3

INDIAN STATISTICAL INSTITUTE
B. Stat. III Year (2004-2005), Differential Equations
Mid-Semestral Examination

Time: 3 hrs: Max. Marks 100: Date: 06-09-2004.

Note: Answer all the questions.

1. Consider the initial value problem:

$$y' = 3y + 1, \quad y(0) = 2.$$

   (a) Solve the above initial value problem.
   (b) In the Picard's method for solving the above equation, find the first four approximations to the solution.

[10]

2. Solve the following differential equations.

   (a) $(2x + 3y + 1)dx + (2y - 3x + 5)dy = 0$.
   (b) $[x\sin(x + y) - \cos(x + y)]dx + x\sin(x + y)dy = 0$.
   (c) $x^2y'' + xy' = 1$.

[15]

3. In each of the following two cases, find the homogeneous differential equation with constant coefficients, of the lowest order, satisfied by all the given functions.

   (a) $y_1(x) = x^2, \quad y_2(x) = e^x, \quad y_3(x) = xe^x$.
   (b) $y_1(x) = e^{-2x}\cos 3x, \quad y_2(x) = xe^{-2x}$.

[10]

4. $u(x)$ is a solution of the differential equation $y'' - 4y' + 29y = 0$ and $v(x)$ is a solution of the differential equation $y'' + 4y' + 13y = 0$. The curves $y = u(x)$ and $y = v(x)$ intersect at the origin and have equal slopes at the origin. Also $u'(\frac{\pi}{2}) = 1$. Determine u and v. [10]

5. Given that the differential equation $y'' + 4xy' + Q(x)y = 0$ has two solutions $y_1(x) = u(x)$ and $y_2(x) = xu(x)$ with $u(0) = 1$, find the general solution of the equation. [10]

P. T. O

6. Show that the change of variable $y = \frac{u'}{q(x)u}$ changes the non-linear first order equation

$$y' + p(x)y + q(x)y^2 = r(x)$$

to the second order linear equation

$$u'' + [p(x) - \frac{q'(x)}{q(x)}]u' - r(x)q(x)u = 0.$$

Use this method to solve:

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

for $x > 0$. [15]

7. Find the singular points of the following equations and find the nature (regular or irregular) of the singular points. If it is a regular singular point, find the indicial equation at the singular point.

   (a) $x^3y'' + (\cos 2x - 1)y' + 2xy = 0.$

   (b) $x^3y'' + (\sin x)y = 0.$

[10]

8. Show that the equation

$$4x^2y'' - 8xy' + (4x^2 + 1)y = 0$$

has only one Frobenius series solution for $x > 0$. Find it and then find the general solution (for $x > 0$). [20]

BIII : 2004-2005
Midterm Examination
Linear Models

Full Marks –50 Date:- 15. 9. 04 Time: 10.30-12.30

1. Consider the Gauss Markov model $(Y, X\beta, \sigma^2 I)$. Show that for every estimable function $c'\beta$ there exists a unique BLUE given by $a_0'Y$ such that $c = X'a_0$ and $a_0 \in \zeta$ (X), column space of X.

[6]

2. Consider the following linear model:

$$Y_1 = \theta_1 + \theta_2 + \theta_4 + \varepsilon_1$$
$$Y_2 = \theta_1 + \theta_3 + 2\theta_4 + \varepsilon_2$$
$$Y_3 = \theta_1 + \theta_2 + \theta_4 + \varepsilon_3$$
$$Y_4 = -\theta_2 + \theta_3 + \theta_4 + \varepsilon_4$$
$$Y_5 = 2\theta_1 + \theta_2 + 2\theta_3 + \varepsilon_5$$

where $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5)' \sim N_5(0, \sigma^2 I)$.

a) Obtain a necessary and sufficient condition for estimability of a linear parametric function.

b) Find the BLUE of $\theta_3 + \theta_4 - \theta_2$.

c) Identify the error functions and hence obtain an expression for Sums of Squares due to error.

d) Obtain an expression for an appropriate test statistic for

$$H_0 : \theta_3 + \theta_4 = \theta_2.$$

[4+6+7+7 = 24]

3. In the context of a chemical balance weighing design problem, explain the role of Hadamard matrices for most efficient estimation of individual weights of a number of objects, assuming that the balance may have inherent bias. Illustrate your answer with reference to an example involving 5 objects and 8 weighing operations.

[8 + 12 = 20]

INDIAN STATISTICAL INSTITUTE

STATISTICAL INFERENCE I

Mid Semester Exam, BStat 3rd Yr.

Maximum you can score is 40.

Date: September 13, 2004 Time: 3 hours

1. Suppose that the p.d.f. of (X_1, X_2) is given by

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

Show that the two statistics $U = X_1 + X_2$ and $V = \frac{X_1}{X_1 + X_2}$ are independent. [7 points]

2. 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 $U \in \mathcal{U}$). [7 points]

3. Consider the class of all symmetric distributions $\mathcal{F}$ with a finite first moment. Let us denote by θ the centre of symmetry of such distributions. Show that there does not exist an estimator, based on a random sample of size n (> 2), which is UMVU for estimating θ for all distributions F in $\mathcal{F}$. [7 points]

4. (a) Suppose that n independent and identically distributed observations $X_1, \ldots, X_n$ $(n \geq 2)$ are available from the distribution of X that is modeled by the usual two parameter $N(\mu, \sigma^2)$ model. Suppose that both μ and σ^2 are assumed to be nonnegative but otherwise unknown. Find the maximum likelihood estimates of μ and σ^2. Justify your answer.

P. T. O

(b) Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $X_1, \ldots, X_m$ are distributed at $N(\mu_1, \sigma^2)$ and $X_{m+1}, \ldots, X_n$ are distributed as $N(\mu_2, \sigma^2)$ (here $m < n$, $m \geq 2$, $n - m \geq 2$) Find the maximum likelihood estimates of μ_1, μ_2, and σ^2 based on $X_1, \ldots, X_n$. [4+3=7 points]

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

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

a) Find the Bayes estimator when $0 < \alpha < 1$ and $0 < \beta < 1$.

b) Find the Bayes estimator when $\alpha = \beta = 1$. [4+3=7 points]

6. Define the functional $T(F)$ corresponding to the distribution function F as the solution of the equation

$$\int \psi(x, \theta) dF(x) = 0.$$

Thus the functional $T(F)$ satisfies

$$\int \psi(x, T(F)) dF(x) = 0.$$

Find the influence function of the above functional.

(We call the above functional the M-estimator functional, and denote estimators obtained by solving the above equation with F replaced by F_n as M-estimators.) [7 points]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: (2004-2005)

Course Name: Introduction to Sociology

B. Stat.- III year.

Date: 17.9.04 Maximum Marks: 30 Duration: 1½ hours

(Number of copies of the question paper required: Minimum Ten (10))

Group- A

Q.1) Write the correct answer matching the two sides: ½ X 6 = 3

a) Auguste Comte coined the term 'Sociology' in
i) 17th century
ii) 18th century
iii) 19th century
iv) 20th century

b) Classical sociology is primarily
i) Vague
ii) Reactionary
iii) Evolutionary
iv) Unscientific

c) Mechanical solidarity is the concept of
i) Georg Simmel
ii) Herbert Spencer
iii) Karl Marx
iv) Emile Durkheim

d) 'Verstehen' (understanding) implies
i) Outsider's view
ii) Critical view
iii) Conservative view
iv) Insider's view

e) 'Positive' era denotes
i) Theological era
ii) Feudal era
iii) Industrial era
iv) Pre-historic era

f) 'Protestant ethic' (according to Marxian view) is
i) Sub-structure
ii) super-structure

P.T.O

Q.2) Write short answers: 1 X 5 = 5

a) Define 'Social Action' after Max Weber.
b) What, according to Max Weber, are the types of 'Social Action'?
c) Why, according to Emile Durkheim, the rate of suicide among the Protestants is higher than among the Catholics?
d) What, according to Karl Marx, are the two basic classes in the capitalistic social formation?
e) Which categories of people, according to Auguste Comte, are in the driving seat in the 'scientific-industrial' society?

Q.3) Write short note: (any two) 5 X 2 =10

a) Distinguish between 'Functional' model and 'Conflict' model in sociological theory.
b) Distinguish between culture and civilization.
c) What are the main features of modern bureaucracy?
d) Distinguish between 'community' and 'society'.

