INDIAN STATISTICAL INSTITUTE

B III Semester I

Statistical Inference I

Backpaper Examination

Date: 13.3.03

Total Points 100

Clearly explain your assumptions and notations used so that all answers are self contained and to the point

1. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample of size n from a $N(\theta, 1)$ distribution.

   (a) Argue that there does not exist any unbiased estimator of e^θ whose variance attains the k-th Bhattacharya lower bound for any k. [6]

   (b) Find the minimum variance unbiased estimator of e^θ. [9]

2. (a) Define CAN and BAN estimators. [3+3]

   (b) Show that under the usual regularity conditions MLE is a BAN estimator. [10]

3. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample of size n from a $Bin(1, \pi)$ distribution.

   (a) Show that no UMP test exists for testing $H_0 : \pi = \pi_0$ againast $H_1 : \pi \neq \pi_0$. [5]

   (b) Find a UMP unbiased test for $H_0 : \pi = \pi_0$ against $H_1 : \pi \neq \pi_0$. [10]

4. Let θ be a real parameter and suppose that the joint density $L(\boldsymbol{X}, \theta)$ of $\boldsymbol{X}$ has monotone likelihood ratio in $T(\boldsymbol{X})$. Show that $T(\boldsymbol{X})$ is a minimal sufficient statistic. [10]

5. Show that if $h(u)$ is a strictly convex function and $\widehat{\theta}$ is an unbiased estimator of θ which is not degenerate, the $h(\widehat{\theta})$ cannot be an unbiased estimator of $h(\theta)$. [1

6. Let X_i be distributed independently as $N(i\Delta, 1), i = 1, \ldots, n$. Find a UMP test for $H_0 : \Delta \leq 0$ against $H_1 : \Delta > 0$. [1

7. Assume that the parameter space in a decision problem is finite, say $\Theta = \{\theta_1, \ldots, \theta_n\}$. Suppose that δ^π is the Bayes rule with respect to a prior distribution π that gives positive probability to every possible value of θ. Show that δ^π is admissible. [10]

8. Suppose δ is a rule with a constant risk function and δ is Bayes with respect to some prior π. Then prove that δ is minimax. [12]

INDIAN STATISTICAL INSTITUTE

First Semestral Backpaper Examination: 2002-2003
B.Stat. (Hons.). 3rd Year
Linear Statistical Models

Date: 20.3.03 **Maximum Marks: 100** **Duration: 3 hours**

- This question-paper carries 110 marks. Answer as much as you can, but maximum you can score is 100.
- You must state clearly any result stated and proved in the class, you may need in order to answer a particular question. Keep the answers brief and to the point.

---

1 (a) Let $\mathcal{V}$ be a subspace of $I\!R^n$. Let X_1 and X_2 be two basis matrices of $\mathcal{V}$. Show that $X_1(X_1'X_1)^{-1}X_1' = X_2(X_2'X_2)^{-1}X_2'$.

(b) Show that for any two matrices X and C (of compatible dimensions) $X'XC = 0$ if and only if $XC = 0$.

(c) Let U be an $n \times r$ matrix. Define a g-inverse of U. Show that any matrix has infinitely many g-inverses. [10+8+12= 30]

2 Let $Y = \underline{\mu} + \epsilon$, $\underline{\mu} \in \mathcal{V} \subseteq I\!R^n$ be a linear model with normally distributed iid errors with variance σ^2(standard notation). Let $\mathcal{W} \subseteq \mathcal{V}$. Let

$$F = \frac{||P_{\mathcal{V}|\mathcal{W}}Y||^2\,(n-p)}{||(I-P_{\mathcal{V}})Y||^2\,(p-r)}$$

where r and p denote the dimensions of $\mathcal{W}$ and $\mathcal{V}$ respectively. Show that F has a $F_{p-r,n-p}(||P_{\mathcal{V}|\mathcal{W}}\underline{\mu}||^2/\sigma^2)$ distribution. [15]

3 (a) Consider a set of experiments given by X. For any experiment $(x \in X)$ the response is given by $y_x = \phi(x) + \epsilon$, where ϵ is normally distributed with mean 0 and unknown variance σ^2. Suppose n such independent experiments are performed at levels $d = (x_1, \ldots, x_n) \in X$. Derive the set of all estimable linear functionals of the regresison function ϕ under the given design d.

(b) Write down the additive model for the response with three treatments under a Latin square design. Obtain the formulae for relevant sums of squares under this design and hence write down the ANOVA table for Latin square designs. [15+15= 30]

P. T. O

4. A consumer organization studied the effect of age of automobile owners on size of cash offers for a used car by utilizing 6 persons (3 male and 3 female) in each of three age groups (young, middle, old) who acted as the owner of a used car. A medium price car was chosen for the experiment and the 'owners' solicited cash-offers from 18 randomly selected auto-dealers. The data is given by

Young (M) : 21, 23, 19
Young (F) : 21, 22, 21
Middle (M) : 30, 29, 26
Middle (F): 26, 29, 27
Old (M) : 25, 22, 23
Old (F) : 23, 19, 20

(a) Obtain the fitted values and residuals.
(b) Set up the ANOVA table. Does any one factor account for most of the total variability in cash offers in the study. Explain.
(c) Test whether or not interaction effects are present; use $\alpha = 0.05$. What is the P-value of the test? [10+20+5= 35]

INDIAN STATISTICAL INSTITUTE

Second Semester Backpaper Examination : 2002-03

B. Stat. III Year

Introduction to Stochastic Processes

Date : 29.07.03 Full Marks : 100 Duration : 3 hours

Justify your steps clearly.

1. $(X_n)_{n\geq 0}$ is a Markov Chain with state space S and transition matrix (p_{ij}).

(a) If a state i is recurrent and i communicates with j then show that j is recurrent.

(b) If a state i has period 3 and i communicates with j then show that j has period 3.

(c) If i is recurrent then show that

$$P(X_n = i \text{ for infinitely many } n \mid X_0 = i) = 1.$$

(5+5+10)

2. Explain 'simple symmetric three dimensional Random Walk' and show that it is transient. (20)

3. Classify the states of the following chain. For those which are recurrent, calculate their period.

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

(8)

[P.T.O.]

4. Consider a renewal process N_t with interarrival distribution having mean μ. State and prove a result concerning the limiting behaviour of $\frac{N_t}{t}$ as $t \to \infty$.

(12)

5. Consider a Nonhomogeneous poisson process with intensity function $\lambda(t) = t$.

(a) What are the chances that upto time 10, no event occurred ?

(b) Given that upto time 10 exactly 2 events occurred, what are the chances that both occurred after time 5 ?

(c) If X is the number of events upto time 6 and Y is the number of events during the time period (4,10) find the joint distribution of (X,Y).

(5+5+10)

6. Consider a pure birth process with birth parameters $\lambda_n = n\lambda$.

(a) Write down the Kolmogorv Forward Equations.

(b) Show that for $1 \le i \le j$

$$P_{ij}(t) = \binom{j \; -1}{j \; -i} e^{-i\lambda t}\left(1 - e^{-\lambda t}\right)^{j-i}$$

(6+14)

***

INDIAN STATISTICAL INSTITUTE
B.Stat. III
Database Management System
Backpaper Examination

30.7.03 Time : 3 hours

Answer any part of any question , mamimum you can score is 100

esign a relational database for the library described below. Yo
esign must include,
i) an ER diagram, 10
ii) entity relations and relational relations in BCNF, 10
nd take care of the queries listed in part iii). 4 X 10 = 40

A library has a lending section and a reading room section. The lending section has a number of members each of whom is g: unique membership number. A member can have only one membershi

A member has to return a borrowed book within thirty days of ate of borrowing. For each subsequent day a fine of rupee one pe er book is imposed on the defaulter. He can borrow a maximum of t a time. He looses his membership if his subscription is overdu ore than 3 months, or, if his fines are not paid within 3 months f the total fine amount exceeds Rs.300 at any time.

The library has a reading room section which can be used by a embers on paying an extra subscription of Rs.20 per month,withir th of each month, and have to pay Rs.10 extra after that.

For each member the library keeps the neccessary information s membership number, name,address, month and year of joining, du f subscription payment, books issued,date of issue,date of retur

---

i) The database is to be accessed and used by the assistant lib o record return and issue books,defaulters of books and fines,

At any time he may like to know the following :

ormulate the following queries in relational algebra or in SQL

list of all members showing name and address by membership numb

a selected membership number showing list of books issued with of issue and due date of return for each book,

list of all members availing reading room facilities,

list of members who have not returned a book which is already d

list of members for whom a fine is due for more than 3 months,

list of members who are defaulters of reading room subscription

list of members whose annual subscription is due next month,

list of members showing reading room status,

a list of members whose fine amount is more than Rs.300,

a list of lending library members who are defaulters of their a: subscription for 3 months.

P.T.

Consider relations r1(X,Y,W), r2(Y,Z,S) and r3(S,T) with primary ke X,Y and S respectively. Assume that r1 has 1000 tuples, r2 has 50(tuples and r3 has 1500 tuples. Estimate the size of the natural jo of r1 and r2 and r3 . Give an efficient strategy for computing the join.

Write equivalent expressions for the following expressions tc improve the efficiency of corresponding queries for r1,r2 and r3.

```
sigma p ( r1 U r2 U r3 )   where p =( Z = 5 )and( W > 1 )and( T <

sigma p ( r1 - r2 )        where p =( Y = 8 )

sigma p ( r1 U r2 )        where p =( Y > 10 )

pie n ( sigma p ( r2 U r3 ) )  where p =( W > 1 )and( T < 3) and
                                     n = r2.S
pie n1 (  pie n2 ( sigma p    ( r1 U r2 U r3)))
                    where p = ( r1.Y = r2.Y - 2)and( r2.S = r3.S + 3
                    and   n2 = ( r1.Y,r2.Y,r2.S,r3.S)
                    and   n1 = ( r1.Y,r3.S)                 (10)
```

Consider that in the relations above there are no primary keys xcept the entire scheme. Let V(Y,r1) be 200, V(Y,r2) be 300, V(S,r2) e 100 and V(S,r3) be 400. Give an efficient strategy for computing th atural join of r1,r2 and r3. (6)

In the context of crash recovery of a transaction Ti explain

a) incremental log with deferred update
b) incremental log with immediate update (8)

Let values of three data items D,E and F be respectively 1500,2000 an 3250 respectively.

Consider the following transactions T1 and T2 and corresponding syste Log.

```
     T1                        System Log
read(D,d1)                   < T1,starts >
d1=d1-500
write(D,d1)                  < T1,D,1500,1000 >
read(E,e1)
e1=e1+500
write(E,e1)                    < T1,E,2000,2500 >
                               < T1,commits >
     T2
read(F,f1)                     < T2,starts >
f1=f1-50
write(F,f1)                    < T2,F,3250,3200 >
                               < T2,commits >
```

hat recovery action is needed when the system comes back if a crash occurs
just after the statement read(E,e1) is executed,
just after the log record < T2,starts > is written to stable storage
just after the log record for write(E,e1) has been written to stable storage,
just after the statement write(E,e1) has been executed,
just after the log record for write(F,f1) has been written to stable storage,
just after the statement write(F,f1) has been executed, (12

will be displayed by the transactions T1 and T2 if their execu is as per the following schedule assuming A and B have values 00 respectively

```
    T1                          T2
read(B)
                             read(B)
                             B=B-50
                             write(B)
                             read(A)
                             A=A+50
                             write(A)
read(A)
display(A+B)
                             display(A+B)                (4)
```

is a serializable schedule ? (4)
will happen if the following schedule is executed?

```
    T1                          T2
--------------------------------------------
  LX(B)
  read(B)
  B=B-1000
  write(B)
  UN(B)
                             LX(A)
                             read(A)
                             LS(B)                       (4)
  LX(A)
```

will be displayed if the following schedule is executed e A = B = 0 are the initial values ?

```
    T1                          T2
-----------------------------------------------
   LS(A)
   read(A)
   LX(B)
   read(B)
   if A = 0 then B = B=1
   UN(B)
   UN(A)
   DISPLAY(A+B)
                                LX(B)
                                read(B)
                                LS(A)
                                read(A)
                                If B=0 then A=A=1
                                UN(A)
                                UN(B)
                                DISPLAY(A+B)             (4)
```

ain what is locking. (4)

INDIAN STATISTICAL INSTITUTE
B.STAT-III (2002-03)
STATISTICAL INFERENCE-II (Back Paper)
Time:3 hours Max marks:100

Date:.,

Note: Answer all questions.

. Let $X_1, X_2, \cdots, X_n$ be iid $N(\theta, \sigma^2)$ random variables. Consider the problem of interval estimation of θ. A family of confidence intervals $\{(\underline{\theta}(\underset{\sim}{x}), \overline{\theta}(\underset{\sim}{x})) : \underset{\sim}{x} \in \mathbb{R}^n\}$ is said to be unbiased confidence interval for θ of level $1-\alpha$ if

$$P_{\theta,\sigma^2}[\ \underline{\theta}(\underset{\sim}{X}) < \theta < \overline{\theta}(\underset{\sim}{X})] \geq 1-\alpha \text{ for all } \theta, \sigma^2$$

and $P_{\theta',\sigma^2}[\ \underline{\theta}(\underset{\sim}{X}) < \theta < \overline{\theta}(\underset{\sim}{X}) \leq 1-\alpha$ for all θ, θ' with $\theta \neq \theta'$ and for all σ^2.

