

INDIAN STATISTICAL INSTITUTE, KOLKATA
MIDTERM EXAMINATION: FIRST SEMESTER 2012 -'13
B.STAT III YEAR

03.09.12

Subject : **Differential Equations**
Time : 2 hours
Maximum score : 30

Instructions:

- *Justify every step in order to get full credit of your answers, stating clearly the result(s) that you use. Points will be deducted for missing arguments. Partial credit will be given for your approach to the problem.*
- *Switch off and deposit your mobile phones to the invigilator during the entire examination.*
- *Total marks carried by the questions turns out to be 65 which is more than 30. The total marks obtained will be multiplied with $\frac{6}{13}$.*

(1) The function $f : S \rightarrow \mathbb{R}$ satisfies Lipschitz condition, that is, $|f(x, y_1) - f(x, y_2)| \leq K |y_1 - y_2|$ where $K > 0$, $S := [-a, a] \times \mathbb{R}$ and $a > 0$. Set $y_0(x) := b$ and $y_{k+1}(x) := b + b'x + \int_0^x (x-t)f(t, y_k(t))dt$ for all $k \geq 0$ and $x \in [-a, a]$ where $b, b' \in \mathbb{R}$.

(a) Prove that $|y_{k+1}(x) - y_k(x)| \leq \frac{|b'|}{\sqrt{K}} \cdot \frac{|\sqrt{K}x|^{2k+1}}{(2k+1)!} + \frac{M}{K} \cdot \frac{|\sqrt{K}x|^{2k+2}}{(2k+2)!}$ for all $k \geq 0$ where

$$M := \sup_{t \in [-a, a]} |f(t, b)|.$$

[9 marks]

(b) Show that y_k converges uniformly to a continuous function z .

[4 marks]

(c) The z in part (b) solves the initial value problem $y'' = f(x, y)$, $y(0) = b$, $y'(0) = b'$ for $|x| \leq a$.

[9 marks]

(2) (a) Check whether $x = 0$ is an ordinary, regular singular or irregular singular point of the differential equation

$$(0.1) \quad x^2 y'' + 5xy' + (3 - x^2)y = 0$$

[1 mark]

(b) Find a Frobenius series (with at least six terms) solution of Equation 0.1.

[7 marks]

(c) Can one obtain another Frobenius series solution which is linearly independent to that in part (b)?

[5 marks]

(3) (a) Let g be a differentiable function and $y_p(x) = \int_0^x g(t) \cos(x-t)dt$. Find y'_p , y''_p and show that y_p satisfies the differential equation

$$(0.2) \quad y'' + y = g'(x)$$

[7 marks]

(b) Using the method of variation of parameters, show that the general solution of Equation 0.2 is $\int_0^x g'(t) \sin(x-t)dt + A \sin(x) + B \cos(x)$ for $A, B \in \mathbb{R}$.

[4 marks]

(c) Find the values of A and B in the general solution (in part (b)) which give the particular solution y_p (in part (a)).

[4 marks]

(4) Find the general solution of $y'' + y = 2 \cos(2x) + e^{-x} - 7x^3$.

[9marks]

(5) Find a particular solution y_1 of the differential equation $\frac{y''-y}{y'-y} = x$ by guessing. Using y_1 , find a linearly independent solution y_2 and write down the general solution.

[6 marks]

INDIAN STATISTICAL INSTITUTE
Mid-Semestral Examination: 2012-13
B-Stat III
Anthropology

Date: 05.09.12
Hours

Maximum Marks:50

Duration: 2

Note: Answer Quesstion No. 1 and any four from the rest, answer should be brief and precise

1. What is Anthropology? What are the different branches of Anthropology? Compare and contrast relationship with other allied Sciences? [3+3+4]
2. What are the significances of Meiotic and Mitotic cell division? [6+4]
3. Illustrate different types of chromosome in man. Describe normal human karyotype and its importance in the study of human genetics. [3+5+2]
4. How chromosomal aberrations occur? Discuss autosomal aberrations with suitable examples. [4+6]
5. Illustrate Mendelian laws of inheritance with examples. [10]
6. Describe rare, autosomal, dominant inheritance in man using hypothetical pedigree. [5+5]
7. What is meant by organic evolution? What are the main theories of organic evolution? [3+7]
8. What are the basic sources of variation of human physical characteristics? What evolutionary factors explain these differences? [5+5]
9. Write short notes on **any 2** form the following: [5X2]
 - a. Allele
 - b. Polymorphism
 - c. Barr body
 - d. Inbreeding

Indian Statistical Institute
Mid Semestral Examination: (2011–2013)
B.Stat.(Hons.) – III year
Economics III

Date: 05/09/2012

Maximum Marks – 50

Duration: 2 hours

Answer *any two* questions.

1. (a) State the assumptions of the Classical Linear Regression Model (CLRM) in a multiple regression set up.
- (b) Show that the Ordinary Least Squares (OLS) estimator of the parameters is consistent and Best Linear Unbiased Estimator (BLUE).

(c) Consider the model

$$y_i = \beta x_i + \varepsilon_i$$

$$\varepsilon_i = u_i + \delta u_{i-1}, \quad |\delta| < 1$$

$$E(u_i) = 0$$

$$E(u_i u_j) = \begin{cases} \sigma_u^2, & i = j \\ 0, & i \neq j \end{cases}$$

Assuming δ and σ_u^2 are known, write down the expression for the GLS estimator of β .

- (d) Suppose you want to test r , $r \leq k$, independent linear restrictions of the form
- $$R\beta = d_{r \times 1}$$

where $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix}$ in $y_{n \times 1} = X\beta + \varepsilon_{n \times 1}$.

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

- (i) $\beta_1 = \beta_2 = \beta_3 = 0$
- (ii) $\beta_1 = \beta_2$ and $\beta_3 = \beta_4$
- (iii) $\beta_1 - 3\beta_2 = 5\beta_3$

[6+8+5+(2×3)=25]

2. (a) A researcher has data on the average annual rate of growth of employment, g , and the average annual rate of growth of GDP, x , both measured as percentages, for a sample of 27 developing countries and 23 developed ones for the period 1985–1995. He runs simple regressions of e on x for the whole sample, for the developed countries only, and for the developing countries only, with the following results:

<i>whole sample</i>	$\hat{g} = -0.56 + 0.24x$ (0.53) (0.16)	$R^2 = 0.04$ $RSS = 121.61$
<i>developed countries</i>	$\hat{g} = -2.74 + 0.50x$ (0.58) (0.15)	$R^2 = 0.35$ $RSS = 18.63$
<i>developing countries</i>	$\hat{g} = -0.85 + 0.78x$ (0.42) (0.15)	$R^2 = 0.51$ $RSS = 25.23$

Now he defines a dummy variable D that is equal to 1 for the developing countries and 0 for the others.

- (i) Explain the role of the dummy variable in estimating the coefficients of the equations for the two types of countries using a single equation.
 - (ii) What are the coefficients of this regression equation?
 - (ii) Compute an appropriate statistic for testing the researcher's hypothesis that the slope and intercept of the equations for the two types of countries are equal. Specify the degrees of freedom and the distribution it follows.
- (b) Describe the procedure of detecting multicollinearity using 'condition number' and 'variance proportions'.

[(4+5+6) + 10 = 25]

3. (a) Explain what is meant by 'autocorrelation'.
- (b) Describe a test for testing the presence of first-order autocorrelation (ρ) in a given time series. Derive the relationship between ρ and the test statistic.
- (c) Show that for a first order autoregressive model with positive coefficient, the autocorrelation function (ACF) declines geometrically.

(d) Consider the following two estimated models:

$$\hat{y}_t = 0.45 - .0041 X_t$$

(-3.96)

$$R^2 = 0.5248, D.W. = 0.8252$$

$$\hat{y}_t = 0.48 + 0.127 y_{t-1} - 0.32 X_t$$

(3.27) (-2.17)

$$R^2 = 0.8829, D.W. = 1.82$$

where figures in parentheses are t -ratios. Comment on the regression results. What are the appropriate estimates of the serial correlation in the two cases?

[2+9+6+8 = 25]

INDIAN STATISTICAL INSTITUTE

Mid – Semestral Examination: 2012-13

B. Stat III Year Geology Elective

Date: 5th September 2012. Maximum Marks: 30. Duration: 2:30 – 5:30 PM.

Note: Answer any five questions

1. What is Nebular disk hypothesis? What are the evidences supporting Nebular disk hypothesis?
(3+3)

2. How iron catastrophe helped in (a) differentiation and (b) formation of early atmosphere of our planet?
(4+2)

3. What is elastic rebound theory? Compare and contrast between (a) Love and Rayleigh wave and (b) body and surface waves?
(2+2+2)

4. Preliminary Reference Earth Model (PREM) (Dziewonski and Anderson, 1981) showing the density and velocity profiles of S – and P – waves through the Earth (see, Figure 1a). Draw the Core mantle boundary? Why S – wave is absent between ~3000 to ~ 5000 km? What can cause S – and P – wave velocities to increase with increase in density? What is S – and P – wave shadow zone? Why we get S – and P – wave shadow zone?
(0.5+0.5+2+1.5+1.5)

You may require: $V_p = \left(\left(\frac{4}{3} \mu + k \right) / \rho \right)^{0.5}$; $V_s = (\mu / \rho)^{0.5}$, where μ = modulus of rigidity; ρ =density; k = modulus of compressibility; and Figure 1b.

5. What is the difference between earthquake epicenter and focus? Recent earthquake epicenters are plotted in Figure 2. Why they show linear trend? What can cause earthquake clusters in Africa (see the boxed area in Figure 2)?
(2+2+2)

6. Emperor-Hawaiian Island seamount chain on the Pacific Ocean is shown in Figure 3. What is hotspot? What kind of igneous rock(s) you expect to find on the volcanic chains? What will be

the expected igneous rock texture(s) and color? Explain. What other information can you extract out from the given figure (hint: velocity, direction, why the kink at ~ 40Ma etc)?

(1+0.5+1+1.5+2)

7. Describe the processes that cause rocks to melt? (6)

8. Consider the compositional characteristics of Mt. St. Helen's rocks given in Figure 4. What can you infer from it? (6)

9. An example of paired metamorphic belt is shown in the Figure 5. The Sierra Nevada Belt shows high pressure and temperature metamorphism while Franciscan belt shows high pressure but low temperature metamorphism. What kind of plate tectonic setting you expect to get this kind of metamorphic activity? Explain using a cross sectional diagram? (2+4)

10. What parameters are used to determine (a) textural and (b) compositional maturity of sedimentary rocks? Why quartz is the most stable mineral i.e. resistant to weathering? What information can you extract from the sedimentary rocks shown in Figure 6? (2+2+2)

11. How can you use the size and sorting of sediments to distinguish between sediments deposited in a glacial environment and those deposited on a desert? (3+3)