Group-B

Q.4) Write short answers: (any two)

a) What is the meaning of research? What are the main characteristic features of research? 1+3= 4
b) Distinguish between sociology and other branches of social sciences. 4
c) Distinguish between observation and experiment. 4

Q.5) Write short note: 2 X 2= 4

a) Research Design
b) Social Research

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2004
B. Stat. (Hons.) III year
Elective: Introduction to Anthropology and Human Genetics

Date: 17.9.04 Maximum Marks: 100 Duration: 3 Hours

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

GROUP A

1(a) Define the discipline of anthropology and mention its distinguishing features. (5)

(b) What are the main divisions of anthropology? What does each one of them deal with? (5)

2. Why man is unique in the animal kingdom? Discuss. (10)

3. What are the changes that took place in the anatomical characteristics of man for having erect posture? (10)

3 (a) Homo sapiens is solely characterized by : (2)
   (i) sagittal crest
   (ii) brachial locomotion
   (iii) erect posture

(b) The uniqueness of anthropology lies in: (2)
   (i) racial classification
   (ii) study of organic evolution
   (iii) holistic approach

(c) Pithecanthropus is classified under: (2)
   (i) Homo erectus
   (ii) Homo habilis
   (iii) Homo neandertal

(d) Man's basic adaptive technique is: (2)
   (i) culture
   (ii) aggressiveness
   (iii) reversible thumb

(e) Primates include: (2)
   (i) prosimians, monkeys, apes and humans
   (ii) amphibians and reptiles
   (iii) molluscs and invertebrates

(Choose the correct answer from three options given in question 4 a through e)

[See overleaf]

5. Describe the characteristic features of order *primate* and suborder *anthropoidea* (10)

6. Write short notes on any two: (5 X 2)
   (i) Australopithecine
   (ii) New world monkey
   (iii) Anthropoid apes

GROUP B

7. Write briefly about types of biological variation. (10)

8. What is Population Structure and Population Composition? Write briefly about the interrelationships between the two. (5 x 2)

9. 'Population Structure plays a significant role in the genetic makeup of human populations' – Discuss (10)

10. What are different causes of human variation? Write briefly about any two causes with examples. (5 x 2)

11. 'Indian populations show wide diversity' - Explain with examples. (10)

12. Write short notes on any two of the following: (2 x 5)

    a. Endogamy – Exogamy b. Genetic Diversity
    c. Phenotype – Genotype d. Marital Distance

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : (2004 - 2005)

B. Stat (Hons) III Year

Economics III

Date : 17 September 2004 Maximum Marks : 25 Duration : 2 Hours

Note: This paper carries 27 marks. Answer as many questions as you like.
The maximum you can score is 25.

1. (i) What is a consistent estimator? [2]
   (ii) Define stationarity for a time series. Check for stationarity of the following process.

$$y_t = 1.5\, y_{t-1} + 0.9\, y_{t-2} + \varepsilon_t$$

   Find the Autocovariance function for this process. [2 + 2 + 2 = 6]
   (iii) Give an algorithm for estimating the parameters of a *Logit* model with one explanatory variable. [4]
   (iv) Define the *Variance Inflation Factor* and explain its usefulness in detecting multicollinearity in a given data set. [2 + 2 = 4]

2. You are given the following data on y, x_1 and x_2.

y	1	1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	1	-1
x_1	1	-1	-1	-1	1	1	1	1	1	1	-1	-1	1	-1	-1	1	-1	1	-1	-1
x_2	-1	-1	-1	1	1	1	1	1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1

Using suitable technique, test whether (i) x_1 and x_2 are useful in predicting y, (ii) x_1 is useful in predicting y, and (iii) x_2 is useful in predicting y. [5 + 3 + 3 = 11]

INDIAN STATISTICAL INSTITUTE

B. Stat. III Year (2004-2005), Differential Equations

Semestral-I Examination

Time: 3 hrs: Max. Marks 100: Date: 29-11-2004.

Note: This paper carries 110 marks. You may answer all the questions, but the maximum you can score is 100.

1. Consider the initial value problem for the system of equations:

$$\begin{aligned} y' &= 2x + z \\ z' &= 3xy + x^2 z \end{aligned}$$

$(y' = \frac{dy}{dx},\ z' = \frac{dz}{dx})$

$$y(0) = 2, z(0) = 0.$$

Starting with the initial functions $y_0(x) = 2$, $z_0(x) = 0$, apply the method of successive approximations and determine $y_3(x)$ and $z_3(x)$. [10]

2. (a) If n is a non-negative integer, show that the Legendre equation

$$(1 - x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + n(n+1)y = 0$$

has a solution $P_n(x)$ which is a polynomial of degree n. Show also that $P_n(x)$ is unique except for constant multiples.

(b) If $m,\ n$ are non-negative integers with $m \neq n$, then show that

$$\int_{-1}^{1} P_m(x)P_n(x)dx = 0.$$

(c) Show that any polynomial $Q(x)$ of degree n can be written as a linear combination

$$Q(x) = \sum_{k=0}^{n} a_k P_k(x)$$

with $a_n \neq 0$.

(d) Show that $P_n(x)$ has n simple zeros, all lying in the interval $(-1, 1)$.

[8+8+6+8=30]

3. (a) Find the eigenvalues of the boundary value problem:

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

and find the corresponding eigenfunctions.

P. T. O

(b) Find the Green's function of the boundary value problem:

$$x^2y'' - 2xy' + 2y = f \text{ on } (1,2),$$

$$y(1) = y(2) = 0.$$

Hence express the solution of the boundary value problem as an integral.

[10+10=20]

4. (a) Solve the initial value problem:

$$X' = \begin{pmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 3 & 6 & 2 \end{pmatrix} X, \quad X(0) = \begin{pmatrix} -1 \\ 2 \\ -30 \end{pmatrix}.$$

(b) Find the general solution of the non-homogeneous system:

$$X' = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} X + \begin{pmatrix} 2 \\ -1 \end{pmatrix} e^t.$$

(c) Solve the Cauchy problem:

$$\frac{\partial u}{\partial x} - 2\frac{\partial u}{\partial y} + 3u = 0, \quad u(0,y) = 1 + \sin y.$$

[10+10+10=30]

5. (a) If L is the length of a closed smooth plane curve, and if A is the area enclosed by the curve, then show that $A \leq \frac{L^2}{4\pi}$. If $A = \frac{L^2}{4\pi}$, then show that the curve is a circle.

(b) Show that the geodesics on a sphere are arcs of great circles.

[8+12=20]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: (2004 - 2005)

B. Stat. – III Year

Subject: Introduction to Sociology

Date: 1.12.04 Maximum Marks: 50 Duration: Two Hours

Group – A

Q.1) Write appropriate answer out of the options given: $1 \times 4 = 4$

a) 'Sanskritization' denotes :
1) Cultural process of social change
2) Structural process of social change
3) Economic process of social change

b) The study of peasantry is:
1) Related to the study of production organization
2) Related to the study of material culture of tribes
3) Related to the study of isolated, autonomous social systems

c) 'Caste' is gaining strength:
1) In the context of ritual observance
2) In the context political organization
3) In the context of industrialization

d) In hierarchical social system
1) There is dissociation between caste and land ownership
2) There is association between caste and land ownership
3) There is no relation between caste and land ownership

Q. 2) Answer **any three** of the following:

a) How do you define 'closed system' and 'open system'? What, according to Andre Beteille, are the major factors that have contributed in a big way to changes in the distribution of power in the village? 2 + 5 = 7

b) What is meant by 'cumulative' inequalities? What are the important factors responsible for cumulative inequalities in rural India? 2 + 5 = 7

c) How do you define the phenomenon of caste-class convergence in rural Bengal in the '50s? Discuss, after Ramkrishna Mukherjee, the phenomenon of caste-class convergence in rural society of Bengal. 2 +5 = 7

P.T.O

d) would you agree that there has been a 'rise of a class of ambidextrous individuals who are equally at ease in handling bureaucratic procedures and personal contacts'? Put forward your argument. 7

e) What do you mean by 'hierarchical' structure of social organization? How does it differ from 'impersonal' or rule based structure of social organization? 3+4=7

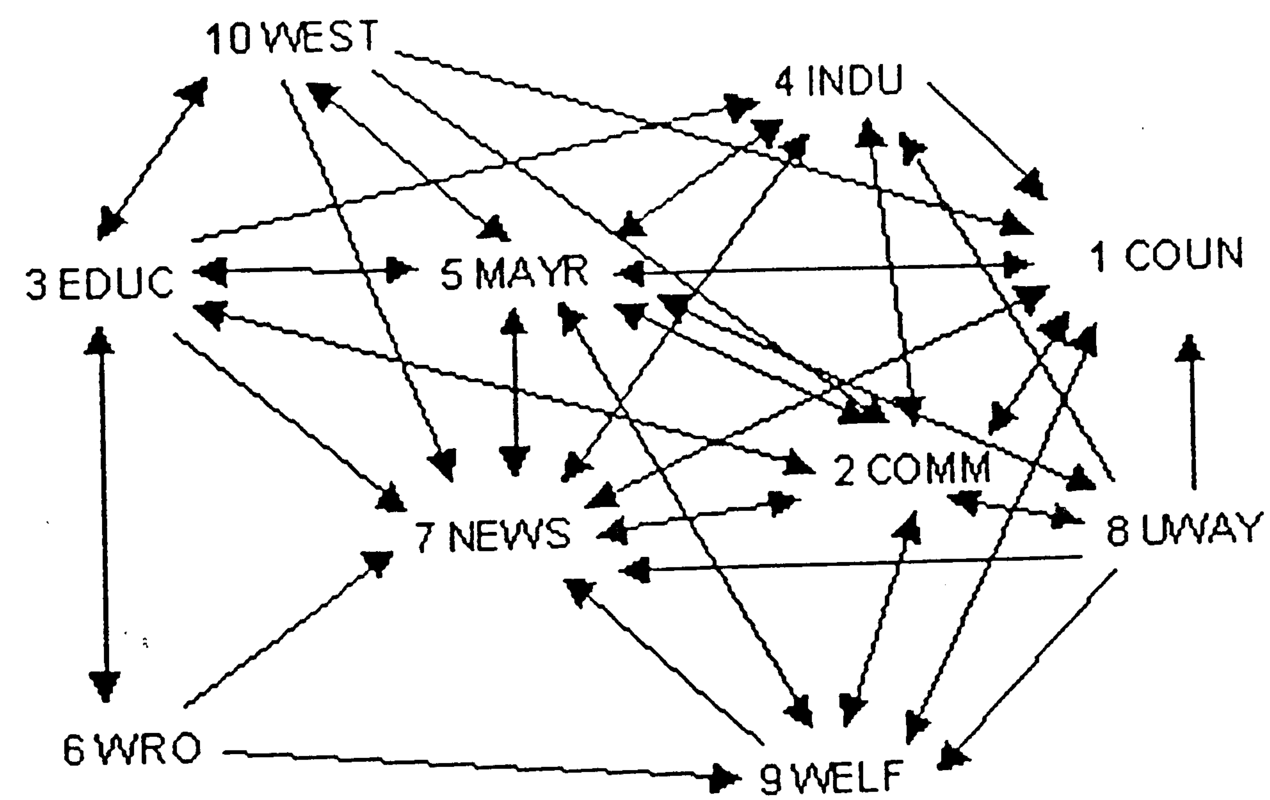
Q. 3) The following digraph (attached herewith) represents the 'information exchange' data among 10 individuals:

a) Transform the data in a 0 – 1 matrix. 4

b) Give the formula of finding the number of reciprocal pairs from the above matrix. What is its value in this digraph? 1+2=3

c) Find the percentage of ordered pairs (i, j) such that distance from i to j is one. 1

d) Obtain the out-degree and in-degree sequences of the digraph using adjacency matrix. 2



