INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD,
KOLKATA 700108

Mid-Sem. Examinations (2015-16).

Subject: Psychology

Date: 31.08.15

B.Stat. – III

Full marks: 25

Time: 1 hour

A. Give reasons (any 5) (5×2=10)

A.1. Psychology is not the study of spirit.

A.2. Individual difference is the most important principle of psychology.

A.3. Before validity, reliability is conducted.

A.4. Statistical approach is a more effective approach in classification of mental disorders.

A.5. Job satisfaction is not the opposite of job dissatisfaction.

A.6. Reward does not always mean reinforcement.

B. Fill up the gaps (any 5) (5×1=5)

B.1. ________ is at the middle level of need-hierarchy theory.

B.2. Behaviour is the function of person , environment and ____________

B.3. __________ can change quality of input and throughput.

B.4. __________ model follows symptomatic approach.

B.5. Sensation refers to __________ of stimulus.

B.6. Perception occurs when sensation carries ______________ .

P.T.O

C. Define the following (any 5) (5×2=10)

C.1. Vigilance

C.2. Reliability

C.3. Validity

C.4. Psychographic profile

C.5. Motivating factors

C.6.Scedule of reinforcement

Indian Statistical Institute

Mid-Semester Examination : 2015-16

Course name: B.Stat. III

Subject name: Sociology

Date: 31.8.2015 *Maximum marks: 30* *Duration: One hour*

Note, if any:

Write short notes on any of the five

a) Perspectives in Sociology
b) Early thinkers of Sociology
c) Sociology of religion and types of religious organizations.
d) Alienation and false consciousness
e) Rural society and its characteristics
f) Urban Sociology and its evolution
g) Social stratification and its various forms
h) Environmental refugee

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination: 2015-16

B. Stat. III Year

Sample Survey

Date: 01/09/2015 **Maximum Marks: 25** **Duration: 3 Hours**

Group – A

Answer any 2 questions each carrying 10 marks.

Records of assignments carry 05 marks.

Answers to Group A and Group B must be given in separate answer books.

1 Develop a theory of confidence interval estimation of a finite population total.

Illustrate its use in working out the size of an appropriate sample to be chosen randomly without replacement from a population of a given size postulating no distributional form of the population values.

2 Discuss briefly the rationales behind the uses of a (i) generalized regression predictor and also of (ii) the Brewer's predictor of a finite population total.

Show when (i) and (ii) may coincide .

3 Explain your concepts of replicated sampling and show its use in deriving Murthy's almost unbiased ratio - type estimator.

Also derive Hartley & Ross's exactly unbiased ratio - type estimator indicating how to unbiasedly estimate its variance.

*****

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination : 2015-2016
B.Stat. III Year
Sample Surveys

Date : 01.09.2015 Maximum Marks : 40 (Gr. A + Gr. B) Duration : 3 hrs (Gr. A + Gr. B)

Use two separate answer sheets for Group A and Group B

Group B (Total Marks : 20)
Notations are as usual.

Answer any 1 from Q.1 to Q.2 and Q.3 is compulsory.

1. (a) In SRSWOR of size n out of N units, derive the expressions of π_i and π_{ij}. Also prove that $Cov(\bar{x},\bar{y})=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})$. **1 + 1 + 3 = 5**

(b) Let $\bar{y}_d$ denote the sample mean based on distinct units obtained in an SRSWR of size n out of N units. Find $V(\bar{y}_d)$. Suggest with proof one unbiased estimator for $V(\bar{y}_d)$ based on distinct units only.
3 + 2 = 5

2. (a) In linear systematic sampling of size n from N = n.k units, where k is an integer, write down with proof an unbiased estimator $\hat{Y}$ of population total Y.

(b) Prove that in above systematic sampling, $V(\hat{Y})=\frac{N}{n}(N-1)S_y^2[1+(n-1)\rho]$.

(ρ is the Intra-class correlation coefficient).

© Write down an unbiased estimator and the variance of that estimator for the population mean of a variable of interest y under Stratified random sampling considering the SRSWOR in each stratum. Derive Neyman's optimum sample size allocation scheme in absence of cost function.
2+ 4 + 4 = 10

3. Samplonia, a small country, say, has exactly 1000 inhabitants, out of which the adult working population size is 341. There are two strata, namely Province Agria and Province Indusia. Suppose the objective is to estimate the mean income. The target population is the working population in Samplonia. Suppose Stratified Random Sampling of size 20 is to be used considering SRSWOR within every stratum. Table 1 below summarizes the population characteristics.

Table 1: Population size and variances of income in Samplonia

Stratum	Working Population Size	Variance in Incomes (S^2)
Province Agria	121	47,110
Province Induston	220	738,676
Country Samplonia	**341**	**929,676**

Calculate the variance of the estimator for population mean income based on the (i) proportional allocation, (ii) optimum allocation, and also considering (iii) complete SRSWOR of size 20. Comment on these results of the magnitudes of variances obtained by these three methods.
10

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : (2015-2016)

B.Stat. 3rd Year

PARAMETRIC INFERENCE

Date: 3 September, 2015 Max. Marks: 100 Duration: $2\frac{1}{2}$ Hours

Answer all questions

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

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

(a) For $n = 2$ verify whether $X_1 + 2X_2$ is sufficient for p.

(b) For $n = 3$ verify whether $X_1 + X_2 + 2X_3$ is sufficient for p.

(c) Find a complete sufficient statistic for p and the UMVUE of $P(X_1 + \cdots + X_{n-1} > X_n)$ for $n = 3$. [4+5+(5+6)=20]

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

4. (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 a U-estimable parametric function $g(\theta)$, an unbiased estimator $\phi(T)$ of $g(\theta)$, based on T, is a UMVUE. [3+10+12=25]

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

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

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

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

6. Let $X_1, \ldots, X_n$ be i.i.d. observations, each following a $U(0, \theta)$ distribution. Suppose it is known that $\theta > 1$. Show that $X_{(n)} = \max\ (X_1, \ldots, X_n)$ is not complete. [11]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination : 2015-2016

B. Stat. (Hons.) III Year 04·09·15

Subject : SQC & OR

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

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

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

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

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

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

Formulate this problem as an LP model for determining the monthly production programme to maximize contribution. [15]

2. For the problem formulated in Question # 2 above, what is the starting basic feasible solution. Show your working completely. [20]

3. The system $Ax = b, \quad x \geq 0$ is given by

$$x_1 + x_2 - 8x_3 + 3x_4 = 2$$
$$-x_1 + x_2 + x_3 - 2x_4 = 2$$

Find

a) a nonbasic feasible solution

b) a basic solution which is not feasible

c) a BFS which corresponds to more than one basis matrix. Write all the corresponding basis matrices.

d) a solution which is neither basic nor feasible.

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

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

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

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

i) What is the $\alpha -$ risk associated with the $\bar{X}$ – chart?

ii) Specifications on this quality characteristics are 358 ± 6. What are your conclusions regarding the ability of the process to produce items within specifications?

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

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

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

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

[35]

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

b). How does the ISO define quality?

c). Distinguish between ***control limits*** and ***specification limits***.

d). Distinguish between ***engineering quality control*** and ***statistical quality control***.

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

6. Choose the best answer:

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 linits.
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.

