UPSC CSE 2017 MAINS PAPER 6 NOVEMBER 03, 2017 STATISTICS OPTIONAL PAPER - I QUESTION PAPER

समय : तीन घण्टे	अधिकतम अंक : 250

प्रश्न-पत्र सम्बन्धी विशिष्ट अनुदेश

(कृपया प्रश्नों के उत्तर देने से पूर्व निम्नलिखित प्रत्येक अनुदेश को ध्यानपूर्वक पढ़ें)

इसमें **आठ** प्रश्न हैं जो दो खण्डों में विभाजित हैं तथा **हिन्दी** और **अंग्रेजी** दोनों में छपे हैं।

परीक्षार्थी को कुल **पाँच** प्रश्नों के उत्तर देने हैं।

प्रश्न संख्या **1** और **5** अनिवार्य हैं तथा बाकी प्रश्नों में से प्रत्येक खण्ड से कम-से-कम **एक** प्रश्न चुनकर **तीन** प्रश्नों के उत्तर दीजिए।

प्रत्येक प्रश्न/भाग के अंक उसके सामने दिए गए हैं।

प्रश्नों के उत्तर उसी प्राधिकृत माध्यम में लिखे जाने चाहिए जिसका उल्लेख आपके प्रवेश-पत्र में किया गया है और इस माध्यम का स्पष्ट उल्लेख प्रश्न-सह-उत्तर (क्यू० सी० ए०) पुस्तिका के मुखपृष्ठ पर अंकित निर्दिष्ट स्थान पर किया जाना चाहिए। उल्लिखित माध्यम के अतिरिक्त अन्य किसी माध्यम में लिखे गए उत्तर पर कोई अंक नहीं मिलेंगे।

यदि आवश्यक हो, तो उपयुक्त आँकड़ों का चयन कीजिए तथा उनको निर्दिष्ट कीजिए।

जब तक उल्लिखित न हो, संकेत तथा शब्दावली प्रचलित मानक अर्थों में प्रयुक्त हैं।

प्रश्नों के उत्तरों की गणना क्रमानुसार की जाएगी। यदि काटा नहीं हो, तो प्रश्न के उत्तर की गणना की जाएगी चाहे वह उत्तर अंशतः दिया गया हो। प्रश्न-सह-उत्तर पुस्तिका में खाली छोड़ा हुआ पृष्ठ या उसके अंश को स्पष्ट रूप से काटा जाना चाहिए।

STATISTICS (PAPER–I)

Time Allowed : Three Hours	Maximum Marks : 250

QUESTION PAPER SPECIFIC INSTRUCTIONS

(Please read each of the following instructions carefully before attempting questions)

There are EIGHT questions divided in two Sections and printed both in HINDI and in ENGLISH.

Candidate has to attempt FIVE questions in all.

Question Nos. **1** and **5** are compulsory and out of the remaining, THREE are to be attempted choosing at least ONE question from each Section.

The number of marks carried by each question/part is indicated against it.

Answers must be written in the medium authorized in the Admission Certificate which must be stated clearly on the cover of this Question-cum-Answer (QCA) Booklet in the space provided. No marks will be given for answers written in a medium other than the authorized one.

Assume suitable data, if considered necessary, and indicate the same clearly.

Unless and otherwise indicated, symbols and notations carry their usual standard meanings.

Attempts of questions shall be counted in sequential order. Unless struck off, attempt of a question shall be counted even if attempted partly. Any page or portion of the page left blank in the Question-cum-Answer Booklet must be clearly struck off.