Group – B

Q. 4) Answer **any two** of the following:

a) What is an interview? Describe the techniques of an effective interview. 2 +3 = 5

b) Give a note on experiences of field level research. 5

c) What is an observation? Examine the significance of observation as method of data collection. 2 +3 = 5

d) What is a questionnaire? Compare the close-ended and open-ended Questionnaire. 2 + 3 = 5

Q. 5) Writes short note on 2½ × 2 = 5

a) Statistical Hypothesis,

b) Data analysis

***

INDIAN STATISTICAL INSTITUTE
First Semestral Examination (2004-2005)
B.Stat. (Hons) III year
Elective: Introduction to Anthropology and Human Genetics

Date: 1.12.04 Maximum Marks: 100 Duration: 3 Hours

Note: Use separate answer script for Group A and Group B.
Answer any five from each group.

GROUP A

1. Choose the correct answer: [5×2]

A. Cultural evolution and biological evolution are two separate phenomena that do not influence each other: True/ False

B. Homeostasis essentially involves maintenance of a stable internal environment of an organism: True/ False

C. Which of the following implies marriage of a woman with more than one man at a time:
(i) Polygyny (ii) Polyandry (iii) Monogamy (iv) Hypergamy

D. Biological Anthropology as a sub-discipline deals with:
(i) The study of primitive man (ii) The comparative study of society (iii) The study of human behaviour (iv) Study of human evolution and biological variation

E. Man's basic adaptive technique is :
(i) Culture (ii) Aggressiveness (iii) Reversible thumb (iv) Non-seasonal sexuality

2. Compare and contrast the concepts and features of mating and marriage in humans. [10]

3. What is genetic adaptation? How does it differ from cultural adaptation? [5 + 5]

Or

Define adaptation and acclimatization. What are the physical environmental stresses characteristic of high altitude? [2 + 8]

4. Discuss the relative merits of typological, populational and clinal models in understanding patterns of human biological variation. [10]

5. (a) 'In 1900 life expectancy at birth for males and females in the US were 46.6 and 48.7 years, respectively and the corresponding figures for 2000 became 74.3 and 80.9 years'. What does this

[P.T.O.]

change mentioned in the statement signify in terms of population ageing. [5]

(b) 'Low birth weight is an indicator of maternal-fetal malnourishment'. Illustrate this statement . [5]

6. Write short notes on any two of the following: [2×5]
(i) Functional adaptation (ii) Senescence (iii) Physical growth and development (iv) Twin study

Group B

7. What is Hardy-Weinberg Equilibrium? Why is it important? (5×2)

8. What is population structure and population composition? Write briefly about the interrelationship between the two. (5×2)

9. Genetic Drift and Founder effect are described as stochastic forces of evolution. Explain. (5×2)

10. Colour blindness in man is due to a sex-linked recessive gene. A survey of 500 men from a local population revealed that 10 were colour blind. (5×2)
    (a) What is the gene frequency of the normal allele in the population?
    (b) What percentage of the females in this population would be expected to be normal?

11. The allele frequencies at the ABO locus in a population are 0.27, 0.08 and 0.65 for the alleles corresponding to blood types A, B and O. The alleles for blood types A and B are co-dominant and the allele for blood type O is recessive.
    Calculate the genotype frequencies at this locus assuming Hardy-Weinberg equilibrium. (5×2)

12. Write short notes on any two of the following: (5×2)

a. Mutation	c. Heritability	e. Genetic Polymorphism
b. Mitochondrial DNA	d. Phenotype-Genotype	

***

INDIAN STATISTICAL INSTITUTE
First Semestral Examination: (2004-2005)
B.Stat. III Year

GEOLOGY
Date: 2.12.04 Maximum Marks: 100, Duration : 3 hours

Answer all the following questions:

1.
a) What is a fossil? How does it occur in nature?
b) Write a short note on the evolution of hominids.
c) Explain how fossils help to understand the age of the host rock.
d) What are the possible causes of biotic explosion during Cambrian Period?

4X4 = 16

2.
a) What is a supernova explosion? How is it related to formation of planetary systems?
b) Distinguish between planet and meteorite.
c) What are the different lines of evidence indicating that the earth has a layered structure? Discuss briefly the possible mode of origin of the internal layering of the earth.
d) What is the cause of earth's magnetic field? Does the moon have a magnetic field? Justify your answer.

5+2+10+3 = 20

3.
a) Distinguish between a rock and a mineral.
b) Enumerate the different steps you would adopt to identify igneous, sedimentary and metamorphic rocks.
c) What are the physical properties that one examines to identify a mineral?
d) Define and draw sketches of 1) Reclined fold, 2) Plunging-upright fold, 3) Vertical fold.

2+4+4+6 = 16

4.Answer the following questions on the background of your experience in the Ghatshila Field Work:
a) How did you distinguish primary sedimentary layering and foliation in the field?
b) How did you find the top direction of the succession at Ghatshila?
c) What is cross-stratification? How is it formed? How does it help to recognize current direction?
d) Mention with sketches the different soft-sediment deformation structures you have noticed in Ghatshila.
e) What is the geological age of the rocks around Ghatshila? Give physical descriptions of three rock types of that region.

2+2+6+4+4 = 18

5.
a) Distinguish between continental and oceanic crust.
b) Why is the Himalayan belt prone to earthquake?
c) How did continental crust evolve?
d) What are sources of the internal heat of the earth?
e) What will be the consequence of total heat loss of the earth?

3+3+6+3+3 = 18

6.
a) What is metamorphism? Arrange the following rocks in order of increasing metamorphic grade: slate, chlorite-mica schist, gneiss, amphibolite.
b) Sort the following rocks into two pairs so that each pair represents a parent rock and its metamorphic equivalent. Identify the parent rock in each pair: basalt, shale, amphibolite, garnet bearing schist.
c) Write a note on contact metamorphism.

6+4+2 = 12

INDIAN STATISTICAL INSTITUTE

Semestral Examination : (2004 –2005)

B. Stat (Hons) III Year

Economics III

Date : 2 December 2004 Maximum Marks : 50 Duration : 3 Hours

Note: This paper carries 60 marks. Answer as many questions as you like. The maximum you can score is 50.

1. Compare between the given pairs in each of the following cases. In each case, also provide examples as appropriate.
   (a) Backward and Forward selection methods for selecting regressors.
   (b) Infinite lag models and Finite Polynomial lag models for explaining lagged dependence.
   (c) Autoregressive (AR) and Moving average (MA) processes in time series.

[5 x 3 = 15]

2. Consider the following model

$$y_t = \alpha + \beta x_t + \varepsilon_t$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + \theta u_{t-1} + u_t$$

where u_t's are white noise terms.
   (a) What are the required restrictions on ρ and θ for the above model to be well defined?
   (b) Find the *Autocovariance* function for ε_t.
   (c) Give an estimation algorithm for the above model.

[2 + 3 + 5 = 10]

[PTO]

3. Consider the following two-equation model

$$y_1 = \gamma_1 y_2 + \beta_{11}x_1 + \beta_{21}x_2 + \beta_{31}x_3 + \varepsilon_1$$
$$y_2 = \gamma_2 y_1 + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3 + \varepsilon_2$$

(a) Verify that, as stated, neither equation is identified.

(b) Establish whether the following restrictions are sufficient to identify the model.

(i) $\beta_{21} = \beta_{32} = 0$ and

(ii) $\beta_{12} = \beta_{22} = 0$.

(c) Suggest an estimation procedure for the model, if identified, for each of the situations in (b) using the *instrumental variables* (IV) method.

(d) Consider the restriction $\beta_{21} = \beta_{31} = \beta_{12} = 0$ for the above model. All variables are measured as deviations from their means. The sample of 25 observations produces the following matrix of sums of squares and cross-products.

	y_1	y_2	x_1	x_2	x_3
y_1	20	6	4	3	5
y_2	6	10	3	6	7
x_1	4	3	5	2	3
x_2	3	6	2	10	8
x_3	5	7	3	8	15

Estimate the two equations by 2SLS and 3SLS methods.

[3 + 2 + 2 + (8 + 10) = 25]

4. (a) What is seasonality in a time series data? Discuss how you take care of seasonality in a linear model.

(b) We want to estimate the inverse demand function for Woolen Sweaters (WS) of the form $P = f(Q)$ where P = price of WS and Q = quantity sold of WS.

Monthly data is available on P and Q and it is desired that the estimated function should be linear and be able to capture seasonal effects.

