INDIAN STATISTICAL INSTITUTE
B.Stat. (Hons.) III year
Sample Surveys
Semester I : 2000-2001
Date of Examination : 11.12.2000

Maximum Marks – 100   Duration :- 3 hrs.

1(a) Define the terms 'inclusion probability of a population unit' and ' joint inclusion probability of a pair of population units' for a given sampling design.

(b) From a population size N , one unit is drawn with probability of selection proportional to its size measure x. The rest of the (n-1) units of the population are selected from the remaining (N-1) units of the population by Simple Random Sampling without replacement. Write down p(s), the probability of obtaining the samples for this design. Hence show that the ratio estimator $\hat{R} = \sum_i \hat{y}_i \big/ \sum_i \hat{x}_i$ is unbiased for $R = \sum_i^N Y_i \big/ \sum_i^N X_i$, the ratio of population totals of two characteristics y and x under this design.

(c) For a fixed effective sample size design of fixed effective size n, simplify the expression $\sum_{i \neq j}^N \sum^N (\pi_{ij} - \pi_i \pi_j)$ where N is the population size.

                                                    (2+2)+(5+4)+5= 18

2(a) For stratified sampling with simple random without replacement sampling in each stratum, obtain an optimum allocation of the sample size under the cost constraint
$$C = C_0 + \sum_{i=1}^k c_i n_i \,,$$
so as to minimize the $Var.(\hat{Y}_{st.})$ where the symbols have their usual meaning. Write down the allocation $(n_1^{opt}, n_2^{opt}, \ldots n_k^{opt})$ when c_i's are all equal.

(b) Consider the above allocation $(n_1^{opt}, n_2^{opt}, \ldots n_k^{opt})$ .Suppose that the actual allocation in practice turns out to be $(n_1^{act}, n_2^{act}, \ldots n_k^{act})$.
(i) Obtain an expression for the relative loss in efficiency measured by
$$\left\{ Var._{act.}(\hat{Y}_{st.}) - Var_{opt.}(\hat{Y}_{st.}) \right\} \Big/ Var._{opt.}(\hat{Y}_{st.})$$
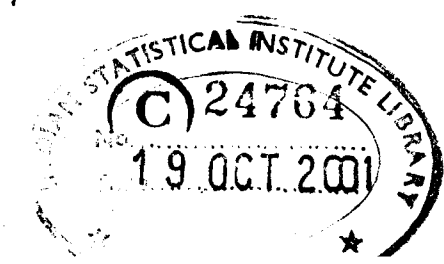where the symbols have the usual meaning and the stratum sizes are assumed large. (ii) Further derive a quick upper bound to the above expression in terms of θ the relative deviation of sample allocations given by $\theta = \underset{1 \leq i \leq k}{Max.} \left| (n_i^{opt} - n_i^{act.}) \big/ n_i^{act.} \right|$.   (10+2)+8+4=24

3(a) Distinguish between Linear and Circular Systematic Sampling techniques.
(b) How is Cluster Sampling different from Systematic Sampling?
    A population consists of N villages which are clusters of households, i^{th} village having M_i households. A sample of k clusters are selected by SRSWOR scheme. Suggest an unbiased estimator for the population mean (of the households) when $\sum_i^N M_i$ is known and give an unbiased estimator of the variance of the estimator. Also suggest a reasonable estimator of the same population parameter when $\sum_i^N M_i$ is unknown.

                                             4 + (2+3+5+2)= 16

P.T.O

This paper carries 120 marks. Answer all questions. Maximum you can score is 100. Precisely justify all your steps.

1. (a) What is an exact differential equation? Check that

$$(2xy^3 + y\cos x)dx + (3x^2y^2 + \sin x)dy = 0$$

is exact and hence solve it. [10]

(b) Check that the equation

$$x^2 y' = 3(x^2 + y^2)\tan^{-1}(y/x) + xy$$

is homogeneous and solve it. [10]

2. (a) Find the general solution of

$$x^2 y'' + xy' - y = x^2$$

by the method of variation of parameters. [10]

(b) Find a particular solution of

$$y'' + 2y' + y = e^{-x}\log x$$

[10]

3. Find the curve that satisfy the following : at each point of the curve, the part of the tangent cut off by the axes is bisected at the point on the curve. [10]

4. Let x_1 and x_2 be two successive positive zeros of a nontrivial solution Y_p of the Bessel's equation :

$$x^2 y'' + xy' + (x^2 - p^2)y = 0$$

Find the normal form of the Bessel's equation and hence, show that

(a) If $0 \le p < 1/2$, $x_2 - x_1 \le \pi$.

