आर.बी.आई.एस.बी. (अनुसंधान अधिकारी – सां.सू.प्र.वि.) P.Y. - 2013

R.B.I.S.B. (R.O. – DSIM) P.Y. – 2013

प्रश्नपत्र II - वर्णनात्मक स्वरुप - सांख्यिकी

PAPER II – DESCRIPTIVE TYPE ON STATISTICS

(अधिकतम अंक – 100) (अवधि – 3 घंटे)

(Maximum Marks – 100) (Duration – 3 Hours)

अनुदेश:

(1) प्रश्नपत्र में **छह खंड** है। उम्मीदवार किसी खंड में से दो से ज्यादा प्रश्नों का चुनाव न करते हुए किन्हीं **पाँच प्रश्नों** के उत्तर दे सकते हैं। यदि उम्मीदवार ने पाँच से ज्यादा प्रश्नों के उत्तर दिए हैं तो उत्तर दिए गए प्रश्नों के क्रमानुसार केवल पहले पाँच प्रश्नों का मूल्यांकन किया जाएगा तथा शेष उत्तरों को अनदेखा कर दिया जाएगा।

(2) **प्रत्येक खन्ड के प्रश्नों के उत्तर अलग उत्तर पुस्तिकाओं/अनुपूरकों (सप्लीमेंट्स) पर दिये जाने चाहिए । अन्य शब्दों में, प्रत्येक उम्मीदवार को उत्तरपुस्तिका के अलावा कम से कम 2 से 4 अनुपूरकों (सप्लीमेंट्स) की आवश्यकता रहेगी।**

(3) अनुपूरक लौटाने से पूर्व उत्तरपुस्तिका के साथ संलग्न करें।

(4) प्रत्येल प्रश्न के 20 अंक है।

(5) उत्तर हिंदी अथवा अंग्रेजी में लिखे जाएं । तथापि सभी प्रश्नों के उत्तर केवल एक ही भाषा में दिए जांए । अंशत: अंग्रेजी तथा अंशत: हिंदी में लिखी गई उत्तर पुस्तिकाओं का मूल्यांकन नहीं किया जाएगा ।

(6) प्रत्येक प्रश्न का उत्तर नए पृष्ठ पर दिया जाना चाहिए तथा प्रश्न की संख्या शीर्ष पर बाईं ओर के हाशिए में लिखी जानी चाहिए ।

(7) एक ही प्रश्न के सभी भागों के उत्तर एक साथ लिखें । दूसरे शब्दों में एक ही प्रश्न के विभिन्न भागों के उत्तर के बीच में किसी अन्य प्रश्न का उत्तर न लिखें ।

(8) नाम, रोल नं. तथा अन्य प्रविष्टियां उत्तर पुस्तिका में केवल निर्धारित स्थान पर ही लिखें तथा इन्हें **उत्तर पुस्तिका और अनुपूरकों पर अन्य कहीं भी न लिखें ।**

(9) उम्मीदवार उत्तर लिखने के लिए केवल नीली अथवा काली स्याही वाले पेन/ बालपोइन्ट पेन का प्रयोग करें ।

(10) उम्मीदवारों को कोई भी संदर्भ पुस्तिकाएँ, पाठ्यपुस्तकें, गणितीय टेबल, इंजीनियरिंग टेबलम अथवा अन्य उपकरण अथवा संचार उपकरण (सेलफोन सहित) नहीं दिये जाएँगे और न ही उन्हें उनके इस्तेमाल की अनुमति दी जाएगी, यहाँ तक कि वे उन्हें अपने पास रख भी नहीं सकेंगे। इस नियम का उल्लंघन करने पर दंड दिया जा सकता है। नॉन-प्रोग्रामेबल इलेक्टॉनिक कॅलक्युलेटर के प्रयोग की अनुमति है।

(11) समस्त कच्चा कार्य (रफ वर्क) उत्तर पुस्तिका के अंतिम 3 या 4 पन्नों पर किया जाना चाहिए।

(12) उत्तरों का मूल्यांकन व्याख्या में तर्क, संक्षिप्तता तथा स्पष्टता के आधार पर किया जाएगा ।

(13) अस्पष्ट लिखाई के लिए अंक काटे जाएंगे ।

खंड - A : संभाविता एवं प्रतिदर्शी

1. (अ) N समूह की जनसंख्या में से n समूह के SRSWOR में, प्रत्येकी M घटक वाले, $\overline{\overline{Y}}$, के अनभिनत आकलक प्राप्त करें (जनसंख्या औसत प्रति घटक) उसका प्रसरण भी प्राप्त करें. प्राप्त प्रसरण का अनभिनत आकलक बताएं.

(ब) निम्नलिखित आंकड़े 25 समूहो के जनसंख्या में से 5 समूहों के यादृच्छिक नमूने हैं, प्रत्येक में दो क्रमिक भूखंड (plot) है. Y का मूल्य नमूना धान की खेती के क्षेत्र है. (areas of sample paddy fields).

समूह (cluster)	1	2	3	4	5
क्षेत्र (Area)	24	16	12	6	25
	21	16	18	3	3

प्रति भूखंड में धान की खेती का औसत क्षेत्र का आकलन उसके प्रसरण के साथ करें.