Specify one statistical problem that you anticipate in estimating such a relationship.

Can you suggest a better specification that takes care of this statistical problem?

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

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2004-05
B. Stat. III Year
Sample Surveys

Date: 6.12.04 | Maximum Marks: 100 | Duration: 3 Hours

Note: 1. Assignment records, carrying 25 marks must be submitted along with the written answer-scripts.
2. Answer any 4 Questions, each carrying 25 marks

Answer any 4 questions.

1. Following the approach of Dalenius show how you may proceed rationally to construct $H(>1)$ strata out of a given population.

(25)

2. Given a sample drawn from a finite population, no matter how chosen, show how a ratio estimator for a finite population total may be treated as an optimal one under certain model postulations.

How will you assess the error in estimation by a ratio predictor ?

(10+15=25)

3. Derive Murthy's almost unbiased ratio estimator for a finite population total.

(25)

4. Given a simple random without replacement sample of n clusters from a finite population of N clusters of unknown sizes, show how you may use information on sampled cluster-wise elements in (i) estimating (a) the population total and (b) the population mean and in (ii) assessing the accuracies in estimation ?

(5+5+8+7=25)

5. Deriving appropriate results show how double sampling is useful in employing stratification principle in efficient estimation when the strata-sizes are initially unknown.

Discuss, from the twin consideration of cost and efficiency, how many initial units and the subsequent strata-specific sample units you should choose.

(15+10=25)

6. In order to estimate the current value of a population mean utilizing the previous year's values show how you should be guided by the consideration of the "current and previous value-based correlation coefficient" in deciding on the extent of 'matched sampling units' in sampling on two successive years.

(25)

***

INDIAN STATISTICAL INSTITUTE

STATISTICAL INFERENCE I

Semestral Examination

BStat 3rd Yr, 2004-2005, Semester I

Maximum you can score is 60.

Date: 08.12.04 Time: 3 hours

1. 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. ($F(\cdot)$ is the distribution function of the length of life of the components). Find the uniformly minimum varianced unbiased estimator of $\bar{F}(t)$. [10 points]

2. 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 situations:
(i) if $E_\theta(X) \geq 0$ for all θ and $b < 0$.
(i) if $E_\theta(X) \leq k$ for all θ and $ak + b > k$. [4+4=8 points]

3. Let $\mathcal{F}$ be the set of all continuous distributions, and let $T(F) = F^{-1}(1/2)$ represent the median functional for a distribution $F \in \mathcal{F}$. Show that the influence function of the median functional is a bounded function. [8 points]

4. Let $X_1, \ldots, X_n$ be a random sample from a uniform $(0, \theta)$ distribution.
(i) Determine if there exists a UMP test for testing $H_0 : \theta = \theta_0$ against $H_0 : \theta > \theta_0$. If so, is the test unique?
(i) Determine if there exists a UMP test for testing $H_0 : \theta = \theta_0$ against $H_0 : \theta < \theta_0$. If so, is the test unique?
(i) Determine if there exists a UMP test for testing $H_0 : \theta = \theta_0$ against $H_0 : \theta \neq \theta_0$. If so, is the
[3+3+4=10 points]

P. T. O

5. Let $X_1, \ldots, X_n$ be a random sample from a $N(\theta, 1)$ distribution. Let $0 < \alpha < 1$, and $\theta_1 < \theta_2 < \theta_3 < \theta_4$. Consider testing the following set of hypotheses:

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

Does there exist a level α UMP test for the above? If yes, derive the test. If not, explain why not.

[10 points]

6. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent random samples from exponential distributions with parameters λ_1 and λ_2 respectively. For the hypothesis

$$H_0 : \lambda_1 = \lambda_2 \quad \text{vs} \quad H_1 : \lambda_1 > \lambda_2$$

derive the uniformly most powerful unbiased (UMPU) test at level α. [10 points]

7. Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with unknown mean μ and unknown variance σ^2. Consider testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. Determine the form of the likelihood ratio test for the above hypotheses and show that the final decision can be taken based on the quantiles of an F distribution. [10 points]

Semestral I
Linear Models
B.stat. III-2004-2005
Full Marks. 100
Answer Question 5. and any three from the rest

10. 12. 04 Time 10:30 - 2:00

Consider the Gauss Markov set up $(Y, X\beta, \sigma^2 I)$. Let $H_0 : C\beta = 0$ be a testable hypothesis. Derive the likelihoo ratio test statistic for testing H_0. Show that this test statistic is equivalent to the test statistic derived from confidence ellipsoid.

[18 + 6 = 24]

The maximum output voltage of a particular type of storage battery is thought to be influenced by the material used in the plates and the temperature in the location at which the battery is installed. An experiment is run in the laboratory with 4 replicates for each of the 4 material and 3 temperature combinations.

(a) Suggest a suitable linear model for the experiment.

(b) How can you logically frame a hypothesis to test for the presence of interaction of material and temperature? Derive an appropriate test statistic for testing the hypothesis.

(c) Suggest a hypothesis and a test statistic for testing "'no difference in material effect"'.

(d) Write down the complete ANOVA table for the model suggested by you.

[2+ (4+11) + 4 + 3 =24]

A set of t drugs, each having d dose levels is administered to subjects divided into b groups. Each dose level of every drug is applied to m subjects of every group. The response is a measure of degree of relief caused by the drug. Write down a suitable linear model with explicit form of the design matrix for this set up and construct the corresponding ANOVA table.(Derivation of the S.S is not needed)

(a) Let $Y_i, i = 1, ...6$ be uncorrelated observations from a normal distribution with $\Sigma = \sigma^2 diag(1, 2, 1, 2, 1, 2)$ such that $E(Y_1) = E(Y_3) = E(y_5) = \beta_1 + \beta_2$, $E(Y_2) = E(Y_4) = E(y_6) = \beta_2 + \beta_3$.

i. Show that β_i 's are not estimable for $i = 1, 2, 3$.

ii. Find the independent error functions. Use them to obtain the BLUE of $\beta_1 - \beta_3$ starting from an u.e $Y_1 - Y_2$ and compute its variance.

[(6 +8) + 2 + (2+4+2) = 24]

P.T.O

4. Consider the usual multiple linear regression model :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i, \quad i = 1, ..., 15.$$

with

$$X = \begin{vmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & -2 \\ 1 & 0 & 0 & 2 \end{vmatrix}$$

Call $\nu_2' = (\beta_2, \beta_3)$.

(a) Obtain an appropiate test statistic in order to test the hypothesis $H_0 : \nu_2 = 0$. Writ down the complete ANOVA table.

(b) Is this an orthogonal design? If so, justify . If not, give reasons and turn it to a orthogonal design with minimum change in the above design matrix X.

(c) In case your answer in (ii) above , is in the negative under the revised design matri suggested by you, modify the test statistic for the above hypothesis H_0.

[(12 +4)+ 3+5 =2

5. Consider the results given below, of an experiment conducted to study the effects of thre feeding treatments on the weight gain of pigs. The initial weight(x) and the weight gain(y were recorded for each pig.

Food	x	y
	38	9.52
	35	8.21
A	41	9.32
	48	10.56
	43	10.42
	39	8.51
	38	9.95
B	46	8.43
	40	8.86
	40	9.20
	48	9.11
	37	8.50
C	42	8.90
	42	9.51
	40	8.76

a) Analyse the data and draw proper conclusions after carefully formulating an appropriate model.

b) Derive and compute 95% simultaneous confidence intervals for pairwise comparisons of food using Tukey's multiple comparison procedure.

[20+(4+4)= 28]

P.T.O

Table A.8 Tukey's method:* Upper α critical values, $q_{v,df,\alpha}$, of the Studentized range distribution

