Backpaper Examination
Design of Experiments
B.Stat. III yr. 2013-2014

Date: 04.08.14 Full marks 100 Time : 3.5 hours

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

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

   (b) Obtain the expression of the relative efficiency of a Randomised Block Design compared to a Completely Randomised Design.

   (c) Show that for a block design any function of treatment effects is estimable only if it is a contrast.

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

   (e) Suppose in a block design with b blocks each of size v, in the first b_1 blocks treatment 1 appears twice and each of the treatments $3, \ldots, v$, appears once. In the remaining b_2 blocks, each of the treatments $1, \ldots, v$ appears once, ($b_1 + b_2 = b$). Is this design connected? Orthogonal? (Give reasons in support of your answer) Obtain the C- matrix of the design and BLUE of any pair wise treatment comparison in terms of the adjusted treatment totals. [6+4+5+ 2+3+5+9=34]

2. (a) Show that there exists a set of $v-1$ mutually orthogonal Latin squares of order v, when v is a prime or prime power.

   (b) Construct a Galois Field of order 8 and hence construct two mutually orthogonal Latin Squares of order 8. [6+ 8+10=24]

3. (a) Suppose that in an Randomised Block design with 5 treatments in 6 blocks, two observations corresponding to block 1; treatment 1 and block 2; treatment 3 are found missing. How will you estimate the missing values to obtain the correct value of the error sums of squares and the corresponding degrees of freedom under the model with the available observations.

   (b) Suppose that the F-test for equality of the effects of the treatments is rejected when the augmented data are analysed using the estimated values obtained in part (a). Do you need to modify your estimates of the missing values to carry out the hypothesis testing? If so, how and if not, justify your answer. [7+7=14]

4. (a) Identify the set of confounded effects in a $(2^6, 2^4)$ experiment where the key block is given below.

0	0	0	0	0	0
1	1	1	1	0	0
0	0	1	1	1	1
0	1	0	1	0	1
0	1	1	0	0	1
1	1	0	0	1	1
1	0	1	0	0	1
1	0	0	1	0	1
0	1	1	0	1	0
0	1	0	1	1	0
0	0	1	1	0	0
1	0	0	1	1	0
1	0	1	0	1	0
1	1	0	0	0	0
0	0	0	0	1	1
1	1	1	1	1	1

(b) Give a balanced confounded scheme for a $(3^3, 3^2)$ factorial experiment(blocks of size 9) in four replications retaining as much information as possible for the main effects and two factor interactions. Compute the loss of information for different factorial effects in the scheme suggested by you. For anyone of the replications chosen by you, construct the key block.

(c) A paper manufacturer is interested to study the effect of three different pulp preparation methods (factor A) and four different cooking temperatures(factor B)on the tensile strength of the paper. The pilot plant is capable of making 12 runs per day and the experimenter wishes to continue this experiment for three days.

i. If the levels of both the factors can be changed easily, suggest a design for the experiment.

ii. If the frequent change to the pulp preparation method is difficult to handle, suggest a modification of your experiment described in part(i).

iii. Clearly state the model for the design suggested in part(ii). Write down the ANOVA table under the model and indicate the test statistics for testing the hypotheses of interest.

iv. Which of the main effects of factor A and B are estimated with more precision?

[6+6+(2+2+2+8+2)=28]

INDIAN STATISTICAL INSTITUTE

Semestral (Backpaper) Examination, 2013-14 Date: 08.08.14

B.Stat Third Year

Time : 3 Hours Statistical Inference II Full Marks: 100

1. Prove or disprove the following statements. [6 × 8 = 48]

(a) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed as $N(\mu, \sigma^2)$ variates. The statistic $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ is the UMVUE for μ because it is unbiased and it is a function of the complete sufficient statistic $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.

(b) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed with distribution function F. If θ is the median of the distribution and $F(\theta - x) + F(\theta + x) < 1$ for all $x > 0$, the sample median cannot be an unbiased estimator of the population median.

(c) If the density function f of a distribution is of the form $f(x_1, x_2, \ldots, x_d) = \psi(|x_1 - \theta_1| + |x_1 - \theta_1| + \ldots + |x_p - \theta_p|)$ for some real valued function ψ, the half-space median and the spatial median of the of the distribution coincide.

(d) Let f_h denote the histogram estimate of a univariate density function f when h is used as the bin width. Assume that f has bounded first derivative. If h varies with n at a rate $O(n^{-\delta})$ for some $\delta \in (0, 1)$, for any fixed x, $f_h(x)$ converges to $f(x)$ in probability.

(e) The function $g(\theta) = 4\sum_{i=1}^{10} |i - \theta| + \sum_{i=1}^{10}(i - \theta)$ has a unique minimizer.

(f) The Nadaraya-Watson estimate (based on Gaussian kernel) of a regression function becomes a nearest neighbor estimate as the bandwidth shrinks to zero, while it becomes a constant as the bandwidth tends to infinity.

(g) The two-sample test based on the Kolmogorov-Smirnov maximum deviation statistic is consistent (power converges to 1 as the sample sizes increase) for general alternative $F \neq G$.

(h) Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed with distribution function F. Let F_n be the empirical distribution function based on $X_1, X_2, \ldots, X_n$. The statistics $T_1 = \int [F_n(x) - F(x)]^2\, dx$ and $T_2 = \int |F_n(x) - F(x)|\, dF(x)$ both have the distribution-free property.

2. Suppose that $\mathbf{X} = (X_1, X_2)'$ has an elliptically symmetric distribution with $E(X_1) = E(X_2) = 0$ and $Corr(X_1, X_2) = \rho$.

(a) Show that $P(X_1 > 0, X_2 > 0) = \frac{1}{4} + \frac{1}{2\pi} sin^{-1}\rho$. [8]

(b) A person generated 20 observations from this distribution. The numbers of observations in the four quadrants were 6, 3, 7 and 4, respectively. Use an appropriate distribution-free method to check whether the underlying distribution is spherically symmetric. [6]

3. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed with density function f, which is symmetric about θ. Show that the Wilcoxon signed rank statistic for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ has the distribution-free property. Will it have the distribution-free property even when f is not symmetric ? Justify your answer. Describe how the Hodges-Lehmann estimate of θ is obtained from this statistic. [4+2+6]

4. Let $X_1, X_2, \ldots, X_m$ be independent and identically distributed with density function f and $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed with density function g, where $f(x) = \theta^{-1} g(x/\theta)$ for all x and some $\theta > 0$. Show that Sukhatme's statistic for testing $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$ has the distribution-free property. Describe how you will construct a 95% confidence interval for θ using this statistic. [5+5]

5. (a) Consider a Poisson distribution with parameter λ. Assuming the independence of the observations generated from this distribution, show that the sequential probability ratio test for $H_0 : \lambda = 1$ against $H_1 : \lambda = 2$ converges with probability 1. [6]

   (b) Consider a bivariate uniform distribution over a square with vertices (θ, θ), $(-\theta, \theta)$, $(-\theta, -\theta)$ and $(\theta, -\theta)$, respectively. Construct a sequential probability ratio test (with $\alpha = \beta = 0.01$) for $H_0 : \theta = 1$ against $H_1 : \theta = 2$. Draw the OC function and the ASN function of this test for θ varying in the interval (0, 2). [4+3+3]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : 2014-2015

B. Stat. (Hons.) III Year

Subject : SQC & OR

25/08/14

Full Marks: 100 **Time: 3hours** **Date of Examination:**

NOTE: This paper carries 106 marks. You may answer any part of any question; but the maximum you can score is 100.

1. A student is planning the coming semester. In particular, he is attempting to allocate the weekly number of hours of study to the individual courses he is taking. The maximum mark of each course is 100. Each hour of study will increase his mark by a certain quantity (starting at zero). Table 1 shows the marginal improvements of the marks given each hour of study (per week) as well as the marks required for passing the course.

Table 1: Input ~~question~~ data for Question 1

	Linear Models	Inference	Sample Surveys	SQC &OR	Elective
Marginal improvement of mark	5	4.5	5.5	3.5	5.5
Pass marks	50	55	60	50	50

For example, if the student were to allocate 15 hours per week to *Linear Models*, then his final mark is expected to be $15 \times 5 = 75$, which means passing the course.

The student's objective is to minimize the total number of hours studies. In addition the following constraints have been identified:-

a. A passing grade must be achieved in ***each course.***
b. Obtain an average grade of at least 64%.
c. The number of hours allocated to SQC & OR should be at least 20% of the number of hours allocated to the four subjects combined.
d. Suppose that the student has the option to sell coffee at Barista in his spare time. This job pays Rs. 250 per hour. The student has a total of 80 hours available for studying and selling coffee.

Formulate the problem so that our student makes at least Rs.2500 per week (and passes each course). [20]

2.a. The following table represents a specific simplex interaction for a maximization problem. All variables are non-negative.

	y_0	$-x_2$	$-x_4$	$-x_5$	$-x_6$	$-x_7$
x_0	620	-5	4	-1	-10	0
x_8	12	3	-2	-3	-1	5
x_3	6	1	3	1	0	3
x_1	0	-1	0	6	-4	0

i. Categorize the variables as basic and non-basic and write the current values of all the variables.

ii. Identify the non-basic variables that have the potential to improve the value of x_0. If each such variable enters the basis, determine the associated leaving variable and the associated change in x_0.

iii. Which non-basic variable(s) will not cause a change in the value of x_0 when selected to enter the basis?

b. Consider the problem P.

$$\begin{aligned} \text{P} : &\ \text{Max } z = cx \\ \text{st} &\ \ Ax = b \\ &\ \ x \geq 0,\ b \geq 0 \end{aligned}$$

How is the optimal solution affected when the cost vector c is replaced by $\lambda c, \lambda > 0$?

[(3+4+2)+6=15]

3.a 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 $\bar{p} = 0.20$.

i. Find the 3σ control limits for the control chart.
ii. When the process is in-control, how often would false alarms be generated?
iii. If the process average deteriorates to $\bar{p} = 0.30$, after how many samples will the control chart be able to detect this shift?
iv. Suggest two methods of reducing the out-of-control ARL.

b Find the probability of detecting a shift of 2σ in the process mean on the first sample following the shift in an $\bar{X}$ – chart with the usual 3-sigma limits and sub-group size $n = 5$.

[(3+5+5+5)+7=25]

4. 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	17	−35	−17	H
x_5	B	E	11	0	2
x_4	C	F	−6	−3	2
x_2	D	G	−5	−2	I

Without performing the simplex iterations, find the missing entries $A, \cdots, I$. Show your calculations.

[15]

5. a. Define quality costs. Give their broad categories for a manufacturing organization.

b. How does the ISO define quality?

c. The Dean's Office handles the admission process of the ISI. Suggest briefly one (or more) SQC tool which they may use for this process.

d. Distinguish between ***control limits*** and ***specification limits***.

e. Distinguish between ***engineering quality control*** and ***statistical quality control***.