Figure 1a

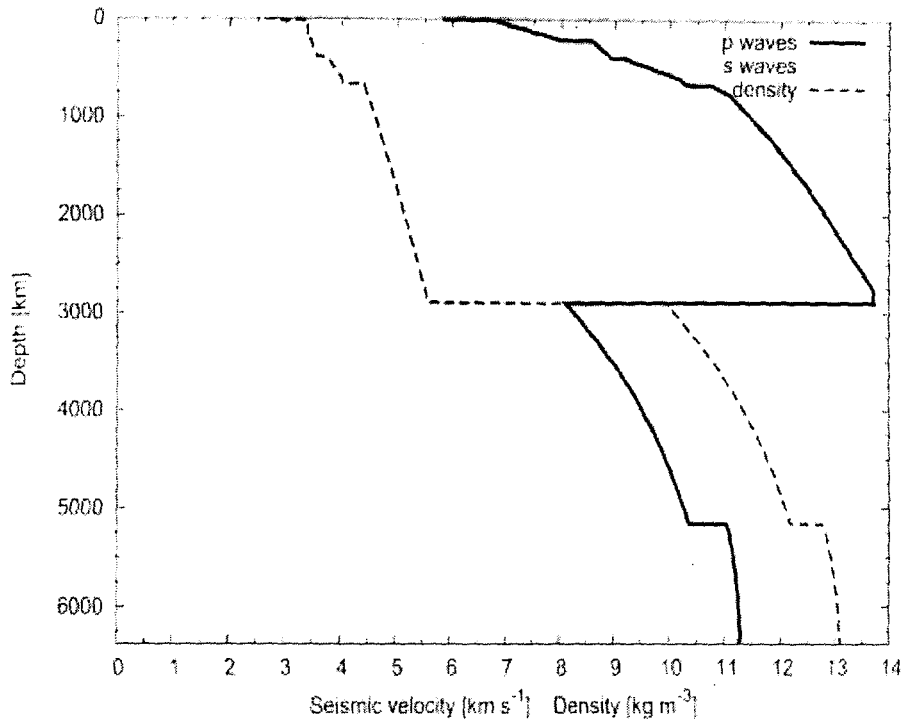


Figure 1b

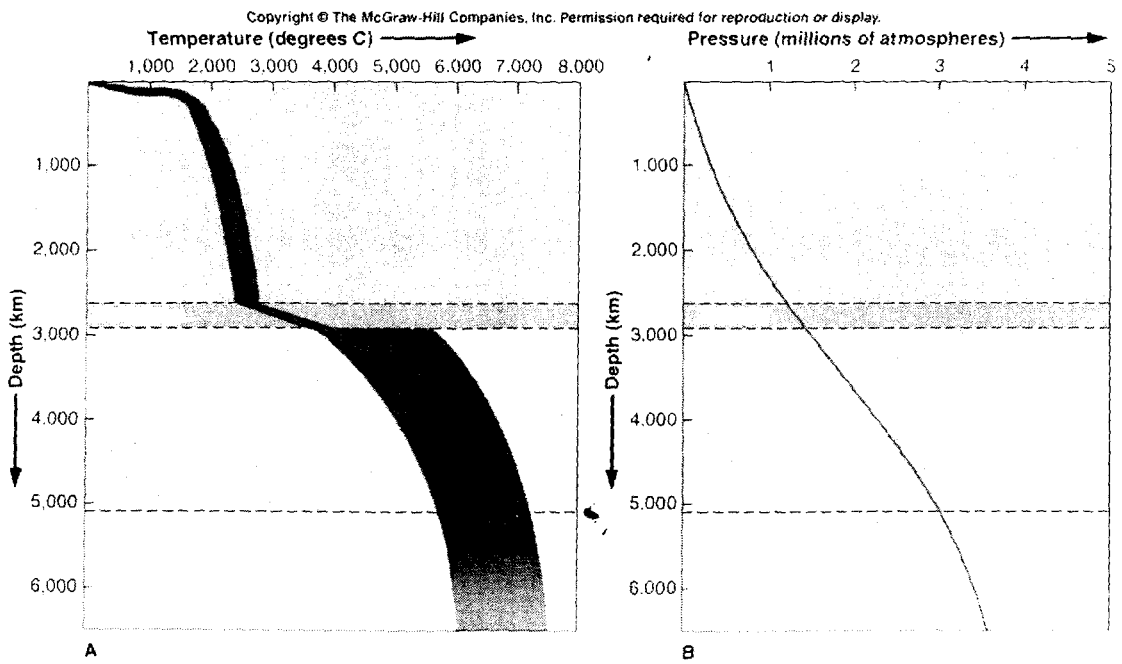


Figure 2

World Seismicity 1977-1992

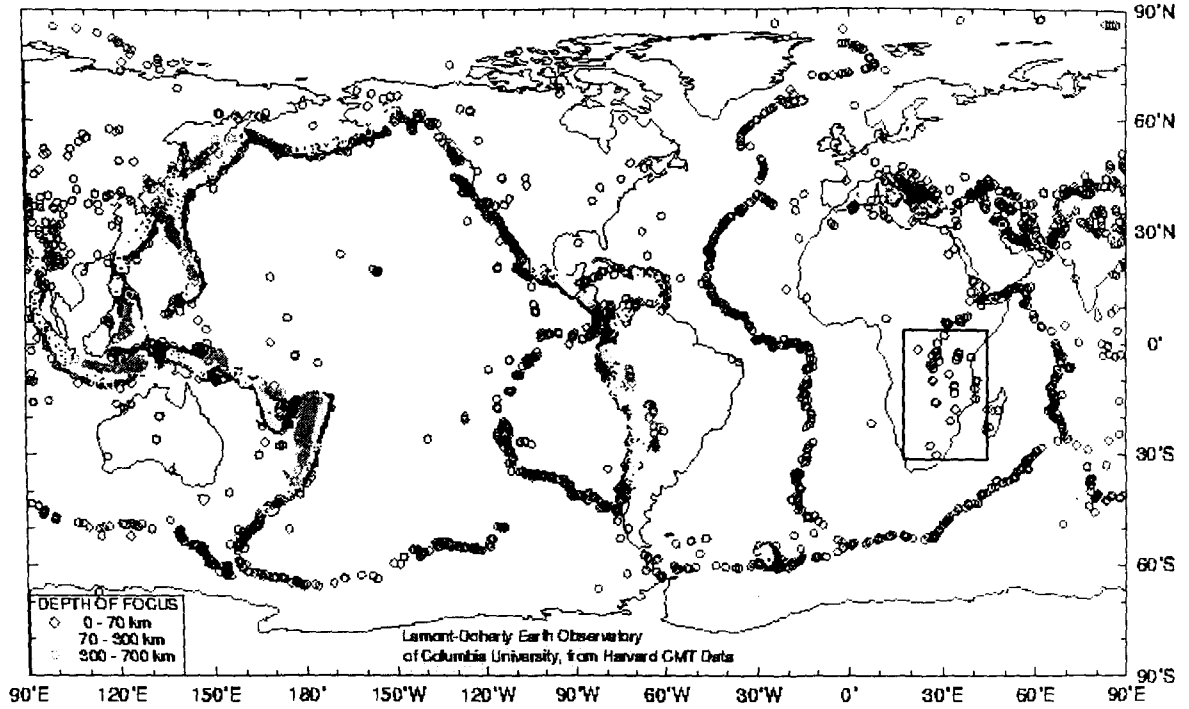


Figure 3

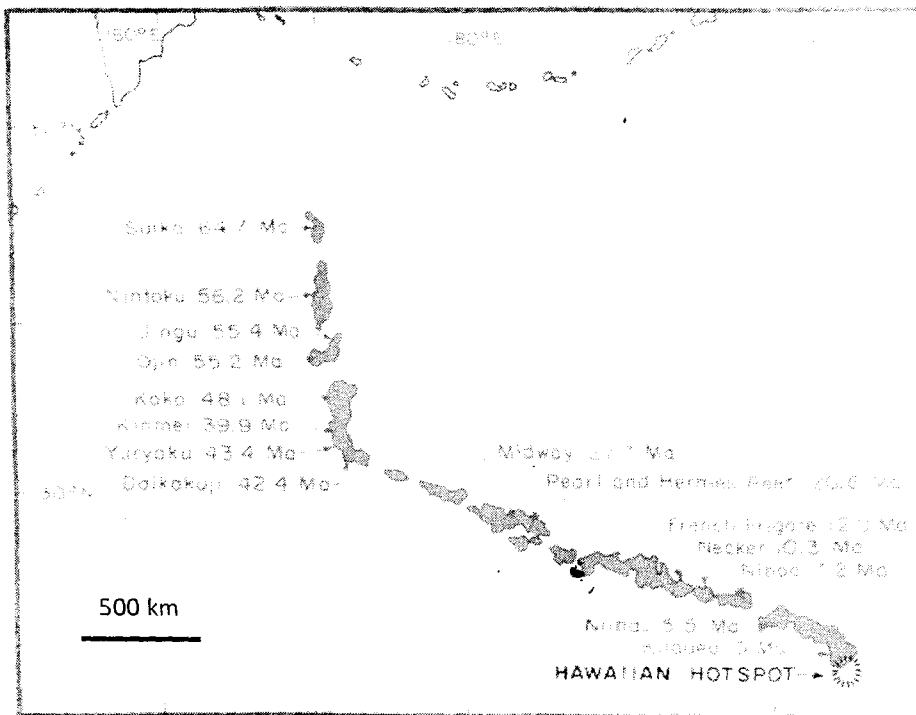


Figure 4

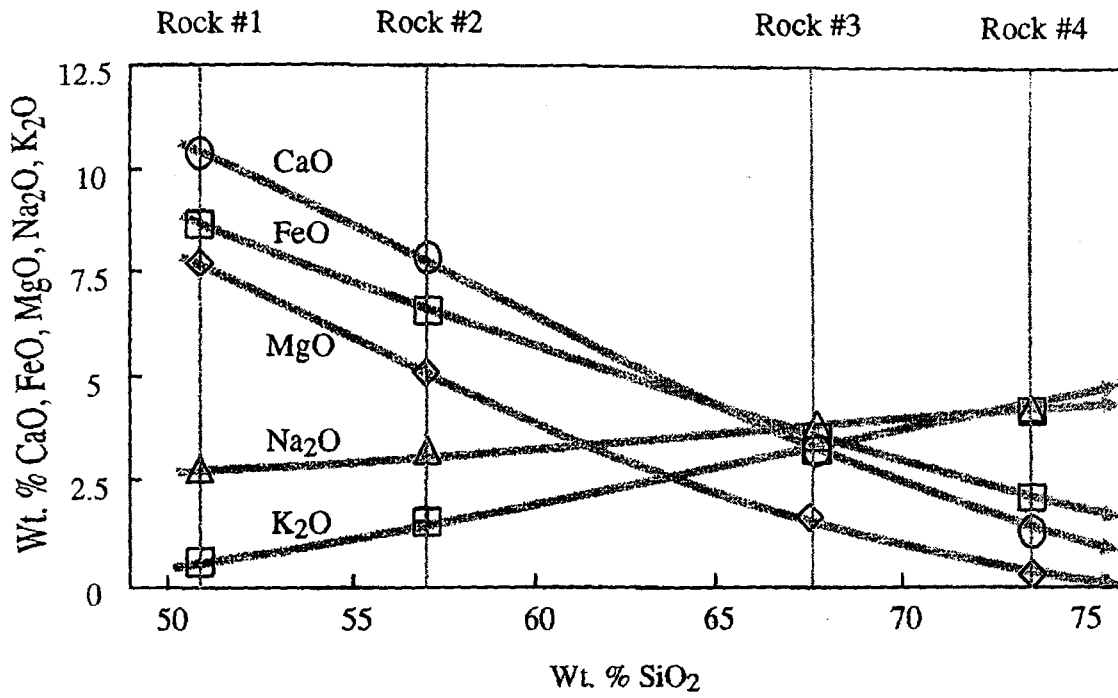


Figure 5

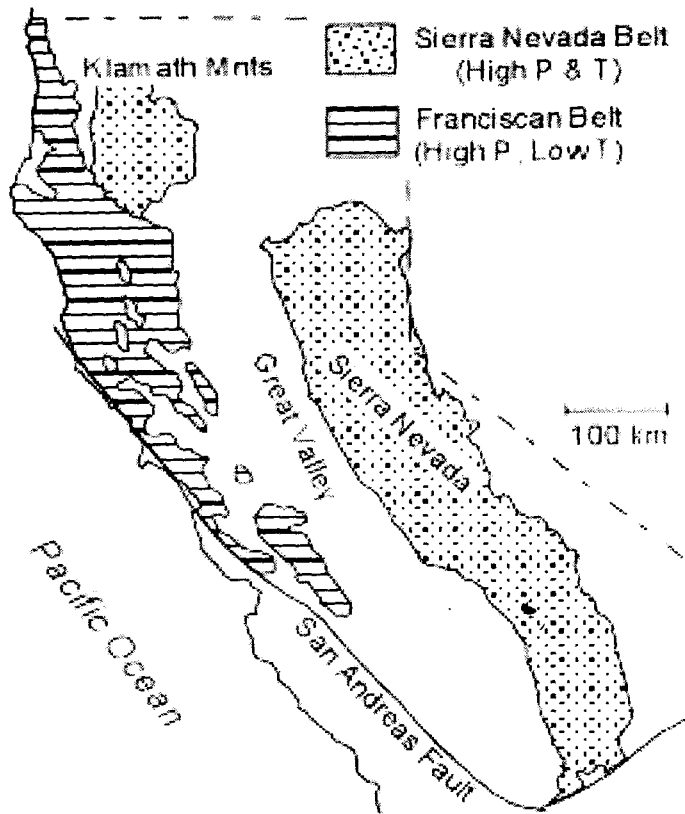
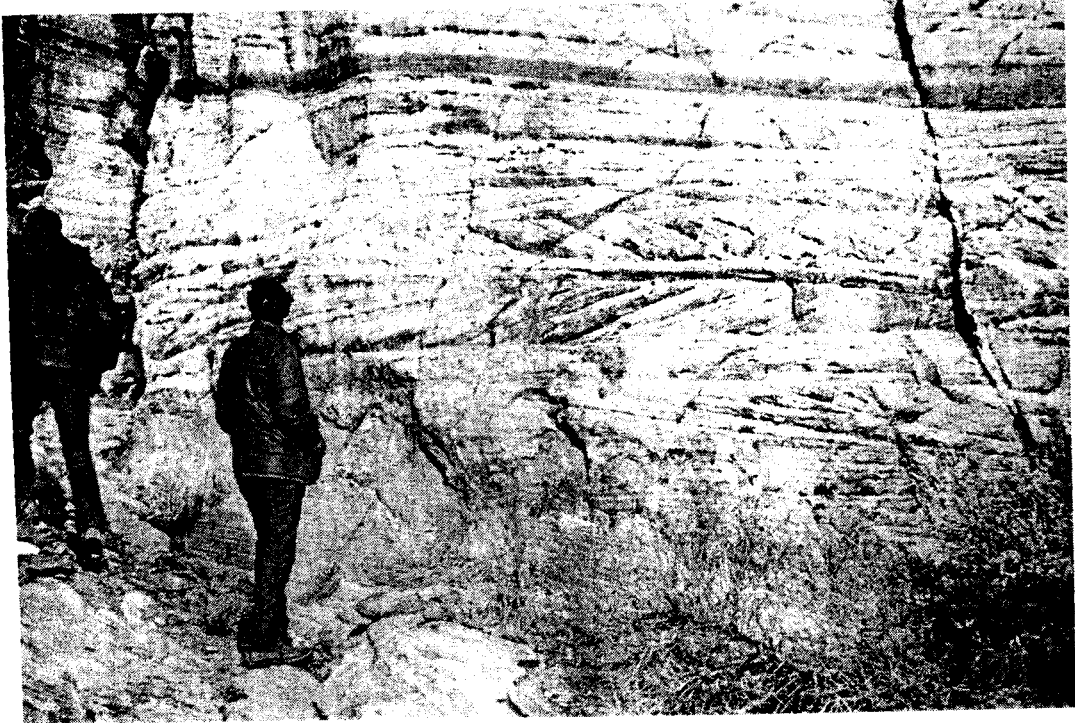


Figure 6



INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2012 - 2013

B.Stat. (Hons.) III Year

Sample Surveys

Date : ~~07~~ 09.2012

Maximum Marks : 100

Duration : 3 Hours

Answer ANY FOUR questions . Marks allotted to each question are given within the parentheses . Standard notations and symbols are used .

1. A simple random sample of size $n = n_1 + n_2$ with mean \bar{y} is drawn from a finite population of size N and a simple random subsample of size n_1 is drawn from it with mean \bar{y}_1 .

Show that (a) $Cov(\bar{y}_1, \bar{y}_2) = -\frac{1}{N} S^2$ where \bar{y}_2 is the mean of the remaining n_2 units in the sample and S^2 is the population variance with divisor $(N-1)$

$$(b) Var(\bar{y}_1 - \bar{y}_2) = S^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]$$

$$(c) Var(\bar{y}_1 - \bar{y}) = S^2 \left[\frac{1}{n_1} - \frac{1}{n} \right]$$

$$(d) Cov(\bar{y}, \bar{y}_1 - \bar{y}) = 0.$$

Repeated sampling implies repetition of the drawing of both the sample and the subsample . (6+6+6+7)=[25]

2. From a population of size 3 an SRSWOR sample of size 2 is drawn . Consider the following estimator

$$\widehat{Y}_{12} = 3 \left(\frac{1}{2} y_1 + \frac{1}{2} y_2 \right), \widehat{Y}_{13} = 3 \left(\frac{1}{2} y_1 + \frac{2}{3} y_3 \right), \widehat{Y}_{23} = 3 \left(\frac{1}{2} y_2 + \frac{1}{3} y_3 \right)$$

where \widehat{Y}_{ij} is an estimator of the population total Y based on the sample that has units (i, j) .

- (a) Prove that \widehat{Y}_{ij} is an unbiased estimator of the population total Y .
(b) Obtain the sampling variance of \widehat{Y}_{ij} .
(c) Hence or otherwise show that $Var(\widehat{Y}_{ij}) < Var(3\bar{y})$ if $y_3(3y_2 - 3y_1 - y_3) > 0$ where \bar{y} is the sample mean .

P.T.O.

(5+10+10)=[25]

3. Suppose a population of size N is divided into L strata and the h th stratum consists of N_h units, $h = 1, 2, \dots, L$. From the h th stratum an SRSWOR sample of size n_h is drawn, $h = 1, 2, \dots, L$.
- Obtain an unbiased estimator of the overall population mean based on the stratified random sample and also obtain an expression for the sampling variance of the estimator.
 - Obtain an unbiased estimator of the sampling variance in (a) above based on the sample data.
 - Obtain an unbiased estimator of the overall population variance based on the stratified random sample.

$$(3+5+5+12)=[25]$$

4. Describe Lahiri's method of drawing a PPS sample. Justify that a sample drawn according to Lahiri's method is really a PPS sample.

$$(10+15)=[25]$$

5. In a sample of 50 households drawn with SRSWOR from a village consisting of 250 households only 8 households were found to possess a bicycle. These had 3, 5, 3, 4, 7, 4, 4 and 5 members respectively. Estimate unbiasedly the total number of households in the village possessing a bicycle as well as the total number of persons in such households. Also estimate the RSE's of these estimates by using the unbiased estimates of their variances.

$$(5+5+7+8)=[25]$$

6. Draw a linear systematic sample of size 2 from the following population of size 5.

Unit No. :	1	2	3	4	5
y :	20	25	35	45	65

- Estimate unbiasedly the population mean and obtain the sampling variance of your estimate.
- How will you modify your method of sampling so that the sample mean provides an unbiased estimator of the population mean? Also obtain the sampling variance of the estimated mean in this case. How does it compare with the sampling variance in (a) above?
- Draw a circular systematic sample of size 2 from the above population and obtain the sampling variance of the estimated mean. How does it compare with the sampling variance in (a) and (b) above?

$$(8 + 8 + 9) = [25]$$

