3

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2011-12

B. Stat. III Year
Sample Surveys

Date: 02.01.2011 **Maximum Marks: 50** **Duration: 3 Hours**

No external equipments to be supplied.

Answer any 4 of the following questions carrying 10 marks each.

Records of work done in Tutorial classes to be submitted on the date of the Examination carry to marks.

1. Establish consistency conditions the inclusion probabilities of the 1^{st} two orders are known to satisfy for any sampling design.

2. Show why positivity of the inclusion probability of every unit of a survey population is a necessary and sufficient condition for the existence of an unbiased estimator of a finite survey population total.

3. Write down the formula for the Hartley- Ross estimator for a finite population total based on an SRSWOR. How will you unbiasedly estimate its variance ?

4. Explain the content and rationale of a complete class theorem in the context of survey sampling. Clearly show its use with a concrete example, referring to PPSWOR sampling of only 2 units to unbiasedly estimate a finite population total.

5. Derive a good predictor for a finite population total separately according to each of predictive, Super-population modelling and Model Assisted approaches.

6. Explain Lahiri's PPS sampling method to choose a unit from a finite population. Write down explicitly the Horvitz-Thompson estimator for a finite population total based on such a PPSWR sample in n draws. Do you recommend its use ? Answer giving reasons.

   Write down its variance and an unbiased estimator thereof giving explicit formulae.

*****

Indian Statistical Institute
Linear Models
B-III, Midsem

Date: Sep 05, 2011 Duration: 2hrs.

The paper carries 40 marks. Attempt all questions. Justify your steps.

1. Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, I)$. Let V be any r-dimensional subspace of $\mathbb{R}^n$, and $\mathbf{b} \in \mathbb{R}^n$. If $\widehat{\mathbf{Y}}$ is the vector in the *affine* linear space $\mathbf{b} + V$ closest to $\mathbf{Y}$. Show that the condition $\boldsymbol{\mu} \in \mathbf{b} + V$ is sufficient for
$$\|\mathbf{Y} - \widehat{\mathbf{Y}}\|^2 \sim \chi^2_{(n-r)}(\text{central}).$$
Is the condition necessary also? [5+5]

2. State and prove the Gauss-Markov theorem. [10]

3. Consider the 2-way classification model with interaction for balanced data:
$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$
for $i = 1(1)I$, $j = 1(1)J$ and $k = 1(1)K$. Assume that the ϵ's are iid $N(0, \sigma^2)$.
   (a) Assume that all the effects are fixed. Write down the first three columns of the standard ANOVA table. [5]
   (b) Now consider α_i's to be random effects having iid $N(0, \tau^2)$ distribution independent of ϵ's. Find the expectation of
$$IK \sum_j \bar{y}^2_{\bullet j \bullet} - IJK\bar{y}^2_{\bullet\bullet\bullet}.$$
[5]

4. Consider the 4×4 Latin square model based on the square The experiment is conducted with cows of 4 types and four age groups. There is exactly one cow of each type, age group combination. Each cow is given one of four different diets according to the following Latin square:

1	2	3	4
4	1	2	3
3	4	1	2
2	3	4	1

Thus if (i, j)-th cell contains k then the k-th diet is given to the cow of i-th type and j-th age group. After a month the yield of milk is measured. It is known that these factors do not interact. Write down a linear model to relate the yield of milk to type, age group and diet. What is the rank of the design matrix? Write down the first two columns (source and d.f.) of the ANOVA table for your model. [5+5]

INDIAN STATISTICAL INSTITUTE

Mid- Semestral Examination: 2011-12

Course Name : B III

Subject Name : **ANTHROPOLOGY**

Date 6/9/11 Maximum Marks: **50**

Duration: 3 hours

Note, if any :

Answer any five questions. Each question carries 10 marks.

1. Define Anthropology. What are the subdivisions of Anthropology?

2. Describe the man's position in Primate taxonomy.

3. What do you mean by adaptation? Briefly describe the adaptation of man at high altitude.

4. What do you mean by demography? Briefly describe the different shapes of age - sex pyramid and their features.

5. Why slow rate of growth is advantageous to children living at high altitude?

6. Write short notes on any two of the following:

   (a) Cold adaptation (b) Growing population (c) Infant mortality (d) Cultural adaptation

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination (2011-2012)

Subject: Introduction to Sociology

B.Stat. III Year

Date: 7th September 2011 Maximum Marks: **25** Duration: Two hours

<u>The figures in the margin indicate full marks</u>

Answer question No. 1 and any three of the rest

1. Write a note on the nature of development of Sociology in the 21st Century in India? 10

2. Write short notes: 2.5 x 2.5 = 5
   a) Observation and experiment.
   b) Interview.

3. Differentiate between Westernization and Modernization. Also point out its impact on Indian society. 2.5 + 2.5 = 5

4. Write any two questions: 2.5 x 2.5 = 5
   (a) What is meant by empowerment of women?
   (b) What is Gram Sabha?
   (c) What is meant by Integrated Rural development Programme?
   (d) What is meant by Social inequality.

5. Define culture. What are the elements of culture? Describe the importance of culture in understanding individual behaviour. 1+1+3= 5

INDIAN STATISTICAL INSTITUTE
Mid-Semestral Examination: 2011-2012

Course Name: BStat III Year

Subject Name: GEOLOGY

Date: 07.09.2011   Maximum marks: 40   Duration: 2 Hrs

Note, if any: ------------

ANSWER QUESTIONS 1, 2 AND ANY TWO FROM THE REST

1. Write short notes on --------------------------- 3 X 5=15
   a. Uniformitarianism
   b. Law of Superposition
   c. Nessosilicates (Island silicates)
   d. Miller's indices for Crystal faces
   e. Phaneritic and aphanitic texture

2. How do we know that the Earth has layered structure? Name the three major layers present within the Earth and the two discontinuity surfaces present between each two layers. State the compositional difference between SIAL and SIMA. ------------------------3+5+1=9

3. Define the three major types of plate boundaries with cartoon/ illustrations. Why elongate chains of mountains are present through out the western coasts of North and South America? --- 6+2=8

4. What do you understand by the term igneous rock? How are they formed? What are the two important variables used for the classification of igneous rocks? Name the igneous rocks having SiO2 (Silica) Content > 66 wt. % - . 52-66 wt% , - 45-52 wt% and < 45 wt % . ---------1+3+2+2=8

5. What do you understand by the term fractional crystallization? Explain with the help of a simple diagram whose end points are A and B. What is eutectic end? ------------------------ 3+5=8

Indian Statistical Institute

Mid Semestral Examination: (2011–2012)

B.Stat.(Hons.) – III year

Economics III

Date: ~~00/00/2011~~ 07.09.11 *Maximum Marks –50* *Duration: 2 hours*

Answer ***any two*** questions.

1. (a) State and explain the assumptions underlying the Classical Linear regression Model (CLRM).

(b) Consider a simple regression model $y_i = \alpha + \beta x_i + \varepsilon_i$, *i=1,2,...,n,* for which all classical assumptions of CLRM are satisfied except that $E(\varepsilon_i) = \lambda z_i$, where z_i is non-stochastic and such that $\sum (x_i - \bar{x})(z_i - \bar{z}) = 0$. Determine the bias (if any) in estimating the parameters by OLS.

(c) Show that $E(e'e) = \sigma^2 (n - k - 1)$, where *e* is the vector of OLS residuals, *k* is the number of regressors and $\text{var}(\varepsilon_i) = \sigma^2$, *i=1,2,...,n.*

[10+8+7=25]

2. (a) What do you mean by a "dummy variable"?

(b) A regression equation explaining household expenditure on recreation (*y*) as a function of income (*x*) was estimated as

$$y_t = -25 + 35D_{t2} + 40D_{t3} - 15D_{t4} + 0.05x_t + e_t$$

$$\quad (10) \quad (16) \quad (15) \quad (8) \quad (0.02)$$

where figures in parentheses are the standard errors, *D*'s are the quarterly dummies with the 1st quarter (January-March) left out.

Determine the values of the estimated coefficients given that

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

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

(c) What is dummy variable trap?

(d) Suppose consumption expenditure y depends on income x, but different levels of income produce different values of marginal propensities to consume (β_i). In particular,

$$\frac{dE(y|x)}{dx} = \begin{cases} \beta_1 \ \ if\ x < Rs.100 \\ \beta_2 \ if\ Rs.100 \leq x < Rs.500 \\ \beta_3 \ \ if\ x \geq Rs.500 \end{cases}$$

Using Dummy variables estimate the marginal propensity to consume for the three groups (assuming that there is no intercept term) given that $\sum yx$ is 30000; 5,00,000 and 10,00,000 for the three groups, respectively, and $\sum x^2$ is 90000; 25,00,000 and 100,00,000 for the three groups, respectively.

[2+8+5+10 = 25]

3. (a) What is multicollinearity?

(b) Describe the procedure of detecting multicollinearity using 'condition number' and 'variance proportions'.

(c) Explain the Principal component and Ridge regression methods in this context.

[3+12+10=25]

Indian Statistical Institute, Kolkata

Midsemestral Examinations : B.Stat. III year

~~Partial~~ *Differential Equations*

Maximum marks : 60 | September 09, 2011 | Time : 3 hours

Answer all questions

1. Find the general solution of the following equations:

   (a) $\frac{dy}{dx} = \frac{3y}{3y^{\frac{2}{3}} - x}$

   (b) $xdy + ydx + 3x^3y^4 = 0$

   (c) $x^2y'' = 2xy' + (y')^2$

   (d) $y'' - 2y' + y = 2x$

   (e) $(x^2 - 1)y'' - 2xy' + 2y = (x^2 - 1)^2$ [$3 \times 5 = 15$]

2. The president and the prime minister order coffee and receive cups of equal temperature at the same time. The President adds a small amount of cool cream immediately and does not drink her coffee until ten minutes later.The prime minister waits 10 minutes, and then adds the same amount of cool cream to his coffee and begins to drink. Who drinks the hotter coffee?

   [6]

3. (a) The logistic model for population growth gives the differential equation $\frac{dp}{dt} = ap - bp^2$ for the population p of a specis as function of time, where a, b are positive constants, b small in comparison to a and the term $-bp^2$ takes care of the competition for survival when the population becomes large. If p_0 is the population at time t_0, find the population $p(t)$ at time t and examine the behaviour of $p(t)$ as $t \to \infty$.

