INDIAN STATISTICAL INSTITUTE

B.Stat. III
Database Management System
Semestral Examination

ate : 09.05.03 Time : 3 hours
Note : Answer any part of any question , mamimum you can score is 100.)

1. Design a relational database for the bank described below. Your design must include,
   i) an ER diagram, 10
   ii) entity relations and relational relations in BCNF, 10
   and take care of the queries listed in part iii). 4 X 10 = 40

A bank has a number of account holders.

For each account holder the bank records the name, address and the name of a referee who must be another account holder.

To each account the bank gives a unique account number. An account holder can not have more than one account.

An account can be of two types i) with cheque facility, and, ii) without cheque facility.

It is necessary to keep a minimum amount of Rs.1000 in an account with cheque facility until the account is closed. For an account without cheque facility it is necessary to keep a minimum amount of Rs.100 only.

The bank also gives loans to persons on monthly repayment terms. The number of monthly instalments per loan and amount per instalment is decided at the time of loan.

Non account holders can also borrow from the bank.

---

iii) The database is to be accessed and used by the manager of the bank only. He updates transactions of account holders and loan repayments. He also keeps note of defaulters of repayment.

At any time he may like to know the following :

(Formulate the following queries in relational algebra or in SQL).

a) a list of all account holders showing name and address,

b) a selected account showing current amount of money,

c) a list of all account holders without cheque facility,

d) a list of account holders whose account has less than the minimum requisite amount + Rs.100,

e) a list of account holders who have also taken loans,

f) a list of borrowers who are defaulters,

g) a list of account holders who are referee to some account holder,

h) a list of borrowers who are going to repay their last instalment,

i) a list of borrowers whose monthly instalment is more than Rs.10000,

j) a list of non account holder who has borrowed more than Rs.50000.

P.T.O

2. Consider relations r1(X,Y,Z), r2(Z,W,S) and r3(S,T) with primary key X,Z and S respectively. Assume that r1 has 2000 tuples, r2 has 1500 tuples and r3 has 1000 tuples. Estimate the size of the natural joi of r1 and r2 and r3 . Give an efficient strategy for computing the join. (4

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

```
sigma p ( r1 U r2 U r3 )   where p =( Y > 3 )and( W < 1 )and( T =

sigma p ( r1 - r2 )        where p =( Z = 4 )

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

pie n ( sigma p ( r2 U r3 ) )  where p =( W < 1 )and( T = 5) and
                                     n = r2.S
pie n1 (  pie n2 ( sigma p   ( r1 U r2 U r3)))
                   where p = ( r1.Z = r2.Z - 3)and( r2.S = r3.S + 2)
                   and   n2 = ( r1.Z,r2.Z,r2.S,r3.S)
                   and   n1 = ( r1.Z,r3.S)                     (10)
```

Consider that in the relations above there are no primary keys except the entire scheme. Let V(Z,r1) be 600, V(Z,r2) be 900, V(S,r2) be 100 and V(S,r3) be 200. Give an efficient strategy for computing the natural join of r1,r2 and r3. (6)

3. In the context of logbased crash recovery of a transaction Ti explain the meaning of

undo(Ti), redo(Ti), <Ti,Start>, <Ti,Commit> (8

Let values of three data items A,B and C be respectively 500,1000 and 1250 respectively.

Consider the following transactions T1 and T2 and corresponding system log.

```
      T1                            System Log
   read(A,a1)                    < T1,starts >
   a1=a1-100
   write(A,a1)                   < T1,A,500,400 >
   read(B,b1)
   b1=b1+100
   write(B,b1)                    < T1,B,1000,1100 >
                                  < T1,commits >
      T2
   read(C,c1)                     < T2,starts >
   c1=c1-50
   write(C,c1)                    < T2,C,1250,1200 >
                                  < T2,commits >
```

What recovery action is needed when the system comes back if a crash occurs

i) just after the statement write(A,a1) is executed,
ii) just after the log record < T2,commits > is written to stable storage,
iii) just after the log record for write(B,b1) has been written to stable storage,
iv) just after the statement write(B,b1) has been executed,
v) just after the log record for write(C,c1) has been written to stable storage,
vi) just after the statement write(C,c1) has been executed, (12)

What will be displayed by the transactions T1 and T2 if their execution order is as per the following schedule assuming A and B have values 500 and 1000 respectively

```
          T1                         T2
   read(A)
   A=A-100                     read(A)
                               temp=A*0.1
                               A=A-temp
                               write(A)
                               read(B)
   write(A)
   read(B)
   B=B+100
   write(B)
   display(A+B)
                               B=B+temp
                               write(B)
                               display(A+B)                  (4)
```

What is a shared lock and what is an exclusive lock ? (4)

What will be displayed after the following schedule ?

```
        T1                        T2
 ------------------------------------------
   LX(B)
   read(B)
   B=B-100
   write(B)
   UN(B)
                             LS(A)
                             read(A)
                             UN(A)
                             LS(B)
                             read(B)
                             UN(B)
                             display(A+B)
   LX(A)
   read(A)
   A=A+100
   write(A)
   UN(A)
   diaplay(A+B)                                     (4)
```

What will happen if the following schedule is executed?

```
        T1                           T2
 ---------------------------------------------
    LX(B)
    read(B)
    B=B-50
    write(B)
                                LS(A)
                                read(A)
                                LS(B)
    LX(A)                                          (4)
```

Explain what is timestamp ordering. (4)

INDIAN STATISTICAL INSTITUTE
B.STAT-III (2002-03)
STATISTICAL INFERENCE-II (Semestral Examinations)
Time:3 hours Max marks:60

Date:6 May, 2003.

Note: Answer as much as you can. The maximum you can score is 60.

1. Let (X_1, Y_1), (X_2, Y_2) be 2 pairs of independent observations from a bivariate c.d.f. $F(.,.)$. Let τ denote the Kendall's coefficient of association between X and Y defined by

$$\tau = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

(a) Show that if F is the bivariate Normal $N_2((0,0), (\sigma_X^2, \sigma_Y^2), \rho)$ -c.d.f. then

$$\tau = \frac{2}{\pi}\arcsin \rho.$$

(b) Suppose that (X_1, Y_1), (X_2, Y_2), $\cdots$, (X_n, Y_n) are *iid* observations from a bivariate c.d.f. F for which the marginals of X and Y have continuous distribution.

(i) Show that $\widehat{\tau} = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} A_{ij}$, where

$$A_{ij} = \begin{cases} +1 & \text{if } (X_i - X_j)(Y_i - Y_j) > 0 \\ -1 & \text{if } (X_i - X_j)(Y_i - Y_j) < 0 \end{cases}$$

is an unbiased estimator of τ.

(ii) Describe how $\widehat{\tau}$ can be used to test the null hypothesis H of independence of X and Y. Find the exact null distribution of $\widehat{\tau}$ when $n = 4$.

[4+(2+4)]

2. Let $X_1, X_2, \cdots, X_n$ be *iid* observations from a continuous c.d.f. F.

(a) Define the one sample Kolmogorov-Smirnov statistics D_n, D_n^+ to test the null hypothesis $H: F = F_0$. Show that the null distribution of D_n, D_n^+ is the same for all continuous c.d.f.'s F_0.

(b) Let $U_1 < U_2 < \cdots < U_n$ denote the order statistic of a sample of size n from uniform distribution $U(0,1)$. Show that

$P(D_n^+ \le c \,|\, H) = P(U_i \ge \frac{i}{n} - c \text{ for } i = 1, 2, \cdots, n)$, for $0 \le c \le 1$.

Hence show that

$$P(D_n^+ \le c \,|H) = n! \int_{a_n}^1 \int_{a_{n-1}}^{u_n} \cdots \int_{a_2}^{u_3} \int_{a_1}^{u_2} du_1 du_2 \cdots du_n$$

where $a_i = max(\frac{i}{n} - c, 0)$ and $0 \le c \le 1$. Illustrate the use of this formula by evaluating $P(D_n^+ \le c \,|H)$ for all values of $0 \le c \le 1$, when $n = 2$.

[5+10]

3. Let $X_1, X_2, \cdots,$ be *iid* random variables with c.d.f. F. Let

$$U_n(X_1, X_2, \cdots, X_n) = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < i_2 < \cdots < i_m \le n} h(X_{i_1}, X_{i_2}, \cdots, X_{i_m})$$

be a U-statistic based on a symmetric kernel h for which $E(h(X_1, X_2, \cdots, X_m)) = 0$ and $E(h^2(X_1, X_2, \cdots, X_m)) < \infty$, $1 \le m \le n$. Let for $1 \le c \le m$, $Cov(h(X_1, X_2, \cdots, X_m), h(X_1, X_2, \cdots, X_c, X_{m+1}, \cdots, X_{2m-c})$ be denoted by σ_c^2.

(a) Show that $\sigma_c^2 \ge 0$ for $1 \le c \le m$.

(b) Show that $\lim_{n \to \infty} nVar(U_n) = m^2 \sigma_1^2$.

Please Turn overleaf

(c) Let $V_n = \sum_{i=1}^{n} E(U_n | X_i)$. Show that $V_n = \frac{m}{n}\sum_{1}^{n} h_1(X_i)$, where
$h_1(a) = E(h(a, X_2, X_3, \cdots, X_m))$.

(d) Show that $E(U_n - V_n)^2 = E(U_n^2) - E(V_n^2)$. Hence show that

$$\sqrt{n}(U_n - V_n) \xrightarrow{P} 0 \text{ , as } n \to \infty.$$

[2+3+3+7]

4. To test the the effectiveness of vitamin B_{12} in treating pernicious anaemia, the treatment is applied to 10 patients suffering from the disease. The following data gives the increase in hemoglobin levels of the subjects after the treatment.

increase in hemoglobin levels(gms/100 cc) of 10 subjects

-1.1 ,2.1, 1.5, -0.7, 3.4, -2.1, 0.6, 3.0, -0.1, 2.5.