INDIAN STATISTICAL INSTITUTE
 Mid-Semestral Examination : 2011 – 12
 B. Stat (3rd Year)
 Linear Statistical Models

Date: 10 September 2012

Maximum Marks: 30

Duration: 1½Hours

1. Show that adding a new explanatory variable (weakly) increases R^2 . [6]

2. Show that the weighted least square estimator $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$ for the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \text{ with } Var(y_i) = \sigma^2 x_i$$

has the form

$$\frac{1}{\sum_i x_i \sum_i \frac{1}{x_i} - n^2} \begin{pmatrix} (\sum_i x_i) \left(\sum_i \frac{y_i}{x_i} \right) - n(\sum_i y_i) \\ (\sum_i y_i) \left(\sum_i \frac{1}{x_i} \right) - n \left(\sum_i \frac{y_i}{x_i} \right) \end{pmatrix}$$

Also show that $Cov(\hat{\beta}) = \frac{\sigma^2}{\sum_i x_i \sum_i \frac{1}{x_i} - n^2} \begin{pmatrix} (\sum_i x_i) & -n \\ -n & \left(\sum_i \frac{1}{x_i} \right) \end{pmatrix}$. [6 + 4 = 10]

3. Obtain the Confidence Interval for $\alpha'\beta$ from the corresponding t-statistic. [6]

4. Consider the model $y_{ij} = \mu + \tau_i + \epsilon_{ij}, i = 1,2,3; j = 1,2,3$:

Write $X, X'X, X'y$ and the normal equations.

What is the rank of X or $X'X$?

Find a set of linearly independent estimable functions. [6 + 2 + 5 = 13]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination 2012 – 13

B.Stat 3rd Year

Linear Statistical Model

Time: 3 Hours

Full Marks: 100

TE - 19.11.12

This paper carries 109 marks. Attempt ALL questions.

The maximum you can score is 100.

1. Consider an agricultural experiment with three alternative HYV seeds {A, B, C} and three types of pesticide {a, b, c} being applied in 18 small plots of land according to the following layout (denote this by D1):

A, a	B, b	C, c	A, a	B, b	C, c
A, b	B, c	C, a	A, b	B, c	C, a
A, c	B, a	C, b	A, c	B, a	C, b

The dependent variable is yield rate (in Kg/sq. mtr), denoted by y .

- Write down a two-way model with interaction representing this design.
- Write down the matrices X , $X'X$, $X'y$ and find rank of X , $r(X)$.
- Give a set of estimable functions in this set up.
- Obtain estimates of all the parameters of the model defined in part (a), including that for σ^2 . Write down the ANOVA table for this model.

$$[4 + (4 + 5 + 3 + 2) + 7 + (7 + 5) = 37]$$

2. In the same experiment as in Q1 above, consider a modified layout:

A, a	B, b	C, c	A, a	B, b	C, c
A, b	B, c	C, a	A, b	B, c	B, c
A, c	B, a	C, b	A, b	B, c	C, b

- What is the modified model for this design?
- Write down the matrices X , $X'X$, $X'y$ and find $r(X)$.
- Give a set of estimable functions in this set up.

- d) Obtain estimates of all the parameters of the model defined in part (a), including that for σ^2 . Write down the ANOVA table for this model.

$$[3 + (3 + 3 + 2 + 2) + 6 + (5 + 4) = 28]$$

3. In the same experiment as in Q1 and Q2 above, Consider a further modified layout in 17 plots of land:

A, a	B, b	C, c	A, a	B, b	C, c
A, b	B, c	C, a	A, b	B, c	B, c
	B, a	C, b	A, b	B, c	C, b

- a) What is the modified model for this design?
 b) Write down the matrices X , $X'X$, $X'y$ and find rank of X , $r(X)$.
 c) Give a set of estimable functions in this set up.
 d) Obtain estimates of all the parameters of the model defined in part (a), including that for σ^2 . Write down the ANOVA table for this model.

$$[3 + (3 + 3 + 2 + 2) + 6 + (5 + 4) = 28]$$

4. In the same experiment as in Q1 above, consider the layout D1 as in Q1. We introduce an additional explanatory factor, amount of water used in the plots denoted by w (in liters/sq. mtr.)

- a) What is the modified model for this design?
 b) Obtain estimates of all the parameters of the model defined in part (a), including that for σ^2 . Write down the ANOVA table for this model.

$$[4 + (6 + 6) = 16]$$

INDIAN STATISTICAL INSTITUTE
First – Semester Examination: 2012-13

B. Stat III Year Geology Elective

Maximum Marks: 50. Duration: Two Hours.

Read carefully: Answer *any one* question from question numbers 1 and 2. Answer *any one* question from question numbers 3 and 4. Question numbers 5 and 6 are *compulsory*.