   (b) A drop of falling rain collects water vapour along its path at a rate proportional to its surface area. If at time $t = 0$, the drop has radius 0 and (of course) velocity 0, show that the drop falls with an acceleration of $\frac{g}{4}$.

   [$6 + 8 = 14$]

4. *Attempt any two of the following.*

   (a) A tank contains S_0 kilogrammes of salt dissolved in 1000 litres of water. Starting at time $t = 0$, water containing .25 kgs of salt per litre enters the tank at the rate of 20 litres per minute, and the well stirred solution leaves the tank at the same rate. Find the concentration of salt in the tank at any time $t > 0$.

   (b) Find the shape assumed by a flexible chain hanging between two points under its own weight.

   (c) A smooth (American) football in the shape of a prolate spheroid 12 inches long and 6 inches

PTO

thick is lying outdoors in the rain. Find the paths along which water will run down its sides.
[6 + 6 = 12]

5. (a) A particle moves along the X−axis subject to a force towards the origin proportional to x starting from rest at time $t = 0$ at the position $x_0 \neq 0$. Find the kinetic energy and the potential energy of the particle at time t. Show that the time average (over the time period of the particle) of its kinetic energy is equal to a half of the total energy of the particle.

(b) Write down the differential equations of motion of a particle moving in a plane under a central force.

[6 + 7 = 13]

13 - III

Date: August 12, 2011

INDIAN STATISTICAL INSTITUTE

Time : $2\frac{1}{2}$ Hours | Statistical Inference - I | Full Marks : 80

Mid-semester Examination

[This question paper carries 90 marks. Answer as many as you can. The maximum you can score is 80.]

1. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where μ and σ^2 both are unknown.
   (a) Find out the UMVUE for μ^3. [4]
   (b) Assuming σ^2 to be known, find the UMVUE for (i) $\Phi((\mu - 10)/\sigma)$ and (ii) $\phi((\mu - 10)/\sigma)$, where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the distribution function and the density function of a standard normal distribution, respectively. [4+4]

2. (a) A dice is thrown repeatedly until we get all the faces at least once. Let X be the minimum number of throws required to get all these faces. Show that the distribution of X is not complete. [5]
   (b) If $X_1, X_2, \ldots, X_n$ are i.i.d. Poisson(θ), show that the UMVUE of $1/\theta$ does not exist. [4]
   (c) If $X_1, X_2, \ldots, X_n$ are observations from a continuous distribution F_1 and $Y_1, Y_2, \ldots, Y_n$ are observations from another continuous distribution F_2, and they are independent, find the UMVUE for $P(X < Y)$. [3]

3. (a) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where μ and σ^2 both are unknown. Find the univariate and the multivariate Cramer Rao lower bound for the variance of an unbiased estimator of μ, and check which one is sharper. [2+2+1]
   (b) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. Poisson(θ). Find the UMVUE of $e^{-3\theta}$. Check whether its variance attains the Cramer Rao lower bound. [3+4]

4. (a) If $X_1, X_2 \ldots, X_n$ are i.i.d. $U(\theta - 1, \theta + 1)$, find a maximum likelihood estimator of θ and check whether it is consistent for θ. [3+3]
   (b) If $X_1, X_2 \ldots, X_{20}$ are independent observations from a distribution with density function $f_\theta(x) \propto e^{-\{2|x-\theta|+(x-\theta)\}}$, find a maximum likelihood estimator of θ and check whether it is unique. [5+1]

5. Suppose that the life time of an electric bulb follows an exponential distribution with mean θ. A person switched 50 bulbs at a time, and after 100 hours, 40 of them were found to be in functioning state.
   (a) Find the maximum likelihood estimate of θ.
   (b) Suppose that the life time of the first 10 bulbs were noted as $50, 55, 60, 65, 70, 75, 80, 85, 90$ and 95, respectively. Write down the modified likelihood function and modify your maximum likelihood estimate.

P.T.O

(c) Describe how the EM algorithm can be used in this problem for finding the maximum likelihood estimate of θ.

(d) Show that this algorithm converges, and it converges to the true solution of the maximum likelihood equation. [2+4+3+3]

6. Prove or disprove the following statements.

(a) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. f_θ. If f_θ belongs to an exponential family with full rank, the joint distribution of $(X_1, X_2, \ldots, X_n)$ is complete. [2]

(b) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. f_θ. If f_θ is not complete, the UMVUE of θ cannot exist. [2]

(c) A minimal sufficient statistic is always complete. [3]

(d) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. f_θ. If f_θ is symmetric about θ, the joint distribution of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is complete. [3]

(e) If T_1 is the UMVUE for $\gamma_1(\theta)$, T_2 is the UMVUE for $\gamma_2(\theta)$ and they are independent, T_1T_2 is the UMVUE for $\gamma_1(\theta)\gamma_2(\theta)$. [3]

(f) If θ is positive, the UMVUE for θ always takes positive values with probability one. [3]

(g) If $X_1, X_2, \ldots, X_n$ are i.i.d. $U(0,\theta)$, for any positive integer $k < n$, we have $E(X_{(k)}^2/X_{(n)}^2) = E(X_{(k)}^2)/E(X_{(n)}^2)$, where $X_{(k)}$ $(k = 1, 2, \ldots, n)$ is the k-th order statistic. [3]

(h) Maximum likelihood estimator is a function of minimal sufficient statistic. [2]

(i) If $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where μ and σ^2 both are positive and unknown, $|\bar{X}|$ is the MLE for μ. [3]

(j) If $X_1, X_2, \ldots, X_n$ are i.i.d. with density function $f(x) \propto e^{-(x-\theta)^{10}}$, a maximum likelihood estimator of θ exists and it is unique. [3]

(k) If $E(T_n) \to \theta + 1$ as $n \to \infty$, T_n cannot be a consistent estimator for θ. [3]

November 14, 2011

INDIAN STATISTICAL INSTITUTE

Statistical Inference -I , B III

ne : 3 Hours | Semestral Examination | Full Marks : 100

Answer as many as you can. The maximum you can score is 100.

1. (a) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where σ^2 is known. Find the UMVUE of $e^{t\mu}$ and check whether its variance attains the Cramer Rao lower bound ? [2+6]

   (b) Check whether the family of distributions $\{f_\theta = U(e^{-\theta}, e^{\theta});\ \theta > 0\}$ is complete. [4]

2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the discrete uniform distribution with p.m.f. $f_N(x) = 1/N,\ x = 1, 2, \cdots, N,\ N \geq 1$, and $\psi(N)$ be a real valued function of N.

   (a) Find an unbiased estimator of $\psi(N)$ based on a single observation. [4]

   (b) Find the UMVUE of $\psi(N)$. [8]

3. (a) If $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\mu, 1)$ truncated at α and β (i.e. $f(x) = Ce^{-(x-\mu)^2/2} I\{\alpha \leq x \leq \beta\}$, where C is an appropriate constant). Show that $\bar{X}$ is the unique MLE for $\psi(\mu) = E(X_1)$ [4]

   (b) Prove or disprove the following statements.

      i. MLE is always consistent. [5]

      ii. If an estimator δ has constant risk (i.e. $E\{L(\delta, \theta) \mid \theta\}$ does not depend on θ), it is a minimax estimator. [3]

4. Assume that $X_1, X_2, \ldots, X_n$ are i.i.d. Ber(θ).

   (a) Show that $\bar{X} = (X_1 + \ldots + X_n)/n$ is a minimax estimator for θ when the loss function is given by $L(\delta, \theta) = (\delta - \theta)^2/\theta(1-\theta)$. [4]

   (b) Consider the loss function $L(\delta, \theta) = (\delta - \theta)^2$ and check whether $\bar{X}$ is (i) admissible (ii) minimax for θ. [4+4]

5. (a) Show that if the Bayes estimator is unique, it is admissible. [3]

   (b) Suppose that $X \sim f_\theta$, where $\Theta = \{1, 2\}$, $f_1(x) = 1;\ 0 \leq x \leq 1$, and $f_2(x) = 1 + sin(2k\pi x);\ 0 \leq x \leq 1$, for k being a positive integer. Consider the uniform prior on Θ and the 0-1 loss function. Find the Bayes estimator and show that the corresponding Bayes risk does not depend on the value of k. [3+6]

6. (a) Assume that $X_1, X_2, \ldots, X_n$ are i.i.d. $N(\theta, \sigma^2)$, where σ^2 is known. Find the shortest expected length confidence interval of level $1 - \alpha$ for θ. [4]

   (b) Assuming that θ has a prior distribution $N(\mu, \tau^2)$, derive the posterior distribution of μ. Find the highest posterior density credible region of level 1 - α and check whether it coincides with the confidence interval obtained in 5(a) as τ increases to infinity. [3+3+2]

7. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. with p.d.f f_θ. Consider a pair of hypotheses $H_0 : \theta \in [\theta_1, \theta_2]$ and $H_1 : \theta \notin [\theta_1, \theta_2]$.

   (a) If ϕ is the test function of a UMPU test of level α and $E_\theta(\phi)$ is a continuous function of θ, show that $E_{\theta_1}(\phi) = E_{\theta_2}(\phi) = \alpha$. [2]

   (b) Show that if $\phi_1, \phi_2, \ldots, \phi_k$ are test functions of k different unbiased tests of size α, so is $\phi_0 = (\phi_1 + \phi_2 + \ldots + \phi_k)/k$. Check whether ϕ_0^2 is an unbiased test of level α^2. [3+2]

   (c) If $f_\theta(x) \propto \exp\{-(x-\theta)^2/2\}$, can you construct a UMP test for testing H_0 against H_1? Justify your answer. [5]

8. (a) Suppose that we have a single observation X from a distribution with p.d.f. f. Find the MP test of size α for testing $H_0 : f(x) = \frac{1}{\pi}(1+x^2)^{-1}$ against $H_1 : f(x) = \frac{1}{2}e^{-|x|}$. Compute the power of this test. [4+3]

   (b) Suppose that we have a single observation X from a distribution with p.d.f. $g(x) = \beta f_1(x) + (1-\beta) f_2(x)$, where $\beta \in [0,1]$, $f_1(x) = \frac{1}{\pi}(1+x^2)^{-1}$ and $f_2(x) = \frac{1}{2}e^{-|x|}$. Describe how you will construct a UMP test for testing $H_0 : \beta \le 0.5$ against $H_1 : \beta > 0.5$? [5]

9. (a) Show that a UMP test is unbiased. [2]

   (b) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $U(0, \theta)$. Consider a pair of hypotheses $H_0 : \theta = \theta_0$ and $H_1 : \theta \ne \theta_0$. Find the size and power function of the test given by $\phi(X_1, \ldots, X_n) = I\{X_{(n)} \ge \theta_0 \text{ or } X_{(n)} \le \theta_0 \alpha^{1/n}\}$. Show that it is a UMP test of level α for testing H_0 against H_1. [5+5]

10. Suppose that $(X_1, Y_1), (X_2, Y_2) \ldots, (X_n, Y_n)$ are independent observations from a bivariate normal distribution with parameters μ_1, μ_2, σ_1^2, σ_2^2 and ρ.