(a) Use Wilcoxon signed rank test to test the hypothesis of no treatment effect against an alternative of positive treatment effect at a level α closest to 0.05 that is attainable by a non-randomised test.
(b)Consider an alternative distribution of the observations $Z_i,\ i = 1, 2, \cdots, 10$ for which $P(Z_1 > 0) = 0.84,\ P(Z_1 + Z_2 > 0) = 0.92,\ P(Z_1 + Z_2 > 0,\ Z_1 + Z_3 > 0) = 0.86$.
Find the power of the test in (a), using Normal approximation against this alternative.

[6+9]

5. Let $Z_1, Z_2, \cdots,$ be be iid random variables such that $P(Z_1 \neq 0) > 0$. Let $S_n = \sum_{1}^{n} Z_i$, $n = 1, 2, \cdots$ and let $a < 0 < b$ be given constants. Let N be a stopping time defined as follows:

$N = n$ if and only if $a < S_m < b$, for $1 \leq m \leq n-1$ and either $S_n \leq a$ or $S_n \geq b$ for $n = 1, 2, \cdots$.

(a) Show that $P(N < \infty) = 1$.Show that $E(N^k) < \infty$, for all $k = 1, 2, 3, \cdots$.
(b) Let t be a real number such that $\phi(t) = E(e^{tZ_1}) < \infty$. Show that

$$E(e^{tS_N}(\phi(t)^{-N}) = 1$$

[6+9]

****

TABLE H. Wilcoxon signed-rank distribution: [redacted] P[T⁺ ≤ v]

v \ N	1	2	3	4	5	6	7
0	.5000	.2500	.1250	.0625	.0313	.0156	.0078
1	1.0000	.5000	.2500	.1250	.0625	.0313	.0156
2		.7500	.3750	.1875	.0938	.0469	.0234
3		1.0000	.6250	.3125	.1563	.0781	.0391
4			.7500	.4375	.2188	.1094	.0547
5			.8750	.5625	.3125	.1563	.0781
6			1.0000	.6875	.4063	.2188	.1094
7				.8125	.5000	.2813	.1484
8				.8750	.5937	.3438	.1875
9				.9375	.6875	.4219	.2344
10				1.0000	.7812	.5000	.2891
11					.8437	.5781	.3438
12					.9062	.6562	.4063
13					.9375	.7187	.4688
14					.9687	.7812	.5312

v \ N	8	9	10	11	12	13	14
0	.0039	.0020	.0010	.0005	.0002	.0001	.0001
1	.0078	.0039	.0020	.0010	.0005	.0002	.0001
2	.0117	.0059	.0029	.0015	.0007	.0004	.0002
3	.0195	.0098	.0049	.0024	.0012	.0006	.0003
4	.0273	.0137	.0068	.0034	.0017	.0009	.0004
5	.0391	.0195	.0098	.0049	.0024	.0012	.0006
6	.0547	.0273	.0137	.0068	.0034	.0017	.0009
7	.0742	.0371	.0186	.0093	.0046	.0023	.0012
8	.0977	.0488	.0244	.0122	.0061	.0031	.0015
9	.1250	.0645	.0322	.0161	.0081	.0040	.0020
10	.1563	.0820	.0420	.0210	.0105	.0052	.0026
11	.1914	.1016	.0527	.0269	.0134	.0067	.0034
12	.2305	.1250	.0654	.0337	.0171	.0085	.0043
13	.2734	.1504	.0801	.0415	.0212	.0107	.0054
14	.3203	.1797	.0967	.0508	.0261	.0133	.0067
15	.3711	.2129	.1162	.0615	.0320	.0164	.0083
16	.4219	.2480	.1377	.0737	.0386	.0199	.0101
17	.4727	.2852	.1611	.0874	.0461	.0239	.0123
18	.5273	.3262	.1875	.1030	.0549	.0287	.0148
19	.5781	.3672	.2158	.1201	.0647	.0341	.0176
20	.6289	.4102	.2461	.1392	.0757	.0402	.0209
21	.6797	.4551	.2783	.1602	.0881	.0471	.0247
22	.7266	.5000	.3125	.1826	.1018	.0549	.0290
23	.7695	.5449	.3477	.2065	.1167	.0636	.0338
24	.8086	.5898	.3848	.2324	.1331	.0732	.0392
25	.8437	.6328	.4229	.2598	.1506	.0839	.0453
26	.8750	.6738	.4609	.2886	.1697	.0955	.0520
27	.9023	.7148	.5000	.3188	.1902	.1082	.0594
28	.9258	.7520	.5391	.3501	.2119	.1219	.0676
29	.9453	.7871	.5771	.3823	.2349	.1367	.0765
30	.9609	.8203	.6152	.4155	.2593	.1527	.0863
31	.9727	.8496	.6523	.4492	.2847	.1698	.0969
32	.9805	.8750	.6875	.4829	.3110	.1879	.1083

r
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

r
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

INDIAN STATISTICAL INSTITUTE
Second Semestral Examination: (2002-2003)

B.Stat/ III year

Design of Experiments

Date: 2.5.03 Maximum Marks: 100 Duration $3\frac{1}{2}$ hours.

Note: Answer as many questions as you can. The question paper carries 110 marks, the maximum you can score is: 100. **Keep your answers brief and to-the-point.**

1. A study is planned to examine whether 4 gasoline additives differ with respect to the reduction in oxides of nitrogen. The investigator has chosen 3 models of cars and for each model, he has 4 cars available. He will use 4 drivers for the entire experiment. He believes that for each model, systematic differences are likely to occur in the cars' perfo-rmance.

   i)Suggest an appropriate design for the experiment.

   ii) Give a linear model and outline the ANOVA table for this experiment, showing the sources of variation, degrees of freedom and expressions for sums of squares.

   iii) Indicate how you would use ANOVA to answer the objective of the study. [3+(2+2+2+2)+4=15]

2. (i)When is a design said to be connected? With 3 treatments and 4 blocks of sizes 2, 4, 3 respectively, give a layout of a block design which is not connected. Compute the C-matrix of this design.

   (ii) State and prove a necessary and sufficient condition for a connected block design to be orthogonal. Show that a connected incomplete block design is necessarily non-orthogonal. With 4 treatments and 3 blocks, give a layout of a connected block design which is orthogonal and another which is not orthogonal. (Each block size is greater than one)

   [(2+2+2)+(4+1)+(2+2)=15]

3. While planning an experiment, explain how you may decide whether a nuisance factor should be used as a blocking factor or as a covariate. Explain your answer with an example of an experimental situation with two nuisance factors where one of the nuisance factors is to be used as a blocking factor and the other as a covariate.

   Consider the model $E(Y) = X\beta + C\gamma$ where Y is the response vector, X is the $n \times p$ design matrix and C is an $n \times k$ matrix with its ith column representing the n observations on ith covariate, $i = 1, \ldots, k$. It is also given that $rank(X) = r$ and $rank(X\ C) = r + k$.

   a) Let $\hat{\beta}_0$ be the least squares solution of the normal equation ignoring the covariates in the above model. Using suitable notations, obtain a solution

P.T.O

of β for the above model, showing the adjustment to be made due to the covariates.

b)Write down the expression for an unbiased estimator of γ. (no proof needed) Find an expression for its variance.

c) Write down the null hypothesis and the test statistic for testing the hypothesis that the adjustment for covariates is profitable. (no derivation required) [(3+2)+6+(2+3)+4=20]

4. An experimenter wants to study the effects of 2 different culture mediums and 3 different temperatures on the growth of a particular virus. The experiment is to be done over 3 days, taking 6 observations each day. It is difficult to switch from one culture medium to another quickly, but the temperatures can be easily varied.

   a)Suggest a possible design for this experiment, explaining how it is different from the usual factorial experiment.

   b) Give the model for analyzing the data from the experiment in (a) and write down (without proof) the test statistics for testing the hypotheses of interest. [8+7=15]

5. A factorial experiment with 3 factors A, B, C, each at 2 levels is to be planned. The experiment will be run over 4 days and in each day, only 4 treatment combinations may be tested. It is desired to have some information on **all** the effects with as much information as possible on the main effects. Among the interactions, interaction BC and ABC are of lesser interest than others.

   (a)Give the layout of a design for this experiment with a suitable confounding scheme.

   (b)Derive the relative efficiencies of estimates of interactions AB and BC compared to the main effect estimates.

   (c)If all the 8 treatment combinations could be observed in each day and the experiment was run over 2 days, what would be the relative efficiencies of main effect A and interaction AB compared to the corresponding estimates from the experiment in (a) above.

   [6+7+7=20]

6. An experiment with 3 factors A, B, C, each at 3 levels, will be carried out in blocks of size 3 each. It is decided to use 2 replicates. Suggest a suitable confounding scheme for this experiment. Give the treatment combinations in the **principal block** (i.e. the block with treatment 000) of **ANY ONE** replicate. [5+5=10]

7. Construct three 4×4 mutually orthogonal Latin Squares (m.o.l.s).

   Show that a set of 3 m.o.l.s of order 5 can always be extended to a complete set of 4 m.o.l.s. [8+7=15]

INDIAN STATISTICAL INSTITUTE
Second Semester Examination : 2002-2003

B. Stat. III Year
Introduction to Stochastic Processes

Date : 28.04.03 | Maximum Marks : 100 | Duration : 3 hours

1. Justify your answers.
2. Quote all theorems you use.

Total Marks : 110, Maximum Marks :100

1. Here is the transition matrix of a Markov Chain with states 1, 2, 3, 4, 5, 6.

$$\begin{pmatrix} 1/10 & 8/10 & 1/10 & 0 & 0 & 0 \\ 0 & 1/10 & 0 & 0 & 9/10 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 3/4 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9/10 & 0 & 0 & 1/10 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

(a) Describe all the communicating classes of the Chain.