[1 × 6 = 6]

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

For $n > 25$

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

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

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

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2015-2016

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

Linear Statistical Models

Date: September 07, 2015 | Maximum Marks: 40 | Duration: 2 hours

---

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

---

1. Consider the linear model

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon, \quad \mathrm{E}(\epsilon) = \mathbf{0}, \quad \mathrm{Cov}(\epsilon) = \sigma^2 \boldsymbol{V},$$
$$\mathbf{X} \text{ is } n \times p,\ \boldsymbol{\beta} \text{ is } p \times 1,\ \mathbf{V} \text{ is a known } n \times n \text{ positive definite matrix.}$$

Suppose we choose to estimate an estimable function $\boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{\beta}$ by its least squares estimate. State, with reasons, a condition involving $\mathbf{X}$ and $\mathbf{V}$ that will force it to be a BLUE. [12]

2. Consider the linear model $\boldsymbol{Y} \sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$. $\mathbf{X}$ is $n \times p$, $\boldsymbol{\beta}$ is $p \times 1$. Let $\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\beta}$ be estimable, where $\boldsymbol{\Lambda}$ is a known $p \times q$ matrix. Suppose we wish to test the hypothesis

$$\mathrm{H}_0 : \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\beta} = \mathbf{0} \text{ versus } \mathrm{H}_1 : \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\beta} \neq \mathbf{0}. \tag{1}$$

Assuming results for testing a linear model against a reduced model, find the distribution of the ANOVA based test statistic for (1), denoted by F. Identify a suitable quantity that can be seen as a measure of deviation of alternative from null, and discuss the role of this quantity in the distribution of F. [8+8=16]

3. Consider a one-way ANOVA model given by

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ j = 1, \ldots, N_i,\ i = 1, \ldots, t,$$
$$\epsilon_{ij}\text{'s are uncorrelated with mean } 0 \text{ and variance } \sigma^2.$$

We denote this model by $(\boldsymbol{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$, where $n := \sum_{i=1}^t N_i$. Let $\mathrm{M}_\alpha := \mathbf{P}_\mathbf{X} - n^{-1}\mathbf{J}_{n\times n}$. Show that with any set of contrasts $\sum_{i=1}^t \lambda_{ri}\alpha_i$, $r = 1, \ldots, t-1$ for which $\sum_{i=1}^t \lambda_{ri}\lambda_{si}/N_i = 0$ for all $r \neq s$, we have a set of $t-1$ orthogonal constraints on $\mathcal{C}(\mathrm{M}_\alpha)$ so that the sum of squares for the contrasts equals $\boldsymbol{Y}^{\mathrm{T}}\mathrm{M}_\alpha\boldsymbol{Y}$. [12]

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

INDIAN STATISTICAL INSTITUTE

First Semester Examination : (2015-16)

B.Stat. 3rd Year

PARAMETRIC INFERENCE

Date: 16 November, 2015 Maximum Marks: 100 Duration: 3 Hours

Answer all questions.

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

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

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

Derive the usual optimum test using a suitable general result (to be stated by you) for exponential family. [7+11=18]

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

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

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

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

3. Let X be a random observable taking values $-1, 0, 1, 2, \cdots$ with probabilities

$$P(X = -1) = p,\ P(X = k) = q^2 p^k,\ k = 0, 1, 2, \cdots$$

where $0 < p < 1$ and $q = 1 - p$. Show that a parametric function $g(p)$ will have a UMVUE if and only if it is of the form $a + bq^2$ for some constants a and b. [11]

4. Suppose X has density $e^{-(x-\theta)}I(x > \theta)$ and the prior density of θ is $\pi(\theta) = [\pi(1+\theta^2)]^{-1}$, $\theta \in R$. Consider a loss function $L(\theta, a) = I(|\theta - a| > \delta)$ and find the Bayes estimate of θ.

[10]

5. Let $X_1, \ldots, X_n$ be i.i.d $N(\theta, 1)$. Consider the uniform prior $\pi(\theta) \equiv 1$.

(a) For the problem of testing $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$, find the posterior probability of H_0 in terms of the observed value t of the statistic $T = \sqrt{n}(\bar{X} - \theta_0)$.

(b) Find a $100(1-\alpha)\%$ HPD credible set for θ and compare it with the classical confidence interval.

[7+7=14]

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

(a) Find a consistent estimator of θ.

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

[5+11=16]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2015–2016

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

Linear Statistical Models

Date: November 18, 2015 **Maximum Marks: 50** **Duration: 2 and 1/2 hours**

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

---

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

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

(a) Consider a contrast in the α_i's. Find, with adequate reasons, its least squares estimate. [7]

(b) We wish to test the following hypotheses

$$\alpha_1 - \alpha_2 + \alpha_3 - \alpha_4 = 0,$$
$$\alpha_1 + \alpha_2 - \alpha_3 - \alpha_5 = 0,$$

with the requirement that the experimentwise error rate is at most α. Describe a suitable test and show from first principle that your test does indeed achieve the requirement on experimentwise error rate. [7+9 = 16]

2. Suppose that we have data on weight gain Y and initial weight x of pigs under four diets. Assume that the number of pigs under each diet is six. Describe how the hypothesis of equality of effects of the diets can be tested. [You must state all your assumptions clearly.] [12]

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

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

[P.T.O.]

Denote by $\mathbf{X}$ the design matrix and by $\mathbf{X}_0$, the submatrix of $\mathbf{X}$, formed by the columns of $\mathbf{X}$ which correspond to the parameters μ, α_i's, and η_j's. Let $\mathbf{M} := \mathbf{P}_{\mathbf{X}}$, $\mathbf{M}_0 := \mathbf{P}_{\mathbf{X}_0}$. Assume that $\mathcal{C}(\mathbf{M}-\mathbf{M}_0) = (\mathcal{C}(\mathbf{X}_0))^{\perp} \cap \mathcal{C}(\mathbf{X})$ is the interaction space. Consider the problem of testing

$$\mu(\sum_{i=1}^{3}\sum_{j=1}^{3} d_{ij}) + \sum_{i=1}^{3}(\sum_{j=1}^{3} d_{ij})\alpha_i + \sum_{j=1}^{3}(\sum_{i=1}^{3} d_{ij})\eta_j + \sum_{i=1}^{3}\sum_{j=1}^{3} d_{ij}\gamma_{ij} = 0$$

which puts a constraint on the interaction space $\mathcal{C}(\mathbf{M}-\mathbf{M}_0)$. Find an expression for the ANOVA based test statistic and its null distribution. [Your expression of the test statistic should be in terms of appropriate sums of squares. You may assume relevant facts about testing in a linear model and relevant linear algebraic facts.] [12+8 = 20]

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

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2015-16

B. Stat. III Year
Sample Survey

Date: 20/11/2015 **Maximum Marks: 25** **Duration: 3 Hours**

Group – A

N.B.: Assignment Records carrying 5 marks are to be submitted within 7 days after examination.

Answers to Group A and Group B must be given in separate answer books.

Answer any 2 questions each carrying 10 marks

1 Motivate the use of a regression estimator in finite populations. For the simple random sampling without replacement (SRSWOR) from a finite population, for the population mean how will you find a measure of error of the regression estimator? How will you estimate this measure of error ? Present explanatory details.

2 What do you mean by the usefulness of two-phase sampling in the context of stratification? Give details about how to use this when an initial simple random sample taken without replacement (SRSWOR) includes a few non-respondents. Explain how many initial units to go for and how many more to be sub-sampled.