(b) If $p = 1/2$, $x_2 - x_1 = \pi$.

P.T.O.

5. A sample survey was conducted to estimate the total yield of wheat in a district. A two stage sampling design was adopted with villages as first stage units and plots within them as second stage units. From each stratum 4 villages were selected with probability proportional to area and with replacement and 4 plots were selected from each selected village with equal probability and without replacement. The data on yield for the sample plots together with information on selection probabilities are given below :

	Sample village	Inverse of probability of selection	Total Number Of Plots	Yield of sample plots 1	2	3	4
a		21.00	256	124	111	135	216
		16.80	288	123	177	106	138
		24.76	222	264	78	144	55
		49.99	69	300	114	68	111
	1	440.21	28	104	182	148	87
	2	660.43	14	108	64	132	156
	3	31.50	240	100	115	50	172
	4	113.38	76	346	350	157	119
I	1	67.68	189	110	281	120	114
	2	339.14	42	80	61	118	124
	3	100.00	134	121	212	174	106
	4	68.07	161	243	116	314	129

Using the above data,

(i) Obtain an unbiased estimate of the total yield of wheat in the district.

(ii) Obtain the unbiased estimate of the variance of the above estimator.

(iii) What are the possible sources of non-sampling errors in the above survey? How do you assess them and control them ?

(12+20+10)= 42

INDIAN STATISTICAL INSTITUTE
First Semestral Examination: (2000-2001)
ELECTIVE GEOLOGY
GEOLOGY
Maximum Marks: 100 Duration: 3 hours
Date :- 20.12.2000

1. How were the elements present in the rocks and minerals of the earth formed? **5**

2. What is the evidence indicating that the present-day earth contains heat energy? What are the sources of this heat energy? **5**

3. Distinguish between continental and oceanic crust? Why do continents stand higher above the ocean floor? **7**

4. Derive the radioactive decay equation. Enumerate the method of finding the age of a rock sample using the decay equation. **10**

5. Distinguish between ductile and brittle deformation. Draw sketches of structures that develop in rocks in response to heterogeneous ductile and brittle deformations. **8**

6. Name the folds represented in following stereograms. Bold and dotted lines are limbs and axial planes respectively. **10**
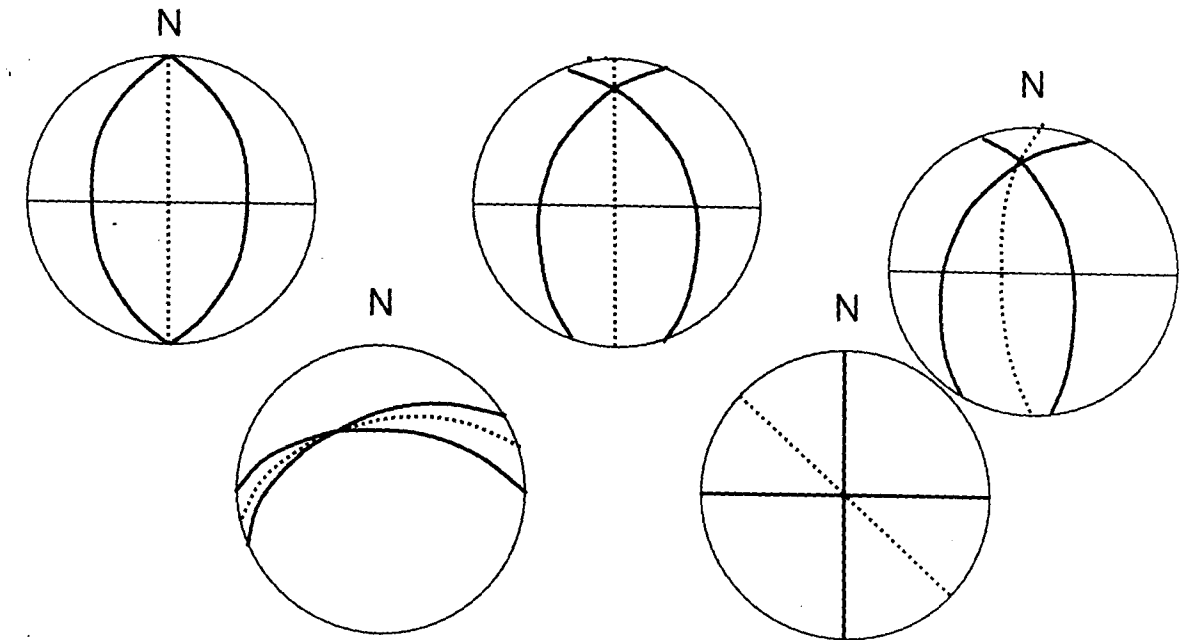
7. Discuss the major postulates of the theory of organic evolution proposed by Darwin. **5**