(b) Explain which are recurrent classes and which are transient – Justify.

(c) For each transient state i and recurrent class C, Calculate the probability of eventually entering C starting from state i.

[7+7+6]

2. Consider the Chandrasekhar Chain describing the number of particles in a region : If there are k particles now, each of them independent of others decides to stay in the region with probability p. Further, independently , a number of particles enter the region, the number having distribution Poisson (λ) . All this happens in one unit of time. X_n is the number of particles in the region after n units of time. Here $0 < p < 1$ and $\lambda > 0$ are fixed.

(a) Obtain formula for p_{ij} $\quad i \geq 0, j \geq 0$.

(b) Obtain a formula for $p_{oj}^{(n)}$.

(c) Calculate $\underset{n}{Lt}\ p_{oj}^{(n)}$.

(d) Explain why $\underset{n}{Lt}\ p_{ij}^{(n)}$ exists for each i. Calculate this limit.

[5+7+3+5]

P.T.O.

3. Consider the descending stairs Chain : State space $= \{0,1,2,....,\}$. If you are at 0 go to i with probability $\alpha_i, i = 0,1,2,.......$. If you are at $i > 0$ then go to $i-1$. Here $\alpha_i > 0$ for each i.

(a) Show that the Chain has stationary distribution iff $\sum i\alpha_i < \infty$.

(b) Assume that the above condition holds. Calculate the transition probabilities for the reverse Chain. [10+10]

4. $(N_t)_{t\geq 0}$ is a homogeneous Poisson process with intensity 1. If k is even = 0, 2, 4 ,..... then events occurring in $[k, k+1)$ are NOT recorded. For k odd, an event occurring in $[k, k+1)$ is recorded, independent of others, with probability $1/k$.

Let X be the number of recorded events in [0,4] and Y be the number of recorded events in [3,6].

(a) Calculate the joint distribution of (X,Y)

(b) Calculate $E(X-Y)$ and Variance of $(X-Y)$. [10+10]

5. A barber shop operated by 2 barbers has room for at most four customers. Potential customers arrive at a Poisson rate of 3 per hour. Service times are independent exponential with mean half an hour.

(a) Formulate this as a Birth and Death Chain , explaining all the parameters.

(b) Calculate the steady state distribution.

(c) What is the expected number of customers in the shop.

(d) What proportion of time are both the barbers busy. [6+6+4+4]

6. Consider a sequence of independent throws of a fair die. Let X_n denote the sum of the first n scores obtained. Show that $\underset{n}{Lt}\, P(X_n \text{ is divisible by } 13)$ exists. Calculate it. [10]

****

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination:(2002-2003)

B.STAT III Year

Database Management Systems

Date: 28.2.03 Time : 3 hours

(You can answer any part of any question . Maximum you can get is 100.)

1. Trains are either local or express but never both. A train has a unique number and name. Stations are either express or local. A station has a name and address. All local trains stop at all stations. Express trains stop only at express stations. For each train at each station there is an arrival and a departure time. For each station for each train the previous and next stations are recorded.

Draw an ER diagram for the above description. Map the ER diagram an appropriate set of relations. (8+7)

From the relations form the following query using relational algebra.

a) Find minimum time taken to travel between station A and station B when you are travelling on one train only. 5

b) Find the list of all stations between station A and station B in all routes where at least one train has a stop. 5

c) Find minimum time taken to travel between station A to station B when you can change trains at intermediate stations. 5

d) Form the above query in SQL. (10 + 10 + 10)

2. From the above ER diagram develop a hierarchichal model and a network model. (5 + 10)

3. A hotel maitains a database for its room reservation and accounting system. The hotel has three types of rooms: dormitory, double- bedding and suite. Each of these type may be airconditioned or not. A guest in a dormitory is charged per bed per day basis, whereas for any other room the charge is on per day basis. So for each dormitory the number of beds available is recorded. Each type of room has a separate charge. The hotel extends a concession of 10% in its room rate for each type of room for a block booking of 10 consequetive days or more. Similarly 10% concession is also given for a group of 10 guests or more. For an individual, the group size attribute value is 1. The database has been developed using a relational database management package and the following relations are maintained.

Guest = (name, address, room-no, room-rate, date-of-entry)
Room = (room-no, room-rate, room-type)
Dormitory = (room-no, room-rate, room-type, no-of-beds)
Bill = (name, room-no, room-type, room-rate, date-of-departure, no-of-days-stayed, group-size, amount-to-be-paid)
eservation = (name, address, reservation-date, group-size, expected-arrival-date, expected-duration-of-stay, percentage-concession-available)

P.T.O

In order to avoid redundancy of data, re-design the relations. If necessary, define new relations and attributes. Some existing relation[s] and attributes may even be dropped if they can be computed from other available data using any standard query language.

Indicate the different dependencies considered, normalization procedures followed and assumptions made during the process of your de[sign].

30

INDIAN STATISTICAL INSTITUTE

BIII – Introduction to stochastic Processes

26-02-03 | Midsemestral Examination | Time : 3 hours
Maximum marks 100

Note :
Write legibly.
Make sure to note the question number in the margin before answering.
Must justify answers/steps.

1. Consider a markov chain with six states 1,2,3,4,5,6 and transition matrix

$$\begin{pmatrix} 1/3 & 0 & 0 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/6 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 1/4 & 0 & 0 & 0 & 3/4 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}$$

Describe the communicating classes of this chain.

For each transient state i and each recurrent class C calculate the probability that the chain eventually enters the class C starting from the state i.

[12+8]

2. Consider the simple symmetric random walk P on the integers – one unit forward or backward, each with probability 1/2. Suppose that $\pi = (\pi_n)_{-\infty<n<\infty}$ is a solution of $\pi P = \pi$.

For $n \geq 2$, calculate π_n in terms of π_0 and π_1.

Show that any non constant solution of $\pi P = \pi$ is unbounded.

Show that there is no stationary initial distribution for the chain.

[8+8+4]

3. Consider a Markov Chain with state space as set of integers. Let $Q_{ij}^{(n)}$ be the probability the chain is in state j on day n starting at i, *before returning to i earlier* . That is

$$Q_{ij}^{(n)} = P(X_m \neq i \text{ for } 1 \leq m < n; X_n = j || X_0 = i)$$

P. T. O

Prove the renewal equation: For any states $i \neq j$ and any integer $n \geq 1$,

$$Q_{ij}^{(n+1)} = \sum_{k \neq i} Q_{ik}^{(n)} P_{kj}$$

You must give a complete proof with justifications.

[15]

4. Consider a chain with three states 1,2,3 and transition matrix

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

Calculate the period of this irreducible chain.

Find the stationary initial distribution π.

[8 + 12]

5. Consider a usual deck of 52 playing cards arranged in a stack. I select a card from the stack at random and interchange it with the top card – that is – the selected card comes to the top and the top card goes to the place of the selected card. I continue doing this. We wish to understand as to what happens in the long run.

(i) Formulate this problem as a markov chain. Clearly explain your state space . How many state are there?

(ii) Explain your transition matrix. Take any state of your choice and write down all the states where you could go in the next step and for each such state the probability of making a transition to that state.

(iii) Show that the chain is doubly stochastic, that is column sums of your transition matrix are also one.

(iv) Show that this is an irreducible and aperiodic chain.

(v) Argue as to why the stack is randomly arranged after several interchanges.

[5× 5 = 25]

Indian Statistical Institute

DESIGN AND ANALYSIS OF EXPERIMENTS, BIII

Mid semester Examination, February 24, 2003, Maximum marks: 40

Answer all questions Time; 3 hours

Figures in the right margin indicate maximum marks

1. a) When is a block design said to be balanced?

   b) Give the layout of any block design for each of the cases below: (each design has 5 treatments (labeled 1, 2, 3, 4, 5), 6 blocks.

   (i) a design which is connected but not balanced

   (ii) a design which is not connected but all contrasts among treatments 1 and 3 and also all contrasts among treatments 2 and 4 are estimable

   (iii) a design which is orthogonal

   (iv) a design which is balanced and connected but not orthogonal.

   [2+2x4=10]

2. Consider a general block design in b blocks and v treatments. Let τ be the vector of treatment effects. Using usual notations, answer the following:

   a)Obtain the expression for the variance-covariance matrix of the estimates of the full set of orthonormal contrasts of τ.

   b) Obtain the expression for sum of squares for treatments(adjusted). Simplify this expression if the design is an orthogonal design.

   [4+3+3=10]

3. Define mutually orthogonal Latin Squares (m.o.l.s.). Can there exist 9 m.o.l.s. of order 9. Give complete argument to justify your answer.

   [3+5=8]

4. A marketing expert for a publishing house wants to measure reader preference for three different covers of the same paperback novel. She has chosen 10 cities and 3 newsstands in each city which are going to see the novel. She wants to use one of two experimental setups described below:

   a) In each city each cover is assigned randomly to one of the 3 newsstands. The number of books sold during a three-week period following the assignment is used to compare the effect of the cover on sale of the novel.

P. T. O

b) In each city each of the 3 newsstands will sell the book using each cover for one week(i.e. the trial extends over 3 weeks) in such a way that during a given week the 3 newsstands in a city will display book with a different cover. The same 3-week period will be used in all cities. Sales figures for each week will be used for the analysis.

For each of the two scenarios described above:

(i) Identify the experimental units

(ii) Give the model for each of the designs and the ANOVA Table showing sources of variation and d.f.