   (a) Is the MLE of ρ consistent ? [3]

   (b) Construct a likelihood ratio test of level α for testing $H_0 : \rho = 0$ against $H_1 : \rho \ne 0$. [5]

   (c) Describe how you will construct a 95% confidence interval for ρ. [4]

INDIAN STATISTICAL INSTITUTE
First Semestral Examination: 2011–12
Linear Statistical Models

B III

Date: Nov 17, 2011 Maximum marks: 60 Duration: 3 hrs

This paper carries 70 marks. Attempt all questions. The maximum you can score is 60. This is a closed book, closed note exam. You can use your own calculator. In any iterative numerical method, it is enough to do at most 3 iterations, with 4 decimal places.

1. Let V, W be subspaces of $\mathbb{R}^n$. $X \sim N_n(\boldsymbol{\mu}, I)$, where $\boldsymbol{\mu} \in \mathbb{R}^n$. Let X_V, X_W be orthogonal projections X onto V, W, respectively. Prove or disprove the following statement:

   "If for all $\boldsymbol{\mu} \in \mathbb{R}^n$ the random variables $\|X_V\|$ and $\|X_W\|$ have the same distribution then we must have $V = W$."

   [5]

2. This problem is about the pre-midsem attendances of the B-I students of last year. There were 37 students and 5 teachers. Assume that each teacher had equal number of classes. The following tables give some summary statistics about the data.

   Attendances (averaged over teachers) for the students:

   23.6, 13.4, 23.4, 12, 14.4, 23.6, 22.2, 14.6, 19.2, 15.4, 21.8, 17.2, 14.8, 19.8, 23.6, 1.4, 7.4, 9.8, 15, 20.6, 19.8, 23.6, 13.8, 20.4, 20.4, 0, 21.4, 5.8, 22.4, 17.4, 21.2, 20.2, 18.8, 0.2, 10.4, 21.6, 19.4.

   Attendances (averaged over students) for the teachers:

   13.32, 19.84, 21.3, 12.65, 15.32.

   Sum of squares of all attendances is 61792.

   Construct a 1-way ANOVA table with teacher as the single factor. Do you accept the null hypothesis that there is no difference among the teacher (w.r.t. attendance) at 95% level? Now carry out a 2-way ANOVA with student and teacher as the factors (no interaction). Test the same hypothesis once again. Could you predict the conclusion of the second test from the first test? [20]

3. Consider a Poisson regression model

   $$y_i \sim \text{Poisson}(\exp(\alpha + \beta x_i)),$$

   where y_i's are independent, and x_i's are fixed and given. Obtain MLE of α, β based on the following data:

P.T.O

i	1	2	3	4	5
x_i	1.1	2.1	3.0	3.2	3.6
y_i	3	5	6	28	27

Suggest how you may estimate the standard errors of your estimators. [15]

4. Consider the following 2×2 contingency table.

60	77
89	57

Compute the odds ratio. Use this to test independence (asymptotically). Also compute a 95% confidence interval based on the asymptotic distribution for the odds ratio. Justify your steps. [2+4+4]

5. Consider the probit model

$$y_{ij} \sim Bernoulli(\Phi(\alpha_i + \beta_j)),$$

where α_i, β_j are unknown parameters. Suggest a set of identifiability constraints with justification. How many free parameters are there (after applying the identifiability constraints)? Outline how you will estimate the parameters under the constraints. [5+3+7]

6. Consider the mixed effects model

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \ldots, n,$$

where α is a fixed effect, β is a random effect. You are given that $\beta \sim N(0, 2^2)$ and $\epsilon_i \sim N(0, 1)$. All these are independent. The x_i's are fixed and known. How will you find BLUE of α and BLUP of β based on data $(x_1, y_1), \ldots, (x_n, y_n)$? [5]

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2011-12
B. Stat. III Year
Sample Survey

Date: 21.11.11 **Maximum Marks: 40** **Duration: 3 Hours**

Answer any 4 questions, each carrying 10 marks.

NB : Records on Tutorials carrying 10 marks on topics covered after the Mid-Semestral Exam. to be submitted on the date of the Semestral Exam.

1. Discuss with an example the need for sampling in two stages. Present a theory to tackle such a situation pointing out how you may unbiasedly estimate a finite population total along with an unbiased estimator for the variance of your estimator indicating your sample selection procedure.

2. Present Brewer's model-assisted predictor for a survey population total explaining a rationale for its use. How will you measure its accuracy following a design –based procedure ?

3. Explain how you may choose a circular systematic sample with probabilities proportional to size-measures. Explain how you may employ Horvitz & Thompson's procedure in estimating the population total from a sample thus drawn. Indicate how you may unbiasedly estimate variance of the estimate of the population total for more than one such independent samples giving procedures in details.

4. Derive Murthy's ratio-type approximately unbiased estimator for the ratio of two finite population totals based on general replicated samples.

5. Discuss the variance function and the cost function and their uses in planning to tackle a problem of non-response partially in an SRSWOR from a survey population following the approach initiated by Hansen & Hurwitz in estimating the mean of a variable of interest.

6. Given U = (1, 2, 3, 4, 5, 6, 7),

   $\underline{X}$ = (6, 2, 3, 4, 5, 1, 4) with usual notations in systematically drawing a sample of size 3 with probabilities proportional to the elements in $\underline{X}$, calculate the inclusion- probabilities of the unit 4 and of the paired units (5,3).

*****

Indian Statistical Institute, Kolkata
Semestral Examinations (Sem II) : B.Stat. III year
Differential Equations

Full Marks : 70 | November 23, 2011 | Time : 3 Hours

Answer any five questions

1. Let the equation of motion of a mechanical system under an impressed external force be

$$M\frac{d^2x}{dt^2}+c\frac{dx}{dt}+kX=F_0\cos wt.$$

(a) In the absence of the external force (i.e. when $F_0=0$), obtain conditions on $b=\frac{c}{2M}$ and $a=\sqrt{\frac{k}{M}}$ under which the system has an oscillatory motion.
(b) Find the general solution of the inhomogeneous equation.

[7 + 7 = 14]

2. (a) Solve the equations
(i)$(6x+4y+3)dx+(3x+2y+2)dy=0$
(ii)$xy'+y=x^4y^3$

(b) Obtain the general solution of the equation $y''+y=\frac{1}{\cos x}$. [4 + 4 + 6 = 14]

3. (a) Show that the differential equation with the initial condition

$$y'=\phi(x,y),\ y(x_0)=y_0$$

is equivalent to the integral equation

$$y=y_0+\int_{x_0}^{x}\phi(t,y)dt,$$

stating all the necessary assumptions.
(b) It is given that the the two equations with continuously differentiable coefficients:

$$y''+p(x)y'+q(x)y=0,\ \text{and}\ y''+r(x)y'+s(x)y=0$$

have the common solutions $y_1(x)$ and $y_2(x)$ which are linearly independent on a non-trivial interval I. Show that $p(x)=r(x)$ and $q(x)=s(x)$ for all $x\in I$.

[7 + 7 = 14]

4. (a) Show that for the equation $y''+q(x)y=0$, where q is a positive continuous function on $(0,\infty)$, any solution has infinitely many zeros on the positive $X-$axis if $\int_1^\infty q(x)dx=\infty$.
(b) If in (a) $q(x)=\frac{k}{x^2}, x>0$, then show that, for $k>\frac{1}{4}$, any solution has infinitely many zeros on the positive $X-$axis. [7 + 7 = 14]

P.T.O.

5. (a) For the Hermite equation $y'' - x^2y = -(2n+1)y$, show that if y_m is a solution for the parameter value $n = m$, then $(D - x)y_m$ is a solution for the parameter value $n = m + 1$.

(b) Show that, for the Hermite polynomials $\{H_n\}$, the generating function is given by

$$\Phi(x,t) = \exp(2xt - t^2).$$

($H_n(x) = (-1)^n e^{\frac{x^2}{2}} h_n(x)$ and $h_0(x) = e^{\frac{-x^2}{2}}, h_1, h_2, \cdots$ are the successive Hermite functions.)

[7 + 7 = 14]

6. (a) Find a Frobenius solution for the Bessel's equation $x^2y'' + xy' + (x^2 - p^2)y = 0$.

(b) The normal form of the Bessel's equation is

$$u'' + (1 + \frac{1-4p^2}{4x^2})u = 0.$$

Show that if u_p is a solution of this equation, then the gap between successive zeros of $u_p(x)$ converges to π as $x \to \infty$. [7 + 7 = 14]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination (2011-2012)

Subject: Introduction to Sociology
B.Stat. III Year

Date: 25th November 2011 Maximum Marks: **50** Duration: Three hours

The figures in the margin indicate allotted marks

Answer any five of the following questions:

1. Make a critical evaluation of Benoy Kumar Sarkar's concept of progress. 10

2. Explain Auguste Comte's positivism OR the Law of Three Stages of Social Development. 10

3. Discuss the purposes and limitations of Statistics in Sociological Research. 10
OR How does Statistics reconcile the difference between the macros and micro level of social reality ?

4. Write short notes on the following issues: 5 x 2 =10
   a) Observation and experiment.
   b) Globalization.
   c) Domestic Violence.
   d) Nature and scope of sociology of gender.
   e) Racial discrimination

5. Briefly discuss the 73rd and 74th Amendment of Indian constitution, and their impact on the role of women in grass-root politics in West Bengal. 10

6. Define culture. What are the elements of culture? Describe the importance of culture in understanding individual behavior. 2+2+6 = 10

7. Write, in brief, on any two of the following: 2 x 5= 10
   (a) What are the major contrasting features of caste and class?
   (b) What is an ethnic group? How would you distinguish it from a minority group? Give an example of a minority group.
   (c) Analyse the Marxian theory of alienation

INDIAN STATISTICAL INSTITUTE
Semester Examination : 2010-11(First Semester)

Course Name: OPTIONAL COURSE FOR B III

Subject Name : GEOLOGY

Date : 25Jul11--: Maximum Marks 50 : Duration : 2 HOUR

Answer question(Q) number 1 and any TWO from Q2 to Q4

1. How would you interpret with the help of primary sedimentary structures the flow regime of the depositing flow? Explain with a diagram.
What is Reynolds number?

