10 students the following results are obtained after an examination:

	Student									
Actual score	1	2	3	4	5	6	7	8	9	10
y_i	61	42	50	58	67	45	39	57	71	53
Guessed score x_i	59	47	52	60	67	48	44	58	76	58

Find the regression estimate $\hat{\bar{Y}}$ of the average score $\bar{Y}$ for the class and estimate its sampling error. 4+(8+14)+(6+8)=(40)

INDIAN STATISTICAL INSTITUTE

B.Stat. (Hons.) III Year : 1997 - 98

SEMESTRAL-I EXAMINATION

Linear Statistical Models

Date: 13.11.1997 Maximum Marks: 100 Time: 3 hours

Note: Attempt each question.

1.(a) Define an inner product space.

(b) Let $\mathcal{V} = \mathbb{R}^n$ be the usual Euclidean space. For any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, define $\| x \|_p = [\sum_{i=1}^{n} |x_i|^p]^{1/p}$ for $p \geq 1$. Show that $\| x \|_p$ is a norm on $\mathbb{R}^n$ for each $p \geq 1$. Moreover, $\| x \|_p$ is generated by an inner product $\langle \cdot, \cdot \rangle_p$ iff $p = 2$.

(c) Let $\mathcal{V} = \mathbb{R}^n$ be as above equipped with the standard inner product $\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$. Let $P: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be another inner product which satisfies $P(x, y) = 0$ iff $\langle x, y \rangle = 0$. Show that $P(x, y) = \sum_{i=1}^{n} \lambda_i x_i y_i$ for some $\lambda_1, \ldots, \lambda_n > 0$.

(5+10+7) = [22]

2. Let $Y \sim N_n(\mu, \Sigma)$ where Σ is positive definite. Let C be a $n \times 1$ vector and A $(n \times n)$ be symmetric. Show that

(a) if $A \Sigma A = A$ then A is nonnegative definite.

(b) Assume A is n.n.d. Then, $C'Y$ and $Y'AY$ are independent if $C' \Sigma A = 0$.

(6+9) = [15]

3. Consider a two-way layout with interactions. Prove that

(a) If the interactions $\{\gamma_{ij}\}$ are all zero for some system of weights $\{v_i, i = 1, 2, \ldots, I\}$, $\{w_j, j = 1, \ldots, J\}$ $(\sum_{1}^{I} v_i = 0, \sum_{1}^{J} w_j = 0)$; then they are zero for every system of weights.

(b) In that case, every contrast in the main effects $\{\alpha_i\}$ or $\{\beta_j\}$ has a value that does not depend on the weights $\{v_i\}$, $\{w_j\}$.

(10+7) = [17]

4. Consider the coordinate free version of the general linear model, i.e., a set of observations Y $(n \times 1)$ follows a linear model if, $E(Y) = \mu \in V$, a linear subspace of $\mathbb{R}^n$ and $Cov(Y) = \sigma^2 I_n$ (I_n is the $n \times n$ identity matrix). Call the model special linear model if further it is assumed that $Y \sim N_n(\mu, \sigma^2 I_n)$. Let $\hat{\mu}$ be the OLS estimate of μ. Then prove that $a'\hat{\mu}$ is the UMVUE for $a'\mu$ in the special linear model. Hence show that $a'\hat{\mu}$ is the BLUE in the general model.

(10+6) = [16]

5. Consider Tukey's non-linear two-way interaction model,

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma \alpha_i \beta_j + e_{ij} \text{ where}$$

(1) e_{ij}'s are i.i.d $N(0, \sigma^2)$.

(2) $\Sigma\alpha_i = \Sigma\beta_j = 0$, $i = 1, 2, \ldots, I$; $J = 1, 2, \ldots, J$.

(a) Suggest an algorithm for obtaining the LS estimate of γ.

(b) Obtain approximate expression for $E(\gamma_{OLS})$ and $Var(\gamma_{OLS})$ when $\gamma = 0$, I and J are sufficiently large.