खण्ड—A / SECTION—A

1. (a) किसी निश्चित समष्टि का दस प्रतिशत एक गम्भीर रोग से पीड़ित है। इस रोग के शक वाले एक व्यक्ति को दो स्वतन्त्र परीक्षण दिए जाते हैं। प्रत्येक परीक्षण से 90% बार सही निदान प्राप्त होता है। प्रायिकता ज्ञात कीजिए कि वह व्यक्ति वास्तव में रोगग्रस्त है, जब दिया गया हो कि—

   (i) दोनों परीक्षण निश्चयात्मक हैं;

   (ii) केवल एक परीक्षण निश्चयात्मक है।

   Ten percent of a certain population suffer from a serious disease. Two independent tests are given to a person suspected of the disease. Each test gives a correct diagnosis 90% of the time. Find the probability that the person really suffers from the disease, given that—

   (i) both tests are positive;

   (ii) only one test is positive. 10

(b) मान लीजिए कि X प्रायिकता घनत्व फलन

$$f(x) = \begin{cases} 2/x^2, & \text{यदि } 1 < x < 2 \\ 0, & \text{अन्यत्र} \end{cases}$$

वाला कोई संतत यादृच्छिक चर है। $Y = X^2$ का संचयी बंटन फलन और प्रायिकता घनत्व फलन ज्ञात कीजिए।

Let X be a continuous random variable with the probability density function

$$f(x) = \begin{cases} 2/x^2, & \text{if } 1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find the cumulative distribution function and the probability density function of $Y = X^2$. 10

(c) मान लीजिए कि $\{X_n\}$ स्वतन्त्र और सर्वसम रूप से बंटित यादृच्छिक चरों का एक ऐसा अनुक्रम है, जिसका समान प्रायिकता फलन

$$P[X = -n] = P[X = n] = \frac{3}{\pi^2 n^2}, \quad n = 1, 2, \cdots$$

है। जाँच कीजिए कि क्या $\{X_n\}$ (i) बृहत् संख्याओं के सबल नियम और (ii) बृहत् संख्याओं के दुर्बल नियम का पालन करता है।

Let $\{X_n\}$ be a sequence of independent and identically distributed random variables with common probability function

$$P[X = -n] = P[X = n] = \frac{3}{\pi^2 n^2}, \quad n = 1, 2, \cdots$$

Check whether $\{X_n\}$ obeys the (i) strong law of large numbers and (ii) weak law of large numbers. 10

(d) मान लीजिए कि $x_1, x_2, \cdots, x_n$, प्रत्येक $\theta - \frac{1}{2}$ और $\theta + \frac{1}{2}$ के बीच एकसमान रूप से स्वतन्त्रतः बंटित, n प्रेक्षण हैं। θ के अधिकतम संभाविता आकलक (MLE) का सामान्य रूप ज्ञात कीजिए। सत्यापित कीजिए कि

$$\frac{2Y_1 + 4Y_n - 1}{6} \quad \text{और} \quad \frac{4Y_1 + 2Y_n + 1}{6}$$

दोनों MLE हैं, जहाँ $Y_1 = \underset{i}{\text{Min}}\, X_i$ और $Y_n = \underset{i}{\text{Max}}\, X_i$.

Let $x_1, x_2, \cdots, x_n$ be n observations, each independently uniformly distributed between $\theta - \frac{1}{2}$ and $\theta + \frac{1}{2}$. Find a general form of maximum likelihood estimator (MLE) of θ. Verify that

$$\frac{2Y_1 + 4Y_n - 1}{6} \quad \text{and} \quad \frac{4Y_1 + 2Y_n + 1}{6}$$

are both MLEs, where $Y_1 = \underset{i}{\text{Min}}\, X_i$ and $Y_n = \underset{i}{\text{Max}}\, X_i$. 10

(e) मान लीजिए कि $(0, \theta)$ पर एकसमान समष्टि से $X_1, X_2, \cdots, X_n$ एक यादृच्छिक प्रतिदर्श है और $M = \text{Max}(X_1, X_2, \cdots, X_n)$ है। दर्शाइए कि M, θ का एक अभिनत परन्तु संगत आकलक है।

Let $X_1, X_2, \cdots, X_n$ be a random sample from uniform population over $(0, \theta)$ and let $M = \text{Max}(X_1, X_2, \cdots, X_n)$. Show that M is a biased but consistent estimator of θ. 10