[(3+2)+3+4+6+6=24]

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

i. Pick out the appraisal quality cost from the following:

a. Fees for an outside auditor to audit the quality management system.
b. Time spent to review customers' drawing before contract.
c. Time spent in concurrent engineering meetings by a supplier.
d. Salary of a metrology lab technician who calibrates the instruments.
e. None of the above.

ii. Which of the following will be considered a failure quality cost?

a. Salaries of personnel testing repaired products.

b. Cost of test equipment.
c. Cost of training workers to achieve production standards
d. Incoming inspection to prevent defective parts coming into stores.
e. All of the above.

iii. Which of the following is not a quality cost:
a. Cost of inspection and test.
b. Cost of routine maintenance of plant and machinery.
c. Cost of routine maintenance of test instruments.
d. Salary of SPC analysts
e. None of the above.

iv. When measurements show a lack of statistical control, the standard error of the average:
a. Is related to the confidence limits.
b. Is a measure of process variability.
c. Is simple to compute.
d. Has no meaning.

v. The PDCA wheel is attributed to
a. Box.
b. Juran.
c. Dodge and Romig.
d. Deming.

vi. Which of the following is **not** a benefit of SPC charting?
a. Charting helps in evaluating of system quality.
b. It helps identify unusual problems that might be fixable.
c. It encourages people to make continual adjustment to processes.
d. Without complete inspection, charting still gives a feel for what is happening.

vii. Who is considered to be the Father of SQC?
a. Box.
b. Juran
c. Deming
d. Taguchi
e. None of the above.

[1 × 7 = 7]

SELECTED VALUES OF THE STANDARD GAUSSIAN CDF

$\Phi(z) = \Pr(Z \leq z)$ when $Z \sim \mathcal{N}(0, 1)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

NOTE: $\Phi(-z) = 1 - \Phi(z)$.

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : (2014-2015)

B.Stat. 3rd Year

PARAMETRIC INFERENCE

Date: 26 August, 2014 | Max. Marks: 100 | Duration: $2\frac{1}{2}$ Hours

GROUP-A

Answer as many questions as you can from Group-A.

The maximum you can score is 82.

1. (a) Define a sufficient statistic and a minimal sufficient statistic.

(b) Let $X_1, \ldots, X_n$ be i.i.d., each following $U(\theta, \theta+1)$, $\theta \in R$. Find a nontrivial sufficient statistic for θ. [4+5=9]

2. Describe the notion of completeness of a statistic. [5]

3. Let the observations $X_1, \ldots, X_n$ be i.i.d. Bernoulli (p) where $0 < p < 1$.

(a) Prove that $T = \sum_{i=1}^{n} X_i$ is complete sufficient for p.

(b) What are the U-estimable functions of p? Justify your answer.

[9+10=19]

4. (a) Let $X_1, \ldots, X_n$ be a random sample from a $N(\theta, \theta^2)$ population where $\theta > 0$. Find a minimal sufficient statistic and show that it is not complete.

(b) Give an example to show that a complete statistic is not necessarily sufficient. [11+7=18]

5. (a) Define a uniformly minimum variance unbiased estimator (UMVUE).

(b) State and prove Rao-Blackwell Theorem.

(c) Let $\boldsymbol{X} = (X_1, \ldots, X_n) \sim P_\theta$, $\theta \in \Theta$ and $T = T(\boldsymbol{X})$ be a complete sufficient statistic for θ. Show that for every U-estimable parametric function $g(\theta)$, there exists a UMVUE of $g(\theta)$.

Also show that an unbiased estimator $\phi(T)$ of $g(\theta)$, based on T, is also a UMVUE. [3+10+10=23]

6. Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ population where $\mu \in R$ and $\sigma^2 > 0$ are both unknown.

(a) Show that $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n (X_i - \bar{X})^2\right)$ is complete sufficient for (μ, σ^2).

(b) Find the UMVUE of μ^3. [8+8=16]

GROUP-B

Answer all questions.

7. Let the observations $X_1, \ldots, X_n$ be i.i.d. Poisson (λ), $\lambda > 0$ and let $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Find $E(S^2 | \bar{X})$. [8]

8. Let $X_1, \ldots, X_n$ be i.i.d. observations, each following a discrete uniform distribution over $\{1, 2, \ldots, N\}$, where N is an unknown positive integer. Suppose it is known that $N \geq 2$. Show that $X_{(n)} = \max\,(X_1, \ldots, X_n)$ is not complete. [10]

INDIAN STATISTICAL INSTITUTE

Backpaper Examination : (2014-2015)

B.Stat. 3rd Year

NONPARAMETRIC AND SEQUENTIAL METHODS

Date: 27 July, 2015 Max. Marks: 100 Duration: 3 Hours

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

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

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

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

[13+10+10=33]

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

[7]

3. Let $X_1, X_2, \ldots$ be i.i.d $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Describe Stein's two-stage sampling procedure for obtaining a bounded length confidence interval for μ with confidence coefficient $(1 - \alpha)$. Prove the results you state. [14]

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

(b) Let $X_i, i = 1, 2, \ldots$ be i.i.d. Bernoulli (θ) where $0 < \theta < 1$. Consider the SPRT for testing $H_0 : \theta = 1/3$ against $H_1 : \theta = 1/2$ where the boundaries satisfy $0 < B < 1 < A < \infty$. Show, without using Stein's lemma, that the SPRT terminates with probability one under $\theta = 2/3$ [14+12=26]

P.T.O

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

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

Indian Statistical Institute

B. Stat. (Hons) III year, 2014-15

First Semester, Periodical examination

Linear Statistical Models

Date: August 27, 2014 Duration: 2 hours Maximum Marks: 50

Note: This paper carries 55 marks. Answer as much as you can.

1. Let $y \sim nor(X\beta, \sigma^2 I)$ where y is an n dimensional random vector, X is a real matrix of n rows and p columns, Rank $(X) = p,\ p < n$,

a. Obtain maximum likelihood estimates for β and σ^2. [7]

b. Obtain unbiased estimate of σ^2. [3]

c. Calculate var($\hat{\beta}$) where $\hat{\beta}$ is m.l.e of β. [2]

d. Let $t'\beta$ be a linear function of elements of β. Find BLUE of $t'\beta$. [8]

e. Let SSR= $\hat{\beta}'X'y$, and SSE= $y'y -$ SSR. Show that SSR and SSE are independent and $\frac{SSE}{\sigma^2} \sim \chi^2_{n-p}$. Find the distribution of $\frac{SSR}{\sigma^2}$. [12]

f. Let $\beta' = (\beta_1, \beta_2, ..., \beta_p)$. Find 95% confidence interval of β_i [3]

g. Suppose the hypothesis $H_0 : w'\beta = m$ is to be tested where w' is a real matrix of s rows and p columns, $s \le p$ and rank(w) = s and m is a given vector. Derive the test statistic for the said problem. [7]

h. Suppose the value of unbiased estimate of σ^2=100, $\hat{\beta}' = (3,5,2)$, $\hat{v}(\hat{\beta}_1) = 28$,

$\hat{v}(\hat{\beta}_2) = 24$, $\hat{v}(\hat{\beta}_3) = 18$, $\widehat{\text{cov}}(\hat{\beta}_1, \hat{\beta}_2) = -16$, $\widehat{\text{cov}}(\hat{\beta}_1, \hat{\beta}_3) = 14$, and $\widehat{\text{cov}}(\hat{\beta}_3, \hat{\beta}_2) = -12$. The hypothesis to be tested is $\beta_1 = \beta_2 + 4 = \beta_3 + 7$. Find the value of the test statistic using (7). [5]

2 State the intercept model with full rank. Write the ANOVA table using the corrected sum of squares. [8]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2014-15

Course Name : B.Stat. 3rd Year

Subject Name : Sample Survey

Date : 28.08, 2014 Maximum Marks : 40 Duration : $1\frac{1}{2}$ *hrs.*

Answer Q5 and any three from Q1 to Q4. Each question carries marks 10.

Notations are as usual.

1. Derive the unbiased estimator $\hat{P}$ of the population proportion P of members of a certain type in a popoulation of size N with a sample of size n drawn by SRSWR as well as by SRSWOR.

   Show that an unbiased estimator of

   $Var(\hat{P})$ for SRSWR is $\dfrac{\hat{P}(1-\hat{P})}{n-1}$ and that for SRSWOR is $\dfrac{\hat{P}(1-\hat{P})}{n-1}\left(1-\dfrac{n}{N}\right)$.

2. Derive the expression of $Cov(\bar{x}, \bar{y})$ for both SRSWR and SRSWOR of size n out of N. Also obatin the unbiased estimators for those.

3. 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.

   Hence show that

   $$-\frac{1}{n-1} \leq \rho_c \leq 1.$$

   Show further that the relative effeciencies of systematic sampling compared to SRSWR and SRSWOR in estimating the population mean are respectively,

   $$\frac{1}{1+(n-1)\rho_c} \quad \text{and} \quad \frac{N-n}{N-1} \times \frac{1}{1+(n-1)\rho_c}.$$

   Hence indicate how the units in the population should be arranged in order that systematic sampling may be highly efficient.

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

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

Prove that if finite population corrections are ignored, then

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

5. For a certain study carried out in Rajasthan during 1980-81, 4 tehsils were selected by SRSWOR from the 12 tehsils of Ajmer Division and from each selected tehsil a certain number of villages was also selected by SRSWOR. Finally, the sheep population in each selected village was counted. The necessary data are presented below. Also, given that the total number of villages in the entire Ajmer Division is 1488.

(i) Estimate the total sheep population in the entire Ajmer Division and also obtain an estimate of the standard error.

(ii) Also estimate the mean sheep population per village in the entire Ajmer Division along with the standard error.

Selected Tehsil	Total No. of villages	No.of sample villages	Sheep population in the selected villages
Behrar	102	10	266, 890, 311, 46, 174, 31, 17, 186, 224, 31
Bairath	105	12	129, 57, 64, 11, 163, 77, 278, 50, 26, 127, 252, 194
Ajmer	200	16	247, 622, 225, 278, 181, 132, 659, 403, 281, 236, 595, 265, 431, 190, 348, 232
Bansur	85	9	347, 362, 34, 11, 133, 36, 34, 61, 249

INDIAN STATISTICAL INSTITUTE

B.Stat.III

Mid-sem exam 2014-15

Physics III

Date: 29·08·14

Maximum Marks: 35

Attempt as many as you can.

1. A police car with velocity c/2 chases a robber's car going with velocity 3c/4. The police officer fires a bullet whose muzzle velocity relative to the gun is c/3. Does the bullet reach its target (a) according to Galileo, (b) according to Einstein?

[3+4]

2. A particle of mass m whose total energy is twice its rest energy collides with an identical particle at rest. If they stick together, what is the mass of the resulting composite particle? What is it's velocity?

[4+3]

3. A truck is travelling along the 45° direction is S frame at (ordinary) speed $\frac{2}{\sqrt{5}}$ c. Find:
   (i) Components u_x and u_y of the ordinary velocity.
   (ii) Components η_x and η_y of the proper velocity.
   (iii) Zeroth component η^0 of the 4-velocity.

   Consider system $\bar{S}$ moving in the x-direction with ordinary speed $\sqrt{\frac{2}{5}}$ c relative to S. Find:

   (i) Components $\bar{u}_x$ and $\bar{u}_y$ of the ordinary velocity.

   (ii) Components $\bar{\eta}_x$ and $\bar{\eta}_y$ of the proper velocity.

[1+2+1+3+2]

4. (a) In a sailboat, the mast is inclined at an angle θ . An observer standing on a dock sees the boat go by at speed v. What angle must this observer make with the mast?

   (b) Jack goes off in a space ship at a speed 3c/5 on his 25th birthday. After 4 years have elapsed on his watch, he turns around and heads back to earth with the same speed to meet his twin sister Jill. How old is each twin at their reunion? Explain why Jack cannot assume that he is at rest and Jill is in motion.

[3+3+1]

5. Light of wavelength 2000 A° falls on a metallic surface. If the work function of the surface is 4.2 eV, what is the
   (i) Kinetic energy of the fastest photoelectrons emitted.
   (ii) Stopping potential.
   (iii) Threshold wavelength for the metal.