(iii) Indicate how you would test whether the covers had the same effect on sales.

(iv) Which of the two designs would you prefer in this situation and why?

[3x4=12]

INDIAN STATISTICAL INSTITUTE
B.STAT-III (2002-03)
STATISTICAL INFERENCE-II (Mid-Semester Examinations)
Time:3 hours Max marks:40

Date:21 February 2003.

Note: Answer as much as you can. The maximum you can score is 40.

1. Let $X_1, X_2, \cdots, X_n$ be iid $N(\theta, 1)$ 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
$$P_\theta[\ \underline{\theta}(\underset{\sim}{x}) < \theta < \overline{\theta}(\underset{\sim}{x})] \geq 1-\alpha \text{ for all } \theta$$
and $P_{\theta'}[\ \underline{\theta}(\underset{\sim}{x}) < \theta < \overline{\theta}(\underset{\sim}{x})\ \leq 1-\alpha$ for all θ, θ' with $\theta \neq \theta'$.
(a) Show that the family of intervals given by $\{(\overline{x} - \frac{z_{\alpha/2}}{\sqrt{n}}, \overline{x} + \frac{z_{\alpha/2}}{\sqrt{n}})\}$ is unbiased level α, where $\overline{x} = \frac{1}{n}\sum_1^n x_i$ and $1-p = \int_{-\infty}^{z_p} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$, for $0 < p < 1$.
(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
$$P_\theta[\ \underline{\theta}(\underset{\sim}{X}) < \theta' < \overline{\theta}(\underset{\sim}{X})] \geq P_\theta[\ \overline{X} - \frac{z_{\alpha/2}}{\sqrt{n}} < \theta' < \overline{X} + \frac{z_{\alpha/2}}{\sqrt{n}}] \text{ for all } \theta, \theta' \text{ with } \theta \neq \theta'.$$
(Such confidence intervals are called uniformly most accurate unbiased confidence intervals of level $(1-\alpha)$.)

[5+5]

2. Suppose that a new post-surgical treatment is being compared with a standard treatment (control) by observing the recovery times of 8 treatment subjects and 7 control subjects.
(a) State for what values of the Wilcoxon rank sum statistic W_S you would reject the null hypothesis of no treatment difference against the alternative of reduced recovery times for the treatment at the significance level α closest to 0.05 that is attainable by a non-randomised test.
(b) Suppose that the observed recovery times (in days) with sample sizes $m = 7$ and $n = 8$ are as follows:

Control: 20, 50, 48, 23, 21, 24, 49.

Treatment: 38, 19, 22. 34, 25, 18, 37, 25.

Find significance probability (P-value) of these results when the Wilcoxon test is used (Tables of null-distribution of Mann-Whitney Statistic is enclosed with the Question paper).
(c) Consider an alternative with F and G being the distributuion of the control and the treatment 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.

[5+5+10]

3. (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 $D_{2,3}$ by complete enumeration.

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

$$XYXYY$$

Compute the value of d of $D_{2,3}$ for this data. Find the significance probability of this data when Kolmogorov-Smirnov test is used.

[4+3+3]

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 $T_a^+ = \#(i,j) : i \leq j$ and $X_i + X_j > a$.

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+2+3+5]

****

k_1	a	$k_2 = 5$	$k_2 = 6$	$k_2 = 7$	$k_2 = 8$	$k_2 = 9$	$k_2 = 10$
5	0	.0040	.0022	.0013	.0008	.0005	.0003
	1	.0079	.0043	.0025	.0016	.0010	.0007
	2	.0159	.0087	.0051	.0031	.0020	.0013
	3	.0278	.0152	.0088	.0054	.0035	.0023
	4	.0476	.0260	.0152	.0093	.0060	.0040
	5	.0754	.0411	.0240	.0148	.0095	.0063
	6	.1111	.0628	.0366	.0225	.0145	.0097
	7	.1548	.0887	.0530	.0326	.0210	.0140
	8	.2103	.1234	.0745	.0466	.0300	.0200
	9	.2738	.1645	.1010	.0637	.0415	.0276
	10	.3452	.2143	.1338	.0855	.0559	.0376
	11	.4206	.2684	.1717	.1111	.0734	.0496
	12	.5000	.3312	.2159	.1422	.0949	.0646
	13	.5794	.3961	.2652	.1772	.1199	.0823
	14	.6548	.4654	.3194	.2176	.1489	.1032
	15	.7262	.5346	.3775	.2618	.1818	.1272
	16	.7897	.6039	.4381	.3108	.2188	.1548
	17	.8452	.6688	.5000	.3621	.2592	.1855
	18	.8889	.7316	.5619	.4165	.3032	.2198
	19	.9246	.7857	.6225	.4716	.3497	.2567
	20	.9524	.8355	.6806	.5284	.3986	.2970
	21	.9722	.8766	.7348	.5835	.4491	.3393
	22	.9841	.9113	.7841	.6379	.5000	.3839
	23	.9921	.9372	.8283	.6892	.5509	.4296
	24	.9960	.9589	.8662	.7382	.6014	.4765
	25	1.0000	.9740	.8990	.7824	.6503	.5235
6	0		.0011	.0006	.0003	.0002	.0001
	1		.0022	.0012	.0007	.0004	.0002
	2		.0043	.0023	.0013	.0008	.0005
	3		.0076	.0041	.0023	.0014	.0009
	4		.0130	.0070	.0040	.0024	.0015
	5		.0206	.0111	.0063	.0038	.0024
	6		.0325	.0175	.0100	.0060	.0037
	7		.0465	.0256	.0147	.0088	.0055
	8		.0660	.0367	.0213	.0128	.0080
	9		.0898	.0507	.0296	.0180	.0112
	10		.1201	.0688	.0406	.0248	.0156
	11		.1548	.0903	.0539	.0332	.0210
	12		.1970	.1171	.0709	.0440	.0280
	13		.2424	.1474	.0906	.0567	.0363
	14		.2944	.1830	.1142	.0723	.0467
	15		.3496	.2226	.1412	.0905	.0589
	16		.4091	.2669	.1725	.1119	.0736
	17		.4686	.3141	.2068	.1361	.0903
	18		.5314	.3654	.2454	.1638	.1099
	19		.5909	.4178	.2864	.1942	.1317
	20		.6504	.4726	.3310	.2280	.1566
	21		.7056	.5274	.3773	.2643	.1838
	22		.7576	.5822	.4259	.3035	.2139
	23		.8030	.6346	.4749	.3445	.2461
	24		.8452	.6859	.5251	.3878	.2811

k_1	a	$k_2 = 7$	$k_2 = 8$	$k_2 = 9$	$k_2 = 10$
7	0	.0003	.0002	.0001	.0001
	1	.0006	.0003	.0002	.0001
	2	.0012	.0006	.0003	.0002
	3	.0020	.0011	.0006	.0004
	4	.0035	.0019	.0010	.0006
	5	.0055	.0030	.0017	.0010
	6	.0087	.0047	.0026	.0015
	7	.0131	.0070	.0039	.0023
	8	.0189	.0103	.0058	.0034
	9	.0265	.0145	.0082	.0048
	10	.0364	.0200	.0115	.0068
	11	.0487	.0270	.0156	.0093
	12	.0641	.0361	.0209	.0125
	13	.0825	.0469	.0274	.0165
	14	.1043	.0603	.0356	.0215
	15	.1297	.0760	.0454	.0277
	16	.1588	.0946	.0571	.0351
	17	.1914	.1159	.0708	.0439
	18	.2279	.1405	.0869	.0544
	19	.2675	.1678	.1052	.0665
	20	.3100	.1984	.1261	.0806
	21	.3552	.2317	.1496	.0966
	22	.4024	.2679	.1755	.1148
	23	.4508	.3063	.2039	.1349
	24	.5000	.3472	.2349	.1574
	25	.5492	.3894	.2680	.1819
	26	.5976	.4333	.3032	.2087
	27	.6448	.4775	.3403	.2374
	28	.6900	.5225	.3788	.2681
	29	.7325	.5667	.4185	.3004
	30	.7721	.6106	.4591	.3345
	31	.8086	.6528	.5000	.3698
	32	.8412	.6937	.5409	.4063
	33	.8703	.7321	.5815	.4434
	34	.8957	.7683	.6212	.4811
	35	.9175	.8016	.6597	.5189

W_{XY} = Mann-Whitney Statistic (here k_1, k_2 are the sample sizes)

... is not sufficient for μ.

[8]

(b) Find the MVUE of θ^2 based on a random sample of size n from the Bernoulli(θ) distribution. Is it an estimator that attains Cramer-Rao lower bound?

[8+4]

P.T.O

INDIAN STATISTICAL INSTITUTE

Mid-Semester ~~Final~~ Examination: Semester II (2002-2003)

Course Name: B. Stat. 3rd year

Subject Name: Statistics Comprehensive

ebruary 18, 2003. Maximum Marks: 100. Duration: 3 hr. 30 min.

Note: Answer all questions.

1. For simple random sampling without replacement, derive (with full justification) a simplified expression for the bias in the estimate $\sum_{j=1}^{N} X_j \sum_{i=1}^{n}(y_i/x_i)/n$ for population total $\sum_{j=1}^{N} Y_j$. Describe how the above estimate can be modified to yield an unbiased estimate of population total. Describe (with justification) a sampling scheme that makes the usual ratio estimate of population total $\sum_{j=1}^{N} X_j(\sum_{i=1}^{n} y_i / \sum_{i=1}^{n} x_i)$ an unbiased estimate.

[7 + 7 + 7 = 21]

2. Suppose that we have one observation X and the null hypothesis states that X has a standard normal distribution while the alternative hypothesis states that X has double exponential distribution with density $(1/4)\exp(-|x|/2)$ for $-\infty < x < \infty$. Derive (with justification) the uniformly most powerful test for these hypotheses at level $\alpha = 0.05$ and calculate the power of this test.