(5+7) = [12]

6. A Nuclear Power agency wants to buy appropriate radioactive reagent for its plant. There are two prospective suppliers A and B. The quality of radio active reagents are considered same if their decay rates are equal. In order to test whether the quality supplied by A and B are same the agency chooses four samples of the radioactive reagent (2 from A and 2 from B) and let them decay for some hours. It is assumed that the decay follows exponential model, $y_t = A e^{-\lambda t + \varepsilon_t}$, where A is the initial amount, λ is the decay constant and $\{\varepsilon_t\}$ is a random noise, with $N(0, \sigma^2)$ distribution, with $\varepsilon_0 = 0$. The following data was obtained:

		Initial (unit)	After t hrs. (fraction of initial weight)
Supplier A	Sample I	1	0.8
	Sample II	1	0.7
Supplier B	Sample I	1	0.75
	Sample II	1	0.9

bcc: Perform an appropriate ANOVA test (preparing the ANOVA table) for this purpose. [18]

INDIAN STATISTICAL INSTITUTE
B.STAT.(HONS. III YEAR:1997-98
SEMESTRAL-I EXAMINATION
DIFFERENTIAL EQUATIONS

Date:18.11.97 Maximum Marks:100 Time: 3 Hours

Note: Answer any five questions.

1.(a) A non-negative continuous function f is defined on the closed interval $[0,1]$ with $f(1)=0$. For each a, $0<a<1$, the line $x=a$ cuts the ordinate set of f into two regions of areas A and B respectively, the area of the leftmost region being A. If $A-B=2f(a)+3a+b$ where b is a constant independent of a, find the function f and the constant b.

(b) Solve the following system of differential equations:

$$\frac{dx}{dt} = -4x - y+t^2$$

$$\frac{dy}{dt} = x-2y+t \qquad (10+10)$$

2.(a) Find the general solution of the differential equation $x^2 y'' + 3x y' + 4y = \frac{1}{x}$ on the interval $x > 0$ given that the corresponding homogenous equation has a particular solution of the form $y = x^m$.

(b) The graph of a nontrivial solution u of $y'' - 3y' - 4y = 0$ intersects the graph of a nontrivial solution v of $y'' + 4y' - 5y = 0$ at the origin. Determine u and v if they have equal slopes at the origin and if $\lim_{x \to +\infty} \frac{v(x)}{u(x)} = \frac{5}{6}$. (10+10)

3. Let $P_n(x)$ be Legendre's polynomial of degree n. Show that for any polynomial $q(x)$ of degree less than n, $\int_{-1}^{1} q(x)\, P_n(x)\, dx = 0$. Deduce from this that $P_n(x)$ has no multiple zeros and all the zeros of $P_n(x)$ lie in the interval $[-1, 1]$. Show that $\int_{-1}^{1} P_n^2(x)\, dx = \frac{2}{2n+1}$ and hence determine the coefficient a_n in the sum $x^{n+1} = \sum_{k=0}^{n+1} a_k P_k(x)$. [20]

contd. .....2.

4.(a) Find a nontrivial polynomial solution of Hermite's equation

$$\frac{d}{dx}\left[e^{-x^2}\ \frac{dy}{dx}\right] + \lambda\, e^{-x^2}\, y = 0$$

in $(-\infty,\infty)$ for $\lambda = 4$.

(b) Let $v(x)$ be the solution of the following initial value problem:

$ay'' + by' + cy = 0,\ y(0) = 0;\ y'(0) = \frac{1}{a},\ a \neq 0$

Using methods of Laplace transforms show that the solution of the initial value problem:

$ay'' + by' + cy = f(x);\ y(0) = 0;\ y'(0) = 0$

is the convolution $f(x) * v(x)$ of $f(x)$ and $v(x)$. (8+12)

5.(a) Determine the curve passing through $(0,0)$ and $(1,0)$ so that the length of the arc of the curve between $(0,0)$ and $(1,0)$ is minimum and the area of the region enclosed between the x-axis and the arc of curve between $(0,0)$ and $(1,0)$ is $\frac{\pi}{8}$.