Either

8. Why are certain places of the earth prone to earthquakes? Why do the positions of the volcanoes of the world often fall in a line? **5**

P.T.O

(c) If $p > 1/2$, $x_2 - x_1 > \pi$. [15]

5. (a) The Legendre's polynomial of degree n is given by the *Rodrigue's formula*

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2 - 1)^n.$$

Show that P_n satisfies the differential equation

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

[Hint : Let $u = (x^2 - 1)^n$. Verify directly that for any $k \geq 1$,

$$(x^2 - 1)u^{(k+2)} - 2(n - k - 1)xu^{(k+1)} - (2n - k)(k + 1)u^{(k)} = 0]$$ [10]

(b) Prove that the Legendre's polynomials $\{P_n\}$ form a sequence of orthogonal functions on $[-1, 1]$. [10]

(c) Evaluate $\int_{-1}^{1} xP_n(x)P_{n-1}(x)dx$. [10]

6. (a) If P and Q are two points in the plane, then in terms of polar coordinates, length of a curve from P to Q is

$$\int_P^Q ds = \int_P^Q \sqrt{dr^2 + r^2 d\theta^2}$$

Find the polar equation of a straight line by minimizing this integral. [10]

(b) Solve the original isoperimetric problem in polar coordinates. That is, choose the origin to be a point on the curve and the polar axis to be the tangent line at that point; then maximize

$$\frac{1}{2} \int_0^\pi r^2 d\theta$$

with the side condition that

$$\int_0^\pi \sqrt{\left(\frac{dr}{d\theta}\right)^2 + r^2} d\theta$$

must be constant. [15]

[Hint : The curvature of a polar curve is given by $\dfrac{|2r'^2 - rr'' + r^2|}{(r'^2 + r^2)^{3/2}}$.]

1

Or

8. Why are magnetic anomalies noted as stripes of alternations of normal and reversed polarities on particular places of the ocean floor basalt? *5*

9. Answer the following questions- *10*

i)   Why are mostly hard parts of the dead organisms preserved as fossils?
ii)   Why is "Formation" mostly used in lithosatratigraphy as working unit?
iii)   Why are fossils noted in the igneous rocks?
iv)   What is "Numerical taxonomy"
v)   What is a biozone?

Either

10. Give an account of the principal sediment transporting agents on the earth. *15*

Or

10. Write short notes on: 15
    a) Sedimentary structures
    b) Fluvial Deposits
    c) Aeolian Deposits
    d) Grain Size Classification for the clastic sediments
    e) Modes of transport of clastics

Or

10. Discuss the principles of stereographic projection with illustration –*15*

Or

10.
    a) Which of the following processes are related with the transport of the clastic sediments:
        i)   Suspension
        ii)   humification
        iii)   traction
        iv)   saltation
    b) Arrange the following sedimentary structures in the order of increasing flow regime
        i)   Plane beds with parting lineation
        ii)   Plane beds without parting lineation
        iii)   Dunes
        iv)   ripples
    c) Relate the environments with the processes:

Environments   Processes

Alluvial Fans   Mass-Flow
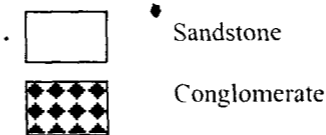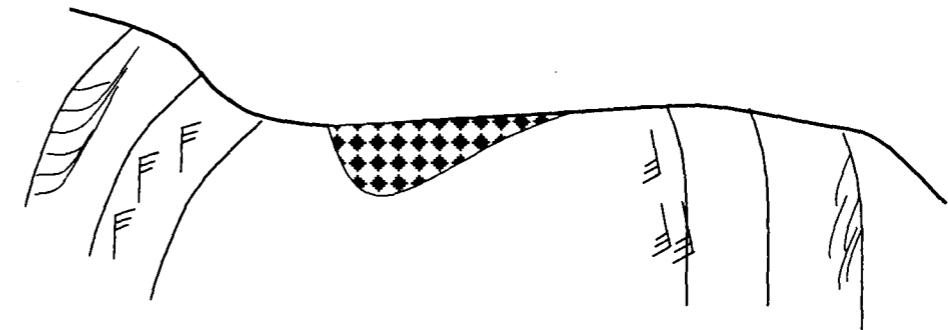Evaporative Lakes in deserts   Aqueous Suspension Fall Out
Lakes   Aqueous Unidirectional Traction
Beach   Aqueous Bi-directional Traction
River Channels   Chemical Precipitation
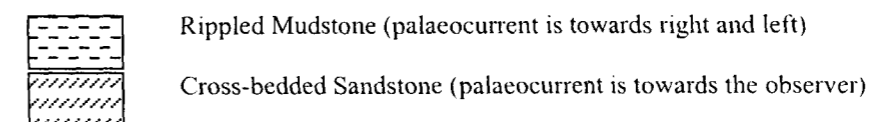
- *3x5=15*