3 Explain the concept of post-stratification. Starting with an SRSWOR present two post-stratified estimators of a finite population mean explaining their rationales. Show a use of post-stratification in handling 'Not-at-homes'.

*****

INDIAN STATISTICAL INSTITUTE
First Semester Examination : 2015-2016
B.Stat. III Year
Sample Surveys

ate : 20/11/15 Maximum Marks : 40 (Gr. A + Gr. B) Duration : 3 hrs (Gr. A + Gr. B)

Use two separate answer sheets for Group A and Group B

Group B (Total Marks : 20)

Notations are as usual.

Answer any 1 from Q.1 to Q.2 and Q.3 is compulsory.

1. (a) In cluster sampling of n clusters out of N clusters, i-th cluster having M_i units, derive an unbiased estimator $\hat{\bar{Y}}_C$ for population mean $\bar{Y}$, assuming that the clusters are selected by probability proportional to the number of units in them with replacement.

(b) Derive $V(\hat{\bar{Y}}_C)$.

(c) Show that $\frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i-\hat{\bar{Y}}_C)^2$ is an unbiased estimator for $V(\hat{\bar{Y}}_C)$, where $\bar{y}_i$ denotes the mean of y values in the i-th selected cluster.

2 + 3 + 5 = 10

2. (a) Derive Des Raj (1956)'s ordered estimator for unbiasedly estimating Y under PPSWOR sampling scheme and also derive its variance and an unbiased estimator for that variance.

(b) Write down Murthy (1957)'s symmetrized estimator from above estimator.

(c) Deduce Horvitz and Thompson (1952)'s unbiased estimator and its variance for the population total of a variable of interest. Show that Yates and Grundy (1953)'s form of this variance is an alternative form of this under fixed effective sample size design.

4 + 1 + 5 = 10

3. In a survey for estimating the impact of a comprehensive development project for the fish-farmers (FFD survey project in West Bengal, 1991 conducted by Indian Statistical Institute, Kolkata) in West Bengal, 8 villages were selected out of 47 in a block in the district of Burdwan by SRSWOR and a sample of households was selected by SRSWOR from each selected village. Following table gives the data about water area (in Satak) collected from the households with pisiculture as principal occupation.

(a) Estimate the total water area of the block along with its coefficient of variation.
(b) Work out the same problem if the villages were selected by SRSWR.

10

Sampled village Sl. No.	Total number of hhs	Total number of sampled hhs	Water area (in Satak) in sampled hhs
3	8	3	50 , 33 , 57
7	6	2	66 , 92
13	9	3	66 , 50 , 33
17	11	4	33 , 45 , 60 , 55
21	13	4	99 , 66 , 50 , 67
31	13	4	60 , 155 , 77 , 80
37	10	3	78 , 33 , 230
41	12	4	78 , 66 , 60 , 77

INDIAN STATISTICAL INSTITUTE
B. Stat. (Hons.) III Year 2015-2016

First Semester Examination

Subject : SQC & OR

Date : **26.11.2015** Full Marks : **100** Duration : **3 hrs.**

This paper carries 115 marks. You may answer as much as you can; but the maximum you can score is 100.

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

$$\mathcal{P}: \text{ Maximize } \quad z = x_1 + 2x_2 + 3x_3 + 4x_4$$

subject to

$$x_1 + 2x_2 + 2x_3 + 3x_4 \leq 20$$
$$2x_1 + x_2 + 3x_3 + 2x_4 \leq 20$$
$$x_i \geq 0 \qquad i = 1, \cdots, 4$$

a) Write down the dual $\mathcal{D}$ of the given problem $\mathcal{P}$.
b) Solve the dual.
c) Determine the solution of the given problem $\mathcal{P}$ using the complementary slackness conditions.

[5+5+10=20]

2. An investor wishes to invest Rs 12,000/- over a 1-year period, although he must decide to invest in market shares and/or gold. The profitability ratios estimated for each investment are based on the possible states of the economy as given in Tables 2.1 below:

Table 2.1: Profitability of investments to face different states of the economy

	High Growth	Average Growth	Stable	Poor Growth
Market Shares	5	4	3	−1
Gold	2	3	4	5

Consider a zero-sum game theory model with two players and answer the following:

a) How should the investor invest the capital of Rs 12,000/- to carry out the optimal strategy?
b) What mean profitability value is obtained?

[8+7=15]

3. A firm which sells and supplies building materials offers its customers a decoration counselling service for bathrooms and kitchens. In a normal operation, 2.5 customers arrive on an average every hour. A design consultant is available to answer customers' questions and to give recommendation about the product. The consultant takes an average time of 10 min per customer.

   a) The service target indicates that one customer who arrives must not wait more than 5 min before being attended on average. Is this target met? If not, what action do you recommend?
   b) If the consultant cuts the average time spent per customer to 8 min, is the service target met?
   c) The firm wishes to evaluate two alternatives:
   - Use one consultant with an average service time of 8 min per customer.
   - Use two consultants, each spending an average service time of 10 min per customer.

   If the firm pays the consultants Rs. 16/h and the expected waiting time before a service for customers is valued at Rs. 25/h, should this firm have two consultants? Give reasons for your answer.

[(5+2)+5+13=25]

4. Show how the following problem can be made separable:

$$\text{Maximize } z = x_1 x_2$$

subject to

$$x_1 + x_2 \leq 100$$
$$x_1, x_2 \geq 0.$$

[10]

5. Answer the following questions:

   a) Consider a double sampling acceptance rejection plan with the following parameters:-

   $$n_1 = 50;\ c_1 = 1;\ r_1 = 4;\ n_2 = 100;\ c_2 = 3;\ r_2 = 4.$$

   Find the probability of acceptance on the second sample a lot that has fraction defective $p = 0.05$.

   b) Suppose that a single sampling acceptance rectification plan with $n = 150, c = 1$ is being used for receiving inspection where the vendor ships the product in lots of size $N = 3000$. Find the AOQL for this plan.

   c) A manufacturing company has two equally important processes namely, machining of valve diameter (R1) and machining of valve thickness (R2). The Quality Department has a major responsibility to judge which area is to be prioritized for improvement. The report is summarized in Table 5.1.

Table 5.1: Data for Question No. 5(c)

Process →	R1	R2
Parameters↓		
USL	3.2741	3.3513
LSL	3.0765	3.1041
μ	3.1830	3.1973
σ	0.0123	0.0542
C_p	2.6775	0.7601
C_{pk}	2.4688	0.5732
$C_p - C_{pk}$	0.2087	0.1869

In view of budgetary constraints, special action for improvement can be initiated on ***only one*** of these processes. As a consultant which process would you recommend and why?

[6+8+11=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 6.1. The specification limits are USL= 0.35, and LSL=0.05.

Table 6.1: Data for Question No. 6

Part Number	Operator I Measurements		Operator II Measurements		Operator III Measurements	
	1	2	1	2	1	2
1	0.20	0.10	0.10	0.10	0.15	0.15
2	0.20	0.30	0.15	0.25	0.25	0.25
3	0.15	0.10	0.20	0.10	0.20	0.15
4	0.30	0.30	0.25	0.30	0.20	0.25
5	0.20	0.15	0.15	0.05	0.15	0.05

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]

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

Semester Examination 2015-16

Course Name : B.Stat III

Subject Name: Psychology

Date: ~~9.11.2015~~ 28.11.2015 Maximum Marks: 50 Duration: 2 hours

Note: Answer should be precise and relevant. Do not write unrelated thing. Write title of figure. Write names of different parts of figure using arrow. Writing should be legible.