(b) Find the extremals for the integral

$$\int y^{1/2} \sqrt{1+(y')^2}\ dx. \qquad (15+15)$$

6. Show that the number of positive zeros of any nontrivial solution of Bessel's equation of order p is infinite. Also show that the difference between two successive zeros of such a solution is less than π if $0 \leq p < \frac{1}{2}$ and greater than π if $p > \frac{1}{2}$. Finally show that the difference between two successive zeros of any nontrivial solution of Bessel's equation of order p tends to π as $x \longrightarrow +\infty$. (20)

INDIAN STATISTICAL INSTITUTE
B.STAT.(HONS.)III YEAR:1997-98
DESIGN OF EXPERIMENTS
SEMESTRAL-II EXAMINATION

Date:27.4.98 Maximum Marks:100 Time: $3\frac{1}{2}$ Hours

Note: Answer Question No.1 and any two questions from the rest. Marks alloted to a question are indicated in brackets [ ] at the end.

1. The following data represent a single replicate of a 2^5 factorial design in blocks of 8 plots each in an experiment to study the tensile strength of rubber. The factors are mix (A), time (B), laboratory (C), temperature (D) and pressure (E). Find out the effects confounded in this design. Take one of them to demonstrate that it is non-estimable. Analyse the data to test significance of the effects of interest, assuming three-factor and higher order interactions are negligible.

Block-1		Block-2		Block-3		Block-4	
Treatments	Yield	Treatments	Yield	Treatments	Yield	Treatments	Yield
(1)	7	be	35	b	34	abce	55
abc	50	a	9	ac	10	bde	40
ce	5	ace	12	bce	25	acde	20
bd	30	bc	30	ae	12	ab	55
ade	15	cd	8	d	10	c	6
acd	11	abcde	65	cde	15	ad	11
abe	62	de	10	abcd	60	e	8
bcde	34	abd	61	abde	65	bcd	34

[8+22=30]

2.(a) Show that at most $(s-1)$ mutually orthogonal latin squares (MOLS) of order s exist. When s is a prime power, give a method of construction of a complete set of MOLS of order s, with a proof that the method works in general.

(b) Describe in details the B-, S-, and the T- methods of multiple comparisons, providing proofs of all the probability statements made. [(2x7)+(3x7) =35]

3.(a) Explain error control and discuss its need. Describe two methods to achieve such control.

(b) Describe the missing plot technique. Prove that it leads to the valid error sum of squares and a valid least square estimator of the parameter vector. Apply the technique to provide expressions for "estimates" of two missing observations, and the expressions for variances of elementary treatment comparisons in a randomised block design for v treatments and b blocks with observations in block $j(1\leq j\leq b)$, and under treatments i and i' $(i\neq i';1\leq i,i'\leq v)$ missing. [10+(5+10+10)=35]

p.t.o.

4.(a) Give a balanced confounding scheme for a 2^5 factorial experiment in blocks of 2^3 plots each. Construct all the key blocks of this balanced confounded design, and describe how the other blocks can be generated, illustrating with respect to one replication only. Give the analysis of variance of this design, explaining clearly how the various sums of squares are to be computed.

(b) Show that the maximum number of factors that can be accommodated in a full factorial experiment with blocks of 2^k plots without confounding main effects, 2-factor, and 3-factor interactions is 2^{k-1}. [(5+8+12)+10=35]

INDIAN STATISTICAL INSTITUTE
B.STAT.(HONS.) III YEAR: 1997-98
SEMESTRAL-II EXAMINATION
STATISTICS COMPREHENSIVE

Date:30.4.98 Maximum Marks:100 Time: $3\frac{1}{2}$ Hours

Note: Use separate Answer Scripts for Group A and Group B.

GROUP-A

Answer any three questions.

1. X_1, X_2, X_3 and Z are iid Bernoulli with $P_r[0]=p$ and $P_r[1]=q=1-p$.

Let $Y = \begin{cases} X_1 & \text{if } X_3 = 1 \\ X_2 & \text{if } X_3 = 0 \end{cases}$

and $C = Y + Z \pmod 2$.

Compute $P_r[\, C = X_i \,]$, i=1,2,3. [4+4+4=12]

2. Given a random binary sequence, how would you select a person from a group of 19 persons with equal probability? Describe your method clearly with a proof. [7+5=12]

3. Huffman's algorithm for coding may produce different coding schemes (all schemes minimize the expected length of the message). Give an example to illustrate this. Of such available schemes, which one would you prefer? Explain the statistical criterion that you would use. [5+7=12]

4. Model the "Blood Stock" situation in a Blood Bank as a Markov Chain, after making suitable assumptions (with proper justification) and then compute the transition probability matrix. [$3\frac{1}{2}+3\frac{1}{2}+5=12$]

GROUP-B

Answer any four questions.

