INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : 2017-2018

B. Stat. (Hons.) III Year

Subject : SQC & OR

Full Marks: 100 **Time: 3 hours** **Date of Examination: 04 Sept 2017**

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

1. A financial organization PQR has a portfolio of investments in shares, bonds and other investment alternatives. Now it has available funds amounting to Rs 2, 00, 000/- which must be considered for new investments. The four investment alternatives that the company is contemplating are given in Table 1 below:

Table 1: Investment data

Financial Details	**A**	**B**	**C**	**D**
Price per share (Rs)	100	50	80	40
Annual rate of return	0.12	0.08	0.06	0.10
Measure of risk per Re invested	0.10	0.07	0.05	0.08

The measure of risk indicates uncertainty associated with the share in terms of its capacity to reach the annual return foreseen; the higher the value, the greater the risk.

PQR has stipulated the following conditions for its investment:

- **Rule 1:** The annual rate of return from this portfolio must be at least 9%.
- **Rule 2:** No value can represent more than 50% of the total investment in Rs.

Use the linear programming model to develop a portfolio of investments which minimizes the risk.

[20]

2. Consider the following LP:

Maximize $x_0 = 4x_1 + x_2 + 3x_3 + 5x_4$

Subject to

$$4x_1 - 6x_2 - 5x_3 + 4x_4 \geq -20$$

$$3x_1 - 2x_2 + 4x_3 + x_4 \leq 11$$

$$8x_1 - 3x_2 + 3x_3 + 2x_4 \leq 23$$

$$x_i \geq 0, \ i = 1, 2, 3, 4$$

One of the simplex iteration tableaus of the above LP is given below:-

	y_0	$-x_1$	$-x_3$	$-x_6$	$-x_7$
x_0	A	33	−38	−17	H
x_5	B	E	11	0	2
x_4	C	F	−6	−3	2
x_2	D	G	−5	−2	I

a) Write down the basis matrix $\mathcal{B}$ corresponding to this tableau.
b) Use the given tableau to find the inverse of the basis matrix $\mathcal{B}$.
c) Without performing the simplex iterations and using the concepts of the simplex algorithm only, find the missing entries $A, \cdots, I$. Show your calculations.

[2+2+16=20]

3. Consider the following problem $\mathcal{P}$:

$$\mathcal{P}\text{: Minimize } z = 5w_1 + 21w_3$$

$$\text{St} \quad \begin{aligned} w_1 - w_2 + 6w_3 &\geq 2 \\ w_1 + w_2 + 2w_3 &\geq 1 \end{aligned}$$

$$w_i \geq 0 \quad i = 1, \cdots, 3$$

Solve $\mathcal{P}$ using the results of Duality Theory. Mention clearly the results that you use.

[15]

4. An electronics company manufactures a special type of cathode ray tubes on a mass production basis. The production rate is 500 tubes per hour. A random sample of size 50 units is taken every hour. A control chart for fraction non-conforming (p) indicates that the current process average is $\bar{p} = 0.20$.
a) Find the 3σ control limits for this control chart.
b) When the process is in-control, how often would false alarms be generated?
c) If the process deteriorates to $\bar{p} = 0.30$, after how many samples will the control chart be able to detect this shift?
d) Suggest three methods of reducing the out-of-control ARL.

[4+5+5+6=20]

5. Answer the following:-
a) Define *Quality Engineering.*
b) What are *Appraisal Costs*?
c) Enumerate the *Seven Tools of Quality Control.*
d) When is the World Quality Day celebrated?

e) Who is considered to be the *Father of SQC?*
f) Who is credited for the development of the Simplex algorithm?
g) Who is reputed to have said: *All models are wrong, but some are useful.*
h) Who is credited for the development of the PDCA wheel?

[3+4+3+1+1+1+1+1=15]

6. 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) 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.

f) 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.

g) 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

h) 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.

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.

j) If ***nothing*** is known concerning the pattern of variation of a set of numbers, we can calculate the standard deviation of this set of numbers and state that the sample mean$\pm$3 the calculated standard deviation will include

i. 89% of all the numbers

ii. 95% of all the numbers.
iii. 99.7% of all the numbers.
iv. None of the above.

k) The j^{th} constraint in the dual of an LPP is satisfied as strict inequality by the optimal solution. The j^{th} variable of the primal will assume a value

i. $\neq 0$.
ii. ≤ 0.
iii. ≥ 0.
iv. $= 0$.

l) If the j^{th} constraint in the primal is an equality, then the corresponding dual variable is

i. $\neq 0$.
ii. ≤ 0.
iii. ≥ 0.
iv. $= 0$.

m) The optimum of an LPP occurs at $X = (1,\ 0,\ 0, 2)$ and $Y = (0,\ \ 1,\ 0,\ 3)$. Then the optimum also occurs at

i. (2, 0, 3, 0).
ii. (1/2 , 1/2, 0, 5/2).
iii. (0, 1, 5, 0).
iv. none of the above

n) Which of the following statements is true with respect to the optimal solution of an LP problem

i. Every LP problem has an optimal solution.
ii. Optimal solution of an LP always occurs at an extreme point.
iii. At optimal solution, all resources are used completely.
iv. If an optimal solution exists, there will always be at least one at a corner point.

o) When measurements show a lack of statistical control, the standard error of the average:

i. Is related to the confidence linits.
ii. Is a measure of process variability.
iii. Is simple to compute.
iv. Has no meaning.

[1 × 15 = 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
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

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2017-18

Course Name : B.Stat. 3rd Year
Subject Name : Sample Surveys
Date : 05/09/, 2017 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. (a) In SRSWR of n draws made from N population units, for two variables of interest y and x, derive $Cov(\bar{y}, \bar{x})$ and an unbiased estimator of it.

   (b) In a village having 800 households, a survey with a SRSWOR of 75 households gave an estimated average monthly expenditure of Rs. 320 on food items with a standard error of Rs. 35. Using this information, for a future survey determine the sample size required to estimate the same characteristic in a neighbouring village such that the permissible margin of error at 95% probability level is 10% of the true value (assume that the coefficient of variation per unit is the same for both the villages).

2. (a) 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.

   (b) 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. (a) In estimating the population ratio $R = \frac{Y}{X}$ through SRSWOR, derive an approximate expression for mean square error of $\hat{R} = \frac{\bar{y}}{\bar{x}}$.

   (b) Show that in SRSWOR, for estimating the population total Y, the usual ratio estimation method will be better than the usual simple average method, if $\rho > \frac{CV(x)}{2CV(y)}$, on assuming that R is positive.

4. The following table gives the data on municipal taxation (in millions of kroner) for the year 1985 for 40 municipalities in Sweden. Draw a 30% sample of municipalities by systematic sampling. Based on your sample give an estimate of the mean taxation along with the standard error and the 95% confidence interval.

Municipality	Revenue	Municipality	Revenue	Municipality	Revenue
1	288	15	422	29	1277
2	199	16	626	30	240
3	196	17	612	31	163
4	159	18	532	32	63
5	536	19	250	33	488
6	134	20	412	34	111
7	623	21	55	35	128
8	517	22	290	36	230
9	96	23	249	37	720
10	467	24	249	38	179
11	277	25	101	39	37
12	155	26	97	40	24
13	386	27	74		
14	241	28	144		

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
Midterm Examination
First Semester, 2017-2018 Academic Year
B.Stat. Third Year
Parametric Inference

Date: 6 September, 2017 Total Marks : 40 Duration: 2 hours

Answer all questions

1. Let $\theta = (\theta_1, \theta_{11}, \theta_{10})$. Let $\theta \in \Theta = (0,1) \times (0,1) \times (0,1)$. We have $n \geq 4$ random variables $X_1, \ldots, X_n$ each of which can take values 0 and 1 only and they have the following joint distribution :

$$\begin{aligned} P_\theta(X_1 = 1) &= \theta_1, \text{ and for } i > 1 \\ P_\theta(X_i = 1 | X_1, \ldots, X_{i-1}) &= \theta_{11} \text{ if } X_{i-1} = 1, \\ &= {}_{10} \text{ if } X_{i-1} = 0. \end{aligned}$$

Find a four-dimensional sufficient statistic for θ based on $X_1, \ldots, X_n$. [5]

2. (a) Define minimal sufficiency. Stating appropriate assumptions, prove that the (vector) natural sufficient statistic in a k-parameter canonical exponential family of distributions is also minimal sufficient. Give an example of an exponential family where the (vector) natural sufficient statistics is not minimal sufficient. Prove your assertion. [1+5+3=9]

   (b) For $\theta \in \Theta = \{\theta : -2 < \theta < 2\}$, let

$$\begin{aligned} p(x|\theta) &= (1 + \theta x) \text{ if } -\frac{1}{2} < x < \frac{1}{2}, \\ &= 0 \text{ otherwise.} \end{aligned}$$

   Let $X_1, \ldots, X_n$ be iid observations from this distribution. Find a minimal sufficient statistics for θ based on the sample. Justify your answer. [4]

3. Let $X_1, \ldots, X_n$ be i.i.d. with discrete uniform distribution over the integers $1, 2, \ldots, N$, where N is an unknown positive integer, i.e $N \in \Theta = \{1, 2, \ldots\}$. Prove that $X_{(n)}$ is complete, where $X_{(n)}$ is the largest order statistic. [3]

4. Let $X_1, \ldots, X_n$ be i.i.d $U(0, \theta)$, where $\theta > 0$ is unknown. Let $X_{(1)}$ and $X_{(n)}$ be the smallest and largest order statistics respectively. Find out $E(\frac{X_{(1)}}{X_{(n)}})$. [3]

5. Let $X_1, \ldots, X_n$ be n i.i.d. observations from a $N(\mu, \sigma^2)$ distribution, with both $\mu \in R$ and $\sigma^2 > 0$ are unknown.

   (a) Find the UMVUE of μ^3. Prove your assertion. [6]

   (b) Find the UMVUE of $P(X_1 \leq 1)$. Prove your assertion. [3]

6. Suppose X has a discrete distribution with p.m.f.

$$\begin{aligned} f_\theta(x) &= \theta \text{ if } x = -1, \\ &= (1-\theta)^2\theta^x \text{ if } x = 0, 1, 2, \ldots \end{aligned}$$

   where $0 < \theta < 1$. Characterize the class $g(\theta)$ of parametric functions for which there exist UMVUE based on X in this case. Justify your answer. You may assume (without proof) the condition to be satisfied by any unbised estimator of zero based on X in this example. [4]

7. Give an example where the knowledge of an ancillary statistic improves the quality of inference about certain parameter in an inference problem. Explain your answer clearly. [3]

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2017-2018

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

Linear Statistical Models

Date: September 7, 2017 | Maximum Marks: 50 | 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 four items having weights $\beta_1, \beta_2, \beta_3$ and β_4. Three out of these four items are measured at a time and each combination is weighed twice. The measurements are denoted by $y_1, \ldots, y_8$.