[10]

3. Derive (with full justification) the uniformly most powerful test at a given level $0 < \alpha < 1$ for testing $H_0 : \theta = 1$ against $H_A : \theta \neq 1$ based on n i.i.d observations $X_1, X_2, \ldots, X_n$ having a common uniform distribution on $[0, \theta]$.

[9]

4. Let $X_1, X_2, \ldots, X_n$ be i.i.d observations with a common Poisson distribution having unknown mean $\theta > 0$. Let T_n be the maximum likelihood estimate of $\exp(\theta)$ based on these n observations. Derive (with appropriate justification - you may state and use any standard result without giving the proof of it) the asymptotic distribution of $n^{1/2}\{T_n - \exp(\theta)\}$ as $n \longrightarrow \infty$. Is T_n a consistent estimate of $\exp(\theta)$ (justify your answer) ?

[5 + 5 = 10]

5. (a) Let X be a single observation from $N(\mu, 1)$. Show that $|X|$ is not sufficient for μ. [8]

   (b) Find the MVUE of θ^2 based on a random sample of size n from the Bernoulli(θ) distribution. Is it an estimator that attains Cramer-Rao lower bound? [8+4]

P.T.O

6. (a) Express the following as the probability of an event and evaluate wit[illegible] integrating by parts:

$$\int_{-\infty}^{\infty}\Phi(x)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}dx,$$

where $\Phi(x)$ is the standard normal cumulative distribution function. [6]

(b) Let X_1 and X_2 be two independent lognormal variables. Show that $\frac{X_1}{X_2}$ and $\sqrt{X_1X_2}$ are both lognormal variables. [4]

(c) If $A \sim W_p(n,\Sigma)$, then show that l^TAl is $\sigma^2\chi^2$ for some σ^2. Use this result to prove $E(A) = n\Sigma$. [4]

7. (a) Let X and Y be two random variables such that

(i) marginally $X \sim N(0,1)$,

(ii) the regression of Y on X is linear,

(iii) $E(Y) = 0$, $V(Y) = 1$, $corr(X,Y) = \rho$.

Find $corr(\Phi(X),Y)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. [10]

(b) If r be the product moment correlation coefficient between x and y and e^2 be the correlation ratio of y on x, then interpret the following cases:

(i) $e^2 = r^2 = 0$.

(ii) $0 \le r^2 < e^2 = 1$.

(iii) $0 = r^2 < e^2 = 1$.

[6]

INDIAN STATISTICAL INSTITUTE

B. Stat (Third year)
Ordinary Differential Equation
Date: 13 December, 2002.
Maximum Marks 100

First Semester Examination (2002-2003)

Maximum Time 3 hrs.

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

All *matrices* used in this paper are square $n \times n$ matrices for arbitrary n unless mentioned otherwise. For a matrix A we also use e^A for $\exp A$. $I_{n\times n}$ 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.

1. (a) Let

$$A = \begin{pmatrix} -1 & 2 & 3 \\ 0 & -2 & 1 \\ 0 & 3 & 0 \end{pmatrix}.$$

Find e^A without using the series expansion. (Hint: Use the differential equation $x' = Ax$).

(b) Let

$$A_\theta = \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}.$$

Show that $e^{A_\theta} = I_{2\times 2}$ if $\theta = 2n\pi$.

(c) Let $\Phi(t)$ be a fundamental matrix solution of the vector equation $x'(t) = Ax(t)$ where A is skew symmetric. Show that $\Phi^T(t)\Phi(t) = C$ where C is a constant matrix. 8 + 4 + 5

2. Convert the following initial value problem to matrix-initial-value problem and find its solution.

$$x''(t) + 2x'(t) - 8x(t) = e^t, x(0) = 1, x'(0) = -4.$$

(Hint: first find the fundamental matrix solution of the homogenous part). 14

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

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

P.T.O.

(b) Let $\Phi(t)$ be a fundamental matrix solution of the equation. Then show that $\Phi(t) = P(t)\Psi(t)$ where for every t, $P(t)$ is invertible matrix, P is periodic of period w and $\Psi(t)$ is a fundamental matrix solution of the equation $z'(t) = Bz(t)$ for some constant matrix B in $M_n(\mathbb{C})$.

(c) Cosider the equation $x''(t) = \alpha(t)x(t)$ where $\alpha(t)$ is a continuous periodic function on $\mathbb{R}$ of period π. Costruct the corresponding first order vector differential equation and show that if its characteristic exponents are r_1, r_2 then $r_1 + r_2 = 0 \mod 2i$. 4 + 5 + 8

4. Let $\mathfrak{sl}_n(\mathbb{C}) \subset M_n(\mathbb{C})$ and $SL_n(\mathbb{C}) \subset GL_n(\mathbb{C})$ respectively be the set of matrices with trace-zero and with determinant 1. Assuming that $\exp : M_n(\mathbb{C}) \longrightarrow GL_n(\mathbb{C})$ is onto, show that $\exp(\mathfrak{sl}_n(\mathbb{C})) = SL_n(\mathbb{C})$. (Hint: Find the determinant of a particular fundamental solution Φ at t of the equation $x'(t) = Ax(t)$ where $A \in \mathfrak{sl}_n(\mathbb{R})$) 14

5. For $\alpha > -1$ the *Laguerre polynomial* of type α and order $k \in \mathbb{N}$, L_k^α is defined by the formula:

$$e^{-x}x^\alpha L_k^\alpha(x) = \frac{1}{k!}\frac{d^k}{dx^k}(e^{-x}x^{k+\alpha}).$$

(a) Check that L_k^α is a polynomial of degree k and $\frac{d}{dx}L_k^\alpha(x) = -L_{k-1}^{\alpha+1}(x)$.

(b) With the help of results in (a) show that the Laguerre polynomials satisfy the following orthogonality property:

$$\int_0^\infty L_k^\alpha(x)L_j^\alpha(x)e^{-x}x^\alpha dx = \frac{\Gamma(k+\alpha+1)}{\Gamma(k+1)}\delta_{jk}.$$

5 + 10

6. (a) For $k \in \mathbb{N}$, let $H_k(x) = (-1)^k e^{x^2}\frac{d^k}{dx^k}(e^{-x^2})$ be the Hermite polynomial. Show that H_k satisfy the following relations:

$$H_k'(x) = 2kH_{k-1}(x), H_k(x) = 2xH_{k-1}(x) - H_{k-1}'(x).$$

(b) Let $h_k(x) = H_k(x)e^{\frac{1}{2}x^2}$ be the Hermite function and $H = -\frac{d^2}{dx^2} + x^2$ be the Hermite operator. Show that H is *self adjoint* and h_k is an eigen function of H with eigen value $2k+1$ (Hint: Use relations in (b) for the last part.) 10 + 9

7. Consider the initial value problem $x' = 1 - 2tx, x(0) = 0$ in the domain

$$R = \{(t,x) : |t| \le \frac{1}{2}, |x| \le 1\}.$$

(a) Show that all the approximations to the solution exist on $t \le \frac{1}{2}$ and converge to the solution of the initial value problem.

(b) Find the successive approximations x_1, x_2, x_3 and prove that $|x(t) - x_3(t)| < .01$.

8 + 8

INDIAN STATISTICAL INSTITUTE
Semesteral Examination – B-III 2002
GEOLOGY
Full marks 100, Duration 3 hours
Date: 9.12.02

1. Name and illustrate the four different types of unconformity. Write the Law of Superposition. 3x4+3=15

OR

2. Classify fold structures on the basis of the orientation of fold axis and axial planes and provide their respective stereograms. 10+5=15

3. What is a geologic time keeper? Name two such "time keepers" or "clocks" and briefly explain how one may have an idea about absolute time with their help. 3+3x2=9

4. What is a Formation? What is a concurrent range zone? What is a chron? 2x3=6

5. Derive the radioactive decay equation.

Samples of two different rock bodies A and B have yielded the following isotopic compositions:

ROCK A	^{87}Rb	^{87}Sr
Sample 1	2	7
Sample 2	4	11

ROCK B	^{87}Rb	^{87}Sr
Sample 1	2	12
Sample 2	3	17

Which of the rocks between A and B is older? 5+5=10

OR

6. In a coalfield, a coal seam is struck at a depth of 400m at the location 'A'. At a location 'B', 300m north of A, the same coal seam is struck at a depth of 400m. Find the strike and dip of the coal seam. At what depth the coal seam will be struck at a location 'C' 300m north-east of A? 5+5=10

7. What are the manifestations of the dynamic nature of the earth? What is the energy resource of the Earth's dynamism? How did the earth acquire this energy? What would be the state of the earth when it will lose all of its energy? 3+3+3+3=12

8. Define lithosphere and asthenosphere. Using the principle of isostasy show that mountains have deep roots and oceanic lithosphere should be thinner than the continental lithosphere. 3+3+3+4=13

9. With a neat diagram, illustrate the internal structure and composition of the Earth and their variability between continental and oceanic areas. 10

OR

10. Describe with neat sketch the different geomorphic parts of a meandering river. 10

P.T.O

11. Describe with neat sketches the method of representing orientation of lines and planes in stereographic projection. How would you determine the dip and the strike of a plane if you were provided with two sets of plunge amount and direction of the apparent dips with help of a stereogram. 10+5=15

12. Write short notes on (any two)– 2x5=10
a) Alluvial fan
b) Lake
c) Fault
d) Cross beds
e) Major postulates of the Darwin's theory of Organic Evolution.

Indian Statistical Institute

First Semestral Examination : (2002-2003)
B.Stat.(Hons.) III Year
Economics III

The question paper carries 110 marks. The maximum you can score is 100.