(a) Show that the family of intervals given by $\{ (\overline{x} - \frac{st_{\alpha/2}}{\sqrt{n}}, \overline{x} + \frac{st_{\alpha/2}}{\sqrt{n}}) \}$ is unbiased level $(1-\alpha)$, where $\overline{x} = \frac{1}{n}\sum_1^n x_i$ and $s = \sqrt{\frac{\sum_1^n (x_i - \overline{x})^2}{n-1}}$ and $t_{\alpha/2}$ is the upper $100\frac{\alpha}{2}\%$ point of Student's $t-$ distribution with $(n-1)$ degrees of freedom.

(b) Show that if $(\underline{\theta}(\underset{\sim}{x}), \overline{\theta}(\underset{\sim}{x}))$ is any other unbiased confidence interval of level $(1-\alpha)$ then

$$P_{\theta,\sigma^2}[\ \underline{\theta}(\underset{\sim}{X}) < \theta' < \overline{\theta}(\underset{\sim}{X})] \geq P_{\theta,\sigma^2}[\ \overline{X} - \frac{z_{\alpha/2}}{\sqrt{n}} < \theta' < \overline{X} + \frac{z_{\alpha/2}}{\sqrt{n}}] \text{ for all } \theta, \theta' \text{ with } \theta \neq \theta'.$$

(Such confidence intervals are called uniformly most accurate unbiased confidence intervals of level $(1-\alpha)$.)

[5+5]

. (a) In a business administration course, a set of lectures was given televised to one group and live to another. In each case an examination was given prior to the lectures and another immediately following them. The difference between the two examination scores for the subjects in the 2 groups were as follows:

Live: 20.3, 23.5, 21.9, 20.0, 26.6, − 9.4, 4.8, − 1.6, 25.0.
TV: 6.2, 15.4, 25.1, 4.6, 28.1, 17.2, 14.1, 31.2, 12.6, 23.4.

Use the two-sided Wilcoxon test to test the hypothesis of no difference at significance level α closest to 0.05 that is attainable by a non-randomised test.

(b) Consider an alternative with F and G being the distributuion of the **Live** and **TV** observations respectively for which

$P(X < Y) = 0.4$ where $X \sim F$ and $Y \sim G$, X and Y are independent.
$P(X < Y, X < Y') = 0.3$, $X \sim F$, $Y, Y' \sim G$, X, Y, Y' are independent
and $P(X < Y, X' < Y) = 0.3$, $X, X' \sim F$, $Y \sim G$, X, X', Y are independent.
Using Normal approximation find the power of the test used in part (a) against this alternative.

[10+10]

. (a) Define the Kolmogorov-Smirnov statistic $D_{m,n}$ for the 2-sample problem. Show that $D_{m,n}$ depends only on the rank of the observations in the combined sample.

P. T. O

(b) Find the exact null distribution of $D_{3,3}$ by complete enumeration.

(c) Suppose the observed arrangement of the $X's$ and $Y's$ when they are arranged in an increasing order is

$$XYXYYX$$

Compute the value of d of $D_{3,3}$ for this data. Find the significance probability of this data when Kolmogorov-Smirnov test is used.

[5+7+8]

4. (a) Define the Wilcoxon signed-rank sum statistic T^+ for the one sample problem.

(b) Find the exact null distribution of T^+ when the sample size $N = 4$.

(c) Show that $T^+ = \sum_i \sum_{j \geq i} U_{ij}$, where $U_{ij} = \begin{cases} 1 & \text{when } X_i + X_j > 0 \\ 0 & \text{otherwise.} \end{cases}$

(d) Let $A_{(1)} < A_{(2)} < \cdots\cdots < A_{(M)}$, be the ordered values of all the Walsh averages $\frac{X_i+X_j}{2}$, $1 \leq i \leq j \leq N$, $M = \frac{N(N+1)}{2}$.

Show that $A_{(i)} \leq a$ if and only if $T_a^+ \leq M - i$, (and hence $A_{(i)} > a$ iff $T_a^+ \geq M - i + 1$) where $T_a^+ \doteq \#\,(i,j) : i \leq j$ and $\frac{X_i+X_j}{2} > a$.

Explain how this fact can be used to get a confidence interval of level $(1-\alpha)$ for the point of symmetry δ of the distribution of X_i.

[2+5+5+8]

5. Let $X_1, X_2, \cdots,$ be *iid* random variables with Normal distribution with unknown mean μ and variance $\sigma^2 = 16$.

(a) Obtain the approximate boundaries of the SPRT for testing $H_0 : \mu = 0$ against $H_1 : \mu = 0.4$ when the type I and type II errors are $\alpha = 0.05$ and $\beta = 0.05$.

(b) When $\mu = 4$ find the approximate values of the OC and the ASN function of this SPRT.

[8+7]

6. (a) Let (X_i, Y_i), $i = 1, 2, \cdots, n$ be *iid* observations from a bivariate distribution for which the marginals of X and Y are continuous. Define the Spearman's rank correlation coefficient r. Describe how this can be used to test the null hypothesis of independence of X and Y.

(b) Find the null distribution of r by complete enumeration when $n = 3$.

[10+5]

****

TABLE B. Wilcoxon rank-sum distribution: $P(W_{XY} \leq a)$ (Continued)

k_1	a	$k_2 = 8$	$k_2 = 9$	$k_2 = 10$	k_1	a	$k_2 = 9$	$k_2 = 10$	k_1	a	$k_2 = 10$
8	0	.0001	.0000	.0000	9	0	.0000	.0000	10	0	.0000
	1	.0002	.0001	.0000		1	.0000	.0000		1	.0000
	2	.0003	.0002	.0001		2	.0001	.0000		2	.0000
	3	.0005	.0003	.0002		3	.0001	.0001		3	.0000
	4	.0009	.0005	.0003		4	.0002	.0001		4	.0001
	5	.0015	.0008	.0004		5	.0004	.0002		5	.0001
	6	.0023	.0012	.0007		6	.0006	.0003		6	.0002
	7	.0035	.0019	.0010		7	.0009	.0005		7	.0002
	8	.0052	.0028	.0015		8	.0014	.0007		8	.0004
	9	.0074	.0039	.0022		9	.0020	.0011		9	.0005
	10	.0103	.0056	.0031		10	.0028	.0015		10	.0008
	11	.0141	.0076	.0043		11	.0039	.0021		11	.0010
	12	.0190	.0103	.0058		12	.0053	.0028		12	.0014
	13	.0249	.0137	.0078		13	.0071	.0038		13	.0019
	14	.0325	.0180	.0103		14	.0094	.0051		14	.0026
	15	.0415	.0232	.0133		15	.0122	.0066		15	.0034
	16	.0524	.0296	.0171		16	.0157	.0086		16	.0045
	17	.0652	.0372	.0217		17	.0200	.0110		17	.0057
	18	.0803	.0464	.0273		18	.0252	.0140		18	.0073
	19	.0974	.0570	.0338		19	.0313	.0175		19	.0093
	20	.1172	.0694	.0416		20	.0385	.0217		20	.0116
	21	.1393	.0836	.0506		21	.0470	.0267		21	.0144
	22	.1641	.0998	.0610		22	.0567	.0326		22	.0177
	23	.1911	.1179	.0729		23	.0680	.0394		23	.0216
	24	.2209	.1383	.0864		24	.0807	.0474		24	.0262
	25	.2527	.1606	.1015		25	.0951	.0564		25	.0315
	26	.2869	.1852	.1185		26	.1112	.0667		26	.0376
	27	.3227	.2117	.1371		27	.1290	.0782		27	.0446
	28	.3605	.2404	.1577		28	.1487	.0912		28	.0526
	29	.3992	.2707	.1800		29	.1701	.1055		29	.0615
	30	.4392	.3029	.2041		30	.1933	.1214		30	.0716
	31	.4796	.3365	.2299		31	.2181	.1388		31	.0827
	32	.5204	.3715	.2574		32	.2447	.1577		32	.0952
	33	.5608	.4074	.2863		33	.2729	.1781		33	.1088
	34	.6008	.4442	.3167		34	.3024	.2001		34	.1237
	35	.6395	.4813	.3482		35	.3332	.2235		35	.1399
	36	.6773	.5187	.3809		36	.3652	.2483		36	.1575
	37	.7131	.5558	.4143		37	.3981	.2745		37	.1763
	38	.7473	.5926	.4484		38	.4317	.3019		38	.1965
	39	.7791	.6285	.4827		39	.4657	.3304		39	.2179
	40	.8089	.6635	.5173		40	.5000	.3598		40	.2406
						41	.5343	.3901		41	.2644
						42	.5683	.4211		42	.2894
						43	.6019	.4524		43	.3153
						44	.6348	.4841		44	.3421
						45	.6668	.5159		45	.3697
										46	.3980
										47	.4267
										48	.4559
										49	.4853
										50	.5147

INDIAN STATISTICAL INSTITUTE

Statistical Inference I

Back Paper Examination

B Stat III, Ist Semester, 2003-2004

Date: March 3, 2004 Time: 3 hrs. Total Points: 100

1. (a) Let $f_\theta(x) = f(x-\theta)$, $-\infty < x < \infty$ represent a location family where θ is a scalar parameter, $-\infty < \theta < \infty$. Show that the Fisher Information in f_θ about θ does not depend on θ.

(b) Let $f_\sigma(x) = \frac{1}{\sigma} f(x/\sigma)$ $-\infty < x < \infty$ represent a scale family where σ is a scalar parameter, $\sigma > 0$. Show that the Fisher Information in f_σ about σ does depend on σ. [6+6=12 points]

2. Suppose that $X_1, \ldots, X_n$ represent a random sample from a distribution with density

$$f(x) = \frac{1}{b} e^{-(x-a)/b}, \; x \geq a, -\infty < a < \infty, b > 0.$$

We will denote this by $E(a, b)$, the exponential distribution with parameters (a, b).

(a) Show that $T_1 = X_{(1)}$ and $T_2 = \sum(X_i - X_{(1)})$ are jointly minimal sufficient for (a, b).

(b) Show that if $E_{a,b}[h(T_1, T_2)] = 0$ for all a, b, then $h(T_1, T_2) = 0$ with probability one. (You can assume that T_1 and T_2 are independent, and are distributed as $E(a, b/n)$ and $\frac{1}{2} b \chi^2_{2n-2}$ random variables respectively). Thus the minimial sufficient statistics are complete. [4+8=12 points]

3. Suppose that a random sample of length of life measurements $X_1, \ldots, X_n$, is to be taken on components whose length of life has an exponential distribution with mean θ. It is often of interest to estimate $\bar{F}(t) = 1 - F(t) = e^{-t/\theta}$, the reliability at time t of a component of this type. Find the uniformly minimum varianced unbiased estimator of $\bar{F}(t)$. [14 points

4. Prove that an estimator $aX + b$ $(0 \leq a \leq 1)$ of $E_\theta(X)$ is inadmissible (with squared error loss) under each of the following conditions.

(i) if $E_\theta(X) \geq 0$ for all θ, and $b < 0$.

(ii) if $E_\theta(X) \leq k$ for all θ, and $ak + b > k$. [6+6=12 points]

P. T. O

Let $X, \ldots, X_n$ be a random sample from a $N(\theta, 1)$ distribution. Let $0 < \alpha < 1$, and $\theta_1 < \theta_2 <$ $< \theta_4$. Consider testing the following set of hypotheses:

$$H_0 : \theta \leq \theta_1 \quad \text{or} \quad \theta_2 \leq \theta \leq \theta_3 \quad \text{or} \quad \theta \geq \theta_4$$
$$H_1 : \theta_1 < \theta < \theta_2 \quad \text{or} \quad \theta_3 < \theta < \theta_4.$$

oes there exist a level α UMP test for the above? If yes, derive the test. If not, explain why ›t. [12 points]

Suppose that $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ represent independent random samples from $N(\xi, 1)$ ıd $N(\eta, 1)$ respectively. Consider testing $H_0 : \eta \leq \xi$ against $H_1 : \eta > \xi$. Show that there exists ı UMP test, which rejects when $\bar{Y} - \bar{X}$ is too large. [14 points]

Suppose that (X, Y) be distributed according to the exponential family density

$$f_{\theta_1,\theta_2}(x, y) = c(\theta_1, \theta_2)h(x, y)exp(\theta_1 x + \theta_2 y).$$

ıow that the only unbiased test for testing $H_0 : \theta_1 \leq a, \theta_2 \leq b$ against $H_1 : \theta_1 > a$, or $\theta_2 > b$, both is $\phi(x, y) = \alpha$. [12 points]

Let $X_1, \ldots, X_n$ be a random sample from a $N(\xi, \sigma^2)$ with both ξ and σ^2 unknown. Show ıat there exists a uniformly most accurate (UMA) lower bound for σ^2. Explain how you would ılculate such a bound with confidence level 0.95 if $n = 16$. [12 points]

INDIAN STATISTICAL INSTITUTE

Statistical Inference I

Mid Semester Exam I, BStat 3rd Yr, September 15, 2003

Answer as many as you can. Maximum you can score is 40.

