INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2013 – 14

B. Stat (3rd Year)

Linear Statistical Models

Date: 26 August 2013 Maximum Marks: 30 Duration: 1½Hours

1. For the linear model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$ with $Var(y_i) = \sigma^2 x_{1i}^2$, for i = 1,…, n and such that $x_{1i} \neq 0$ for all i
   (a) Derive the best estimators for $\beta_0, \beta_1, \beta_2$ *and* σ^2 . Find the variance of $(\beta_0, \beta_1, \beta_2)'$.
   (b) If by mistake we assume homoscedasticity, what will be the properties of the estimators (expectation and variance). [(4 + 3) + (3 + 4) = 14]

2. In the context of a linear model with usual notation, obtain the Confidence Interval for C$\boldsymbol{\beta}$ from the corresponding F-statistic. [6]

3. Consider the agricultural experiment with output(y) affected by fertilizer A in three varieties {A1, A2, A3} and pesticide B in two varieties {B1, B2} with each combination being applied on two plots each. We are also told that there is no pure interaction effect between A1 and B1.

   Write the linear model corresponding to this experiment, stating all relevant assumptions.
   State a complete set of estimable functions for this model.
   Find the normal equations for the above model and the estimators.
   [3 + 3 + 4 = 10]

INDIAN STATISTICAL INSTITUTE
Mid- Semester Examination: 2013-14
B. Stat. III Year
Sample Survey

Date: 30.08.2013 **Maximum Marks: 50** **Duration: 3 Hours**

Answer any 4 questions, each carrying 10 marks.

Records of Assignments to be submitted on the date of the examination carry 10 marks.

1 Given a Non-census sampling design, show why you may not find a uniformly minimum variance unbiased estimator for a finite population total.

2 Find an admissible estimator which is homogeneous linear and unbiased for a finite population total explaining a necessary and sufficient condition for its existence.

3 Show how you may find a sufficient statistic from a sequence of sample survey observations.

Explain in details an advantage in deriving a sufficient statistic in a sample survey.

4 Give an account of different possible ways of drawing a systematic sample in each of the four usual kinds.

5 Obtain an unbiased estimator for the covariance between the means of two real variables based on a simple random sample taken without replacement in n draws from a finite population of size $N\ (> n > 2)$.

6 Explain fully D.B. Lahiri's method of drawing a PPS sample of size one.

N.B: Answers beyond the first 4 will be cancelled.

*****

INDIAN STATISTICAL INSTITUTE

Mid – Semestral Examination: 2013-14

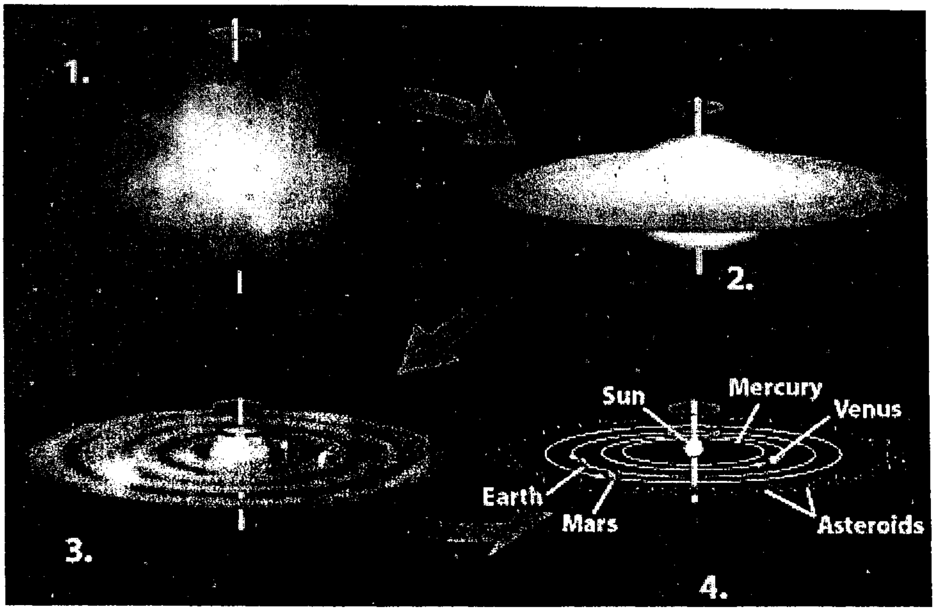
B. Stat III Year Geology Elective

Maximum Marks: 30. **Duration: 2.00 Hours.**

02/09/13

Answer <u>all</u> the questions. Be brief and to the point.

Question 1:



The figure above depicts four stages (1, 2, 3 and 4 as shown in the figure) of evolution of the solar system (Nebular disk hypothesis).

a) Give three evidences supporting the Nebular disk hypothesis? (3)

b) During which stage would the processes that create the Sun's energy begin? (1)

c) During which stage would there be the onset of significant collection of solar particles and gas? (1)

d) The formation of Earth and other planetary bodies through the processes of condensation and accretion was essentially complete _____ years ago. (1)

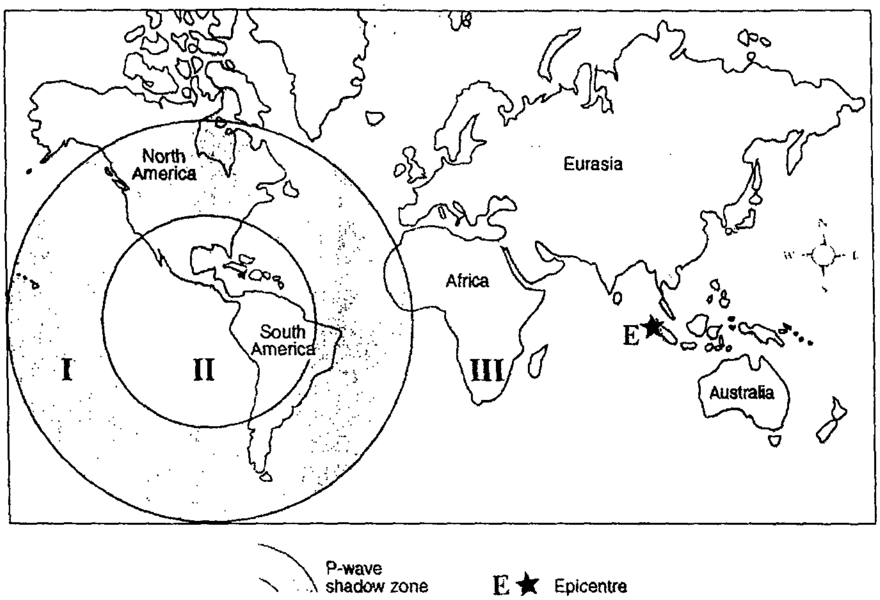
A. 460 million
B. 4.60 million
C. 4.60 billion
D. 46.0 billion

Question 2:

a) Explain elastic rebound theory? (1)

b) Why we get S – and P – wave shadow zone? (2)

c) If our planet is completely solid then do you expect to get S –wave shadow zone? (1)

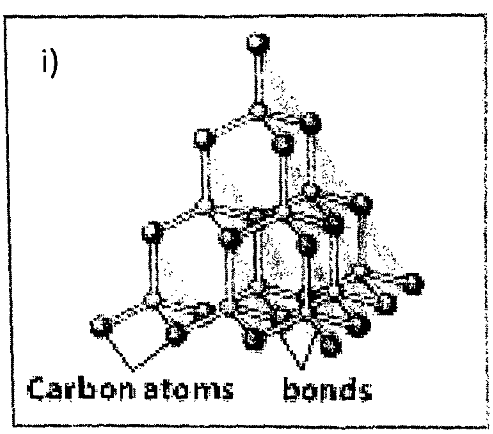


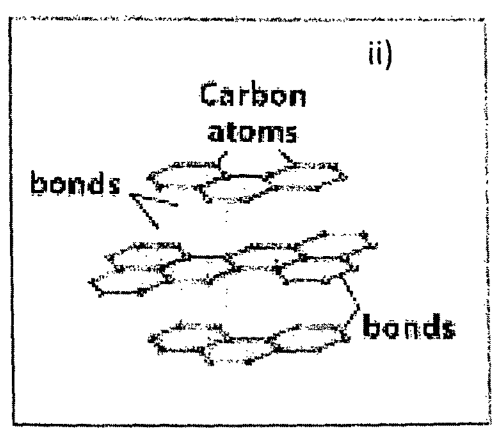
d) The map above shows the P-wave shadow zone resulting from the Indonesian earthquake of December 26, 2004. Which of the locations, I, II and III would have received the P-waves? (1)