Answer any *four* questions (Select from 1-5)

Date : 9.12.02 Time :3 h

1. (a) Consider the model

$$y_i = \beta x_i + \varepsilon_i, \quad i = 1,2,\dots,n.$$
$$\varepsilon_i = u_i + \delta u_{i-1}, \quad |\delta| < 1$$
$$E(u_i) = 0$$
$$E(u_i u_j) = \begin{cases} \sigma_u^2, i = j \\ 0, i \neq j. \end{cases}$$

(i) Show that the LS estimator of β is unbiased.

(ii) Assuming δ and σ_u^2 are known, obtain the GLS estimator of β.

(b) Consider the model $y_i = \beta x_i + \varepsilon_i$, for which all the regular assumptions hold except that $\text{var}(\varepsilon_i) = \sigma^2 x_i (x_i > 0)$. Derive the BLUE of β and its variance, given a sample of n observations.

(c) Consider the model $y_i = \alpha + \beta x_i + \varepsilon_i, i = 1,2,\dots,n,$ where x is stochastic, x and ε are independent. Examine the consequences (in terms of unbiasedness and efficiency of $\hat{\alpha}$ and $\hat{\beta}$) of applying OLS to estimate the equation.

[12+5+8=25]

2. (a) What are distributed lag models? Explain.
(b) Describe the geometric lag model and rationalize the model in terms of
(i) Adaptive Expectation Model ;
(ii) Partial Adjustment Model.
(c) Describe how one can estimate such a model.

[4+(8+8)+5=25]

3. (a) What are 'errors in variables' models? Do such models violate the assumptions of the classical linear regression model? Illustrate.

P.T.O

(b) Consider the model with two explanatory variables:

$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$.

Suppose X_1, X_2 and Y are measured with errors and we measure

$x_1 = X_1 + u_1, x_2 = X_2 + u_2$ and $y = Y + v$.

Assume that u_1, u_2 and v are mutually uncorrelated and also uncorrelated with X_1, X_2 and Y. Are the OLS estimates of the parameters consistent? Justify your answer.

(c) How does one overcome problems , if any, in this model ?

(d) Describe the use of discrete grouping variable in this context.

[6+10+4+5=25]

4. (a) What do you mean by 'dummy variables'?

(b) Suppose

$$y = \begin{cases} \alpha_1 + \beta_1 x + \lambda_1 z + \varepsilon & for \quad period \ 1 \\ \alpha_2 + \beta_2 x + \lambda_2 z + \varepsilon & for \quad period \ 2 \end{cases}$$

Using dummy variables, how would you test the hypothesis that only the coefficient of x changes between the two periods?

(c) In a situation where the dependent variable of a linear regression model is binary in nature, what are the problems in using a linear probability model? Describe an appropriate model to handle the situation.

[3+7+15=25]

5. (a) In the context of simultaneous equations system what is 'identification problem'? When is a system identified?

(b) Discuss the identifiability status of each of the following equations systems :

(i) $y_{1t} = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 y_{2t} + \varepsilon_{1t}$

$y_{2t} = \alpha_1 y_{1t} + \alpha_2 x_{4t} + \alpha_3 x_{5t} + \varepsilon_{2t}$

(ii) $\beta_{11} y_{1t} + \beta_{12} y_{2t} + \gamma_{11} x_{1t} + \gamma_{12} x_{2t} = \mu_{11}$

$\beta_{21} y_{1t} + \beta_{22} y_{2t} + \gamma_{21} x_{1t} + \gamma_{22} x_{2t} = \mu_{21}$

with the restrictions $\gamma_{11} = 0, \gamma_{12} = 0, \gamma_{22} = 0, \beta_{21} + \gamma_{21} = 0$.

(c) Describe briefly (mention only the steps involved) a single equation and a system method of estimating a simultaneous equations system, stating the appropriateness of each of these methods in terms of the identifiably status. [5+12+8=25]

6. Practical exercise. [10]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2002-2003

B.Stat. (Hons.). 3rd Year

Linear Statistical Models

Date: December 5, 2002 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) Give an example of a coordinate-free standard linear model. Discuss the distributional assumptions for such models.

(b) Consider an experimental setup consisting of four experiments x_1, x_2, x_3, x_4 respectively. Suppose that the regression function ϕ satisfy the constraints (i) $\phi(x_1) - \phi(x_2) + \phi(x_3) - \phi(x_4) = 0$ and, (ii) $\phi(x_1) + \phi(x_2) - \phi(x_3) - \phi(x_4) = 0$. Assume the standard linear model for the response. What is the minimum number of experiments necessary to (linear) unbiasedly estimate ϕ and the error variance σ^2. Suggest one such design and derive the required linear unbiased estimates. [6+14= 20]

2 (a) Let $Y = \mu + \epsilon$, $\mu \in \mathcal{V}$ be a linear model with normally distributed errors (standard notation). Show that the least squares estimate $\hat{\mu}$ and $||Y - \hat{\mu}||^2$ are jointly complete sufficient for (μ, σ^2).

(b) Drop the normality assumption in 2(a) and assume that $\mu = X\beta$ where $\mathcal{V} = col(X)$ (X is of full rank). State and prove the Gauss-Markov theorem in this context.

(c) Let $\tau : \mathbb{R}^p \to \mathbb{R}^n$ $(p < n)$ be given by $\tau(\beta) = X\beta$ for some $n \times p$ matrix X which may not be of full column-rank. Derive all the liner maps $\tau^- : \mathbb{R}^n \to \mathbb{R}^p$ satisfying the condition: $\tau \circ \tau^- \circ \tau = \tau$. [15+10+15=40]

3 (a) Show that corresponding to any set of linear hypotheses of the form $H_0 : C\beta = 0$, where rows of C are not necessarily estimable there exists an equivalent hypothesis $H_0^e : C^e\beta = 0$ which is testable (including the possibility that H_0^e might be an empty hypothesis).

P. T. O

(b) Let $y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$, $1 \leq i \leq v$, $1 \leq j \leq b$, denote a balanced two way lay-out. Express the layout in the form $Y = X_0\mu + X_1\underline{\alpha} + X_2\underline{\beta} + \epsilon$ with suitable rearrangement of the observations in a vector form. Here $\underline{\alpha}^t = (\alpha_1, \ldots, \alpha_v)$ and $\underline{\beta}^t = (\beta_1, \ldots, \beta_b)$ respectively. Further, let P_i denote the projection onto the column space of X_i, $0 \leq i \leq 2$. Show that $(P_1 - P_0)(P_2 - P_0) = 0$.
(c) Hence write down the ANOVA table for balanced two-way layout explicitly describing the formulae of various sum of squares. [10+7+8=25]

4. In an experiment to investigate the effect of color of paper (blue,green,orange) on response rates for questionnaires distributed in supermarket parking lots 15 lots, were chosen in a city and each color was randomly assigned to five of these lots. The response rates (in percentages) follow. Assume one-way ANOVA model is appropriate.

Blue : 28, 26, 31, 27, 35

Green : 34, 29, 25, 31, 29

Orange : 31, 25, 27, 29, 28

(a) Obtain fitted values.
(b) Conduct a test to determine whether or not the mean response rates for the three colors differ. Use $\alpha = 0.05$ as level of significance.
(c) Obtain Bonferroni type joint confidence interval for the three pairwise contrasts with confidence coefficient at least 95%.
(d) Obtain a 95% prediction interval for the response if a blue paper is used for the questionnaire in an independently chosen parking lot. [5+7+7+6=25]

INDIAN STATISTICAL INSTITUTE

B III Semester I

Statistical Inference I

Semestral Examination

Total Points 60

Date : 2 Dec 2002 Time : 3 Hours

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

1. (a) State Generalized Neyman–Pearson Lemma. [3]

Let $p(x|\theta), \theta \in \Theta$ be a real parameter family. For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ a test ϕ^0 is said to be locally most powerful (LMP) size-α if for some $\epsilon (> 0)$

i. $\phi^0 \in \Phi(\alpha) = \{\phi : E_{\theta_0}\phi(X) = \alpha\}$

ii. $E_\theta\phi^0(X) \geq E_\theta\phi(X), \forall\phi \in \Phi(\alpha)$ and for every $\theta : \theta_0 < \theta < \theta_0 + \epsilon$.

Assume continuous distribution.

(b) Let $p(x|\theta)$ be such that, for every test ϕ, $\frac{d}{d\theta}E_\theta\phi(X)$ exists in the neighbourhood of θ_0 and is continuous. Let $\phi^0 \in \Phi(\alpha)$ and $\frac{d}{d\theta}E_\theta\phi^0(X)|_{\theta_0}$ is maximum among all $\phi \in \Phi(\alpha)$. Then prove that for every $\phi \in \Phi(\alpha)$ we would have an $\epsilon(> 0)$ such that $E_\theta\phi^0(X) \geq E_\theta\phi(X), \forall\theta : \theta_0 < \theta < \theta_0 + \epsilon$. [4]

Hence to construct LMP size-α test, we need to maximize $\frac{d}{d\theta}E_\theta\phi(X)|_{\theta_0}$ subject to $E_{\theta_0}\phi(x) = \alpha$.