[2+1+2]

6. An electron has a speed of 500 m/s with an accuracy of 0.004%. Calculate the certainty with which we can locate the position of the electron.

[4]

7. Consider a wave function at t=0:

$$\Psi(x,0) = \begin{cases} A\dfrac{x}{a} & 0 \le x \le a \\ A\dfrac{b-x}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

(i) Normalize Ψ (i.e. find A in terms of a and b).
(ii) Sketch Ψ as a function of x.
(iii) Where is the particle most likely to be found?
(iv) What is the probability of finding the particle to the left of a?
(v) What is the expectation value of x ?

[2+1+1+1+2]

INDIAN STATISTICAL INSTITUTE
MID-SEMISTRAL EXAMINATION
2014-2015

Course name: Elective Course for B. Stat III Year
Subject name: Psychology
Date: 29.8.2014
Maximum marks: 30
Duration: 1 hour

---

1. **Fill in the blanks: Any six (6×1)**

   (a) Psychology is the study of ______________

   (b) Brain hemispheres are connected by ______________________

   (c) Hormone that triggers sympathetic nervous system is ______________

   (d) Shifting attention between only two potential stimuli is called ________________________

   (e) Resting potential of a neuron is maintained at ________________.

   (f) Arousal and wakefulness is maintained by ________________________ of the brain.

   (g) Drugs that suppress brain activities are called ____________________

   (h) Paradoxical sleep is the other name for _____________________________

2. **Write short notes: Any four (2×4)**

   (a) Physiology of drug addiction

   (b) Species-typical behaviour

   (c) Circadian rhythm

   (d) Vigilance

   (e) Cognitive Psychology

   (f) HPA axis

3. **Indicate whether the following statements are true or false. Justify your answers with adequate reasons: Any eight (2×8)**

   (a) If a specific brain area is damaged, its functions can sometimes be assumed by corresponding area of the opposite hemisphere.

   (b) Attention process is determined by stimulus characteristics only.

(c) Function of sympathetic nervous system and parasympathetic nervous system are opposite.

(d) Externally taken psychotropic drugs have no effect on brain functions as these depressants/stimulants already exist in the brain processes.

(e) Prolonged stress causes immense harm to body and mind.

(f) NREM sleep is essential to survival.

(g) Behaviour is regulated by human nature only.

(h) Action potential is an all-or-none process.

(i) Certain drugs mimic certain neurotransmitters.

(j) Repetition of a stimulus helps gain attention.

Indian Statistical Institute

Mid-semestral examination : 2014-15

Course name: B.Stat.III

Subject name: Sociology

Date: 29.08.14 *Maximum marks:* 20 *Duration: One hour*

Note, if any:

Please answer any of the ~~four~~ five

a) Discuss the various perspectives in Sociology. Which one do you think important and why?
b) What do you mean by electronic community? Discuss its advantages and disadvantages.
c) Explain the concept of church, sect and cult from the perspective of Sociology.
d) What are the causes of the changes in the caste system?
e) Define rural sociology and its importance in Indian society.
f) Define social stratification and its various forms.
g) Explain briefly the usefulness of questionnaire in social research
h) Write a short note on deviance.

INDIAN STATISTICAL INSTITUTE

First Semester Examination : 2014-15

Course Name : **B.Stat. 3rd Year**
Subject Name : **Sample Surveys**
Date : Oct 31, 2014 **Total Marks : 50** **Duration : 3 hrs.**

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

1. (a) Given any sampling design p defined on a finite population $U = (1, 2, \ldots, N)$, prove that
$\sum_{i=1}^{N} \pi_i = E(\nu(s))$, and
$\sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \pi_{ij} = Var(\nu(s)) + E(\nu(s))(E(\nu(s)) - 1)$,
where π_i and π_{ij}s are the first and second order inclusion probabilities and $\nu(s)$ is the effective sample size of a sample s.

   (b) For SRSWR of n draws from N units, find $E(\nu(s))$.

(5 + 5 = 10)

2. (a) In estimating the population ratio $R = \frac{Y}{X}$ through SRSWOR, show that an approximate expression for mean square error of $\hat{R} = \frac{\bar{y}}{\bar{x}}$ is $M_1(\hat{R}) = \left(\frac{N-n}{Nn}\right) \frac{S_y^2 + R^2 S_x^2 - 2R\rho S_x S_y}{X^2}$.

   (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.

(5 + 5 = 10)

3. (a) Define regression estimator for estimating population mean through SRSWOR and state if it is biased or unbiased.

   (b) Derive an approximate expression for its mean squared error, and suggest how to estimate it.

(3 + 7 = 10)

4. (a) Derive the condition under which, for estimating the population total Y, the estimator $\frac{1}{n} \sum_{r=1}^{n} \frac{y_r}{p_r}$ in sampling with probability proportional to x with replacement is better than the estimator $\frac{N}{n} \sum_{r=1}^{n} y_r$ in SRSWR .

   (b) Now assume that the finite population is a random sample from a superpopulation with the properties $Y_i = BX_i + e_i$, $E(e_i|X_i) = 0$, $E(e_i^2|X_i) = \sigma^2$, $E(e_i e_j|X_i, X_j) = 0$. Now derive the condition under which the estimator $\frac{1}{n} \sum_{r=1}^{n} \frac{y_r}{p_r}$ is better than the estimator $\frac{N}{n} \sum_{r=1}^{n} y_r$, as described in (a) above.

(5 + 5 = 10)

5. (a) Prove that the Horvitz and Thompson's estimator is an unbiased estimator for the population total of a variable of interest, provided the first order inclusion probabilities are positive for all the units.

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

(5 + 5 = 10)

6. (a) Describe Hansen and Hurwitz's technique of dealing with non-response errors to estimate the population mean $\bar{Y}$ in SRSWOR of size n out of N units. Find the variance of this estimator.

(b) In the above technique, find optimum sample size for a given value V_0 of the variance and a given cost function with rates defined as, C_0 = cost per unit for 1st attempt of data collection, C_1 = cost of editing and processing data for per unit in response class and C_2 = cost of interviewing and processing data per unit of non-response class.

(5 + 5 = 10)

7. The following figures of expenditures (Rs.) relate to a group of 10 households. Take an SRSWOR of 4 households and give an estimate of the per capita last month's expenses of these 10 households. Obtain the estimated standard error and the coefficient of variation of your estimate.

(10)

Serial no.	HH size	Exp. (Rs.) last month
1	7	3470.35
2	6	2716.80
3	5	1873.75
4	4	1693.20
5	3	1393.55
6	6	2198.74
7	2	3178.35
8	5	2708.75
9	6	1873.60
10	4	2175.80

Random number table

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

Indian Statistical Institute

B. Stat. (Hons) III year, 2014-15

First Semester, Final Examination

Linear Statistical Models

Date: 03·11·14 Duration: 210 minutes Maximum Marks: 100

Note: This paper carries 104 marks. Answer as much as you can.

1. Consider the following model

$\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ where X is a given *nxp* matrix, $\underline{\varepsilon} \sim Nor(\underline{0}, \sigma^2 I_n)$, $\underline{y}' = (y_1, y_2, ..., y_n)$, and $\underline{\beta}' = (\beta_1, \beta_2, ..., \beta_p)$. It is given that Rank($X$) $= r < p < n$. Let G be a generalized inverse of $X'X$.

(a) Form normal equations and estimate $\underline{\beta}$. Let the estimate be denoted by $\underline{\beta}^0$.

(b) Let the error sum of squares be represented by SSE. Show that $SSE/\sigma^2 \sim \chi^2$ distribution with degrees of freedom $(n-r)$.

(c) Define estimable function $\underline{q}'\underline{\beta}$ of parameter vector $\underline{\beta}$.

(d) Show that the BLUE of $\underline{q}'\underline{\beta}$ is $\underline{q}'\underline{\beta}^0$ if $\underline{q}'\underline{\beta}$ is estimable.

(e) Show that $\underline{q}'\underline{\beta}$ is estimable if and only if $\underline{q}'H = \underline{q}'$ where $H = GX'X$.

(f) Derive the unbiased estimate for σ^2 if the model is restricted to $Q\underline{\beta} = \underline{\delta}$, Q is a sxp matrix,

Rank (Q)=s$\leq r$, and $Q\underline{\beta} = \underline{\delta}$ are component-wise estimable.

[6+4+2+12+7+7=38]

2. Consider the model stated in problem 1 along with the notations mentioned. Consider the hypothesis $H_1 : W'\underline{\beta} = \underline{m}$ where W' is a kxp matrix, Rank(W)=k$\leq r$, W and $\underline{m}$ are given.

(a) When is it possible to test H_1?

(b) Show that, under H_1, λ follows an F distribution with degrees of freedom $(k, n-r)$ where $\lambda = (W'\underline{\beta}^0 - \underline{m})'(W'GW)^{-1}(W'\underline{\beta}^0 - \underline{m})/kA,\ \ A = SSE/(n-r).$ [2+7=9]

(P.T.O)

3. Let us consider the 1-way classification model

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, ε_{ij} s are i.i.d normal $(0, \sigma^2)$, i=1,2,3, j=1,2.

(a) Write the ANOVA table for the above.

(b) How do you estimate $\alpha_1 - \alpha_2$? Justify your answer. [7+5=12].

4. Let us consider the model

$y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$, k=1,2,..., n_{ij}, j=1,2,...,b, i=1,2,...,a,

$\varepsilon_{ijk} \sim nor(0, \sigma^2)$ for all $i, j.k$. and ε_{ijk}'s are independent.

(a) Write normal equations for the said model.

(b) Write ANOVA tables for the above.

(c) Write the ANOVA table under the restriction $n_{ij} = 1 \; \forall ij$. [7+10+3=20]

5. Describe the 2-wayclassification model with interactions for balanced data, and write the corresponding ANOVA table. [12]

6. (a) Describe the model for logistic regression.

(b) Describe the procedure for estimation of parameters in the logistic regression model. [5+8=13]

INDIAN STATISTICAL INSTITUTE
B. Stat. (Hons.) III Year 2014-2015

First Semester Examination

Subject : SQC & OR

Date : **07.11.2014** Full Marks : **100** Duration : **3 hrs.**

This paper carries 112 marks. You may answer as much as you can; but the maximum you can score is 100.

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

$$\mathcal{P}: \text{ Minimize } z = 2x_1 + 3x_2 + 5x_3 + 2x_4 + 3x_5$$

$$\text{St} \quad \begin{aligned} x_1 + x_2 + 2x_3 + x_4 + 3x_5 &\geq 4 \\ 2x_1 - 2x_2 + 3x_3 + x_4 + x_5 &\geq 3 \end{aligned}$$

$$x_i \geq 0 \quad i = 1, \cdots, 5$$

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

[15]

2. Two companies A and B sell two brands of baby-food. Both the companies advertise using radio, TV, newspaper and the Web. Depending on the cleverness and intensity of the advertisement campaign, each company can capture a portion of the market from the other. The following matrix summarizes the percentage of the market captured or lost by company A.

		Company B			
		Radio	TV	Newspaper	Web
Company A	Radio	1	2	−2	2
	TV	3	1	2	3
	Newspaper	−1	3	2	1
	Web	−2	2	0	−3

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

[15]

3. In a car-wash facility, information gathered indicates that cars arrive for service according to a Poisson distribution with mean 5 per hour. The time for washing and cleaning each car varies but is found to follow an exponential distribution with mean 10 minutes per car. The facility cannot handle more than one car at a time. (However, the manger wants that there be enough parking space to accommodate all arriving cars).

   a) How many parking spaces should be made available for cars that arrive at the facility?
   b) Determine the number of parking spaces required so that an *arriving* car will be able to park at least 80% of the time.
   c) Find the percentage of time the facility is idle.
   d) Find the expected waiting time from the moment a car arrives until it leaves the facility.