   (a) Write a suitable linear model for the measurements.

   (b) Find the class of estimable linear parametric functions.

   (c) Find, for each estimable linear parametric function, an expression for its BLUE.

   (d) Find an expression for the residual sum of squares. [4+4+6+6=20]

2. Consider a balanced one-way ANOVA model given by

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ j = 1, \ldots, 6,\ i = 1, \ldots, 4,$$

   with usual assumptions on the ϵ_{ij}'s. Find the vector of fitted values subject to $\alpha_1 - 2\alpha_2 + \alpha_3 = 0$ and $\alpha_2 + 2\alpha_3 - 3\alpha_4 = 0$. [10]

3. Consider a balanced two-way ANOVA model with interaction given by

$$Y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + \epsilon_{ijk},\ k = 1, \ldots, N,\ i = 1, 2,\ j = 1, 2,$$

   where ϵ_{ij}'s are i.i.d. are i.i.d. $N(0, \sigma^2)$. Suppose we wish to test the hypothesis $H_0 : \gamma_{ij} = 0 \,\forall\, i, j$ *versus* $H_1 : H_0$ is false.

   (a) Explain how you can formulate this problem as one of testing a linear model against a reduced model.

   (b) Find the residual sum of squares for the full model and its degrees of freedom (df).

   (c) Find the residual sum of squares for the reduced model and its df.

   [3+(4+3)+(7+3)=20]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2017-2018
Course Name: BStat III
Subject Name: Physics III

Date: 8 September 2017 **Maximum Marks: 40** **Duration: 2 hrs**

Note: Answer as many questions as you can. Each question carries 10 marks.

1. (a) Show that an observer moving with relativistic speed will measure a shorter length of a rod as compared to its length in rest frame. For convenience, you can consider motion along x-axis only.

   (b) Using 4-vector notation, prove that the 4-velocity is a Lorentz-invariant quantity. Can you interpret it physically?

   [6+(3+1)=10]

2. Two objects A and B travel with velocities $\frac{4}{5}c$ and $\frac{3}{5}c$ respectively (with respect to a stationary observer sitting on the earth) along the same straight line in the same direction. How fast (with respect to the stationary observer) should another object C travel between them, so that it feels that both A and B are approaching C at the same speed?

   [10]

3. (a) Derive the expression for kinetic energy of a relativistic point particle moving with velocity v and show that it can be expressed as

$$K = m_0c^2 \left[\frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right],$$

   where symbols have their usual meanings.

   (b) Hence prove that it reduces to the usual Newtonian expression $\frac{1}{2}m_0v^2$ for low velocity limit.

   [7+3=10]

4. A particle of mass m oscillates relativistically along x-axis under a force $F = -m\omega^2 x$ with an amplitude of oscillation b. Show that the period of oscillation is given by

$$T = \frac{4}{c} \int_0^b \frac{\gamma}{\sqrt{\gamma^2 - 1}} dx,$$

   where $\gamma = 1 + \omega^2(b^2 - x^2)/2c^2$.

   [10]

P.T.O

5. The electromagnetic field tensor is given by the components of a traceless antisymmetric matrix

$$F^{\mu\nu} \equiv \begin{bmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{bmatrix}$$

Derive the good old expressions for (i) Gauss law and (ii) Ampere's law using the relativistic electrodynamics formula $\frac{\partial F^{\mu\nu}}{\partial x^\nu} = \mu_0 J^\mu$, where J^μ is the 4-conserved current.

[4+6=10]

INDIAN STATISTICAL INSTITUTE
B. Stat. (Hons.) III Year 2017-2018

First Semester Examination

Subject : SQC & OR

Date : 17 11 17 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:

$$\mathcal{P}: \quad \text{Minimize} \quad Z = 2x_1 + x_2 - x_3$$

subject to

$$x_1 + x_2 - x_3 = 1$$
$$x_1 - x_2 + x_3 \geq 2$$
$$x_2 + x_3 \leq 3$$
$$x_1 \geq 0;\ x_2 \leq 0;\ x_3 \text{ unrestricted in sign.}$$

a) Write down the dual $\mathcal{D}$ of the above problem.
b) Find (if it exists) a feasible solution for the primal $\mathcal{P}$.
c) Give a feasible solution (if it exists) for the corresponding dual problem.
d) State with reasons whether the Weak Duality Theorem holds for this pair of primal - dual problems.
e) State with reasons whether this pair of primal - dual problems admits of an optimal solution.

[9+2+2+3+3=20]

2) Customers arrive at a retail outlet at a rate of 30 customers per hour. The total time that customers spend in the store contributes to their dissatisfaction. A wasted customer hour has been estimated to cost Rs 10. Management has now two options: (i) either employ one fully trained fast clerk who is able to serve 50 customers per hour; or (ii) two less trained slower clerks, who can handle 30 customers per hour each. Each of the two clerks would have his own waiting line (as in a supermarket). Each of the slow clerks asks for Rs 6 per hour as wages, while the fast clerk asks for Rs 16 per hour.

a) Should we hire the two slower clerks or the one fast clerk?
b) A new applicant for the job offers his services. The company tried him out and it turned out that he is able to handle 75 customers per hour. Based on the result obtained in (a) above, what is the maximal amount that we can pay him?

[10+5=15]

3) Two breakfast food manufacturers A and B are competing for an increased market share. The payoff matrix, shown in the following table, describes the increase in market share for A and decrease in market share of B.

A	*B*			
	Give Coupons	*Decrease Price*	*Maintain Present Strategy*	*Increase Advertising*
Give Coupons	1	2	−2	2
Decrease Price	3	1	2	3
Maintain Present Strategy	−1	3	2	1
Increase Advertising	−2	2	0	−3

Determine the optimal strategies for both the manufacturers and the value of the game.

[15+5=20]

4) Control charts for $\bar{X}$ and R are in use with the following parameters:

	$\bar{X}$ chart	R chart
UCL	363.0	16.81
Central Line	360.0	8.91
LCL	357.0	1.64

The sample size is $n = 9$. Both charts exhibit control. Assume that the quality characteristic is normally distributed.

a) What is the α – risk associated with the $\bar{X}$ – chart?

b) Specifications on this quality characteristics are 358 ± 6. Find the C_{pk} index for this process.

c) Find the proportion of non-conformance being produced, if any.

d) Suppose the mean shifts to 357. What is the probability that the shift will not be detected on the first sample following the shift?

e) What would be the appropriate control limits for the $\bar{X}$ – chart if the probability of type I error were to be 0.01?

f) Since both charts exhibit control, the supervisor suggested that the sample size be reduced to $n = 5$. What will be the new control limits (and the central lines) for the two charts?

[3+2+3+3+3+6=20]

5) Answer the following questions:-

a) A consumer has received a special consignment of lot size 100. Find the probability of accepting the lot if it contains 2% defective and he decides to use a sampling plan with $n = 10$ and $c = 1$.

b) Consider the following acceptance sampling plans:-

Plan	N	n	c
I	1000	240	2
II	1000	170	1
III	1000	100	0

Suppose that the AQL has been fixed at 2.2%.

i. As a consumer which plan would you prefer to use?
ii. Find the probability of accepting lots of 1% defective for each of these plans.
iii. If you are the producer, which plan would you prefer to use?

[7+(5+6+2)=20]

6) Answer the following questions:-

a) Caroline is a wine buff and bon vivant. Her wine rack holds 240 bottles. She notices that she seldom fills the rack to the top but sometimes after a good party the rack is empty. On average it seems to be about 2/3rds full.
Many wines improve with age. After reading an article about this, Caroline starts to wonder how long, on average, she has been keeping her wines. She went back through a few months of wine invoices and estimates that she has bought, on average, about eight bottles per month. But she certainly does not know when she drank which bottle and so there seems to be no way she can find out, even approximately, the average age

of the bottles she has been drinking. She seeks your assistance in estimating the average amount of time a bottle stays in her cellar. What is your answer?

b) Show how the following problem can be made separable:

$$\text{Maximize } z = x_1x_2 + x_3 + x_1x_3$$

subject to

$$x_1x_2 + x_2 + x_1x_3 \leq 10$$
$$x_i \geq 0, \quad i = 1, 2, 3.$$

[5+10=15]

7) Answer the following questions briefly:-

a) Distinguish between Type A and Type B OC curves of acceptance sampling plans.

b) Why should we be interested in obtaining the optimal solution of the primal by solving the dual?

c) Sis Proseguer is a company which describes itself as ***Leader in Cash Logistics***. It provides comprehensive range of Cash in Transit and ATM Services for banks and commercial establishments. You have been recruited at a hefty salary to help the company improve its services. Mention three OR problems that you can take up to help the company.

[4 + 3 + 3 = 10]

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	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					
Sample, n	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

First Semester Examination : 2017-18

Course Name : B.Stat. 3rd Year
Subject Name : Sample Surveys
Date : 20/11/17, 2017 Total Marks : 60 Duration : 3 hrs.

Answer Q.5 and any 2 from Q.1 to Q.4.
Notations are as usual.

1. (a) Derive Des Raj's ordered unbiased estimator for Y based on a PPSWOR of size n drawn from N units having the size measure variable x.

   (b) Derive the variance of this estimator and obtain an unbiased estimator of the variance.

(20)

2. (a) Deduce the procedure of choosing a sample of units of size n by systematic sampling with probability proportional to size. State how the Horvitz and Thompson's (HT) estimator can be employed to estimate the populational total Y of a variable of interest y based on a sample drawn by this scheme.

   (b) State what is the difficulty of variance estimation in this scheme and show how to overcome it.

(20)

3. (a) Explain how Hansen and Hurwitz's technique of double sampling can be utilized in non-response situation to estimate the population mean $\bar{Y}$ in SRSWOR of size n out of N units.

   (b) Find the variance of this estimator.

(20)

4. (a) What is called a super-population model ? Show with an 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 ?

   (b) Define non-sampling error. Consider an illustrative mathematical model relating to non-sampling error and based on that model, examine if the sample mean of the observed values is an unbiased estimator of the true population mean that is intended to estimate.