2. (a) मान लीजिए कि $X_1, X_2, \cdots, X_n$ प्रायिकता घनत्व फलन $f_\theta(x)$ वाले किसी बंटन से स्वतन्त्रतः और समरूपतः बंटित प्रेक्षण हैं। यदि $Y = g(X_1, X_2, \cdots, X_n)$, θ का कोई अनभिनत आकलन हो, तो

$$\text{var}_\theta(Y) \geq \frac{1}{nE_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right]}$$

इस असमता का प्रयोग प्रतिदर्श माध्य को प्वासों बंटन के माध्य का एकसमान निम्नतम प्रसरण अनभिनत आकलक दर्शाने के लिए कीजिए।

Let $X_1, X_2, \cdots, X_n$ be independently and identically distributed observations from a distribution with probability density function $f_\theta(x)$. If $Y = g(X_1, X_2, \cdots, X_n)$ is an unbiased estimate of θ, then

$$\text{var}_\theta(Y) \geq \frac{1}{nE_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right]}$$

Use this inequality to show that the sample mean is a uniformly minimum variance unbiased estimator of mean of a Poisson distribution. 20

(b) शेबीशेव असमता का कथन कीजिए और उसको सिद्ध भी कीजिए। अतएव या अन्यथा सिद्ध कीजिए कि (i) $P\left[X \le \frac{\lambda}{2}\right] \le \frac{4}{\lambda}$ और (ii) $P[X \ge 2\lambda] \le \frac{1}{\lambda}$, यदि X कोई धनात्मक यादृच्छिक चर हो जिसके माध्य और प्रसरण दोनों λ के बराबर हों।

State and prove Chebyshev's inequality. Hence or otherwise, prove that (i) $P\left[X \le \frac{\lambda}{2}\right] \le \frac{4}{\lambda}$ and (ii) $P[X \ge 2\lambda] \le \frac{1}{\lambda}$, if X is any positive random variable having both mean and variance equal to λ. 15

(c) किसी समुदाय में $(a+b)$ संभावित मतदाताओं में से a गर्भपात के पक्ष में और $b(b < a)$ इसके विरुद्ध हैं। मान लीजिए कि गर्भपात के वैधीकरण के लिए बहुमत की इच्छा के निर्धारण हेतु वोट डलवाई जाती है। यदि इन $(a+b)$ संभावित मतदाताओं में से $n(n < b)$ यादृच्छिक लोग मत देने नहीं जाते हैं, तो क्या प्रायिकता है कि जो गर्भपात के विरुद्ध हैं वे जीत जाएँगे?

In a community of $(a+b)$ potential voters, a are for abortion and $b(b < a)$ are against it. Suppose that a vote is taken to determine the will of the majority with regard to legalizing abortion. If $n(n < b)$ random persons of these $(a+b)$ potential voters do not vote, what is the probability that those against abortion will win? 15

3. (a) X_1, किसी ग्राहक द्वारा किसी बैंक में सेवापटल पर पंक्ति में लगने से सेवापूर्णता तक का समय है और X_2 सेवापटल तक पहुँचने से पूर्व पंक्ति में प्रतीक्षा का समय है $(X_1 \ge X_2)$. (X_1, X_2) का संयुक्त प्रायिकता घनत्व फलन

$$f(x_1, x_2) = \begin{cases} e^{-x_1}, & 0 \le x_2 \le x_1 < \infty \\ 0, & \text{अन्यत्र} \end{cases}$$

द्वारा दिया गया है। $Y_1 = X_1 + X_2$ और $Y_2 = X_1 - X_2$ मानिए। Y_1 और Y_2 के संयुक्त प्रायिकता घनत्व फलन को प्राप्त कीजिए। क्या Y_1 और Y_2 स्वतन्त्र रूप से बंटित हैं? कारण बताइए।

X_1 is the time that a customer takes from getting on line at a service desk in a bank to completion of service and X_2 is the time to wait in line before reaching the service desk $(X_1 \ge X_2)$. The joint probability density function of (X_1, X_2) is given by

$$f(x_1, x_2) = \begin{cases} e^{-x_1}, & 0 \le x_2 \le x_1 < \infty \\ 0, & \text{elsewhere} \end{cases}$$