2. (अ) Strong Law of Large Numbers (SLLN) के बारे में बताएं.

यह सिध्द करें कि स्वतंत्र यादृच्छिक के प्रत्येक अनुक्रम $\{X_n\}$, एक समान परिबध्द प्रसरण के साथ चरवस्तु SLLN का पालन करता है.

(ब) इस बात का परीक्षण करें कि क्या Strong Law of Large Numbers (SLLN) स्वतंत्र और एक समान वणटन यादृच्छिक चार वस्तुओं के अनुक्रम $\{X_n, n \geq 1\}$, सामान्य संभाव्यता सघन कार्य के साथ सहमत है.

$$f(x) = \begin{cases} 7x^{-8} & x \geq 1 \\ 0 & otherwise \end{cases}$$

3 (अ) निम्नलिखित को परिभाषित करें.

(i) संभाविता में अभिसरण (Convergence)

(ii) r^{th} mean में अभिसरण (convergence)

(iii) अतिनिश्चित अभिसरण

यदि $X_n \xrightarrow{p} X$ और g निरंतर कार्य है, यह दर्शाये कि $g(X_n) \xrightarrow{p} g(X)$

(ब) $\{X_n, n \geq 1\}$ **सामान्य सघन कार्य के साथ iid यादृच्छिक चर वस्तुओं का अनुक्रम है,**

$$f(x) = \begin{cases} \dfrac{5^6}{(X+5)^6} & , \quad x > 0 \\ 0 & \quad otherwise \end{cases}$$

$\overline{X}_n \xrightarrow{p} \frac{5}{4}$ **जहां** $\overline{X}_n$ **नमूना माध्य (sample mean) है.**

दर्शाये कि

खंड B: रेखीय प्रतिरूप और अर्थ सांख्यिकी

4 (अ) बहुविध समाश्रयन में अवश्य संतुष्टि देनेवाली मान्यताओं को स्पष्ट करें. संक्षिप्त में बताएं कि मान्यताओं के उल्लंघन को कैसे रोका जा सकता है.

(ब) मान लीजिए समाश्रयन model $y = \beta_0 + \beta_1 x + e$ **और n अवलोकन (observation).**

(i) यह साबित करें कि regression sum of squares = $\hat{\beta}_1^{\,2} \sum(x_i - \overline{x})^2$.

(ii) 10 अवलोकनों के आंकड़े दर्शाते हैं कि :

$\sum(y_i - \hat{y}_i)^2 = 3.267$, $\sum y_i = 63.3$, $\sum y_i^{\,2} = 423.49$ $\sum x_i = 139$, $\sum x_i^{\,2} = 2239$

X **और y में घनात्मक सहसंबंध मानते हुए प्राचल (parameters)** β_0 **और** β_1 **का आकलन करें.**

5 (अ) k उपचार के प्रभाव के समानता के परीक्षण के लिए प्रसरण model के एक मार्गीय विश्लेषण (one way analysis) का विवरण यह ध्यान में रखते हुए दें कि i^{th} **उपचार पर n, observation,** i =1,2,…,k. **मान्यताओं को स्पष्ट करें तथा परीक्षण प्रक्रिया को प्राप्त करें. ANOVA टेबल लिखे.**

(ब) चार जनसंख्याओं में से 16 टिप्पाणियों (observations) के यादृच्छिक नमूने चयनित किए गए. ANOVA टेबल का भाग नीचे दिया गया है.

विचरण का स्त्रोत (Source of Variation)	d.f.	Squares का जोड़	Squares के जोड़ का औसत	F
अभिक्रिया (Treatments)	-	-	-	80
त्रुटि (Error)	-	-	-	
कुल (Total)	-	1500		

समुचित d.f. के साथ, सार्थक 5% के स्तर पर F का मूल्य = 2.76.

(i) अविध्यमान मूल्यों को भरकर टेबल को पूरा करें.

(ii) सार्थक 5% स्तर पर आपका निष्कर्ष क्या है?

6 (अ) (i) सूचकांक के बारे में आप क्या जानते हैं? इसका प्रयोजन क्या है?

(ii) समय उलटाव तथा गुणक उलटाव (time reversal & factor reversal) के परीक्षण की क्या आवश्यकताए हैं? यह दर्शाये कि Fisher's ideal index number से दोनों परीक्षण हो सकते हैं.

(ब) निम्नलिखित आंकड़ों का प्रयोग करते हुए Fisher' price index number को परिकलित (compute) करते हुए इस बात को सत्यापित करें कि इससे {6 (अ)} में उल्लेखित दोनों परीक्षण संतोषजनक ढंग से हो सकते है.