(20)

5. For estimating the total yield of paddy (Y) in a district, a **stratified two-stage** sampling scheme was adopted. Within each stratum 4 villages were selected with **PPSWR**, size being measure of the geographical area of that stratum. Within each selected village, 4 plots were selected by **circular systematic** sampling. For every selected plot, the yield of paddy in kg (y) was collected. The following data gives the collected sample data. Based on this sample data, estimate Y unbiasedly and obtain its standard error ($\widehat{SE}$). Also obtain the relative standard error ($RSE = \frac{\widehat{SE}}{|\hat{Y}|}$) as a percentage.

(20)

Yield of paddy for the sample plots

Stratum	Sample village	Inverse of probability ($1/p_i$)	Total no. of plots	Yield of paddy (in kg)			
				Plot 1	Plot 2	Plot 3	Plot 4
1	1	440.21	28	104	182	148	87
	2	660.43	14	108	64	132	156
	3	31.50	240	100	115	50	172
	4	113.38	76	346	350	157	119
2	1	21.00	256	124	111	135	216
	2	16.80	288	123	177	106	138
	3	24.76	222	264	78	144	56
	4	49.99	69	300	114	68	111
3	1	67.68	189	110	281	120	114
	2	339.14	42	80	61	118	124
	3	100.00	134	121	212	174	106
	4	68.07	161	243	116	314	129

Indian Statistical Institute
Semestral Examination
First Semester, 2017-2018 Academic Year
B.Stat. Third Year
Parametric Inference

Date : 24 November, 2017 Full Marks : 60 Duration: $3\frac{1}{2}$ Hours

Answer all questions

1. Suppose $X_1, \ldots, X_n$ are iid with a Bernoulli distribution with unknown parameter θ where $0 < \theta < 1$.
   (a) Find the UMVUE of $\theta(1-\theta)$ based on the observations. [4]
   (b) Does the estimator in part (a) attain the Rao-Cramer lower bound for variance of unbiased estimators of θ ? Justify your answer. [3]
   $\theta(1-\theta)$

2. Suppose $\mathbf{X} = (X_1, \ldots, X_p)'$ has a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matric $\sigma^2 I_{p\times p}$ where $\sigma^2 > 0$ is unknown. Show that $\frac{\mathbf{X}}{||\mathbf{X}||}$ is independent of $||\mathbf{X}||$. [4]

3. Suppose $X_1, \ldots, X_n, \ldots$ are iid random variables with common density $f(x, \theta)$ where $\theta \in R$ is an unknown parameter. Suppose there exists an unbiased estimator T of θ based on sample of size m i.e. $E_\theta T(X_1, \ldots, X_m) = \theta$, $\forall \theta \in R$. Also assume that T has finite variance for each θ. Can you find an estimator of θ based on $X_1, \ldots, X_n$ which is consistent for θ ? Justify your answer. [5]

4. 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 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. [5]

5. Suppose $X_1, \ldots, X_n$ are iid with a uniform distribution on $(0, \theta)$ where $\theta > 0$ is unknown. Based on the sample we want to test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ where $\theta_0 > 0$ is a fixed quantity. Does there exist a

UMP level-α test for this problem for any $0 < \alpha < 1$? Justify your answer. [8]

6. Consider the problem of testing $H_0 : X \sim f_0$ against $H_1 : X \sim f_1$ where X is a random variable and f_0 and f_1 are densities of two distinct probability distributions. Show that for $0 < \alpha < 1$, the power of the most powerful test of level α is strictly bigger than α. [6]

7. (a) Suppose $\mathbf{X}$ has a distribution P_θ with density $f(\mathbf{x}, \theta)$, $\theta \in \Theta$, an open interval on R so that the family $\{f(.,\theta), \theta \in \Theta\}$ has monotone likelihood ratio in some statistic $T(\mathbf{X})$. Suppose that we have a test of the form

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

with $E_{\theta_0}\phi(\mathbf{X}) = \alpha$ where $0 < \alpha < 1$ and $\theta_0 \in \Theta$. Show that $\phi(\mathbf{x})$ is Uniformly Most Powerful test of level α for testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$. [6]

(b) Suppose $X_1, \ldots, X_n$ are iid with an exponential distribution with mean θ. Consider the problem of testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$. Show that for any $0 < \alpha < 1$, there does not exist a UMP level-α test for this problem. [5]

8. Suppose you have $(m + 1)$ functions $f_1, \ldots, f_{m+1}$ such that for each $i = 1, \ldots, m + 1$ f_i is a function from R^n to R where $n \geq 1$. Suppose $c_1, \ldots, c_m$ are given constants. Suppose for some real numbers $k_1, \ldots, k_m$, $\phi^*(x)$ satisfies the following conditions :

$$a) \qquad \phi^*(\mathbf{x}) = \begin{cases} 1, & \text{if } f_{m+1}(\mathbf{x}) > \sum_{i=1}^{m} k_i f_i(\mathbf{x}) \\ 0, & \text{if } f_{m+1}(\mathbf{x}) < \sum_{i=1}^{m} k_i f_i(\mathbf{x}) \end{cases}$$

(b) $\int \phi^*(\mathbf{x}) f_i(\mathbf{x}) d\mathbf{x} = c_i$ for all $i = 1, 2, \ldots, m$ and (c) $0 \leq \phi^*(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in R^n$. Then prove that ϕ^* maximizes $\int \phi(\mathbf{x}) f_{m+1}(\mathbf{x}) d\mathbf{x}$ among all test functions $0 \leq \phi(\mathbf{x}) \leq 1$ subject to $\int \phi(\mathbf{x}) f_i(\mathbf{x}) d\mathbf{x} = c_i$ for $i = 1, \ldots m$. [6]

9. Suppose $X_1, \ldots, X_n$ are iid with a normal distribution with both the mean(μ) and variance(σ^2) parameters unknown. Find the likelihood ratio test of level α for testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ where μ_0 is a fixed constant and $0 < \alpha < 1$. [4]

10. Suppose, given $\theta \in (0,1)$, $X_1, \ldots, X_n$ are iid Bernoulli(θ) random variables for each $n \geq 1$. Assume that θ has a uniform prior distribution on the open interval $(0,1)$. Show that the conditional probability of the event $\{X_{n+1} = 1\}$ given that $X_i = 1$ for all $1 \leq i \leq n$, increases to 1 as $n \to \infty$. [4]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2017–2018
B.Stat. (Hons.) 3rd Year. 1st Semester
Linear Statistical Models

November 28, 2017 Maximum Marks: 50 Duration: 3 hours

- This question paper carries 55 points. Answer as much as you can. However, the maximum you can score is 50.
- You should provide as much details as possible while answering a question. You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.
- You may use scientific calculator for numerical calculation.

---

1. Consider the following balanced two-way ANOVA model without interaction, for a single replicate and one covariate:

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

We wish to test the hypothesis $H_0 : \alpha_1 = \alpha_2 = \alpha_3$. Derive for this hypothesis the ANOVA based test statistic. [Note. You may need appropriate conditions, to be stated by you, on the z_{ij}'s. Also, you may assume without proof, and use, expressions for the fitted values of the Y_{ij}'s, for both balanced two-way ANOVA model without interaction and one-way ANOVA model.] [12]

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

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

Consider the contrast $\psi_1 := 4\alpha_1 - 3\alpha_2 - \alpha_3$.

(a) Obtain a contrast $\psi_2 \equiv \lambda_1\alpha_1 + \lambda_2\alpha_2 + \lambda_3\alpha_3$ which is orthogonal to ψ_1.

(b) Suppose we wish to obtain *simultaneous* confidence intervals for the contrasts ψ_1 and ψ_2 with the goal that the overall confidence coefficient be at least $1 - \alpha$.

[P.T.O.]

(i) Describe how using Scheffe's method of multiple testing you can achieve this.

(ii) Explain why your method achieves the goal of overall confidence coefficient being at least $1 - \alpha$. [3+(6+9)=18]

[You may assume without proof the following: *For any symmetric and idempotent $p \times p$ matrix* $\mathbf{A}$ *and fixed* $x \in \mathbb{R}^p$, $\mathbf{A}x \neq \mathbf{0}$, $\sup \mid u^{\mathrm{T}}x \mid = (x^{\mathrm{T}}\mathbf{A}x)^{1/2}$, *where the supremum is taken over all* $u \in \mathcal{C}(\mathbf{A})$ *satisfying* $\|u\| = 1$. *Moreover, the supremum is attained at* $u_0 := \mathbf{A}x/[\|\mathbf{A}x\|]$.]

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 one of three courses, denoted by $1, 2, 3$, in their first semester. All three courses in the first semester are large and are divided into three sections each. Each section has a different instructor and not all sections necessarily have the same number of students. Assume that the number of students in the j-th section of the i-th course equals N_{ij}, $i = 1, 2, 3; j = 1, 2, 3$. In the survey, the response provided by each student is opinion (measured on a scale of 1 through 10) about his instructor's use of the computer. The data consist of all the responses. Explain how you will test the hypothesis that all sections within the course 1 have the same opinions. [Note. You must state explicitly your assumptions, formulation, and the results you use.] [12]

4. Twenty-four animals were randomly assigned to four different diets and blood samples were taken in a random order. The blood coagulation time was measured. Some of the summary statistics are given in the table below. Determine which of the six pairs of diets lead to different coagulation times ($\alpha = 0.05$). [$F_{0.95,3,20} = 3.0984$, $F_{0.99,3,20} = 4.9382$, $F_{0.95,1,20} = 4.3512$, and $F_{0.99,1,20} = 8.0959$] [13]

Blood coagulation times for 24 animals corresponding to each of four different diets (A–D)				
Diet →	A	B	C	D
Group size	4	6	6	8
Group mean	61	66	68	61
Group SS (corrected)	10	40	14	48
Overall mean	64			
Overall SS (corrected)	340			

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2017-2018
Course Name: BStat III
Subject Name: Physics III

Date: 1 December 2017 **Maximum Marks: 75** **Duration: 3 hrs**

Note: Answer as many questions as you can. Each question carries 15 marks.

1. (a) Using 4-vector notation, prove that the 4-momentum is a Lorentz-invariant quantity. Can you interpret its fourth component and find out the mass-energy relation $E^2 = p^2c^2 + m_0^2c^4$ from there?

   (b) An atomic clock is a classic example of time dilation. Its working principle is based on the fact that an atom makes transition between two of its internal energy states and the frequency change due to this transition can be measured in spectroscopes. The average kinetic energy of thermal motion of the atom is $\frac{3}{2}kT$.
   Given that the Boltzmann constant $k = 1.38 \times 10^{-23} m^2 Kg s^{-2} K^{-1}$ and a typical atom's mass $M \approx 1.67 \times 10^{-27} Kg$, find out the observed change in frequency due to this transition at room temperature.
   Note: You don't need to find out the exact numerical values. Results upto correct orders will do.