Question 1. (Total marks 10)

- a) What is a mineral? (2)
- b) Give one example each of a silicate mineral, sulfide mineral, oxide mineral and carbonate mineral? (2)
- c) Do you think that water and coal are minerals? Justify your answer. (2 + 2 = 4)
- d) Write the formula of silicon-oxygen tetrahedron. (1)
- e) Draw silicon-oxygen tetrahedron. (1)

Question 2. (Total marks 10)

- a) Draw clearly the arrangements of the silicon-oxygen tetrahedrons in single chain, double chain and sheet silicate structures. (3 x 1 = 3)
- b) What will be the Si:O ratio for single chain, double chain and sheet silicate structures? (3 x 1 = 3)
- c) Match the silicate structures with the representative mineral groups given in the table below. (4 x 1 = 4)

Silicate Structures	Mineral groups
a. Single chain	I. Pyroxene
b. Double chain	II. Mica
c. Sheet silicate	III. Clay minerals
d. Framework silicate	IV. Quartz
	V. Feldspar
	VI. Amphibole

Question 3. (Total marks 10)

- a) What is weathering? (2)
- b) What is differential weathering? (1)
- c) Differentiate between mechanical and chemical weathering. (2)
- d) How does mechanical weathering help in chemical weathering? (1)
- e) How do sinkholes form? Write the relevant balanced chemical equation. (1 + 1 = 2)
- f) How does a placer deposit form? Give an example of a placer deposit. (1 + 1 = 2)

Question 4.

(Total marks 10)

- a) Define porosity and permeability. What is the unit of porosity? (2 + 1 = 3)
- b) Where do you think clay and gravel will locate in the porosity – permeability plot (draw in Figure 1 in your answer sheet)? (1)
- c) Write the Darcy equation. Explain each term of the equation (give units). (1 + 2 = 3)
- d) Consider Figure 2 that shows the geographic position, distance, and the total head at each well. Find out hydraulic gradient and groundwater flow direction from data given in Figure 2. (2 + 1 = 3)

Question 5.

(Total marks 4 x 2.5 = 10)

Briefly compare any four of the following pairs

- a) Agnatha and Ganthostomata.
- b) Polyphyletic and Monophyletic Groups.
- c) Lobe fin and Ray fin fishes.
- d) Amniota and non-Amniota.
- e) Morphological Diversity and Morphological Disparity.
- f) Diapsid skulls and Synapsid skulls.
- g) Pelvic girdle structures of Saurischian and Ornithischian dinosaurs.

Question 6. Answer all the questions:

(Total marks 20)

- a) Would you consider an Egyptian mummy a fossil? Justify your answer. (2)
- b) What are the different states of preservations of fossils? (5)
- c) What is molting? (2)
- d) If a rock initially contained 10 milligrams of a radioactive parent element when it first crystallized, how much of it remains after 4 half-lives? (2)
- e) How do evaporite minerals form? Name the evaporite mineral that helps in trapping petroleum. Name the evaporite mineral that is used in fertilizer production in Peru and Chile. (2+1+1=4)
- f) In a sedimentary succession you are getting ammonite fossils. What kind of depositional environment do you infer? (1)
- g) In a sedimentary succession you are getting thick coal seams. What kind of depositional environment do you infer? (1)
- h) In which kind of environment are aeolian processes important. Give an example. (1)
- i) Why do fossils mostly indicate relative ages of their host rocks? (2)

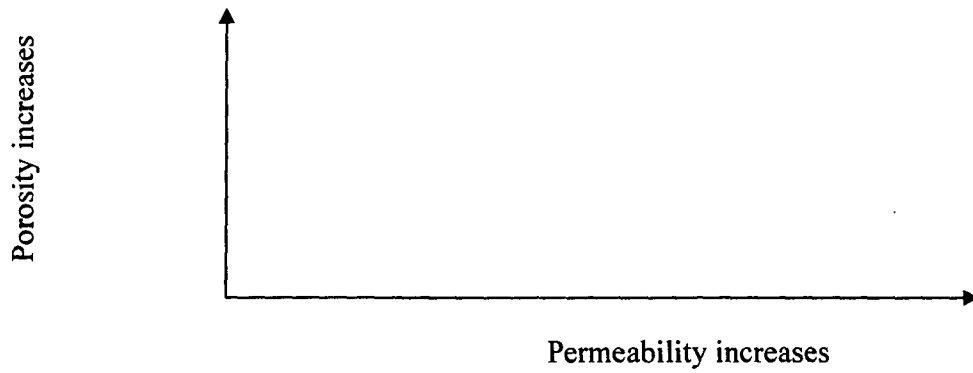


Figure 1

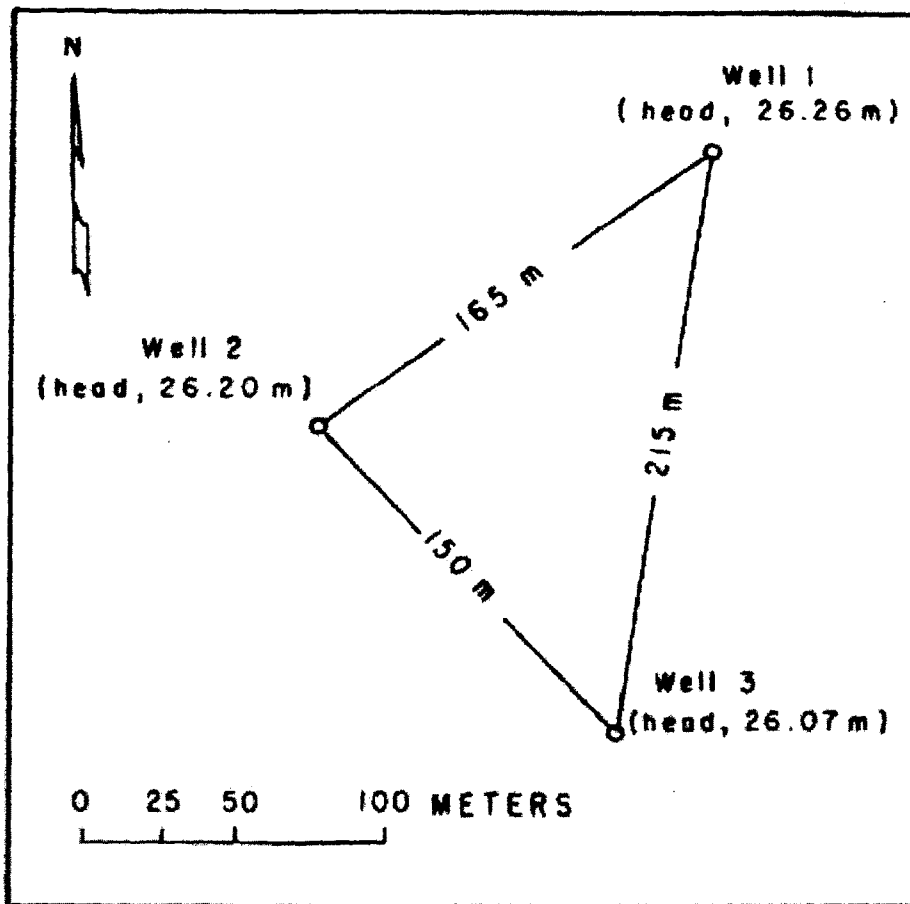


Figure 2

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2012-13

B-Stat III

Introduction to Anthropology and Human Genetics

Date: 21/11/12

Maximum Marks: 40

Duration: 2 Hours

Answer question no. 1 and any 3 (three) from the rest

1. Choose (by ticking) the right answer: [1X10]
 - (a) Humans belong to order Primates: True/False
 - (b) Linea aspera is present in human femur: True/False
 - (c) Humans are not a culture-bearing animal: True/False
 - (d) Humans have 23 pairs of autosomes: True/False
 - (e) Modern humans have around 30,000 genes: True/False
 - (f) Australopithecines emerged in the earliest stage of hominid evolution:
True/ False
 - (g) New world monkeys belong to order Prosimi: True/False
 - (h) Australopithecus have been discovered only from Africa: True/False
 - (i) First primate appeared 80 million years ago: True/False
 - (j) Anatomically modern humans are scientifically called *Homo sapiens*:
True/False
2. Which features distinguish humans from other members of order Primates?
[10]
3. (a) Draw a diagram showing evolutionary sequence of the order Primates in relation to time (in millions of years). [5]
(b) Describe the important changes that have taken place in the physical structure of humans due to attainment of erect posture and bipedal locomotion. [5]

4. After graduation, you and your 19 friends build a raft, sail to a deserted island, and start a new population, totally isolated from the world. Two of your friends carry (that is, are heterozygous for) the recessive *cf* allele, which in homozygotes causes cystic fibrosis. Assuming that the frequency of this allele does not change as the population grows, what will be the incidence of cystic fibrosis on your island? [10]

5. Illustrate with suitable examples how interaction between nature and nurture determines biological characteristics in humans. [10]

6. Write short note on any 5 (five) of the following: [2X 5]

(a) Adaptability

(b) Senescence

(c) Physical growth

(d) Lamarckism

(e) Demography

(f) Culture

(g) Mitochondria

(h) Homeostasis

(i) Health

(j) Morbidity

Indian Statistical Institute
First Semestral Examination: (2012–2013)
B.Stat.(Hons.) – III year
Economics III

Date: 21.11.12

Maximum Marks 100

Duration: 3 hours

Answer any *three* questions. The total is **105** marks. The maximum you can score is **100**. Marks allotted to each question are given within parentheses at the end of the question.

1. (a) (i) Consider the following model:

$$C_t = \nu_0 + \nu_1 Y_t + \varepsilon_{1t} \quad (\text{Consumption function})$$

$$I_t = \delta_0 + \delta_1 Y_{t-1} + \varepsilon_{2t} \quad (\text{Investment function})$$

$$Y_t = C_t + I_t. \quad (\text{Income identity})$$

Reduce the three-equation model to a single equation and examine if any assumption of the CLRM is violated. Give reasons for your answer.

- (ii) Show that the OLS estimator of the parameters is biased but consistent.

- (b) In the multiple regression model $y = X\beta + \varepsilon$, if ε 's are correlated with the regressors, what is the appropriate method of estimation? Briefly describe the method.

- (c) 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 error and we measure

$$x_1 = X_1 + u_1, \quad x_2 = X_2 + u_2 \quad \text{and} \quad y = Y + v.$$

Assume that u_1, u_2 and v are mutually uncorrelated and also uncorrelated with

X_1, X_2 and Y . Is the OLS estimator consistent? Justify your answer.

[(8 + 6) + 9 + 12 = 35]

2. (a) Explain what is meant by heteroscedasticity. Also discuss the consequences of its presence for the estimation of a regression model.

- (b) Describe the Goldfeld–Quandt test for detecting heteroscedasticity and explain why it may detect heteroscedasticity only under certain conditions.

(c) What is multicollinearity? Explain.

(d) Define 'condition number' and explain how it is used in detecting multicollinearity.

(e) Describe some remedial measures to deal with the multicollinearity problem.

[5 + 6 + 2 + 12 + 10 = 35]

3. (a) What are distributed lag models? Define "impact multiplier" and "equilibrium multiplier".

(b) Describe the geometric lag model and rationalize the model in terms of (i) Adaptive Expectation hypothesis and (ii) Partial Adjustment hypothesis.

(c) What is "Koyck transformation"? Explain its relevance in the context of estimation of a Partial Adjustment Model.

(d) Describe the Almon polynomial lag model.

[7 + (9+9) + 5 + 5 = 35]

4. (a) Describe the general structural form of a simultaneous equations model (SEM) explaining all the terms that you use. Obtain the reduced form of the equation system.

(b) Discuss the identifiability status of each of the equations in the following SEM.

$$y_{1t} = \beta_2 y_{2t} + \gamma_1 x_{1t} + \gamma_2 x_{2t} + \gamma_3 x_{3t} + \varepsilon_{1t}$$
$$y_{2t} = \alpha_1 y_{1t} + \delta_1 x_{4t} + \delta_2 x_{5t} + \varepsilon_{2t}$$

(c) Explain why ordinary least squares (OLS) would yield an inconsistent estimator of the parameters of the first equation in (b) above.

(d) Describe briefly the 2SLS and 3SLS methods of estimating a simultaneous equations system, stating the appropriateness of each of these methods in terms of the identifiability status.

[10 + 10 + 5 + 10 = 35]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : 2012 - 2013

B.Stat.(Hons.) III Year

Sample Surveys

Date : 23.11.2012

Maximum Marks :100

Duration : 3 Hours

Answer ANY FOUR questions . Marks allotted to each question are given within the parentheses . Standard notations and symbols are used .

