INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2016-17

Course Name : B.Stat. 3rd Year

Subject Name : Sample Survey

Date : August 29, 2016 Maximum Marks : 30 Duration : 2 hrs.

Answer Q4 and any two from Q1 to Q3. Each question carries marks 10.

Notations are as usual.

1. In SRSWR of size n out of N, derive $Cov(\bar{y}, \bar{x})$ and an unbiased estimator of it.

   Obtain with proof an estimator of $\bar{Y}$ possibly better than usual sample mean for SRSWR.

2. Show that the variance of the linear systematic sample mean is

$$\frac{\sigma^2}{n}\left[1 + (n-1)\rho_c\right],$$

   where σ^2 is the population variance and ρ_c is the intraclass correlation coefficient.

   Show with proper proof how the units in the population should be arranged in order that systematic sampling will be much more efficient than SRSWOR.

3. Write down the unbiased estimator for population mean in a stratified random sampling with SRSWOR for within stratum sampling and the variance of that.

   Derive Neyman's optimum allocation and Bowley's proportional allocation formulae for startified random sampling.

   Prove that if finite population corrections are ignored, then

$$V(\hat{\bar{Y}}_{Neyman}) \leq V(\hat{\bar{Y}}_{prop}) \leq V(\hat{\bar{Y}}_{srswor}).$$

Please Turn Over

4. The following figures relate to a group of 15 households.

Serial No.	HH Size	Expenditure last month (Rs.)
1	8	5470.35
2	6	2716.80
3	5	1873.75
4	4	1693.20
5	3	1393.55
6	6	2398.74
7	2	3153.35
8	5	2708.75
9	7	2873.60
10	6	3775.80
11	8	5027.25
12	3	1175.28
13	4	2952.15
14	2	1032.27
15	2	2075.41

Based on an SRSWOR sample of size 6 drawn from above population data, give an estimate of the monthly per capita expenditure of those 15 households. Also give an estimate of the standard error, and the coefficient of variation in percentage.

Clearly, write down your formulae.

Random Number Table

97	50	71	35	65	67	15	45	73	19	17	60	68	38	50
96	17	27	35	82	80	77	28	97	11	26	72	21	88	96
21	48	84	49	72	93	48	66	75	82	36	33	57	97	35
85	12	90	36	72	81	62	73	40	20	38	10	81	34	44
49	57	40	54	64	88	97	69	37	12	94	45	86	74	66
97	43	79	37	60	96	75	39	46	33	42	41	29	83	73
80	71	51	15	59	55	24	80	49	12	61	68	40	44	58
40	81	81	93	32	35	60	29	42	53	38	35	54	67	73

INDIAN STATISTICAL INSTITUTE

Midsemester Examination : (2016-2017)

B. Stat 1st Year

Statistical Methods - I

Date: 30. 08. 2016 Maximum marks: 60 Duration: 2 hours.

Note: This paper carries 65 points. Maximum you can score is 60.

1. (a) In a (hypothetical) study, it was found that the percentage of incidence of hypertension was less for people who regularly exercised. The study concluded that regular exercise prevents hypertension. Give reasons for your agreement or disagreement.

   (b) Dividing the two groups (people who regularly exercise and who don't) into young and old, illustrate with hypothetical numbers how the percentage of incidence of hypertension can be higher for people who regularly exercise whether young or old, even when the overall percentages agree with the study.

   (c) Explain briefly how will you design a statistical study to check their conclusion? [3+10+5]

2. Suppose there are two sets of values of a variable x with n_1 and n_2 values respectively.

   (a) Let the geometric means of the two sets be G_1, and G_2 respectively. Show that the geometric mean of all $n_1 + n_2$ values taken together will lie between G_1 and G_2.

   (b) Suppose the harmonic means of the two sets are H_1 and H_2 respectively. Will the harmonic mean of all $n_1 + n_2$ values taken together will lie between H_1 and H_2? Justify your answer. [5+7]

3. Using the Cauchy-Schwarz inequality or otherwise, prove the following results:

   (a) Gini's mean absolute difference cannot be greater than $\sqrt(2)$ times the standard deviation of a variable.

   (b) If the measure of Kurtosis $\beta_2 = \frac{m_4}{m_2^2}$ and the measure of skewness $\beta_1 = \frac{m_3^2}{m_2^3}$ for a variable, then $\beta_2 \geq \beta_1 + 1$. [5+10]

4. The first four moments of a distribution about the value 4 (i.e. $\frac{1}{n}\sum_{i=1}^{n}(X_i - 4)^r, r = 1, \ldots, 4$) are -1.5, 17, -30 and 108, respectively. Find the first four central moments. [10]

5. Assignments [10]

INDIAN STATISTICAL INSTITUTE

B.STAT.

III Year

Mid-Semester Examination

2016

Full Marks: 25

Answer any 5 of the 6 questions given

Marks are provided in the parentheses

Time: 1 hr. 30 minutes

31·08·2016

1. Anthropology is the holistic study of man. Briefly and precisely illustrate this statement. [5]

2. Provide a brief evaluation of the nature and extent of relationship between Biological Anthropology and Statistics. [5]

3. Choose the right answer (by encircling the right option) [½ x 10 = 5]

(a) Physical anthropology is the study of human...................... and human ....:
[Bones; Nature/ Evolution; Variation/ Language; Culture/ Pottery; Stone tools]

(b) Which of the following is **not** a uniquely human activity:
[Gossiping with friends/ Making a book shelf/ Growing corn/ Climbing a tree]

(c) Which of the following would a physical anthropologist not study to know about humans?
[Pottery and stone tools/Bones and teeth/ Disease and nutrition/
Genes and reproduction]

(d) Physical anthropology and biological anthropology are equivalent: True/ False

(e) An anthropologist researches the Turkana pastoralist of Kenya, investigating both the genetic changes that allow them to easily digest milk, and the roles that the dairy animals have played in their history. The anthropologist has adopted which of the following methods?
[Sociolinguistics/ Biocultural approach/ Interdisciplinary science]

(f) Physical anthropologists' contemporary focus largely rests on Africa where humans evolved: [True/ False]

(g) Biological anthropology deals with all aspects of human biology, both past and present: [True/ False]

(h) The four branches of anthropology are:

[Linguistic anthropology; cultural anthropology; palaeontology; biological anthropology/

Biological anthropology; cultural anthropology; linguistic anthropology, archaeological anthropology or Palaeoanthropology/

Physical anthropology; ethnography; cultural anthropology; archaeology]

(i) Biological anthropology may best be defined as:

[The study of human disease and illness/The study of human culture/The study of primates/ The study of humans as biological organisms]

(j) Osteology is the study of:

[Human survival strategies/ Human languages/ Human skeleton/ Human brain]

4. Define 'culture'. State why man is called a bio-cultural being? Explain what is meant by cultural variation? [1+2+2]

5. Mention various stages of hominid evolution in their proper chronological order along with their space and time dimension. [5]

6. What is the significance of field work in Biological Anthropology? What are the features of a good anthropological field researcher? [2+ 3]

INDIAN STATISTICALK INSTITUTE

Mid-Semester Examination; Geology Optional course, BIII,

Full Marks 40 Time 2hrs

31 · 8 · 2016

1. How do sediments move once they are lifted into the flow? Explain with the help of a simple diagram. -- 4
   Describe the formation of igneous, sedimentary and rocks. How do minerals become rocks, and rocks become soil? ------------------------------2+2

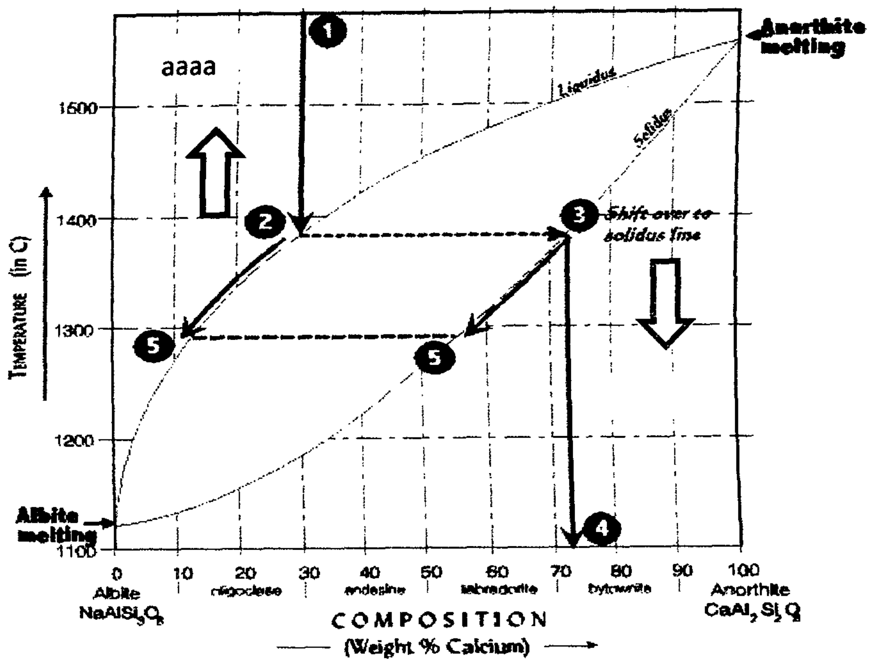
   Which of the following best defines a mineral and a rock? --------------- 2
   A) A rock has an orderly, repetitive, geometrical, internal arrangement of minerals; a mineral is a lithified or consolidated aggregate of rocks.
   B) A mineral consists of its constituent atoms arranged in a geometrically repetitive structure; in a rock, the atoms are randomly bonded without any geometric pattern.
   C) In a mineral the constituent atoms are bonded in a regular, repetitive, internal structure; a rock is a lithified or consolidated aggregate of different mineral grains.
   D) A rock consists of atoms bonded in a regular, geometrically predictable arrangement; a mineral is a consolidated aggregate of different rock particles.

2. What do you understand by the term "Mohs' hardness scale". Arrange the minerals in order of their hardness and chemical formulae. Why do the colour and streak of a mineral may vary? -- 3+2+5=10

Mineral	Hardness	Chemical formula
Diamond	6	Pb,Ag,As,Sb,Cu)S
Corundum	1	$KAl_3Si_3O_{10}(OH)_2$
Quartz	7	C
Feldspar	10	$KAlSi_3O_8 - NaAlSi_3O_8 - CaAl_2Si_2O_8$
Mica	9	Al_2O_3
Beryl 5.5	7.5-8	$Mg_3Si_4O_{10}(OH)_2$
Talc	2.5	$Be_3Al_2(Si\ O_3)_6$
Galena	2.5-2.75	SiO_2

P.T.O

3. Study the diagram below and state



a. What do you understand by the term "solidus" and "liquidus"?
Fill in the blanks

b. The upward arrow indicates the field of ----------- physical state of the material concerned

c. The downward arrow indicates the field of ----------------- physical state of the material concerned

d. The grey field indicates a mixture of ------------ and -------------- physical states of the material concerned

e. Explain what is represented by the numbers 1 to 5 and the lines with arrows present in the diagram.

f. What is Gibb's phase rule? ------------------------------------ 2+1+1+1+3+2 = 10

4. What are the three different types of plate boundaries? Why are the Andes and the Rocky Mountains are noted on the western sides of North and South America? Why the volcanos of the world are arranged in chains? How do the earth quake waves help us to decipher the layered structure of the earth? -- 3+1.5+1.5+4 = 10

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination : (2016-2017)

B. Stat. II Year

Physics III

Date: 31.08.16 Maximum Marks 30 Duration 2 hour

1.(a) In an inertial frame a child was born and died after 1 hour at same place. Consider another inertial frame having relative velocity $\frac{4c}{5}$. Calculate the life span of the child with respect to an observer in the later frame.

(b) A person standing at the rear of a train fires a bullet towrads the front of the train. The speed of the bullet, as measured in the frame of the train, is 0.5C and the proper length of the car is 400m. The train is moving at 0.6C as measured by observers in the ground.
What do ground observers measure for
i) the length of the railroad car,
ii) the speed of the bullet,
iii) the time required for the bullet to reach front of the car.

$2 + (2 + 2 + 2)$

2(a) With respect to an inertial observer, two particles from far ends are moving towards each other. The distance between the particles is decreasing at a rate of $\frac{7c}{6}$, where the velocity of one of the particles is $\frac{2c}{3}$ with respect to the inertial observer. Find the relative velocity between the particles.

(b) If two clocks separated by a distance D are synchronized in an inertial frame, show that in another inertial frame in which the clock moves along the line joining them with velocity v, the reading of the clock in front is behind the cloch in the rear by $\frac{Dv}{c^2}$.

$3 + 5$

3(a) In an inertial frame, two events have the space time coordinate $\{x_1, y, z, t_1\}$ and $\{x_2, y, z, t_2\}$ respectively. Let $x_2 - x_1 = 3c(t_2 - t_1)$. Consider another inertial frame which moves parallel to x-axis with velocity u w.r.t. the first one. Find the value of u for which the the events are simultaneous in the later frame.

(b) Let a particle is rest in an inertial frame and m_0 is the rest mass of a particle. A force is applied on the particle and let after time t it's velocoty is v. Find the kinetic energy of the particle.

3 + 5

4.(a) Show that for photon, the wavelength is given by $\lambda = \frac{h}{p}$, where p is the momentum and h is Planck constant.