(c) Show that the test given by

$$\begin{aligned} \phi^0(x) &= 1 \text{ if } \tfrac{d}{d\theta}p(x|\theta)|_{\theta_0} > \lambda p(x|\theta_0) \\ &= 0 \text{ if } \tfrac{d}{d\theta}p(x|\theta)|_{\theta_0} < \lambda p(x|\theta_0) \end{aligned}$$

where λ is such that $\int \phi^0(x)p(x|\theta_0)dx = \alpha$ is LMP size-α for $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. [7]

(d) Let $X_1, \ldots X_n$ be i.i.d

$$p(x|\theta) = \frac{1}{\pi}\frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty$$

Show that LMP size-α test for $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ is given by

$$\begin{aligned} \phi^0(x) &= 1 \text{ if } \textstyle\sum_{i=1}^n \left\{\frac{x_i}{1+x_i^2}\right\} > k \\ &= 0 \text{ if } \textstyle\sum_{i=1}^n \left\{\frac{x_i}{1+x_i^2}\right\} < k \end{aligned}$$

where k is determined from the size condition. [6]

P. T. O

2. (a) Define minimal sufficiency. [3]

(b) Let $\boldsymbol{X} = (X_1, \dots X_p)' \sim N_p(\boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} = (p\theta_1, 0, \dots, 0)'$ and

$$\Sigma = \begin{bmatrix} (p-1)+\theta_2^2 & -\epsilon' \\ -\epsilon & I_{p-1} \end{bmatrix}$$

$\epsilon'_{1\times(p-1)} = (1, \dots 1)$. Find minimal sufficient statistics for $\boldsymbol{\theta} = (\theta_1, \theta_2)$ [12]

3. (a) Define consistency of a sequence of estimators. [2]

(b) If $\{U_n\}$ be a sequence of estimators of a parameter θ satisfying $lim_{n\to\infty} Var_\theta U_n = 0$ and $lim_{n\to\infty} Bias_\theta U_n = 0$. Further let $a_1, a_2, \dots$ and $b_1, b_2, \dots$ be sequences of constants satisfying $lim_{n\to\infty} a_n = 1$ and $lim_{n\to\infty} b_n = 0$. Show that the sequence $W_n = a_n U_n + b_n$ is a consistent sequence of estimators of θ. [7]

4. Show that an MP size-α test is admissible if there does not exist a test with power 1. [8]

5. Let $\boldsymbol{X} = (X_1, \dots X_n)$ be a random sample of size n from a negative binomial distribution. Find MVUE of $Pr(X_i = r) = \binom{k+r-1}{r} \theta^r (1-\theta)^k$. [8]

INDIAN STATISTICAL INSTITUTE
First Semestral Examination : 2002-2003
B.Stat.(Hons.) III Year
Sample Surveys

Date : 28.11.2002 **Maximum Marks : 100** **Duration : 3 Hours**

Answer Question No. 6 and ANY THREE questions from the rest . Marks allotted to each question are given within the parentheses . Standard notations and symbols are used

1.A company intends to interview a simple random sample of employees who have been with it for more than five years . The company has 1000 units of money to spend , and each interview costs 10 units . There is no separate list of employees with more than 5 years service , but a list can be compiled from the files at a cost of 200 units . The company can either (a) compile the list and interview a simple random sample drawn from the eligible employees or (b) draw a simple random sample of all employees , interviewing only those eligible . The cost of rejecting those not eligible in the sample is assumed to be negligible .

Show that for estimating a total over the population of eligible employees , plan (a) gives a smaller variance than plan (b) only if $C_j < 2\sqrt{Q_j}$, where C_j is the coefficient of variation of the item among eligible employees and Q_j is the proportion of non-eligibles in the company . You may ignore the fpc .

[25]

2.(a) If the sample size required to estimate the proportion of workers in a population with an RSE of $\alpha\%$ is n in SRSWR , determine the sample size required to estimate the proportion of non-workers with the same precision .

(b) What do you mean by Murthy's unordering principle ? Show that the ordered estimator of the population total due to Des Raj based on a PPSWOR sample of n units can be improved upon by using Murthy's unordering principle .

Derive an expression for the variance of Murthy's unordered estimator based on a PPSWOR sample of size 2 and also obtain an unbiased estimator of the variance .

(5+2+5+10+3) = [25]

3. (a) Show that if fpc be ignored ,

$$V_{ran} \geq V_{prop} \geq V_{opt}$$

where V_{ran} , V_{prop} and V_{opt} denote respectively the variances of the estimated mean based on unstratified simple random sampling , stratified random sampling with proportional allocation and stratified random sampling with optimum allocation for a given total sample size.

(b)Discuss how one can estimate the gain in efficiency due to stratification on the basis of a stratified simple random sample .

(12+13) = [25]

4. If π_i and π_{ij} denote respectively the first order and second order inclusion probabilities for a given sampling design p(s) , show that

P.T.O.

(a) $\sum_{i=1}^{N} \pi_i = E[\nu(s)]$, where ν(s) denotes the effective sample size ;

(b) $\sum_{i \neq j} \sum_{=1}^{N} \pi_{ij} = \text{Var}[\nu(s)] + \nu(\nu-1)$ where $\nu = E[\nu(s)]$ is the expected effective sample size .

(c) If $\nu = [\nu] + \theta$ where $[\nu]$ denotes the greatest integer contained in ν and $0<\theta<1$, show that

$$\nu(\nu-1) + \theta(1-\theta) \leq \sum_{i \neq j=1}^{N} \sum \pi_{ij} \leq N(\nu-1) . \qquad (3+7+15) = [25]$$

5. (a) Describe a sampling scheme under which the conventional ratio estimator of the population mean becomes unbiased . Describe how, under certain conditions to be stated by you , the above sampling scheme can be modified so as to give a πps sampling scheme .

(b) Show that the Horvitz-Thompson estimator based on the above modified πps sampling scheme is more efficient than the Hansen-Hurwitz estimator based on PPSWR sampling scheme based on same size measures and involving the same number of draws. (5+5+15) =[25]

6. To estimate the total number of words (Y) in an English dictionary , 10 out of 26 letters were selected with PPSWR , size being the number of pages devoted to a letter and for each selected letter , two pages were selected with SRSWOR . The relevant sample data are given in the following table .

Sl. No.	Sample letters	No. of pages devoted	No. of words in sample page	
			1	2
1.	S	131	34	27
2.	C	97	27	26
3.	N	21	44	38
4.	S	131	24	29
5.	F	43	25	32
6.	J	7	42	48
7.	U	18	24	21
8.	P	85	53	24
9.	A	49	47	55
10.	D	54	38	57

(Total number of pages in the dictionary is 980)

(a) Estimate unbiasedly Y and obtain an estimate of its RSE .

(b) Estimate also the efficiency of the above method of sampling compared to that of drawing 20 pages from the dictionary with SRSWR . (10 + 15) = [25]

INDIAN STATISTICAL INSTITUTE

B.Stat. (III) : 2002-2003

Economics III

Mid-semestral Examination

Date : 23.9.02 | Maximum Marks : 100 | Time : 3 hours

Answser any *four* questions

1. (a) State the asssumptions underlying the classical multiple linear regression model.

   (b) Find the Least squares estimator $\hat{\beta}$ from the regression model $y_{n\times 1} = X_{n\times(k+1)}\,\beta_{(k+1)\times 1} + \epsilon_{n\times 1}$. Show that $\hat{\beta}$ is best linear unbiased.

   (c) Derive the estimator of the variance of ϵ_i.

   (d) Suppose y_i = log (production per workers), x_i = log (wage rate)

   subscript i refers to the ith firm in an industry A. A model of the form $y_i = \alpha + \beta x_i + \epsilon_i$ $(i = 1, 2, ..., n)$ is used to estimate β (the elasticity of substitution between labour and capital). The least squares results are as follows :

$$y_i = -0.4 + 1.0x_i + e_i \qquad (n = 52)$$

   ↓ (standard error = 0.1)

   Show that $R_A^2 = 2/3$.

   [5+6+8+6=25]

2. (a) A regression equation explaining household expenditure on recreation as a function of income was estimated as

$$y_t = \underset{(10)}{-25} + \underset{(16)}{35D_{t2}} + \underset{(15)}{40D_{t3}} - \underset{(8)}{15D_{t4}} + \underset{(0.02)}{0.05x_t} + e_t$$

   where figures in parentheses are the standard errors, y : expenditure on recreation, x : income and the D's are quarterly dummies with the 1st quarter (January - March) left out. Determine the values of the estimated coefficients given that

   (i) the equation contains four seasonal dummies and no constant term.

   (ii) the equation contains four seasonal dummies and a constant term, but the coefficients of the seasonal dummies represent deviations from the annual average and their sum is equal to zero.

P. T. O

(b) Suppose consumption expenditure y depends on income x, but different levels of income produce different values of marginal propensity to consumer (β_i). In particular,

$$\begin{aligned}\frac{d\,E(y \mid x)}{dx} &= \beta_1 \text{ if } x < \text{Rs.}1000/- \\ &= \beta_2 \text{ if Rs.}1000/- \leq x < \text{Rs.}5000/- \\ &= \beta_3 \text{ if } x \geq \text{Rs.}5000/-.\end{aligned}$$

Using dummy variables how can your represent the regression equation? (Assume that there is no intercept term in the equation). Estimate the marginal propepsity to consume for the three groups.

(c) What is 'dummy variable trap'?

(d) Illustrate how dummy variables can be used to allow for differences in both 'slope' and 'intercept' terms for a bivariate regression model in two situations.

[8+10+3+4=25]

3. (a) Explain what is meant by 'autocorrelation'.

(b) Describe a test for first order positive autocorrelation in a given time series.

(c) Show that for a first order autoregressive model with positive coefficient the autocorrelation function (ACF) declines geometrically.

(d) Consider the two models :

$$\hat{y}_t = 0.45 - \underset{(-3.96)}{0.0041}\, x_t, \qquad R^2 = 0.5284, \quad D.W. = 0.8252$$

and

$$\hat{y}_t = 0.48 + \underset{(3.27)}{0.0127}\, y_{t-1} - \underset{(-2.17)}{0.0321}\, x_t, \qquad R^2 = 0.8829, \quad D.W. = 1.82$$

where figures in parentheses are the t-ratios.