   [(4+4)+7=15]

2. (a) Derive the expression for kinetic energy of a relativistic point particle moving with velocity v and show that it can be expressed as

$$K = m_0c^2\left[\frac{1}{\sqrt{1-v^2/c^2}} - 1\right],$$

   where symbols have their usual meanings.

   (b) A and B leave from a common point and travel in opposite directions with a relative speed $4c/5$. When B's clock shows that a time T has elapsed, he sends out a light signal. When A receives the signal, how much time has elapsed in A's clock? Calculate first (i) in A's frame and then (ii) in B's frame, and show that you arrive at same results.

   [7+(4+4)=15]

3. (a) What do you mean by group velocity in Quantum Mechanics? Can you show that in classical limit it results in the usual energy-momentum relation $E = \frac{p^2}{2m}$?

   (b) A one-dimensional free particle wave function is given by

$$\psi(x,t) = (2\pi\hbar)^{-1/2}\int_{-\infty}^{\infty} \exp\left[i\left(p_x x - \frac{p_x^2}{2m}t\right)/\hbar\right]\phi(p_x)dp_x,$$

where the symbols have their usual meanings. Show that the expectation value of the position $\langle x \rangle$ and that of the momentum $\langle p_x \rangle$ can be related by the expression

$$\langle x \rangle = \langle x \rangle_{t=t_0} + \frac{\langle p_x \rangle}{m}(t - t_0).$$

(c) Consider two operators $\hat{A}$ and $\hat{B}$ satisfying the commutation relation $[\hat{A}, \hat{B}] = i\hbar$. Using Cauchy-Schwarz inequality, prove that the uncertainties in simultaneous measurement of $\hat{A}$ and $\hat{B}$ satisfy the inequality

$$\Delta\hat{A}\Delta\hat{B} \geq \frac{\hbar}{2}$$

[3+5+7=15]

4. (a) Prove that $\langle \hat{x}\hat{p}_x \rangle$ and $\langle \hat{p}_x\hat{x} \rangle$ are both non-Hermitian operators. Can you construct a Hermitian operator out of these two?

(b) For a one-dimensional quantum linear harmonic oscillator, the Hamiltonian operator is given by

$$\hat{H} = \frac{\hat{p}_x^2}{2m} + \frac{1}{2}m\omega^2\hat{x}^2,$$

where $\hat{x}$ and $\hat{p}_x$ are two Hermitian operators satisfying the usual commutation algebra. Using raising and lowering operators

$$\hat{a}_\pm = \mp\frac{i}{\sqrt{2}}\left[\frac{\hat{p}_x}{(m\hbar w)^{1/2}} \pm i\left(\frac{mw}{\hbar}\right)^{1/2}\hat{x}\right]$$

show that the energy eigenvalues of the system are discrete and are given by $E_n = (n + 1/2)\,\hbar w$.

[(3+2+2)+8=15]

5. (a) A general angular momentum operator is defined in Cartesian coordinates by $\hat{J} \equiv (\hat{J}_x, \hat{J}_y, \hat{J}_z)$. Prove the following:
(i) $[\hat{J}_x, \hat{J}_y] = i\hbar\hat{J}_z$
(ii) $[\hat{J}_z, \hat{J}^2] = 0$
(iii) $[\hat{J}_z, \hat{J}_\pm] = \pm\hbar\hat{J}_\pm$
where $\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y$ are the ladder operators.

(b) Prove that the two independent states of a spin-$\frac{1}{2}$ particle are normalized and orthogonal to each other. Hence show that the spin components $(\hat{S}_x, \hat{S}_y, \hat{S}_z)$ can uniquely be represented in terms of three (2×2) matrices, called Pauli spin matrices.

[(3 × 3)+(2+4)=15]

6. **Note: Optional question on advanced topics.**

(a) The electromagnetic field tensors are given by the components of two traceless antisymmetric matrices

$$F^{\mu\nu} \equiv \begin{bmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{bmatrix}$$

and $G^{\mu\nu} \equiv [\vec{E}/c \to \vec{B}, \vec{B} \to -\vec{E}/c]$.
Derive the good old expressions for (i) Ampere's law and (ii) Faraday's law using the relativistic electrodynamics formulas $\frac{\partial F^{\mu\nu}}{\partial x^\nu} = \mu_0 J^\mu$, $\frac{\partial G^{\mu\nu}}{\partial x^\nu} = 0$, where J^μ is the 4-conserved current.

(b) In relativistic quantum mechanics, Fermions are described by Dirac equation with the Hamiltonian

$$\hat{H} = -i\hbar c\vec{\alpha}.\vec{\nabla} + \beta m_0 c^2,$$

where $\vec{\alpha} \equiv (\alpha_1, \alpha_2, \alpha_3)$ along three spatial directions; m_0 is the rest mass of the Fermion and c is the speed of light in vacuum.
Prove that all the α_i's and $\hat{\beta}$ are Hermitian, traceless matrices with eigenvalue $= \pm 1$.

[(5+4)+6=15]

INDIAN STATISTICAL INSTITUTE

First Semester Paper Examination:2017-2018

Subject: Introduction to Sociology

B. Stat. III Year

Date: 01.12.17 Maximum Marks: 50 Duration: Three hours

The figures in the margin indicate full marks

Answer any five of the following questions:

Q 1. Elucidate Max Weber's theory of authority. 10

Q 2. What is gender socialization?
Describe the agents of gender socialization. 3+7= 10

Q 3. Explain Karl Marx's theory of historical materialism. 10

Q 4. Discuss the impact of the 73rd and 74th Constitutional Amendment on grassroots democracy. 10

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

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

Q 7. Write **short notes** (any five) on the following: 2 x 5= 10

(a) Social Statistics & Social Dynamics
(b) Marx's idea of class.
(c) Traditional and affectual action.
(d) Primitive communism and scientific communism.
(e) Social movements.
(f) Research Design.
(g) The Marxian theory of alienation.

INDIAN STATISTICAL INSTITUTE

Back Paper Examination : 2017-18

Course Name : **B.Stat. 3rd Year**
Subject Name : **Sample Surveys**
Date : 27/12/2017, **2017** **Total Marks : 100** **Duration :** $3\frac{1}{2}$ **hrs.**

Answer Q.8 and any 4 from Q.1 to Q.7.
Notations are as usual.

1. In SRSWOR of n draws made from N population units, derive $Var(\bar{y})$ and $Cov(\bar{y}, \bar{x})$, and unbiased estimators of those.

(20)

2. Show how the sample size can be determined for a survey to be conducted to estimate the population mean of a continuous variable y by sampling the units through SRSWOR, given that the population size is N and the permissible margin of error at 95% probability level is 10% of the true value.

(20)

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

   State when the stratified random sampling can perform better than usual SRSWOR.

   (b) Obtain the allocation of sample sizes by the proportional allocation and optimum allocation method.

(20)

4. In estimating the population ratio $R = \dfrac{Y}{X}$ through SRSWOR, derive an approximate expression for mean square error of $\hat{R} = \dfrac{\bar{y}}{\bar{x}}$.

(20)

5. Describe D.B.Lahiri's method of drawing a random sample. Prove that this method yields a PPS sample.

(20)

6. (a) Define the regression estimator for estimating the population mean $\bar{Y}$ through SRSWOR. Determine if it is biased or unbiased. Accordingly derive its variance or mean sqaured error.

   (b) Derive the condition under which it performs better in comparison to the usual simple average estimator.

(20)

7. Explain how Politz and Simmon's 'at-home-probabilities' technique can be utilized in the non-response situation to estimate the population mean $\bar{Y}$ by SRSWR scheme. Obtain an estimator for the variance of this estimator.

(20)

8. Data for a small population having 40 units and exhibiting a fairly steady rising trend are given in the following table. Calculate the relative efficiency of linear systematic sampling as compared to SRSWOR in estimating $\bar{Y}$ when the sample size is 4.

(20)

Values of a variable y for a population of 40 units

Unit	y	Unit	y	Unit	y	Unit	y
1	0	11	10	21	22	31	39
2	1	12	11	22	25	32	43
3	2	13	13	23	29	33	46
4	1	14	12	24	30	34	50
5	4	15	12	25	32	35	53
6	5	16	15	26	35	36	52
7	7	17	14	27	33	37	57
8	7	18	17	28	38	38	59
9	9	19	20	29	40	39	63
10	8	20	23	30	41	40	62

Indian Statistical Institute
Supplementary Semestral Examination
First Semester, 2017-2018 Academic Year
B.Stat. Third Year
Parametric Inference

Date : 28.12.2017 | Full Marks : 100 | Duration: 4 Hours

Date : 28/12/2017

Answer all questions

1. State and prove the Neyman-Fisher Factorization Theorem for sampling from a discrete distribution. [15]

2. Stating appropriate conditions, prove the Rao-Cramer Inequality for variance of unbiased estimators. [15]

3. State and prove the Lehmann-Scheffe Theorem in the context of minimum variance unbiased estimation. [10]

4. State and prove the Neyman-Pearson Lemma in the context of testing a simple H_0 against a simple H_1. [25]

5. Derive a general form of the UMPU level-α test for testing a simple null against a two-sided alternative in the context of sampling from a one-parameter exponential family of distributions. Prove your answer assuming the generalized Neyman-Pearson Lemma. [15]

6. Suppose $X_1, \ldots, X_n$ are iid Uniform$(0, \theta)$ with $\theta > 0$. Show that $X_{(n)}$ is a consistent estimator of θ where $X_{(n)}$ is the maximum of the observed sample. [10]

7. Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Find the UMVUE of μ^2. Prove your answer. [10]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination (Supplementary): 2017–2018
B.Stat. (Hons.) 3rd Year. 1st Semester
Linear Statistical Models

December 29, 2017 | Maximum Marks: 50 | Duration: 3 hours

- This question paper carries 55 points. Answer as much as you can. However, the maximum you can score is 50.
- You should provide as much details as possible while answering a question. You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.
- You may use scientific calculator for numerical calculation.

---

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

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

We wish to test the hypothesis $\mathrm{H}_0 : \beta = 0$. Derive for this hypothesis the ANOVA based test statistic. [Note. You may need appropriate conditions, to be stated by you, on the x_{ij}'s. Also, you may assume without proof, and use, expressions for the fitted values of the Y_{ij}'s, for balanced one-way ANOVA model.] [12]

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

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},\ j = 1, \ldots, N,\ i = 1, 2, 3, 4,$$
$$\epsilon_{ij}\text{'s} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2).$$

Consider the contrasts $\psi_1 := \tau_1 + \tau_2 - \tau_3 - \tau_4$ and $\psi_2 := \tau_1 - \tau_2 + \tau_3 - \tau_4$. Suppose we wish to test the hypotheses $\mathrm{H}_{01} : \psi_1 = 0$ and $\mathrm{H}_{02} : \psi_2 = 0$ *simultaneously* so that the probability of rejecting at least one of these hypotheses, when both are true, is at most α.