पदार्थ (Commodity)	आधार वर्ष (Base year)		वर्तमान वर्ष (Current year)	
	मूल्य	मात्रा	मूल्य	मात्रा
P	5	60	8	60
Q	2	100	2	120
R	5	50	6	70
S	8	40	10	30

खंड C : सांख्यिकी निष्कर्ष

7 (अ) मान लीजिए T_0, $g(\theta)$ **का UMVUE है तथा v, O का अनभिनत आकलक (unbiased estimator) है. तब यह सिध्द कीजिए कि** T_0, UMVUE **केवल उसी स्थिति में हैं, यदि** $E\,(V\,T_0) = 0 \;\forall \theta \in \Theta$**. मान लीजिए कि दुसरी स्थिति** $S(\theta)$ **और** $E\,(V^2) < \infty$ **के सभी अनभिनत आकलकों (unbiased estimators) के लिए मौजूद है.**

(ब) **(i) मान लीजिए** $X_1, X_2, ..., X_n$ be iid rvs with $N(\mu, \sigma^2)$, μ **और** σ **अज्ञात (unknown)** $P[X_1 \geq k]$ **जहां k कोई वास्तविक संख्या है, का UMVUE पता करें.**

(ii) मान लीजिए कि $X_1, X_2, ..., X_n$ be iid rvs **with discreet uniform.**

$$f(X = x) = \begin{cases} \frac{1}{N} & ; \quad x = 1,2...N, \quad N \geq 1 \\ 0 & ; \quad otherwise \end{cases}$$

e^N **के UMVUE का पता लगाए.**

8 **(अ)** **(i) मान लीजिए कि** $X_1, X_2, ..., X_n$ are iid rvs with $U(\theta, 2\theta)$, $\theta \in \Theta$.
θ **के MLE का पता लगाएं.**

(ii) **मान लीजिए** $P(X = -2) = \frac{\theta}{3}, P(X = -1) = \frac{\theta}{3}$

$P(X = 0) = 1 - \theta, \quad P(X = 1) = \frac{\theta}{3}$

निम्नलिखित आंकड़ों से θ **के mle का पता लगाएं.**

-1, -2, -1, -2, 0, 0, 1, 1, -1, 1.

(ब) **मान लीजिए कि** $X_1, X_2, ..., X_n$ be iid rvs with $N(\theta, 1)$. $g(\theta) = \theta^2$, $\theta \in \Theta$ **के लिए भट्टाचार्य बाउण्ड (Bhattacharya bound) प्राप्त करें.**

9 **(अ)** **मान लीजिए** $X_1, X_2, ...X_{k_1}$ **तथा** $Y_1, Y_2, ..., Y_{k_2}$ are iid rvs. $B(n_1, p_1)$ **तथा** $B(n_2, p_2)$ **के साथ क्रमशः, जहां** n_1 **तथा** n_2 **अज्ञात है,** $H_1 : p_1 > p_2$ **के सामने** $H_0 : p_1 = p_2$ **के परीक्षण के लिए Neyman Structure Test का पता लगाएं.**

(ब) **दो शहरों (मुंबई तथा दिल्ली) के 10 परिवारों के यादृच्छिक नमूनों से यह पता चला कि उनके साप्ताहिक चिकित्सा खर्चे निम्नानुसार हैं.**

मुंबई (X)	:	20	16	10	9	18	21	8	11	27	24
दिल्ली (Y)	:	12	19	23	17	26	22	15	13	14	25

Wald-Wolfowitz run test का प्रयोग करते हुए यह सत्यापित करें कि $\alpha = 0.05$, **यदि दोनों जनसंख्याओंमें एक ही प्रकार का वितरण (underlying distributions) है.**

$[For\ \alpha = 5\%, C_\alpha = 6]$

खंड D: प्रसंभात्य प्रक्रियाएं

10 (अ) एक सुंदर सिक्के को बारंबार उछाला गया, जिसका परिणाम $Y_0, Y_1, Y_2, \ldots$ **आया. जो कि 0 या 1, प्रत्येकि आधि संभाविता के साथ.** $n \geq 1$**, के लिए मान लीजिए** $X_n = Y_n + Y_{n-1}$, $(n-1)^{th}$ **में यह संख्या एक और** n^{th} **बार उछाले है. प्रत्येक संभाविता के लिए अभिव्यक्ति (expression) का पता लगाएं.**

$$P[X_{n-1} = 1], P[X_n = 2, X_{n-1} = 1], P[X_n = 2 / X_{n-1} = 1], P[X_{n-1} = 1, X_{n-2} = 0],$$
$$P[X_n = 2, X_{n-1} = 1, X_{n-2} = 0], P[X_n = 2 | X_{n-1} = 1, X_{n-2} = 0]$$

क्या X_n, **का Markov Property है? न्यायोचित ठहराएं.**

(ब) (i) एक अन्तरिम संभाविता matrix p दुगना Stochastic होता है, यदि प्रत्येक कॉलम का जोड़ एक के बराबर होता है. यदि ऐसा Markov Chain को कम नहीं किया जा सकता, जो सामयिक (aperiodic) नहीं हैं तथा जिसमें od m+1 states **है, बताते हैं कि सीमित संभाविता (limiting probabilities)** $1/m+1$ **के द्वारा दी जाती है.**

(ii) मान लीजिए सुंदर सिक्के के प्रथम n flips में पायी गई heads कि संख्या N_n **है तथा** X_n, N_n mod 5, **तब** $\lim_{n\to\infty}$ P [N_n**, 5 के गुणांक में है.]**

11 (अ) दो जुआरी A और B, सिकके के एक के बाद एक स्वतंत्र रूपसे होनेवाले उछाल पर दांव लगाते है. जिसमें head दिखने की संभावना p है. यदि सिक्के का head ऊपर होता है तो जुआरी A, जुआरी B से एक रुपया जीतता है. और यदि सिक्के का tail ऊपर होता है तो जुआरी B, जुआरी A से एक रुपया जीतता है. इस प्रकार दो जुआरियों के पास एक निश्चित मात्रा में रुपये रहते हैं, मान लीजिए N. यह खेल दोनों में से किसी भी जुआरी के कंगाल होने पर रुक जाता है.