(b) Show that if a photon with wavelength λ is scattered by an electron in metal, by angle θ, then

$$\lambda' - \lambda = \frac{h}{mc}(1 - cos\theta)$$

λ' being the wavelength of the scattered photon.

2 + 6

(In all the problems, c represents the velocity of light in vacuum)

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination (2016-2017)

Subject: Introduction to Sociology

B. Stat. III Year

Date: 31·08·16 Maximum Marks: 20 Duration: Two hours

The figures in the margin indicate full marks

Answer any four of the following questions:

Q 1. What is Sociology? Delineate the difference between Common Sense and Sociology. 2+3 =5

Q 2. What is Rural Sociology? What is meant by self-sufficient Village? 2+3 =5

Q 3. What is Development? Critically analyse the main points in the debate on models of development in India. 2+3 =5

Q 4. What is case study? Explain its use in social research. 2+3 =5

Q 5. Distinguish between inductive and deductive forms of logic. Provide suitable examples. 2+3 =5

Q 6. Explain Survey research. Point out its advantages and weakness. 2+3 =5

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2016-2017

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

Linear Statistical Models

Date: September 01, 2016 | Maximum Marks: 35 | Duration: 2 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. Consider the linear model $\boldsymbol{Y} \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$, $\mathbf{X}$ is $n \times p$, $\boldsymbol{\beta}$ is $p \times 1$. Let $\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\beta}$ be estimable, where $\boldsymbol{\Lambda}$ is a known $p \times q$ matrix, $\mathrm{rank}(\Lambda) = q\,(< p)$. Also, let $\mathrm{rank}(\mathbf{X}) = r$ and suppose $r < n$. Suppose we wish to test the hypothesis $\mathrm{H}_0 : \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{0}$ *versus* $\mathrm{H}_1 : \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\beta} \neq \mathbf{0}$.
   (a) Explain how you can formulate this problem as one of testing a linear model against a reduced model.
   (b) Show that the ANOVA based test statistic for this problem is given by
   $$F := \frac{\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{\Lambda}\left[\boldsymbol{\Lambda}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\boldsymbol{\Lambda}\right]^{-}\boldsymbol{\Lambda}^{\mathrm{T}}\hat{\boldsymbol{\beta}}/q}{MSE},$$
   where $\hat{\boldsymbol{\beta}}$ is a solution of the system of normal equations. [You may assume and use results related to testing a linear model against a reduced model.]
   (c) Find from first principles the null distribution of F. [3+8+6=17]

2. Consider a balanced one-way ANOVA model given by
   $$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ j = 1, \ldots, N,\ i = 1, \ldots, 5,$$
   $$\epsilon_{ij}\text{'s are i.i.d. } \mathrm{N}(0, \sigma^2).$$
   Suppose we wish to test the hypothesis $\mathrm{H}_0 : \alpha_1 = \alpha_2, \alpha_1 = \alpha_3$ *versus* $\mathrm{H}_1 : \mathrm{H}_0$ *is false.* Derive the ANOVA based test statistic for this problem. [9]

3. Consider a balanced two-way ANOVA model without interaction given by
   $$Y_{ijk} = \mu + \alpha_i + \eta_j + \epsilon_{ijk},\ k = 1, \ldots, N,\ i = 1, \ldots, 2,\ j = 1, \ldots, 3, \tag{1}$$
   $$\epsilon_{ijk}\text{'s are mutually uncorrelated with mean } 0 \text{ and variance } \sigma^2\,(> 0).$$
   Let $\boldsymbol{Y} := (Y_{111} \ldots Y_{11N}\, Y_{121} \ldots Y_{12N} \ldots\ldots Y_{2,3,1} \ldots Y_{2,3,N})^{\mathrm{T}}$ and we define $\boldsymbol{\epsilon}$ in a similar manner. We express the model in (1) as $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is a suitable $6N \times 6$ matrix. Find, with adequate reasons, the BLUE of $\mathbf{X}\boldsymbol{\beta}$. [9]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : 2016-2017

B. Stat. (Hons.) III Year

Subject : SQC & OR

Full Marks: 100 **Time: 3hours** **Date of Examination: 05 Sept 2016**

NOTE: *This paper carries **105** marks. You may answer any part of any question; but the maximum you can score is **100**.*

1. A machine shop has one drill press and five milling machines, which are used to produce an assembly consisting of two parts P1 and P2. The productivity of each machine for the two parts is given below:

	Production time in minutes per piece	
Part	**Drill**	**Mill**
P1	3	20
P2	5	15

It is desired to maintain a balanced loading on all machines such that no machines runs for more than 30 minutes per day longer than any other machine. Assume that the milling load is split evenly among all five machines. All the machines run for 8 hours each working day.

The profit obtained by selling one piece is given below:

Item	Profit (Rs/item)
P1	10
P2	20
Assembly	50

The machine shop wants to maximize its profit. Formulate this as an LP problem.

[15]

2. Consider the following set of equations:

$$x_1 + x_2 + 4x_3 + 2x_4 + 3x_5 = 8$$
$$4x_1 + 2x_2 + 2x_3 + x_4 + x_5 = 4$$
$$x_i \geq 0, \qquad i = 1, \cdots, 5$$

Find

a) a basic feasible solution
b) a basic solution which is not feasible
c) a solution which is neither basic nor feasible.

d) a system of equations which produces infinitely many solutions.

e) a system of equations which does not admit of a solution.

f) the number of possible basic solutions.

[2+2+3+2+2+1=12]

3. Consider the following LP:

Maximize $x_0 = -5x_1 - 21x_3$
Subject to

$$x_1 - x_2 + 6x_3 \geq 2$$

$$x_1 + x_2 + 2x_3 \geq 1$$

$$x_i \geq 0, \quad i = 1, 2, 3.$$

A student wants to solve the above problem by the simplex method. Find the initial basic feasible solution for him.

[13]

4. The viscosity of an aircraft primer paint is an important quality characteristic. The product is produced in batches, and as each batch takes several hours to produce, the production rate is too slow to allow sample sizes greater than one. The viscosity of the previous 15 batches is shown below:

Table 4.1: Viscosity of Aircraft Primer Paint

1	33.75
2	33.05
3	34.00
4	33.81
5	33.46
6	34.02
7	33.68
8	33.27
9	33.49
10	33.20
11	33.62
12	33.00
13	33.54
14	33.12
15	33.84

Find the control limits for an appropriate control chart to monitor this process.

[20]

5. ABCD Ltd. is a company that packages materials such as liquid soap, deodorant and cooking oil for large customers. The company fills the liquids into bottles or cans in its

filling lines, seals and labels them, packages them in cartons and ships them to customers' warehouses. Sometimes, they do some mixing operations before or during filling the liquids into the containers. Customers' quality requirements include the appearance and strength of bottles and cans and a very demanding set of specifications for fill volume or fill weight of the contents of the bottles and cans. Their Quality Engineer has collected the following costs related data for the past six months:

Table 5.1: Quality Cost Data

Sl. No.	Account Head	Amount (Rs)
1.	Company-owned material loss (because of overfill, changeovers, spills)	507,000
2.	Rework labour	214,200
3.	Customer-owned material loss (This category involves costs billed by the customers. ABCD 'pay back' their customers for over using their materials, which occurs during "reconciliation" meetings.)	107,000
4.	Cost of producing off-quality products	77,300
5.	Cost of raw material destroyed because of off-quality rejection	24,700
6.	Filled stock destroyed	4,700
7.	Cost of disposing crushed cans	3,300
8.	Cost of re-inspecting products that have been re-worked	3,300
9.	Chemical QA	164,600
10.	Physical QA	162,800
11.	Inspection and testing	119,000
12.	Materials & services consumed during testing	30,000
13.	Calibration of test equipment	12,500
14.	Process control	119,000
15.	New products review	28,700
16.	Improvement projects	33,800
17.	Quality reporting	8,500

Sales revenue for the corresponding period was Rs 24,287,000; while the profit for the corresponding period was Rs 1,005,000.

a) Find the different quality costs based on the above data.
b) Express quality costs as a percentage of sales.
c) Express failure cost as a percentage of profit.
d) If there were no quality costs, what would be the profit of the company during these six months.
e) On which two areas should the company direct its efforts to reduce quality costs.

[12+2+2+2+2=20]

6. Consider the following LP:

Maximize $W = y_1 + 2y_2 + 3y_3$

Subject to

$$y_1 + y_2 \leq 2$$
$$y_1 - y_2 + y_3 \geq 1$$
$$-y_1 + y_2 + y_3 = -1$$
$$y_1 \text{ unrestricted in sign}; y_2 \geq 0;\ y_3 \leq 0$$

a) Write down the dual of the above problem.

b) Does the Weak Duality Theorem hold for this pair of primal-dual problems? Give reasons.

c) Does this pair of primal-dual problems admit of optimal solutions? Give reasons.

[10+3+3=16]

7. Choose the best answer. (You need not copy the statements.)

a) The chart which aggregates poor quality outcomes to show management which are the most important problems is the:

i. p chart
ii. Pareto chart
iii. R chart
iv. c chart

b) The distribution of the sample range is?

i. Normal
ii. always Poisson
iii. not normal
iv. always binomial

c) In an actual quality control problem, the first test would be on the:

i. mean
ii. variation
iii. number of defects
iv. average number of defects per unit

d) What does the phrase "in control" mean with respect to processes?

i. An in-control process is one in which the proportion of output that is defective falls within the agreed-upon range
ii. An in-control process is one in which the process width (i.e., 6σ) is substantially wider than the specification width (i.e., the upper specification limit minus the lower specification limit)
iii. An in-control process is statistically stable; it is free of assignable or non-random variation

iv. An in-control process is statistically stable; it is free of unassignable or random variation

e) Using the terminology of statistical control, the variation within a stable system
   i. is random variation.
   ii. results from common causes.
   iii. is predictable within a range.
   iv. all of the above.

f) A customer service hotline has received an average of 7 complaints a day for the last 25 days. What type of control chart should be used to monitor this hotline?
   i. p-chart
   ii. c-chart
   iii. u-chart
   iv. $\bar{X} - R$ chart

g) Using the terminology of statistical control, the variation outside the control limits on an $\bar{X} - R$ chart
   i. is viewed as uncontrollable.
   ii. is assumed to have been caused by special or assignable causes.
   iii. indicates that the system is probably out of control.
   iv. ii and iii.

h) One type of error a manager can make is to blame a worker for an undesirable variation that is caused by the system. Refer to this as a type I error. Another type of error a manager can make is to blame the system when a worker caused the undesirable variation. Refer to this as a type II error. If a company changed the basis for the upper and lower limits on a control chart from three standard deviations to two standard deviations
   i. the number of type I errors would increase.
   ii. the number of type II errors would increase.
   iii. the number of both types of errors would increase.
   iv. the number of both types of errors would decrease.
   v. there is no basis for choosing an answer.

i) A predictable range of variation in the output of a particular worker occurs on a routine basis. This variation represents
   i. common cause variation and is uncontrollable.
   ii. common cause variation and is controllable.
   iii. assignable cause variation and is uncontrollable.
   iv. assignable cause variation and is controllable.
   v. none of these.

[1X9=9]

PTO for Table

Appendix VI Factors for Constructing Variables Control Charts

Observations in Sample, n	Chart for Averages: Factors for Control Limits			Chart for Standard Deviations: Factors for Center Line		Chart for Standard Deviations: Factors for Control Limits				Chart for Ranges: Factors for Center Line		Chart for Ranges: Factors for Control Limits				
	A	A_2	A_3	c_4	$1/c_4$	B_3	B_4	B_5	B_6	d_2	$1/d_2$	d_3	D_1	D_2	D_3	D_4
2	2.121	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606	1.128	0.8865	0.853	0	3.686	0	3.267
3	1.732	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276	1.693	0.5907	0.888	0	4.358	0	2.575
4	1.500	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088	2.059	0.4857	0.880	0	4.698	0	2.282
5	1.342	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964	2.326	0.4299	0.864	0	4.918	0	2.115
6	1.225	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	5.078	0	2.004
7	1.134	0.419	1.182	0.9594	1.0423	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.204	5.204	0.076	1.924
8	1.061	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751	2.847	0.3512	0.820	0.388	5.306	0.136	1.864
9	1.000	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707	2.970	0.3367	0.808	0.547	5.393	0.184	1.816
10	0.949	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.687	5.469	0.223	1.777
11	0.905	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.811	5.535	0.256	1.744
12	0.866	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.922	5.594	0.283	1.717
13	0.832	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	1.025	5.647	0.307	1.693
14	0.802	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	1.118	5.696	0.328	1.672
15	0.775	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	1.203	5.741	0.347	1.653
16	0.750	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	1.282	5.782	0.363	1.637
17	0.728	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	1.356	5.820	0.378	1.622
18	0.707	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496	3.640	0.2747	0.739	1.424	5.856	0.391	1.608
19	0.688	0.187	0.698	0.9862	1.0140	0.497	1.503	0.490	1.483	3.689	0.2711	0.734	1.487	5.891	0.403	1.597
20	0.671	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470	3.735	0.2677	0.729	1.549	5.921	0.415	1.585
21	0.655	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	1.605	5.951	0.425	1.575
22	0.640	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448	3.819	0.2618	0.720	1.659	5.979	0.434	1.566
23	0.626	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438	3.858	0.2592	0.716	1.710	6.006	0.443	1.557
24	0.612	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	1.759	6.031	0.451	1.548
25	0.600	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	1.806	6.056	0.459	1.541

For $n > 25$

$$A = \frac{3}{\sqrt{n}}, \quad A_3 = \frac{3}{c_4\sqrt{n}}, \quad c_4 \simeq \frac{4(n-1)}{4n-3},$$

$$B_3 = 1 - \frac{3}{c_4\sqrt{2(n-1)}}, \quad B_4 = 1 + \frac{3}{c_4\sqrt{2(n-1)}},$$

$$B_5 = c_4 - \frac{3}{\sqrt{2(n-1)}}, \quad B_6 = c_4 + \frac{3}{\sqrt{2(n-1)}}.$$

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2016–2017
B.Stat. (Hons.) 3rd Year. 1st Semester
Linear Statistical Models

November 07, 2016 — Maximum Marks: 50 — Duration: 3 hours

- This question paper carries 58 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.
- You may use scientific calculator for numerical calculation.