(i) Describe how using Scheffe's method of multiple testing you can achieve this.

(ii) Explain why your method achieves the goal of bounding the probability of rejecting at least one of H_{01} and H_{02} when both are true, by α. [8+10=18]

[P.T.O.]

[You may assume without proof the following: *For any symmetric and idempotent* $p \times p$ *matrix* $\mathbf{A}$ *and fixed* $\boldsymbol{x} \in \mathbb{R}^p$, $\mathbf{A}\boldsymbol{x} \neq \mathbf{0}$, $\sup \mid \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} \mid = (\boldsymbol{x}^{\mathrm{T}}\mathbf{A}\boldsymbol{x})^{1/2}$, *where the supremum is taken over all* $\boldsymbol{u} \in \mathcal{C}(\mathbf{A})$ *satisfying* $\|\boldsymbol{u}\| = 1$. *Moreover, the supremum is attained at* $\boldsymbol{u}_0 := \mathbf{A}\boldsymbol{x}/[\|\mathbf{A}\boldsymbol{x}\|]$.]

3. Suppose we have collected data on hospital expenses by patients sampled from three districts each of three states of India. We denote the states by $1, 2, 3$. Assume that the number of patients sampled from the j-th district of the i-th state equals N_{ij}, $i = 1, 2, 3; j = 1, 2, 3$. Explain how you will test the hypothesis that hospital expenses in all districts within each state are equal. [Note. You must state explicitly your assumptions, formulation, and the results you use.] [12]

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

Backpaper Examination: 2017–2018

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

Linear Statistical Models

December 29, 2017 | 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:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ j = 1, \ldots, n,\ i = 1, \ldots, 4,$$
$$\epsilon_{ij}\text{'s} \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma^2).$$

We wish to test the hypothesis $\mathrm{H}_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ against $\mathrm{H}_1 : \mathrm{H}_0$ *is false*. Argue that the level-α ANOVA based test rejects H_0 if

$$\frac{N\sum_{i=1}^{3}(\bar{Y}_{i\cdot} - Y)^2}{2\,MSE} > F_{1-\alpha,2,4(N-1)},\ \ \bar{Y}_{i\cdot} := \frac{\sum_{j=1}^{n} Y_{ij}}{n},\ \ Y := \frac{\sum_{i=1}^{3}\sum_{j=1}^{n} Y_{ij}}{3n},$$

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:

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

Suppose $\sum_{i=1}^{s} \lambda_i \alpha_i$ is a contrast. We wish to test the hypothesis $\mathrm{H}_0 : \sum_{i=1}^{s} \lambda_i \alpha_i = 0$ against $\mathrm{H}_1 : \sum_{i=1}^{s} \lambda_i \alpha_i \neq 0$. Find the ANOVA based test statistic for H_0 and its null distribution. [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.] [12+8=20]

[P.T.O.]

3. 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 $\psi := \sum_{i=1}^{s}\sum_{j=1}^{t} d_{ij}\ \gamma_{ij}$. Show that ψ is estimable iff $\sum_{j=1}^{t} d_{ij} = 0\ \forall\, i = 1,\ldots,s$, $\sum_{i=1}^{s} d_{ij} = 0\ \forall\, j = 1,\ldots,t$. Show, moreover, that BLUE of ψ, when it is estimable, is given by $\sum_{i=1}^{s}\sum_{j=1}^{t} d_{ij}\bar{Y}_{ij\cdot}$. [10+10=20]

4. 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 $\mathrm{H}_0 : \beta = 0$ be tested. [You must state all your assumptions clearly.] [20]

5. Suppose we have collected data on hospital expenses by patients sampled from three districts each of three states of India. We denote the states by $1, 2, 3$. Assume that the number of patients sampled from the j-th district of the i-th state equals N_{ij}, $i = 1, 2, 3; j = 1, 2, 3$. Explain how you will test the hypothesis that hospital expenses in all districts within each state are equal. [Note. You must state explicitly your assumptions, formulation, and the results you use.] [12]

6. 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
B. Stat. (Hons.) III Year 2017-2018

Back Paper Examination

Subject : SQC & OR

Date : 01/01/2018 **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/liter)	0.30	0.15	0.75	600 tonnes
Milling (Kiloliter /machine shift)	2.00	3.00	5.00	100 machine shifts
Packing (Kiloliter /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]

socio-economic variables.

International Statistical System: Comparison of major macro variables – National Income/ GDP.

Selected topics from: Purchasing power parity; Indicators relating to Energy, environment, Gender, Industry, National accounts, Social Statistics and Trade. (10 lectures)

Demography:
Sources of demographic data - census, registration of vital events. Rates and ratios. Measures of mortality. Life Table - construction and applications. Stable and stationary population. Measures of fertility and reproduction. Standardization of vital rates. Population growth curves, population estimates and projections. Measures of migration. Use of demographic data for policy formulation. (18 lectures)

Reference Texts for Economic Statistics

1. P.H. Karmel and M. Polasek: *Applied Statistics for Economists.*
2. R.G.D. Allen: *Price Index Numbers.*
3. N. Kakwani: *Income Inequality and Poverty.*
4. L.R. Klein: *An Introduction to Econometrics.*
5. J.S. Cramer: *Empirical Econometrics.*
6. M.D. Intrilligator: *Econometric Models, Techniques and Applications.*

Reference Texts for Official Statistics

1. M.R. Saluja: *Indian Official Statistical Systems.*
2. CSO (MOSPI) Publication: *Statistical System in India.*
3. United Nations publications
4. RBI: *Handbook of Statistics for the Indian Economy* (various years)
5. *Economic Survey,* Govt. of India, Ministry of Finance (various years)

Reference Texts for Demography

1. R. Ramkumar: *Technical Demography.*
2. K. Srinivasan: *Demographic Techniques and Applications.*
3. B.D. Mishra: *An Introduction to the Study of Population.*
4. H.S. Shryock: *The Methods and Materials in Demography.*

Statistical Quality Control and Operations Research

Statistical Quality Control (SQC):
Introduction to quality: Concept of quality and its management – quality planning, quality control and quality improvement; concept of variations and its impact, relevance of exploratory data analysis, run plot, lag plot, frequency distribution and other QC tools. (5 lectures)

Measurement System: Introduction to measurement system; types of measurement; measurement validity; measurement errors and their estimation. (5 lectures)

Use of Control Chart: Introduction to control chart, control chart for variables and attributes - X-MR chart, $\bar{X}$-R chart, $\bar{X}$-s chart, p-chart, np-chart and c-chart; u-chart, CUSUM chart, EWMA chart; process capability analysis. (8 lectures)

Acceptance Sampling: Introduction to acceptance sampling; concept of AQL, LTPD, producer's risk and consumer's risk; single sampling plan and its OC function; acceptance rectification plan - concept of AOQ, AOQL ATI, acceptance sampling tables; concept of double and multiple sampling plan; average sample number. (7 lectures)

Operations Research (OR):
Introduction to Operations Research (3 lectures)

Optimization Theory: Mathematical modeling and concept of optimization problems: linear, nonlinear and integer programming problems; formulation and application of optimization problems; convex analysis in optimization theory; linear programming problem - graphical method to solve linear programming problem, simplex algorithm, sensitivity analysis, solution procedure of two person zero-sum games; optimality conditions and duality theory; nonlinear programming problem and its classification. (19 lectures)

Queuing Theory: Queuing system in practice and importance in Operations Research; pure birth process, birth and death process; introduction to M/M/1 and M/M/C queues; finite queuing system; application of queuing system and limitation. (6 lectures)

Concluding remark: Synthesizing Statistical Quality Control and Operations Research. (1 lecture)

Reference Texts

1. *Statistical Quality Control*- E.L. Grant & R.S. Leavenworth, McGraw-Hill, N.Y.
2. *Quality Control and Industrial Statistics* - A. J. Duncan, Irwin, Homewood, Ill
3. *Introduction to Statistical Quality Control*- D.C. Montgomery, Wiley, N.Y.
4. *Exploratory Data Analysis*- J. W. Tukey, Addison-Wesley
5. *Principles of Quality Control*- Jerry Banks, John Wiley
6. *Defect Prevention* – Victor E Kane, Marcel Dekker, New York
7. *Juran's Quality Control Handbook*-J. M. Juran & F. M. Gryne, McGraw Hill.
8. *Introduction to Linear Optimization*, D. Bertsimas and J. N. Tsitsiklis, Athena, Scientific, Belmont, Massachusetts, 1999.
9. *Linear and Nonlinear Programming*, D. G. Luenberger, Second Edition, Addison-Wesley, Reading, MA, 1984.
10. *Linear Programming* - G. Hadley, Addison Wesley.
11. *Linear Programming* - K. G. Murty, John Wiley
12. *Linear Programming and Network Flows*, M. S. Bazaraa and J. J. Jarvis, John Wiley & Sons, Inc., New York,.
13. *Nonlinear Programming: Theory and Algorithms*, M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, New York, NY: John Wiley & Sons Inc.
14. *Introduction to Operations Research*. Hillier and Lieberman, McGraw-Hill, Boston., MA.
15. *Numerical Optimization with Applications*, S. Chandra, Jayadeva and Aparna Mehra, Narosa Publishing House (2009).

INDIAN STATISTICAL INSTITUTE
First Semester Back Paper Examination: 2017-2018
Course Name: BStat III
Subject Name: Physics III

Date: 02 January 2018 Maximum Marks: 40 Duration: 2 hrs

Note: Answer all questions. Marks will be given for any positive attempt.

1. (a) Show that an observer moving with relativistic speed will measure a longer time interval from his frame as compared to the time interval in the rest frame. For convenience, you can consider motion along x-axis only.
   (b) Prove that the norm of a 4-vector $(X_\mu X^\mu)$ is a Lorentz invariant quantity.

   [6+5=11]

2. (a) Derive the relativistic velocity addition formula.
   (b) Hence prove that the speed of any signal can not exceed c – the speed of light in vacuum.