इसे Markov Chain के रूपमें वर्णित करें. तथा इसके stationery distribution का पता लगाए. क्या यह Limit distribution है?

(ब) मान लीजिए एक सीधी रेखा के ऊपर 1, 2, 3, 4 चिन्हित किया जाता है. मान लो X, Markov Chain है जिसकी दाहिनी ओर जाने की संभावना 2/3 है तथा बायी ओर जाने की संभावना 1/3 है. इसमें यह शर्त है कि यदि X_n**, 1 से बायी ओर जाने की कोशिश करता है या 4 से दाहिनी ओर जाता है, तो वही रुक जाता है. यह पता करें कि Markov Chain प्रत्येक पॉइंट पर कितना समय बिताता है.**

12 टाइम सीरिज मॉडेल को $Y_t = 2a\,Y_{t-1} - a^2 Y_{t-2} + e_t$, के द्वारा दर्शाया जाता है. जहां e_t एक सफ़ेद noise process है जिसका प्रसरण σ^2 है.

(i) ए के मूल्य का निर्धारन करें जिसके लिए process stationery है.

(ii) auto covariances υ_k for $k \geq 2$ का परिणाम निकालें.

(iii) यह बताए कि auto covariances के कार्य को $\upsilon_k = Aa^k + k\,B\,a^k$ के रूप में लिखा जा सकता है. A तथा B के कुछ मूल्यों के लिए आपको constants a तथा σ^2 के रूपमें स्पष्ट करना होगा. अत: यह दिखाये की ऑटो correlation ρ_k -

$$\rho_k = \left[1 + \left(1 - a^2\right)k / \left(1 + a^2\right)\right] a^k \quad k = 0,1,2,..$$

(iv) ऊपर दिए गए A R (2) process के रूप में underlying time series को मॉडेल के रूप मे परीक्षण के लिए, परीक्षण प्रक्रिया बताएं.

खंड E: बहुविचरविश्लेषण

13 (अ) लंबकोणीय गुणक प्रतिमान (orthogonal factor model) का उसकी सभी मान्यताओं के साथ विवरण दें तथा निम्नलिखित शब्दों को पारिभाषित करें.

(i) गुणक भरण (Factor loading)

(ii) विशिष्ट प्रसरण (Specific variance)

(iii) सामुदायिकता (Communality)

गुणक आवर्तन (Factor Rotation) के बारे में संक्षिप्त में स्पष्ट करें.

(ब) निम्नलिखित सहप्रसरण (covariance) matrix Σ के लिए एक factor model को मानते हुए loading matrix L तथा विशिष्ट प्रसरण (specific variances) प्राप्त करें. इसके लिए प्रमुख घटक प्रक्रिया (principal component method) का प्रयोग करें.

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

प्रथम घटक द्वारा बताए गए अनुरूप कुल प्रसरण के अनुपात को प्राप्त करें.

14 (अ) प्रामाणिक सहसंबंध (canonical correlations) तथा प्रामाणिक विचरण (कनोनिकल variates) को परिभाषित करें. यह दर्शाये कि प्रथम प्रामाणिक सहसंबंध मेट्रिक्स $\rho_{12 \cdot}$ की किसी भी प्रविष्टि के absolute value से बड़ा है,

जहां $\rho = \left(\begin{array}{c|c} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{array}\right)$

Vector $\underset{(p+q)\times 1}{X} = \left(\begin{array}{c} X^{(1)}{}_{p\times 1} \\ \hline X^{(2)}{}_{q\times 1} \end{array}\right)$ **का सहसंबंध मेट्रिक्स है.**

(ब) निम्नलिखित विभक्त सहसंबंध मेट्रिक्स के लिए प्रथम प्रामाणिक सहसंबंध तथा उसके सहयोगी प्रामाणिक विचार जोड़ी (associated canonical variate pair) प्राप्त करें.

$$\rho = \left(\begin{array}{cc|cc} 1 & 1/2 & 2/3 & 2/3 \\ 1/2 & 1 & 2/3 & 2/3 \\ \hline 2/3 & 2/3 & 1 & 1/2 \\ 2/3 & 2/3 & 1/2 & 1 \end{array}\right)$$

15 (अ) मान लीजिए p_1, p_2**, दो बहुचर जनसंख्यओं** π_1 and π_2 **के पूर्व संभावनाओं को दर्शाते है तथा** $f_1(x), f_2(x)$ **उनके संभावित सघन कार्य को दर्शाते है. उस वर्गीकरण नियम को प्राप्त करें जो Total probability of misclassification (TPM) को कम कर सकें. ताकि, जब** f_i**,** p-variate normal mean vector $\mu_i(i=1,2)$ **तथा सामान्य प्रसरण मेट्रिक्स** $\sum$ **है तब न्यूनतम TPM नियम का पता लगाए.**