1. (a) Let $f_\theta(x) = f(x - \theta)$, $-\infty < x < \infty$ represent a location family where θ is a scalar parameter, $-\infty < \theta < \infty$. Show that the Fisher Information in f_θ about θ does not depend on θ.

(b) Let $f_\sigma(x) = \frac{1}{\sigma} f(x/\sigma)$ $-\infty < x < \infty$ represent a scale family where σ is a scalar parameter, $\sigma > 0$. Show that the Fisher Information in f_σ about σ does depend on σ. [3+3=6 points]

2. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ represent n independent and identically distributed observations drawn from an exponential distribution with mean θ. (The probability density function of an exponentially distributed random variable with mean θ is given by $f(x) = \theta^{-1}e^{-x/\theta}$, $x \geq 0$). Let

$$g(\boldsymbol{X}) = \frac{X_1}{X_1 + \ldots + X_n}.$$

(a) Show that $g(\boldsymbol{X})$ is an ancillary statistic.

(b) Using part (a) or otherwise, find $E(g(\boldsymbol{X}))$. [4+4=8 points]

3. Consider the parametric family $F_\theta, \theta \in \Omega$, and let Δ be the class of all statistics δ with $E(\delta^2) < \infty$. Let $\mathcal{U}$ be the set of all unbiased estimators of zero in Δ (i.e. if $\delta \in \mathcal{U}$, then $E(\delta) = 0$. Show that a necessary and sufficient condition for δ to be a UMVU estimator for its expectation $g(\theta)$ is that

$$E(\delta U) = 0$$

for all $U \in \mathcal{U}$ and all $\theta \in \Omega$. (Notice since $E(U) = 0$, $Cov(\delta, U) = E(\delta U)$, so that the condition essentially says that δ is uncorrelated with every $U \in \mathcal{U}$). [10 points]

4. Suppose that a random sample of length of life measurements $X_1, \ldots, X_n$, is to be taken on components whose length of life has an exponential distribution with mean θ. It is often of

P. T. O

interest to estimate $\bar{F}(t) = 1 - F(t) = e^{-t/\theta}$, the reliability at time t of a component of this type. Find the uniformly minimum varianced unbiased estimator of $\bar{F}(t)$. [10 points]

5. Suppose that n independent and identically distributed observations $X_1, \ldots, X_n$ are available from the distribution of X that is modeled by the usual two parameter $N(\mu, \sigma^2)$ model. Show that neither the MLE nor the UMVUE of σ^2 is admissible for σ^2 with respect to squared error loss.

[8 points]

6. Suppose that n independent and identically distributed observations $X_1, \ldots, X_n$ are available from a Bernoulli distribution with parameter p. Find the Bayes estimator for p based on a Beta(α, β) prior on (0, 1), and the loss function

$$L(\theta, \delta) = \frac{(\delta - \theta)^2}{\theta(1-\theta)}.$$

[6 points]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2003-2004
B. Stat (Hons.) - III Year
Sample Surveys

Date : 17.09.2003 Maximum Marks : 40 Duration : 3 Hours

1.(a) Distinguish between 'sampling and non-sampling errors'. In a sample survey, describe four sources of non-sampling errors.

(b) What is a 'pilot survey'. Mention four advantages of a pilot survey.

(3+3)=(6)

2.(a) For a Probability Proportional to Size With Replacement (PPSWR) selection of n units from a Population of N units with size measures X_i on units U_i, $i = 1,2,\ldots,N$ by Lahiri's Method, show that the selection probability of U_i is indeed X_i/X, where $X = \sum_1^N X_i$.

(b) For the above scheme, write down the values of inclusion probability for the ith Unit and joint inclusion probability for the pair of units U_i & U_j.

(4+(2+3))=(9)

3.(a) Write down the variance of the unbiased estimator of the population total for SRSWOR scheme. Also derive an unbiased estimator of this variance.

(b) Distinguish between 'Linear and Circular Systematic Sampling'.

((1+3)+4)=(8)

EITHER

4. A population is divided into two strata of sizes 30 and 20. From the first stratum 3 units are chosen by simple random sampling without replacement and the corresponding y-values of a study variate are found to be 121, 141 and 102. From the second stratum 2 units are selected using PPSWR scheme and the following data on y and size measure x is observed :

	Units Selected	
	1	2
y	314	441
x	61	82

It is also known that the total of X values for the 20 units of stratum 2 is 1498.

(a) Estimate the population mean $\bar{Y}$ unbiasedly.

(b) Also obtain an unbiased estimate of the sampling error of your estimate in (a) above.

(7+10)=(17)

P.T.O.

OR

The following data show the stratification of all the farms in a district by farm size and the average acres of wheat per farm in each stratum.

Farm Size (acres)	No. of farms N_i	Average wheat (acres) $\bar{Y}_i$	Standard deviation S_i
0-40	394	5.4	8.3
41-80	461	16.3	13.3
81-120	391	24.3	15.1
121-160	334	34.5	19.8
161-200	169	42.1	24.5
201-240	113	50.1	26.0
241-	148	63.8	35.2

For a sample of 100 farms (selected by random sampling WOR) compute the sample sizes in each stratum under (i) proportional allocation and (ii) Neyman's optimum allocation. Compare these two allocations with a direct (unstratified SRSWOR) sample of size 100, using Variance criterion.

(2+3+12)=(17)

***

INDIAN STATISTICAL INSTITUTE

B. Stat (Third year) — First semester

Ordinary Differential Equation — Mid-semester Examinatio

Instructor: Rudra Sarkar — Date: September 19, 2003

Maximum Marks 100 — Maximum Time 3 hrs

Answer as many questions as possible. But maximum marks can be obtained without answerin all (check the marks in the margin).

1. Let $L(y) = y'' + a_1 y' + a_2 y$ for two constants a_1, a_2.

(a) Let ϕ_1 and ϕ_2 be two linearly independent solution of $L(y) = 0$ on an interval . Let J be a subinterval of I. Prove that ϕ_1 and ϕ_2 are also linearly independent on J.

(b) For ϕ_1 and ϕ_2 as in (a) show that they can not have a common zero in I.

(c) Let ϕ be a solution of $L(y) = 0$ defined on $\mathbb{R}$ which satisfies the bounda conditions $\phi(0) = \phi(2\pi), \phi'(0) = \phi'(2\pi)$. Prove that ϕ is periodic of period 2π. 5+ 2+

2 Let $\phi_1, \ldots, \phi_n$ be n linearly independent solution of $L(y) = 0$. Let $\psi_i(x) = e^{\frac{a_1}{n}x}\phi_i($ for $i = 1, \ldots, n$. Show that:

(a) $\psi_1, \psi_2, \ldots, \psi_n$ are also linearly independent.

(b) Wronskian of $\psi_1, \psi_2, \ldots, \psi_n$ is constant.

(c) Conclude that if ψ_i are soultions of $M(y) = 0$ where

$$M(y) = y^{(n)} + b_1 y^{(n-1)} + \cdots + b_n y$$

for some constants $b_1, \ldots, b_n$, then $b_1 = 0$. 5+3-

3. Obtain the general solution of

$$y''(x) - y(x) = \sum_{n=1}^{k} \frac{\sin nx}{n^2}.$$

(Hint: split the differential equation in k-many differential equations and use annihila method.)

4. Let

$$L(y)x = a_0(x)y'' + a_1(x)y' + a_2(x)y = 0 \text{ for } x \in I,$$

P.T.

where $I \subset \mathbb{R}$ is an interval, a_o, a_1, a_2 are continuous functions on I and $a_0(x) \neq 0$ for any $x \in I$. If u and v are two twice-differentiable functions on I, prove that
(a)

$$uL(v) - vL(u) = a_0(x)\frac{d}{dx}W(u,v)(x) + a_1(x)W(u,v)(x),$$

where W is the Wronskian.

(b) Using (a) show that if u, v are two solutions of $L(y) = 0$ then

$$W(u,v)(x) = k\exp\left[-\int_{x_0}^{x}\frac{a_1(s)}{a_0(s)}ds\right]$$

where k is a constant. Find k. 5+5

5.(a) Prove that a general solution of a differential equation of order 2 has exactly two independent constants.
(b) Consider the algebras $\mathcal{D}$ and $\mathcal{P}$ of all (complex) constant coefficient diferential opertaors and all polynomials with complex coefficients respectively. Find an algebra isomorphism between $\mathcal{D}$ and $\mathcal{P}$. 5+5

6. Consider the differential equation $L(y) = Ae^{\alpha x}$ where L is the constant coefficient differential operator

$$L(y) = y^{(n)} + a_1 y^{(n-1)} + \cdots + a_n y$$

and A, α are constants. Compute a particular solution of the differential equation in case α is a root of the characteristic polynomial p of L of multiplicity k. 10

7. Consider the equation $y'' + \alpha(x)y = 0$ where α is a real valued continuous function on an open interval I. Let ϕ and ψ be two treal-valued linearly independent solutions of the above equation. Show that between any two succesive zeros of ϕ there is a zero of ψ. 10

8. Consider the differential equation $L(y) = 0$ where L is the constant coefficient differential operator

$$L(y) = y^{(n)} + a_1 y^{(n-1)} + \cdots + a_n y.$$

(a) Show that the multiplicity of a zero of a nontrivial solution ϕ is less than n.
(b) Show that the zeroes are isolated (i.e. set of zeros do not have a limit point). 15

9. Let m, n be two positive integers and $m < n$. Consider the differential equation

$$\frac{d}{dx}\left\{(1-x^2)\frac{dy}{dx}\right\} + \alpha(\alpha-1)y = 0.$$

Prove that
(a) $P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n$ is a solution of the above equation for $\alpha = n+1$.
(b) $\int_{-1}^{1} x^m P_n(x)dx = 0$.

(c) $\int_{-1}^{1} x^n P_n(x)dx = \frac{2^{n+1}(n!)^2}{(2n+1)!}$.
(Hint for b and c: Use expression for P_n as in (a). Justify $(\frac{d}{dx})^* = -\frac{d}{dx}$.) 7+6+7

10. Find the first three terms of the Legendre series $f(x) = e^x$ where Legendre series of a suitable function f is given by

$$f(x) = \alpha_0 P_0 + \alpha_1 P_1 + \alpha_2 P_2 + \ldots.$$

(Hint: $\{P_n\}$ is an orthogonal family for the inner-product $\langle f, g\rangle = \int_{-1}^{1} fg$. P_0, P_1, P_2 are known.) 10

11. Let n be a positive integer. Consider the Hermite equation

$$L_n(y) = y'' - 2xy + 2ny = 0.$$

let $H_n(x) = (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2}$.
Prove that:
(a) $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$.
(b) $H_n'(x) = 2xH_n(x) - H_{n-1}(x) = 2nH_{n-1}(x)$.
(c) H_n is a solution of $L_n(y) = 0$. 3+4+3

12. For postive integer m, n let H_m, H_n be as in the previous problem and $h_i = H_i(x)e^{-\frac{x^2}{2}}$ for $i = m, n$. Show that h_n satisfies the differential equation:

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

Use this to show that $\int_{-\infty}^{\infty} h_n(x)h_m(x)dx = 0$. Find

$$\int_{-\infty}^{\infty} h_n(x)h_n(x)dx.$$

3+5+7

INDIAN STSTIATICAL INSTITUTE

Mid-Semester Examination : (2003-2004)

B.Stat. (Hons.) III Year

Elective : Introduction to Anthropology and Human Genetics

Date : 26.9.03 **Maximum Marks : 100** **Duration 3 Hours**

Note : Use separate answerscript for Group A and Group B. Answer ***any five*** questions from each group.

GROUP A

1. What is Anthropology ? Describe its different branches. [10]
2. What do you mean by the term *systematics* and *nomenclature* ? Discuss with examples. [10]
3. Describe the characteristic features of order *primate* and suborder *anthropoidea* with examples.[10]
4. Describe in brief the anatomical, physiological and behavioral changes that have occurred in the evolutionary development of mankind. [10]
5. Describe the characteristic features that made man 'unique' [10]
6. Compare the anatomical/ morphological features of *Homo erectus* with *Homo sapiens.* [10]
7. Write in brief different theories of *evolution of man.* [10]
8. Discuss in detail Darwin's theory of natural selection with its criticism. [10]
9. What is *marriage*? Discuss its different forms. [10]
10. Define *Culture* ? Illustrate its qualities. [10]

GROUP B

1. Define a cell. Describe the functions of different organelles of the cell. (3+7)
2. What is the significance of Meiosis? How does it differ from Mitosis? If two daughter cells resulting from Meiosis, do they have the same amount of genetic material, do they also have the same genotype? (5+3+2)
3. Describe the normal human karyotype? Illustrate how Turner's syndromes arise. (5+5)

P.T.O

4. Discuss Mendel's law of independent assortment. A married couple both unable to taste PTC, have taster child; (a) what is the probability that their 2^{nd} child is also a taster? (b) what are th probabilities of every possible configuration of four children in this family? (5+2+3)

5. A man of blood group 'O' marries a woman of blood group 'A', the wife's father is blood group 'O' What is the probability that their children will belong to blood group 'O'? Justify your answer wi reasons. (4+6)

6. Name a rare disease in human which is caused by single, autosomal recessive gene. Explain how recessive genetic disorder may skip its inheritance over one or even two generations. (3+7)

7. In his theory of evolution by natural selection, Darwin assumed that individuals in a population va but he could not explain the origin of those variations, because the principles of genetics had n been discovered. How would you explain briefly the origin of variations? (10)

8. Write short notes on any two of the following: (5+5
   (a) Polygenic inheritance
   (b) Albinism
   (c) Homozygous and Heterozygous
   (d) Haploid and Diploid

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2003-2004

B. Stat. - III Year

Introduction to Sociology & Sociometry

Date : 26.09.03 | Maximum Marks : 20 | Duration : One hour

Note : There will be two groups (A & B) in question paper.
But the answer sheet should be separate for each group.