1.(a) The sample size required to estimate the proportion of workers in a population with relative standard error $\alpha\%$ is n in a simple random with replacement sampling . Determine the sample size required to estimate the population proportion of non-workers with the same precision .

(b) From a simple random without replacement sample of n units a random sub-sample of m units is duplicated and added to the original sample . Show that the mean based on $(n+m)$ units is an unbiased estimator of the population mean and its variance is greater than or equal to the variance of the mean based on the original sample of n units.

(10 + 15) = [25]

2. (a) Find the bias of the estimator $\hat{Y}_R = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i} \cdot X$ of the population total Y of the study variable y under simple random without replacement sampling of n units and an unbiased estimator of the bias where X is the population total of the auxiliary variable x . Hence find an unbiased estimator of Y utilizing the information on the auxiliary variable.

(b) Explain why it is not generally possible to estimate unbiasedly the sampling variance of the estimated mean based on a single systematic sample . What do you mean by interpenetrating network of sub-samples ? Explain how this technique can be utilized in estimating unbiasedly the sampling variance of the estimated mean in case of circular systematic sampling .

(13 + 12) = [25]

3. Derive approximate expressions of the mean square errors of the ratio and the regression estimator of the population mean under simple random without replacement sampling . Compare the precisions of the two estimators .

(10+10+5)=[25]

P.T.O.

4. Suppose a population consists of N first stage units and the i th first stage unit consists of M_i second stage units, $i = 1, 2, \dots, N$. Suppose a sample of n first stage units is drawn from the population by simple random without replacement sampling and if the i th first stage unit is selected, a sample of m_i second stage units is selected again by simple random without replacement sampling scheme. Obtain an unbiased estimator of the population total on the basis of the sample drawn and derive an expression for its sampling variance. Also obtain an unbiased estimator of the sampling variance of the unbiased estimator of the population total.

(5+10+10)=[25]

5.(a) Describe how you would unbiasedly estimate the population proportion of individuals possessing a certain attribute based on a stratified simple random without replacement sample. Derive an expression for the sampling variance of the suggested estimator.

(b) Derive Neyman's optimum allocation rule under the set-up in (a) and also an expression for the variance of the estimated proportion under Neyman's optimum allocation.

(5 + 8 + 7 + 5)=[25]

6. In a demographic survey , it is proposed to use stratified with replacement sampling taking the districts in a region as strata . The relevant data are given in the following table .

District Sl. No.	No. of villages (N_h)	Average population per village (\bar{Y}_h)	Standard deviation (σ_h)
1.	1953	487	564
2.	1664	829	931
3.	1381	822	996
4.	1174	1083	1167
5.	531	1956	1940
6.	1391	664	625
7.	1996	456	779
8.	1951	372	556
9.	3369	339	591

- (a) Assuming that the cost of enumeration and tabulation per person is $\frac{1}{4}$ th of a rupee and the overhead cost is Rs.10,000 , determine the optimum values of sample sizes n_h 's for the strata that would minimize the sampling variance of the estimator of the overall population mean for a given expected total cost of Rs.80,000 .
- (b) For the same value of the total sample size n obtained in (a) find the values of n_h 's when the allocation is made in proportion to $N_h\sigma_h$ and obtain the cost -efficiency of the procedure as compared to that of (a) .

(10 + 15) =[25]

INDIAN STATISTICAL INSTITUTE
 First Semestral Examination: (2012-2013)
 MS (Q.E.) II Year
Macroeconomics II

Date: 30.11.12

Maximum Marks 60

Duration 3 hours

Group A

Answer any two

1. Robinson Crusoe lives forever in his island economy which experiences productivity shock, $\ln z_t$ (i.i.d. with zero mean). In period t he gives labour l_t (which yields disutility), produces output y_t , consumes c_t and maximizes expected present value of lifetime utility:

$$E_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t, l_t) \right] = E_0 \left[\sum_{t=0}^{\infty} \beta^t \{ \ln c_t - l_t^\sigma \} \right] \quad (\sigma > 1; 0 < \beta < 1), \quad \text{subject to}$$

(i) $c_t + k_{t+1} = y_t = z_t k_t^\alpha l_t^{1-\alpha}$, ($0 < \alpha < 1$ and capital, k , depreciates fully after one period).

Let $v(k_t, z_t)$ denote the value function which satisfies the following *Bellman* equation

$$(ii) \quad v(k_t, z_t) = \max_{c_t, k_{t+1}, l_t} \{ \ln c_t - l_t^\sigma + \beta E_t \{ v(k_{t+1}, z_{t+1}) \} \}$$

An educated guess about the solution of (ii) is that $v(\cdot)$ takes the following form:

(iii) $v(k_t, z_t) = \theta_0 + \theta_1 \ln k_t + \theta_2 \ln z_t$ { $\theta_0, \theta_1, \theta_2$ are constants to be determined.}

(a) Using (i) and (iii), maximize (ii) with respect k_{t+1} (or c_t) and l_t .

Find θ_1 and θ_2 as well as optimal values of k_{t+1} and l_t in terms of α, β, σ and y_t .

Does the optimal value of k_{t+1}/y_t or that of l_t depend on k_t and z_t ? Explain.

(b) Suppose there was no shock up to period 0 and the economy was in equilibrium

at an output level y^* . Suppose further that at period zero there is a *one-time* productivity shock, i.e. $\ln z_0 = \varepsilon > 0$ and $\ln z_t = 0$ for all $t \neq 0$.

Find $\ln y^*$ and trace the time path of y_t for all $t \geq 0$.

[9 + 6] = [15]

2. How is a real business cycle model tested empirically? Describe briefly the methodology used.

[15]

3. Describe in detail a model of CAPM with the absolute minimum of assumptions, namely that there be no opportunities for pure arbitrage. What is the expression of risk premium of an asset? Derive it and explain.

[15]

Group B

Answer all

1. In the Blanchard-Yaari model with cohort dependent wage, what would be the effect of a sharper decline in wage with respect to age on the steady state capital accumulation? Explain.

Consider the same model, but now with zero population growth, cohort independent wages and open to international asset market with a constant rate of interest. Can you show that the aggregate savings in this model is negatively related to the level of assets? [Savings = Total income – Total Consumption, where Total income = wage income + interest income on assets. Also assume that the steady state exists in the model.]

(10+5)

2. Show that with investments having convex adjustment costs, the capital stock exhibits smooth transitional dynamics even when the country is small in the international capital market facing a constant rate of interest.

Find out the conditions on the production function and on the investment adjustment cost function under which Tobin's marginal q would be equal to the average q .

(10+5)

**INDIAN STATISTICAL INSTITUTE
FIRST SEMESTER BACKPAPER EXAMINATION 2012-2013**

Course Name : BSTAT III

Subject Name : Statistical Inference I

Date : 26 12 12 Maximum Marks : 50 Duration : 3 hours

Note : There are 5 problems carrying 10 points each.

1) Suppose X_{ij} , $i = 1, \dots, k$, $j = 1, \dots, n_i$, are independent $N(\mu_i, \sigma^2)$ observations. Let $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ and $\bar{X}_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}/n$ where $n = \sum_{i=1}^k n_i$.

a) Show that $\sum_{i=1}^k n_i(\bar{X}_i - \bar{X}_{..})^2$ and $\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$ are independent.

b) Derive their distributions.

2) Let X follow negative binomial(p, m), i.e.,

$$P(X = x|p, m) = \binom{m+x-1}{x} p^m (1-p)^x, \quad x = 0, 1, \dots; \quad p \in (0, 1); \quad m = 1, 2, \dots$$

Assume m to be fixed.

a) Show that p^{-k} has a UMVUE for any positive integer k .

b) Find the information $I(p)$ based on X .

3) Let X_1, \dots, X_n be iid $N(\mu, \sigma^2)$. Set $\tau = \frac{1}{2\sigma^2}$. Let $\theta = (\mu, \tau)$. Set a prior on θ as

$$\pi(\tau, \mu) d\tau d\mu = \tau^{\alpha-1} \exp(-\lambda\tau) d\tau d\mu$$

for some $\alpha, \lambda > 0$. Find the Bayes estimators of σ w.r.t. the loss functions

a) $(d - \sigma)^2$ and

b) $\frac{(d - \sigma)^2}{\sigma^2}$.

4) Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be iid $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. To test $\mu_1 = \mu_2$ vs $\mu_1 \neq \mu_2$, we reject for large values of $|S|$ where $S = \frac{\sqrt{n}(\bar{X} - \bar{Y})}{\sqrt{S_1^2 + S_2^2 - 2S_{12}}}$. Here $S_1^2 = \sum (X_i - \bar{X})^2$, $S_2^2 = \sum (Y_i - \bar{Y})^2$ and $S_{12} = \sum (X_i - \bar{X})(Y_i - \bar{Y})$.

a) Show that the critical region for this level α test is $|S| > \frac{1}{\sqrt{n-1}} c$ where $c = F^{-1}(1 - \alpha/2)$ and F is the c.d.f. of the T distribution with $(n - 1)$ degrees of freedom.

b) Show that the power function is increasing in $\frac{|\mu_1 - \mu_2|}{\sigma}$ where $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$.

5) Consider the Gaussian linear model $Y = X\beta + \epsilon$, $Y, \epsilon \in \mathbb{R}^n$, X is a $n \times p$ matrix of rank p ($p < n$) and ϵ_i are iid $N(0, \sigma^2)$, $i = 1, \dots, n$. We want to test $\sigma = \sigma_0$ vs $\sigma \neq \sigma_0$. Show that the acceptance region for the LRT test is given by

$$c_1 \leq \sigma_0^{-2} \|Y - X\hat{\beta}\|^2 \leq c_2$$

where $F(c_1) - F(c_2) = 1 - \alpha$ and $c_1 - c_2 = n \log(c_1/c_2)$. Here F denotes the cdf of χ_{n-p}^2 and $\hat{\beta}$ is the MLE of β .

INDIAN STATISTICAL INSTITUTE, KOLKATA
BACK PAPER EXAMINATION: FIRST SEMESTER 2012 -'13
B.STAT III YEAR

Subject : **Differential Equations**
Time : 2 hours 30 minutes
Maximum score : 70

Instructions:

- *Justify every step in order to get full credit of your answers, stating clearly the result(s) that you use. Points will be deducted for missing arguments. Partial credit will be given for your approach to the problem.*
- *Switch off and deposit your mobile phones to the invigilator during the entire examination.*

- (1) Show that $(1 - x^2)y'' - 2xy' + \lambda(\lambda + 1)y = 0$ has a non-identically zero polynomial solution if and only if $\lambda \in \mathbb{N}$.

[12 marks]

- (2) Find the general solution of $x'' - 4x = t^2e^{2t}$ using system of linear differential equations.

[12 marks]

- (3) Solve for $f(t)$ in the following problem:

$$f(t) = t + e^{2t} + \int_0^t e^{-\theta} f(t - \theta) d\theta$$

[10 marks]

- (4) A rope with uniform density has its ends tied to two opposite walls. Find the equation of the curve along which the rope hangs so that its potential energy is minimized.

[8 marks]

- (5) Solve the vibrating string problem if the initial shape is given by the function $f(x) = \lambda - \cosh\left(\frac{2x-\pi}{\pi}\right)$ for $0 \leq x \leq \pi$ where $\lambda = \frac{1+e^2}{2e}$.

[8 marks]

- (6) Use the series expansion of operator method to solve the differential equation $y'' + 6y' + 10y = e^{-3x}(x^3 + 7x^2 - 1)$.

[10 marks]

- (7) Use the method of variation of parameters to solve the differential equation $y''' + y' = \operatorname{cosec} x$.

[10 marks]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination

Semester II : 2012-2013

B.Stat.(Hons.) III Year

Introduction to Stochastic Processes

Date : 19.02.13

Maximum Score : 40

Time : $2\frac{1}{2}$ Hours

Note : This paper carries questions worth a total of 52 MARKS. Answer as much as you can. The MAXIMUM you can score is 40.

1. Let $\{X_n, n \geq 0\}$ be a Markov Chain on a state space S . Prove each of the following, starting from the definition of a Markov Chain.