[3+6+3+3=15]

4. Answer the following questions briefly:-

   a) Why should we be interested in obtaining the optimal solution of the primal by solving the dual?
   b) What does the notation $(M/E_3/4):(12/36/SIRO)$ mean in the context of queuing theory?
   c) Define ***Quadratic Programming***? Give an example.
   d) Define a ***Separable Function***? Give an example.

[4+5+3+3=15]

5. Answer the following questions:

   a) If a quality characteristic is centered and normally distributed, then find the expected proportion of Non-Conformance (NC) in terms of the process capability index C_p.
   b) 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 $c = 1$.
   c) Describe the operation of a double sampling plan.
   d) Show that $AOQ = p(1 - AFI)$, where the symbols have their usual meaning.

[7+8+5+5=25]

6. In a study to isolate both gage repeatability and gage reproducibility, three operators use the same gage to measure 20 parts twice each. The data are given on the next page in Table 1. The specification limits are USL= 60, LSL=5.

   a) Estimate the gage repeatability, gage reproducibility, total variability and product variability.
   b) Find the precision-to-tolerance ratio for this gage. Is it adequate?
   c) Do you think training to the operators is called for? Give reasons for your answer.

[12+4+4=20]

Table 1: Data for Question No. 6

Part Number	Operator I Measurements		Operator II Measurements		Operator III Measurements	
	1	2	1	2	1	2
1	21	20	20	20	19	21
2	24	23	24	24	23	24
3	20	21	19	21	20	22
4	27	27	28	26	27	28
5	19	18	19	18	18	21
6	23	21	24	21	23	22
7	22	21	22	24	22	20
8	19	17	18	20	19	18
9	24	23	25	23	24	24
10	25	23	26	25	24	25
11	21	20	20	20	21	20
12	18	19	17	19	18	19
13	23	25	25	25	25	25
14	24	24	23	25	24	25
15	29	30	30	28	31	30
16	26	26	25	26	25	27
17	20	20	19	20	20	20
18	19	21	19	19	21	23
19	25	26	25	24	25	25
20	19	19	18	17	19	17

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

i. 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
a) $\neq 0$. b) ≤ 0. c) ≥ 0. d) $= 0$.

ii. If the j^{th} constraint in the primal is an equality, then the corresponding dual variable is
a) unrestricted in sign.
b) restricted to ≥ 0.
c) restricted to ≤ 0.
d) always 0.

iii. The optimum of an LPP occurs at $X = (1,\ 0,\ 0, 2)$ and $Y = (0,\ 1,\ 0,\ 3)$. Then the optimum also occurs at
a) (2, 0, 3, 0).
b) (1/2 , 1/2, 0, 5/2).
c) (0, 1, 5, 0).
d) none of the above.

iv. If in any simplex iteration the minimum ratio rule fails, then the LPP has
a) non-degenerate BFS.
b) degenerate BFS.
c) unbounded solution.
d) infeasible solution.

v. AOQL means
a) Average outgoing quality level.
b) Average outgoing quality limit.
c) Average outside quality limit.
d) Anticipated optimum quality level.

vi. Consumer's risk of 10% means that
a) The probability that a sampling plan will reject "good" material is 10%.
b) The probability that a sampling plan will accept "poor" material is 10%.
c) The acceptable quality level of the lot is 10%.
d) The unacceptable quality level of the lot is 10%.

vii. 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 ***at least***
a) 89% of all the numbers.
b) 95% of all the numbers.
c) 99.7% of all the numbers.
d) None of the above.

viii. Which of the following statements is true with respect to the optimal solution of an LP problem:
a) Every LP problem has an optimal solution.
b) Optimal solution of an LP always occurs at an extreme point.
c) At optimal solution, all resources are used completely.
d) If an optimal solution exists, there will always be at least one at a corner point.

ix. Who is reputed to have said: *All models are wrong, but some are useful.*
a) Deming
b) Shewhart
c) Taguchi
d) None of the above

x. Two person zero-sum game means that the
a) sum of losses to one player equals the sum of gains to other.
b) sum of losses to one player is not equal to the sum of gains to other.
c) both (a) and (b).
d) None of the above.

xi. Game theory models are classified by the

a) number of players.
b) sum of all pay-offs.
c) number of strategies.
d) all of the above.

xii. Customer behavior in which she moves from one queue to another (in the hope of reducing her waiting time) in a multiple channel situation is
a) balking.
b) reneging.
c) jockeying.
d) alternating.

xiii. Which of the following is **not** a key operating characteristic for a queuing system?
a) utilization factor.
b) per cent idle time.
c) average time spent waiting in the system.
d) none of the above.

xiv. When is the World Quality Day celebrated?
a) First Monday of January.
b) Fourth Tuesday of April.
c) Third Sunday of December.
d) Second Thursday of November.

[14 × 1/2 = 7]

Selected Values of the Standard Gaussian CDF

$\Phi(z) = \Pr(Z \leq z)$ when $Z \sim \mathcal{N}(0,1)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

NOTE: $\Phi(-z) = 1 - \Phi(z)$.

Appendix VI Factors for Constructing Variables Control Charts

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

For $n > 25$

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

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

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

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : (2014-2015)

B.Stat. 3rd Year

PARAMETRIC INFERENCE

Date: 11 November, 2014 Maximum Marks: 100 Duration: 3 Hours

Answer all questions.

1. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with mean θ. Find the UMP level α test for testing $H_0 : \theta \leq \theta_0$ against the alternative $H_1 : \theta > \theta_0$ where $\theta_0 > 0$ is some specified value. Give a direct proof of your result without using the general theorem on UMP test for MLR families of distributions. [12]

2. Let the observations $X_1, \ldots, X_n$ be i.i.d $N(\theta, 1)$. Consider the problem of testing $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$.

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

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

Show that the power function of this test is an increasing function of $|\theta|$. [7+5+7=19]

3. (a) Consider the problem of testing $H_0 : X$ has densiy f_0 against $H_1 :$ X has density 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.

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

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

Show that the power function of this test is nondecreasing in θ. [7+9=16]

4. Derive a condition for equality in Cramer-Rao Inequality. Show that a minimum variance bound (MVB) estimator of a parametric function $g(\theta)$ must be a sufficient statistic. [6]

5. Let $X_1, \ldots, X_n$ be i.i.d $N(\theta, 1)$.

(a) For some specified $u \in R$, find the UMVUE of $P(X_1 < u)$.

(b) Suppose it is known that $-1 \leq \theta \leq 1$. Find the UMVUE of θ. Show that there exists an estimator of θ which is better than the UMVUE in the sense of having less MSE. [7+(4+9)=20]

6. Let $X_1, X_2, \ldots$ be i.i.d. random variables with a common density involving an unknown real parameter θ. Suppose there exists an unbiased estimator T of θ based on a sample of size 2, i.e., $E_\theta T(X_1, X_2) = \theta$ for all θ. Assume that $Var_\theta[T(X_1, X_2)] < \infty$ for all θ.

Find an estimator V_n of θ based on $X_1, \ldots, X_n$ such that V_n is consistent for θ. [10]

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

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

where $\alpha > 0, \beta > 0$ are known. Find the Bayes estimate of σ^2 for the loss function $L(\sigma^2, a) = (a - \sigma^2)^2/\sigma^4$. [It is given that the mean of an inverse-gamma (α, β) distribution is $\beta/(\alpha - 1)$.] [9]

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

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

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

B. Stat III Year
Final Semester Examination
Elective Course on Psychology

Full Marks: 20
Time: 1 hour

---

1. **Fill in the blanks: Any five (5×1)**

   i. Perceptual organization is the process of establishing relations among potentially ________elements of the object in perception.

   ii. Elements which are _______ in structure or have common characteristics will be grouped together.

   iii. Forgetting is the ______, permanent or temporary, of the ability to _____ or recognize something learned earlier.

   iv. _________ occurs due to some changes of biological processes, which occur, with time. This results in ___________, alzheimer's disease etc.

   v. Psychophysics is the study about relation between change in _______ and psychological scale.

2. **Write short notes: Any three (3×5)**

   (a) Validity

   (b) Decay theory

   (c) Classical conditioning

   (d) Causes of forgetting

   (e) Learning curve

   (f) Laws of learning

Indian Statistical Institute

First-semester examination: 2014-15

Course name: B.Stat. III

Subject name: Sociology

Date: 14.11.2014. *Maximum marks: 60* *Duration: 2.5 hour*

Note, if any:

Answer two of the following four questions (1 through 5) in not more than 200 words.
Each answer carries a maximum of 10 marks.

1. What is qualitative research and what are the methods in qualitative research?
2. Briefly describe concept of social exclusion in reference to Indian society.
3. Describe Jajmani system in rural India. How is it related to social stratification?
4. Characteristics of industry and its impact on society. Do they vary with the type of industry?
5. Outline Max Weber's rationalization of protestant ethic and spirit of capitalism.

Provide short definitions of any eight of the following:
8 questions of 5 marks each

6. Short notes

a) Alienation
b) Criminology
c) Durkhiem's classification of suicide
d) Ethnic group
e) Endogamy & exogamy
f) Environmental refugee
g) Exemplary and ethical prophecy
h) Features of urban society
i) False consciousness
j) Impact of development on tribal women
k) Polygyny and polyandry
l) Poverty line in India
m) Reservation policy in India
n) Taboo

Indian Statistical Institute

First Semester Examination: 2014

Elective: Introduction to Anthropology

Date : 14/11/2014 Full Marks: 50 Time : Two and a half hours

Answer All the Questions

[Figures in parentheses are given marks]

1. Choose the right answer for (a) thru (o) from among the given options: (15 x 1 =15)

   (a) New world monkeys belong to order Prosimi: (i) True (ii) False

   (b) Relationship between learned human behavior and human biology is: (i) a product of evolution (ii) a programmed phenomenon (iii) doesn't exist at all

   (c) An evolutionary theory explained how humans and other species came to exist through the operation of natural selection. The theory in question is propounded by: (i) G. J. Mendel (ii) C. R. Darwin (iii) J.B.S. Haldane

   (d) Virtually all biologists and anthropologists are in agreement with the correctness of basic elements of Darwin's theory: (i) True (ii) False

   (e) Humans have closest relationship with: (i) Rhesus monkeys and Howler monkeys (ii) Orangutans and Gibbons (iii) Chimpanzee and Gorilla

   (f) In higher primates, strong bond between mother and offspring becomes essential owing to the necessity of: (i) sensory inputs (ii) teaching and learning (iii) physiological prerequisites to survival

   (g) The earliest fossil remains for human ancestors are between: (i) 6-7 million years old (ii) 7-8 million years old (iii) 8-9 million years old

   (h) Between 4.2 million and 1 million years ago, a diverse group of creatures lived in south and east Africa having bipedalism and were relatively small-brained. They were: (i) Dryopithecids (ii) pre-australopithcines (iii) australopithecines

   (i) An important distinguishing feature of *Homo habilis* from australopithecines is: (i) Somewhat larger brain and use of simple tools (ii) Being herbivores (iii) Epicanthic fold

   (j) *Homo erectus* fossils were found from: (i) Europe, Africa and Asia (ii) Australia (iii) Antarctica