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$. Find the joint probability density function of Y_1 and Y_2. Are Y_1 and Y_2 independently distributed? Give reason. 20

(b) मान लीजिए कि X_1, X_2, X_3 और X_4 स्वतन्त्र और समरूपतः बंटित यादृच्छिक चर हैं और प्रत्येक का प्रायिकता घनत्व फलन

$$f(x) = \begin{cases} 3(1-x)^2, & 0 < x < 1 \\ 0, & \text{अन्यथा} \end{cases}$$

है। $Y = \text{Min}(X_1, X_2, X_3, X_4)$ का संचयी बंटन फलन और प्रायिकता घनत्व फलन ज्ञात कीजिए।

Let X_1, X_2, X_3 and X_4 be independent and identically distributed random variables each having probability density function

$$f(x) = \begin{cases} 3(1-x)^2, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the cumulative distribution function and probability density function of $Y = \text{Min}(X_1, X_2, X_3, X_4)$. 15

(c) कोल्मोगोरोव-स्मिरनोव परीक्षण प्रयुक्त करते हुए निर्धारित कीजिए कि क्या नीचे दिए गए प्रतिदर्श आँकड़े, माध्य 4 वाले घातांकी बंटन से आए हैं। प्रतिदर्श में प्रेक्षण हैं

0·7, 5·5, 5·7, 4·3, 0·8, 1·1, 5·1, 4·9, 2·0, 3·9, 5·3, 4·2, 5·2, 5·8, 6·0

(दिया गया है, $D_{13,\ 0\cdot05} = 0\cdot361$, $D_{14,\ 0\cdot05} = 0\cdot349$, $D_{15,\ 0\cdot05} = 0\cdot338$)

Using Kolmogorov-Smirnov test, determine whether the sample data given below come from an exponential distribution with mean 4. The observations in the sample are

0·7, 5·5, 5·7, 4·3, 0·8, 1·1, 5·1, 4·9, 2·0, 3·9, 5·3, 4·2, 5·2, 5·8, 6·0

(Given, $D_{13,\ 0\cdot05} = 0\cdot361$, $D_{14,\ 0\cdot05} = 0\cdot349$, $D_{15,\ 0\cdot05} = 0\cdot338$) 15

4. (a) मान लीजिए कि $X_1, X_2, \cdots, X_n$ एक ऐसे बंटन का यादृच्छिक प्रतिदर्श है, जिसका प्रायिकता घनत्व फलन

$$f(x) = \theta x^{\theta - 1}, \quad 0 < x < 1, \ \theta > 0$$

है। θ का अधिकतम संभाविता आकलक (MLE) और न्यूनतम प्रसरण अनभिनत आकलक (MVUE) प्राप्त कीजिए।

Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with probability density function

$$f(x) = \theta x^{\theta - 1}, \quad 0 < x < 1, \ \theta > 0$$

Find the maximum likelihood estimator (MLE) and minimum variance unbiased estimator (MVUE) of θ. 20

(b) यदि यादृच्छिक चर X, प्रसामान्य बंटन $N(\mu, 1)$ का अनुसरण करता हो, तो $Y = \frac{1-\Phi(X)}{\phi(X)}$ की प्रत्याशा ज्ञात कीजिए, जहाँ Φ और ϕ, मानक प्रसामान्य बंटन $N(0, 1)$ के क्रमशः संचयी बंटन फलन और प्रायिकता घनत्व फलन को निर्दिष्ट करते हैं।

If the random variable X follows normal distribution $N(\mu, 1)$, then find the expectation of $Y = \frac{1-\Phi(X)}{\phi(X)}$, where Φ and ϕ denote the cumulative distribution function and probability density function respectively of $N(0, 1)$, the standard normal distribution. 15