(a) $P(X_9 \neq X_8 | X_0 = x_0, X_1 = x_1, \dots, X_5 = x_5, X_6 = x) = P(X_3 \neq X_2 | X_0 = x)$, for any $x_0, x_1, \dots, x_5, x \in S$.

(b) $P(X_n = y | X_0 = x_0, \dots, X_{n-2} = x_{n-2}, X_{n-1} = x, X_{n+1} = z, X_{n+2} = z_2, \dots, X_{n+m} = z_m)$
 $= P(X_1 = y | X_0 = x, X_2 = z)$, for any $n \geq 1, m \geq 2$ and $x_0, \dots, x_{n-2}, x, z, z_2, \dots, z_m \in S$.

(4+4)=[8]

2. Consider a MC $\{X_n, n \geq 0\}$ on the state space $S = \{1, 2, 3, 4, 5, 6\}$ and having transition

probability matrix $P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}$.

- (a) Classify the states into recurrent and transient states, providing justification for your answer.

(a) Find the probability ρ_{25} .

- (b) If the initial distribution of the chain is the uniform distribution on S , then what is the probability that the state 5 will be visited infinitely many times?

(4+4+4)=[12]

3. Consider the renewal chain where successive replacements have i.i.d life times with common distribution given by $P(L = x) = \frac{1}{x(x+1)}$, $x = 1, 2, \dots$ and X_n denotes the age of the machine in operation on the n th day, $n = 0, 1, 2, \dots$

(a) Show that the state space is irreducible with all states recurrent.

(b) Assuming usual notations, find $E_1(T_1)$ and $E_4(T_1)$.

(4+2×4)=[12]

4. Show that for a Markov chain with a finite number of transient states, there are constants $C < \infty$ and $\alpha < 1$ such that for every transient state y , one has $p_{xy}^{(n)} < C\alpha^n$ for any $n \geq 1$ and any state x .

[8]

5. Consider a gene with two alleles A and a . Under the *Wright-Fisher Model*, each generation consists of $2N$ genes and the $2N$ genes of a generation are obtained by drawing a random sample of size $2N$ with replacement from the $2N$ genes of the previous generation. Denote by X_n the number of A alleles in the n th generation.

(a) Write down the transition probabilities for the Markov chain $\{X_n\}$ and hence argue that it is an absorbing chain.

(b) Show that for every state x , $E_x[X_1] = x$ and more generally, $E_x[X_n] = x$ for all n .

(c) Hence (or otherwise), find the absorption probabilities.

(4+4+4)=[12]

Midterm Examination
 Statistical Inference II
 B. Stat. Third Year
 Second Semester
 2012-2013 Academic Year

Date : 21.02.13 Maximum Marks: 40 Duration : $2\frac{1}{2}$ hours.

Answer as many questions as you can. The maximum you can score is 40.

1. Suppose $X_{(\infty)} = X_1, X_2, \dots$ be an i.i.d. sequence of observations under both H_0 and H_1 . Let H_0 and H_1 be two hypotheses concerning each X_i such that

$$f(x_i|H_0) = 1 \quad 0 < x_i < 1$$

and

$$f(x_i|H_1) = 2 \exp(-2x_i) \quad x_i > 0$$

and suppose $0 < \alpha, \beta < 1$ are the pre-assigned errors of the two kinds with $\alpha + \beta < 1$. Derive the SPRT for this problem and without using Stein's Lemma, can it be shown that this SPRT terminates with probability 1? Explain. [2+2=4]

2. Suppose $X_{(\infty)} = X_1, X_2, \dots$ be an i.i.d. sequence of observations from a continuous probability distribution with the following p.d.f.

$$f(x_i) = \begin{cases} \frac{\gamma \rho^\gamma}{x_i^{\gamma+1}} & \text{if } x_i \geq \rho \\ 0 & \text{if } x_i < \rho. \end{cases}$$

Suppose ρ is known and we want to do a SPRT for $H_0 : \gamma = \gamma_0$ versus $H_1 : \gamma = \gamma_1 (> \gamma_0)$ with pre-assigned $\alpha, \beta > 0$ and $\alpha + \beta < 1$ and pre-assigned decision boundaries A and B .

- (a) Derive an approximate expression for the OC function for such an SPRT. [5]
- (b) Suppose $\alpha = 0.05$, $\beta = 0.05$, $\gamma_0 = 2$, $\gamma_1 = 3$ and $\rho = 4$. Is it true that on an average, the minimum number of observations required under H_0 to carry out the SPRT for this problem is 10? Explain. [6]
3. Let $\{X_n, n \geq 1\}$ be independent, identically distributed non-negative integer valued random variables and suppose it has a mass function $\{p_k\}$, i.e., $P\{X = k\} = p_k, k \geq 0$. Let $P_{X_1}(s) = E(s^{X_1})$ for $0 \leq s \leq 1$ be the generating function of X_1 . Let N be independent of $\{X_n, n \geq 1\}$ and suppose N is a non-negative integer valued random variable with a mass function $\alpha_j, j \geq 0$, i.e., $P\{N = j\} = \alpha_j$. Suppose $E(s^N) = P_N(s)$ denote the generating function of N . Define $S_0 = 0$ and $S_N = X_1 + \dots + X_N$, for $N \geq 1$.

- (a) Derive an expression for the generating function of S_N [7]
- (b) Using the result in part (a), show that $E(S_N) = E(N)E(X_1)$. [3]

4. Let X_1, X_2, \dots be i.i.d. random variables from the normal distribution $N(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$, $\sigma > 0$ are both unknown. Stein's two-stage method describes a procedure to construct a fixed-width confidence interval of μ as $J_N = [\bar{X}_N - d, \bar{X}_N + d]$ of width $2d$ where

$$N \equiv N(d) = \max \left\{ m, \left\lceil \frac{t_{m-1}^2 S_m^2}{d^2} \right\rceil + 1 \right\}$$

m is the pilot size of the procedure, S_m^2 is the estimate of σ^2 based on the pilot size m and t_{m-1}^2 is the upper $50\alpha\%$ of the t -distribution with $m - 1$ degrees of freedom.

- (a) Show that Stein's two-stage sequential procedure terminates with a probability 1, i.e., $P_{\mu, \sigma}(N < \infty) = 1$. [5]
- (b) Show that $P_{\mu, \sigma}\{\mu \in J_N\} = E_{\mu, \sigma}[2\Phi(\sqrt{N}d/\sigma) - 1]$. [3]
- (c) Show that $\text{Var}(\bar{X}_N) = \sigma^2 E_{\mu, \sigma}(N^{-1})$. [2]
5. Let X_1, X_2, \dots be i.i.d. real-valued random variables such that the MGF of each X_i is finite. Let $E(X_i) = \mu$ and $\bar{X}_n = n^{-1} \sum_{j=1}^n X_j$. Then, show that given any $a < \mu$, there exists $0 < \rho < 1$ such that $P(\bar{X}_n \leq a) \leq \rho^n$, for $n = 1, 2, \dots$ [5]

6. Consider an iid sample X_1, X_2, \dots, X_{10} from a continuous distribution F . Suppose ranks are given to the sample values thus obtained with the convention that the smallest sample value gets rank 1, the second smallest gets rank 2 and so on with the largest one getting a rank of 10. What is the probability that the sum of the ranks of X_1, \dots, X_5 is strictly larger than the sum of the ranks of X_6, \dots, X_{10} ? [4]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2012-2013

Course: B.Stat. III

Subject: Database Management Systems

Date: 23.02.13

Maximum Marks: 40

Duration: 3 hours

1. A socio-economic survey has been done in a set of remote tribal villages to study their agricultural practices. A village is identified by a name, which may be considered unique. Each village has certain number of houses identified by the house-no.

Each house belongs to a certain tribe like Santhal, Munda, Onrao etc. In each house, the number of persons involved in agriculture may vary. Some house may have non agricultural income also.

A village has three types of lands defined as HF (highly-fertile), MF (moderately-fertile) and UF (Un-fertile). UF lands are used for raising fruit trees only. Each plot of land is identified by plot number, land type and the cropping practice.

Cropping practice signifies whether a land is used for mono-cropping or multi-cropping (cultivated more than once in a year).

Village authority, for each village, keeps an account of the crops cultivated in that village in a year and the average yield for each of them in tons/acre. Basic cereals like paddy and wheat are cultivated on HF type lands only.

From the above description draw an ER/EER diagram. From the diagram derive a set of relation applying the standard mapping rules.

20+20=40

INDIAN STATISTICAL INSTITUTE
MID-SEMESTRAL EXAMINATION
Second Semester 2012-2013

B.STAT III year. Design of Experiments

February 26, 2013, Total marks 30. Duration: One and half hour

Answer all questions.

Keep your answers brief and to the point.
Marks will be deducted for rambling answers.

1. A laboratory gets its chemicals from 4 suppliers and wants to design an experiment to compare the chemicals obtained from the different suppliers. There are 4 scientists who can carry out the experiment, each scientist can do a maximum of 4 experiments during the entire study and a total of 4 experiments can be done in one day. All experiments in a day can be conducted under homogeneous conditions, but conditions from day to day may vary. Suggest a design for the experiment under each of the following assumptions;
 - (i) All scientists are equally skilled in performing the experiment.
 - (ii) All scientists are **not** equally skilled in performing the experiment.
 - (iii) All scientists are **not** equally skilled in performing the experiment. Moreover, there are 4 instruments for carrying out the experiment and these instruments are not all similar. No instrument can be used more than once in a day. [2 + 2 + 3 = 7]
2. (a) Define a connected block design.
 - (b) For a connected block design with t treatments, show that $(C + \frac{\mathbf{u}\mathbf{u}'}{w})^{-1}$ exists and is a generalized inverse of the C matrix, where \mathbf{u} is any $t \times 1$ vector with positive elements and $\mathbf{1}$ is the $t \times 1$ unit vector. [2 + (2 + 3) = 7]
3. (a) Define an orthogonal block design.
 - (b) Give an example of a connected and orthogonal block design in 4 treatments and 3 blocks. (You may choose the block sizes and replication numbers.) Justify your answer, correctly quoting all necessary results needed in your justification.
 - (c) For a block design, consider the reduced normal equations $C\tau = Q$ for estimating the treatment effects τ under the usual additive linear model. Show that if the design is connected and orthogonal, then $\hat{\tau} = r^{-d}T$ is a solution to this equation, where r^{-d} denotes a diagonal matrix with the reciprocals of the replication numbers in the diagonals and T is the $t \times 1$ vector of treatment totals.
 - (d) Hence, or otherwise, show that for a connected orthogonal block design, the adjusted treatment sum of squares reduces to the expression $T r^{-d} T - \frac{T' \mathbf{1}^2}{n}$ where n is the total of all observations in the design. [2 × 4 = 8] P.T.O.