A. I only
B. I and II only
C. I and III only
D. II and III only

e) What is meant by PP wave and PKP wave? (1)

Question 3:





a) Using the figure above that illustrates the bonding of carbon atoms forming the crystal structures of diamond (i) and graphite (ii) can you explain the difference in hardness between these two minerals (diamond: Mohs scale of mineral hardness 10; graphite: Mohs scale of mineral hardness 1.5)? (1)

b) Which list correctly shows the order of abundance of silicon, oxygen and iron in the Earth's crust? (1)

	Most Abundant		Least Abundant
A.	Silicon	Iron	Oxygen
B.	Oxygen	Silicon	Iron
C.	Silicon	Oxygen	Iron
D.	Oxygen	Iron	silicon

c) Give one example each, of simple substitution; coupled substitution and omission solid solution. (3)

d) Give one example each of ring silicate and framework silicate minerals. (1)

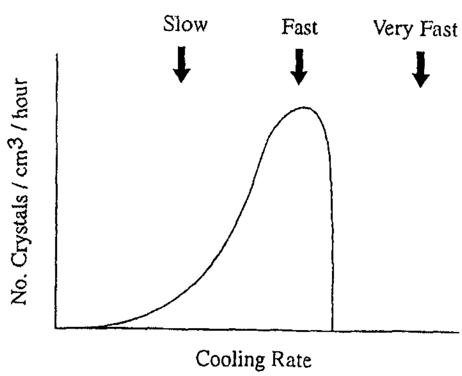
Question 4:

a) Where do you expect to get decompression melting, flux induced melting and melting due to elevated geothermal gradient? (3)

b) Settling velocity (V_s) of a spherical particle (of radius R) with density r_p through a fluid of density r_f and viscosity *mu* can be calculated using the formula $\left(V_s = \frac{2}{9}\left(\frac{r_p - r_f}{mu}\right) gR^2 \right)$. Consider an olivine crystal with a diameter of 0.1 cm and density of 3.3 g/cm^3 precipitating from a mafic magma of density 2.65 g/cm^3 and viscosity 1000 poise. What will be settling velocity of the crystal? (1)

c) What are the two major magma transport mechanisms? (1)

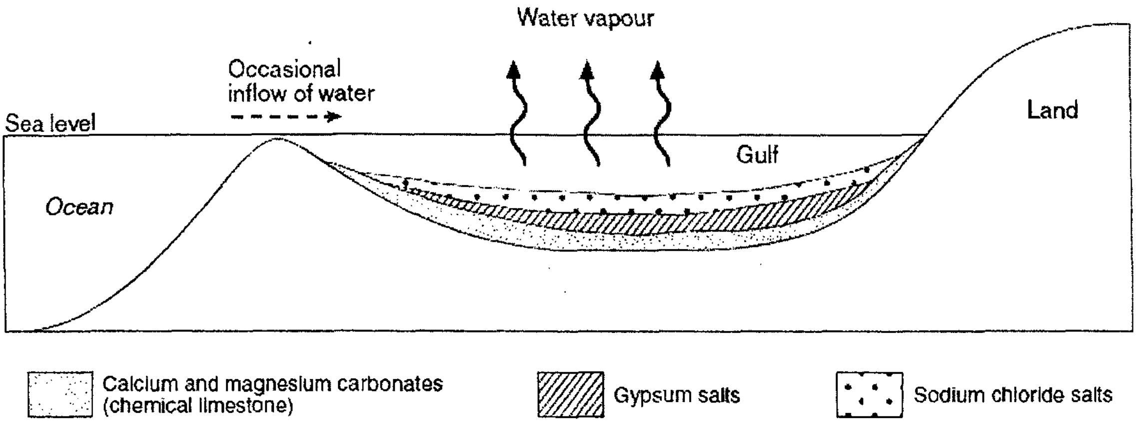
d) The figure below shows the relationship between cooling rate and rate of crystal formation. What will be the expected crystal sizes from a slow cooling and rapid cooling melt? (1)



Question 5:

a) What parameters are used to determine (a) textural and (b) compositional maturity of sedimentary rocks? (2)

b) Why quartz is the most stable mineral i.e. resistant to weathering? (2)



c) What type(s) of deposit is/are expected to form in the shallow gulf that gets occasional influx of seawater as shown in the diagram? (1)

A. Metamorphic
B. Evaporite
C. Magmatic
D. None of the above

d) Which of the following processes is most likely to preserve land plants as a large coal deposit? (1)

A. Dead plants are heated and dried by the tropical sun.
B. Eruption of lava flow that covers and heat the plant matter.
C. Sea level rises and plant matter is buried by sediments.
D. Chemical and biological weathering occurs, causing organic decay.

Indian Statistical Institute

Mid-Semesters Examination: 2013-14

Course Name: B. Stat. III year

Subject Name: Anthropology and Human Genetics

Date: 02 September 2013, Maximum Marks: 60 Duration: 3 hours

---

01. How do you define Anthropology? What are the distinctive features of Anthropology?

02. What are the anatomical and morphological features of the class Mammalia?

03. Why is man unique in the animal kingdom?

04. State Darwin's theory of evolution. What are the weaknesses of this theory?

05. What are the stresses on man at high altitude? How do human beings cope with altitude stress?

06. Write short notes on (any two)
    a) Adaptation and Acclimatization
    b) Primate
    c) Adaptation to hot climate
    d) Theory of use and disuse.

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2013-2014

B. Stat. (Hons.) 3rd Year. 1st Semester

Statistical Inference I

Date: September 06, 2013 **Maximum Marks: 50** **Duration: 2 and 1/2 hours**

- This question paper carries 55 points. Answer as much as you can. However, the maximum you can score is 50.

- You should provide as much details as possible while answering a question. You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.

---

1. Let $X := (\min(T, C), I(T \leq C))$ where T, C are independent with $P(T = j) = p_j$, $P(C = j) = r_j$; $j = 0, \ldots, N$. Let $\boldsymbol{p} := (p_1, \ldots, p_N)$, $\boldsymbol{r} := (r_1, \ldots, r_N)$. Suppose $(\boldsymbol{p}, \boldsymbol{r})$ varies freely over $\Theta := \{(\boldsymbol{p}, \boldsymbol{r}) : p_j > 0, r_j > 0, 0 \leq j \leq N, \sum_{j=0}^{N} p_j = \sum_{j=0}^{N} r_j = 1\}$ and N is known. Show that $(\boldsymbol{p}, \boldsymbol{r})$ is identifiable. [8]

2. For N independent and identically distributed (i.i.d.) observations from the family $\{\text{Multinomial}(1, \theta_1, \theta_2, \theta_3) : \theta_1, \theta_2, \theta_3 \in (0,1) \text{ and } \sum_{j=1}^{3} \theta_j = 1\}$, find a minimal sufficient and complete statistic. [8]

3. Suppose $X_1, \ldots, X_n$ are i.i.d. Poisson(θ), $\theta > 0$ variables. Define $\widehat{\theta}_n = \sum_{i=1}^{n} X_i/n$. Show that the bias of $\exp(-\widehat{\theta}_n)$, as an estimator of $\exp(-\theta)$, is positive for all θ and n, and that it goes to zero as n tends to infinity. [7+4=11]

4. Let $X_1, \ldots, X_n$ $(n > 2)$ be i.i.d. $\mathrm{U}(\theta - \sigma, \theta + \sigma)$, where $\theta \in \mathbb{R}, \sigma > 0$ are both unknown. Find the UMVUE of θ/σ. [**Note** You may assume the algebraic form of complete and sufficient statistic for this problem and standard facts about $\mathrm{U}(0, 1)$ order statistics.] [12]

5. (a) Suppose X follows a distribution with pdf denoted by $f(x, \theta)$, where $\theta \in \Theta$. Let $T(X)$ be a statistic. State a version of Cramér-Rao inequality that provides a lower bound to $\mathrm{Var}_\theta(T(X))$ assuming only $\mathrm{E}_\theta(T^2(X)) < \infty$ and making no other assumption about $T(X)$. You have to state all the assumptions clearly.

(b) Assume in (a) that $f(x, \theta)$ is the pdf of logistic distribution with location parameter θ and $\Theta = \mathbb{R}$. Verify the assumptions stated in (a).