11. The following diagram shows the disposition of the rocks in a vertical profile. Reconstruct the sequence of geologic events mentioning the principles used for each step. *10*

Sandstone

Conglomerate

12. Reconstruct the sedimentary environment from the following geological section: *10*

100 meter

Rippled Mudstone (palaeocurrent is towards right and left)

Cross-bedded Sandstone (palaeocurrent is towards the observer)
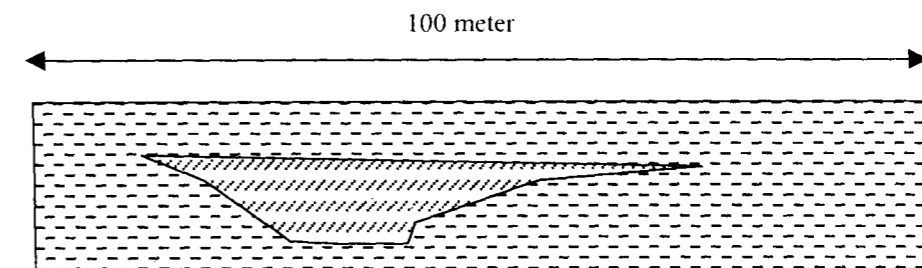
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD, CALCUTTA – 35

B.STAT. III YEAR.SEMESTRAL I EXAMINATION (2000-2001)

ECONOMICS - III

Date : 20.12.2000   Duration: 3 hrs.

Maximum Marks : 100

Answer any five questions :

1.   Describe the features of an under-developed economy with special reference to India.
                                                                        (20)

2.   (a)   Why is an underdeveloped economy caught in a low level equilibrium trap ?

      (b)   Why it is called 'low level' equilibrium trap ?

      (c)   What are the conditions conducive to such trap ?

      (d)   How can an economy get out of such a trap ?
                                                        (5 x 4 = 20)

3.   For each of the following cases, examine whether the statement is true, false or
      uncertain ? Explain your answer properly.

      (a)   If a relatively poor person loses income to a relatively rich person, the mean
             absolute deviation must rise.

      (b)   If everybody's income increases by a constant dollar amount, inequality must
             fall.

      (c)   The Lorenz Curve must necessarily lie in the lower triangle of the diagram,
             bounded by the 45 degree line at the top and the horizontal axes at the bottom.

      (d)   In the solow model, a change in the population growth has no effect on the long
             run rate of growth of per capital income.
                                                        (5 x 4 = 20)

4.   Suppose a country saves 20% of its gross national product and has a capital outpur
      ratio 4.

      (a)   Using the Harrod-Domar model, calculate the rate of growth of total fross
             national product.

P.T.O
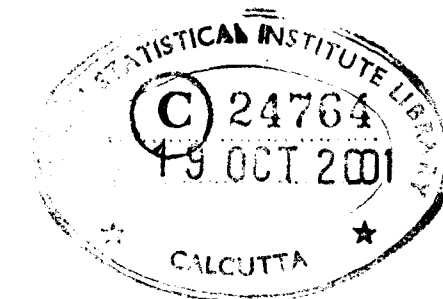
(b)   If the population growth were 3% per year and the economy wants to achieve a growth rate of 4% per year, what would its savings rate have to be ?

(c)   Estimate the per capita growth rates in case of (b).

$$(8+6+6 = 20)$$

5.   Consider two economies indexed by i = 1, 2 and described by

$$Y_i(t) = [K_i(t)]^\theta \qquad \theta > 1$$

$$\dot{K_i}(t) = s_i Y_i(t)$$

where Y =   total output

K =   capital stock

s =   marginal propensity to save

θ =   Effect of capital on output

(a)   Suppose that the two economies have the same initial value of K, but $s_1 > s_2$

Show that Y_1/Y_2 is continually rising.

(b)   How would Y_1/Y_2 behave if $K_1(o) > K_2(o)$ and $s_1 > s_2$?

(c)   What would be the behaviour of Y_1/Y_2 when $s_1 = s_2$, but $K_1(o) > K_2(o)$?