Group-A

Q. 1) a) Relate the sociological thinkers with the concept introduced by them: 2½

Thinker	Concept
1) Emile Durkheim	Sympathetic Introspection ( )
2) Auguste Comte	Organic Solidarity ( )
3) Max Weber	Social Darwinism ( )
4) Karl Marx	Social Statics ( )
5) Herbert Spencer	Dialectical Materialism ( )

b) Write the correct answer: 2½

i) The rise of sociology took place in 16^{th}/19^{th}/20^{th}/ century.
ii) The 'Positive era' (according to Auguste Comte) will be dominated by 'Military' / 'Scientists & Industrialists'.
iii) Emile Durkheim's conception of society is related to Functionalism/ Conflict theory.
iv) Max Weber's conception of social change is related to 'Cultural interpretation'/ 'Economic interpretation'.
v) Feudal society is principally based on agriculture/ industry.

Q. 2) Write short notes (any two): 3x2

i) Bureaucracy
ii) Social action
iii) Culture
iv) Community

P.T.O

Q. 3) Write, in brief, any one of the following: 4

i) What are the major contrasting features of caste and class systems?
ii) What are the main characteristics of functional theory?
iii) What is the essential content of Robert K. Merton's theory of functionalism?

Group- B

Q. 4) Answer the following: 5

Six persons ***a, b, c, d, e,*** and ***f*** are friends. But ***f*** is jealous. He watches who goes to whom and himself goes only to ***a*** and ***d***. No one comes to him. Again, ***b, c,*** and ***e*** go to each other. On the other hand, ***a*** and ***d*** go to ***b***.

i) Arrange the above description in the form of a (0,1) matrix;
ii) Also show the above data graphically; and,
iii) obtain in-degrees and out-degrees of each friend.

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : (2003 –2004)

B. Stat (Hons) III Year

Economics III

Date : 26 September, 2003. Maximum Marks : 100 Duration : 3 Hours

Note: This paper carries 116 marks. Answer as many questions as you like. The maximum you can score is 100.

1. Discuss the advantages or otherwise of using generalized inverses in analyzing linear models. In this context, comment on the merits and demerits of some popular generalized inverses like the *minimum norm*, *least squares* and the *Moore – Penrose* inverse. [10 + 10 = 20]

2. Consider the Cobb – Douglas production function $y = Ak^{\alpha}L^{\beta}$, where y = output, K = capital employed and L = labour.
   (a) How would you estimate the model given data on (y, K, L)?
   (b) How would you test for constant returns to scale and increasing returns to scale? Compute the test statistics explicitly in each case and state the testing rule.
   (c) Under the assumption of constant returns, what are the estimators for α and β? [5 + (6 + 8) + 7 = 26]

3. Consider the model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where y = monthly rainfall data over several years, x_1 = 1 if the month is July, 0 otherwise and x_2 = 1 if the month has 31 days, 0 otherwise. We have 10 observations on (y, x_1, x_2). Find the estimators of α, β_1, β_2 explicitly. You can make any relevant assumption. [14]

P. T. O.

4. (a) What are the problems associated with estimating the model

$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$ and $x_2 = x_1 + u$ with $u \sim N(0,1)$?

(b) Would your answer change in any way if $u \sim N(0,100)$? [7 + 3 = 10]

5. Consider the linear model $y = \alpha + \beta x + \varepsilon$ where the data on variable x is mean centered ($\bar{x} = 0$ for the given data). Compare the performance of (i) principal component regression with one regressor, $y = \gamma z + \varepsilon$, where z is the principal component with the higher eigenvalue, and (ii) ridge regression.
Do the comparison in terms of goodness of fit. [16]

6. (a) Suppose you have estimated the CLRM model $y = X\beta + \varepsilon$ and let the t-statistic corresponding to the variable x_j be given by t_j. Show that if $|t_j| < 1$ then $\bar{R}^2$ will improve if we re-estimate our model after removing the variable x_j from the set of regressors.

(b) Generalize (a) to the statement that if the F-statistic for a subset of the regressor variables is < 1 then $\bar{R}^2$ will improve if we drop these variables from the set of regressors.

(c) Is the converse to (a) true? [12 + 10 + 8 = 30]

INDIAN STATISTICAL INSTITUTE

Mid-Term Examination: (2003-04)

ELECTIVE GEOLOGY

Date : 26. 9 . 03. Maximum Marks :100 Duration : 3 hours

B.Stat-III yr.

Answer Q.1 and any 4 questions from the rest.

1. Write the correct answer.

A) The mineral(s) present in Dunite is (are)
   a. olivine+quartz+orthoclase
   b. granite+basalt
   c. kyanite+sillimanite
   d. mainly olivine

B) Fossils are expected in
   a. sedimentary rocks
   b. igneous rock
   c. metamorphic rocks
   d. meta-basalt

C) Which of the following metamorphic mineral can be used as refractory material
   a. Biotite
   b. Chlorite
   c. Kyanite
   d. Muscovite

D) Ichnofossils are
   a. body fossils
   b. trace fossils
   c. plant fossil only
   d. vertebrate fossils only

E) Paraconformity is marked by
   a. angular discordance among strata
   b. lithological discordance among strata
   c. erosional contacts among strata
   d. time discordance among two groups of strata

F) Prolonged transportation in water produces
   a. well sorted clasts
   b. poorly sorted clasts
   c. highly angular clasts
   d. none of the above

G) Trough cross bedding is produced by
   a. migration of straight crested dunes
   b. migration of lunate crested dunes
   c. migration of linguoid crested dunes
   d. migration of sinuous crested dunes

P. T. O

H) The matrix of a sandstone is
a. mechanically deposited
b. chemically precipitated
c. produced by mechanical weathering of framework grains
d. produced by chemical weathering of framework grains.

I) A length transverse imbrication is formed in subaqueous transport of
a. platy clasts
b. prolate clasts
c. equant clast
d. none of the above

J) Pyrite is a
a. sulphide
b. sulphate
c. native
d. oxide

2. What is geology? Write the laws of superposition and faunal succession. Define (with sketches) angular unconformity, disconformity and nonconformity. ------------
5+6+3x3=20

3. What is a mineral? How do minerals differ from rocks? Name two silicate minerals. Write the names of two minerals present in granite and basalt. What do you understand by the term "Bowen's reaction series"? --------
---------- 4+4+4+8=20

4. What is a fossil? Why generally hard parts of organisms are preserved as fossils? How does a fossil (or more than one variety of fossils) help to determine the age of the host rock? Do fossils provide absolute ages of the host rock? Name another method (apart from paleontological methods) that determines the age of a rock or mineral.
------------ 5+5+5+3+2 =20

5. Write short notes on
a. Primary and secondary seismic waves
b. Soil profile
c. Sphericity
d. Size classification for clastic grains ------------------- 5x4=20

6. A) Describe different modes of mechanical weathering with necessary diagrams.
B) Describe the geomorphologic elements of the braided and the meandering streams with labeled diagrams. --------------------------- 10x2=20

INDIAN STATISTICAL INSTITUTE

B. Stat (Third year) First Semester Examination (2003-2004)
Ordinary Differential Equation Date: November 27, 2003

Maximum Marks 80 Maximum Time 2 hrs. 30 Minutes

Answer as many questions as possible. The maximum marks you can score is 80.

In this paper $I_{n\times n}$ is the $n \times n$ identity matrix. The set of all $n \times n$ matrices with real and complex entires are denoted by $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ respectively. The set of invertible elements in M_n is denoted by GL_n. $SO(n)$ is the set of special orthogonal $n \times n$ matrices.

1. Consider the following differential equation:

$$x^{(n)} = F(x, x', x'', \cdots, x^{(n-1)}),$$

where F is a continuous function.

(a) Show that if $x = \sin t$ is a solution then $x = \cos t$ is also a solution.

(b) For a solution ϕ (defined on entire $\mathbb{R}$) of the above equation let us call the parametric curve $\tilde{\phi} : \mathbb{R} \longrightarrow \mathbb{R}^n : t \mapsto (\phi(t), \phi'(t), \cdots, \phi^{(n-1)}(t))$ *the phase curve* of ϕ in the *phase space*. Show that only one phase curve can pass through a point x_0 in the phase space $\mathbb{R}^n$. 3+7

2. (a) Consider the equation in Question 1 for $n = 1$. Show that if for a solution ϕ defined for an inetrval I, $\phi(a) = \phi(b)$ for $a, b \in I$ then ϕ can be extended as a periodic solution defined on entire $\mathbb{R}$. What is its period ?

(b) For a continuous map f defined on $\mathbb{R}$, let P_f be the set of all periods of f. Show that P_f is a closed subset of $\mathbb{R}$ and a subgroup of $(\mathbb{R}, +)$.

(c) Suppose ϕ is a nonconstant solution of the equation in Question 1 for an arbitrary $n \in \mathbb{N}$. If ϕ is periodic of periods T_1 and T_2 with $T_2 > T_1$ then $T_2 = mT_1$ for some $m \in \mathbb{N}$. (Hint: What are closed subgroups of $\mathbb{R}$?)

(d) Conclude from (c) that the phase curve of a soltion to $x^{(n)} = F(x, x', x'', \cdots, x^{(n-1)})$ is either a point, or has no self intersection or is a closed smooth curve (like a circle). 3+6+6+ 5

3. (a) Show that ∞ (infinity) is not a regular singular point of the Bessel equation: $t^2x'' + tx' + (t^2 - p^2)x = 0$.

(b) Put the initial value problem $x'' + 9x' + 14x = \frac{1}{2}\sin t, x(0) = 0, x'(0) = 1$ into matrix form.

(c) Consider the vector equation $x'(t) = A(t)x(t)$ where $A(t + w) = A(t)$ for all $t \in \mathbb{R}$.

P.T.O

Define the set of characteristic multipliers of A and show that this set is well defined.

5+3+5

4. Let $J_p(t) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\Gamma(p+k+1)} \left(\frac{t}{2}\right)^{p+2k}$ be the Bessel function of the first kind of order p. Assume the following relations:

$$tJ_p'(t) = -tJ_{p+1}(t) + pJ_p(t) = tJ_{p-1}(t) - pJ_p(t).$$

Prove the following:

(a) If for $\alpha \neq 0$, $J_p(\alpha) = 0$ then $J_{p+1}(\alpha) \neq 0$.

(b) Between any two positive roots of $J_p(t)$ there is a zero of $J_{p+1}(t)$.

(c) Between any two positive roots of $J_{p+1}(t)$ there is a zero of $J_p(t)$.

Hint: Show $(t^p J_p(t))' = t^p J_{p-1}(t)$ and $(t^{-p} J_p(t))' = -t^{-p} J_{p+1}(t)$.

5+5+5

5. (a) For $A \in M_n(\mathbb{R})$ let e^A be $\Phi(1)$ where $\Phi(t)$ is the fundamental matrix solution of the initial value problem $X'(t) = AX(t), X(0) = I_{n\times n}$. Show that $e^{\text{trace } A} = \det e^A$ using only this definiton of e^A.

(b) Let A be a skew symmetric matrix in $M_n(\mathbb{R})$. Show that $e^A \in SO(n)$. Establish (by an example) that for $B \in M_2(\mathbb{R})$, $e^B = I_{2\times 2}$ does not imply that $B = 0$.

(c) Let $A, C \in M_n(\mathbb{R})$. Let $\Phi(t)$ be a fundamental matrix solution of the vector equation $X'(t) = AX(t)$.

Show that $\Phi^T(t)\Phi(t) = C$ if and only if $e^{tA} \in SO(n)$.

(d) Let

$$A = \begin{bmatrix} 5 & 4 \\ 1 & 2 \end{bmatrix}.$$

Find e^A using the definition given in (a) above (i.e without using any other definition of e^A).

5+5+5+5

6. Let $f(t,x)$ be a continuous function on the strip $S_\alpha := \{|t| \leq \alpha, |x| < \infty\}$ and $|f(t,x_1) - f(t,x_2)| \leq k|x_1 - x_2|$ for some constant k and for all $(t,x_1), (t,x_2)$ in S_α.