(c) Obtain the lower bound stated in (a) explicitly. [4+6+6=16]

***** ** *Best of Luck!* ***** **

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2013 – 14

B. Stat (3rd Year)

Linear Statistical Models

Date: 11 November 2013 Maximum Marks: 100 Duration: 3 Hours

1. Cite a commonly known theoretical model from your *Economic and Official Statistics* course and write down a statistical version of it.
   State a relevant hypothesis for this model and derive the test statistic.

   [6 + 8 = 14]

2. In an agricultural experiment, three varieties of fertilizers (A, B and C) are used in six otherwise identical plots each. The output (y) is measured for the 18 plots.
   (a) Write an appropriate linear model for this experiment, stating all relevant assumptions.
   (b) If we want to test whether fertilizer A is better than both B and C, state the appropriate hypothesis. Derive the test statistic for this test and find its probability distribution. [7 + (4 + 7) = 18]

3. Consider a linear model $y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i$ with $Var(y_i) = \sigma^2$ and $Correlation\ (y_i, y_j) = -0.05$, for $i \neq j$; i, j = 1,...,25.
   Derive the best estimators (MVUE) for $\beta_0, \beta_1\ and\ \sigma^2$. Find the variance of $(\beta_0, \beta_1)'$.
   How would your answer change if $Correlation\ (y_i, y_j) = 0.05$?

[6 + 8 + 6 = 20]

4. A Mobile phone manufacturer is trying to increase their sales by offering free 4G internet dongle and / or "A R Rahman Live in Concert" DVD set (selection of customers for each offer are made independently). The offers are made to a random selection of potential customers visiting their shops. The output of interest is whether a purchase is made.

(a) Write down an appropriate statistical model, stating relevant assumptions, to analyse the data from this experiment. Also write down the relevant estimators and test statistics.

(b) If instead of only the purchase decision, the actual price of the handset sold is also recorded, how will you modify your model and analysis to this situation?

(c) It is also believed that the age and annual income of customers affect the purchase decision. So the shop has collected such data on each potential customer. How can you modify your analysis to test these beliefs?

(d) In (a) above, if some customers are observed purchasing more than one phone how can this output be modeled and analysed?

[(7 + 7) + 12 + 12 + 10 = 4[

INDIAN STATISTICAL INSTITUTE

First-Semester Examinations: 2013-14

Course Name : B. STAT. III YEAR

Subject Name : Differential Equations

Date: November 13, 2013, Maximum Marks: 60, Duration: 3 hrs.

Any result that you use should be stated clearly.

1. Solve the equations
   a) $dx + xdy = e^{-y}\sec^2 y\, dy$.
   b) $(\sin x \sin y - xe^y)dy = (e^y + \cos x \cos y)dx$. 3+3
2. Find the general solution of $(1+x^2)\frac{d^2y}{dx^2} + 2x\frac{dy}{dx} - 2y = 0$ in terms of power series in x. Can you express this solution by means of elementary functions? 6+1
3. Define Legendre polynomial $P_n(x)$ of degree n. Hence show that Legendre polynomial satisfy the relation $(n+1)P_{n+1}(x) - (2n+1)x\,P_n(x) + n\,P_{n-1}(x) = 0,\ n = 1, 2, 3, \ldots$ 2+7
4. Show that the equation $x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} - 16y = 0$ transforms into an equation with constant coefficients by changing the independent variable given by $x = e^z$. Hence find the general solution of the given equation. 3+3
5. Using the method of variational calculus, deduce Euler's differential equation for extremization (stationarity) of the integral

$$I = \int_0^4 \left[x\frac{dy}{dx} - \left(\frac{dy}{dx}\right)^2\right]dx$$

   for fixed values of $y(0), y(4)$ and find out the 'stationary' function $y(x)$ with boundary conditions $y(0) = 0$ and $y(4) = 3$. 5+2
6. Find the general solution of
   a) $xzp + yzq = xy$,
   b) $pxy + pq + qy = yz$, where $p = \frac{\partial z}{\partial x}, q = \frac{\partial z}{\partial y}$.
   c) Find the partial differential equation by eliminating a and b from $z = ax\,e^y + \frac{1}{2}a^2 e^{2y} + b$. 3+4+3
7. a) For the Bessel's equation of order n

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

   find out the Frobenius series solution $J_n(x)$ about the regular singular point which is bounded at the origin $(x = 0)$.
   b) Prove that $x\,J_n'(x) = n\,J_n(x) - x\,J_{n+1}(x)$. 8+6
8. The Logistic model for population growth gives the differential equation $\frac{dp}{dt} = ap - bp^2$ for the population p of a spices 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 behavior of $p(t)$ as $t \to \infty$. 6
9. Solve by method of undetermined co-efficients
   $\frac{d^2y}{dx^2} - 3\frac{dy}{dx} + 2y = \sin 2x$. 6

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2013–2014

B. Stat. (Hons.) 3rd Year. 1st Semester

Statistical Inference I

Date: November 22, 2013 **Maximum Marks: 60** **Duration: 3 hours**

• This question paper carries 66 points. Answer as much as you can. However, the maximum you can score is 60.

• You should provide as much details as possible while answering a question. You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.

---

1. Suppose (X, τ) is a random vector such that (i) the conditional distribution of X given τ is $\mathrm{N}(\theta, \sigma^2/\tau)$ and (ii) the distribution of τ is Gamma with scale parameter 2 and shape parameter 1/2. It is known that the marginal distribution of X is $\mathrm{C}(\theta, \sigma)$, Cauchy with location parameter θ and scale parameter σ, and that the conditional distribution of τ given X is exponential with scale parameter $2\sigma^2/[\sigma^2 + (x-\theta)^2]$. Use these facts to discuss how you can find the MLE of (θ, σ) based on a random sample of size n from $\mathrm{C}(\theta, \sigma)$, both θ and σ being unknown, by employing EM algorithm. What is an initial choice for (θ, σ) and why? [You should give all relevant details.] [10]

2. Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) observations from $\mathrm{U}(0, \theta)$. Let $\mathbf{X} := (X_1, \ldots, X_n)$. Show that for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ a UMP test is given by the following: $\phi(\mathbf{X}) = 1$ when $X_{(n)} > \theta_0$ or $X_{(n)} \leq \theta_0 \alpha^{1/n}$ and $\phi(\mathbf{X}) = 0$ otherwise. [9]

3. Let $X_1, \ldots, X_n$ denote the incomes of n persons chosen at random from a certain population. Suppose that each X_i has the Pareto density $f(x, \theta) = c^\theta \theta x^{-(1+\theta)}$, $x \geq c$ where $\theta > 1$ and $c > 0$. Let c be known.

(a) Express mean income μ in terms of θ.

(b) Find the UMP test for testing $\mathrm{H}_0 : \mu = \mu_0$ against $\mathrm{H}_1 : \mu > \mu_0$.

(c) Use the central limit theorem to find an approximation to the critical value of the test in part (b). [4+7+4=15]

4. Let $X_1, \ldots, X_n$ be i.i.d exponential variables with location parameter 0 and unknown scale parameter θ, $\theta > 0$. Show that there does not exist a UMP test for testing $\mathrm{H}_0 : \theta = \theta_0$ against $\mathrm{H}_1 : \theta \neq \theta_0$. [9]

[P.T.O.]

5. Suppose that we have two independent random samples: $X_1, \ldots, X_m$ from a Beta$(\mu, 1)$ population and $Y_1, \ldots, Y_n$ from a Beta$(\theta, 1)$ population. Find the LRT of $\mathrm{H}_0 : \theta = \mu$ versus $\mathrm{H}_1 : \theta \neq \mu$. Show that this test is a one-to-one function of the statistic

$$T(\boldsymbol{X}, \boldsymbol{Y}) \stackrel{def}{=} \frac{\sum_{i=1}^{m} \log X_i}{\sum_{i=1}^{m} \log X_i + \sum_{j=1}^{n} \log Y_j}.$$

Find the null distribution of $T(\boldsymbol{X}, \boldsymbol{Y})$. [7+5=12]

6. Suppose $X_1, \ldots, X_n$ are i.i.d. exponential variables with unknown location parameter θ and unknwon scale parameter σ. Let $T := \sum_{i=1}^{n}(X_i - X_{(1)})$. Find the shortest $1 - \alpha$ confidence interval for σ based on T. [11]

$* * * * **$ *Best of Luck!* $* * * * **$

INDIAN STATISTICAL INSTITTUTE
First Semestral Examination: 2013 -2014
Anthropology and Human Genetics - Human Genetics
B. Stat. III year (Electives)

Maximum Marks: 40

Date: 25-11-2013 Duration: 2 hours

*Instructions: Attempt any **four** questions. All questions carry equal marks.*

1. (a). Explain the importance of Hardy-Weinberg Equilibrium (HWE)? In case in a population the observed gene (allele) frequencies of a particular bi-allelic locus follow HW expectations, does this imply the population satisfies the assumptions of HW equilibrium?