   (k) There were several different forms of archaic *Homo sapiens*, including Neanderthals between: (i) 10,000 and 15,000 years ago (ii) 30,000 and 35,000 years ago (iii) 2 million and 3 million years ago

(l) Currently considered the most probable theory concerning transition from *Homo erectus* to *Homo sapiens* is: (i) the hybrid theory (ii) the multiregional theory (iii) the replacement theory

(m) Anatomically modern humans are scientifically called *Homo sapiens*: (i) True (ii) False

(n) The concept of "Race" currently doesn't have any biological validity but does have enormous historical and socio-political importance: (i) True (ii) False

(o) Haplorhini is the suborder of the order primate to which: (i) lemurs and lorises belong (ii) tarsiers, monkeys, apes and humans belong (iii) only the apes belong

2. Delineate briefly the major bio-behavioral trends in human evolution (10)

3. (a) Why is fossil evidence important in understanding human ancestry? (2)

   (b) Was Homo erectus the first human ancestor to walk upright? (3)

   (c) Compare prosimians with anthropoids in respect of bio-behavioral adaptive tendencies. (5)

4. (a) What is meant by clinal distribution of biological traits? (2)

   (b) Provide suitable examples of clinal distributions in humans. (2) (3)

   (c) It is observed that "there is a clear relationship between melanin, ultraviolet light and skin cancer". Illustrate the above statement. (5)

5. Write short notes on any five (5) of the following: (2 x 5)

(a) Adaptation

(b) Senescence

(c) Physical growth

(d) Demography

(e) Culture

(f) Homeostasis

(g) Family

(h) Marriage

B. Stat III Year
Final Semester Examination
Elective Course on Psychology

Full Marks: 20
Time: 1 hour

---

1. Fill in the blanks: Any five (5×1)

i. Perceptual organization is the process of establishing relations among potentially ________elements of the object in perception.

ii. Elements which are _______ in structure or have common characteristics will be grouped together.

iii. Forgetting is the ______, permanent or temporary, of the ability to or recognize something learned earlier.

iv. _________ occurs due to some changes of biological processes, which occur, with time. This results in ___________, alzheimer's disease etc.

v. Psychophysics is the study about relation between change in and psychological scale.

2. Write short notes: Any three (3×5)

(a) Validity
(b) Decay theory
(c) Classical conditioning
(d) Causes of forgetting
(e) Learning curve
(f) Laws of learning

Indian Statistical Institute
Statistics Comprehensive
B-III, Midsem

Date: Feb 23, 2015

Duration: 2hrs.

Each question carries 8 marks. Attempt all questions. The maximum you can score is 40. Justify all your steps.

1. X, Y, Z, W are independent Bernoulli random variables with parameters p, q, r, s respectively. Define U as

$$\begin{aligned} X + W &\quad \text{mod } 2 \text{ if } Z = 0 \\ Y + W &\quad \text{mod } 2 \text{ if } Z = 1 \end{aligned}$$

   Find $Prob(U = X)$ and $Prob(U = Z)$.

2. Draw a sample of size 1 from 3 persons with equal probability using a possibly biased coin.

3. n samples are drawn from $U[a, 2a + 1]$. Suggest a "good" estimate for a. Justify your answer.

4. A random variable X can take only integer values ≥ 3. Construct (if possible) a distribution for X such that for any two integers $3 \leq m < n$

$$P(X > n | X > m) = P(X > n - m).$$

   Justify your answer.

5. A random variable takes three values 1, 2, 3 with unknown probabilities p, q, r, respectively where $q = 2p$, $p + q + r = 1$ and $p, q, r > 0$. Let V be the vector space of all complete statistics based on a sample of size 1. Find dimension of V. Also find (if possible) a complete statistic that is not suffcient for (p, q, r).

6. Consider $X_1, X_2, ...$ iid $N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are both unknown. You are required to construct a 95% confidence interval (CI) for μ, that is guaranteed to be ≤ 1 in length (irrespective of σ^2). Does the "usual" CI using t-statistic meet this requirement? Justify your answer. Here is another way you may proceed: First compute the "usual" CI using t-statistic based on only $X_1, ..., X_5$:

$$(\bar{X}_5 \pm t_* \frac{S_5}{\sqrt{5}}).$$

   Here t_* is the appropriate cut-off value from $t_{(4)}$-distribution. If the length of this CI is ≤ 1, then report it, and stop. Otherwise compute the smallest integer $n > 5$ such that

$$\frac{2t_* S_5}{\sqrt{n}} \leq 1,$$

   and collect extra data $X_6, ..., X_n$. Suggest a 95% CI for μ based on $X_1, ..., X_n$ that meets our requirement.

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination (Second Semester) : 2014-15

Course : B.Stat.-III Year

Subject : Design of Experiments

Date: 24/02/2015 Maximum Marks: 70 Duration: One and half hour

NOTE: (i) This paper carries 77 marks. Question number 5 is compulsory. Answer as much as you can from the remaining 56 marks but the maximum you can score is 49. The marks are indicated in [ ] on the right margin.

(ii) The symbols and notations have the usual meaning as introduced in your class.

1. Define/describe the followings with suitable example wherever feasible:
Experiment, Factor & levels, Random factor. [2×3 = 6]

2. Two different pesticides, A and B say, are compared for their effectiveness and both are applied in two different forms: spray (A_1 and B_1) and powder (A_2 and B_2). A control (that is, no pesticide), C say, is included in the experiment to establish any effectiveness of the pesticides at all. Each of the five treatments is applied to r uniformly infested plots of land using a *completely randomized design*. In order to detect existing differences between treatments, we consider the following three sets of contrasts Γ_1 to Γ_4.

(a) Are contrasts in each set orthogonal? Explain.
(b) What treatment comparisons are suggested by each set of contrasts?
(c) Write down an appropriate model for the design.

Set	Contrasts	Treatment (code)*				
		$C(1)$	$A_1(2)$	$A_2(3)$	$B_1(4)$	$B_2(5)$
1	Γ_1	4	-1	-1	-1	-1
	Γ_2	0	1	1	-1	-1
	Γ_3	0	1	-1	0	0
	Γ_4	0	0	0	1	-1
2	Γ_1	4	-1	-1	-1	-1
	Γ_2	0	1	-1	1	-1
	Γ_3	0	1	0	-1	0
	Γ_4	0	0	1	0	-1
3	Γ_1	-2	1	0	1	0
	Γ_2	0	-1	-1	1	1
	Γ_3	0	-1	0	1	0
	Γ_4	2	0	-1	0	-1

*: These codes are used to identify effects, e.g., effect of C is denoted by τ_1.

(d) For the model suggested by you, is τ_i estimable? Is $\mu + (\tau_1 + \tau_3 + \tau_5)/3$ estimable? Explain. (2+2×3+2+2) = [12]

P.T.O.

3. A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist decides to use a design, with the bolts of cloth considered as a source of variation. She selects all the five available bolts and applies all four chemicals in random order to each bolt. The resulting tensile strengths follow.

Chemical	Bolt				
	1	2	3	4	5
1	-2	-3	-3	0	-4
2	1	-4	-4	0	-2
3	4	3	7	2	3
4	2	1	4	4	2

Perform an analysis of variance of the data and draw appropriate conclusions. Which chemical(s) will you recommend for use to improve strength and why?

($F_{0.05}(3,12)=3.49$, $F_{0.05}(4,12)=3.26$, $F_{0.01}(3,12)=5.95$, $F_{0.01}(4,12)=5.41$)

[7+6 = 13]

4. Consider a block design in v treatments. Let $\hat{\boldsymbol{\tau}}$ be a solution to the reduced normal equations $\boldsymbol{C\tau} = \boldsymbol{Q}$.

(i) For $\boldsymbol{l}'\hat{\boldsymbol{\tau}}$ to be an unbiased estimator of $\boldsymbol{l}'\boldsymbol{\tau}$, show that we must have $\boldsymbol{l}'C^{-}C = \boldsymbol{l}'$, where C^{-} is a g-inverse of the C-matrix. Consequently or otherwise, prove that only treatment contrasts are estimable.

(ii) Prove that $\boldsymbol{l}'C^{-}C = \boldsymbol{l}'$ if and only if $\boldsymbol{l}'$ belongs to the row space of the C-matrix.

(iii) If the block design is **connected** then show that the diagonal elements of C are positive.

(iv) For the C-matrix of a connected design, show that $C_1 = C + a\,J_v$, where $a \neq 0$ is a real number and J_v is the $v \times v$ matrix of all 1's, admits an inverse C_1^{-1}, and also show that C_1^{-1} is a generalized inverse of the C-matrix of the design.

(v) With C_1^{-1}, as in (iv), a generalized inverse for the C-matrix of a connected design, let $\hat{\boldsymbol{\tau}}$ be a solution to the above reduced normal equations. Find $E(\mathbf{1}'\hat{\boldsymbol{\tau}})$, where $\mathbf{1}'$ is the $1 \times v$ unit vector.

[(3+1)+3+5+8+5 = 25]

5. Assignments.

[21]

Critical values $r_{0.05}(p, df)$ for Duncan's multiple range tests

df *p*->	2	3	4	5	6	7	8	9	10	11	12	13	14
1	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969	17.969
2	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085	6.085
3	4.501	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516	4.516
4	3.926	4.013	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033	4.033
5	3.635	3.749	3.796	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814	3.814
6	3.460	3.586	3.649	3.680	3.694	3.697	3.697	3.697	3.697	3.697	3.697	3.697	3.697
7	3.344	3.477	3.548	3.588	3.611	3.622	3.625	3.625	3.625	3.625	3.625	3.625	3.625
8	3.261	3.398	3.475	3.521	3.549	3.566	3.575	3.579	3.579	3.579	3.579	3.579	3.579
9	3.199	3.339	3.420	3.470	3.502	3.523	3.536	3.544	3.547	3.547	3.547	3.547	3.547
10	3.151	3.293	3.376	3.430	3.465	3.489	3.505	3.516	3.522	3.525	3.525	3.525	3.525	3.525
11	3.113	3.256	3.341	3.397	3.435	3.462	3.480	3.493	3.501	3.506	3.509	3.510	3.510	3.510
12	3.081	3.225	3.312	3.370	3.410	3.439	3.459	3.474	3.484	3.491	3.495	3.498	3.498	3.498
13	3.055	3.200	3.288	3.348	3.389	3.419	3.441	3.458	3.470	3.478	3.484	3.488	3.490	3.490
14	3.033	3.178	3.268	3.328	3.371	3.403	3.426	3.444	3.457	3.467	3.474	3.479	3.482	3.484
15	3.014	3.160	3.250	3.312	3.356	3.389	3.413	3.432	3.446	3.457	3.465	3.471	3.476	3.478
16	2.998	3.144	3.235	3.297	3.343	3.376	3.402	3.422	3.437	3.449	3.458	3.465	3.470	3.473
17	2.984	3.130	3.222	3.285	3.331	3.365	3.392	3.412	3.429	3.441	3.451	3.459	3.465	3.469
18	2.971	3.117	3.210	3.274	3.320	3.356	3.383	3.404	3.421	3.435	3.445	3.454	3.460	3.465
19	2.960	3.106	3.199	3.264	3.311	3.347	3.375	3.397	3.415	3.429	3.440	3.449	3.456	3.462
20	2.950	3.097	3.190	3.255	3.303	3.339	3.368	3.390	3.409	3.423	3.435	3.445	3.452	3.459
21	2.941	3.088	3.181	3.247	3.295	3.332	3.361	3.385	3.403	3.418	3.431	3.441	3.449	3.456
22	2.933	3.080	3.173	3.239	3.288	3.326	3.355	3.379	3.398	3.414	3.427	3.437	3.446	3.453
23	2.926	3.072	3.166	3.233	3.282	3.320	3.350	3.374	3.394	3.410	3.423	3.434	3.443	3.451
24	2.919	3.066	3.160	3.226	3.276	3.315	3.345	3.370	3.390	3.406	3.420	3.431	3.441	3.449
25	2.913	3.059	3.154	3.221	3.271	3.310	3.341	3.366	3.386	3.403	3.417	3.429	3.439	3.447
26	2.907	3.054	3.149	3.216	3.266	3.305	3.336	3.362	3.382	3.400	3.414	3.426	3.436	3.445
27	2.902	3.049	3.144	3.211	3.262	3.301	3.332	3.358	3.379	3.397	3.412	3.424	3.434	3.443
28	2.897	3.044	3.139	3.206	3.257	3.297	3.329	3.355	3.376	3.394	3.409	3.422	3.433	3.442
29	2.892	3.039	3.135	3.202	3.253	3.293	3.326	3.352	3.373	3.392	3.407	3.420	3.431	3.440
30	2.888	3.035	3.131	3.199	3.250	3.290	3.322	3.349	3.371	3.389	3.405	3.418	3.429	3.439