---

1. Consider the following balanced two-way ANOVA model with interaction and one covariate:

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + \beta z_{ijk} + \epsilon_{ijk},\ k = 1, \ldots, N,\ i = 1, \ldots, 3,\ j = 1, \ldots, 3,$$
$$\epsilon_{ijk}\text{'s} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2).$$

We wish to test the hypothesis $\mathrm{H}_0 : \gamma_{ij} = 0\ \forall\ i, j$. Derive for this hypothesis the ANOVA based test statistic and its null distribution. [Note. You need appropriate conditions, to be stated by you, on the z_{ijk}'s.] **[12+6=18]**

[You may assume without proof, and use, expressions for the fitted values of the Y_{ijk}'s, for both balanced two-way ANOVA models without interaction and with interaction.]

2. Consider the following balanced two-way ANOVA model without interaction:

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \epsilon_{ijk},\ k = 1, \ldots, N,\ i = 1, \ldots, 5,\ j = 1, \ldots, 3,$$
$$\epsilon_{ijk}\text{'s} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2).$$

We wish to obtain *simultaneous* confidence intervals for the contrasts $\psi_1 := \alpha_1 - \alpha_2 + \alpha_3 - \alpha_4$ and $\psi_2 := \alpha_1 + \alpha_2 - \alpha_3 - \alpha_5$ with the goal that the overall confidence coefficient be at least $1 - \alpha$.

(a) Describe how using Scheffe's method of multiple testing you can achieve this. **[6]**

(b) Explain why your method achieves the goal of confidence coefficient being at least $1 - \alpha$. **[9]**

[You may assume without proof the following: *For any symmetric and idempotent* $p \times p$ *matrix* $\mathbf{A}$ *and fixed* $\boldsymbol{x} \in \mathbb{R}^p$, $\sup \mid \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} \mid = (\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x})^{1/2}$, *where the supremum is taken over all* $\boldsymbol{u} \in C(\mathbf{A})$ *satisfying* $\|\boldsymbol{u}\| = 1$.]

3. Suppose at a university a student survey is carried out to ascertain the reaction to instructors' usage of a new computing facility. We suppose that all freshmen have to take English or Geology or Chemistry in their first semester. All three courses in the first semester are large and are divided into sections, each section with a different instructor and not all sections necessarily having the same number of students. In the survey, the response provided by each student is opinion (measured on a scale of 1 through 10) of his instructor's use of the computer. Some of the summary statistics are given in the table below. (Each column represents a section of a course.) Decide if all sections within the English course have the same opinions. [$F_{0.95,2,16} = 3.6337$ and $F_{0.99,2,16} = 6.2262$, where F_{γ,n_1,n_2} is the γ-the quantile of F-distribution with n_1, n_2 degrees of freedom.] [12]

Subject →	English			Geology			Chemistry		
Section →	1	2	3	1	2	3	1	2	3
Section size →	3	2	5	2	3	1	4	3	2
Section mean →	3	8	5	4	9	8	3	2	7
Section SS (corrected) →	6	2	16	8	2	0	14	2	2
Class (subject) size →	10			6			9		
Subject mean →	5			7.1667			3.5556		
Subject SS (corrected) →	54			40.8333			50.2222		
Total no. of students →	25								
Overall mean →	5								
Overall SS (corrected) →	192								

4. Blood sugar levels (mg/100g) were measured on 10 animals from each of five breeds. Some of the summary statistics are given in the table below. Determine which of the ten pairs of breeds have different blood sugar levels ($\alpha = 0.05$). [$F_{0.95,4,45} = 2.5787$, $F_{0.99,4,45} = 3.7674$, $F_{0.95,1,45} = 4.0566$, and $F_{0.99,1,45} = 7.2339$] [13]

Blood Sugar Levels (mg/100 g) for 10 Animals from Each of Five Breeds (A–E)					
Breed →	A	B	C	D	E
Group size	10	10	10	10	10
Group mean	117	118	129.9	114	134.5
Group SS (corrected)	756	1122	1138.9	910	1558.5
Overall mean	122.6800				
Overall SS (corrected)	8968.9				

INDIAN STATISTICAL INSTITUTE

First Semester Examination : 2016-17

Course Name : **B.Stat. 3rd Year**
Subject Name : **Sample Surveys**
Date : 09-11-, **2016** **Total Marks : 60** **Duration : 3 hrs.**

Answer Q.7 and any 4 from Q.1 to Q.6.
Notations are as usual.

1. What is called a super-population model ? Show with example how to estimate the population total Y of a variable of interest y and its error respectively by model-based and model-assisted approaches ?

   State the idea about the predictive approach for estimation of Y and give an example.

   (10)

2. (a) Describe D.B.Lahiri's method of drawing a random sample.

   (b) Prove that this method yields a PPS sample.

   (10)

3. (a) State the form of Horvitz and Thompson's (HT) estimator for the population total of a variable of interest and show that it is unbiased.

   (b) Derive the variance of HT estimator. Derive Yates and Grundy's form of the variance of HT estimator, clearly stating the conditions.

   (10)

4. Describe systematic sampling with probability proportional to size. How can the populational total Y of a variable of interest y be estimated based on a sample drawn by this scheme ? Describe the difficulty in this estimation and show how to overcome it.

   (10)

5. Derive an unbiased estimator for Y based on a two-stage SRSWOR-SRSWOR sampling. Show with proof how to estimate its error.

   (10)

6. Describe Politz and Simmon's technique of using 'at-home-probabilities' to estimate the population mean $\bar{Y}$ by SRSWR scheme. Obtain an estimator for the variance of this estimator.

   (10)

7. The following figures relate to the yield of a crop in quintlal for 12 plots of land (in acre). Take a **PPSWOR** of 4 plots with land area as size measure. Based on your sample, give an estimate of the **total yield of the crop** grown in these 12 land areas. Also provide an estimate of its standard error and coefficient of variation (cv) in percentage.

(20)

Plot Serial no.	Land area (Acre)	Yield of the crop (Quintal)
1	8.3	5620.33
2	6.0	2819.81
3	5.7	1763.72
4	4.1	1923.20
5	3.5	1381.55
6	6.2	2458.71
7	2.9	3043.35
8	5.4	2838.79
9	7.1	2774.69
10	6.7	3789.85
11	8.0	5187.27
12	3.2	1045.23

Random number table

25	19	17	50	50	46	26	92	62	41	27	66	85	60	70
54	61	41	41	91	88	83	30	32	75	59	73	58	58	83
97	50	71	35	65	67	15	45	73	92	17	60	68	38	50
96	17	27	35	82	80	77	28	97	11	26	72	12	88	96
21	48	84	49	72	93	48	66	75	82	36	33	77	97	35
85	12	59	36	72	81	36	73	14	82	33	10	81	34	44
49	57	40	54	64	88	97	69	43	12	94	45	86	74	66
17	43	79	37	60	96	75	39	46	33	42	41	29	83	73
80	47	51	15	59	55	24	80	49	12	61	68	60	44	58
40	71	81	93	23	52	60	49	42	53	38	35	28	67	73

INDIAN STATISTICAL INSTITUTE

Semestral Exam Optional Course – Geology, Time -2 hours

Date 11.11.2016, No. of Student – 7

---

1. Choose the correct answer from the following list 1x5=5

i) Detrital sedimentary rocks are classified (named) based on the

a) colours of the cementing minerals

b) grain sizes of the detrital particles

c) compositions of soluble minerals

d) degree of compaction and lithification.

ii) Flint, chert, and jasper are microcrystalline forms of.

a) quartz; (SiO_2)

b) hematite ($Fe2O_3$)

c) halite (NaCl)

d) calcite ($CaCO_3$)

iii) What major change occurs during metamorphism of limestone to marble?

a) calcite grains recrystallize to larger and interlocked grains.

b) clays crystallize to micas, forming a highly foliated, mica-rich rock

c) limestone grains react to form quartz and feldspars

d) calcite grains are dissolved away leaving only marble crystals.

iv) For undisturbed, horizontal strata of sedimentary rocks, their age

a) increases from top to bottom

b) decreases from top to bottom

c) can be determined from their color

d) is same.

v) The biozone defines

a) relative age of the host rock

b) absolute age of the host rock

c) mineral age of the host rock

d) whole-rock age of the host rock

2. Define a Triclinic crystal system. --------------------- 5

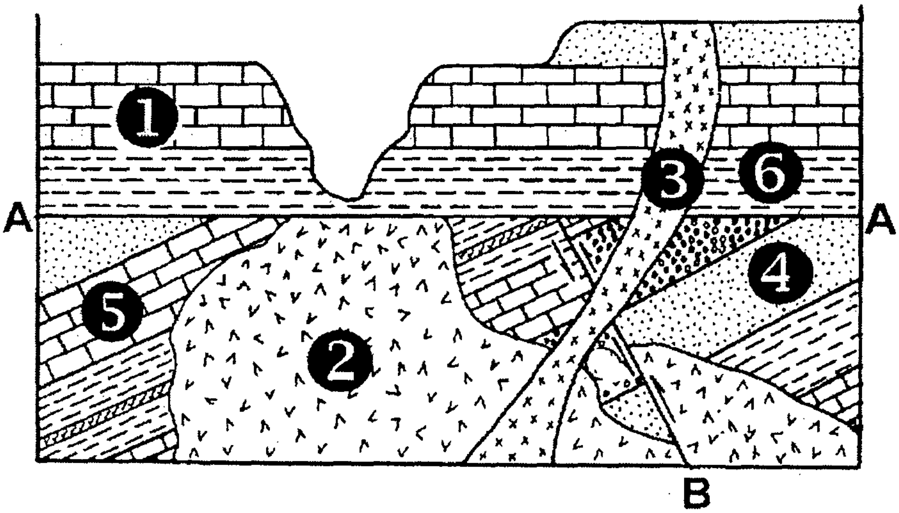
3. Define cyclosilicates citing one mineral as example. --------------------- 5

4. Write the classification of folds based on dip isogons with illustrations..---- 10

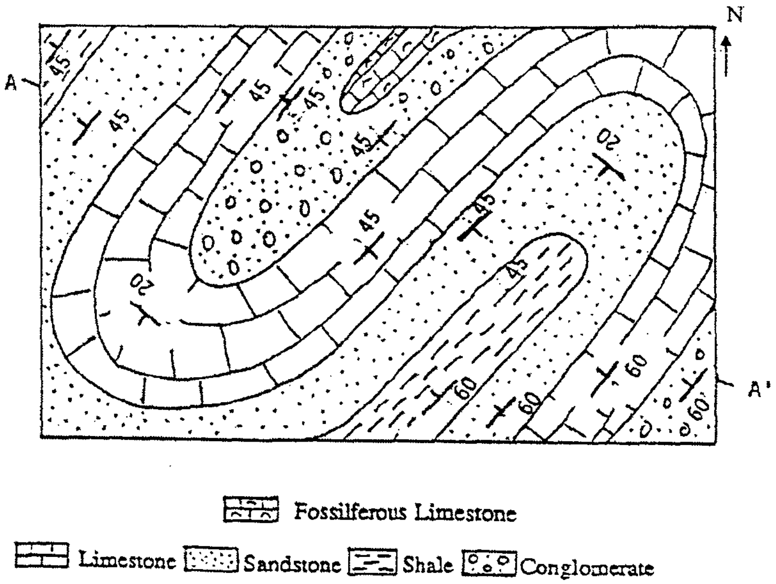
5. Write a short note on "Bowen's reaction series". ------------------ 5

6. Write the geological history of the road-cut section given below. ----------- 5

1 and 5 = limestone of different ages; 2,3 = igneous intrusions; 4 = shale, 6 = sandstone;



6. Interpret the given geological map considering the area as a plane land. -- 5



7. What are the physical conditions necessary for formation of a coral reef? -- 5

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2016-2017

Subject: Introduction to Sociology

B. Stat. III Year

Date: 11th November 2016 Maximum Marks: 50

Duration: Three hours

The figures in the margin indicate full marks

Answer any five of the following questions:

Q 1. Examine the gender perspective on development. 10

Q 2. Write any five questions (briefly): 2 x 5 =10

(a) What is Gram Sabha?
(b) What is meant by empowerment of women?
(c) What is the connotation of Bargadar?
(d) What is Questionnaire?
(e) What is schedule?
(f) Define status and role.
(g) Mention two marriage laws, enacted in post-Independence-Period, which have influenced the upliftment of women's Status in family?
(h) What is the meaning of measurement in research?