df	α	v = 2	3	4	5	6	7	8	9	10	12	14	16	18	20
2	0.01	14.0	19.0	22.3	24.7	26.6	28.2	29.5	30.7	31.7	33.4	34.8	36.0	37.0	37.9
	0.05	6.08	8.33	9.80	10.9	11.7	12.4	13.0	13.5	14.0	14.7	15.4	15.9	16.4	16.8
	0.10	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9.41	9.72	10.3	10.7	11.1	11.4	11.7
3	0.01	8.26	10.6	12.2	13.3	14.2	15.0	15.6	16.2	16.7	17.5	18.2	18.8	19.3	19.8
	0.05	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.95	10.3	10.7	11.0	11.2
	0.10	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29	7.67	7.98	8.25	8.48	8.68
4	0.01	6.51	8.12	9.17	9.96	10.6	11.1	11.5	11.9	12.3	12.8	13.3	13.7	14.1	14.4
	0.05	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.21	8.52	8.79	9.03	9.23
	0.10	3.01	3.98	4.59	5.03	5.39	5.68	5.93	6.14	6.33	6.65	6.91	7.13	7.33	7.50
5	0.01	5.70	6.98	7.81	8.42	8.91	9.32	9.67	9.97	10.2	10.7	11.1	11.4	11.7	11.9
	0.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.32	7.60	7.83	8.03	8.21
	0.10	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82	6.10	6.34	6.54	6.71	6.86
6	0.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.48	9.81	10.1	10.3	10.5
	0.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.79	7.03	7.24	7.43	7.59
	0.10	2.75	3.56	4.07	4.44	4.73	4.97	5.17	5.34	5.50	5.76	5.98	6.16	6.32	6.47
7	0.01	4.95	5.92	6.55	7.02	7.39	7.70	7.98	8.21	8.43	8.80	9.11	9.38	9.62	9.84
	0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.81	5.99	6.15	6.42	6.65	6.84	7.00	7.15
	0.10	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28	5.53	5.74	5.91	6.06	6.20
8	0.01	4.74	5.64	6.21	6.63	6.97	7.24	7.48	7.69	7.88	8.20	8.46	8.70	8.90	9.08
	0.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.17	6.39	6.57	6.72	6.86
	0.10	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4.99	5.13	5.36	5.56	5.73	5.87	6.00
9	0.01	4.60	5.43	5.96	6.35	6.66	6.92	7.14	7.33	7.50	7.79	8.03	8.23	8.41	8.57
	0.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.98	6.19	6.36	6.51	6.64
	0.10	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01	5.23	5.42	5.58	5.72	5.85
10	0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.88	7.05	7.21	7.48	7.71	7.90	8.07	8.22
	0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.83	6.03	6.19	6.34	6.47
	0.10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91	5.13	5.32	5.47	5.61	5.73
11	0.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.25	7.46	7.64	7.80	7.94
	0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.71	5.90	6.06	6.20	6.33
	0.10	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84	5.05	5.23	5.38	5.51	5.63
12	0.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.68	6.82	7.07	7.28	7.46	7.62	7.76
	0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.26	5.39	5.61	5.80	5.95	6.08	6.20
	0.10	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78	4.99	5.16	5.31	5.44	5.55
14	0.01	4.21	4.89	5.32	5.63	5.88	6.09	6.26	6.41	6.55	6.77	6.97	7.13	7.27	7.40
	0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.46	5.64	5.78	5.91	6.03
	0.10	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68	4.88	5.05	5.19	5.32	5.43
16	0.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.56	6.74	6.90	7.03	7.15
	0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.35	5.52	5.66	5.79	5.90
	0.10	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61	4.80	4.97	5.11	5.23	5.33
18	0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.41	6.58	6.73	6.85	6.97
	0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.27	5.43	5.57	5.69	5.79
	0.10	2.45	3.10	3.49	3.77	3.98	4.16	4.31	4.44	4.55	4.75	4.90	5.04	5.16	5.26
20	0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.28	6.45	6.59	6.71	6.82
	0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.20	5.36	5.49	5.61	5.71
	0.10	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51	4.70	4.85	4.99	5.10	5.20
24	0.01	3.96	4.55	4.91	5.17	5.37	5.54	5.68	5.81	5.92	6.11	6.26	6.39	6.51	6.61
	0.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.10	5.25	5.38	5.49	5.59
	0.10	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44	4.63	4.78	4.91	5.02	5.12

Table A.8 Tukey's method:* Upper α critical values, $q_{v,df,\alpha}$, of the Studentized range distribution

df	α	v = 2	3	4	5	6	7	8	9	10	12
2	0.01	14.0	19.0	22.3	24.7	26.6	28.2	29.5	30.7	31.7	33.4
	0.05	6.08	8.33	9.80	10.9	11.7	12.4	13.0	13.5	14.0	14.7
	0.10	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9.41	9.72	10.3
3	0.01	8.26	10.6	12.2	13.3	14.2	15.0	15.6	16.2	16.7	17.5
	0.05	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.95
	0.10	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29	7.67
4	0.01	6.51	8.12	9.17	9.96	10.6	11.1	11.5	11.9	12.3	12.8
	0.05	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.21
	0.10	3.01	3.98	4.59	5.03	5.39	5.68	5.93	6.14	6.33	6.65
5	0.01	5.70	6.98	7.81	8.42	8.91	9.32	9.67	9.97	10.2	10.7
	0.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.32
	0.10	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82	6.10
6	0.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.48
	0.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.79
	0.10	2.75	3.56	4.07	4.44	4.73	4.97	5.17	5.34	5.50	5.76
7	0.01	4.95	5.92	6.55	7.02	7.39	7.70	7.98	8.21	8.43	8.80
	0.05	3.34	4.16	4.68	5.06	5.36	5.61	5.81	5.99	6.15	6.42
	0.10	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28	5.53
8	0.01	4.74	5.64	6.21	6.63	6.97	7.24	7.48	7.69	7.88	8.20
	0.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.17
	0.10	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4.99	5.13	5.36
9	0.01	4.60	5.43	5.96	6.35	6.66	6.92	7.14	7.33	7.50	7.79
	0.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.98
	0.10	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01	5.23
10	0.01	4.48	5.27	5.77	6.14	6.43	6.67	6.88	7.05	7.21	7.48
	0.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.83
	0.10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91	5.13
11	0.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.25
	0.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.71
	0.10	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84	5.05
12	0.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.68	6.82	7.07
	0.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.26	5.39	5.61
	0.10	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78	4.99
14	0.01	4.21	4.89	5.32	5.63	5.88	6.09	6.26	6.41	6.55	6.77
	0.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.46
	0.10	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68	4.88
16	0.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.56
	0.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.35
	0.10	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61	4.80
18	0.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.41
	0.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.27
	0.10	2.45	3.10	3.49	3.77	3.98	4.16	4.31	4.44	4.55	4.75
20	0.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.28
	0.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.20
	0.10	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51	4.70
24	0.01	3.96	4.55	4.91	5.17	5.37	5.54	5.68	5.81	5.92	6.11
	0.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.10
	0.10	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44	4.63

INDIAN STATISTICAL INSTITUTE
First Semester Back Paper Examination: 2004-05
B. Stat III Year
Sample Surveys

Date: 7.1.05 Maximum Marks: 100 Duration: 3 Hours

Answer any four questions.

1. Derive appropriate formulae to show how you may employ two-phase sampling in a regression based estimator for a population mean.

   Apply the twin consideration of cost and efficiency in deciding on how many sample units you should choose in the two phases.

   (15+10=25)

2. Derive results to show how to unbiasedly estimate a finite population mean if there be a partial non-response in an initial sample followed by another admitting full response.

   Find optimal sampling and sub-sampling sizes in suitable ways.

   (15+10=25)

3. How will you unbiasedly estimate the population mean from a linear systematic sample? Assuming that the population size is an integral multiple of the sample size, find a formula for the variance of your estimator to show how it is affected by the 'intra-systematic sample correlation coefficient'.

   (12+13=25)

4. If a sample of clusters be chosen by PPS sampling with replacement and each chosen cluster is independently sub-sampled every time it is selected, find an unbiased estimator for the population mean and variance of the estimator.

   Show how you may unbiasedly estimate the variance even if a single systematic sample be drawn independently from each chosen cluster every time it is selected.

   (25)

5. Illustrate the application of 'double sampling' in stratification and discuss associated results on unbiased estimation of the population mean and the variance of the estimator.

   (25)

***

ꟼ STATISTICAL INSTITUTE
First Semester Back Paper Examination 2004-05

B. STAT. III YEAR
Linear Statistical Models

Date: 11.01.2005 Maximum Marks : 100 Duration : $3\frac{1}{2}$ Hours

Answer all questions.

1. Consider a two way classification model with one observation per cell.

   a) Write down to ANOVA table. (Deviation of different S.S is not needed).

   b) Can you introduce the usual interaction effects of the two factors in the model and carry out a test for their presence in the model. Justify your answer.

   c) Modify the model to a nonadditive one (with proper justification) to introduce some form of interaction of the factors.

   d) Derive an appropriate test statistic to test the absence of nonadditive parameter. Derive its distribution under the null hypotheses.

[6+2+3+14=25]

2. Consider a one way classification model :

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$
$$i = 1, \ldots k$$
$$j = 1 \ldots n_i$$

   a) Show that $\sum C_i \alpha_i$ is not estimable unless $\sum C_i = 0$. Derive an appropriate test statistic to test equality of $\alpha_i' s$, $i = 1, \ldots k$

   b) Consider $H_0 : \frac{\mu+\alpha_1}{C_1} = \frac{\mu+\alpha_2}{C_2} = \frac{\mu+\alpha_k}{C_k}$ where

   $C_i' s$ are known constants. Derive an appropriate test statistic to test H_0.

Contd...2/-

c) Discuss Scheffe's multiple comparison procedure with reference to this model.

d) If one regressor variable x is included in the model above, how will you carry out the test for equality of $\alpha_i' s$, $i = 1, \ldots k$?

[2+5+7+6+10=30]

3. a) Consider the standard gauss - Markov set up $(Y, X\beta, \sigma^2 I)$. Prove that $\underline{a}'\underline{Y}$ is the BLUE of the estiable pametuce function $\underline{p}'\beta$ iff $Cov(\underline{a}'\underline{Y}, \underline{\ell}'\underline{Y}) = 0$ for every linear zero function $\underline{\ell}'\underline{Y}$.

b) Derive a sufficient condition for equality of the OLSE with the BLUE under $(Y, X\beta, \Sigma)$.

[6+6=12]

4. Write down a random effects model and the corresponding ANOVA for a two way classification set up with r (>1) observation per cell. Indicate appropriate hypothesis of testing and corresponding test statistics.

[8]

5. A mechanical engineer is testing the performance of three machines. Each machine can be operated at two power settings. Furthermore, a machine has three stations at which the product is formed. An experiment is conducted where each machine is tested at both power settings. Three observations one taken from each station. The results are shown below. Analyse the experiment assuming that all the factors are fixed.