(a) Prove the uniquness of the solution of the initial value problem:

$$x'(t) = f(t,x), x(0) = x_0, |t| \leq \alpha.$$

(Hint: Assume that the successive approximations tend to a solution. Show that any other solution is also the limit of the sequence of successive approximations.)

(b) In the above initial value problem if $f(t,x) = (3t^2+1)\cos^2 x + (t^3 - 2t)\sin 2x$ then show that it has solution for any $t \in \mathbb{R}$.

10+5

INDIAN STATISTICAL INSTITUTE
First Semester Examination : 2003-2004

B. Stat. (Hons.) III Year
Sample Surveys

Date : 1.12.2003 | Maximum Marks : 100 | Duration : 3 Hours

Answer all questions.

1.(a) Distinguish between sampling and non-sampling errors in a survey. List down the various sources of non-sampling errors and discuss briefly the methods of assessment and control of these errors.

(b) A survey was conducted in a village consisting of 625 households (hh) by covering a sample of 50 hh using a simple random sampling (srs) without replacement scheme to estimate the average monthly expenditure $(\overline{Y})$ on salt. The estimate was found out to be Rs.4.20 with a standard error of 0.47. Using this information, determine the sample size needed to estimate the same characteristic in a neighbouring village on the basis of a sample to be selected by srs with replacement scheme such that the length of the confidence interval at 95% confidence level is 20% of the true value of $(\overline{Y})$. State clearly the assumptions involved in finding out the sample size.

(2+4+6)+(8+2)=(22)

2.(a) What do you understand by the term 'Intra Class Correlation Coefficient'? A population consists of 13 cluster each of size 6. Find, giving an explanation, the bounds for the Intra Cluster Correlation Coefficient among the elements of a cluster.

(b) Suppose that n clusters are selected at random and with replacement from a population consisting of N clusters the ith cluster being of size μ_i, $i = 1,2,\ldots,N$.

(i) Write down the conventional unbiased estimator for the population mean $\overline{Y}$, when $\sum_1^N \mu_i$ is known.

(ii) When $\sum_1^N \mu_i$ is unknown, derive the 'almost unbiased ratio-type estimator' due to Murthy and Nanjamma for estimating the population mean $\overline{Y}$.

(iii) Obtain the inclusion probability π_i of the ith cluster for the sampling design which makes the estimator. $\sum_1^n \mu_i \bar{y}_i \Big/ \sum_1^n \mu_i$ unbiased for $\overline{Y}$, where the summation is over the sampled clusters selected without replacement.

(3+4)+(3+8+4)=(22)

[P.T.O.]

3.(a) Explain, with mathematical justification, how 'two-stage sampling' can be considered as a compromise between 'sampling of second stage units directly' and 'cluster sampling'.

(b) Assuming a suitable cost function when m second stage units are selected from each of the n selected first stage units of equal size M, obtain the optimum values of n and m so as to minimize the Variance of the sample mean, when simple random sampling without replacement is used at both stages. Also obtain the 'optimum' Variance of the sample mean.

(c) For the sampling design in (b), how do you estimate the loss due to 'two stage sampling' compared to 'unistage sampling'.

(5+(2+3+4)+8)=(22)

4. A sample survey was conducted in a district to estimate the total industrial output. A stratified two-stage sampling design was adopted with urban blocks as first stage units and factories within them as second stage units. From each stratum, 4 blocks were selected with probability proportional to size and with replacement and 4 factories were selected from each selected block with equal probability and without replacement. The data on output (in suitable units) for the sampled factories together with information on selection probabilities are given below:

Using this data

(i) Obtain an unbiased estimate of the total industrial output in the district.
(ii) Also obtain an unbiased estimate of the variance of the estimator used in (i).

Stratum	Sample block	Inverse of Probability of selection	Total No. of factories	Output of sampled factories			
				1	2	3	4
	1	440.21	28	104	182	148	87
I	2	660.43	14	108	64	132	156
	3	31.50	240	100	115	50	172
	4	113.38	76	346	350	157	119
	1	21.00	256	124	111	135	216
II	2	16.80	288	123	177	106	138
	3	24.76	222	264	78	144	55
	4	49.99	69	300	114	68	111
	1	67.68	189	110	281	120	114
III	2	339.14	42	80	61	118	124
	3	100.00	134	121	212	174	106
	4	68.07	161	243	116	314	129

(12+22)=(34)

***

INDIAN STATISTICAL INSTITUTE

1st Semestral Examination (2003-04)

B.Stat. (Hons.) 3rd year

Linear Statistical Models

December 9, 2003 Maximum time: $3\frac{1}{2}$ hours Maximum marks: 60

[*This test is closed book. The total number of marks of all the questions is 66, and the maximum you can score is 60. Except for question 3, you can use any result that was proved in class.*]

1. Show that all the components of β in the model $(\boldsymbol{y}, \boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ are estimable if and only if $\boldsymbol{X}$ has full column rank, and that in such a case, every LPF is estimable. [6]

2. Four items having weights β_1, β_2, β_3 and β_4 are weighed eight times in different combinations. The measurements follow the model

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \end{pmatrix},$$

with uncorrelated errors having zero mean and variance σ^2. We denote this model by $(\boldsymbol{y}, \boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$.

(a) Show that every linear parametric function (LPF) of the model is estimable.

(b) When is $\boldsymbol{l'y}$ the BLUE of its expectation?

(c) When is $\boldsymbol{l'y}$ a linear zero function (LZF)?

(d) What is the BLUE of β_1?

(e) What is the variance of the BLUE of β_1?

(f) Find ~~an~~ the usual unbiased estimator of σ^2. [1+2+2+1+2+2=10]

3. Given n observations from the simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon$ with zero-mean uncorrelated normal errors with variance σ^2, show that a $100(1-\alpha)\%$ confidence band for the regression line is

$$\left[\widehat{\beta}_0 + \widehat{\beta}_1 x - \sqrt{2F_{2,n-2,\alpha}\widehat{\sigma}^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{n(\overline{x^2} - \bar{x}^2)}\right)},\right.$$

$$\left.\widehat{\beta}_0 + \widehat{\beta}_1 x + \sqrt{2F_{2,n-2,\alpha}\widehat{\sigma}^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{n(\overline{x^2} - \bar{x}^2)}\right)}\right],$$

P.T.O

where $\bar{x}$ and $\overline{x^2}$ are the observed average of x and x^2, respectively, and $F_{2,n-2,\alpha}$ is the $1-\alpha$ quantile of the F distribution with 2 and $n-2$ degrees of freedom. Show that the band is the narrowest where the explanatory variable is equal to its sample average. [6+2=8]

4. Given n observations satisfying the linear model

$$E(\boldsymbol{y}) = \beta_0\mathbf{1} + \beta_1\boldsymbol{x}_1 + \beta_2\boldsymbol{x}_2 + \beta_3\boldsymbol{x}_3 + \boldsymbol{\epsilon}, \quad D(\boldsymbol{y}) = \sigma^2\boldsymbol{I},$$

you have to test the the hypothesis $\mathcal{H}_0 : (\beta_1 : \beta_2 : \beta_3) = c(\beta_{10} : \beta_{20} : \beta_{30})$, where β_{10}, β_{20} and β_{30} are specified constants and c is fixed but unspecified. Assume that all the β-parameters are estimable.

   (a) Show that $\mathcal{H}_0$ is a linear hypothesis.

   (b) If you have a statistical package which can do the usual computations for multiple linear regression, describe how you will use it to carry out the likelihood ratio test for $\mathcal{H}_0$. [3+4=7]

5. Consider the usual model for one-way classified data with n_i observations in the ith group. The following data summary are available for groups $i = 1, \ldots, t$: sample size (n_i), sum of y-values (a_i) and sum of squares of y-values (b_i).

   (a) Describe the ANOVA table and a test for the hypothesis of no difference among group means, by expressing all the quantities in terms of the *available* data.

   (b) If b_1 is missing, can you still carry out a test of the hypothesis stated in part (a)? Justify your answer.

   (c) If a_1 is missing, can you still carry out a test of the hypothesis stated in part (a)? Justify your answer. [3+3+3=9]

6. Given the balanced two-way classified data model (with more than one observation per cell) with independent and normally distributed errors, suppose that we have to test if there is significant interaction between the first treatment and the second block. Formulate this problem as a testable hypothesis involving the model parameters, and give a t-statistic for testing it, with justification. [5]

7. A set of t drugs, each having d dose levels, are administered to subjects divided into b blocks. Each dose level of every drug is applied to m subjects of every block, while the allocation is completely random. The response is a measure of degree of relief caused by the drug.

   (a) Write down a suitable nested model for this set-up and derive the ANOVA table.

   (b) How will you test the hypotheses that the various dose levels of Drug 1 do not have different effects? [2+3=5]

8. Consider the analysis of covariance model,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{z}\eta + \boldsymbol{\epsilon}, \quad E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad D(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I},$$

where the matrix $\boldsymbol{X}$ is the usual design matrix of two-way classified data with single observation per cell, the elements of $\boldsymbol{\beta}$ are the corresponding effects and the vector $\boldsymbol{z}$ is a covariate vector with covariate effect η.

   (a) Find the BLUEs of η and $\boldsymbol{X\beta}$, and the fitted values.

   (b) Give simple expressions for $Var(\hat{\eta})$ and $D(\widehat{\boldsymbol{X\beta}})$. [3+4=7]

9. Consider three-way classified count data from a survey, where the respondents are classified by Gender G (male/female), Drug use status D (never used/used at least once) and Community C (rural/urban).

   (a) Describe the saturated log-linear model for this data with suitable constraints.

   (b) Interpret all the parameters of the model subject to these constraints.

   (c) Describe the homogeneous association model, using the parameters described in part (a).

   (d) Describe very clearly how you would test for the goodness of fit of the model of part (c). [2+2+2+3=9]

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2003-2004

B. Stat. (Hons.) III Year
Statistical Inference I

Date : 5.12.2003 | Maximum Marks : 60 | Duration : 3 Hours

1. Let X be a random variable taking values $-1, 0, 1, 2, \ldots\ldots$ with probabilities

$$P_\theta[X=-1]=\theta,\ P_\theta[X=x]=(1-\theta)^2\theta^x,\ x=0,1,\ldots\ldots$$
$$0<\theta<1$$

Show that X is a boundedly complete statistic for the above family of distributions, but X is not complete.

[8]

2. Let $f_\theta(x)$ be the probability density function of a random variable X indexed by the parameter θ such that $\{x \mid f_\theta(x) > 0\}$ does not depend on θ. Suppose that for a random sample $\underset{\sim}{X}=(X_1,\ldots\ldots,X_n)$ of n observations from this distribution, there exists a function $T(\underset{\sim}{X})$ such that for any two realizations $\underset{\sim}{x}=(x_1,\ldots\ldots\ldots,x_n)$ and $\underset{\sim}{y}=(y_1,\ldots\ldots\ldots,y_n)$ of $\underset{\sim}{X}$,

$f_\theta(x_1,\ldots\ldots,x_n)=Kf_\theta(y_1,\ldots\ldots,y_n)$ if and only if $T(\underset{\sim}{x})=T(\underset{\sim}{y})$. Then $T(\underset{\sim}{X})$ is a minimal sufficient statistic for θ. (Here K is a constant independent of θ, and $f_\theta(x_1,x_2,\ldots\ldots,x_n)$ is the joint density of $X_1,X_2,\ldots\ldots,X_n$ at $x_1,x_2,\ldots\ldots\ldots,x_n$).

[8]

3. Consider the squared error loss $L(\theta,\delta)=(\delta-\theta)^2$. Suppose that δ is an admissible estimator of θ with respect to the above loss function. Let $R(\theta,\delta)$ be the risk function of the estimator δ. Suppose δ' is any other estimator satisfying $R(\theta,\delta)=R(\theta,\delta')$ for all θ.
Show that $\delta=\delta'$ w.p.l. (with probability 1).

[5]

4. Consider the $E(a,b)$ family where the density is given by

$$f(x)=\frac{1}{b}e^{-(x-a)/b} \qquad x\geq a.$$
$$-\infty<a<\infty$$
$$b>0$$

[P.T.O.]

Find the UMP level α test for testing

$$H_0 : a = a_0, \quad b = b_0 \qquad vs \qquad H_1 : a < a_0, \quad b < b_0$$

based on a random sample of size n.
(UMP : uniformly most powerful).

[6]

5. Let X have a geometric distribution with probability mass function

$$f_p(x) = p(1-p)^x \quad , \quad x = 0,1,2,.....$$

Based on a single observation from this distribution, derive the critical function for the UMP unbiased test of level $\alpha = 0.2$ for $H_0 : p = 1/2$ against $H_1 : p \neq 1/2$.