Comment on the regression results. What are the appropriate values of the serial correlation in the two cases?

(e) Describe the Breusch-Godfrey test for testing fourth order serial correlation.

[2+5+5+8+5=25]

4. (a) In a linear regression model $y = X\beta + \epsilon$, consider estimation of $\beta_{k\times 1}$ subject to $r(< k+1)$ linearly independent restrictions $R\beta = d$. Show that the constrained LS estimated of β (call it β^*) is given by

$$\hat{\beta}^* = \hat{\beta} - (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} [R\hat{\beta} - d],$$

where $\hat{\beta}$ is the OLS estimator.

(b) In the regression model $y = 4 + 2.5x_2 - 1.5x_3$ (no. of observations = 5), the explained sum of squares = 26.5, total sum of squares = 28 and $R^2 = 0.9464$.

(i) Test the joint significance of x_2 and x_3. What are the appropriate degrees of freedom? [The critical value of the relevant statistic at 5% level of significance = 19.00].

(ii) Given that $(X'X)^{-1} = \begin{bmatrix} * & * & * \\ * & 1 & -1.5 \\ * & -1.5 & 2.5 \end{bmatrix}$,

where * denotes figures not available, and X is the matrix with columns 1, x_2 and x_3, test the hypothesis that the coefficients of x_2 and x_3 are equal in magnitude, but opposite is sign. Determine the appropriate degrees of freedom. [The critical value of the relevant test statistic at 5% level of significance is 18.5]

[Hint : $RRSS - URSS = (\hat{\beta} - \hat{\beta}^*)' (X'X) (\hat{\beta} - \hat{\beta}^*)$].

(c) Suppose you want to test r independent linear restrictions of the form $R\beta = d_{r\times 1}$, where $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$ in $y_{n\times 1} = X\beta + \epsilon_{n\times 1}$.

Write down R and d to incorporate the following cases : ($k = 4$)

(i) $\beta_1 = \beta_2 = \beta_3 = 2$

(ii) $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$

(iii) $\beta_1 - 3\beta_2 = 5\beta_3$

(iv) $\beta_1 + 3\beta_2 = 0$

[7+14+4=25]

5. (a) What is multicollinearity? What are the consequences of multicollinearity?

(b) Describe a procedure for detection of multicollinearity.

(c) Describe the method of 'principal components' regression.

[5+10+10=25]

INDIAN STATISTICAL INSTITUTE
First Semestral Examination: (2002-2003)
B.Stat. III

ELECTIVE GEOLOGY

Date : 23.09.02 Maximum Marks : 100 Duration : 3 hours

1. What is a supernova explosion? How is it related to formation of planetary systems? *2 + 4 = 6*

2. Discuss briefly the internal structure of the earth and the possible mode of origin of the internal layering. *4 + 4 = 8*

3. What does the term "plate" mean in geology? Show different types of plate boundaries with the help of suitable diagrams. *2 + 6 = 8*

4. Distinguish between continental and oceanic crust. Why we do not get oceanic crust older than 200 million years? *3 + 2 = 5*

5. How are magmas generated within the earth? Distinguish between a rock and a mineral. Enumerate the different steps you would adopt to identify igneous, sedimentary and metamorphic rocks? *3 + 2 + 5 = 10*

6. Describe with illustrations the different modes of clastic transport. *10*

7. Describe sub-aqueous sedimentary bedforms and their records with illustrations. *10*

8. Write short notes on (any three): *3 + 3 + 3 = 9*

a) Sorting
b) Sphericity and roundness
c) Grain size classification
d) Transporting agents
e) Mechanical and Chemical Weathering

9. The majority of the rock-forming minerals are - borates/ molybdates/ silicates - state which one of the three is correct. *1*

10. Name one rock-forming mineral, each from the Nesosilicates, Sorosilicates and Phyllosilicates and write their gross chemical composition and structure. *10*

11. Name three major minerals that you expect in Basalt. *6*

12. Name the major minerals that appear in the discontinuous reaction series of Bowen with gradual fall of temperature from ~1000 °C to ~ 600 °C. *10*

13. What is the most common element in earth's crust by weight %? *1*

14. What is the most common element in earth's crust by weight %? *1*

15. Why the most common element of crust and that of the whole earth are different? *5*

INDIAN STATISTICAL INSTITUTE

B. Stat (Third year) | Year 2002-2003
Ordinary Differential Equation | Mid-semester Examination
Instructor: Rudra Sarkar | Date: September 19, 2002.
Maximum Marks 90 | Maximum Time 3 hrs.

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

1. (a) Prove that a general solution of a differential equation of order 2 has exactly two independent constants.
(b) Suppose a solution $f = \phi(x, c_1, c_2)$ of a differential equation involves two constants c_1 and c_2. Show that constants c_1 and c_2 are independent if and only if the Wronskian of $\frac{\partial\phi}{\partial c_1}, \frac{\partial\phi}{\partial c_2}$ is not equal to zero. 5+10

2. Let $L(y) = y'' + a_1y' + a_2y$ be a constant coefficient differential operator of order 2 and degree 1. Let $L(\phi) = 0$ on an interval I contaning a point x_0 for some complex valued C^2-function ϕ. Define $||\phi(x)||$ by

$$||\phi(x)||^2 = |\phi(x)|^2 + |\phi'(x)|^2.$$

(a) Show that for all $\in I$,

$$||\phi(x_0)||e^{-k|x-x_0|} \leq ||\phi(x)|| \leq ||\phi(x_0)||e^{k|x-x_0|}$$

for some constant k. Find a suitable k related to L.
(b) From (a) conclude that given an initial condition, there can be at most one solution of $L(y) = 0$ satisfying the initial condition. 7+3

3. 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

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

$$L(y) = y^{(n)} + a_1y^{(n-1)} + \cdots + a_ny$$

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

P. T. O

5. Let $\{\phi_1, \phi_2\}$ and $\{\psi_1, \psi_2\}$ be two sets of linearly independent solutions of

$$y'' + a_1(x)y' + a_2(x)y = 0$$

on an interval I, where a_1, a_2 are continuous on I.

(a) Show that ϕ_1 and ϕ_2 can not have a common zero on I.

(b) Find relation between the Wronskians $W(\phi_1, \phi_2)$ and $W(\psi_1, \psi_2)$.

(c) Show that a_1 and a_2 can be uniquely determined from solutions ϕ_1 and ϕ_2. 3+10+10

6. Consider the equation $y''+\alpha(x)y = 0$ where α is a real valued continuous function on an open interval I. Show that the zeros of a nontrivial solution ϕ are simple (i.e. ϕ' is nonzero at that point) and isolated (i.e. zeros do not have a limit point). 3+9

7. Consider the equation $y'' + a_1(x)y' + a_2(x)y = 0$, where a_1, a_2 are continuous functions on $\mathbb{R}$ of period $\eta > 0$.

Let ϕ_1 and ϕ_2 are two solutions satisfying

$$\phi_1(0) = 1 \quad \phi_2(0) = 0$$
$$\phi_1'(0) = 0 \quad \phi_2'(0) = 1.$$

Show that $\phi_1(x + \eta) = \phi_1(\eta)\phi_1(x) + \phi_1'(\eta)\phi_2(x)$. 10

8. Show that -1 and 1 are regular singular points of the equation

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

Find the indicial polynomial and its roots corresponding to the point $x = 1$. 4+6

9. Consider the *Laguerre equation*:

$$xy'' + (1 - x)y' + \alpha y = 0$$

(a) Show that it has a regular singular point at $x = 0$.

(b) Compute the indicial polynomial.

(c) Show that for any nonegative integer n, $L_n(x) = e^x \frac{d^n}{dx^n}(x^n e^{-x})$ satisfies the Laguerre equation for $\alpha = n$. 2+3+15

INDIAN STATISTICAL INSTITUTE

B III Semester I

Statistical Inference I

Midterm Examination

Total points 30

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

ate: September 16, 2002

Time: 3 Hours

1. Let X be one observation from the p.d.f.

$$f(x|\theta) = \left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|}, \quad x = -1, 0, 1; \quad 0 \le \theta \le 1$$

(a) Is X a complete sufficient statistic? [3]

(b) Is $|X|$ a complete sufficient statistic? [3]

2. Suppose p.d.f. of (X_1, X_2) is given by

$$f(x_1, x_2) = \frac{x_1^{\theta_1-1}x_2^{\theta_2-1}exp\left(-\frac{x_1+x_2}{\lambda}\right)}{\Gamma(\theta_1)\Gamma(\theta_2)\lambda^{\theta_1+\theta_2}}, \quad x_1, x_2 > 0, \ \theta_1, \theta_2 > 0$$

Apply Basu's Theorem to show that the two statistics $X_1 + X_2$ and $\frac{X_1}{X_1+X_2}$ are independent. [8]

3. Let X be a random variable whose pmf under H_0 and H_1 is given by

x	1	2	3	4	5	6	7
$f(x\|H_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$f(x\|H_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

Find the most powerful test for H_0 versus H_1 with size $\alpha = 0.04$. Compute the probability of type II error for this test. [5]

4. Let $X_1, \ldots, X_n$ be a random sample for a *Binomial*(m, p) population where p is known and m unknown. Indicate how to obtain MLE of m noting the fact that m must be an integer. [6]

5. Consider a single observation X from a Cauchy distribution with an unknown location parameter θ,

$$f(x|\theta) = \frac{1}{\pi[1 + (x - \theta)^2]}, \ \infty < x < \infty$$

Consider testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$. Does there exist a UMP test of these hypotheses at any specified level of significance α $(0 < \alpha < 1)$? Justify your answer. [5]