(ब) मान लीजिए कि $n_1 = 30$ **तथा** $n_2 = 22$ observations, **दो random vectors** X_1 **तथा** X_2**, पर किए जाते है, जिनके पास bivariate normal distribution की संभावना, सामान्य covariance मेट्रिक्स** $\sum$ **के साथ है परंतु संभवतः प्रथक औसत वेक्टर** μ_1 **तथा** μ_2 **है.**

नमूना औसत वेक्टर्स तथा pooled covariance matrix के विपरीत है, निम्नानुसार है:

$$\bar{X}_1 = \begin{pmatrix} -0.0065 \\ -0.039 \end{pmatrix}, \bar{X}_2 = \begin{pmatrix} -0.02483 \\ -0.0262 \end{pmatrix}$$

$$S_p^{-1} = \begin{pmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{pmatrix}$$

(i) $H_1 : \mu_1 \neq \mu_2$ **के सामने अनुमान (hypothesis)** $H_0 : \mu_1 = \mu_2$ **का परीक्षण करें.**

(ii) Mahalanobis sample distance D^2. **प्राप्त करें.**

(iii) वर्गीकरण के लिए न्यूनतम TPM नियम प्राप्त करें.

(iv) Observation $X' = (-0.21 \quad -0.044)$ **से** π_1 or π_2**.**

खंड F: संख्यात्मक विश्लेषण तथा कंप्यूटर तकनीक

16 (अ) यदि $X^{(k)}$ एकघात (linear equations) की पद्धति X का सन्निकरण (Approximation) है, तो यह बताए कि Gauss-Seideliteration प्रक्रिया का प्रयोग करते हुए X का अगला सन्निकरण (next approximation) $X^{(k+1)}$, $X^{(k+1)} = x^{(k)} + v^{(k)}$ द्वारा प्रापता किया जा सकता हैं, जहां $v^{(k)} = (D + L)^{-1} r(k); r(k) = b - AX^{(k)}$ और डी - विकरणी भाग (Diagonal part), L, A का निचला त्रिकोणीय भाग. k=0, 1, …n.

इस प्रक्रिया का प्रयोग करते हुए निम्नलिखित समीकरण पद्धति (system of equations) के लिए $X^{(1)}$ प्राप्त करें.

$$2x_1 - x_2 = 7$$
$$-x_1 + 2x_2 - x_3 = 1$$
$$- x_2 + 2x_3 = 1$$

प्रारम्भिक सन्निकरण (approximation) $X^{(0)} = [0 \ \ 0 \ \ 0]'$ के रूप में लें.

(ब) Newton-Cotesवर्गाकारन सूत्र (quadrature formula) का परिणाम निकालें.

$$\int_a^b f(x)dx = \sum_{k=0}^{n} \lambda_k f(x_k)$$

तथा आंतरगणन (Interpolation) प्रक्रिया का प्रयोग करते हुए निम्नलिखित form मे Simpson's rule का परिणाम निकाले -

$$\int_a^b f(x)dx = \frac{b-a}{6}[f(a) + 4f((a+b)/2) + f(b)]$$

अत: निम्न पूर्णसांख्यिक का मूल्यांकन करे.

$$\int_0^1 \frac{x^2}{\cos(x)} dx$$

17 (अ) (i) अग्रसारित अंतर परिचालक (forward difference operator) Δ को परिभाषित करें तथा यह प्रमाणित करें कि

(a) $\Delta \log f(x) = \log(1 + \frac{\Delta f(x)}{f(x)})$,

(b) $\Delta^n \sin(ax+b) = (2\sin\frac{ah}{2})^n \sin[ax+b+n(\frac{ah+\pi}{2})]$

(ii) Interval of differencing **being unity, निम्नलिखित का मूल्यांकन करें.**

(अ) $(2\Delta+3)(E+2)(3x^2+x+2)$

(ब) $\Delta^2 E^3 x^3$

(ब) **x के मूल्य के लिए** f(1) =4, f(2) = 5, f(7) =5, f(8) =4 **के रूप में दिए गए हैं. लागरंगे के आंतर गणन सूत्र (Lagrange's interpolation formula) का प्रयोग करते हुए** f(6) **के मूल्य तथा x के मूल्य, जिसके लिए** f(x) **अधिकतम या न्यूनतम है, का पता लगाए.**

18 (अ) निम्नलिखित प्रश्नों का जवाब दें.

(i) नेटवर्किंग में TCP/IP संदर्भ मॉडेल की व्याख्या करें.

(ii) संपर्क (संचार माध्यम) क्या है? Point to point संपर्क तथा प्रसार संपर्क की व्याख्या करें.

(ब) निम्नलिखित प्रश्नों के उत्तर दें.

(i) Relational Database Management System में relation का आप क्या अर्थ निकालते है. रिलेशन से विशिष्ट attribute column तथा tuples के चयन की प्रक्रिया को स्पष्ट करें.

(ii) सर्च की 40 के रूप में लेकर निम्नलिखित संख्याओं के लिए Binary Search Technique को स्पष्ट करें.

12, 18, 24, 30, 35, 38, 40, 45, 50.