[8]

6. $X_1,........,X_m$ and $Y_1,........,Y_n$ are independent random samples from exponential distributions with rate parameters λ_1 and λ_2 respectively.
For the hypotheses

$$H_0 : \lambda_1 = \lambda_2 \quad vs \quad H_1 : \lambda_1 > \lambda_2$$

derive the uniformly most powerful unbiased test at level α.
(The density of an exponential random variable with rate parameter λ is $f(x) = \lambda e^{-\lambda x}, x > 0.$)

[10]

7. Let $X_1,........,X_n$ be a random sample from a double exponential distribution with location parameter θ.

$$f_\theta(x_1,......,x_n) = \frac{1}{2^n}\exp\left(-\sum_{i=1}^{n}|x_i - \theta|\right), \quad -\infty < x_i < \infty$$

is the joint density.

Consider testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. You may assume that one continuous derivative of the power function of any test exists and the derivative may be taken inside the integral.

(a) Find the locally most powerful (LMP) test of size α for the above hypotheses for $n \geq 2$ and $0 < \alpha < 1$, and give the form of the critical region.

(b) In general (not necessarily just for the above model), can you describe a way of finding approximate critical values for an LMP test for large n?

[4+3=7]

8. Consider a random sample of size n from a uniform $(0,\theta)$ distribution. Let $X_{(i)}$ be the ith order Statistic and Y be the sample range.

(a) Show that if ξ is given by the relation $\xi^{n-1}[n-(n-1)\xi] = \alpha$, then $\left(Y, \frac{Y}{\xi}\right)$ is a confidence interval for θ with confidence coefficient $(1-\alpha)$.

(b) Show that $\left(X_{(n)}, \frac{X_{(n)}}{\alpha^{\frac{1}{n}}}\right)$ is also a confidence interval for θ with confidence coefficient $1-\alpha$. How does the width of this confidence interval compare with that in part (a)?

[4+4=8]

***

Indian Statistical Institute
First Semester Examination: 2003-2004

B.Stat. (Hons.) III Year
Introduction to Sociology & Sociometry

Date : 12.12.2003 | Maximum Marks: 50 | Duration: Two Hours

Note: Answer Groups A and B in separate booklets.

Group - A

Q. 1) Answer <u>any two</u> of the following: (10 x 2)=(20)

a) What are the three occupational classes in rural India as considered by R. K. Mukherjee? What is their meaning with respect to production relations?

b) What, according to F. G. Bailey, were the possible reasons for investment in land by the beneficiaries of the newly emerging economy in the village Bisipara?

c) What are the major features of caste-class divergence as observed by Andre Beteille with reference to Sripuram village at Tanjore district of Tamilnadu?

d) Discuss after D. N. Dhanagare the impact of globalization on peasant economy in developing countries.

Q. 2) Write short notes on <u>any two</u> (5x2)=(10)

a) Participant Observation

b) Experimental Design

c) Concept

d) Sanskritization

P.T.O

Group- B

Q.3) A small village consists of eleven households serially numbered as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11. Though small, all of them do not go for help to each other. Among them Sl. Nos. 1, 5, 7 and 9 help each other as and when required. So also do Nos. 6 and 9. But No. 10 goes to Nos. 7 and 9 for help, while No. 4 goes to Nos. 9 and 10. Sl. No. 6 goes to Nos. 3, 4, 5 and 7. Again, Nos. 3, 4, 11 and 8 go for help among themselves also. Lastly, it is also observed that No. 2 goes to Nos.5, 6, 7 and 11.

a) With the usual notations, obtain the values of 's_o' and 'm' for this village and standardize 's_o' with respect to 'm'.

b) Verify whether every household in this village can approach all other households for help either directly or through intermediaries.

c) Identify the 'most centrally located' household(s).

(3+6+3)=(12)

Q. 4) In a city, 500 persons were asked whether they preferred English as medium of instruction and 300 replied 'yes'. Out of 300 replying 'yes', 150 belonged to upper class families and the remaining 150 to lower class. Reporting this data the survey report claims that 'class' carries no influence upon preference for English. Mr. J, a journalist, supports the claim, but Mr. Z, a sociologist, does not. Whom would you support? (It is known that, out of 500 persons 200 belongs to upper class and 300 to lower class.) Give reason(s).

(4+4)=(8)

***

Indian Statistical Institute

1st Semester Examination : (2003-2004)

Biology-I, B.Stat. II yr

Date 12.12.03 Full marks: 50 Duration: 2hr 30min

(Answer any five, each question carries equal marks)

1. A geneticist crossed wild black-colored mice with white-colored mice and all progeny were black. These progeny were inter-crossed to produce an F2 generation, which consisted of 198 black and 72 white mice. Explain this result with reference to the laws of genetics.

2. How do you define a gene and an allele? Write down the genotypes that correspond to the A, B, O and AB blood groups. Clearly describe the dominance-recessive relationships among the alleles.

3. (a) If a man and a woman are heterozygous for a gene and if they have three children, what is the chance that all the three children will also be heterozygous?

(b) In a family of six children what is the chance that at least three are girls?

4. (a) What is "codon"? How many different "codons" and amino acids are present in our cells? How are amino acids and codons related in protein synthesis?

(b) Define with example: Phenotype, trait, haploid and diploid cells.

5. Describe digestion and metabolism of fat in respect of energy production.

6. Consider an autosomal locus with two alleles A and a. In four different populations, the frequencies of the allele A are 0.35, 0.3, 0.25 and 0.32 respectively. Develop the phylogenetic relationship between the four populations based on allele frequencies at this locus.

INDIAN STATISTICAL INSTITUTE

Semestral I Examination: (2003-2004)
B.Stat (Hons.) III Year
Economics III

Date: 12 December 2003 | Maximum Marks: 100 | Duration: 3 Hours

Note: This paper carries 115 marks. Answer as many questions as you like.

The maximum you can score is 100.

1. Discuss alternative distributed lag models (finite and infinite) used in econometric analysis. Give intuitive and/or economic justifications for such models. [13+10=23]

2. Discuss the estimation methods for the following two situations. Carefully state the assumptions you need.

   (i) $y_i = \alpha + \beta x_i + \varepsilon_i$,

   $$E(\varepsilon_i) = 0, V(\varepsilon_i) = \begin{cases} \sigma_1^2 \text{ if i odd} \\ \sigma_2^2 \text{ if even} \end{cases}, \quad r(\varepsilon_i, \varepsilon_j) = \rho^{|j-i|}, \; 0 < \rho < 1; \; \text{i, j} = 1,\ldots,2n$$

   (ii) $y_i = (\alpha + \beta x_i)\, \varepsilon_i$, $\varepsilon_i \sim Lognormal\,(0, \sigma^2)$, i=1,…,n. [13+10=23]

3. Compare between the two tests/methods mentioned in each of the following three cases.

   (i) Durbin – Watson statistic and Von-Neuman ratio for detecting autocorrelation.

   (ii) Principal component method and Ridge regression for solving the multicollinearity problem.

   (iii) Probit and Logit methods for regressing dummy dependent variables.

   [8 × 3=24]

4. 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 per-capita nonfarm housing stock

P.T.O

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 wage of unskilled construction workers
s = ratio of wages of skilled to unskilled workers

The follwing equations were fitted with the data on the above variables (standard error in parentheses). Interpret each as depicting the demand factors or supply factors or both, and discuss whether each of the equations makes sense and is a 'good' equation.

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

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

$$\frac{H_g}{H} = .225\,\frac{R}{p} - .00608r \qquad R^2 = .714$$
$$\quad (.046) \quad (.00451)$$

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

[20]

5. (i) Define rank and order conditions for identifiability in a simultaneous equations system.

(ii) Discuss the identifiability of the parameters in the follwing system

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

Where C + I = Y, P+W = Y and $I = k - k_{-1}$. Further, C, I, W, P, Y, k are endogenous variables.

(ii) Suggest an estimation procedure for the above model using the I.V. method.

[8+10+6= 24]

INDIAN STATISTICAL INSTITUTE
Semestral Examination: (2003-04)
B-Stat-III, GEOLOGY.
Date: 12.12. 03, Marks: 100, Duration: 3 hours
Answer Group – A and B in separate answer scripts

Group - A

1. Select the right answer: ------------------------(5)

A) The fold axis of a recumbent fold
I. Is horizontal
II. Is inclined
III. Is vertical
IV. can have any orientation

B) Schistocity is
I. a sedimentray Feature
II. a structural Feature
III. an igneous Feature
IV. None of the above

C) Vertical displacement of a faulted strata will be minimum in case of
I. a normal fault
II. a reverse Fault
III. a strike-slip fault
IV. None of the above

D) Garnet is
i. a metamorphic mineral
I. an igneous mineral
II. an authigenic mineral
III. None of the above

E) Mica-schist is a metamorphosed product of
I. Shale
II. Sandstone
III. Limestone
IV. None of the above

2. The folds having their bedding shape constant is known as _________ [Parallel folds / Similar folds]? Draw a picture of that fold. (2.5+2.5)=(5)

3. Draw a fold belonging to Class 1A. For Class 1A folds the dip isogons are _______ [converging/ diverging/ parallel]? (2.5+2.5)=(5)

4. Write short notes on:----------------------------(2.5+2.5)=(5)
I. Small circle and Great circle
II. Strike, Dip and Apparent Dip

P.T.O

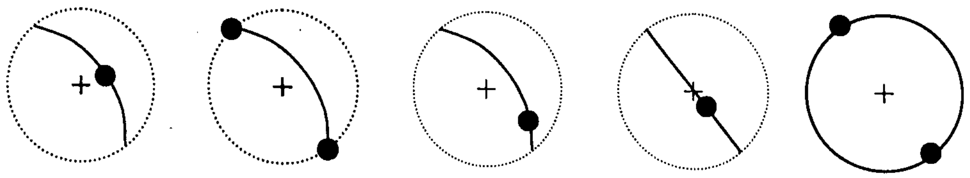
Group - B
(Answer any four)

1. Write down the major postulates of Darwin's theory of organic evolution. What were the paleontological evidences cited by Darwin to support the above theory. (10+10)=(20)

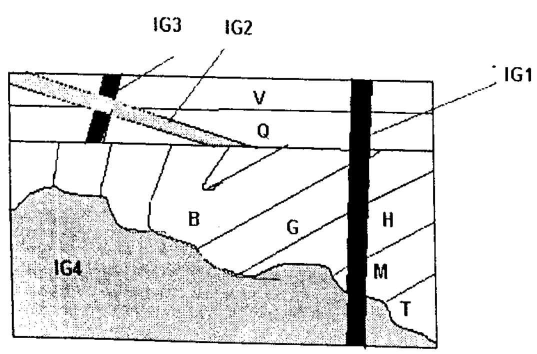
2. What is pangaea? What do you understand by the term "Gondwana"? Name the geological evidences that prove the existence of Gondwanaland. What is "*Glossopteris*" and how does it help to support the theory of plate tectonics? --------------- 4+4+4+(4+4)=(20)

3. A). The bedding plane normals of a cylindrical fold lies on a great circle in a stereogram – explain with suitable block diagram. --- (10)
B). What steps will you follow to rotate a line on a plane whose normal is given using the stereo-net. Justify your method – -- (10)

4. Name the folds depicted in the following setereograms showing the axial planes (soild curves) and fold axes (dots) and draw the representative block diagrams depicting their three dimensional form: --- 4+4+4+4+4=(20)



5. From the diagram below, 'IG's are igneous rocks while all others are sedimentary rocks. The fold is a synform. Write down the sequence of rock strata according to their relative ages. (Write the oldest rock at the bottom). Also, state the nature of unconformity surfaces that you think are present in the diagram. -- (20)



INDIAN STATISTICAL INSTITUTE
First-Semester Examination: 2003-2004

B.Stat. (Hons.) III Year
Introduction to Anthropology and Human Genetics

Date : 12.12.2003 Maximum Marks : 100 Duration 3 Hours

Note : Use separate answerscripts for Group A and Group B. Answer ***any five*** questions from each group. All questions carry equal marks.

GROUP A

1. Assess the role of Statistics in bio-anthropological studies. (10)
2. Explain briefly the different theories of evolution of man. (10)
3. Describe, in brief, the anatomical, morphological and biological changes that have occurred in the evolutionary development of mankind. (10)
4. (a)Phenotype = Genotype + Environment. Explain the statement with suitable examples. (5)
(b)Distinguish between fecundity and fertility. What are the major factors influencing fertility variation among human population groups? (2+3)
5. (a) What are the major health problems found to be experienced by populations inhabiting high altitude? Does physical environment play any role in the occurrence of such problems? (3+2)
(b) Briefly discuss, the on causes and consequences of variation among and within human populations in respect of physical growth, with suitable examples. (5)
6. Illustrate the interrelationship among genes, modernization and adiposity, with examples from recent human population biological studies. (10)
7. Choose the correct answer: (2 × 5)
   (a) Individuals born in a family which is known as his 'Family of Procreation': True / False
   (b) Polygyny is a form of marriage in which a man has more than one wife at a time: True / False
   (c) What kind of kin is the wife's brother? Consanguinal / Affinal
   (d) A family which consists of a married couple and their unmarried children can be defined as 'Extended family' : True / False
   (e) Clan or *Gotra* is a group in which members claim descent from a common ancestor, but cannot demonstrate it : True / False

P.T.O.