Q 3. What is research design? Explain the important functions of research design. 4 + 6 =10

Q 4. Discuss Durkheim's theory of suicide. 10

Q 5. Write short notes on any four: 2.5 x 4 = 10

(a) Social Statistics & Social Dynamics
(b) Caste system
(c) Modernization & Westernization
(d) Social movements
(e) Social mobility

(f) Domestic violence.

Q 6. What are the essential steps in a research procedure in Sociological study? Explain briefly 10

Q 7. Choose the correct answers: 2 x 5 =10

(a) "The quality of Life" was written by:
- (i) Ramkrishna Mukherjee
- (ii) Andre Beteille
- (iii) Parthanath Mukherji

(b) Who is the father of Sociology?
- (i) August Comte
- (ii) Ramkrishna Mukherjee
- (iv) Darwin

(c) What is the basis for the demarcation of the family into matriarchal and patriarchal?
- (i)Organization
- (ii) Authority
- (iii) Residence

(d) "Census sampling" means:
- (i) Covering 5% of the universe
- (ii) Covering 20 % of the universe
- (iii) Taking all the respondents available

(e) The concept of 'Sanskritization and Westernization' is attributed to:
- (i) M.N.Srinivas
- (ii) Ramkrishna Mukherjee
- (iii) Andre Beteille

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : (2016-2017)

B. Stat. III Year

Physics III

Date: 11. 11. 2016 Maximum Marks: 70 Duration: 3 hours

Answer any five questions

1 (a) Consider the following linear operator $\hat{A}$ acting on a two dimensional complex Hilbert space

$$\hat{A} = aI + \vec{n}.\sigma$$

where $\vec{n}.\sigma = \sum n_i\sigma_i$ $(i = x, y, z)$, a is a real number, $\vec{n}$ is a vector in R^3, σ's are Pauli matrices and I is identity operator.

i) Find the eigenvalues of A.

ii) Under what condition A is a projection operator?

iii) Under what condition A is a density operator?

(b) Let the state of a spin-1/2 system is represented by the following density operator

$$\hat{\rho} = \frac{1}{2}[I + \vec{n}.\sigma] \qquad |\vec{n}| \leq 1$$

Find the probability of finding up spin for spin measurement along some vector $\vec{m}$.

[(3 + 3 + 3) + 6]

2. (a) Consider two self adjoint operators $\hat{A}$ and $\hat{B}$ representing two observables of a quantum system. Establish the following relation;

$$\Delta_\psi \hat{A}.\Delta_\psi \hat{B} \geq \frac{1}{2}|\langle\psi|[\hat{A}, \hat{B}]|\psi\rangle|$$

where $\Delta_\psi \hat{A}$ is the standard deviation of $\hat{A}$ for quantum state $|\psi\rangle$ and similarly for $\hat{B}$.

(b) From the above result, establish Hiesenberg's uncertainty relation for position and momentum and discuss its implication.

[10 + 5]

3 Consider a particle in the potential $V(x) = \infty$ for $x \leq 0$ and $x \geq L$ and $V(x) = 0$ for $0 \leq x \leq L$. The wave function of the particle at $t = 0$ is the following;

$$|\psi(x)\rangle = \frac{2}{\sqrt{L}} \cos\left(\frac{3\pi x}{L}\right) \sin\left(\frac{4\pi x}{L}\right)$$

(i) If energy measurement is performed on the particle in this state, what are the possible energy values? Find the corresponding probabilities.
(ii) Find the wave function at $t = T$.
(iii) Argue why momentum eigen function for such system cannot be a possible state.

[8 + 5 + 2]

4 (a) Write the Hamiltonian operator $\hat{H}$ for a simple harmonic oscillator with angular frequency ω.

(b) Two operators $\hat{a}$ and $\hat{a}^\dagger$ can be defined in such a way that

$$\hat{H} = \left[\hat{a}^\dagger \hat{a} + \frac{1}{2} I\right] \omega\hbar$$

where $\hbar$ is the Planck constant.
i) Show that $[\hat{a}, \hat{H}] = \hbar\omega\hat{a}$
ii) Find the eigenvalues of $\hat{H}$ by finding the eigen values of $\hat{a}^\dagger \hat{a}$.
iii) Show that for nth energy eigen state

$$\Delta\hat{X}\Delta\hat{P} = \left(n + \frac{1}{2}\right)\hbar$$

where $\hat{X}$ and $\hat{P}$ are position and momentum operators respectively.

[2 + (2 + 6 + 5)]

5 (a) Let $[\hat{H}, \hat{\pi}] = 0$, where $\hat{H}$ is the Hamiltonian of the quantum system and $\hat{\pi}$ is the Parity operator. Let $|n\rangle$ is the non-degenerate eigenstate of $\hat{H}$ with eigen value E_n.

$$\hat{H}|n\rangle = E_n|n\rangle$$

Then show that $|n\rangle$ is also a Parity eigenstate.

(b) Consider the following two qubits state shared between Alice and Bob

$$|\psi^-\rangle = \frac{1}{\sqrt{2}}\left[|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B\right]$$

where $|0\rangle$ and $|0\rangle$ are eigenstates of σ_z.
i) Show that the state can not be a product state.
ii) If Alice measures spin along some vector $\vec{n}$ and Bob measures spin along some vector $\vec{m}$, what is the probability that their results are correlated?

[3 + (4 + 8)]

6 (a) Write the self adjoint operators corresponding to the angular momentum observable along x, y, z, axes in terms of position and momentum operators. Prove

the following commutation relations
$[\hat{L}_x, \hat{L}_y] = i\hbar\hat{L}_z$ and $[\hat{L}^2, \hat{L}_z] = 0$.

(b) Show that eigenvalues of $\hat{L}^2$ are $l(l+1)$ for some number l with $2l+1$ to be integer and possible eigenvalues of $\hat{L}_z$ are $-l, -l+1, -l+2, ...l-1, l$.

(c) Consider a particle with angular momentum l. Then for any simultaneous eigen state of the operators of $\hat{L}^2$ and $\hat{L}_z$, show that the expectation values satisfy the following relations:-

$$\langle \hat{L}_x \rangle = \langle \hat{L}_y \rangle = 0$$

[4 + 6 + 5]

INDIAN STATISTICAL INSTITUTE
B. Stat. (Hons.) III Year 2016-2017

First Semester Examination

Subject : SQC & OR

Date : **15.11.2016** Full Marks : **100** Duration : **3 hrs.**

This paper carries 120 marks. You may answer as much as you can; but the maximum you can score is 100.

1. Consider the following LP:

Maximize $Z = x_1 + 4x_2 + 3x_3$

subject to

$$2x_1 + 3x_2 - 5x_3 \leq 2$$
$$3x_1 - x_2 + 6x_3 \geq 1$$
$$x_1 + x_2 + x_3 = 4$$
$x_1 \geq 0;\ x_2 \leq 0;\ x_3$ unrestricted in sign.

a) Write down the dual of the above problem.
b) Is the vector $(x_1, x_2, x_3) = (0, 0, 4)$ feasible for the primal?
c) Give a feasible solution for the corresponding dual problem.
d) Use complementary slackness conditions to obtain the primal and dual optimal solutions.

[9+1+4+6=20]

2. Assume that there are only two producers A and B of oil in the world who are competitors of each other. Assume that there are only three possible production levels: 2, 4 and 6 million barrels/day. So depending on the two producers' (players') decision, total world production is 4, 6, 8, 10 and 12 million barrels/day. Accordingly, world prices are 35, 30, 25, 15 or only 10\$/barrel. The production costs of one barrel are considered to be \$2 for producer A and \$4 for producer B.

a) Find the profit of the two producers under the different production scenarios.
b) The intention is to maximize the difference between the profit of a producer and that of its competitor. Consider the case as if it were a zero-sum game with two players. Write down the pay-off matrix.
c) Determine the optimal strategy.

[9+6+5=20]

3. A computer network shares a printer. Jobs arrive at a mean rate of 2 jobs per minute and follow a Poisson process. The printer prints 10 pages per minute and the mean number of pages per job is 4. There is a 3-second idle time between one job and the next. The service time follows an exponential distribution. Calculate the following:

   a) The percentage of time that the printer is available.
   b) The mean queue length.
   c) The service rate required for the mean time in the system to be under 3 minutes.

[5+5+5=15]

4. A process producing coaxial cables is being monitored by a "number of defects per unit" chart. The process average has been calculated as 0.10 defects per unit. Three sigma control limits are employed and samples of size 200 are taken on a daily basis.

   a) Calculate the upper and lower control limits for the chart.
   b) If the process mean were to shift suddenly to 0.15 per unit, what is the probability that this shift would be detected on the 5^{th} subsequent day?
   c) What is the expected number of samples until an out-of-control signal is received when the process is actually in control?

[4+5+6=15]

5. Suppose that a vendor ships components in lots of size 5000. A single sampling plan with $n = 50$ and $c = 2$ is being used for receiving inspection. Rejected lots are screened, and all defective items are reworked and returned to the lot.

   a) Find the probability of acceptance when the vendor's process average is 0.005.
   b) Management has objected to the use of the above sampling procedure and wants to use a plan with an acceptance number $c = 0$, arguing that this is more consistent with their zero-defect program. Design a single-sampling plan with $c = 0$ that will give a 0.90 probability of rejection when the vendor's process average is 0.103.
   c) If you were to draw the OC curve for the given sampling plan $n = 50, c = 2$ and the sampling plan you have designed in part (b) above, how would they compare? (You do **not** need to draw the actual OC curves.)
   d) Suppose that incoming lots are 0.5% nonconforming. What is the probability of rejecting these lots under the sampling plan you have designed in part (b) above? Calculate the ATI at this point for both the plans. Which plan would you choose to adopt?

[5+2+2+(1+4+1)=15]

6. Ten parts are measured three times by the same operator in a gage capability study. The data are given in Table 6.1. The specifications on the part are at 100 ± 15.

a) Analyze the data appropriately and check whether the operator is having any difficulty in using this gage?
b) Estimate total variability and product variability.
c) What percentage of total variability is due to the gage?
d) Find the precision-to-tolerance ratio for this gage. Is it adequate?

Table 6.1: Data for Question No. 6

Part Number	**Measurements**		
	1	**2**	**3**
1	100	101	100
2	95	93	97
3	101	103	100
4	96	95	97
5	98	98	96
6	99	98	98
7	95	97	98
8	100	99	98
9	100	100	97
10	100	98	99

[10+4+1+2=17]

7. Answer the following questions briefly:-

a) Distinguish between Type A and Type B OC curves of acceptance sampling plans.
b) What are the properties of a good measuring instrument?
c) Define ***Quality Engineering***.
d) Define the ***Taguchi Capability Index***.
e) Why should we be interested in obtaining the optimal solution of the primal by solving the dual?
f) Define ***Quadratic Programming***. Give an example.

[3 × 6 = 18]

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Appendix VI Factors for Constructing Variables Control Charts

Observations in Sample, n	Chart for Averages: Factors for Control Limits			Chart for Standard Deviations: Factors for Center Line		Factors for Control Limits				Chart for Ranges: Factors for Center Line		Factors for Control Limits				
	A	A_2	A_3	c_4	$1/c_4$	B_3	B_4	B_5	B_6	d_2	$1/d_2$	d_3	D_1	D_2	D_3	D_4
2	2.121	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606	1.128	0.8865	0.853	0	3.686	0	3.267
3	1.732	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276	1.693	0.5907	0.888	0	4.358	0	2.575
4	1.500	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088	2.059	0.4857	0.880	0	4.698	0	2.282
5	1.342	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964	2.326	0.4299	0.864	0	4.918	0	2.115
6	1.225	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	5.078	0	2.004
7	1.134	0.419	1.182	0.9594	1.0423	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.204	5.204	0.076	1.924
8	1.061	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751	2.847	0.3512	0.820	0.388	5.306	0.136	1.864
9	1.000	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707	2.970	0.3367	0.808	0.547	5.393	0.184	1.816
10	0.949	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.687	5.469	0.223	1.777
11	0.905	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.811	5.535	0.256	1.744
12	0.866	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.922	5.594	0.283	1.717
13	0.832	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	1.025	5.647	0.307	1.693
14	0.802	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	1.118	5.696	0.328	1.672
15	0.775	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	1.203	5.741	0.347	1.653
16	0.750	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	1.282	5.782	0.363	1.637
17	0.728	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	1.356	5.820	0.378	1.622
18	0.707	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496	3.640	0.2747	0.739	1.424	5.856	0.391	1.608
19	0.688	0.187	0.698	0.9862	1.0140	0.497	1.503	0.490	1.483	3.689	0.2711	0.734	1.487	5.891	0.403	1.597
20	0.671	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470	3.735	0.2677	0.729	1.549	5.921	0.415	1.585
21	0.655	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	1.605	5.951	0.425	1.575
22	0.640	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448	3.819	0.2618	0.720	1.659	5.979	0.434	1.566
23	0.626	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438	3.858	0.2592	0.716	1.710	6.006	0.443	1.557
24	0.612	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	1.759	6.031	0.451	1.548
25	0.600	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	1.806	6.056	0.459	1.541

For $n > 25$

$$A = \frac{3}{\sqrt{n}}, \quad A_3 = \frac{3}{c_4\sqrt{n}}, \quad c_4 \simeq \frac{4(n-1)}{4n-3},$$

$$B_3 = 1 - \frac{3}{c_4\sqrt{2(n-1)}}, \quad B_4 = 1 + \frac{3}{c_4\sqrt{2(n-1)}}.$$

$$B_5 = c_4 - \frac{3}{\sqrt{2(n-1)}}, \quad B_6 = c_4 + \frac{3}{\sqrt{2(n-1)}}.$$

INDIAN STATISTICAL INSTITUTE

First Semester Examination : (2016-2017)

B.Stat. 3rd Year

PARAMETRIC INFERENCE

Date: 18 November, 2016 Maximum Marks: 100 Duration: 3 Hours

Answer all questions.

1. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with mean θ.

(a) Find the UMP level α test for testing $H_0 : \theta \leq \theta_0$ against the alternative $H_1 : \theta > \theta_0$ where $\theta_0 > 0$ is some specified value. Give a direct proof of your result without using the general theorem on UMP test for MLR families of distributions.

(b) Consider now the problem of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

Show that for any $0 < \alpha < 1$, there does not exist a UMP level α test for this problem.

Derive the usual optimum test using the general result (to be stated by you) for exponential family.

[12+(6+8)=26]

2. Let $\boldsymbol{X}$ have distribution P_θ with a density $f(\boldsymbol{x}|\theta)$, $\theta \in \Theta$, an open interval of R, so that the family $\{f(\cdot|\theta), \theta \in \Theta\}$ has monotone likelihood ratio in some statistic $T(\mathbf{x})$. Consider a test of the form

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if } T(\boldsymbol{x}) > c \\ 0, & \text{if } T(\boldsymbol{x}) < c. \end{cases}$$

Show that the power function of this test is nondecreasing in θ. [10]

3. Let $X_1, \ldots, X_n$ be i.i.d. $U(0, \theta)$ variables and $X_{(n)} = \max(X_1, \ldots, X_n)$. Fix $0 < \alpha < 1$.

(a) Show that for testing $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta > \theta_0$, any test ϕ_1 satisfying $\phi_1(\boldsymbol{X}) = 1$, if $X_{(n)} \geq \theta_0$ and $E_{\theta_0}\phi_1(\boldsymbol{X}) = \alpha$, is UMP level α.

P.T.O

(b) Assume that the test ϕ_2, given by

$$\phi_2(\boldsymbol{X}) = \begin{cases} 1, & \text{if } 0 < X_{(n)} < \theta_0\alpha^{1/n} \\ 0, & \text{if } \theta_0\alpha^{1/n} < X_{(n)} < \theta_0 \end{cases}$$

is UMP level α for testing $H_0 : \theta = \theta_0$ against the alternative $H_2 : \theta < \theta_0$.

Use this result and the result stated in part (a) to find a UMP level α test for testing $H_0 : \theta = \theta_0$ against the alternative $H : \theta \neq \theta_0$. [8+8=16]

4. (a) Derive a condition for equality in Cramer-Rao Inequality. Show that a minimum variance bound (MVB) estimator of a parametric function $g(\theta)$ must be a sufficient statistic.

(b) Let $X_1, \ldots, X_n$ be i.i.d. with common p.d.f.

$$f(x|\theta) = \begin{cases} \theta x^{\theta-1}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

where $\theta > 0$. Find the Cramer-Rao lower bound for the variance of an unbiased estimator of $1/\theta$ and find an estimator which attains this lower bound. [6+6=12]

5. Let $(X_1, \ldots, X_n)$ be a random sample form a $U(\theta - 1/2, \theta + 1/2)$ distribution where $\theta \in R$ is unknown. Let $X_{(j)}$ be the j-th order statistic. Show that $(X_{(1)} + X_{(n)})/2$ is a consistent estimator of θ. [8]

6. (a) What is the difference between Bayesian inference and classical inference with respect to evaluation of performance of an estimator?

(b) Describe how the posterior distribution can be used for estimation of a real parameter θ. How do you measure the accuracy of an estimate?

Given a loss function $L(\theta, a)$, how do you find an optimum estimate in the Bayesian paradigm? [4+6=10]

7. (a) Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$ variables. Consider a $N(\mu, \tau^2)$ prior for θ where μ and τ^2 are known. Find the posterior distribution of θ and also a 95% highest posterior density (HPD) credible set for θ.

(b) Let $X_1, \ldots, X_n$ be a random sample from a $U(\theta - 1/2, \theta + 1/2)$ distribution, $-\infty < \theta < \infty$. Consider the noninformative prior $\pi(\theta) = 1$. Find a 95% HPD credible interval for θ. [9+9=18]

INDIAN STATISTICAL INSTITUTE

Supplementary
Back paper/Examination:2016-2017

Subject: Introduction to Sociology

B. Stat. III Year

Date: 21/12/ 2016 Maximum Marks: 50 Duration: Three hours

The figures in the margin indicate full marks

Answer any five of the following questions:

Q 1. What is development? What is the liberal perspective on Development? 10

Q 2. Analyse the nature of class inequality in India? 10

Q 3. Discuss the impact of the 73rd and 74th Constitutional Amendment on grassroots democracy 10

Q 4. Discuss the concept of social structure as a model. 10

Q 5. Outline the steps in testing a hypothesis. 10

Q 6. What is the main source of data for estimation of income poverty in India? How is income poverty measured? 5 + 5 = 10

Q 7. What is the difference between census surveys and sample surveys? If we want to test for differences in farm incomes across farm size, what type of sampling would you suggest? 5 +5 = 10

INDIAN STATISTICAL INSTITUTE

Backpaper Examination: 2016–2017
B.Stat. (Hons.) 3rd Year. 1st Semester
Linear Statistical Models

26.12.2016 | 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.
- You may use scientific calculator for numerical calculation.

---

1. Consider the following balanced one-way ANOVA model:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ j = 1, \ldots, N,\ i = 1, \ldots, 3,$$
$$\epsilon_{ij}\text{'s} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2).$$

We wish to test the hypothesis $\mathrm{H}_0 : \alpha_1 + \alpha_2 - 2\alpha_3 = 0$ against $\mathrm{H}_1 : \alpha_1 + \alpha_2 - 2\alpha_3 \neq 0$. Argue that the level-$\alpha$ ANOVA based test rejects H_0 if

$$\frac{N(\bar{Y}_{1\cdot} + \bar{Y}_{2\cdot} - 2\bar{Y}_{3\cdot})^2}{6\,MSE} > F_{1-\alpha,1,3(N-1)},$$

where F_{γ,n_1,n_2} is the γ-the quantile of F-distribution with n_1, n_2 degrees of freedom. [15]

2. Consider the following balanced two-way ANOVA model without interaction:

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \epsilon_{ijk},\ k = 1, \ldots, N,\ i = 1, \ldots, s,\ j = 1, \ldots, t.$$
$$\epsilon_{ijk}\text{'s} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2).$$

Describe how equality of the η_j's can be tested. [Your expression of the test statistic should be in terms of appropriate sum of squares. You may assume relevant facts about testing in a linear model and relevant linear algebraic facts.] [20]

3. Consider the following balanced one-way ANOVA model with t treatments and one covariate:

$$Y_{ij} = \mu + \alpha_i + \beta x_{ij} + \epsilon_{ij},\ j = 1, \ldots, N,\ i = 1, \ldots, t.$$
$$\epsilon_{ij}\text{'s} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2).$$

Describe how can the hypothesis $H_0 : \beta = 0$ be tested. [You must state all your assumptions clearly.] [20]

4. Consider the following balanced two-way ANOVA model with interaction:

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + \epsilon_{ijk},\ k = 1, \ldots, N,\ i = 1, \ldots, s,\ j = 1, \ldots, t,$$
$$\epsilon_{ijk}\text{'s} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2).$$

Consider a linear parametric function $\psi := a\mu + \sum_{i=1}^{s} b_i\alpha_i + \sum_{j=1}^{t} c_j\eta_j + \sum_{i=1}^{s}\sum_{j=1}^{t} d_{ij}\gamma_{ij}$. Show that ψ is estimable iff $a = \sum_{i=1}^{s}\sum_{j=1}^{t} d_{ij}$, $b_i = \sum_{j=1}^{t} d_{ij}\ \forall i = 1, \ldots, s$, $c_j = \sum_{i=1}^{s} d_{ij}\ \forall j = 1, \ldots, t$. Show, moreover, that BLUE of ψ, when it is estimable, is given by $\sum_{i=1}^{s}\sum_{j=1}^{t} d_{ij}\bar{Y}_{ij\cdot}$. [11+9=20]

5. Suppose at a university a student survey is carried out to ascertain the reaction to instructors' usage of a new computing facility. We suppose that all freshmen have to take English or Geology or Chemistry in their first semester. All three courses in the first semester are large and are divided into sections, each section with a different instructor and not all sections necessarily having the same number of students. In the survey, the response provided by each student is opinion (measured on a scale of 1 through 10) of his instructor's use of the computer. Some of the summary statistics are given in the table below. (Each column represents a section of a course.) Decide if all sections within each of the courses have the same opinions. [$F_{0.95,6,16} = 2.7413$ and $F_{0.99,6,16} = 4.2016$, where F_{γ,n_1,n_2} is the γ-the quantile of F-distribution with n_1, n_2 degrees of freedom.] [12]

Subject →	English			Geology			Chemistry		
Section →	1	2	3	1	2	3	1	2	3
Section size →	3	2	5	2	3	1	4	3	2
Section mean →	3	8	5	4	9	8	3	2	7
Section SS (corrected) →	6	2	16	8	2	0	14	2	2
Class (subject) size →	10			6			9		
Subject mean →	5			7.1667			3.5556		
Subject SS (corrected) →	54			40.8333			50.2222		
Total no. of students →	25								
Overall mean →	5								
Overall SS (corrected) →	192								

6. Three methods of packaging frozen foods are being compared. The response variable is ascorbic acid (mg/100g). Some of the summary statistics are given in the table below. Determine using Scheffe's method which of the three pairs of group means are significant at the 0.05 level. [$F_{0.95,2,18} = 3.5546$ and $F_{0.95,1,18} = 4.4139$] [13]

Ascorbic acid (mg/100g) for three methods of packaging			
Method→	A	B	C
Group size	7	7	7
Group mean	17.1514	19.3114	23.5300
Group SS (corrected)	27.2413	5.0581	22.6676
Overall mean	19.9976		
Overall SS (corrected)	202.3126		

INDIAN STATISTICAL INSTITUTE
B. Stat. (Hons.) III Year 2016-2017

Supplementary Examination
(in lieu of Semestral Examination held on November 15, 2016)

Subject : SQC & OR

Date : 26.12.2016 **Full Marks : 100** **Duration : 3 hrs.**

1. a) Define Quality as per ISO standards.
   b) Describe the operation of a double sampling plan

[5X2=10]

2. A process producing coaxial cables is being monitored by a "number of defects per unit" chart. The process average has been calculated as 0.10 defects per unit. Three sigma control limits are employed and samples of size 200 are taken on a daily basis.

   a) Calculate the upper and lower control limits for the chart.
   b) What is the expected number of samples until an out-of-control signal is received?
   c) If the process mean were to shift suddenly to 0.15 per unit, what is the probability that this shift would be detected at the n^{th} subsequent day?

[4+6+5=15]

3. Solve the following problem:

$$\text{Maximize } x_0 = -5x_1 - 21x_3$$

subject to the following constraints:

$$x_1 - x_2 + 6x_3 \geq 2$$
$$x_1 - x_2 + 2x_3 \geq 1$$
$$x_i \geq 0, \qquad i = 1, 2, 3.$$

[15]

4. Consider a single sampling acceptance rectification plan. Suppose that the consignments come in lots of size 10,000. A random sample of 89 units is inspected; and the consignment is considered to be acceptable if the number of defectives in the sample is at most 2. For such a plan, find the AOQ and the ATI if the vendor's process operates at 1%.

[15]

5. Derive the waiting time distribution for the $(M/M/1)$: $(\infty/\infty/FCFS)$ queue.

[15]

6. A company manufactures three grades of paints: Venus; Diana and Aurora. The plant operates on a three-shift basis and the following data are available from the production records:

Requirement of Resources	Grade			Availability (capacity/month)
	Venus	Diana	Aurora	
Special additive (kg/litre)	0.30	0.15	0.75	600 tonnes
Milling (kl/machine shift)	2.00	3.00	5.00	100 machine shifts
Packing(kl/shift)	12.00	12.00	12.00	80 shifts

There are no limitations on other resources. The particulars of sale forecasts and estimated contribution of overheads and profits are given below:

	Venus	Diana	Aurora
Maximum possible sales per month (kiloliters)	100	400	600
Contribution (Rs./Kiloliter)	4000	3500	2000

Due to commitments already made, a minimum of 200 kiloliters per month of Aurora has to be necessarily supplied the next year.

Just as the company was able to finalize the monthly production program for the next 12 months, an offer was received from a nearby contractor for hiring 40 machine shift per month of milling capacity for grinding Diana paint, that could be spared for at least a year. However, due to additional handling at the contractor's facility, the contribution from Diana will get reduced by Re 1/- per liter.