PERCENTAGE POINTS OF THE T DISTRIBUTION

Tail Probabilities

	One Tail	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	
	Two Tails	0.20	0.10	0.05	0.02	0.01	0.002	0.001	
D	1	3.078	6.314	12.71	31.82	63.66	318.3	637	1
E	2	1.886	2.920	4.303	6.965	9.925	22.330	31.6	2
G	3	1.638	2.353	3.182	4.541	5.841	10.210	12.92	3
R	4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	4
E	5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	5
E	6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	6
S	7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7
	8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	8
O	9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	9
F	10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	10
	11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	11
F	12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	12
R	13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	13
E	14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	14
E	15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	15
D	16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	16
O	17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	17
M	18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	18
	19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	19
	20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	20
	21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	21
	22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	22
	23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	23
	24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	24
	25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	25
	26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	26
	27	1.314	1.703	2.052	2.473	2.771	3.421	3.690	27
	28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	28
	29	1.311	1.699	2.045	2.462	2.756	3.396	3.659	29
	30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	30
	32	1.309	1.694	2.037	2.449	2.738	3.365	3.622	32
	34	1.307	1.691	2.032	2.441	2.728	3.348	3.601	34
	36	1.306	1.688	2.028	2.434	2.719	3.333	3.582	36
	38	1.304	1.686	2.024	2.429	2.712	3.319	3.566	38
	40	1.303	1.684	2.021	2.423	2.704	3.307	3.551	40
	42	1.302	1.682	2.018	2.418	2.698	3.296	3.538	42
	44	1.301	1.680	2.015	2.414	2.692	3.286	3.526	44
	46	1.300	1.679	2.013	2.410	2.687	3.277	3.515	46
	48	1.299	1.677	2.011	2.407	2.682	3.269	3.505	48
	50	1.299	1.676	2.009	2.403	2.678	3.261	3.496	50
	55	1.297	1.673	2.004	2.396	2.668	3.245	3.476	55
	60	1.296	1.671	2.000	2.390	2.660	3.232	3.460	60
	65	1.295	1.669	1.997	2.385	2.654	3.220	3.447	65
	70	1.294	1.667	1.994	2.381	2.648	3.211	3.435	70
	80	1.292	1.664	1.990	2.374	2.639	3.195	3.416	80
	100	1.290	1.660	1.984	2.364	2.626	3.174	3.390	100
	150	1.287	1.655	1.976	2.351	2.609	3.145	3.357	150
	200	1.286	1.653	1.972	2.345	2.601	3.131	3.340	200
	Two Tails	0.20	0.10	0.05	0.02	0.01	0.002	0.001	
	One Tail	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	

Tail Probabilities

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : (2014-2015)

B.Stat. 3rd Year

NONPARAMETRIC AND SEQUENTIAL METHODS

Date: 25 February, 2015 Max. Marks: 100 Duration: $2\frac{1}{2}$ Hours

1. A six-sided die with numbers from 1 to 6 is thrown n times. How do you test for the hypothesis that the die is fair? [7]

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

(a) What is the Glivenko-Cantelli Theorem on the sample distribution function and how do you use it to construct a test for the hypothesis $H_0 : F = F_0$ (for various possible alternatives) where F_0 is a specified distribution function?

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

(c) Show that the Kolmogorov-Smirnov statistics D_n^+ and D_n^- are identically distributed under H_0. [10+9+8=27]

3. Let $X_1, \ldots, X_n$ be a random sample from a population with a continuous distribution having unknown (unique) median θ. Consider the problem of testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. Find an asymptotic level α $(0 < \alpha < 1)$ test for this problem and show that this test is consistent for any alternative $\theta > 0$. [6+9=15]

4. Let $X_1, \ldots, X_n$ be a random sample from a population with a continuous distribution that is symmetric about an unknown median θ. Describe the Wilcoxon signed rank test for testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. Find the variance of the test statistic T_+ under H_0. [5+14=19]

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

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

(b) Suppose that for the testing problem of part (a), we reject H_0 for large values of a rank statistic of the form

$$S = \sum_{i=1}^{n} a(R_i)$$

where $a(1) \leq a(2) \leq \cdots \leq a(m+n)$ are known values (not all the same) and R_i is the rank of Y_i among all $(m+n)$ observations.

Consider a level α test based on S $(0 < \alpha < 1)$. Show that this test is unbiased. Also show that this test is of level α for testing the null hypothesis $H : \theta \leq 0$ against $H_1 : \theta > 0$.

(c) Construct an upper confidence bound for θ with confidence coefficient $(1-\alpha)$ $(0 < \alpha < 1)$ based on the differences $Y_j - X_i$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. [10+9+13=32]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2014-15

Course Name : B. STAT. III YEAR

Subject Name : Differential Equations

Date: 26·02·2015 **Maximum Marks: 40,** **Duration: $2\frac{1}{2}$ hrs.**

Any result that you use should be stated clearly.

1. Let $f(x,y)$ be a continuous function and satisfy a Lipschitz condition in R i.e.
$$|f(x,y_1) - f(x,y_2)| \leq K|y_1 - y_2|,\ K > 0$$
for all $(x,y_1),(x,y_2) \in R$. If a solution of the initial value problem
$$\frac{dy}{dx} = f(x,y),\ y(x_0) = y_0$$
exist, then its unique. 6

2. The President and the Prime Minister order coffee and receive cups of equal temperature at the same time. The President adds a small amount of cool cream immediately, but does not drink his coffee until 10 minutes later. The Prime Minister waits 10 minutes, and then adds the same amount of cool cream and begins to drink. Who drinks the hotter coffee? Solve it using differential equations. 8

3. Find the orthogonal trajectories of the family of hypocycloids $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$, where a is a variable parameter. 4

4. Solve the simultaneous equation $\frac{d^2x}{dt^2} - 3x - 4y = 0, \frac{d^2y}{dt^2} + y + x = 0.$ 4

5. If $y_1 = e^x$ is a solution of $x\frac{d^2y}{dx^2} - (2x+1)\frac{dy}{dx} + (x+1)y = 0$, find the general solution. 4

6. Find the general solution of $x^2\frac{d^2y}{dx^2} - x(x+2)\frac{dy}{dx} + (x+2)y = x^3$, given that $y = x, y = xe^x$ are two linearly independent solutions of the corresponding homogeneous equation. 5

7. Solve the equation $\frac{\partial z}{\partial x} + 3\frac{\partial z}{\partial y} = 5z + \tan(y - 3x).$ 4

8. Using the method of variational calculus, deduce Euler's differential equation for extremization (stationarity) of the integral
$$I = \int_0^4 [xy' - (y')^2]dx,$$
for fixed values of $y(0), y(4)$ and find out the 'stationary' function $y(x)$ with boundary conditions $y(0) = 0,\ y(4) = 3.$ 4+2

9. Solve and find the singular solution of the differential equation
$$\sin\left(x\frac{dy}{dx}\right)\cos y = \cos\left(x\frac{dy}{dx}\right)\sin y + \frac{dy}{dx}.$$ 5

Date: 26.2.2015
Time: 2 hours

Statistical Methods in Genetics
B-Stat (3rd Year)
Mid Semester Examination 2014-15

This paper carries 35 marks. Answer all questions.

1. If it is reported that the heterozygosity at an autosomal locus is 0.825, what can be said about the number of alleles at this locus? [5]

2. Consider a disorder controlled by two unlinked autosomal biallelic loci with alleles (A_1,a_1) and (A_2,a_2), respectively. Suppose an individual is affected if and only if he/she has genotype $A_1A_1a_2a_2$.

(a) If p_1 is the conditional probability of an offspring being affected given that exactly one of the parents is affected and p_2 is the conditional probability that an offspring is affected given both parents are unaffected, show that $p_2=p_1^2$.

(b) In a study on trios (two parents and an offspring), 30 families are selected with exactly one affected parent and 50 families are selected with both parents unaffected. In the first set of families, 6 offspring are affected, while in the second set of families, 2 offspring are affected. Obtain the maximum likelihood estimate of the prevalence of the disease. [7 + 7]

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 proportional to the initial genotype frequencies, explain whether the allele frequencies at the locus reach non-trivial equilibrium values. [6]

4. Consider data on *A-B-O* blood groups for a random sample of individuals chosen from an inbred population. Describe an EM algorithm to obtain the maximum likelihood estimators. of the frequencies of the alleles *A*, *B*, *O* and the inbreeding coefficient. [10]

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination

B. Stat. - III Year, 2014-2015 (Semester - VI)

Design and Analysis of Algorithms

Date: 27.02.2015 Maximum Marks: 60 Duration: 2 hours 30 minutes

---

Note: Answer as much as you can, but the maximum you can score is 60.

This is a cheat sheet based examination. You can write whatever you want on an A4 sized sheet of paper and bring to the exam hall. Write your name and roll number on the sheet. Cheat sheets can not be shared.

(Q1) We say that a function $f(n)$ is polylogarithmically bounded if $f(n) = O(\log^k n)$ for some constant k. Now, prove that any positive polynomial function grows faster than any polylogarithmic function. A positive polynomial function is of the form n^c for any constant $c > 0$. [8]

(Q2) Let X and Y be two arrays of n distinct integers each. X and Y are sorted in an increasing order. Now, consider $Z = X \cup Y$; you can assume that Z has $2n$ distinct elements. Find the median of Z efficiently. [8]

[Hints: Doing it in linear time is too trivial!]

(Q3) Let $\mathcal{A}$ be an array of n integers. Each $a_i \in \mathcal{A}$ lies in the range $[0, n^3 - 1]$. Design an efficient algorithm to sort $\mathcal{A}$. [10]

(Q4) Given a sequence of numbers $A = \{a_1, a_2, \ldots, a_n\}$, design an algorithm to find the length of the largest subset such that for every $i < j$, $a_i < a_j$. Find the complexity of the algorithm. An example of such a largest subset follows: $A = \{11, 17, 5, 8, 6, 4, 7, 12, 3\}$. The length of the largest subset is 4 and the set is 5, 6, 7, 12. [10]

[Hints: Can dynamic programming be used?]