A. Give reasons (any 5) (5×4= 20)

A.1. Difficulty in memorizing occurs when non-sense syllables are used.

A.2. Learning curve is S-shaped.

A.3. Some information is lost in short term memory.

A.4. Man wants to forget stressful events.

A.5. Brain is the seat of cognition.

A.6. Man intends to perceive as a whole.

A.7. Illusion is not psychiatric disorder.

A.8 Memorizing ability can be assessed through questionnaire based survey and experiment.

P.T.O

B. Fill up the gaps (any 5) (5×1=5)

B.1. In experiment on memory, inter stimulus gap is _____ seconds.

B.2. _____ suggested presence of visual sketchpad.

B.3. Hallucination is the _______ of psychiatric disorder.

B.4. Echoic memory briefly holds mental representations of ______ stimuli.

B.5. In ________nearby things pass quickly, while far off objects appear stationary.

B.6. _________is the seat of emotion.

B.7. Validity is what the ____________ intends to measure.

B.8. Reliability of psychological test is prelude of ________.

C. Define the followings (any 5) (5×5= 25)

C.1. Questionnaire

C.2. Monocular cues

C.3. Brain localization

C.4. Illusion

C.5. Lymbic system

C.6. Neuron

C.7. Forgetting

Indian Statistical Institute

First-Semester Examination: 2015-16

Course name: B.Stat. III

Subject name: Sociology

Date: 28.11.2015 *Maximum marks: 50* *Duration: Two hours*

Note, if any:

Write short notes on any of the four (each question carries 5 marks)

1. Durkheim's idea on suicide
2. Exemplary prophecy and ethical prophecy
3. Jajmani system in India
4. Characteristics and impact of industry on society
5. Sociology of work and occupation
6. Ethics of research
7. Definition of Culture
8. Veblen's definition and impact of conspicuous consumption

Answer any two of the following within 250 words each

Each answer carries 15 marks

1. Outline Weber's description of "the Protestant Ethic", as well as the way in which he derives the concept of "Spirit of Capitalism".
2. What is qualitative sociology? Compare its advantages and disadvantages with quantitative research methods.
3. How and why social movements arise? What are the stages and types of social movements?
4. Briefly outline constitutional safeguards for Scheduled Castes, Scheduled Tribes and Other Backward classes in India.
5. Define Social exclusion. Describe types of social exclusion and its impact on Dalits in India.

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2015-16

Course Name : Optional Geology

Subject Name: Geology

Date: 30.11.2015 **Maximum Marks: 40** **Duration: 2 Hrs**

ANSWER ANY 4 (FOUR) QUESTIONS

1. What do you understand by the term 'Clastic rocks'? Can you categorize evaporites as clastic rocks? How would you interpret with the help of primary sedimentary structures the flow regime of the depositing flow? Explain with a diagram. What do you understand by the term stream power? 2+1+5 + 2

2. A coarse grained igneous rock contains Quartz, K-feldspar, Na-feldspar and minor amount of amphibole and mica and has 75% silica. Name the rock. Does this rock contain iron and magnesium as major chemical elements? Does this rock commonly show 'aphanitic' texture? What is the extrusive equivalent of this rock? What do you understand by the term "Bowen's reaction series"? Does silica percentage control the viscosity of the parent magma?
1+1+1+1+5+1

3. What do you understand by the term "Gondwana"? Why do we get fossils of marine invertebrates from the Himalayan sedimentary rocks? Do you expect any fossil to be present in an igneous rock? Justify your answer.
4+3+1+2

4. What are the important factors for metamorphism of an existing rock? Can a metamorphic rock be metamorphosed again? Define Barrovian zones of metamorphism
4+1+5

5. Why fossilized coral reefs are used to illustrate the physical conditions of deposition of their host rocks? Why the sutures of the ammonites are important to understand their evolution and phylogeny? Write a note on the symmetry of echinoid test (shell). 4+3+3

INDIAN STAISTICAL INSTITUTE

Mid – Semester Examination: 2015 – 16

B.Stat - III year.

Course Name : Elective Subject

Date: 31.08.2015

Subject Name : Geology   Maximum Marks : 40   Duration : 2 Hours

Note, if any :

Answer question number 1 and any questions

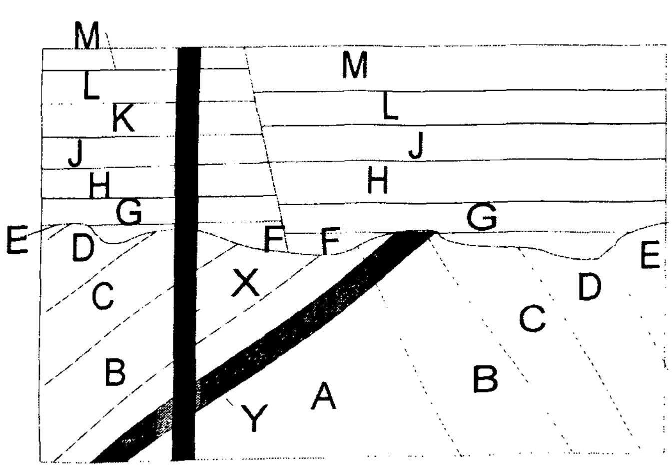
1. From the river-cut section given below, write the

   i. succession of the beds starting from the oldest to the youngest -- 6

   ii. nature of the unconformity -- 2

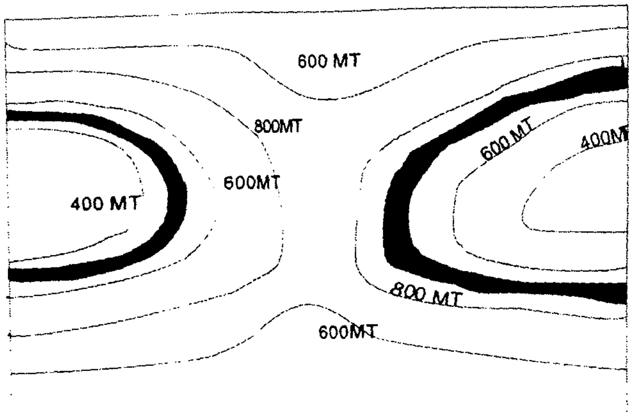
   iii. nature of the folds -- 2

   iv. And the nature of the fault -- 2



   v. In the diagram below, write the approximate amount of dip of the black colored bed. -- 3

P.T.O



2. How do we predict that the earth has a layered structure? ------ 3
3. What kind of silicate structure is present within the feldspar group of minerals? Elaborate your answer with illustrations. -- 5
4. What is a monophyletic group? --3
5. Why land living plants and animals of prehistoric time are found as fossils in the present day continents that are widely separated by oceans? ---------------------- 2
6. What is the basis/principle of Magnetostratigraphy? ----------------- 3
7. What is a "Formation" and what is a "Biozone"? ----------------------- 4
8. Why fossils mostly provide the relative ages of their host rocks? How can we use radio-isotopes to find out the absolute age of rocks? --------------------------- 5

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination

B. Stat-III Year, 2015-16 (Semester-I)

Design and Analysis of Algorithms

Date: **February 22, 2016** Maximum Marks: **60** Duration: **2 Hours**

Note: The question paper carries a total of 73 marks. You can answer as much as you can, but the maximum you can score is 60.