(c) मान लीजिए कि $X_1, X_2, \cdots, X_n$ प्रसामान्य समष्टि, जिसका माध्य μ और प्रसरण 25 है, से लिया गया एक यादृच्छिक प्रतिदर्श है। $n = 16$ के प्रतिदर्श आमाप के लिए और साधारण शून्य परिकल्पना $H_0 : \mu = 5$ के, संयुक्त वैकल्पिक परिकल्पनाओं *(i)* $H_1 : \mu > 5$ तथा *(ii)* $H_1 : \mu < 5$ के विरुद्ध, परीक्षण के लिए एकसमानतः शक्ततम (UMP) परीक्षण प्राप्त कीजिए।

Let $X_1, X_2, \cdots, X_n$ be a random sample from a normal population with mean μ and variance 25. For a sample size of $n = 16$, find the uniformly most powerful (UMP) test for testing the simple null hypothesis $H_0 : \mu = 5$ against the composite alternative hypotheses *(i)* $H_1 : \mu > 5$ and *(ii)* $H_1 : \mu < 5$. 15

खण्ड—B / SECTION—B

5. (a) Y_1, Y_2, Y_3, Y_4 चार स्वतन्त्र चर हैं, जहाँ $E(Y_1) = E(Y_3) = \theta_1 + \theta_3 + \theta_4$, $E(Y_2) = E(Y_4) = \theta_1 - \theta_2$ और $V(Y_i) = \sigma^2$ $(i = 1, 2, 3, 4$ के लिए)। सत्यापित कीजिए कि क्या $\theta_1 + \theta_3$ और $\theta_2 + \theta_3 + \theta_4$ आकलनीय हैं। यदि ऐसा है, तो उनके BLUE प्राप्त कीजिए। BLUE के प्रसरणों को भी प्राप्त कीजिए।

Y_1, Y_2, Y_3, Y_4 are four independent variables with $E(Y_1) = E(Y_3) = \theta_1 + \theta_3 + \theta_4$, $E(Y_2) = E(Y_4) = \theta_1 - \theta_2$ and $V(Y_i) = \sigma^2$ (for $i = 1, 2, 3, 4$). Verify whether $\theta_1 + \theta_3$ and $\theta_2 + \theta_3 + \theta_4$ are estimable. If so, obtain their BLUEs. Also obtain the variances of the BLUEs. 10

(b) मान लीजिए कि $X = [X_1, X_2, X_3]'$ का त्रिचर प्रसामान्य बंटन $N_3(\mu, \Sigma)$ है, जहाँ $\mu' = [1, 1, 1]$ और

$$\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

दर्शाइए कि $(X_2 - X_1)^2 + (X_3 - X_2)^2$ काई-वर्ग (chi-square) बंटन रखता है। स्वतन्त्रता की कोटियों का कथन कीजिए।

Let $X = [X_1, X_2, X_3]'$ have trivariate normal distribution $N_3(\mu, \Sigma)$, where $\mu' = [1, 1, 1]$ and

$$\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

Show that $(X_2 - X_1)^2 + (X_3 - X_2)^2$ has a chi-squared distribution. State the degrees of freedom. 10

(c) सन्तुलित एकदिशिक यादृच्छिक प्रभाव प्रतिमान

$$Y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, 2, \cdots, a; \; j = 1, 2, \cdots, r$$

के लिए, जहाँ $a_i \sim N(0, \sigma_a^2) \; \forall i$ और $e_{ij} \sim N(0, \sigma_e^2) \; \forall i, j$, हों, दर्शाइए कि

$$P[\hat{\sigma}_a < 0] = P[F_{a(r-1),\, a-1} > 1 + n\tau]$$

जहाँ $\tau = \dfrac{\sigma_a^2}{\sigma_e^2}$.