   (b). Which of the following genotype frequencies of AA, Aa and aa, respectively, satisfy the Hardy-Weinberg principle?

   i. 0.25 0.50 0.25
   ii. 0.36 0.55 0.09
   iii. 0.49 0.42 0.09
   iv. 0.64 0.27 0.09
   v. 0.29 0.42 0.29

2. (a). Derive estimates of standard error by Maximum Likelihood (ML) method in case of a genetic trait governed by single locus with two alleles with dominance: e.g. PTC (Phenyl Thio Carbamide) taste sensitivity. It is observed that some people taste PTC and some others do not. The ability to taste is genetically controlled by a gene at a single locus with two alleles. Assume 'T' (tasters) and 't' (non-tasters) – as two alleles where, 'T' is dominant and allele 't' is recessive.

   (b). In a population of northeast region the frequency of alleles determining the ABO blood type groups were estimated as A = 0.209, B = 0.129, and O = 0.660. What are the expected gene frequencies? OR

   (c). Consider a population in which 36% of the people exhibit a trait due to a homozygous recessive mutation. What is the allele frequency for the dominant allele if a population is in Hardy-Weinberg equilibrium?

3 (a). What is population structure? Knowledge of population structure is essential for understanding the gene dynamics in human populations. Explain?

(b). By drawing a pedigree show the inbreeding coefficient of an offspring whose parents are double cousins?

4. (a). Mutation ('μ') is an important stochastic factor that can influence the gene frequency in a population. Derive the expression relating to the change in gene frequency in a population as a result of mutation ('μ')?

(b). Sanfillipo syndrome is an inborn error of metabolism. Affected children develop quite normally in the early years of their life, but the disease express in their early teens and later teen age, the disease grows worse as time goes on. These affected children will have difficulties in learning how to speak and exhibit behavioural disorders: including temper tantrums, hyperactivity, destructiveness, and aggression. The affected children gradually lose mobility and become unresponsive and usually die in their late teens or twenties. It has been found that the disorder is an autosomal recessive trait that affects 1 in 50,000 new-born children among unrelated parents. If so, please answer the following situations!

i. What is the expected frequency of affected children among the offspring of first-cousin (marriage) mating?

ii. What is the relative risk (ratio) of the disease among the offspring of first cousin relatives to that among the offspring of unrelated individuals?

5. (a). What is Genetic Distance? Write briefly about Nei's Genetic Distance

(b). The table below shows Mahalanobis D^2 distances calculated among 5 regional subpopulations (CY, IY, P1, HF and P2) of a tribal population based on 14 anthropometric measurements among males (upper matrix) and females (lower matrix). Based on the two Distance Matrices given below construct a dendrogram (clustering tree) separately for males and females by following 'minimum distance method'?

Table: Mahalanobis D^2 distance estimates between 5 regional subpopulations of a tribe (Upper matrix – males, Lower matrix – females)

Sub Population (Female)	Sub population (Male) CY	IY	P1	HF	P2
CY	---	1.64	1.73	2.5	2.3
IY	3.1	---	2.1	3.2	2.3
P1	1.2	3.6	----	0.5	0.9
HF	3.3	2.0	2.0	----	0.34
P2	2.3	3.0	1.6	0.6	---

INDIAN STATISTICAL INSTITUTE

First – Semester Examination: 2013-14

B. Stat III Year Geology Elective

Date: 25.11.2013 **Maximum Marks: 30** **Duration: Two Hours**

Question 1. **(Total marks 6)**

a) How does mechanical weathering help in chemical weathering? (2)
b) How does a placer deposit form? Give an example of placer deposit. (1 + 1 = 2)
c) How sinkholes form? Write the relevant balanced chemical equation. (1 + 1 = 2)

Question 2. **(Total marks 6)**

a) Show that $t_{1/2} = \frac{0.693}{\lambda}$, where $t_{1/2}$ is the half life and λ is the decay constant. (2)
b) After 500 years, a sample of Radium-226 has decayed to 80.4% of its original mass. Find the half-life of Radium-226. (4)

Question 3. **(Total marks 6)**

a) Would you consider an Egyptian mummy as a fossil? Justify your answer. (2)
b) Match the following columns A and B: (4)

Column A	Column B
Jurassic sauropod	Rhynchosaur
Phanerozoic	Trilobite
Molting	Eon
	Barapasaurus
Series	Epoch
	Bivalve

Question 4. **(Total marks 6)**

All Assemblage Zones are Concurrent Range Zones. But all Concurrent Range Zones are not Assemblage Zones. Justify the statement with labelled diagram(s). (2+2+2=6)

Question 5. **(Total marks 6)**

What is phyletic gradualism? How does it differ from the concept of "Punctuated Equilibrium"? (2+4=6)

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2013-14
B. Stat. III Year
Sample Survey

Date: 27.11.2013 **Maximum Marks: 50** **Duration: 3 Hours**

Answer any 4 questions, each carrying 10 marks.

Assignment Records to be submitted on the date of the examination carry 10 marks.

1 Explain the principle one should bear in mind giving reasons, while constructing (A) strata and (B) single – stage clusters in sampling finite populations.

2 Work out Neyman's allocation rules in stratified sampling showing more journal ones of which these are but mere special cases.

3 Suppose, given some positive normed size- measures, a PPSWR sample is taken in n draws and each time a sampling unit appears in the sample, it is independently sub-sampled three times each by SRSWOR method, show how you should unbiasedly estimate a finite population total and present an unbiased estimator of the variance of your proposed estimator for the total.

4 Derive Hartley and Ross's unbiased estimator for a finite population total. Also derive an unbiased estimator for the variance of this estimator.

5 Using usual notations give a reasonable estimator, along with its variance estimator, for a finite population mean based on simple random samples taken suitably without replacement in two stages with appropriate numbers of draws.

*****

Indian Statistical Institute, Bangalore Centre
B.Math (Hons.) III Year, First Semester
Backpaper Examination - 2013-2014
Complex Analysis

Time: 3 hours | January 7, 2014 | Instructor: Bhaskar Bagchi

Full Marks : 100.

1. (a) If a holomorphic function is injective on its domain then show that it can't have any essential singularity.

   (b) If $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is an injective meromorphic function then show that $f(z) \equiv \frac{az+l}{cz+d}$ for some constants a, l, c, d.

   [10+10=20]

2. (a) Define the residue of a holomorphic function at an isolated singularity.

   (b) State and prove the residue theorem.

   [5+15=20]

3. (a) If $f : \Omega \to \mathbb{C}$ is a non-constant holomorphic function then show that for each $z \in \Omega$ there is a positive integer m_z such that f is m_z - to - one on some neighborhood of z.

   (b) Prove that we can't have $m_z = 2$ for all $z \in \Omega$.

   [12+8=20]

4. Use the fundamental theorem of Gauss to compute $\int\limits_{-\infty}^{\infty} e^{itx}\, e^{-x^2/2} dx$ for $t \in \mathbb{R}$. (You may use without proof the fact that for $t = 0$, the value of this integral is $\sqrt{2\pi}$.)

   [20]

5. Let $\Omega = \{x + iy : x > 1, y \in \mathbb{R}\}$. Define $\zeta : \Omega \to \mathbb{C}$ by $\zeta(z) = \sum\limits_{n=1}^{\infty} n^{-z}$.

   (a) Prove that this series converges locally uniformly on Ω. Hence deduce that ζ is holomorphic on Ω.

   (b) Prove that $\zeta(z) \neq 0$ for z in Ω.

   [12+8=20]

INDIAN STATISTICAL INSTITUTE

Backpaper Examination: 2013–2014

B.Stat. (Hons.) 3rd Year. 1st Semester

Statistical Inference I

Date: January 22, 2014 **Maximum Marks: 100** **Duration: 3 hours**

- Answer all the questions.
- You should provide as much details as possible while answering a question. You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.