1. Let $f_1(n)$ and $f_2(n)$ be two non-negative functions in n, where n is a positive integer. Suppose $f_1(n) = O(f_2(n))$.
   (i) Show that $f_2(n) = \Omega(f_1(n))$.
   (ii) Prove or disprove : $2^{f_1(n)} = O(2^{f_2(n)})$.

(5+8=13))

2. (a) Design an efficient algorithm which, given two strings A and B, finds the first occurrence (if any) of B in A. Derive the complexity of the algorithm.
   (b) The input is two strings of characters $A = a_1 a_2 \cdots a_n$ and $B = b_1 b_2 \cdots b_n$. Design an $O(n)$time algorithm to determine whether B is a cyclic shift of A. In other words, the algorithm should determine whether there exists an index k such that $a_i = b_{(k=i) \bmod n}$ for all i, $1 \leq i \leq n$.

((10+3)+7=20))

3. Show how to compute the square of a 2×2 real matrix with only five multiplications.

(10))

4. Design an algorithm for finding the LCM of two given positive integers. Derive the complexity of your algorithm.

(7+3=10))

5. The input is a set S containing n real numbers and a real number x.
   (a) Design an algorithm to determine whether there are two elementsof S whose sum is exactly x. The algorithm should run in time $O(n \log n)$.
   (b) Suppose now that the set S is given in a sorted order. Design an $O(n)$time algorithm for the problem.

(10+10=20))

Indian Statistical Institute

B. Stat. Third Year : 2015-2016
Mid-Semester Examination
Subject: Number Theory
Date: 23/02/2016 Time: 2 hours Marks: 60

Note: Notations used are as explained in the class.

1. Show that if $(a, b) = 1$ then $(a + b, a^2 - ab + b^2) = 1$ or 3. [8]

2. Show that 24 is the largest integer divisible by all integers less than its square root. [10]

3. Let p be a prime factor of $a^2 + 2b^2$. Show that if p does not divide both a and b then the congruence $x^2 \equiv -2 \pmod{p}$ has a solution. Also show that $x^2 \equiv -2 \pmod{p}$ is solvable if and only if $p = 2$ or $p \equiv 1$ or $3 \pmod{8}$. [6 + 6]

4. Show that if p is a prime then there exist $\phi(p - 1)$ primitive roots modulo p. [12]

5. Show that if Q is a positive odd integer then

$$\left(\frac{2}{Q}\right) = (-)^{\frac{Q^2-1}{8}}.$$

[10]

6. Evaluate the value of

$$\left(\frac{7411}{9283}\right).$$

[8]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination

BStat III Year

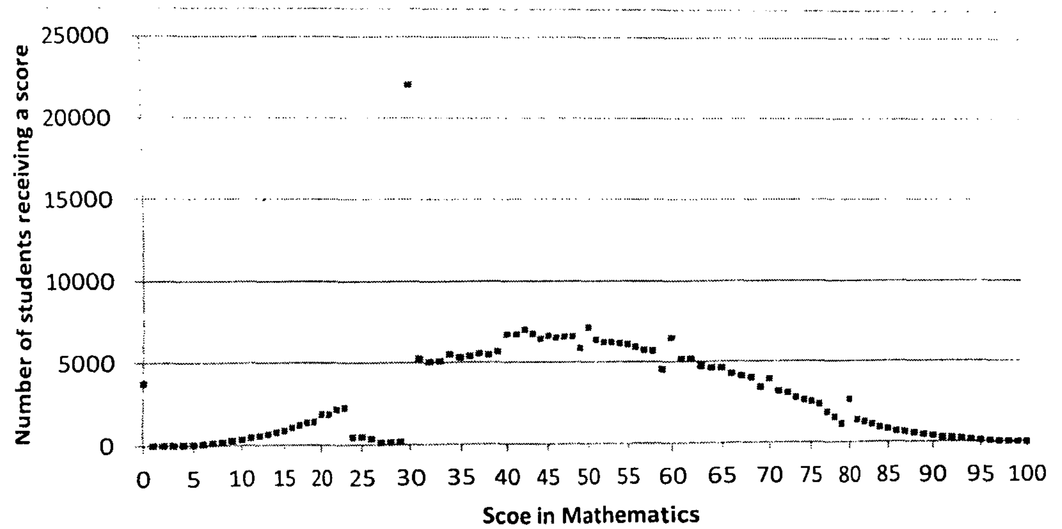
2015-16, II Semester

Statistics Comprehensive

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

Maximum time: 2 hours | 24th February, 2017 | Maximum score: 100

1. The performance of students in a mathematics examination is summarized in the scatter plot given below, where the number of students receiving a particular score is plotted against that score.



   a. Examine the plot carefully and explain the unusually high values at the scores of 0 and 30.

   b. Explain the sudden drop in the number of students from score 23 to 24.

   c. Identify any other notable feature of the plot and explain plausible reasons for the same.

   d. Draw a free-hand 'normal plot' of the data.

2. Suppose X is a sample of size 1 from the Poisson distribution with mean λ. Two estimators of $e^{-2\lambda}$ are proposed:

$$T(X) = \begin{cases} +1 & \text{if } X \text{ is even,} \\ -1 & \text{if } X \text{ is odd,} \end{cases} \qquad S(X) = \begin{cases} 1 & \text{if } X \text{ is even,} \\ 0 & \text{if } X \text{ is odd.} \end{cases}$$

Which estimator would you prefer and why?

P.T.O

3. a. Give an example of ten data points so that their mean is larger than the mode but smaller than the median.

   b. Give an example of ten data points so that their mode is larger than the median but smaller than the mean.

   c. You have a sample of size 19 from the discrete uniform distribution over the set

$$\{n, n+1, n+2, \ldots, 3n-1, 3n\}.$$

   The observed values happen to be 7, 13, 10, 13, 12, 5, 14, 14, 9, 10, 7, 7, 8, 12, 14, 13, 6, 8 and 6. Suggest three different MLE's of n and justify your answers. Are any of the MLE's unbiased?

4. A population consists of three males and four females. A sample of size $n = 3$ has to be drawn without replacement, so that both the genders are represented. Compute the first and second order inclusion probabilities.

5. Consider data $\boldsymbol{y}$ following the linear model $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta}$ is of the form $\begin{pmatrix}\beta_0 \\ \boldsymbol{\beta}_1\end{pmatrix}$. Here, β_0 is the intercept and $\boldsymbol{\beta}_1$ is the vector of other regression coefficients. You wish to test whether $\boldsymbol{\beta}_1$ is proportional to a specified vector $\boldsymbol{\gamma}$.

   a. Show that the testing problem can be reformulated as a linear hypothesis.

   b. Explain how you will carry out the test, for a given data set, by using any standard package for linear regression. Your answer should be in the form of detailed instructions to a practitioner who can handle and manipulate data sets but does not know statistics.

BIII-2015-16

Midterm Examination

Design of Experiments

Full Marks 35 Date: 25th February, 2016 Time: 10.30-12.30

1. For a general block design with b blocks and v treatments, show that Rank (C matrix) + b = Rank(D matrix) + v where the C-matrix and the D-matrix have their usual meanings. (You can assume the structure of the X'X matrix)

(7)

2. Consider v treatments and all possible $\binom{v}{k}$ sets of k treatments. Assume a general block design d with $\binom{v}{k}$ blocks, for which each such set is a block of size k. Find out the replication of each treatment in d. Obtain the C- matrix of the design d. Is it a connected design? Is it orthogonal? Obtain the BLUE of an estimable treatment contrast in terms of the adjusted treatment totals. What will be the variance of the BLUE you obtained?