1. An urn contains "a" white and "b" black balls. Balls are drawn one by one till exactly "s" white balls are observed. Find expected number of balls to be drawn under both with and without replacement scheme of sampling. [6+7=13]

2.(a) In repeated Bernoulli trials with success prob. p, generate an event with probability $8p^2q^2$.

(b) In repeated Bernoulli trials with success prob. p, consider the following game:

Continue the trial till 4 successes are observed. If, however, the number of trials in doing so exceeds 11, then 5 successes (instead of 4) are to be observed.

(i) Display the above game in the form of a random walk.

(ii) Suggest unbiased estimates of p and p^{-1} when the game actually ends in 13 trials.]5+8=13]

p.t.o.

3. Suppose $X \sim N(0,1)$ under H_o while $X \sim$ Double exponential $(f(x) = \frac{1}{2} e^{-|x|}, -\infty < x < \infty)$. Carry out a level-$\alpha$ MP test based on a single observation and work out an expression for the power. Verify that the test is unbiased. [4+5+4=13]

4. (a) In normal samples, work out an expression for the large sample s.e. of the sample coeff. of variation (v) in terms of the population coefficient of variation (V).

Also indicate a procedure for testing $H_o: V = V_o$ vs $H_1: V \neq V_o$ in large samples.

(b) What is $\sin^{-1}\sqrt{p}$ - transformation? How is it useful in testing homogeneity of several binomial proportions? [(5+2)+(2+4)=13]

5. (a) $(X_1, X_2, \ldots, X_k)$ have a k-variate non-singular joint distribution with mean vector $\underset{\sim}{\mu}$ and dispersion matrix Σ. Under linear regression, examine the nature of the regression coefficients when $\Sigma = \underset{\sim}{\mu}\,\underset{\sim}{\mu}' + D_{iag}(\pi_1, \pi_2, \ldots, \pi_k)$, $\pi_i > 0\ \forall\, i$. Comment on the result.

(b) In a bivariate Dirichlet distribution, show that the regression of one on the other is linear. [8+5=13]

Neatness - 2
Home Work Assignments - 10.

INDIAN STATISTICAL INSTITUTE
B.STAT.(HONS.) III YEAR: 1997-98
SEMESTRAL-II EXAMINATION
INTRODUCTION TO STOCHASTIC PROCESSES

Date:4.5.98 Maximum Marks:140 Time: 4 Hours

Note: This paper carrics problems worth a total of 170 pts. Answer as many as you can. Maximum you may score is 140 pts.

1. Consider a markov chain with 15 possible states. Show that if x and y are two distinct states such that x leads to y, then one must have $p_{xy}^{(n)} > 0$ for some $n \leq 14$. [10]

2. Consider the markov chain on $\mathcal{S} = \{1,2,3,4,5,6\}$ with transition matrix

$$\begin{pmatrix} 0 & 0 & 1/2 & 0 & 1/4 & 1/4 \\ 0 & 1/3 & 0 & 2/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 2/3 & 1/3 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 1/4 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(i) Compute $p_{16}^{(3)}$

(ii) Find $P_1(T_6=3)$

(iii) Classify the states as recurrent and transient, with reasons

(iv) Find ρ_{52}

(v) Find the stationary distribution (s) for the chain.

(vi) Find $\lim_{n\to\infty} \frac{1}{n} E(N_2^n | X_0=5)$, where N_2^n = no. of visits to state 2 upto time n. [10x6]

3. Prove that an irreducible positive recurrent markov chain is time-reversible (in stationary state) if and only if, for every $k \geq 1$ and every choice of states $x_1,\ldots,x_k$,

$$p_{x_1x_2} \cdot p_{x_2x_3} \cdot \ldots \cdot p_{x_{k-1}x_k} \cdot p_{x_kx_1} = p_{x_1x_k} \cdot p_{x_kx_{k-1}} \cdot \ldots \cdot p_{x_3x_2} \cdot p_{x_2x_1}$$

[10]

4. Consider the markov chain on $\mathcal{S} = \{0,1,2,\ldots\}$ with transition probabilities given by

$$p_{x,x+1} = \left(\frac{x+1}{x+2}\right)^{\alpha} = 1 - p_{x,0}, \text{ for all } x \in \mathcal{S}$$

(where α is some positive real).

p.t.o.