---

1. Let $X := (\min(T,C), I(T \leq C))$ where T, C are independent with $P(T = j) = p_j$, $P(C = j) = r_j$; $j = 0, \ldots, N$. Let $\boldsymbol{p} := (p_1, \ldots, p_N)$, $\boldsymbol{r} := (r_1, \ldots, r_N)$. Suppose $(\boldsymbol{p}, \boldsymbol{r})$ varies freely over $\Theta := \{(\boldsymbol{p}, \boldsymbol{r}) : p_j > 0, r_j > 0, 0 \leq j \leq N, \sum_{j=0}^{N} p_j = \sum_{j=0}^{N} r_j = 1\}$ and N is known. Show that $(\boldsymbol{p}, \boldsymbol{r})$ is identifiable. [12]

2. Denote by Gamma(α, β), the gamma distribution with scala parameter α and shape parameter β; $\alpha, \beta \in (0, \infty)$. For n independent and identically distributed (i.i.d.) observations from the family $\{\text{Gamma}(\alpha, \beta) : (\alpha, \beta),\ \alpha, \beta \in (0, \infty)\}$, let $T := \sum_{i=1}^{n} X_i$, $U := \max(X_1, \ldots, X_n) / \min(X_1, \ldots, X_n)$. Use Basu's theorem to show that T and U are independent. [12]

3. Suppose $X_1, \ldots, X_n$ are i.i.d. with distribution function $F \in \mathcal{F}$, the family of all probability densities on $\mathbb{R}$. Find the UMVUE of $\theta(F) := P_F(X_1 \leq a_0) = F(a_0)$, a_0 known. [12]

4. Let $X_1, \ldots, X_n$ be i.i.d. according to a density $f(x, \theta)$ which is positive for all x. Write $\boldsymbol{X} = (X_1, \ldots, X_n)$. Show that the variance of any unbiased estimator $\delta(\boldsymbol{X})$ of θ satisfies

$$\text{Var}_{\theta_0}(\delta(\boldsymbol{X})) \geq \frac{(\theta - \theta_0)^2}{\left\{ \int_{-\infty}^{\infty} \frac{[f(x,\theta)]^2}{f(x,\theta_0)}\, dx \right\}^n - 1}, \quad \theta \neq \theta_0. \qquad [12]$$

5. Let $X_1, \ldots, X_n$ be i.i.d N$(\theta, 1)$ variables, $\theta \in \mathbb{R}$ is unknown. Show that there does not exist a UMP test for testing H$_0 : \theta = \theta_0$ against H$_1 : \theta \neq \theta_0$. [11]

[P.T.O.]

6. Let $X_1, \ldots, X_n$ denote the incomes of n persons chosen at random from a certain population. Suppose that each X_i has the Pareto density $f(x, \theta) = c^\theta \theta x^{-(1+\theta)}$, $x \geq c$ where $\theta > 1$ and $c > 0$. Let c be known.

(a) Express mean income μ in terms of θ.

(b) Find the UMP test for testing $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$.

(c) Use the central limit theorem to find an approximation to the critical value of the test in part (b). [4+7+4=15]

7. Suppose that we have two independent random samples: $X_1, \ldots, X_m$ are exponential scale parameter θ, and $Y_1, \ldots, Y_n$ are exponential scale parameter μ. Find the LRT of $H_0 : \theta = \mu$ versus $H_1 : \theta \neq \mu$. Show that this test is equivalent to the statistic

$$T(\boldsymbol{X}, \boldsymbol{Y}) \stackrel{def}{=} \frac{\sum_{i=1}^m X_i}{\sum_{i=1}^m X_i + \sum_{j=1}^n Y_j}.$$

Find the null distribution of $T(\boldsymbol{X}, \boldsymbol{Y})$. [7+7=14]

8. Suppose $X_1, \ldots, X_n$ are i.i.d. exponential variables with unknown location parameter θ. Let $T := X_{(1)}$. Find the shortest $1 - \alpha$ confidence interval for θ based on T. [12]

****** *Best of Luck!* ******

INDIAN STATISTICAL INSTITUTE

First Semester Back Paper Examination: 2013 – 14

B. Stat (3rd Year)

Linear Statistical Models

Date: 24·01·14 Maximum Marks: 100 Duration: 3 Hours

1. For a linear regression model, show that adding a new explanatory variable (weakly) increases R^2. [10]

2. Show that the weighted least square estimator $\widehat{\boldsymbol{\beta}} = (\widehat{\beta_0}, \widehat{\beta_1},)'$ for the model

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

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

Also show that $Cov(\widehat{\boldsymbol{\beta}}) = \frac{\sigma^2}{\sum_i x_i \sum_i \frac{1}{x_i} - n^2}\begin{pmatrix} (\sum_i x_i) & -n \\ -n & \left(\sum_i \frac{1}{x_i}\right)\end{pmatrix}$. [9 + 6 = 15]

3. For a linear model, obtain the Confidence Interval for $\boldsymbol{a}'\boldsymbol{\beta}$ from the corresponding t-statistic. [10]

4. For the linear model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$ with $Var(y_i) = \sigma^2 x_{1i}^2$, for i = 1,…, n and such that $x_{1i} \neq 0$ for all i

(a) Derive the best estimators (MVUE) for $\beta_0, \beta_1, \beta_2$ *and* σ^2 . Find the variance of $(\beta_0, \beta_1, \beta_2)'$.

(b) If by mistake we assume homoscedasticity, what will be the properties of the estimators (expectation and variance). [(6 + 4) + (6 + 9) = 25]

Indian Statistical Institute

First Semestral Examination: (2013–2014)

B.Stat.(Hons.) – III year

Economics III

Back Paper

Date: ~~00/00/2013~~ 28.01.2014 *Maximum Marks –**100*** *Duration: 3 hours*

Answer ***all*** questions.

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

   (b) Describe the Ordinary Least Squares (OLS) method of estimating the parameters of a CLRM.

   (c) Show that the OLS estimator is the Best Linear Unbiased Estimator (BLUE).

[12+8+5=25]

2. (a) What is multicollinearity?

   (b) How does one detect multicollinearity? Explain.

   (c) Describe some remedial measures to deal with this problem.

[3+12+10=25]

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

   (b) Describe a test for testing the presence of first order autocorrelation (ρ) in a given time series. Derive the relationship between ρ and the test statistic.

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

   (d) Consider the two models:

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

(-3.96) (3.27) (-2.17)

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

P.T.O.

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

[2+9+6+8 = 25]

4. (a) Describe the Two Stage least Squares (2SLS) procedure for estimating a simultaneous equations system.

(b) Explain the Instrumental Variables (IV) approach for estimating a single equation of a simultaneous equations model. Derive the 2SLS estimator as an IV estimator.

(c) Describe the K-class estimator. What are the conditions under which the K-class estimator is consistent? What are the different types of estimators belonging to this category?

[10+10+5=25]

INDIAN STATISTICAL INSTITUTE

Mid-semestral Examination, Second Semester, 2013-14

B.Stat Third Year

Date - 24.02.14

Time: $2\frac{1}{2}$ Hours | Nonparametric Inference | Full Marks: 60

1. (a) Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ and $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ be two sets of independent observations from two bivariate continuous distributions F and G. For any fixed $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)' \in R^2$, find the UMVUE of $\theta(\boldsymbol{\alpha}, F, G) = P(\boldsymbol{\alpha}'\mathbf{X} > \boldsymbol{\alpha}'\mathbf{Y})$, where $\mathbf{X} \sim F$, $\mathbf{Y} \sim G$ and they are independent. [5]

   (b) Show that a random variable $\mathbf{X}$ is spherically symmetric about $\boldsymbol{\mu}$ if and only if its characteristic function $\phi(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{X}})$ is of the form $\phi(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu}}\,\psi(\|\mathbf{t}\|)$ for some function ψ. [6]

   (c) If $\mathbf{X}$ is spherically symmetric about $\boldsymbol{\mu}_1$, $\mathbf{Y}$ is spherically symmetric about $\boldsymbol{\mu}_2$ and they are independent, show that

   (i) $\mathbf{X} - \mathbf{Y}$ is sperically symmetric about $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

   (ii) $P(\boldsymbol{\alpha}'\mathbf{X} > \boldsymbol{\alpha}'\mathbf{Y})$ is maximum when $\boldsymbol{\alpha} \propto (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. [2+5]

2. (a) Show that the univariate kernel density estimate with triangular kernel function can be viewed as a limiting case of the average shifted histogram density estimate. [6]

   (b) Consider the following data set consisting of 12 pairs of observations on x and y.