6+2+2=10

2. Distinguish between the following pairs using adequate illustrations

5x4 = 20

i) True dip and apparent dip
ii) Parallel and Similar fold
iii) Normal and Reverse Fault
iv) Reclined fold and Plunging-upright fold

3. Write short notes on 4x5= 20

i) Detrital sedimentary rocks
ii) Maturity of sedimentary rocks
iii) Graded beds
iv) Current ripples
v) Ramsay's classification of folds

4. a) What is a fossil? Explain how fossils help to determine the age of their host rocks. Would you consider the tracks and trails marking the movements of prehistoric animals as fossils? Justify your answer. How would you differentiate – a fossilized brachiopod and a fossilized bivalve shell?
---3+3+1+3 = 10

b) What do you understand by the term Biological Mass extinction? 'Mass extinction events created the path of new forms of lives' -- Justify the statement. Why traces of ancient corals reefs are treated as indicator of the paleo-climate and paleo-latitudinal positions during the time of their formation?. ---------------------3+3+4 = 10

Indian Statistical Institute

First Semestral Examination: (2011–2012)

B.Stat.(Hons.) – III year

Economics III

Date: 25.11.2011 *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) Suppose the model $y = X_{n\cdot(k+1)}\beta + \varepsilon$, in a CLRM set-up, has been misspecified as $y = Z_{n\cdot(r+1)}\delta + u$, where $r<k$ and Z is a subset of X. Then show that the OLS estimator $\hat{\delta}$ is a biased estimator of β, and that the estimator of the residual variance from the misspecified model is biased upwards.

(b) Describe a procedure to test r, $r \le k$, independent linear restrictions of the form $R\beta = d_{r\times 1}$

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

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

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

(d) Consider the model $y_t = \alpha + \beta x_t + \varepsilon_t$, where x is stochastic. Examine the consequences (in terms of unbiasedness and efficiency of the estimators of α and β) of applying the OLS method to estimate the equation in the following situations:

(i) x and ε are independent.

(ii) x and ε are contemporaneously uncorrelated.

(iii) x and ε are contemporaneously correlated.

[12 + 6 + 8 + 9 = 35]

P.T.O.

/

2. (a) Explain what is meant by heteroscedasticity, and describe the consequences of its presence in a regression model.

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

(c) Explain what is meant by 'autocorrelation'.

(d) 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.

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

(f) Consider the following two estimated models:

$$\hat{y}_t = 0.45 - .0041\ X_t \qquad\qquad \hat{y}_t = 0.48 + 0.127\, y_{t-1} - 0.32X_t$$

$$(-3.96) \qquad\qquad\qquad\qquad (3.27) \qquad (-2.17)$$

$$R^2 = 0.5248,\ D.W.\ 0.8252 \qquad\qquad R^2 = 0.8829,\ D.W. = 1.82$$

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? Explain.

[5 + 6 + 2 + 8 + 4 + 10 = 35]

3. (a) What are distributed lag models?

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

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

(d) Describe the Almon polynomial lag model. In an empirical exercise, how does one determine the lag length and the degree of the polynomial? Explain.

[4+ (8+8) + 5 + 10 = 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) Write down the general rules (order and rank conditions) of determining the identification status of a structural equation.

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

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \varepsilon_{1t}$$

$$x_{3t} = \alpha_1 y_t + \alpha_2 z_{1t} + \alpha_2 z_{2t} + \varepsilon_{2t}.$$

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

(e) Describe briefly (mention only the steps involved) a single equation and a system method of estimating a simultaneous equations system, stating the appropriateness of each of these methods in terms of the identifiably status.

[9 + 8 + 6 + 4 + 8=35]

1

INDIAN STATISTICAL INSTITUTE

First Semester Examination [2011-2012]
B.Stat III year

te 25.11.2011

Introduction to Anthropology and Human Genetics

Human Genetics

Maximum marks 40 | Duration 3 hours

Attempt any **FIVE** questions

1. What are different components of population structure? Write the relationship between demographic and genetic structures? 3+5

2. What is Hardy-Weinberg law? Explain its importance in (human) population genetics? 3+5

3. The blood group MN, haptoglobin are examples of genetic traits by a single autosomal locus with two alleles that are codiminant. Derive the allele frequency estimators by Maximum Likelihood Method and by Gene counting method. 5+3

4a. In an investigation of the Rhesus blood types in a population the reactions to C-antiserum and c-antiserum were recorded. This reaction is determined by two co-dominant alleles C and c, and the phenotypic distribution was given below the table. Calculate the allelic frequencies of 'C' and 'c'? 4

CC	Cc	cc	Total
1013	2764	1723	5500

4b. Mutation rate of phenylketonuria (PKU) is 10^{-4} and the incidence in a population is 1 in 10,000. If there is no selection find the incidence of PKU after 100 generations? 4

5. Write briefly about non-random mating? In case a population practices inbreeding derive the allele frequencies deviations from H-W expectation as a result of inbreeding? 4+4

6. Write short notes on **any two** of the following: 4+4
   a. Genetic Diversity
   b. Mutation
   c. Inbreeding Coefficient
   d. Heterozygosity

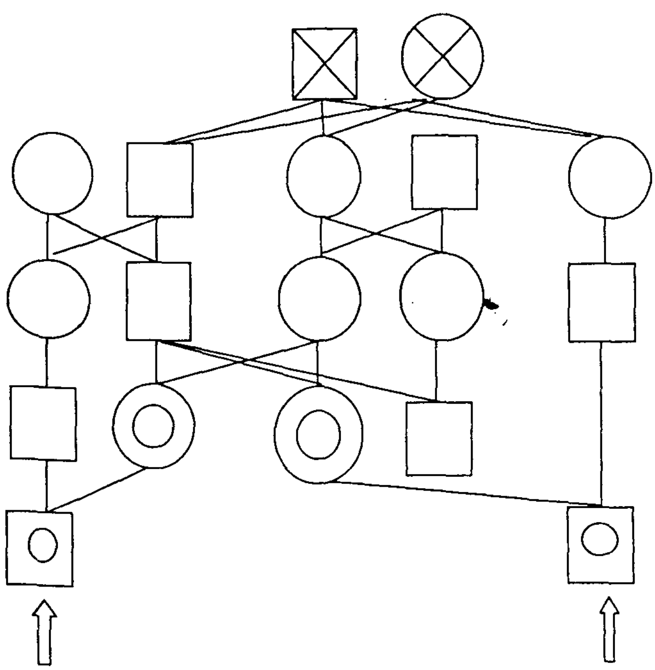
P.T.O

7. Genetic distances (Nei's D_A and F_{ST}) have been calculated among 5 regional subpopulations (PL, MY, PG, KK and PD) of a tribal population based on 15 STR (autosomal) loci (lower matrix) and based on mtDNA (upper matrix). Based on the two Distance Matrices given below construct dendrogram (clustering tree) separately on any method (minimum distance, pair group method or average linkage method etc.): What do you infer from the two trees. 4+4

Table 7: Nei's D_A distance (lower matrix) and F_{ST} distance (upper matrix) of 5 regional subpopulations of a tribe (Upper matrix – mtDNA, Lower matrix – STR autosomal loci).

Sub Population (STR loci)	Sub population (Male) PL	MY	PG	KK	PD
PL	---	1.68	4.31	15.51	15.16
MY	14.48	---	3.15	8.36	17.13
PG	7.23	11.14	----	16.22	20.84
KK	8.46	12.33	4.92	----	31.13
PD	7.31	10.89	3.96	5.0	----

8. During a field study in a tribal population, information on type of marriages between husbands and wives were collected and a five generation pedigree was drawn (given below) showing the consanguineous marriages and the inbred children that involved 20 families. Calculate the inbreeding coefficient of the two inbred children (arrow mark) in the given pedigree? 4+4



B - III

INDIAN STATISTICAL INSTITUTE
First Backpaper Examination: 2011-12
Linear Statistical Models

Date: 23/12/11 Maximum marks: 100 Duration: 3 hrs

This paper carries 100 marks. Attempt all questions. The maximum you can score is 100. This is a closed book, closed note exam. You can use your own calculator. In any iterative numerical method, it is enough to do at most 3 iterations, with 4 decimal places.

1. Let V, W be subspaces of $\mathbb{R}^n$. $X \sim N_n(\mathbf{0}, \sigma^2 I)$, where $\sigma^2 > 0$. Let X_V, X_w be orthogonal projections X onto V, W, respectively. Prove or disprove the following statement:

   If for all $\sigma^2 > 0$ the random variables $\|X_V\|$ and $\|X_W\|$ have the same distribution then we must have $V = W$.

   [5]

2. State the second fundamental theorem of least squares. Use it to test

$$H_0 : \mu = 1 \quad Vs. \quad H_1 : \mu \neq 1$$

   based on $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$, where μ, σ^2 are both unknown. Show that this test is equivalent to the usual t-test. [15]

3. Consider the following contingency table. Asymptotically test the goodness-of-fit of the model of conditional independence of gender and education level given location. Write down the steps clearly.