आर.बी.आई.एस.बी. (अनुसंधान अधिकारी - सां.सू.प्र.वि.) P.Y.2013
R.B.I.S.B. (R.O. – DSIM) P.Y.2013

PAPER II – DESCRIPTIVE TYPE ON STATISTICS

(Maximum Marks – 100) (Duration – 3 Hours)

Instructions :

(1) The question paper consists of **six sections**. The candidate may attempt ***any five*** questions selecting not more than two from any section. In case the candidate answers more than five questions, only the first five questions in the chronological order of question numbers answered will be evaluated and the rest of the answers ignored.

(2) **QUESTIONS FROM EACH SECTION SHOULD BE ANSWERED ON SEPARATE ANSWER-SCRIPT/SUPPLEMENTS. In other words, a candidate may require/use minimum 2 to 4 supplements, in addition to Answer script.**

(3) Supplement should be attached to the answer script, before returning.

(4) Each question carries 20 marks.

(5) Answers must be written either in English or in Hindi. However, all the questions should be answered in one language only. Answer-books written partly in English and partly in Hindi will not be evaluated.

(6) Each question should be answered on new page and the question number must be written on the top in left margin.

(7) The answers of parts of the same question, if any, should be written together. In other words, the answer of another questions should not be written in-between the Parts of a question.

(8) The Name, Roll No. and other entries should be written in the answer-scripts at the specified places only and these **should not be written anywhere else in the answer script and supplements**.

(9) Candidate should use only Blue or Black ink pen**/**ballpoint pen to write the answers.

(10) No reference books, Text books, Mathematical tables, Engineering tables, other instruments or communication devices (including cellphones) will be supplied or allowed to be used or even allowed to be kept with the candidates. Violation of this rule may lead to penalties. Use of non-programmable electronic calculator is permitted.

(11) ALL ROUGH WORK MUST BE DONE IN THE LAST THREE OR FOUR PAGES OF THE ANSWER SCRIPT.

(12) Answers will be evaluated on the basis of logic, brevity and clarity in exposition.

(13) Marks will be deducted for illegible hand-writing.

A: Probability and Sampling

1. (a) In a SRSWOR of n clusters from a population of N clusters each containing M elements, obtain an unbiased estimator of $\overline{\overline{Y}}$ (population mean per element) . Also obtain its variance. State the unbiased estimator of the variance obtained.

   (b) The following data relates to a random sample of 5 clusters from a population of 25 clusters each consisting of 2 consecutive plots. The value of y are the areas of sample paddy fields.

Cluster	1	2	3	4	5
Area	24	16	12	6	25
	21	16	18	3	3

   Estimate the average area under paddy per field together with its variance.

2. (a) State the strong law of large numbers (SLLN). Show that every sequence $\{X_n\}$ of independent random variables with uniformly bounded variances obeys SLLN.

   (b) Examine whether the strong law of large numbers hold for the sequence $\{X_n, n \geq 1\}$ of independent and identically distributed random variables with a common probability density function.

$$f(x) = \begin{cases} 7x^{-8} & x \geq 1 \\ 0 & otherwise \end{cases}$$

3. (a) Define i) Convergence in probability
   (ii) Convergence in r^th mean
   (iii) almost sure convergence.

   If $X_n \xrightarrow{p} X$ and g is continuous function show that $g(X_n) \xrightarrow{p} g(X)$

   (b) $\{X_n, n \geq 1\}$ is a sequence of iid random variables with a common density function

$$f(x) = \begin{cases} \dfrac{5^6}{(X+5)^6} & , \quad x > 0 \\ 0 & otherwise \end{cases}$$

   Show that $\overline{X}_n \xrightarrow{p} \frac{5}{4}$ where $\overline{X}_n$ is the sample mean.

B: Linear Models and Economic Statistics

4. (a) State the assumptions that must be satisfied in multiple regression. Briefly state how to check violation of assumptions.

(b) Assume regression model $y = \beta_0 + \beta_1 x + e$ and n observations.

(i) Prove that regression sum of squares = $\hat{\beta}_1^{\,2} \sum (x_i - \bar{x})^2$

(ii) Data on 10 observations gives:

$\sum (y_i - \hat{y}_i)^2 = 3.267$, $\sum y_i = 63.3$, $\sum y_i^{\,2} = 423.49$ $\sum x_i = 139$, $\sum x_i^{\,2} = 2239$

Estimate parameters β_0 and β_1 assuming positive correlation between x and y. Find $\operatorname{var}(\hat{\beta}_0)$ and coefficient of determination.

5. (a) Describe one way analysis of variance model for testing equality of k treatment effects, there being n_i observations on the i^{th} treatment; i =1,2,…,k. State the assumptions and obtain testing procedure. Write ANOVA table.

(b) A random sample of 16 observations was selected from each of four populations. Part of ANOVA table is given below.

Source of Variation	d.f.	Sum of squares	Mean sum of squares	F
Treatments	-	-	-	80
Error	-	-	-	
Total	-	1500		

Value of F at 5% level of significance with appropriate d.f. =2.76.

(i) Fill in the missing values and complete the table.
(ii) What is your conclusion at 5% level of significance?

6. (a) (i) What do you mean by an index number? What purpose does it serve?

(ii) What are the requirements of time reversal and factor reversal tests? Show that Fisher's ideal index number satisfies both the tests.

(b) Using the following data compute Fisher's price index number and verify that it satisfies both the tests mentioned in (a).