(d)   Compare the three cases : (a), (b) and (c).

$$(5 \times 4 = 20)$$

6.   Let an economy be described by

$$Y(t) = K(t)^\alpha \ H(t)^{1-\alpha} \qquad 0 < \alpha < 1$$

$$\dot{K}(t) = s_k \ Y(t)$$

$$H(t) = s_H \ Y(t)$$

Where Y   =   Total production

K   =   Total physical capital stock

H   =   Total human capital stock

s_k =   Fraction of income used in the formation of physical capital

s_H =   Fraction of income used in the formation of human capital

(a)   Show that regardless of the initial levels of K and H (as long as both are positive) the ratio K/H converges to some balanced growth path level $\left(\frac{K}{H}\right)^*$

(b)   Once K/H has converged to (K/H)* . what are the growth rates of K, H and Y ?

(c)   How, if at all, does the growth rate of Y on the balanced growth path depend on s_k and s_H ?

(d)   Suppose $\frac{K}{H}$ starts off at a level that is smaller than $\left(\frac{K}{H}\right)^*$. Is the initial growth rate of Y greater that, less than, or equal to its growth rate on the balanced growth path ?

$$(5 \times 4 = 20)$$

7.   Review your understanding of the development of a dual economy in the light of any dual economy model.

$$(20)$$

B.STAT. (Hons.) –III YEAR (2000-2001)

II SEMESTRAL EXAMINATION

STATISTICAL INFERENCE II

Date : 30.04.01   Duration : 3.5 hrs.

Maximum Marks : 100

Answer any four questions.

1. Let X_1, ..., X_n be i.i.d. observations with a common continuous distribution symmetric about its unknown median θ.

    a) Construct a 100 $(1- \alpha)$% confidence interval for θ on the basis of the Wilcoxon signed rank statistic.

    b) Find the asymptotic distribution of the Wilcoxon signed rank statistic under the hypothesis $H_0 : \theta = 0$

    (13 + 12 = 25)

2. Consider two independent samples $X1$, ..., X_n and Y_1,..... Y_n from two continuous distributions F and G respectively where $G(X) = F(X- \theta)$ for all X and some unknown θ. Our problem is to test $H_0 : \theta = 0$ vs. $H1 : \theta > 0$.

    a) Describe the Mann-Whitney U test and the Wilcoxon rank sum test for the above problem and show that these two tests are equivalent.

    b) Suppose that we reject H_0 for large values of a rank statistic of the form $S = \sum_{i=1}^{n}$
        Where $a(1) \leq ..... \leq a(m+n)$ are known values (not all equal) and Ri is the rank of Yi among all (m+n) observations.

        Show that the test based on S has a monotone power function in θ. What are the implications of this result in the context of testing $H'_0: \theta \leq 0$ vs. $H_1 : \theta > 0$

    (13 + 12 = 25)

P.T.C

3. a) Let X1, ..., Xn be i.i.d. observations with a common continuous distribution F with unknown median θ. Assume that F has support over the whole realline. Show that the sign test for testing Ho : θ = 0 is consistent for both one-sided and two-sided alternatives.

b) Describe the concept of Pitman's asymptotic relative efficiency (ARE) of tests.

c) Let X1,.... Xn be a random sample from a population having an N (θ, 1) distribution. Consider the problem of testing Ho : θ = 0 vs. H1 : θ > 0. Compute Pitman's ARE of the sign test relative to the usual parametric test based on the sample mean for a sequence of alternatives $\theta_n = \frac{\delta}{\sqrt{n}}$, $\delta > 0$. Assume that the asymptotic distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n}$ sign X_i under this sequence of alternatives is N $(\sqrt{2/\pi}\ \delta,\ 1)$.

(10+6+9 = 25)

4. State and Prove the "Fundamental Identity of Sequential Analysis" and describe how it can be used to find the approximate expressions for the OC and ASN functions of an SPRT.

(8 + 17 = 25)

5. a) Describe Stein's double sampling procedure to obtain a bounded length confidence interval for a normal mean μ when the variance σ^2 is unknown.

b) Let X1, X2, ... be i.i.d. observations from a Bin (1, θ) distribution, 0< θ <1. Our problem is to obtain an unbiased estimate of $1/\theta$. Suggest a sequential procedure and find an unbiased estimate that attains the Cramer-Rao lower bound for the sequential case.

(11 + 14 = 25)

INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD, CALCUTTA – 35

B.STAT. (Hons.) – III YEAR (2000-2001)

II SEMESTRAL EXAMINATION

SOCIOLOGY

Date : 11.05.01   Duration : 2 hrs.