For a balanced one-way random effects model

$$Y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, 2, \cdots, a; \; j = 1, 2, \cdots, r$$

where $a_i \sim N(0, \sigma_a^2) \; \forall i$ and $e_{ij} \sim N(0, \sigma_e^2) \; \forall i, j$, show that

$$P[\hat{\sigma}_a < 0] = P[F_{a(r-1),\, a-1} > 1 + n\tau]$$

where $\tau = \dfrac{\sigma_a^2}{\sigma_e^2}$. 10

(d) कोई अन्वेषक सरल यादृच्छिक प्रतिचयन द्वारा 10 एक-एकड़ वाले भूखण्डों को चुनती है और प्रत्येक भूखण्ड पर पेड़ों की संख्या (y) गिनती है। उसके पास वृक्षारोपण के हवाई चित्र भी हैं, जिससे वह सम्पूर्ण वृक्षारोपण के प्रत्येक भूखण्ड पर पेड़ों की संख्या (x) का आकलन कर सकती है। अतः वह $\mu_x = 19 \cdot 7$ को जानती है और चूँकि दोनों गणनाएँ मूल के सन्निकटतः आनुपातिक हैं, वह आनुपातिक विधि का प्रयोग μ_y के आकलन के लिए करती है। आँकड़े निम्नलिखित पैदा करते हैं :

$N = 1000$ (वृक्षारोपण आमाप), $n = 10$ (SRS द्वारा लिया गया)

y_i = एक-एकड़ वाले भूखण्ड पर वास्तविक पेड़ों की गिनती, $i = 1, 2, \cdots, 10$

x_i = प्रत्येक भूखण्ड का हवाई आकलन

$\bar{y} = 22 \cdot 10$, $\bar{x} = 20 \cdot 80$

भूखण्ड	1	2	3	4	5	6	7	8	9	10
वास्तविक संख्या प्रति एकड़ (Y)	25	15	22	24	13	18	35	30	10	29
हवाई आकलन (X)	23	14	20	25	12	18	30	27	8	31

प्रति एकड़ पेड़ों की औसत संख्या और आकलन की मानक त्रुटि का प्राक्कलन कीजिए।

An investigator selects 10 one-acre plots by simple random sampling and counts the number of trees (y) on each plot. She also has aerial photographs of the plantation from which she can estimate the number of trees (x) on each plot of the entire plantation. Hence, she knows $\mu_x = 19 \cdot 7$ and since the two counts are approximately proportional through the origin, she uses a ratio estimate to estimate μ_y. The data yield the following :

$N = 1000$ (plantation size), $n = 10$ (taken by SRS)

y_i = The actual count of trees in one-acre plots, $i = 1, 2, \cdots, 10$

x_i = The aerial estimate for each plot

$\bar{y} = 22 \cdot 10$, $\bar{x} = 20 \cdot 80$

Plot	1	2	3	4	5	6	7	8	9	10
Actual No. per acre (Y)	25	15	22	24	13	18	35	30	10	29
Aerial estimate (X)	23	14	20	25	12	18	30	27	8	31

Estimate the average number of trees per acre and the standard error of the estimate. 10

(e) संतुलित अपूर्ण खण्डक अभिकल्प (BIBD) की संकल्पना का वर्णन कीजिए। BIBD के अस्तित्व के लिए क्या-क्या प्रतिबन्ध हैं? यदि $V = \{1, 2, 3, 4, 5, 6, 7\}$, तो कुछ खण्डक बनाइए जिनका क्रम 3 हो और V में अवयवों का प्रत्येक युग्म ठीक-ठीक एक ही खण्डक में हो।

Explain the concept of balanced incomplete block design (BIBD). What are the conditions for existence of a BIBD? If $V = \{1, 2, 3, 4, 5, 6, 7\}$, then form a number of blocks, each of order 3, such that each pair of elements in V is contained in exactly one block. 10

6. (a) निम्न सहसम्बन्ध आव्यूह के लिए प्रथम विहित सहसम्बन्ध और इससे जुड़े विहित चर युगल को प्राप्त कीजिए :

$$\rho = \left[\begin{array}{cc|cc} 1 & \frac{1}{2} & \frac{2}{3} & \frac{2}{3} \\ \frac{1}{2} & 1 & \frac{2}{3} & \frac{2}{3} \\ \hline \frac{2}{3} & \frac{2}{3} & 1 & \frac{1}{2} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{2} & 1 \end{array}\right]$$