Formulate this problem as an LP model for determining the monthly production program to maximize contribution.

[15]

7. Solve the game whose pay-off matrix is given below:

		Player B			
		B_1	B_2	B_3	B_4
Player A	A_1	1	2	−2	2
	A_2	3	1	2	3
	A_3	−1	3	2	1
	A_4	−2	2	0	−3

[15]

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
-0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414

Appendix VI Factors for Constructing Variables Control Charts

Observations in Sample, n	Chart for Averages			Chart for Standard Deviations						Chart for Ranges							
	Factors for Control Limits			Factors for Center Line		Factors for Control Limits				Factors for Center Line		Factors for Control Limits					
	A	A_2	A_3	c_4	$1/c_4$	B_3	B_4	B_5	B_6	d_2	$1/d_2$	d_3	D_1	D_2	D_3	D_4	
2	2.121	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606	1.128	0.8865	0.853	0	3.686	0	3.267	
3	1.732	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276	1.693	0.5907	0.888	0	4.358	0	2.575	
4	1.500	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088	2.059	0.4857	0.880	0	4.698	0	2.282	
5	1.342	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964	2.326	0.4299	0.864	0	4.918	0	2.115	
6	1.225	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	5.078	0	2.004	
7	1.134	0.419	1.182	0.9594	1.0423	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.204	5.204	0.076	1.924	
8	1.061	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751	2.847	0.3512	0.820	0.388	5.306	0.136	1.864	
9	1.000	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707	2.970	0.3367	0.808	0.547	5.393	0.184	1.816	
10	0.949	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.687	5.469	0.223	1.777	
11	0.905	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.811	5.535	0.256	1.744	
12	0.866	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.922	5.594	0.283	1.717	
13	0.832	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	1.025	5.647	0.307	1.693	
14	0.802	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	1.118	5.696	0.328	1.672	
15	0.775	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	1.203	5.741	0.347	1.653	
16	0.750	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	1.282	5.782	0.363	1.637	
17	0.728	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	1.356	5.820	0.378	1.622	
18	0.707	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496	3.640	0.2747	0.739	1.424	5.856	0.391	1.608	
19	0.688	0.187	0.698	0.9862	1.0140	0.497	1.503	0.490	1.483	3.689	0.2711	0.734	1.487	5.891	0.403	1.597	
20	0.671	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470	3.735	0.2677	0.729	1.549	5.921	0.415	1.585	
21	0.655	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	1.605	5.951	0.425	1.575	
22	0.640	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448	3.819	0.2618	0.720	1.659	5.979	0.434	1.566	
23	0.626	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438	3.858	0.2592	0.716	1.710	6.006	0.443	1.557	
24	0.612	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	1.759	6.031	0.451	1.548	
25	0.600	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	1.806	6.056	0.459	1.541	

For $n > 25$

$$A = \frac{3}{\sqrt{n}}, \quad A_3 = \frac{3}{c_4\sqrt{n}}, \quad c_4 \cong \frac{4(n-1)}{4n-3},$$

$$B_3 = 1 - \frac{3}{c_4\sqrt{2(n-1)}}, \quad B_4 = 1 + \frac{3}{c_4\sqrt{2(n-1)}},$$

$$B_5 = c_4 - \frac{3}{\sqrt{2(n-1)}}, \quad B_6 = c_4 + \frac{3}{\sqrt{2(n-1)}}.$$

INDIAN STATISTICAL INSTITUTE

Backpaper Examination : (2016-2017)

B.Stat. 3rd Year

PARAMETRIC INFERENCE

Date: 30 December, 2016 Maximum Marks: 100 Duration: 3 Hours

Answer all questions.

1. Describe the notions of sufficiency, minimal sufficiency and completeness of a statistic. Illustrate with examples. [13]

2. Show that if a minimal sufficient statistic exists then a complete sufficient statistic is also minimal sufficient. [7]

3. Let $X_1, \ldots, X_m$ be a random sample from $N(\mu_1, \sigma^2)$ and let $Y_1, \ldots, Y_n$ be a random sample from $N(\mu_2, \sigma^2)$ where $\mu_1 \in R$, $\mu_2 \in R$ and $\sigma^2 > 0$ are all unknown.

(a) Find a complete sufficient statistic for (μ_1, μ_2, σ^2).

(b) Find the UMVUE of $1/\sigma$.

(c) Find the UMVUE of $(\mu_1 - \mu_2)/\sigma$ [8+8+8=24]

4. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$ population where μ is known.

(a) Find a UMP level α test for testing $H_0 : \sigma^2 \leq \sigma_0^2$ against the alternative $H_1 : \sigma^2 > \sigma_0^2$ where $\sigma_0^2 > 0$ is some specified value. Give a direct proof of your result without using the general theorem on UMP test for MLR families of distributions. [12]

(b) Consider the problem of testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$. Show that for any $0 < \alpha < 1$, there does not exist a UMP level α test for this problem.

Derive the usual optimum test using a suitable general result (to be stated by you) for exponential family. [6+11=17]

(c) Show that there exists a minimum variance bound (MVB) estimator of σ^2. Does there exist an MVB estimator of σ? Justify your answer. [9]

5. Let $X_1, \ldots, X_n$ be i.i.d $N(0, \sigma^2)$, $\sigma^2 > 0$. Consider an inverse-gamma (α, β) prior distribution having density

$$g(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)}e^{-\beta/\sigma^2}, \ \sigma^2 > 0$$

where $\alpha > 0, \beta > 0$ are known. Find the Bayes estimate of σ^2 for the loss function $L(\sigma^2, a) = (a - \sigma^2)^2/\sigma^4$. [It is given that the mean of an inverse-gamma (α, β) distribution is $\beta/(\alpha - 1)$.] [10]

6. Let X be a random variable with probability distributions under a simple hypothesis H_0 and a simple alternative H_1 given in the following table.

Values of X	1	2	3	4	5	6
Probability under H_1	0.15	0.15	0.15	0.15	0.15	0.25
Probability under H_0	0.15	0.20	0.15	0.10	0.10	0.30

Suppose that we want to find a most powerful (MP) test of level 0.3. Is the MP test of level 0.3 unique ? Justify your answer. [8]

INDIAN STATISTICAL INSTITUTE

Supplementary Examination : (2016-2017)

B.Stat. 3rd Year

PARAMETRIC INFERENCE

Date: 30 December, 2016 Maximum Marks: 100 Duration: 3 Hours

Answer all questions.

1. Describe the notions of sufficiency, minimal sufficiency and completeness of a statistic. Illustrate with examples. [12]

2. Show that if a minimal sufficient statistic exists then a complete sufficient statistic is also minimal sufficient. [7]

3. Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$ population where μ is known.

(a) Find a UMP level α test for testing $H_0 : \sigma^2 \leq \sigma_0^2$ against the alternative $H_1 : \sigma^2 > \sigma_0^2$ where $\sigma_0^2 > 0$ is some specified value. Give a direct proof of your result without using the general theorem on UMP test for MLR families of distributions. [12]

(b) Consider the problem of testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$. Show that for any $0 < \alpha < 1$, there does not exist a UMP level α test for this problem.

Derive the usual optimum test using a suitable general result (to be stated by you) for exponential family. [6+11=17]

(c) Show that there exists a minimum variance bound (MVB) estimator of σ^2. Does there exist an MVB estimator of σ? Justify your answer. [9]

4. Let $\boldsymbol{X}$ have distribution P_θ with a density $f(\boldsymbol{x}|\theta)$, $\theta \in \Theta$, an open interval of R, so that the family $\{f(\cdot|\theta), \theta \in \Theta\}$ has monotone likelihood ratio in some statistic $T(\mathbf{x})$. Consider a test of the form

$$\phi(\boldsymbol{x}) = \begin{cases} 1, & \text{if } T(\boldsymbol{x}) > c \\ 0, & \text{if } T(\boldsymbol{x}) < c. \end{cases}$$

Show that the power function of this test is nondecreasing in θ. [10]

5. Let $X_1, \ldots, X_n$ be i.i.d $N(0, \sigma^2)$, $\sigma^2 > 0$. Consider an inverse-gamma (α, β) prior distribution having density

$$g(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)}e^{-\beta/\sigma^2}, \ \sigma^2 > 0$$

where $\alpha > 0, \beta > 0$ are known. Find the Bayes estimate of σ^2 for the loss function $L(\sigma^2, a) = (a - \sigma^2)^2/\sigma^4$. [It is given that the mean of an inverse-gamma (α, β) distribution is $\beta/(\alpha - 1)$.] [10]

6. Let X be a random variable with probability distributions under a simple hypothesis H_0 and a simple alternative H_1 given in the following table.

Values of X	1	2	3	4	5	6
Probability under H_1	0.15	0.15	0.15	0.15	0.15	0.25
Probability under H_0	0.15	0.20	0.15	0.10	0.10	0.30

Suppose that we want to find a most powerful (MP) test of level 0.3. Is the MP test of level 0.3 unique ? Justify your answer. [8]

7. Let $X_1, \ldots, X_n$ be a random sample from a distribution with density $e^{-(x-\theta)}I(x > \theta)$, $\theta \in R$.

(a) Find a consistent estimator of θ.

(b) Find an MP test of level α for testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ where $\theta_0 < \theta_1$ are two specified values.

[5+10=15]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination

BStat III Year

2016-17, II Semester

Statistics Comprehensive

This test is open books, open notes. No electronic device for calculation/communication is allowed. Answer any four out of the five questions. All the questions carry equal marks.

Maximum time: 2 hours | 20^{th} February, 2017 | Maximum score: 100

1. With 40 independent observations from a Bernoulli distribution with probability of success p, you have to test:

$$H_0:\ p = 0.5 \qquad \text{vs.} \qquad H_1:\ p = 0.6.$$

   Is it possible to perform this test with level of significance 1% and power at least 95%?

2. Find, if possible, a sequence of 10 binary digits (0 or 1) whose autocorrelation of lag 1 is 0.

3. You have an unbiased coin and you may toss it independently, as many times as you need. A population has 5 units having size measure 5, 3, 1, 1 and 1. You have to draw a sample of size 2 using PPSWOR.

   (a) Describe a procedure to choose a sample by using the coin.

   (b) How many tosses you expect to perform for choosing the sample?

4. The number of times a student falls seek in a year (Y) has the Poisson distribution. The mean of the distribution (λ) depends on the gender (X) of the student ($X = 1$ for male, 0 for female) and attendance percentage (Z) of the student in the previous year. You have to quantify this dependence relation, by using data on the triplet (Y, X, Z) for n students.

   (a) Identify a suitable Generalized Linear Model (GLM) for this dependence relation, by using the canonical link function.

   (b) Write down the likelihood function explicitly in terms of the regression parameters.

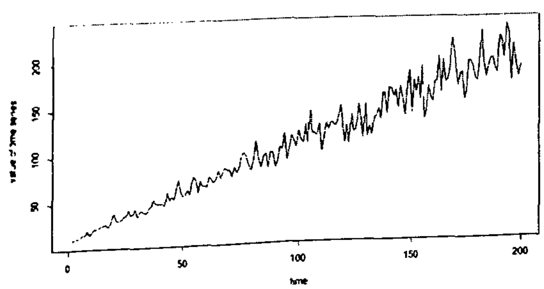
   (c) Describe an iterative algorithm to obtain the MLE.

5. The time plot of 200 consecutive samples of a simulated time series $\{X_t\}$ is shown below. Which of the following difference equations is the most likely model that could have produced it? Explain. Draw free-hand sketches of two time plots that could have come from the other two models.

   A. $X_t = a + bt + Z_t,\ \ Z_t\ iid \sim N(0, \sigma^2)$

   B. $(X_t - \mu) = (X_{t-1} - \mu) + Z_t,\ \ Z_t\ iid \sim N(0, \sigma^2)$

   C. $X_t = (a + bt)e^{Z_t},\ \ Z_t\ iid \sim N(0, \sigma^2)$



INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : (2016-2017)

B.Stat. 3rd Year

NONPARAMETRIC AND SEQUENTIAL METHODS

Date: 21 February, 2017 Max. Marks: 100 Duration: $2\frac{1}{2}$ Hours

1. Let $X_1, \ldots, X_n$ be a random sample from a population with unknown continuous distribution function F.

(a) Describe the Kolmogorov-Smirnov test for the hypothesis $H_0 : F = F_0$ (for various possible alternatives) where F_0 is a specified distribution function.

(b) Show that the Kolmogorov-Smirnov statistic D_n^- is distribution free.

[6+9=15]

2. Let $X_1, \ldots, X_n$ be a random sample from a population with a continuous distribution having unknown (unique) third quartile θ. Consider the problem of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$ where θ_0 is some specified value. Suggest a consistent test for this problem and prove its consistency. [15]

3. Let $X_1, \ldots, X_n$ be a random sample from a population with a continuous distribution which is symmetric about its unknown median θ.

(a) Describe the Wilcoxon signed rank test for testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. Find the variance of the test statistic T_+ under H_0.

(b) Express the statistic T_+ as a weighted sum of two U-statistics. Also find a confidence interval for θ with a given confidence coefficient. Assume that the distribution of T_+ is Symmetric under $\theta = 0$.

[(5+15)+(10+8)=38]

4. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two random samples drawn independently from two populations with continuous distribution functions F and G respectively. Assume that F and G are strictly increasing.