(Q5) In the *SELECTION* (median finding) algorithm studied in class, we divided the input elements into groups of 5; and found out by our analysis that the algorithm runs in linear time. Deduce what happens to the time complexity of the *SELECTION* algorithm if the input was divided into groups of 3. [10]

(Q6) Given an array $\mathcal{X}$ containing n *unique* real numbers, design and analyze an efficient algorithm that finds out a number in $\mathcal{X}$ that is neither minimum nor maximum. [4]

(Q7) Given n elements, the second maximum can be found trivially with $2n - 3$ comparisons in the worst case: $n - 1$ comparisons for the first maximum and $n - 2$ comparisons for the second maximum. Try to find out an efficient method (that takes less than $2n - 3$ comparisons) to find the second maximum. Note that, here we are not interested in the asymptotic complexity but the exact count in the worst case. [10]

(Q8) Show that there is no comaprison-based sort whose running time is $O(n)$ for at least a constant fraction of the $n!$ inputs of length n. You can assume that the input is distinct. [10]

INDIAN STATISTICAL INSTITUTE

End Semestral Examination

B. Stat. - III Year, 2014-2015 (Semester - VI)

Design and Analysis of Algorithms

Date: 27.04.2015 Maximum Marks: 100 Duration: 3 hours

Note: This is a 4-page question paper with 12 questions, each of 10 marks. Answer as much as you can, but the maximum you can score is 100. Answer parts of a question together.

If you are asked to design an algorithm of a specific time complexity and your time complexity is more (i.e., less efficient), you will get partial credit.

This is a cheat sheet based examination. You can write whatever you want on an A4 sized sheet of paper and bring to the exam hall. Write your name and roll number on the sheet. Cheat sheets can not be shared.

(Q1) We want to use a binary counter to count from 0 to n. You can think of a binary counter as an array X storing values 0 and 1; and any number is represented as a string of 0s and 1s. As an example, 11 is represented as 1011; to count 12 using this counter, we have to change the configuration of the binary counter from 1011 to 1100. So in this example, we have to change 3 bit positions.

Now, the size or number of bits k of the binary counter to count numbers upto n is $\lfloor \log n \rfloor + 1$. To increment the counter by 1, we can use the following routine (the routine is given for your help):

Method *Increment Counter*
Input: A counter $X[0, 1, \ldots, k-1]$ of size $k = (\lfloor \log n \rfloor + 1)$ to count from 0 to n with any number x_i between 0 to $n-1$.
Output: The number $x_i + 1$ stored in the counter X.

```
j ← 0;
while ( X[j] = 1 )
    X[j] ← 0;
    j ← j + 1;
endwhile
X[j] ← 1;
```

So, in the worst case we have to change $O(k) = O(\log n)$ bits for incrementing by one. As, we have n numbers in all, the time complexity is $O(nk) = O(n\log n)$.

Do you think this is a tight time complexity analysis of the method? If yes, justify your answer; else, try improving the time complexity by doing a tighter analysis of the same method. [10]

[Hints: Write the binary representation of numbers between 0 to 7 – they are 000, 001, 010, 011, 100, 101, 110, 111. Look at the LSB (Least Significant Bit). How many bit changes are there? Then, look at the MSB (Most Significant Bit). How many bit changes are there? Can you now generalize?]

(Q2) (a) Design and analyse an $O(n \log k)$ time algorithm to generate a sorted list of n numbers from k sorted lists, each of size n/k.

(b) Consider a min-heap $\mathcal{H}$ of n distinct elements. Can you say anything definite about where the maximum element will lie in $\mathcal{H}$?

[8+2=10]

(Q3) You are given a time series data $S = \{s_1, \ldots, s_i, \ldots, s_n\}$, where each element is unique. The *span* of any s_i, $i > 1$, is an index $j < i$, such that $s_j > s_i$ and all s_k's between s_j and s_i are less than s_i. If no such j exists for an i, then the span of s_i is set to 0. Design and analyse an $O(n)$ time algorithm to find the span of each $s_i \in S$, $i > 1$. [10]

(Q4) Let A be an unsorted array on n integers. We need to search for a value a in A. Consider the following randomized strategy: choose a random index $i \in [1, n]$ and check if $A[i] = a$. If $A[i] = a$, then we terminate; otherwise, we continue the search by choosing a new random index into A. We continue choosing random indices into A until we find an index j such that $A[j] = a$, or until we have checked every element of A. Notice that, we choose from the whole set of indices every time, so that we may examine a given element more than once.

(a) Suppose, there is only one index i such that $A[i] = a$. What is the expected number of indices that needs to be tried before a is found and the algorithm terminates?

(b) Suppose, there are $k \geq 1$ indices i such that $A[i] = a$. What is the expected number of indices into A that needs to be tried before the algorithm terminates? Your answer should be a function of n and k.

(c) Suppose, there is no index i for which $A[i] = a$. What is the expected number of indices into A that must be tried before all elements of A have been checked and the algorithm terminates?

$[3 + 3 + 4 = 10]$

[3 + 3 + 4 = 10]

(Q5) Let $G = (V, E)$ be a graph with $V = \{v_1, v_2, \ldots, v_n\}$. (v_i, v_j) is an edge in G if and only if the numbers i and j differ by exactly 1, i.e., the edges in G are $(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n)$. Each node $v_i \in V$ has a positive integer weight w_i. A subset of the nodes in V is called an independent set if no two of them are joined by an edge. Design and analyse an algorithm that finds an independent set of maximum weight. Prove its correctness. [10]

[Hints: We are not trying to find the maximum cardinality independent set but the one with the maximum weight. Greedy won't help! Can you locate overlapping subproblems?]

(Q6) Let S be a set of n disjoint line segments (i.e. non-intersecting line segments) whose left endpoints lie on the line $x = 0$ and the right endpoints lie on the line $x = 1$. These line segments partition the open vertical strip between $x = 0$ and $x = 1$ into $n + 1$ regions. Your problem is to report the region in which a given query point will lie.

(a) Design and analyse an efficient algorithm that creates a data structure to store the $n + 1$ regions of the partition of the open vertical strip between $x = 0$ and $x = 1$.

(b) Given a query point q, design and analyse an efficient algorithm that finds the region in which q lies.

[6+4=10]

[Hints: Draw the diagram and proceed.]

(Q7) Let S_1 be a set of disjoint horizontal line segments and S_2 be a set of disjoint vertical line segments. $|S_1 \cup S_2| = n$. Design a plane sweep algorithm that reports all the intersections in $S_1 \cup S_2$ in $O(n \log n + k)$ time, where k is the number of such intersections. [10]

(Q8) Let P be a simple polygon of n vertices. We model a surveillance camera as a point that can see $360°$ around it and can see upto infinity. Cameras can only be placed in the interior of P, and not on the boundary or the vertices of P. We say that P is *guarded* when each point inside P can be seen by a camera.

(a) Prove that $n - 2$ cameras are sufficient to *guard* P.

(b) Design and analyse an algorithm to find the locations inside P where the cameras can be placed such that P is guarded. Prove your result.

[3+7=10]

(Q9) Let $G = (V, E)$ be a strongly connected directed graph with positive edge weights, and $v_0 \in V$ be a particular node.

(a) Design and analyse an efficient algorithm to find the shortest path between all pairs of vertices in V that go through v_0.

(b) Prove the correctness of your algorithm.

[7+3=10]

(Q10) Given a set P of n points in the plane and a positive integer k, find a *k-clustering* of P. A *k-clustering* of P is a partition of P into k non-empty groups such that the *spacing* between the clusters is maximized. The *spacing* between any two clusters is the minimum distance between any pair of points in different clusters.

(a) Design and analyse an efficient algorithm to find a *k-clustering* of P.

(b) Let $C_1, C_2, \ldots, C_k$ be the k clusters that your algorithm returns. Prove that $C_1, C_2, \ldots, C_k$ forms a k-clustering of P.

[5+5=10]

(Q11) Let $G = \{V, E\}$ be a directed acyclic graph with weight/length $w(u, v)$ on edge (u, v). Find the average path length from a source vertex s to a destination vertex t. The average path length is defined as the total length of all paths from s to t divided by the total number of distinct paths. [10]

(Q12) Let P be a set of n points in the plane. We form a simple, undirected, weighted graph $G = (V, E)$ on P as follows. Each point $p_i \in P$ is a vertex $v_i \in V$. The edge set consists of all the $\binom{n}{2}$ edges (v_i, v_j), $i \neq j$, possible between the vertices of V. The weight of an edge (v_i, v_j) is the Euclidean distance between p_i and p_j.

(a) Show that the minimum spanning tree of G consists only of the edges of the Delaunay triangulation of P.

(b) Using the above result, design and analyse an efficient algorithm to compute the minimum spanning tree of G.

[5+5=10]

INDIAN STATISTICAL INSTITUTE

Semestral Examination (Second Semester) : 2014-15

Course : B.Stat.-III Year

Subject : Design of Experiments

Date: 05/05/2015 Maximum Marks: 65 Duration: 2 hours 15 minutes

NOTE: (i) This paper carries 75 marks. Answer as much as you can but the maximum you can score is 65. The marks are indicated in [ ] on the right margin.

(ii) The symbols and notations have the usual meaning as introduced in your class.

1. List the three fundamental principles of design of experiments briefly explaining the role of any two of them in the planning of an experiment. [6]

2. Consider the following two block designs:

(a)

Block	Treatments		
1	1	2	3
2	3	1	2
3	2	3	1
4	4	6	5
5	6	5	4
6	5	4	6
7	3	5	7

(b)

Block	Treatments		
1	1	2	3
2	3	1	2
3	2	3	1
4	4	6	5
5	6	5	4
6	5	4	6

Find out if the preceding two designs are connected (justify your answers).

Write out the C matrix of design (a).

Identify a set of estimable functions of the treatment effects for design (b).

[2×2+6+4 = 14]

3. When are two Latin squares said to be orthogonal? What is meant by a set of "mutually orthogonal Latin squares (MOLs)"? Show that the maximum number of MOLs of order v is $(v-1)$.

Construct EITHER three (3) MOLs of order 8, (given that the elements of GF(8) are: 0, 1, α, $1+\alpha$, α^2, $1+\alpha^2$, $\alpha+\alpha^2$, $1+\alpha+\alpha^2$ & use $\alpha^3+\alpha+1$ as the irreducible/minimal polynomial) OR ELSE construct a pair of MOLs of order 12 (given that the elements of GF(4) are: 0, 1, α, $1+\alpha$ & use $\alpha^2+\alpha+1$ as the irreducible/minimal polynomial).

[2+2+3+10 = 17]

4. Consider the linear model $y_{n\times1} = X_{n\times p}\theta_{p\times1} + \varepsilon_{n\times1}$, $D(\varepsilon) = \sigma^2 I_n$.

Suppose the matrix X is partitioned and written as $X_{n\times p} = \begin{bmatrix} x_1' \\ X_a \end{bmatrix}$, where x_1' is the first row of the X matrix and X_a is the $(n-1)\times p$ submatrix of rows 2 through n of X. If the observation y_1 is missing and the data is augmented by an algebraic quantity z, find an explicit expression for $\hat{z}$, the value of z that gives correct error sum of squares as would be computed based on available data, i.e. $y_2, y_3, \cdots, y_n$.

Suppose in a RCBD involving 4 treatments applied in 4 blocks, an observation in block-1 corresponding to the treatment-1 is missing. Use your result to estimate the missing observation given that the grand total, total for treatment-1 and total of block-1 based on available data are 17, 1 and 6 respectively.