Commodity	Base year		Current year	
	Price	Quantity	Price	Quantity
P	5	60	8	60
Q	2	100	2	120
R	5	50	6	70
S	8	40	10	30

C: Statistical Inference

7 (a) Let T_0 be the UMVUE of $g(\theta)$ and V be the unbiased estimator of 0. Then prove that T_0 is UMVUE if and only if $E\,(V\,T_0) = 0 \;\forall \theta \in \Theta$. Assume that second moment exist for all unbiased estimators of $g(\theta)$ and $E\,(V^2) < \infty$.

(b) (i) Let $X_1, X_2, ..., X_n$ be iid rvs with $N(\mu, \sigma^2)$, μ and σ are unknown. Find the UMVUE of $P[X_1 \geq k]$, where k is some real number.

(ii) Let $X_1, X_2, ..., X_n$ be iid rvs with discrete uniform.

$$f(X = x) = \begin{cases} \frac{1}{N} & ; \quad x = 1,2...N, \quad N \geq 1 \\ 0 & ; \quad otherwise \end{cases}$$

Find UMVUE of e^N

8. (a) (i) Let $X_1, X_2, ..., X_n$ are iid rvs with $U(\theta, 2\theta)$, $\theta \in \Theta$. Find MLE of θ.

(ii) Let $P(X = -2) = \frac{\theta}{3}, P(X = -1) = \frac{\theta}{3}$

$$P(X = 0) = 1 - \theta, \quad P(X = 1) = \frac{\theta}{3}$$

Find the mle of θ from the following data
-1, -2, -1, -2, 0, 0, 1, 1, -1, 1.

(b) Let $X_1, X_2, ..., X_n$ be iid rvs with $N(\theta, 1)$. Obtain Bhattacharya bound for $g(\theta) = \theta^2 \quad , \quad \theta \in \Theta$

9. (a) Let $X_1, X_2, ... X_{k_1}$ and $Y_1, Y_2, ..., Y_{k_2}$ are iid rvs with $B(n_1, p_1)$ and $B(n_2, p_2)$ respectively, where n_1 and n_2 are known. Find Neyman Structure test for testing $H_0 : p_1 = p_2$ against $H_1 : p_1 > p_2$.

(b) A random sample of 10 households selected from two cities (Mumbai and Delhi), revealed that their weekly medical expenses were as under:

Mumbai (X)	:	20	16	10	9	18	21	8	11	27	24
Delhi (Y)	:	12	19	23	17	26	22	15	13	14	25

Using the Wald-Wolfowitz run test, verify at $\alpha = 0.05$, if the two population have the same underlying distributions.
$[For \ \alpha = 5\% , C_\alpha = 6]$

D: Stochastic Processes

10 (a) A fair coin is tossed repeatedly with results $Y_0, Y_1, Y_2, ...$ that are 0 or 1 with probability half each. For $n \geq 1$ let $X_n = Y_n + Y_{n-1}$ be the number of one's in the $(n-1)^{th}$ and n^{th} tosses. Find the expression for each of the probabilities.
$P[X_{n-1} = 1]$, $P[X_n = 2, X_{n-1} = 1]$, $P[X_n = 2 / X_{n-1} = 1]$, $P[X_{n-1} = 1, X_{n-2} = 0]$,
$P[X_n = 2, X_{n-1} = 1, X_{n-2} = 0]$, $P[X_n = 2 | X_{n-1} = 1, X_{n-2} = 0]$
Whether X_n has a Markov Property? Justify.

(b) (i) A transition probability matrix P is said to be doubly stochastic if the sum over each column equals one. If such a Markov Chain is irreducible, aperiodic and consists of m+1 states show that limiting probabilities are given by $1/m+1$.

(ii) Let N_n be the number of heads observed in the first n flips of a fair coin and X_n equal to N_n mod 5, find $\lim_{n\to\infty}$ P [N_n is multiple of 5]

11 (a) Two gamblers A and B bet on successive independent tosses of a coin that lands heads up with probability p. If the coin turns up heads gambler A wins a Rupee from Gambler B and if the coin turns up tails gambler B wins a Rupee from Gambler A. Thus the total number of Rupees among the two gambler stays fixed say N. The game stops as soon as either gambler is ruined. Describe this as a Markov Chain and found its stationary distribution. Is it the limit distribution?

(b) Consider the points 1,2,3,4 marked on a straight line. Let X_n be the Markov Chain that moves to the right with probability $2/3$ and to the left with probability $1/3$ and subject to the condition that if X_n tries to go the left from 1 or to the right from 4 it stays there. Find the limiting amount of time the chain spends at each point.

12 A time series model is specified by $Y_t = 2a\, Y_{t-1} - a^2 Y_{t-2} + e_t$.
where e_t is a white noise process with variance σ^2 .

(i) Determine the values of a for which the process is stationary.

(ii) Derive the auto covariances υ_k for $k \geq 2$

(iii) Show that the autocovariance function can be written in the form $\upsilon_k = Aa^k + k\ B\ a^k$

for some values of A and B which you should specify in terms of the constants a and σ^2 . Hence show that autocorrelation ρ_k

$\rho_k = [1 + (1-a^2)k/(1+a^2)]a^k \qquad k = 0,1,2,..$

(iv) Give a testing procedure for testing whether underlying Time series can be modeled as AR(2) process given above.