Maximum Marks :100

Note : There are 8 questions, out of which you are required to answer 4 questions. Each questions carries 25 marks. Answers to each question are expected to be about 2 pages.

1. What are the major structural changes, in respect of output and labour force, that occur in an economy as income rises ?

2. What transformation of the peasantry do we expect with the development of capitalism ?

3. What was the social and cultural milieu of Prasanta Chandra Mahalanobis, founder of the Indian Statistical Institute ? To what extend was he influenced by Rammohan Roy and Rabindranath Tagore ?

4. How far was the Indian Statistical Institute a nation-building exercise on the part of its founder, Prasanta Chandra Mahalanobis ?

5. Suppose that you want to conduct a sociological survey in villages of a tehsil to estimate the level of social interaction between households. Discuss how you plan a survey formulating clearly the problem and describing the data to be collected and the sources of non-sampling errors.

6. What is the title of the book on religion written by Emile Durkheim ? Briefly explain the content of his theory on religion.

7. what are the disadvantages of taking up an empirical sociological study without a-priori hypothesis ?

8. Describe briefly the research tools you want to use in a village-level study.

ANTHROPOLOGY

Date : 11.05.01   Duration : 2 hrs.

Maximum Marks : 40

Answer any four questions from Group A and one question from Group B.

All question carry equal marks. Answer in separate answer scripts.

GROUP – A

1.   How do you define Anthropology ?

       What are the distinguishing features of Anthropology ?

2.   Why is man unique in the animal kingdom ?

3.   Critically examine the Darwin's theory of evolution.

       How does it differ from Lamarckian theory of evolution ?

4.   Write short notes on any <u>four</u> of the following :

       a)   Fertility and Fecundity
       b)   Caste and Class
       c)   Marriage and Mating
       d)   Adaptation and Acclimatization
       e)   Stable and stationery population
       f)   Mammals

5. Define :

   a) allele and trait
   b) Polymorphism and balanced polymorphism
   c) Genotype and phenotype

   How is allele frequency changed with and without natural selection ?

GROUP – B

6. Phenylketonuria, a metabolic disease in humans, is caused by a recessive allele, two heterozygous carriers of the allele marry and plan a family of five children : what is the chance that (a) all their children will be normal ? (b) four children will be normal and one affected with phenylketonuria ? (c) at least three children will be normal (d) first child will be a normal girl ?

7. a) In a population, frequency of L^M and L^N alleles were 0.78 and 0.22, respectively. If the population mates randomly with respect to M-N blood types, what are the expected frequencies of the M, MN and N phenotypes ?

   c) Consider a population in Hardy-Weinberg equilibrium. Suppose at an autosomal locus, there are two segregating alleles, A and a. The frequency of heterozygotes in the population is 0.5. What are the frequencies of each of the alleles ?

P. T. O

INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD, CALCUTTA – 35

B.STAT. (Hons.) – III YEAR (2000-2001)

II SEMESTRAL EXAMINATION

ECONOMICS - IV

Date : 9.5.2001   Duration : 3 hrs.

Maximum Marks : 100

Answer any five questions. All questions carry equal marks.

1. In the classical normal linear regression model, how do you predict (i) the expected value of the conditional distribution of Y (ii) an individual value of the dependent variable, given specified values of the regressors ? Show how one can set up probability limits for the values. Discuss, in this context, how one can test whether or not a fresh observation $(Y_{n+1}, X_{1\,n+1}, \ldots X_{K\,n+1})$

   satisfies the same structural relation as the n observations in the sample.

2. Consider the following linear regression model with non-stochastic regressors

   $$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad , \quad i = 1, 2, \cdots, n$$
   $$E(u_i) = 0 \quad \text{for all } i$$
   $$E(u_i u_j) = \begin{cases} \sigma^2 & \text{if } i = j \\ c & \text{if } i \neq j \end{cases}$$

   where all the variables have mean zero.

   If β_1 is estimated from the regression of y on x_1 with x_2 omitted, show that the resulting estimate is, in general, biased but has a smaller variance than that of the estimate obtained with x_2 included. When will the mean-square error $E(\hat{\beta}_1 - \beta_1)^2$ be

   smaller for the regression with x_2 omitted than with x_2 included ?