4. (a) Define mutually orthogonal Latin squares.
- (b) Construct two mutually orthogonal Latin squares of order 5, (given that the elements of $GF(5)$: $0, 1, 2, 2^2, 2^3$.)
- (c) Construct, if possible, 10 mutually orthogonal Latin squares of order 11, such that each square has symbols $0, 1, \dots, 10$ and these symbols appear in the natural order along the principal diagonal of **each** square. [2+2+4=8]

INDIAN STATISTICAL INSTITUTE
Mid-Sem Examination, 2nd Semester, 2012-13
Statistics Comprehensive, B.Stat 3rd Year

Date: March 1, 2013

Time: 3 hours

**This paper carries 80 marks. Answer all questions.
Use separate answer scripts for each group.**

Group A

1. (a) When a least squares linear regression of Y is performed on X_1, X_2, \dots, X_n , the regression coefficient of X_n is $\widehat{\alpha}_n$. Alternatively, when a least squares linear regression of Y is first performed on X_1, X_2, \dots, X_{n-1} and the residuals are then linearly regressed on X_n using least squares, the regression coefficient of X_n is $\widehat{\beta}_n$. Explain whether $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ are necessarily equal.
(b) Suppose the monthly expenditure of B.Stat students is modeled by a probability density $f(x)$ that is proportional to $\exp\{-(x-b)^4\}$, $x \geq a$; $b > a$. Explain how you would simulate the mean monthly expenditure under this model of those students who spend in the interval $[E_1, E_2]$ using a random number generator from $U(0, 1)$.
[7 + 8]
2. (a) Suppose that the time to completion of a chemical reaction is distributed as exponential with mean λ . A set of n independent reactions is carried out and observed till time t_0 . If X_1, X_2, \dots, X_n be the times to completion of these reactions, where X_i is recorded as t_0 if the i^{th} reaction is not completed at time t_0 , $i = 1, 2, \dots, n$; obtain the m.l.e. of λ . Is the m.l.e sufficient for λ ?
(b) Suppose we want to estimate the difference in the proportions of airline passengers in Kolkata and Delhi who prefer window seats to aisle seats using an asymptotic 95% equal tail confidence interval. What is the minimum combined sample size required such that the error in estimation based on the confidence interval is at most 0.03?
[8 + 7]

3. Suppose X_1, X_2, \dots, X_m is a random sample from a p-variate normal population with mean μ_1 and dispersion matrix Σ ; while Y_1, Y_2, \dots, Y_n is an independent random sample from a p-variate normal population with mean μ_2 and dispersion matrix Σ . Obtain an unbiased estimator of the Mahalanobis Distance between the two populations as a linear function of its m.l.e. [10]

Group B

4. Consider a linear regression set-up:

$$y_i = \alpha x_i + \beta z_i + \alpha\beta u_i + e_i ; i = 1, 2, \dots, n$$

where, e_i s are random errors with mean 0 and the same variance. Describe an estimation procedure to obtain the least squares estimates of α and β . [12]

5. Construct an example of a test of hypothesis set-up H_0 versus H_1 where H_0 is rejected, and when H_0 is interchanged with H_1 , the new H_0 is also rejected at the same level of significance. [16]
6. Consider a group of 45 Type 2 Diabetes patients classified by two categories: gender (G) and age (A) as follows:

G \ A	≤ 40	> 40	
Male	10	15	25
Female	12	8	20
	22	23	45

Describe an algorithm of selecting a random sample of size 3 from the above population such that:

- (a) each gender and each age-group is represented in the sample;
 (b) the probability of inclusion of any individual in the sample is the same [12]

INDIAN STATISTICAL INSTITUTE
Second Semestral Examination : 2012-13
B. Stat. Third Year
Statistical Inference II

Date : 26.4.2013

Total Marks: 60

Duration :- 3 hours

Answer all questions

1. Let X_1, X_2, \dots, X_n be i.i.d. $N(\mu, \sigma^2)$ where both μ and σ are unknown. Consider the one-sided Kolmogorov-Smirnov statistic with μ and σ estimated by \bar{X} and S respectively, i.e consider

$$\hat{D}_n^+ = \sup_t \{F_n(t) - \Phi((t - \bar{X})/S)\},$$

where $F_n(\cdot)$ is the empirical distribution function. Show that the distribution of \hat{D}_n^+ does not depend on μ and σ . [5]

2. What is the probability of observing an A-run of length at least 5 in a random arrangement of 8 A's and 8 B's in a line ? Prove your assertion. [5]

3. Consider the two sample problem where one has a random sample X_1, \dots, X_m from a distribution F and an independent random sample Y_1, \dots, Y_n from a distribution G , where F and G are both assumed to be continuous.

(a) Describe the Wald-Wolfowitz runs test for testing $H_0 : F(x) = G(x)$ for all $x \in R$ versus $H_1 : F(x) \neq G(x)$ for at least one $x \in R$. Will this test criterion be suitable for testing against the alternative that the Y 's are stochastically larger than the X 's ? Explain.

(b) Consider the case when $m = n$. Consider the same testing problem as in part (a) and assume that actually $F(x) \neq G(x)$ for some $x \in R$. Show that the Kolmogorov-Smirnov two-sided test rejects H_0 with probability tending to 1 when n tends to infinity. You may assume the fact that the upper $\alpha\%$ point of the Kolmogorov-Smirnov statistic $D_{n,n}$ under H_0 tends to 0 as $n \rightarrow \infty$.

(c) Now suppose that $F(x) = G(x)$ for all $x \in R$. Find the limiting value of $P(D_{n,n}^+ > \lambda \sqrt{\frac{2 \log n}{n}})$ as $n \rightarrow \infty$, where $\lambda > 0$ and $D_{n,n}^+$ is a two-sample one-sided Kolmogorov-Smirnov statistic. [3+3+5]

4. Suppose X_1, \dots, X_m are iid with distribution F and Y_1, \dots, Y_n are iid with distribution G , where the X and Y samples are independent of each other while F and G are continuous but unknown. It is further known that $G(x) = F(x - \theta)$ for all $x \in R$, for some unknown constant $\theta \in R$.

(a) Consider the Hodges-Lehmann estimator of θ . Call it $\hat{\theta}$. Show the following :

- i. Distribution of $\hat{\theta} - \theta$ is the same for all $\theta \in R$.

- ii. $\hat{\theta}$ is symmetrically distributed about θ if $F(\cdot)$ is symmetric about some point $\mu \in R$.
- (b) Describe the Mann-Whitney U-test for testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$. Show that the probability of rejecting H_0 under $\theta > 0$ is a nondecreasing function of θ .
- (c) Describe the van der Waerden test for testing $H_0 : \theta = 0$ versus $H_1 : \theta > 0$. Prove that the null distribution of the test criterion is symmetric around 0. [(2+4)+4+3]
5. Suppose X_1, \dots, X_n are iid with a distribution F and Y_1, \dots, Y_n are iid with a distribution G , where the X and Y samples are independent of each other and F and G are continuous but unknown. You are told further that $G(x) = F(\theta x)$ for all $x \in \mathcal{R}$, for some unknown constant $\theta \in (0, \infty)$ and F has median 0. Describe the Siegel-Tukey test procedure for testing $H_0 : \theta = 1$ versus $H_1 : \theta > 1$. Show that under H_0 , the distribution of the test statistic is independent of the unknown F , as long as F is continuous. [1+3]
6. (a) Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. F where F is a bivariate distribution and the random variables X and Y are continuous and independent under F . Derive the asymptotic distribution of (suitably normalized) Kendall's sample tau coefficient.
- (b) Do you think Kendall's sample tau coefficient can be useful to test for randomness against the alternative of an upward trend in a time series data? Explain. [7+3]
7. Write a short note on the concept of Asymptotic Relative Efficiency of tests due to Pitman. [5]
8. (a) Stating appropriate assumptions, prove Wald's inequalities connecting the boundaries (A, B) and strength (α, β) of an SPRT for testing a simple null hypothesis against a simple alternative hypothesis.
- (b) Let $\{X_i\}_{i \geq 1}$ be a sequence of random variables and suppose the joint density of (X_1, \dots, X_m) is $p_{jm}(x_1, \dots, x_m)$ under hypothesis $H_j, j = 0, 1$. Let $0 < B < 1 < A < \infty$ and consider the SPRT of strength (α, β) that stops when $\lambda_m = \frac{p_{1m}}{p_{0m}} \geq A$ or $\leq B$ for the first time. H_0 is rejected if $\lambda_n \geq A$, accepted if $\lambda_n \leq B$ and no decision is made if $n = \infty$, where n is the stopping time of the test. Assume $\alpha > 0$ and $\beta > 0$. Will the Wald inequalities connecting A, B, α and β remain true? Justify your answer. Observe that we are not assuming that the X_i 's are iid. [4+3]

INDIAN STATISTICAL INSTITUTE
Semestral Examination – Semester II : 2012-2013
B.Stat. (Hons.) III Year
Introduction to Stochastic Processes

Date : 06.05.13

Maximum Score : 60

Time : 3½ Hours

Note : Qn.6 carries 15 marks and is compulsory. Qns.1-5 carry a total of 60 marks. You may answer as many as you want, but the maximum you can score in Qns.1-5 is 45.

1. Consider a Markov chain $\{X_n\}$ on state space $S = \{1, 2, 3, 4, 5\}$ with transition matrix

$$P = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

- (a) Find all possible stationary distributions, if any, for the Markov chain $\{X_n\}$.
 (b) Assuming that the chain starts with initial distribution $(0, \frac{5}{12}, 0, \frac{7}{12}, 0)$, find the expected time taken by the chain to return to its initial state. (6+4)=[10]

2. Let x be an essential state for a Markov chain.

(a) Define what is meant by the period of the state x .

Denote $\rho_{xx}^{(n)} = P_x\{T_x = n\}$ and let B denote the set of all $n \geq 1$ such that $\rho_{xx}^{(n)} > 0$.

(b) Show that $\rho_{xx}^{(n)} > 0$ if and only if n is a finite sum of elements from the set B .

(c) Conclude that the period of x equals the greatest common divisor of the elements of the set B . (2+6+4)=[12]

3. Let $\{X_n, n \geq 0\}$ be a Markov chain on a state space S and with transition probabilities $p_{xy}, x, y \in S$. Define $Y_n = (X_n, X_{n+1})$ for $n \geq 0$.

(a) Show that $\{Y_n, n \geq 0\}$ is a Markov chain on $S' = \{(x, y) \in S \times S : p_{xy} > 0\}$ and find its transition probabilities.

(b) Show that if $\{X_n, n \geq 0\}$ is irreducible and aperiodic, then so is $\{Y_n, n \geq 0\}$.

(c) Assuming that $\{X_n, n \geq 0\}$ has a stationary distribution π , show that the chain $\{Y_n, n \geq 0\}$ also has a stationary distribution and find it. (4+6+4)=[14]

4. 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 states. Assume also that every transient state leads to every state in S . Denote the transition probabilities of this chain by $p_{xy}, x, y \in S$.

Fix a transient state \bar{x} and consider the Markov chain $\{\tilde{X}_n, n \geq 0\}$ on the same state space with transition probabilities $\tilde{p}_{xy} = p_{xy}$ for all transient states x , while $\tilde{p}_{0\bar{x}} = \tilde{p}_{d\bar{x}} = 1$.

(a) Show that the chain $\{\tilde{X}_n, n \geq 0\}$ is irreducible and hence argue that it has a unique stationary distribution $\tilde{\pi}$.

(b) Show that for the original chain $\{X_n\}$, the expected time till absorption, given that it started from the state \bar{x} , equals $(\tilde{\pi}_0 + \tilde{\pi}_d)^{-1} - 1$. (6+6)=[12]

5. Show that if x is a recurrent state for a Markov chain $\{X_n\}$, then for every $m \geq 1$, $P_x\{X_{m+1} \neq x, X_{m+2} \neq x, \dots, X_{m+n} \neq x\} \rightarrow 0$ as $n \rightarrow \infty$. Prove also that, in case x is positive recurrent, the above convergence is uniform in m . (6+6)=[12]

6. Let $\{N_t, t \geq 0\}$ denote a Poisson Process with rate α .

(a) For $0 < u < s < t$, find the conditional joint distribution of (N_u, N_s) given $N_t = n$.

(b) Given that $N_t = 4$, find the conditional probability of the two events: (i) at least one of the four arrivals happened after time $2t/3$ and (ii) neither the waiting time for the first arrival nor the time between any two successive arrivals before t exceeded $t/2$.

(c) Let $\{N'_t, t \geq 0\}$ be another Poisson process with rate ν , independent of $\{N_t, t \geq 0\}$. Find the probability that there are three arrivals for the process $\{N'_t\}$ between the first and second arrivals for the process $\{N_t\}$. (3+(2×4)+4)=[15]

INDIAN STATISTICAL INSTITUTE
SEMESTRAL EXAMINATION
 Second Semester 2012-2013

B.STAT III year. Design of Experiments

Total marks 70. Duration: Three hours
 Answer all questions.

Keep your answers brief and to the point.