E: Multivariate Analysis

13 (a) Describe an orthogonal factor model with all its assumption and explain the following terms (i) Factor loading (ii) specific variance (iii) communality. Explain in brief what is factor rotation.

(b) For the following covariance matrix Σ, obtain the loading matrix L and specific variances assuming one factor model and using principal component method.

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Obtain the proportion of total variance explained by first factor.

14 (a) Define canonical correlations and the canonical variates. Show that the first canonical correlation is larger than the absolute value of any entry in the matrix ρ_{12}. Where $\rho = \left(\begin{array}{c|c} \rho_{11} & \rho_{12} \\ \hline \rho_{21} & \rho_{22} \end{array}\right)$ is a correlation matrix of vector. $\underset{(p+q)\times 1}{X} = \begin{pmatrix} X^{(1)}{}_{p\times 1} \\ \hline X^{(2)}{}_{q\times 1} \end{pmatrix}$

(b) Obtain first canonical correlation and its associated canonical variate pair for the following partitioned correlation matrix.

$$\rho = \left(\begin{array}{cc|cc} 1 & 1/2 & 2/3 & 2/3 \\ 1/2 & 1 & 2/3 & 2/3 \\ \hline 2/3 & 2/3 & 1 & 1/2 \\ 2/3 & 2/3 & 1/2 & 1 \end{array}\right)$$

15 (a) Let p_1, p_2 denote the prior probabilities of two multivariate populations π_1 and π_2 and $f_1(x), f_2(x)$ denote their probability density functions. Obtain the classification rule that minimizes (TPM) total probability of misclassification. Hence derive minimum TPM rule when f_i is p-variate normal with mean vector $\mu_i(i=1,2)$ and common variance matrix Σ.

(b) Suppose that $n_1 = 30$ and $n_2 = 22$ observations are made on two random vectors X_1 and X_2 which are assumed to have bivariate normal distribution with a common covariance matrix Σ but possibly different mean vectors μ_1 and μ_2. The sample mean vectors and inverse of a pooled covariance matrix are as follows:

$$\overline{X}_1 = \begin{pmatrix} -0.0065 \\ -0.039 \end{pmatrix}, \overline{X}_2 = \begin{pmatrix} -0.02483 \\ -0.0262 \end{pmatrix}$$

$$S_p^{-1} = \begin{pmatrix} 131.158 & -90.423 \\ -90.423 & 108.147 \end{pmatrix}$$

(i) Test the hypothesis $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$.

(ii) Obtain the Mahalanobis sample distance D^2.

(iii) Obtain the minimum TPM rule for classification.

(iv) Classify the observation $X' = (-0.21 \quad -0.044)$ to π_1 or π_2.

F: Numerical Analysis and Basic Computer Techniques

16. (a) If $X^{(k)}$ is the approximation to the solution X of the system of linear equations AX = b, then show that the next approximation $X^{(k+1)}$ of X using Gauss-Seidel iteration method is obtained by $X^{(k+1)} = x^{(k)} + v^{(k)}$, where $v^{(k)} = (D + L)^{-1}$ r(k); r(k) = b – $AX^{(k)}$ and D the diagonal part, L the lower triangular part of A. k=0, 1, …n.

Using this method obtain $X^{(1)}$ for the following system of equations

$$2x_1 - x_2 = 7$$
$$-x_1 + 2x_2 - x_3 = 1$$
$$-x_2 + 2x_3 = 1$$

Take initial approximation as $X^{(0)} = [0\ \ 0\ \ 0]'$

(b) Derive the Newton-Cotes Quadrature formula

$$\int_a^b f(x)\,dx = \sum_{k=0}^{n} \lambda_k f(x_k)$$

and deduce the Simpson's rule in the form

$$\int_a^b f(x)\,dx = \frac{b-a}{6}[f(a) + 4f((a+b)/2) + f(b)]$$

using method of interpolation. Hence evaluate the following integral $\int_0^1 \frac{x^2}{\cos(x)}\,dx$

17 (a) (i) Define forward difference operator Δ and prove that

(a) $\Delta \log f(x) = \log(1 + \frac{\Delta f(x)}{f(x)})$,

(b) $\Delta^n \sin(ax+b) = (2\sin\frac{ah}{2})^n \sin[ax + b + n(\frac{ah+\pi}{2})]$

(ii) Evaluate the following, interval of differencing being unity:

(a) $(2\Delta+3)(E+2)(3x^2+x+2)$
(b) $\Delta^2 E^3 x^3$

(b) The following values of the function f(x) for values of x are given as f(1) =4, f(2) = 5, f(7) =5, f(8) =4. Find the value of f(6) and also the value of x for which f(x) is maximum or minimum using Lagrange's interpolation formula.

18 (a) Answer the following questions:

(i) Explain TCP/IP reference model in networking.
(ii) What is a link (transmission media)? Explain point-to-point link and broadcast link.

(b) Answer the following questions:

(i) What do you understand by a relation in Relational Database Management System? Explain how to select specific attribute columns and tuples from a relation.
(ii) Explain Binary search technique for the following numbers by taking search key as 40:

12, 18, 24, 30, 35, 38, 40, 45, 50.