		Level 1	Level 2	Level 3
Location 1	Male	107	204	50
	Female	240	67	40

   and

		Level 1	Level 2	Level 3
Location 2	Male	121	187	56
	Female	224	127	34

   [20]

4. Suppose that you have a data set:

$$(x_1, Y_1), \ldots, (x_n, Y_n),$$

   where x_i's are fixed and known, and Y_i's are independent Binomial(50, $p(x_i)$) for $i = 1, 2, \ldots, n$. You have two competing generalised linear models to choose from:

$$p(x) = \Phi(\alpha + \beta x)$$

P.T.O.

and

$$p(x) = \frac{1}{1 + \exp(\alpha + \beta x)}, \quad i = 1, 2, ..., n.$$

Suggest some technique (graphical/analytical) to choose one over the other. Justify your answer. [10]

5. Consider a 3-way contingency table with I rows, J columns and K layers. Consider the log-linear model [12][23][13] in the standard notation. Show that under this model all the odds ratios are the same. Is the converse true? Justify your answer. [5+10]

6. Consider the following 2×2 contingency table.

50	67
97	55

Compute the odds ratio. Use this to test independence (asymptotically). Also obtain a 95% confidence interval based on the asymptotic distribution for the odds ratio. Justify your steps. [2+8]

7. Consider the probit model

$$y_{ij} \sim Bernoulli(\Phi(\alpha_i + \beta_j)),$$

where α_i, β_j are unknown parameters. Suggest a set of identifiability constraints with justification. How many free parameters are there under the constraints? Outline how you will estimate the parameters under the constraints. [15]

8. Consider the Gauss-Markov set up

$$Y = X\beta + \epsilon,$$

where $\epsilon \sim (\mathbf{0}, \sigma^2 I)$. Let $E(L'Y) = 0$. Let $K'Y$ be the BLUE of its expectation. Show that $L'K = 0$. [10]

December 27, 2011

B-III

INDIAN STATISTICAL INSTITUTE

Statistical Inference -I

Time : 3 Hours | Semestral (Backpaper) Examination | Full Marks : 100

Answer as many as you can. The maximum you can score is 100.

1. (a) Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x,\theta) = a(x)\theta^x/b(\theta), x = 0, 1, \cdots$, where $0 < \theta < \rho$ for some $\rho > 0$. Find the minimal sufficient statistic and check whether it is complete. [2+2]

   (b) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $U(0,\theta)$. Check whether the following statistics are consistent estimators for θ : (i) $T_1 = \frac{n+1}{n}X_{(n)}$ (ii) $T_2 = X_{(n-1)}$ (iii) $T_3 = X_{(1)}$. [2+4+2]

2. (a) Consider the p.d.f. $f_\theta(x) = \exp^{-x-\theta}(1 + \exp^{-x-\theta})^{-2}; \; -\infty < x < \infty$. Does $\{f_\theta, -\infty < \theta < \infty\}$ belong to an exponential family ? Does it have the MLR property ? [2+2]

   (b) Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$. Find the MLE for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. [2+6]

3. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the discrete uniform distribution with p.m.f. $f_N(x) = 1/N, \; x = 1, 2, \cdots, N$.

   (a) Show that the family $\{f_N, N \geq 1\}$ is complete. [2]

   (b) Find the minimal sufficient statistic and check its completeness. [2+3]

   (c) Find an unbiased estimator for N based on a single observation. [2]

   (d) Using Rao-Blackwellization or otherwise find the UMVUE of N. [3]

   (e) Check whether the family $\{f_N, N \geq 2\}$ is complete. [2]

   (f) Construct a likelihood ratio test for $H_0 : N = N_0$ against $H_1 : N > N_0$. [3]

   (g) Is this a UMP test for $H_0 : N \leq N_0$ against $H_1 : N > N_0$? Justify your answer. [3]

4. (a) Let X be a random variable whose p.m.f. under H_0 and H_1 is given by

x	1	2	3	4	5	6	7
$f(x\|H_0)$	.01	.01	.01	.01	.01	.01	.94
$f(x\|H_1)$	.06	.05	.04	.03	.02	.01	.79

   Find the MP test with size 0.04 and compute the probability of Type II error for this test. [4+2]

   (b) Check whether the family of distributions $\{f_\theta(x) = \frac{\theta}{\pi}\frac{1}{\theta^2+x^2}, \; -\infty < x < \infty, \; \theta > 0\}$ has an MLR. Find the minimal sufficient statistic and check whether its distribution has an MLR. [2+2+2]

   (c) Suppose that X is a random observation from a Cauchy distribution with location parameter θ and scale parameter 1. Assume that we are testing $H_0 : \theta = 0$ against $H_1 : \theta = 1$. For a sample of size one, show that $\phi(x) = 1$ if $1 < x < 3$ is an MP test. Calculate the size and the power of the test. Is this a UMP test for $H_0 : \theta = 0$ against $H_1 : \theta > 0$? Justify your answer. [2+2+2+2]

P.T.C

5. Consider the problem of testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ when we have a random sample of size n from $N(\theta, \sigma^2)$ where σ^2 is known.

   (a) Construct a likelihood ratio test of size α and show that for all $\theta \neq \theta_0$, its power converges to 1 as the sample size n tends to infinity. [3+3]

   (b) Is this test unbiased ? Justify your answer. [3]

   (c) Can you construct an MP test for this problem? Justify your answer. [3]

6. (a) Let $H_0 : f(x) = 1,\ 0 < x < 1$ and $H_1 : f(x) = 4x,\ 0 < x < 1/2$ and $(4 - 4x),\ 1/2 < x < 1$. Obtain MP level α test based on a single observation and compute its power. [3+3]

   (b) Let $f_\theta(x) = \theta(2x) + (1 - \theta)2(1 - x)$ for $0 < x < 1$, $0 \leq \theta \leq 1$. On the basis of a single observation obtain a UMP level α test for testing $H_0 : \theta \leq 1/2$ vs $H_1 : \theta > 1/2$. Find the power function of this test. [4+2]

7. (a) Let $X_1, X_2, \ldots, X_n$ be a sample of size n from $U(0, \theta)$. Find the shortest expected length confidence interval for θ with confidence coefficient $1 - \alpha$. [4]

   (b) Prove or disprove the following statements.

      i. A UMP test, if exists, is unique. [4]

      ii. Minimax estimator, if unique, is admissible. [2]

      iii. If an admissible estimator has constant risk, it is minimax. [2]

8. Suppose that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are i.i.d. observations from a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and known dispersion matrix $\boldsymbol{\Sigma}$. Assume that the parameter space $\Theta = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2\}$ contains only two points. Consider the uniform prior on Θ and the 0-1 loss function.

   (a) Find the minimax estimator for $\boldsymbol{\mu}$. [6]

   (b) Show that its expected risk is a decreasing function of the Mahalanobis distance between the two distributions. [2]

   (c) Show that this expected risk converges to 0 as the sample size n tends to infinity. [2]

INDIAN STATISTICAL INSTITUTE
First Semester Backpaper Examination: 2011-12

B. Stat. III Year
Sample Surveys

Date: 29.12.11 **Maximum Marks 50** **Duration: 3 hrs.**

Answer any 4 questions each carrying the same score.

1 Give two examples of cluster sampling in single stages. Indicate their uses in estimating survey population totals.

2 Explain the principle in forming strata in sampling finite populations. Describe Rao, Hartley & Cochran's (RHC) sampling method. Is it a case of stratified sampling? Gibe reasons for your answer.

3 Giving details clearly bring out the differences among linear regression, generalized regression, difference and generalized difference estimators of a finite population total. How do you measure their accuracy levels ?

4 Developing appropriate theoretical results explain why in two stage sampling with PPSWR sampling in the 1^{st} stage it does not matter in variance estimation if systematic sampling is implemented in the second stage.

5 Show that Brewer's predictor for a finite population total is nothing but a generalized regression estimator for the same.

6 In drawing a circular systematic sample of size 4 from a population of 7 units labelled 1, …., 7 with sampling interval 2calculate the inclusion probability of the unit 3 and of the pair of the units 1 and 5

INDIAN STATISTICAL INSTITUTE
First Semester Backpaper Examination: 2011-12

B. Stat. III Year
Differential Equations

Date: 30.12.11 **Maximum Marks : 100** **Duration: 3 Hours**

1 A chain 4 feet long starts with 1 foot hanging over the edge of a table. Neglect friction, and find the time required for the chain to slide off the table.

[10]

2 Solve the following differential equations

a) $e^y\,dx + (x\,e^y + 2y)\,dy = 0$

b) $x\,\frac{dy}{dx} - 3y = x^4$

c) $y'' + y = \frac{1}{\cos x}$

[4 + 4 + 6 = 14]

3 a) Find the normal form of the differential equation $y'' + P(x)y' + Q(x)y = 0$.

b) Show that any nontrivial solution of the Bessel equation

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

has infinitely many zeros on the positive real axis.

[6 +12 = 18]

4 a) Let y_1 *and* y_2 be linearly independent solutions of the equation

$y'' + p(x)y' + Q(x)y = 0$ on an interval [a,b]. Show that the Wronskian $W(y_1, y_2) = y_1(x)\,y_2'(x) - y_1'(x)y_2(x)$ has no zeros on [a,b].

b) Find a particular solution of the equation $y'' + 4y = \tan 2x$

[8 +10 = 18]

5 Consider the Legendre equation

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

P.T.O

a) Show that $P_l(x) = \frac{1}{2^l l!} \frac{d^l}{d_x l} \ (x^2 - 1)^l$

is a solution of Legendre equation for the parameter value $p = l$, a non-negative integer.

b) Assume the orthogonality relations

$$\int_{-1}^{1} (P_l\,(x))^2 dx = \frac{2}{2l+1} \quad , \quad \int_{-1}^{1} P_{l'}(x) dx = 0$$

For l, l' nonnegative integer and $l \neq l'$. Given a continuous function f on [-1, 1], find the polynomial Q_n of degree less than or equal to n which minimizes the integral

$\int_{-1}^{1} (f\,(x) - \,Q_n\,(\text{x}))^2 dx$

[8 +8 = 16]

6 a) Obtain the Frobenious solution J_p of the Bessel's equation.

b) Show that

i) $\frac{d}{dx} \left[x^p \, J_p(x)\right] = \; x^p J_{p-1}\,(x)$

ii) $J_{p-1}\,(x) + \; J_{p+1}(x) = \; \frac{2p}{x} \; J_p\,(x)$.

[10+ 5 +5 = 20]

7 State the fundamental existence and uniqueness theorem for the ordinary differential equation $\frac{dy}{dx} \; = \; \varphi\ (x, y)$

[4]

✿✿✿✿✿

INTRODUCTION TO STOCHASTIC PROCESSES
B. STAT. IIIRD YEAR SEMESTER 2
INDIAN STATISTICAL INSTITUTE

Mid-semestral Examination
Time: 2 Hours Full Marks: 35
Date: February 20, 2012

This is an OPEN NOTE examination. You can use **any note written in your own handwriting**, but you are not allowed to share them. Printed, typed, photocopied, carbon copied, cyclostyled notes are not allowed.

1. Give examples (with justification) of Markov chains with following properties:
   (a) all states are inessential;
   (b) has no stationary distribution;
   (c) has infinitely many linearly independent stationary distributions.