[10+3 = 13]

5. How the factorial effect $F_1^1 F_2^2$ is defined based on analytical as well as geometric definitions of factorial effects for a 3^2 factorial design involving two quantitative factors F_1 and F_2? What are the degrees of freedom associated with each of these? Given the following average response for the 3^2 treatment combinations, replicated twice based on a RCBD, compute the sum of squares due to $F_1^1 F_2^2$ based on both the definitions.

Treatment combinations	**(0 0)**	**(0 1)**	**(0 2)**	**(1 0)**	**(1 1)**	**(1 2)**	**(2 0)**	**(2 1)**	**(2 2)**
Average response	-3	-3	1	0	1	4	-2	4	-2

If we have an independent estimate for error variance as 1.44 based on 9 observations then test for the significance of both the above effects. Use the D-matrix $\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix}$ and the F-distribution table given in the next page for your computations.

[5+2+2×2+2×2 = 15]

6. A 2^6 factorial experiment with factors A, B, C, D, E, F was performed in 8 blocks of 8 treatment combinations each. A particular block consisted of the treatment combinations

a, abcde, bd, abef, def, ce, acdf, bcf.

i) Obtain the principal block.

ii) What effects are confounded with blocks?

[3+7 = 10]

PERCENTAGE POINTS OF THE T DISTRIBUTION

Tail Probabilities

	One Tail	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	
	Two Tails	0.20	0.10	0.05	0.02	0.01	0.002	0.001	
D	1	3.078	6.314	12.71	31.82	63.66	318.3	637	1
E	2	1.886	2.920	4.303	6.965	9.925	22.330	31.6	2
G	3	1.638	2.353	3.182	4.541	5.841	10.210	12.92	3
R	4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	4
E	5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	5
E	6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	6
S	7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7
	8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	8
O	9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	9
F	10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	10
	11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	11
F	12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	12
R	13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	13
E	14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	14
E	15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	15
D	16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	16
O	17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	17
M	18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	18
	19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	19
	20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	20
	21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	21
	22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	22
	23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	23
	24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	24
	25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	25
	26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	26
	27	1.314	1.703	2.052	2.473	2.771	3.421	3.690	27
	28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	28
	29	1.311	1.699	2.045	2.462	2.756	3.396	3.659	29
	30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	30
	32	1.309	1.694	2.037	2.449	2.738	3.365	3.622	32
	34	1.307	1.691	2.032	2.441	2.728	3.348	3.601	34
	36	1.306	1.688	2.028	2.434	2.719	3.333	3.582	36
	38	1.304	1.686	2.024	2.429	2.712	3.319	3.566	38
	40	1.303	1.684	2.021	2.423	2.704	3.307	3.551	40
	42	1.302	1.682	2.018	2.418	2.698	3.296	3.538	42
	44	1.301	1.680	2.015	2.414	2.692	3.286	3.526	44
	46	1.300	1.679	2.013	2.410	2.687	3.277	3.515	46
	48	1.299	1.677	2.011	2.407	2.682	3.269	3.505	48
	50	1.299	1.676	2.009	2.403	2.678	3.261	3.496	50
	55	1.297	1.673	2.004	2.396	2.668	3.245	3.476	55
	60	1.296	1.671	2.000	2.390	2.660	3.232	3.460	60
	65	1.295	1.669	1.997	2.385	2.654	3.220	3.447	65
	70	1.294	1.667	1.994	2.381	2.648	3.211	3.435	70
	80	1.292	1.664	1.990	2.374	2.639	3.195	3.416	80
	100	1.290	1.660	1.984	2.364	2.626	3.174	3.390	100
	150	1.287	1.655	1.976	2.351	2.609	3.145	3.357	150
	200	1.286	1.653	1.972	2.345	2.601	3.131	3.340	200
	Two Tails	0.20	0.10	0.05	0.02	0.01	0.002	0.001	
	One Tail	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	

Tail Probabilities

Date: 8.5.2015
Time: 3 hours

INDIAN STATISTICAL INSTITUTE

Statistical Methods in Genetics – I
B-Stat (3rd Year) 2014-2015
Final Examination

This paper carries 50 marks. Answer all questions.

1(a) Consider the following genotype data at a triallelic locus on 500 randomly chosen individuals in a population:

Genotype	Frequency
AA	182
AB	170
AC	66
BB	48
BC	34

Test whether the locus is in Hardy-Weinberg Equilibrium.

(b) Consider a disorder controlled by L unlinked autosomal biallelic loci such that an individual is affected if and only if he/she is homozygous recessive in at least one of these loci. If both parents are heterozygous at all L loci, what is the probability that an offspring will be affected? [8+4]

2. Suppose genotype data are available on a random set of individuals at an autosomal biallelic locus in an inbred population. If an analysis ignores the fact that the population is inbred, what is the impact on the maximum likelihood estimators of the allele frequencies at the locus? Show that the variances of the above estimators are the minimum in the absence of inbreeding in the population. [6+6]

3(a) Consider an autosomal biallelic locus with alleles A and a. Suppose the fitness corresponding to the genotypes AA, Aa and aa are f_1, f_2 and f_3, respectively. Show that the allele frequencies do not reach non trivial equilibrium values if $f_2 < min\{f_1,f_3\}$.

P.T.O.

(b) Consider a disease controlled by an autosomal biallelic locus. If an individual is affected, show that irrespective of the disease model, it is equally likely for the aunt and the grandson of the individual to be affected.

[6+7]

4(a) Consider a case-control association study at an autosomal biallelic locus comprising 50 cases and 50 controls randomly chosen from a population. If the allelic Odds Ratio based on the above sample is 1.5, justify that the allele-based test for association based on Odds Ratio does not provide significant evidence of association at level 0.05.

(b) Many genetic disorders are governed by body mass index (*BMI*). Suppose the following data are available on *BMI* and genotypes at an autosomal biallelic locus for a set of 100 individuals randomly chosen from a population.

Genotype	No. of individuals	Minimum *BMI*	Maximum *BMI*
AA	10	17.8	20.4
Aa	40	20.7	25.1
aa	50	25.6	31.2

Is it possible to infer on association between the locus and *BMI* based on the above data?

[6+7]

INDIAN STATISTICAL INSTITUTE

Second Semestral Examination: 2014-15

Course Name : B. STAT. III YEAR

Subject Name : Differential Equations

Date: May 8, 2015 **Maximum Marks: 60,** **Duration: 3 hrs.**

1. Find complete integral and singular solution of the nonlinear partial differential equation $2xz - \frac{\partial z}{\partial x}x^2 - 2\frac{\partial z}{\partial y}xy + \frac{\partial z}{\partial x}\frac{\partial z}{\partial y} = 0.$ 6
2. Using the Laplace transform technique, solve the following initial value problem
$$t\frac{d^2y}{dt^2} + \frac{dy}{dt} + ty = 0 \text{ with } y = 1, \frac{dy}{dt} = 0 \text{ at } t = 0.$$ 6
3. Solve the initial boundary value problem using Laplace technique
$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1,\ t > 0$$
with boundary conditions: $u(0,t) = 1, u(1,t) = 1, t > 0$ and initial condition: $u(x,0) = 1 + \sin \pi x, 0 < x < 1.$ 10
4. Show that Legendre polynomial $P_n(x)$ of degree n satisfies the relations
   a) $P'_{n+1}(x) - 2xP'_n(x) + P'_{n-1}(x) - P_n(x) = 0,$
   b) $(n+1)P_{n+1}(x) - (2n+1)xP_n(x) + nP_{n-1}(x) = 0, n = 1, 2, 3, \cdots$ 6+6
5. Calculate the regular singular point of the equation $(x - x^2)\frac{d^2y}{dx^2} + (1 - x)\frac{dy}{dx} - y = 0$. Find out the Frobenius series solution about that regular singular point. 2+7
6. Let $f(x,y)$ and $g(x,y)$ have continuous first partial derivatives in a simply connected domain D in R^2. Show that the system $\frac{dx}{dt} = f(x,y), \frac{dy}{dt} = g(x,y)$ has no closed trajectory in D if $f_x + g_y$ has the same sign in D. 5
7. Find the values of r at which bifurcation occur and classify them as saddle-node, transcritical, pitchfork bifurcation of $\frac{dx}{dt} = rx - \frac{x}{1+x^2}$. Sketch the bifurcation diagram of equilibrium points x^* vs. r. 8
8. By constructing a Lyapunov function, show that the system $\frac{dx}{dt} = -x + 4y, \frac{dy}{dt} = -x - y^3$ has no closed orbits. 4
9. Let the equation of motion of a mechanical system be $m\frac{d^2x}{dt^2} + a\frac{dx}{xt} + kx = 0$. Obtain conditions on $\lambda = \sqrt{\frac{k}{m}}$ and $b = \frac{a}{2m}$ under which the system has an oscillatory motion and critical damping motion. 4

INDIAN STATISTICAL INSTITUTE

Semestral Examination, 2nd Semester, 2014-2015

B.Stat. 3rd Year

NONPARAMETRIC AND SEQUENTIAL METHODS

Date: 11 May, 2015 Maximum Marks: 100 Duration: 3 Hours

Answer all questions.

1. Consider Wald's sequential probability ratio test (SPRT) for simple hypotheses with target strength (α, β) in the case of i.i.d. observations where α and β are small positive numbers with $\alpha + \beta < 1$.

(a) Show that Wald's approximations for the boundaries of an SPRT are conservative with respect to error probabilities.

(b) Find approximate expressions for average sample number (ASN) under H_0 using Wald's approximation and Wald's first equation. [10+6=16]

2. Let $X_i, i = 1, 2, \dots$ be i.i.d $N(\theta, \sigma^2)$ where σ^2 is known. Consider the SPRT for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ $(\theta_1 \neq \theta_0)$ where the boundaries satisfy $0 < B < 1 < A < \infty$. Show, without using Stein's lemma, that the SPRT terminates with probability one under all $\theta \neq \frac{1}{2}(\theta_0 + \theta_1)$. [12]

3. Let $X_1, X_2, \dots$ be i.i.d $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Describe Stein's two-stage sampling procedure for obtaining a bounded length confidence interval for μ with confidence coefficient $(1-\alpha)$. Prove the results you state. [13]

4. (a) Describe the concept of Pitman's asymptotic relative efficiency of tests.

(b) Let $X_1, \dots, X_n$ be i.i.d $N(\mu, 1)$ where μ is unknown. We want to test $H_0 : \mu = 0$ against $H_1 : \mu > 0$. Find the asymptotic relative efficiency of the sign test relative to the test based on the sample mean. Assume that the distribution of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Sign(X_i)$ under $\mu = \delta/\sqrt{n}$ converges to $N(\sqrt{2/\pi}\delta, 1)$. [6+11=17]

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

Define projection $\hat{U}_n$ of the U-statistic U_n and find its expression. Show that under suitable conditions $\sqrt{n}(U_n - \hat{U}_n) \xrightarrow{P} 0$ and hence find the asymptotic distribution of $\sqrt{n}(U_n - \theta)$. (Assume that $n\text{Var}(U_n) \to m^2\sigma_1^2$ where σ_1^2 has its usual meaning.) [6+12=18]

6. Let $X_1, \ldots, X_n$ be a random sample from a population with a continuous distribution F which is symmetric about its unknown median θ. Consider the use of Wilcoxon signed rank statistic T_+ for testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$.

(a) Express T_+ as a weighted sum of two U-statistics.

(b) Find the asymptotic distribution of T_+ under H_0.

(c) Show that the Wilcoxon signed rank test is consistent for any alternative under which $P(X_1 + X_2 > 0) > \frac{1}{2}$. [9+8+7=24]