(a) Consider the Mann-Whitney U test of level α for testing $H_0 : F = G$ against the alternative $H_1 : F > G$. Show that this test is unbiased. Also show that this test is of level α for testing the null hypothesis $H : F \leq G$ against $H_1 : F > G$.

(b) Show that the Wilcoxon rank-sum statistic is symmetrically distributed under $H_0 : F = G$. Show this without using the symmetry of the distribution of the Mann-Whitney U-statistic. [12+8=20]

5. Consider the two-sample scale problem. What is the Freund-Ansari-Bradley test for this problem? Show that this test is equivalent to the test based on a test statistic which may be thought of as sum of "ranks" of the first sample observations for some appropriate assignment of ranks to the observations in the combined sample. [12]

INDIAN STATISTICAL INSTITUTE

Periodical Examination

B. STAT III YEAR

DESIGN AND ANALYSIS OF ALGORITHMS

Date : 22.02.2017 Maximum Marks : 40

Duration : 3 Hours

Question 1: Solve the following Recurrence Relations: (5×1)

(a) $T(n) = T(n/2) + O(\log n)$

(b) $T(n) = 2T(n/2) + O(\log n)$

(c) $T(n) = 3T(n/2) + O(n)$

(d) $T(n) = 4T(n/2) + O(n)$

(e) $T(n) = 8T(n/2) + O(n^2)$

Question 2: Are the following statements True or False? Justify your answers. (5×1)

(a) $n \log n = O(n \log^2 n)$.

(b) $n\sqrt{(n)} = \Omega(n \log^2 n)$

(c) $n\sqrt{(n)} = \Theta(n \log^2 n)$

(d) $O(n^{2n}) = O(2^n)$

(e) $O(n^{2n}) = \Omega(2^n)$

(Answer either Question 2 or Question 3)

Question 2: Multiply $x = 11001010$ and $y = 10000111$ using divide and conquer method. Write down the recurrence relation for your algorithm for any n. Solve your recurrence relation. (5+2+3)

Question 3: Write a linear time algorithm to find the convex hull of two disjoint convex polygons. Use this algorithm to find convex hull of a set of n points in R^2. Analyze your algorithm. (5+2+3)

Question 4: Median of n numbers can be found in $O(n)$ time. Devise a divide and conquer based sorting algorithm that uses median find algorithm. Write down the recurrence relation of your algorithm and analyze the running time. (5+2+3)

Question 5: Clearly show the steps of dijkstra's algorithm in the graph shown in Figure 1. The steps should explain how to augment the heap data structure to perform ExtractMin() and DecreaseKey() in $O(\log |V|)$ time, where $|V|$ denotes the number of vertices.

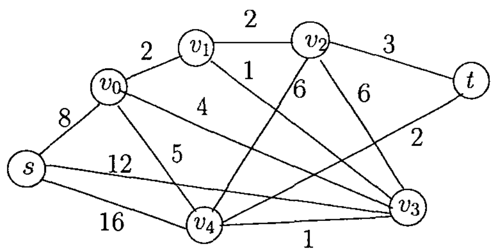


Figure 1: Shortest Path

(b) Prove that, in Dijkstra's algorithm if the last vertex popped from the heap is v whose distance from s is $\delta(s, v)$ then there exists no path from s to v with distance lesser than $\delta(s, v)$. (5+5)

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2016 - 17

Course Name : B. STAT. III Year
Subject : Differential Equations

Date: 23. 02. 2017 Full Marks: 40 Duration: $2\frac{1}{2}$ hrs.

Any result that you use should be stated clearly.

1. Find all solutions of $x(1+x)\frac{dy}{dx} + y = x(1+x)^2 e^{-x^2}$ on the interval $(-1, 0)$. Prove that all solution approach 0 as $x \to -1$ but that only one of them has a finite limit as $x \to 0$. (6)

2. Show that the Euler's equation is a second order linear differential equation for an extremal of
$$I = \int_{x_1}^{x_2} \{a(x)(y')^2 + 2b(x)yy' + c(x)y^2\}dx$$
where $y' = \frac{dy}{dx}$. (5)

3. Find the general solution of the following equations,
(a) $\frac{dx}{dt} = x + 2y + t - 1, \qquad \frac{dy}{dt} = 3x + 2y - 5t - 2.$ (6)

(b) $4x^2\frac{d^2y}{dx^2} - 3y = 0.$ (5)

(c) $\frac{d^2x}{dt^2} + 16x = 8\sin 4t$ with initial conditions $x = \frac{1}{2}$ and $\frac{dx}{dt} = 0$ when $t = 0$. (5)

4. (a) Let f be continuous and satisfy a Lipschitz condition in R. If a solution of the initial-value problem
$$\frac{dy}{dx} = f(x, y), \qquad y(x_0) = y_0$$
exists, then it is unique.
(b) In the above initial value problem if $f(x, y) = \frac{|y|}{x}, x_0 = 1, y_0 = 0$, then examine the uniqueness of the solution. (5+3)

5. A spherical raindrop, starting from rest, falls under the influence of gravity. It gathers in water vapor (assumed at rest) at a rate proportional to its surface. If the initial radius of the raindrop is r_0 and r is its radius at time t, show that its acceleration at time t is $\frac{g}{4}(1 + \frac{3r_0^4}{r^4})$ where g is the acceleration due to gravity. (5)

INDIAN STATISTICAL INSTITUTE

B. Stat. – III
Mid-Semestral Examination
STATISTICAL METHODS IN GENETICS

Time: 2 hours Total marks: 50

February 23, 2017

1. Consider an autosomal codominant locus with alleles, A and a. Suppose the proportions of the allele A in the two populations, both in Hardy-Weinberg equilibrium, are p and P, respectively.

(a) If the two populations are mixed in proportions m and (1-m), what would be the proportions of the various genotypes in the mixed population?

(b) If the individuals in the mixed population practice random mating, what would be the equilibrium genotype proportions? [6+12=18]

2. Many plants, e.g., cotton, practice both random-mating and selfing. Suppose (D,H,R) denote the frequencies of genotypes (AA, Aa, aa) at an autosomal codominant locus in such a population. Let w denote the fraction of the population practicing selfing in each generation and (1-w) denote the fraction practicing random-mating. Would this population ever reach an equilibrium in respect of genotype proportions? [State your assumptions clearly and give reasons for your answer.] [20]

3. Suppose in a random sample of n individuals from a population of size N, the observed frequencies of genotypes MM, MN and NN at an autosomal biallelic locus are, respectively, n_1. n_2 and n_3). If F denotes the inbreeding coefficient of this population, obtain the maximum likelihood estimator of F, and its variance. [9+3=12]

INDIAN STATISTICAL INSTITUTE
First Mid-Semestral Examination: 2016-17

Subject Name : **Number Theory** Date: 23/02/17
Course Name : B.Stat. III yr. Max. Score: 40 Duration: 2 Hours 30 min

Note: Attempt all questions. Marks are given in brackets. Total score is 50. **State results** clearly which you want to use. Use **separate page** for each question.

Problem 1. For all primes p of the form $4k+1$ there exists an integer c such that $c^2 = -1 \mod p$. [6]

Problem 2. Define Legendre symbol $(\frac{n}{p})$ where p is a prime. Show that

$$\sum_{n=1}^{p-1} \left(\frac{n}{p}\right) = 0$$

where p is an odd prime. Hence or otherwise show that if both p and $p+2$ are primes then $\sum_{n=1}^{p-1} n\left(\frac{n}{p}\right) = 0$. [2+ 6+6 = 14]

Problem 3. State the Extended Euclidean Algorithm. Show that the number of division is $O(\log_2 n)$ while we run EEA for two positive integers $a, b \leq n$. [2+8 =10]

Problem 4. Let $n > 9$ be an odd positive composite integer. We write $n-1 = 2^k m$ for some exponent $k \geq 1$ and some odd integer m. Let

$$B = \{x \in \mathbb{Z}_n^* : x^m = 1 \text{ or } x^{m2^i} = -1 \text{ for some } 0 \leq i < k\}.$$

Show that $|B| \leq \phi(n)/2$. [10]

Problem 5. Prove that Z_n^* is a cyclic group where $n = p^2$ for some prime p. [10]

BIII-2016-17

Midterm Examination

Design of Experiments

Full Marks 35 Date: 24th February, 2017 Time: 2.30p.m- 4 p.m

1. For a general block design, derive the variance –covariance matrix of the vector of adjusted treatment totals. Hence obtain the variance of the BLUE of an estimable treatment function.

(4+3 =7)

2. Consider a Randomised Block design with v treatments denoted by 1,2, …,v , in b blocks. Obtain a new design D by deleting the ith treatment form the ith block of the RBD, i=1,2……,v. Answer the following questions considering the new design D.

a) Is the design connected ? Justify.
b) Is the design orthogonal? Justify.
c) List a set of maximum number of independent treatment contrasts.
d) Obtain the C- matrix, explicit form of the Q- vector and hence the explicit form of the BLUE of $\tau_1 - \tau_2$.

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

3. a) Write down the ANOVA for a Latin Square design. Hence obtain an expression to measure the efficiency of a Latin Square design with respect to a Randomised block design.

b) If the two treatments A and C got interchanged in the 2nd row of a standard Latin Square of order 3, can you estimate $\tau_A - \tau_C$ from this new row-column design, under the usual fixed effects additive model considering general mean, row effect, column effect and treatment effects? If so, obtain an unbiased estimator, if not, justify your claim.

((3+3)+6=12)

INDIAN STATISTICAL INSTITUTE

Semestral Examination, 2nd Semester, 2016-2017

B.Stat. 3rd Year

NONPARAMETRIC AND SEQUENTIAL METHODS

Date: 24 April, 2017 Maximum Marks: 100 Duration: 3 Hours

Answer all questions.

1. Consider Wald's sequential probability ratio test (SPRT) for a simple hypotheses H_0 and a simple alternative H_1 in the case of i.i.d. observations $X_1, X_2, \ldots$ with target strength (α, β) and boundaries A and B satisfying $0 < B < 1 < A < \infty$. Let n be the stopping time of the SPRT and $Z_i = \log(f_1(X_i)/f_0(X_i))$, $i \geq 1$, $f_j(\cdot)$ denoting the density of a single observation under H_j, $j = 0, 1$.

State the *Fundamental Identity of Sequential Analysis* and describe how it can be used to find approximate expressions for the OC and ASN functions of an SPRT. Prove your results. [2+20=22]

2. Consider the setup of Question 1. Let H be a hypothesis under which $X_1, X_2, \ldots$ are i.i.d. and $P_H(Z_1 = 0) < 1$.

(a) If $E_H(Z_1) = 0$, show that the OC under H is approximately equal to $a/(a-b)$ where $a = \log A$ and $b = \log B$.

(b) Suppose that $P_H(Z_1 < c) < 1$ for any real number c. Show that $P_H(n < \infty) = 1$. [6+6=12]

3. Let $X_1, X_2, \ldots$ be i.i.d $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Describe Stein's two-stage sampling procedure for obtaining a bounded length confidence interval for μ with confidence coefficient $(1-\alpha)$. Prove the results you state.

Let n be the stopping time of Stein's procedure. Prove that n is finite with probability one.

Show that, given any two positive numbers α and β, however small, one can use Stein's two-stage procedure to derive a test for $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$ for which probabilities of Type I error and Type II error are less

P.T.O

than or equal to α and β respectively. Here $\mu_0 < \mu_1$ are two specified values of μ. [13+4+7=24]

4. Describe the concept of Pitman's asymptotic relative efficiency of tests. Illustrate through an example involving a nonparametric test. [6+10=16]

5. Consider a U-statistic U_n for unbiased estimation of $\theta = \theta(F)$ based on a kernel $h(x_1, \ldots, x_m)$ and $n(\geq m)$ i.i.d. observations from a distribution F.

Define projection $\hat{U}_n$ of the U-statistic U_n and find its expression. Show that under suitable conditions $\sqrt{n}(U_n - \hat{U}_n) \xrightarrow{P} 0$ and hence find the asymptotic distribution of $\sqrt{n}(U_n - \theta)$. (Assume that $n\text{Var}(U_n) \to m^2\sigma_1^2$ where σ_1^2 has its usual meaning.) [8+10=18]

6. Let $X_1, \ldots, X_n$ be a random sample from a population with a continuous distribution F which is symmetric about its unknown median θ. Consider the use of Wilcoxon signed rank statistic T_+ for testing $H_0 : \theta = 0$.

Assume that T_+ can be expressed as $T_+ = \sum\sum_{1 \leq i \leq j \leq n} I(X_i + X_j > 0)$ and use this to find the asymptotic distribution of T_+ under H_0. [8]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

B. STAT III YEAR

DESIGN AND ANALYSIS OF ALGORITHMS

Date : 27.04.2017 Maximum Marks : 50

Duration : 2 hours 30 minutes.