Machine Selection	1			2			3		
	1	2	3	1	2	3	1	2	3
Power Setting 1	34.1	33.7	36.2	32.1	33.1	32.8	32.9	33.8	33.6
	30.3	34.9	36.8	33.5	34.7	35.1	33.0	33.4	32.8
	31.6	35.0	37.1	34.0	33.9	34.3	33.1	32.8	31.7
Power Setting 2	24.3	28.1	25.7	24.1	24.1	26.0	24.2	23.2	24.7
	26.3	24.3	26.1	25.0	25.1	27.1	26.1	27.4	22.0
	27.1	28.6	29.9	26.3	27.9	23.9	25.3	28.0	24.8

[25]

***

INDIAN STATISTICAL INSTITUTE

First Semestral Back Examination: (2004-2005)

B.Stat. III Year

GEOLOGY

Date: 13 January 2005 *Maximum Marks: 100* *Duration : 3 hours*

Answer all the following questions:

1. What is a rock? Describe three principal types of rocks. *20*
2. Distinguish between plutonic and volcanic rock. *5*
3. What are the physical properties that one examines to identify a mineral? *10*
4. What is a fossil? In which kind of rock do fossils occur? *7*
5. Describe briefly the internal structure of the earth. *15*
6. How did the Himalaya evolve? *10*
7. What is earthquake? Why does it happen? *10*
8. What is the geological age of the rocks around Ghatshila? Give physical descriptions of thre rock types of that region. *13*
9. Distinguish between continental and oceanic crust. *5*
10. Name the different planets of the Solar System. *5*

INDIAN STATISTICAL INSTITUTE

B-Stat III, Semester I, 2004-2005

STATISTICAL INFERENCE I

Date: 13.1.05 Backpaper Examination Time: 3 hours

Answer as many as you can. Total points 100

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

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

We will denote this by $Exp(a, b)$, 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 that they are distributed as $Exp(a, b/n)$ and $\frac{1}{2}b\chi^2_{2n-2}$ random variables respectively). [5+10=15 points]

2. Suppose that $X_1, \ldots, X_n$ represent a random sample from the Poisson distribution with mean θ. We are interested in estimating $P[X = 0]$. Find the uniformly minimum varianced unbiased estimator of $P[X = 0]$. [15 points]

3. Suppose the random variable X has a double exponential distribution with density $f_\xi(x) = \frac{1}{2}e^{-|x-\xi|}, -\infty < x < \infty$.

(a) Based on one observation from $f_\xi(x)$, find a most powerful test of level α for testing $H_0 : \xi = 0$ against $H_1 : \xi = \xi_1$ $(\xi_1 > 0)$. Is the test unique?

Now suppose that $X_1, \ldots, X_n$ represents a random sample of size n (≥ 2) from the above distribution. Consider testing $H_0 : \xi = 0$ against $H_1 : \xi > 0$. The first derivative of the power function exists, is continuous, and may be taken inside the integral sign.

P. T. O

(b) Find the LMP size α test for the above hypothesis for $n \geq 2$ and $0 < \alpha < 1$, and give the form of the critical region.

(c) For the general model $f_\theta(x)$, can you describe an approximate way of finding approximate critical values for the LMP test for large n. [8+8+4=20 points]

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

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

Show that the only unbiased test for testing $H_0 : \theta_1 \leq a, \theta_2 \leq b$ against $H_1 : \theta_1 > a$, or $\theta_2 > b$, or both is $\phi(x,y) = \alpha$. [20 points]

5. Consider a geometric random variable with probability mass function $f_p(x) = p(1-p)^x$, $x = 0, 1, \ldots$ Based on a single observation X from this distribution, derive the critical function for the UMP unbiased test of level $\alpha = 0.2$ for $H_0 : p = 1/2$ against $H_1 : p \neq 1/2$. [15 points]

6. Suppose that $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ represent independent random samples from $N(\xi, 1)$ and $N(\eta, 1)$ respectively. Consider testing $H_0 : \eta \leq \xi$ against $H_1 : \eta > \xi$. Show that there exists a UMP test, which rejects H_0 when $\bar{Y} - \bar{X}$ is too large. [15 points]

Indian Statistical Institute

Statistical Inference II

B. Stat (hons.) III (2004-2005)

Midterm Examination

Maximum score: 60 *Duration: 3hrs.*

This paper bears 70 marks. The maximum you can possibly score is 60. You can use any result done in class, provided that you state it correctly before using it. This is a closed book, closed note examination.

1. State whether the following statements are true or false. Prove the true statements and disprove/provide counterexample for the false ones.

   (a) $X_1, \ldots, X_{10}$ are iid with mean μ. Then there exists a U-statistic of degree 1 to estimate μ^2 unbiasedly.

   (b) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an iid sample from some bivariate distribution. Let S_i's be the concommitants of the Y_i's with respect to the X_i's. If X_i's are independent of the Y_i's then S_1 has discrete uniform distribution over $\{1, \ldots, n\}$.

   (c) Let $X_1, \ldots, X_n$ be iid F, and $Y_1, \ldots, Y_n$ be iid G. The X's are independent of the Y's. Let R_i's be the ranks of the X's among themselves, and S_i's those of Y's among themselves. Let $D_i = R_i - S_i$. Then D_i's are mutually independent.

[5+5+5]

2. Consider the following set up. We have I groups, and J observations from each:

$$X_{ij}, \quad 1 \leq i \leq I, \quad 1 \leq j \leq J.$$

We have a location model:

$$X_{i1}, \ldots, X_{iJ} \sim F(\cdot - \theta_i) \quad \forall i,$$

where F is some unknown continuous distribution function. We want to test

$$H_0 : \theta_1 = \cdots = \theta_I$$

against

$$H_1 : \theta_1 \leq \cdots \leq \theta_I, \text{ with at least one strict inequality}$$

The Jonckheere-Terpstra test uses the test statistic

$$B = \sum_{1 \leq i < j \leq I} U_{ij},$$

where

$$U_{ij} = \#\{(r,s) : X_{ir} < X_{js}\}.$$

   (a) Suggest whether we should reject H_0 for large values of B, or for small values.

   (b) Compute $E(B)$ under H_0. Argue that B is distribution-free under H_0.

(c) Find $E(U_{1,2} \cdot U_{1,3})$ under H_0.

[2+5+10]

3. We have iid data $X_1, ..., X_n$ $(n \geq 2)$. Consider the U-statistic based on the kernel $h(x, y) = xy - y^2$. Compute its Hajek projection. You may assume finite existence of moments of all orders. [10]

4. Prove that the one-sided sign test is consistent. [5]

5. Suppose that we have iid data X_i's of size n from an unknown continuous distribution with finite variance σ^2. We estimate σ^2 using

$$T = \sum (X_i - \overline{X})^2/n.$$

Compute the ideal bootstrap estimate for the bias of T. Suppose that we denote this ideal bootstrap bias estimate by $B\hat{IAS}_{boot}(T)$. Then is it true that

$$T - B\hat{IAS}_{boot}(T) \text{ is unbiased for } \sigma^2?$$

Justify your answer. [10+3]

6. Suppose we have data $X_1, ..., X_m$ iid F and $Y_1, ..., Y_n$ iid G, where F, G are both unknown and continuous. We know that F is a scaled version of G, *i.e.*,

$$F(x) = G(\theta x),$$

for some unknown θ. We want to test

$$H_0 : \theta = 1 \text{ vs } H_1 : \theta > 1.$$

One standard test is Mood test which uses the test statistic

$$T = \sum_{i=1}^{m+n} \left(i - \frac{m+n+1}{2}\right)^2 Z_i,$$

where $Z_i = 1$ if the observation with rank i (in the pooled data) is an X. Otherwise, $Z_i = 0$.

(a) Is T a *linear* rank statistic? Justify you answer.

(b) Should we reject H_0 for large values of T or for small values?

(c) Compute $E(T)$ under H_0.

[3+2+5]

BIII-2004-2005
Midterm Examination
Design Of Experiments

Full Marks –50 Date : 24.2.05 Time: 10.30-12.30

1. Define a completely randomized design (CRD) and a randomized block design(RBD) with example. How will you judge efficiency of RBD with respect to CRD?

4 + 4 =8

2. What do you mean by structural connectedness of a block design? Give an example of a structurally connected non-orthogonal design. Show that a block design is connected iff all the treatment contrasts are estimable.

3 +3 +4 =10

3. From the incidence matrix of a block design given below, obtain the C- matrix find the estimable treatment contrasts and the degrees if freedom associated with the adjusted treatment and adjusted block sums of squares. Is the design orthogonal?

$$N = \begin{matrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0. \end{matrix}$$

6 + 3 +4+5=18

4. Define a Latin Square Design with example. Write down the underlying model and obtain the expression of sums of squares due to treatments.

3 + 11= 14

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination: 2004-2005, B.Stat. III Year

Statistics Comprehensive

Date: 28.2.2005 Maximum Marks: 100 Duration: 3 hours

Answer all questions.