GROUP B

1. Point out the similarities and differences between mitosis and meiosis-I? What is the significance of meiosis-I? (10)

2. What is karyotype? State the number of chromosomes normally found in each group A-G, in the diploid set of the human male and human female explaining the basis by which the classification has been made. (2+8)

3. Define Hardy-Weinberg principle. Suppose the estimated frequencies of alleles A, B and O in a random sample of 1000 individuals from a population are, respectively, 0.21, 0.34 and 0.45. If marriages take place at random in this population, what are the expected frequencies of the four blood groups in this population? (4+6)

4. Give an example of single, rare, completely recessive inheritance, which is caused by an autosomal gene. Write the salient features of such kind of inheritance. (4+6)

5. What do you mean by chromosomal aberration? Give examples. (10)

6. In his theory of evolution, Darwin noted variation within species but could not explain the reasons of such variation because the laws of genetic principles were unknown. Could you genetic reasons for such variation? (10)

7. Write notes on: any two
   (a) Down's syndrome
   (b) Linkage
   (c) Multifactorial inheritance
   (d) Crossing over
   (e) Prokaryotes

(5 ×2

***

INDIAN STATISTICAL INSTITUTE

1st Semester Backpaper Examination (2003-04)

B.Stat. (Hons.) 3rd year

Linear Statistical Models

Maximum time: 3 hours   Date: 5.2.04   Maximum marks: 100

[This test is closed book. The total number of marks of all the questions is 100, and the maximum you can score is 45. Except for question 2, you can use any result that was proved in class.]

1. Four items having weights β_1, β_2, β_3 and β_4 are weighed eight times, each item being weighed twice. The measurements follow the model

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \end{pmatrix},$$

with uncorrelated errors having zero mean and variance σ^2. We denote this model by $(\boldsymbol{y}, \boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$.

   (a) Show that every linear parametric function (LPF) of the model is estimable.
   (b) When is $\boldsymbol{l}'\boldsymbol{y}$ the BLUE of its expectation?
   (c) When is $\boldsymbol{l}'\boldsymbol{y}$ a linear zero function (LZF)?
   (d) What is the BLUE of β_1?
   (e) What is the variance of the BLUE of β_1?
   (f) Find an unbiased estimator of σ^2. [2+3+3+2+3+3=16]

2. Let $\boldsymbol{z}$ be any vector of linear zero functions (LZF) of the linear model $(\boldsymbol{y}, \boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ such that its elements are uncorrelated, each having variance σ^2, and any other linear zero function can be written (with probability 1) as a linear combination of the elements of $\boldsymbol{z}$.

   (a) Show that the number of elements of $\boldsymbol{z}$ is $\rho(\boldsymbol{I}) - \rho(\boldsymbol{X})$.
   (b) Show that $\boldsymbol{z}'\boldsymbol{z}$ is equal to the sum of squares of the least squares residuals of the model. [7+5=12]

P.T.O

3. If simultaneous confidence intervals have to be provided for the means of all the observed responses in linear regression (assuming normal errors), which of the three following confidence intervals should be used: (a) Bonferroni, (b) Scheffé, (c) Maximum modulus-t? Why? [5]

4. Suppose that you have data from two linear regression models each with one independent variable, $(\boldsymbol{y}_1, \boldsymbol{X}_1\boldsymbol{\beta}_1, \sigma^2\boldsymbol{I})$ and $(\boldsymbol{y}_2, \boldsymbol{X}_2\boldsymbol{\beta}_2, \sigma^2\boldsymbol{I})$. Let $\boldsymbol{\beta}_i = (\beta_{0i} : \theta_i)'$, $i = 1, 2$. The objective is to test the hypotheses of (a) *equality* of the regression lines, that is, $\mathcal{H}_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and (b) *parallelity* of the regression lines, that is, $\mathcal{H}_0 : \theta_1 = \theta_2$. How will you formulate the testing problems and proceed to solve them? Is it possible to form a single analysis of variance table which would lead to both the tests? [3+6+3=12]

5. Let there be n_i observations of the response (arranged as the $n_i \times 1$ vector $\boldsymbol{y}_i$) for a given combination of the explanatory variables $(\boldsymbol{x}_i)$, $i = 1, \ldots, m$, $n_1 + \cdots + n_m = n$. The plan is to check the adequacy of the model $(\boldsymbol{y}, \boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ through a formal test of lack-of-fit, assuming normal errors. Here, $\boldsymbol{y} = (\boldsymbol{y}_1' : \cdots : \boldsymbol{y}_m')'$ and

$$\boldsymbol{X}' = (\boldsymbol{x}_1 \otimes \boldsymbol{1}_{n_1 \times 1} : \cdots : \boldsymbol{x}_m \otimes \boldsymbol{1}_{n_m \times 1}).$$

Assume that $m > r = \rho(\boldsymbol{X})$.

(a) Show that the model $(\boldsymbol{y}, \boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$ is a restricted version of another model, where the response for every given $\boldsymbol{x}_i$ is allowed to have an arbitrary mean.

(b) Obtain the error sum of squares under the unrestricted model (*pure error sum of squares*).

(c) Identify the restriction of part (a) as the hypothesis of adequate fit, and obtain an expression for the sum of squares for deviation from the hypothesis (*lack of fit sum of squares*).

(d) Construct the ANOVA table.

(e) Describe the GLRT for lack of fit. [3+3+3+2+1=12]

6. A set of t drugs, each having d dose levels, are administered to subjects divided into b blocks. Each dose level of every drug is applied to m subjects of every block, while the allocation is completely random. The response is a measure of degree of relief caused by the drug.

(a) Write down a suitable nested model for this set-up and derive the ANOVA table.

(b) How will you test the hypotheses that the t drugs do not have different effects?

(c) How will you test the hypotheses that the dose levels of none of the drugs have different effects? [3+3+3=9]

7. The following table gives data, on measured hemoglobin content in the blood of brown trout that were randomly allocated to four troughs. The fish in the four troughs received food containing various concentrations of sulfamerazine, 35 days prior to measurement. Assuming that the response (hemoglobin content) follows the usual one-way classified data model with normal errors, test the hypothesis that sulfamerazine has no effect on hemoglobin content of trout blood, using this data. [10]

Sulfamerazine content (grams per 100 pounds of fish)	Hemoglobin in Brown Trout blood (grams per 100 ml of blood)									
0	6.7	7.8	5.5	8.4	7.0	7.8	8.6	7.4	5.8	7.0
5	9.9	8.4	10.4	9.3	10.7	11.9	7.1	6.4	8.6	10.6
10	10.4	8.1	10.6	8.7	10.7	9.1	8.8	8.1	7.8	8.0
15	9.3	9.3	7.2	7.8	9.3	10.2	8.7	8.6	9.3	7.2

8. Describe the Analysis of covariance table for two-way classified data with single observation per cell when there is a single covariate. Adapt Tukey's one-degree of freedom test for interaction to the present situation, and indicate explicitly the test statistic, along with its null distribution. [4+11=15]

9. Consider three-way classified count data from a survey, where the respondents are classified by Gender G (male/female), Drug use status D (never used/used at least once) and Community C (rural/urban).

(a) Describe the saturated log-linear model for this data with suitable constraints.

(b) Interpret all the parameters of the model subject to these constraints.

(c) Describe the unsaturated model corresponding to conditional independence of G and D given C, using the parameters described in part (a).

(d) Describe very clearly how you would test for the goodness of fit of the model of part (c) without using iterations. [1+2+2+4=9]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2003-2004

B.Stat. (Hons.) 3rd Year. 2nd Semester

Statistical Inference II

Date: February 16, 2004 **Maximum Marks: 30** **Duration: 1 hour and 30 minutes**

- Answer all the questions.
- You must state clearly any result stated and proved in class you may need in order to answer a particular question.

---

1. (a) Define consistency of a test. (b) State an appropriate theorem that can be used to establish consistency of a test. (c) Establish consistency of the sign test. [3+4+5=12]

2. Show that Wilcoxon signed rank statistic is consistent for stochastically positive alternatives. [10]

3. Find the asymptotic relative efficiency of the sign test relative to the t-test for logistic distribution. [8]

************** *Best of Luck!* **************

INDIAN STATISTICAL INSTITUTE

Mid-semester Exam: (2003-2004)

B. Stat. III

Database Management System

Date: 20.2.04 Maximum Marks: 50 Duration: 2 hours

Note: Answer ALL questions.

1. A survey is conducted in a set of tribal villages to study their agricultural practices. A village is identified by a unique name. Each village has a number of houses identified by a house-no. Each house belongs to a certain tribe like Santhal, Munda, Onrao etc. In each house the no. of persons involved in agriculture may vary. Some of the houses has some non-agricultural income also. A village has three types of land namely, HF(highly-fertile), MF(moderately-fertile) and UF(unfertile). Each plot of land, identified by a plot no., has attributes land type, cropping practice (monocropping or multi-cropping) and plot area. The village authorities keep accounts of the crops cultivated in the villages.

Draw an E-R diagram for the above problem and construct all the tables.
Write SQL code to create the tables with appropriate constraints. [20]

2. Consider a hostel mess that has several members. A member is identified by a unique member-id. The name, address, course-name and the advance deposit of a member are also maintained by the mess. There are three types of breakfast, four types of lunch and dinner. A member can opt for any one of these types for a day. He can even skip a meal (breakfast or lunch or dinner). To compute the cost of a meal, a register of all the commodities (e.g., rice, wheat, vegetables, fish, egg etc.) is maintained. A commodity has a unique code and a name. Its relative quantity use (%) in a meal type is also to be kept. At the end of a month the total cost on each commodity is fed to the system. The cost of a meal type is computed by summing up the total cost on commodities, used in the meal type, and then dividing it by total number of that meal type.

Draw an E-R diagram for the above problem and construct all the tables.
Write SQL code to create the tables with appropriate constraints. [30]

Indian Statistical Institute
Introduction to Stochastic Processes
B. Stat 3rd Year
Midterm

Monday 23.2.2004, 10:30-1:30 Total Points $5 \times 6 = 30$

Answers must be justified with clear and precise arguments.

1. In a Markov chain with n states let $(x_1, \cdots, x_n)$ be a solution of the system of linear equations $x_j = \sum P(j,l)x_l$.

(a) Show that if $x_j \leq 1, \forall j$ then the states for which $x_r = 1$ form a closed set.

(b) If states E_j and E_k belong to the same irreducible set then $x_j = x_k$.

(c) In a finite irreducible chain the solution $\{x_j\}$ reduces to a constant.

2. A Markov chain has the following transition probability matrix:

$$\begin{pmatrix} 3/12 & 2/12 & 1/12 & 3/12 & 1/12 & 2/12 \\ 1/12 & 1/12 & 3/12 & 1/12 & 4/12 & 2/12 \\ 0 & 0 & 3/4 & 1/4 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 0 & 0 & 2/3 & 1/3 \end{pmatrix}$$

(a) If the initial distribution is $(1/2, 1/2, 0, 0, 0, 0)$ then find the expected time to leave the transient states.

(b) Find the probability of absorption into the classes $(3,4)$ and $(5,6)$ if the initial state is 1.

3. A die is consecutively turned from one face to any of the four neighbouring faces with equal probability and independently of the preceding turns. Suppose $X_0 = 6$. Find $\lim_{n\to\infty} P^n(6,6)$.

4. Consider a random walk starting from $S_0 = 0$, with $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$, $S_n = \xi_1 + \cdots + \xi_n$. Find an expression for $P^{(n)}(0,0)$ and decide if the chain (i.e. the random walk) is positive recurrent or null recurrent. (You may assume that the chain is recurrent and use the Stirling formula,

$$n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n}$$

meaning that the ratio of the two sides goes to 1 as $n \to \infty$.)

5. Suppose for a certain branching process $\mu < 1$. Show that if $X_0 = 1$, then the expected number of individuals that ever exist in the population is $\frac{1}{1-\mu}$.

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination 2003-2004

B. Stat. (Hons.) III Year
Design of Experiments

Date: 25.2.04 **Maximum Marks : 50** **Duration : 2 Hours**

Answer all questions.

1\. a) Define a completely Randomized Design (CRD) and characterize the set of all estimable parametric functions.

b) Show that for a CRD the treatment sums of squares can be partitioned into sums of squares due to V-1 mutually orthogonal treatment contrasts and for V=4, r_1= 4, r_2= 5, r_3=4, r_4= 5 construct such a set of mutually orthogonal contrasts.

$\overline{2+3} + \overline{6+7} = 18$

2 a) Define with example a Randomized Block Design and obtain an expression for its efficiency with respect to a completely randomized design.

b) Starting with a Latin Square Design with 5 treatments you delete one arbitrary row and consider the resultant design as a block design only, with the columns as blocks. Give the layout of such a design. Is this design (i) connected ? (ii) orthogonal ? (iii) balanced ? Obtain an expression for the average variance of all elementary treatment contrasts.

$\overline{3+3} + \overline{2+3+3+3+3} = 20$

3\. From the incidence matrix of a block design given below obtain (i) the set of all estimable elementary treatment contrasts (ii) the degrees of freedom associated with the adjusted treatment sums of squares and the adjusted block sums of squares. Suppose that $\tau_i - \tau_j$ is estimable, then show that the variance of its BLUE will be between $\frac{2\sigma^2}{\lambda_{max}}$ and $\frac{2\sigma^2}{\lambda_{min}}$ where λ_{max} and λ_{min} are the maximum and the minimum non zero eigenvalues of the C - matrix.