4.(i) Find $P_0(T_0 > n)$.

(ii) Use (i) to find values of α for which the chain is transient/ null recurrent/positive recurrent.

(iii) Show that $\lim_{n\to\infty} p_{xy}^{(n)}$ exists for all x,y and find the limits. [10x3]

5. Show that if $\{X_t, t \geq 0\}$ is a Poisson process with rate λ, then, for any $t_0 > 0$, the process $\{Y_t, t \geq 0\}$ defined as $Y_t = X_{t_0+t} - X_{t_0}$, is a Poisson process with rate λ. Hence find the distribution of

$$S_{X_{t_0}+1} - t_0.$$ [10]

6. Let $\{X_t, t \geq 0\}$ be a Poisson process with rate λ.

(i) Find $P(S_2 > 2S_1+1)$.

(ii) Find $E(S_1/X_1=3)$

(iii) Let $Z = t - S_{X_t}$. Find the distribution of Z.

(Try to find $P(Z > x)$ for $x < t$).

(iv) For any $t > 0$, find the expected value of the time between the last event prior to time t and the first event after time t. (Use 0.5 and (iii) above). [4x10]

7. Show that if $\{Z_t, 0 \leq t \leq 1\}$ is a Gaussian process with zero mean and covariance function

$$C(s,t) = s(1-t) \text{ for } 0 \leq s \leq t \leq 1$$

then $X_t = \{(t+1)\, Z_{t/(t+1)}, t \geq 0\}$ is a standard BM. [10]

INDIAN STATISTICAL INSTITUTE

ECONOMICS - IV (B. Stat. 3rd Year, 1997 - 98)

Time : 3 Hours | Semestral Examination | Date : 8.5.98

The paper carries 66 marks. Attempt ALL questions. The maximum you can score is 60.

1. Briefly describe any two solutions to the problem of multicollinearity in a linear model. [6]

2. Explain the consequence of excluding a relevant variable from the set of two regressors in a linear model. [4]

3. You are given data on the following variables:

   H_g = per-capita gross rate of nonfarm residential construction (in constant Rs.).

   H = end-of year par-capita nonfarm housing stock (in constant Rs.)

   y_p = permanent income

   p = index of residential construction costs (in real terms)

   R = index of rent (in real terms)

   r = average quality of new dwellings

   N = average size of households

   m = index of relative building material prices

   U = relative wages of unskilled construction workers

   s = ratio of wages of skilled to unskilled workers.

   The following equations were fitted with the data on the above variables. Interpret each as depicting the demand factors or supply factors or both and discuss whether the equation makes sense and is a 'good' equation.

$$H_g = -2.49p + .438y_p - 8.34r - .282H \qquad R^2 = .621$$
$$(.589)\quad (.092)\quad (4.47)\quad (.07)$$

$$p = -.043H_g - .034m + 21.6U + 54.9s \qquad R^2 = .804$$
$$(.053)\quad (.124)\quad (4.3)\quad (13.7)$$

$$\frac{H_G}{H} = .255\frac{R}{p} - .00608r \qquad R^2 = .714$$
$$(.046)\quad (.00451)$$

$$H = -59.1p + 4.09y_p - 510r + 385N \qquad R^2 = .723$$
$$(17.7)\quad (.62)\quad (146)\quad (802)$$

[10]

4. Suppose the true equation is $Y = \alpha_0 + \alpha_1 X + \alpha_2 X^2 + u$. Suppose we estimate instead the equation $Y = \beta_0 + \beta_1 X + v$ by OLS. Express $E(\widehat{\beta_1})$ in terms of α_0, α_1 and α_2. [6]

5. An investigator tried to determine the relationship between true property values and assessed values. For this, he regressed the 'ratio of assessed value to sales price' on sales price and found a negative slope. Why might this procedure give misleading results? How should you analyze the basic data to avoid this difficulty? Under what circumstances, if any, could you defend the analysis actually made? [8]

6. Suppose you have apriori convictions about the algebraic sign of certain parameters in an econometric equation. How would you estimate the parameters subject to the restrictions that the parameters in question have the signs that you believe to be correct? (Take a simple 'regressing y on x' situation.) [6]

7. (a) Define rank and order condition for identifiability in a simultaneous equation model.