1. (a) You are to design a completely randomized experiment with 5 treatments and 24 experimental units. How many times should you use each treatment under the following objectives:
 - (i) You would like to minimize the average variance of BLUEs of all pairwise treatment contrasts.
 - (ii) You would like to minimize the average variance of BLUEs of the pairwise contrasts between treatment 1 with all other treatments. [3 + 5 = 8]
2. A block design with 9 treatments labeled 1, ..., 9 and 9 blocks numbered 1, ..., 9 is shown below with the blocks written as columns:

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

- a) Consider the usual additive model for block designs and let τ_i denote the effect of treatment i , $i = 1, \dots, 9$. Which of the following treatment contrasts are estimable from this design? **Briefly** justify your answers.
 - i) $\tau_3 + \tau_7 - 2\tau_8$
 - ii) $\tau_1 + 2\tau_4 - \tau_5 - \tau_6 - \tau_9$
 - iii) $\tau_1 - 2\tau_2 + \tau_4 - 2\tau_3 + 2\tau_8$
 - b) Suppose we consider a block design with 3 treatments and 3 blocks given by blocks 1, 5 and 9 above. Is this design orthogonal? **Briefly** justify your answer. [$(3 \times 2) + 2 = 8$]
3. An experiment is conducted as a Latin square design with 4 treatments. The observation from row 2, column 1, corresponding to treatment 3 is missing.
 - (i) Suppose you would like to keep the analysis simple by using the standard analysis of Latin square designs. Derive an expression for the estimate of the missing observation so that you could do that.

- (ii) What would you do if the usual analysis with your estimate (obtained in (i)) used in place of the missing observation leads to an **insignificant** value of the F statistic for testing equality of treatment effects? **Briefly** justify your answer. [5+3=8]
4. Consider a 2^4 factorial experiment run as a block design with 2 blocks as shown below in usual notation:
 Block 1: (1), a, b, ab, Block 2: c, ac, bc, abc, d, ad, bd, abd.
 (a) Check which of the following effects are estimable: A, B, AB, ABC. [2 × 4 = 8]
5. Consider a 2^5 factorial experiment in 4 blocks of size 8 each. Some of the treatment combinations in the **principal** block are : *ab, cd, ade*.
 (i) Write down the complete principal block.
 (ii) Write down **only** the block containing the treatment combination *b*.
 (iii) Identify all the confounded effects (no proofs required). [4+2+2=8]
6. You are to construct a 4^2 factorial experiment in blocks of size 4 each such that no main effect contrast is confounded. For this construction, show **only** the principal block. [8]
7. Suppose you need to construct a factorial experiment with 4 factors each at 3 levels in 2 replicates. Each replicate should have 9 blocks of size 9 each, such that the effects ABC , AB^2D are confounded in replicate 1 and AB^2C , AB^2D confounded in replicate 2.
 (i) Identify all the confounded interaction components.
 (ii) Write down the **principal block in replicate 1** only.
 (iii) Compare the information on A , ABC and AB^2D as obtained from this design with the information on these effects that would have been obtained from a design where each replicate could consist of one single block of size 81 with all treatment combinations appearing once each. [3 + 2 + 3 = 8]
8. In a chemical reaction the amount of catalyst used and the level of temperature maintained during the experiment affect the duration of the reaction. You would like to compare the effects of three amounts of catalysts and 3 temperature levels. Each day 9 observations can be taken under identical conditions and the experiment is to be conducted over 3 days.
 (a) Give the layout of a design for this experiment. (model not required).
 (b) You are told that during experimentation, it is easy to change the amount of catalyst but it is harder to change the level of temperature. In this case, suggest a design which will be easier to implement than the one in (a). [2+2=4]
9. In-class Presentation [10 marks]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2012-2013

Course: B.Stat. III

Subject: Database Management Systems

Date: 13/05/2013

Maximum Marks: 90

Duration: 3 hours

Answer all questions

1. A car insurance company maintains a database for its customers, agents, claims against cars insured and the settlements of such claims.

The company allocates a unique customer number to each of its customers and maintains the name, address and phone numbers of a customer against the customer number.

Similarly, for each agent, the insurance company maintains his agent-number (unique), name, address and phone numbers.

The record of each car insured with the company is retained with its registration number (which is unique), engine number (which is unique), chassis number (which is unique), make, model, type of body, cubic capacity, year of manufacturing, price, seating capacity, customer number, agent number, date of insurance (the date of insurance renewal will be one year from the date of insurance) and the amount of premium paid per year.

Detail record of claims for different accidents is also maintained by the company where each claim is assigned with a unique claim-number, date of claim and the amount of claim. Claims are classified as:

1. Settled-claims which have the date of settlement and the amount paid
2. Pending-claims are the claims those are yet to be settled. They have no extra attribute.

One customer may own more than one car. All such cars may not be insured by the same agent. One agent may be associated with more than one customer. One car is supposed to have only one owner.

(a) Draw an ER/EER diagram for the above problem.

(b) Map them to appropriate relations.

(c) Write SQL Query to solve the following problems:

- (i) Print all the Agent Names who have sold at least one insurance of every make of car insured.
- (ii) Print the Agent Name and total commission amount of the agent who earned the highest commission for 1997.

20+20+8+5=53

2. (a) List the inference rules for functional dependencies. Prove the transitive rule.

(b) Consider the following relation:

car_sale (car_no, date_sold, salesman_no, commission, discount_amount).

Assume a car may be sold by multiple salesmen and hence

{car_no, salesman_no} is the primary key. Additional dependencies are:

date_sold → discount_amount

salesman_no → commission

What normal form is the relation?

Explain your answer.

10+5=15

3. Consider the three transactions T_1 , T_2 and T_3 and the schedules S_1 and S_2 given below.
(a) Draw the serializability (precedence) graphs for S_1 and S_2 and state whether each schedule is serializable or not.

(b) If a schedule is serializable, write down the equivalent serial schedule(s).

$T_1: r_1(X); r_1(Z); w_1(X);$

$T_2: r_2(Z); r_2(Y); w_2(Z); w_2(Y);$

$T_3: r_3(X); r_3(Y); w_3(Y);$

$S_1: r_1(X); r_2(Z); r_1(Z); r_3(X); r_3(Y); w_1(X); w_3(Y); r_2(Y); w_2(Z); w_2(Y);$

$S_2: r_1(X); r_2(Z); r_3(X); r_1(Z); r_2(Y); r_3(Y); w_1(X); w_2(Z); w_3(Y); w_2(Y);$

5+5=10

4. Consider the following relational schema:

faculty (name, dept_name)

department (dept_name, building)

committee (comm_name, name)

Form the following queries using relational algebra/any calculus:

- (a) Find the name of all the faculties who are in any one of the committees that Professor XYZ is in.
- (b) Find the name of all the faculties who are in at least all those committees that Professor Smith is in.
- (c) Find the name of all the faculties who are in exactly (i.e., no more and no less) all those committees that Professor XYZ is in.
- (d) Find the name of all the faculties who have offices in at least all those buildings that Professor XYZ has offices in.

2+3+4+3=12

INDIAN STATISTICAL INSTITUTE

Supplementary Examination: 2012-2013

Course: B.Stat. III

Subject: Database Management Systems

Date: 26/06/2013

Maximum Marks: 60

Duration: 3 hours

Answer all questions

1. Consider the following information for a University Database:

- i. Every teacher has a unique number, a name, designation and an area of specialisation.
 - ii. Each teacher is associated with only one Department.
 - iii. Each Department identified by a unique name, has an address consisting of a unique building name and floor number.
 - iv. One of the teachers belonging to a Department serves as its Head of the Department.
 - v. Every research scholar identified by a unique roll number has a name, year of enrolment and one supervisor.
 - vi. Each research scholar is either a teaching assistant or a research assistant.
 - vii. Each teaching assistant is associated with the Department of his supervisor.
 - viii. University runs various projects sponsored by different funding agencies.
 - ix. Every project has a unique number, sponsor name, a starting date, a completion date, and a budget.
 - x. The project also has a principal investigator who is one of the teachers of the University and is (the project) assigned to that Department.
 - xi. A few other teachers from any Department may also be associated with each project as a member.
 - xii. A teacher can manage and / or work on multiple projects.
 - xiii. A research assistant may allow to maximum one project.
- a. From the above description draw an ER/EER diagram and map them to appropriate set of relations.
- b. Write the queries using SQL / Relational algebra/calculus (any 5).
- i. Find the name of the research scholars not associated with any projects.
 - ii. Find the name of the teachers who are only principal investigator of one or more projects, but not a member of any projects.
 - iii. Find the Department having the maximum number of teaching staffs with "xyz" specialisation.

- iv. Find the name of the Supervisor having maximum no of research scholars.
- v. Find the specialisation having maximum number of teachers.
- vi. Find the building name where "abc" Department is located.

20+10 + 3x5=45

2. Define Transaction. Assume that immediate modification is used in a system, Show by example, how inconsistent database state could result if log records for a transaction are not out put to a stable storage prior to committing that transaction.

8

3. How to test if a relation is in First Normal form and in Third Normal Form? Explain with example.

7



INDIAN STATISTICAL INSTITUTE
BACK PAPER EXAMINATION
Second Semester 2012-2013

B.STAT III year. Design of Experiments

Total marks 100. Duration: Three hours

Answer all questions.

Keep your answers brief and to the point.

Marks will be deducted for rambling answers.

1. (a) Define a connected block design.
(b) Give the layout of a design with 6 treatments and 7 blocks such that all contrasts between treatments 1, 2 and 3 are estimable, all contrasts between treatments 4,5 and 6 are estimable but no contrast between the set of treatments 1,2,3 and the set of treatments 4,5,6 are estimable.
(b) Write down a necessary and sufficient condition for a connected block design to be orthogonal. Prove the necessity part of the statement. [3+5+7=15]
2. (a) Construct GF(8). Hence construct 2 mutually orthogonal Latin squares of order 8.
(b) Show that there exists a maximum of $v - 1$ mutually orthogonal Latin squares of order v . [(7+7)+6=20]
3. (a) Construct a 3^3 experiment in 3 replicates, each replicate consisting of 9 blocks of size 3 each. Give the confounding scheme you use.
(b) Construct a design for this experiment if each block can have size 9. Give the confounding scheme you use in this case.
(c) Compare the designs in (a) and (b) in terms of information on different factorial effects. [5+5+5=15]
4. (a) For a non-orthogonal block design write down the expression for the C -matrix in usual notation. Derive the expectation and dispersion matrix of $C\hat{\tau}$.
(b) Show that all treatment contrasts can be estimable from the block design if $\text{rank}(C) = v-1$.
(c) Show that if $\text{rank}(C) = v-1$ then $C + uu'$ is positive definite where u is any vector of positive elements. [10+5+5=20]
5. (a) What are the advantages of a factorial experiment?
(b) Construct a factorial experiment in blocks of size 8, with 3 factors where 2 factors have 2 levels and one factor has 4 levels. [5+10=15]

6. An experiment for comparing 4 treatments was carried out as a randomized block design. The observations for treatments 2 and 3 in block 3 were lost.

Show how these missing values may be estimated from the available observations. [15]

Statistical Inference II

Maximum Marks: 100

Duration :- 3 hours **DATE - 02-07-1**

Answer all questions

1. Show, stating appropriate assumptions, that the null distribution of Kolmogorov-Smirnov one-sample goodness of fit test criterion is independent of the assumed null distribution from which the data are generated. [14]
2. Consider the one sample problem: X_1, \dots, X_n are iid F where F is continuous with unknown unique median M . We want to test $H_0 : M = 0$ versus $H_1 : M \neq 0$. Show that the sign test is consistent for this testing problem. [14]
3. Consider the two-sample testing problem of testing equality of two continuous distributions against the alternative that one distribution is stochastically larger than the other. Suppose two samples of size n each are drawn from the two distributions and the one sided Kolmogorov-Smirnov test statistic $D_{n,n}^+$ is used. Find $P_{H_0}(D_{n,n}^+ \geq \frac{k}{n})$, for $k = 0, 1, \dots, n$, where under H_0 the two distributions are identical. [12]
4. State and prove Stein's lemma about the termination property of an SPRT. [18]
5. State and prove the fundamental identity of sequential analysis. [12]
6. Stating appropriate assumptions, prove the optimality of SPRT in terms of its average sample number under both H_0 and H_1 among tests whose type I and type II errors are bounded above by α and β respectively where $0 < \alpha, 0 < \beta$ and $\alpha + \beta < 1$. [20]
7. Write a short note on testing for randomness using runs. [10]