Note : *Answer as many questions as you like. The whole question paper carries 110 marks. The maximum you can score is 100. All Markov Chains considered in this paper are homogeneous in time.*

1. Consider a $Markov\ Chain$ with state space $S = \{0, 1, 2, 3, 4\}$ and the following transition matrix P

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{2} & 0 \\ \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{8} & 0 & 0 & \frac{7}{8} \end{pmatrix}$$

(a) Determine all the recurrent and transient classes.
(b) Find the absorption probabilities at 0 from each of the transient states.
(c) Find the following limits :

$$\lim_{n \to \infty} p_{14}^{n}, \quad \lim_{n \to \infty} p_{21}^{n}, \quad \lim_{n \to \infty} p_{23}^{n}$$

[5+10+15]

2. (a) Define an essential state. Show that in a finite $Markov\ Chain$ a state is transient if and only if it is not essential.
(b) Define the period of a state. Show that a finite irreducible $Markov\ chain$ is aperiodic if and only if there exists $n \geq 1$ such that $p_{ij}^{n} > 0$ for all i, j.

[(2+3)+10]

3. Consider a $Markov\ Chain$ with state space $S = \{0, 1, 2, \cdots, m\}$ and transition probability matrix

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & \cdots & p_m \\ p_m & p_0 & p_1 & \cdots & p_{m-1} \\ p_{m-1} & p_m & p_0 & \cdots & p_{m-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ p_1 & p_2 & p_3 & \cdots & p_0 \end{pmatrix}$$

where $0 < p_0 < 1$ and $p_0 + p_1 + \cdots + p_m = 1$ Show that $\lim_{n \to \infty} p_{ij}^{n}$ exists and find this limit for all i, j.

[10]

Please turn overleaf

---

3.   a)   What do you understand by the problem of heteroscedasticity in the context of a single equation regression model ?   Discuss at least two tests for detecting the presence of heteroscedasticity in a regression model.

   b)   Show that the least square estimator of the slope coefficient in the linear regression of Y on X from grouped observation is always less efficient than the corresponding estimator from ungrouped data.

4.   "Autocorrelated disturbances without lagged Y values do not produce biased estimators, even in small samples ; lagged Y values with random disturbances will give ordinary least square, (OLS) estimators which are consistent though biased in finite samples; the combination of the two problems, however, throws OLS off-beam and gives inconsistent estimators".

   Examine the statement in the context of OLS estimation of the regression coefficients in a linear regresion equation of Y_t on X_t and Y_{t-1}. Also, examine in this context, whether the conventional Durbin-Watson test is applicable in an equation containing lagged Y values among the explanatory variables.

5.   Give an account of the errors-in-variables model.   Show that the least squares estimator of the slope parameter in a two variable linear regression model will be an underestimate no matter how large a sample is available.   Derive the maximum likelihood estimator of the regression coefficients mentioning clearly the assumptions made.

6.   What explanations are usually given for the existence of autocorrelation in the disturbances of a linear regression model? Discuss alternative methods of estimating the regression coefficients in a linear regression model when the disturbances follow a first-order autoregressive process with an unknown autocorrelation coefficient.   What can you say about the properties of these estimators in the (i) Small Samples and (ii) Large Samples?

Answer all questions:

1. An equireplicate experiment was carried out to compare 4 treatments, using a completely randomised design and one covariate. It is assumed that the regression of the response y on covariate x is linear. There were a total of 32 observations. The following (in usual notation) were computed:

   $\bar{y}_{1.} = 18.37, \bar{y}_{2.} = 22.57, \bar{y}_{3.} = 21.87, \bar{y}_{4.} = 18.19$

   $\bar{x}_{1.} = 16.25, \bar{x}_{2.} = 15.62, \bar{x}_{3.} = 17.5, \bar{x}_{4.} = 16.62.$

   $\sum_i \sum_j (x_{ij} - \bar{x}_{i.})^2 = 2713.3, \sum_i \sum_j (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}) = -572.59, \text{MSE(observed)} = 4.72$

   The experimenter wants to study if the average effect of treatments 1 and 3 differs from average effect of treatments 2 and 4. She also wants to study all pairwise treatment comparisons.

   a) What are the contrasts of interest? Obtain their estimates and their standard errors.

   b) To find simultaneous 95 per cent confidence intervals for the above preplanned contrasts, would you prefer Bonferroni's method or Scheffe's method of multiple comparison? Justify your answer.

   c) Give expressions for the confidence intervals by the above two methods.

   Compute the intervals for the above contrasts by the method chosen by you in (b) above. Would Tukey's method be applicable in this case?

   $[(4+4+4)+3+(6+4)=25]$

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

   a) Show how these missing observations may be estimated. (Final simplified form of estimate is not required, but initial derivations must be shown.)

   b) If these two estimates are used in place of the missing observations and then the augmented data is analysed, will the usual F test for treatments be biased? Justify your answer with mathematical proof.