   [3 × 2 = 6]

2. Prove the following identity, where the symbols have their usual meaning:
$$f_{ij}^{(n)} = \sum_{k=0}^{n} {}_j p_{ii}^{(k)} \, {}_i f_{ij}^{(n-k)}.$$
[6]

3. Consider the simple random walk $\{X_n\}$ on $\mathbb{Z}$ starting at 0 where you move one step right with probability $\frac{1}{2}$ and one step left with probability $1/2$. Define $R_n = |X_n|$. Show that $\{R_n\}$ is a Markov chain and obtain the transition matrix. [6]

4. Say that a state i has property $\mathcal{P}$ if for some invariant measure ν, we have $\nu_i > 0$. Show that $\mathcal{P}$ is a solidarity property. [6]

5. The lone barber in the barber shop near the hostel takes T minutes to serve a customer, where T follows Geometric distribution with parameter $\frac{1}{4}$. At most one customer arrive at the shop at the end of every minute with probability of arrival being $\frac{3}{4}$. The shop has 3 chairs for the waiting customers. If all chairs are occupied, any new arriving customer is turned away. Once the barber finishes serving a customer, he starts serving a new one if anybody is waiting and if nobody is waiting, as soon as one arrives. All the arrivals and service times are independent of each other.

   Let X_n be the number of customers in the shop (including the one being served, if any) at the end of nth minute. Show that $\{X_n\}$ is a Markov chain and calculate its transition matrix. [6]

   For $i = 0, 1, 2, \ldots, 4$, let $u(i)$ be the probability that starting with i customers (including the one being served), the shop will be empty before it has reached its capacity. Observe that $u(0) = 1$ and $u(4) = 0$. Show that, for $i = 1, 2, 3$,
$$u(i) = \frac{1}{16}u(i-1) + \frac{3}{8}u(i) + \frac{9}{16}u(i+1).$$
[3]

   Define $v(i) = u(i) - u(i-1)$ for $i = 1, 2, 3, 4$ and obtain a recurrence relation for v. Solve for u and v. [4]

Midterm Examination
Statistical Inference II
B. Stat. Third Year
Second Semester
2011-2012 Academic Year

Date : 23.02.12 Maximum Marks: 40 Duration : 2 hours

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

1. (a) Let $X_1, \ldots, X_n$ be iid observations from some unknown distribution F. We want to test $H_0 : F(x) = F_0(x)$ for all $x \in \mathcal{R}$ versus $H_1 : F(x) \geq F_0(x)$ $\forall x \in \mathcal{R}$ with $F(x) > F_0(x)$ for at least one $x \in \mathcal{R}$. Here $F_0(x)$ is a completely known distribution function (not necessarily continuous). Suggest the appropriate Kolmogorov-Smirnov test for this problem and show that the test is consistent. [1+4=5]

   (b) Let $X_1, \ldots, X_n$ be iid with a distribution $F \in \mathcal{F}$, where $\mathcal{F}$ is the class of continuous strictly increasing distributions on the real line which are symmetric around zero. Let F_n be the empirical distribution function of the original sample and $\bar{F}_n$ is that based on $-X_1, \ldots, -X_n$. Show that the distribution of $\tilde{D}_n = \sup_x |F_n(x) - \bar{F}_n(x)|$ is independent of F as long as $F \in \mathcal{F}$. Can $\tilde{D}_n$ be used to test for some aspect of the distribution F ? Explain your answer. [4+1=5]

2. Let $X_i, i = 1, \ldots, n$ be iid F where $F(x) = G(x - \delta)$ for all $x \in \mathcal{R}$, where $G(.)$ is a continuous distribution symmetric around zero and δ is unknown.

   (a) Write down the Hodges-Lehmann estimator $\hat{\delta}$ for δ. [1]

   (b) Show that the distribution of $\hat{\delta}$ is symmetric around δ. [6]

3. Suppose $X_1, \ldots, X_m$ are iid with a distribution F and $Y_1, \ldots, Y_n$ 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.

   (a) Consider first the case when $m = n$. 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. [2+4=6]

   (b) In the general case, with m and n arbitrary, derive the formula for a $100(1 - \alpha)\%$ confidence interval for the unknown θ, where $\alpha \in (0, 1)$. [5]

4. Suppose $X_1, \ldots, X_m$ are iid with distribution F and $Y_1, \ldots, Y_n$ 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 $x \in \mathcal{R}$, for some unknown constant $\theta \in (-\infty, \infty)$.

(a) 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. [5]

(b) Describe the van der Waerden's 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+3=5]

(c) Consider the case when $m = n$ and H_0 is true and you are using the two-sided or one-sided two-sample Kolmogorov-Smirnov test procedure. Find the value of $P(D_{6,6} < 0.5)$. Prove that $P(\sqrt{\frac{n}{2}}D^+_{n,n} > \lambda) \to e^{-2\lambda^2}$ as $n \to \infty$ where $\lambda > 0$. [2+4=6]

INDIAN STATISTICAL INSTITUTE

Mid-Sem Examination, 2^{nd} Semester, 2011-12

Statistics Comprehensive, B.Stat 3^{rd} Year

Date: February 27, 2012 Time: 2 hours

This is an open notes examination. Answer any four questions. This paper carries 40 marks.

1. (a) If the maximum of a set of observations is increased and the minimum is decreased, explain whether the mean deviation about the mean will necessarily increase.

   (b) Consider a set of bivariate observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where x and y do not have a perfect linear relationship. Suppose two new observations (x_{n+1}, y_{n+1}) and (x_{n+2}, y_{n+2}) are introduced in the dataset. Obtain a necessary and sufficient condition on these two observations such that the least squares regression lines of y on x and of x on y both remain unchanged with the introduction of the two new observations. [5+5]

2. (a) Using a random number generator from $U(0, 1)$, explain, with suitable justification, how you would simulate two observations such that each of them is distributed as chi-squares with 3 degrees of freedom and the correlation coefficient between them is 0.36.

   (b) If $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is a random sample from a bivariate normal distribution with parameters $(\mu, -\mu, \sigma^2, \sigma^2, \rho)$, obtain the m.l.e.s of μ, σ^2 and ρ. Examine whether the m.l.e of ρ is consistent for ρ. [5+5]

3. (a) Based on a random sample $X_1, X_2, \ldots, X_n$ from an exponential distribution with mean λ, obtain the UMVUE of $e^{-1/\lambda}$.

   (b) Suppose X is an observation from a Geometric distribution with parameter p. Construct a UMP test of size 0.05 for $H_0 : p \geq 1/3$ vs $H_1 : p < 1/3$. [5+5]

4. (a) Suppose $\mathbf{x}$ is a random vector distribured as $N_p(0, \sigma^2 I)$. If A is an idempotent symmetric matrix with rank r $(< p)$, obtain the distribution of $\dfrac{\mathbf{x}'A\mathbf{x}}{\mathbf{x}'\mathbf{x}}$

(b) Given India's recent string of poor performances in test cricket, a survey was carried out in four metros as to whether the team requires immediate restructuring. The data obtained were as follows:

Metro	# interviewed	# positive responses
Mumbai	1300	1015
Delhi	1500	1186
Kolkata	1200	938
Chennai	1400	1079

Do the data provide evidence that the opinion is similar in the four metros? [4+6]

5. (a) Consider a linear model $(Y, X\beta, \sigma^2 I)$.

   i. Suggest a suitable error model such that the LAD estimator of β is also its m.l.e.

   ii. Suppose the model has an additional constraint $A\beta = 0$. Show that a parametric function $p'\beta$ is estimable in this model if and only if there exists a vector l such that $X'l - p \in \mathcal{C}(A')$.

   (b) Suppose a population comprises k varieties of objects with N_i objects of the i^{th} variety, $i = 1, 2, \ldots, k$. If n (≥ 2) objects are drawn randomly from the population without replacement, compute the correlation coefficient between the proportions of two distinct varieties in the sample. [6+4]

INDIAN STATISTICAL INSTITUTE
Mid-semestral Examination: 2011-12
B.Stat 3rd year Design and Analysis of Experiments

29 February, 2012 Time: Two hours

Answer all questions, keep your answers brief. Maximum you can score is 40

1. Answer True or False and give a BRIEF justification for your answer in each case.

   a) For a connected block design with v treatments, rank of the estimation space is $v-1$.

   b) Any incomplete binary block design may or may not be an orthogonal design.

   c) If all pairwise treatment contrasts are estimable from a block design then all treatment contrasts are estimable.

   d) A randomized block design with 5 treatments will remain connected even if *any* four observations are missing.

   e) For a connected block design and under the usual model, the BLUE of a contrast $l'\tau$ is of the form $\lambda' Q$, where λ is a $v \times 1$ vector, and Q is the vector of adjusted treatment totals as usual. $[5 \times 2 = 10]$

2. Let d be any connected and orthogonal block design with b blocks. Suppose a new design d^* is constructed from d where the jth block of d^* consists of all treatments in the jth block of d repeated twice, $j = 1, \ldots, b$. Will d^* be connected? Will d^* be orthogonal? [3+3=6]

3. For a connected block design prove that $(C + \mathbf{1}\mathbf{1}')^{-1}$ exists and that it is a generalized inverse of the C matrix, where $\mathbf{1}$ is the vector with all elements unity. [3+3=6]

4. An experiment is planned with 7 treatments, and 21 experimental units are available. Suppose these 21 units are heterogeneous with respect to two factors and can be classified into 3 rows and 7 columns such that units in a row (or column) are homogeneous.

   a) Write down a model for analyzing data from such a row-column design.

   b) Suppose the following row-column design is planned with 3 rows and 7 columns, the treatments being labeled $1, 2, \ldots, 7$

7	1	2	3	4	5	6
1	2	3	4	5	6	7
3	4	5	6	7	1	2

   Check if all treatment contrasts will be estimable from the above design. (You may assume the algebraic expression of the C- matrix) [3+5=8]

5. Construct GF(9). [Hint: Show that: x^2+x+1 is a factor of the cyclotomic polynomial]. Hence construct 2 mutually orthogonal Latin squares of order 9. [5+5=10]

6. BONUS QUESTION: Let A be an orthogonal array OA(24, 13, 2, 2) with 2 symbols, +1 and -1, written as a 13×24 array. Define $X = [\mathbf{1} \;\; A']$, where $\mathbf{1}$ is the vector with all elements unity. Obtain $X'X$. [3]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2011-12

Course Name: B. Stat. III Year

Subject Name: Database Management Systems

Date: 02/03/2012 **Maximum Marks: 40** **Duration: 3hours**

1. Consider the following information for a University database:
   a. Every teacher has a unique teacher no. a name, a designation and an area of specialisation.
   b. Each teacher is associated with only one department.
   c. A department has a unique name and address consisting of a unique building name and floor no.
   d. One of the teachers belonging to a department serves as its Head.
   e. University runs various projects sponsored by different funding agencies.
   f. Every project has a unique project no. sponsor name, a starting date, a completion date and a budget.
   g. Each project has a principal investigator who is one of the teachers of the university.
   h. Few other teachers may also be associated with each project as member.
   i. Teachers from other University may also be associated with each project as external member.
   j. External members name along with their affiliation is unique.
   k. A teacher can manage and / or work on multiple projects.
   l. Each project is assigned to the department where its principal investigator works. The other members of the project may, however, belong to different departments.
   m. Each research scholar of the university is either a teaching assistant or a research assistant.
   n. A teaching assistant is associated with the department of his / her supervisor.
   o. A research assistant may, however, be associated with a project which may not belong to the department where his / her supervisor is working.
   p. Research assistant is allowed to work in only one project.
   q. Each scholar has a unique roll no, name and a year of enrolment.