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	-6	-5	-3	-2	0	1	1	2	2	3	3	4
y_i	5	4	2	0	0	-2	1	-2	-1	-1	-3	-3

$$\left[\sum_{i=1}^{12} x_i = 0,\ \sum_{i=1}^{12} y_i = 0,\ \sum_{i=1}^{12} x_i^2 = 118,\ \sum_{i=1}^{12} y_i^2 = 74,\ \sum_{i=1}^{12} x_i y_i = -87\right]$$

   If Gaussian kernel is used to construct the Nadaraya Watson estimate $\hat{f}_h(\cdot)$ of the regression function $f(\cdot)$, find the limiting values of $\hat{f}_h(1.5)$ when the smoothing parameter h (i) tends to infinity (ii) shrinks to zero. [2+3]

   (c) For any $\delta \in (0,1)$, the central region of a distribution with a bounded density function f is defined as $C(f,\delta) = \{\mathbf{x} : f(\mathbf{x}) \geq \delta \sup_{\mathbf{t}} f(\mathbf{t})\}$. Show that for any $\delta \in (0,1)$, the probability of the central region of a d-dimensional normal distribution shrinks to zero as the dimension increases. [5]

3. (a) Consider a game between two players 'A' and 'B', where 'A' tosses a coin first. Then 'B' tosses that coin again, and he wins if he gets the same outcome as 'A'. Otherwise, 'A' wins the game. In a series of 30 games, if 'B' wins 20 times, use an appropriate statistical method to test (at 5% nominal level) the fairness of the coin. Show that if the coin is not unbiased the power of your test converges to 1 as the number of games increases. [5+5]

   (b) Derive the expectation, the variance and the large sample distribution of the Wilcoxon signed rank statistic. [2+3+6]

   (c) Describe how the Hodges-Lehmann estimate of the median of a population is obtained from the Wilcoxon signed rank statistic. [5]

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

Mid-semestral Examination
Time: 2 Hours Full Marks: 50
Date: February 25, 2014

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. For a branching process with probability generating function $\phi(s) = p/(1-qs)$, where $0 < p < 1$ and $p+q=1$, find out the probability generating function and expectation of the total progeny size. [6+3=9]

2. A particle moves on a circle through points marked (in a clockwise order) as 0, 1, 2, 3, 4. At each step it moves to its neighbouring point - to the right with probability p and to the left with probability $1-p$. Let X_n denote its location after n steps. Show that $\{X_n\}$ is a Markov chain and find its transition matrix. Find the Cesaro limit of the powers of the transition matrix. [2+3+4=9]

3. Let $\boldsymbol{P}$ be the transition matrix on the state space $\{0, 1, 2, \cdots\}$ with the first column $(q_0, q_1, \ldots)$ and for $i \geq 0$, we have $p_{i,i+1} = 1 - q_i$. Show that the chain is irreducible iff $q_i < 1$ for all i and $q_i > 0$ for infinitely many i. Assuming the chain to be irreducible, show that the chain is are transient if and only is $\sum q_i < \infty$. [8+8=16]

4. Give example, if possible, of a Markov chain with state space $\{0, 1, 2\}$ such that $0 \rightsquigarrow 2$ but $p_{02}^{(n)} = 0$ for $n = 1, 2$. [8]

5. Show that an irreducible idempotent transition matrix must necessarily have all rows same. (Hint: Think of getting a matrix with all rows same.) [8]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2013-2014

Course: B.Stat. III

Subject: Database Management Systems

Date: 26/02/2014 | Maximum Marks: 40 | Duration: 2 hours

Answer all questions

1. Suppliers supply zero or more parts to different projects taken by XYZ Manufacturing Company at various cities. Each Supplier has a unique name. More than one Suppliers may be located in the same city. Each part has a unique number. In addition they may be of various colours. A project may have zero or more suppliers.

   (i) Draw an ER/EER schema diagram *stating any assumptions you make.*

   (ii) Map them to appropriate relations

   (iii) Write SQL statement to create tables **with proper constraints.**

   (iv) Form the following queries using relational algebra:

      a. Name of suppliers who supply all "YELLOW" parts

      b. Name of suppliers who supply some "WHITE" parts

      c. Name of suppliers who supply parts to those projects that are located in the same city where they belong.

10 + 06 + 10 + (5 + 4 + 5) = 40

BIII-2013-2014
Midterm Examination
Design of Experiments

Full Marks 30 date: 27·02·14 Time: 10.30-12.00

1.Consider aCRD with 5 treatments of which one is the standard treatment(denoted by 0) and four are new treatments (denoted by 1,2,3 and 4). The total number of available experimental units is 12. If the main interest is on the pair wise comparison of the standard and new treatment effects which of the following design will you choose and why?

Design 1: $r_0=4,$: $r_1=2,$: $r_2=2,$: $r_3=2,$: $r_4=2$
Design 2: $r_0=3,$: $r_1=3,$: $r_2=2,$: $r_3=2,$: $r_4=2$
Design 3: $r_0=2,$: $r_1=3,$: $r_2=3,$: $r_3=2,$: $r_4=2$

6

2. For a general block design, obtain the rank of the information matrix X'X (you can assume the form of X'X)

6

3.Suppose in a block design with b blocks each of size v, in the first b_1 blocks treatment 1 appears twice and each of the treatments 3,.....,v, appears once. In the remaining b_2 blocks, each of the treatments 1,......,v appears once, ($b_1 + b_2=b$).
Is this design connected? Orthogonal? (Give reasons in support of your answer)
Obtain the C- matrix of the design and BLUE of any pair wise treatment comparison in terms of the adjusted treatment totals.

2+2+7+7=18

INDIAN STATISTICAL INSTITUTE

Mid-Sem Examination, 2^{nd} Semester, 2013-14

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

Date: February 28, 2014 Time: 2 hours 15 minutes

This paper carries 50 marks. Answer all questions.
Use separate answer scripts for each group.

Group A

1. Suppose X, Y and Z are independent and identically distributed binary random variables such that

$$P(X = 0) = P(Y = 0) = P(Z = 0) = 0.5.$$

Define a random variable W as follows:

$$W = \begin{cases} X & \text{if } Z = 0 \\ Y & \text{if } Z \neq 0 \end{cases}$$

Compute $P(W = 0)$ and $P(W = X)$ [2+4]

2. Consider a rare disease with prevalence 1 in 1000. The disease is determined through a clinical test that has an error probability of detection (or non-detection) equal to 0.01. What is the probability that an individual has the disease given that the test is positive? Comment on the probability value you obtain. [6+3]

3. Suppose $X_1, X_2, \ldots, X_n$ be a random sample from $Ber(p)$. Is it possible to construct a test for $H_0 : p = 0.5$ vs $H_1 : p = 0.75$ such that the probabilities of Type I error and Type II error are both less than or equal to 0.05? Justify your answer. [10]

Group B

4. The Michaelis-Menten model of enzyme kinetics relates the rate V of a reaction to the concentration S of a substrate. Based on n independent enzymatic reactions, consider a regression model as follows:

$$V_i = \frac{\beta_0\ S_i}{\beta_1\ +\ S_i} + e_i\ ;\ i = 1, 2, \ldots, n$$

where, e_is are random errors with mean 0 and the same variance. Describe a suitable algorithm to estimate β_0 and β_1. Show all computational steps clearly. [10]

5. Suppose the positions of pea plants are uniformly distributed over a circular field with radius r and coordinates of the centre (0,0). If $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ denote the coordinates of n randomly selected pea plants, obtain the maximum likelihood estimator of r. Examine whether this estimator is (i) sufficient, (ii) unbiased and (iii) consistent for r. [2+2+3+3]

6. Suppose the times to completion of drying fully loaded clothes by dryers of a specific brand are distributed as exponential. However, the exact times to completion are not observed, but recorded at the end of every minute (i.e., if drying is completed in between 30 minutes and 31 minutes, the time will be recorded as 31 minutes). If the recorded times to completion of three dryers of this brand to dry fully loaded clothes are 34 minutes, 41 minutes and 39 minutes, obtain the maximum likelihood estimate of the median time to completion. [5]

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

Semestral Examination

Time: 3 Hours   Full Marks: 65

Date: April 29, 2014

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. Consider a discrete time branching process $\{X_n\}$ with offspring size taking values 0, 1, 2 and 3 with probability 1/8, 3/8, 3/8 and 1/8 respectively. Show that $\{X_n\}$ is a Markov chain and clearly state its state space and transition probability matrix. [7]