(2+ 7+ 3+ 3 +3+2=20)

3.Consider a CRD with $n= 4t+3$ observations with $2t$ treatments, for some integer t. If you want to minimize the average variance of the BLUEs of all pairwise comparison of treatment effects, how many number of observations will you allocate to the treatments? Justify your claim.

(8)

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination : (2015-2016)

B.Stat. 3rd Year

NONPARAMETRIC AND SEQUENTIAL METHODS

Date: 26 February, 2016 Max. Marks: 80 Duration: $2\frac{1}{2}$ Hours

1. A set of four coins is tossed n times. How do you test for the hypothesis that all the four coins are fair? [7]

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

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

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

(c) Show that the Kolmogorov-Smirnov statistics D_n^+ and D_n^- are identically distributed under H_0. [6+9+8=23]

3. 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 $P(X_1 + X_2 > 0) > \frac{1}{2}$. [10+9+8=27]

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

(a) 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 $H_1 : \theta > 0$.

(b) Show that the Hodges-Lehmann estimate of θ is symmetrically distributed about θ if either (i) F is symmetric or (ii) $m = n$. [10+13=23]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2015 - 16

Course: B.STAT. III Year
Subject: Differential Equation

Date: 27/2/16 **Maximum Marks: 40** **Duration: $2\frac{1}{2}$hrs.**

Any result that you use should be stated clearly

1. Find the general solution of the following equations:
a) $y'' - 7y' + 12y = e^{2x}(x^3 - 5x^2)$
b) $x^2y'' + 3xy' + 10y = 0$
c) $xy' + y = x^4y^3$ [3+3+3]

2. A rock of mass m is thrown upward from the surface of the earth with initial velocity v_0. If air resistance is assumed to be proportional to the velocity, with constant of proportionality k, and if the only force acting on the rock is a constant gravitational force, find the maximum height it reaches. When does the rock reach this height? [3+2]

3. A rabbit starts at the origin and runs up the y-axis with speed m. At the same time a dog, running with same speed m starts at the point (c,0) and pursues the rabbit. Find y as a function of x. [5]

4(a). Find the Frobenius series solutions to the Gauss's Hypergeometric equation

$$x(x-1)y'' + [c - (a+b+1)x]y' - aby = 0$$

where a, b and c are constants, about the regular singular point $x = 0$.
(b). Prove the convergence of these series solutions. [10+3]

5. Find the Frobenius series solutions of the equation

$$x^2y'' + xy' + (x^2 - \tfrac{1}{4})y = 0$$ [3]

6. A 50 gallons tank initially contains 10 gallons of fresh water. At $t = 0$, a brine solution containing 1 pound of salt per gallon is poured into the tank at the rate of 4 gallons per minute, while the well-stirred mixture leaves the tank at the rate of 2 gallons per minute. Find (a) the time required for overflow to occur and (b) the amount of salt in the tank at the time of overflow. [3+2]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination

Second semester

B. Stat - Third year 2016

Random graphs

Date: 29th February, 2016

Maximum Marks: 30

Duration: 2 hours

Anybody caught using unfair means will immediately get 0. Please try to explain every step. Only class notes are allowed in the exam.

(1) Show that in an $\mathrm{ER}_n(\lambda/n)$ ($\mathrm{ER}_n(p)$ denotes the Erdös Rényi random graph with parameters n and p),

$$P(1 \nleftrightarrow 2 \,|\, |\mathcal{C}(1)| = l) = 1 - \frac{l-1}{n-1},$$

where $|C(1)|$ denotes the size of the cluster containing vertex 1. [5 points]

(2) Consider an $\mathrm{ER}_n(\lambda/n)$. We say that the distinct vertices (i, j, k) form an occupied triangle when the edges (i, j), (j, k) and (k, i) are all occupied (equivalently, an edge is present). Note that (i, j, k) is the same triangle as (i, k, j) and other permutations. Compute the expected number of occupied triangles.

[5 points]

(3) We know that the extinction probability of Galton Watson branching process

$$\eta = \lim_{n\to\infty} P(Z_n = 0) = \lim_{n\to\infty} \phi_n(0)$$

where ϕ_n is the generating function of the number of particles in the n^{th} generation. If $p_0 > 0$ then show that for every $s \in [0, 1)$, $\phi_n(s) \to \eta$ as $n \to \infty$. [8 points]

(4) For a branching process with i.i.d. offspring X having mean offspring $\mu < 1$ show that

$$E[T_p] = \frac{1}{1-\mu}$$

where T_p is the total progeny of the process.

[5 points]

(5) Show that for (X, Y) random variables, $P(Y \le z) \le P(X \le z)$ for all z if and only if there exists a coupling $(\widehat{X}, \widehat{Y})$ of (X, Y) such that $P(\widehat{X} \le \widehat{Y}) = 1$.

Let $\lambda, \mu \ge 0$ such that $\lambda \le \mu$. Let $X \sim \mathrm{Poi}(\lambda)$ and $Y \sim \mathrm{Poi}(\mu)$. Then show that $P(Y \le z) \le P(X \le z)$. [7 points]

Indian Statistical Institute

B. Stat. Third Year : 2015–2016
Semester Examination
Subject: Number Theory

Date: 26/04/2016 Time: 3 hours Marks: 100

Note: Notations used are as explained in the class.

1. Let a and b be positive integers such that $ab-1$ divides a^2+b^2. Show that $\frac{a^2+b^2}{ab-1}$ is 5. [12]

2. Prove that there are infinitely many primes by considering the sequence $2^{2^1}+1, 2^{2^2}+1, 2^{2^3}+1, \cdots$. [8]

3. Evaluate the value of
$$\left(\frac{6278}{9975}\right).$$
[10]

4. If f and g are multiplicative functions, then prove that $f*g$ (the convolution of f and g) is also multiplicative. [10]

5. Let a, b and c be nonzero integers such that abc is squarefree. Then prove that $ax^2+by^2+cz^2=0$ has a nontrivial rational solution if and only if both of these conditions are satisfied: (1) All of a, b, c don't have the same sign, (2) $-ab$ is a square $\bmod c$, $-bc$ is a square $\bmod a$ and $-ac$ is a square $\bmod b$. [15]

6. Prove that any positive integer n can be written as sum of four squares. [12]

7. Prove that there exist positive constants a and b, $a<1<b$ such that
$$a\left(\frac{x}{\log x}\right) < \pi(x) < b\left(\frac{x}{\log x}\right)$$
for sufficiently large x. [15]

8. Let d be a positive integer which is not a square. Show that if (p,q) is a positive integer solution of $x^2-dy^2=1$, then $\frac{p}{q}$ is a convergent of the continued fraction expansion of $\sqrt{d}$. [8]

9. Prove that among the rational numbers, the only ones that are algebraic integers are the integers $0, \pm1, \pm2, \pm3\cdots$. Give an example to show that factorization in the field $\mathbb{Q}(\sqrt{-6})$ is not unique. [7 + 3 = 10]

INDIAN STATISTICAL INSTITUTE

Second Semester Examination: 2015 - 16

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

Date: 26. 04. 2016 Maximum Marks: 80 Duration: 3 hrs.

Any result that you use should be stated clearly.