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

Introduction to Stochastic Processes

B. Stat. Third Year Semester 2

Indian Statistical Institute

Semestral Examination

Time: 3 Hours   Full Marks: 50

Date: April [illegible], 2012

This is an OPEN NOTE examination. [illegible] notes that you have brought with you [illegible] printed, photocopied, [illegible] allowed. [illegible] not allowed.

1. Consider [illegible] with [illegible] (0, 1) and $p_{\dots}$ [illegible] q_i [illegible] $\sum$ [illegible] Show that P is irreducible. Show that [illegible]
Now, further assume [illegible] Show that P is positive recurrent iff $\sum$ [illegible] $\prod$ [illegible] [2+[illegible]]

2. If the chain $\{X_n\}$ [illegible] the chain $\{X_n, X_{n+1}\}$ is also Markov chain [illegible] [illegible]

3. The [illegible] during the exam. Occurrences [illegible] process of intensity [illegible] per hour [illegible] Poisson process [illegible] false before [illegible] What [illegible] on its first visit? What [illegible] [illegible]

4. Suppose $\{X_n : n \geq 0\}$ [illegible] Markov chain on the state space $S = \{1, 2, \dots, m\}$ [illegible] the chain is ergodic. Let the stationary distribution be π [illegible] vector with all entries 1 and [illegible] Show that π [illegible] [illegible]
(Hint: Use Kolmogorov [illegible])

5. ST's cricket [illegible] IPL [illegible] among three states: 1 (completely fit), 2 [illegible] and 3 (major injury causing total absence [illegible] of the team, who has done a course in Stochastic Processes [illegible] can be modelled by a continuous time Markov chain [illegible] given by the matrix

$$\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix}$$

and mean holding times in each [illegible] days respectively. What will be the long run proportion of days that ST [illegible]?
NA is willing to pay ST [illegible] in states 1, 2 and 3 respectively. Based on ST's long run [illegible] will NA be willing to bid for ST in the auction? [3+4=7]

6. Consider an M [illegible] inter-arrival times have mean $1/\alpha$ and service times [illegible] number of customers being served at time t. Show that [illegible] Show that its stationary [illegible] [3+4=7]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination: 2011-12

Course: B.Stat. – III (Second Semester)

Subject: Database Management

Date:25/04/2012 Maximum Marks: 60 Duration: 3hr

Answer all questions

1. The railway network of a country runs three types of trains - Local, Passenger and Express. An Express Train is also of two types – Ordinary and Superfast. Each Train is identified by a name and a number, both of which are unique. For each train the originating station and the final destination are known. Each station of the railway network has a unique name. The number of platforms available in the station is also known. Each station knows the arrival and departure time of each train stopping at that station. Each station also maintains for each train, the name and distance of the previous and next station on the route where the concerned train stops. Each station has multiple facilities like: retiring room, tea stall, drinking water, refreshment room etc. All stations may not have all the facilities.

   Draw an ER/EER diagram and map to appropriate relations.

   10+10=20

2. Consider the following relation:

   car_sales (car_id, date_sold, salesman_id, salesman_commission_percentage, discount_amount)

   Assume a car may be sold by multiple salesmen, and thus (car_id, salesman_id) is the primary key. Additional dependencies are:

   date_sold → discount_amount

   salesman_id → salesman_commission_percentage

   Based on the given primary key, and dependencies is the relation in

   a) 1NF,

   b) 2NF,

   c) 3NF or

   d) BCNF?

   Justify your answer.

   e) *If necessary, generate the normalized relation(s) free from partial and transitive dependencies.*

   $4 \times 2.5 + 15 = 15$

3. a) Explain the desirable properties of a transaction.

   b) What transaction related information are to be stored in the log file for the purpose of crash recovery?

   c) Compare binary locks to exclusive/shared locks. Which type of lock is preferable? Justify your answer.

$2 + 3 + 3 = 08$

4. Consider the following schema where parts may be of multiple colour:

   Supplier_master(<u>sid</u>, sname, address)

   Parts_master(<u>pid</u>, pname)

   Colour(<u>cid</u>, colour_name)

   Parts_colour(<u>pid</u>, <u>cid</u>)

   Catalog(<u>sid</u>, <u>pid</u>, price)

   *Write the following queries in **Relational Algebra** and in **SQL**.*

   a) Find only the "RED" coloured parts name.

   b) Find multicoloured "RED" and "YELLOW" parts name.

   c) Find the name of the suppliers who supply only multicoloured parts.

$(5 + 3 + 3) + (3\times 2) = 17$

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2011-12
B.Stat 3rd year Design and Analysis of Experiments

27 April , 2012 Time: Three hours

Answer all questions. Keep your answers brief. Maximum you can score:60

1. a) Give an example of an orthogonal block design and one example of a non-orthogonal block design, both designs being connected with 4 treatments. Justify your answer.

   b) Consider a block design d with 4 treatments. If the C-matrix of d has rank 3, show that all treatment contrasts are estimable in d.

   c) Consider a single replicate of a 2^3 experiment with 2 blocks of size 4 each. Can there be more than one factorial effect confounded with the blocks? Justify your answer.

   d) Consider a single replicate of a 3^2 experiment in factors A and B and having two blocks as given by:

   $$\text{Block 1} : 00,\ 12,\ 21,\ 22, \quad \text{Block 2} : 01,\ 02,\ 10,\ 11,\ 20.$$

   Examine whether all contrasts belonging to the component AB^2 of the 2-factor interaction are estimable in this design.

   e) For a 2^4 experiment, write down the contrasts belonging to the main effect of the first factor and the four-factor interaction effect. If the underlying design for this experiment is unconfounded, then give the expressions for the sums of squares for these two effects.

   [$5 \times 3 = 15$]

2. (a) Define a set of mutually orthogonal Latin squares with 8 symbols. Define an orthogonal array OA(64, 8, 8, 2).

   (b) At most how many squares can there be in a set of mutually orthogonal Latin squares of order 8? Justify your answer.

   (c) Indicate the construction of 2 mutually orthogonal Latin squares of order 8.

   (d) If you want to construct an orthogonal array OA(64, k, 8, 2) what is the maximum value of k that you can accommodate? State, without proof, any result you use.

   [$4 \times 3 = 12$]

3. An experiment with 5 treatments is carried out in a Latin Square design.

   (a) Write down the usual model for analyzing the data from this experiment.

   (b) Suppose, by mistake, the experimenter forgot to record the observation corresponding to treatment 1 in row 5 and column 5, and all other 24 observations are available. Write down the model for the available observations.

   (c) Estimate the missing observation from the available data in such a way that if we use this estimated value in place of the missing observation and then analyze the augmented data using the model in (a) above, then the resulting normal equations and error sum of squares will be the same as under the model in (b).

(d) Prove that your estimate indeed satisfies the requirement of (c) above.

(e) Suppose that the available observations are augmented with your estimated value in place of the missing value and analyzed using model in (a) above. Now, if the F-test rejects the null hypothesis of equality of treatment effects, what will be your conclusion? (state,without proof, any relevant result you use in support of your answer)

[1+1+4+4+2=12]

4. (a) Give an example of an experimental situation where one factor is used as a blocking factor, one factor as a covariate and there is one treatment factor of interest.

(b) Consider the model $E(Y) = X\beta + Z\gamma$ where Y is the response vector, X is the $n \times p$ design matrix and Z is an $n \times k$ matrix with its ith column representing the n observations on the ith covariate, $i = 1 \ldots, k$. Under suitable assumptions on the ranks of X and $(X\ Z)$, answer the following:

(i) Let β_0 be the least squares solution of the normal equations ignoring the covariates in the above model. Using suitable notations, obtain an estimate for β under the above model, showing the adjustments to be made due to the covariates.

(ii) Obtain an unbiased estimator of γ and an expression for its dispersion matrix.

(iii) How will you test the hypothesis $H_0 : \gamma = 0$ against $H_1 : \gamma \neq 0$? [$4 \times 3 = 12$]

5. (a) An experiment with 3 factors A, B, C, each at 3 levels is to be carried out in 4 replicates, each replicate consisting of blocks of size 9 each. Suggest suitable factorial effects that may be confounded in the different replicates so that the loss of information on each main effect and on each two-factor interaction is zero.

(b) Give the treatment combinations in the principal blocks of each replicate.

(c) Prove that in your design in (a), no main effect or two-factor interaction is confounded. What is the loss of information on each component of the 3-factor interaction compared to an unconfounded design?