2. For an irreducible finite state Markov chain, show that the period is given by the number of eigenvalues of unit modulus. [9]

3. Let $\{X_n\}$ be an i.i.d. sequence of unit exponential random variables. Show that $\sum_n X_n/\lambda_n$ diverges a.e. iff $\sum_n \lambda_n$ diverges. [9]

4. For a nonhomogeneous Poisson process with intensity parameter $\alpha(t)$, let T_1 be the time until the first occurrence and T_2 be the time between the first two occurrences. Find the joint distribution of (T_1, T_2). [9]

5. In a friendly football match, the goalkeeper of the B-III team played very well, but became nervous every time he let in a goal. Initially, the probability that he let in a goal in a period of length h was $\lambda h + o(h)$, but once he let in k goals, it modified to $(\lambda + \alpha k)h + o(h)$. Identify the process of number of goals let in by the keeper till time t and calculate the probability that he conceded at least two goals in a match of T time units. [10]

6. In Progressive Institute of General Studies, the exams are held out of a maximum marks of 5 and the students can obtain nonnegative integer valued scores. A score of 0 or 1 is considered as fail, while the rest are considered as pass, with a score of 4 or 5 is considered as distinction. Anyone getting distinction in the initial exam is allowed to keep that score as the official one and not repeat. However, the authorities have decided against failing any student and ask all students without distinction to repeat the exams till he obtains a pass score. The official score of a student is the maximum of his score in the initial exam and that in the final exam, but it is truncated at 3, if he repeats the exams and the maximum exceeds 3. The score a student forms a Markov chain where the initial score is i with probability proportional to $6 - i$, for $i = 0, 1, 2, 3, 4, 5$ and the transition matrix is given by

$$\begin{pmatrix} \frac{1}{3} & \frac{4}{15} & \frac{1}{5} & \frac{2}{15} & \frac{1}{15} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{6} & \frac{1}{6} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Give the probability distribution of the official score obtained by a student. [15]

Continued Overleaf ...

7. Due to the stress of coping with the excessive pressure of studies, a BStat third year student begins to experience migraine of random severities, which attack at times following a homogeneous Poisson process of rate λ. The severities are i.i.d. random variables with common distribution function H and independent of the Poisson process. If the severity exceeds the threshold c, the student visits the medical centre, where the excellent doctors cure her immediately of the present attack, but the migraines continue to return as before. Let $X(t)$ denote the number of visits to the medical centre by the student to get migraine treatment till time t. Describe the process $\{X(t) : t \geq 0\}$. [6]

INDIAN STATISTICAL INSTITUTE

Semestral Examination, 2013-14

Date: 02.05.14

B.Stat Third Year

Time : $3\frac{1}{2}$ Hours

Statistical Inference II

Full Marks: 100

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

1. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be independent and identically distributed with the density $f(\mathbf{x}) \propto |\boldsymbol{\Sigma}|^{-1/2} \psi\{-\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\theta})\|\}$, where $\boldsymbol{\theta} \in R^d$, $\boldsymbol{\Sigma}$ is a $d \times d$ real non-singular matrix, $\|\cdot\|$ denote the usual Euclidean norm, and ψ is a real measurable function defined on $[0,\infty)$. Assume that the distribution has finite second moments.

   (a) Define $\mathbf{Y} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\theta})$, where $\mathbf{X} \sim f$. If $\mathbf{H}$ is a $d \times d$ orthogonal matrix, show that the distribution of $\mathbf{HY}$ does not depend on $\mathbf{H}$. [4]

   (b) Define $\bar{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$. Show that $E(\bar{\mathbf{X}}) = \theta$ and $Var(\bar{\mathbf{X}}) = \frac{C}{n}\,\boldsymbol{\Sigma}$, where C is a positive constant. [2+6]

   (c) Show that the constant C in (b) is given by $C = E(\|\mathbf{Y}\|^2)/d$. [4]

   (d) If $\psi(z) = \exp(-z)$, using polar transformation or otherwise, find the distribution of $\|\mathbf{Y}\|$. Find $E(\|\mathbf{Y}\|^2)$ and hence check whether $\mathbf{S} = (n-1)^{-1}\sum_{i=1}^{n}(\mathbf{X}_i-\bar{\mathbf{X}})(\mathbf{X}_i-\bar{\mathbf{X}})'$ is unbiased for $\boldsymbol{\Sigma}$. [4+4]

   (e) Show that the coordinate wise median $\tilde{\mathbf{X}}$ computed from $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is unbiased for $\boldsymbol{\theta}$. [6]

   (f) Show that $\boldsymbol{\theta}$ is the spatial median and also the half-space median of the distribution. [3+3]

2. Suppose that $X_1, X_2, \ldots, X_m$ are independent and identically distributed with density function f and $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed with density function g, where $f(x-\theta) = g(x)$ for all x and some $\theta \geq 0$.

   (a) Show that the Mann Whitney statistic $T_{MW} = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{n} I(X_i < Y_j)$ is a consistent estimator of $P(X < Y)$, where $X \sim f$, $Y \sim g$ and they are independent. [6]

   (b) Is T_{MW} the uniformly minimum variance unbiased estimator for $P(X < Y)$? Give a brief justification for your answer. [2]

   (c) For testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$, show that the power of the two sample test based on T_{MW} converges to 1 as $\min\{m,n\}$ tends to infinity. [4]

   (d) Consider a linear rank statistic of the form $T = \sum_{i=1}^{N} a_i Z_i$, where $N = m+n$, $a_i = \Phi^{-1}(\frac{i}{N+1})$ and Z_i is the indicator variable that takes the value 1 if the i-th order statistic associated with $\{X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n\}$ comes from f. Here $\Phi(\cdot)$ is the distribution function of the standard normal variate. Check whether the null distribution of T (i.e., the distribution of T when $\theta = 0$) is symmetric. [4]

3. In the Institute for Scientific Investigation, a student failed to score the pass mark 35 in an examination in the final semester, and the student was asked to discontinue the course. Mentioning illness as the main reason for poor performance, the student requested the authority for another opportunity. After carrying out an investigation, the Dean of student affairs discovered that the semestral question paper was not moderated. As a result, the exam was scrapped, and all students were asked to appear for a new examination. The scores of the students in these two examinations are given below.

Student	1	2	3	4	5	6	7	8	9	10	11	12
Semestral exam	64	30	52	70	42	87	69	91	75	79	58	47
Re-exam	70	45	60	75	51	84	76	90	71	77	68	59

(a) Use a suitable nonparametric test to check whether the overall performance of the students was better in the re-examination. [6]

(b) Use a suitable nonparametric distribution-free test to check whether the students having lower scores in the semestral exam were benefited more by the re-examination. [8]

(c) Describe how the two-sample Kolmogorov-Smirnov statistic can be used to test the independence between two sets of scores. Will the resulting test have the distribution-free property? Justify your answer. [2+4]

4. (a) Consider a discrete distribution with $P(X=1)=\beta/(2-\beta)$ and $P(X=2)=2(1-\beta)/(2-\beta)$, were $0<\beta<1$. Show that there does not exist any unbiased estimator of β based on a fixed sample size n, but it is possible to construct an unbiased estimator of β if an appropriate sequential method is used. [3+3]

(b) Consider the uniform distribution $U(\theta,\theta+1)$, where $\theta\in(0,\ 1)$ is unknown. Suppose that we want to test $H_0:\theta=0.4$ against $H_1:\theta=0.6$. Show that it is not possible to construct any test based on fixed sample size, which has probabilities of type I error and type II error both equal to zero, but it is possible to construct a sequential test which has this property. Draw the OC function and the ASN function for that sequential test for θ varying in the domain (0,1). [2+2+4+4]

(c) Consider a univariate normal distribution with unknown mean μ and unknown variance σ^2.

(i) Describe how you will construct a confidence interval of a fixed length l_0 for μ, which has the coverage probability 0.95. Give an appropriate mathematical justification for your proposed method. [4+6]

(ii) Describe how you will construct a sequential test of strength (α,β) for $H_0:\mu=0$ against $H_1:\mu\neq 0$ considering $\mathcal{Z}=\{(\mu,\sigma):0<|\mu|<0.05\ \sigma\}$ as the zone of indifference. [8]

INDIAN STATISTICAL INSTITUTE
Semester Examination: (2013-2014)
B. Stat. III Year
Database Management Systems

Date: 05.05.14 Maximum Marks: 100 Duration: 3.5 Hours

1. For the relation R=(A,B,C,D,E,F), the following dependencies are specified :
   $A \rightarrow BC$, $BC \rightarrow D$, $CD \rightarrow E$, $EA \rightarrow F$
   Considering the above dependencies,
   a) derive all possible candidate keys of R;
   b) explain whether the relation R is free from partial and transitive dependencies;
   c) if necessary, decompose R into a set of normalized relations removing the partial and transitive dependencies;
   d) if a new multivalued dependency D->->F is introduced replacing the functional dependency $EA \rightarrow F$, what would be the new set of normalized relations?