   [6+4=10]

3. (a) What is the difference between group velocity and phase velocity? In which case they are equal and why?
   (b) The wave function for a plane wave is given by

$$\psi(x,\ t) = A \exp[i(p_x x - Et)/\hbar],$$

   where symbols have their usual meanings. Using the operator relations $\hat{p} = i\hbar\nabla$ and $\hat{E} = i\hbar\frac{\partial}{\partial t}$, derive the Schrödinger equation of motion for this plane wave.

   [(2+2)+6=10]

4. A general angular momentum operator is defined in Cartesian coordinates by $\hat{J} \equiv (\hat{J}_x, \hat{J}_y, \hat{J}_z)$. Prove the following:
(i) $[\hat{J}_x, \hat{J}_y] = i\hbar\hat{J}_z$
(ii) $[\hat{J}_z, \hat{J}^2] = 0$
(iii) $[\hat{J}^2, \hat{J}_\pm] = 0$
where $\hat{J}_\pm = \hat{J}_x \pm i\hat{J}_y$ are the ladder operators.

[3 × 3=9]

INDIAN STATISTICAL INSTITUTE
Mid – Semestral Examination: 2017-18
B.Stat III Geology Elective

Date: 08/01/2018 **Maximum Marks: 40** **Time: 2 hours**

(Answer all questions. <u>Answer should be brief and to the point</u>)

1. Describe the processes that convert: the sediments into sedimentary rocks. **2**

2. In which kind of rocks would you expect to get sink holes? Why? 1+2=**3**

3. N.L. Bowen reproduced the crystallization process of magma in laboratory. In several of his laboratory experiments he commonly found that "crystals of olivine are surrounded by crystals of pyroxenes". What conclusions can be drawn from this observation? How the chemical composition of plagioclase feldspar gets modified with magma temperature in continuous reaction series? 2+2=**4**

4. The most likely way of generating magma is by partial melting of the existing rocks. Explain 'partial melting of existing rocks'. **3**

5. Explain with an example how progressive chemical weathering can lead to formation of ore. **3**

6. Where would you plan to mine for manganese nodules? How do these nodules form? 1+1=**2**

7. What is an aquifer? Mention two properties of a good aquifer. 1+2=**3**

8. Define porosity. Why porosity decreases with depth. 1+2=**3**

9. Biological activity may help and accelerate chemical weathering. Critically evaluate the statement. **3**

10. Explain why Hawaii beach-sand is mostly green in color. **2**

11. Describe the changes that take place in the parent rock during metamorphism? **3**

12. Write short notes on *any two* of the following: (a) 'sources of heat needed for metmorphism', (b) 'foliation and lineation', and (c) 'uniform stress and differential stress'. 3+3=**6**

13. Why quartz is the common mineral in most of the clastic sedimentary rocks? **2**

14. Name one living fossil. **1**

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2017-2018

B. Stat. (Hons.) 3rd Year. 2nd Semester

Nonparametric and Sequential Methods

Date: February 19, 2018 | Maximum Marks: 40 | Duration: 2 hours

---

- Answer all the questions.
- **You should present all your arguments while answering a question.**
- You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.

---

1. Suppose $X_1, \ldots, X_n$ are i.i.d. with a common distribution function given by $F(x - \theta)$, where $\theta \in \mathbb{R}$ is unknown and $F \in \Omega_0$ also is unknown, where

$$\Omega_0 := \{F : F \text{ is an absolutely continuous cdf with unique median at } 0\}.$$

We wish to test $H_0 : \theta = 0$ against $H_1 : \theta > 0, F \in \Omega_0$. Consider the sign test for this problem. Show that this test is strictly unbiased, that is, show that for admissible values of α, $P_{H_0}(S \geq k) = \alpha$ and $P_G(S \geq k) > \alpha$ for any G in Ω_{alt}, where S is the test statistic. Also, find the Hodges-Lehmann point estimator of θ based on S.

[9+5=14]

2. Suppose X_i's, $i \geq 1$, are i.i.d. with a common distribution function given by $F(x - \theta)$, where $\theta \in \mathbb{R}$ is unknown and $F \in \Omega_s$ also is unknown, where

$$\Omega_s := \{F \in \Omega_0 : F \text{ is symmetric around } 0\text{, i.e., } F(x) + F(-x) = 1 \,\forall\, x\},$$

with Ω_0 as in Q. 1 above. We wish to test $H_0 : \theta = 0$ against $H_1 : \theta > 0, F \in \Omega_s$. Consider the Wilcoxon signed rank test for this problem. Show that this test is consistent against all alternatives for which $P(X_1 + X_2 > 0) > 1/2$. [13]

3. Suppose X_i's, $i \geq 1$, are i.i.d. with a common distribution function F such that $\mathrm{E}_F(X_1^4) < \infty$, $\mu(F) := E_F(X_1) \neq 0$ and $\sigma^2(F) := \mathrm{Var}_F(X_1) > 0$. We wish to estimate $\mu^2(F) = \mathrm{E}(X_1 X_2)$. An unbiased estimator of $\mu^2(F)$ is given by

$$U_n := \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} X_i X_j.$$

Show by use of projection that $\sqrt{n}(U_n - \mu^2(F)) \xrightarrow{d} \mathrm{N}(0, 4\mu^2(F)\sigma^2(F))$ as $n \longrightarrow \infty$.

[13]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2017-18

BSTAT III YEAR

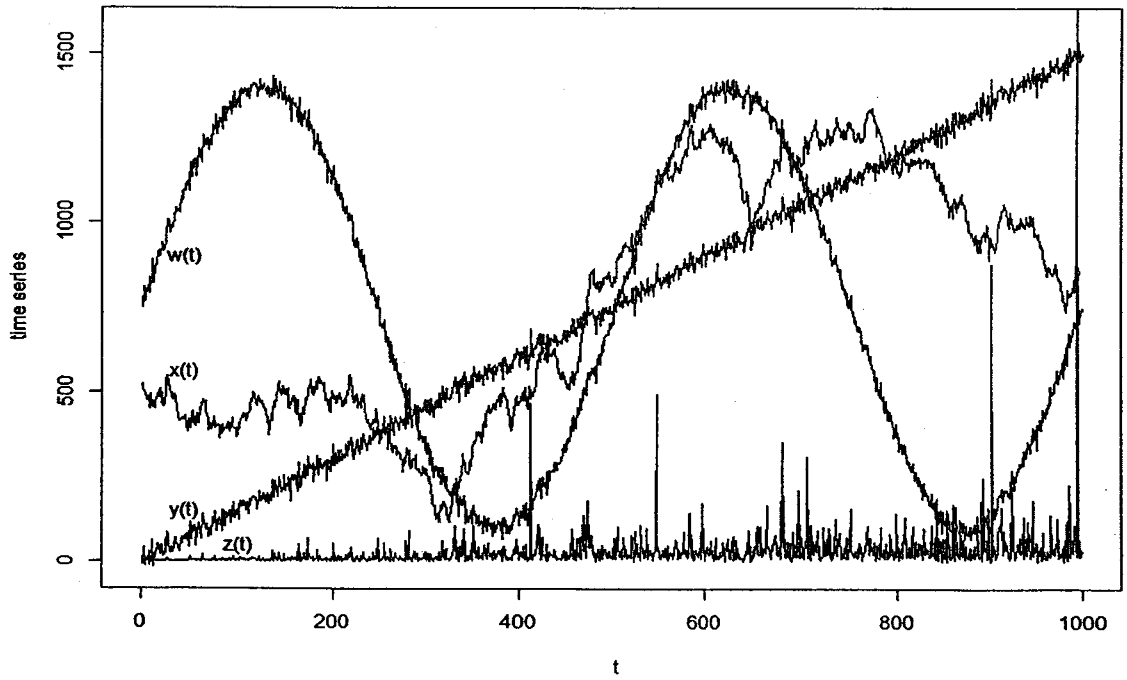
Statistics Comprehensive

Date: 20 February 2018 | Maximum marks: 100 | Duration: 2 hours

This examination is open book, open notes. Answer any five out of the six questions, each carrying 20 marks.

1. Suppose the joint distribution of the random variables Y, X_1 and X_2 is multivariate normal, the correlation of Y with $\beta_0 + \beta_1 X_1 + \beta_2 X_2$ is equal to the multiple correlation of Y with X_1 and X_2, and $E(Y) = E(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$. Can it be claimed that $\beta_0 + \beta_1 X_1 + \beta_2 X_2$ is the regression of Y on X_1 and X_2? Explain.

2. The following figure shows the overlaid time plots of four time series, $w(t)$, $x(t)$, $y(t)$ and $z(t)$ for $t = 1, \ldots, 1000$. For which of the four series would it be reasonable to look for a stationary model, after first order differencing of the given series at lag 1? Explain.



3. Specify precisely the Generalized Linear Model (GLM), where the response has the exponential distribution, and the canonical link function is used. Suggest a good reason why, instead of using this link function, one might prefer the link function log( ) (popular in the field of Survival Analysis). Write down the likelihood (in the simplest possible form) of the regression parameters under the GLM with the latter link function, assuming there are n paired observations on the response and a scalar predictor. Explain how you can maximize it for estimating these parameters.

P.T.O

4. In a district, there are 5 villages with population 200, 500, 300, 100 and 300. You have to draw a PPSWOR sample of size 2 by using an unbiased coin. You can toss it independently as many times as you like. Show every step of your procedure to get the sample.

5. There are eight similar looking balls of which seven have the same weight, while the other one is heavier. You are given an ordinary balance with two sides. Here is a method to detect the heavier ball: *Draw two balls randomly from the eight and compare their weights in the balance. If one side is heavier, the ball is detected. Else, draw another two balls randomly from the remaining and follow the same procedure.* It is evident that the maximum number of weighing is 4. Find out the expected number of weighing.

6. X follows the Poisson distribution with parameter λ. We would like to estimate $e^{-2\lambda}$ from this single observation.

   (a) Find the MVUE.

   (b) Suggest a "better" estimate, explaining your notion of "better".

INDIAN STATISTICAL INSTITUTE

Periodical Examination

B. STAT III YEAR

DESIGN AND ANALYSIS OF ALGORITHMS

Date : 21.02.2018 Maximum Marks : 40

Duration : 2 Hours 30 minutes

(Although the paper carries a total marks of 60, the maximum marks you can score is 40)

Question 1: Solve the following Recurrence Relations: (5×2)

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

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

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

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

(e) $T(n) = T(n-1) + O(n^2)$

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

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

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

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

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

(e) $O(2^{2\log n}) = \Omega(n^3)$

Question 3: Given an array A of n integers in the range $[0, k]$, where $k = O(\log n)$. Design an algorithm to find the most frequent integers in A. Analyse your algorithm. (Hint: may try to use AVL tree) [10]

P.T.O

Question 4: Given an array A of n integers. Write an $O(n)$ time algorithms to build a min-heap of A. Analyze running time of your algorithm. [5+5]

Question 5: 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 6: Show all the steps of inserting the following keys into a AVL tree one by one. After each insertion, if the AVL-tree is not balanced then make necessary rotation/rotations to balance it. While rotating the tree mention which AVL-tree rotation you have used and why?
The keys are $\{39, 22, 63, 13, 27, 52, 25, 23, 53\}$. [10]

Date: 22.2.2018
Time: 2 hours

Statistical Methods in Genetics
B-Stat (3rd Year)
Mid Semester Examination 2017-18

This is an open notes examination.
The paper carries 30 marks. Answer all questions.