1. Find the general solution of

$$\frac{dx}{dt} + 4x + 3y = t, \qquad \frac{dy}{dt} + 2x + 5y = e^t.$$

(7)

2. Derive the Euler's differential equation for an extremal of

$$I = \int_{x_1}^{x_2} f(x, y, \frac{dy}{dx}) dx.$$

Hence, find the extremals for the integral

$$\int \frac{\sqrt{1 + (\frac{dy}{dx})^2}}{y} dx.$$

(9)

3. Solve the partial differential equation, $(1 + y)\frac{\partial z}{\partial x} + (1 + x)\frac{\partial z}{\partial y} = z.$ (7)

4. Consider the two-dimensional system

$$\frac{dx}{dt} = -x^4 + 5\mu x^2 - 4\mu^2, \quad \frac{dy}{dt} = -y.$$

Determine the critical points and the bifurcation diagram for this system. Draw the phase portraits for various values of μ and draw the bifurcation diagram. (9)

5. Let $f(x, y)$ be a continuous function and satisfy a Lipschitz condition in R. Prove that, if a solution of the initial value problem

$$\frac{dy}{dx} = f(x, y), \quad y(x_0) = y_0$$

exists, then it is unique. (5)

6. By constructing a Liapunov function, show that the system $\frac{dx}{dt} = -2y + yz - x^3$, $\frac{dy}{dt} = x - xz - y^3$, $\frac{dz}{dt} = xy - z^3$ has no closed orbits. (6)

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. (9)

8. (a) Show that Legendre polynomial $P_n(x)$ of degree n satisfies the relation:

$$(2n+1)P_n(x) = P'_{n+1}(x) - P'_{n-1}(x).$$

(b) Find the general solution of the equation

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

(5+5)

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

$$y'' + 4y' + 8y = \cos 2t$$

(where ($'$) denoting differentiation with respect to t), given that $y = 2$ and $y' = 1$ when $t = 0$.
(b) Show that $L[x\cos ax] = \frac{p^2-a^2}{(p^2+a^2)^2}$ and use this result to find $L^{-1}[\frac{1}{(p^2+a^2)^2}]$, where L denotes the Laplace transform of a function $f(x)$ defined as $L[f(x)] = F(p)$, (p being a parameter) and L^{-1} denotes the inverse Laplace transform. (5+5)

10. (a) Prove that

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

where λ_m, λ_n are the positive zeros of the Bessel function $J_p(x)$, p being an integer.
(b) If $f(x)$ is defined by

$$f(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ \frac{1}{2} & x = \frac{1}{2} \\ 0 & \frac{1}{2} < x \leq 1 \end{cases}$$

show that $f(x) = \sum_{n=1}^{\infty} \frac{J_1(\frac{\lambda_n}{2})}{\lambda_n J_1(\lambda_n)^2} J_0(\lambda_n x)$ where λ_n are the positive zeros of $J_0(x)$. (5+3)

INDIAN STATISTICAL INSTITUTE

Semestral Examination

Second semester

B. Stat - Third year 2016

Random graphs

Date: 26th April, 2016

Maximum Marks: 60

Duration: 3 hours

Anybody caught using unfair means will immediately get 0. Please try to explain every step. Only class notes are allowed in the exam.

(1) Clustering measures the degree to which neighbours of vertices are also neighbours of one another. For instance, in Facebook, your friends tend to know each other therefore the amount of clustering is high. In general clustering is high when network has many triangles. The quantity that measures the amount of network clustering is the clustering coefficient. We define the following quantities for Erdös-Renyi random graphs $ER_n(p)$ with $p = \lambda/n$. Let the number of wedges and triangles be defined as

$$W_G = \sum_{1 \le i,j,k \le n} \mathbb{1}_{ij,jk \text{ is occupied}}$$

and

$$\Delta_G = \sum_{1 \le i,j,k \le n} \mathbb{1}_{ij,jk,ik \text{ is occupied}}.$$

The clustering coefficients C_G is defined as

$$C_G = \frac{\Delta_G}{W_G}.$$

Show that $W_G/n \longrightarrow \lambda^2$ in probability using the second moment method. Also show that

$$nC_G \xrightarrow{d} \frac{6}{\lambda^2} Y,$$

where Y follows the Poisson distribution with parameter $\lambda^3/6$.

[5+10=15points]

(2) Fix $\lambda = a \log n$ and let M denote the number of isolated edges in $ER_n(p)$ with $p = \lambda/n$, that is, the edges that are occupied but for which the vertices at either end have no other neighbours.

(a) Show that $M \xrightarrow{P} \infty$ when $a > 1/2$.

(b) Show that $M \xrightarrow{P} 0$ when $a < 1/2$.

[8+7=15 points]

(3) Consider the generalised random graph $GRG(\mathbf{w})$ with the weights $\mathbf{w} = (w_i)_{i \in [n]}$ satisfying the standard assumptions:

(a) Let U be an uniformly chosen vertex and $W_n = w_U$ the weight corresponding to this vertex. Assume there exists a W such that $W_n \xrightarrow{d} W$.

(b) Also assume $\lim_{n\to\infty} E[W_n] = E[W] > 0$.

We have seen in the lectures that degrees of m (fixed) uniformly chosen vertices in $[n]$ are asymptotically independent. Now suppose in addition to the above assumptions that there exists $\varepsilon > 0$ such that $\varepsilon < w_i < \varepsilon^{-1}$ for every i. Show that we can couple the degree sequence $(D_i)_{i\in[m]}$ to an independent vector $(\hat{D}_i)_{i\in[m]}$ such that

$$P\left((D_i)_{i\in[m]} \neq (\hat{D}_i)_{i\in[m]}\right) = o(1)$$

whenever $m = o(\sqrt{n})$. Hence conclude that even the degrees of a growing number of vertices can be coupled to independent degree sequence.

[12+3=15 points]

(4) Consider the configuration model $CM_n(\mathbf{d})$ with degree sequence $\mathbf{d} = (d_i)_{i\in[n]}$ satisfying the assumptions:

(a) Let U be an uniformly chosen vertex and $D_n = d_U$ the degree corresponding to this vertex. Assume there exists a D such that $D_n \xrightarrow{d} D$.

(b) Also assume $\lim_{n\to\infty} E[D_n] = E[D] > 0$.

Let $(D_i^{er})_{i\in[n]}$ be the degree sequence of the erased configuration model. Define

$$p_k^n = \frac{1}{n}\sum_{i\in[n]} \mathbf{1}_{\{d_i=k\}} \text{ and } P_k^n = \frac{1}{n}\sum_{i\in[n]} \mathbf{1}_{\{D_i^{er}=k\}}.$$

If $p_k = P(D = k)$ then show that under assumptions (a) and (b) that $(P_k^n)_{k\geq 1}$ converges in probability to $(p_k)_{k\geq 1}$.

[You can assume the result holds for the case when one assumes an extra condition $\lim_{n\to\infty} E[D_n^2] = E[D^2]$, the case which was done in lectures.] [15 points]

INDIAN STATISTICAL INSTITUTE

Semestral Examination, 2nd Semester, 2015-2016

B.Stat. 3rd Year

NONPARAMETRIC AND SEQUENTIAL METHODS

Date: 29 April, 2016 Maximum Marks: 100 Duration: 3 Hours

Answer all questions.

1. Consider Wald's sequential probability ratio test (SPRT) for a simple hypotheses H_0 and a simple alternative H_1 with target strength (α, β) in the case of i.i.d. observations.