1. a) Aerial observations y_1, y_2, y_3, y_4 are made of the angles $\theta_1, \theta_2, \theta_3, \theta_4$ of a quadrilateral on the ground. Assuming that these observations are subject to independent errors with zero means and common variance σ^2, derive least squares estimates of the θ_i's and obtain an unbiased estimate of σ^2. (Note that here $\sum \theta_i = 360^0$).

   b)The expected distance covered by a car(of a particular brand) after pushing the brake is assumed to be a linear function of the speed at which the car was traveling. Assuming normal distribution of this distance with the corresponding variance being proportional to the square of the speed, find the m.l.e's of the linear functions. Check that they can be viewed as a form of least squares estimate. [8+7=15]

2. a)A sample of size 4 is to be drawn from a population of size 8. It is known that the first and last units have unusually high and low values respectively of the study variable y. So, it was decided to draw a simple random sample of size 2 from units $2, 3, \ldots, 7$ and include units 1 and 8 in every sample. Find an unbiased estimator of the population mean and obtain the variance of this estimator.

   b) A library contains an unknown number N of books serially numbered from 1 to N. on a certain day, n books are issued from the library and their serial numbers noted. Treating the issued books as a simple random sample from the population of N books, show that an unbiased estimate of N is given by $2m - 1$, where m is the mean of the serial numbers of the issued books. If Y be the serial number of the last issued book, find the distribution of Y. [5+10=15]

3. Let independent random variables $Y_1, \ldots, Y_m$ be such that Y_i has the binomial distribution with parameters n_i and p_i, $i = 1, \ldots, m$. Corresponding to response y_i, let the observations on k covariates be x_{ij}, $j = 1, \ldots, k$, $i = 1, \ldots, m$. For the linear logistic link function, obtain the sufficient statistic and derive an expression for the Fisher information for the parameters appearing in the linear function. How are the MLE's of these parameters obtained ? [15]

4. Suppose that we have a dependent variable Y and p independent variables $X_1, X_2, \ldots, X_p$ such that the correlation between Y and X_i is 0.4 and the correlation between X_i and X_j is also 0.4 for all i, j with i different from j. Obtain the multiple correlation coefficient between Y and $X_1, X_2, \ldots, X_p$ and the partial correlation coefficient between Y and X_1 fixing the effects of $X_2, \ldots, X_p$. What happens to this multiple and partial correlation coefficients as p tends to ∞ and give some explanation for that. [15]

5. Consider i.i.d observations $X_1, X_2, \ldots, X_n$ with a common uniform distribution on $[0, \theta]$ where $\theta > 0$ is an unknown parameter. Derive the maximum likelihood estimator and the uniformly minimum variance unbiased estimator for θ based on these observations with full justification. Compare their mean squared errors. [15]

6. Consider the null hypothesis H_0 that says that the observation X has a Poisson distribution with mean 3 against the alternative hypothesis H_A that says that the

observation X has geometric distribution with $P(X = x) = (1/3)(2/3)^{x-1}$ for $x = 1, 2, 3, \ldots$. Derive the form of the most powerful test with level α for any given $0 < \alpha < 1$. Obtain the power of this most powerful test. [15]

7. Class presentation during review. [10]

INDIAN STATISTICAL INSTITUTE

Periodical Examination – Semester II : 2004-2005

B.Stat. (Hons.) III Year

Introduction to Stochastic Processes

: 02.03.05 | **Maximum Score : 80 pts** | Time : 3 Hours

e : This paper carries questions worth a **total** of **96 POINTS**. Answer as much as you The **MAXIMUM** you can score is **80 POINTS**.

Consider the Markov Chain with state space $S = \{1, 2, 3, 4, 5, 6\}$ and transition probability matrix given by $P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$.

(a) Classify the states into recurrent and transient states, giving adequate justification for your answer.
(b) Assuming usual notations, find $f_{31}^{(3)}$. (8+8)=[16]

Show that, for a Markov Chain with d states, if i and j are any two states (not necessarily distinct) such that i leads to j, then $f_{ij}^{(n)} > 0$ for some $n \leq d$. [10]

Show that, for a Markov Chain with only finitely many transient states, there exist constants $C > 0$ and $0 < \lambda < 1$, such that, for any state i and any transient state j, $p_{ij}^{(n)} \leq C\lambda^n$ for all $n \geq 1$. [10]

Consider a Galton-Watson Branching Chain $\{X_n, n \geq 0\}$ with the usual assumptions. Suppose that the number of offsprings produced by each individual can be 0, 1, 2 with probabilities $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{2}$ respectively.
(a) Assuming that $X_0 \equiv 1$, find the expected size of the nth generation and the probability of eventual extinction.
(b) What would be the probability of eventual extinction, if X_0 were a Poisson random variable with parameter $\lambda = 2$? (12+8)=[20]

Consider the Markov Chain with state space $S = \{1, 2, 3, 4\}$ and the transition probability matrix given by $P = \begin{pmatrix} 0 & \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$.

(a) Argue that that chain has a unique stationary distribution and find the stationary distribution.
(b) Given that chain starts in the state 2, what would be the expected time until it returns to the state 2? (10+6)=[16]

Consider the Markov Chain with the state space $S = \{0, 1, 2, \ldots\}$ and the transition probabilites given by $p_{i0} = \frac{1}{i+2}$, $p_{i,i+1} = 1 - p_{i0}$ for all $i \in S$.
(a) Assuming usual notations, find $f_{00}^{(n)}$.
(b) Show that all the states are recurrent.
(c) Examine whether the states are positive recurrent. (6+8+10)=[24]

INDIAN STATISTICAL INSTITUTE

Mid-semester Exam: (2004-2005)

B. Stat. III

Database Management System

Date: 4. 3. 05 Maximum Marks: 50 Duration: 2 hours

Note: The questions carry total **57** marks. Answer as much as you can.

1. Write short notes on: [4x4 = 16]
   (a) Data abstraction
   (b) Database manager
   (c) Indexed sequential file system
   (d) Data independence.

2. Explain the following terms with examples: [4x4 = 16]
   (a) Entity and Entity-set
   (b) Relationship and Relationship-set
   (c) Existence dependency, dominant entity and subordinate entity
   (d) Attribute, super key, candidate key and primary key.

3. A car insurance company maintains a database for its customers to keep track of the claims and their settlements. The company allocates a unique customer number to each of its customers and maintains the name and address of a customer against his/her customer number. The record of each car insured is stored with its year of manufacture, model name and car number. One customer may own more than one car. Detail records of claims are also maintained by the company where claim number, date of claim, date of settlement, amount claimed and amount paid are recorded. A customer makes a claim in case of an accident.

   Draw ER-diagram for the above problem. Reduce all the relations from it and make the normalization, if necessary, and write down the SQL statements to create the tables with suitable constraints. [13+12 = 25]

BIII-2004-2005
Semestral Examination
Design Of Experiments

Full Marks – 100 | Date: 4.5.05 | Time: 10.30-2.00

Answer Question no.5 and any three from the rest. Marks will be deducted for unnecessary long answers.

1. A manufacturing company has 4 makes of machines A,B,C,D which are used to manufacture metal brackets. Each machine requires one skilled worker for its operation. The company wishes to determine which machine is most efficient in providing the brackets. Suppose 5 workers are available for the operation of the machines and each can be used for four trials.

a) Suggest a suitable experiment, ensuring pairwise comparisons among the machines and write down the underlying model.

b) Suppose at the analysis stage it is found that the observation on the acceptable brackets produced using machine A operated by one of the workers is missing. How will you estimate the missing value to obtain the correct value of the error sums of squares under the model assumed by you in (a) above.

c) Can you use the estimated value obtained above to test the equality of performance of the four machines? Justify your answer with mathematical proof.

[(3+2)+6+14=25]

2. a)What is meant by a connected balanced block design?

b) Consider a block design with 6 treatments taking all the combinations with 4 treatments at a time in each block. Is the design connected? Is it balanced? Give reasons.

c) Derive a necessary and sufficient condition for a connected block design to be orthogonal and check whether the design in b) above is orthogonal?

d) Construct a complete set of Mutually Orthogonal Latin square designs of order 4.

[3+ (3+4) + (6+2) + 7=25]

3.a) Show that in a 2^n experiment the total sums of squares due to treatments can be split into the sums of squares due to main effects and interactions orthogonally each carrying one d.f.

b) Consider a factorial experiment with 4 factors A, B, C, D each at 2 levels.Suppose on each day only 4 treatment combinations can be tested. It is desired to have some

P.T.O

information on all the factorial effects and as much information as possible on the main effects and two factor interactions AB, CD and BC. Give a layout of the design for this experiment using suitable confounding scheme if the experiment can be run for eight days.

c) Indicate how will you calculate the sums of squares due to main effects and interactions.

d)Calculate the efficiency of the estimates of main effects and the interactions.

[5+ 12+ 6+2=25]

4. a) What is meant by a balanced confounding scheme in a factorial experiment.

b) Consider a factorial experiment with three factors each at three levels. Give a balanced confounding scheme using 9-plot blocks with minimum number of replications and ensuring no loss of information for main effect contrasts.For only one replication chosen by you, write down the key-block composition and indicate one treatment combination for other blocks indicating how to complete these blocks.

c) Suppose an experiment is to be performed to see the effects of lighting conditions(factorA) and the speed of a rotating drum(factor B) on a subject's ability to focus on the center of the drum. It is easy to change the speed of the rotation of the drum by the turn of a dial, however it takes time to set up a lighting condition. Suppose 6 subjects are available to collect the data.

(i) Suggest a suitable design for this experiment

(ii) Clearly state the underlying model and write down the ANOVA table showing the sources of variaton with corresponding d.f, and different sums of squares.

(iii) Which of these two factors will have more precision in the suggested design.

[2+ (4+ 5)+ (2+2+8+2)=25]

5. Below is given the yields in gm per plot for three varieties of cotton seed using a block design. Varieties are shown in brackets and rows represent blocks.

(1) 57	(2) 37	(3) 68	(1) 48
(2) 42	(3) 63	(1) 58	(2) 39
(1) 53	(3) 66	(3) 61	(2) 35

Test for equality of treatment effects . Obtain the BLUE of $2\tau_2 - \tau_3 - \tau_1$ and an estimate of its standard error.

[17+3+5= 25]

INDIAN STATISTICAL INSTITUTE

Semester Exam: (2004-2005)

B. Stat.(Hons) III

Database Management System

Date: 6.5.05 Maximum Marks: 100 Duration: 3 hours

The questions carry total **105** marks. Answer as much as you can.