1.Suppose genotype data are available at an autosomal biallelic locus for three randomly chosen sets of individuals from three independent populations. Explain how you would test whether the allele frequencies at this locus are identical in the three populations. [12]

2. Consider a recessive disorder controlled by an autosomal biallelic locus with alleles A and a. In a study on trios (two parents and an offspring), 40 families are selected with exactly one affected parent and 80 families are selected with both parents unaffected. In the first set of families, 8 offspring are affected, while in the second set of families, 3 offspring are affected. Obtain the maximum likelihood estimate of the prevalence of the disease. [10]

3. Suppose the initial genotype frequencies at an autosomal biallelic locus are according to Hardy-Weinberg equilibrium proportions. If the fitness coefficients corresponding to the genotypes are inversely proportional to the initial genotype frequencies, explain whether the allele frequencies at the locus reach non-trivial equilibrium values. [8]

Mid Semester II Examination
Design of Experiments
B.Stat. III yr. 2017-2018

Date: 23.02.2018 Full marks 50 Time : 2 hours

1. Suppose that in a completely randomised experiment a control treatment is being compared with 9 test treatments. Given that the total number of 96 experimental units(e.u) are available for the experiment, how many e.u.s you should allocate to the control and test treatments, if your objective is to minimise the total variance of the BLUEs of all the elementary contrasts involving control treatment effect and test treatment effect. Give the necessary derivation of the results in support of your allocation. [12]

2. Show that for a block design $l'\tau$ is estimable iff l' belongs to the row space of the C-matrix of the design. [8]

3. Give an example of a disconnected, orthogonal block design with 6 treatments involving $b(\geq 4)$ blocks such that the rank of the C- matrix of the proposed design is 4. Show that your proposed design satisfies the underlined properties. [3+6+6=15]

4. Prove or disprove the following statement:

   Consider a binary, equireplicate, proper block design d with v treatments in b blocks, such that every pair of treatments occurs together in exactly one block in the entire design. Assume that $r \geq 2,\ v > 2k > 5$. Let $B_1, B_2, \ldots B_b$ denote the b blocks and S_i denote the set of blocks where treatment i occurs. Consider that $B_1 \in S_1$ and $B_2 \in S_2$. Consider a new block design d_0 derived from d by deleting treatment 1 and treatment 2 from the blocks B_1 and B_2 respectively, while retaining the other blocks of d as they are. The design d_0 allows estimability of all possible treatment contrasts. [15]

Date: 23.4.2018
Time: 3 hours

INDIAN STATISTICAL INSTITUTE

Statistical Methods in Genetics
B-Stat (3rd Year) 2017-2018
Semester Examination

This paper carries 60 marks.

1. Explain each of the following terms with a suitable example:

(a) Snyder's Ratios
(b) heterozygote advantage
(c) segreagation ratio
(d) additive QTL model [3 x 4]

2. The genotype distribution of a random set of 100 individuals at an autosomal biallelic locus in an inbred population is as follows:

Genotype	Frequency
AA	52
Aa	33
aa	15

Obtain a 95% confidence interval for the allele frequency of *A* based on its maximum likelihood estimate. [12]

3. Suppose, in every generation of a certain population, a fraction α practises self mating, while the remaining $(1-\alpha)$ fraction of the population practises random mating. The initial genotype frequencies at a biallelic locus in this population are D_0, H_0 and R_0. Examine whether the genotype frequencies reach equilibria. If so, what are the equilibrium values? If not, provide suitable justification. [12]

P.T.O.

4(a) Consider a dominant disorder controlled by an autosomal biallelic locus. If an individual is affected, who among the following relatives is most likely to be affected: a parent, a sib or an offspring?

(b) Explain whether the genotypes of a pair of sibs conditioned on the genotype of one of the parents are mutually independent. [8 + 4]

5(a) Consider a case-control association study at an autosomal biallelic locus based on a combined sample size of 100 unrelated individuals. What should be the minimum allele-based Odds Ratio (greater than 1) such that the large sample test for association based on Odds Ratio provides significant evidence of association at level 0.05?

(b) Consider data on a quantitative trait for 50 randomly chosen individuals from a population along with their genotypes at a triallelic marker locus (with alleles *A*, *B* and *C*). However, there was no individual in the sample with genotype *CC*. Explain how you would test for genetic association between the marker locus and the quantitative trait, clearly stating the relevant test statistics and the underlying distributions in the absence of genetic association. [6 + 6]

INDIAN STATISTICAL INSTITUTE

End-Semestral Examination: 2017-18

Subject Name : **Number Theory** Date: 23/04/18

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

Note: Attempt all questions. Marks are given in brackets. Total score is 56. **State results** clearly which you want to use. Use **separate page** for each question.

Problem 1. Let $n = pq$ be product of two primes $p, q \equiv 3 \mod 4$. Show that the function $f : QR_n \to QR_n$ mapping y to $y^{\frac{\phi(n)+4}{8}} \mod n$ is well defined and it is a bijection. [3+5=8]

Problem 2. Let p be an odd prime. Prove that if g is primitive root (a generator of the multiplicative group $\mathbb{Z}_p^*$) then it must be quadratic non-residue modulo p. Find all forms of primes p, for which the set of all primitive roots is exactly the set of all quadratic non-residues modulo p. [3+5= 8]

Problem 3. Let $\xi = [a_0, a_1, \ldots]$ and $\{\frac{h_n}{k_n}\}$ denote the convergent of ξ. Prove that the sequence $\langle x_j := \frac{h_{n-1}+jh_n}{k_{n-1}+jk_n} \; 0 \le j \le a_{n+1}\rangle$ is increasing if n is odd, decreasing if n is even. [8]

Problem 4. Let $\nu(x) = \sum_{p \le x} \log p$ and $\psi(x) = \sum_{p^\alpha \le x} \log p$. Show that $\lim_{x\to\infty} \frac{\psi(x)}{x} = 1$ if and only if $\lim_{x\to\infty} \frac{\nu(x)}{x} = 1$. [8]

Problem 5. Let d_n denote the number of divisor of n. Show that $\sum_{t|n} d_t^3 = (\sum_{t|n} d_t)^2$. [8]

Problem 6. Let $x > 1$ be real and $\Theta(x) = \prod_{p \le x} p$. Show that $\Theta(x) \le 4^x$. [8]

Problem 7. Describe RSA encryption system. Show that the factorization of RSA modulus N can be done efficiently from the knowledge of the secret key exponent d. [3+5=8]

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2017 - 18

Course: B. STAT. II & III Subject: Differential Equations
Date: 23.04.2018 Maximum Marks: 60 Duration: 3hrs.

Any result that you use should be stated clearly

1. (a) Find the Laplace transform of

$$\frac{cos\sqrt{t}}{\sqrt{t}}$$

(b) Solve

$$\frac{dx}{dt} + ky = a \sin kt, \qquad \frac{dy}{dt} - kx = a \cos kt$$

using Laplace transform technique, given that $x(0) = 0,\ y(0) = b$.
(c) Using Laplace transform, solve the following initial boundary value problem:

$$\text{Partial differential equation} : \frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < l,\ t > 0$$

$$\text{Boundary conditions} : u(0,t) = u(l,t) = 0$$

$$\text{Initial conditions} : u(x,0) = \lambda \sin(\frac{\pi x}{l}), \quad \frac{\partial u(x,0)}{\partial t} = 0.$$

[5+6+7]

2. (a) Prove that the system

$$\frac{dx}{dt} = -2y + yz - x^3, \ \frac{dy}{dt} = x - xz - y^3, \ \frac{dz}{dt} = xy - z^3$$

has no closed orbits.
(b) Consider the system

$$\frac{dx}{dt} = \mu x + x^3 - x^5.$$

P.T.O

i) Find algebraic expressions for all the fixed points as μ varies.
(ii) Sketch the bifurcation diagram of fixed points x^* vs. μ. Be sure to indicate all the fixed points and their stability.
(iii) Calculate μ_c, the parameter value at which the nonzero fixed points are born in a saddle-node bifurcation. [4+(3+4+2)]

3. (a) Find the Frobenius series solution(s) of the equation

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

about the regular singular point $x = 0$.
(b) 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. The initial radius of the raindrop is r_0 and its radius at time t is r. Find the acceleration of the raindrop at time t. [7+5]

4. (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 n is an integer and $J_n(x)$ is the Bessel function of order n.

(b) Prove that

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

where λ_m, λ_n are the positive zeros of the Bessel function $J_p(x)$ and p is an integer. [4+8]

5. Prove that
(i)

$$nP_n(x) = x\frac{dP_n(x)}{dx} - \frac{dP_{n-1}(x)}{dx} \quad (n \geq 1),$$

where $P_n(x)$ is the Legendre polynomial of degree n.
(ii)

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x) \quad (n \geq 1),$$

where $H_n(x)$ is the Hermite polynomial of degree n. [3+2]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

B. STAT III YEAR

Design and Analysis of Algorithms

Date: 26.04.2018 Maximum Marks : 50 Duration : 2.5 hours

(You can answer any part of any question. Maximum you can score is 50)

1. (a) 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 algorithm should use the union by rank heuristic for the union operation of two trees. (You must clearly show every step of the execution of your algorithm).

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

   (b) Clearly show the steps of execution 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.

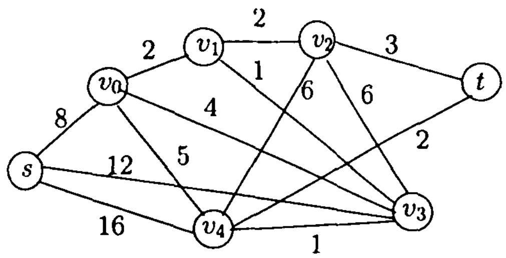


Figure 1: Shortest Path

5+5

2. **Subset Sum Problem** is stated as follows: Given a set X of positive integers in an array $X[1..n]$, and an integer T, test whether any subset of X sums to T.