(a) Find approximate expression for average sample number (ASN) under H_0 using Wald's approximations and Wald's first equation.

(b) Let n be the stopping time of any sequential test procedure (for testing H_0 against H_1) of strength (α, β) with finite $E(n|H_i)$, $i = 0, 1$, which terminates with probability one under H_0 and H_1. Show that $E(n|H_0)$ is greater than or equal to the approximate value of the ASN under H_0 for the SPRT, as obtained in part (a) [7+13=20]

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

Let n be the stopping time of Stein's procedure. Prove that n is finite with probability one. Also show that the sample mean $\bar{X}_n$ is an unbiased estimator of μ. [13+4+5=22]

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

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

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

5. Consider independent pairs of observations (X_i, Y_i), $i = 1, \ldots, n$ from a continuous bivariate distribution. Define sample Kendall's Tau and find its asymptotic distribution under independence of X and Y. [10]

6. Consider a regression problem in which $Y_i = \beta x_i + e_i$, $i = 1, \ldots, n$ where $e_1, \ldots, e_n$ are continuous random variables with median zero and $x_1, \ldots, x_n$ are known constants satisfying $x_1 < x_2 < \cdots < x_n$. Suppose that we test $H_0 : \beta = 0$ against $H_1 : \beta > 0$ by rejecting H_0 for large values of the statistic T_n where T_n is the Kendall's Tau coefficient between $\boldsymbol{x}$ and $\boldsymbol{Y}$.

Consider the above test with level α. Show that this test is unbiased. Also show that this test is of level α for testing the null hypothesis $H : \beta \leq 0$ against $H_1 : \beta > 0$. [7+7=14]

INDIAN STATISTICAL INSTITUTE

Semestral Examination

B. Stat-III Year, 2015-16 (Semester-I)

Design and Analysis of Algorithms

Date: **May 03, 2016** Maximum Marks:100 Duration: **3 Hours**

Note: The question paper carries a total of 122 marks. You can answer as much as you can, but the maximum ›u can score is 100.

1. Find two non-negative functions $f_1(n)$ and $f_2(n)$, both monotonically increasing, such that $f_1(n) \neq O(f_2(n))$ and $f_2(n) \neq O(f_1(n))$.

(8)

2. (a) The input is d sequences of elements such that each sequence is already sorted, and there is a total of n elements. Design an $O(n \log d)$ time algorithm to merge all the sequences into one sorted sequence.
   (b) Establish the lower bound on the time complexity of a comparison-based sorting algorithm.

(10+10=20)

3. (a) Let $G = (V, E)$ be a connected undirected weighted graph. Write an efficient algorithm to find a minimum weight spanning tree of G.
   (b) Let $G = (V, E)$ be a connected undirected weighted graph and $e \in E$. Design an efficient algorithm to generate spanning tree of G which contains e and has minimum weight among all spanning trees of G which contain e. Derive the time complexity of your algorithm.
   (c) Design an efficient algorithm for the following problem: Given n positive integers $d_1, d_2, \ldots, d_n$, such that $d_1 + d_2 + \ldots + d_n = 2n - 2$, construct a tree with n vertices of degree exactly $d_1, d_2, \ldots, d_n$.

(8+(7+5)+10=30)

4. Describe the Ford-Fulkerson algorithm for finding a maximum flow in a flow network. Discuss the correctness and efficiency of the algorithm.

(10+10=20)

5. (a) Let P be a simple polygon (not necessarily convex) enclosed in a given rectangle R, and q be an arbitrary point inside R. Design an efficient algorithm to find a line segment connecting q to any point outside of R such that the number of edges of P that this line intersects is minimum.
   (b) Describe an $O(n \log n)$ time algorithm to compute the convex hull of n given points on the plane. Show that your algorithm runs in $O(n \log n)$ time.

(10+(8+6)=24)

6. (a) When do you say that a problem is $NP - complete$?
   (b) Prove that the $3 - colouring$ problem is $NP - complete$.

(6+14=20)

BIII-2015-16

Semestral Examination, 2nd Semester

Design of Experiments

Full Marks 100 Date: 6th May, 2016 Time: 10.30-2:00p.m

Answer Question 1 and any three from the rest.

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

a)A chemist wants to compare three treatments. The experimental material he plans to use comes from four different manufactures. There is sufficient experimental material from each manufacturer for 12 experimental units.

b)A marketing expert for a publishing house wants to measure reader preference for four different covers of the same paperback novel. She has chosen 8 cities and 5 newsstands in different locations in each city which are going to sell the novel.

c)An experimenter wants to compare the usefulness of four different word processing softwares using four different PCs, four secretaries and four different texts. He wants to eliminate the differences from all sources of variation, using the minimum number of observations. [(2+5+2)+ (2+5+2)+ (2+6+2) = 28]

2.For a block design **prove or disprove** the following statements:

a) For any block design, structural connectedness of the design is equivalent to the rank definition of connectedness.

b)For a connected block design, none of the diagonal elements of the C-matrix can be zero.

c)Any disconnected block design is necessarily non orthogonal. [8x3=24]

3.a)Show that there can exist at most $v - 1$ Mutually Orthogonal Latin Squares (M.O.L.S) of order v.

b) Construct a complete set of M.O.L.S of order 5.

c) Suppose a need arises to replicate Latin squares of order v in p locations, with the same set of treatments and the same set of columns, but different sets of rows specific to the location. Write an appropriate model for the observations and complete the ANOVA table, showing the different columns and the corresponding entries. [(4+8+ 4+8)=24]

4. a) Suppose that in an Randomised Block design with 5 treatments in 6 blocks, observation corresponding to treatment 3 in block 1 is found missing.

(i)How will you estimate the missing value to obtain the correct value of the error sums of squares?

(ii) When can you use the estimate obtained in (i) above to arrive at a correct conclusion while testing the equality of the treatment effects? Justify your answer.

(iii) How will you modify the estimates of the missing values to obtain the correct expression for the treatment sums of squares.

b) Consider a completely randomized experiment to compare the effectiveness of two different pesticides A and B in two different forms: spray(A_1 and B_1) and powder (A_2 and B_2) . A control(no pesticide), say C is also included in the experiment in order to establish any effectiveness of the pesticides. Each of these five treatments C, A_1 , A_2, B_1 and B_2 is applied to r more or less uniformly infested plots.

(i) Suggest a set of four mutually orthogonal, practically meaningful contrasts of the treatment effects.

(ii) Develop Scheffe's method of multiple comparison procedure in the context of a completely randomized design..

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

5.a) In a $(2^7, 2^4)$ experiment, some treatments of the 16-plot key block are given as (bcdf, abd, dfg, bef, abce, cefg, , acf, cde,........). With proper justification complete the key block **OR** identify the complete set of confounded effects.

b) An experimenter is interested to investigate the breaking strength of dinnerware manufactured by using four chemical compounds(factor A) and baking it at three temperatures (factor B). Three furnaces, each of which is big enough to hold four dinnerwares are available for the experiment, which is run for five days.

(i) If the frequent change to the furnace temperature is difficult to handle, sug' 'st a suitable design for your experiment.

(ii)Clearly state the model for the design suggested in part (i) and derive the BLUE of the and the variances of the pairwise comparison of the effects of levels of factor A and Factor B .

(iii)Which of the main effects of factor A and B are estimated with more precision? Give justification of your claim. [(8+(3+(5+5)+3)=24]