1. Consider a software development company that runs several projects. Each project has a unique identification number other than its description, start date, duration, project leader's name etc. The employees in the organization work in these projects. The details like employee number, name, date of join, salary, address etc. are kept for every employee. The employee number is unique for every employee. The company uses a number of computers each having a unique serial number, model name etc. for software development. One project may require more than one computers. One computer may be shared by many projects. The number of hours one computer being used in a project is important for accounting purpose.

   Draw ER diagram for the above problem and reduce it to a set of relations.

   [10+5=15]

2. (a) Define functional dependency and explain with an example.
   (b) State reflexivity, augmentation and transitivity rules.
   (c) Let a scheme R = (A, B, C, G, H, I) and the set F of functional dependencies $\{A \to B, A \to C, CG \to H, CG \to I, B \to H\}$. Verify whether AG is a super/candidate key.

   [5+5+10=20]

3. (a) Give the justification of normalization with the help of an example.
   (b) Define BCNF and 3NF.
   (c) Explain loss-less join decomposition. Suppose we decompose the scheme R = (A, B, C, D, E) into (A, B, C) and (A, D, E). Verify whether the decomposition is loss-less under the set of functional dependencies $F = \{A \to BC, CD \to E, B \to D, E \to A\}$.

   [5+5+10=20]

P.T.O.

4. Consider the following schemes:

Branch = (b-name, assets, b-city);
Loan = (b-name, loan-no., issue date, c-name, amount)
Customer = (c-name, street, c-city);
Deposit = (b-name, ac-no., c-name, balance).

Write relational algebra and tuple relational calculus expressions for the following problems:

(a) Find all customers having a loan from Shyambazar branch and the cities in which they live.
(b) Find all customers having a loan, an account, or both at the Shyambazar branch.

[5+5=10]

5. Consider the schemes in Question No. 4.
Write SQL statements for the following problems:
(a) Find the branch that has issued maximum total amount of loan in the month of January. (Nesting of aggregate functions is NOT allowed)
(b) Find the total deposit amounts of customers in each city.

[5+5=10]

6. Describe deferred and immediate database modification techniques for database recovery with the help of examples.

[5+5=10]

7. Explain the term 'conflicting instruction'.
Consider the following two transactions T and T′ in a schedule S. Find a conflict equivalent serial schedule S′ with proper justification at each step.

T	T′
Read(A)	
Write(A)	
	Read(A)
	Write(A)
Read(B)	
Write(B)	
	Read(B)
	Write(B)

[4+6=10]

8. Write short notes on (a) Hierarchical database and (b) Network database.

[5+5=10]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination: (2004–2005)

Statistical Inference II

B. Stat. (hons.) III

May 10, 2005

Maximum marks: 50 Duration: 3hrs.

This is a closed-book, closed-note examination.

1. You are given 5 closed boxes. You know that two of them has a coin in each. One coin is of value Rs 5, the other Rs 2. You do not know which are those boxes. The other boxes are empty. You have to pay Re 1 to open each box. You may stop after opening k boxes ($0 \le k \le 5$). Your gain is (the total value of coins discovered $-k$). Using dynamic programming principle (DPP) derive an explicit strategy to maximise your expected gain. What is the value of the maximum possible expected gain? [8+2]

2. Consider the usual linear regression model
$$\mathbf{y}_{n\times 1} = X_{n\times p}\beta + \epsilon_{n\times 1},$$
where $\epsilon \sim N_n(0, \sigma^2 I)$. There is a technique called Ridge Regression where we estimate β by
$$\hat{\beta} = (X'X + \lambda I)^{-1}X'\mathbf{y},$$
for some tuning parameter $\lambda > 0$. Write down a cross-validation algorithm to choose a suitable value for λ based on iid data $\{\mathbf{x}_i, y_i : 1 \le i \le 1000\}$. [5]

3. What is Wald's approximation? Suppose that we have iid data with unknown density f. We want to test $H_0 : f = f_0$ against $H_1 : f = f_1$, where f_0, f_1 are known densities. We want an SPRT with error probabilities α and β chosen so that the test is unbiased. We use Wald's approximation to derive the test. Is it true that the approximate test is unbiased? [2+3]

4. Consider sequentially testing a simple H_0 against a simple H_1. Suppose that we have prior p_i for H_i ($p_0 + p_1 = 1$). Each sample costs c, and we incur loss W_i if we reject H_i wrongly. Consider the Bayes risk of a sequential test ϕ
$$R(p_0, W_0, W_1, c, \phi) = \sum p_i(c \cdot ASN_i(\phi) + W_i\alpha_i(\phi)),$$
where ASN_i is the expected sample size under H_i, and α_i's are the error probabilities. Let
$$R(p_0, W_0, W_1, c) = \inf\{R(p_0, W_0, W_1, c, \phi) : \phi \text{ uses at least one sample.}\}$$
Show that $R(p_0, W_0, W_1, c)$ is a continuous function of W_0 when the other arguments are held fixed. Is it true that $R(p_0, W_0, W_1, c)$ is a concave function of c when p_0, W_0, W_1 are held fixed? Justify your answer. [2+3]

P. T. O

ate the sequential Cramer-Rao bound. Suppose that in a production e the items are iid, and each item is "good" with probability $1/\theta$ and "bad" with probability $1-1/\theta$. We have $\theta > 1$. Describe Haldane's inverse sampling scheme to estimate θ. Prove that it attains the sequential Cramer Rao bound. [2+2+6]

6. We want to apply the Propp-Wilson algorithm to the Markov chain using the following transition matrix

$$\begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}.$$

Explicitly mention your choice of the random maps and the corresponding probabilities. Also prove that for this choice the algorithm will eventually coalesce with probability one. [5+5]

7. If $X_1, \dots, X_n$ are iid with continuous density f, and you want to estimate f using kernel density estimation method. Suppose that you are using a kernel $K(x)$, that you can simulate from. Suggest how you can use the kernel density estimate to perform bootstrapping. [2]

8. $X_1, \dots, X_n$ are iid with unknown continuous distribution function G. Derive a nonparametric 95% (approximate) confidence interval for G using the idea of one sample Kolmogorov-Smirnov test. [3]

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2004-05
B. Stat. III Year
Introduction to Stochastic Processes

Date: 13.5.05 Maximum Marks: 120 Duration:3 Hours

The paper carries questions worth a total of 135 marks.
Maximum you can score is 120 marks.
Notations have same meanings as in class.

1. Let $\{X_n, n \geq 0\}$ be a Markov Chain on a state space S and with transition probabilities $p_{ij}, i, j \in S$. Define $Y_n = (X_n, X_{n+1})$ for $n \geq 0$.

(a) Show that $\{Y_n, n \geq 0\}$ is a MC on $S' = \{(i,j) \in S \times S : p_{ij} > 0\}$ and that if $\{X_n, n \geq 0\}$ is irreducible and aperiodic, then so is $\{Y_n, n \geq 0\}$.

(b) Show that if $\{X_n, n \geq 0\}$ has stationary distribution π, then so also does $\{Y_n, n \geq 0\}$.

10+10=20

2. Let $\{X_n, n \geq 0\}$ be a Markov Chain on $S = \{0,1,\dots,d\}$ for which 0 and d are absorbing states, while the rest are transient, communicating with each other. Fix a transient state i and let $\{\widetilde{X}_n, n \geq 0\}$ be the MC with the same transition probabilities as $\{X_n, n \geq 0\}$ except that $p_{0i} = p_{di} = 1$.

(a) Show that $\{\widetilde{X}_n, n \geq 0\}$ has a unique stationary distribution π.

(b) Show that for the chain $\{X_n, n \geq 0\}$, the mean time till absorption, given that it started from state i, equals $(\pi_0 + \pi_d)^{-1} - 1$.

10+10=20

3. Let $\{X_n, n \geq 0\}$ be an irreducible recunent MC on the state space of all integers. Think of it as the motion of a particle. A telescope observing the particle can only observe the states $0,1,\dots\dots,d$. Let $\{Y_n, n \geq 0\}$ denote the successive observations on the telescope.

P.T.O.

(a) Show that $\{Y_n\}$ is a MC with a unique stationary distribution π.

(b) Show that for any two distinct states i, j in $\{0,1,.....,d\}$, the expected number of visits to the state j by the original particle between any two successive returns to i, equals π_j / π_i.

(c) For simple symmetric random walk m integers, find the expected member of visits to a state j between any two successive returns to state i.

10+10+10=30

4.(a) Find the periods of the states for the MC $\{X_n, n \geq 0\}$ on $S = \{1,2,3,4\}$ with transition matrix

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

(b) What are the closed irreducible sets for the MC $\{X_{2n}, n \geq 0\}$? Justify your answer.

10+10=20

5. $\{X_t, t \geq 0\}$ is a Poisson process with intensity λ.

(a) For $s, t > 0$, find $E(X_s X_t)$

(b) Given $X_t = 3$, find the expected time between the first and the second events.

(c) If $\{Y_t\}$ is another Poisson process with intensity μ, independent of $\{X_t, t \geq 0\}$, find the expected member of events for the $\{Y_t, t \geq 0\}$ process between occurence of any two successive events of the $\{X_t, t \geq 0\}$ process.

5+10+10=25

6.(a) Define the Yule process.
For a Yule process, write down the forward equations and then solve them.

(b) For any $0 < x < t$, find the expected number of people at time t of age less than x, given that the Yule process started with 1 individual.

10+10=20

***