(Although the paper carries a total marks of 55, the maximum marks you can score is 50)

Question 1: Show step by step construction of MST using Union find operation in the graph of 8 vertices with following cost function on the edges. (Your union find should clearly show and use the depth of the parent for merging two trees). 10

$\{\{v_1, v_2\} = 1; \{v_1, v_3\} = 15; \{v_1, v_4\} = 20; \{v_1, v_5\} = 30; \{v_1, v_6\} = 12;$
$\{v_2, v_3\} = 5; \{v_2, v_4\} = 7; \{v_2, v_5\} = 9; \{v_2, v_8\} = 3; \{v_3, v_4\} = 4;$
$\{v_4, v_5\} = 5; \{v_5, v_6\} = 3; \{v_6, v_7\} = 2; \{v_7, v_8\} = 4\}; \{v_1, v_8\} = 2\}$

Question 2: Run edit distance algorithm using memoization table to find edit distance between "person" and "lesson". 10

Question 3: State and prove lower bound for Convex hull of a set of n points in plane. (You can assume that the lower bound of sorting is already given). (2+10)

Question 4: N is a set of n integers, among which only k integers are distinct. Write an sorting algorithm that sorts N in $O(n \log k)$ time. Analyze the running time of your algorithm. (8+5)

Question 5: A and B are two sets, each having exactly n integers. For $a \in A$, let $Pred(a)$ = largest integer $b \in B$ such that $b < a$. Write an $O(n \log n)$ time algorithm for finding $Pred(a)$ for all $a \in A$. Analyze the running time of your algorithm. (5+5)

INDIAN STATISTICAL INSTITUTE
Second Semester Examination 2016-17
B.Stat. III yr.
Design of Experiments

Date: 02.05.2017 Maximum Marks 100 Duration : 3 hours

Answer **all** questions given below. Keep your answers brief and to the point.

1. For the two scenarios described below (i) identify the sources of systematic variation, (ii) suggest a good design indicating the layout of treatments to the experimental units to fulfill the objective of the experimenter and (iii) write the appropriate model for the observations collected.

   (a) A scientist is interested to study the effectiveness of three chemicals and water in extracting sulpher from soils in a specific region. Six different types of soils have been collected from the region for this experiment. Each soil specimen is large enough to be used in six equal parts.

   (b) An educationist wants to compare the effect of three different methods of teaching on the performance of students. For this purpose, four different regions have been chosen. In each region three different schools have been selected and three different teachers of each school try out these methods and give their feed back. [(3+7+4) +(3+7+4)=28]

2. Prove or disprove the following statements.

   (a) An orthogonal block design is necessarily equireplicate.

   (b) For a connected block design $\left(C + a\frac{r\,r'}{n}\right)$; $a > 0$, is always nonsingular and its inverse is a G -inverse of the C matrix. (The notations used here carry the usual meaning.)
   [10+10=20]

3. (a) Consider a factorial experiment with 5 factors A,B,C,D and E , each at 2 levels. Suppose on each day only 8 treatment combinations can be tested. It is desired to have some information on all the factorial effects and as much information as possible for all the main effects, two factor effects, and at least for two three factor effects namely, ACD and BDE. Suggest a suitable confounding scheme, indicating the complete set of confounded effects, if the experiment can be run for eight days. Construct the Key block of any one replication. [8+8=16]

   OR

   (b) Suggest a balanced confounding scheme with minimum number of replications for a factorial experiment involving four factors, each at three levels, to be split in nine blocks, each of size nine, so that none of the main effects and two factor intercation components are confounded. Construct the Key block of any one replication suggested in your scheme. [8+8=16]

4. Suppose that in a Randomised Block Design with 6 treatments and 4 blocks, the observation corresponding to treatment 2 in block 4 is missing. How will you estimate this missing value to obtain the correct error sums of squares? How will you modify this estimate to obtain the correct expression for treatment sums of squares? [6+6=12]

5. Construct a complete set of mutually orthogonal latin squares of order 7. (The method of construction is not needed) [12]

6. Suppose an experiment is to be performed to determine the effect of temperature (Factor A) and the heat treatment time (Factor B) on the strength of normalized steel. Three temperatures and four times are selected. It is difficult to change the temperature setting frequently. Suppose the experiment is to be run for three days to collect the data.

   (a) Suggest a suitable design for this experiment.

   (b) Main effects of which factor can be estimated with more precision in your suggested design? Write down the test statistic for testing equality of the effects of the levels of this factor.

[4+(2+6)=12]

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2016 - 17

Course: B. STAT. III Year Subject: Differential Equation
Date: 05.05.2017 Maximum Marks: 70 Duration: 3hrs.

Any result that you use should be stated clearly

1. Find the general solution of
   $2(z + px + qy) = yp^2$, where $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$. 8

2. (a) Find the limit cycle of the system

$$\frac{dx}{dt} = y + x(1 - x^2 - y^2), \quad \frac{dy}{dt} = -x + y(1 - x^2 - y^2).$$

   (b) By constructing a Lyapunov function, show that the system

$$\frac{dx}{dt} = -x + 4y, \quad \frac{dy}{dt} = -x - y^3$$

   has no closed orbits. (7+7)

3. Find the values of r at which bifurcation occur and classify them as saddle-node, transcritical, pitchfork bifurcation of $\frac{dx}{dt} = rx - \frac{x}{1+x^2}$. Sketch the bifurcation diagram of equilibrium points x^* vs. r. (10)

4. (a) Find the general solution of the equation

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

   (b) Find the Frobenius series solution(s) of the equation

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - \frac{1}{4})y = 0,$$

   about the regular singular point $x = 0$. 7+7

5. (a) Use the method of Laplace transform to solve the equation

$$y'' + 2y' + y = 3te^{-t},$$

where ($'$) denoting differentiation with respect to t (Given that $y = 4$ and $y' = 2$ when $t = 0$).
(b) If $L[f(x)] = F(p) = \ln(1 - \frac{1}{p^2})$ then find $f(x)$, where p is the parameter of Laplace transform. 6+5

6. (a) Prove that

$$e^{\frac{x}{2}}(t - t^{-1}) = J_0(x) + \sum_{n=1}^{\infty} J_n(x)[t^n + (-1)^n t^{-n}],$$

where $J_n(x)$ is the Bessel function of order n and n is an integer.
Hence show that $J_n(x+y) = \sum_{k=-\infty}^{\infty} J_{n-k}(x)J_k(y)$.
(b) Prove that

$$\int_0^1 xJ_p(\lambda_m x)J_p(\lambda_n x)dx = \begin{cases} 0 & if\ m \neq n, \\ \frac{1}{2}J_{p+1}(\lambda_n)^2 & if\ m = n, \end{cases}$$

where λ_m, λ_n are the positive zeros of the Bessel function $J_p(x)$ and p is an integer.

(c) Find the first three terms of the Legendre series of the function $f(x) = e^x$. (3+2)+5+3

INDIAN STATISTICAL INSTITUTE

End-Semester Examination: 2016-17

Subject Name : **Number Theory** Date: 05/05/17

Course Name : B.Stat. III yr. Max. Score: 50 Duration: 3 Hours

Note: Attempt all questions. Marks are given in brackets. Total score is 55. **State results** clearly which you want to use. Use **separate page** for each question. p will always mean odd prime.

Problem 1. Prove that for all integers $k > 1$, the sum $\sum_{n=1}^{k} \frac{1}{n}$ is not an integer. [5]

Problem 2. Let $*$ denote the convolution operation on arithmetic functions. Let $\sigma_r(n) = \sum_{d|n} d^r$ and $N(n) = n$. For an arithmetic function f, let $f^{(s)} = f * \cdots * f$ ($s-1$ times convoluted). For example, $f^{(3)} = f * f * f$. Show that (i) $N^{(4)} = N.\sigma_0^{(2)}$ and (ii) $\phi * \sigma_0 = \sigma_1$. [5+5= 10]

Problem 3. Show that

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1).$$

[10]

Problem 4. Let m be a positive integer, not of the form 1, 2, 4, p^r and $2p^r$. Show that the group $\mathbb{Z}_m^*$ is not cyclic. [10]

Problem 5. Let (e, N) denote the pubic key of RSA and (d, p, q) is its corresponding secret key. Suppose $e < \phi(N)$, $d < \frac{1}{3}N^{1/4}$ and $q < p < 2p$. Describe a strategy to obtain d from the public key and justify. Also describe a possible variant of RSA (keeping d small as above) for which your strategy would not work. [7+3 = 10]

Problem 6. Show that every infinite continued fraction (over positive integers) converges and converges to an irrational number. [8+2 = 10]

B. STAT. (HONS.) – III (2016-17)
STATISTICAL METHODS IN GENETICS
END-SEMESTER EXAMINATION
MAXIMUM MARKS = 100

Date: 05/05/2017

(1) In respect of a rare human disorder that is known to be determined by an autosomal biallelic locus, an investigator collected data from a large number of nuclear families (parents and children), each family being ascertained through an affected child. (The affected child through whom the family is ascertained is called the proband.) The investigator noted that none of the parents in these families was affected. For each family, the total number of children and the number of affected children were noted. To estimate the segregation probability, i.e., Prob{offspring is affected | parental genotypes}, the investigator discarded the data of the probands and obtained an estimate of the segregation probability by simply computing the proportion of affected children in the reduced data set. Do you think that the estimate so obtained is a maximum likelihood estimate? Justify your answer, stating clearly any assumptions that underlie your justification. [30]

(2) Show that the proportion of heterozygotes at an autosomal biallelic locus is halved in each generation of selfing.

[5]

(3) In respect of the ABO blood groups in a population which is in Hardy-Weinberg equilibrium, what is the probability that an individual with AB blood group will have an AB mother? [5]

(4) Under the multistage model of carcinogenesis, show that the hazard rate is a polynomial in c, where c denotes the carcinogen concentration. (Make appropriate assumptions, but state the assumptions clearly.)

[30]

(5) Consider an autosomal biallelic locus. If in a random-mating population, in which the allele frequency of A is p, the relative fitnesses of AA, Aa and aa are 1, 1-h ad 1-s, respectively ($0<h,s<1$), and if the allele A mutates to allele a at the rate m per generation, what will be the equilibrium allele frequency of a, if this allele is very rare in the population?

[20]

(6) For an autosomal biallelic locus, will the Hardy-Weinberg law hold if in the initial generation the relative frequencies of the three genotypes are different for males and females? Give reasons for your answer ad state your assumptions clearly.

[10]

INDIAN STATISTICAL INSTITUTE

Backpaper Examination : (2016-2017)

B.Stat. 3rd Year

NONPARAMETRIC AND SEQUENTIAL METHODS

Date: 10 July, 2017 Max. Marks: 100 Duration: 3 Hours

1. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two random samples drawn independently from two populations with continuous distribution functions F and G respectively. Assume that $G(x) = F(x - \theta)$ for all x and some θ.

(a) Consider the problem of testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. Describe the Mann-Whitney U test and the Wilcoxon rank sum test for this problem and show that these two tests are equivalent.

(b) Consider the Mann-Whitney U test of level α for testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. Show that this test is unbiased. Also show that this test is of level α for testing the null hypothesis $H : \theta \leq 0$ against $H_1 : \theta > 0$.

(c) Find the mean and variance of the Mann-Whitney U statistic under $H_0 : \theta = 0$.

[13+10+10=33]

2. Describe the concept of Pitman's asymptotic relative efficiency of tests.

[7]

3. Let $X_1, X_2, \ldots$ be i.i.d $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Describe Stein's two-stage sampling procedure for obtaining a bounded length confidence interval for μ with confidence coefficient $(1 - \alpha)$. Prove the results you state.

[14]

4. (a) State and prove Stein's result on the termination property of an SPRT.

(b) Let $X_i, i = 1, 2, \ldots$ be i.i.d. Bernoulli (θ) where $0 < \theta < 1$. Consider the SPRT for testing $H_0 : \theta = 1/3$ against $H_1 : \theta = 1/2$ where the boundaries satisfy $0 < B < 1 < A < \infty$. Show, without using Stein's lemma, that the SPRT terminates with probability one under $\theta = 2/3$

[14+12=26]

P.T.O

5. Consider a U-statistic U_n for unbiased estimation of $\theta = \theta(F)$ based on a kernel $h(x_1, \ldots, x_m)$ and $n(\geq m)$ i.i.d. observations from a distribution F.

Define projection $\hat{U}_n$ of the U-statistic U_n and find its expression. Show that under suitable conditions $\sqrt{n}(U_n - \hat{U}_n) \xrightarrow{P} 0$ and hence find the asymptotic distribution of $\sqrt{n}(U_n - \theta)$. (Assume that $n\text{Var}(U_n) \to m^2\sigma_1^2$ where σ_1^2 has its usual meaning.) [7+13=20]

INDIAN STATISTICAL INSTITUTE
Back Paper Examination 2016-17
B.Stat. III yr.
Design of Experiments

Date: 12.07.17 Maximum Marks 100 Duration : 3 hours

Answer **all** questions given below. Keep your answers brief and to the point.

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 $N = \frac{\underline{r}\,\underline{k}'}{n}$ is connected and orthogonal. (The notations have their usual meaning.)

   [6+4+(4+3)= 17]

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

   (b) Construct the complete set of mutually orthogonal latin squares of order 4.

   (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]

P.T.O

4. (a) 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.

   (b) 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. [(4+2+3+3)+(3+4)=19]

5. (a) An experiment is performed to determine the effect of pulp preparation method(A) and cooking temperature(B) on the tensile strength of paper. Three temperatures and four heating times are selected. The pilot plant is capable of making 12 runs per day and the experimenter wishes to continue this experiment for five days.

      i. If the frequent change of the pulp preparation method 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.

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