(d) For a 3^3 factorial experiment, give a suitable confounding scheme in 4 replicates if blocks of size 3 are to be used and no main effect is confounded in any replicate and each two-factor interaction component is confounded in 2 replicates. (Actual composition of blocks not needed.) [$4 \times 3 = 12$]

INDIAN STATISTICAL INSTITUTE
Second Semestral Examination : 2011-12
B. Stat. Third Year
Statistical Inference II

Date : 30.4.2012 Maximum Marks: 60 Duration :- 3 hours

Answer all questions

1. (a) Stating appropriate assumptions, prove Wald's ineqalities 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 the joint density of $(X_1, \ldots, X_m)$ be $p_{jm}(x_1, \ldots, x_m)$ under hypothesis H_j, $j = 0, 1$. Let $0 < B < 1 < A < \infty$ and consider the SPRT of strength (α, β) that stops first time when $\lambda_m = \frac{p_{1m}}{p_{0m}} \geq A$ or $\leq B$. 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. Note that we are not assuming that the X_i's are iid and it is not guaranteed that the SPRT terminates with probability 1 under H_0 or H_1. [6+4=10]

2. Stating appropriate assumptions, prove Wald's First Equation in the context of sequential analysis. Using this, find expressions for the average sample size needed by SPRT to reach a decision (under both the null and the alternative) in a simple vs simple testing situation . [7+7=14]

3. Consider the $N(\mu, \sigma^2)$ distribution, where both μ and σ are unknown. Let $\alpha \in (0, 1)$. Under a suitable sampling scheme, derive a confidence interval of confidence coefficient at least $1 - \alpha$ for μ, of width at most $2l$ where $l > 0$. Jusify your answer. Can you propose an unbiased estimate of μ of variance at most 1 ? Prove your answer. [6+4=10]

4. Write expressions for the Jackknife estimator of bias of an estimator and Jackknife bias-adjusted estimator. Can it be expected that in large samples the Jackknife bias-adjusted estimator will reduce the order of bias of the original estimator ? Explain your answer. [2+1+3=6]

5. Let $X_1, \ldots X_n$ be iid $F(x - \theta)$ where F is a continuous distribution having a continuous density f which is symmetric around 0 and $f(0) > 0$ and $Var_F(X_i) = \sigma_F^2 < \infty$. We want to test $H_0 : \theta = 0$ versus $H_1 : \theta > 0$. Derive a general expression for the ARE of the sign test relative to the t-test in this problem and evaluate it when $F = N(0, 1)$ and F is the double exponential distribution with density $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$. [10]

6. Consider $X_1, \ldots, X_n$ iid with distribution F where $F \in \mathcal{F}$, $\mathcal{F}$ being some class of distributions.

(a) Define the U-statistic U_n for unbiased estimation of $\theta = \theta(F)$ based on a symmetric kernel $h(x_1, \ldots, x_m)$.

(b) Can the asymptotic distribution of U-statistic be used to derive the asymptotic null distribution of the Wilcoxon Signed Rank test statistic? Explain your answer. You assume that X_i's are iid from a continuous distribution F, symmetric around its unknown (unique) median θ and the null hypothesis specifies θ as 0.

(c) Suppose now that $X_1, \ldots, X_n$ are iid Bernoulli($\frac{1}{2}$). Find a sequence a_n and a constant b such that $a_n(s_n^2 - b)$ converges to a non-degenerate limit distribution. Explain your answer. Here $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. [1+6+3=10]

INDIAN STATISTICAL INSTITUTE
Backpaper Examination
Second Semester 2011-12
B. Stat. Third Year
Statistical Inference II

25.6.12

Maximum Marks: 100 Duration :- 3 hours

Answer all questions

1. Show, stating appropriate assumptions, that the null distribution of Kolmogorov-Smirnov one-sample goodness of fit test criterion is indepndent of the assumed null distribution from which the data are generated. [12]

2. Consider the one sample problem: $X_1, \ldots, X_n$ 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 prolem. [12]

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 $k = 0, 1, \ldots, n$, where under H_0 the two distributions are identical. [15]

4. State and prove Stein's lemma about the termination property of an SPRT. [15]

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 the concept of Asymptotic Relative Efficiency of tests due to Pitman. [14]

INDIAN STATISTICAL INSTITUTE

Back paper Examination: 2011-12

B.Stat 3rd year Design and Analysis of Experiments

26.06.2012 Time: Three hours

Answer all questions, keep your answers brief. **Total marks:100**

1. (a) State and prove a necessary and sufficient condition for a connected block design to be orthogonal.

   (b) For a connected block design prove that $(C+\mathbf{1}\mathbf{1}')^{-1}$ exists and that it is a generalized inverse of the C matrix, where $\mathbf{1}$ is the vector with all elements unity.

   (c) Explain the difference between covariates and blocking factors in a block design.

   (d) What is meant by partial confounding in the context of factorial experiments and why is it useful?

   (e) In the context of a 3^3 factorial experiment, obtain the expressions for the contrasts belonging to the different components of the three factor interaction effect.

   $[5 \times 4 = 20]$

2. (a)Construct GF(9).

   (b) Give a method for constructing a set of mutually orthogonal Latin squares of order 9 using the elements of GF(9).

   (c)Construct 2 mutually orthogonal Latin squares of order 9 .

   (d) Hence indicate how you will construct an orthogonal array OA(81, 10, 9,2).

   $[4 \times 5 = 20]$

3. a) Explain the advantages of using a factorial experiment compared to separate varietal experiments, one for each factor.

   b) You are to plan an experiment to study the effect of 2 different culture-mediums and 3 different times on the growth of a particular virus. Each day, six observations can be taken under identical conditions and the experiment is to continued for 3 days. Describe a design for this factorial experiment.

   c) Write down the model for analyzing data from this experiment.

   d) In this context, write down a contrast belonging to main effect 'culture-medium' and a contrast belonging to the interaction effect of 'culture-medium' with 'time'.

   e) The experimenter says that each day, he will first prepare any one type of the culture-medium and then take observations with it for the three different times, then prepare the second type of culture-medium and take observations with it as before. If observations are collected this way, will the model in (b) above be applicable? Justify your answer. $[5 \times 4 = 20]$

4. a) Construct a 2^4 experiment in 4 replicates, each replicate consisting of 4 blocks of size 4 each, so that some information is available on all the effects with no loss of information on main effects.

P.T.O.

b) Obtain the loss of information on the different effects in the design constructed in (a) above.

c) Construct a 2×4 factorial experiment in 4 replicates, each replicate consisting of 4 blocks each of size 2, such that there is no loss of information on main effects.

[7+5+8=20]

5. a) In a Randomized block design with 5 treatments and 8 blocks, the observation from block 2 corresponding to treatment 4 is lost. Estimate this missing observation from the available data.

b) An experiment with 2 factors, each at 3 levels, is to be planned. Each day 9 observations are to be taken and the experiment is to be conducted for 2 days. Suggest a suitable design for this experiment.

c) Suppose it is found that each day only 3 observations can be taken but the experiment can be continued for 6 days. Suggest a suitable design for this experiment. What is the loss of information on the different effects from this design compared to the one in (b) above? [5+5+10=20]

b) Obtain the loss of information on the different effects in the design constructed in (a) above.

c) Construct a 2×4 factorial experiment in 4 replicates, each replicate consisting of 4 blocks each of size 2, such that there is no loss of information on main effects.

[7+5+8=20]

5. a) In a Randomized block design with 5 treatments and 8 blocks, the observation from block 2 corresponding to treatment 4 is lost. Estimate this missing observation from the available data.

b) An experiment with 2 factors, each at 3 levels, is to be planned. Each day 9 observations are to be taken and the experiment is to be conducted for 2 days. Suggest a suitable design for this experiment.

c) Suppose it is found that each day only 3 observations can be taken but the experiment can be continued for 6 days. Suggest a suitable design for this experiment. What is the loss of information on the different effects from this design compared to the one in (b) above? [5+5+10=20]

Introduction to Stochastic Processes
B. Stat. IIIrd Year Semester 2
Indian Statistical Institute

Backpaper Examination

Time: 3 Hours Full Marks: 100

Date: June 28, 2012

This is an OPEN NOTE examination. You may use any handwritten notes that you have brought with you, but you are not allowed to share them. Also no notes in printed, photocopied, cyclostyled or any other form (including electronic storage) is allowed. Calculators and any other computing devices are strictly not allowed.

1. Recently a sudden increase in the number of extra-curricular activities among the students has been observed. However, due to the availability of training in diverse extra-curricular forms, the students have a wide variety of choice, which may be divided into three categories: sports (state 1), music (state 2) and drama (state 3). Spoilt for a choice, it is difficult for the student to make up her mind and each day, the interest of a student fluctuates among these three choices. Further, due to the unfortunate requirements of the exam, the student also has to occasionally devote time to study (state 4). However, due to the severe stress associated with the studies, the student has to move to one of the extra-curricular activities at the first possible opportunity. These succession can be modelled as a Markov chain with transition matrix

$$\begin{pmatrix} 2/5 & 1/4 & 1/4 & 1/10 \\ 7/30 & 1/3 & 1/3 & 1/10 \\ 7/30 & 1/3 & 1/3 & 1/10 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix}.$$

If the student starts with music, what is the probability that she will study before going to sports? If the student starts with sports, what is the probability that she will attend music session before studying? [10+10=20]

2. Let P be an idempotent irreducible transition matrix (not necessarily with finitely many states). Show that P is necessarily aperiodic. Show that P must have all rows same. Find a stationary distribution for P. [6+7+3=16]

3. **Renyi traffic model.** At time 0 cars are positioned along an infinite highway, modelled by the entire real line, according to a Poisson process with rate α. Assume the initial position of the nth car is X_n. Each car chooses a velocity $V_n (\in (-\infty, \infty))$ independently and identically of all other cars and proceeds to travel at that fixed velocity. (A negative velocity means the car is moving to left and a positive velocity denotes a movement to right.) Assume collisions are impossible; if necessary, cars pass through each other! Let $N(\cdot)$ be the point process of the positions at time 10. Obtain the distribution of N.
Let T_n be the time the nth car passes through 0. Show that $\sum \delta_{T_n}$ is also a Poisson point process and find its mean function. [10+10=20]

4. The Death Eaters arrive at the shop Borgin and Burkes' in Knockturn Alley for trading in dark objects according to a Poisson process with intensity λ. When a Death Eater arrives, Mr. Borgin disappears behind the counter with his customer to finalize the illicit deal. The deal takes a random amount of time with mean μ, to finish. All customers arriving in that duration are lost. What fraction of customers are lost in the long run? [10]

5. A system has m machines and one mechanic to repair the faulty machines. The machines work for an Exp(a) duration before breaking down and the mechanic requires an Exp(b) amount of time to fix the broken machines. The life times of the machines and the repair times are independent. Show that the number of machines working at time t form a continuous time Markov chain. Find its stationary distribution. [5+7=12]

6. Consider an $M/M/s$ queue, where the inter-arrival times have mean $1/a$ and service times have mean $1/b$. Find out the necessary and the sufficient conditions for the queue length to be stable.
Now, further assume that $s = 1$, that is, there is only one server and also the buffer size is 1. Find out the expected time required to turn away a customer starting from an empty system. [10+12=22]