Obtain first canonical correlation and its associated canonical variable pair for the following correlation matrix :

$$\rho = \left[\begin{array}{cc|cc} 1 & \frac{1}{2} & \frac{2}{3} & \frac{2}{3} \\ \frac{1}{2} & 1 & \frac{2}{3} & \frac{2}{3} \\ \hline \frac{2}{3} & \frac{2}{3} & 1 & \frac{1}{2} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{2} & 1 \end{array}\right]$$

20

(b) दर्शाइए कि $C\beta = 0$ से प्रतिबन्धित प्रतिमान $y = X\beta + \varepsilon$ में β का आकलक

$$\hat{\beta}_C = \hat{\beta} - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C\hat{\beta}$$

है, जहाँ $\hat{\beta} = (X'X)^{-1}X'y$ है।

ध्यान रखिए कि y, $n \times 1$; X, $n \times k$; β, $k \times 1$ और C, $q \times k$ है, जिसकी कोटि $q \leq k$ है।

Show that in the model $y = X\beta + \varepsilon$ subject to $C\beta = 0$, the estimator of β is

$$\hat{\beta}_C = \hat{\beta} - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C\hat{\beta}$$

where $\hat{\beta} = (X'X)^{-1}X'y$.

Note here that y is $n \times 1$, X is $n \times k$, β is $k \times 1$ and C is $q \times k$ of rank $q \leq k$. 15

(c) बहु-उपादानी प्रयोग और कुछ एक एकल-उपादान परीक्षणों में भेद कीजिए। बहु-उपादानी प्रयोग में संकरण का क्या तात्पर्य है? संकरित प्रभावों पर सूचनाओं की हानि के मूल्य पर भी संकरण को क्यों तरजीह दी जाती है?

Distinguish between a factorial experiment and a number of single-factor experiments. What is meant by confounding in a factorial experiment? Why is confounding preferred even at the cost of loss of information on the confounded effects? 15

7. (a) क्रमबद्ध प्रतिचयन विधि को स्पष्ट कीजिए और उसके गुण तथा अवगुण बताइए। आप समष्टि योग के आकलन के प्रतिदर्श प्रसरण का आकलन कैसे करेंगे? दर्शाइए कि क्रमबद्ध प्रतिचयन, समूह प्रतिचयन का एक विशेष उदाहरण किस प्रकार है।

Explain the method of systematic sampling and give its merits and demerits. How do you estimate the sampling variance of the estimate of the population total? Show how systematic sampling is a particular case of cluster sampling. 20

(b) मान लीजिए कि $n_1 = 11$ और $n_2 = 12$ प्रेक्षण दो यादृच्छिक सदिशों X_1 और X_2 पर लिए गए हैं, जो कि उभयनिष्ठ सहप्रसरण आव्यूह Σ परन्तु सम्भवतः विषम माध्य सदिशों μ_1 तथा μ_2 के साथ द्विचर प्रसामान्य बंटन रखने वाले अवधारित हैं। प्रतिदर्श माध्य सदिश और संयुक्त सहप्रसरण आव्यूह नीचे दिए गए हैं :

$$\bar{X}_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \bar{X}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad S_{\text{संयुक्त}} = \begin{bmatrix} 7 & -1 \\ -1 & 5 \end{bmatrix}$$

महलानोबिस प्रतिदर्श दूरी D^2 और फिशर का रैखिक विविक्तकर फलन निकालिए। किसी भी समष्टि π_1 या π_2 को प्रेक्षण $X_0' = (0, 1)$ नियत कीजिए।

Suppose $n_1 = 11$ and $n_2 = 12$ 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 pooled covariance matrix are

$$\bar{X}_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \bar{X}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad S_{\text{pooled}} = \begin{bmatrix} 7 & -1 \\ -1 & 5 \end{bmatrix}$$

Obtain Mahalanobis sample distance D^2 and Fisher's linear discriminant function. Assign the observation $X_0' = (0, 1)$ to either population π_1 or π_2. 15

(c) मान लीजिए कि $y_1 = \beta_1 - \beta_2 + e_1$, $y_2 = \beta_3 + e_2$ और $y_3 = \beta_1 + \beta_2 + e_3$, जहाँ e_1, e_2 और e_3 स्वतन्त्र और $N(0, \sigma^2)$ के साथ समान बंटन यादृच्छिक चर हैं। $(2\beta_1 - \beta_2)$ का सर्वोत्तम रैखिक अनभिनत आकलक (BLUE) प्राप्त कीजिए। क्या यह रैखिक प्राचलिक फलन परीक्षणीय है?