(b) Discuss the identifiability of the parameters in the following system:

$$\begin{aligned} C &= \alpha_0 + \alpha_1 W + \alpha_2 P + U_1 \\ I &= \beta_0 + \beta_1 P + \beta_2 k_{-1} + U_2 \\ W &= \gamma_0 + \gamma_1 Y_{-1} + U_3 \end{aligned}$$

where $C + I = Y$, $P + W = Y$, $I = k - k_{-1}$ (after eliminating any of the three endogeneous variables). C, I, W, P, Y, k are endogeneous.

(c) What is I.V. method of estimation in a linear model?

(d) Suggest estimation procedure for (b) using the I.V. method.

[6 + 7 + 7 + 6 = 26]

INDIAN STATISTICAL INSTITUTE
B.STAT.(HONS.) III YEAR: 1997-98
SEMESTRAL-II EXAMINATION
STATISTICAL INFERENCE-II

Date:6.5.98 Maximum Marks:100 Time: 3 Hours

Note: Answer any four questions from Group A, and answer any three questions from Group B.

GROUP-A

Answer any four questions. [13x4=52]

1. Let $X_1, X_2, \ldots X_n$ be iid $N(\mu,2)$. Find the minimum length 95% confidence interval for μ using

$$T(X_1, X_2, \ldots, X_n; \mu) = (\bar{X} - \mu)$$

2. Using Stein's two-stage procedure give essential steps for finding a $(1-\alpha)100\%$ confidence interval of fixed length L for μ of a $N(\mu,\sigma^2)$ population.

3. Clearly stating regularity conditions, if any, show that

Let $\lim_{n\to\infty} P[N > n]=0$, where N is the step at which the sampling process is stopped in an SPRT procedure.

4. Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu,\sigma^2)$. Find a family of uniformly most accurate unbiased $(1-\alpha)100\%$ confidence interval for μ.

5. Let $X_1, X_2, \ldots, X_n$ be iid as X. Compute the expected value of $P[X_{(r)} \leq X \leq X_{(s)}]$. If 3 observations are to be drawn from a N(0,1) population, what do you expect the numerical values of the observations be?

6. Let $X_1, X_2, \ldots, X_n$ be iid as X, and $Y_1, Y_2, \ldots, Y_m$ be iid as Y. Compute the variance of $U +(= \#(X > Y))$.

GROUP-B

Answer any three questions. [16x3=48]

1. Determine the size of a sample to be drawn from a $N(\mu,4)$ population for testing H_o: $\mu=12$, against H_1: $\mu=8$ with the probabilities of Type I error and Type II error being $\alpha=0.05$, $\beta=0.01$ respectively.

2. Simulate data from Bernoalli ($\theta=0.3$). Use an SPRT for testing H_o: $\theta=0.2$ against H_1: $\theta=0.5$ when the two error rates are $\alpha = \beta = 0.05$.

p.t.o.

3. Find a confidence inteval with confidence level 0.9 for the 60th percentile of the distribution; the data are

   38.3, 35.6, 41.2, 39.9, 43.9, 40.5,
   38.6, 36.7, 38.1, 46.8.

4. A laboratory technician proposes to reduce costs by feeding experimental rats a diet that costs much less than the diet currently being used. The economy move will be made if the technician can show that there is no weight loss compared to the standard diet. Twenty-four rats were used in the experiment and assigned at random to either the new diet or the standard diet. The gains in weight were measured in grams

   Standard diet: 22.51,19.14,11.43,16.83,23.55,22.30
   19.73, 21.22,16.05,20.46,25.32,17.53

   New diet : 11.48,13.73,20.32,14.76,11.86,17.70,
   14.53,12.36,14.53,13.07,22.03,13.63.

   How does the new diet compare with the standard diet? State clearly your H_0 and H_1 and test using Wilcoxon-Mann-Whitney test procedure.

5. The following are the correlations computed on the basis of 25 observations on x_1, x_2, and x_3

	x_1	x_2	x_3
x_1	1	0.365	0.526
x_2		1	0.243
x_3			1

   Obtain a 95% confidence inteval for $\rho_{12.3}$.