(In each case you have to give proper argument)

$$N = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$\overline{3 + 3 + 3 + 3} = 12$

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination: 2003–2004
B. Stat. (Hons.). 3rd Year
Statistical Inference 2

Date: April 23, 2004 **Maximum Marks: 55** **Duration: 3 and 1/2 hours**

- This question-paper carries 70 points. Answer as much as you can, but maximum you can score is 55.
- You must state clearly any result stated and proved in class you may need in order to answer a question.

1. Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. having exponential distribution with scale parameter $\sigma > 0$. Identify a suitable pivot T, and find a $1-\alpha$ confidence interval based on T having shortest length among all possible $1-\alpha$ confidence intervals based on T. [2+9=11]

2. Consider Spearman's rank correlation for a random sample of size n from a bivariate distribution $F(x, y)$. Find its asymptotic distribution when the marginals of F are assumed to be continuous and independent. [12]

3. Suppose $X_1, \ldots, X_n$ is a random sample from an unknown continuous distribution F. Consider the Kolmogorov-Smirnov one-sample statistic, based on $X_1, \ldots, X_n$, for testing $H_0 : F = F_0$ against $H_1 : F \neq F_0$. Show that its distribution under H_0 does not depend on F_0. [9]

4. Suppose $X_i, i = 1, 2, \ldots$, are independent and identically distributed (i.i.d.) having exponential distribution with scale parameter $\sigma > 0$. Consider the sequential probability ratio test (SPRT) for testing $H_0 : \sigma = 1$ against $H_1 : \sigma = 2$, where Wald's approximation for the boundaries are used with target strength (α, β), $\alpha = \beta = 0.05$. Compare the approximate average sample numbers under H_0 and H_1 with the minimum sample size required by the most powerful test for testing H_0 against H_1 with error probabilities at most α and β. [Hint: You can use (i) normal approximation to chi-square distribution, and (ii) $P(Z \leq 1.645) = 0.95$, where $Z \sim N(0, 1)$.] [12]

5. Suppose $X_i, i = 1, 2, \ldots$, are i.i.d. $\sim$ Poisson(θ), $\theta \in \Theta = (0, \infty)$. Consider the problem of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, where $\theta_0 < \theta_1$ and both are known. Show, without using Stein's lemma, that the stopping time corresponding to the SPRT for this problem terminates with probability one under any $\theta \in \Theta$. [9]

6. Suppose $X_i, i = 1, 2, \ldots$, are i.i.d. $\sim N(\theta, \sigma^2)$, $\theta \in \mathbb{R}, \sigma > 0$, both unknown. (a) Show that based on a fixed-sample size procedure it is *not* possible to obtain an estimator T of θ such that $\sup_{\theta,\sigma} E([T-\theta]^2) \leq L$, where $L \in (0, \infty)$ is known. (b) Describe briefly how you will obtain an estimator W of θ based on a sequential procedure such that $\sup_{\theta,\sigma} E([W-\theta]^2) \leq L$. [Note: For part (b), you are required only to give the main steps of your argument, but not any proof.] [10+7=17]

****** *Best of Luck!* ******

Indian Statistical Institute
B. Stat 3rd Year (2003-04)
2nd Semestral Examination
Introduction to Stochastic Processes

Date: 27.4.2004, 10:30-1:30 Total Points 80

Answers must be justified with clear and precise arguments. If you are using any theorem proved in class the conditions for the validity of that theorem must be verified. The maximum you can score is 70.

1. Let P be the transition probability matrix of an irreducible finite state space Markov chain and suppose $P^2 = P$. Show that $P_{ij} = P_{jj}, \forall i, j$. 8 pts.

2. Consider a regular $(2r+1)$ polygon with vertices $V_1, \cdots, V_{2r+1}$. At each V_k there is a nonnegative mass $w_k^{(1)}$, where the $w_k^{(1)}$'s satisfy $w_1^{(1)} + \cdots + w_{2r+1}^{(1)} = 1$. Obtain new masses $w_k^{(2)}$ as follows

$$w_k^{(2)} = \frac{1}{2}(w_{k-1}^{(1)} + w_{k+1}^{(1)}),$$

and so on, identifying w_{2r+2} with w_1 and w_{-1} with w_{2r+1}. Does $\lim w_k^{(n)}$ exist? If yes find it. 8 pts.

3. Consider a Markov chain with

$$p_{ij} = \binom{N}{j} \pi_i^j (1-\pi_i)^{N-j}, j = 0, 1, \cdots, N,$$

where

$$\pi_i = \frac{1 - e^{-(2ai/N)}}{1 - e^{-2a}}, a > 0.$$

Note that $0, N$ are absorbing states. 8 + 8 = 16 pts.

(i) Verify that e^{-2aX_t} is a martingale.

(ii) Show that the probability of absorption into state N starting at $k, 0 < k < N$ is

$$\frac{1 - e^{-2ak}}{1 - e^{-2aN}}.$$

4. (a) A continuous time Markov chain has two states labelled 0 and 1 respectively. The waiting time in both states are i.i.d. Exp(λ). The process starts in state 0 at time 0. Let $N(t)$ be the number of times the process has changed states in time $(0, t]$. Find the probability distribution of $N(t)$.

(b) A continuous time Markov chain has two states labelled 0 and 1 respectively. The waiting time in state 0 is Exp(λ) and the waiting time in state 1 is

 P.T.O

Exp(μ). The process starts in state 0 at time 0. Let $P_{0,n}(t)$ and $P_{1,n}(t)$ be the probabilities that there are n changes of states and the process is in states 0 and 1 at time t respectively. Find differential equations for $P_{0,n}(t)$ and $P_{1,n}(t)$. 8 + 8 = 16 pts.

5. (a) For a linear growth (i.e. $\lambda_n = n\lambda, \mu_n = n\mu$) birth and death process $X(t)$ with $\lambda = \mu$, write

$$u(t) = P(X(t) = 0 | X(0) = 1).$$

Find an integral equation for u by conditioning on the time of the first event in $(0, t]$ and from this integral equation show that $u(t)$ satisfies

$$u'(t) + 2\lambda u(t) = \lambda + \lambda u^2(t), u(0) = 0.$$

(b) Solve the above differential equation to show that

$$u(t) = \frac{\lambda t}{1 + \lambda t}.$$

Comment on the limiting behaviour of the system as $t \to \infty$. 8 + 8 = 16 pts.

6. (a) For a renewal process with distribution $F(t)$ consider $P_O(t)$ = P(number of renewals in $(0, t]$ is odd). Conditioning on the time of occurence of the first event show that

$$P_O(t) = F(t) - \int_0^t P_O(t-u)dF(u).$$

(b) Find $P_O(t)$ for a Poisson process with rate λ, and verify that it satisfies the above equation. 8 + 8 = 16 pts.

INDIAN STATISTICAL INSTITUTE

Semester Exam: (2003-2004)

B. Stat. III

Database Management System

Date: 29.4.04 Maximum Marks: 100 Duration: 3 hours

Answer all the questions

OPEN BOOK TEST

1. A survey is conducted in a rural area to study the employment of the villagers. A village is identified by a unique name and has a known area. Each village has a number of houses identified by a house-number. For every house, its address, covered area, open area are also known. In each house the members, identified by their names, are involved in different types of employment and have different yearly income. For every household-member, his or her name, age, gender are also obtained.

A Relational database is to be designed with the above data. For this, draw an E-R diagram and construct the tables. (You need not enter any data. List the tables along with their attributes) [15]

2. In an organization, the Salary Section maintains the following tables:

employee – contains details of employees like employee number (unique), employee name, surname, sex, age, address, section code, monthly basic salary, date of joining, designation;

section – contains details of sections like unique section code, section name, location;

pf – contains Provident Fund details like employee number, monthly P.F. contribution, P. F. balance.

P.T.O

Write SQL statements for the following: [5 x 5 = 25]

(a) List details of employees who have served less than five years in the company in increasing order of their ages. (Hint: Get System date and use Date Functions)

(b) Compute the total and average gross monthly salary of all the employees. (gross salary = basic salary + DA + HRA, where DA and HRA are 10% and 30% of basic salary respectively).

(c) List the section names for employees who are 'Computer operator'.

(d) Count section-wise number of employees for sections with more than 20 employees.

(e) Increase by 10% the basic salary of the employees who have served more than ten years in the company.

Write PL/SQL blocks for the following:

(f) Split the table ***employee*** into two. Employees with monthly basic salary >= Rs.10,000 will go to table "HIG" and employees with monthly basic salary < Rs.10,000 will go to table "LIG". (You may use CURSOR and FETCH command) [7]

(g) Calculate the projected Income Tax for each employee and write into a separate table against the employee number. The I.T. has to be computed on annual gross income (10% and 8% of gross for male and female respectively). An employee gets a rebate of 20% of the P.F. contribution on I.T. [8]

3. Consider the following relation scheme for a teaching organisation.

Teacher(T_name, <u>T_roll</u>, T_dept)
Department(<u>D_no</u>, D_name, D_location, D_head)
Course(C_name, <u>C_no</u>, C_location, C_dep)
Teaches(<u>T_roll, C_no</u>)
Student(<u>T_roll, S_name</u>)

Primary keys have been underlined.

The organisation has branches in several locations and in each location there are several departments. Each department has a name(D_name), a head(D_head) and a unique identification number(D_no). D_location is the name of the location where the department is located. There are several teachers in the organisation, each teacher is identified by a unique identification number T-roll and belongs to one department whose number is in T_dept. T_name stores the name of the teacher. D_head is the T_roll of the head. Each course has a name (C_name), a unique identification number C_no, location where offered C_location and identification number of the organising depatrment C_dep. Teaches relation has identification number of the teacher(T_roll) who teaches the course identified by C_no. A course may be taught by more than one teacher and one teacher may teach several courses. Student relation has teacher's roll number and names of students taught by him.

Write an expression in relational algebra for each of the following queries.

(1) Retrieve names of teachers who belong to 'CSSC' department.
(2) For every course offered at 'Bangalore' list the course name, organising department name along with name of the head.
(3) Retrieve names of teachers who have no students.
(4) List names of heads who have at least one student.

[4x5 = 20]

4. A library maintains a catalogue of books and a list of members who can borrow books from the library. Each book is assigned one unique accession number. Several copies of the same book may be available but they will have different accession numbers. Each member will have one unique identification number. For storing information on books and members a relation scheme has been designed as follows.

BM_scheme=(Book_Title, Book_Author, Book_Accession_No, Book_Status, Member_Name, Member_Dept, Member_Id)

Book_status can take one of the following values.
IO if the book is issued out,
or AV if the book is available.

(1) List the shortcomings of the above relation scheme.
(2) List the functional dependencies of the above scheme.
(3) Find the BCNF decomposition of the above scheme, explaining the procedure in each step.
(4) Is the decomposition loss-less? Justify your answer.
(5) Is the decomposition dependency preserving? Justify your answer.

[5x3 = 15]

5. Consider the relations r1(A,B,C), r2(C,D,E) and r3(E,F) with primary keys A, C and E respectively. Assume that r1 has 1000 tuples, r2 has 1500 tuples and r3 750 tuples. Estimate the size of r1 |X| r2 |X| r3 and give an efficient strategy for computing the join.

[10]

INDIAN STATISTICAL INSTITUTE
Second Semestral Examination: 2003-04
B. Stat. III Year
Design of Experiments

Date: 5.5.04 | Maximum Marks: 100 | Duration: 3½ Hours

Answer Q.6 and any four from the rest.

1.(a) Define "connectedness' of a block design.

(b) Show that a necessary and sufficient condition for all elementary treatment contrasts to be estimated with the same precesion from a "connected" block design with v treatments is that the C-matrix of the design has $(v-1)$ equal non-zero eigen values.

(c) Form an analysis of variance appropriate to the block design whose incidence matrix $N = 2J_{v\times b}$ and compare it with that of a design whose incidence matrix is $J_{v\times b}$ in terms of average variance of elementary treatment contrasts.

2+10+8=20

2.(a) Define a pair of Mutually Orthogonal Latin Squares (MOLS) if order v.

(b) Show that a complete set of MOLS can always be constructed when v is a prime or prime power.

(c) Construct a pair of MOLS of order 12.

2+8+10=20

3. A study is to be conducted to find the work schedule that leads to the best job satisfaction for a group of technicians. Four schedules were used in a study:

A : 4-day week, day shift

B : 4 day week, evening shift

C : 5 day week, day shift

D : 5 day week, evening shift

Four technicians are available for the study and each is supposed to use each work schedule for a week.
(a) What are the possible sources that can influence job satisfaction ?

(b) Suggest an appropriate design to conduct this study and write down the underlying model. Give a layout of the design.

(c) Suppose that a questionnaire is administered at the end of each week and a job satisfaction score in week 4 for technician 4 is missing. How will you analyse the resulting data to obtain correct error sums of squares?

[P.T.O]