(3+2+5+4=14)

2.

Time Slot	T1	T2	T3	T4
1		read(A)		
2	write(A)			
3				read(B)
4			write(B)	
5				read(A)
6		write(A)		
7			write(A)	

The above concurrent schedule involves four transactions T1 to T4 using two data items A and B for read and write.

a) Draw a precedence graph and check whether the above schedule is conflict serializable.

b) Draw the labeled precedence graphs for the two data items separately and check whether they are individually view serializable.

c) Also draw all possible composite labeled precedence graphs to determine whether the schedule is view serializable considering both the data items together.

(3+(5x2)+6=19)

3. Considering the concurrent schedule given in Question 2, determine whether the schedule is executable under two-phase locking protocol with upgrade facility? What would be the possible serial order of execution?

(7+4=11)

4. Consider that the schedule given in Question 2 is executed using timestamp ordering protocol where a transaction is rolled back for any time conflict. If TS(T) signifies the timestamp of any transaction T, timestamp ordering of the three transactions given in Question 2 is:
   TS(T1) > TS(T2) > TS(T3) > TS(T4)
   What would be the status of the four transactions at the end of the schedule and what would be the value of the read and write timestamps of the different data items. Assume that the initial value of all the read and write timestamps are less than TS(T4).

(8x4=32)

5. Three transactions T_o, T_1 and T_2 are executed as shown below. All the three transactions are manipulating the same data item A. If a crash occurs in one of the six places (1 to 6) as indicated in the schedule, explain the recovery action the system would undertake in each case, if it follows a deferred update log maintenance strategy with standard 'redo' and 'undo' routines when the log contains no check point. Now, if a check point is inserted in the log after the commit of T_o, explain what change will occur in the recovery process at the crash points 5 and 6. In each case show the content of the log as well.

<u>Schedule</u>:

T_o: read (A)
A=A-1
write(A)
-------------------(1)
T_1: read (A)
A=A+2
write(A)
-------------------(2)
commit
-------------------(3)
T_0 : commit
-------------------(4)
T_2 : read (A)
A=A*5
write(A)
-------------------(5)
commit
-------------------(6)

(8x3=24)

※

Semestral II
Design of Experiments
B.Stat. III yr. 2013-2014

Date: 13·05·14 | Full marks 100 | Time : 3.5 hours

Answer **any four** questions given below. Keep your answers brief and to the point. Marks will be deducted for unnecessarily long answers.

1. (a) Develop Bonferroni's method of multiple comparison procedure with respect to a Completely Randomised Design.

   (b) Show that every diagonal element of the C matrix of a connected block design is positive.

   (c) Prove that a block design with treatment-block incidence matrix N having the form $N = \frac{r\,k'}{n}$ is connected and orthogonal. (The notations have their usual meaning.)

   (d) Show that for a connected block design $\left(C + a\frac{rr'}{n}\right)$; $a > 0$, is always nonsingular and its inverse is a G -inverse of the C matrix.

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

2. (a) Prove that the number of mutually orthogonal Latin Squares of order v is at most $v-1$.

   (b) Let $S = \{L_1, L_2, \ldots, L_t\}$ be a complete set of mutually orthogonal Latin squares where

$$L_1 = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 4 & 5 & 6 & 7 & 1 \\ 3 & 4 & 5 & 6 & 7 & 1 & 2 \\ 4 & 5 & 6 & 7 & 1 & 2 & 3 \\ 5 & 6 & 7 & 1 & 2 & 3 & 4 \\ 6 & 7 & 1 & 2 & 3 & 4 & 5 \\ 7 & 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \quad \text{and} \quad L_2 = \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 5 & 6 & 7 & 1 & 2 \\ 5 & 6 & 7 & 1 & 2 & 3 & 4 \\ 7 & 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 7 & 1 \\ 4 & 5 & 6 & 7 & 1 & 2 & 3 \\ 6 & 7 & 1 & 2 & 3 & 4 & 5 \end{matrix} .$$

   Write down the other members of the set S above.

   (c) Suppose that p Latin Squares of order v have been used in p different locations with the same set of treatments but different sets of v^2 experimental units specific to the locations. Write down a suitable model for analysing the data and the corresponding ANOVA table indicating test statistics to test relevant hypotheses. (You do not have to derive the expressions of different sums of squares in the ANOVA table)

   [5+ 8+(2+10)=25]

3. The effect of four different lubricating oils (A, B, C, D) on fuel economy in diesel truck engines is being studied. Fuel economy is measured using brake-specific fuel consumption after the engine has been running for 15 minutes. Five different truck engines are available for the study.

(a) Suggest a suitable design to carry out the experiment ensuring estimability of the pairwise comparisons of the lubricating oils. Write down the underlying model.

(b) Suppose that at the analysis stage it has been found that the observation on the fuel consumption specific to the oil brand A and the truck 4 is missing. How will you estimate the missing value to obtain the correct value of the error sum of squares and the corresponding degrees of freedom under the model with the available observations.

(c) If the estimated value obtained in part (b) is used in place of the missing value and the augmented data are analysed to test the equality of the effects of the different oil brands, will that be correct? Justify your answer with mathematical proof.

[(3+2)+5+ 15=25]

4. (a) Show that for a 2^n factorial experiment in a confounding scheme $(2^n, 2^k)$ with blocks of size 2^k, where none of the main effects and two factor interactions is confounded, $n \leq 2^k - 1$.

(b) Give a balanced confounded scheme for a $(2^5, 2^3)$ factorial experiment in five replications retaining as much information as possible for the main effects and two factor interactions. Compute the loss of information for different factorial effects in the scheme suggested by you. For any one of the replications chosen by you, construct the key block and indicate one treatment for each of the other blocks in that replication.

(c) For a 3^3 factorial experiment define the main effects and interaction effects contrasts and the corresponding degrees of freedom. Show that the intra and inter effect contrasts are orthogonal.

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

5. (a) An experiment is performed to determine the effect of temperature(A) and heat treatment time(B) on the strength of normalised steel. Three temperatures and four heating times are selected. The pilot plant is capable of making 12 runs per shift and the experimenter wishes to continue this experiment for shifts.

   i. If the frequent change of the temperature setting is difficult to handle, suggest a suitable design for the experiment.
   ii. Clearly state the model for the design suggested in part(i).
   iii. Write down the BLUEs of pairwise comparisons of the effects of temperature(A) and pairwise comparisons of the effects of heating times(B) (you do not have to derive the BLUEs). Compute the variances of the BLUEs.
   iv. Which of the main effects of factor A and B are estimated with more precision?

(b) From the data given below compute the sums of squares due to the interaction component AB^2C.

Replication I					
block I		block II		block III	
Treatment	yield	Treatment	yield	Treatment	yield
0 2 2	58	1 2 2	89	1 1 1	82
1 1 2	96	1 1 0	88	0 0 0	71
2 2 0	91	2 2 1	77	2 2 2	75
1 0 0	96	0 1 1	69	0 1 2	61
0 0 1	59	1 0 1	87	1 2 0	73
1 2 1	82	2 1 2	83	2 0 1	72
2 1 1	89	0 0 2	80	0 2 1	64
0 1 0	71	0 2 0	77	2 1 0	79
2 0 2	89	2 0 0	87	1 0 2	86
Replication II					
1 0 2	83	0 1 0	80	2 0 2	91
2 1 2	89	1 2 0	93	0 1 2	68
2 0 1	83	0 0 2	85	1 2 2	92
0 0 0	76	2 2 2	103	0 2 0	72
2 2 0	83	2 1 1	104	1 0 0	85
1 2 1	87	1 0 1	101	2 1 0	82
0 1 1	79	0 2 1	75	1 1 1	96
1 1 0	87	2 0 0	95	0 0 1	54
0 2 2	82	1 1 2	89	2 2 1	83

[(2+2+(2+2+3+3)+1)+10=25]