Let $y_1 = \beta_1 - \beta_2 + e_1$, $y_2 = \beta_3 + e_2$ and $y_3 = \beta_1 + \beta_2 + e_3$, where e_1, e_2 and e_3 are iid $N(0, \sigma^2)$. Obtain best linear unbiased estimator (BLUE) of $(2\beta_1 - \beta_2)$. Is this linear parametric function testable? 15

8. (a) एक समबाहु त्रिकोण की तीन भुजाएँ 5 व्यक्तियों द्वारा मापी गई थीं और जिसके परिणाम नीचे दिए गए है :

भुजा	व्यक्ति				
	A	*B*	*C*	*D*	*E*
a	5·44	5·41	5·43	5·42	5·43
b	5·43	5·41	5·42	5·43	5·44
c	5·45	5·42	5·43	5·43	5·44

क्या (i) विभिन्न व्यक्तियों द्वारा किए गए मापों में और (ii) त्रिकोण की भुजाओं में कोई महत्त्वपूर्ण अन्तर है?

Three sides of an equilateral triangle were measured by 5 persons with the following results :

Sides	Persons				
	A	*B*	*C*	*D*	*E*
a	5·44	5·41	5·43	5·42	5·43
b	5·43	5·41	5·42	5·43	5·44
c	5·45	5·42	5·43	5·43	5·44

Is there any significant difference between (i) measurements by the persons and (ii) the sides of the triangle? 20

(b) n प्रेक्षणों के आधार पर एक सरल समाश्रयण प्रतिमान $y = \beta_0 + \beta_1 x + e$ विचारिए। दिखाइए कि वर्गों का समाश्रयण योग $\hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$ है। मान लीजिए 10 प्रेक्षणों के आधार पर नीचे लिखे आँकड़े मिलते हैं :

$$\sum_i (y_i - \hat{y}_i)^2 = 3{\cdot}267, \quad \sum_i y_i = 63{\cdot}3, \quad \sum_i y_i^2 = 423{\cdot}49, \quad \sum_i x_i = 139, \quad \sum_i x_i^2 = 2239$$

X और Y के बीच धनात्मक सहसम्बन्ध मानते हुए β_1 का आकलन कीजिए। इसके साथ ही निर्धारण गुणांक भी प्राप्त कीजिए।

Consider a simple regression model $y = \beta_0 + \beta_1 x + e$ based on n observations. Show that the regression sum of squares is $\hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$. Suppose the summarized data on 10 observations yield

$$\sum_i (y_i - \hat{y}_i)^2 = 3{\cdot}267, \quad \sum_i y_i = 63{\cdot}3, \quad \sum_i y_i^2 = 423{\cdot}49, \quad \sum_i x_i = 139, \quad \sum_i x_i^2 = 2239$$

Estimate β_1 assuming positive correlation between X and Y. Also obtain coefficient of determination. 15

(c) समष्टि संपूर्ण के हॉर्विट्ज़-थॉम्पसन आकलक (HTE) के प्रसरण का व्यंजक व्युत्पन्न कीजिए। इस आकलक को प्राप्त करने में सामने आने वाली कठिनाई को इंगित कीजिए। येट्स-ग्रंडी प्रसरण आकलक ज्ञात कीजिए।

Derive the expression for the variance of Horvitz-Thompson estimator (HTE) of population total. Point out the difficulty faced in estimating it. Find Yates-Grundy variance estimator. 15

★ ★ ★