   Write a dynamic programming based algorithm for solving Subset Sum Problem. Analyze running time of your algorithm. 7+3

3. State and prove lower bound of the time complexity for the comparison based sorting problem with n integers. (2+10)

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)

5. Let $\mathcal{H}$ be a finite collection of hash functions that map a given universe $\mathcal{U}$ of keys into the range $\{0, 1, ..., m-1\}$. Such a collection is said to be universal if for each pair of distinct keys $x, y \in \mathcal{U}$, the number of hash functions for which $h(x) = h(y)$ is precisely $\frac{\mathcal{H}}{m}$. In other words, with a hash function randomly chosen from $\mathcal{H}$, the chance of a collision between x and y when $x \neq y$ is exactly $\frac{1}{m}$.

   (a) Let h be chosen from a universal collection of hash functions and is used to hash n keys into a table of size m, where $n < m$. Show that the expected number of collisions involving a particular key x is less than 1.

   (b) Let us choose our table size m to be prime. We decompose a key x into $r + 1$ bytes (i.e., characters, or fixed-width binary sub-strings), so that $x = < x_0, x_1, x_2, ..., x_r >$; the only requirement is that the maximum value of a byte should be less than m. Let $a = < a_0, a_1, a_2, ..., a_r >$ denote a sequence of $r + 1$ elements chosen randomly from the set $\{0, 1, ..., m-1\}$. Define a corresponding hash function h_a where $h_a(x) = \sum_{i=0}^{r} a_i x_i \ mod \ m$.

   Show that $\mathcal{H} = \bigcup_a \{h_a\}$ is a universal class of hash functions. 5+5

6. Define Discrete Fourier Transform (DFT) and Inverse DFT (IDFT). Show that the algorithm for solving an IDFT can be used for solving DFT, and vice-versa. 1+1+8

INDIAN STATISTICAL INSTITUTE
Second Semester Examination: 2017-2018
B. Stat. (Hons.) 3rd Year. 2nd Semester
Nonparametric and Sequential Methods

April 30, 2018 | Maximum Marks: 60 | Duration: 3 hours

- This question paper carries 62 points. Answer as much as you can subject to the restriction in question 2. However, the maximum you can score is 60.
- You should provide as much details as possible while answering a question. You must state clearly any result stated (and proved) in class you may need in order to answer a particular question.
- You may use scientific calculator for numerical calculation.

---

1. Let

$$\Omega_s := \{f : f \text{ is a pdf with unique median at } 0 \text{ and } f(x) = f(-x)\,\forall\, x\}.$$

Suppose $X_1, \ldots, X_n$ are i.i.d. with a common pdf given by $f(x-\theta)$, where $\theta \in \mathbb{R}$ is unknown and $f \in \Omega_s$ also is unknown. We wish to test $H_0 : \theta = 0$ against $H_1 : \theta > 0, f \in \Omega_s$. Consider the sign test and the Wilcoxon signed rank test for this problem. Let S and T denote the corresponding test statistics. Show that the asymptotic relative efficiency of S relative to T is given by

$$\frac{f^2(0)}{3\left(\int f^2(x)\,dx\right)^2}.$$ [15]

2. Answer *any one* of parts (a), (b) and (c):

(a) Find the variance of Spearman's rank correlation r_s under the assumption of independence. [7]

(b) Find the projection of Kendall's tau which can be used to obtain its asymptotic null distribution. [7]

(c) Consider k populations which differ only in their locations. Suppose a random sample of size n_i is drawn from the i-th population ($i = 1, \ldots, k$), and suppose, moreover, that these k samples are independent. Develop a suitable nonparametric test for equality of the locations. [Note. You are not expected to derive the asymptotic null distribution.] [7]

[P.T.O.]

3. Suppose that X_i, $i \geq 1$, are i.i.d. observations from $\mathrm{N}(0, \sigma^2)$, where $\sigma > 0$ is unknown. We wish to test $\mathrm{H}_0 : \sigma = \sigma_0$ versus $\mathrm{H}_1 : \sigma = \sigma_1$, where $0 < \sigma_0 < \sigma_1$ are known. Suppose $0 < \alpha, \beta < 1$. Let $\tau = \sigma_1/\sigma_0$.

(i) Consider the most powerful size-α test for H_0 versus H_1 whose power is at least $1 - \beta$. Show, using an approximation for the quantile of chi-square distribution based on a suitable normal approximation for the chi-square distribution, that the required number n of sample satisfies

$$n \geq \frac{2(z_{1-\alpha} + z_{1-\beta}\tau^2)^2}{(\tau^2 - 1)^2}, \qquad (*)$$

where for $0 < p < 1$, z_p is the p-the quantile of standard normal distribution.

(ii) Consider now the SPRT for testing H_0 versus H_1 having target strength (α, β) and with Wald's approximation for the boundaries. Write down approximate expressions for the average sample numbers (ASN) under H_0 and H_1 in terms of α, β, τ.

(iii) Assume now that $\tau = 2.0$. Choose $\alpha = \beta = 0.05$. It is known that $z_{0.95} = 1.645$. Compute the quantity on the right-hand side of $(*)$ and the ASN's in (ii). Offer your comments on the findings. [6+4+(3+3)=16]

4. Suppose that X_i, $i \geq 1$, are i.i.d. observations from a normal distribution with unknown mean $\theta \in \mathbb{R}$ and unknown variance $\sigma^2 > 0$. We wish to estimate θ by T such that $\mathrm{E}_{\theta,\sigma}(T - \theta)^2 \leq M \ \forall \ (\theta, \sigma)$, where M is a known positive quantity. Obtain, with reasons, one such estimator. [Note. You must justify all the steps of your argument.] [14]

5. Suppose that F is an unknown distribution function with density function $f = F'$. Let $X_1, \ldots, X_n$ be i.i.d. observations from this distribution.

(i) Explain what is meant by a kernel density estimator of f, based on $X_1, \ldots, X_n$.

(ii) Show, under suitable conditions, to be stated by you, that variance of the estimator in (i) converges to zero as n goes to infinity. [3+7=10]

INDIAN STATISTICAL INSTITUTE
Second Semester Examination 2017-18
B.Stat. III yr.
Design of Experiments

Date: 04 · 05 · 2018 | Maximum Marks 100 | Duration : 3.5 hours

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

1. (a) Consider an experiment where three different washing solutions are being compared to study their effectiveness in retarding bacteria growth in five-gallon milk containers. The analysis is done in a laboratory using five trials per day. Suggest a "good" design if the experiment is run for six days, clearly indicating the layout of the design and the appropriate model for analysing the data. Obtain the C- matrix of the design suggested by you and best linear unbiased estimator (BLUE) of the pairwise differences of the effects of washing solutions.

   (b) Show that for a connected block design with b blocks and v treatments, the variance -covariance matrix for $Q_i + \frac{r_i G}{n}$, $i = 1, \ldots, v$ is non-singular, where the notaions have their usual interpretations in a block design. (You can assume the form of the variance -covariance matrix of Q).

   (c) Prove or disprove: It is always possible to construct a connected incomplete orthogonal design with v treatments in b blocks. (You can assume the definition of orthogonality and connectedness of a block design). [(5+9) +6+5)=25]

2. (a) How many standard Latin Squares of order 4 are there? List all of these.

   (b) Prove or disprove: Given any standard Latin Square L_1 of order 4, it is always possible to find another Latin Square L_2 of order 4 such that L_1 and L_2 are mutually orthogonal.

   (c) An industrial engineer is investigating the effect of three assembly methods A, B, and C on the assembly time for a color television. Suppose the experiment is being carried out in 4 different factories. For this purpose, three operators are chosen at random from each factory and are engaged in three different work-periods suitable to each factory. Suggest a "good " design for this experiment, clearly indicating the layout of the experiment and the appropriate model for analysing the data. Write down the ANOVA table, indicating the test statistic for testing relevant statistical hypothesis. [6+8+(4+7)=25]

3. (a) Identify the confounded factorial effects in a 2^6 factorial experiment run in 2^3 blocks, where one of the block compositions is given by the set of treatments { def, af, be, abd, cd, ace, bcf, abcdef}.

   (b) For a 3^3 factorial experiment, define the main efffects and interaction effect contrasts with corresponding degrees of freedom. Show that any two factorial effect contrasts are mutually orthogonal.

   (c) Is it possible to construct a confounded $(3^5, 3^2)$ factorial experiment with blocks of size 3^2, where there is no loss of information for the main effects and two factor interactions? Justify your answer. Hence or otherwise find an upperbound for the number of factors that can be accommodated in such an experiment with blocks of size 9 . [7+(6+4)+(4+4)=25]

P.T.O

4. Consider the following two way design with five treatments A, B, C, D and E where the batches and days are two sources of nuisance variation in the experiment. In each cell, the observations are given within bracket, " -" indicating that the corresponding observation is missing.

Batch	Day					Row total
	1	2	3	4	5	
1	A(8)	B(7)	D(1)	C(7)	E(3)	26
2	C(11)	E(2)	A(7)	D(3)	B(8)	31
3	B(4)	A(9)	C(10)	E(-)	D(5)	28
4	D(6)	C(8)	E(-)	B(6)	A(10)	30
5	E(4)	D(2)	B(3)	A(8)	C(8)	25
column total	33	28	21	24	34	total (146)
trt. total	A(42)	B(28)	C(44)	D(17)	E(9)	

Estimate the two missing values, obtain the value of exact error sum of squares and treatment sum of squares with corresponding degrees of freedom. [8+4+13=25]

5. An experiment is carried out by a paper manufacturer who is interested in studying the effects of three different pulp preparation methods and four different cooking temperatures on the tensile strength of paper. The experiment is run for three days. For operational convenience, on each day, first a batch of pulp is prepared at random by one of the three methods and then the batch is divided into four parts, subjecting each part to one of the four tempeartures, selcted at random. The process is repeated until all the methods of pulp preapartions are undertaken on each day. The data are given below.

	Day 1			Day 2			Day 3			Row total
Pulp preparation method	1	2	3	1	2	3	1	2	3	
Temperature(^0F)										
200	30	34	29	28	31	31	31	35	32	281
225	35	41	26	32	36	30	37	40	34	311
250	37	38	33	40	42	32	41	39	39	341
275	36	42	36	41	40	40	40	44	45	364
column total	138	155	124	141	149	133	149	158	150	

State an appropriate model for the analysis of above data. Obtain the BLUEs of the pairwise differences of the effects of pulp preparation methods as well as the effects of temperatures. Write down the ANOVA table and hence obtain the standard errors of the above BLUEs.

[3+4+4+14=25]