   $[(9)+6=15]$

3. There are 3 formulations for manufacturing ceramic dishes and an experiment is to be planned to compare the effects of these 3 formulations on the quality of the dishes. It is believed that the quality will be also affected by the oven used in baking the dishes, the quality of the raw material used and the operator doing the baking. The experiment with these 3 formulations will be repeated in 4 factories. All these 4 factories get raw material from the same three suppliers.

   a) Suggest a suitable design for this experiment, assuming that enough operators and ovens are available in each factory.

---

4. A person who owns r umbrellas distributes them between home and office according to the following routine. If it rains when he leaves for office or leaves office for home, he takes an umbrella with him (provided, of course, there is one at that location to take). Assume that the probability that it rains at the time he starts any of his trips is $p\,(0 < p < 1)$, independent of all other things. Denote by X_n the number of umbrellas available to him when he starts his n^{th} trip.

   (a) Find the transition matrix of this $MC\,\{X_n, n \geq 0\}$. Show also that it is an irreducible aperiodic chain.

   (b) Show that $\underset{\sim}{\pi} = (\pi_0, \pi_1, \cdots, \pi_r)$ given by

   $$\pi_0 = \frac{1-p}{1-p+r} \quad \text{and} \quad \pi_1 = \pi_2 = \cdots = \pi_r = \frac{1}{1-p+r}$$

   is the unique stationary initial distribution for this MC.

   $[12+8]$

5. There are two external P&T lines connected to the NE20M telephone exchange of the ISI Campus. P&T calls arrive at this exchange according to a Poisson Process at the rate of 10 calls per hour. Any incoming call is accepted if at least one line is free, otherwise (i.e. if both line are busy) the call is lost. The durations of the calls are i.i.d. exponential random variables with mean 3 minutes and are independent of the incoming Poisson Process. Let X_t, $t \geq 0$, denote the number of busy lines at time t, $t \geq 0$.

   (a) Set this up as a birth and death process and find the infinitesimal parameters.

   (b) Find the steady state probability of both lines being busy.

   $[6+9]$

6. Write down the forward differential equations for the continuous time Markov process $\{X_t, t \geq 0\}$ with state space $S = \{0, 1, 2\}$ and the infinitesimal parameters given by

   $$Q = \begin{pmatrix} -1 & 1 & 0 \\ 3 & -6 & 3 \\ 0 & 1 & -1 \end{pmatrix}$$

   Solve these equations to find $P_{00}(t)$, $P_{01}(t)$ and $P_{02}(t) = 1 - P_{00}(t) - P_{01}(t)$ explicitly.

   $[5+15]$

   ----

b)Give the layout of your design. State the model for analysing the data collected and gi
the ANOVA table showing the different sources of variation and the corersponding degrees
freedom.   [4+(5+3+3)=1

4. A chemist needs a design for an experiment for which she has two factors of interest: t
   temperature at which the response is taken and the amount of catalyst used. She can control
   temperature at 4 levels and can use 3 different amounts of catalyst. She can take 12 observati
   on each day and will take observations for 5 days.

   a)If levels of both factors may be changed easily, suggest a suitable design for this experimen

   b)If it is found that the temperature levels cannot be changed easily, suggest a modification
   your design suggested in (a) above so that this practical difficulty may be overcome.

   c)For the design in (b) above, clearly state the suitable model and give the method of testir
   the hypotheses of interest. Which of the two factors will be compared with greater precision?
                                                                    [3+3 +(4)+(4+3)=1

5. a)Why is partial confounding to be preferred over total confounding in factorial experiments?

   b) A factorial experiment with 3 factors A, B, C; each at 2 levels, is to be planned. Each d
   only 4 treatment combinations may be tested. It is desired to have some information on a
   the effects with as much information as possible on the main effects. Among the interaction
   interactions BC and ABC are of lesser interest than the others. Give the layout of the design f
   this experiment using a suitable confounding scheme if the experiment can be run for 4 days.

   c)Derive the relative efficiencies of estimates of interactions AB, BC and ABC compared to th
   main effect estimates.

   d) If all the 8 treatment combinations could be observed in each day and the experiment w
   to be run over 2 days, what would be the relative efficiencies of estimates of main effect A a
   interaction AB compared to the corresponding estimates from the experiment in (b) above?
                                                                    [3+5+(4+4)+ (4) =2

6. An experiment with 3 factors A, B, C each at 3 levels, needs to be carried out in blocks of si
   9. It is decided to use 4 replicates.

   Give a suitable confounding scheme for this experiment using partial confounding. Give t
   treatment combinations in the **principal block** of <u>any one</u> replicate.   [4+4=